2015 Electronic Imaging

SCIENCE AND TECHNOLOGY

TECHNICAL SUMMARIES

www.electronicimaging.org

Conferences and Courses
8–12 February 2015

Location
Hilton San Francisco, Union Square
San Francisco, California, USA

IS&T imaging.org

SPIE.

2015 Symposium Chair

Sheila S. Hemami

Northeastern Univ.
(USA)

2015 Symposium Co-Chair

Choon-Woo Kim

Inha Univ.
(Republic of Korea)

2015 Short Course Chair

Majid Rabbani

Eastman Kodak Co.
(USA)

Join us in celebrating
www.spie.org/IYL

Contents

9391: Stereoscopic Displays and Applications XXVI . . . 3
9392: The Engineering Reality of Virtual Reality 2015 . . . 25
9393: Three-Dimensional Image Processing, Measurement (3DIPM), and Applications 2015 . . . 33
9394: Human Vision and Electronic Imaging XX . . . 45
9395: Color Imaging XX: Displaying, Processing, Hardcopy, and Applications . . . 72
9396: Image Quality and System Performance XII . . . 89
9397: Visualization and Data Analysis 2015 . . . 111
9398: Measuring, Modeling, and Reproducing Material Appearance 2015 . . . 117
9399: Image Processing: Algorithms and Systems XIII . . . 139
9400: Real-Time Image and Video Processing 2015 . . . 161
9401: Computational Imaging XIII . . . 174
9402: Document Recognition and Retrieval XXII . . . 184
9403: Image Sensors and Imaging Systems 2015 . . . 189
9404: Digital Photography and Mobile Imaging XI . . . 199
9405: Image Processing: Machine Vision Applications VIII . . . 209
9406: Intelligent Robots and Computer Vision XXXII: Algorithms and Techniques . . . 225
9407: Video Surveillance and Transportation Imaging Applications 2015 . . . 234
9408: Imaging and Multimedia Analytics in a Web and Mobile World 2015 . . . 258
9409: Media Watermarking, Security, and Forensics 2015 . . . 265
9410: Visual Information Processing and Communication VI . . . 277
9411: Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2015 . . . 283

IS&T/SPIE
Electronic Imaging
SCIENCE AND TECHNOLOGY

8–12 February 2015
Hilton San Francisco, Union Square
San Francisco, California, USA

Click on the Conference Title to be sent to that page

Conference 9391: Stereoscopic Displays and Applications XXVI

Monday - Wednesday 9–11 February 2015

Part of Proceedings of SPIE Vol. 9391 Stereoscopic Displays and Applications XXVI

9391-1, Session 1

Enhancement of the effective viewing window for holographic display with amplitude-only SLM

Geeyoung Sung, Jungkwuen An, Hong-Seok Lee, Sun Il Kim, Song Hoon, Juwon Seo, Hojung Kim, Wontaek Seo, Chil-Sung Choi, U-in Chung, Samsung Advanced Institute of Technology (Korea, Republic of)

1. Background, Motivation and Objective:

In recent years, the necessity of holographic display that can reproduce natural three-dimensional (3D) images with midair interactions has begun to increase. To overcome the limitation of a viewing angle of a digital Fresnel hologram due to the relatively large pixel-pitch of a spatial light modulator (SLM), the viewing window is formed only near the observer's eye positions in one of the current holographic display approaches. In this approach, the size of viewing window is defined by the pixel pitch, wavelength, and observer distance. However, the effective viewing window is much smaller than the theoretical one, because the amplitude-only SLM inherently produces significant optical noise such as twin noise and zero order non-diffracting (DC) noise.

The objective of this study is to enlarge the size of the effective viewing window without additional hardware components.

2. Description of Our Concept:

To enhance the effective viewing window at the Fourier plane, there are many kinds of methods including the multiple light sources or the diffusive surface algorithm. The multiple light source method has shortcoming of increasing system complexity. The diffusive surface method is to apply the random phase to the generated hologram in order to generate diffusive 3D images. Using this method, we can observe 3D images within the entire region of theoretical viewing window, but this is available only when we use the complex SLM instead of the amplitude-only SLM (used in our experiment). The complex SLM that can modulate both phase and amplitude does not make a conjugate holographic image, while the amplitude-only SLM makes the conjugate holographic image in addition to the original 3D image. Thus the image quality of reconstructed original 3D images is degraded by the diffusive conjugate holographic image.

We propose the effective viewing window enhancement method for a holographic display with an amplitude-only SLM by using algorithmic approach. The basic concept is the superposition principle of holography. The multiple computer generated hologram (CGH) can be displayed on SLM, and multiple 3D images are reconstructed at different positions within viewing window simultaneously.

3. Experimental Results and Summary:

We have implemented the holographic display using an amplitude-only SLM, field lens and laser light sources. From this experimental setup, we can observe the holographic 3D image in the frustum formed by the field lens through the viewing window located in the Fourier plane of the hologram. To enhance the effective viewing window, we generate multiple CGHs with an observer's eye positions, and then overlap them to make the final CGH. Multiple 3D images can be reconstructed in different positions within the theoretical viewing window from the CGH displayed on SLM. This makes the enlargement of viewing zone that can observe the holographic images. The multiple holograms can be also made for enhancement of the viewing window along both horizontal and vertical direction (2D enlargement viewing zone). We confirmed that the experimental results and the simulation based on Rayleigh-Sommerfeld theory match well.

9391-2, Session 1

A full-parallax 3D display with restricted viewing zone tracking viewer's eye

Naoto Beppu, Tomohiro Yendo, Nagaoka Univ. of Technology (Japan)

We propose a Three-Dimensional (3D) display that we can see 3D images of 360 degrees around the display, and reproduce Full-parallax. It is auto-stereoscopic, requiring no special viewing glasses. In real space vision of object is different by position of view point. If the image depending on position of view point is seen on display, viewer can see object stereoscopically. Viewer can see only image depending on position of view point by using directional ray. In the case of a display that we can see 3D images of 360 degrees around it, the ray is scanned by a spinning screen. It turns the ray of a specific from a spinning screen into a viewer. This display project the time division image for screen from the projector. Therefore, it displays the image depending on the point of view, and presents 3D images to viewers 360 degree around the display. This technique need images that the number of viewpoints. If you projection of the horizontal and vertical parallax for 360 degrees around the display at the same time, it need many images. However, it lacks feasibility. Because there is a limit to what current technique can high-speed projection of the number of images. Therefore, we present way to improve the usability of the display by to focus the ray for the viewing area.

We can increase the image that can be displayed for one area, and we realize the horizontal and vertical parallax for 360 degrees around the display at the same time by the using it. A display consists of a high-speed video projector, a spinning mirror that tracking viewer's eye, and a spinning polygonal cylinder. A spinning mirror is located at the center of a spinning cylinder, and the image is projected by a high-speed video projector from underneath toward it. The ray that is projected a high-speed video projector is reflected toward a spinning cylinder by oblique spinning mirror. And, this mirror rotates 360 degree to track viewer. Therefore, it can direct the ray toward the viewing area in 360 degrees around a display. A spinning cylinder scans the ray from a spinning mirror in horizontal direction by rotation. We reproduce the switching of point of view in the horizontal direction, and realize horizontal parallax by the using it.

Each mirror of a spinning cylinder is different the axial tilt of the vertical direction. Accordingly, the ray that reflected by each surface toward to different direction. And, it is scanned for vertical direction by a spinning cylinder. We reproduce the switching of point of view in the vertical direction, and realize vertical parallax by the using it.

This display switches the image corresponding to a spinning cylinder rotation angle. A spinning cylinder is rotating at high speed. Therefore a display needs the projector that can high speed project. We realize a high-speed video projection by modifying an off-the-shelf projector to use DLP that faster than an existing DLP with FPGA circuitry.

We confirmed usability of this method by simulation.

9391-50, Session Key

A stereoscope for the PlayStation generation *(Keynote Presentation)*

Ian H. Bickerstaff, Sony Computer Entertainment Europe Ltd. (United Kingdom)

After many years of waiting, virtual reality will soon be available for home use. Smart phones have given us small, high quality displays and accurate movement tracking while the games industry has given us the necessary

real-time graphics power to drive these displays. In addition, advances in technologies such as free-form optics, and binaural audio processing have arrived at just the right time.

More than just viewing images on a screen, the aim of ventures such as Sony Computer Entertainment's Project Morpheus is to produce a system that convinces the wearer that they have been transported to another place, and the display system is a vital component. Ever since the beginning of photography, equipment has been created to achieve this goal: an 1850's Brewster stereoscope contains many design features found in the latest HMDs. In both, near ortho-stereoscopic viewing conditions ensure that subjects appear life sized and with realistic depth placement. Unlike a monitor or cinema screen, images are always seen from an optimum viewing position with keystone distortion and vertical parallax kept to a minimum. A far greater range of depth can be viewed comfortably on a head-mounted display than is possible on a conventional screen.

Unlike Victorian stereoscopes, the latest devices offer a wide field of view using techniques first pioneered by Eric Howlett with his Leep VR system in the early 1980s. Screen edges are pushed into peripheral vision so that the concept of a stereo window is no longer relevant. Pincushion distortion is used to increase visual acuity in the centre of vision, mimicking the characteristics of the human eye.

To complete the illusion, high frequency data from accelerometers and gyros are fused with lower frequency camera data to provide accurate, low latency tracking of the viewer's head position and orientation. Ingenious new techniques create the illusion of zero latency, drastically reducing the potential for any viewer disorientation.

So how do we create images for these displays? From a software point of view, the main challenge is to achieve guaranteed high frame rates while avoiding pixel aliasing. Using stereography to manage 3D settings is not required though. In fact any unexpected departure from ortho-stereoscopic viewing could lead to viewer disorientation.

One challenge relevant to this conference is how to photograph and display real-world subjects in a virtual reality system. Even basic 360-degree photography is difficult enough without capturing in the three dimensions necessary for these displays. Multi-camera rigs generate image stitching errors across the joins caused by the very parallax necessary for binocular depth cues. An even more fundamental problem is how these images should be encoded. How can the parallax baked into an image be correct for every viewing direction? It is surprising how despite the maturity of conventional 3D photography, capturing 360 degree 3D images is still in its infancy.

Virtual reality technology is developing faster now than ever before but the more we discover, the more we realise how little we actually know. This presents an enormous opportunity for everyone to define the knowledge that will be so important in the future.

9391-3, Session 2

3D UHDTV contents production with 2/3 inch sensor cameras

Alaric C. Hamacher, Sunil P. Pardeshi, Kwangwoon Univ. (Korea, Republic of); Taeg Keun Whangboo, Gachon Univ. (Korea, Republic of); Sang-IL Kim, Kwangwoon University (Korea, Republic of); SeungHyun Lee, Kwangwoon Univ. (Korea, Republic of)

CONTEXT:

Most of todays UHDTV Cameras are large single CMOS sensor cameras. Because large sensors require longer lenses to achieve the same framing like in HD, the present lens systems represent a huge limitation especially on the field of broadcast for 2D and 3D. This is due to the fact that most equipment comes from the cinema industry.

Grass Valley has recently introduced a new box camera model LDX4k, which allows the use of conventional broadcast lenses and achieves a 4k image resolution via pixel shifting.

OBJECTIVE:

The purpose of this article is to evaluate the main differences between the used sensor technologies and to propose a practical setup for stereoscopic UHDTV acquisition, overcoming the lens limitations due to large single CMOS sensors. The demonstration setup covers camera control, lens control, monitoring and recording with two LDX4k cameras.

METHOD:

The present study examines previous research to compare the technical imaging quality of large single CMOS sensor cameras with the quality of 3D pixel shifted HD sensor cameras. During the beginning of HD Panasonic has successfully introduced pixel shifted imaging sensors to the broadcast marked. A similar approach can be observed today with UHDTV.

Following the technical specifications, the present article presents a practical 3D setup. It evaluates which components are needed for practical control and handling of the two box cameras. It demonstrates which broadcast devices can be used and presents a workflow for processing the recorded contents.

RESULTS:

While achieving similar image quality to most large single CMOS sensor cameras, two main results can be observed: first, the presence of a 2/3 inch sensor allows the use of traditional broadcast lenses, to create an aesthetics similar to HD, especially useful in sports and other subjects where long focal lengths are required. Second, if processing with raw sensor image data is considered, the article demonstrates how 4k postproduction can be handled within a HD workflow.

NOVELTY:

Creation of UHDTV Contents has shifted towards larger sensors and camera bodies, making it more difficult to create setups for capturing stereoscopic 3D. The new 4k box camera models allow the setup of compact 3D systems with smaller lenses for 3D UHDTV content production. The present article demonstrates how 3D UHDTV production can take advantage of this new technology.

9391-4, Session 2

Integral three-dimensional capture system with enhanced viewing angle by using camera array

Masato Miura, Naoto Okaichi, Jun Arai, Tomoyuki Mishina, NHK Japan Broadcasting Corp. (Japan)

CONTEXT:

Integral imaging is advantageous in three-dimensional (3-D) television system. One advantage is that integral imaging can provide full parallax and motion parallax without the need for special glasses. Also, it is possible to capture objects under natural light.

OBJECTIVE:

An integral imaging system requires an extremely large number of pixels for capturing and displaying 3-D images at a high resolution, a wide viewing angle, and a wide reconstructible depth range. We recently applied an ultra-high definition video system called 8k Super Hi-Vision (SHV) to our latest 3-D television system prototype to improve the quality of 3-D images. However, there is no video system with a much higher pixel count than 8k SHV, so it is not expected to further improve the quality of 3-D images when applied to a single video device.

METHOD:

We applied a camera array consisting of numerous cameras to increase the pixel count of an integral-imaging-based capture system without being limited by a single device. This system allows us to increase the pixel count according to the number of cameras, and the viewing angles of the 3-D images can be enlarged as a result of the increased number of pixels.

We can directly reconstruct the captured 3-D image by using a display system with the same system parameters as the capture system. We developed a method to stitch together the 3-D images in order to

reconstruct the captured 3-D images by using a display system with different system parameters.

RESULTS:

We developed an integral-imaging-based capture system consisting of a lens array with 213 x 138 lenses, a field lens, and 7 high-definition cameras in a delta arrangement, where we arranged 2, 3, and 2 cameras for the top, middle, and bottom layers, respectively. We note that the viewing angles of 3-D images are ideally 3 times wider than that captured by a single camera; however, we captured 3-D images with a partial overlap among the cameras for a smooth stitch together of the 3-D images. Consequently, the viewing angles were approximately 2.5 times wider than that captured by a single camera.

We also developed an integral-imaging-based display system consisting of a display with 3840 x 2160 pixels and a lens array with 213 x 138 lenses. We corrected the positioning errors of the cameras by using a calibration pattern to stitch together the 3-D images with a high level of accuracy. As a result, the 3-D images with smooth motion parallax over the whole viewing zone were reconstructed.

NOVELTY:

R. Martinez-Cuenca et al. reported on an integral imaging system using a multiple-axis telecentric relay system to enlarge viewing angles of 3-D images [1]; however, all of the pixels of their system would not contribute to enlarge the viewing angles. Our system allows us to effectively enlarge the viewing angles according to the number of cameras because a lens-shift method is applied to each camera in order to capture the same spatial region among the cameras.

[1] R. Martinez-Curnca et al., Opt. Express 15, 16255-16260 (2007).

9391-5, Session 2

A stereoscopic lens for digital cinema cameras

Lenny Lipton, Leonardo IP (United States); John A. Rupkalvis, StereoScope International (United States)

An Improvement to the monochrome anaglyph is the subject of this paper. The scope of the discussion is limited to anaglyph drawings, especially line drawings, cartoons, technical illustrations or the like. Such hardcopy illustrations use a nominally white background, or in some cases, a nominally black background printed on a medium, usually paper, with cyan and red ink. Anaglyph drawings are in use, for example, for comic book and technical illustrations and for phantograms. This paper will describe a technique that allows for a total elimination of crosstalk, a frequent defect in the medium because of the inability to perfectly match inks and anaglyphoscope filters. Photometric techniques are not employed but rather a heuristic approach is described which is readily accessible to the individual worker or workers in the printing industry, especially those who lack prior experience with the method. The technique is achieved in printing using the originally selected inks.

9391-6, Session 2

A novel optical design for light field acquisition using camera array

Mei Zhang, Geng Zheng, Zhaoxing Zhang, Institute of Automation (China); xuan cao, Institute of Automation, Chinese Academy of Sciences (China)

CONTEXT:

Light field capture techniques have received considerable attention in recent research. A number of lightfield camera designs are proposed using single image sensor. For example, Lytro and RayTrix implemented lightfield cameras based on placing lenticular optics in front of image sensor. Adobe group used multi-aperture objective lenses. MIT and MERL groups proposed aperture encoding mechanism. However, due to the limited size of image sensor chip and optical design, the disparity of the lightfield captured using these single sensor camera systems is very small.

Stanford group pioneered an implementation of lightfield capture systems using camera array. The physical displacement among multiple cameras can be large thus eliminating the problem of small disparity in lightfield capture. However, these light field capture systems based on camera array architecture often employ discrete imaging sensors and associated optics. The optical performance and structure integrity are compromised.

OBJECTIVE:

We propose a novel optical design approach that incorporates an integrated optical design to optimize all optical parameters for all sensors towards a common goal. Instead of design and use the same optics for each and every sensor in the array, we customize the design for each optical channel to maximize the image quality, coverage area, among other design targets. We also integrate the optical design of all imaging channels into a single monolithic piece, incorporating it with structural components to minimize the overall footprint of the camera array, and enhance system reliability and assembly precision. The captured light field images form all imaging channels have the same object size with uniform image quality, thus greatly improve the quality of 3D light field reconstruction.

METHOD:

We design an integrated miniature camera array with eight optical channels to acquire object image about 80mm away (Figure 1). The semi-height of object is set 80mm, while the size of CMOS is 0.1 inches with pixel of 2?m. The optical design procedure is described as follow:

(1) Designing an imaging optical system with four spherical lenses to meet the requirement of the system in ZEMAX. It is shown that only 80% coverage area on the CMOS sensor of marginal camera is left for 3D surface reconstruction. (2) Setting five configurations for different objective distance requirement based on the optical system in step (1). (3) Setting both surface of the first lens different for every configuration while making other structure parameters multi-used. (4) Setting the second surface as aspheric surface to get better performance, and changing the first lens as plastics material to make easy manufacture for such miniature optical lens. (5) Making optimization with unified image size for the whole optical multi-configuration system.

RESULTS:

After integrated optical design, we can see that only the first lens has different curvature, while other lenses have same parameters (Figure 2). Considering the difficulty of process, only first four high order coefficient of the aspheric surface are used. The size of images of all configurations is all around 0.85 mm to fully match the CMOS. The MTF of every configuration below 250 lp/mm is above 0.3 (Figure 3), which means all aberrations are well corrected. The integrated optical design realizes uniform magnification for all cameras without introducing too much fussy modifies. All optical channels are fixed in a tube (Figure 4) with compact structure, high reliability and assembly precision.

NOVELTY:

Most camera array system aligns discrete identical cameras to make 3D image acquisition, which introduce considerable diversity in image quality and structure integrity problems. We propose an integrated design method to realize a camera array with uniform high performance. This work can significantly improve the performance 3D imaging acquisition with camera array to fulfill 3D surface reconstruction with high quality.

9391-7, Session 2

Real-time viewpoint image synthesis using strips of multi-camera images

Munekazu Date, Hideaki Takada, Akira Kojima, Nippon Telegraph and Telephone Corp. (Japan)

Video communications with high sense reality is needed to make natural connection between users at different places. One of the key technologies to achieve sense of high reality is image generation corresponding to individual

user's view point. Generally, generation of viewpoint image requires advanced image processing such as 3D modeling including segmentation and stereo-matching, light field image reconstruction and so on. These image processing is usually too heavy to use in real-time and low-latency purpose.?We have developed a system, which synthesize images of remote user, background scenery and CG objects [1]. It could mostly reproduce stably position images using head tracking of local and remote users and achieved higher reality comparing to previous video communication systems. However there was a feeling of wrongness in background scenery, because it was produced only by changing images taken by multi cameras and it was incorrect when user is not on the line of the camera. Therefore real-time image generation algorism is needed when user is not on the line of cameras. In this paper we proposed a real-time viewpoint image generation method using simply blending multiple still camera images and achieved a prototype for real-time full-frame operation for stereo HD videos using GPU processing. The users can see his individual viewpoint image for left-and-right and back-and-force movement toward the screen.

As multi-camera images, we used still-camera images which are taken at same horizontal intervals using dolly rail. A strip of image, which is the most closest to the rays from the user's eye, are cropped from each still images. We formed two images by putting the image strips side by side. Then we mixed these two images with weight, which is in proportion to the closeness of the ray form the user's eye. We applied the same image generation algorithm for user's left and right eyes. To avoid doubled images, the interval of cameras was decided considering the depth range of photographic subject. It calculated using DFD (depth fused 3-D) visual effect [1]. A prototype system is made using a PC with a graphic card which has 12GB memory and GPU. All images taken by still camera is loaded to the memory of the graphic card. Making strips and mixing was made by GPU processing. We used 180 still-camera images and generate viewpoint of the images corresponding the user position detected by Kinect. Two stream of natural HD video could be generated. The speed of calculation was enough fast for full-frame operation.?As a result, we can confirm the viewpoint image generation using life-sized screen images without doubled image. The boundary of strips could not be detected and the generated images were so natural.

We achieved real-time synthesis of a viewpoint 3D image from images taken by multi camera. Our algorithm is very simple and promising for video communication with high sense of reality.

Reference:

[1] M. Date, H. Takada, S. Ozawa, S. Mieda, and A. Kojima, "Highly Realistic 3D Display System for Space Composition Telecommunication," Proc. of IEEE IAS Annual Meeting, 2013-ILDC 440 (2013).

9391-8, Session 3

Interactive stereo games to improve vision in children with amblyopia using dichoptic stimulation

Jonathan H. Purdy, Univ. of Bradford (United Kingdom); Alexander Foss, Nottingham Univ. Hospitals NHS Trust (United Kingdom); Richard M. Eastgate, The Univ. of Nottingham (United Kingdom); Daisy MacKeith, Nottingham Univ. Hospitals NHS Trust (United Kingdom); Nicola Herbison, The Univ. of Nottingham (United Kingdom); Anthony Vivian, Nottingham Univ. Hospitals NHS Trust (United Kingdom)

CONTEXT:

Amblyopia, often referred to as Lazy Eye, is abnormal visual development in the brain during childhood causing poor vision in one eye. Amblyopia affects 2-3% of the population and conventional treatment in children involves patching the 'good' eye for hours each day which has a detrimental effect on the child's ability to use their eyes together.

OBJECTIVE:

A system of dichoptic stimulation for the treatment of amblyopia has been developed called I-BiTTM. This involves a common background to both eyes with the area of interest only presented to the lazy eye. A temporally interlaced stereoscopic system was used to separate the images to the two eyes. The treatment system involves either playing a 2D scrolling action game or watching a DVD. The primary objective of the study was to investigate whether there was any difference in visual acuity in patients treated using the stereoscopically separated system and the equivalent content presented equally to both eyes.

METHOD:

The treatment system was a standard PC with a 23" stereo capable monitor and commercial wireless NVidia shutter glasses. Subjects were treated individually and seated a fixed distance from the monitor. The game was a modified version of a 2D scrolling shooting game with multiple levels. The player had to avoid obstacles, collect tokens and shoot enemies; they also received a time bonus for finishing the levels within a fixed time. The DVD system included a common background frame with a central area where any commercially available visual media (Film, TV, or Animation) could be displayed and viewed. In the treatment (I-BiTTM) version of the game the enemies and obstacles were only rendered to the image that was presented to the lazy eye. In the I-BiTTM treatment version of the DVD system the media is only visible to the Lazy Eye. For the none-treatment game all content is rendered to both eyes simultaneously however shutter glasses are still worn by the subject.

RESULTS AND CONCLUSION

The study was carried out on 75 children aged between 4 and 8 who had strabismic, anisometropic or mixed amblyopia. These patients were randomised to 24 patients receiving the dichoptic stimulation DVD (group 1), 26 patients received dichoptic stimulation games (group 2) and the third group receiving non-treatment games (group 3). At the end of treatment, visual acuity improved in 19 patients in the DVD (group 1), 19 patients in the treatment games (group 2) and in 14 patients in the non-treatment games group. The proportion of patients showing a clinically significant increase of visual acuity was 46% in the DVD group, 50% in the treatment games group and 36% in the non-treatment games group.

Unlike patching the study showed a high acceptability of the treatment to the patients with only 1 patient who voluntary withdrew due to problems with transport.

NOVELTY:

This is the first randomised control trial of dichoptic stimulation for the treatment of amblyopia in children. The study showed that the visual acuity improved in all groups and that the treatment has very good compliance.

9391-9, Session 3

Stereoscopic visualization of 3D volumetric data for patient-individual skull base prosthesis prior to manufacturing

Justus F. Ilgner M.D., Martin Westhofen M.D., Univ. Hospital Aachen (Germany)

Introduction: In recent years, 3D volumetric data from taken from axial CT scans have been successfully used to manufacture patient-individual prosthetic parts, which replace bony defects of the outer skull and the skull base next to the ear and nose that have been inflicted either by trauma, inflammation or tumor growth. Yet, approving the 3D rendered model proposed by the manufacturer prior to making the actual prosthesis is the surgeon's responsibiity. This is often difficult, as complex anatomy with numerous important structures (nerves, vessels etc.) must be respected. While pdf reader software allows displaying 3D volumetric data monoscopically, stereoscopic visualisation would be helpful to detect errors in the proposed model at an early stage.

Material, Patients and Methods: Since 1999, we have implanted 23 patients (17m, 6f) aged 20 to 81 years with 25 individually CAD-manufactured ceramic prosthesis made of Bioverit II(R) (3di, Jena, Germany). 16 implants

were used to replace the frontal bone and frontal skull base, 1 for the lateral portion of the temporal bone and 1 for the laryngeal cartilage. in 7 cases we managed to close an uncovered part of the vestibular organ which is exposed due to a degenerative process in bone turnover. Particularly in complex cases, we usually order a hardcopy of the proposed 3D model and surrounding bone as monoscopic rendering of the 3D model in a pdf reader levaes uncertainties concerning the intraoperative fit. As a pilot study, we specifically rendered left and right views of the 3D model in these complex cases and displayed them on a conventional stereoscopic monitor using either shutter glasses at 120 Hz or line-polarized Tv monitor using polarized filter glasses.

Results: In two-dimensional defects, particularly on the outer surface of the cranial skull, defect margins were clearly visible on the monoscopic 3D rendered image by the pdf reader. From these, the implant was ordered directly from the manufacturer. In cases with complex implants which cover more than one surface or need to be split in several parts, stereoscopic rendering gave the surgeon a better understanding of the proposed prosthesis form and therefore enabled him to suggest corrections to the manufacturer prior to implant making.

Conclusion: In patient-individual manufacturing of skull base prosthetic materials, perfect fit is the key to success. In complex defects, the manufacturing of an intermediate model by rapid prototyping prior to making the final prosthesis is as yet mandatory, but also costs additional money and time. Stereoscopic rendering of the virtual 3D model helps to detect modelling flaws at an early stage. For future perspective, it has the potential to make intermediate hardcopy models become obsolete and therefore shorten the implant manufacturing process and save costs.

9391-10, Session 3

Visual perception and stereoscopic imaging: an artist's perspective

Steve Mason, Yavapai College (United States)

CONTEXT: This paper is a follow-up to my last February's IS&T/SPIE Convention exploration into the relationship of stereoscopic vision and consciousness (90141F-1). It was proposed then that by using stereoscopic imaging people may consciously experience, or "see," what they are looking at and thereby help make them more aware of the way their brains manage and interpret visual information. Stereoscopic environmental imaging was suggested as a way to accomplish this. This paper is the result of further investigation, research, and the creation of such imaging.

OBJECTIVE: For this coming year's conference I would like to put on a show of images that are a result of this research and will allow viewers to experience for themselves the effects of stereoscopy on consciousness. Using dye-infused aluminum prints for the intense colors achieved, I hope to not only raise awareness of how we see but also explore the difference between the artist and scientist?art uses the visual experience, not only empirical thinking, to further the viewer's awareness of the process of seeing.

METHOD: A show of aluminum prints will be exhibited in a corner of the display area. ChromaDepth® 3D glasses have been employed for the spontaneity achieved, reasons for which have already been discussed in my previous SPIE papers 64921I-1 and 90141F-1 (glasses will be provided with work). As will be shown in this paper, the artist must let go of preconceptions and expectations, despite what the evidence and experience may indicate in order to see what is happening in his work and to allow it to develop in ways he/she could never anticipate?this process will then be revealed to the viewer in a show of work. It is in the experiencing, not just thinking, where insight is achieved. Directing the viewer's awareness during the experience using stereoscopic imaging allows for further understanding of the brain's function in the visual process.

RESULTS: A cognitive transformation occurs, the preverbal "left/right brain shift," in order for viewers to "see" the space. Using what we know from recent brain research, these images will draw from certain parts of the brain when viewed in two dimensions and different ones when viewed stereoscopically, a shift, if one is looking for it, which is quite noticeable. People who have experienced these images in the context of examining their own visual process have been startled by the effect these have on how they perceive the world around them. For instance, when viewing the mountains on a trip to Montana, one woman exclaimed, "I could no longer see just mountains, but also so many amazing colors and shapes"?she could see beyond her preconceptions of mountains to realize more of what was really there, not just the objects she "thought" to be there.

NOVELTY: The awareness gained from experiencing the artist's perspective will help with creative thinking in particular and overall research in general. Perceiving the space in these works, completely removing the picture-plane by use of the glasses, making a conscious connection between the feeling and visual content, and thus gaining a deeper appreciation of the visual process will all contribute to understanding how our thinking, our normal brain processes, get in the way of our seeing what is right in front of us. We fool ourselves with illusion and memory?understanding this process and experiencing these prints can potentially help one come a little closer to reality.

References will include works by Steven Pinker, Eric Kandel, Semir Zeki, Joseph Campbell, Margaret Livingstone, and William Blake, amongst others.

9391-11, Session 3

Assessing the benefits of stereoscopic displays to visual search: methodology and initial findings

Hayward J. Godwin, Univ. of Southampton (United Kingdom); Nicolas S. Holliman, The Univ. of York (United Kingdom); Tamaryn Menneer, Simon P. Liversedge, Univ. of Southampton (United Kingdom); Kyle R. Cave, Univ. of Massachusetts Amherst (United States); Nicholas Donnelly, Univ. of Southampton (United Kingdom)

CONTEXT: Visual search involves the examination of the environment in an attempt to detect a target object. This can include everyday tasks such as searching for your keys on a messy desk, as well as safety-critical tasks such as airport X-ray baggage screening (searching for bombs and threat items) and medical search tasks such as radiographic image screening.

OBJECTIVE: The goal of the present project was to determine whether visual search performance could be improved by presenting objects on different depth planes to one another. Here we present an overview of the methodology involved in assessing visual search performance that we have developed, including the software and hardware setup, along with an overview of our initial findings.

METHOD: We developed a software suite for the creation of interlaced stereoscopic 3d images using C#/.net. The images formed the stimulus set for a series of experiments where participants searched through displays containing overlapping objects. The level of overlap was manipulated as part of three conditions: No Overlap (0% Overlap), Low Overlap (45% Overlap) and High Overlap (90% Overlap). We anticipated that presenting the overlapping objects on different stereoscopic depth planes to one another would enable participants to segment (i.e., segregate) and identify overlapping objects more easily than when the objects were presented on a single stereoscopic plane. We recorded eye-movement behaviour to gain detailed insights into the information-processing taking place as participants searched through overlapping displays.

RESULTS: We found that participants required more time to search the displays as the level of overlap in the displays increased. Participants were also less likely to detect targets at higher levels of overlap. Examination of the eye movement data revealed that, when overlap was high, participants were less likely to fixate (i.e., look at) targets and less likely to detect targets when they did fixate them, compared with the conditions with less overlap. Crucially, and in line with our predictions, the presence of depth information in the displays did not improve response accuracy, but did reduce the time needed to search the displays and make a correct response. This reduction was only evident in the High Overlap condition, suggesting that depth information only aids performance in highly complex displays.

NOVELTY: There are two key novel contributions to the present work. First,

we have developed (and will describe in detail) the methodology that was used to assess visual search performance in displays containing depth using modern eye-tracking technology. Our goal in doing so is to foster the development of further experimentation using our methodology. The second key novel contribution is the finding that the presence of depth information in the displays speeds search when there is a high degree of overlap and complexity in the displays.

9391-12, Session 4

Small form factor full parallax tiled light field display

Zahir Y. Alpaslan, Hussein S. El-Ghoroury, Ostendo Technologies, Inc. (United States)

CONTEXT

There have been many attempts at bringing glasses-free 3D displays to market and some of the reasons for the failure of these products can be summarized as: low resolution, lack of vertical parallax, lack of focus cues and large form factor. This paper introduces a prototype full parallax light field display with small form factor, high resolution and focus cues.

OBJECTIVE

Small pixel pitch, emissive displays with high brightness and low power consumption are required to enable new light field display designs. Ostendo's QPI technology is an emissive display technology with 10 μm pitch pixels, high brightness and low power consumption. The current tiled light field display prototype is a proof of concept that demonstrates possibility of a small form factor, high resolution, full parallax, mobile light filed displays.

METHOD

We created a very small form factor full parallax tiled light field display using a 4x2 array of light field display tiles. Each individual display tile makes use of the new display technology from Ostendo called Quantum Photonic Imager (QPI). We combined custom designed micro lens array layers with monochrome green QPIs to create very small form factor light field display tiles. Each of these small light field display tiles can address 1000 x 800 pixels placed under an array of 20 x 16 lenslets each with 500 μm in diameter.

RESULTS

These small light field displays are tiled with small gaps to create a tiled display of approximately 48 mm (W) x 17 mm (H) x 2 mm (D) in mechanical dimensions. The tiled display addresses total of 6.4 mega pixels with an array of 80 x 32 lenslets to create a light field with more than 120 mm depth of field, 30° viewing angle @ 60 Hz refresh rate.

NOVELTY

To our knowledge this is the first full parallax light field display with such a small form factor and large depth of field. This display and its subsequent iterations will play a big role in enabling mobile full parallax light field displays.

9391-13, Session 4

Load-balancing multi-LCD light field display

Xuan Cao, Zheng Geng, Mei Zhang, Xiao Zhang, Institute of Automation (China)

CONTEXT:

Multi-LCD architecture provides an off-the-shelf solution for 3D display by operating LCDs as dynamic parallax barrier[Jacobs et al. 2003] or space light modulator[G. Wetzstein, et al. 2012]. However, the quality of displayed light field (LF) in these reported literatures was not satisfying.

OBJECTIVE:

The capacity of generating full-fledged LF in small number layers (e.g. 3) of LCD is deficient, due to limited degree of freedom to reproduce a dense target LF(e.g. 8x8). Simply adding additional LCDs can improve the LF display quality slightly, at a price of significantly reducing the brightness of image, because extra LCDs result in much lower light transmittance.

To solve this problem, we propose a load-balancing Multi-LCD light field display design architecture to 1)enhance the LF reproducing capability of Multi-LCD and improve LF display quality without adding extra LCD layers, and 2)increase the luminosity of image ensuring a comfortable visual perception level in indoor lighting environment. The final goal is to construct a real-time indoor LF display system with higher level of LF fidelity.

METHOD:

For enhancing the LF reproducing capability of Multi-LCD, our main strategy is to reduce the decomposition load for each frame. We construct an over-determined equation for the LF decomposition and define the ratio of rank to columns in transform matrix "load factor". Our experimental results show that adding one more LCD layer makes very limited contribution to reduce the "load factor" but may reduce display brightness and increase computational cost.

Inspired by the temporal multi-frame decomposition method developed by [G. Wetzstein, et al. 2012], we propose a new spatial decomposition strategy that utilize multiple small areas (zones). The target LF is divided into multiple zones and all zones are reproduced in sequence. For each sub-zone of target LF, the whole multi-LCD pixel units are driven to reproduce the sub-LF. By resolving the target LF, the Multi-LCD's LF reproducing capability is further explored. As a result, higher quality of LF display can be achieved without adding extra LCDs. Different with static directional backlight in [G. Wetzstein, et al. 2012], we design a dynamic directional backlight, alternately changing the direction of backlight, to provide directional light source for each sub-LF.

RESULTS:

We build a three-layer VH242H LCDs system at resolution of 1920x1080. The 8x8 target LF decomposition is implemented on one NV Geforce 690 card with no less than PSNR of 30dB. The display system provides enough luminosity and runs in indoor lighting environment like general TV. We also develop a camera array to provide real-time LF stream for our load-balancing display system to reappear the real scene at 24 fps.

NOVELTY:

1) A novel multi-layer and multi-zone joint optimization and decomposition architecture for Multi-LCD display: reducing the decomposition load and achieving higher LF display quality(PSNR) and better virtual perception in indoor lighting environment without adding extra LCDs.

2) A dynamic directional backlight: combining one LCD panel and lenticular lens but no requirement on slanting lenses; increasing the brightness of backlight by allowing multiple pixels under each lens to illuminate; establishing an independent angle coordinate system for each cylindrical lens and the direction of backlight can be changed dynamically.

3) A real-time light field capturing and display system: developing a camera array controlled by distributed MCUs to provide LF stream of the real scene; the proposed load-balancing LF decomposition not only improves the LF fidelity but also matches multi-GPUs operational mode well.

Reference

[1] G. Wetzstein, D. Lanman, M. Hirsch, R. Raskar. Tensor Displays: Compressive Light Field Synthesis using Multilayer Displays with Directional Backlighting. Proc. of SIGGRAPH 2012 (ACM Transactions on Graphics 31, 4), 2012.

Project Link: http://web.media.mit.edu/~gordonw/TensorDisplays/

ACM DL Link: http://dl.acm.org/citation.cfm?id=2185576

[2] Jacobs, A., M Ather, J., Winlow, R., Montgomery, D.,Jones, G., Willis, M., Tillin, M., Hill, L., Khazova,M., Stevenson, H., AND Bourhill, G. 2D/3Dswitchable displays. Sharp Technical Journal, 4, pp 1–5, 2003.

http://sharp-world.com/corporate/info/rd/tj4/pdf/4.pdf

Conference 9391:
Stereoscopic Displays and Applications XXVI

9391-14, Session 4

Light field display simulation for light field quality assessment

Rie Matsubara, Zahir Y. Alpaslan, Hussein S. El-Ghoroury, Ostendo Technologies, Inc. (United States)

CONTEXT

This paper is focused on light field reconstruction, however, it takes light field reconstruction one step further by introducing the viewer in to the simulation. This is important because it can help pave the way for future light field compression performance evaluation methods.

OBJECTIVE

The purpose of this research is to create the perceived light image from a viewer's perspective. This enables:

1- A light field display designer to predict perception related problems before building the light field display.

2- Predict the result of any image processing and/or compression related artifacts before having an actual display.

METHOD

Our light field display simulation software that takes in to account the viewer's position, pupil diameter, viewer focus position and display parameters to create a simulation of the observed light field image.

RESULTS

Our light field simulation software has been used in designing full parallax light field displays that range in resolution from mega pixels to giga pixels. We have also used it in identifying and eliminating compression artifacts related to compressed rendering. The simulation results follow the real world observations very closely.

NOVELTY

Previous publications in the field of light field reconstruction have been mainly involved in creating images at different focal planes or creating 3D reconstruction of the objects. However, this research includes the viewer characteristics such as viewing distance, pupil diameter and focus location to create a more realistic evaluation of a light field created by a display.

9391-15, Session 4

Integration of real-time 3D capture, reconstruction, and light-field display

Zhaoxing Zhang, Zheng Geng, Tuotuo Li, Institute of Automation (China); Renjing Pei, Chinese Academy of Sciences, Institute of Automation (China); Yongchun Liu, Nanjing Univ. of Aeronautics and Astronautics (China); Xiao Zhang, Jiangsu Univ. (China)

CONTEXT

Effective integration of 3D acquisition, reconstruction (modeling) and display technologies into a seamless systems provides augmented experience of visualizing and analyzing real objects and scenes with realistic 3D sensation. Applications can be found in medical imaging, gaming, virtual or augmented reality and hybrid simulations. Although 3D acquisition, reconstruction, and display technologies have gained significant momentum in recent years, there seems a lack of attention on synergistically combining these components into a "end-to-end" 3D visualization system.

OBJECTIVE

We plan to design, build and test an integrated 3D visualization system in order to capture real-time 3D light-field images, perform 3D reconstruction to build 3D model of the captured objects, and display the 3D model on a large screen in the autostereoscopic multi-view way. By means of the seamless involvement of the real-time 3D light-field acquisition, 3D reconstruction and 3D display modules, our whole "end-to-end" system will offer an ideal solution for the various "end-to-end" 3D visualization applications mentioned above.

METHOD

(1) We have designed and built camera arrays with 8 or 16 cameras to acquire light-field 3D images of objects from various perspectives. The Structure from Motion (SFM) and Bundle Adjustment (BA) algorithms are implemented in the calibration process to minimize the calibration errors and ensure quality of reconstructed 3D model.

(2) We have developed a set of 3D dense surface reconstruction (3D-DSR) algorithms that is able to recover high resolution (pixel wise) 3D surface profile of the scene while preserving feature details and discontinuities. The algorithms are based on total variation L1 (TV-L1) optimization techniques and an optimization energy function is designed converting the inherent non-convex problem into two separate optimization processes that can be solved in alternative steps using duality theory.

(3) We post-process the reconstructed model including de-noising the background, smoothing the surface and so on. Then we render full set of light-field images required by the light-field 3D display based on the post-processed model and feed these images to the autostereoscopic light-field 3D display in real-time.

Experiments and simulation are performed and the results are presented to validate our proposed algorithms and demonstrate the successful integration of 3D capture, reconstruction and 3D display modules.

RESULTS

We have developed and implemented the design of a 3D acquisition and reconstruction modules based on camera array and GPU-powered algorithms to offer real-time full frame 3D reconstruction at around 5 fps. We use the fully textured 3D reconstruction model to generate multi-perspective 3D light-field images. These light-field images are then fed to an autostereoscopic multi-view 3D display module with up to 64 views (720p resolution per view) and a screen with the size of 1.2m by 0.7m. We are thus able to demonstrate the live 3D objects captured, processed, and displayed in this novel "end-to-end" 3D visualization system.

NOVELTY

Although there are a number of innovations in the development of 3D key components (3D capture, 3D reconstruction, and 3D display), we believe that a major innovation of our system resides on its seamless integration of all these components into an "end-to-end" 3D visualization system. The ability to perform real-time 3D reconstruction enable us to incorporate 3D camera and 3D display systems that are designed to have different numbers of views. Instead of directly acquiring different perspectives of the captured object with large numbers of cameras, as our previous publications have shown, we can now use much less cameras to generate textured 3D model, from which we can get free viewpoints of this reconstructed model. In addition, we widely use GPU resources to efficiently implement our time-consuming calibration, reconstruction and 3D image generation algorithms.

9391-500, Session Plen1

Analyzing Social Interactions through Behavioral Imaging

James M. Rehg, Georgia Institute of Technology (United States)

No Abstract Available

9391-16, Session 5

A large 1D retroreflective autostereoscopic display

Quinn Y. Smithwick, Disney Research, Los Angeles (United States); Nicola Ranieri, ETH Zürich (Switzerland)

Conference 9391:
Stereoscopic Displays and Applications XXVI

CONTEXT:

Three-dimensional stereoscopic projection in ride attractions currently requires riders to wear 3D glasses. Glasses require cleaning, are uncomfortable and need adjustment after rapid movements. Autostereoscopic projection would greatly enhance the experience and sense of immersion.

A way to provide each viewer autostereoscopic imagery is two projectors proximal to the eyes and a collaborating retroreflective screen [1,2]. Figure 1. Left/right eye images projected towards a retroreflective screen are sent back towards the projector, so each eye views only its corresponding projector's imagery. Only horizontal retroreflection is needed, and the screen can be vertically scattering, so s3D images are viewable below the projectors. [3.4,5,6]. Figure 2.

Preliminary research, figure 3, found personal projectors' fields of view (fov) (and retroreflected imagery) too small to provide an immersive experience. Projectors need calibration, and corner-push pin homography is too slow and imprecise for a large number of projectors.

OBJECTIVE

We aim to produce a wide fov autostereoscopic display. Picoprojectors modified with wide angle lenses are mounted unobtrusively over each viewing location. Pin-cushion distortion correction, rectification and cross-talk reduction are implemented for proper stereo fusion. The 1D retroreflective screen consists of retroreflector, anisotropic diffuser, and embedded fiber-optic array with optical sensors. Each projector's structured lighting is detected for calibration.

METHOD

A 10' tall by 6' wide screen is constructed using cornercube retroreflective sheets (Reflexite) with an anisotropic diffuser (Luminit , 40°vx0.2°h) overlay. Thirty optical fibers in a rectangular grid pierce the screen. Coupled photodetector (Taos) outputs are sampled by a microcontroller pair (Arduino) communicating to a computer via USB. Figure 4a,b.

Picoprojectors (Optoma PK320) modified with two stacked 0.5x wide angle lenses are placed 15' from the screen to meet 36° fov THX specification for immersive viewing [7]. Figure 4c.

Each projector projects graycode structured light detected by the screen's photodetector array. From this, Matlab programs estimate each projector's distortion coefficients and the homography matrix elements.

Using these parameters, GLSL shaders implemented on the graphics cards' GPU pre-distort frames decoded from SBS stereo movies, and implement cross-talk cancellation by subtracting weighted neighboring views [8].

RESULTS

Modifiing picoprojectors increased the fov from 15° to 60°, providing screen coverage. Projector pair calibration and image rectification/alignment was successful; each projector reprojecting light to identified fiber positions, figure 5a and producing undistorted image pairs, figure 5b,c. Wide fov immersive movies were viewable with easily fuseable bright stereo imagery, figure 6. Cross-talk cancellation was effective with slight contrast reduction. Screen seams are noticable but not distracting nor affecting stereo fusion.

CONCLUSIONS

We produced wide fov autostereoscopic imagery on a large 1D retroreflective screen by modifying overhead pico projectors with wide angle lenses. The screen's embedded sensor array allowed automated system calibration needed for many projectors. Projected structured lighting calibration [9] was included and extended to compute distortion correction. The projector's ability to travel with the ride vehicle projecting onto retroreflective screens provides the possibility of wide-spread stereoscopic imagery throughout an attraction.

9391-17, Session 5

Time-sequential lenticular display with layered LCD panels

Hironobu Gotoda, National Institute of Informatics (Japan)

Lenticular imaging is the technology that is frequently incorporated in auto-stereoscopic displays. A typical lenticular display consists of a lenticular sheet attached to an LCD panel and a light source called backlight. Here we propose an extension of lenticular display where additional layers of LCD panels are inserted between the lenticular sheet and backlight. The LCD panel nearest to the viewer (i.e., the one to which the lenticular sheet is attached) is made to function as an electronic shutter controlling the paths of light while the other panels show images that are periodically changing. This new configuration of LCD panels enables us to increase the angular resolution of the light field to be shown by the display, and thus to enhance the depth of field perceived by potential viewers.

The layered structure of LCD panels reminds us of multilayer displays investigated extensively in recent years. In fact, the proposed display looks like a multilayer display with an additional lenticular sheet attached to its surface. However, these two types of displays are fundamentally different. The main difference lies in the existence of shutter in our proposal. The shutter size is equal to the lens pitch of lenticular sheet, and each shutter controls whether the corresponding lenslet is "active" or "inactive". When a lenslet is active, some portions of LCD panels in the background are visible through the lenslet from the viewers. We choose the set of active lenslets so that the portions visible through the lenslets will never overlap. The active set changes periodically, and accordingly images shown on the panels also change.

The periodic control of shutter and images enables us to improve the angular resolution of the light field to be shown by the display. To what extent the resolution is improved depends both on the number of LCD panels and on the number of active set in each period. In the simplest implementation, where 2 LCD panels are used (one for shutter and one for image) in the display, and 3 active sets appear in each period, the angular resolution is multiplied only by 3. However, if we use more than 3 panels and 5 active sets, 8-times finer resolution can be achieved. The resolution could be improved further by allowing the shutter to modulate the intensity of light. According to the depth of field analysis of lenticular display, the depth of field (i.e., the range of 3D images that can be reproduced at maximum spatial resolution) is known to be proportional to the angular resolution. Our extension of lenticular display can provide deeper depth of field, which reduces the visual artifacts that are present in conventional lenticular displays.

We finally report on the prototype implementation of the proposed display. The display consists of 2 LCD panels periodically showing 3 light fields at 20Hz. The angular resolution is approximately 2 to 4 times finer than that of conventional lenticular display using the same lenticular sheet and LCD panels. We are working further to increase the number of LCD panels to 3.

9391-18, Session 5

Dual side transparent OLED 3D display using Gabor super-lens

Sergey Chestak, Dae-Sik Kim, Sung-Woo Cho, SAMSUNG Electronics Co., Ltd. (Korea, Republic of)

Transparent 3D display can be used to create attractive visual effects by combining the displayed 3D image with some real objects, visible through the display. It is apparent that to build transparent autostereoscopic 3D display one can use transparent 2D display with conventional parallax means like parallax barrier or lenticular lens array. Lenticular 3D array does not block the light and provides brighter stereoscopic image but high optical power of the lenticular lenses makes not possible to see through the lenticular sheet. Actually it works like one-dimensional diffuser. Optical power of lenticular lens in 3D display cannot be easily compensated without loss of functionality. We have found a method to overcome this problem.

To solve the problem we have applied second (additional) lenticlar sheet placed upon the opposite side of the transparent OLED, confocal and parallel with the first one. Thus the light-emitting layer of the OLED display appears to be sandwiched between two confocal lenticular sheets. Although the optical effect of additional lenticular sheets is not a kind of compensation effect, such structure does not destroy the image of the objects behind it. A combination of two confocal lenticular sheets is known as Gabor super-lens. Gabor super-lens configured as above is capable of

producing erected real image of the object1. If the object is located at some distance behind the display, its image appears at the same distance in front of the display. Therefore viewer can see the object behind the display clearly, but the object appears not in its real position but floating in front of the display. This floating image can be combined with 3D image, produced by 3D display.

Another novel feature of the display is dual side operation in 3D. On the reverse side of the display viewer can see the orthoscopic 3D image mirrored in respect to the front side image.

To our knowledge the described configuration of autostereoscopic 3D display has never been discussed before and its capabilities of displaying 3D image, floating image and dual side operation are of interest.

We have fabricated dual side 4-views 3D display, based on transparent OLED panel with resolution 128x160 pix and Gabor super-lens. Advantages and constrains of new display configuration will be demonstrated and discussed in the manuscript.

[1] G. Hembd-Sölner, R.F. Stevens, M.C. Hutley "Imaging properties of the Gabor superlens" J. Opt. A: Pure Appl. Opt. 1 (1999) 64-102

(http://iopscience.iop.org/1464-4258/1/1/013/pdf/1464-4258_1_1_013.pdf)

9391-19, Session 5

360-degree three-dimensional flat panel display using holographic optical elements

Hirofumi Yabu, Osaka City Univ. (Japan); Yusuke Takeuchi, Osaka City University (Japan); Kayo Yoshimoto, Osaka Univ. (Japan); Hideya Takahashi, Osaka City Univ. (Japan); Kenji Yamada, Osaka Univ. (Japan)

CONTEXT: Many 360-degree 3D display systems have been proposed for intuitive communication tools. However, most of them have complex mechanisms which are rotating screen, rotating mirror, many projectors and so on. To construct and adjust those systems is difficult because of their configurations. Therefore, it is difficult to observe 360° 3D images in many places for many people.

OBJECTIVE: We propose the 360-degree 3D display system whose configuration is simple. Our system consists of a liquid crystal display (LCD) and holographic optical elements (HOEs). Many HOEs are used as lenses to control lights of multiple pixels of an LCD to multiple viewpoints. HOEs can be produced on the thin polygonal glass plate. The size of proposed system is about the same with the size of an LCD because glass plate is just placed on an LCD. Nowadays, many flat panel displays are used as TV, mobile phone and so on in everywhere. Therefore, our system has potential to be used in a lot of situations by many people easily. Additionally, to display large 3D images and to increase viewpoints, we divided images for multiple viewpoints into stripe images and distributed them on the display. A lot of line images make one stripe image for one particular viewpoint. Therefore, observers can see the large 3D image around the system.

METHOD: The proposed system generates multiple viewpoints on a circle above the system. Observers can see the 3D image by looking down the system. The LCD displays the 2D image which composed of multiple parallax images. All images are divided into stripe images when they are synthesized to one image. The size of one line of stripe images is same with one HOE. Each HOE concentrates light from pixels which display stripe image to its viewpoint. HOEs for one particular viewpoint is located at intervals of one HOE. In gaps of those HOEs, other HOEs for other viewpoints are located. By placing HOEs like that, large 3D images are displayed. If we can use the high resolution LCD, large and high quality parallax images can be displayed at all viewpoints.

RESULTS: We constructed the prototype system to verify the effectiveness of the proposed system, we used 22.2 inch 4K monitor as an LCD. We used 10 trapezoid shaped glass plates for HOEs. The width of one HOE is 0.249 mm. Those glass plates are put side by side with their legs to shape the decagon whose center is missing. Each glass plate has the area for displaying 2 parallax images. The parallax image is trapezoid shape. Its height is 80.72 mm, upper base is 33.14 mm and the under base is 85.6 mm. We displayed the 3D image by using this system and confirmed that 20 viewpoints are generated and parallax images are observed at all viewpoints around 360° of the system.

NOVELTY: The 360-degree 3D display is proposed. Its configuration is very simple. We confirmed that prototype system provides 20 parallax images around the system.

9391-49, Session Key

What is stereoscopic vision good for? *(Keynote Presentation)*

Jenny C. A. Read, Newcastle Univ. (United Kingdom)

Stereoscopic vision has been described as "one of the glories of nature". Humans can detect disparities between the two eyes' images which are less than the diameter of one photoreceptor. But when we close one eye, the most obvious change is the loss of peripheral vision rather than any alteration in perceived depth. Many people are stereoblind without even realising the fact. So what is stereoscopic vision actually good for? In this wide-ranging keynote address, I will consider some possible answers, discussing some of the uses stereo vision may have in three different domains: in evolution, in art and in medicine.

Stereo vision incurs significant costs, e.g. the duplication of resources to cover a region of the visual field twice, and the neuronal tissue needed to extract disparities. Nevertheless, it has evolved in many animals including monkeys, owls, horses, sheep, toads and insects. It must therefore convey significant fitness benefits. It is often assumed that the main benefit is improved accuracy of depth judgments, but camouflage breaking may be as important, particularly in predatory animals. I will discuss my lab's attempts to gain insight into these questions by studying stereo vision in an insect system, the praying mantis.

In humans, for the last 150 years, stereo vision has been turned to a new use: helping us reproduce visual reality for artistic purposes. By recreating the different views of a scene seen by the two eyes, stereo achieves unprecedented levels of realism. However, it also has some unexpected effects on viewer experience. For example, by reducing the salience of the picture surface, it can affect our ability to correct for factors such as oblique viewing. The disruption of established mechanisms for interpreting pictures may be one reason why some viewers find stereoscopic content disturbing.

Stereo vision also has uses in ophthalmology. The sub-photoreceptor level of stereoacuity referred to in my opening paragraph requires the entire visual system to be functioning optimally: the optics and retina of both eyes, the brain areas which control eye movements, the muscles which move the eyes, and the brain areas which extract disparity. Thus, assessing stereoscopic vision provides an immediate, non-invasive assessment of binocular visual function. Clinical stereoacuity tests are used in the management of conditions such as strabismus and amblyopia as well as vision screening. Stereoacuity can reveal the effectiveness of therapy and even predict long-term outcomes post surgery. Yet current clinical stereo tests fall far short of the accuracy and precision achievable in the lab. At Newcastle we are exploiting the recent availability of autostereo 3D tablet computers to design a clinical stereotest app in the form of a game suitable for young children. Our goal is to enable quick, accurate and precise stereoacuity measures which will enable clinicians to obtain better outcomes for children with visual disorders.

9391-20, Session 6

Subjective contrast sensitivity function assessment in stereoscopic viewing of Gabor patches

Johanna Rousson, Jérémy Haar, Barco N.V. (Belgium); Ljiljana Platiša, Univ. Gent (Belgium); Bastian Piepers, Tom R. Kimpe, Barco N.V. (Belgium); Wilfried Philips, Univ. Gent (Belgium)

CONTEXT: Knowledge about the contrast sensitivity function (CSF) is crucial to properly calibrate medical, especially diagnostic, displays. Although 2D imaging remains more widespread than 3D imaging in diagnostic applications, 3D imaging systems are already being used and studies revealed that they could improve diagnostic performance, e.g. lead to earlier breast cancer detection. Nevertheless, very few studies have examined the CSF in 3D viewing even for the most basic parameters such as binocular disparity and 3D inclination.

OBJECTIVE: Assessing the CSF over a range of spatial frequencies in terms of: 2D orientations, depth planes (inverse of binocular disparities), and 3D inclinations (rotation of the 2D Gabor patch around the horizontal axis of the considered depth plane). For the purpose of validation, we measured also the 2D CSF and examined the statistical difference between our measurements and the well-established 2D CSF model of P.G.J. Barten. Finally, for our experimental data, we investigated the relationship between our 2D CSF and the 3D CSFs of different binocular disparities and 3D inclinations.

METHOD: We conducted subjective experiments following a 3-down 1-up staircase with nine human observers tested for normal visual acuity and stereovision. In the staircase experiment, the contrast of the stimulus was either decreased or increased depending on the observer's response to the preceding stimulus: target visible or target invisible. The stimuli were computer-generated stereoscopic images comprising a 2D Gabor patch as the target. The experiment was performed for seven different frequencies (0.4; 1; 1.8; 3; 4; 6.4; 10) expressed in cycles per degree (cpd), two planar orientations of the Gabor patch (vertical denoted by 2D:0 and diagonal denoted by 2D:45), two depth planes (the plane of the display, DP:0, and the depth plane lying 171 mm behind the display plane, DP:171), and two different 3D inclinations (3D:0 and 3D:45).The stimuli were 1920x1200 pixel large images displayed on a 24 inch full HD stereoscopic surgical display using a patterned retarder. The experiments were performed in a controlled environment with an ambient light of 0.8 lux.

RESULTS: Results of Friedman test suggested that, for all considered spatial frequencies, no significant differences were found between the CSF measured for pure 2D setup (2D:0, 3D:0 and DP:0) and the 2D CSF extracted from the mathematical model developed by P.G.J. Barten. Additionally, for 2D:0, 3D:0, and frequencies of 0.4 cpd and 1 cpd, medians of the measured contrast sensitivities and results of the Friedman test indicated that the CSF was significantly lower for DP:171 than for DP:0; for spatial frequencies above 1cpd, the CSF was not affected by the change in binocular disparity. For all planar orientations at DP:171, no differences were found between 3D:0 and 3D:45; thus CSF was not affected by 3D inclinations. Finally, for all 3D inclinations at DP:171, no differences were found between the two planar orientations (2D:0 and 2D:45).

NOVELTY: To the authors' knowledge, this is the first report of an experimental study exploring the impact/effects on the CSF of the common parameters of 3D image displaying, especially, binocular disparity and 3D inclination.

9391-21, Session 6

An objective method for 3D quality prediction using visual annoyance and acceptability level

Darya Khaustova, Orange SA (France); Olivier Le Meur, Univ. de Rennes 1 (France); Jerome Fournier, Emmanuel Wyckens, Orange SA (France)

CONTEXT

To assess the video quality of stereoscopic contents, the measure of spatial and temporal distortions became incomplete because of the added depth dimension. Improperly captured or rendered stereoscopic information can induce visual discomfort, which has an impact on overall video 3D QoE independently of image quality [1, 2]. Therefore, any 3D service should minimize visual discomfort perceived by its customers while watching stereoscopic contents. Several software for stereoscopic quality monitoring are available on the market (StereoLabs tool, Cel-Scope, Sony MPE-200 etc.). However, they are able just measure technical parameters (view asymmetries, disparities) relevant to comfort issues but nothing is known about their impact on human perception. Therefore, for stereoscopic quality monitoring it would be practical to avoid time consuming subjective tests and characterize objectively an impact of technical parameters on human perception.

OBJECTIVES

The first objective of the present study is to present and validate a new objective model that predicts objectively the impact of technical parameters relevant to visual discomfort on human perception. The model consists of three color categories that characterize a detected problem in accordance with evoked perceptual state:

- Green – no problem perceived.
- Orange – problem is acceptable, but induces visual annoyance.
- Red – unacceptable problem level.

Perceptual state reflects the viewers' categorical judgment based on stimulus acceptability and induced visual annoyance. Therefore, each perceptual state can comply with one or several perceptual thresholds (visual annoyance, acceptability). Among detected problems can be various views asymmetries, excessive disparities etc.

The boundary between "Green" and "Orange" categories defines the visual annoyance threshold (inspired by impairment scale [3]) and between "Orange" and "Red" categories defines the acceptability threshold (see Figure 1.a). By default, acceptability threshold level is defined as 50% e.g. 50% of viewers would rank the detected problem as unacceptable. The acceptability threshold might be also adapted based on service requirements. For example 80% of acceptability for cinema, 50% for 3DTV at home etc. Any selected percentage of acceptability defines the upper bound of Red category. The visual annoyance perceived by a percentage of viewers is quantified by the upper bound of Orange category. Once thresholds limiting the boundaries are known, it is possible to predict perceptual state objectively: if the detected impairment is higher than its acceptability threshold, then category is Red; otherwise if it is higher than its visual annoyance threshold, then Orange, otherwise Green.

The thresholds can be adopted from the state of the art studies or determined via subjective test using any standard method. However, in both cases it is important to make sure that thresholds were received under the same conditions (screen size, viewing distance, and 3D technology) as the target 3D system. Also the thresholds can be defined directly if objective model is used as subjective scale. This method allows defining both perceptual thresholds in the same subjective test. Therefore, the second objective of the current study is to verify whether it is possible to use the objective model as a subjective scale to assess both perceptual thresholds in the same time.

METHOD

Five different view asymmetries (focal length mismatch, vertical shift, and rotation, green and white level reduction) were used to validate the proposed model via series of subjective experiments. Three uncompressed stereoscopic scenes with controlled camera parameters were processed to introduce 5 degradation levels for each asymmetry. Furthermore, 30 subjects evaluated proposed stereoscopic scenes using modified SAMVIQ methodology [4], where Continuous Quality Scale was substituted with Color Scale.

The Color Scale was constructed from the proposed objective model as categorical scale with labels (Categories: [Green; Orange; Red], correspondent labels are [Acceptable, not annoying; Acceptable, but annoying; Unacceptable]), which is illustrated in Figure 1b. Basically, the constructed subjective Color Scale is composed of two scales: acceptability scale and visual annoyance scale.

In order to assign a category to the viewed stereoscopic stimulus, observers could use following two-steps algorithm: (1) Evaluate if the stimulus is acceptable. If true, pass to the second step; if false, choose Red category. (2) Evaluate if the stimulus is visually annoying. If true – Orange category; false – Green.

Collected subjective judgments were compared to objectively estimated categories. Objective predictions were made using acceptability and visual annoyance thresholds obtained for the same stimuli during another series of subjective experiments [5].

Conference 9391:
Stereoscopic Displays and Applications XXVI

RESULTS

Subjective test results suggest that it is possible to classify detected problem to one of the objective categories using corresponding acceptability and visual annoyance thresholds. For example, if amount of vertical shift exceeded acceptability threshold, the stimulus was classified to Red category by viewers. If it is above acceptability threshold but below visual annoyance threshold it was classified to Orange category in subjective test. Therefore, our initial hypothesis is supported by the series of subjective tests. Besides, it was established that acceptability and visual annoyance thresholds can be obtained in the same test. However, the acceptability threshold obtained in simple subjective test, where observers are only asked to judge a stimulus acceptable or not do not match acceptability threshold obtained with Color Scale. This can be explained by appearance of the Orange category, where some observers change their mind from "not acceptable" in favor of "acceptable, but annoying".

REFERENCES

[1] Tam, W. J., Stelmach, L. B., and Corriveau, P. J., "Psychovisual aspects of viewing stereoscopic video sequences," (1998), pp. 226-235.

[2] Kaptein, R. G., Kuijsters, A., Lambooij, M. T. M., Ijsselsteijn, W. A., and Heynderickx, I., "Performance evaluation of 3D-TV systems," San Jose, CA, USA, (2008), pp. 680819-11.

[3] ITU, "Recommendation ITU-R BT.500-13. Methodology for the subjective assessment of the quality of television pictures.," Broadcasting service (television), 2012.

[4] ITU, "Recommendation ITU-R BT.1788.Methodology for the subjective assessment of video quality in multimedia applications. ," Broadcasting service (television), 2007.

[5] Chen, W., "Multidimensional characterization of quality of experience of stereoscopic 3D TV," PhD Thesis, IRCCyN, 2012.

9391-22, Session 6

Disparity modification in stereoscopic images for emotional enhancement

Takashi Kawai, Daiki Atsuta, Sanghyun Kim, Waseda Univ. (Japan); Jukka P. Häkkinen, Univ. of Helsinki (Finland)

No Abstract Available

9391-23, Session 6

Preference for motion and depth in 3D film

Brittney A. Hartle, York Univ. (Canada); Arthur Lugtigheid, Univ. of Southampton (United Kingdom); Ali Kazimi, Robert S. Allison, Laurie M. Wilcox, York Univ. (Canada)

CONTEXT:

While heuristics have evolved over decades for the capture and display of conventional 2D film, it is not clear these always apply well to stereoscopic 3D (S3D) film. Further, S3D filmmakers have an additional variable, the range of binocular parallax in a scene, which is controlled by varying the camera separation (IA). While the current popularity of S3D film has prompted research on viewer comfort, relatively little attention has been paid to audience preferences for filming parameters in S3D.

OBJECTIVE:

Here we evaluate observers' preferences for moving S3D film content in a theatre setting. Specifically, we examine preferences for combinations of camera motion (speed and direction) and stereoscopic depth (IA).

METHOD:

S3D footage of a professional dancer performing a six-second dance phrase was captured using a 3D camera rig mounted on a dolly and track system. The dancer was filmed using three IAs (0, 10, 65 mm), two directions of camera motion (lateral, towards the dancer) and three speeds (0.11, 0.15, 0.45 m/s). The resulting 18 video clips corresponded to unique combinations of IA, speed, and direction. Clips were combined to create all pairings, each separated by a brief blank interval. The pairs were put in random sequences in a playlist.

Eighty students participated in a paired comparison task. They were grouped in either the motion in-depth condition (n=39) or the lateral motion condition (n=41). Testing was conducted in a 16-seat screening room with a Christie Digital Mirage projector, Xpand active eyewear, and a 10 x 20 ft screen. Prior to testing all observers completed a stereoscopic vision test. On each trial, observers viewed a pair of clips and indicated which they preferred using a remote device; responses were recorded electronically.

RESULTS:

The amount of stereoscopic depth (IA) had no impact on clip preference regardless of the direction or speed of camera movement. Preferences were influenced by camera speed, but only in the in-depth condition where observers preferred faster motion.

NOVELTY:

Viewer preferences for moving S3D content may not depend on the same factors that influence visual comfort. Contrary to earlier studies indicating that slower speeds are more comfortable for viewing S3D content, we found that observers preferred faster camera movements in depth. While most studies of visual comfort focus on factors affecting visual fatigue, here we considered aesthetic preference for individual shots. The discrepancy between these results indicates that there may be different influences or temporal dynamics at play. In our study, viewers may have experienced more compelling motion or improved vection (apparent self motion) in scenes with faster motion; such effects may have led to enhanced preference. Given the differences between the visual comfort literature and the preference results reported here, it is clear that viewer response to S3D film is complex and that decisions made to enhance comfort may, in some instances, produce less appealing content. Filmmakers should be aware of this balance between comfortable and preferable S3D content to provide the ideal viewing experience.

9391-24, Session 7

Microstereopsis is good, but orthostereopsis is better: precision alignment task performance and viewer discomfort with a stereoscopic 3D display

John P. McIntire, Paul R. Havig II, Air Force Research Lab. (United States); Lawrence K. Harrington, Ball Aerospace & Technologies Corp. (United States); Steve T. Wright, U.S. Air Force (United States); Scott N. J. Watamaniuk, Wright State Univ. (United States); Eric L. Heft, Air Force Research Lab. (United States)

Two separate experiments examined user performance and viewer discomfort during virtual precision alignment tasks while viewing a stereoscopic 3D (S3D) display. In both experiments, virtual camera separation was manipulated to correspond to no stereopsis cues (zero separation), several levels of microstereopsis (20, 40, 60, and 80%), and orthostereopsis (100% of interpupillary distance). Viewer discomfort was assessed before and after each experimental session, measured subjectively via self-report on the Simulator Sickness Questionnaire (SSQ). Objective measures of binocular status (phoria and fusion ranges) and standing postural stability were additionally evaluated pre and post-sessions. Overall, the results suggest binocular fusion ranges may serve as useful objective indicators of eyestrain, fatigue, and/or discomfort from S3D viewing, perhaps as supplemental measures to standard subjective reports. For the group as a whole, the S3D system was fairly comfortable to view, although roughly half of the participants reported some discomfort, ranging from mild to severe, and typically with the larger camera separations. Microstereopsis conferred significant performance benefits over the no-stereopsis conditions, so microstereoscopic camera separations might be of great utility for non-critical viewing applications. However, performance was

best with near-orthostereoscopic or orthostereoscopic camera separations. Our results strongly support the use of orthostereopsis for critical, high-precision manual spatial tasks performed via stereoscopic 3D display systems, including remote surgery, robotic interaction with dangerous or hazardous materials, and related teleoperative spatial tasks.

9391-25, Session 7

Effects of blurring and vertical misalignment on visual fatigue of stereoscopic displays

Sangwook Baek, Chulhee Lee, Yonsei Univ. (Korea, Republic of)

Several errors may cause visual fatigue in stereo images. In this paper, we investigate two issues, which may produce such visual fatigue. When two cameras are used to produce 3D video sequences, vertical misalignment is a problem. This problem may not present in professionally produced 3D video programs. However, in many low-cost 3D programs, it is still a major cause for 3D visual fatigue. Recently, efforts have been made to produce 3D video programs using smart phones or tablets, which may have the vertical alignment problem. Also, in 2D-3D conversion techniques, the simulated frame may have blur effects, which can also introduce visual fatigue in 3D programs.

To investigate the relation between the two errors (vertical misalignment and blurring in one image), we performed subjective tests using simulated 3D video sequences that include stereo video sequences with various vertical misalignment and several levels of blurring in one image.

We used the viewing conditions and laboratory environments for subjective assessment described in ITU-R Recommendation BT.500 and ITU-T Recommendation P.910. The viewing distance was set to 3H (three-times the display height). The employed display was a 55" LCD 3DTV and its resolution was 1920?1080 pixels. It uses passive glasses with a film-type patterned retarder (FPR) and can play the top-bottom and side-by-side formats. In each subjective test, 24 non-experts participated. Prior to the subjective tests, some screening tests were conducted: a vision test using a Snellen Chart (corrected eyesight > 1.0), a color blindness test and a stereoscopic acuity test using a Randot Stereo test (stereotest-circle level > 4). After the experiments, all the evaluators' votes were screened. If an evaluator was rejected, additional subjective tests were performed to replace the evaluator. Prior to the formal experiment, a short training session was given in order that evaluators would be familiar with impairment characteristics and the rating method.

Analyses of the subjective data (visual fatigue) and the two errors showed strong relationship with some saturation effects. In this paper, we present in-depth analyses along with some objective models that can predict visual fatigue from vertical misalignment and blurring.

9391-26, Session 7

Subjective and objective evaluation of visual fatigue on viewing 3D display continuously

Danli Wang, Yaohua Xie, Yang Lu, Institute of Software (China); Xinpan Yang, Institute of Software, Chinese Academy of Sciences (China); Anxiang Guo, Shanxi Electric Power Research Institute (China)

CONTEXT: In recent years, three-dimensional (3D) displays become more and more popular. However, they also have some disadvantages which may prevent their applications in more fields. One of the major disadvantages is the visual fatigue caused by viewing 3D displays. Therefore, many researchers devoted to evaluation methods for visual fatigue[1]. Most of the existing researches use either subjective or objective methods. Recently, these two kinds of methods were combined in some researches[2]. However, little research has combined subjective methods and objective ones to evaluate the visual fatigue in a continuous viewing 3D process.

OBJECTIVE: The aim of this study is to find the laws of the visual fatigue when viewing 3D contents continuously. The first part is how the tendency of visual fatigue in the entirely continuous viewing process differs from a discrete process. The second part is how the change of percentage of eye closure (PERCLOS) in the entirely continuous process differs from a discrete process. The third part is how the blink frequency (BF) changes with time in the continuous process. Finally, we aim to build a model to predict visual fatigue using objective indicators measured during the viewing process, and then verify it using subjective score.

METHOD: The experimental scheme of this method mainly includes an entirely continuous viewing process. During the process, subjects view stereo contents on a polarized 3D display and rate their visual fatigue whenever it changes. The contents consist of several images displayed randomly. A specialized tool, which uses a 10-point scale, is designed to rate visual fatigue subjectively. At the same time, an infrared camera coupled with an infrared light is used to record the eye movements of the subject. The recorded videos are later used to extract objective indicators such as PERCLOS and BF. We design an algorithm to extract these indicators nearly automatically. Before and after the viewing process, two questionnaires are used respectively to collect the subjective evaluation of visual fatigue for every subject. Each questionnaire contains fourteen questions related to different symptoms of visual fatigue. The results of subjective score and PERCLOS are then compared to those in reference [1]. Finally, a predictive model for visual fatigue in such a condition is built and verified.

RESULTS: Experimental results showed that the subjective score of visual fatigue increased with viewing time although it fluctuated sometimes. The symptoms of visual fatigue were generally more serious after viewing 3D contents than before. Blink frequency and PERCLOS had similar tendency to the visual fatigue, i.e., they also increased with time. Both subjective score and PERCLOS have greater values in the continuous viewing process than the discrete process described in an existing research. Based on the above results, a model was built to predict visual fatigue from PERCLOS during continuous viewing processes.

NOVELTY: First, subjective score and objective indicators are combined in an entirely continuous viewing process, during which the subjects are not interrupted. Second, the changes of subjective score and objective indicators with time are analyzed to find the tendency of visual fatigue, and the relationship between these indicators. Finally, a predictive model is built for visual fatigue in such a condition, and then verified with subjective score.

REFERENCES:

[1] Matthieu Urvoy, Marcus Barkowsky, Patrick Le Callet. How visual fatigue and discomfort impact 3D-TV quality of experience: a comprehensive review of technological, psychophysical, and psychological factors. Ann. Telecommun.2013, 68:641-655

[2] Wang Danli, Wang Tingting, Gong Yue. "Stereoscopic visual fatigue assessment and modeling". Proc. ST&A/SPIE, Stereoscopic Displays and Applications XXV, (2014).

9391-27, Session 7

Study of objective evaluation indicators of 3D visual fatigue based on RDS related tasks

Yi Huang, Yue Liu, Bochao Zou, Yongtian Wang, Dewen Cheng, Beijing Institute of Technology (China)

CONTEXT: Three dimensional (3D) display has witnessed rapid progress in recent years because of its highly realistic sensation and sense of presence to humanist users. The fatigue issues of stereoscopic display have been studied by many researchers. The purpose of our study is to investigate the objective evaluation parameters and methods of 3D visual fatigue based on random dot stereogram (RDS) related tasks.

OBJECTIVE: Stereoscopic visual fatigue is caused by the conflicts between

accommodation and convergence when users watch 3D images or videos. In order to study the objective evaluation parameters associated with 3D visual fatigue, an experiment is designed in which subjects are required to accomplish a task realized with RDS. The aim of designing the task is to induce 3D visual fatigue of subjects and exclude the impacts of depth cues. The visual acuity, critical flicker frequency, reaction time and correct judgment rate of subjects during the experiment are recorded and analyzed. Combination of the experimental data with the subjective evaluation scale is adopted to verify which parameters are closely related to 3D visual fatigue.

METHOD: 15 participated subjects aged between 20 and 35 was first screened for visual and stereo acuity. A refractometer was used to measure the visual acuity and a flicker fusion frequency machine was used to measure Critical Flicker Frequency (CFF). A 46-inch patterned retarder 3D display was used to present RDS which is generated by software to the users. The dot density of the stereograms was 50dots/deg? and its pattern is an arrow in the middle of the screen. The arrow has eight different parallaxes (-0.5°, -0.4°, -0.3°, -0.2°, +0.2°, +0.3°, +0.4°, +0.5°) and four different directions (Up, Down, Left, Right). The subjects wearing the polarized glasses sat 2 meters in front of the 3D display and judged whether the arrow is behind or in front of the screen by pressing number “1” or “2” on the keyboard, then pressed the cursor key to judge the direction of the arrow. Our program can record each keyboard operation and calculate the accuracy of the subject as well as the reaction time of each displayed RDS. The experiment for all subjects was conducted for 6 periods and each period lasts for 10 minutes. CFF was measured before and after each period. The visual acuity was measured and the subjects were also asked to fill in a subjective scale questionnaire about the symptoms related to visual fatigue after the last period of the experiment.

RESULTS: Experimental results show that the visual acuity and the time consumption of judging each RDS have no significant difference during the experiment. The correct rate of the task has decreased from 96.42% to 91.89%, while the CFF gradually decreased from 36.68Hz to 35.42Hz. The result is in line with the result of subjective evaluation. Analysis of variance (ANOVA) showed significant difference across the correct rate ($p<0.05$), but the CFF values didn’t show significant difference. This means that the correct rate of the task can be used as an objective parameter to evaluate the 3D visual fatigue.

NOVELTY: We design a task that excludes the interference of depth cues. When performing the designed task, the diopter, the critical flicker frequency, the reaction time and the correct rate of subjects were recorded during the experiment. Analysis of the experimental data show that the trends of the correct rate are in line with the result of subjective evaluation. In our future study more apparatuses will be applied to measure such parameters as ECG, EOG, EEG and EMG to find more dependable objective evaluation parameters.

9391-40, Session PTues

Enhancement of viewing angle with homogenized brightness for autostereoscopic display with lens-based directional backlight

Takuya Mukai, Hideki Kakeya, Univ. of Tsukuba (Japan)

CONTEXT:

We previously proposed an autostereoscopic display with directional backlight using Fresnel lens array. It, however, has a drawback that the viewing angle with homogenized brightness on its display is narrow horizontally and vertically.

OBJECTIVE:

The purpose of our study is to enhance the viewing angle with homogenized brightness on the display vertically and horizontally.

METHOD:

The system we have been developing was originally composed of a dot matrix light source and a convex lens array and a LCD panel. The orientation of directional light can be controlled when luminous areas on the light source change. By synchronizing the alternations of the directional light to the right eye and the left eye with the alternations of images for the right eye and the left eye, the viewers can observe a stereoscopic image. By use of a sensor device to track the viewer`s eye positions, he/she can keep on observing a 3D image while he/she is moving.

The conventional system, however, had non-uniform brightness and narrow viewing zone free from the crosstalk as its drawbacks. We have already proposed the methods to mitigate these problems. One way we have proposed to solve the non-uniform brightness is to place a vertical diffuser in front of the lens array. It is possible to blend the darker parts and the brighter parts on the lens array so that the brightness of the display becomes more uniform. One way we have proposed to expand the viewing zone free from the crosstalk is to attach a large aperture convex lens onto the surface of the lens array. It can expand the viewing zone by reducing the influence of the field curvature of the elemental lenses.

The remaining problem is that the viewing angle with homogenized brightness on the display is narrow vertically and horizontally. The peripheral part of the display region looks darker compared with the central part.

In this paper two optical approaches to enhance the viewing angle with homogenized brightness are proposed. The first one is to place two mirror boards on the upper end and the lower end between the lens array and the LCD panel horizontally. The second one is to move the large aperture convex lens from the surface of the lens array to just behind the LCD panel.

RESULTS:

By the first method, it is possible to reflect the directional light vertically and to make the upper and the lower of the display region brighter, which leads to enhancement of the vertical viewing angle. The second one is to utilize the directional light to viewer’s eyes from the light source more efficiently, which leads to the enhancement of horizontal and vertical viewing angles.

NOVELTY:

We have introduced the autostereoscopic display system employing the methods to improve uniformity of brightness and to expand the viewing zone free from the crosstalk so far. However, there still is a drawback that the viewing angle with homogenized brightness is narrow. In this paper the methods to solve it and the result when they are employed to the system are newly introduced.

9391-41, Session PTues

Effect of Petzval curvature on integral imaging display

Ganbat Baasantseren, National Univ. of Mongolia (Mongolia); Densmaa Batbayr, Ulaanbaatar State University (Mongolia); Lodoiravsal Choimaa, National Univ. of Mongolia (Mongolia)

Integral imaging (InIm) is an interesting research area in the three-dimensional (3-D) display technology because it is simple in structure and it can show full color, and full parallax. However, the InIm display has drawbacks such as shallow depth, narrow viewing angle, and resolution. InIm display uses simplest lens arrays, so displayed 3-D image has distortions because of lens array. A dominating distortion is a Petzval curvature. If a flat surface is used to examine the image with the on-axis image being focused, the off-axis images will be defocused, with a degree of defocusing growing with the square of the image height. The image plane is not truly a plane because of Petzval curvature. The main effect of Petzval curvature is that depth plane of InIm display becomes the curved plane array because each lens creates one curved plane. A radius of Petzval curvature of a thin lens depends on a focal length and a refractive index. In the first, we analyzed an effect of the Petzval curvature in InIm display. In a simulation, we used a thin lens which a focal length is 2.3 mm and a size of the lens is 1 mm because this lens is widely used InIm display. According to the result, the radius of Petzval curvature is equal to 3.4 mm and does not depend on the object distance and constants for a lens. When a distance between a lens array and a display is 1.27 mm, we create 9 integrated points on a central depth plane. From the results, a size of the smallest integrated

point is 0.36 mm and a size of the largest integrated point is 0.41 mm. The central depth plane is not plat and curved because of the Petzval curvature. It reduces a depth range and increases size of integrated pixel. In addition, shape of integrated pixels is not rectangular anymore. In real case, the light emitting angle of the 2-D display panel is large, so elemental lenses create many unwanted integrated pixels. The Petzval double-lens can compensate Petzval curvature, but it is difficult to be applied to elemental lenses. Therefore, InIm display using conventional lens array does not display the correct 3D image. The full explanation and the verification of the proposed method will be provided at the presentation.

9391-42, Session PTues

Data conversion from multi-view cameras to layered light field display for aliasing-free 3D visualization

Toyohiro Saito, Keita Takahashi, Mehrdad P. Tehrani, Toshiaki Fujii, Nagoya Univ. (Japan)

To achieve highly realistic 3-D visualization with motion parallax, 3-D displays are expected to present not only a pair of stereoscopic images but also many images corresponding to different viewing directions. 3-D displays that can output tens of images to different viewing directions, i.e. light field displays, have been developed. However, recent 3-D displays using parallax barriers or lenticular lens have a fundamental limitation in 3-D reproduction quality. There is a trade-off between the number of viewing directions and the resolution for each viewing direction because all directional images share a single display device with finite pixels. Meanwhile, a new display, namely layered light field display, is not limited by this trade-off and expected to simultaneously support many viewing directions and high resolution for each direction.

Below summarizes the principle of the layered light field display. The display consists of a backlight and several light attenuating layers placed in front of the backlight. The transmittance of each pixel on each layer is individually controlled. The light rays emitted to different directions from a single point on the backlight, pass through different points on the layers. Consequently, the displayed images can be direction dependent. The transmittances of the layers' pixels can be obtained inversely from the expected observation for each viewing direction. Therefore, given a set of multi-view images as input, we solve an optimization problem to obtain the layers' transmittances.

We have already developed a CG simulator of the layered light field display and evaluated the quality of displayed images (output quality) using real multi-view images as input. We have found that the output quality highly depends on the configuration of input images. For example, assume that input images are given from evenly-spaced viewing directions. The layers' transmittances are then calculated from the input images and finally fed to the display's simulator. Our evaluation has shown that aliasing artifacts are occasionally observed from the directions without input images. To prevent aliasing artifacts, it is necessary to limit the disparities between neighboring input images into one pixel (Plenoptic sampling theory's condition).

Aliasing-free condition is easily satisfied if the targets are CG objects. However, if the targets are real objects, it is not always possible to capture multi-view images densely enough for satisfying the aliasing-free condition. To tackle this problem, we propose to use image based rendering techniques for synthesizing sufficiently dense virtual multi-view images from photographed images. We have already developed a system that can generate a free-viewpoint image from multi-view images captured by 25 cameras arranged in a 2-D (5 x 5) array. We will demonstrate evaluations on the output quality of this display given sparsely photographed multi-view images, where the images captured by the 25 cameras are firstly used to synthesize sufficiently dense virtual multi-view images, secondly the synthesized images are used to obtain the layers' transmittances, and finally layers' transmittances are fed to the simulator to display real objects without aliasing artifacts.

9391-43, Session PTues

Free-viewpoint video synthesis from mixed resolution multi-view images and low resolution depth maps

Takaaki Emori, Nagoya Univ. Graduate School of Engineering (Japan); Mehrdad Panahpour Tehrani, Keita Takahashi, Nagoya Univ. (Japan); Toshiaki Fujii, Nagoya Univ. Graduate School of Engineering (Japan)

Multi-view video can be used for several applications such as auto-stereoscopic 3D display and free-viewpoint video synthesis. Streaming application of multi-view and free-viewpoint video is potentially attractive but due to the limitation of bandwidth, transmitting all multi-view video in high resolution may not be feasible. Our goal is to propose a new streaming data format that can be adapted to the limited bandwidth and capable of free-viewpoint video streaming using multi-view video plus depth (MVD) 3D video format.

We have already considered multi-view video streaming scenario where all of the views are transmitted in a lowered resolution and the requested view in the high resolution. However, when the user changes the viewpoint, the newly requested view cannot be received immediately in the high resolution but it is available only in the low resolution. Therefore, we proposed to compensate the lack of resolution during the delay time by using the low resolution image of the current viewpoint, and the low and high resolution images of the previous viewpoint. The basic idea was to compensate high frequency components from the previous view for upsampling the current view.

In this research, we extend this idea to the free-viewpoint video streaming scenario, assuming a 1D horizontal camera arrangement. Given a requested free-viewpoint, we use the two closest views and corresponding depth maps to perform free-viewpoint video synthesis. Therefore, we propose a novel data format consisting of all views and corresponding depth maps in a lowered resolution, and the two views in the high resolution, which are the closest to the requested viewpoint.

When the requested viewpoint changes, the two closest viewpoints will change, but one or both of views are transmitted only in the low resolution during the delay time. Therefore, the resolution compensation is also required in this scenario. However, the requirement is different from multi-view streaming, because the two views are not individually displayed, but blended in order to synthesize a free-viewpoint video. According to the location of the free-viewpoint, there are three cases, (1) both views are high resolution, (2) one view is high resolution and another view is low resolution, and (3) both views are low resolution. Note that all views have corresponding depth maps in the low resolution.

Our approach for maintaining the quality of the free-viewpoint video during the delay time is to jointly optimize the upsampling of lowered resolution views and blending operation for free-viewpoint video synthesis. We examined our data format for this scenario when conventional upsampling method such as bicubic upsampling is combined with View Synthesis Reference Software (VSRS), provided by MPEG. By the conference, we expect to obtain the optimal joint upsampling and view synthesis method which can outperform previous works.

9391-44, Session PTues

Formalizing the potential of stereoscopic 3D user experience in interactive entertainment

Jonas Schild, Consultant (Germany); Maic Masuch, Entertainment Computing Group, University of Duisburg-Essen (Germany)

CONTEXT:

Playing a digital game with stereoscopic 3D vision is often expected to

automatically expedite the perceived quality of a game. In contrast, this works propose that the actual potential holds no automatic benefit but depends on a significant amount of adaptation. The question is how the adaptation can be framed to provide an optimal development and whether an impact on user experience can be controlled. The challenge in framing the actual potential is thereby twofold: the complexity of technical and perceptive affordances as well as the variety and subtlety of effects found in user studies.

OBJECTIVE:

This work provides an accessible way for understanding both the nature as well as the complexity of S3D potential regarding content development and user experience. We further regard our approach as a solid basis for future discussion we regard necessary to progress towards shaping S3D games as a distinct medium. Regarding the increasing availability of stereoscopic displays and hence growing amount of content, we support future quantification of our approach.

METHOD:

The development framework and the user reaction scheme are derived from three game development cases and four user studies conducted during the past five years. The cases have iteratively explored effects in user experience in existing games as well as in custom-developed game prototypes. The results of the cases and studies have been published at international conferences. This work performs a meta interpretation of these and other recent external studies to derive a formalization of the overall potential.

RESULTS:

Besides the development framework and the user reaction scheme the overall potential is proposed as follows: the experience consists of an engagement part and an involvement part. The engagement describes initial fascination about the stereo effect and complexity to use the technology and align it with human factors. It is a function of these effects which depends on the used display technology and human factors but is rather independent of the actual game content. The involvement describes user experience once a user is immersed and further involved over a longer period of time. The main found effects are a higher spatial presence but also an increase in visual discomfort or simulator sickness on the negative side. Secondary, indirect effects in involvement may occur in user performance, emotional reaction and user behavior as a function of the spatial presence in combination with the actually designed game content and human factors, not so much with the display technology.

NOVELTY:

The proposed formalization allows developers and researchers to address particular issues such as cumbersome adjustment of display technology and more importantly focus on how a long-term benefit can occur by designing appropriate content. The presented approach hints towards more actively using the higher spatial presence, e.g., in explorative tasks, and motivates optimizing software explicitly for stereoscopic user as a unique user group with potentially other decisions and behavior during interactions. Controlling this effect is difficult in practice but provides a substantial starting point for understanding S3D games as a distinct medium with its own potential.

9391-45, Session PTues

Development of binocular eye tracker system via virtual data

Frank Hofmeyer, Sara Kepplinger, Technische Univ. Ilmenau (Germany); Manuel Leonhardt, Nikolaus Hottong, Hochschule Furtwangen Univ. (Germany)

The success of stereoscopic representations is limited by the effect it has to the user. Nevertheless, S3D movies causing illness to the audience are still produced. Despite the improved knowledge of the filmmakers there are still fundamental issues in S3D production. One of these issues is the marginal consideration of the subjective assessment of the recipients' perception.

There are several software tools available to check stereoscopic parameters while filming or editing, including stereo camera assistance systems and real time correction. Mostly based on mathematical specifications, the implementation of objectively measure able factors about stereoscopic film perception is required to improve the final S3D experience.

The goal is to have an objective measuring tool providing required perception data. Hence, our aim is the development of a sufficient accurate and precise binocular eye tracking system. This includes a hardware prototype as well as software modules for optimized data interpretation.

Up to now purchasable eye tracking systems are not precise enough for issues concerning stereo 3D perception. The accuracy of viewing direction measurements of offered eye tracking systems is around 0.5 and 1 degrees, including carefully executed calibration. This is not accurate enough to triangulate the actual focused depth of the viewer due to the given ability of the human disparity perception. Besides this, there are other hurdles based on physiological conditions, for example, the size of the fovea, the discrepancy between the optical axis and the real viewing direction, or fast eye movements (tremor, drift, micro saccades). In addition to the physiological challenges, there are problems from a technical point of view, such as the resolution and recording frequency of tracking cameras.

Due to the complex coherence of these parameters, their single influences can hardly be isolated by an experiment with real test participants. Furthermore, the participants' condition (e.g., fatigue) and learning effects could affect test results. Hence, we had to find another way to validate a certain improvement of our software algorithms.

Influences of different parameters can be examined already in an early development stage with the usage of an ideal computer generated model of the human eyes. By the creation of virtual camera views of the pupil during the fixation of an exactly predefined point in the 3D room the eye tracking process can be simulated. Here, different parts of the tracking algorithms can be monitored for their usefulness and performance. This happens by rendering of technically faultless images from the eye cameras' view, without any lens distortions or image noise, with absolutely consistent illumination and without any disturbing shades or mirroring.

Subject of a first test was to approve the basic capabilities of our modeling approach, according to a typical eye-tracking setup, implemented equal in virtual modeling as well as a real installation. Here, first results show that it is possible to use the virtual model in order to explore the crucial thresholds of camera resolution and frame rate useful for the tracking algorithms leading to optimal results of the pixel computation in an easy and fast way.

9391-46, Session PTues

Two CCD cameras stereoscopic position measurement for multi fiber positioners on ground-based telescope

Zengxiang Zhou, Hongzhuan Hu, Jianping Wang, Jiaru Chu, Zhigang Liu, Univ. of Science and Technology of China (China)

Parallel controlled fiber positioner as an efficiency observation system, has been used successfully in LAMOST for four years. And there are several survey telescope project will be proposed in ngCFHT and rebuilt telescope Mayall. The fiber positioners on the telescope focal plane were calibrated the initial position and error compensation curve after installed in the focal plane. In the past, the single 2K CCD camera was used to capture one fourteenth total by one time. Now we tried to study a new CCD measuring system scheme by two cameras, the system could acquire positioners 3D coordinates by stereoscopic vision measurement. However the two cameras will bring some problems during the measurement, such as lens distortion correction, two camera position calibration and two cameras' high order equation fitting. The paper will present the detail of measuring system design and system solution, discuss the results in the experimental. With the stereo vision and image processing method, we studied the influencing factor for accuracy measurement. Finally we present baseline parameters for the fiber positioner measurement as a reference of next generation survey telescope design.

Conference 9391:
Stereoscopic Displays and Applications XXVI

9391-47, Session PTues

Usability of stereoscopic view in teleoperation

Wutthigrai Boonsuk, Eastern Illinois Univ. (United States)

CONTEXT: Recently, there are tremendous growths in the area of 3-D stereoscopic visualization. The 3-D stereoscopic visualization technology has been used in a growing number of consumer products such as the 3-D televisions and the 3-D glasses for gaming systems. Proponents of this technology argue that the stereo view of 3-D visualization increases user immersion and entertainment as more information is gained with the 3-D vision as compare to the 2-D view. However, it is still uncertain if additional information gained from the 3-D stereoscopic visualization can actually improve user performance in real world situations such as in the case of teleoperation.

OBJECTIVE: This paper examines whether the use of 3-D stereoscopic view might improve user performance in remote activities over typical 2-D viewing.

METHOD: Experiments were conducted using head mounted display (HMD) with separated video images on left and right eyes to create stereo vision of the environment. The videos were fed from two raspberry PI cameras attached on an RC car. User performance to control the car with 3-D and 2-D observation of the environment was recorded. Several variables such as different driving situations (obstacle vs. clear) and speed were evaluated. The response times in each situation were measured and analyzed to compare user performance utilizing 3-D versus 2-D visualizations.

RESULTS: Results from the pilot experiment showed that user performance was generally improved in tasks that require depth perception when stereo (3-D) rather than the 2-D visualizations were used. However, for situations where the 3-D effect is minimal such as the car ran into a flat wall, the user performance between 3-D view and 2-D view was not significantly different.

NOVELTY: Several studies had compared user performance in 3-D and 2-D visualization systems. However, most experiments in existing studies were simulated in virtual environment. There has been considerable doubt if results from these virtual environments can be applied to the real world situations. Thus, this study is conducted in order to alleviate this doubt. To the best knowledge of the author, this study is the first detailed investigation of usability test for teleoperated objects with stereoscopic view in real world environment.

9391-48, Session PTues

Using binocular and monocular properties for the construction of a quality assessment metric for stereoscopic images

Mohamed-Chaker Larabi, Univ. of Poitiers (France); Iana Iatsun M.D., XLIM-SIC (France)

More and more people are getting access to 3D. Despite the decrease of 3D industry income, 3D is still attracting the interest of spectators. Before being delivered to public, any 3D content has to be encoded, compressed and transmitted. All these treatments can impact the quality of the final product. Thus, it is essential to have a measurement tool allowing the estimation of the perceived quality and the quality of experience of a given stereoscopic content. To date, several trials have been proposed in the literature. Most of them are relying on a perceptual weighting of already published 2D quality metrics. Unfortunately, most of the communicated results are not satisfactory because it does not take into account the inner properties of the binocular perception such as the binocular fusion and rivalry.

In this work, we propose a full-reference metric for a quality assessment of stereoscopic images using two different properties and trying to mimic the fusion process happening in visual brain. Our metric takes in consideration the fact that most of the stereoscopic content is encoded in an asymmetric way. Therefore, it estimates the probability of fusion that can be obtained perceptually using left and right views. This probability will help in a sense that if the quality difference between left and right views is high, this will lead to binocular rivalry resulting in a global discomfort decreasing the quality of experience. At the contrary, if the quality of both views is close, the binocular fusion will happen and the quality of experience will depend upon the level of impairment.

The comparison between both views is performed only on the most salient area. For our model of visual saliency detection, we propose to exploit monocular depth in addition to 2D salient features. The idea lies in the fact that there are similarities between 2D and 3D attention behavior. Moreover, depth can be accurately predicted from a single 2D view. Therefore, our model performs a fusion between 2D saliency map obtained using efficient methods from the literature and monocular depth. The latter is estimated based on low level vision features and without any a priori information about scene starting from the proposal of Palou et al. It is based on the detection of T-junction points identifying occlusions. Our approach applies on one of the views depending on the notion of master eye that is implemented based on the dominant saliency.

The proposed metric has been tested on publicly available datasets, and its results show a good correlation with human judgment.

9391-51, Session PTues

Dynamic mapping for multiview autostereoscopic displays

Jing Liu, Univ. of California, Santa Cruz (United States); Tom Malzbender, Cultural Heritage Imaging (United States); Siyang Qin, Bipeng Zhang, Che-An Wu, James Davis, Univ. of California, Santa Cruz (United States)

Multiview autostereoscopic displays have several image artifacts which prevent widespread adoption. Crosstalk between adjacent views is often severe, stereo inversion occurs at some head positions, and legacy 2-view content is difficult to display correctly. We introduce a method for driving multiview displays, dynamically assigning views to hardware display zones, based on potentially multiple observer's current head positions. Rather than using a static one-to-one mapping of views to zones, the mapping is updated in real time, with some views replicated on multiple zones, and some zones left blank. Quantitative and visual evaluation demonstrates that this method substantially reduces crosstalk.

9391-501, Session Plen2

What Makes Big Visual Data Hard?

Alexei (Alyosha) Efros, Univ. of California, Berkeley (United States)

There are an estimated 3.5 trillion photographs in the world, of which 10% have been taken in the past 12 months. Facebook alone reports 6 billion photo uploads per month. Every minute, 72 hours of video are uploaded to YouTube. Cisco estimates that in the next few years, visual data (photos and video) will account for over 85% of total internet traffic. Yet, we currently lack effective computational methods for making sense of all this mass of visual data. Unlike easily indexed content, such as text, visual content is not routinely searched or mined; it's not even hyperlinked. Visual data is Internet's "digital dark matter" [Perona,2010] -- it's just sitting there! In this talk, I will first discuss some of the unique challenges that make Big Visual Data difficult compared to other types of content. In particular, I will argue that the central problem is the lack a good measure of similarity for visual data. I will then present some of our recent work that aims to address this challenge in the context of visual matching, image retrieval, visual data mining, and interactive visual data exploration.

Conference 9391:
Stereoscopic Displays and Applications XXVI

9391-28, Session 8

Multi-view stereo image synthesis using binocular symmetry based global optimization

Hak Gu Kim, Yong Ju Jung, Soosung Yoon, Yong Man Ro, KAIST (Korea, Republic of)

CONTEXT

In autostereoscopic displays, multi-view synthesis is one of the most important solutions to provide viewers with many different perspectives of the same scene, as viewed from multiple viewing positions. It is often difficult to generate visually plausible multi-view images because multi-view synthesis can induce very large hole regions (i.e., disocclusion regions) in the warped images at a distant target view position.

In existing literature, many view synthesis algorithms have used exemplar-based hole filling methods [1]. However, such methods could lead to poor results because of the limitation of the greedy way of filling the image with local patches [2]. Especially, in the extrapolated view, hole regions are much larger compared to those of the interpolated view, obtained from several reference images. In recent years, the global optimization technique has been proposed to mitigate the limitation of local optimization [3, 4]. The method based on global optimization not only considers similarity at the current position, but also uses the relation between neighboring patches. However, the global optimization technique alone is still weak in filling large holes for autostereoscopic displays, one of the reasons being that these methods do not consider binocular symmetry of the synthesized left- and right-eye images. Excessive binocular asymmetry can lead to binocular rivalry and binocular suppression [5, 6]. Consequently, it can cause the failure of binocular fusion and lead to poor viewing quality along with the already persisting large hole filling problem.

OBJECTIVE

Hole filling algorithms for multi-view stereo images in view synthesis should consider binocular symmetry as well as spatial consistency. Binocular asymmetry is one of critical factors in determining the level of visual discomfort and also has an effect on the overall visual quality of synthesized multi-view stereo images.

In this paper, we propose multi-view stereo image synthesis using binocular symmetric hole filling method that employs the global optimization framework with Markov random field (MRF) and loopy belief propagation (BP). Specifically, the proposed symmetric hole filling method can achieve the structural consistency in the target view (e.g., left image) and binocular symmetry between stereo images by using the information of already filled regions in the adjacent view (e.g., right image).

METHOD

The proposed method consists of two parts: 1) We design a new cost function in MRF to enforce the left- and right-image symmetry in the synthesized multi-view stereo image for binocular symmetry. Using the disparity map at a target view (e.g., left image) position, we can find specific regions in an adjacent view (e.g., right image) that corresponds to the hole regions to be filled in the target view. So, in order to calculate a more exact similarity between the target patch with some hole pixels and labels of MRF that consist of all NxM patches from the source region (i.e., non-hole region) in an adjacent view, we replace some hole pixels of target patch with corresponding pixels in an adjacent view and calculate the matching cost to find the patch with maximum similarity. 2) We use an efficient loopy BP to reduce high computational cost. Here, we additionally apply depth constraints against nodes and candidate labels to fill the hole region with only background information. It not only improves the performance of view synthesis, but also simultaneously helps in reduction of computational cost.

RESULTS

To evaluate the performance of the proposed method, we conducted subjective assessment experiments with diverse synthesized stereo images generated by the proposed and existing hole filling methods. In the subjective assessment, all experimental settings followed the guideline sets in the recommendations of the ITU-R BT.500-11 and BT.2021 [7, 8]. As a result, we obtained differential mean opinion scores (DMOS) of the overall visual quality for each test dataset. The subjective result of the proposed method was then compared with those of the existing methods. The results obtained after comparison showed that the performance of the proposed method outperformed those of existing methods.

NOVELTY

In the field of multi-view synthesis, most of the previous works have their own drawbacks, one of the most potent ones being the negligence of binocular asymmetry. Binocular asymmetry is one of the most crucial factors which influence the level of visual discomfort and overall viewing quality of stereoscopic images. Recently, in order to fill huge hole regions in the synthesized multi-view images, there have been a few methods [3, 4] that employ image completion based on the global optimization. However, they could cause binocular asymmetry and degrade the overall visual quality of multi-view stereo image in stereoscopic viewing. To maximize the overall visual quality of the synthesized multi-view stereo image in the stereoscopic viewing, we designed a novel hole filling method of multi-view stereo image synthesis that effectively combined two concepts: the global optimization framework and the binocular symmetry.

9391-29, Session 8

Depth assisted compression of full parallax light fields

Danillo Graziosi, Zahir Y. Alpaslan, Hussein S. El-Ghoroury, Ostendo Technologies, Inc. (United States)

CONTEXT

Full parallax light field displays require extremely high pixel resolution and huge amounts of information. Recent modulation devices achieve the desired pixel resolution but have difficulty coping with the increased data bandwidth requirements. To address this issue, new compression methods are necessary to reduce data bandwidth in order to accommodate high data demands.

OBJECTIVE

The 3D video coding format adopted by MPEG utilizes multiview texture and depth images. In this work, we use the same 3D video coding format of multiple views with associated per-pixel depth maps and propose a method to compress full parallax light fields. In contrast to current MPEG deployments that only target linearly arranged cameras with small baselines, our compression and synthesis method can deal with extremely high number of views arranged in a 2D array.

METHOD

Our proposed method consists of view selection inspired by plenoptic sampling followed by transform-based view coding, using view synthesis prediction to code residual views. First an adaptive selection of the camera views in the 2D camera array is performed to cover as much of the scene with minimum camera overlapping. The selected views have their texture and respective depth maps coded as reference views, and then are used by a modified multiple reference depth image based rendering (MR-DIBR) algorithm to synthesize the remaining views. The difference between the original views and the synthesized views is then coded and sent along with the references to the display.

RESULTS

We perform simulations using computer generated light fields and place a 3D object at different distances to the 2D capture array. We investigate the number of views necessary to reconstruct the light field using the view synthesis reference software, VSRS, and using the proposed MR-DIBR, and discuss the efficacy and limitations of our view selection algorithm. We also compare our compression results with proposals of full parallax light field encoding that utilize established video compression techniques, such as H.264/AVC, H.264/MVC, and the new 3D video coding algorithm, 3DV-ATM. In our experiments we obtain better rate-distortion performance and better preservation of the light field structures perceived by the viewer. In our analysis, we present different quality measurements based on peak-signal to noise ratio (PSNR) and also structural similarity (SSIM) to verify our results.

Conference 9391:
Stereoscopic Displays and Applications XXVI

NOVELTY

The novelty of our coding method for full parallax light field content is the utilization of a standard 3D video coding format. We developed a view selection algorithm and a unique 2D view synthesis algorithm to enable the utilization of multiview plus depth for full parallax light field coding. Additional to our previously published work on this subject, we describe the compression method used to code single reference images, corresponding depth maps and residual views. In this work we determine the conditions necessary for view sub-sampling of full parallax light fields and provide the optimal rate-distortion performance of view compression according to the scene content.

9391-30, Session 8

A 3D mosaic algorithm using disparity map

Bo Yu, Hideki Kakeya, Univ. of Tsukuba (Japan)

CONTEXT:

The effect of mosaic is usually used to protect witness, personal information, or intellectual property right. The effect of mosaic is needed for 3D videos also when they include such information.

OBJECTIVE:

There are two well-known methods to create mosaics in 3D videos. However, for 3D videos, the method of mosaic processing that does not lose image quality has not yet been discussed. To tackle this problem, we propose a new method to create 3D mosaics by using a disparity map.

NOVELTY:

One of the conventional methods to create 3D mosaics is to duplicate the area of mosaic from one viewpoint (left or right view) to the other viewpoint (Method 1). This method, however, has a problem that the mosaic area looks completely flat, for the areas of mosaic are the same in the left view and the right view. The other method is to create the mosaics separately in the left view and the right view (Method 2). Though certain depth can be perceived in the area of mosaics with this method, 3D perception is not good enough because the pixels of different depths are blended in each block of mosaic pattern.

METHOD:

To begin with, a mosaic pattern is created in the left view by averaging the pixel values in each block. Then, according to the disparity map, the corresponding position of pixels are found in the right view, which are averaged to make the mosaic pattern for the right view. After all blocks of mosaics are calculated, there remain some pixels that are not processed because of occlusions etc. We use a median filter to fill in these gaps. In the proposed method (Method 3) the mosaic of the image from one viewpoint is made with the conventional method, while the mosaic of the image from the other viewpoint is made based on the data of the disparity map so that the mosaic patterns of the two images can give proper depth perception to the viewer.

RESULTS:

We carried out subjective experiments comparing the conventional methods (Method 1 and Method 2) and the proposed method (Method 3). Three methods were ranked according to the quality of 3D perception by the subjects. In order to examine the results of the subjective experiments, statistical analyses were done according to Mann-Whitney U test. There was a significant difference between Method 1 and Method 3. When the size of block was 10 x 10, there also was a significant difference between Method 2 and Method 3. On the other hand, when the size of block was 20 x 20, there was no significant difference between them.

9391-31, Session 8

Post inserted object calibration for stereo video rectification

Weiming Li, Samsung Advanced Institute of Technology (China); Zhihua Liu, Kang Xue, Samsung R&D Institute China-Beijing, SAIT China Lab (China); Yangho Cho, Samsung Advanced Institute of Technology (Korea, Republic of); Xiying Wang, Gengyu Ma, Haitao Wang, Samsung R&D Institute China-Beijing, SAIT China Lab (China)

Stereoscopic display creates 3D imagery by sending different image views to the viewer's left and right eyes. Vertical disparity of corresponding image features across the views is a major source of visual discomfort. They can be corrected by stereo rectification, which is important for stereoscopic image processing. Most of existing stereo rectification methods assume a unique epipolar geometry for a stereo image pair. However, this work addresses a real-world problem where this assumption is not true. Specifically, we address the issue when the input stereo video has been enhanced with post inserted objects (PIOs) such as TV logos, captions, and other graphic objects. Since these PIOs are artificially created with 'standard' epipolar geometry, each pair of corresponding PIOs are aligned to the same row. When transformed by stereo rectification that aims at correcting errors in video contents, the PIOs become distorted and result in extra vertical disparity.

To solve this problem, this work proposes a PIO calibration method, which extracts, fills, and re-inserts the PIOs during stereo rectification. This is challenging since PIOs are difficult to extract due to their generic image appearance. The method proposes different PIO extractors to deal with PIO type variations that may appear in real stereo TV programs.

(1) A temporal feature based PIO extractor is used for PIOs whose pixels remain still compared with video content. Examples are TV station logos.

(2) A caption PIO extractor is used to detect inserted captions. It first extracts text strokes and then fit caption models to them. Stereo disparity is used to reject false detection since true captions are usually in the front.

(3) A graphic PIO extractor for generic graphic objects. It first segments the image into objects and then computes matching costs for each object along the same row in the other view. A PIO classifier is used based on the matching cost profile.

A stereo video rectification procedure is also designed with PIO calibration. The procedure includes video shot detection. The PIO calibration is only performed in the shots where vertical disparity is detected. This avoids unnecessary processing and video jitters due to inconsistent epipolar geometry estimations between frames. After the PIOs are extracted, they are removed and filled by interpolation. Finally, the PIOs are reinserted after the stereo image pair is rectified.

The proposed method is tested with real stereo videos with various types of PIOs. The test videos include stereo errors from various un-matched stereo-camera configurations. On average, the proposed method achieves a PIO detection rate (DR) of 86% at a false alarm rate (FAR) of 11%.

To the best of our knowledge, this is among the first work that addresses the PIO calibration issue. It is useful in:

(1) Stereoscopic displays in TV and mobile devices.

This work avoids errors in PIO regions for stereo rectification methods.

(2) Auto-stereoscopic displays.

When converting a stereo video to a multi-view video used for auto-stereoscopic displays, this work separates the PIOs and avoids their induced errors in depth-map estimation and multi-view image synthesis.

Conference 9391:
Stereoscopic Displays and Applications XXVI

9391-32, Session 9

A new type of multiview display

René de la Barré, Fraunhofer-Institut für Nachrichtentechnik Heinrich-Hertz-Institut (Germany); Silvio Jurk, Technical Univ. Berlin (Germany); Mathias Kuhlmey, Fraunhofer-Institut für Nachrichtentechnik Heinrich-Hertz-Institut (Germany)

The authors present a new render strategy for multiview displays which departs from the geometric relations of common 3D display design. They show that a beam emitted from a sub-pixel can be rendered with an individual camera direction determined by an algorithm. This algorithm also uses, besides the parameters of the display design, the desired viewing distance and the allowed increments of camera angle.

This enables high flexibility in the design of autostereoscopic displays. One advance is the possible reduction of the distance between the display panel and the image splitter. This enables a durable fixation of these two parts and eliminates inner reflections by directly bonding with optical glue. Another advance is the chance for realizing light weight glasses-free multiview displays by using a thinner carrier for the image splitter raster.

In contrast to the common multiview display, the new design approach enables a continued viewing zone without the usually diamond shaped sweet spots. The benefit lies in the equalized presentation of 3D content for users with strong different intra-pupillary distances. Particularly in the case of large enhanced viewing distances, the mixing of views is an advance because it enhances the viewing zone in depth.

The full range of render options enables shortening and widening the viewing distance continuously depending on the target input. In respect to the render effort, the number of images or render directions used in the presentation can be controlled. On the one hand, a number of views higher than the number of directional differently arranged elements eliminates the diamond shaped sweet spots and leads to a continuous viewing zone. On the other hand, a lower number of views leads to an approximate reconstruction of viewing zones. The missing smoothness in this reconstruction reduces the range for the change of viewing distance.

Finally, the optical quality in viewing zones with changes in viewing distance is dependent on the number of image directions used in the presentation. The complete view on the algorithm shows that the conventional design approaches, with analogies between the number of viewing zones and the number of views, are special cases in that algorithm.

The authors show the ability of the algorithm to continuously expand or compress the width of the multiplexed stripe images behind an image splitter element if different viewing distances are required. Accordingly, the algorithm for controlling the rendering angle and the multiplexing with artifact filtering is generic, as well as for integral and multiview design approaches using an image splitter raster. Even through the rays which represent the views that start in the subpixel position will be emitted slightly divergently, the algorithm will deliver the correct render directions. The paper introduces this approach for autostereoscopic displays with horizontal parallax and slanted image splitters. It shows the results of the ray simulation on the new 3D display design and the required content distribution. The authors compare these simulation results with the resulting views on a display setup realizing different viewing distances by the proposed content generation, as well as its filtering and multiplexing.

9391-33, Session 9

Compact multi-projection 3D display using a wedge prism

Soon-gi Park, Chang-Kun Lee, Byoungho Lee, Seoul National Univ. (Korea, Republic of)

CONTEXT:

Three-dimensional (3D) displays are widening their applications as display technologies are developed [1]. Among them large-sized 3D displays can provide immersive experience to observers because it can provide wide field of view. For this reason, 3D displays can be applied not only to entertainment area but also industrial field for training employees such as flight simulators as well as exhibition of products.

OBJECTIVE:

For achieving large-sized 3D displays, projection-type displays are generally used for implementation of autostereoscopic displays. A multi-view display made of multiple projection optics is a representative large-scale 3D display [2], and its principle is very similar to the integral floating system. Multi-projection systems based on integral floating method have advantages in realization of high-quality 3D images because of each view image is created by one projection-unit. As the resolution of view image is identical to the resolution of projection units, the resolution of the 3D system can be preserved as conventional 2D displays. However, a projection system requires large space to expand the size of projected images. For solving this problem, we propose a compact multi-projection system based on integral floating method with wave-guided projection.

METHOD:

Wave-guided projection can be a solution for reducing the projection distance by multiple folding of an optical path [3]. Because a wedge prism, which is used as a wave guide, does not change the effective optical path length, the size of projected image can be preserved. Also, other optical elements such as collimating lens can be used with the wedge prism. The proposed system is composed of multiple projection-units and an anisotropic screen made of collimating lens combined with a vertical diffuser. Multiple projection-units are aligned in line at the entrance of wedge prism. As projected images propagate through the wedge prism, it is reflected at the surface of prism by total internal reflections, and the final view image is created by collimating lens at the viewpoints.

RESULTS:

In this wave-guided projection system, the view image is generated on the line of optimal viewing distance like a conventional integral-floating-based multi-projection displays. The optimal viewing distance is decided by the optical configuration of the collimating lens and projection-units. By the Lens maker's equation and effective optical path length of the wave guide, the optimum viewing distance can be calculated according to the focal length of collimating lens. Also, the interval of the projection-unit is adjusted with the interval of viewpoint by the magnification of the collimating lens. In the experimental system, we employed three projection-units made of conventional off-the-shelf projection display.

NOVELTY:

By adopting a wedge prism, required projection distance can be effectively reduced for the multi-projection optics. We believe that the proposed method can be useful for implementing a large-sized autostereoscopic 3D system with high quality of 3D images using projection optics. In addition, the reduced volume of the system will alleviate the restriction of installment condition, and will widen the applications of a multi-projection 3D display.

REFERENCES

[1] B. Lee, "Three-dimensional displays, past and present," Phys. Today 66, 36–41 (2013).

[2] M. Kawakita, S. Iwasawa, M. Sakai, Y. Haino, M. Sato, and N. Inoue, "3D image quality of 200-inch glasses-free 3D display system," Proc. SPIE 8288, 82880B (2012).

[3] A. Travis, T. Large, N. Emerton, and S. Bathiche, "Wedge Optics in Flat Panel Displays," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (IEEE 2011), pp. 1-16

9391-34, Session 9

Integral 3D display using multiple LCDs

Naoto Okaichi, Masato Miura, Jun Arai, Tomoyuki Mishina, NHK Japan Broadcasting Corp. (Japan)

CONTEXT:

We have been developing an integral three-dimensional imaging system that enables a user to see spatial images without special glasses. To enhance

the quality of the integral three-dimensional images, a display device that has a lot of pixels is required. However, there is no display device that has a number of pixels greater than that of 8K Super Hi-Vision (SHV), so it is difficult to improve the quality of spatial images by using a single display device.

OBJECTIVE:

The objective is to improve the quality of the integral three-dimensional images by combining multiple LCDs and displaying with a number of pixels higher than that for one LCD.

METHOD:

When merely placing multiple LCDs in parallel, it is not possible to connect images without gaps caused by bezel of the LCDs. Thus, we developed a technology to combine multiple display images without gaps by placing a magnifying optical system in front of each display and enlarging the display images. Display images are imaged on a diffusion plate that is placed in front of the magnifying optical system, which is composed of two lens arrays and a concave Fresnel lens. The two lens arrays image an erect equal-magnification image, and by using a concave Fresnel lens in front of the two lens arrays, the display image is enlarged and imaged on the diffusion plate. The number of pixels is increased by combining multiple magnified images. An IP image is reconstructed by placing a lens array in front of the combined image. Because the image on the diffusion plate is distorted by the magnifying optical system, an image processing for compensation is performed with a computer, and multiple magnified images are combined seamlessly without distortion.

RESULTS:

We constructed a prototype of the display device by using four HD LCDs. Thus, the IP image was 4 times larger than that reconstructed by a single display. The pixel pitch of the HD display we used was 55.5 ?m. The number of elemental lenses was 212 horizontally and 119 vertically. The spatial image pixel count was 25,228, and the viewing angle was 28°.

NOVELTY:

So far, we have been successful in reconstructing a spatial image of approximately 100,000 pixels with the integral imaging system with SHV imaging devices by using the pixel-offset method [1]. However, it is difficult to increase the number of pixels by using a single display device. By adding the display and the magnifying optical system, it is possible to increase the number of pixels in proportion to the number of displays. Additionally, by using this structure, it is possible to make the overall structure of the device thinner than with the method of using a projector, so its application to home television can be expected in the future.

REFERENCE:

[1] Jun Arai, Masahiro Kawakita, Takayuki Yamashita, Hisayuki Sasaki, Masato Miura, Hitoshi Hiura, Makoto Okui, and Fumio Okano, “Integral three-dimensional television with video system using pixel-offset method,” Opt. Express 21, 3474-3485 (2013)

9391-35, Session 9

A super multi-view display with small viewing zone tracking using directional backlight

Jin Miyazaki, Tomohiro Yendo, Nagaoka Univ. of Technology (Japan)

A super multi-view display provides three-dimensional images by emitting a lot of rays of different colors depending on the direction from each point on the display. It provides smooth motion parallax without special glasses, and it is expected that the observer is free from the visual fatigue caused by the accommodation-vergence conflict. However, a huge number of pixels are required on a display device because high-density rays are required for good quality images and each ray needs corresponding pixel. We proposed a new method to reduce the required number of pixels by limiting rays emitted to only around observer’s pupils.

The display is based on the lenticular method. As stated above, the rays should be shot out to only around observer’s pupils. Therefore, the lenticular lens of which viewing zone angle is narrowed is used. But, due to the characteristics of the lenticular lens, the same image is seen repeatedly from different positions out of the viewing zone. It is called side lobe. Because of the side lobes, the rays for one eye enter the other eye. To suppress these side lobes, we proposed the lenticular lens is illuminated by the directional light. The direction of the directional light has to be changed to follow the observer’s eye.

However, a previous prototype display didn’t have a function controlling the direction of the backlight. Therefore, it fixed the direction of the directional light when we conduct an experiment. In this paper, we designed a backlight part which can control the directional light. In the case of the previous prototype display was composed of cylindrical lens and light source, it was the problems that involuntary rays interfered in adjacent lenses and the arrangement of light sources had to be on arc. These problems were due to the curvature of field. We designed a new structure which was added Fresnel lens and cylindrical lens in order to reduce an effect of the curvature of field. Fresnel lens decreases the angle between rays around the display edge and optical axis of cylindrical lens by concentrate the directional light in a viewpoint. Another cylindrical lens puts imaging plane close to optical axis. A liquid crystal panel (LC panel) with backlight is used as light source whose lighting pattern is controllable. A new prototype display controls the direction of the directional light by switching the display image of the LC panel.

We confirmed the principle of the proposed method by computer simulation. We decided the viewing distance is 1.0[m] in the simulation. And, we decided the viewing angles were 0[deg] and 10[deg]. A camera moved the line parallel to the display. The simulation results confirmed that we could see image corresponding from each viewpoint by the change of the photographed images. In addition, it confirmed suppression of the side lobes by couldn’t see the image outside the viewing zone. It was obtained similar results when there is a viewpoint in the front and a slant. By these results, we showed the effectiveness of the proposed method.

9391-36, Session 10

Real object-based 360-degree integral-floating display using multiple depth camera

Munkh-Uchral Erdenebat, Erkhembaatar Dashdavaa, Ki-Chul Kwon, Hui-Ying Wu, Kwan-Hee Yoo, Chungbuk National Univ. (Korea, Republic of); Young-Seok Kim, Korea Electronics Technology Institute (Korea, Republic of); Nam Kim, Chungbuk National Univ. (Korea, Republic of)

Conventional 360 degree integral-floating display (IFD) which is the combination of a volumetric light field display and an integral imaging technique, proposed by M.-U. Erdenebat et al., is a distinguished solution for the limited viewing angle of the integral imaging display that presents the full-parallax three-dimensional (3D) image in unlimited 360 degree viewing zone; however the vertical viewing angle was too narrow. Thereafter, the enhanced 360 degree IFD with wide vertical viewing angle using the anamorphic optic system (AOS) which is the vertically curved convex rotating mirror has been conducted. Both display systems are based on the computer-generated integral imaging technique which generates the elemental image arrays (EIAs) virtually through the computer graphics. Therefore, displaying the 3D image based on the real world object in 360 degree viewing zone is a very important issue for this kind of display.

In this paper, a novel 360 degree IFD system using multiple depth camera is proposed. The depth cameras acquire the depth and color information of selected real world object simultaneously from all viewing directions around the object. Then, the 3D virtual representations of the real object captured from the corresponding viewing directions are reconstructed respectively based on the acquired depth and color information by each depth camera, in the user computer. The special point registration function combines the multiple representations with each other by detecting the same points

which are reconstructed in the several representations and merge that points as single point. So eventually, the entire 3D object space including the depth and color information from all viewing directions is created. For the newly created 3D object space, the EIAs are generated for each given angular step of the AOS, and uploaded into the random-access memory of the digital micromirror device (DMD) in the generated sequence.

In the display part, the DMD projector projects two-dimensional EIAs sequentially in high-speed and a lens array directly integrates the initial 3D perspectives for each EIA. Through the double floating lenses, the initial 3D perspectives are relayed to the center of the spinning mirror where motor rotates it in synchronization with the DMD projection. 3D perspectives which relayed on the rotating mirror are tailored with each other by the horizontal direction and entire 3D reconstruction of the real world object is successfully displayed in the unlimited 360 degree viewing zone with wide vertical viewing angle. Also, the displayed entire 3D image is provided with perfect depth cues of human visual perception. The proposed method is verified by optical experiment.

9391-37, Session 10

Multi-layer 3D imaging using a few viewpoint images and depth map

Hidetsugu Suginohara, Hirotaka Sakamoto, Satoshi Yamanaka, Mitsubishi Electric Corp. (Japan); Shiro Suyama, Univ. of Tokushima (Japan); Hirotsugu Yamamoto, Utsunomiya Univ. (Japan) and The Univ. of Tokushima (Japan)

CONTEXT: The autostereoscopic display that has multiple display screens in the depth direction has been proposed. Images at different depth cause the difference between retinal images of each eye, it means disparity, and the viewer can perceive 3D image. Additionally, the multi-layer display has a real depth, so that it can show 3D image more naturally than single flat panel display with parallax barrier or lenticular lens array.

OBJECTIVE: In previous studies, a lot of multiple viewpoint images, so called Light Field, are needed to make multi-layer images that enable stereoscopic view on wide viewing angle. Because they try to display Light Field input on fewer screens, it takes time to do optimization and gives rise to some blurring.

METHOD: In this paper, we propose a multi-layer 3D imaging that makes multi-layer images from single 2D image with depth map. At first, we render multiple viewpoint images from 2D image using depth map. We get an initial solution by dividing 2D image input according to depth map so that we can do the simpler optimization to make multi-layer images from multiple viewpoint images. To confirm our method, we develop a prototype stacked two in-plane switching LCDs. We remove polarizer films between two LCDs to modulate rays from backlight by adding polarization angle on each LCD. And we use FPGA to enable synchronous image outputs to two LCDs.

RESULTS: We can get multi-layer images with less computation time, almost 1/10, compared with the optimization without using an initial solution according to depth map. Subjective test results show that it is enough to make multi-layer images using 3-viewpoints images to give a pop-up 3D image perception to the viewer. It improves the sharpness of the display image because fewer multiple viewpoint images cause less blurring on the optimization process. To avoid narrow viewing angle that caused by making multi-layer images from fewer multiple viewpoint images, we track the head motion of the viewer using Microsoft Kinect and update screen images in real time so that the viewer can maintain correct stereoscopic view within +/- 20 degrees area and can percept motion parallax at the same time.

NOVELTY: Using depth map, we succeed to make multi-layer images by less calculation resources. Additionally, the prototype we have developed can display sharper image with motion parallax on wide viewing area. Our system can display not only computer graphics but also natural image and video without difficulty because it requires only 2D image with depth map as inputs.

9391-38, Session 10

Evaluation of vision training using 3D play game

Jungho Kim, Soon Chul Kwon, Kwang-Chul Son, SeungHyun Lee, Kwangwoon Univ. (Korea, Republic of)

The present study aimed to examine the effect of the vision training, which is a benefit of watching 3D video images (3D video shooting game in this study), focusing on its accommodative facility and vergence facility. Both facilities, which are the scales used to measure human visual performance, are very important factors for man in leading comfortable and easy life. This study was conducted on 30 participants in their 20s through 30s (19 males and 11 females at 24.53 ± 2.94 years), who can watch 3D video images and play 3D game. Their accommodative and vergence facility were measured before and after they watched 2D and 3D game.

It turned out that their accommodative facility improved after they played both 2D and 3D games and more improved right after they played 3D game than 2D game. Likewise, their vergence facility was proved to improve after they played both 2D and 3D games and more improved soon after they played 3D game than 2D game. In addition, it was demonstrated that their accommodative facility improved to greater extent than their vergence facility. While studies have been so far conducted on the adverse effects of 3D contents, from the perspective of human factor, on the imbalance of visual accommodation and convergence, the present study is expected to broaden the applicable scope of 3D contents by utilizing the visual benefit of 3D contents for vision training.

9391-39, Session 10

Partially converted stereoscopic images and the effects on visual attention and memory

Sanghyun Kim, Waseda Univ. (Japan); Hiroyuki Morikawa, Aoyama Gakuin Univ. (Japan); Reiko Mitsuya, Takashi Kawai, Waseda Univ. (Japan); Katsumi Watanabe, The Univ. of Tokyo (Japan)

This study contained two experimental examinations of the cognitive activities such as visual attention and memory in viewing stereoscopic (3D) images. For this study, partially converted 3D images were used with binocular parallax added to a specific region of the image.

In Experiment 1, change blindness was used as a presented stimulus. The visual attention and impact on memory were investigated by measuring the response time to accomplish the given task. In the change blindness task, an 80 ms blank was intersected between the original and altered images, and the two images were presented alternatingly for 240 ms each. Participants were asked to temporarily memorize the two switching images and to compare them, visually recognizing the difference between the two. The stimuli for four conditions (2D, 3D, Partially converted 3D, Partially converted 3D distractor) were randomly displayed for 24 participants. While the 3D condition had a variety of parallax information, the partially converted 3D condition had binocular disparity provided only to the converting section in the image to convert the 2D images into 3D images. On the other hand, partially converted 3D distractor was used to examine the impact of visual attention at the wrong target on cognitive performance. The results of Experiment 1 showed that partially converted 3D images tend to attract visual attention and are prone to remain in viewer's memory in the area where moderate crossed parallax has been added.

In order to examine the impact of a dynamic binocular disparity on partially converted 3D images, an evaluation experiment was conducted that applied learning, distractor, and recognition tasks for 33 participants. The learning task involved memorizing the location of cells in a 5 ? 5 matrix pattern using two different colors. Two cells were positioned with alternating colors, and one of the white cells was moved up, down, left, or right by one cell width. Experimental conditions was set as a partially converted

Conference 9391:
Stereoscopic Displays and Applications XXVI

3D condition in which a white cell moved diagonally for a certain period of time with a dynamic binocular disparity added, a 3D condition in which binocular disparity was added to all white cells, and a 2D condition. The correct response rates for recognition of each task after the distractor task were compared. In the distractor task, participants were asked to follow the cross displayed in the left, center, or right of the screen with their eyes, interfering with the short-term memory related to the learning task. In the recognition task, a game pad for responding was placed at the hands of the participants. The participants were asked to press the right button if they judged that the pattern presented for recognition was the same as the one included in the learning task and to press the left button if they judged that it was not. The results of Experiment 2 showed that the correct response rate in the partial 3D condition was significantly higher with the recognition task than in the other conditions.

Conference 9392: The Engineering Reality of Virtual Reality 2015

Monday - Tuesday 9–10 February 2015

Part of Proceedings of SPIE Vol. 9392 The Engineering Reality of Virtual Reality 2015

9392-1, Session 1

Game-day football visualization experience on dissimilar virtual reality platforms

Vijay K. Kalivarapu, Anastacia MacAllister, Anthony Civitate, Melynda T. Hoover, Iowa State Univ. (United States); Phillip Thompkins, Jesse Smith, Univ. of Maryland, Baltimore County (United States); Janae Hoyle, Tufts Univ. (United States); Eliot Winer, Shubang Sridhar, Jonathan Schlueter, Gerrit Chernoff, Iowa State Univ. (United States); James H Oliver, Iowa State University (United States)

American college football season typically only lasts three months. However, most recruiting efforts happen off-season. College athletics departments, hence, do not always have opportunities to bring recruits and demonstrate first-hand the football home game experience. Recruits are traditionally offered physical visits to football facilities such as an empty stadium, and are shown video footage of past home games. Off-campus recruiting efforts have even fewer options since they lack the possibility of physical facility visits. These traditional routes may provide a recruit an overall idea of a football game experience, but it requires tremendous imaginative power to fully comprehend what it feels like to be cheered on by a crowd of 50,000 spectators in a stadium on game day.

The college football administration and the Virtual Reality Applications Center at Iowa State University teamed up to address the challenges above, and developed an immersive six-sided CAVE (C6) VR application for on-campus recruiting efforts, and a portable Head Mounted Display (HMD) application for off-campus recruiting efforts.

Research Question 1 (development): Typical video games come with features such as dynamic frustum culling, and non-essential elements in the environment can be rendered with a lower LOD. Unfortunately, immersive VR environments do not have such privileges and applications should scale reasonably and perform consistently across immersive displays as well as portable HMDs, while maintaining interactive frame rates. Given the application's focus on re-living a game-day experience, its implementation presents a formidable challenge especially with requiring detailed geometry/animation elements such as the stadium, the players, marching bands, audience, cheerleaders, etc.

Research Question 2 (validation):

User studies will be performed to explore the differences in immersion, presence, and general experience between the lower-cost HMD and a multi-million dollar C6. Should a peripheral like an HMD be comparable in immersion to the C6, using a lower-cost HMD can be justified to provide game day experience for football recruitment. Furthermore, this will provide the basis for testing similar trends in other VR areas and applications.

The application development was made using the game engine Unity 3D. A third-party plugin, GetReal3D from Mechdyne was used for clusterization. The application was deployed in the C6, a 24-projector 4K-resolution display system powered by a 48-node graphics cluster. Each node houses two Nvidia Quadro 6000 GPUs, one per eye. The portable HMD application was developed using the Oculus Rift. The VR application development is currently in its final stages of preparation for user studies. Upon approval from the Institutional Review Board (IRB), participants will be sought and randomly assigned a view mode (video, C6 or Oculus). Using a Witmer & Singer questionnaire, their immersion experience will then be measured. Conclusions will then be drawn on whether: a) VR environments offer advantages over traditional football recruiting methods, and b) a high-end multi-million dollar C6 offers immersion experience comparable to a low-cost HMD.

In the full paper, description of the methods used in application development and analysis of results from user studies will be discussed.

9392-2, Session 1

archAR: an archaeological augmented reality experience

Bridgette Wiley, Jürgen P. Schulze, Univ. of California, San Diego (United States)

We present an application for Android phones or tablets called "archAR" that uses augmented reality as an alternative, portable way of viewing archaeological information from UCSD's Levantine Archaeology Laboratory. archAR provides a unique experience of flying through an archaeological dig site in the Levantine area and exploring the over 1000 artifacts uncovered there. Using a Google Nexus tablet and Qualcomm's Vuforia API, we use an image target as a map and overlay a three?dimensional model of the dig site onto it, augmenting reality such that we are able to interact with the plotted artifacts. The user can physically move the Android device around the image target and see the dig site model from any perspective. The user can also move the device closer to the model in order to "zoom" into the view of a particular section of the model and its associated artifacts. This is especially useful, as the dig site model and the collection of artifacts are very detailed. The artifacts are plotted as points, colored by type. The user can touch the virtual points to trigger a popup information window that contains details of the artifact, such as photographs, material descriptions, and more. Clicking on "Open in Maps" uses the coordinates of the artifact to plot it in Google Maps. Clicking on "Open in Earth" plots it in Google Earth.

We wrote archAR for Android version 4.3. Development was done in C++ and Java, using the Java Native Interface (JNI) as a bridge between the two languages. archAR uses the augmented reality library Vuforia, which is written in C++. The dig site model was created in 3ds Max, based on ArcGIS maps created in the field. We exported it as an OBJ file and read it into the application session via a custom?written C++ function. For rendering, we use OpenGL ES 2.0. It is responsible for rendering the dig site model overlaid onto the image target. We have written custom GLSL shaders to support basic lighting of the dig site model and coloring of the artifact points.

To pick an artifact we use the Vuforia API. Our implementation projects the 3D coordinates of the artifact in virtual space onto the 2D screen of the Android device. Using the artifact's projected point, we find the closest artifact. The artifact with the closest distance to the screen point that the user touched becomes the selected artifact.

One concern in using augmented reality applications is user experience. The "rotate" mode in archAR allows the user to explore the dig site in an ergonomically improved way. Instead of having to bend over the image target and stress the back muscles, the image target can be hung on a wall and the 3D model projected in front of it. We rotate the 3D model 90 degrees to allow this to occur. We rather accidentally discovered that this mode also allows for much improved subterranean viewing of models that project below the ground, as this dig site does.

9392-3, Session 1

Photorealistic 3D omni-directional stereo simulator

Dirk Reiners, Carolina Cruz-Neira, Univ. of Arkansas at Little Rock (United States); Carsten Neumann, Univ. of Arkansas at Little Rock (United States) and Univ. of Arkansas at Little Rock (United States)

Conference 9392:
The Engineering Reality of Virtual Reality 2015

Most existing aircraft and vehicle simulators are based on jet simulators for flying at high speeds and high altitudes. These simulators do not address the training needs for aircrafts or vehicles that need to operate at much closer distances from objects in the environment. Furthermore, operators of these vehicles need to obtain a much higher amount of 3D situational awareness of their surroundings.

We present an innovative approach to enable aircraft and vehicle simulators with photorealistic visual quality and omnidirectional stereo in spherical environments without the need of any user-location technology. Our approach provides high quality rendering and true stereo separation at a pixel level throughout the entire image.

The innovative features of our image generator are: accurate 360-degree surround stereoscopic immersive display, and unprecedented visual quality at near real-time performance rates. We have incorporated near-by detailed object rendering for elements such as power lines, trees, and edges of small landing pads. Specifically, our research results include: a real-time ray tracer incorporating omnidirectional stereoscopic display, a set of rendering algorithms to display small-detail visuals, specifically power lines, and a set of rendering methods for real-time shadows and time of day. Our prototype shows that it is possible to adapt ray tracing algorithms to run at real-time interactive speeds incorporating 360-degrees 3D stereo to generate imagery that achieves a strong feeling of immersion, depth perception, and precision for operating in close proximity to objects like radio towers, other vehicles, or an urban landscape.

Our work enables the development of a new class of simulators for the close-range maneuvers needed for ground vehicles and helicopter flying with increased capabilities for training in situations and scenarios that are currently not possible in existing simulators due to limited 3D perception, surround view, and visual quality.

9392-4, Session 1

Composing a model of outer space through virtual experiences

Julieta C. Aguilera, Adler Planetarium & Astronomy Museum (United States)

This paper frames issues of trans-scalar perception in visualization, reflecting on the limits of the human senses, particularly those which are related to space, and showcases planetarium shows, presentations, and exhibit experiences of spatial immersion and interaction in real time.

9392-5, Session 1

How to avoid simulation sickness in virtual environments during user displacement

Andras Kemeny, Renault Technocentre (France) and Ecole Nationale Supérieure d'Arts et Métiers (France); Florent Colombet, Thomas Denoual, THEORIS (France)

Driving simulation and Virtual Reality (VR) share the same technologies for visualization, head movement tracking and 3D vision as well as similar difficulties when rendering the displacements of the observer in virtual environments, especially when these displacements are carried out using driver commands, including steering wheels, joysticks and nomad devices. High values for transport delay, the time lag between the action and the corresponding rendering cues or visual-vestibular conflict, due to the discrepancies perceived by the human visual and vestibular systems when driving or displacing using a control device, induces the so-called simulation sickness.

While the visual transport delay can be efficiently reduced using high frequency frame rate, the visual-vestibular conflict is inherent to VR, when not using motion platforms. In order to study the impact of displacements on simulation sickness, we have tested various driving scenarios in Renault's 5-sided ultra-high resolution CAVE. First results indicate that low speed displacements without longitudinal and lateral accelerations are well accepted and a worst case scenario is corresponding to rotational displacements in well detailed graphical environments. These results will be used for optimization technics at Arts et Metiers ParisTech for motion sickness reduction in virtual environments for industrial, research, educational or gaming applications.

9392-6, Session 2

Development of simulation interfaces for evaluation task with the use of physiological data and virtual reality applied to a vehicle simulator

Mateus R. Miranda, Diana G. Domingues, Alessandro Oliveira, Cristiano J. Miosso, Carla Silva Rocha Aguiar, Thiago Bernardes, Henrik Costa, Luiz Oliveira, Univ. de Brasília (Brazil); Alberto C.G.C. Diniz, Universidade de Brasília (Brazil)

This paper, with application in modeling simulation games and collecting experimental data, aims the description of an experimental platform for evaluating immersive games. The platform proposed in this paper is embedded in an immersive environment, in a CAVE of Virtual Reality and consists of a base frame with actuators with three degrees of freedom, sensor array interface and physiological sensors. Physiological data of breathing, galvanic skin resistance (GSR) and pressure in the hand of the driver and a subjective questionnaire were collected during the experiments.

This work includes presenting the theoretical background used in a project focused on Engineering Software, Biomedical Engineering and Creative Technologies.

The case study involves the evaluation of a vehicular simulator. As the integration of simulation software with immersion system interferes directly with the actions of the driver, the evaluation was performed by correlation between the analysis of their physiological data obtained before, in a period of rest and during the simulation with and without moviments at the simulator; also by the use of images captured through time at simulation and data collected by the subjective questionnaire.

9392-7, Session 2

An indoor augmented reality mobile application for simulation of building evacuation

Sharad Sharma, Bowie State Univ (United States); Shanmukha Jerripothula, Bowie State Univ. (United States)

Augmented Reality enables people to remain connected with the physical environment they are in, and invites them to look at the world from new and alternative perspectives. There has been an increasing interest in emergency evacuation applications for mobile devices. Nearly all the smart phones these days are Wi-Fi and GPS enabled. In this paper, we propose a novel emergency evacuation system that will help people to safely evacuate a building in case of an emergency situation. It will further enhance knowledge and understanding of where the exits are in the building and safety evacuation procedures. . It is a fast and robust marker detection technique inspired by the use of ARToolkit. It was tested using matrix based markers used by standard mobile camera. We show how the application is able to display a 3D model of the building using markers and web camera. The system gives a visual representation of a building in 3D space, allowing people to see where exits are in the building through the use of a smart phone or laptop. Pilot studies are being conducted with the system showing its partial success and demonstrate the effectiveness of the application in emergency evacuation. Our computer vision methods give good results when the markers are close to the user, but accuracy decreases the further the cards are from the camera.

9392-8, Session 2

Programmable immersive peripheral environmental system (PIPES): a prototype control system for environmental feedback devices

Chauncey E. Frend, Michael J. Boyles, Indiana Univ. (United States)

Improved virtual environment (VE) design requires new tools and techniques that enhance user presence. Despite being relatively sparsely studied and implemented, the employment of environmental devices (e.g. those that provide wind, warmth, or vibration) within the context of virtual reality (VR) provides increased presence. Many previously created peripheral environmental devices (PEDs) are not sufficiently prescriptive for the research or development community and suffer from a steep development or software integration learning curve. In this paper, we introduce a peripheral environmental device (PED) control system, called the "PIPE System". This work extends unpublished, rough prototype versions of the same system. The PIPE achieves enhanced user presence, while also lowering the barrier of entry for engineers and designers. The system is low cost, requires little hardware setup time, and promotes easy programming and integration into existing VEs using the Unity development engine.

Summarily, the PIPE System consists of a number of computationally-controlled electronic gates that interface between the VE and PEDs. It has three major components: (1) the support requirements for PEDs, (2) behavior control of PEDs, and (3) software level event sensing. Like its predecessors as well as the PEDs, the PIPE relies on common wall socket voltage for operation. Virtual effects (e.g. wind zones, warmth zones, and vibration events) are sensed in software, communicated from the VE to the PIPE, and then rendered from the connected PEDs. The PIPE is arranged in a multi-channel structure where each channel controls a single PED. This design offers scalability in adapting PIPE channels to existing VR systems. For example, a large-format CAVE VR system may require more channels than a smaller HMD VR system. Each channel makes use of an ATMega328 microcontroller that interfaces with an MOC3020 opto-isolator and a BTA 16-600B triac component topology. Communication from the VE dictates to each channel microcontroller appropriate behaviors for the electronic gates. Full implementation details are provided including circuit schematics and code implementations. The software mechanism for supporting real-time communication between the VE and the PIPE is discussed.

The latest prototype of the PIPE has been tested with household fans and electric heaters. Testing with other heat sources and vibration motors is currently underway. Empirical evaluation of the PIPE was focused on the ease of development using Unity. An established Unity VE known as "Tuscany" and a custom, recently developed driving simulation VE known as "Car Tutorial" have been enhanced using the PIPE. Both enhanced VEs were tested using an Oculus Rift HMD and a large-format reconfigurable CAVE. The PIPE is still being developed and will continue to improve over the months ahead. Immediate next steps include: adding support for wireless communication between VR systems and the PIPE, testing with additional developers using their existing VEs, and constructing additional PEDs. VR systems often strive to present VEs that resemble the real world, but developers are limited by VR systems that cannot simulate environmental conditions. The PIPE better equips developers to use existing VR systems with a broader range of environmental effects.

9392-9, Session 2

Explorations in dual-view, co-located VR

Silvia P. Ruzanka, Benjamin C. Chang, Rensselaer Polytechnic Institute (United States)

One of the major features of projection-based VR is that it allows multiple users to share both the same virtual space and the same physical space. However, the use of user-centered stereoscopy means that only one user actually has an accurate view of the scene, and in general only one user at a time can interact. Using modified polarized projection, we developed a system for two co-located users with independent views and interaction, in a monoscopic view using head-tracking and multi-screen panoramic projection. Giving users the ability to actively collaborate or explore different spaces simultaneously opens up new possibilities for VE's. We present prototype interaction designs, game designs, and experimental artworks based on this paradigm, pointing towards future developments in VR toward physically co-located designs that allow for co-presence, fostering innovative collaborations.

9392-10, Session 2

From CAVEWoman to VR diva: breaking the mold

Carolina Cruz-Neira, Univ. of Arkansas at Little Rock (United States)

One of the main ground-breaking developments in the history of virtual reality (VR) was the creation of the CAVE virtual reality system. It was first introduced in the early 90s and it is still one of the key technologies that define the field. What it is not so well known in circles outside the core research communities is that this technology was primarily conceived, designed, implemented and put to work outside the research labs by a woman, the author of this paper. After the development of the CAVE, her work expanded to spread the use of VR technology into a wide range of disciplines, ranging from deeply scientific and engineering areas to rigorous humanities applications, to creative art experiences. Being a woman, she brings a pragmatic perspective on what VR is, how the supporting tools and technologies need to be designed to simplify its use, and to enable unexpected groups to explore VR technology as a new medium. This paper presents a set of truly interdisciplinary VR projects that were made possible by having a strong technical expertise rooted in a pragmatic feminine interpretation of the technology and its capabilities. Examples of these projects are: turning a low-power wireless embedded software scientific project into a dance performance, using field work of religious studies about religious rituals of 15th Century India as a testbed for a distributed computing software architecture, and extending a scripting language framework to support storytelling to document the sad events of 9/11. The paper also discussed the successes and struggles to gain acceptance and credibility as a VR researcher with this unconventional approach to the technology.

9392-11, Session 3

The use of virtual reality to reimagine two-dimensional representations of three-dimensional spaces

Elaine Fath, Indiana University Bloomington (United States)

A familiar realm in the world of two-dimensional art is the craft of taking a flat canvas and creating, through color, size, and perspective, the illusion of a three-dimensional space. Using well-explored tricks of logic and sight, impossible landscapes such as those by surrealists de Chirico or Salvador Dalí seem to be windows into new and incredible spaces which appear to be simultaneously feasible and utterly nonsensical. As real-time 3D imaging becomes increasingly prevalent as an artistic medium, this process takes on an additional layer of depth: no longer is two-dimensional space restricted to strategies of light, color, line and geometry to create the impression of a three-dimensional space. A digital interactive environment is a space laid out in three dimensions, allowing the user to explore impossible environments in a way that feels very real. In this project, surrealist two-dimensional art was researched and reimagined: what would stepping into a DeChirico or a Magritte look and feel like, if the depth and distance created by light and geometry were not simply single-perspective illusions, but fully formed and explorable spaces? 3D environment-building

software is allowing us to step into these impossible spaces in ways that 2D representations leave us yearning for. This art project explores what we gain--and what gets left behind--when these impossible spaces become doors, rather than windows. Using sketching, Maya 3D rendering software, and the Unity Engine, the 1920's surrealist art movement was reimagined as a fully navigable real-time digital environment. The surrealist movement and its key artists were researched for their use of color, geometry, texture, and space and how these elements contributed to their work as a whole, which often conveys feelings of unexpectedness or uneasiness. The end goal was to preserve these feelings while allowing the viewer to actively engage with the space.

9392-12, Session 3

Theory review and interaction design space of body image and body schema (BIBS) for embodied cognition in virtual reality

Xin Tong, Diane Gromala, Simon Fraser Univ. (Canada); Owen Williamson, Monash Univ. (Australia); Christopher D. Shaw, Ozgun E. Iscen, Simon Fraser Univ. (Canada)

In this paper, we introduce six fundamental ideas in designing interactions in VR that are derived from BIBS literature, demonstrating a specific idea of embodied cognition. We discuss how this theory motivates our VR research projects, from mature to an evolving prototype for girls who undergo Scoliosis surgery. This discussion supports explorations in BIBS theories and how they influence an approach of VR interaction design.

Six ideas about embodied cognition that are supported by the current BIBS literature were explored and guided our VR research:

1. Mind is embodied. We think with our bodies rather than solely relying on our brains, even though we are not always -- or often -- consciously aware of it.

2. Proprioception plays an important role in embodiment and consciousness.

3 The sense of self and perception of others are strongly informed by embodiment.

4. Plasticity is involved in both body image and body schema; they are interdependent systems, not exclusive categories.

5. Such operations of BIBS do not become apparent to conscious awareness until there is a reflection on our bodily situations brought upon by certain limit-situations such as pain.

6. As body image and body schema are also shaped by pre-reflexive processes, it is difficult to assess or evaluate BIBS-related issues based solely on reflective methods, such as interviews or surveys. Therefore, experimental situations need to be created that address both the human subject's both reflective (how they express [verbalize] how they feel while moving) and pre-reflective/proprioceptive processes (how their body actually moves).

Based on our experiences, we argue that incorporating those ideas into design practices requires a shift in the perspective or understanding of the human body, perception and experiences of it, all of which affect interaction design in unique ways. The dynamic, interactive and distributed understanding of cognition, where the interrelatedness and plasticity of BIBS play a crucial role, guides our approach to interaction design.

For instance, in our most recent research, we work with teenage girls, before, during and after surgery for treating scoliosis, in order to help them to transform their distorted body image and restore its balance with their body schema. Teenage girls, who arguably care more about their body schema (the collection of process that registers the posture of the body's spatial properties) and are in the important stage of formalizing self-recognition and self-awareness, usually suffer more from both physical pain and psychological pressure of their body image (conscious ideas of the aesthetics and sexual attractiveness about their own body) than other patients with similar diseases. Therefore, we outlined our VR design based on BIBS ideas before actually developing virtual environment. Though we have not fully developed this VR or tested it with our users, this case shows how BIBS can be incorporated into the VR interaction design stage. BIBS literature not only facilitates developing expressive VR designs, but also provides these designs with a perspective through which we can explore the experience of the human body in multiple dimensions. Both the literature and VR design developed in this paper is based on six BIBS ideas which emphasize interaction design and embodied cognition in VR.

9392-13, Session 3

Embodied information behavior, mixed reality systems, and big data

Ruth G. West, Univ. of North Texas (United States)

As information overwhelm becomes a thing of the past, big data becomes a standard for every aspect of human endeavor spanning the arts, sciences and commerce. In parallel, computation has transitioned from specialized systems to personal computing to ubiquitous and persistent computing embedded in tools, wearables and dedicated systems. The question now becomes not so much how to deal with the bigness of big data, but how to design experiences tailored to the human sensorium and embodiment to enable us to reap the benefits of bid data. This panel and review paper will address current and future approaches and design principles for the application of mixed reality and immersive experiences, and embodied information behavior in relation to big data.

9392-14, Session 3

GoTime: a storytelling platform for SAGE2

Todd Margolis, Univ. of California, San Diego (United States)

Traditionally large format display systems are used (outside of cinema) for interactively displaying multimedia content such as images, videos and computer graphics. Although the means by which this media is rendered is varied, the data is generally procedurally shown on demand as users navigate through a system. This serves us very well for applications that require an analytical interface for knowledge discovery. However, there has been an unfulfilled need for scripting the automated playback of content for presentation purposes. In this paper, I will describe an application designed to satisfy the need for composing presentations for large format tiled displays. I will explain the design criteria, supported media types, typical usage scenarios as well how this platform expands upon traditional presentation software to take advantage of the unique capabilities of large format display systems. GoTime is intended for supporting immersive storytelling through traditional slideshows as well as non-linear techniques that empower presenters with the flexibly to pause presentations in order to interactively delve further into ideas in response to audience feedback.

9392-15, Session 4

System for augmented reality authoring (SARA): a new way of authoring augmented reality systems

Bhaskar Bhattacharya, Eliot Winer, Iowa State Univ. (United States)

CONTEXT

One of the biggest problems with Augmented Reality (AR) today is that authoring an AR based system requires a lot of effort and time by a person or a team of people with a wide variety of skills. Consider the case of assembly guidance by AR where the system aids in building a product. Registering each individual virtual object (image/video/3D model) with the real environment and providing correct feedback to the end-user are

just some of the difficult operations that must be performed. If multiple "steps" in a process need to be augmented, each individual step needs to be authored. In addition, once a process is completed, it may require adjustment due to future changes. These need to be reflected in the AR system as quickly as possible. If extensive effort and time is involved to author these steps, this process becomes unmanageable. Thus, we are in the situation we currently see. That AR is more of a gimmick with systems taking months or longer to deploy. There is thus a need for an authoring tool that can quickly generate the required AR output in real or near real-time.

OBJECTIVE

The intent of this research is to allow users with low levels of technical knowledge a tool by which they can create AR Guided Assembly Systems through their demonstrations of the assembly process.

METHOD

Imitation is a recognized form of learning/teaching. With this in mind a novel method is proposed using a depth camera to observe an expert performing the assembly task to be augmented. It is assumed that each physical part of the assembly has a corresponding 3D model available. For engineered systems, these models are almost always created and available. Using developed computer vision and image processing techniques, the expert's hand(s) and assembly parts are tracked providing information of the location and form of the assembly. Initial augmenting elements are animated arrows for attention direction and animated 3D models for correct orientation and positioning of each part. In this way, each individual step is understood and the necessary augmentation is provided.

RESULT

A prototype version of SARA has been developed and examples of the tool will be provided such as an example product will be developed with three different approaches and analyzed by SARA. For each approach a different AR system will be built with no changes required in SARA.

CONCLUSION

A gap in research has been identified pertaining to AR authoring. A novel method has been developed to solve this problem using imitation principles as a basis for learning. This allows users with limited technical knowledge to demonstrate assembly tasks and create an AR system with limited input.

FUTURE WORK

With the speed and ease with which assembly tasks are authored, research and industry can now begin to compare through user studies the different methodologies of product assembly. Since SARA captures hand pose for each assembly task, intelligent path planning can be provided to minimize delay between steps.

9392-16, Session 4

Free-body gesture tracking and augmented reality improvisation for floor and aerial dance

Tammuz Dubnov, Cheng-i Wang, Univ. of California, San Diego (United States)

This paper describes recent developments of an augmented reality system for floor and aerial dance that uses a depth-sensing camera (MS Kinect) for tracking dancers' movement to generate visual and sonic elements of the show. Dealing with free movement in space, the tracking of performer's body could not be done reliably using skeleton information. In an earlier system we employed a single infra-red (IR) marker placed on the dancer's hand or leg, with a Hidden Markov Model (HMM) providing recognition of gestures and their tracking in time (time-warping). Although the system worked well in detecting and tracking quite complex movements in frontal recording conditions, the performance was significantly degraded in cases of marker disappearance caused by turns, spins or occlusion by aerial apparatuses. Another limiting factor was that the system required manual switching between scenes to indicate when to initiate a detection of a gesture within a long performance sequence. It is desirable for the system to be able to identify gestures in a single long recording of a complete dance, without breaking it up into individual scenes.

In the current paper we introduce several improvements to the previous system: We provide a more robust model that uses multiple IR markers with a predictive model that is able to overcome problems of occlusion (marker disappearance and reappearance), as well as crossing trajectories and other disturbances. In order to deal with rotation issues, we included an additional feature for tracking spins (in which markers disappear and reappear rapidly) by integrating a small motion-sensing accelerometer to track body orientation. To deal with gesture segmentation, we introduce a new model called Variable Markov Oracle (VMO). This model represents gestures in terms of a graph of repeated sub-segments within a long sequence, such as features derived from a complete dance. VMO also has additional advantages, such as direct mapping of the performance timeline to detected movement primitives (gestural subsequences) that are found in the long choreography sequence. Detection of gesture is done by finding a combination of sub-segments in a recording that best matches a live input.

We also outline the steps and challenges that arise when putting on a production that uses this system during the stages of choreography, rehearsals and the final performance. By using the new model for the gesture we can unify an entire choreography piece, and potentially an entire show, with the model dynamically tracking and recognizing gestures and segments of the piece. This gives the performer also additional freedom to improvise and have the system respond autonomously in relation to the performer within the framework of the rehearsed choreography, an ability that goes beyond queued projection-design and other interactive systems known to date. In overall, these improvements allow more freedom of movement and help dramatically increase the versatility and precision of the augmented virtual reality created on stage.

9392-17, Session 4

Marker-less AR system based on line segment feature

Yusuke Nakayama, Hideo Saito, Keio Univ. (Japan); Masayoshi Shimizu, Nobuyasu Yamaguchi, Fujitsu Labs., Ltd. (Japan)

In general marker-less AR system, the camera pose is estimated from feature correspondences between the 3D model and its 2D image. For obtaining the feature correspondences, the feature points matching is used. However, in the scene where few feature points matching are detected, the 2D-3D point correspondences cannot be obtained, this estimation of the camera pose will fail. In this case, we cannot perform AR, therefore scene feature is needed.

Line segments can be considered as an alternative scene feature to solve this problem. A lot of line segments are detected especially in man-made situation even where few feature points are detected. Therefore, a marker-less AR system which uses line segments feature is suitable to such man-made environment. However, comparing with points matching, finding matching of line segments is not easy task. Line segments have their own disadvantages, such as inaccurate locations of line endpoints, fragments of same line, less distinctive appearance of line segments, etc.

Therefore, the purpose of our research is dealing with these line segments problems and performing marker-less AR based on line segments for the situation where feature points matching cannot be obtained.

In our proposed method, line segment matching is used instead of feature point matching. We adopted a fast line feature descriptor, Line-based Eight-directional Histogram Feature (LEHF) for line segment matching. LEHF computes differential value by taking a constant number of points around the line segment and makes eight-directional gradient histograms to describe the line segment. Therefore, LEHF provides line segment matching even if line segments have less distinctive appearance.

Moreover, we represent line segment as their directional vector and one point on the line, that is, we do not need information of line endpoints. Thanks to this representation, there is no problem if the line endpoints are in inaccurate location or the line segments are divided into some parts.

Here is the overview of our proposed method. First of all, a 3D line segment database is constructed as a 3D model. This database is generated from

multiple images of the target scene taken from RGB-D camera. It contains positions of 3D line segments and kd-trees of LEHF features from multiple angle. Next, 2D line segments from an input image are detected and LEHF features are extracted. Nearest-neighbor matching between these LEHF features and the database is performed and the 2D-3D line segment correspondences are obtained. Finally, from this correspondences, the camera pose of the input image is estimated by RPnL which is a method for solving the Perspective-n-Lines problem. RPnL needs no information of line endpoints. Then, we can overlap CG onto the camera image using the estimated camera pose. We have experimentally demonstrated the performance of the proposed method by comparing the proposed method with methods based on point features as landmark. The experimental result shows that our proposed method using line segment matching can estimate the camera pose and perform AR even in the situation which has only a few feature points.

9392-18, Session 4

On the usefulness of the concept of presence in virtual reality applications

Daniel R. Mestre, Aix-Marseille Univ. (France)

Virtual Reality (VR) leads to realistic experimental situations, while enabling researchers to have deterministic control on these situations, and to precisely measure participants' behavior. However, because more realistic and complex situations can be implemented, important questions arise, concerning the validity and representativeness of the observed behavior, with reference to a real situation. One example is the investigation of a critical (virtually dangerous) situation, in which the participant knows that no actual threat is present in the simulated situation, and might thus exhibit a behavioral response that is far from reality. This poses serious problems, for instance in training situations, in terms of transfer of learning to a real situation. Facing this difficult question, it seems necessary to study the relationships between three factors: immersion (physical realism), presence (psychological realism) and behavior.

Immersion refers to the sensorimotor coupling between a participant and a virtual environment, including sensorial vividness and real-time interaction capacities of the VR setup. Immersion is thus described as the quantifiable, physical, aspects of the simulation, and can be qualified as the potentialities of the VR setup to isolate the participant from the real world. A large number of studies have investigated immersive determinants of performance (e.g. latency of the real-time loop, display resolution...). The general hypothesis is that immersive properties of the virtual environments, by isolating the participant from stimulation emanating from the real environment, and by replacing them with stimulations from the virtual environment, will promote optimal behavioral control within the virtual environment. However, if immersion is a necessary condition for behavioral performance in a virtual environment, it is not a sufficient condition for the expression of a behavior that is representative of the actual behavior in real conditions. Confronted with this problem, in areas of simulation and teleoperation, from early stages of virtual reality research, a concept was introduced, that seems to address the question of the "ecological validity" of behaviors observed in VR setups. This is the concept of "presence", traditionally described as the feeling of "being there" in the virtual world. In other words, presence is related to the fact that the participant feels "concerned" by what is happening in the virtual scenario. Presence thus refers to a psychological, attentional and cognitive state, in which the participant, immersed within a virtual environment, behaves in accordance with the affordances provided by this environment, as if what is happening in the virtual environment was real.

In this framework, presence appears as a consequence and a condition complementary to immersion, necessary for the relative validity of virtual environments. Presence is supposed to enable participants to express representative behaviors, as compared to a real situation. We will present here results from recent experimental studies, focusing on the ecological validity of (the behavior observed during) virtual reality exposure. The general hypothesis is that 1) contextual and psychological factors will influence the feeling of presence (from low-level –physiological- to high-level –cognitive- levels of behavioral responses to virtual events) and 2) behavioral presence will result in representative behavior, with reference to a real situation.

9392-23, Session 4

Bringing scientific visualization to the virtual reality environments using VTK and VRUI

William R. Sherman, Indiana Univ. (United States)

The Visualization Toolkit (VTK) is one of the most commonly used libraries for the visualization and computing in the scientific community. The VTK provides classical and model visualization algorithms to visualize structured, unstructured and point datasets on desktop, mobile, and web environment.

Immersive interfaces (virtual reality) have been demonstrated to have positive impact on the ability to explore and connect with data over the use of standard desktop displays.

We combine the strength of VTK with immersive interfaces through an extention to VTK that enables immersive software libraries such as Vrui and FreeVR to quickly take advantage of the plethora of features within VTK.

Thus, The objective of this work is bring high-quality scientific visualization computing and rendering capabilities to the virtual reality environments in a way that's easier to maintain and develop by the developers. By bringing the VTK into the virtual environment created by the domain specific tools such as VRUI and FreeVR, we are providing tools necessary to build interactive, 3D scientific visualizations to the developers.

9392-19, Session PTues

Building the metaverse

Ben Fineman, Internet2 (United States)

Advances in immersive head mounted displays are driving us toward a liminal point in terms of adoption and utilization of virtual environments. These trends will have profound implications for the fundamental nature how we share information and collaborate. We are poised to begin the creation of an open, interoperable, standards-based "Metaverse" of linked virtual environments that leverage these technologies. Standards exist today, but many unsolved problems remain in areas including security, identity, and distributed architecture. This session will discuss current challenges facing the creation of the Metaverse, as well as an overview of efforts working on solutions.

To date, information sharing on the Internet has primarily been done via the World Wide Web in formats that mimic printed documents – that is, web pages that are two-dimensional and scroll vertically. Real-time collaboration technologies have enabled richer collaboration through the use of high quality video and audio, but these technologies have also been limited to two-dimensional displays. Three-dimensional displays will enable information sharing and collaboration opportunities that we have not yet imagined today. In 2008, the National Academy of Engineering (NAE) identified virtual reality as one of 14 Grand Challenges awaiting solutions in the 21st century.

In time, many of the Internet activities we now associate with the 2D Web will migrate to the 3D spaces of the Metaverse. This does not mean all or even most of our web pages will become 3D, or even that we'll typically read web content in 3D spaces. It means that as new tools develop, we'll be able to intelligently mesh 2D and 3D to gain the unique advantages of each, in the appropriate context . The research and education community is ideally suited to explore the opportunities enabled by these new capabilities.

Current State of the Technology:

Environment:

Real-time three-dimensional rendering has reached unparalleled levels of realism due to increases in processing power. Examples of realistic 3D environments are now widely available, primarily for entertainment applications, but increasingly for educational and training simulation.

Conference 9392:
The Engineering Reality of Virtual Reality 2015

Interface:

Many aspects contribute to the realism of a user's interface to virtual environments, including sight, sound, and eventually all of the senses. High quality surround sound reproduction has been long understood, but until recently immersive three-dimensional visual reproduction has been unavailable outside of expensive dedicated studio environments. Recent advances in head mounted, stereoscopic, motion-tracking displays are the primary driver for impending popularization of virtual worlds. The most notable example is the Oculus Rift, recently acquired by Facebook for $2B after raising $91M as a startup, and whose technology is currently available for purchase in beta form.

Avatar:

Users will be represented in virtual environments via avatars. Historically these avatars have been relatively unsophisticated and not suitable for immersive collaboration. Advances in computational power and motion sensing input devices have enabled realistic looking avatars that track movements and facial expressions without cumbersome peripherals.

Architecture:

To date, all widely deployed virtual environments have leveraged centralized servers. This is not scalable for a Metaverse scenario – it would be akin to hosting every website on the same cluster of servers. Instead, a distributed peer-to-peer architecture is required. Models of these kinds of architectures have been established but not deployed in the wild.

Standards and Interoperability:

Many standards have already been established, and many mirror the architecture of the World Wide Web. VRML exists as the standard for 3D spaces as HTML does for web pages. Open Cobalt exists as a protocol standard upon which diverse servers and clients can base interactions. VWURLs exist as locators for virtual world locations like URLs for websites. Identity standards are required but lacking, and may be addressed by Shibboleth. Despite the diverse standards that exist, production implementations leveraging these standards are still lacking.

Efforts towards solutions:

The advanced networking consortium Internet2 is facilitating the Metaverse Working Group, focused on implementing standards-based and interoperable open virtual environments.

9392-20, Session PTues

A passage for transmutation and transition

Hyejin Kang, Indiana Univ. (United States)

An interest of the origin and the energy of life has allowed me to make artwork continually in painting and other 2D media. I tried to develop unique organic imagery and wanted to expand more possibility in artistic expression continually.

In Unity which enables people to navigate in 3D space, I can realize to create more powerful experience about the topic beyond visual expression in painting. The new technology, interactive 3D navigating in Unity expands the possibilities of expression and experience in artwork.

In Unity, the key of virtual physical experiences is navigating. Unity provides diverse physical movements. I focus on having a sense of 'transition' and 'falling down'. In this work, there are two different worlds and those are connected in a specific spot. Through transporting from one world to another, I'd like to provide a sense of 'transition' between two worlds for audience. In each world, for navigating, users fall down from the top to the bottom which is the surface of water. I'd like to deliver the beauty of abstract painting and decorative aesthetic in 3D space. The concept of navigating this work is 'moving painting'. Based on visual imagery in my painting, the 3D modelings are designed and the color palette is chosen. For textures, I choose subtle change of colors and light effects. Repeating theme objects in 3D space creates another way of visualizing my artistic theme from my paintings. While navigating, users can appreciate visual aesthetic of 3D object and environment in different perspectives and angles.

I have a transcendental desire to go forward further beyond all of limitations within my existence even though I'm a human being living specific time and space. I created two different worlds; one indicates a world of the primordial state of life and the other is about a pure abstract world to show the natural rule of cosmos. Through a journey wandering the worlds and passing over from a world to another as a metaphor of living, I'm eager to show a 'transition' which enables to reach ultimate transcendence.

In the first scene, I'd like to create a unique world where the first life showed up. I guess lots of energies, passions and coincidences interwound together for the birth of life. I'm curious about the beginning time. I imagined the moment. I'd like to visualize my imagination, and share the atmosphere and the energy of the moment with audience in an immersive media. To express the environment of the beginning, I make a space which is similar to inner body.

In the second scene, I make difference between the first and the second, so I choose different background colors and textures with the first scene's one. Also, the modelings are more abstract than the first one.

Through this work, I strongly hope that audience can experience fully something compelling in their body. I believe I can deliver vivid experience to audience through interactive 3D navigation work.

In presentation, I will explain how I developed and improved my artwork in Unity from painting and time based media so that we can discover new potential in Unity and discuss ways of using Unity in creativity.

9392-21, Session PTues

A framework of augmented reality for a geotagged video

Kyoung Ho Choi, Mokpo National Univ. (Korea, Republic of)

In this paper, we present a framework for the implementation of augmented reality services for a geotagged video. In the proposed framework, a live video is captured by a mobile phone and the corresponding geo-coordinates and angular parameters are recorded at the same time. Based on the parameters, the information about buildings around the center position of the mobile phone is obtained from a database called V-World (http://map.vworld.kr) that is an open platform built by a Korean government and freely available through the online. More specifically, MPEG-7 is used for the description of camera parameters for a captured video and a detailed description about a chosen building, e.g., the name and the height of the building, is generated and overlaid on the building. To determine the building selected by a user among buildings in the surrounding area, a segmentation process is performed and the height of extracted buildings is measured and compared with buildings in the database. To distinguish buildings located closely each other, a probabilistic network approach is adopted in the proposed framework. The proposed approach is based on a database freely available through the online and a new database is not required to be constructed for the service, which is very essential to build augmented reality services. Experimental results show that the proposed framework can be used to build augmented reality applications for a geotagged video.

9392-22, Session PTues

The application of virtual reality in medical study

Shan Yu, Indiana Univ. (United States)

Presently, with the adoption and development of multi-sensory, as well as, user interfaces, they are manipulated into diverse fields, not only some high-tech fields but also are committed to entertainment devices. Taking advantage of such devices, we are going to present an interactive physical piano embedded with multi-sensory, which are built with LEGO building blocks and works for users sitting in the front of desks or facing computers for a long time to help them take exercises with the whole bodies, accordingly to the kinematics of the human body theories. To prove the effectiveness of our thoughts, we built a demo and after the first round test, as well as, surveys, it got primary victory.

9392-24, Session PTues

Reduce blurring and distortion in a projection type virtual image display using integrated small optics

Tatsuya Hasegawa, Tomohiro Yendo, Nagaoka Univ. of Technology (Japan)

Head Up Display (HUD) is being applied to automobile. HUD displays information as far virtual image on the windshield. For example, speed meter, tachometer and so on. Therefore driver can get the information without refocus between scenery on the road and some meters. Existing HUD usually displays planar information such as running speed of automobile. If the image corresponding to scenery on the road like Augmented Reality (AR) is displayed on the HUD, driver can efficiently get the information. For example in routing assistance by car navigation system, guide sign should be displayed on the intersection which should be turn next. To actualize this, HUD covering large viewing field is needed. However existing HUD cannot cover large viewing field. In the conventional ways to generate virtual images that uses single axis optics for whole viewing field, length of light path would be long due to limitation of F number of the optics. Thus the display system must be large and it is difficult to mount on the automobile. Therefore to realize thin HUD covering large viewing field, we have proposed system consisting of projector and many small diameter convex lenses. Base principle is using a convex lens array and elemental images similarly to Integral Photography (IP). Each lens generates virtual image in front by corresponding elemental image. One lens of IP is seen as just one pixel; in contrast, one lens of the proposed method shows a part of virtual image. Observing through lens array, one virtual image can be seen by gathering these virtual images. By displaying virtual image using this principle, system becomes compact because each lens of lens array has short focal length. If elemental images are displayed directly by a two-dimensional display device, resolution of the virtual image is low because the number of the pixel per an elemental image is a few. Therefore to keep resolution of virtual image; in this system elemental images are generated by optical replicate projected image from projector. However observed virtual image has blurring and distortion by this way. In this paper, we propose two methods to reduce blurring and distortion of images. First, to reduce blurring of images, distance between each of screen and lens comprised in lens array is adjusted. We inferred from the more distant the lens from center of the array is more blurred that the cause of blurring is curvature of field of lens in the array. Thus each position of screen is moved to position that minimizes spot diameter of light focusing. Second, to avoid distortion of images, each lens in the array is curved spherically. We inferred from the more distant the lens from center of the array is more distorted that the cause of distortion is incident angle of ray. Thus to minify incident angle of ray, each lens of the array is curved spherically in such a way as to cross optical axis of all lens at view point. We designed convex lens array and confirmed effectiveness of both methods.

Conference 9393: Three-Dimensional Image Processing, Measurement (3DIPM), and Applications 2015

Tuesday - Thursday 10–12 February 2015

Part of Proceedings of SPIE Vol. 9393 Three-Dimensional Image Processing, Measurement (3DIPM), and Applications 2015

9393-20, Session PTues

Crosstalk characterization of PMD pixels using the spatial response function at subpixel level

Miguel Heredia Conde, Klaus Hartmann, Otmar Loffeld, Zess Univ. Siegen (Germany)

Time-of-Flight cameras have become one of the most widely-spread low-cost 3D-sensing devices. Most of them do not actually measure the time the light needs to hit an object and come back to the camera, but the difference of phase with respect to a reference signal. This requires special pixels with complex spatial structure, such as PMD pixels, able to sample the cross-correlation function between the incoming signal, reflected by the scene, and the reference signal. The complex structure, together with the presence of in-pixel electronics and the need for a compact readout circuitry for both pixel channels, suggests that systematic crosstalk effects will come up in this kind of devices. For the first time, we take profit of recent results on subpixel spatial responses of PMD pixel to detect and characterize crosstalk occurrences. Well-defined crosstalk patterns have been identified and quantitatively characterized through integration of the inter-pixel spatial response over each sensitive area. We cast the crosstalk problem into an image convolution and provide deconvolution kernels for cleaning PMD raw images from crosstalk. Experiments on real PMD raw images show that our results can be used to undo the low-pass filtering caused by crosstalk in high contrast image areas.

9393-21, Session PTues

Unified crosstalk measurement method for various distances on multi-view autostereoscopic displays

Bernd Duckstein, René de la Barré, Thomas Ebner, Roland Bartmann, Silvio Jurk, Ronny Netzbandt, Fraunhofer-Institut für Nachrichtentechnik Heinrich-Hertz-Institut (Germany)

What is the addressed scientific topic or problem?

In this paper a procedure for crosstalk (CT) measurements on space-multiplexed multi-user autostereoscopic 3D displays with so-called Viewing Distance Control (VDC) is presented. VDC makes use of a rendering method which allows shifting of the viewing distance for multi-view displays by using a novel distribution of content. Methods for CT measurements to date cannot be used as the measurements have to be performed at distances that are not defined in the standard procedures. The measuring procedures used so far are not applicable, as neither a measurement process nor any test images are defined for the use at different viewing distances.

As separate CT-measurement specifications for two-view and multi-view autostereoscopic displays already exist, the authors propose a unified measurement process. This process is supposed to utilize both, the equipment, as well as the physical arrangement of measuring subject and instrument that are used so far. It has to be considered that, due to the basic functional principles, several quality measurement and evaluation criteria for 3D displays have emerged. Different autostereoscopic display technologies lead to different measurement procedures. A unified method for analyzing image quality features in 3D displays, requiring no enhanced effort but offering comparable results, is desirable. Although the technical 3D display solutions are often quite diverse in their visual appeal, they all aim at the same objective: realistic, high quality and flawless reproduction of spatial imagery. So to improve the comparability could help the 3D community to evaluate results of research and production more efficiently. This could lead to a definition of display quality which as a result could help to better identify improvements or drawbacks in new developments or modifications.

What are the challenges and barriers faced?

First, the similarities of the existing procedures that, when used, allow to find a value for CT of the different display methods (two-view, multi-view) have to be identified. The already established methods imply the problem that different measurement procedures lead to hardly comparable results.

Second, there is a challenge in proving, that these similarities are valid for a wide variety of possible variants of displays.

Why this is important for the 3D community?

The 3D community is a grouping of researchers and developers working theoretical and practical on hardware topics like display technology and software tasks, with optical physics and psychological aspects. Since the standards of metrology are too diverse, it is hard, even for experts, to fairly and reliable benchmark the existing systems. Unified and simplified description methodologies are important to help all these different professional groups to find a common language.

What is the original method proposed to address this problem or issue?

As an example, the publication of the ICDM addressing a very wide field of 2D measurement procedures. In chapter 17 in particular methods to evaluate 3D displays, several measuring approaches for 3D and autostereoscopic 3D alone are presented. These are very thoroughly and accurately adapted and created procedures that help to evaluate autostereoscopic displays. But the approaches for the different technologies still make it hard for a user utilizing them to compare and evaluate the results received. The authors prove that it is sufficient and appropriate to dismiss a measurement condition and that a more homogeneous measurement procedure for several classes of displays can evolve.

What is the novelty comparing to the state of art?

The architecture of multi-view autostereoscopic displays assigns a nominal viewing distance (NVD). By using VDC, such displays can also be viewed from different distances. Though the conventional stereoscopic crosstalk evaluation method cannot be applied at these distances. State of the art measurement procedures consider autostereoscopic displays that are used and measured at a fixed distance. Therefore these standards offer no solution for a change of distance due to sub-pixel based view point adaption. There is no option to display the content of one single image at one specific position as it is the case when using standard image input, for example eight single images for an eight-view multi-view display. The analysis of a very small display area at several measuring distances leads to comparable results. The method can be used on multiple types of autostereoscopic screens and systems. The method is based on interlaced views represented by black and white test images. These test images generate nearly uniform crosstalk within the entire screen area when observed from the nominal viewing distance. This also applies for x-y-z tracked single-user autostereoscopic displays. However, in multi-view autostereoscopic displays a deviation of the viewing distance from the NVD results in a lack of multi-view channels. As a consequence, the crosstalk condition does not lead to viewing channels visible across the entire screen. Therefore, an alternative method for measuring display quality is required.

The authors present a modification of the common crosstalk measuring method by examining the geometrical structure of the light beams when a viewing distance control is applied. They show that a beam emitted orthogonally from the center of the screen is consistently visible—independent from the observing distance. In parallel, the light distribution shows almost similar relative functions of intensity. Deviations are thought to originate from beam expansion and beam shaping caused by the cylindrical lens element. Therefore, one intensity function characterizes the distance-dependent quality of the optical system at a particular location. This quality information can also be obtained for all laterally located lens elements, since their design proportions are similar.

The article clearly illustrates that sufficient results can be achieved, when a single view within the zone of one cylindrical lens element is used for the measurement. Furthermore, the authors show that the same view can be used in the immediate adjacent area to increase the desired signal. The proposed approach makes use of the fact that the visual property in a single channel can be investigated when the width of the observed measuring field is small enough. The measurement result characterizes the optical property of the optical system in a selected viewing distance.

The proposed crosstalk measurement method differs in several points from the IDMS recommendations. In particular, the designated eye position is not equal to the optimal viewing distance. Rather, the measurement distance can be selected arbitrarily. Hence, the crosstalk is described as a function of distance.

What is the efficiency of the method (presentation of results and comparison with the state of art)?

The intention of the authors was to find a way to measure and evaluate the developed VDC. The presented procedure allows to classify and quantify the quality of this technology. Early results show that, over a wide range of viewing distances, the measurable angular distribution of luminance is constant, taking into account the loss of intensity in relation to the measuring distance. When comparing the angular light distribution of luminance emission of a multi-view screen at different distances and with only about 1 percent at the center of the screen surface is active, the distribution and the luminance relation is almost identical (Figure 1).

9393-22, Session PTues

Registration between point clouds and image data based on moment invariant features

Liu Sheng, Chang'an Univ. (China)

3D laser scanning is an advanced technology as an important tool to obtain the spatial data, which has been developed in recent years.The method of the registration between point clouds and image data belongs to the fields of 3D image acquisition and generation techniques.

9393-23, Session PTues

An evaluation method of 3D road slope deformation trend based on image analysis and three-dimensional laser scanning technology

Zhiwei Wang, Chang'an Univ. (China)

In recent years, the use of terrestrial laser scanning technology in engineering surveys is gaining an increasing interest due to the advantages of non-contact, rapidity, high accuracy and large-scale. This technique delivers millions of accurate 3D points (mm-level accuracy) with a very high point density in a short time frame (up to 1 million points per second), which makes it a valuable alternative or complementary technique for classical topographical measurements based on Total Station or digital photogrammetry. The terrestrial laser scanning can still delivers very accurate points even in the situations where other topographical techniques are difficult or impossible to use.

Where, the digital photogrammetry is inapplicable in some extreme conditions, such as the drilling tunnels, but the laser scanning is applicable in these complex situations. The measurement with a Total Station is also an option, but the advantage of the laser scanning is obvious: instead of focus on the rather limited number of specified points, the laser scanning delivers millions of 3D points in a complete monitored tunnel section.

Recently, the improvements of this technology regarding the speed, accuracy, software algorithms and the fall in price have introduced a high potential of large scale applications in highly demanding engineering environments such as tunnels, bridges and heritage buildings. Tunnels, in particular those of a long length, create great challenges for surveyors due to their elongation to obtain the satisfactory geometry of the scanned data.

Road slope disasters continue to occur in recent years, due to the complexity of the road slope disaster, it is difficult to make effective monitoring and research. To safely and effectively monitor the safe status of the road slope, three parameters of slope are measured by advanced three-dimensional laser scanning technology: e.g. the earthwork calculation based on triangular mesh, the slope calculation based the point cloud data, contour distribution based on mesh grid. And then the state of the slop is respectively analyzed from the three aspects. Based on a single measurement parameter to determine the state of the slope is one-sided. To overcome the one-sidedness, a fuzzy comprehensive evaluation method is proposed, which can determine a more comprehensive trend of slope deformation, it is important to warn the slope disaster. The method set the factors of changes of the amount of earthwork, slope and contour distribution as evaluation factors, then the membership of each evaluation factor is determined using three factors share in the proportion of all factors, then, the slope state is evaluated by using the model proposed. This method can be more comprehensive to determine the slope deformation trends, and it is important to warn the slope disaster.

Before the 3D measurement, the slope image is processed: for each 2-D part, the color and edge information is used to get planes, and then the different planes are constructed into 3D slope.

The core technology of terrestrial laser scanning is the LiDAR technique, which is used to obtain the distance of each object point from the lens. The acronym LiDAR stands for Light Detection and Ranging. The laser system produces and emits a beam (or a pulse series) of highly collimated, directional, coherent and in-phase electromagnetic radiation. When the light reflected by the surface of an object is received, the system can calculate the range by the flight time and acquire the reflectivity of the surface. There are two different methods of range determination: phase and pulse. The former is more accurate in range but suffers from a limited range. Alternatively, the latter can measure in a greater range. Therefore, the latter is implemented in the most TLS used for the measurement of civil construction.

9393-24, Session PTues

About using Pockels cell for time-of-flight imaging

Frédéric Truchetet, Le2i - Lab. d'Electronique Informatique et Image (France) and Univ. de Bourgogne (France); Jing Min Teow, Mei Chen Tay, Univ. Teknologi Petronas (Malaysia)

No Abstract Available

9393-25, Session PTues

Towards automated firearm identification based on high resolution 3D data: Rotation-invariant features for multiple line-profile-measurement of firing pin shapes

Robert Fischer, Fachhochschule Brandenburg (Germany); Claus Vielhauer, Fachhochschule Brandenburg (Germany) and Otto-von-Guericke Univ. Magdeburg (Germany)

SCIENTIFIC TOPIC AND APPLICATION CONTEXT

The traditional examination of toolmarks impressed on shot cartridges and bullets plays an important role in criminalistic examinations. The main objective is to link a spent cartridge or bullet to a specific firearm, or exclude that a specific gun has provoked the markings. The underlying concept is based on two main hypotheses. Firstly, markings that are provoked by a weapon are consistent and reproducible. Secondly, it is possible to differentiate between the individual markings of two different weapons [1]. The application of 3D acquisition techniques for digital crime scene analysis promises the possibility to extract very fine features for a more reliable classification. Furthermore, 3D systems provide extended surface characteristics and allow the application of features based on topography information. Thus, the introduction of devices for 3D surface measurement combined with pattern recognition approaches promise to make the procedure of forensic firearm examinations more reproducible and reliable.

TECHNICAL CHALLENGES

The application of confocal microscopy is described as a promising technique for acquiring firearm-related toolmarks. At the same time challenges are arising, especially important is the handling of noisy topography data. This renders a direct warping and comparison of the 3D data difficult. Our presented approach attempts to overcome this limitation by using a configurable amount of profile-lines to describe the 3D shape of firing-pin impressions. Due to the circular appearance of cartridge bottoms, another important challenge is the application of features with rotation invariance. Those allow an examination without previous registration of the samples with respect to the alignment of rotation. Therefore, it would help to decrease required user interaction and eliminate a possible source of error during registration stage. The use of features with rotation invariance for automatic firearm identification systems has been motivated by a lot of authors. Our approach is adopting this technique for utilization on high resolution 3D scans of cartridge bottoms. Furthermore, our presented approach is extending the concept of profile-lines based on point/line-scanning-measurement devices, to the application of multiple profile-lines using area-measurement devices.

RELATED WORK

In the final paper we will give a comprehensive review of the line-profile measurement related works from Smith et al., Bachrach et al., Sakarya et al., Bolton-King et al., and Suapang et al. In the final paper we will present a comprehensive review of the rotation-invariant related works from Li et al., Geradts et al., Leng et al., Thumwarin et al., and Ghani et al. Thumwarin et al. [13] introduced a method that achieves a correct classification rate of 87.53%. Ghani et al. [16,17,18,19] introduced the application of different order geometric moments for firearm identification. According to the authors least recent publications the method achieves a correct classification rate between 96% and 65.7%.

OUR PROPOSED METHOD

We introduce two features for topography measurement of firing-pin impressions. These features combine a freely configurable amount of multiple profile-lines to describe the firing-pin shapes. Detail-scans of firing-pin impressions are covering an area of approx. 2.5x2.5mm acquired by using a 20-fold magnification and 0.5µm z-pitch. The dot-distance of the digital result is equal to 1.3µm (approx. 18000ppi). The test set contains samples of three different ammunition manufactures, three weapon models and six individual guns. Each possible combination of weapon model, weapon instance and ammunition type is represented by four samples this results in an amount of 72 cartridge samples. Preprocessing is applied including image preprocessing, as well as necessary segmentation and lateral alignment of data. Subsequently the features are calculated using different configurations. The fused feature set is subsequently used for classification. The performance of the proposed method is evaluated by using 10-fold stratified cross-validation. Our evaluation goal is three-fold; during the first part (E1) we evaluate the features degree of freedom regarding different angles of rotation. During the second part (E2) we evaluate how well it is possible to differentiate between two 9mm weapons of the same mark and model. This is extended during the third part (E3) here we evaluate the discrimination accuracy regarding a set of six different 9mm firearms, at which each two of the guns are of same mark and model.

The MAPn feature is a composition of n straight path-lines regarding different angles. For example, the application of the MAP8 feature results in 4 individual path-lines. At first a polar mapping is used to collect the height values along each of these path-lines in a vector. The length of each vector is equal to 2 times the outer circles radius. Additionally, all of the n vectors are concatenated to obtain one global path. During the second step nine statistics are calculated for each local and the global path. At the current sate we use min, max, span, incremental-length, cumulative-length, mean, variance, standard-deviation and rms. Generally the MAPn feature results in n/2+1 straight path-lines with 9 statistical values for each path. Using example of MAP8 this results into 45 values. A formalization of the MAPn feature is provided in appendix A Table 1. An illustration of the MAP feature using configurations MAP4, MAP8, MAP16 and MAP32 is given in the appendix B Figure 1.

The MCPn feature is a composition of n circular path-lines regarding different radii. At first a polar mapping is used to collect the height values along each of the circular path-lines in a vector. Additionally, all of the n vectors are concatenated to obtain one global path. During the second step nine statistics are calculated for each local and the global path. At the current sate we use min, max, span, incremental-length, cumulative-length, mean, variance, standard-deviation and rms. Generally the MCPn feature results in n+1 circular path-lines with 9 statistical values for each path. Using example of MCP10 this results into 99 values for each sample. A formalization of the MCPn feature is provided in appendix A Table 1. An illustration of the MCP feature using configurations MCP5, MCP10, MCP15 and MCP20 is depicted in appendix B Figure 2.

PRELIMINARY RESULTS

All samples are acquired with random alignment of rotation and no subsequent registration is applied, due to that we assume the features to be fully invariant to rotation (E1). This will be further evaluated in the final paper by rotating different subsets of samples. Regarding the individualization of two different firearms of the same brand and model (E2) the best results: for Walther P99 100%, for Ceska 75B 100%, and for Beretta 92FS 91.66% correct classification rate using feature MCP15 and a RotationForest classifier. The results for MAP32 feature are 83.33%, 91.66%, and 87.5% correct classifications using same classes and classifier. Regarding the individualization of six different guns (E3) the best result of 86.11% is achieved using the MCP15 feature and a RotationForest classifier. The utilization of MAP32 feature results in a correct classification rate of 62.5%. A simple feature level fusion of MAP32 and MCP15 features results in classification accuracy of 84.72%. Compared to the state of the art our approach achieves a similar level of accuracy e.g. [13] 87.53% or the least recent publications of Ghani et al. [18,19] with 65.7% and 75.4%. In contrast to [13,18,19] we use real 3D features. A more comprehensive presentation and discussion of experimental results with regards to E1, E2 and E3 will be presented in the final paper. Up to now feature values are simply fused by concatenating them into one vector. No ranking of features or feature selection are done so far. For the final paper we plan an extensive feature analysis with respect to their individual impact on the classification result. Furthermore, a feature selection seems to be very promising for increasing the performance.

9393-26, Session PTues

Continuous section extraction and over underbreak detection of tunnel based on 3D laser technology and image analysis

Xin Zhang, Chang'an Univ. (China)

In order to ensure safety, long term stability and quality control in modern tunneling operations, the acquisition of geotechnical information about encountered rock conditions and detailed installed support information is required. The limited space and time in an operational tunnel environment make the acquiring data challenging. The laser scanning in a tunneling environment, however, shows a great potential.

The surveying and mapping of tunnels are crucial for the optimal use after construction and in routine inspections. Most of these applications focus on the geometric information of the tunnels extracted from the laser scanning data. There are two kinds of applications widely discussed: deformation measurement and feature extraction.

The traditional deformation measurement in an underground environment is performed with a series of permanent control points installed around the profile of an excavation, which is unsuitable for a global consideration of the investigated area. Using laser scanning for deformation analysis provides many benefits as compared to traditional monitoring techniques. The change in profile is able to be fully characterized and the areas of the anomalous movement can easily be separated from overall trends due to the high density of the point cloud data. Furthermore, monitoring with a laser scanner does not require the permanent installation of control points, therefore the monitoring can be completed more quickly after excavation, and the scanning is non-contact, hence, no damage is done during the installation of temporary control points.

The main drawback of using the laser scanning for deformation monitoring is that the point accuracy of the original data is generally the same magnitude as the smallest level of deformations that are to be measured. To overcome this, statistical techniques and three dimensional image processing techniques for the point clouds must be developed.

For safely, effectively and easily control the problem of Over Underbreak detection of road and solve the problem of the roadway data collection difficulties, this paper presents a new method of continuous section extraction and Over Underbreak detection of road based on 3D laser scanning technology and image processing, the method is divided into the following three steps: based on Canny edge detection, local axis fitting, continuous extraction section and Over Underbreak detection of section. First, after Canny edge detection, take the least-squares curve fitting method to achieve partial fitting in axis; Then adjust the attitude of local roadway that makes the axis of the roadway be consistent with the direction of the extraction reference, and extract section along the reference direction; Finally, we compare the actual cross-sectional view and the cross-sectional design to complete Overbreak detected. Experimental results show that the proposed method have a great advantage in computing costs and ensure cross-section orthogonal intercept terms compared with traditional detection methods.

9393-27, Session PTues

Efficient Edge-Awareness Propagation via Single-Map Filtering for Edge-Preserving Stereo Matching

Takuya Matsuo, Shu Fujita, Norishige Fukushima, Yutaka Ishibashi, Nagoya Institute of Technology (Japan)

Accurate depth maps are needed for a great variety of applications (e.g. 3D object modeling, object recognition and depth-image-based rendering) in the field of computer vision. Stereo matching is a method to estimate depth maps, and can be separated into local methods and global methods. To obtain the accurate depth maps, global stereo matching is preferred, but the global methods are not suitable for real-time applications. For this reason, we focus on the local stereo matching, in this paper.

Generally, the local stereo matching consists of three steps: calculation of matching cost, cost aggregation and post filtering. We can obtain accurate depth maps if we adopt edge-preserving filtering in the cost aggregation. However, the computational cost is high since we must perform filtering for every disparity ranges (i.e., 256 times in a byte accuracy). The issue of the computational cost can be solved if we exploit box filtering for the cost aggregation. The box filter can be performed by making an integral image, which can be made computationally quite efficient, and its cache-efficiency is high. Nevertheless, the accuracy of the estimated depth map becomes low.

To solve the issue of the trade-off, therefore, we propose an efficient framework for real-time stereo matching to obtain accurate depth maps, which are edge-preserved.

The basic idea of our framework is that iterates three steps of local stereo matching. In this regard, the box filter is applied to the cost aggregation step, and the edge-preserving filter is utilized for the post filtering step. The essential point is that we propagate the edge-awareness by feeding back the refined depth map to the calculation of matching cost. Specifically, the matching cost is the sum of the already computed by the first matching cost and the additional cost which shows the nearness of the refined depth map in the previous loop. Moreover, the spatial domain of the box filter is gradually decreased. Due to this iteration, the refined depth map is reflected into the next estimation; hence the edge-awareness is propagated. As a result, small error regions are firstly removed, and large error regions are gradually improved.

In our experiment, we compared the accuracies of the depth maps estimated by our method with that of the state-of-the-art stereo matching methods. The experimental results show that the accuracy of our method is comparable to the state-of-the-art stereo matching methods. For example, the error rate (the error threshold is 1.0, and the evaluated region is non-occluded region) of our result with 5 iterations is approximately 3.56% when the dataset is "Teddy". The result ranks in the 13th of the all stereo matching methods in Middlebury Stereo Evaluation site. Especially, our result ranks in the 1st among the local stereo matching methods.

In this paper, we proposed an efficient iterative framework which propagates edge-awareness by single time edge-preserving filtering. The contribution of our work is that we made edge-preserving stereo matching more efficient, and result in the top among the local stereo matching.

9393-28, Session PTues

Disparity fusion using depth and stereo cameras for accurate stereo correspondence

Woo-Seok Jang, Yo-Sung Ho, Gwangju Institute of Science and Technology (Korea, Republic of)

CONTEXT

Recently, three-dimensional (3D) content creation has received a lot of attention due to the financial success of many 3D films. Accurate stereo correspondence is necessary for efficient 3D content creation.

OBJECTIVE

The objective of this paper is to obtain accurate depth information using depth and stereo images. Over the past several decades, a variety of stereo-image-based depth estimation methods have been developed to obtain accurate depth information. However, accurate measurement of stereo correspondence from natural scene still remains problematic due to difficult correspondence matching in several regions: textureless, periodic texture, discontinuous depth, and occluded areas. Usually, depth cameras are more

effective in producing high quality depth information than the stereo-based estimation methods. However, depth camera sensors also suffer from inherent problems. Especially, they produce low resolution depth images due to challenging real-time distance measuring systems. Such a problem makes depth cameras not practical for various applications. Thus, in the work, we propose a disparity fusion method to make up for weakness of stereo-based estimation and carry out better correspondence matching by adding a depth camera to the stereo system.

METHOD

The proposed method is initially motivated by global stereo matching. The information of the depth camera is included as a component of the global energy function to acquire more accurate and precise depth information. Depth camera processing is implemented as follows: 1) Depth data is warped to its corresponding position of the stereo views by 3D transformation. 3D transformation is composed of two processes. First, the depth data is backprojected to the 3D space based on the camera parameters. Then, the backprojected data in the 3D space is projected to the target stereo views. 2) Depth-disparity mapping is processed. Due to the different representation of the actual range of the scene, correction of the depth information is required. 3) We perform joint bilateral upsampling to interpolate the low resolution depth data. This method used high resolution color and low resolution depth images to increase the resolution of the depth image. The processed information of the depth camera is applied as the additional evidence for data term of disparity fusion energy function.

RESULTS

In order to evaluate the performance of the proposed method, we compare our proposed method with the stereo-image-based and depth upsampling method. The results indicate that the proposed method outperforms other comparative methods. The visual comparison of the experimental results demonstrates that the proposed method can represent the depth detail and improve the quality in the vulnerable areas of stereo matching.

NOVELTY

This paper presents a novel stereo disparity estimation method exploiting a depth camera. The depth camera is used to supplement crude disparity results of stereo matching. The camera array is determined to reduce the inherent problems of each depth sensor. Furthermore, the disparity acquisition algorithm deals with fully unsolved problems by fusing and refining the depth data. This increases the precision and accuracy of the final disparity values by allowing large disparity variation.

9393-1, Session 1

Object matching in videos using rotational signal descriptor

Darshan Venkatrayappa, Philippe Montesinos, Daniel Diep, Mines Alès (France)

In this paper, we propose a new approach for object matching in videos. By applying our novel descriptor on points of interest we obtain point descriptors or signatures. Points of interest are extracted from the object using a simple color Harris detector. This novel descriptor is issued from a rotating filtering stage. The rotating filtering stage is made of oriented anisotropic half-gaussian smoothing convolution kernels. Further, the dimension of our descriptor can be controlled by varying the angle of the rotating filter by small steps. Our descriptor with a dimension as small as 36 can give a matching performance similar to that of the well known SIFT descriptor. The small dimension of our descriptor is the main motivation for extending the matching process to videos. By construction, our descriptor is not euclidean invariant, hence we achieve euclidean invariance by FFT correlation between the two signatures. Moderate deformation invariance is achieved using Dynamic Time Warping (DTW). Then, using a cascade verification scheme we improve the robustness of our matching method. Eventually, our method is illumination invariant, rotation invariant, moderately deformation invariant and partially scale invariant.

9393-2, Session 1

Depth propagation for semi-automatic 2D to 3D conversion

Ekaterina V. Tolstaya, Petr Pohl, SAMSUNG Electronics Co., Ltd. (Russian Federation); Michael N. Rychagov, Samsung Advanced Institute of Technology (Russian Federation)

Nowadays, huge gain of interest to 3D stereoscopic video forced multimedia industry to produce more and more of stereo content. This need attracted much attention to conversion techniques, i.e. producing 3D stereo content from common 2D monocular video. Algorithms for creating stereo video from mono stream can be roughly divided in two main subcategories: fully automatic conversion which are implemented most often via SoC inside TV, and semi-automatic (operator-assisted) conversion, using special application, tools for marking videos, serious quality control, and so on. In this paper, we will focus on the most challenging problem that arises in semi-automatic conversion: temporal propagation of depth data.

Operator assisted pipeline for stereo content production usually involves manual drawing of depth for some selected reference frames (key-frames), and subsequent depth propagation for stereo rendering. An initial depth assignment is done sometimes by drawing just disparity scribbles, and after that restored and propagated using 3D cues. In others, full key-frame frame depth is needed.

The bottleneck of this process is quality of propagated depth: if the quality is not high enough, a lot of visually disturbing artifacts appear on final stereo frames. The propagation quality strongly depends on the frequency of manually assigned key-frames, but drawing a lot of frames requires more manual work and makes production slower and more expensive. That's why the very crucial problem is error-free temporal propagation of depth data through as many as possible frames. Optimal key-frame distance for desired quality of outputs is highly dependent on properties of video sequence. We recommend using automatic selection of key frames based on down-sampled between-frame motion (optical flow) analysis. The description of such key-frame detection algorithm is beyond scope of this article.

A problem of dense video tracking can be addressed by application of one of many motion estimation (optical flow) algorithms [1-3] with application of post-filtering, most often based on bilateral or trilateral filtering [4-5]. Most of current motion estimation algorithms are not ready for larger displacement and cannot be used for interpolation over more than few frames. The integration of motion information leads to increasing motion errors, especially near objects edges and in occlusion areas. Bilateral filtering can cover just small displacements as well, but this can be applied in occlusions area.

In current paper, we propose an algorithm for propagation of full frame (dense) depth from key frames to other frames. Our method utilizes nearest preceding and nearest following frames with known depth information (reference frames). The propagation of depth information from two sides is essential as it allows to solve most occlusion problems correctly. At first step, RGB video frames are converted to YCrCb color space. This allows treating luminance and chrominance channels differently. To accommodate fast motion that process is done using image pyramids. Three pyramids for video frames and two pyramids for key depth information are created. Iterative scheme starts on the coarsest level of pyramid with matching two reference frames to current video frame. Initial depth is generated by voting over patch correspondences using weights dependent on patch similarity and temporal distances from reference frames. On the next iterations, matching between images combined from color and depth information is accomplished. Owing to performance reasons, only one of chrominance channels (Cr or Cb does not make big difference) with given depth for reference frames is replaced, and thus depth estimate is obtained for current frame. On each level, we perform several matching iterations with updated depth image. A Gaussian kernel smoothing with decreasing kernel size is used to blur depth estimate on every level. This removes small noise coming from logically incorrect matches. Low resolution depth result for current

frame is up-scaled and the process is repeated for every pyramid level and ends at finest resolution, which is the original resolution of frames and depth.

Image matching is based on the coherency sensitive hashing (CSH) method [6]. First, images are filtered with bank of filters, using Walsh-Hadamard kernels. After that, having a vector of filtering results for each filter, a hash code (corresponding, indeed, to the whole patch) is constructed. Hash code is an integer of 18 bits, and having hashes for each patch, a greedy matching algorithm selects patches with the same hash, and compute patch difference, as difference between vectors with filtering results. The matching error used in search for best correspondence is a combination of filter output difference and spatial distance with dead zone (no penalty for small distances) and limit for maximum allowed distance. Spatial distance between matching patches was introduced to avoid unreasonable correspondences between similar image structures from different part of image. Such matches are improbable for relatively close frames of video input.

Disclosed results are compared with motion based temporal interpolation. The proposed algorithm keeps sharp depth edges of objects even in situations with fast motion or large occlusions. It also handles well many situations, when the depth edges don't perfectly correspond with true edges of objects. On the other hand, motion-based algorithm provides more temporally stable tracking for motions recognized by optical flow and is better in handling zooming and rotation of scene or objects. Our algorithm is parallelizable and suitable for GPU implementation. With partial GPU implementation the running time is about 3 seconds for 0.5 megapixels frame (960x540 pixels).

In general, some situations remain difficult for propagation: the low contrast videos, noise, and small parts of moving objects, since in this case background pixels inside patch occupy the biggest part of the patch and contribute too much in voting. However, in case when background does not change substantially, small details can be tracked quite acceptably. The advantages given by CSH matching include the fact that it is not true motion, and objects on the query frame can be formed from completely different patches, basing only on their visual similarity to reference patches.

REFERENCES

[1] Werlberger M. and Pock T. and Bischof H., " Motion estimation with non-local total variation regularization", Proc.CVPR 2010, 2464-2471, (2010)

[2] Sun D., Roth S., Black M.J., "Secrets of Optical Flow Estimation and Their Principles," Proc. CVPR 2010, 2432-2439 (2010).

[3] Pohl P., Sirotenko M., Tolstaya E., Bucha V., "Edge preserving motion estimation with occlusions correction for assisted 2D to 3D conversion", Proc. SPIE 9019, Image Processing: Algorithms and Systems XII, 901906 (2014), 1-8, (2014)

[4] Varekamp C., and Barenbrug B., "Improved depth propagation for 2d to 3d video conversion using key-frames", 4th European Conference on Visual Media Production, 2007. IETCVMP

[5] Muelle M.r, Zill F.y, KauffP.,"Adaptive cross-trilateral depth map filtering", In 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2010 (pp. 1-4). IEEE.

[6] Korman S., Avidan S., "Coherency Sensitive Hashing", Proc. ICCV 2011, 1607-1614 (2011)

9393-3, Session 1

Exploiting time-multiplexing structured light with picoprojectors

Mario Valerio Giuffrida, Giovanni M. Farinella, Sebastiano Battiato, Univ. degli Studi di Catania (Italy); Mirko Guarnera, STMicroelectronics (Italy)

A challenging task in computer vision is the depth estimation of the objects located in a scene.

Since the information about the distance of the objects is lost in a picture, several techniques have been developed to cope with this problem. Some of them use specific hardware for estimating this information, such as laser scanner or ultrasounds. Other techniques involve specific hardware beside vision applications. A stereo camera system can be employed for the task, by exploiting the parallax phenomenon. Using two cameras placed in front of the scene at a small distance to each other, it is possible to compute the disparity between two matching points in order to estimate the depth value associated to that location. Another way to compute the depth is using structured light, where special light patterns are projected on the scene. The estimation is done by analyzing how those patterns get distorted when they hit the objects in the scene.

In the literature there are three kinds of light patterns that can be used for the purpose.1 The first method is called direct coding, which employs a special light pattern identifying univocally each pixel. A second technique for the structured light is the spatial neighborhood, where a light patterns is shot on the scene and the depth information of each pixel is found in a small spatial range. For instance, Kinect exploits this technique by shooting a specific infrared pattern, which allows to estimate the depth in real time. The last methods is the time-multiplexing and it shoots different light patterns in the time to estimate the object's distance. It has been widely proven that both direct coding and spatial neighborhood can be used for real-time depth estimation, whereas time-multiplexing cannot be perform so fast. Since it requires several patterns to calculate the object's distance, it would not be feasible to be fast enough for real-time applications.

We present a hardware and software framework, using a picoprojector provided by ST Microelectronics Catania, inspiring our work from [2]. The software has been developed in C language, using the well-known computer vision library OpenCV. The hardware framework uses a picoprojector connected to a common notebook, shooting light patterns on the objects.

Picoprojectors have two main problems making the depth estimation more tricky: the former one is the image distortion and the latter one is about flickering. The technology used for the projection, laser steering beam, is the main cause of flickering. The way of working is similar to CRT, thus a proper synchronization between camera acquisition and structures projection is necessary.

We employ the time-multiplexing structured light technique3 ,4 projecting gray code patterns5 on the objects. The gray code is a binary numeral system, developed by Frank Gray (Figure 1). It has the property that two successive values differ by one bit, namely the hamming distance between them is just one. Because of the error due to illumination, noise and projector lens distortions, the gray code is a better alternative to the classic binary coding, since it gives stabler results in the depth estimation. The employed approach shoots many gray code patterns on the objects, by projecting specific black and white bands on the scene. For each point in the image, it could fall either in a black or white area, labeling pixels with a sequence of ones and zeros by analyzing their positions for each pattern. The correspondent decimal value is used to compute the disparity map and estimate the depth value corresponding to each pixel. In order to improve our results, we employ homography rectification to compensate the picoprojector's distortions. The proposed time multiplexing structured light with picoprojectors can be used in different contexts, such as 3D object scanning. An example of our framework is shown in Figure 2. The rightmost image represents the actual depth estimation. Noise reduction methods can be employed to reduce the noise in areas where shadow is cast.

Further details about the mathematical background, current implementation and analytic results will be deepen discussed in the final paper.

9393-4, Session 2

Joint synchronization and high capacity data hiding for 3D meshes

Vincent Itier, William Puech, Lab. d'Informatique de Robotique et de Microelectronique de Montpellier (France); Gilles Gesquière, Lab. des Sciences de l'Information et des Systèmes (France); Jean-Pierre Pedeboy, Stratégies S.A. (France)

Three-dimensional (3-D) meshes are already profusely used in lot of domains. In this paper, we propose a new high capacity data hiding scheme for vertex cloud. Our approach is based on very small displacements of vertices, that produce very low distortion of the mesh. Moreover this method can embed three bits per vertex relying only on the geometry of the mesh. As an application, we show how we embed a large binary logo for copyright purpose.

9393-5, Session 2

Digitized crime scene forensics: automated trace separation of toolmarks on high-resolution 2D/3D CLSM surface data

Eric Clausing, Claus Vielhauer, Otto-von-Guericke Univ. Magdeburg (Germany) and Fachhochschule Brandenburg (Germany)

1. Scientific Topic & Application Context

Toolmark investigation is a main aspect of criminal forensics. In this case toolmarks can be everything from clearly visible crowbar impacts to subtle microscopic traces of illegal opening attempts on locking cylinders. Especially for the latter of both a time-consuming, difficult manual process must be performed to detect, acquire and interpret relevant traces in the form of toolmarks. Such an investigation process is nowadays performed manually by forensic experts with almost no technical support except for a classic light microscope. By partially transfering this analysis to a high-resoluted 3D domain and by providing technical support for detection, segmentation, separation and interpretation of toolmarks, one can significantly improve the efficiency and reproducibility of the classic forensic process.

2. Technical Challenge & Relevance

A main challenge when dealing with high-resolution 3D surface data, is the precise detection and segmentation of relevant regions. In case of relevant regions overlapping each other and rather complex texture of the surface, the task becomes even more difficult. Especially in the case of mechanically fabricated and frequently used metal components (such as the components of a locking cylinder), these surfaces are naturally cluttered with a vast number of toolmarks of either relevant (toolmarks originating from illegal opening methods) or irrelevant (e.g. toolmarks originating from fabrication) origin. In most cases these toolmarks form complex formations with relevant traces overlapping and distorting each other. Even for the highly experienced and well trained forensic expert, the differentiation of single traces is not a simple task to solve. With our approach from [CV14] we present a method to automatically differentiate between relevant and irrelevant toolmarks on high-resolution 2D/3D data, to segment relevant toolmarks as precise as possible and to visualize results in an adequate way for forensic experts. This includes the adequate 2D/3D acquisition of the surfaces as proposed in [CKD1], and the pre-processing and preparation of the acquired high-resolution data (as proposed in [CKD2]) to allow for the application of our segmentation approaches from [CV14]. As the segmentation approach alone is not capable of further distinguishing single traces in the masked regions, we have to further expand the segmentation approach to allow for a separation of overlapping traces to receive a set of single traces, rather than a trace region consisting of a number of single traces.

3. State of the Art & Our Novel Approach

As basis we use findings from 2D/3D acquisition, pre-processing and segmentation in [CKD1], [CKD2] and [CV14]. In [CKD1] we propose an acquisition method which allows for a stepwise partial scanning of the whole key pin surface in about 45 partial scans. In [CKD2] we introduced a new digital SIFT-based assembling (SIFT, see [Low04]) of the key pin surface as a whole digital representation of the lateral surface of the key pin tip in topography, intensity and color. In [CV14] we propose a segmentation approach fusioning a texture classification with GLCM (Gray-Level-Cooccurrences-Matrices; [HSD73]) with an adapted pixel-based Gabor filtering approach (Gabor filtering, see [Dau85][Dau88]). We utilize known orientations and patterns of e.g. fabrication toolmarks to specifically dampen or amplify certain trace structures on certain positions of the 3D surface and thereby differentiate fabrication marks from relevant toolmarks. With this fusion we receive a quite precise segmentation on pixel-level with an accuracy of over 90%. However, the automated segmentation approach tends to comprehend multiple single traces into one masked region. But as we stated in [CKD2], information on shape and dimension of each single trace is essential for the steps of trace type determination (i.e., differentiation of traces of wear and traces of illegal opening attempts) and the determination of the opening method (i.e., determination of the technique most probably used for an opening attempt). For this purpose, it is necessary to further split the masked regions into separated traces. To allow for such a separation, we use a graph-based analysis of the binary masks resulting from [CV14]. With the help of different skeletonizing algorithms, we are able of approximating a graph to the segmentation mask. By analyzing the junction points and surface features along the graph paths, we are able of finding connecting regions and by that separate single traces from each another. As surface features we use a balanced set of statistical, roughness and texural features. The extraction of the surface features is performed on the intensity, topography and color representation of the investigated surfaces. The interpretation of the surface features is performed with the help of pattern recognition and thresholding.

4. Preliminary Results & Progress

Our approach is tested on a preliminary test set of about 1,300 single traces of key pins originating from locking cylinders opened with five different opening methods (Normal Key Usage, Single Pin Picking, Raking, Impressioning and Pick Gun). The surfaces are acquired with the 3D Confocal Laser Scanning Microscope Keyence VK-X 110 in three surface representations (intensity, topography, color). For evaluation, we use a comparison of manually segmented surface data (ground truth) to the results of our proposed automated segmentation and separation approach. In [CV14], we achieve a trace segmentation accuracy of over 90% on pixel-level. By adding the separation step after segmentation, we expect to gain at least 3-4% in segmentation accuracy. For the steps of trace type determination and determination of the opening method, we are already able of achieving almost perfect results of 99%-100% in [CKD2], but with segmentation and separation performed manually. With introduction of the automated separation, we expect to achieve comparable, (almost) perfect classification results for the given test set. The final paper will present and discuss detailed experimental results, both regarding the segmentation accuracy and the entire classification regarding opening methods of our enhanced scheme in relation to earlier results.

5. References

[CKD1] Eric Clausing, Christian Krätzer, Jana Dittmann, and Claus Vielhauer. A First Approach for the Contactless Acquisition and Automated Detection of Toolmarks on Pins of Locking Cylinders Using 3D Confocal Microscopy, In Proceedings of the on Multimedia and security (MM&Sec '12), ACM New York, NY, USA, pages 47-56, 2012.

[CKD2] Eric Clausing, Christian Krätzer, Jana Dittmann, and Claus Vielhauer. A first approach for digital representation and classification of toolmarks on locking cylinders using confocal laser microscopy. In SPIE Security + Defence: Optics and Photonics for Counterterrorism, Crime Fighting and Defence VIII, 854609. SPIE, 2012.

[CV14] Eric Clausing, and Claus Vielhauer. Digitized locksmith forensics: automated detection and segmentation of toolmarks on highly structured surfaces. Proc. SPIE 9028, Media Watermarking, Security, and Forensics 2014, 90280W (February 19, 2014); doi:10.1117/12.2036945.

[Dau85] J. Daugman. Uncertainty relations for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. In Journal of the Optical Society of America, volume 2, pages 1160–1169, 1985.

[Dau88] J. Daugman. Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression. In IEEE Trans on Acoustics, Speech, and Signal Processing, volume 36 (7), pages 1169–1179, 1988.

[HSD73] R. Haralick, K. Shanmugam, and I.Dinstein. Textural features for image classification. In IEEE Transactions on Systems, Man, and Cybernetics SMC, volume 3 (6), pages 610–621, 1973.

[Low04] D.G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. In International Journal of Computer Vision 60 (2), pp. 91-110, 2004.

9393-6, Session 2

Say No to Flat Face: A Non-uniform Mapping Technique for Precise Depth Representation

Wenxiu Sun, Zibin Wang, Lenovo (Hong Kong) Ltd. (Hong Kong, China)

Transmitting compactly depth maps from a sender can enable a multitude of imaging functionalities at a receiver, such as 3D reconstruction. While the depth map---projections of 3D geometry onto a chosen 2D image plane---can nowadays be readily captured by inexpensive depth sensors, they are often suffered by quality degradation after standard compression. For example, in some cases a 3D human face would become flat after compression which is undesirable for receiver. The 3D human face contains important details which is crucial for 3D reconstruction as well as other 2D/3D applications.

To preserve the important depth map details while achieving high compression rate, in this paper we consider to `tune' the depth map globally such that the details are easily preserved after standard compression at the sender and then `re-tune' them back at the receiver. Generally, the depth map is stored as a 8-bit integer number, 0-255. Based on the following observations from depth map, a) the important details may occupy only a limited range, which would be lost after compression, b) some not-so-important parts (such as floor) may occupy a large range, c) the depth intensities may not fully occupy the range from 0-255, we propose to apply a non-linear transformation globally to `tune' the depth map intensities, such that the range of important details is enlarged while the other non-important ranges are shrinked. The non-linear transformation function is calculated based on the depth distribution probability together with the importance factor. Intuitively, the more important the depth range is, the more the depth range should be scaled up, and vice versa. However, all the depth intensities should stilled be limited to 0-255. The non-linear function is constructed in a closed form and content adaptively. The importance factor can be identified in many ways, such as saliency detection, or can be specified by user. At the receiver side, an inverse non-linear transformation is applied to `re-tune' to the original range. Therefore, the non-linear function need to be sent to the receiver together with the compressed depth map in order to apply the inverse transformation. To reduce the number of bits used to transmit this function, we further propose to simply it to a piece-wise linear function. Using the H.264 codec for compression of depth map video, experimental results show that the compressed depth map by our algorithm can reduce MSE (mean squared error) by 31.6% compared to the traditional compression at the same compression rate.

9393-7, Session 3

3D puzzle reconstruction for archeological fragments

Frédéric Truchetet, Univ. de Bourgogne (France) and Le2i - Lab. d'Electronique, Informatique et Image (France); Florian Jampy, Le2i - Lab. d'Electronique, Informatique et Image (France); Antony Hostein, Univ. Paris 1 Panthéon Sorbonne (France); Eric Fauvet, Olivier Laligant, Le2i - Lab. d'Electronique, Informatique et Image (France)

No Abstract Available

9393-8, Session 3

Stereo matching with space-constrained cost aggregation and segmentation-based disparity refinement

Yi Peng, Ge Li, Ronggang Wang, Peking Univ. (China); Wenmin Wang, Peking University, Shenzhen Graduate School (China)

Stereo matching is a fundamental and hot topic in computer vision. Usually, stereo matching is mainly composed of four stages: cost computation, cost aggregation, disparity optimization and disparity refinement. In this paper we propose a novel stereo matching method with space-constrained cost aggregation and segmentation-based disparity refinement. There are mainly three technical contributions in our method: first, applying space-constrained cross-region in cost aggregation stage; second, utilizing both color and disparity information in image segmentation; third, using image segmentation and occlusion region detection to aid disparity refinement. State-of-the-art methods are used for cost computation and disparity optimization stages. The rank of our platform is the 2th in the Middlebury evaluation.

9393-9, Session 3

A real-time 3D range image sensor based on a novel tip-tilt-piston micromirror and dual frequency phase shifting

Øystein Skotheim, Henrik Schumann-Olsen, Jostein Thorstensen, Anna N. Kim, Matthieu Lacolle, Karl H. Haugholt, Thor Bakke, SINTEF (Norway)

Structured light is a robust and accurate method for 3D range imaging in which one or more light patterns are projected onto the scene and observed with an off-axis camera. Commercial sensors typically utilize DMD- or LCD-based LED projectors, which produce good results but have a number of drawbacks, e.g. limited speed, limited depth of focus, large sensitivity to ambient light and somewhat low light efficiency.

We present a 3D imaging system based on a laser light source and a novel tip-tilt-piston micro-mirror. Optical interference is utilized to create sinusoidal fringe patterns. The setup allows fast and easy control of both the frequency and the phase of the fringe patterns by altering the axes of the micro-mirror. For 3D reconstruction we have adapted a Dual Frequency Phase Shifting method which gives robust range measurements with sub-millimeter accuracy.

The use of interference for generating sine patterns provides high light efficiency and good focusing properties. The use of a laser and a bandpass filter allows easy removal of ambient light. The fast response of the micro-mirror in combination with a high-speed camera and real-time processing on the GPU allows highly accurate 3D range image acquisition at video rates.

9393-10, Session 3

A no-reference stereoscopic quality metric

Alessandro R. Silva, Centro Federal de Educação Tecnológica de Goiás (Brazil); Mylène C. Q. Farias, Univ. de Brasilia (Brazil); Max E. Vizcarra Melgar, Univ. de Brasília (Brazil)

Although a lot of progress has been made in the development of 2D objective video quality metrics, the area of 3D video quality metrics is still in its infancy. Many of the proposed metrics are simply adaptations of 2D

quality metrics that consider the depth channel as an extra color channel. In this paper, we propose a 3D no-reference objective quality metric that estimates 3D quality taking into account spatial distortions, depth quality, excessive disparity , and temporal information of the video. The metric is resolution and frame-rate independent. To estimate the amount of spatial distortion in the video, the proposed metric uses a blockiness metric. The contribution of motion and excessive disparity to 3d quality is calculated using a non-linear relative disparity measure that use the correlation of two local calculated histograms, and a frame-rate proportional motion measure. The metric's performance is verified against the COSPAD1 database. The MOS predicted using the proposed metric obtained good correlation values with the subjective scores. The performance was on average better than the performance of two simple 2D full reference metrics: SIMM and PSNR.

9393-11, Session 4

Coarse to fine: toward an intelligent 3D acquisition system

Frédéric Truchetet, Vincent Daval, Olivier Aubreton, Univ. de Bourgogne (France)

No Abstract Available

9393-12, Session 4

Mesh saliency with adaptive local patches

Anass Nouri, ENSICAEN (France); Christophe M. Charrier, Olivier Lézoray, Univ. de Caen Basse-Normandie (France)

3D object shapes (represented by meshes) include both areas that attract the visual attention of human observers and others less or not attractive at all. This visual attention depends on the degree of saliency exposed by these areas. In this paper, we propose a technique for detecting salient regions in meshes.

To do so, we define a local surface descriptor based on local patches of adaptive size and filled with a local height field. The saliency of mesh vertices is then defined as its degree measure with edges weights computed from adaptive patch similarities. Our approach is compared to the state-of-the-art and presents better results. A study evaluating the influence of the parameters establishing this approach will also be carried out.

9393-13, Session 4

Phase-aware-candidate-selection for Time-of-Flight Depth Map Denoising

Thomas Hach, ARRI AG (Germany); Tamara N. Seybold, Arnold & Richter Cine Technik GmbH & Co. Betriebs KG (Germany); Hendrik Böttcher, Technische Univ. München (Germany)

1. INTRODUCTION

The visual representation of a scene is typically captured by an RGB camera. In many sophisticated applications, additional information about the 3D position of pixels is requested. This 3D information can be captured using a Time-of-Flight (TOF) camera, which measures a so-called depth map, thus each pixel's distance from the camera, by phase measurement. For instance, this type of 2D plus depth information, an RGB image with matching depth map,1 can be used for enhanced green screen compositing which is solely done by color keying techniques to date.

While this technology is beneficial for the foresaid use cases, its limitations are low resolution compared to RGB sensors, phase ambiguity which leads to a limited range of distinct depth and strong noise. Fig. 1 shows a typical depth map taken with a PMD TOF camera which is degraded by two types of noise. The first is usually assumed as Gaussian noise and the second appears like Salt and Pepper noise. As the noise level of today's TOF-cameras is significant, the depth data cannot be used in most applications without a denoising step.

Typical denoising for TOF-data is done by joint-bilateral filtering (JBF).2–5 Although the JBF significantly reduces Gaussian noise, the typical solution for Salt and Pepper noise is median filtering. However, median filtering smoothes the image and therefore leads to a loss of details.

We propose to use a new method to remove the salt and pepper noise that is not based on the effect but deploys the physical TOF-specific causes leading to the effect. The cause of the salt and pepper noise in PMD TOF images is phase ambiguity and thus we propose a novel phase-aware pre-processing filter.

This phase-aware pre-processing filter is motivated by investigating the physical causes of the Salt and Pepper noise, the phase jumps. Supported by noise measurements, we define a confidence interval for each pixel and apply a pre-filter, that allows data points to be located outside the non-ambiguity range. The subsequent step removes the Salt and Pepper noise correcting only the affected pixels and thus without smoothing details. The remaining noise has a distribution that corresponds to the initial noise model and thus can effectively be reduced using usual denoising methods. We show that our approach reduces the overall error retaining more details and contrast.

2. NOISE CHARACTERISTICS OF A PMD TOF-CAMERA

The noise characteristic can be approximated by a Gaussian distribution which is visualized in Fig. 2.

In the TOF case, Salt and Pepper noise is not a different noise contribution, instead it is caused by the inherent Gaussian noise. Fig. 3 illustrates this issue. When a signal S is located close to the borders of the non-ambiguity range, in our example 7.5 m, the Gaussian noise leads to values exceeding the signal range. As the measurement is a phase measurement, this leads to a modulo effect with respect to the non-ambiguity range. That means, value C is mapped to value N near the opposite border of the signal range. Value N is very far away from the true signal S and hence appears as Salt and Pepper noise.

Instead of directly estimating the true signal S, we first recover value C in our method. This value is the physically correct depth value affected by Gaussian noise only. This pre-processing trick is crucial, as now the filters originally optimized for Gaussian noise distributions can be used adequately.

3. PHASE-AWARE-CANDIDATE-SELECTION

Our proposed pre-processing filter, the Phase-Aware-Candidate-Selection filter (PACS), estimates the value C based on the spatio-temporal neighborhood ? ? T. For the case that the current noisy depth value N is near the limits of the TOF range, PACS determines the corresponding candidate value C using the non-ambiguity range dnar (e.g. 7.5 m).

Then, PACS calculates weighted Euclidean distances d for N and C using equations (2) and (3) with G, a Gaussian weighting function.

To take advantage of the nearly noiseless edge information of the color image, the weighting function is based on the color intensities I. The last step is the decision for the final depth intensity D at the position p.

4. EXPERIMENTS

To evaluate the effectiveness of our approach, we use our pre-processing filter followed by a typical spatio-temporal joint- bilateral filter (TJBF) and compare it to the result achieved with TJBF without PACS. As median filtering is usually applied for images degraded by Salt and Pepper noise, we additionally include the results of a spatio-temporal median filter.6

We optimized the parameters of TJBF with and without PACS pre-filtering by an RMSE parameter sweep on a static sequence with respect to previously generated ground truth. The corresponding RMSE values with and without PACS are listed in Tab. 1. TJBF with PACS achieves clearly lower RMSE values than both TJBF without PACS and the median filter. Besides

the best RMSE values, Fig. 4 visually proves that even more details are retained. The clear outlines of the mannequin as well as the gap between left arm and body show sharp edges and high contrast.

5. CONCLUSION

The pre-processing algorithm PACS has shown convincing results. The elimination of RAN works on static sequences as well as on dynamic sequences. Median filtering is able to suppress RAN and achieves an average RMSE value of 0.430 m on the static sequence in contrast to PACS with 0.650 m. However, median filtering causes oversmoothing along edges in the depth map. This effect is prevented by PACS because PACS performs just a re-arrangement of the values instead of smoothing the depth map.

The influence of PACS on the main filter is the next evaluation step. While the denoising result of TJBF without PACS has an average RMSE of 1.327 m, TJBF achieves a reduction to 0.258 m using PACS. The observation of TJBF without PACS is again oversmoothing because of the high denoising attempt which was necessary to eliminate RAN.

9393-14, Session 4

Camera model compensation for image integration of time-of-flight depth video and color video

Hiromu Yamashita, Shogo Tokai, Shunpei Uchino, Univ. of Fukui (Japan)

In this paper, we describe a consideration of a method of camera calibration for TOF depth camera with color video camera to combine their images into colored 3D models of a scene. Mainly, there are two problems for the calibration to combine them. One is stability of the TOF measurements, and another is deviation between the measured depth values and actual distances to the points on objects based on a geometrical camera model. To solve them, we propose a method of two-steps calibration. In the 1st step, we analyzed a statistical model of measured depth values, and suppress a variation of them. In the 2nd step, we estimated an optimum offset distance and intrinsic parameters for the depth camera to match both measured depth values and ideal depth. For the estimation, we used the Zhang's calibration method for the intensity image by the depth camera and the color video image of a checker board pattern. Using this method, we can get the 3D models which are matched between depth and color information correctly and stably. We also explain effectiveness of our method by showing several experimental results.

9393-15, Session 5

A practical implementation of free viewpoint video system for soccer games

Ryo Suenaga, Kazuyoshi Suzuki, Tomoyuki Tezuka, Mehrdad Panahpour Tehrani, Keita Takahashi, Toshiaki Fujii, Nagoya Univ. (Japan)

Free viewpoint video generation is a technology that enables users to watch 3D objects from their desired viewpoints. Practical implementation of free viewpoint video for sports events is highly demanded. However, a commercially acceptable system has not yet been developed. The main obstacles are insufficient user-end quality of the synthesized images and highly complex procedures that sometimes require manual operations. In this work, we aim to develop a commercially acceptable free viewpoint video system. A supposed scenario is that soccer games during the day can be broadcasted in 3D, even in the evening of the same day.

Our work is still ongoing. However, we have already developed several techniques to support our goal. First, we captured an actual soccer game at an official stadium (TOYOTA stadium) where we used 20 full-HD professional cameras (CANON XF305). Due to significant brightness difference between shadowed and exposed areas in a sunny day, we carefully set camera parameters such as gain, exposure, and aperture to have almost equal capturing conditions among the cameras and to clearly capture players in the field. The cameras were placed among the seats in the stadium about 20m over the field level. The cameras were divided into two groups of 10 cameras. The first 10 cameras covered behind a goal roughly arranged in a 1D arc configuration. The other 10 cameras were placed along one of the side lines roughly in a 1D line configuration. In each group, 10 cameras were horizontally spaced with about 10m intervals. All cameras were faced into a common converging point in the soccer field. Each group of cameras was individually synchronized using generator-locking (GEN-LOCK) signals distributed through physical cables.

Second, we have implemented several tools for free viewpoint video generation as follow. In order to facilitate free viewpoint video generation, all cameras should be calibrated. We calibrated all cameras using checker board images and feature points on the field (cross points of the soccer field lines). We extract each player region from captured images by using background subtraction given an estimated background image. The background image is estimated by observing chrominance changes of each pixel in temporal domain. Using the extracted regions of players, we roughly estimated the 3D model of the scene, where the field is set to X-Y plane of the world coordinate and players' shapes are reconstructed approximately using cylindrical models. The estimated 3D model of the scene is further converted to a depth map given the camera parameters, which can be used for depth image-based rendering (DIBR). Additionally, we have developed a user interface for visualizing free viewpoint video generation of a soccer game using a game engine (Unity), which is suitable for not only commercialized TV sets but also devices such as smartphones. We are currently working hard to complete an end-to-end prototype free viewpoint video system. We expect to present this system at the conference venue.

9393-16, Session 5

Observing atmospheric clouds through stereo reconstruction

Rusen Oktem, Univ. of California, Berkeley (United States); David M. Romps, Univ. of California, Berkeley (United States) and Lawrence Berkeley National Lab. (United States)

Observing cloud life cycles and obtaining measurements on cloud features is a significant problem in atmospheric cloud research. Scanning radars have been the most capable instruments to provide such measurements, but they have shortcomings when it comes to range and spatial and temporal resolution. High spatial and temporal resolution is particularly important to capture the variations in developing convection. Stereo photogrammetry can complement scanning radars with the potential to observe clouds as distant as tens of kilometers and to provide high temporal and spatial resolution, although it comes with the calibration challenges peculiar to various outdoor settings required to collect measurements on atmospheric clouds. This work explores the use of stereo photogrammetry in atmospheric clouds research, focusing on tracking vertical motion in developing convection.

Two different stereo settings in Miami, Florida and in the plains of Oklahoma are studied. Oceanic clouds off the coast of Miami, Florida are captured with the former setting, and a calibration method that does not require stationary landmarks but uses the horizon and the epipolar constraint is developed for it. The accuracy of the calibration is verified by comparing 3D reconstructed cloud base heights against data obtained from a colocated ceilometer, and by comparing the reconstructed horizontal cloud motions against data from nearby radiosondes. A feature extraction and matching algorithm is developed and implemented to identify cloud features of

interest. FAST (Features from Accelerated Segment Test) feature extractor is used to obtain cloud features in one image, and these extracted features are matched in the corresponding image by use of a normalized cross-correlation index within a block search around the epipolar line. A two-level resolution hierarchy is exploited in feature extraction and matching. 3D positions of cloud features are reconstructed from matched pixel pairs, and the matched pixel pairs of successive time frames are further processed to obtain vertical velocities of developing turrets for analyzing deep-convective dynamics. Cloud tops of developing turrets in shallow to deep convection are tracked in time and vertical accelerations are estimated from the reconstructed 3D positions. A clustering method is required to identify pixel contributions from separate turrets which cannot be discerned in 2D images.

Our results show that stereophotogrammetry provides a useful tool to obtain 3D positions of cloud features as far as 50 km away. This enables observing cloud life cycles and tracking the vertical acceleration of turrets exceeding 10 km height, which provides a significant opportunity in atmospheric research.

9393-17, Session 5

Robust stereo matching based on probabilistic Laplacian propagation with weighted mutual information

Junhyung Kim, Seungchul Ryu, Seungryong Kim, Kwanghoon Sohn, Yonsei Univ. (Korea, Republic of)

[CONTEXT]:Stereo matching methods are prone to a false matching problem in an uncontrolled environment, e.g., radiometric source variations, camera characteristics, and image noises. To alleviate this problem, a number of methods have been proposed focusing on development of a robust cost function. However, such a robust cost function based methods cannot guarantee to overcome inherent limitation under severe radiometric variations.

[OBJECTIVE]:Recently, a progressive scheme has been considered an alternative approach to address the inherent ambiguities of the stereo matching. The Laplacian propagation is one of the most popular progressive schemes, which enables solving these problems by propagating unambiguous pixels, called as ground control points (GCPs), into ambiguous neighboring pixels. However, the conventional Laplacian propagation has two critical problems. First, it is very sensitive to errors in GCPs since it assumes GCPs to be ground-truth. If selected GCPs are erroneous, the errors are propagated and consequently degrade final dense disparity map. Second, when CGPs do not exist around discontinuities, it may lead to the lack of information needed for appropriately guiding the subsequent propagation process.

[METHOD]:To alleviate these problems, this paper proposes a probabilistic Laplacian propagation (PLP) in which GCPs are stochastically distributed according to the reliability of CGPs. Our approach proposes weighted mutual information (WMI) to compute initial GCPs, which is an enhanced version of mutual information with bilateral weight term. Moreover, to impose the reliability for GCPs, a difference ratio between the highest cost and the second highest cost in cost volume domain is encoded, and it is used to the confidence measure for initial GCPs. According to the reliability, some highly confident GCPs are selected and propagated using probabilistic Laplacian model in which the propagation of less reliable GCP is suppressed.

[RESULTS]:The performance of the proposed PLP is evaluated on the Middlebury datasets in Intel Core i7-2600 CPU 3.40GHz PC. The PLP was compared with the state-of-the-art optimization based robust stereo matching methods such as the MI, Census transform, and the ANCC. The experimental results showed that conventional optimization based methods provide unsatisfactory disparity maps since errors in an initial cost volume cannot be suppressed in optimization process. However, the proposed PLP outperforms the compared methods in terms of both disparity accuracy (in average 13% reduced errors) and computational complexity (in average 3 times faster), which indicates that the proposed PLP can be an alternative approach robust to radiometric distortions.

[NOVELTY]:First, it is the novel attempt to employ a progressive framework for the stereo matching under radiometric variations. Second, a novel confidence measure is proposed for stereo pairs taken under different radiometric conditions, and it is used to solve a probabilistic laplacian propagation model. Third, the WMI is proposed as a robust cost function to overcome the limitation of the conventional mutual information which is known to be robust but inaccurate at an object boundary.

9393-18, Session 5

Structure-aware depth super-resolution using Gaussian mixture model

Sunok Kim, Changjae Oh, Youngjung Kim, Kwanghoon Sohn, Yonsei Univ. (Korea, Republic of)

CONTEXT

Depth acquisition is a fundamental challenge in image processing and computer vision, and it is widely used in various applications including image-based rendering, 3D modeling, object detection, and tracking. Recently, an active depth sensor, e.g., a time-of-flight sensor, has been used to obtain a depth map. Due to the physical limitations, however, the depth map from the active sensor has low-resolution and acquisition noises.

OBJECTIVE

Conventional approaches in depth super-resolution have exploited corresponding high-resolution color image as a depth cue, which assumes that the nearby pixels with similar color may belong to the same depth [1-4]. However, it may blend depth values of different objects having similar color information, which causes depth bleeding and texture transferring artifacts.

METHOD

In this paper, we formulate the depth super-resolution problem as a probabilistic optimization. An efficient depth prior is proposed using Gaussian mixture model, which jointly encodes the color and depth structures. Since the proposed model shows an implicit form, a fixed-point iteration scheme is adopted to address the non-linearity of the smoothness constraints derived from the proposed depth prior. In each iteration, our model measures a new affinity weight that provides crucial information to enhance the noisy and low-resolution depth map.

RESULTS

We have validated our approach on the Middlebury dataset [5] and real-world dataset [6]. We compared our method with joint bilateral upsampling(JBU) [1], guided filter(GF) [2], 3D-JBU [3], weighted mode filter(WMF) [4], and anisotropic total generalized variation(ATGV) [6]. Among 23 dataset, our method outperformed state-of-the-art methods both quantitatively and qualitatively. Average of root-mean-squared-error of our results was 2.60, while JBU, GF, 3D-JBU, WMF, and ATGV showed 4.87, 4.69, 4.18, 2.81, and 2.73, respectively. A qualitative comparison shows that the proposed method alleviates depth bleeding and texture transferring artifacts effectively. Further experimental results can be found in the supplemental materials.

NOVELTY

1) The proposed model globally considers color and depth structures to estimate high-resolution depth map. In each iteration, a new affinity weight is constructed in feature space which considers both color and estimated depth values in the previous iteration. Thanks to the adaptive feature space, our model does not suffer from depth blending problem although two different objects have similar color. This property considerably reduces the depth bleeding and texture transferring artifacts which are frequently appeared in the conventional methods.

2) We show that the proposed method may be thought of as iteratively reweighted least squares in which the reweighted terms are generated from the previous solution. Such analysis demonstrates that why our model reconstructs sharp edges in the depth continuities.

REFERENCES

[1] J. Kopf et al., "Joint bilateral upsampling," TOG, 2007.

[2] K. He et al., "Guided image filtering," ECCV, 2010.

[3] Q. Yang et al., "Spatial-depth super resolution for range images," CVPR, 2007.

[4] D. Min et al., "Depth video enhancement based on weighted mode filtering," TIP, 2012.

[5] http://vision.middlebury.edu/stereo

[6] D. Ferstl et al.,"Image Guided Depth Upsampling using Anisotropic Total Generalized Variation," ICCV, 2013.

9393-19, Session 5

A new fast-matching method for adaptive compression of stereoscopic images

Alessandro Ortis, Sebastiano Battiato, Univ. degli Studi di Catania (Italy)

No Abstract Available

Conference 9394: Human Vision and Electronic Imaging XX

Monday - Thursday 9–12 February 2015

Part of Proceedings of SPIE Vol. 9394 Human Vision and Electronic Imaging XX

9394-1, Session Key

Up periscope!: Designing a new perceptual metric for imaging system performance *(Keynote Presentation)*

Andrew B. Watson, NASA Ames Research Ctr. (United States)

No Abstract Available

9394-40, Session Key

Cognitive psychology meets art: exploring creativity, language, and emotion through live musical improvisation in film and theatre *(Keynote Presentation)*

Monica Lopez-Gonzalez, La Petite Noiseuse Productions (United States) and Maryland Institute College of Art (United States) and Johns Hopkins Univ. (United States)

Creativity is defined as a mental phenomenon that engages multiple cognitive processes to generate novel and useful solutions to problems. The systematic psychological study of creativity began most significantly in 1950 with J.P. Guilford's experiments that quantified the various improvisatory outcomes created in response to test items. Since then, various types of ever-evolving behavioral experiments have tested problem-solving skills and the role of memory through the perception and creation of visual and auditory mental imagery. Two core creative thinking modes have been identified: long-term deliberate methodical problem solving vs. short-term spontaneous problem solving. The advent of brain-imaging techniques in the 1990s has further spawned a broad range of experiments within the arts; most focusing on professional artists as a way to understand the highly skilled brain during simplified moments of artistic creation and positing a significant correlation between prefrontal cortex attenuation and the spontaneous creative act. While behavioral models have been proposed integrating the multiple activities (e.g. technical issues, emotional responses) arising within and the socio-cultural effects surrounding the long-term creative process in various artistic disciplines, no systematic study exists of short-term improvisatory behavior in response to emotional stimuli within such ecologically valid contexts as live film and theatre.

I use musical improvisation within cinema and theatre to explore the process of spontaneous creative thinking and emotion perception, particularly as it pertains to the in-the-moment expressive translation of scenic variables such as actors' movements and dialogue within film and theatre to musical language. Staging live original film screenings and theatrical productions, I capture musicians while they improvise live in direct reaction to the actors' body language-changes and/or conversations within a scene. In this paper I discuss two novel projects that were prepared in order to be screened, performed, and recorded on stage with live improvised music by professional jazz musicians: a film titled "Moments" and a theatrical play titled "The Final Draw" that both explore the six universal human emotions (anger, disgust, fear, happiness, sadness, and surprise).

Using clearly marked emotional segments as regions of interest, determined by the combination of facial expressions and language semantics from the recorded footage, I performed scene analyses of both the film and theatrical productions and examined their respective improvised musical scores. Several observations are made: First, visual emotional content has a direct effect on improvised music. For example, a happy scene elicits the creation of happy music. Second, musical variables such as mode, tempo, rhythm, and dynamics are used to create clear musical emotional contexts to the visual and spoken narrative. For example, major and minor scales and fast tempi are implemented within the happy musical score. Third, salient non-emotional scenographic elements within the narrative are also simultaneously musically translated. For example, within the happy scene of Moments, water falling in a fountain is present within the visual frame and the improvised music mimics its sound effects. Fourth, the musicians engaged in interactive communication between each other that included various common musical exchanges such as melodic and rhythmic imitation, transposition, repetition, and motivic development. I argue that the use of improvisatory music adds an interesting layer of structure to both film and theatre narratives. Each of these musical scores takes on two distinct roles: (1) the primary role in which emotion and meaning are congruent with sound and image as in an ascending motion in tandem with an ascending melodic sequence to consequently interpret, enhance and elicit a particular emotion in the viewer, and (2) a secondary role in which sound becomes another character invented in-the-moment by the musician(s) that guides the perceiver through several possible narratives ex- or implicitly intended by the filmmaker and playwright. Both of these roles simultaneously played out elucidate many unanswered questions pertaining to how audiovisual information, perception, emotion, and the generation of new material all interact in one given moment to produce a coherent artistic object.

From these novel data and methods, I propose a cognitive feedback model of spontaneous creative emotional innovation that integrates music, spoken language, and emotional expression within the context of live music scoring in film and theatre.

9394-2, Session 1

Use of a local cone model to predict essential CSF behavior used in the design of luminance quantization nonlinearities *(Invited Paper)*

Scott J. Daly, Dolby Labs., Inc. (United States); Seyedalireza Golestaneh, Arizona State Univ. (United States)

Early work in designing luminance quantization nonlinearities includes the familiar gamma domain from the CRT era, and density (log luminance) quantization for hardcopy. For newer displays having higher dynamic ranges, these simple nonlinearities result in contour and other quantization artifacts, particularly in the dark end of the tonescale. Knowing the visual system exhibits behavior that ranges from square root behavior in the very dark, gamma-like behavior in dim ambient, cube-root in office lighting, and logarithmic for daylight ranges, nonlinearities were developed based on luminance JND data that used bipartite fields [1,2] and including cone models [3]. More advanced approaches considered general spatial frequency behavior, as opposed to solely the dominating edge of bipartite fields, and used the frequency-descriptive Contrast Sensitivity Function (CSF) modelled across a large range of light adaptation to determine the luminance nonlinearity. The DICOM medical imaging standard GSDF (grayscale standard display function, also referred to as electro-optical transfer functions OETF) [4] is a key example of this approach. It is based on the crispening effect [5] combined with the Barten spatial CSF model [6] to achieve a best-case visual system performance for worst-case engineering design, and is widely used. However, the DICOM nonlinearity is limited to the luminances 0.05 to ~4000 cd/m^2 and 10 bits. Newer High Dynamic Range (HDR) displays have exceeded both ends of this range, especially at the dark end, and 12-bit systems have been implemented as well.

A new approach [7] using the Barten light-adaptive CSF model has been developed that improves on DICOM by instead of tracking the sensitivity as a function of light adaptation level at a single frequency (4 cpd), it rather tracks the sensitivity of the CSF for the most sensitive spatial frequency,

which changes with adaptation level (Fig. 1). The maximum sensitivity across all frequencies for any adapted luminance level can easily be visualized by taking a typical surface plot of a light-adapted CSF (Figure 2A) and rotating its viewpoint to be from the luminance axis, so that the upper hull indicates the maximum sensitivity across all frequencies (Fig 2B). As in DICOM, an OETF nonlinearity is built up through a procedural process of summing JNDs calculated from the sensitivity function shown in Fig 2B, starting from the designed system minimum luminance. The resulting OETF is intended for a future visual signal format that is designed to be used across HDR and Wide Color Gamut (WCG) consumer TVs, home theater, and large scale cinema applications. The luminance range goes from 0.0 to 10,000 cd/m2, and with flexible bit-depths of 10 and 12.

There is a large body of neuroscience work that models the neural processing as an ideal observer using Bayesian analysis [8] where in certain cases the essential limit of performance is the transducer stage [9]. In addition, other work has found unification of cone models with appearance and detection [10]. With this viewpoint, we explored whether the cone photoreceptor's limits can be responsible for the maximum sensitivity of the CSF as a function of light adaptation; despite the CSF's frequency variations and that the cone's nonlinearity is a point-process. The crispening concept leads to the use of a cone model continuously adapted to its local neighborhood, referred to as the local cone model [3] shown in Figure 3 along with its global adaptation counterpart. We used a modified Naka-Rushton equation to tune the model to human psychophysics [11, 12]. In addition, a flare term was explored since light levels below 0.001 cd/m^2 were of interest. Rather than using procedural techniques, we tried to constrain our approaches to calculus, using partial derivatives to track the CSF max sensitivity ridgeline, and taking derivatives of the local cone model in the calculation of its contrast sensitivity. We found parameters* of this local cone model that could fit the max sensitivity of the Barten CSF, across all frequencies, and are within the ranges of parameters commonly accepted for psychophysically-tuned cone models , and an example using a flare term is shown in Fig. 4. The flare term was not needed, but can be used to allow more limited ranges of parameters to fit the CSF model.

In summary, a linking of the spatial frequency and luminance dimensions has been made for a key neural component. This provides a better theoretical foundation for the designed visual signal format using the aforementioned OETF. Further, since actual light adaptation levels are directly modelled in the semi-saturation constant of this model, it can allow for better application of the signal to differing ambient viewing conditions. This can lead to new ambient adjustment approaches that are certainly better than the confusing picture/contrast settings [13] of today's displays. Future work will investigate a more functional cone model [14] in a similar analysis to more fully account for the role of rod vision in the mesopic range [15].

* The local cone model's primary deviation from the max sensitivity derived from the CSF model is that it begins to decline at the highest luminances (around 1000 nits) while the CSF model does not (Figure 4). We were able to remove this decline and provide a better fit to the CSF model by allowing the Valeton - van Noren exponent to be set to 0.4, as opposed to 0.7 . However, the data on CSFs for light adaptation greater than 1000 nits is very limited, and we don't have strong confidence in the Barten CSF for that range. For the plot shown here, we chose to constrain the local cone model to constants that were commonly used in cone models in the literature. Other variations such as n < 0.7 and the case with no flare term will also be shown in the full paper.

REFERENCES

1. G. Ward: "Defining Dynamic Range". SID Symposium Digest of Technical Papers, 59.2 , 39(1), pp.2168-0159. Blackwell Publishing, 2008.

2. R. Mantiuk, et al (2005) Predicting Visible Differences in high-dynamic-range images: model and its calibration. SPIE proc 5666, Electronic Imaging Conference: Human Vision and Electronic Imaging, pp204-214.

3. M. Sezan, K. Yip, and S. Daly (1987) "Uniform perceptual quantization: applications to digital radiography". IEEE Trans. Systems, Man, and Cybernetics. V 17 #4. 622-634."

4. NEMA Standards Publication PS 3.14-2008, Digital Imaging and Communications in Medicine (DICOM), Part 14: Grayscale Standard Display Function, National Electrical Manufacturers Association, 2008.

5. P. Whittle, "Increments and decrements: luminance discrimination", Vis Res. V 26, #10, 1677-1691, 1986.

6. P. G. J. Barten (1999) Contrast Sensitivity of the Human Eye and its Effects on Image Quality, SPIE Optical Engineering Press: Bellingham, WA.

7. S. Miller, S. Daly, and M. Nezamabadi (2013) Perceptual Signal Coding for More Efficient Usage of Bit Codes," May/June 2013 issue of the SMPTE Motion Imaging Journal

8. D. Ganguli and E. P. Simoncelli (2014) Efficient sensory coding and Bayesian decoding with neural populations. Neural Computation, 2014.

9. H. B. Barlow et al (1987) human contrast discrimination and the threshold of cortical neurons. JOSA A 4:2366-2371.

10. Z. Xie and T. Stockham (1989) Toward the unification of three visual laws and two visual models in brightness perception, IEEE Trans SMC, V19, #2, 379-387.

11. J. M. Valeton and D. van Norren (1983) Light adaptation of primate cones: an analysis based on extracellular data, Vis Res. V23, #12, 1539-1547.

12. R. Normann, et al (1983) Photoreceptor contribution to contrast sensitivity: applications in radiological diagnosis. IEEE trans SMC-13, #5, 944-953.

13. R.W. Hunt (2010) The challenge of our unknown knowns. IS&T 18th annual Color Imaging Conference, 280-284.

14. J. H. van Hateren and H. P. Snipe (2007) Simulating human cones from mid-mesopic up to high-photopic luminances, JOV 7(4) 1-11.

15. McCann J. J., and Benton J. L., "Interactions of the long-wave cones and the rods to produce color sensations." JOSA. . 59, 103–107 ((1969)).

9394-3, Session 1

Display device-adapted video quality-of-experience assessment

Abdul Rehman, Kai Zeng, Zhou Wang, Univ. of Waterloo (Canada)

Over the past years, we have observed an exponential increase in the demand for video services. Video data dominates Internet video traffic and is predicted to increase much faster than other media types in the years to come. Cisco predicts that video data will account for 79% of Internet traffic by 2018 and mobile video will represent two-thirds of all mobile data traffic by 2018. Well accustomed to a variety of multimedia devices, consumers want a flexible digital lifestyle in which high-quality multimedia content follows them wherever they go and on whatever device they use. This imposes significant challenges for managing video traffic efficiently to ensure an acceptable quality-of-experience (QoE) for the end user, as the perceptual quality of video content strongly depends on the properties of the display device and the viewing conditions. Network throughput based video adaptation, without considering a user's QoE, could result in poor video QoE or wastage of bandwidth. Consequently, QoE management under cost constraints is the key to satisfying consumers and video monetization services.

One of the most challenging problems that needs to be addressed to enable video QoE management is the lack of objective video quality assessment (VQA) measures that predict perceptual video QoE based on viewing conditions across multiple devices based. There is a lack of publicly available subject-rated video quality assessment database that investigates the impact on perceptual video quality under the interaction of display device properties, viewing conditions, and video resolution. In this work, we performed a subjective study in order to collect subject-rated data representing the perceptual quality of selected video content in different resolutions viewed on various display devices under varying viewing conditions. A set of raw videos sequences, consisting of 1920x1080 and 1136x640 resolutions, was compressed at five distortion levels to obtain bitstreams compliant to H.264 video compression standard. The decompressed distorted video sequences were scored by subjects under the viewing conditions.

The perceptual quality of video content depends on the sampling density of the signal, the viewing conditions, and the perceptual capability of the observer's visual system. In practice, the subjective evaluation of a given video varies when these factors vary. We propose a full-reference

video QoE measure that provides real-time prediction of the perceptual quality of a video based on human visual system behaviors, video content characteristics (such as spatial and temporal complexity, and video resolution), display device properties (such as screen size, resolution, and brightness), and viewing conditions (such as viewing distance and angle). We compared the performance of the proposed algorithm to the most popular and widely used FR-VQA measures that include Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), Multi-scale Structural Similarity (MS-SSIM), MOtion-based Video Integrity Evaluation(MOVIE), and Video Quality Metric (VQM). Experimental results have shown that the proposed algorithm adapts to the properties of the display devices and changes in the viewing conditions significantly better than the state-of-the-art video quality measures under comparison. Additionally, the proposed video QoE algorithm is considerably faster than the aforementioned perceptual VQA measures and fulfills the need for real-time computation of an accurate perceptual video QoE index and a detailed quality map.

9394-4, Session 1

About subjective evaluation of adaptive video streaming *(Invited Paper)*

Samira Tavakoli, Univ. Politécnica de Madrid (Spain); Kjell E. Brunnström, Acreo Swedish ICT AB (Sweden); Narciso García, Univ. Politécnica de Madrid (Spain)

There are different real world scenarios when the network bandwidth used to deliver video content could be fluctuated so that the playout buffer could fill more slowly or even evacuated, causing video playback interruption until adequate data has been received. Consequently, providing a high-quality playback experience would not always be guaranteed. HTTP adaptive video streaming (HAS) is a relevant advancement to cope the bandwidth fluctuations. Having available segmented video in multiple quality representation (called chunk), it makes it possible to switch the video quality during the playback in order to adapt to current network conditions.

By recent high usage of HAS, various studies have been carried out in this area mainly focused on technical enhancement of adaptation algorithms to potentially supersede the previous proprietary adaptation approaches. However, to optimize the Quality of Experience (QoE) of HAS users, it is fundamental to study the influence of adaptation related parameters evaluated by the human subjects. On the other hand, although various recommendations has been provided by ITU to formalize subjective evaluation methods, the novelties of HAS technology, in relation to the conventional broadcasting services, entails the research for new assessment methodology which fits HAS specification.

Since the adaptation process could span during longer intervals, using typical methodologies like Absolute Category Rating (ACR) mainly designed for evaluation of short sequences may not be appropriate. From another side, in practice the adaptation events do not happen continually but 'during' the video display. Accordingly, assessing the QoE of 'an adaptation event while watching a video' rather than 'adaptation event only' could be more appropriate. In this regard, the difference between quality of the adaptation event and acceptance of the general service would be also highlighted. Other standardized methods such as single stimulus continuous quality evaluation would seem to fit with the objective of HAS QoE evaluation. However the recency and hysteresis effects of the human behavioral response while evaluating the time-varying video quality could lead to provide an unreliable evaluation through this methodology.

Considering above issues, following factors were considered for this study:

1. Switching pattern. Regarding switching behavior among different video quality representations, the perceptual impact of two key aspects was considered to study: the frequency and amplitude (quality level difference) of the switches. This was done by comparing the quality variation in smooth and rapid way while employing short and long chunk size (2 sec and 10 sec).

Different states where the quality adaptation should take place can be summarized as when the bitrate has to be lowered due to restricted network conditions, and when it has to be increased due to better condition. Taking the aforementioned aspects into account, for each of these cases different switching strategies as the possible behavior of the client were considered to apply on the test sequences.

2. Content effect. Several previous works have addressed the content dependency in perception of video quality adaptation. This dependency includes objective characteristics of the content (especially spatial-temporal information), motion patterns and genre of the video. In spite of these findings, the relationship between these factors and perception of quality adaptation is not clarified. To investigate this issue, seven source videos (SRC) in different content type such as movie, sport, documentary, music video and newscast were selected to study. The selected contents were different regarding the amount of motion, scene change and camera record, in addition to spatial-temporal complexity. They were all originally in 1080p and 24/25fps that were played in 25 fps in the test.

3. Evaluation methodology. To investigate the impact of evaluation methodology on observers' assessment, two different experiments (denoted as 'UPM' study) were performed employing a methodology that has been previously developed to evaluate the quality degradation in long test sequences. Using this method, subjects continuously viewed about 6 min videos including the adaptation strategies –providing processed video sequence (PVS) with different duration depending on the adaptation scenario- which were applied in every 8 seconds non-impaired video interval. During the non-impaired segment, subjects rated the quality of the previous impaired segment. To study the influence of audio presence on evaluation of adaptation strategies, one of the experiments was done showing only the video (denoted as 'NoAudio'- 22 observers), and the other in presence of audio (denoted as 'Audio'- 21 observers). The experimental results were eventually compared with our previous experiment (denoted as 'Acreo') where a single stimulus ACR method was employed (this study was published in SPIE-2014).

To produce the adaptive streams (quality levels), the characteristics of the streams which are in practice provided by streaming companies for living room platform were considered. Consequently, in total four streams encoded in 5000, 3000, 1000 and 600 kbps, and in 720p resolution were provided. By considering the QoE of adaptive streams to evaluate, in total 12 adaptation-related test scenarios were considered to study.

Using ACR as voting scale, subjects were asked to assess the quality of PVS by answering two questions: the overall quality of PVS (answer: Excellent to Bad), and if they perceived any change in the quality (answer: increasing, decreasing, no change). At the end of evaluating the 12 PVSs of each video, the subjects were rated the overall quality of whole (6 min) video.

After subjects screening, statistical analysis on the mean opinion scores (MOS) of the experiments was done. The selected results are presented below.

• Correlation between overall quality of SRC and Mean of 12 PVSs' MOS in NoAudio and Audio experiments was 99% and 94% respectively.

• Correlation and P value obtained from ANOVA (?=0.05) calculated for the pair of Audio-No Audio experiments was 93% and 0.78 (no significant difference) respectively, and for the pair of UPM-Acreo was 90% and 0.208 (no significant difference).

• After applying the linear transformation of Acreo's data to UPM's, the difference between studies vanished. The analysis of combined cross-lab data revealed the visible content influence on results specially in switching frequency. However, this effect was not significant considering the switching amplitude.

• The significant influence of audio presence on evaluation of test conditions applied on some of the content was observed.

9394-5, Session 1

A transformation-aware perceptual image metric

Petr Kellnhofer, Max-Planck-Institut für Informatik (Germany); Tobias Ritschel, Max-Planck-Institut für Informatik (Germany) and Univ. des Saarlandes (Germany); Karol Myszkowski, Hans-Peter Seidel, Max-Planck-Institut

für Informatik (Germany)

Predicting human visual perception of image differences has several applications such as compression, rendering, editing and retargeting. Current approaches however, ignore the fact that the human visual system compensates for geometric transformations, e.g. humans are able to see, that an image and a rotated copy are identical. Image metrics however, will report a large, false-positive difference. At the same time, if the transformations become too strong or too spatially incoherent, comparing two images indeed gets increasingly difficult. Between these two extremes, we propose a system to quantify the effect of transformations, not only on the perception of image differences, but also on saliency and motion parallax.

Our approach is based on analysis of the optical flow between two images. These can represent either two views on a static scene or a scene with motion captured at different points in time. First, we use the optical flow as an input to our method and fit local homographies which serve as a description of local transformations. It is assumed, that given a high enough image resolution and a small enough local region, elementary surfaces of most of the objects can be approximated by planes while others are discarded from the analysis. Second, we decompose such local homographies represented by 4x4 matrices into more meaningful signals representing elementary transformations in projective space – translation, rotation, scaling, shearing and perspective change. That allows us to apply the principle of mental transformation, which is best explained for mental rotation. Here, time needed by a human to align two rotated images is linearly dependent on the rotation angle. We extend this idea to all remaining signals and conduct a perceptual experiment to obtain relations of individual transformation magnitudes to the time needed for mental processing. This we use to obtain the total local transformation difficulty map.

The magnitude of local transformation cannot describe global complexity of the transformation field between two images. If one image contains two large regions, e.g. two cars, which are rigidly moved relative to each other it will be easy to understand the transformations even if the translation is large. On the other hand, hundreds of butterflies incoherently flying in all directions will be very hard to match. We utilize the entropy as a novel measure of global transformation complexity. It describes the information quantity in the signal and we relate this amount of information that the human brain has to process to the time that it takes. We compute the entropy of individual transformation signals. This intuitively leads into counting how many different transformations occur in the image. The processing of meaningful signals instead of e.g. optical flow itself is a key to successful entropy computation. A simple global rotation appears as a very incoherent optical flow when interpreted as local translations but can be explained by single rotation angle in our representation.

Putting together the transformation magnitude and the entropy we form a transformation difficulty metric which can be applied to applications where unaligned images are compared by a human. We present an image quality metric for un-aligned images. Local differences of computationally aligned images predicted by conventional metric – DSSIM – are scaled by the predicted difficulty to conduct local mental alignments, resulting in the expected detection threshold elevation. This way, artifacts in significantly transformed (scaled, rotated,...) regions or in regions with a very incoherent flow (e.g. shuffled puzzle pieces) are reported as less significant than others in easier regions. We conducted a preliminary user study where users are presented with pairs of images with randomly introduced distortions and are tasked to find them. We found a statistically significant correlation (Pearson's $r = 0.701$, $p < 0.05$ by t-test) of our difficulty metric with the probability of distortion detection measured in the experiment.

A second application of our metric that we present is a saliency prediction. We extend Itty's saliency model by a transformation saliency component. Unlike in the traditional motion-aware modification of this model we do not limit ourselves to translation directly available in the optical flow but instead we use our perceptually more meaningful elementary transformation signals. We find that our saliency predictor is able to properly detect attention increase for objects deviating from motion pattern of the surrounding. As an example we show rotation of a single puzzle piece in an image taken from a different viewpoint where competing saliency models are distracted by the large global transformation from the viewpoint change while our approach properly understands such change as a single coherent global event and increases the saliency for the local piece that deviates from such assumption.

Other applications of our difficulty metric could be optimization of surveillance systems where camera viewpoints in the presentation are ordered so that it is easier to understand geometrical relations in the scene or re-photography and re-rendering where reproduction of the same camera shot is done so that the difference of alignment is not reduced to minimize a strictly geometrical residual but to minimize the effort required by a viewer to find correct relations between both images.

9394-6, Session 1

Designing a biased specification-based subjective test of image quality *(Invited Paper)*

Amy R. Reibman, AT&T Labs. Research (United States)

No Abstract Available

9394-7, Session 1

Towards a model for personalized prediction of quality of visual experiences

Yi Zhu, Technische Univ. Delft (Netherlands); Ingrid Heynderickx, Eindhoven Univ. of Technology (Netherlands); Judith A. Redi, Technische Univ. Delft (Netherlands)

Increasing online video consumption has raised the need for an objective model that can estimate the Quality of the user's Visual Experience (QoVE) at any point of the video delivery chain, to ensure user satisfaction and increase it when needed. QoVE is a concept typically used to describe the level of delight or annoyance of a user with a video [1]. it is a multifaceted quality concept, not limited to the perceptual quality of a video [1, 2], but involving also personal as well as contextual factors. Up to now, objective approaches to measure QoVE have mainly focused on the perceptual quality, i.e., estimating how annoying artifacts of a video (e.g., blockiness, blur, ringing) are for the user [3]. Such approaches (e.g., MOVIE [4], VQM [5] or VSSIM [6]) have achieved great performance in predicting human perceptual quality, but ignored other dimensions of QoVE (e.g., enjoyment, endurability or involvement) as well as other influencing factors. Up until now, to the best of the authors' knowledge, models estimating the full complexity of QoVE do not exist.

Attempts towards such a model should consider features additional to artifact visibility. Recently, it has been proved that QoVE is affected, among others, by emotions [7] and social context [8]. In fact, three main categories of factors (i.e., system factors, human factors and context factors) potentially influence QoVE [1]. System factors relate to technical properties of a video service (e.g., bitrate, bandwidth, or genre), and they can typically be inferred from the video itself or from the knowledge of the delivery system properties. Human factors refer to the individual characteristics of a user (e.g., age, gender, interest), and as such are independent from the video itself, as in the case of context factors, which refer to the environment, in which a user watches a video (e.g., presence of co-viewers, environmental noise, lighting conditions etc.).

This paper aims to address how some characteristics of the system, human and context factors influence QoVE. More specifically, the following research questions will be answered:

• How do different influencing factors contribute to different QoVE aspects (e.g., perceptual quality, enjoyment or involvement)?

• How to combine this knowledge into a single estimate of QoVE?

To reduce overall complexity, we here will focus on specific aspects of all three classes of influencing factors. System features will be extracted

directly from the videos (e.g., perceptual quality [9] and affective content[10]). User features that will be included are personal interest and demographics (e.g., age, gender, and cultural background). For the context factors we particularly focus on the social context (i.e., the presence/absence of co-viewers). Subsequently, a feature selection procedure will check the contribution of each feature to the different aspects of QoVE, and based on the results of this feature selection a more extended model for QoVE will be proposed.

Input user and context data, as well as video material and quality evaluations needed for this extended model of QoVE will be obtained from the empirical study that we reported before at the SPIE Human Vision and Electronic Imaging XIX conference (2014) [8]. In this study, we investigated the impact of social context (i.e., the physical presence/absence of co-viewers) and user characteristics (for 60 participants) on five aspects of QoVE (i.e.,. enjoyment, endurability, satisfaction, involvement and perceived visual quality) for videos with different bitrates. The results showed that watching videos with friends increased participants' level of enjoyment and enhanced the endurability of the experience. In addition, the low bitrate level obviously had a negative effect on perceptual quality, but did not affect any other aspect of the QoVE.

The proposed model will be reported at the conference and in the final conference paper. We believe the major contributions of this paper are that: 1) we will identify not only features from the video itself, but also features related to the user as being important for QoVE, and 2) the model of QoVE will not only predict the perceptual quality of a video, but also other aspects of QoVE. The latter model May provide insights in how system features may be balance against user and/or contextual features in the eventual QoVE.

9394-8, Session 1

Quality labeled faces in the wild (QLFW): a database for studying face recognition in real-world environments *(Invited Paper)*

Lina J. Karam, Tong Zhu, Arizona State Univ. (United States)

The varying quality of face images is an important challenge that limits the effectiveness of face recognition technology when applied in real-world applications. Existing face image databases do not consider the effect of distortions that commonly occur in real-world environments. This database (QLFW) represents an initial attempt to provide a set of labeled face images spanning the wide range of quality, from no perceived impairment to strong perceived impairment for face detection and face recognition applications.

Types of impairment include JPEG2000 compression, JPEG compression, additive white noise, Gaussian blur and contrast change. Subjective experiments are conducted to assess the perceived visual quality of faces under different levels and types of distortions and also to assess the human recognition performance under the considered distortions. One goal of this work is to enable automated performance evaluation of face recognition technologies in the presence of different types and levels of visual distortions.

This will consequently enable the development of face recognition systems that can operate reliably on real-world visual content in the presence of real-world visual distortions.

Another goal is to enable the development and assessment of visual quality metrics for face images and for face detection and recognition applications.

9394-9, Session 1

Parameterized framework for the analysis of visual quality assessments using crowdsourcing

Anthony Fremuth, Velibor Adzic, Hari Kalva, Florida Atlantic Univ. (United States)

The ability to assess the quality of new multimedia tools and applications relies heavily on the perception of the end user. In order to quantify the perception, subjective tests are required to evaluate the effectiveness of new technologies. However, the standard for subjective user studies requires a highly controlled test environment and is costly in terms of both money and time. To circumvent these issues we are utilizing crowdsourcing platforms such as CrowdFlower and Amazon's Mechanical Turk. The crowdsourcing method has been shown to be cost effective [2] and faster than traditional subjective tests [3], however, the test administrator loses a large portion of control over the testing environment and opens testing to potentially unreliable users. In this paper we are presenting a framework that uses basic and extended set of parameters to achieve increased reliability and allow for in-depth analysis of test results.

With the knowledge of limited environmental control, we gain some knowledge of the subjects and settings by introducing a survey that collects traditional demographic data such as age and gender. In addition, subjects were also asked to provide information about their weekly gaming habits, media consumption, and description of the platform through which the test was taken. The results of this survey were explored in order to determine differences in video ratings. To a lesser extent, this survey was also able to be used to determine unreliable users that were inconsistent or gave several outlandish responses.

The design of our subjective test closely matches methods outlined in previous studies [1][5] by using the Double Stimulus Impairment Scale (DSIS)[6]. DSIS primarily involves the use of a reference video A which is then followed by an impaired video B. These pairs of videos are then directly followed by an impairment scale with ranks 0 through 10 which represent "very annoying" to "imperceptible", respectively. In our subjective tests, the user is first introduced to instructions on how to proceed throughout the test and is given a set of instructions on how to rate each video. Next, the user is exposed to a random sequence of ten DSIS pairs and the answers are recorded along side the initial survey data. In addition, the subjective test was available to be retaken as many times as was desired by the user. Over the course of the subjective tests, we recorded 2543 ratings from 271 unique users.

To characterize the reliability of the subjects, we removed unreliable ratings by the identification of outliers and requiring completion of at least 50% of the test. Beyond this, subjects were only rejected if many his or her survey responses were literally impossible. Due to the structuring of the test, traditional reliability coefficients such as Kendall's W and Spearman's rank-order correlation coefficient could not be determined from the population as a whole. To resolve this, each unique user was compared to the mean video score for each video in the subject's sequence. The resultant reliability coefficients for each user were averaged with the methods outlined in [4] in order to characterize the average user reliability.

Beyond the initial exploration of user reliability, we also successfully modeled the rating data using ANOVA using the survey results as explanatory factors. In accordance with [5], we found that factors such as age, hours of video watched weekly, and favorite video genre had no significant impact on the users ratings. However, we were able to shed light on environmental and subject factors that influence the variability in the video ratings. The auxiliary information was collected using a survey at the beginning of the tests. Survey is part of the test that is available online on our web page [6]. Using Tukey's honest significant difference (HSD) test [7] at the 95% confidence level, for example we concluded that those who game 6 or more hours on a PC or console rate higher, on average, than those who game for 1 to 5 hours a week. Likewise, it is found that there is no significant difference in the mean opinion scores of those who game 6 or more hours as opposed to 0 hours per week and those who game 0 hours as opposed to 1 to 5 hours per week.

With the set of parameters that are representing baseline information about the tests as well as extended set of parameters representing auxiliary information it is possible to create more effective tests to regulate any bias in the subject's ratings. Furthermore, a framework for the analysis of results can be established that models the results according to desired change in parameters. The full paper will present one such model that can be used in broad spectrum of crowdsourcing tests for visual quality assessment, but also has potential application for psychophysical experiments and similar tests.

REFERENCES

[1] Baroncini, V., Ohm, J.-R. and Sullivan, G., "Report of subjective test results of responses to the joint call for proposals (cfp) on video coding technology for high efficiency video coding (HEVC)," JCTVC-A204 (2010).

[2] Hoßfeld, T., Seufert, M., Hirth, M., Zinner, T., Tran-Gia, P. and Schatz, R, "Quantification of YouTube QoE via crowdsourcing," In IEEE International Symposium on Multimedia (ISM), 494-499 (2011).

[3] Figuerola Salas, Ó., Adzic, V., Shah, A. and Kalva, H., "Assessing internet video quality using crowdsourcing," In Proceedings of the 2nd ACM international workshop on Crowdsourcing for Multimedia, 23-28 (2013).

[4] Feldt, L. S. and Charter, R. A., "Averaging internal consistency reliability coefficients," Educational and Psychological Measurement, 66(2), 215-227 (2006).

[5] Hoßfeld, T., Keimel, C., Hirth, M., Gardlo, B., Habigt, J., Diepold, K. and Tran-Gia, P., "Best practices for QoE crowdtesting: QoE assessment with crowdsourcing," IEEE Transactions on Multimedia, vol.16, no.2, 541-558 (2014).

[6] ITU-R., "BT. 500-13: Methodology for the subjective assessment of the quality of television pictures" (2012).

[7] Multimedia Lab Crowdsourcing Subjective Quality Test. (2014). http://mlab.fau.edu/mlab/subjective-crowdflower/.

[8] Hervé, A. and Williams, L., "Tukey's honestly significant difference (HSD) test," In Encyclopedia of Research Design. Thousand Oaks, CA: Sage, 1-5 (2010).

9394-10, Session 1

What do you think of my picture?: Investigating factors of influence in profile images context perception *(Invited Paper)*

Filippo Mazza, Ecole Centrale de Nantes (France); Matthieu Perreira Da Silva, Patrick Le Callet, Univ. de Nantes (France); Ingrid E. J. Heynderickx, Eindhoven University of Technology (Netherlands)

Multimedia quality assessment evolved greatly in last decade. At the beginning only low-level technical features (i.e. resolution) have been taken into account. Over time more complex elements have been taken in consideration for predicting subjective evaluations. These considerations started from the fact that low level features are not enough and human-oriented semantics are requiredcite{Datta2008}. Some years ago aesthetic considerations have been introduced. Aesthetic analysis adopted mainly elements taken from photography literature, as contrast, composition and colors.

Studies on multimedia quality paved the way to a series of related studies. Those too investigate the relationship between subjective evaluations and multimedia characteristics. This is the case for example for studies on visual and social interestingness of images cite{Hsieh2014}, online image diffusion cite{Totti2014} and image memorability cite{Isola2011}. While low and high level features can have an impact on overall quality, they also impact human perception regarding the message that is conveyed by the image itself. Perception is deeply biased by the interpretation of the message conveyed by the image. This is especially important in multimedia where people are present, as our brain unconsciously processes informations to get a first idea of the depicted person. Deeper researches should be made on these considerations. Some steps have been made in this direction, encompassing broader concept called "image psychology" cite{Fedorovskaya2013}. In fact, cognitive processes related to multimedia content may influence overall perception, even at an unconscious level.

In our previous work cite{Mazza2014a} we digged into this new concept stressing the importance of focusing on social biases conveyed inside images by the content itself. In particular we focused on how different portrait typologies influence candidates' choice in a resumes selection process. In this work we continue our analysis on high level portrait images factors of influence digging into which elements in particular influence context perception. By context we refer here to the overall message that the picture is believed to convey. To make an example, we expect a portrait of a person in business suit more likely to be perceived as related to a working context and more likely to be preferred as professional portrait.

To this extent we gathered subjective evaluations asking participants to categorize portrait images respect to perceived context. Images have been chosen from known image databases related to research or popular image services as Flickr, paying attention to pictures' license (i.e. Creative Commons). Proposed categories were images for purposes related to friends (A), to work (B) and to romantic affairs (C). We made this choice considering the most common categories of social networks that exist nowadays (i.e. Facebook, Linkedin and Meetic). This task is quite common considering the huge adoption of profile images online.

As we needed a large amounts of subjective evaluations, we adopted crowdsourcing. This technique allows to gather fastly a large number of subjective evaluations and has been already adopted with success for different research purposes cite{Ribeiro2011,Keimel2012}. Participants have been gathered through dedicated online platforms and worked remotely to our task from all over the world. The experiment run on a dedicated framework we developed. We asked both a first and a second choice in category selected for each image, to have a better understanding of preferences. Between proposed images we put also some pictures without any person depicted; experiment instructions clearly asked to mark these images as non-profile pictures. The purpose of this request was to spot eventual outliers that either did not understand instructions or that did not accomplish carefully our task.

Influent factors analysis has been carried out exploiting machine learning. We adopted different supervised learning algorithms to predict subjective categorization preferences, as Fisher Discriminant Analysis (FDA), Classification Trees and Neural Networks. Differences in prediction accuracy while pruning selectively adopted features outlined the importance of each factor in our analysis. Features adopted in algorithms were both low and high level features. While many of those have been taken from literature, as in cite{Datta2006} and cite{Totti2014} for the low level and high level respectively, we introduced also new features we supposed to be important. Added features are related mostly to content interpretation, like subject dress typology.

These elements related to cognition are usually very easy to be evaluated by humans but very hard to be extracted through computer vision algorithms. For this reason we preferred to exploit the power of distributed human intelligence with crowdsourcing, asking to assess high level features online. This approach differs from other researches adopting instead computer vision, and allowed us to consider content cognition in analysis.

Our results underline that especially some of these features influence prediction accuracy, suggesting that low level features marginally influence context perception.

9394-11, Session 2

Texture, illumination, and material perception

Sylvia C. Pont, Andrea Jacoba van Doorn, Maarten W. A. Wijntjes, Jan J. Koenderink, Technische Univ. Delft (Netherlands)

In this paper we will present an overview of our research into perception and biologically inspired modeling of illumination (flow) from 3D textures and the influence of roughness and illumination on material perception. Here 3D texture is defined as an image of an illuminated rough surface.

In a series of theoretical and empirical papers we studied how we can estimate the illumination orientation from 3D textures of matte, globally flat samples. We found that the orientation can be estimated well by humans and computers using a gradient based approach. This approach makes use of the dipole-like structures in 3D textures that are the results of illumination of bumps / throughs.

For 3D objects, the local illumination direction varies over the object, resulting in illumination flow over the object. We have shown that the

illumination flow can be used to do shape and illumination inferences. Moreover, in perception experiments it was found that adding 3D texture to a matte spherical object helps to judge the direction and diffuseness of its illumination.

The illumination and surface shape characteristics again interact with the perception of the material. We have shown that they can even interact such that a matte surface looks shiny. Currently we are running experiments into perception of real rough surfaces for surface shape and illumination variations.

9394-12, Session 2

Effects of contrast adjustment on visual gloss of natural textures

Jing Wang, Thrasyvoulos N. Pappas, Northwestern Univ. (United States)

Visual gloss is an important attribute for texture perception and in the recognition of the natural of materials. As one of the surface optical characteristics, gloss depends on both intrinsic properties (material surface geometry) and extrinsic condition (illumination direction, viewing direction etc.). There are known works studying the relation between perceived gloss and the statistics of the surface geometry [1,2,3,4]. However, purely photometric statistics like skewness could not provide diagnostic information about gloss in complex environment [3,4]. Therefore it is necessary to investigate the relation of visual gloss with more perceptual attributes. Similar to gloss, contrast perception of natural images also relies on the structure of texture and lighting. Thus there should be some close interaction between contrast and perceived gloss. And it is the task of this research to investigate them.

In this paper, we analyze the effect of contrast adjustment on subjective gloss under diverse illumination angles. The manipulation of local band-limited contrast is based on a novel yet simple s-curve transformation. We demonstrate that the contrast enhancement will add apparent gloss to glossy material surfaces and compensate for the gloss difference caused by different illumination angles. The analysis is based on subjective experiments with stimuli from not only natural textures but also synthesized Lambertian surface.

METHOD:

Contrast perception is closely related to the spatial frequency content [5,6]. In our work, each image is first divided into sub-bands by cosine log filter banks to get band-limited images. For each bandpass-filtered image, we introduce an s-curve transformation to adjust the local band-limited contrast.

The specific s-transformation is developed from the ?-curve proposed by Wijntjes and Pont in [4]. According to Wijntjes and Pont, geometrical property of Lambertian surface can be changed by ? transforming the luminance histogram. However we discovered that simple photometric ? transformation does not fit complex geometry of real-world texture. As shown in Figure 1, s-curve is a conjunctive function of ?-curve and its counter-diagonal transformed form. Equation (1) illustrates the transformation function for input value I in each sub-band k

Different from ?-transformation, which controls the value of skewness of luminance histogram, s-curve stretches the spread of histogram in each sub-band such that we could get locally contrast-enhanced image. Figure 2 shows the comparison results between the two transformations on the same input natural texture.

The stretch value 1/? of the S-curve is the sole control parameter of our contrast modification experiments. The larger 1/? is, the more spread the luminance histogram, and the more enhanced contrast we could get.

EXPERIMENTS

1. Dataset description

In our experiments we used original textures from the CUReT database [7]. The images were converted to gray-scale and 220 ?220 pixels square images were extracted from the center of the textures. For each texture image, we have different illumination directions with fixed frontal viewing direction. In addition to CUReT database, we also test on Lambertian surfaces by generating Brownian noise images with a power spectrum of 1/f^2 and a random phase spectrum.

2. Experiment setups

We carried on two subjective experiments. The preliminary experiment aims to get the relation between local contrast and visual gloss. For each stimulus, we showed the subjects an original and a series of s-transformed images with varying stretch values in random order and asked them to rearrange the images from the glossy to matte. As a comparison, we also evaluate subjects' performance on ?-transformed images.

In the second session, each texture owns three different illumination directions with the same frontal viewing direction. To determine the amount of contrast increase needed to equalize the perceived gloss of textures under different illumination condition, we placed the subjects with two images of the same texture side by side, one original in oblique illumination direction with fixed contrast and one contrast-adjustable image in frontal illumination associated with a sliding control. Subjects were required to change the geometry of the test image using the sliding bar until they perceive that the two images have the same gloss.

3. Preliminary results

15 observers participated in the first session subjective tests on ranging the images according to the visual gloss to 10 textures from CUReT dataset. In Figure 3, we use Pearson's linear correlation to evaluate the subjective performance. The correlation coefficients value between stretch value 1/? adjusting local contrast and subjective gloss ranges from 0.8 to 0.9 in s-curve transformation, indicating a strong positive correlation between local contrast and gloss. In comparison, ?-curve transformation for skewness does not present direct relation to the subjective gloss ranking.

In the full paper we will also disclose experimental results on synthesized Lambertian surfaces. The goal of the second experiments is to test whether contrast enhancement could compensate for the visual difference in gloss caused by different lighting angles. However, equal subjective gloss does not necessarily mean equal contrast perception. The result will give us more insights in their iteractions.

NOVELTY:

We focus on analyzing the effect of contrast adjustment on gloss perception of natural textures. Compared with simple statistics on luminance histogram, contrast conveys more structural properties with gloss in human perception dimensions. The manipulation on the textures is based on a novel yet simple s-curve transformation. By carrying on subjective tests, this paper brings more insights in the mutual effect between contrast and subjective gloss.

REFERENCE:

[1] Motoyoshi, Isamu, Shin'ya Nishida, Lavanya Sharan, and Edward H. Adelson. "Image statistics and the perception of surface qualities." Nature 447, no. 7141 (2007): 206-209.

[2] Ho, Yun-Xian, Michael S. Landy, and Laurence T. Maloney. "Conjoint measurement of gloss and surface texture." Psychological Science 19, no. 2 (2008): 196-204.

[3] Anderson, Barton L., and Juno Kim. "Image statistics do not explain the perception of gloss and lightness." Journal of vision 9, no. 11 (2009): 10.

[4] Wijntjes, Maarten WA, and Sylvia C. Pont. "Illusory gloss on Lambertian surfaces." Journal of Vision 10.9 (2010): 13.

[5] Peli, Eli. "Contrast in complex images." JOSA A 7, no. 10 (1990): 2032-2040.

[6] Haun, Andrew M., and Eli Peli. "Perceived contrast in complex images." Journal of vision 13, no. 13 (2013): 3.

[7] "CUReT: Columbia-Utrecht Reflectance and Texture Database." [Online]. Available: www1.cs.columbia.edu/CAVE/software/curet.

9394-13, Session 2

A subjective study and an objective metric to quantify the granularity level of textures

Mahesh M. Subedar, Lina J. Karam, Arizona State Univ. (United States)

Texture granularity is an important attribute that quantifies the size of the primitives in texture images. A texture primitive is defined as the smallest repetitive object that one can recognize in a texture. Texture granularity can be used in image retrieval, texture segmentation, texture synthesis and rate control for image compression applications.

In this paper we are presenting a texture database GranTEX along with the MOS scores to compare the granularity level of the textures. The GranTEX database consists of 30 textures with low (10 textures), medium (10 textures), and high (10 textures) levels of granularity. A subjective study is conducted using the GranTEX database to calculate the MOS scores for the granularity level of the textures in images. There were 12 subjects who participated in these experiments. The subjects were asked to judge the granularity level of the textures and rate it as either low, medium or high.

This paper also presents a granularity metric that automatically provides a measure of the perceived granularity level of a texture. The texture image is first decomposed into 5 levels of undecimated wavelet transform with three subbands (low-low or LL, horizontal high or HH, and vertical high or VH) at each level. At each level, the LL (LL) subband is obtained by applying a low-pass filter in both the horizontal and vertical directions; the HH and VH subbands are obtained by applying a high-pass filter in the horizontal direction and vertical direction, respectively. The size of the primitive is calculated by identifying the peaks and measuring the distances between peaks in the HH and VH subbands. The peaks are detected by calculating the slope of the intensity transitions subject to a minimum height requirement for the peaks in the signal. The average size of the primitives is estimated along each row in the HH subband and along each column in the VH subband by computing the distance between the detected peaks in the signal and averaging them across rows and columns, respectively. Then we find the size of the primitives at the level where visually important structural details are maintained. For this we use SSIM index between LL subband and original image. If the SSIM index is above the empirically calculated threshold then the LL subband still maintains the important structural details and next level is analyzed. If the SSIM index is below the threshold then the calculated average primitive sizes in HH and VH band are considered.

The performance results of the proposed objective granularity metric for the GranTEX database in terms of the Pearson and Spearman correlation coefficients are presented. These results show that the proposed metric achieves a high correlation with the subjective scores.

9394-14, Session 2

Texture synthesis models and material perception in the visual periphery

Benjamin Balas, North Dakota State Univ. (United States)

Algorithms for texture synthesis are increasingly being used as a tool for describing the perceptual representation of texture appearance in the human visual system (Balas, 2006). Aside from their obvious utility as a feature vocabulary for texture perception, these models have recently been used as a means of explaining a broader class of visual processes including visual crowding (Balas, Nakano & Rosenholtz, 2009), visual search (Rosenholtz et al, 2011), and aspects of visual attention (Rosenholtz, Huang, & Ehinger, 2012). A unifying conjecture underlying a great deal of recent research is that human vision is subject to a number of information bottlenecks that limit the fidelity with which visual appearance can be encoded. Texture representations offer a useful strategy for coping with these bottlenecks insofar as they make it possible to compress substantial information about visual appearance so long as the location of specific features is not critical. Obviously texture perception itself relies relatively little on positional information, and human performance in many of the tasks listed above also seems to be consistent with the application of texture representations. Texture synthesis algorithms thus offer a rich set of candidate models that are sufficiently expressive to cope with the complexity of natural stimuli as well as a means of developing psychophysical experiments that make it possible to test the limits of texture representations as a means of accounting for human performance in a wide range of visual tasks.

Presently, we chose to investigate the extent to which human material perception in the visual periphery could be accounted for by the representation of visual textures by the descriptors available in non-parametric and parametric models of texture synthesis. We asked observers to complete a 4AFC material categorization task using images drawn from the Flickr Materials Database (Sharan et al., 2013) in two distinct scenarios: (1) Observers were asked to classify images as depicting either water, metal, wood, or stone when viewed at fixation and two locations in the visual periphery (~10 and 20 degrees). (2) Observers were asked to classify synthetic versions of the original images presented at fixation. Synthetic images for this experiment were created using a non-parametric "texture quilting" method (Efros & Freeman, 2001) as well as the Portilla-Simoncelli algorithm (Portilla & Simoncelli, 2000). In each case, we calculated confusion matrices describing both how often observers correctly classified each material as well as which errors they tended to make (e.g. confusing water for metal). If the texture models considered here were adequate for describing performance in the visual periphery, the confusion matrices for viewing the original textures in the periphery should strongly resemble those obtained from viewing synthetic images at fixation. Instead, we found substantial disagreement between these sets of confusion matrices, suggesting that these models are not an adequate description of how materials properties are represented in the visual periphery. We discuss these results in the context of alternative proposals for the representation of material properties and suggest methods for further exploring the vocabulary of material perception using rich representations of texture appearance.

9394-15, Session 3

Feature maps driven no-reference image quality prediction of authentically distorted images *(Invited Paper)*

Deepti Ghadiyaram, The University of Texas at Austin (United States); Alan Bovik, The Univ. of Texas at Austin (United States)

No Abstract Available

9394-16, Session 3

Combining full-reference image visual quality metrics by neural network *(Invited Paper)*

Vladimir V Lukin, National Aerospace University (Ukraine); Nikolay N Ponomarenko, Oleg I. Ieremeiev, National Aerospace Univ. (Ukraine); Karen O Egiazarian, Jaakko T Astola, Tampere University of Technology (Finland)

A task of assessing full-reference visual quality of images is gaining more popularity nowadays. Availability of tools able to estimate visual differences between a pair of images adequately allows improving performance for many applications. In particular, this relates to lossy image compression, watermarking, image denoising, etc.

Recently, a lot of full-reference visual quality metrics (indices) has been designed. Most of them are based on taking account some heuristics

concerning human visual system (HVS). It is practically impossible to prove optimality of a given metric for a given application analytically because HVS is still studied not well enough. Due to this, metric verification is carried out using large databases of test images. Many observers (volunteers) are attracted to obtain averaged estimates of image visual quality (mean opinion score - MOS). Correlation between the obtained array of MOS and the corresponding array of given metric values allows characterizing correspondence of a considered metric to HVS. Larger correlation (Spearman and Kendall rank order correlations are often used) approaching to unity shows better correspondence between a metric and human perception of image quality. For the largest existing database TID2013 intended for metric verification, Spearman correlation is about 0.85 for the best existing HVS-metrics. Therefore, ways to improve the metrics are intensively studied now.

One simple way to improve efficiency of assessing visual quality of images is to combine several metrics designed independently by different groups of researchers with obtaining some "integrated" metric. For such combining under condition of availability of enough data for learning, it is possible to use different tools as neural networks, support vector machines, data clustering approaches, etc. Our work addresses possibility of using neural networks for the aforementioned purpose.

Neural networks are powerful tools for solving similar tasks but sometimes they are not reliable enough. Quality of neural network combiner of several inputs depends upon several factors including quality of learning data set, size of learning sample, number of inputs, network type, configuration and parameters (as number of layers and number of neurons in each layer for perceptron, activation function type, etc.). It is desired to avoid under- and overlearning as well as to provide getting into global extremum of goal function. It is also not possible to expect that the learned network will perform well for data having other range than those ones used at learning stage.

As leaning data, we used metric sets for images of the database TID2013 that are used as network inputs and that have been calculated earlier. As network outputs the values of MOS obtained for TID2013 have been used. 1500 out of 3000 images of the database TID2013 selected randomly have been used at learning stage whilst other 1500 images have been exploited for assessing quality of neural network based HVS-metric. The use of this large database that contains images with 24 types of distortions and 5 levels of distortions allows to hope that learning sample correspond well to most practical situations of image distortions.

The first stage of NN-metric design is to choose NN inputs. To select known metrics that can serve as inputs, statistical analysis has been carried out. For each type of distortions present in TID2013, Spearman correlation for more than 20 metrics have been calculated between metric values for a given type of distortion and metric values for images corrupted by additive white Gaussian noise (distortion type #1 in the database). Then, for each type of distortion, those metrics have been left for which this Spearman correlation is the highest or close to the highest (this means that these metrics are adequate for a given type of distortions). Finally, 6 metrics have been selected which "cover" well all types of distortions. This means that one of these six metrics is the best for ant type of distortions: FSIMc, PSNR-HMA, PSNR-HVS, SFF, SR-SIM, VIF. The metric values have been used as inputs without any preliminary processing as, e.g., normalization.

We have analyzed several different types of neural networks and their topology. The obtained results show that NN type and its configuration influence the obtained data less than selection of input metrics. The results for most configurations and types of neural networks are in the range 0.915..0.925 – these are values of Spearman correlation between NN output and MOS for 1500 images of verification sample. Meanwhile, exclusion of one metric from six used ones (e.g., VIF) leads to essential drop in efficiency – correlation reduces to 0.90.

As the result of NN learning, for the best configuration the Spearman correlation between NN output and MOS for the verification set of database TID2013 reaches 0.93 which is considerably better than for any particular metric set as input (FSIMc is the best among them). Analysis of the designed metric efficiency is carried out, its advantages and drawbacks are shown.

9394-17, Session 3

Geometrical and statistical properties of vision models obtained via maximum differentiation *(Invited Paper)*

Jesus Malo, Universitat de Valencia (Spain); Eero Somincelli, New York Univ. (United States)

Realistic models of visual information processing in the brain should have a multi-layer structure mimicking the sequence of stages in the neural pathway (Mante et al. 2005; Freeman and Simoncelli 2011; Coen and Schwartz 2013). This should be the case too in image quality assessment where it makes sense to analyze factors of increasing complexity (e.g. luminance, contrast, and structure) in different processing stages (Wang et al. 2004; Seshadrinathan 2008; Laparra et al. 2010). This required multi-layer structure is consistent with the emergence of interesting features when training deep networks (Salakhutdinov and Hinton 2012; Malo 2013) and multi-stage generative models (Hyvarinen et al. 2009; Gutmann 2012) with natural images. In all the above cases, each layer can be understood as a cascade of a linear+nonlinear operation. Criteria to train these structures is important for the vision science, the image processing, and the machine learning communities.

In this work we present a simple psychophysical method to fit the these models. The method consist of relating the free parameters of the model with different instances of image metrics (Malo et al. 2006; Laparra et al. 2010), and deciding among these using the straightforward Maximum Differentiation (MAD) psychophysical method (Wang and Simoncelli 2008). Here, we apply this idea to fit multi-layer models in which the stages have either statistical or perceptual roots, which makes sense under the Efficient Coding hypothesis (Lyu and Simoncelli 2009; Malo and Laparra 2010).

The geometrical properties of the models obtained through multiple MAD competitions are analyzed in terms of their ability to reproduce subjective distortion (non-negligible perceptual distance) and to generate visual metamers (negligible perceptual distance). The statistical behavior of the models is analyzed in terms of the redundancy reduction (mutual information among the coefficients). Preliminary results (only two observers) lead to models with distance estimation well aligned to human opinion, sensible metamer generation (in frequency and masking terms), and redundancy reduction along the stages.

9394-18, Session 3

Relations between models of local masking in natural images and perceptual quality *(Invited Paper)*

Md Mushfiqul Alam, Pranita Patil, Martin T. Hagan, Damon M Chandler, Oklahoma State Univ. (United States)

Image quality assessment algorithms which mimic the actions of the human visual system often harness the effects of visual masking phenomena where an image acting as a mask hides different processing artifacts (targets). The masking algorithms used in such image quality assessment techniques have been commonly derived from the studies of simple unnatural masks (such as, sine-wave gratings) rather than complex natural masks. Furthermore, it is not clear for which distortion types and distortion levels masking affects the human judgments of image quality. Very recently, a large database of local masking has been developed where the contrast thresholds for detecting vertically oriented 3.7 c/deg log-Gabor noise targets have been measured on a total of 1080 natural image patches, resulting local masking maps of 30 natural images [Alam, Vilankar, Field, and Chandler, JOV, 2014]. Such a database has the potential to facilitate the development of improved local masking models for natural images. In this paper we describe three different approaches for predicting the local detection thresholds, namely, local feature based model, Watson and Solomon masking model [Watson and Solomon, JOSA, 1997], and a convolutional neural network based model. Preliminarily, we evaluate the performance of the three models at their

simplest forms. In addition, we describe the results of predicting the quality scores of the high-quality images from the CSIQ image quality database [Larson and Chandler, JEI, 2010] using the local masking map obtained from the local masking database.

9394-19, Session 3

Building structural similarity database for metric learning *(Invited Paper)*

Guoxin Jin, Thrasyvoulos N. Pappas, Northwestern Univ. (United States)

Human vision system has great advantage on recognizing the structural similarity between two images, regardless the existence of distortions thereof. Although the structural similarity metrics have played important roles in image compression, image retrieval and image segmentation applications, more efforts should be devoted to the research for further developing the objective structural similarity and quality metrics [1]. In the literature, most of the structural similarity metrics compute and combine visually significant features in heuristic means [2, 3, 4, 5, 6, 7]. Recent research shows the necessaries of considering the aspect of metric learning [8]. In order to perform supervised learning for the metrics, building a pair-wised database, where each entry is consisted of two images and a quantitative measurement (score) indicating the similarity between them, will be very helpful.

However, there is no such database available for testing the structural similarity metrics. On one hand, the databases used in state-of-the-art metric learning approaches [9, 10, 11] are seldom focusing on structural similarity. Rather, they are designed for face recognition [12], hand-writing [13], voice recognition [14], etc. One database related to structural similarity is CUReT [15]. However, the score in those databases are generally binary. In other words, for each image in the database, a label is assigned, such that when the similarity between two images are under consideration, the labels of them can only indicate whether or not do they belong to the same class (similar) or to the different class (dissimilar). On the other hand, the widely used visual quality databases, such as LIVE [16] and TID13 [17], considered only distortions on the entire image. If one seeking to test a block based metric for structural similarity, those quality databases are not appropriate.

In this paper, we will introduce an approach to build pair-wise database using Matched-texture coding (MTC) [18, 19] subjective coding mode. In the MTC subjective coding mode, a score between 0 to 10, where 0 indicates dissimilar and 10 indicates same, is given by the subjects to each pair of image blocks. Here the pair-wised data collected will be used to train the structural similarity metrics [8] in MTC to improve the coding efficiency. However, the database can be used to any other structural similarity researches.

In MTC, an image is partitioned and coded in raster scan order. The image block to be encoded at a particular stage is called the target block. The target is coded by either pointing to a previously coded block or by JPEG as the baseline coder. For the former case, MTC will select several potential candidate blocks from the coded and reconstructed region. Then the lighting of the candidates are modified [19] in accordance with the lighting of the target. For the subjective coding mode, the candidates and the target will be shown to the subjects. Before doing that, however, one must takes the masking effect caused by the surrounding pixel context into consideration. We embed the lighting corrected candidate blocks into the context of the target block with blending algorithm [20], in order to conceal spatial discontinuity. For each target, multiple candidates may be available for the subjects in the same time. When a subject points (using the cursor) to a particular candidate, the embedded version of this candidate and the original target embedded within the same context are presented side-by-side. The subjects will evaluate the structural similarity between the two embedded versions and give a score to each candidate.

The subjective evaluation of the structural similarity behaves like a monotonic function proportional to the similarity of the pair, only when the actual similarity is high enough [1]. However, given two pairs of dissimilar images, it is hard to ask the subjects to evaluate which pair is more dissimilar than the other. This suggests that the structural dissimilarity is not a monotonic function. Luckily, it is not common in the practical application that the dissimilarity is considered. Therefore in our database construction process, the question exposed to the subjects is 'What is the degree of similarity between the images shown below?' The subjects are restricted to select four verbal scales: highest (10), high (9), good (8) , and bad(0). Moreover, it is important to include the target itself as a potential candidates in the test. If a subject gives low score (good (8) or bad (0)) to the target itself, all the scores of the candidates of this target will be discard for untruthfulness.

The data collected during the subjective coding mode contain three typical distortions, which may not affect the perception similarity but could cause significant differences in the feature space; And it is the task of the objective metric to deal them accordingly. They are: (1) lighting distortions introduced by imperfect lighting correction of the candidates, (2) blocking artifacts resulted from of the baseline coder (JPEG), and (3) wrapping distortions caused by blending during the context embedding.

Although the subjective tests are running on a web-server [21] which can be accessed remotely through world-wide-web and collect data from any subject, the data collected from IP address out of our lab domain are discarded for now. The main reason is that the tests are still under development and the web-server is set up only for experimental uses that higher amount of accessing is unstable. We will make this web-server mature before the end of 2014. Moreover, there is no instructions for the common users of viewing conditions at this stage. There are 6 subjects by now. Five out of six subjects are experts in the image processing area. ViewSonic VX2250wm-LED (21.5 inch, 1920 ? 1080) and Macbook Pro monitors without retina display (13.3-inch 1280 ? 800) were used for the test. The viewing distance of the test is around 3H, where H is the height of the LCD monitor.

As a result, for each entry in the database, there are a pair of images, one target and one candidate. A score, either 10, 9, 8, or 0, is associated with the pair. The data can be collected quickly by the proposed approach. Given an input image with resolution 1024 ? 1024 and the block size 32 ? 32, if the average MTC successful replacement rate is 20% and average 4 potential candidates are selected for each target, then about 800 entries can be updated to the database by coding this image.

9394-41, Session PTues

Do curved displays make for a more pleasant experience?

Nooree Na, Kyeong-Ah Jeong, Hyeon-Jeong Suk, KAIST (Korea, Republic of)

This study investigates the optimum radius of curvature (R, mm)of a monitor. For the examination, a bendable monitor prototype was used to enable subjects to adjust the display radius manually. Each subject was instructed to search for an optimal radius with respect to individual preference and comfort. Six different themes were applied for the display content. Also, each subject reported the radius when distortion occurred. We found that 600 R to 700 R is optimum for a 23-inch display, and 700 R to 800 R is appropriate for 27-inch. However, when the curvature monitor was smaller than 600 R, a majority reported visual distortion regardless of the monitor size. Moreover, in a validation test, subjects read the texts faster on the curved display than on the flat display.

9394-42, Session PTues

The importance of accurate convergence in addressing stereoscopic visual fatigue

Christopher A. Mayhew, Stephen M. Bier, Vision III Imaging, Inc. (United States)

Visual fatigue (asthenopia) continues to be a problem in extended viewing of stereoscopic imagery. The Authors submit that poor or badly converged imagery contributes to this problem. In 2013, the Authors reported to the SPIE that a surprisingly high number of 3D feature films sampled were post

produced with poor attention to convergence accuracy. The films were released as stereoscopic blu-rays designed for television viewing. Selection and placement of the convergence point can be an "artistic" call on the part of the editor. However, a close examination of the films sampled revealed that in most cases the intended point of convergence was simply missed. The errors maybe due to the fact that some stereoscopic editing tools lack the means to allow a user to set a pixel accurate point of convergence. Compounding this further is the fact that a large number of stereoscopic editors do not believe in, or may not be aware of the importance of having an exact point of convergence.

The Authors contend that having a pixel accurate point of convergence at the start of any given stereoscopic scene will significantly improve the viewer's ability to fuse left/right images quickly and because fatigue is cumulative, extended viewing is improved.

To test this premise, the Authors conducted a series of controlled screenings with twenty participants with ages ranging from 18 to 60 years. A series of static random dot stereoscopic (RDS) images with differing points of convergence were presented to participants. The amount of time each viewer required to fuse and identify the object[s] within the imagery was measured.

A second series of full motion stereoscopic video imagery was also tested. The films subject matter consisted of three people talking in an exterior wooded area. Four identical 10-minute video films were assembled that included between 4 to 6 edits per minute. The first film was post produced with pixel accurate convergence. The other three films were variations of the first with differing points of convergence. Each participant viewed the films randomly over a four-day period. Each participant completed a questionnaire after each viewing. This paper details the test process and reports on its findings the conclusion of which supports the Author's original premise.

9394-43, Session PTues

Improvement in perception of image sharpness through the addition of noise and its relationship with memory texture

Xiazi Wan, Hiroyuki Kobayashi, Naokazu Aoki, Chiba Univ. (Japan)

Currently, some professional photographers intentionally add granular noise to images captured using digital cameras to improve the texture quality of digital photographs. This is because granular noise resembles the graininess of a silver-halide film or silver-halide photograph[1].

Some studies have suggested the possibility of improving image quality through the addition of noise[2]. Other studies have suggested improvements in likeability by noise addition[3]. In addition, there are studies that highlight the relationship between noise addition and memory of texture[4], with the authors of one study proposing that textures can be recalled in association with familiar objects, in a similar manner as memory color[5].

Further, Kurihara et al. investigated the effects of noise on the sharpness of an image by adding white noise to one- (1D) and two-dimensional (2D) single-frequency sinusoidal gratings as stimuli and comparing them with rectangular and checkerboard patterns[6]. The results of his experiment indicate that the addition of noise improves the sharpness more effectively when low-frequency stimuli, rather than high-frequency stimuli, are used. Addition of noise is not particularly effective in images that contain clear edges, such as rectangular and checkerboard patterns. This effect is more pronounced with 2D, rather than 1D patterns.

This study further develops the ideas of the preceding study by evaluating natural color images as opposed to black-and-white, single-frequency patterns. We aim to verify if the results observed for simple patterns also apply to color images. In addition, we further elucidate the mechanisms involved in increasing sharpness through the addition of noise by discussing the relationship between the perception of image sharpness and memory texture.

Because sinusoidal gratings can be regarded as blurred rectangular or checkerboard patterns, we used focused original images and their blurred images as stimuli, and then added white noise to these images.

We used rice crackers and cookies because they are familiar items that have various textures. Photographs were captured under diffuse light. All images were adjusted to the actual size of the objects.

We measured the power spectra of the sample images and chose four out from them. The buckwheat cookie (d) had more high-frequency components than the round rice cracker (a) and the square cookie (c). The deep-fried rice cracker (b) did not have many high-frequency components, but it did have many lines and a distinct texture.

The stimuli used for the evaluation were prepared using the following process: first, the original images were blurred at two different levels; next, white noise was added at two different levels (RMS10 and 20).

Including images without noise, there were nine images to be evaluated for each sample; therefore, in total, there were 36 images.

The evaluation was carried out under a natural color evaluation lamp (Toshiba Corp., Tokyo, Japan; 5000K, 700lx), at the distance of distinct vision (25-30cm in the normal eye).

Images were evaluated using a normalized ranking method. The research participants were asked to rearrange images in the following manner: "A sharp image implies that an image appears to be clear, and in focus. Please sort these images in descending order of sharpness for each sample at the distance of distinct vision." Seventy-eight subjects, between the ages of 10 and 60 years, participated in this experiment. Based on their answers, we calculated the sharpness values.

Overall, the effect of noise is more evident in strongly blurred images than in weakly blurred images. In particular, for the flat, round rice cracker (a) and the square cookie (c), the original low-frequency and flat-textured images, sharpness increases with noise level to the extent that the blurred images appear similar to the non-blurred images. On the other hand, for the buckwheat cookie (d), the high-frequency image, noise was not as effective. For image (b), which has many lines and a distinct texture but a low-frequency image, noise had only a minor effect in contrast to the effects seen in (a) and (c). It is possible that the addition of noise erased the lines on the surface of the deep-fried rice cracker (b) resulting in a loss of texture.

The results show that the effect of noise in improving image sharpness is frequency dependent; in other words, its effects are more evident in blurred images than in sharp images, and in strongly blurred images more than weakly blurred images. This is consistent with the results from previous studies.

In terms of the relationship between the perception of image sharpness and memory texture, according to the study on memory texture, there is a texture that is appropriate for each object (which can be divided into "white noise" or "1/f noise"). We hypothesized that adding noise to a blurred image causes the subject to recall the texture of that object from memory and to simultaneously perceive sharpness. We tested the hypothesis by adding white noise and 1/f noise to a variety of objects to see if noise that has a larger effect on sharpness improvement matches the noise of memory texture. Since previous studies did not discuss the individual difference, this study discusses the relationship between the individual differences in the perception of image sharpness and that in memory texture.

References

1) T. Kurihara, Y. Manabe, N. Aoki, H. Kobayashi, "Digital image improvement by adding noise: An example by a professional photographer," J. Imaging Sci. Technol., 55, 030503(2011)

2) B. W. Keelan, "Handbook of image quality: Characterization and Prediction," Marcel Dekker Inc., (2002)

3)?Y. Kashibuchi, N. Aoki, M. Inui, H. Kobayashi, "Improvement of description in digital print by adding noise," J. Soc. Photogr. Sci. Technol. Japan, 66(5), 471–480(2003)

4)?H. Kobayashi, Y. Zhao, N. Aoki, "Memory Texture as a mechanism of improvement in preference by adding noise," 2014 Electronic Imaging, 9014-13(2014)

5)?N. Takesue, N. Aoki, H. Kobayashi, "Memory texture and its relationship with object's texture," PPIC08, June 27, 2008(Tokyo)

6)?T. Kurihara, N. Aoki, H. Kobayashi, "Analysis of sharpness increase and decrease by image noise," J. Imaging Sci. Technol., 55, 030504(2011)

9394-44, Session PTues

Depth image enhancement using perceptual texture priors

Duhyeon Bang, Hyunjung Shim, Yonsei Univ. (Korea, Republic of)

Due to the rapid advance of 3D imaging technology, 3D data and its processing algorithms receive increasing attentions in recent years. To acquire the 3D data of a real scene, we can use a depth camera, providing a depth image of the scene in real time. However, because of the limited power consumption, the depth camera presents its inherent limitations: the low signal-to-noise ratio, low precision and poor distance resolution. As a result, the noisy raw data from the depth camera is inappropriate to serve the high quality 3D data. To subside the depth noise, the smoothness prior is often employed for depth image processing, but it discards the geometric details so to yield the poor distance resolution. Unfortunately, the geometric details are important to achieve the realism in 3D contents. To this end, we propose a new depth image enhancement method for recovering the surface details inspired by the analysis in human visual perception (HVP). Instead of processing the entire depth image, we focus on some target regions where human recognizes their surface details based on texture priors. For that, we construct the high quality normal and texture database of various surfaces and develop a pattern recognition technique, which automatically recognizes the surface topography of target texture and classifies them into several categories. Given the classification results, it is possible to match the noisy input with the good alternative in the database. By replacing the input normal with the matched sample, we achieve the depth image enhancement with the unnoticeable changes.

Recently, the studies in HVP have been widely applied in image processing algorithms for enhancing the image quality. They include the analysis on how human filters the noise, discriminates two different noise levels and perceives a wide range of color and tones. Similarly, we present a new perceptual-based depth image enhancement method that uses the pattern recognition driven by how human perceives the surface topography. To learn the human ability of recognizing the surface details from texture, we review several existing literatures. According to them, the roughness is one of the important characteristics of surface topography and the pattern density is a major component of the roughness perception. Inspired by their analysis, we decide to use the pattern density as our primary feature for classifying the textures.

After classifying the textures, we replace the input noisy normals of target region by the high quality normals from the pre-collected database. Because the normals encode the geometric details, the proposed method consequently enhances the depth image quality by increasing the distance resolution. Previous work increases the contrast of image to enhance the roughness perception while the over-increment introduces the excessive artifacts in image. When the same idea is applied to depth image enhancement, such an artifact becomes even more serious due to severe noises in raw depth data. Because our method replaces the normals, it does not add artifacts.

The proposed method manipulates the surface normals in order to enhance the depth image quality using a pair of depth and color input image and the pre-collected normal and texture database. We train the classifiers to categorize various surfaces based on textures. After that, we segment the input color image into several local regions upon corresponding depth image, apply classifiers to select the target regions and identify their best match from the database. Then, we replace the input noisy normals by the selected normals. Our method consists of 1) training the classifiers, 2) classifying the input data and 3) replacing normals and combining the depth and normal data for improving the surface details. First, we construct the database composed of the high quality normal and its texture. We use the pattern density as a primary feature for classifying the surface topography; other features like the shape and color are also included. To define the pattern density, we compute the frequency response of the image. Given a set of features, we train a nearest neighbor classifier. Our training objects include a cushion, sofa, rug and other goods easily seen in living room. Second, we evaluate the local regions of the test input data using classifiers, determine the target regions and assign each of them into one of pre-determined classes. In this way, the noisy input normals are matched to the high quality normals in the database. Finally, we replace the noisy normals by corresponding high quality normals and combine them with the depth image by the existing method. Since this conventional method is effective to improve the depth quality using the high quality normal map, we apply the same scheme to merge our selected normal and input depth map. As a result, our method automatically recovers the surface details and provides the enhanced depth image from the perceptual perspective.

From the preliminary experiments, we confirm that the frequency response of pattern density is important factor to characterize the object surface. First, we conduct a user study to derive the human mechanism of perceiving the surface topography. Observers were asked to choose which one is rougher between two images. They are randomly selected from ten example images, which span natural fabric materials with varying pattern densities. In addition, we ask the participants to describe the criteria to compare them. From their responses, we find the important factors to judge the roughness: the pattern density, the amount of high frequency component, and color association. Second, we observe that the Fourier coefficients of image are distinctive enough to characterize the pattern density in our test examples. From these observations, we confirm the potential feasibility of our method. In the final manuscript, we present the experimental results with various natural objects with different patterns. Then, we conduct the subjective and objective quality assessments for evaluating our method.

The proposed method can be effective to enhance the 3D image quality because the high quality normals guided by HVP improve the geometric details. We expect that our work is effective to enhance the details of depth image from 3D sensors and to provide a high-fidelity virtual reality experience.

9394-45, Session PTues

A perceptual masking model for natural image with additive defects and detail loss based on adjacent visual channel inhibition

Yucheng Liu, Jan P. Allebach, Purdue Univ. (United States)

Masking is a perceptual effect in which the contents of the image inhibit the observer's ability to see a given target signal hidden in the image. Masking has been investigated extensively over the years due to its wide variety of applications in image analysis and processing tasks such as image compression, digital watermarking, and image quality evaluation. In the scenario of image quality evaluation, image contents are usually considered to be masks; and distortions in the image resulting from all aspects of the imaging process are considered to be the target signal. Understanding masking effects is a key step to solving the problem of determining how the image content influences the observer's response to deterioration in the image.

In this work, we focus on several types of typical image defects: blur, wavelet quantization errors, banding, AWGN, JPEG2000 transmission errors and JPEG artifacts, all of which result from common image compression, transmission, and rendering techniques. We have learned from previous work ([1,2]) that image defects affect the observer's perception in two distinct ways: additive impairment and detail loss. Additive impairment refers to the type of distortion that adds redundant information that doesn't exist in the original image (e.g. AWGN, banding etc.), and detail loss refers to another type of distortion that destroys the image structure and results in loss of image information (e.g. blur, quantization artifacts). These two types of distortions in the image may act jointly, in which case they play the role of masks to each other, in addition to the mask placed by the image content. In order to separate the modeling of these two types of defects, we decouple the additive impairment from the detail loss in the first stage of our image analysis pipeline.

We also learned from previous work ([3]) that different types of content in the image also affect the visibility of distortion differently, and using a universal set of model parameters for all types of image content tends to overestimate the masking effect in regions with simple image structure (edge), or underestimate the masking effect in regions with complex structure (grass, foliage). As a result, it is also necessary to optimize

the model for different types of image structures. We will follow the classification convention from previous work ([3]), and focus on three types of image masks with different levels of texture complexity: edges, structures, and textures.

The visual significance of additive impairment and detail loss are then modeled individually based on adjacent visual channel inhibition, which we proposed in previous research ([4]). In this work, we extend the model to include a cross-defect interaction term to model the mutual masking effect between the two types of defects described above. The model parameters are then optimized for the three types of image masks mentioned above, based on psychophysical experiment results collected in a controlled lab environment. In the experiment, the subjects are asked to make a yes/no decision regarding the visibility of distortions in the presence of a natural image mask. The threshold of visibility for each type of distortion on different types mask is learned from repeated experiments with a wide range of different combinations of distortion and mask. The model parameters are then trained to minimize the difference between the predicted thresholds and the measured ones.

Finally, the validity of the model is verified through cross-validation. And we also test the validity of the model in local masking by applying the model to a full natural image, where we conduct autonomous patch-based mask classification, and then adopt model parameters corresponding to the mask type. The full natural images are also inspected by human subjects. Comparison is made between the predicted result and ground truth to verify the performance of the model.

Reference

[1] Li, Songnan, et al. “Image quality assessment by separately evaluating detail losses and additive impairments.” Multimedia, IEEE Transactions on 13.5 (2011): 935-949.

[2] Chandler, Damon M. “Seven challenges in image quality assessment: past, present, and future research.” ISRN Signal Processing 2013 (2013).

[3] Damon M, Chandler, Gaubatz Matthew D, and Hemami Sheila S. “A patch-based structural masking model with an application to compression.” EURASIP Journal on Image and Video Processing 2009 (2009).

[4] Liu, Yucheng, and Jan P. Allebach. “A computational texture masking model for natural images based on adjacent visual channel inhibition.” IS&T/SPIE Electronic Imaging. International Society for Optics and Photonics, 2014.

9394-46, Session PTues

Influence of high ambient illuminance and display luminance on readability and subjective preference

Katrien De Moor, Norwegian Univ. of Science and Technology (Norway); Börje Andrén, Acreo Swedish ICT AB (Sweden); Guo Yi, Acreo Swedish ICT AB (Sweden) and KTH Royal Institute of Technology (Sweden); Kjell E. Brunnström, Acreo Swedish ICT AB (Sweden) and Mid Sweden Univ. (Sweden); Kun Wang, Acreo Swedish ICT AB (Sweden) and KTH Royal Institute of Technology (Sweden); Anton Drott, David S. Hermann, Volvo Car Corp. (Sweden)

Introduction

Many devices, such as tablets, smartphones, notebooks, fixed and portable navigation systems are used on a (nearly) daily basis, both in in- and outdoor environments. These environments can however differ greatly from each other, and have different requirements for enabling good and pleasurable user experiences. As a result, devices need to be adaptive and adaptable to a wide range of use contexts and conditions. It is often argued that contextual factors, such as the ambient illuminance (e.g., bright light on a sunny summer day vs. in-house artificial light in the evening) in relation to characteristics of the display (e.g., surface treatment, size, screen reflectance, and display luminance) may have a strong influence on the use of such devices and corresponding user experiences. Yet, the current understanding of these influence factors [1] is still rather limited. In this work, we therefore focus in particular on the impact of lighting on readability, visual performance and affective state.

In this paper, we share results from a subjective study aimed at evaluating two displays with different characteristics in conditions that simulate bright outdoor lighting conditions. More concretely, the paper addresses the following research questions: (1) which of our combinations of ambient illuminance levels and display luminance levels enables the best readability and visual performance? (2) Is there a difference between these different combinations in terms of subjective evaluation (i.e., affective state, acceptability, visual fatigue)? (3) Is there a difference between the displays used in the investigation?

Method and test setup

To answer these questions, we conducted a controlled lab study (N=18) with a within-subjects design. Four ambient illuminance levels and three display luminance settings were combined into 7 experimental conditions, which were evaluated for two different displays: one with a glossy, low-reflectance surface (Display A) and one matt display with a standard anti-glare surface (Display C). The labels in the figures below represent the different conditions; the numbers are respectively the reflected luminance and the display luminance, e.g. Condition 7 1536/680 means the condition number 7 has the reflected luminance 1536 cd/m2 and the display luminance 680 cd/m2. The reflected luminance level is the luminance reflected from a Kodak grey card of reflectance 18%. The ambient light levels were based on a pre-study [2] of outdoor lighting conditions. The selection of the conditions was however not straightforward, as we had to find a balance between high ambient illuminance levels, display luminance, screen reflectance on the one hand, and the comfort of the test subjects on the other hand. Too bright lighting conditions may affect the test subjects too much (without wearing sunglasses) and make it hard for them to further participate in the tests. For this reason, very bright lighting combinations were not included in this study. The order of the conditions in terms of severity is as follows: 7, 5, 3, 6, 4, 1, 2, 0 (baseline). According to our pre-study, the highest condition (7) corresponds to a daylight situation in which many people would tend to put their sunglasses on. Furthermore, the number of conditions was limited in order to not fatigue the test subjects too much. The aim was to have enough conditions to find a functional relationship, so that higher illuminance values can be extrapolated.

For every condition, readability was evaluated by means of different measures and tasks. These included the Freiburg visual acuity test (http://michaelbach.de/fract/index.html), a reading speed / reading comprehension task and a subjective evaluation. For the reading task, the participant was always asked to read English text excerpts (approximately 100 words) with a similar readability index (based on the Flesch Reading Ease Score) on the display and to answer a question about the content of the text afterwards. The reading time was captured as a performance measure and the content-related question was included as an additional measure (reading comprehension). For the subjective evaluation, the following subjective measures were included: emotional valence or pleasure, using the 7-point pictorial measure of valence from the Self-Assessment Manikin scale [3]; a measure of annoyance (3 items, reduced to one variable after reliability analysis using Cronbach’s alpha with ?= .894); visual fatigue (based on 6 items which were reduced to one variable, Cronbach’s ?= .907) measured on a 10-point scale [4]; acceptability (binary scale) and a measure of annoyance due to possible reflections in the screen (5-point scale, an adaptation of the Degradation Category Rating scale [5]). Before and during the subjective tests a number of measurements of e.g. luminance, illuminance levels, screen reflectance were performed. It took around 50 to 60 minutes for one test subject to complete the test (including the briefing) and the order of the conditions was randomized for the different participants.

Preliminary results

We have just finished the data gathering and the analysis phase is currently ongoing. Our preliminary inspection of the data indicates that even though there are some differences between the different conditions and displays in terms of visual performance and subjective evaluation, the lighting combinations were not too demanding for the test subjects. The texts were evaluated as readable in all conditions and in 90.6 % of the cases, the conditions under which the reading task was performed were evaluated as

good or good enough to perform the task (i.e., acceptable). Conditions 7, 5 and 6 are mainly responsible for the 9.6% of non-acceptability. On average, most time was needed to perform the reading task under conditions 7, 5 and 3. The subjective evaluations point in the same direction.

Figure 1a (left, display A) and 1b (right, display C): average values for self-reported pleasure, visual fatigue and annoyance for the different conditions (error bars show 95% confidence levels)

Figure 1a (display A, glossy screen surface) and 1b (display C, matt screen surface) show the average scores for pleasure, visual fatigue and annoyance for the different combinations of reflected luminance and display luminance. Overall, it can be observed that visual fatigue and annoyance are relatively low, while the self-reported pleasure (measured on a 9-point scale) is relatively high for all conditions and for both displays. However, the results mirror the level of contrast between reflected luminance and display luminance. Higher contrast corresponds with slightly higher self-reported pleasure, lower frustration and visual fatigue. Similarly, more reflections are reported as the contrast decreases (Figure 2). However, there are some differences between the matt (display C) and glossy (display A) display, which will be analysed in detail.

Figure 2: evaluation of the effect of possible reflections in the screen (mean values), ranging from 5 (imperceptible) to 1 (very annoying). The error bars show 95% confidence levels.

In terms of preference, the majority (61%) of the test subjects indicated a preference for the glossy display A. 33% preferred display C and one participant did not have any preference for one or the other.

Conclusions

A subjective experiment was performed evaluating two displays with different characteristics in conditions that simulate bright outdoor lighting conditions. We can see a negative effect of increased ambient illuminance and reduced relative contrast on readability and comfort (subjective evaluations). To further investigate this effect and to identify the boundaries in terms of what is (un)comfortable and (un)acceptable, more demanding illuminance levels need to be considered. However, this is challenging to handle for the concern of the test subject.

REFERENCES

[1] U. Reiter, K. Brunnström, K. Moor et al., [Factors Influencing Quality of Experience] Springer International Publishing, 4 (2014).

[2] B. Andrén, K. Brunnström, and K. Wang, "Readability of displays in bright outdoor surroundings," Proc of SID Display Week 2014, June 1-6, 2014, San Diego, USA, paper: P-37, (2014).

[3] P. J. Lang, [Behavioral treatment and bio-behavioral assessment: computer applications.] Ablex, Norwood(1980).

[4] S. Benedetto, V. Drai-Zerbib, M. Pedrotti et al., "E-Readers and Visual Fatigue," PloS one, 8(12), e83676 (2013).

[5] ITU-T, "Recommendation P.910 (04/08) Subjective video quality assessment methods for multimedia applications," (2008).

9394-47, Session PTues

A no-reference bitstream-based perceptual model for video quality estimation of videos affected by coding artifacts and packet losses

Katerina Pandremmenou, Univ. of Ioannina (Greece); Muhammad Shahid, Blekinge Institute of Technology (Sweden); Lisimachos P. Kondi, Univ. of Ioannina (Greece); Benny Lovstrom, Blekinge Institute of Technology (Sweden)

No Abstract Available

9394-48, Session PTues

Saliency detection for videos using 3D FFT local spectra

Zhiling Long, Ghassan AlRegib, Georgia Institute of Technology (United States)

Bottom-up spatio-temporal saliency detection identifies perceptually important regions of interest in video sequences. Such detection may help video processing (e.g., coding, classification, summarization, etc.) to be more effective and efficient. The center-surround model proves to be useful for visual saliency detection [1]. With this model, saliency is determined by comparing a pixel against its neighbors with regard to some certain features. A pixel being very different from its neighbors is considered to be salient.

In this work, we adopt a specific center-surround model based framework for videos, i.e., saliency detection by self-resemblance (SDSR) [2]. SDSR computes for each video frame pixel some local features named as local steering kernels (LSK). Similar to most saliency detection techniques available in the literature, the LSK features are calculated in the time-space domain. In this research, we are interested in exploring with spectral domain local features, in an attempt to capture the intrinsic patterns of the localized frequency components of the signal. Our features are extracted by applying the 3D Fast Fourier Transform (FFT) to localized time-space cubes obtained by using sliding windows. We use 3D FFT mainly for two reasons: 1) application of 3D FFT to video signals does not involve complicated parameter settings, which is a common shortcoming many available techniques suffer from; and 2) FFT calculation is very efficient and fast. Although FFT was employed in several previous techniques for visual saliency detection, it was normally utilized in a global manner in 2D form [3, 4].

The FFT magnitude map has been commonly adopted for spectral analysis. However, when we use local FFT magnitude maps as features for the SDSR style saliency detection, the results are not satisfactory. As revealed in some previous study [4], phase information plays an important role in saliency detection as well. But FFT phase maps are usually very noisy, not suitable for analysis in their original form. To overcome this problem, we directly use the complex FFT values as features, so that complete spectral information is available. Such complex spectra are computed locally and then compared using a matrix cosine similarity distance, as with SDSR but modified to accommodate the complex values.

SDSR uses all surrounding neighbors to evaluate the self-resemblance (or the similarity between a center and its surrounding counterparts), without differentiating between the spatial neighbors and the temporal neighbors. This implies that the two types of neighbors contribute equally to the video saliency. However, as pointed out in the literature, the human visual system (HVS) is more sensitive to moving objects than static ones [5]. In other words, temporal changes tend to contribute more to video saliency than spatial variations. Therefore, we break the time-space neighbors into two separate groups: spatial and temporal. The matrix cosine distances are calculated for each group respectively before being combined via a weighted summation. In accordance with the aforementioned HVS property, we weigh the temporal results differently than the spatial results.

The proposed technique is applied to CRCNS, a public database, which includes 50 videos of diversified contents [6]. Both visual and numerical evaluations will be conducted to demonstrate its performance. In Fig. 2, we provide some example saliency maps obtained using different methods in our preliminary experiments. The results show that our method outperforms the others by generating more accurate detections with fewer false alarms at the same time. In our final paper, we will also provide numerical results such as the area under the receiver operating characteristics curve (AUC) and the Kullback-Leibler (KL) divergence.

References

1. D. Gao, V. Mahadevan and N. Vasconcelos. On the plausibility of the discriminant center-surround hypothesis for visual saliency. Journal of Vision, 8(7):13, 1–18, 2008.

2. H. Seo and P. Milanfar. Static and space-time visual saliency detection by self-resemblance. Journal of Vision, 9(12):15, 1-27, 2009.

3. X. Hou and L. Zhang. Saliency detection: A spectral residual approach. IEEE Conference on Computer Vision and Pattern Recognition, 2008.

4. C. Guo and L. Zhang. A novel multiresolution spatiotemporal saliency detection model and its applications in image and video compression. IEEE Transactions on Image Processing, 19(1), 185-198, 2010.

5. L. Itti. Automatic foveation for video compression using a neurobiological model of visual attention. IEEE Transactions on Image Processing, 13(10), 1304-1318, 2004.

6. L. Itti. Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. Visual Cognition, 12(6), 1093-1123, 2005.

9394-49, Session PTues

Perceived interest versus overt visual attention in image quality assessment

Ulrich Engelke, Commonwealth Scientific and Industrial Research Organisation (Australia); Patrick Le Callet, Univ. de Nantes (France); Hantao Liu, The Univ. of Hull (United Kingdom)

No Abstract Available

9394-50, Session PTues

A tone mapping operator based on neural and psychophysical models of visual perception

Praveen Cyriac, Marcelo Bertalmio, David Kane, Javier Vazquez-Corral, Univ. Pompeu Fabra (Spain)

High Dynamic Range (HDR) imaging techniques involve capturing and storing real world radiance values that span many orders of magnitude. However, common display devices can usually reproduce intensity ranges only up to two orders of magnitude. There- fore, in order to display an HDR image on a low dynamic range (LDR) screen, the dynamic range of the image needs to be compressed without losing details or introducing artifacts, and this process is called tone mapping. A good tone mapping operator (TMO) must be able to produce an LDR image that matches as much as possible the perception of the real world scene. For this reason, many successful TMOs are based on models of some aspects of the human visual system [11], and also the metrics to evaluate TMOs relate to data and models on visual perception [1, 9, 14].

In this work we follow the approach proposed by Ferradans et al. [5] and present a two stage TMO. The first stage is a global method (i.e. pixel location isn't considered) for range compression based on perceptual models for light adaptation and contrast percep- tion, and the second stage performs local contrast enhancement and color induction using neural activity models for the visual cortex.

For the first stage of our TMO, we take into account the following three approaches:

1. Gamma correction depending on luminance distribution and background (e.g. [7]). Ongoing research in our lab [6] demonstrates that subjects preference for gamma compression may be predicted by computing the degree of histogram equalization in the lightness distribution. Here lightness refers to the perception of monochromatic light in the scene and is computed by a pixel-wise gamma function of onscreen luminance, for which the exponent varies with mean luminance of the stimulus and surround.

2. Working in the HDR color space [4], which is an HDR extension of CIELAB.

3. Using models of photoreceptor response curves [11].

For the second stage we adopt the neural model [3], which is an extension of the contrast and color enhancement method of [2] used in [5], with larger capabilities in terms of redundancy reduction and the ability to reproduce assimilation phenomena. Both [3] and [2] are closely related to the Retinex theory of color vision [8] and to the perceptually inspired color correction approach of [12].

For the validation of our tone mapping approach we consider three different metrics, all comparing an original HDR image with its LDR tone mapping result in terms of perception of details and/or colors:

1. Metric [1], based on psychophysical data on contrast visibility.

2. Metric [14], a combination of a measure of structural fidelity (based on [13]) with a measure of deviation from the statistics of natural images, which the authors claim are a good approximation of perceptual quality.

3. Metric [10], an HDR extension of the iCID metric [9], which in turn was derived from the color extension of the perceptual metric [13].

The first two metrics compare only the luminance of the two images, while the third one also estimates color distortions. Preliminary results show that our proposed method compares well under all three metrics with several state of the art approaches.

References

[1] T.O. Aydin, R. Mantiuk, K. Myszkowski, H.-P. Seidel, "Dynamic range independent image quality assessment", Proc. ACM SIGGRAPH 08, pp. 1-10, 2008.

[2] M. Bertalm ?ıo, V. Caselles, E. Provenzi, and A. Rizzi, "Perceptual color correction through variational techniques", IEEE Trans. Image Processing, vol. 16, no. 4, pp. 1058-1072, 2007.

[3] M. Bertalm ?ıo, "From Image Processing to Computational Neuroscience: A Neural Model Based on Histogram Equation", Front. Comput. Neurosci. 8:71, 2014.

[4] M. Fairchild, P-H Chen, "Brightness, Lightness, and Specifying Color in High- Dynamic-Range Scenes and Images", proc. SPIE 7867, Image Quality and System Performance, 2011.

[5] S. Ferradans, M. Bertalm ?ıo, E. Provenzi, and V. Caselles, "An analysis of visual adaptation and contrast perception for tone mapping", IEEE Trans. Pattern Anal. Mach. Intell., 33(10), pp. 2002-2012. 2011

[6] D. Kane, M. Bertalm ?ıo, "Perceptual histogram equalization predicts the preferred gamma transform for any given natural image", SPIE (abstract submitted)

[7] C. Liu, M. Fairchild, "Re-measuring and Modeling perceived image contrast under different levels of surrounding illumination", Color and Imaging Conference vol.2007, no. 1, pp. 66-70, 2007.

[8] E. Land, J. McCann, "Lightness and Retinex Theory", J. Optical Soc. Am., vol. 6, no. 1, pp. 1-11, 1971.

[9] J. Preiss, F. Fernandes, P. Urban, "Color-Image Quality Assesment: From Prediction to Optimization", IEEE Trans. Image Processing, vol. 23, no.3, pp. 1366-1378, 2014.

[10] J. Preiss, M. Fairchild, J. Ferwerda, P. Urban, "Gamut Mapping in a High-Dynamic- Range Color Space", Proc. SPIE 9015, 2014.

[11] Reinhard, E., Heidrich, W., Debevec, P., Pattanaik, S., Ward, G., & Myszkowski, K. (2010). "High dynamic range imaging: acquisition, display, and image-based light- ing." Morgan Kaufmann, 2010.

[12] A. Rizzi, C. Gatta, D. Marini, "A New Algorithm for Unsupervised Global and Local Color Correction", Pattern Recognit. Lett., vol. 124, pp. 1663-1677, 2003.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Trans. Image Processing, vol. 13, no.4, pp. 600-612, 2004.

[14] H. Yeganeh, Z. Wang, "Objective Quality Assessment of Tone-mapped Images", IEEE Trans. Image Processing, vol. 22, no.2, pp. 657-667, 2013.

9394-51, Session PTues

Illuminant color estimation based on pigmentation separation from human skin color

Satomi Tanaka, Chiba Univ. (Japan); Akihiro Kakinuma, Naohiro Kamijo, Hiroshi Takahashi, Ricoh Co., Ltd. (Japan); Norimichi Tsumura, Chiba Univ. (Japan)

Human has the visual system called “color constancy” intending that the perceptive colors of same object are maintained across various light sources, and nevertheless, the object’s physical color is strongly dependent on the illuminant color. Almost all recent digital cameras have this visual system as “white-balance function”. White-balance function corrects the color of digital images by using the information of estimated illuminant color.

Various color constancy algorithms have been proposed. The effective method is to take a color image with an object whose color is previously known. Tsukada et al. proposed the method that uses the human facial color included in a digital color image. This method adapts the subject’s facial color to the averaged skin color of Japanese woman acquired in advance. However, this method has wrong estimation results because of difference among individual facial colors.

In this paper, we propose the novel color constancy algorithm that uses subject’s facial color in image and doesn’t depend on individual difference. We use the skin color analysis method that was presented by Tsumura et al. to analyze the subject’s skin color. This method separates the skin color from the facial image into 3 color components; the color of melanin, hemoglobin and shading. In this research, we confirmed that the same component can be extracted from 3 separated components among different person. Using this property of human skin color, we proposed the illuminant color estimation method from facial color image, and furthermore, we realized white-balance function among individual skin color.

Our illuminant color estimation method is performed by following steps. First, we separated the skin color to 3 components of melanin, hemoglobin and shading by using the skin color analysis. Next, we calculated the minimum vector of pigment components as standard bias vector from the facial image taken under the xenon lamp. To estimate the illuminant color, we assumed that the bias vectors consisted the minimum components of melanin and hemoglobin under same light source are maintained among the different Japanese people. We performed the experiment to calculate the bias vectors from facial images of 10 Japanese persons, and confirmed that the bias vectors extracted were same value.

Then, we calculated the bias vectors from facial images taken under various light sources such as halogen lamp, and estimated the color of light sources by subtracting the standard bias vector from the bias vectors. In this way, the bias vectors extracted from the pigment components enable us to represent the color of light sources. Finally, we exclude the color shift of the facial image caused by the change of light source from standard light source, and realized the white-balance function.

In this paper, we performed the experiment to correct facial images and evaluate the effect on our proposed method. We took the 4 Japanese persons under 5 light sources, and took the ground truth images under standard light source. We color corrected 20 facial images by using the proposed method and also conventional method. To compare the results of 2 methods, we use the color difference between the corrected image and ground truth. The results of this experiment represent the superiority of the proposed method against the conventional method regarding white-balance function.

In conclusion, we presented the novel color constancy algorithm based on skin color analysis. We used the skin color of subject’s face. We assume the property that the bias vectors extracted facial images taken under same light source are maintained among the different people. In this paper, our white-balance function can be achieved when Japanese people is subject in the image. We expect that the proposed method can be adapted to Southeast Asians like Japanese. However, Caucasian and African American are not taken into account in this work. Therefore, we will endeavor on the proposed color constancy method for various race people as the future work.

9394-52, Session PTues

Evaluation of color encodings for high dynamic range pixels

Ronan Boitard, Technicolor S.A. (France); Rafal K. Mantiuk, Bangor Univ. (United Kingdom); Tania Pouli, Technicolor S.A. (France)

No Abstract Available

9394-53, Session PTues

Using false colors to protect visual privacy of sensitive content

Serdar Ciftci, Middle East Technical Univ. (Turkey); Pavel Korshunov, Ecole Polytechnique Fédérale de Lausanne (Switzerland); Ahmet O. Akyuz, Middle East Technical Univ. (Turkey); Touradj Ebrahimi, Ecole Polytechnique Fédérale de Lausanne (Switzerland)

The widespread usage of digital video surveillance systems has increased the concerns for privacy violations. According to a 2013 security industry report, Britain has a CCTV camera for every 11 people and decade old reports dating as early as 2004 suggest that an average Briton is caught on camera 300 times a day – a number which is likely to have since increased significantly. The situation is similar in many other countries as far as the rapid deployment of video surveillance systems are concerned. Such recording of personal data combined with video analytics solutions and ease of access of personal information through social networks give rise to an unprecedented ability to collect privacy sensitive content on individuals. The invasive nature of video surveillance systems makes it difficult to find an acceptable balance between privacy of the public under surveillance and security related features of such systems.

Many privacy protection tools have been proposed for preserving privacy, ranging from simple methods such as masking, blurring, and pixelization to more advanced solutions such as scrambling, warping, and morphing. Tools for protection of visual privacy available today lack either all or some of the important properties which may be summarized as being simple and efficient (can potentially operate in real-time), secure (the ability to undo privacy only through a secret key), reversible (by authorized personnel), flexible (independent of specific formats and encodings), adjustable (the degree of protection can be changed), visually pleasing (viewing protected content should not be disturbing), and selective (should only protect privacy but allow non-private information to be extracted).

Therefore, in this paper, we propose a simple yet effective method for privacy protection based on false color visualization, which maps color palette of an image into a different color palette, possibly after a compressive point transformation of the original pixel data, distorting the details of the original image. This method does not require any prior face detection or other sensitive regions detection and, hence, unlike typical privacy protection methods, it does not rely on inaccurate computer vision algorithms. It is also secure as the look-up tables can be encrypted, reversible as table look-ups can be inverted, flexible as it is independent of format or encoding, adjustable as the final result can be computed by interpolating the false color image with the original with different degrees of interpolation, visually pleasing as it does not create distracting visual artifacts, and selective as it preserves the semantic structure of the input.

To demonstrate the feasibility of the approach, we apply the objective framework for privacy evaluation on a public FERET dataset. Such objective evaluation assumes an automated video surveillance system that relies on video analytics for its operation. This assumption is reasonable, given how widespread and largely deployed the surveillance systems are today, which forces them to rely on video analytics in order to reduce the cost of the surveillance and increase their scalability. In such systems, privacy can be considered as well protected if the performance of a privacy intrusive video analytic approach, such as face recognition, drops below an acceptable

level. Therefore, we run face detection and recognition algorithms on the images from the FERET dataset using different color palette/compression functions combinations and show that faces in false colored images retain intrinsic features of a face, while making them unrecognizable, which ensures the protection of privacy. We also evaluate the security of the approach by simulating an attack on the protected content.

In a more conventional surveillance systems human subjects are typically the end users of privacy protection tools, which means that the performance of such tools should also be evaluated subjectively. Therefore, the performance of our approach is also evaluated using subjective assessment using a specifically designed task based evaluation methodology that analyzes the tradeoff between the preservation of privacy offered by privacy protection tools and the intelligibility of activities under video surveillance. To that end, human subjects rate the degree of privacy protection versus the amount of visible non-sensitive information.

The effectiveness of the objective framework for assessing the performance of the proposed solution is then compared to subjective evaluations via curve fitting and statistical analysis.

9394-54, Session PTues

The visual light field in paintings of museum Prinsenhof: comparing settings in empty space and on objects

Tatiana Kartashova, Technische Univ. Delft (Netherlands); Susan te Pas, Utrecht Univ. (Netherlands); Sylvia C. Pont, Huib de Ridder, Technische Univ. Delft (Netherlands); Marga Schoemaker, Museum Prinsenhof Delft (Netherlands)

Lighting a scene while rendering is admittedly one of the most intricate and resources-consuming operations in the computer graphics field. In the meantime, artists cope with such tasks for centuries what resulted in an accumulation of certain techniques helping to expose the light in space naturally. To what extent are human observers sensitive to painters' solutions? This research will be performed on paintings of the museum Prinsenhof to study pictorial light perception of art works in a quantitative instead of qualitative manner. In our experiments we exploit the method of a visual probe. Specifically, subjects are requested to interactively set the lighting (direction, intensity, diffuseness) of a matte white sphere such that it fits into (an empty part) of the scene (probe superposed on the painting), in Experiment 1; or as if it is illuminated in the same manner as the object (probe in separate window next to part of the painting), in Experiment 2.

We will analyze whether the settings in empty parts and on objects of these paintings are congruent with each other, whether they vary over the painting, and between observers. Results of the full experiment will be presented at SPIE.

9394-55, Session PTues

Using V1-based models for difference perception and change detection

Pei Ying Chua, DSO National Labs. (Singapore); K. Kwok, DSO National Labs. (Singapore) and Temasek Labs. (Singapore)

Using V1-based models, it is possible to investigate the features of human visual processing that influence difference perception and change detection. V1-based models were based on basic colour opponency and receptive field properties within the human visual system. The studies conducted demonstrate the various applications and key considerations of human vision models.

Difference Perception: Human feedback on the perceived difference between pairs of naturalistic images were compared against models' predictions. Performance was enhanced through tuning of colour and contrast sensitivity. Separately, in order to study performance with arti ficial images, models' predictions for pairs of thermal images were also compared against the known ground-truth temperature. Crucially, models that performed well with naturalistic images did not work well with thermal images, and vice versa.

Change Detection: Models' performances were evaluated by their sensitivity and accuracy in correctly detecting the presence or absence of changes in the given scene. Models tuned to higher spatial frequencies were both more accurate and sensitive, likely because high spatial frequencies represent the fine details critical for change detection. Modelling smaller receptive fields also increased sensitivity to small changes. Allowing the model to shift attention and consider cues from all regions of the image equally also improved performance.

9394-20, Session 4

Effect of daylight on atmosphere perception: comparison of a real space and visualizations *(Invited Paper)*

Mariska G. M. Stokkermans, Yuexu Chen, Technische Univ. Eindhoven (Netherlands); Michael J. Murdoch, Ingrid M. L. C. Vogels, Philips Research Nederland B.V. (Netherlands); Ingrid E. J. Heynderickx, Technische Univ. Eindhoven (Netherlands) and Philips Research Nederland B.V. (Netherlands)

Introduction

Artificial interior lighting has a major impact on the perceived atmosphere in a space. Changing the light's intensity, spectral distribution or spatial distribution can make a space look more cozy, relaxing, exciting or even threatening [1]. Hence, understanding the exact relationship between light characteristics and perceived atmosphere is crucial for those application contexts where a dedicated atmosphere is desired. For example, one would like to have an exciting atmosphere in entertaining environments, a relaxing atmosphere in wellness environments, an activating atmosphere in classrooms or a brand-specific atmosphere in retail. In the latter case, it has been demonstrated that an appropriate interior light ambience contributes to the perceived quality of a store [2].

In many environments daylight, naturally entering the space, is added to the specifically designed light ambience, often in an uncontrolled way. In the particular case of retail, for example, maintaining a specific atmosphere in all brand stores is not straightforward because of for instance the variation in the amount of daylight due to the location of the store (mall versus outdoor shopping street) or daytime versus evening opening hours. As daylight may affect among others the physical light intensity, spectral distribution and spatial distribution of the light in a space, it is expected to have a considerable impact on the perceived atmosphere. On the other hand, the human visual system is also known to adapt to some of these effects, and so, adaptation may suppress to some extent the expected impact of daylight on the perceived atmosphere. Surprisingly little evidence on the extent to which daylight influences the perceived atmosphere of a light ambience exists in literature. Therefore, the goal of this study is to measure the influence of the presence of daylight on the perceived atmosphere of an interior light ambience.

This research question is addressed both in a real space and via visualizations. Previous research has demonstrated the relevance and accuracy of visualizations of lighting environments to perform studies on light perception [3], [4]. However, these studies only used artificial lighting, and so, didn't include a contribution of daylight. Adding daylight to these visualizations may cause additional challenges because of the limited dynamic range and luminance output of a display. By conducting our research in a real space as well as via visualizations, we hope to prove the perceptual accuracy of visualizations also in the presence of daylight.

Method

The research question is addressed with an empirical study, using a 2 (daylight: no daylight – daylight) by 2 (mediation: real space – visualization) by 5 (light ambience) full-factorial within-subjects design. The five distinct light ambiences that were created varied in type of spatial distribution (diffuse or directional), intensity, color temperature and color of the accent lighting. Participants (N = 28) were asked to assess the perceived atmosphere of these light ambiences with the 11-items atmosphere questionnaire, developed by Vogels [5]. Each item had to be assessed on a 7-point scale, ranging from "not applicable at all" to "very applicable". The items were then aggregated into the following four atmosphere dimensions: coziness, liveliness, tenseness and detachment. Additionally, to increase our insights in possible relations between light perception and atmosphere perception, participants also assessed all light ambiences on brightness (dim – bright), uniformity (non-uniform – uniform) and color temperature (warm – cool) using a 7-point scale.

For the no-daylight conditions, the real space and the modelling pipeline of the visualizations were the same as used in our previous studies [3], [4]. Daylight was added in the real space by opening a window-blind (3.2 x 2.2 m), located behind the participant. As this window was located at the north side (northern hemisphere) of the building, mainly diffuse daylight entered the space. A similar condition was created in the visualizations by including a window in the virtual space, through which daylight entered. The visualizations were shown on a 46" TV in a lab space without ambient lighting.

Results

The data were analyzed by two Linear Mixed Model (LMM) analyses.
In the first LMM, we analyzed the effect of adding daylight to the light ambiences in the real space. The results of this LMM showed a significant main effect ($p < .05$) of light ambience for all perceptual attributes. Further, a significant main effect of daylight was found for uniformity, detachment and tenseness. As shown in Figure 1, uniformity and detachment showed a small increase, while tenseness showed a small decrease when daylight was added to the artificial light ambience. Furthermore, we found significant interaction effects between daylight and light ambience for the attributes color temperature, liveliness, and tenseness. For each of these attributes the effect of daylight was significant only for some of the light ambiences. However, the direction of the effect for those light ambiences showing significant differences was equal: results showed a decrease in color temperature and tenseness and increase in liveliness when daylight was present. For brightness and coziness no significant effects were found.

In the second LMM we tested the effect of mediation for all light ambiences in both daylight and no-daylight conditions. Similar to the first LMM, results showed a significant main effect ($p < .05$) of light ambience for all perceptual attributes. Further, a significant main effect of mediation was found for brightness, uniformity and detachment. The visualizations were assessed to be less bright, less uniform and less detached than the real space (as shown in Figure 1). Additionally, for uniformity and detachment interaction effects between mediation and light ambience were found. For those light ambiences that showed a significant effect of mediation, the direction of the effect was equal to the main effect. For color temperature and uniformity an interaction effect between daylight and mediation was found. Regarding color temperature the light ambiences were perceived as slightly cooler in the visualizations compared to the real space in the daylight condition, while this effect was reversed for the no-daylight condition. For uniformity the light ambiences in the visualizations were perceived significantly less uniform than in the real space with daylight, while without daylight no significant difference between the visualizations and the real space was found. No significant effects were found for coziness, liveliness, and tenseness. Overall, even though some findings were statistically significant, the differences between the real space and the visualizations – for both daylight and no-daylight conditions – were very marginal and did not exceed 0.5 on a 7-point scale for most light ambiences.

Conclusions

This study showed a small influence of adding daylight to a light ambience on the perceived light and atmosphere of an environment. The most pronounced and consistent differences were found for the uniformity, tenseness and detachment. The addition of daylight led to a small increase in uniformity and detachment, while tenseness decreased slightly. The fact that daylight had little or no impact on brightness and color temperature suggests that adaptation may play an important role in the perception of a light ambience. However, the present study only took diffuse daylight from the north side into account, whereas daylight from the south side may cause larger differences in the actual light distribution, due to its contribution of direct sunlight.

Furthermore, we can conclude that for the majority of perceptual attributes the visualizations including daylight have a similar perceptual accuracy compared to visualizations without daylight. Regarding the overall perceptual accuracy, we found that similar to our previous studies 'higher level' attributes related to perceived atmosphere have a higher accuracy than 'lower level' light perception attributes.

References

[1] I. M. L. C. Vogels, M. de Vries, and T. A. M. Erp van, "Effect of coloured light on atmosphere perception," in Interim meeting of the international colour association - Colour Effects & Affects, 2008.

[2] J. Baker, D. Grewal, and A. Parasuraman, "The influence of store environment on quality inferences and store image," J. Acad. Mark. Sci., vol. 22, no. 4, pp. 328–339, 2009.

[3] U. Engelke, M. G. M. Stokkermans, and M. J. Murdoch, "Visualizing lighting with images: converging between the predictive value of renderings and photographs," in Human vison and electronic imaging, 2013.

[4] M. J. Murdoch and M. G. M. Stokkermans, "Effects of image size and interactivity in lighting visualization," in Human vison and electronic imaging, 2014.

[5] I. M. L. C. Vogels, "Atmosphere Metrics: a tool to quantify perceived atmosphere," in International symposium creating an atmosphere, 2008.

9394-21, Session 4

The role of natural lighting diffuseness in human visual perception *(Invited Paper)*

Yaniv Morgenstern, Univ. of Minnesota, Twin Cities (United States); Wilson S. Geisler, The Univ. of Texas at Austin (United States); Richard F. Murray, York Univ. (Canada)

The pattern of the light that falls on the retina is a conflation of real-world sources such as illumination and reflectance. Human observers often contend with the inherent ambiguity of the underlying sources by making assumptions about what real-world sources are most likely. Here we examine whether the visual system's assumptions about illumination match the statistical regularities of the real world. We used a custom-built multidirectional photometer to capture lighting relevant to the shading of Lambertian surfaces in hundreds of real-world scenes. We quantify the diffuseness of these lighting measurements, and compare them to previous biases in human visual perception. We find that (1) natural lighting diffuseness falls over the same range as previous psychophysical estimates of the visual system's assumptions on diffuseness, and (2) natural lighting almost always provides lighting direction cues that are strong enough to override the human visual system's well known assumption that light tends to come from above. A consequence of these findings is that what seem to be errors in visual perception are often actually byproducts of the visual system knowing about and using reliable properties of real-world lighting when contending with ambiguous retinal images.

9394-22, Session 4

The influence of lighting on visual perception of material qualities *(Invited Paper)*

Fan Zhang, Huib de Ridder, Sylvia Pont, Technische Univ. Delft (Netherlands)

As part of the EU PRISM project, one of our goals is to understand the

relations between the material and lighting perception of real objects in natural scenes. In this paper, we present an experiment to investigate the influence of different canonical lighting conditions on the perception of material qualities. This was investigated with a novel material probe and systematical changes of the lighting condition of the materials in stimulus images. Matte, velvety, specular and glittery are selected as four canonical BRDF modes. Images of the stimulus and the probe are optically mixed photos taken from the same matte, velvety, specular and glittery bird-shaped objects, but different lighting conditions and viewing angles. A matching experiment was conducted by asking participants to simply match material qualities of the object in the probe to that of the stimulus. We found a decrease of performance when the lighting was different for stimulus and probe. For example, the glittery mode was quite independent from the other modes under quite diffuse lighting, while sometimes mistakenly perceived as specular under more directed lighting condition. In current on-going experiments, we investigate the interactions between lighting and material in a systematical manner by using canonical lighting conditions and again mixing canonical materials. The details of complete experiment and analysis of the results will be presented at SPIE.

This work has been funded by the EU FP7 Marie Curie Initial Training Networks (ITN) project PRISM, Perceptual Representation of Illumination, Shape and Material (PITN-GA-2012-316746)

9394-23, Session 4

Effect of fixation positions on perception of lightness *(Invited Paper)*

Matteo Toscani, Justus-Liebig-Univ. Giessen (Germany)

No Abstract Available

9394-24, Session 4

Perception of light source distance from shading and shadows *(Invited Paper)*

Roland W. Fleming, Justus-Liebig-Univ. Giessen (Germany)

No Abstract Available

9394-57, Session Key

Next gen perception and cognition: augmenting perception and enhancing cognition through mobile technologies *(Keynote Presentation)*

Sergio R. Goma, Qualcomm Inc. (United States)

No Abstract Available

9394-25, Session 5

Reducing observer metamerism in wide-gamut multiprimary displays

David Long, Mark D. Fairchild, Rochester Institute of Technology (United States)

For over 100 years, the cinema has afforded filmmakers the opportunity to share intentional visual experiences with moviegoers, the result of their mastering photographic technologies to render creative efforts on screens both big and small. The model of distribution for these artistic endeavors has been mostly reliable and predictable. Color stimuli generated on screen during a movie's production can survive to the ultimate audience with little distortion in presentation and interpretation. In the era of film capture and projection, filmmakers harnessed the tools of color reproduction, learning film's palette and gamut and shaping each frame through lighting, art direction and optical printing to yield the desired impact. Even for traditional electronic distribution via video broadcast, the color technologies of CRT phosphors proved robust to the visual sensitivities of most viewers. However, reproduction of color in cinematic reproduction has always been a well-crafted illusion within the human visual system.

More than 150 years after Maxwell first proposed the theory of trichromatic color reproduction, all practical motion-imaging systems continue to rely on metamerism wherein a particular integrated stimulation of the three cone types found on the human retina is sufficient to reproduce the sensation of color of any real object, regardless of higher order spectral composition. This simplified treatment, though effective, is necessarily restrictive in light of emerging trends such as the introduction of wide-gamut display technologies.

Concerns in the trichromatic reproduction model are emerging in the form of variability in observer metamerism. Controlled metameric matches of color within the display for a single observer may prove not to be matches for another observer with slightly different color-matching functions or may prove inconsistent even for single observers as they age. Dye-based film systems and phosphor-based CRT displays are generally forgiving in the metamerism illusion across disparate observers. Broad spectral representation within each colorant limit the chances for quantal integration differences within the cones amongst a diverse population. But evolving technologies for both large and small screens employ new physics with LED, OLED and laser illumination engines. As they promote widened color gamuts, these displays are decidedly more limited in their spectral composition. Previous research confirms that spectrally selective primary sets necessary for expanding color gamut exacerbate observer variability. In related work, the Society of Motion Picture and Television Engineers is exploring alternatives to standard observer colorimetry for calibrating newer video mastering displays employing these same physics.

Multispectral imaging systems offer different options for co-optimization of increased palette and reduced observer variability. In the ideal case, narrow-bandwidth, high-spectral-resolution systems would be conceived to accomplish the goals of controllable spectral capture and reproduction of target stimuli. By combining near-monochromatic characteristics at a high sample rate across the visible electromagnetic spectrum, many sufficiently complex stimuli could be rigorously rendered. However, these solutions require optically-complex, processing-intensive systems that often compromise spatial quality for spectral precision. The present research seeks to confirm that an abridged primary display system is capable of minimizing observer metamerism while delivering enhanced color gamut.

The design of such a system, though, must be deliberate, assessed against meaningful objective criteria for color reproduction, metamerism reduction and spectral gamut. Recent work at Rochester Institute of Technology (RIT) has focused on three critical components of the multispectral system ecosystem: 1) Observer color-matching-function demographics, 2) observer variability index definitions and 3) abridged multispectral display optimization. Previous publications outline many of the results of objectives 1 and 2. Here, summary of efforts to build and test a prototype system are provided. The intended purpose of the RIT multispectral display is to confirm current understanding of variabilities amongst real observers and to provide evidence for potential in metamerism reduction versus emerging cinema display technologies.

The starting objective for design of the RIT multi-primary display (MPD) was to deliver meaningfully reduced observer variability versus traditional 3-channel RGB systems. Candidate primary spectra were originally simulated via parametric optimization as opposed to being restricted to a heuristic selection from a set of available commercial color filters. The final design was implemented using materials then that performed most closely to the ideal computational models. In this manner, deficiencies in available filter sets could be quantified versus optimized results. To keep the mathematics simplified in the constrained computational optimization, a generalized Gaussian transmission profile, T(?), was modeled for each potential primary filter. The entire system was illuminated using one of two measured source spectra common in cinema applications, one a typical large-venue xenon arc lamp and the other a consumer-grade mercury

arc UHP lamp. Across N total primaries for the system, the transmission profiles were varied in both peak transmission wavelength, ?, and peak-width, ?. Variations investigated for primary design included the number of primaries (N=3 through 8), starting guess for Gaussian parameters and spectral domain permitted for iteration of each primary's characteristics (each primary having it's peak wavelength constrained to a window of wavelengths versus permitting any monotonic array of peak wavelengths for the N primaries between constrained spectral endpoints of 380 and 720nm).

The spectral objective for the primary parameter optimizations was minimized observer metamerism in reproduction of a set of a priori training spectra. A collection of candidate spectral sets were investigated and compared for training the reproduction model. These sets included

1) MacBeth Color Checker (24 samples)

2) MacBeth Color Checker DC (240 samples)

3) US Patent No. 5,582,961 "Kodak/AMPAS" test spectra (190 samples)

4) Munsell sample spectra (1269 samples)

5) SOCS spectral database (53,350 samples)

6) select high metamerism color set (65 samples)

Further, the accumulated reflection spectra from each of the above sets was illuminated via one of four common cinema lighting sources for the simulations.

1) CIE Illuminant D65

2) CIE fluorescent, F2

3) CIE Illuminant A

4) Measured Hydrargyrum Medium-arc Iodide lamp, HMI

In the case of all training set spectra and all illuminants, optimization was performed with one permutation followed by performance verification with each of the other permutations.

Previously published indices were used to quantify observer metamerism magnitude and observer variability. For the former, color difference formulae between 2 stimuli are used. For the latter, three-dimensional CIELAB color error vectors are used. These indices are based on observer color-matching function demographics published from three previous efforts. The first set is produced by Fedutina and Sarkar, et al. and represents a statistical compartmentalization of the original Stiles-Burch observer experiments. The second set represents a sampling of the CIE2006 model published by CIE TC1-36 for observer response as a function of age and field of view. The final set derives from Heckaman, et al's work with Monte Carlo simulation of physiological transmission and absorption characteristics for the ocular media and l, m and s cones. Stimuli pairs input to the calculations derive from any established reference spectrum and a corresponding reproduction spectrum on the MPD.

The observer metamerism magnitude is the maximum individual observer average color difference across all the patches in a stimuli set. In this manner, the observer metamerism represents the on-average poorest color matching observer from the population of CMFs for the patchset. A slight variation of this index is based on measurement of the worst color difference patch across all observers in the given CMF set. This is thus the worst color match achieved across a full set of stimuli in the patchset considering all candidate observers. To minimize either of these metrics suggests a move towards improving the color match between two stimuli for all observers in a population and thus a minimization of observer metamerism magnitude.

Observer variability indices represent the mean CIELAB ellipsoid volume constructed from CMF-based error vectors in L*, a* and b* for each patch in a stimuli set. For the present work, covariance analysis is used to construct the ellipsoid volumes from individual observer CIELAB error vectors with a 90% statistical significance.

The objective for the present simulations was to identify the most robust training spectra, illuminant and optimization parameters to develop an idealized MPD design with the most effective number of primaries across the larger set of validation stimuli. The primary spectra modeling progressed in 2-stages. In a first screening simulation, spectral matches were predicted by way of simple linear reconstruction transforms as primary spectra candidates were iterated via the Gaussian parameters. The optimization was allowed to progress until a minimization of observer metamerism was achieved for the original stimuli versus the reconstructed stimuli. Once primary spectra for each training scenario were determined, verification stimuli sets were used to assess the metamerism robustness of the spectral reconstructions. For the second stage of simulation, primary spectra were retained from the screening models for each permutation. However, the reconstructed spectra for each stimuli set in this variation were computed via a fully constrained nonlinear optimization, permitting much better spectral reconstructions to be produced at the cost of computing speed.

As general result from the above processes, the primaries synthesized from varying the training patchset are significantly different across each permutation, offering a fairly strong signal in the modeling. The Kodak/AMPAS test spectra generate the most robust training results when all other patchsets are verified from these optimized primary spectra. This is further proven true in both the screening psuedoinversion models and the subsequent nonlinear optimizations, though an order of magnitude improvement in all indices is realized with the addition of the constrained spectral optimization. Varying starting guess parameters for the Gaussian curveshapes makes very little difference in the results. A small improvement is seen when the iterating peak wavelengths are permitted to vary subject to a monotonic vectorization versus each primary being binned in a restricted spectral span. The latter technique was hypothesized to be beneficial to enforcing full spectrum coverage across all visible wavelengths in the design though proved somewhat restrictive to the observer metamerism objective function. Relative to the number of primaries necessary to produce optimum metamerism reduction, N = 7 and 8 were shown to generate notable performance benefits versus systems with 6 or fewer total primaries. Finally, in regards to training illuminant, the HMI source yielded the most robust results while the CIE F2 source was poorest. For the ultimate RIT MPD design, however, a permutation consisting of all 4 light sources as training was implemented in conjunction with the Kodak/AMPAS reflectances.

Figure 1 summarizes the best simulated Gaussian primaries for an N = 8 design cascaded with the source spectrum of the consumer UHP lamp. In this plot are also shown modeled primaries utilizing commercially available color filters closest in performance to the Gaussian models. Upon further system characterization with actual measured filters acquired for the MPD, an alternate configuration of only 7 primaries was ultimately produced for laboratory experiments, consisting of the measured spectra shown in Figure 2. Currently, experiments are underway with observers to confirm the performance of the 7-channel system versus traditional 3-channel cinema displays (including laser systems) in minimizing observer variability when generating matches to reference spectra.

9394-26, Session 5

Gamut extension for cinema: psychophysical evaluation of the state of the art and a new algorithm

Syed Waqas Zamir, Javier Vazquez-Corral, Marcelo Bertalmio, Univ. Pompeu Fabra (Spain)

A) The need for gamut extension (GE):

The term "color gamut" refers to a set of colors that a display device is able to reproduce. State of the art digital cinema projectors have a wide gamut, but often the movies arriving at these new projectors have had their gamuts reduced so as to prevent issues with old or regular projectors (that have a smaller gamut). Therefore, in order to make use of the full potential of new projectors in terms of color, a process called gamut extension (GE) is needed, transforming colors from a smaller source gamut to a larger destination gamut. With the rise in digital display technology [1] it is foreseeable that in a few years the gamut of a standard digital cinema projector will be increased to a dramatic extent, often surpassing the color gamut capabilities of cameras. GE has received much less attention than gamut reduction, and to the best of our knowledge there aren't any works in the literature that deal with GE specifically for cinema.

B) State of the art approaches and challenges in GE:

The state of the art approaches presented in [3] perform GE by stretching out (linearly or non-linearly) the input signal to a wider gamut in order to

have a boost in the saturation of colors. This way of approaching the GE problem is simple but prone to several issues. One major problem is that the GE procedure may alter the artistic intent of the content's creator. Also, memory colors such as the blue of the sky and the green of the grass may look unrealistic with the application of GE, requiring a special treatment. Another challenge is to preserve skin tones, always a key issue in movie postproduction, but many Gamut Extension Algorithms (GEAs) fail to do so, as reported in [6].

C) First contribution: psychophysical evaluation of GEAs in cinema conditions:

As our first contribution we present the results of an evaluation which is, to the best of our knowledge, the first psychophysical evaluation of GEAs done specifically for cinema, using a digital cinema projector in cinematic (low ambient light) conditions. To this end, we evaluate GEAs in the following experimental setups:

1) Mapping from a small gamut to the DCI-P3 gamut: we map the source images from a small gamut, slightly smaller than the Rec. 709 gamut (which is the standard for HDTV) to the larger DCI-P3 gamut (the standard for digital cinema). We choose an input gamut smaller than Rec. 709 so as to have clear color differences in the reproductions, thus making the experimental task easier for observers. In this setup take part both color experts and non-experts.

2) Mapping from Rec. 709 to DCI-P3: also with experts and non-experts.

3) Comparison only in some regions of interest: experts mark some regions of interest and non-experts perform the test considering image results inside those regions only.

For each setup observers are asked to choose which of two different gamut-mapped versions of an image is more faithful to the wide gamut ground truth. We perform this subjective assessment using 8 observers with normal color vision, and on a set of 19 images that have a wide variety of spatial and chromatic characteristics. To avoid noise and add reliability to the psychophysical data, we repeat the experiment three times on each subject on three different days (to avoid learning). We analyze the psychophysical data using the work of Morovic [5], that is based on Thurstone's law of comparative judgement [8]. We also compare the results of these tests with those of current metrics for gamut mapping and color fidelity in HDR images in order to identify the best performing color difference metric for the GE problem.

D) Second contribution: a novel GE algorithm:

Zamir et al. [9] proposed a spatial GEA that adapts the image energy functional of Bertalmio et al. [2] to expand the color gamut. The work of [2] is closely related to the Automatic Color Enhancement (ACE) model of Rizzi et al. [7], and both the algorithms of [2] and [7] mimic global and local perceptual properties of the human visual system and are very closely related to the Retinex theory of color vision [4]. But the GEA of [9] produces results with loss of saturation and over-enhancement of contrast, which in turn makes it rank low in the psychophysical tests explained in section C.

As our second contribution we propose a new GEA based on [9], introducing several simple but key modifications that eliminate the problems with saturation and contrast just mentioned. The preliminary tests show that this new GEA generates results that are perceptually more faithful to the wide-gamut ground truth, as compared with the other four state of the art algorithms reported in [3]. Moreover, on visual inspection it can be observed that our GEA performs well in terms of skin tones and memory colors, with results that look natural and which are free from artifacts.

References:

[1] http://spectrum.ieee.org/consumer-electronics/audiovideo/lasers-coming-to-a-theater-near-you. [Online; accessed 29-July-2014].

[2] M. Bertalmio, V. Caselles, E. Provenzi, and A. Rizzi. Perceptual color correction through variational techniques. IEEE Transactions on Image Processing, 16(4):1058–1072, 2007.

[3] J. Laird, R. Muijs, and J. Kuang. Development and evaluation of gamut extension algorithms. Color Research & Application, 34(6):443–451, 2009.

[4] E. H. Land. The retinex theory of color vision. Scientific American, 237(6):108–128, 1977.

[5] J. Morovic. To Develop a Universal Gamut Mapping Algorithm. PhD thesis, University of Derby, UK, 1998.

[6] J. Morovic. Color gamut mapping, volume 10. Wiley, 2008.

[7] A. Rizzi, C. Gatta, and D. Marini. A new algorithm for unsupervised global and local color correction. Pattern Recognition Letters, 24:1663–1677, 2003.

[8] L. L. Thurstone. A law of comparative judgment. Psychological Review, 34(4):273–286, 1927.

9394-27, Session 6

Modeling the importance of faces in natural images

Bin Jin, Gökhan Yildirim, Cheryl Lau, Appu Shaji, Ecole Polytechnique Fédérale de Lausanne (Switzerland); Maria V. Ortiz Segovia, Océ Print Logic Technologies (France); Sabine Süsstrunk, Ecole Polytechnique Fédérale de Lausanne (Switzerland)

No Abstract Available

9394-28, Session 6

Bridging the gap between eye tracking and crowdsourcing

Pierre Lebreton, Technische Univ. Berlin (Germany); Evangelos Skodras, Univ. of Patras (Greece); Toni Mäki, VTT Technical Research Ctr. of Finland (Finland); Isabelle Hupont Torres, Instituto Tecnológico de Aragón (Spain); Matthias Hirth, Julius-Maximilians-Univ. Würzburg (Germany)

Visual attention constitutes a very important feature of the human visual system (HVS). Every day when watching videos, images or browsing the Internet, people are confronted with more information than they are able to process and analyse only part of the information in front of them [Google_2013]. The understanding of how people scan images or videos is of high interest [Jac91]. Many applications can be found, from perceptual coding to layout optimization of websites. A large amount of research has been done for the estimation of visual attention and fixations using mathematical algorithms to model human vision. These models are mostly bottom-up, designed for images, and may not work on new types of contents such as website or medical images, which require top-bottom analysis [ITTI]. Building such kinds of models is particularly challenging and subjective evaluation is the only reliable way to have accurate data. However, these types of studies require specific equipment (e.g. IR-based eye trackers), which can be highly expensive. Moreover, one of the challenges with visual attention studies is that they require a large amount of tests to go beyond individuals [Hansen_2010]. They are time consuming, may be intrusive and require specific setup that can be only obtained in a laboratory environment [ICCE].

In recent years crowdsourcing has become a particularly hot topic to scale subjective experiments to a large crowd in terms of numbers of test participants. It enables increasing the diversity of observers in terms of nationalities, social background, age, etc. and performing the evaluation in the environment the test participants are used to (out of the lab). The ability to reach a high number of participants behaving naturally can be a great added value of crowdsourcing to visual attention studies.

This paper describes a novel framework with the aim to bridge these two fields, by enabling crowd-sourced eye tracking experiments. The framework is divided into two parts: a client and a server side. On the client side, the test participant uses his personal computer to access a web page which records his face using the webcam of his device and streams it to a distant server via WebRTC. The location of every mouse click within the page is also transmitted to the server. On the server side, the video of the participant is processed offline. Once the position of the face is found, the positions of the eye centers and the eyelids are located. The displacement vectors between

the localized 'moving' points and stable facial points are used to form features. The mapping between features and screen coordinates of user fixations is derived using regression techniques. Those screen coordinates correspond to mouse click events: it is assumed that when a test participant clicks on a specific location of the screen and hits a button or any other graphical user interface object, the test participant is fixating at the specific object while clicking. A continuous calibration between eye position and click position during the entire length of the experiment is then possible. This simplifies the problem to interpolating every fixation between known key points.

To evaluate the performance of the proposed framework, a preliminary crowdsourcing experiment using voluntary testers was conducted. The experiment was divided into two parts. The first part was a game where test participants had to click on moving objects on the screen, serving as the calibration phase, through which the mapping function was derived. Then, in the second part of the test, participants were presented with a web page comprising of disks with different colors and were asked to click on the red ones. This latter part was used to measure the proposed system's accuracy by following the 'inverse' procedure; given the eye position data and the mapping function built, the ability of the algorithm to estimate the click positions was validated.

While the resolution of the video recorded by the webcam was only of 640x480 and VP8 compression at 1 Mbps was applied to the videos for matching bandwidth constraints of the crowdsourcing context, the performance of the proposed framework is promising. Across the current 8 test participants, the prediction errors are in average of 13.3% of the user's screen width and 16.1% of the user's screen height with a respective standard deviation of 9 and 4.5. It is planned to extend the experiment to more than 40 participants, using open calls in social networks and paid recruiting on commercial crowdsourcing platforms. The performances are good considering that the evaluation was done on extrapolated data: the validation followed the calibration whereas in the optimal application scenario, it is planned to estimate fixation between every calibration key point. Such condition was motivated by the fact that in some scenarios, user clicks may be rather sporadic. A clear limit of the framework is the uncontrolled lighting condition in the user side, which is a key aspect to ensure good performance and should be checked in a future through computer vision techniques. Likewise, the degree of head movement is one of the most influential factors which determines the system's accuracy. To conclude, although the proposed web-cam based gaze trackers present inferior performance compared to active light approaches, it can be applicable to applications where accuracy can be traded off for low cost, simplicity and flexibility.

[ITTI] Itti L, K. C. (2001), "Computational modelling of visual attention", Nature reviews. Neuroscience, 2(3) pp. 194-203.

[ICCE] Hupont, P. Lebreton, T. Mäki, E. Skodras, M. Hirth, "Is affective crowdsourcing reliable?", 5th IEEE International Conference on Communications and Electronics, ISBN 978-1-4799-5051-5, Vietnam, 2014.

[Jac91] Robert J. K. Jacob. 1991. The use of eye movements in human-computer interaction techniques: what you look at is what you get. ACM Trans. Inf. Syst. 9, 2 (April 1991), 152-169.

[Google_2013] Sandra P. Roth, Alexandre N. Tuch, Elisa E.D. Mekler, Javier A. Bargas-Avila, Klaus Opwis, "Location matters, especially for non-salient features – An eye-tracking study on the effects of web object placement on different types of websites", International Journal of Human-Computer Studies, vol. 71, pp. 228-235, 2013.

[Hansen_2010] Hansen, D. W., & Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32(3), 478-500.

9394-29, Session 6

Visual saliency in MPEG-4 AVC video stream

Marwa Ammar, Marwen Hasnaoui, Mihai Mitrea, Télécom SudParis (France); Patrick Le Callet, Univ. de Nantes (France)

Visual saliency maps already proved their efficiency in a large variety of image/video communication, covering from selective compression and channel coding to watermarking. Such saliency maps are generally based on different characteristics (like color, intensity, orientation, motion, ...) computed from the pixel representation of the visual content1.

The present paper resumes and extends our previous work2 devoted to the definition and evaluation of a saliency map solely extracted from the MPEG-4 AVC stream syntax elements.

The MPEG-4 AVC saliency map definition is structured at three levels.

First, the static and temporal features are extracted directly at the stream level, from each and every 4x4 macroblock in the I and P frames, respectively. The considered static features are: (1) the intensity, computed from the residual luma coefficients, (2) the color, extracted from the chroma residual coefficients and (3) the orientation, given by the variation (gradient) of the intra directional prediction modes. The motion feature is considered to be the variation (gradient) of the differences in the motion vectors.

Secondly, individual saliency maps are computed for the four above-mentioned features (intensity, color, orientation and motion). The saliency maps are obtained from feature maps following three incremental steps. First, the outliers are detected and eliminated. Second, an average filtering with fovea size kernel is applied. Finally, the maps are normalized to belong to a dynamic range of [0, 1].

Third, a static saliency map is obtained by merging the intensity, color and orientation maps according to an equal weighted addition.

The final saliency map is obtained by fusioning the static and the motion maps according to a weighted addition over the static saliency map, the motion saliency map and their element-wise multiplication.

This saliency map is validated by both a confrontation with the ground-truth and by its integration into a robust watermarking application.

In order to evaluate the coherency between the above-defined saliency map and the ground-truth, we considered a corpus of 8 video sequences, of 10 s each - this corpus is available for free downloading3. This video corpus comes across with density fixation maps obtained by averaging the visual attention of 30 human observers, captured by an EyeTracker device. The differences between the MPEG-4 AVC saliency map and the EyeTracker density fixation maps are evaluated by computing the KLD (Kullback-Leibler divergence) and the AUC (area under the ROC curve), according to the Borji's code4. The KLD results in an average value of 0.83; for comparison, two state-of-the-art methods acting in the pixel domain 5,6 result in KLD values of 1.01 and 1.06 respectively (of course, these values are computed on the same corpus and with the same KLD implementation). In such an experiment, the lower the KLD value, the closer the saliency map to the density fixation map. When computing the AUC, the binary density fixation map was obtained by applying a threshold equal to a half of its maximal value. The AUC values obtained for the MPEG-4 AVC saliency map and for the same two state-of-the-art algorithms are 0.84, 0.69 and 0.79 respectively. Note that the closer to 1 the AUC value, the better the saliency map.

The applicative performances are evaluated under a robust m-QIM watermarking framework7. The video corpus consists of 6 videos sequences of 20 minutes each. They were encoded in MPEG-4 AVC Baseline Profile (no B frames, CAVLC entropy encoder) at 512 kb/s. The GOP size is set to 8 and the frame size is set to 640x480. The MPEG-4 AVC reference software (version JM86) is completed with software tools allowing the parsing of these elements and their subsequent usage, under syntax preserving constraints. For prescribed data payload (of 30, 60, and 90 bits/second) and robustness (BER of 0.07, 0.03, and 0.01 against transcoding, resizing

and Gaussian attacks respectively), the saliency information increase the transparency by an average value of 2 dB.

To conclude with, the present study demonstrates that the visual saliency map can be extracted directly from the MPEG-4 AVC stream syntax elements. When compared to the state-of-the-art methods acting in the pixel domain, average relative gains of 0.2 and 0.13 are obtained in KLD and AUC, respectively.

However, the main benefit of the MPEG-4 AVC saliency map is its reduced complexity: note that by computing the saliency map directly from the MPEG-4 AVC stream, complex operations like DCT transformation, prediction, motion estimation and compensation are avoided. This way, a gain by a factor of 20 is obtained in the speed (corresponding to the following hardware configuration: Intel(R) Xeon W3530, 2.8 GHz, 12GB of RAM).

[1] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," IEEE Trans. Pattern Anal. Mach. Intell., vol. 20, 1998.

[2] Ammar, M., Mitrea, M., & Hasnaoui, M. (2014, February). MPEG-4 AVC saliency map computation. In IS&T/SPIE Electronic Imaging (pp. 90141A-90141A). International Society for Optics and Photonics, 2014.

[3] ftp://ftp.ivc.polytech.univ-nantes.fr/IRCCyN_IVC_Eyetracker_SD_2009_12/

[4] https://sites.google.com/site/saliencyevaluation/evaluation-measures

[5] Ming-Ming Cheng, Jonathan Warrell, Wen-Yan Lin, Shuai Zheng, Vibhav Vineet, Nigel Crook. "Efficient Salient Region Detection with Soft Image Abstraction", ICCV 2013.

[6] Seo, H. J., & Milanfar, P. (2009, June). Nonparametric bottom-up saliency detection by self-resemblance. In Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on (pp. 45-52). IEEE.

[7] M. Hasnaoui, M. Mitrea, "Multi-symbol video watermarking", Singal Processing: Image Communication, in Press (http://dx.doi.org/10.1016/j.image.2013.07.007), 2013.

9394-30, Session 6

Learning visual balance from large scale datasets of aesthetically highly rated images

Ali Jahanian, S. V. N. Vishwanathan, Jan P. Allebach, Purdue Univ. (United States)

Psychological studies show that visual balance is an innate concept for humans, which influences how we perceive visual aesthetics and cognize harmony. In visual design, for instance, balance is also a key principle that helps designers to convey their messages. Photographers, specifically, create visual balance in the spatial composition of photos through photo cropping. Learning visual balance from the work of professionals in design and photography may help to enable the automatic design applications in layout creation, content retargeting, cropping, and quantifying aesthetics.

In prior work, visual balance is defined as "looking visually right" [1] and is studied under the "theory of rightness in composition" [2-3]. Balance is considered in two general categories: symmetry and asymmetry [1], which in any case relates to harmony [4]. One of the central theories around balance is perhaps Arnheim's structural net [5], in which he hypothesizes that there are nine hotspots (including the center) on any visual artwork, and identifies their locations. Although in prior work, Arnheim's net is studied through psychophysical experiments, by asking participants' opinions about special arrangements of visual elements in paintings and photos, to the best of our knowledge, the present work is the first design mining framework on large scale datasets of images for evaluating Arnheim's theory.

In this paper, we examine Arnheim's hypothesis about balance through design mining a dataset of 120K images. These images are obtained from a professional photography website, 500px [6], and have at least 100 user likes (some of these images have several thousands of likes). Because visual balance is stimulated by the interaction of visual elements such as lines, color, texture, and orientation in an image, we run our design mining on the images' saliency maps. This decision is justified according to the fact that this same set of visual elements comprise the underlying features in the saliency map models. Having computed the saliency maps of the images, we then model these maps as a mixture of Gaussians by utilizing GMM and EM techniques. In our modeling, we initially position the Gaussians in the same places that Arnheim locates his visual balance hotspots on an image. We describe the adaption of GMM and the scalability considerations in processing large scale datasets in our framework.

Our inferred Gaussians align with Arnheim's hotspots, and confirm his structural net, specifically the center and the symmetry of the net. Moreover, because our design mining framework is able to associate the semantic tags (e.g. landscape, architecture, fashion, etc.) of the images with the derived Gaussian mixtures, we are able to cluster some general visual balance templates and to establish their linkage to the tags. This may lead to recommendations for visual layout of photos and visual design in general.

References

[1] S. Lok, S. Feiner, and G. Ngai, "Evaluation of visual balance for automated layout," In Proceedings of the 9th International Conference on Intelligent User Interfaces, ACM (2004), 101–108.

[2] P. J. Locher, "An empirical investigation of the visual rightness theory of picture perception," Acta Psychologica, Elsevier, vol. 114, pp. 147-164, 2003.

[3] I. C. McManus, K. Stöver, and D. Kim. "Arnheim's Gestalt theory of visual balance: Examining the compositional structure of art photographs and abstract images," i-Perception, vol. 2, no. 6, pp. 614, 2011.

[4] F. Samuel, and D. Kerzel, "Judging whether it is aesthetic: Does equilibrium compensate for the lack of symmetry?," i-Perception, vol. 4, no. 1, pp. 57, 2013.

[5] Art and visual perception: A psychology of the creative eye, Rudolf Arnheim, University of California Press, 1954.

[6] 500px. http://www.500px.com/.

9394-31, Session 6

Assessing the influence of combinations of blockiness, blurriness, and packet loss impairments on visual attention deployment

Alexandre F. Silva, Mylène C. Q. Farias, Univ. de Brasília (Brazil); Judith A. Redi, Technische Univ. Delft (Netherlands)

In this work, our focus is on better understanding the influence of artifacts on visual attention deployment, towards the improvement of objective quality metrics. Different from previous studies, which investigated artifacts as well as their impact on visual attention in isolation (e.g., only blockiness or only packet loss), we are interested in measuring the impact of combinations of such artifacts on video saliency. In fact, the perceptual impact of combined artifacts itself has been poorly studied in the past, except for studies on combination of artifacts in SD video. We tracked eye-movements of subjects in a quality scoring subjective experiment. The test sequences used in the experiment contain a set of combinations of the most relevant artifact found in digital video application: packet-loss, blockiness and blurriness.

9394-32, Session 7

Hue tinting for interactive data visualization

Jonathan I. Helfman, Agilent Technologies, Inc. (United States)

A general problem in Data visualization is 'How to best choose colors to show numbers?' 'Hue tinting' addresses a more specific version of this problem: 'How to best choose colors to show numbers while 1) using hue to mark only relevant aspects of the data and 2) minimizing color-related distortions?'

People naturally use color to identify objects and separate objects from backgrounds visually. People should be able to use color in data visualization in a similar way: to identify features in their datasets. Unfortunately, most visualization systems use color in a way that is rarely meaningful. You can supply your own colormap or switch between a large set of default colormaps, but none of these may be a good match for the dataset. To help make the colors meaningful, some systems let you warp or distort the colormap in an attempt to match the non-linearities in the colormap's brightness function with non-linearities in the distribution of data values, in an ad-hoc and error-prone manner. 'Hue tinting' is a set of visualization interactions that make it possible to use color in ways that are meaningful and specific to visualization tasks. Hue tinting provides a direct method for identifying the data values associated with a range of colors and the colors associated with a range of data values. Like colorizing a black-and-white photograph, Hue tinting lets you use color to select, identify, and mark relevant portions of your data without distorting the brightness of the underlying grayscale visualization.

Hue tinting is related to previous work on choosing color for static visualizations (see Ware 1988, Travis 1991, Healy 1996, Rogowitz and Treinish 1996, Borland and Taylor 2007, Moreland 2009). Because hue tinting is an interactive technique, it is also related to previous work on interactive data visualization (Hanrahan 2000, Heer 2005, etc.). Hue tinting is motivated by examples in the field of electronic measurement, where false color visualizations such as spectrograms and persistence spectra are commonly integrated with the user interfaces of complex measurement instruments, which are used by highly skilled electronic and electrical engineers (Witte 1993, Agilent 2012).

To model and quantify brightness distortions, 'brightness functions' of colormaps are defined and calculated. Standard line fitting techniques are used to measure the linearity of brightness functions. Brightness functions are also used to calculate Tintmaps, which are used to add hue to a grayscale colormap without introducing brightness distortions. Prototypes have been implemented in MATLAB and shown to electrical engineers, designers, and planners of electronic measurement test equipment.

'Hue tinting' is a new methodology for using color in visualization that minimizes the major problems associated with using color to show numbers (brightness distortion, visual segmentation, false edges). Because Hue tinting is interactive, it can be used to minimize or eliminate problems associated with matching colors in an image against a key or scale. Hue tinting also simplifies certain common measurements, such as identifying and validating the range of data values associated with a feature. Hue tinting provides a direct way to use color to show numbers in a way that is meaningful for any combination of dataset and task.

9394-33, Session 7

On the visualization of tetrachromatic images

Alfredo Restrepo, Edisson Maldonado, Univ. de los Andes (Colombia)

By a tetrachromatic image we mean here either an RGB+UV or a NIR+RGB image.

In the same way that a colour blind person cannot differentiate certain pairs of colours that normal trichromats can, it is very likely that normal trichromats cannot differentiate pairs of colours that a tetrachromatic animal does. Likewise, a camera with four types of photoreceptor can differentiate between light spectra that give rise to equal responses in an RGB camera. The question then arises of how can a normal trichromat person attempt to see a tetrachromatic image?

We propose 2 techniques, both of which use the time dimension. This makes it difficult to use the techniques to visualise tetrachromatic videos so, we deal here only with the visualisation of tetrachromatic (static) images.

In one of the visualisation techniques we use "movies" (as in a Matlab movie) made of frames that are trichromatic renderings of tetrachromatically-processed versions of the tetrachromatic image to be visualised, where a parameter in 4-Runge space varies across the versions.

In the other technique, we give a "dynamic texture" to regions of pixels having a UV or, alternatively an IR component. The texture can be for example a superposed "granularity" (texture) that "vibrates" in time, where the density of the grains depends on the amount of UV (or NIR) present in the tetrachromatic colour.

For each of the techniques and for different types of images, we investigate good choices of the parameters involved.

9394-34, Session 7

Evaluating the perception of different matching strategies for time-coherent animations

Javier Villegas, The Univ. of Arizona (United States); Ronak Etemadpour, Oklahoma State Univ. (United States); Angus G. Forbes, Univ. of Illinois at Chicago (United States)

Non-photorealistic animations can be generated directly from video data. We explore an analysis/synthesis approach to generating animations, where features are extracted from the video and then information from those features is used to create new graphic elements. However, a main issue with our approach, as with previous approaches, is that the frames of the video are analyzed independently of each other, leading to an unpleasing non-continuity in the elements that form the resulting animation. That is, elements can appear or disappear suddenly, or move abruptly and erratically. In this paper, we propose and evaluate the use of combinatorial optimization algorithms to match the elements used for synthesis on each frame in order to reduce artifacts and to create animations that are perceived as temporally coherent.

Many current image-based non-photorealistic techniques focus on imitating the appearance of hand painted styles. These techniques therefore do not have a ground-truth reference for how a moving version should look. Moreover, new approached to stylized video (in addition to those that explicitly try to solve the problem of maintaining temporal coherence) can be used to define the way in which elements should move and evolve. In this work, we present three terms that can be used to describe the perceptual qualities of the animation sequences generated through our approach. We introduce the terms "smoothness," "cohesion," and "accuracy" to assess the differences between various parameter matching algorithms. Smoothness is a desirable quality of the animation that implies that the image elements should change slowly with time. Abrupt changes on position, size, shape, etc., are distracting and destroy the feeling of continuity. Cohesion is the spatial equivalent of smoothness. In a cohesive animation, features that are close to each other should behave in a similar way. Accuracy is an indication of how well the recreated image resembles the original analyzed picture. We introduce a set of metrics that quantify these terms and that can be applied to parametric spaces of any finite dimension. We then compare the objective description of the generated sequences (using these metrics) with a user study in order to demonstrate that these terms correlate to user perception.

Different strategies for the generation of time-coherent animations from video data can be found in the literature; most of them focus on the problem of placing and propagating marks, such as brush strokes, through time. However, few of them use a combinatorial matching approach. Our

work is original since we propose the use of different matching optimization criteria as part of the creative control. Moreover, we explore the perceptual differences that are obtained when different optimization criteria are used.

Another contribution of our work is the introduction of metrics that can be used to describe the strengths and weakness of each of the matching strategies we explored. These metrics may be useful for future evaluations of stylized video. Moreover, we also describe and discuss our user study that proved that our objective metrics are indeed perceptually relevant. For our user study we created an artificial sequence that was used as a ground truth, in which no matching algorithm was required. Then, we recreated alternative sequences by filling up intermediate frames using different matching algorithms. All the resulting animations were presented to a group of 30 students and we asked them to rate each sequence in terms of "smoothness," "cohesion," and "accuracy" on a five-step Likert scale. We then asked them to perform pairwise comparisons between matching techniques and we evaluated the results for statistical significance.

Based on the results of our work, we propose different scenarios where each optimization algorithm can be used most effectively in creative projects.

9394-35, Session 7

Shared digital spaces

Bernice E. Rogowitz, Visual Perspectives Research and Consulting (United States); Paul Borrel, IBMEC/RJ (Brazil)

This research explores how to create the impression that people who are geographically separated are sharing the same space. SKYPE has revolutionized the ability to provide instant video connection between different locations, and video conference technology goes even further to provide a sense of context. Our goal is to go beyond seeing each other's environment, to create a shared space in which the person has a strong impression that they are physically sharing the room with someone far away. This paper explores the factors that influence the perception of sharing the same space, and constructs several technical scenarios that embody these principles. We explore, in particular, the application where a child is in a hospital far from family and friends. A shared digital environment would help mitigate this isolation. We present a novel scheme, based on sensors, computer graphics and projection technology, which is designed to provide a truly immersive experience without recourse to typical virtual reality methods such as head-mounted displays. The goal is to create a shared experience that allows people in separate geographies to feel that they are in the same physical space.

9394-36, Session 8

Examples of challenges and opportunities in visual analysis in the digital humanities *(Invited Paper)*

Holly E. Rushmeier, Yale Univ. (United States); Ruggero Pintus, Ctr. for Advanced Studies, Research and Development in Sardinia (Italy); Ying Yang, Yale Univ (United States); Christiana Wong, David Li, Yale Univ. (United States)

The massive digitization of books and manuscripts has converted millions of works that were once only physical into electronic documents. This conversion has made it possible for scholars to study large bodies of work, rather than individual texts. This has offered new opportunities for scholarship in the humanities. A lot of work has used optical character recognition and focused on the textual content of books. Machine learning techniques such as topic modeling have been applied to learn about topical trends over time, and attribute authorship to texts. New work is emerging that is analyzing the visual layout and content of books and manuscripts. We present two different digital humanities projects in progress that present new opportunities for extracting data about the past, with new challenges for designing systems for scholars to interact with this data. The first project we consider is the layout and spectral content of medieval manuscripts. The second in the analysis of representations in the complete archive of Vogue magazine over 120 years.

The first project we consider is a collaboration at Yale between Computer Science, Digital Collections, English and Classics entitled DIgitally Enabled Scholarship of Medieval Manuscripts. Within this work, we have two initiatives. The first is analyzing the visual layout in versions of the Book of Hours. Multiple copies of the Book of Hours are by design nearly identical in content. Of interest to scholars though is how the individual copies vary, with the inclusion of secular poetry or music. The emergence of such variations is important in understanding the history of the concept of a book. Finding variations without automated analysis is a long and tedious visual inspection process. We are developing algorithms to segment pages into various types of text blocks and illustrations. We then can present the scholar with samples of particular visual features that are found in a large collection of manuscripts. Challenges in this research thread are developing methods for humanist scholars to indicate the visual features they are interested in finding, and presenting the results in a meaningful manner. The second initiative is multispectral scanning of manuscripts to examine material variations. Changes in pigments and form can provide clues about where the manuscripts were produced. Multispectral images, including data from the ultraviolet and infrared bands, can provide insight into where changes in materials occur. Changes in pigment can be quite subtle, and determining what variations are significant depends on scholarly interpretation. The challenges in this thread are narrowing down many thousands of possible changes to those that may be of interest to scholars, and allowing scholars to view and make notes about these changes.

The second project is the analysis of the Vogue magazine data archive. This archive was provided to Yale with pages that were segmented and annotated. The archive provides a unique view of attitudes about and representations of women. Textual analysis has already shown and validated clear trends in editorial policy and the viewpoints of each editor-in -hief. Visual analysis can help form some hypotheses of both the perceptions of women and perhaps how women perceive themselves. In early work we have developed a system to browse the segmented page images and the metadata associated with them. We used the segmented pages in an initial study making use of face finding algorithms to collect statistics about images of women. Three effects were studied. First, we used the ratio of the facial area to the area of the full image as a proxy for measuring faceism. In analyzing covers, faceism appears to increase in the 1960's and 70's, and then decrease again in the 1980's. Trends in faceism varied for images within the magazine versus the cover images. Second we used the ratios between various facial measurements, such as distance between eyes and size of eyes, to look for variations of the "perfect face" as indicated by faces on the magazine cover. Finally, the position of faces in different classes of images (i.e. covers, fashion shoots, advertisements) were studied to look for trends in photographic composition. In general strong trends were hard to find, but the data show promise for further investigation if robust algorithms for identifying additional visual features can be found.

While the digital humanities projects described here are quite different, they lead to a definition of a common tool set that can be used across the humanities. They also provide guidance to the understanding of the potential and limitations of visual analysis that should be included in modern humanities education.

9394-37, Session 8

From A to B via Z: strategic interface design in the digital humanities *(Invited Paper)*

Milena Radzikowska, Mount Royal College (Canada); Stan Ruecker, Illinois Institute of Technology (United States); Stefan Sinclair, McGill Univ. (Canada)

It is well known that change of any kind, although it can be exhilarating, can also be challenging. Take for example an organization where employees are faced with a condition of positive change, both for the organization

and for them. They may have had years of experience in doing their jobs and being evaluated according to a given set of criteria. Now, they will find themselves faced with a new set of tasks and revised evaluation criteria that are at least to a certain extent unknown. This circumstance puts pressure on the employees and the organization, and the question is whether or not the positive changes will be sufficiently beneficial to warrant the additional stress (Jansson 2013).

If we consider the Humanities and the Digital Humanities as organizations, it is possible to understand that both are experiencing additional stress due to change. In the case of the Humanities, the agents of change are increasingly associated with the Digital Humanities in the form of the innovations they are producing. In the Digital Humanities, the stress arises from the nature of invention, where new practices, systems, data stores, and standards are part of normal scholarly activity. In this paper, we discuss an approach to interface design that we have developed over the past fifteen years, as our awareness has grown that change is a process that is most successful when it can be managed as part of a slowly shifting sense of identity (Paulsen et al. 2009).

We first became aware of the problem during the communication process between digital humanists working on research teams as designers and digital humanists or computer scientists working as programmers (Ruecker et al, 2008). Recognizing that both groups routinely make significant intellectual contributions to a project (e.g. Blandford et al. in review; Mehta 2009), there is nonetheless often still a tendency for designers to imagine systems that initially strike programmers as too difficult, too unusual, or simply too odd, often perhaps unnecessarily so. There are several reasons for this tendency, which can be healthy in the sense that it implies the need for justification of choices that are off the current default, but can be less than useful in constantly challenging the need for innovation in design research.

First is that most modern programming languages come with libraries of design elements that are pre-programmed so that the development process is faster, easier, and less prone to error. To abandon these elements is to make the programming process slower, harder, and more prone to error. However, for an experimental system, especially a system that is predicated on the need for interactive visualization, components that come “out of the box” are less likely to provide evidence related to research questions about novelty of information presentation and use. For example, every interactive browsing tool needs a search component. However, not every experimental system requires one, since the need is already well established and does not need to be established anew with each new prototype.

Second is that the academic programming community has a strong belief in the need to see a working system in order to properly understand and evaluate its potential. The alternative is “vaporware” or designs that have not been realized as prototypes. Unfortunately, the result of this emphasis is that as little time as possible is spent on design, which represents a delay in the process of creating a working prototype. What can go unrecognized is that design and functionality are inextricably connected (e.g. Ruecker et al. 2007), so that studies attempting to isolate the two are often less convincing than might otherwise be the case.

Third, as mentioned above, is that change itself is a source of additional stress, anxiety, and confusion. While this is true within the research team, it can be an even greater factor within the user community. For many Humanists, the scholarly effort spent on Digital Humanities experiments is simply wasted effort, since it is clear that existing practices, systems, data stores, and standards are sufficient to the task at hand, which is the hermeneutic or interpretive approach to the cultural objects of study, whether those be poetry, novels, plays, historical objects or documents, and so on. If there is a need at all for digital enhancements, then the process should consist of digital enhancements to approaches that have stood the various tests of academic rigour. Many Digital Humanists agree, and have focused their efforts on developing new systems that extend, for example, access to source materials.

The process looks something like this (Fig. 1). The researcher interested in an experimental system consults potential users of the system in order to understand their current best practices. For research purposes, we can describe those as situation A. An interesting set of research questions might be addressed by inventing a new system that involves a significant degree of change. We will call this constellation of technology, data, and processes situation B. However, the ideal for the user is not B, which is too far removed to be useful under present circumstances; better would be position A+1. In fairness, A+1 can be reasonably justified from every perspective except the production of significant research in the area of experimental systems. It costs less time and effort, involves fewer risks, and has the potential to produce immediate benefits for current users.

However, it also resembles incremental change rather than either experimental research or its natural concomitant, disruptive innovation. Therefore, in an attempt to sufficiently shift the relationship between identity and current best practices, we provide position Z, a design radical enough to be as far as possible outside the current bounds of discourse. In the process of considering Z, the scholarly community of Humanists and Digital Humanists alike has tended to shy back to A+1, spend some time in beginning to recognize its limitations, then asking instead for research in the area of position B.

We illustrate this process with reference to a range of current and past research projects, including Metadata Open New Knowledge (MONK), the mandala browser, and visualizations for decision support.

References

Blandford, Ann, Sara Faisal, Carlos Fiorentino, Alejandro Giacometti, Stan Ruecker, Stéfan Sinclair, Claire Warwick, and the INKE Research Group [submitted] “Programming is an Interpretive Act with Implications for the User: the Case of the Bubblelines Comparative Visualization of Search Results.” Digital Humanities Quarterly.

Jansson, Noora. (2013) “Organizational change as practice: a critical analysis”, Journal of Organizational Change Management, Vol. 26(6), pp.1003 – 1019.

Mehta, Paras, Amy Stafford, Matthew Bouchard, Stan Ruecker, Karl Anvik, Ximena Rossello, and Ali Shiri. “Four Ways of Making Sense: Designing and Implementing Searchling, a Visual Thesaurus-Enhanced Interface for Multilingual Digital Libraries.” Proceedings of the Chicago Colloquium on Digital Humanities and Computer Science. 1(1). 2009.

Paulsen, Neil, Diana Maldonado, Victor J. Callan, Oluremi Ayoko, (2009) “Charismatic leadership, change and innovation in an R&D organization”, Journal of Organizational Change Management, Vol. 22 (5), 511 – 523.

Ruecker, Stan, Milena Radzikowska and Susan Liepert. “The Introduction of Radical Change in Human-Computer Interfaces.” Presented at An-Institut Deutsche Telekom Laboratories, Technische Universität Berlin. Feb 19, 2008.

Ruecker, Stan, Stéfan Sinclair, and Milena Radzikowska. “Confidence, Visual Research and the Aesthetic Function.” Partnership: the Canadian Journal of Library and Information Practice and Research. 2(1). 2007.

9394-38, Session 8

Big data, social computing and civic engagement: What can digital humanities and visualization scientists teach each other? *(Invited Paper)*

Lyn Bartram, Simon Fraser Univ. (Canada)

No Abstract Available

9394-39, Session 8

Introducing digital humanities in an engineering school: new data; new approaches *(Invited Paper)*

Sabine Süsstrunk, Ecole Polytechnique Fédérale de Lausanne (Switzerland)

No Abstract Available

9394-56, Session 8

Can brain changes from art training reveal commonalities between the mechanisms of drawing and of music?

Lora T. Likova, The Smith-Kettlewell Eye Research Institute (United States)

No Abstract Available

Conference 9395: Color Imaging XX: Displaying, Processing, Hardcopy, and Applications

Monday - Thursday 9–12 February 2015

Part of Proceedings of SPIE Vol. 9395 Color Imaging XX: Displaying, Processing, Hardcopy, and Applications

9395-1, Session 1

Optimizing color fidelity for display devices using vectorized interpolation steered locally by perceptual error quantities

Marina M Nicolas, Fritz Lebowsky, STMicroelectronics (France)

High-end PC monitors and TVs continue to increase their native display resolution to 4k by 2k and beyond. Subsequently, uncompressed pixel amplitude processing becomes costly not only when transmitting over cable or wireless communication channels, but also when processing with array processor architectures. We recently presented a block-based memory compression architecture for text, graphics, and video enabling multi-dimensional error minimization with context sensitive control of visually noticeable artifacts. The underlying architecture was limited to small block sizes of 4x4 pixels. Although well suitable for random access, its overall compression ratio ranges between 1.5 and 2.0. To increase compression ratio as well as image quality, we propose a new hybrid approach that operates on larger block sizes. A simple search for pixel amplitudes that can be interpolated in combination with a vector carrying orientation information reduces the number of pixels that need to be transmitted. A detailed comparison with the prior algorithm highlights the effectiveness of our new approach, identifies its current limitations with regard to high quality color rendering, and illustrates remaining visual artifacts.

9395-2, Session 1

Demosaicking algorithm for the Kodak-RGBW color filter array

Mina Rafi Nazari, Eric Dubois, Univ. of Ottawa (Canada)

Digital cameras capture images through different Color Filter Array (CFA) and reconstruct the full color image. Each CFA pixel only captures one primary color information and other primary components are estimated with demosaicking algorithm. Demosaicking algorithms and CFA design methods are both crucial steps to restore an optimal output.

Since the traditional CFAs contain Red, Green and Blue filters, they are not robust to the noise. A new demosaicking algorithm using Kodak-RGBW is presented in this paper. The RGBW reduce the amount of noise using White pixels and the Kodak-RGBW is one of the most popular templates due to the large number of White filters comparing to the other RGBW templates. The new proposed demosaicking algorithm reduce the overlap between luma and chroma components.

Review of prior works

Digital cameras use different CFAs to capture images and employ different demosaicking scheme to find the full color image. A CFA is employed to measure three primary colors such that each CFA sensor is responsible for capturing only one tristimulus value at a certain pixel location. CFA patterns differ in the number and formation of pixels; the Bayer pattern is the most popular one with a 2x2 sensor array consisting two green, one blue and one red pixel [1].

Demosaicking process consists of receiving an incomplete CFA image and calculating missing information to retrieve the color image. A basic demosaicking scheme relies on bilinear interpolation of neighboring pixels color information [2].

There is a large number of demosaicking schemes employing the Bayer structure in the frequency domain. One of the most recent and well known algorithms which outperform previous approaches employs Least Squares Optimization technique [3].

We have developed a new demosaicking algorithm using the Kodak-RGBW pattern as a four-channel color filter array to enhance the quality of the display and reduce the noise in the sense of human vision perception [4]. The additional filter array is spectrally nonselective and isolate luma and chroma information.

Proposed Approach

We have studied and optimized the reconstruction techniques for a new sampling structure with four color components. The Kodak-RGBW CFA pattern is a 4x4 template containing eight white pixels, four green pixels, two blue and two red. This study involves the design of an appropriate demosaicking method and is applied on the Kodak image dataset. Since four-channel CFAs usually improve signal-to-noise ratio performance, we attempt to model the demosaicking steps using Kodak-RGBW pattern and simulate non-adaptive and adaptive demosaicking algorithms in Matlab software. A detailed optimization of filter parameters and the region of support will be addressed.

Results

This research reconstructs the 24 Kodak images using Kodak-RGBW pattern to compare them with the previous patterns.

The results contains the PSNR comparison between the least squares luma-chroma demultiplexing (LSLCD) method using Bayer, adaptive and non-adaptive demosaicking scheme using RGBW pattern for some sample images as well as the average PSNR over 24 kodak images.

Based on the results, the proposed adaptive demosaicking scheme using the RGBW pattern gives us a comparable set of PSNR with the LSLCD demosaicking scheme using Bayer pattern. On the other hand, the visual results show improvement in the reconstructed details and its colors. The visual results prove the RGBW pattern gives less false colors than Bayer pattern on these details. Then, future work will involve improving the filter design using the least-square methodology, as in [2].

References

[1] O. Losson, L. Macaire, and Y. Yang, " Comparison of Color Demosaicing Methods", vol. 162 of Advances in Imaging and Electron Physics, chapter 5, pp. 173 – 265, Elsevier, 2010.

[2] B. Leung, G. Jeon, and E. Dubois, "Least-squares lumachroma demultiplexing algorithm for bayer demosaicking," IEEE Transactions on Image Processing, vol. 20, no. 7, pp. 1885–1894, July 2011.

[3] E. Dubois, "Frequency-domain methods for demosaicking of bayer-sampled color images", Signal Processing Letters, IEEE, vol. 12, no. 12, pp. 847–850, Dec 2005.

[4] J. Adams, A.Deever, J. Hamilton, M. Kumar, R. Palim and B. Pillman,"Single capture image fusion with motion consideration", Computational Photography, pp. 63 -81, Oct 2010.

9395-3, Session 1

Subjective comparison of brightness preservation methods for local backlight dimming displays

Jari Korhonen, Claire Mantel, Søren O. Forchhammer, DTU Fotonik (Denmark)

Local backlight dimming is rapidly becoming a popular technique in high quality LCD screens for professional and home use. The basic idea of local backlight dimming is simple: the screen is divided in several segments, each with an individually adjustable backlight element. Different backlight intensities can be chosen for each segment, following the luma of the corresponding regions in the input image. For the television industry, the benefits of local backlight dimming are twofold: first, it improves contrast,

as the light leaking through dark pixels can be reduced by reducing the backlight intensity; and second, it saves energy, since full backlight intensity is not needed for segments with dark content.

In practical local backlight dimming displays, the resolution of the backlight segments is significantly lower than the display resolution. In addition, backlight segments are not optically isolated from each other, and this is why the light from each backlight element is mixed on a passive diffuser plate located between backlight elements and pixels of liquid crystals. Therefore, each pixel receives light from several backlight segments. In many typical real-life images there may be bright and dark pixels located close to each other, and since the variation in the backlight intensity tends to be very smooth, a compromise between light leakage in dark pixels and reduced intensity of bright pixels (clipping) is often needed. This makes local backlight dimming a very challenging optimization problem.

To date, scientific research in local backlight dimming has mainly focused on algorithms determining the backlight levels. Less attention has been paid to subjective comparison of techniques for enhancing luminance of clipped pixels, i.e. brightness preservation, in backlight dimming displays. Subjective evaluation studies of tone mapping algorithms have been reported in the related literature. However, the problem definition for brightness preservation with local backlight dimming is clearly different from classical tone mapping, since conventional tone mapping algorithms do not consider the local variations in the dynamic range of the output device, caused by non-uniform backlight intensity. To our knowledge, this study is the first systematic subjective quality assessment study comparing different brightness preservation algorithms for local backlight dimming displays.

Since the visible variations between different brightness preservation schemes are often subtle, we have assumed that differences may be difficult to quantify via conventional methods based on numerical ratings. Pairwise comparisons could provide more accurate results, but the problem is the large amount of pairs that need to be tested. This is why we have chosen a rank ordering method, where test subjects sort different versions of an image into an order of preference. The results of the subjective study show that in most cases locally adapted linear enhancement produces the most preferred visual outcome. This approach preserves the local uniformity well, but at the cost of reduced global uniformity and overall brightness. However, results show also some dependency on the content: for high contrast images, hard clipping seems to produce more acceptable results than for low contrast images.

9395-4, Session 1

Shading correction of camera captured document image with depth map information

Chyuan-Tyng Wu, Jan P. Allebach, Purdue Univ. (United States)

Nowadays, camera modules have become more popular and convenient in consumer electronics and office products. As a consequence, people have many opportunities to use a camera-based device to record a hardcopy document in their daily lives. However, unlike a traditional flatbed scanner or a portable sheet-feed scanner, it is easy to let undesired shading into the captured document image through the camera. Sometimes, this non-uniformity may degrade the readability of the contents. In order to mitigate this artifact, some solutions have been developed. But most of them are only suitable for particular types of documents. For example, some methods are designed to correct the shading of documents with a uniform bright color margin, and some are developed to handle documents with a certain shape.

In this paper, we introduce a content-independent and shape-independent method that will lessen the shading effects in captured document images. We want to reconstruct the image such that the result will look like a document image captured under a uniform lighting source. Our method utilizes the 3D depth map of the document surface and a look-up table strategy. The 3D information will be the key contribution to do the shading correction; and the look-up table method can reduce the real-time computational complexity.

We will first discuss the model and the assumptions that we used for this approach. Initially, we simply used a Lambertian reflectance model for our system. However, the results did not fully take the advantage of our algorithm. We did some experiments and comparisons to find a better model that would generate a more satisfactory outcome from our method. Theoretically, our model can enhance the contrast of the images. This model is the foundation of the shading correction, and it deals with single channel image. For RGB color images, we perform a color space transformation in advance and apply it to the brightness related channel with the same model. Properly choosing the color space to use is another important component of this method. Otherwise, the reconstructed image will not be color balanced.

Then, the process of creating the look-up table will be described in this paper. This look-up table takes consideration of both the performance of the camera device and the lighting condition. Basically, it will be a three dimensional table in order to handle any shape of the input document. Before doing the shading correction process, we need to create the table first. Even though the table generation is time-consuming, this one-time off-line preparation will not decrease the efficiency of the real-time process. The table look-up step is pixel-wise; and it depends on the position of each pixel in the 3D world coordinate system. This is where we utilize the additional depth map information of the document surface.

We implement this algorithm with our prototype 3D scanner, which also uses a camera module to capture a 2D image of the object. Some experimental result will be presented to show the effectiveness of our method. We test our algorithm with both flat and curved surface document examples to show how this approach can mitigate the unpleasant shading on them. Some intermediate experimental outcomes that help us choose the most suitable approaches are also included.

9395-5, Session 2

A robust segmentation of scanned documents

Hyung Jun Park, Ji Young Yi, SAMSUNG Electronics Co., Ltd. (Korea, Republic of)

The image quality of reprinted documents that were scanned at a high resolution may not satisfy human viewers who anticipate at least the same image quality as the original document. Moiré artifacts without proper descreening, text blurred by the poor scanner modulation transfer function (MTF), and color distortion resulting from misclassification between color and gray may make the reprint quality worse. To remedy these shortcomings from reprinting, the documents should be classified into various attributes such as image or text, edge or non-edge, continuous-tone or halftone, color or gray, and so on. The improvement of the reprint quality could be achieved by applying proper enhancement with these attributes. In this paper, we introduce a robust and effective approach to classify scanned documents into the attributes of each pixel. The proposed document segmentation algorithm utilizes simple features such as variance-to-mean (VMR), gradient, etc in various combinations of sizes and positions of a processing kernel. We also exploit each direction of gradients in the multiple positions of the same kernel to detect as small as 4-point text.

While many studies claimed that their suggested methods detect both halftones and text separately, they could not demonstrate that their approaches discern between halftone patterns generated in a resolution of low lines per inch (LPI) and small text or fine lines that are located in a cramped condition. The characteristic of low LPI halftone patterns is quite similar to the one of small text or fine lines in terms of signals in the spatial domain. That is, patterns resulting from alternative positions of thin lines and narrow spaces are not different from the low LPI halftone patterns in the conventional window size. To overcome this similarity, we apply the pre-smoothing filters to input documents in all color channels. Although our segmentation algorithm is established to work in the YCbCr color space, the performance in any other color space is the same as the one in the YCbCr color space. The classification is done in the hierarchy tree structure. First, we classify the scanned documents into text or image at the top level. Then, text and image are sub-classified into halftone or continuous-tone at the second level. For text, halftone and continuous-tone are divided into

color or gray at the third level. For image, halftone is divided into edge or non-edge, and continuous-tone image is not divided further. Experimental results show that our proposed algorithm performs well over various types of the scanned documents including the documents that were printed in a resolution of low LPI.

9395-6, Session 2

Text Line Detection Based on Cost Optimized Local Text Line Direction Estimation

Yandong Guo, Microsoft Corp. (United States); Yufang Sun, Purdue Univ. (United States); Peter Bauer, Hewlett-Packard Co. (United States); Jan P. Allebach, Charles A. Bouman, Purdue Univ. (United States)

Text line detection is a critical step for applications in document image processing. In this paper, we propose a novel text line detection method. First, the connected components are extracted from the image as symbols. Then, we estimate the direction of the text line in multiple local regions. This estimation is, for the first time, formulated in a cost optimization framework. We also propose an efficient way to solve this optimization problem. Afterwards, we consider symbols as nodes in a graph, and connect symbols based on the local text line direction estimation results. Last, we detect the text lines by separating the graph into subgraphs according to the nodes' connectivity. Preliminary experimental results demonstrate that our proposed method is very robust to non-uniform skew within text lines, variability of font sizes, and complex structures of layout. Our new method works well for documents captured with flat-bed and sheet-fed scanners, mobile phone cameras and with other general imaging assets.

9395-7, Session 2

Color image enhancement based on particle swarm optimization with Gaussian mixture

Shibudas Kattakkalil Subhashdas, Bong-Seok Choi, Ji-hoon Yoo, Yeong-Ho Ha, Kyungpook National Univ. (Korea, Republic of)

In the recent years, enhancement of low illuminated color images has become one of the hottest research issues. At present, researchers have developed various low illuminant color image enhancement methods based on retinex theory, color and depth histogram and space variant luminance map. Although these methods improve the image quality, many of them are incapable to address all the issues such as washout appearance, information loss and gradation artifacts.

This paper proposes a color image enhancement method which uses particle swarm optimization (PSO) to have an edge over other contemporary methods. To enhance a low illuminant color image, the brightness of an image is redistributed by PSO and chroma is also compensated according to the brightness. PSO is a swarm intelligence technique inspired by the social behavior of bird flocking and fish schooling. Compared to other evolutionary computation algorithms, PSO exhibit better convergence and computational simplicity unlike selection, crossover and mutation operations in genetic algorithm (GA). Therefore, PSO is a successful tool in solving many real-world problems.

The proposed method first transforms the input sRGB image to the CIEL*a*b* color space and then it divides the luminance histogram of the input image in CIEL*a*b* space into a group of Gaussian which is called Gaussian mixture. Each Gaussian in this mixture represents a set of pixels with closely related luminance value. The brightness of each pixel set can be changed independently by varying the position of the corresponding Gaussian. Therefore the brightness of the whole image can be improved by rearranging the Gaussian. Here, PSO is employed to control the Gaussian shift. PSO finds the optimal Gaussian shift to enhance the brightness of the image. Colorfulness and discrete entropy are used as the fitness function for PSO. The entire rearranged Gaussian combines to form the brightness enhanced histogram. Subsequently, the enhanced histogram is subdivided into sub-histograms based on the intersection point of rearranged Gaussian. Then histogram equalization is applied to each sub histogram to compensate the contrast of the image. The resulting histogram is used as the reference to adapt the luminance component of input image in CIEL*a*b* by histogram matching method. The saturation of the resultant image is low compared to the input image. Therefore, chroma compensation which is based on the relative chroma ratio of the input image is applied to the resultant image. The relative chroma ratio is to consider sRGB color gamut boundary in the CIEL*a*b* color space.

Experimental results show that the proposed method produces a better enhanced image compared to the traditional methods. Moreover, the enhanced image is free from several side effects such as washout appearance, information loss and gradation artifacts. This method also has an edge over other contemporary methods in terms of colorfulness and discrete entropy. The proposed method has been tested with different variants of PSO such as adaptive particle swarm optimization (APSO), heterogeneous multi swarm particle swarm optimization (MsPSO) and cooperative particle swarm optimization (CPSO) and the efficiency of these PSO algorithms to enhance the image is compared.

9395-8, Session 2

Image enhancement for low resolution display panels

Rakshit S. Kothari, Eli Saber, Rochester Institute of Technology (United States); Marvin Nelson, Michael A. Stauffer, Dave Bohan, Hewlett-Packard Co. (United States)

This paper presents a real time automatic image correction technique for improved picture/video quality in low resolution display panels. The primary objective of this project is to develop an algorithm which operates under real-time constraints (milliseconds) without affecting other imaging processes on the display panel hardware processor. Display panels, particularly lower end printers or scanners, are comprised of sparsely arranged LEDs with an offset which induces artifacts such as jagged edges and bright red dots. Due to the aforementioned offset in the LED pattern, diagonally adjacent red LEDs produce bright red dots in regions of high contrast, such as text, numbers or sharp edges in graphics. When perceived by the human observer, the artifacts result in discontinuous and jagged edges and poor image quality. The algorithm is designed as a two stage process, starting with the identification of pixels which could cause such artifacts followed by a color correction scheme to combat the perceived visual errors. Artifact inducing pixels are identified by a threshold on a vector color gradient of the image. They are further masked by a second threshold based on pixel levels. The algorithm then iterates through these identified pixels and activates neighboring blue and green LEDs as per the estimated color correction based on partitive spatial color mixing. For visual color continuity, the color levels of the adjacent LEDs are estimated as per the nearest pixel. Since the distance of the observer from the display panel is assumed to be greater than half a foot, the distance between the LEDs is considered negligible and a point source model is assumed for color mixing. Red dot artifacts occur as singlet, couplets or triplets and consequently three different correction schemes are explored. The algorithm also ensures that artifacts occurring at junctions, corners and cross sections are corrected without affecting the underlying shape or contextual sharpness. Red dot artifacts in relatively bright neighborhoods are perceptibly lower due to higher spatial color mixing and visual continuity. Based on the threshold, we compute the average brightness using which we classify local regions as 'bright neighborhoods' and conversely, 'dark neighborhoods'. The proposed algorithm mentioned in this paper tackles red dot artifacts in dark neighborhoods. The performance of our algorithm shall be benchmarked on a series of 30 corrected and uncorrected images using a psycho-visual metric recorded from human observations. The experiment will be conducted at a standard viewing distance from the display panel while

various images, corrected or otherwise, will be shown progressively. This algorithm is designed as a general purpose color correction technique for display panels with an offset in their LED patterns and can be easily implemented as an isolated real time post processing technique for the output display buffer on low end display panel processors without any higher order processing or image content information. All the above mentioned benefits are realized through software and don't require any upgrade or replacement of existing hardware.

9395-9, Session 2

Video enhancement method with color-protection post-processing

Youn Jin Kim, SAMSUNG Electronics Co., Ltd. (Korea, Republic of); Youngshin Kwak, Ulsan National Institute of Science and Technology (Korea, Republic of)

Improving image quality has been of a great importance for digital consumer electronic devices and there have been various efforts on image enhancement. Those methods, however, are mostly designed in the context of still images and have been simply used on video as well. There are a couple of aspects that should be taken into account for video enhancement. For example, most video contains compression artifacts still remained after being decompressed and the color domain is usually required to be converted into that of a display device such as RGB in the end. The current study is aimed to propose a post-processing method for video enhancement by adopting a color-protection technique. The color-protection intends to attenuate perceptible artifacts due to over-enhancements in visually sensitive image regions such as low-chroma colors, including skin and gray objects. In addition, reducing the loss in color texture caused by the out-of-color-gamut signals is also taken into account. The proposed color-protection post-processing method is comprised of the three main steps: contents analysis, gain revising, and out-of-color-gamut (OCG) signal removal. The contents analysis block provides a sensitivity probability map (SPM) for a given input image and the initial gain level of video enhancement can be revised in the gain revising block. A value of SPM for each pixel in a given input video scene is estimated as the following three steps: a color space conversion from YCbCr to chroma and hue, detecting the low-chroma color, and adjusting a detection threshold and merging. The OCG removal block re-maps OCG signal; it is bypassed if the pixel values are placed inside the display destination gamut. The OCG removal block intends to preclude the scalability distortion occurred during a device color space conversion by means of an adaptive chroma mapping whilst preserving hue and lightness of the input color signal.

The performance and merits of the proposed color-protection method can be classified into the three main features: noise, chromatic adaptation, and color texture. 1) Attenuating exaggerated transitions by the means of gain revising in the low-chroma regions can be the first merit of the color-protection method and suppressing the undesirable noise boost-up can also be come up with. 2) Maintaining the adaptive white point of a scene during color processing is the second merit of the method; therefore chromatic adaptation of the human visual system, which may be involved in perceptual contrast and dynamic range as well, should not be confounded. Since the adaptive white point of an image relies upon neutral color contents, preserving chromaticness in the low-chroma region should be substantial. 3) Preserving the fine details in highly chromatic texture region can be the third. Such high frequency information is often lost and seen blurred when a device color space conversion is not carefully designed. The OCG removal block in the color-protection mainly contributes to this part.

9395-10, Session 2

Fast Algorithm for Visibility Enhancement of the Images with Low Local Contrast

Ilya V. Kurilin, Samsung Advanced Institute of Technology (Russian Federation); Ilia V. Safonov, National Research Nuclear Univ. MEPhI (Russian Federation); Michael N. Rychagov, Sergey S. Zavalishin, Samsung Advanced Institute of Technology (Russian Federation); Donghyeop Han, SAMSUNG Electronics Co., Ltd. (Korea, Republic of); Sang Ho Kim, Samsung Digital City (Korea, Republic of)

There are several classes of images which require enhancement but cannot be improved without user interaction in spite of significant progress in development of automatic image correction techniques. One of examples of such images can appear during digital copying. Sometimes, high-quality hardcopies possess strongly different contrasts between foreground and background for various areas of the documents: black text on dark-gray background and white text on light-gray or color background. Such hardcopies have a very wide dynamic range of tones as High Dynamic Range (HDR) scenes. When the originals with wide dynamic range of tones on conventional MFP-devices with color depth 24 bpp are scanned, Low Dynamic Range (LDR) image with poor local and sometimes global contrast. A visibility of a part of text and graphics should be improved.

For contrast improvement in each local area, specific tone transformation function can be defined. Our idea is the following: let's define S-shaped curve that is able to vary adaptively parameters depending on local and global tones distribution. We use cubic Hermit spline as S-shaped curve. In order to calculate starting point and ending point, global histogram of brightness is analyzed. A shape of a spline depends on local distribution of background and foreground pixels in a local region. A threshold by Otsu for brightness can be used for proper setting up tangent vectors. In order to prevent forming of visible artifacts, it is necessary to provide smooth alteration spline shape for adjacent local areas and neighbor pixels. Thresholds Ki for overlapped blocks and store Ki in matrix M is properly calculated. Further, we apply low-pass filter for smoothing the matrix of thresholds. Each pixel of source image is transformed by unique curve that has tangents depending on parameter K that is extracted from the matrix M with application of bilinear interpolation.

There are a lot of effective approaches for speeding up of proposed technique. Calculation of cubic Hermit spline for each pixel has high computational cost. Instead, 2D lookup table (LuT) with size 256x256 pixels can be pre-calculated. Size of the initial 2D LuT is 64 kB; it can be too big for embedded implementation. We decimate the LuT by 4 times, i.e. 2D LuT with size 64x64 is stored and corresponding values from the LuT by bilinear interpolation are taken. In order to decrease number of calculations, estimation of global and local thresholds for downsampled grayscale copy of initial RGB image should be carried out. For smoothing matrix M, we use box-filter with kernel size 5x5. There is effective algorithm for box-filter based on summed area table (also known as integral image). In addition, the algorithm can be parallelized effectively for modern multi-core CPU by means of OpenMP.

Test pattern for comparison of suggested approach with existing well-known techniques is (developed. Proposed method outperforms all analyzed solutions: it improves all local areas, significantly increases average and minimal Michelson contrast (also known as visibility metrics) and does not lead to any visual artifacts.

Application area of proposed method is not limited by the improvement of scanned images only. Nowadays, one of important tasks is correction of photos and video affected by fog/haze. Described algorithm can be applied without any modifications for enhancement of such type of images. Achieved dehazing via implementation of disclosed algorithm is demonstrated in the paper. Initial image with fog looks dull and ugly but enhanced image looks nice and vivid.

9395-11, Session 2

Online image classification under monotonic decision boundary constraint

Cheng Lu, Jan P. Allebach, Purdue Univ. (United States); Jerry K. Wagner, Brandi Pitta, David Larson, Hewlett-Packard Co. (United States); Yandong Guo, Purdue Univ (United States)

All-in-one (AIO) devices are now widely used for various purposes. They typically provides scanning, copying and printing services. One major issue for an AIO device is its copy quality. In order to optimally process different types of input images, multiple processing pipelines are included in an AIO device. Each of these processing pipelines is purposefully designed for one type of image. In our application, the AIO device includes three processing pipelines which target three types of images, respectively. Theses three types are: pure text, picture, and mix. An incorrect choice of processing pipeline (mode) when copying an image will lead to significantly worse output quality. For example, if a pure text image is chosen to be copied under the picture mode, the characters on the output image will have poor edge sharpness and low visual contrast both of which harm the reading experience. Different types of misclassifications do not cause equally severe image quality degradations. This suggests that we need to apply a discriminative training strategy, which will be discussed later. Given this problem, an embeded-firmware-friendly automatic classification algorithm is required before sending the input image to its corresponding processing pipeline. There has been a significant amount of research on the image classification topic and some methods have been proposed specifically to address this issue. Dong et al. introduced several features which are statistically significant to classify different types of images. He used a simple threshold method to the make final classification decision. However, this method relies on having a 300 dpi high resolution input image, which is computationally expensive. This method also lacked a basis in a statistical machine learning method. Lu et al.proposed a classification algorithm for digital copier based on Support Vector Machine. This method allows limited quick decision capability, which may speed up classification for mix images. However, it applies only offline training. Therefore it excludes AIO users from training the AIO device in the future to expand the training set and thus customize the product's behavior.In this paper, we present an algorithm which includes online SVM training that allows the algorithm to be improved and customized by the users as the input images accumulate. At the same time, it enables quick decision for mix images which is the most frequently copied image type and significantly speeds up classification for picture documents. We will also introduce a method to make online SVM training and quick decision compatible with each other.

9395-12, Session 3

An evaluation of the transferability of Munsell's colour notation methodology to modern inkjet printing technology

Melissa K. Olen, Adrian Geisow, Carinna E. Parraman, Univ. of the West of England (United Kingdom)

Munsell's work is widely regarded as the first well-founded attempt to formulate a perceptually uniform colour space. His original work was partly based on psychophysical principles – for example the employment of spinning discs to determine complementary colours – and partly the result of artistic training [1]. This research examines Munsell's experimental design, specifically his method of determining the arrangement of colours according to hue, value and chroma. It seeks to assess the transferability of Munsell's colour methodology to modern digital print technologies and colourants. Current inkjet technologies exhibit several notable differences in their approaches to colour mixing when compared to the processes used in defining Munsell's colour notation. Whereas the available pigments limited Munsell at the time, modern inkjet printing provides a wider range of producible colour, particularly when incorporating direct channel and multiple pass printing methods [2]. Furthermore, for his notation, Munsell used traditional paint pigments and divided the hue dimension into five primary colours rather subjectively based on a preference for a decimal system [3]. In contemporary digital printing, inkjet ink sets contain a greater number of primary ink colourants unable to be equally divided around the hue circle. These inks are also significantly higher in chroma, resulting in the Munsell notation representing only a portion of colours producible by inkjet technologies.

While extensive research and development has gone into establishing methods for measuring and modelling the modern colour gamut, we aim to examine the transferability of the Munsell system to modern inkjet colorants and printing technology following a similar approach to his original method. Here we seek to reintegrate the psychophysical and artistic principles used in Munsell's early colour studies, while breaking away from potentially compromising algorithmic calculations, as a means of exploring anomalies in both Munsell's and modern colour representation. Following Munsell's design and implementation, our experimental design replicates the use of Clerk-Maxwell's spinning disks [4] in order to examine the effects of colour mixing with the six principle colours of our inkjet printing system. The resulting data will be used in determining not only chroma distances from the neutral axis, but also in analysing hue distribution and placement. Assessment of the results is conducted qualitatively though physical observation, as well as quantitatively by comparison against spectral colour analysis in CIE L*a*b*. This work revisits Munsell's project in light of know issues, and formulates questions about how we can reintegrate Munsell's psychophysical approach and artistic requirements for colour description and mixing into modern colour science, understanding, and potential application. The ultimate objective of this research is to develop a model defined by the pigment colours of the inkjet inks, which allows it to be organic in shape and without compromise from computational processes.

[1] Munsell, A.H., "A Pigment Color System and Notation," AM J PSYCHOL, 23(2), 236-244 (1912)

[2] Olen, M. and Parraman, C., "Exploration of alternative print methodology for colour printing through the multi?layering of ink," Proc. AIC 12(2), 573-576 (2013)

[3] Kuehni, R, "The Early development of the Munsell system," COLOUR RES APPL, 27(1), 20-27 (2002)

[4] Rood, O.N., [Modern Chromatics], D. Appleton & Co., New York, 167-170 (1879)

9395-13, Session 3

Effect of ink spreading and ink amount on the accuracy of the Yule-Nielsen modified spectral Neugebauer model

Radovan Slavuj, Ludovic G. Coppel, Jon Yngve Hardeberg, Gjøvik Univ. College (Norway)

To control a printer so that the mixture of inks results in a determined color in a specific visual environment requires a spectral reflectance model that estimates reflectance spectra from nominal ink surface coverage. The Yule-Nielsen modified spectral Neugebauer (YNSN) is widely used and one the most accurate model for coloured halftones especially when the coverage space is divided into different cells leading to the so-called cellular YNSN. This paper investigates the dependence of the model accuracy on the ink amount that is controlled in different manner in today's printer. For an inkjet printer, we show that the performance of the YNSN model strongly depends on the maximum ink amount and that ink limitation that eventually reduces the printable gamut are required to get an acceptable performance of the model.

In a cellular implementation, this limitation mainly occurs for high coverage prints, which impacts on the optimal cell design. For inkjet prints, apparent coverages derived from both the Murray-Davis and Yule-Nielsen models show very large ink spreading resulting in that the printed dots are larger than their nominal surface coverage (dot gain). Because inkjet printing is a non-impact printing process, this leads to the hypothesis that ink dots have a smaller thickness than fulltone ink film since the ink volume on each pixel

is constant. Measured spectral reflectance curves show similar trends to those of ink film with increasing thickness, which supports the hypothesis.

The lesser accuracy of the YNSN model when using maximum ink thickness can therefore be explained with the fact that the patches with lower coverage have a mean ink thickness very different from that of the full ink fulltone patch. This observation implies that the Yule-Nielsen n-factor is accounting for varying ink thickness that results in non-linear relationships between the spectral reflectance at different ink coverage. This can partially explains why the n-factor is often determined as being different for different inks on the same substrate. Since the effect of significant ink spreading on the dot spectral reflectance will be related to the relative and not the absolute dot gain since the dot thickness is proportional to the dot area for constant volume. This means that the effect will be the largest when the dots are small and the dot gain large.

Analysis of the results suggests that the performance of the YNSN model can be improved by incorporating the change of ink thickness related to the mechanical dot gain in the optimization of the apparent coverage and n. This can be easily implemented by modifying the spectral reflectance of the ink according to the ink thickness determined as a fraction of full tone ink thickness that is directly related to the relative dot gain. An extended Yule-Nielsen model is therefore proposed in which the ink reflectance is related to the estimated apparent coverage using the Kubelka-Munk model after estimation of the scattering and absorption coefficients of individual inks. Preliminary results show a significant increase of the prediction accuracy and a significant lowering of the n-factor. This impacts on the separation of the mechanical and optical dot gain and open for simplified spectral reflectance model calibrations.

9395-14, Session 3

The precise prediction model of spectral reflectance for color halftone images

Dongwen Tian, Shanghai Research Institute of Publishing & Media (China); Fengwen Tian, Shanghai Maritime Univ. (China)

In order to predict the spectral reflectance of color halftone images, we considered the scattering of light within paper and the ink penetration in the substrate and proposed the color spectral reflectance precise prediction model for halftone images. The paper based on the assumption that the colorant is non-scattering and the assumption that the paper is strong scattering substrate. By the multiple internal reflection between the paper substrate and the print-air interface of light, and the light along oblique path of the Williams-Clapper model, we propose this model for taking into account ink spreading, a phenomenon that occurs when printing an ink halftone in superposition with one or several solid inks. The ink-spreading model includes nominal-to-effective dot area coverage functions for each of the different ink overprint conditions by the least square curve fitting method and the network structure of multiple reflection. It turned out that the modeled and the measured colors agree very well, confirming the validity of the used model. The new model provides a theoretical foundation for color prediction analysis of recto-verso halftone images and the development of prints quality detection system.

9395-15, Session 3

Ink thickness control based on spectral reflectance model

Dongwen Tian, Shanghai Research Institute of Publishing & Media (China)

In the color printing process, the thickness and uniformity of ink have a great affect on the color reproduction. The ink thickness uniformity is an important parameters of measuring the quality of printing. The paper based on the expansion of the ink, optical properties of paper, the internal lateral spread of light in paper and the spectral reflectance prediction model, we introduce two factor parameters which are the initial thickness of the inks and the factor of ink thickness variation. A spectral model for deducing ink thickness variations of printing on the paper substrate is developed by the least square method and the spectrum reflectance of prints which measures the ink thickness variations. The correctness of the conclusions are verified by experiment.

9395-16, Session 3

Yule-Nielsen based multi-angle reflectance prediction of metallic halftones

Vahid Babaei, Roger D. Hersch, Ecole Polytechnique Fédérale de Lausanne (Switzerland)

Spectral prediction models are widely used for characterizing classical, almost transparent ink halftones printed on a diffuse substrate. There are however applications where the inks or the substrate are not diffuse. Metallic-ink prints, for example, reflect a significant portion of the light in specular direction. We are interested in predicting reflectances and colors that appear at different illumination and observation geometries.

In the present paper, we study the Yule-Nielsen spectral Neugebauer (YNSN) model for predicting the reflectances of metallic halftones calibrated separately at the desired illumination and observation angles. At each geometry, we observe the Yule-Nielsen n-value in order to better understand the interaction of light and metallic halftones.

We use the OKI DP-7000 printer (also known as ALPS MD) to print metallic inks. We consider only the 4 available metallic inks on this printer: metallic cyan (c), metallic magenta (m), gold (y) and silver (s). In order to generate metallic halftones, we apply discrete line juxtaposed halftoning [1]. Due to limited printer dot positioning precision, we consider in all our experiments an effective resolution of 100 dpi.

In order to examine the predication accuracy of the YNSN model for metallic halftones, we consider a test set formed by variations of the nominal surface coverages of the four metallic inks in 20% intervals while constraining them such that the sum of the area coverages remains one, i.e. c + m + y + s = 100%, where c, m, y and s denote the respective area coverages of the cyan, magenta, yellow and silver inks.

The overall prediction accuracy of the YNSN model at different illumination and capturing orientations is within an acceptable range. In order to analyze how the optimal n-values behave as a function of the measuring geometry, we examine one halftone sample with 60% metallic magenta and 40% silver area coverage. The prediction for this particular halftone shows that there are different optimal n-values for different measuring geometries. For example, negative n-values are optimal for the 45as-15 (45°:60°) geometry. In addition, 5 out of the 11 measuring geometries have an optimal n-value in the interval (0, 1) which never occurs in traditional color reproduction.

In order to better understand the effect of different n-values on the prediction of halftone reflectances, we use the graphical representation of the Yule-Nielsen function [2]. This representation enables displaying the attenuation of a halftone of a constant dot size as a function of the corresponding attenuation of the fulltone. Different n-values correspond to different attenuation curves.

For a halftone composed of two colorants, we denote Ra the spectral reflectance of the non-silver patch with area coverage of a (0\< a\<1) and R1 the spectral reflectance of same patch with full area coverage. The silver ink halftone has area coverage 1-a. Its solid ink reflectance is R0. Note that Ra, R1 and R0 are all functions of wavelength. The Yule-Nielsen equation for such a halftone can be expressed as

Ra/R0=[(1-a) + a*(R1/R0)^(1/n)]^n (1)

The halftone attenuation relative to the silver ink Ra/R0 is a function of the fulltone attenuation Ra/R0 = f(R1/R0) and it can be plotted according to the spectral measurements of R0, Ra and R1. These reflectances provide as many Ra/R0 and R1/R0 attenuations as values contained in the measured spectra.

Within the attenuation graph, the straight line with n = 1 represents the “Neugebauer function”, a linear relationship between no attenuation and

attenuation by the fulltone ink according to the area coverage of that ink. Varying the n-value has the effect of making the YN function nonlinear. The traditional interval of n-values [1, 100 (infinity)) covers a small portion of the function's potential range. In this interval, the reflectance of halftones is lower than the one predicted by the Neugebauer function. Extending this interval to negative values helps spanning a greater portion of that range [3]. The most interesting case is the interval (0, 1) where the Yule-Nielsen function yields an inverse effect, i.e. the halftone exhibits a higher reflectance than the one predicted by the Neugebauer function. This is the case for example when the metallic particles of the inks are coarse and reflect specularly also at non-specular directions.

For the considered metallic halftone, the Yule-Nielsen graph follows a straight line at all geometries. This is in contrast with traditional print on paper where due to optical dot gain the halftone attenuation follows a curve. This might be an indication of a very moderate or no optical dot gain.

We design an experiment to demonstrate the effect of halftoning on the texture of a metallic sample and on its multi-angle reflecting behavior. We print a fulltone (100%) sample in a single pass and the same sample by sequentially printing halftone lines side by side. The microscopic picture of the "multi-pass" sample shows the misregistration between the successive printed lines. These misregistrations are at the origin of small unprinted areas and of partial superpositions of the halftone lines. Hence, the surface of the multi-pass solid sample becomes uneven and prone to shadowing.

We also compare their multi-angle reflectances. Without the effects of multi-pass printing, these two samples would have the same reflection spectra at all illumination and observation angles. However, they are different and the difference depends on the illumination and capturing geometry.

Depending on the measuring geometry, the halftone exhibits different contributions of shadowing and misregistration. Shadowing results in darker measurements while misregistration creates unprinted paper areas that lighten the reflectance. The optimal n-value seems to reflect these two inverse effects. Since the effect of shadowing and misregistration is different for each geometry, the n-value also differs.

References

[1] V. Babaei and R. D. Hersch, "Juxtaposed color halftoning relying on discrete lines", IEEE Trans. Image Process., Vol. 22, No. 2, pp. 679–686 (2013).

[2] M. Hébert, "Yule-Nielsen effect in halftone prints: graphical analysis method and improvement of the Yule-Nielsen transform," Proc. SPIE 9015, Color Imaging XIX: Displaying, Processing, Hardcopy, and Applications, 90150R-1–10, (2014).

[3] A. Lewandowski, M. Ludl, G. Byrne and G. Dorffner, "Applying the Yule-Nielsen equation with negative n," JOSA A, vol. 23, no. 8, pp. 1827–1834 (2006).

9395-17, Session 3

Multichannel DBS halftoning for Improved texture quality

Radovan Slavuj, Marius Pedersen, Gjøvik Univ. College (Norway)

There is still limited number of halftoning algorithms proposed to work in a multichannel printing environment [1, 2]. Any halftoning should provide dot placement in visually unnoticeable manner and should maintain local average color accuracy. Direct-Binary-Search (DBS) [3] is often considered as one of the best halftoning algorithms. Channel independent DBS halftoning applied to CMYK or multichannel systems would yield noise texture which reduces image quality and possess a problem to color accuracy due to the unwanted overlaps[4,5]. One can tackle the problem of color accuracy by weighing different overlaps [6] and the problem of reduced image quality by a channel dependent DBS. [4,5,7]. Tradeoff between image quality and color accuracy needs to be made when implementing any hafltoning algorithm in the printing workflow. In order to increase the gamut and save ink, [8] a model is developed that has Neugebauer Primaries (NPs) as the output instead of individual dots and uses modified version of Error Diffusion (ED) to place NPs. This way, it is possible to have color accuracy per area and satisfying image quality.

We aim to create multichannel DBS that takes the printing method's characteristics and benefits, and uses them as constraints for a channel dependent DBS halftone algorithm. One of the specifics of the multichannel printing is that it provides greater variability of choices (e.g. metamers) to select from. This means that one color can be printed using different combination of inks. The selection of these colors for the halftoning is set here based on the visibility of its texture. For the initial stage, we have used approach by Lee and Allebach [4] of monochrome DBS for overall dot position, and coloring based on an approach similar to that of hierarchical color DBS proposed by He [5]. Multichannel printer (CMYKRGB configuration) is already equipped with what would be secondary colors in conventional CMYK configuration. Regarding dot positioning, this advantage can result in textures that are less visible (e.g. selecting blue color instead of combination of the cyan and magenta). Subsequently, this is extendable to any of the NPs (e.g. instead of using magenta yellow and cyan, red and cyan could be used) that would reduce density or visibility of the texture. In comparison, our method is similar to introduction of diluted solutions (light cyan, magenta, and black) in printing that provide greater gradation in high and mid-tones.

The number of inks used to create an NP is limited to three channels, although it is possible to extend to whatever number is needed for a particular application. Also we try to minimize dot on dot (or NP over NP) printing and to have both individual and combined texture uniformity.

Preliminary results show that by comparing with channel independent DBS halftoning applied to multichannel printing, we are able to reduce individual and combined texture visibility, to reduce noise and unwanted overlaps and to provide better gradation, especially in dark regions.

As we could expect that such approach would yield a small color difference on the expense of image quality, our future work would be to integrate this algorithm with printer models that can make this decision based on a given threshold for color accuracy and selected image quality attributes.

References

[1] Zitinski,P.J., Nystrom, D., Gooran, S. (2012), Multi-channel printing by Orthogonal and Non-Orthogonal Halftoning, in Proceedings 12th International AIC Congress, pp. 597-604

[2] Gerhardt, J. and Hardeberg, J. Y. (2006), Spectral colour reproduction by vector error diffusion. In Proceedings CGIV 2006, pages 469–473, 2006

[3] Lieberman,D.J. and Allebach,J.P. (2002), "A dual interpretation for direct binary search and its implications for tone reproduction and texture quality," IEEE Trans. Image Process., vol. 9, no. 11, pp. 1950–1963, Nov. 2000.

[4] Lee,J. and Allebach,J.P.(2002), "Colorant-based direct binary search halftoning," J. Electron. Imaging, vol. 11, pp. 517–527, Oct. 2002.

[5] He,Z. (2010), Hierarchical Colorant-Based Direct Binary Search Halftoning, IEEE Transactions on Image Processing, Vol. 19, No. 7, July 2010, pp. 1824-1836

[6] Emmel, P. and Hersch, R. D. (2002), Modeling ink spreading for color prediction, J. Imaging Sci. Technol. 46(2), (2002).

[7] Maria V. Ortiz Segovia, Nicolas Bonnier, and Jan P. Allebach (2012), Ink Saving Strategy Based on Document Content Characterization and Halftone Textures. Proc. SPIE 8292, Color Imaging XVII: Displaying, Processing, Hardcopy, and Applications, Jan. 2012

[8] Morovi?, J., Morovi?, P. and Arnabat, J. (2012), HANS: Controlling Ink-Jet Print Attributes Via Neugebauer Primary Area Coverages, IEEE Transactions On Image Processing, Vol. 21, No. 2, FEBRUARY 2012

9395-18, Session 3

Color dithering methods for LEGO-like 3D printing

Pei-Li Sun, Yuping Sie, National Taiwan Univ. of Science and Technology (Taiwan)

The development of 3D printing technology has accelerated over the past

few years. More and more commercial products have brought to the market at lower price than ever. However, the popularity of 3D printer is not growing that fast as the printing speed are very slow and most of them cannot print different colors on one product. The 3D printing market now has two extremes: hi-end market needs high resolution and various materials for industrial applications, whereas low-end market needs rapid prototyping for product design. Color and printing speed are more important than spatial resolution in the low-end market. However, no commercial 3D printer to date can fulfill the needs.

To increase the printing speed with multi-colors, a LEGO-like 3D printing can be considered. In the type of 3D printing, each brick (pixel or photopolymer ink drop) is big enough to occupy a physical space. The speed is increased by reducing spatial resolution of the printing, and full color printing is achievable by using multiply print heads but the color cannot be fully mixed by conventional halftoning or continuous-tone dye sublimation techniques. The LEGO-like 3D printing might adapt to either fused deposition modeling (FDM) or photopolymer inkjet technologies. The aim of this study is to develop color dithering methods for the type of printing.

Three cases were considered in this study:

(1) Opaque color brick building: The dithering is done by color error diffusion on the surface of the product. It is close to a color mosaic solution. Our simulation results show that using RGBWCMYK 8-color printing will maximize color gamut and spatial smoothness. However, RGBWK 5-color printing also provides acceptable result. If only 4-primary is applicable, RGBK is better than CMYK.

(2) Transparent color brick building: To optimize its color appearance, top 4 layers behind the surface should be used in color mixing. In this case, CMYW 4-color printing will be the minimize requirement where the white bricks must be opaque.

(3) Translucent color brick building: In reality, dark colors cannot be reproduced by stacking few color bricks. Translucent color bricks therefore are preferable for 3D color printing. However, its colors cannot be predicted accurately by simple mathematics. In CMYW translucent color brick building, the color gamut will be very small if we use certain color order to print the top 4 layers. The color gamut can be expanded by considering all possible combination and ignoring the order. However, it will dramatically increase the complexity of its color characterization. For instant, a MCY stack (i.e. yellow appears on the surface) is more yellowish and brighter than that of YCM. And a vivid yellow can be produced by WYYY stack. Multi-dimensional LUTs are used to predict colors with different stacking orders. Simulation results show the proposed multi-layer dithering method can really improve the image quality of 3D printing.

9395-19, Session 3

Design of irregular screen sets that generate maximally smooth halftone patterns

Altyngul Jumabayeva, Yi-Ting Chen, Purdue Univ. (United States); Tal Frank, Indigo Ltd. (Israel); Robert A. Ulichney, Hewlett-Packard Co. (United States); Jan P. Allebach, Purdue Univ. (United States)

With the emergence of high-end digital printing technologies, it is of interest to analyze the nature and causes of image graininess in order to understand the factors that prevent high-end digital presses from achieving the same print quality as commercial offset presses. In this paper, we report on a study to understand the relationship between image graininess and halftone technology. With high-end digital printing technology, irregular screens can be considered since they can achieve a better approximation to the screen sets used for commercial offset presses. This is due to the fact that the elements of the periodicity matrix of an irregular screen are rational numbers, rather than integers, which would be the case for a regular screen. To understand how image graininess relates to the halftoning technology, we recently performed a Fourier-based analysis of regular and irregular periodic, clustered-dot halftone textures [1]. From the analytical results, we showed that irregular halftone textures generate new frequency components near the spectrum origin; and that these frequency components are low enough to be visible to the human viewer, and to be perceived as a lack of smoothness. In this paper, given a set of target irrational screen periodicity matrices, we describe a process, based on this Fourier analysis, for finding the best realizable screen set. We demonstrate the efficacy of our method with a number of experimental results.

For a given irrational target periodicity matrix screen set, our method searches over the space of realizable periodicity matrices i.e., those that have elements that are rational numbers. Within this space, those screen sets whose tile vectors are sufficiently close to those of the target set, in terms of the Euclidean distance between them, are chosen for further consideration. For each candidate screen set, we next consider all sums and differences between integer multiples of the tile vectors of the screen and integer multiples of the vectors corresponding to the printer lattice. These frequency domain alias terms that fall within the passband of the human viewer, and which have a sufficiently large amplitude are responsible for the perceived lack of smoothness in the printed halftone patterns.

From our Fourier analysis, we can identify two sources of attenuation of the frequency domain alias terms. The first is the Fourier spectrum of the ideal halftone dot shape that would be rendered by an analog device. The second is the Fourier spectrum corresponding to the shape of the printer-addressable pixels. Due to the relatively small size of these pixels, this effect is much weaker than the attenuation due to the spectrum of the ideal halftone dot shape. Taking into account both the locations and amplitudes of each Fourier component, we compute an integrated measure of the root-mean-squared (RMS) fluctuation, with a visual weighting function to predict the perceived lack of smoothness in the printed halftone pattern. The realizable screen set that minimizes the RMS fluctuation can be chosen as the best solution. However, in practice, aspects of the non-ideal press behavior that are not included in our model can have a significant impact on the quality of the printed halftone patterns. Thus, it is best to chose a small set of candidate screen sets that all produce low RMS fluctuation, and to print halftone patterns with these screen sets to determine which set will ultimately perform best.

[1] Y-T. Chen, T. Kashti, T. Frank, R. Ulichney, and J. Allebach, "Fourier-based Analysis of Regular and Irregular Periodic, Clustered-dot Halftone Textures," International Congress of Imaging Science (ICIS), Tel Aviv, Israel, 12-14 May 2014.

9395-45, Session PTues

Representation of chromatic distribution for lighting system: a case study

Maurizio Rossi, Fulvio Musante, Politecnico di Milano (Italy)

For the luminaire manufacturer, the measurement of the lighting intensity distribution emitted by lighting fixture (LID) is based on photometry. So light is measured as an achromatic value of intensity and there is no the possibility to discriminate the measurement of white vs. coloured light. At the Light Laboratory of Politecnico di Milano a new instrument for the measurement of spectral radiant intensities distribution for lighting system has been built: the gonio-spectra-radiometer. This new measuring tool is based on a traditional mirror gonio-photometer with a CCD spectra-radiometer controlled by a PC.

Beside the traditional representation of photometric distribution we have introduced a new representation where, in addition to the information about the distribution of luminous intensity in space, new details about the chromaticity characteristic of the light sources have been implemented.

Some of the results of this research have been applied in developing the line of lighting system "My White Light" (the research project "Light, Environment and Humans" funded in the Italian Lombardy region Metadistretti Design Research Program involving Politecnico di Milano, Artemide, Danese, and some other SME of the Lighting Design district), giving scientific notions and applicative in order to support the assumption that colored linear fluorescent lamps, can be used for the realization of interior luminaries that, other than just have low power consumption and long life, may positively affect the mood of people.

9395-20, Session 4

Introducing iccMAX: new frontiers in color management

Maxim W. Derhak, Onyx Graphics (United States); Phil Green, Gjøvik Univ. College (Norway); Tom Lianza, Photo Research, Inc. (United States)

The ICC profile and associated colour management architecture has been a part of the colour imaging landscape since it was first introduced in 1995. Owing to its interoperable format, unambiguous processing and open standardisation, it is now universally used to connect colour devices.

The ICC.1 architecture and specification have evolved since to address new requirements, including from the graphic arts, photography and motion picture industries, and the current version ISO 15076-1:2010 represents an optimal solution for many existing workflows. However, it has become clear that the underlying concept of connecting through a fixed, colorimetric connection space is not capable of delivering solutions to more advanced requirements that have emerged, and will continue to emerge in the future. It has very limited capabilities for the processing of spectral data.

ICC is responding to these new requirements by extending the ICC.1 architecture. The new iccMAX functionality is a major step change that goes far beyond the previous incremental approach of adding and deleting tags from the specification. Some of the new features are: a spectral Profile Connection Space to connect spectral data; a spectral viewing conditions tag which permits arbitrary illuminants and observers to be defined; a Material Connection Space that supports identification and visualisation of material amounts in addition to colour; support for bidirectional reflectance distribution functions and 3D rendering; support for processing of bi-spectral fluorescence data; an extended named colour profile that incorporates spectral, bi-spectral and BRDF processing; and a color encoding profile that does not carry a transform but a pointer to a reference encoding. iccMAX also incorporates a flexible processing element that gives profile creators the ability to incorporate arbitrary transforms yet continue to benefit from unambiguous CMM processing.

An iccMAX CMM will be backward-compatible and will recognise and correctly process v2 and v4 profiles. However, iccMAX profiles are not expected to be compatible with v4 CMMs. ICC is providing a reference CMM implementation in the form of open source code, which will aid developers and researchers who wish to adopt iccMAX. An iccMAX application is not required to support all the new features but just the sub-set appropriate to a particular workflow domain.

Some iccMAX applications already exist, and researchers are encouraged to download the specification and reference implementation and incorporate in their workflows.

9395-22, Session 4

Baseline gamut mapping method for the perceptual reference medium gamut

Phil Green, Gjøvik Univ. College (Norway)

Gamuts of source and destination colour encodings are invariably different, and as a result some form of gamut mapping is required when transforming between source and destination encodings in a colour managed workflow. In the original ICC.1 colour management architecture for, transforms between individual media and the ICC Profile Connection Space (PCS) were constructed independently of each other. As a result, the PCS connects encodings but not gamuts, and it was necessary for the destination profile to make an assumption about the gamut of the source encoding.

This problem was addressed in ICC v4 by adopting a common reference gamut. It is recommended that the Perceptual Intent transform in a source profile renders to this reference gamut and the Perceptual Intent transform in a destination profile renders from this gamut. Perceptual Reference Medium Gamut (PRMG) was derived from work on surface colour gamuts standardised in ISO 12640-3.

ICC has provided profiles which render to the PRMG from common source encodings such as sRGB. Preferred renderings are incorporated into this source-to-PCS transform, so that the image in the Perceptual PCS represents the desired reproduction on a large-gamut output device. However, as yet there is no agreed baseline method for rendering from the PCS to different output systems. Such a method is anticipated to further improve the consistency between different transforms when using v4 profiles.

The method described in this paper uses a face/vertex gamut boundary descriptor (GBD), and the computation of the face/vertex list from an ICC input or output profile or other colour transform is described. The face/vertex GBD method is not limited to the convex hull of the gamut surface, but instead describes the triangulation of a set of points which lie on the gamut surface, including any concavities.

By defining both source and destination gamuts as a set of cognate triangles, the mapping from an intersection with a triangle to a cognate triangle in another gamut can be computed. The Barycentric coordinates of the intersection points relative to the planes of each triangle are used to determine whether a point lies inside a given triangle, and to compute the intersection with the cognate triangle.

Two candidate algorithms are described, based on the CUSP2CUSP and GCG algorithms previously described by the author.

Preliminary results of the two methods will be shown, but the goal of the current work is to encourage other researchers to contribute candidate algorithms and participate in the evaluation.

9395-23, Session 4

False-colour palette generation using a reference colour gamut

Phil Green, Gjøvik Univ. College (Norway)

False-coloured images are widely used to convey information. Arbitrary colours are assigned to regions of an image to aid recognition and discrimination of features in the image. Examples include different materials, object classifications, geographic features, morphology, texture, and tissue types.

Except in a small number of cases such as maps, false-colour colour coding schemes have no semantic content and can thus be freely assigned according to the preferences of the palette or image creator.

The requirements of a false-colour palette are three-fold: first, the colours must be visually distinct, in order to maximise recognition and discrimination of features. Second, the palette should be extensible, so that distinctiveness is preserved regardless of the number of colours in the palette; and finally, the palette should be capable of being reproduced consistently on different media. For many applications a further requirement is that the palette should maximise difference between colours for users with colour vision deficiencies as well as those with normal colour vision.

Existing approaches to false-colour palette selection are usually based on selection in colorant space (e.g. display RGB) or in a uniform colour space (e.g. CIELAB), in the latter case maximising the colour difference between palette colours. Both approaches suffer from the gamut limitation of reproduction systems, and this can cause problems when a false-colour image is required to be consistently reproduced on different media.

A method of defining a false-colour palette using a reference colour gamut is described in this paper.

In a false-colour palette, hue will be the main differentiating factor between colours. At a given hue angle, chroma should be as large as possible in order to maximise the perceptibility of the hue. In any real colour gamut, the lightness of the colour with maximum chroma at a given hue will be fixed, and hence there is limited freedom in the use of different lightnesses.

The palette is selected from colours on the surface of a reference gamut which approximates a gamut of real surface colours. Several candidate gamuts exist for this purpose. The ISO Reference Colour Gamut (ISO 12640-3) was found in Li, Luo, Pointer and Green (2014) to be slightly larger than the most recent ly accumulated surface colour reflectance data, and

is therefore a good basis for false colour palette selection. The use of this gamut ensures that selected colours can readily be mapped into the gamut of real colour reproduction devices using conventional colour management methods.

An extensible palette is generated by first selecting the gamut vertices at 0, 90, 180 and 270 degrees of CIELAB hue angle. Next, colours are selected at 12 intermediate hue angles. Finally a further 16 hue angles are chosen which lie between the ones already selected, and are at different lightness levels than the gamut vertices to avoid confusion with other colours in the palette. This approach generates a maximum of 72 colours which are readily distinguishable from each other.

The palette colours are defined in media-relative colorimetry in CIELAB colour space, and are converted to device coordinates for reproduction by simply applying the ICC profile for the reproduction device.

9395-24, Session 4

Color correction using 3D multi-view geometry

Dong-Won Shin, Yo-Sung Ho, Gwangju Institute of Science and Technology (Korea, Republic of)

Recently, interest on three-dimensional (3D) image has been raised from day to day. It is used in various fields such as education, advertising, medical, and gaming. 3D images make you feel the sense of the depth which cannot be found on 2D images. This sort of sense makes you feel like the object exists right in front of you.

We use the multiview system at several categories in overall 3D image processing. For example, we use it at the acquisition of the depth image; 3D image includes color and depth images. The method of the depth acquisition consists of active and passive methods. Especially in the passive method, the multiview color images are used for stereo matching; stereo matching is to recover quantitative depth information from a set of input images, based on the visual disparity between corresponding points. However, there is color mismatching problem in the multiview system. It means color inconsistency among multiviews. It's caused by an inconsistent illuminance or inevitable errors and noises resulting from imaging devices. The color relationship among views affects the performance of stereo matching. So, if the target view's color differs from that of the source view, it is difficult to get the accurate disparity. For another multiview example, we can consider the 3D reconstruction. If the color difference is reflected in the reconstructed object, we are hard to get a smooth object in the color sense.

In order to resolve this kind of color mismatching problem, we use the 3D multiview geometry. As an offline process, we employee the camera calibration to understand the geometry between source and target view. We can get camera projection matrix as a result of camera calibration method. Next, we need color and depth images from both source and target views to begin an online process. The correspondences matching between two views is employed by using the 3D geometry. Then, we acquire the mathematical relationship between correspondences by using polynomial regression; polynomial regression is a form of linear regression in which the relationship between the independent variable x and the dependent variable y is modelled as an n-th order polynomial. From polynomial regression method, we can get the translation matrix which can calculate adjusted color values. After translating all the pixels in the target view by using this matrix, we can get the final adjusted image.

As a result of proposed method, we acquired multiview 2D images which has similar color distribution and constructed the smooth 3D object in the color sense via the proposed color correction method to the 3D object reconstruction.

9395-25, Session 4

Real-time subsurface scattering volume rendering for reproduction of realistic skin color in 3D ultrasound volume

Yun-Tae Kim, Sungchan Park, Kyuhong Kim, Jooyoung Kang, Jung-Ho Kim, SAMSUNG Electronics Co., Ltd. (Korea, Republic of)

1. Background, Motivation, and Objective

The evolution of ultrasound imaging from 2D to 3D imaging has enhanced diagnostic accuracy. The objective of this study is to enable realistic ultrasound volume rendering with similar color to a real human skin in order to provide psychological stability and enjoyment to pregnant women. The conventional volume rendering method used the reflection of the light using the local illumination model such as Phong reflection model. However, because the human skin has the reflection and translucent components of the light when the light ray reaches in the skin, the more similar rendering method to skin color is required.

2. Methods

We propose a real-time and realistic subsurface scattering rendering method for translucent ultrasound volume. First, the dual-depth map method is used to calculate the distance between intersection points of surface points and the light ray in the user and light viewpoint. A plane perpendicular to the direction of light is defined in the light viewpoint and this plane has the same resolution as the final image. The light ray is casted simultaneously from the light source point to the pixels on the plane. The intersection points between the light ray casted from the light source and surface points of the object are calculated. Also, using the same method as the intersection points calculated in the light viewpoint, the eye ray is casted from the eye source point to the pixels on the plane. The intersection points between the eye ray casted from the eye source and surface points of the object are calculated. Finally, the distance between the intersection points of the light ray and the eye ray is calculated.

Second, we propose the subsurface scattering method using the Sum-of-Gaussian (SOG) function in order to represent the realistic human skin color. This SOG function models the degree of light scattering physically according to the actual thickness of the each layer of the skin. The attenuation of the blue channel according to the distance between the intersection points of the surface points is the largest, the green channel attenuation is medium, and the red channel attenuation is designed to be gradually decreased according to the distance between the intersection points. The distance between intersection points calculated by the dual-depth map method is applied to the attenuation parameter of the SOG function. The translucent effect of the ultrasound volume is represented differently according to the penetration depth of the light ray.

3. Results and Discussion

In the experiments, we used an NVIDIA GeForce GTX Titan BLACK graphics card and Intel Core i7-4960X CPU. CUDA parallel coding is used to implement the volume rendering based on ray casting. The resolution of the volume data is 200x200x200 voxels. The speed of the volume rendering is the 60.8 fps (frame per second). The experimental results show that the proposed subsurface scattering rendering method achieves better and more realistic skin color reproduction than the conventional volume rendering methods using the light reflection model.

9395-26, Session 4

Vague color image enhancement on fractional differential and improved retinex

Xin Zhang, Weixing Wang, Xiaojun Huang, Lingxiao Huang, Zhiwei Wang, Chang'an Univ. (China)

Using aerial color images or remote sensing color images to obtain the earth surface information is an important way for gathering geographic

information. High-resolution color images provide more accurate information, and it is convenience for city planning, mapping, military inspecting, changing detection and GIS update. High-resolution color aerial images include innumerable data information, and how to quickly and accurately get special geo-information is becoming more significant, hence, the road extraction from high-resolution color aerial images is one of the hot research topics in the road recognition research area.

In order to improve visibility in the poor quality weather for aerial color images, a changing scale Retinex algorithm based on fractional differential and depth map is proposed. After a new fractional differential operation, it requires the image dark channel prior treatment to obtain the estimated depth map. Then according to the depth map, Retinex scales are calculated in each part of the image. Finally the single scale Retinex transform is performed to obtain the enhanced image. Experimental results show that the studied algorithm can effectively improve the visibility of an aerial color image without halo phenomena and color variation which happens if using a similar ordinary algorithm. Compared with the He's algorithm and others, the new algorithm has the faster speed and better image enhancement effect for the images that have greatly different scene depths.

The fractional derivative is different from the integral derivative, and the image application in the signal processing based on the fractional derivative is a very new research topic. In particular, in the field of the image processing by using fractional differential, the related research papers were published only over the past few years. There is the obvious auto-correlation among the pixel grey values in a pixel neighborhood, and the auto-correlation presents itself as a complex textural feature. It is very effective to enhance textures by the fractional derivative. Then, it introduces the fractional differential into the road traffic video image enhancement in a bad weather environment, which will be a meritorious try.

Whatever, in a digital image, there is a large relevance among the pixel grey values in a pixel neighborhood. The pixels are auto-correlated, their fractal geometric information show up the properties of complex textural features, and the fractional differential is one of Fractal Geometry Mathematical Foundations. It is natural for one to think about what effect is gained when applying the fractional differential to detect textural features in an image. In this paper, a new fractional differential algorithm, which can improve detailed texture information remarkably, is studied.

The Retinex algorithm was introduced in 1971 by Edwin H. Land, who formulated "Retinex theory" to explain it. Many researchers demonstrate the great dynamic range compression, increased sharpness and color, and accurate scene rendition which are produced by the Multiscale Retinex with Color Restoration, specifically in aerial images under smoke/haze conditions. Overall, the Retinex performs automatically and consistently superior to any of the other methods. While the other methods may work well on occasion cooperative images, it is easy to find images where they perform poorly. The Retinex performs well on cases where the other methods are clearly not appropriate. In order to use Retinex algorithms into the studied images, a single scale Retinex algorithm is extended into a multiple scale Retinex algorithm by referring previous research work.

9395-28, Session 5

Challenges in display color management (DCM) for mobile devices

Reza Safaee-Rad, Qualcomm Inc. (Canada); Jennifer L. Gille, Milivoje Aleksic, Qualcomm Inc. (United States)

Systematic and effective display color management for displays on mobile devices is becoming increasingly more challenging. A list of the main challenges we face includes:

1. Significant differences in display technologies: LCD, OLED (RGB, Pentile, WRGB), QD_LCD, ...

2. Significant display color response and tone response variability

3. Significant display color gamut variability: sub-sRGB, sRGB/HDTV, Adobe, OLED gamut (110% NTSC to 140% NTSC)

4. Significant content gamut variability: HDTV REC709, UHDTV2020, Adobe RGB, content with non-standard color gamut (mobile cameras), new Blue Ray content gamut, ...

5. Managing mixed contents with different gamuts during TV broadcast transitional period in the next few years—going from REC709 to REC2020

6. Significant variability in viewing conditions: indoor (dark and dim) and outdoor (bright, dim and dark)

7. Display huge power consumption issue

The first section of this paper will provide general descriptions of the above challenges, their characteristics and complexities. The second part will discuss various potential solutions to effectively perform display color management in the context of the above challenges. By display color management we mean: To manage video/image content-colors and display-color rendering characteristics to achieve the best perceptually pleasing images on a display in a systematic and consistent manner. This includes managing the following:

- Display tone response and gray tracking

- Display white point

- Display color correction—including RGB color crosstalk

- Display gamut mapping

- Content gamut mapping—single and mixed content

9395-29, Session 5

White balance for mobile device displays: navigating various image-quality demands

Jennifer L. Gille, Qualcomm Inc. (United States); Reza Safaee-Rad, Qualcomm Inc. (Canada); Milivoje Aleksic, Qualcomm Inc. (United States)

Mobile device displays are subject to multiple standards and non-standards (Rec. 709, Adobe RGB, sub-sRGB), multiple formats (888, 666), and multiple display technologies (LCDs, OLEDs). How to achieve a desired white point without sacrificing other image-quality factors, such as brightness, can be difficult to conceptualize in color space.

For an RGB display with given primaries, changing the white point of the display means re-balancing the primaries, which can only be done by reducing channel outputs. However, reducing channel output means reducing white luminance (with only a few exceptions), which has an impact on image quality.

White point is often specified by Color Temperature (CT), referring to locations on the Black Body (BB) radiation curve. Generally, lower CTs are considered to be "warmer", and higher CTs "cooler".

Changing from a display's native CT to a target CT may incur an unacceptable luminance loss. However, if the definition of color temperature is extended to include Correlated Color Temperature (CCT), a range of solutions is generated that may include an acceptable alternative solution. Points defined by their CCT without being on the BB curve may have a relative coloration, for instance a "neutral" white that is neither yellowish nor bluish, but ever-so-slightly greenish.

Finding an optimum white point is a 3D color space problem, although specifying it in terms of CT or CCT seems to turn it into a 2D problem. Searching in even 2D space with multiple image-quality criteria (CT, CCT coloration, white luminance) can be very confusing.

A further complication arises if the white-point-correcting function is meant to be applied to the gamma-compressed RGB values rather than in RGB that is nominally linear with XYZ.

A target white point defined by a CCT value inherently represents a one-to-many mapping problem. Our challenge was to find a set of meaningful, functional, useful constraints that could lead to a closed-form 1D solution.

The following two constraint extremes were proposed to address the above problem in white point calibration and adjustment:

(a) No relative coloration in the target white point (i.e., on the BB curve), and

(b) Minimum luminance loss for the target white point.

Since these are the endpoints of a line segment in (x,y), we have a 1D space to search for a given display and target CCT, parameterized by the locations of the extremes.

We discuss these white point calibration issues in the display manufacturing context, where additional factors of time and ease of operation compel simple, robust solutions. We addressed

(a) Presenting the 1D solution space for a given native-to-target white point conversion in a simple form that allows exploration

(b) Showing the image-quality impacts of luminance loss and relative coloration numerically and in visualizations

And assessed

(c) The number of display measurements necessary in order to carry out the adjustments in either linear or gamma-compressed space

We present our solutions as embodied in the display white point correction and optimization portion of Qualcomm's Display Color Management system (QDCM), part of the Snapdragon Display Engine (SDE).

9395-30, Session 5

A comparative study of psychophysical judgment of color reproductions on mobile displays between Europeans and Asians

Kyungah Choi, Hyeon-Jeong Suk, KAIST (Korea, Republic of)

With the internationalization of markets and growing user demand in the emerging economies, a cultural-specific approach should be adopted when designing the color appearance on the display devices. The users' psychophysical judgment of display colors is influenced by several factors such as religion, natural environment, race, and nationality. A pan-cultural approach has often resulted in cultural faux pas, and herein it is argued that a cross-cultural perspective is imperative for display color reproduction. In this regard, this study aims to investigate differences in display color preferences among Europeans and Asians.

This study involved a total of 50 participants, consisting of 20 Europeans (9 French, 6 Swedish, 3 Norwegians, and 2 Germans) and 30 Asians (30 Koreans). Their ages ranged from early to late adulthood, with an average age of 22.81 years. For visual examination, stimuli were shown on a smartphone with a 4.8 inch display. A total of 18 nearly whites were produced, varying from 2,470 to 18,330 K in color temperature. The 11 ambient illuminants produced ranged from 2,530 to 19,760 K in color temperature, while their illumination level was approximately 500 lux in all cases. The display stimuli were also examined with the light turned off. In all, the subjects were asked to evaluate the optimal level of the 18 display stimuli under 12 illuminants using a five-point Likert scale.

For analysis, the total scores of the 18 stimuli were averaged for each of the illuminants. The display stimuli with the highest average score for each illuminant were considered as the most optimal display color temperatures. In both cultural groups, it was found that the illuminant color temperature and the optimal display color temperature are positively correlated, which is in good agreement with previous studies. However, when the light was turned off (0 lux), the optimal display color temperatures were 5,950 K in Europe and 7,500 K in Asia. It is interesting to note that the Europeans preferred a lower color temperature compared to the standard white point D65, while the Asians preferred a higher color temperature. Regression analysis was performed in each group in order to predict the optimal display color temperature (y) by taking the illuminant color temperature (x) as the independent variable. The derived formula is as follows: y=?+?*log(x), where ?=-8770.37 and ?=4279.29 are for Europe (R2=.95, p\<.05), while ?=-16076.35 and ?=6388.41 are for Asia (R2=.85, p\<.05). The regression curve fitted for Europe lay below the curve fitted for Asia. An independent samples t-test was performed between each of the constant values in order to verify if the errors between the two equations were significantly different (?: t(20)=2.01, p\<.05; ?: t(20)=2.23, p\<.05).

In conclusion, the study reveals that the display color temperature perceived to be ideal increases as the illuminant color temperature rises; however, Europeans preferred a lower color temperature compared to Asians under every illuminant condition. The findings of this study could be used as the theoretical basis from which manufacturers can take cultural-sensitive approach to enhancing their products' values in the global markets.

9395-31, Session 5

Perceived image quality assessment for color images on mobile displays

Hyesung Jang, Choon-Woo Kim, Inha Univ. (Korea, Republic of)

Unlike TVs, viewing environments for mobile dipslays are quite diverse. Legibility of text information and image details is one of key performance indices of mobile devices. It is affected by viewing conditions as well as characteristics of mobile displays. Intensity and color temparature of illumination are major attributes of viewing conditions. Also, distance and angle to mobile display influence legibility of information. Attributes of mobile displays affecting legibilty would be resoltion (often called as ppi), contrast, and brighness of displays etc. In this paper, a quantitative measure of legibility for mobile displays is proposed. Both of text and image information are considered for constructing a legibility measure. Experimental results indicate that values of the proposed measure are highly correlated with results of human visual experiements.

9395-32, Session 6

Illumination estimation based on estimation of dominant chromaticity in nonnegative matrix factorization with sparseness constraint

Ji-Heon Lee, Ji-hoon Yoo, Jung-Min Sung, Yeong-Ho Ha, Kyungpook National Univ. (Korea, Republic of)

In image capture a scene with nonuniform illumination has a significant effect on the image quality. Various color constancy algorithms are already exist to remove the chromaticity of illuminants in an image for improving image quality. Most of the existing color constancy algorithms assume that the illumination is constant throughout the scene and the images are modeled by combining an illuminant component and reflectance features. Recently, NMFsc(nonnegative matrix factorization with sparseness constraints) was introduced to extract illuminant and reflectance component in an image. In NMFsc, Sparseness constraint is used to separate illuminant from an image since illuminant varies slowly compared to reflectance component. The low sparseness constraint is used as an illumination basis matrix and the high sparseness constraint is used as a reflectance basis matrix for separating illuminant and reflectance component in an image. Since the sparseness constraint is determined in a global scene, NMFsc has an illuminant estimation error for images with large uniform chromaticity area. Therefore, in this paper, a NMFsc based illuminant estimation method that considers the dominant chromaticity is proposed. To reduce the effect of dominant chromaticity in an image, an image is firstly converted to chromaticity color space and divided into two regions by K-means algorithm. Then the image region with low standard deviation in chromaticity space is determined as dominant chromaticity in an image. Subsequently, illumination estimation of each region is performed by using non-negative matrix decomposition and sparse constraints. Finally, the overall illumination is estimated by combining each region with weighted sum. The performance of the proposed method is evaluated by using angular error for Ciurea 11,346 image data set. Experimental results illustrate that the proposed method reduces the angular error over previous methods.

9395-33, Session 6

Clarifying color category border according to color vision

Takumi Ichihara, Yasuyo G. Ichihara, Kogakuin Univ. (Japan)

Clarifying color category border according to color vision

We usually recognize color by two kinds of processes. In the first, the color is recognized continually and a small difference in color is recognized. In the second, the color is recognized discretely. This process recognizes a similar color of a certain range as being in the same color category. The small difference in color is ignored. Recognition by using the color category is important for communication using color. For example, a color category is used when you plot different data as red and blue in a graph and then say to someone "I want you to take a look at the red data".

It is known that a color vision defect confuses colors on the confusion locus of color. However, the color category of a color vision defect has not been thoroughly researched. If the color category of the color vision defect is clarified, it will become an important key for color universal design.

Previous research provides the following two conclusions about the color category of a color vision defect.

First, the color name is not usable in the color category of a color vision defect. If the color category of the color vision defect is decided from color name as for normal color vision, color vision defect will confuse color categories with each other.

Second, the border of the color category of a color vision defect becomes approximately parallel to the color confusion locus. But this conclusion is based on an experiment in which only a color stimulus was only changed with respect to hue. An experiment in which both value and chroma are changed is necessary.

In this research, we classified color stimuli into four categories to check the shape and the border of the color categories of varied color vision. The number of color stimuli were 116, of which 85 were the same value and chroma of the object color and 31 color stimuli were monochromatic light from 400 nm to 700 nm. An experiment was conducted to find the border between categories.

The number of subjects were 26: 4 males and 3 females who passed the Ishihara color test, 5 males with protanomaly, 8 males with protanopia, 1 male with deuteranomaly, and 5 males with deuteranopia.

The degree of similarity between stimuli was recorded, and then converted into distance data between the stimuli for multidimensional scaling. The distance data was classified in four categories by cluster analysis. These categories were used to calculate the border between categories.

The experimental result was as follows. The border of protanopia is the following three on the CIE 1931 (x, y) chromaticity diagram: y=-0.3068x+0.4796, y=-0.2339x+0.41334, y=-0.3135x+0.4023.

The border of deuteranopia is the following three on the CIE 1931 (x, y) chromaticity diagram: y=-0.8375x+0.72703, y=-0.8152x+0.6243, y=-0.6667x+0.5061.

9395-34, Session 6

Investigation of the Helmholtz-Kohlrausch effect using wide-gamut display

Semin Oh, Youngshin Kwak, Ulsan National Institute of Science and Technology (Korea, Republic of)

The Helmholtz-Kohlrausch effect describes that perceived brightness of a color is affected not also by luminance but also by the purity. That is, perceived brightness can increase as a chroma of the object increases. The aim of this study is to investigate whether the Helmholtz-Kohlrausch effect exists among the images having various chroma levels.

Firstly, one image containing various colors was selected. Then this image was adjusted to have various average Chroma and Brightness. To adjust the average Brightness and Chroma of the image, the CIECAM02 J and C of each pixel of the original image were linearly adjusted. The image was adjusted to have 4 different average C values and 5 different average J values. The average CIECAM02 C values of test images were19.45, 29.17, 38.89 and 48.61 and the average CIECAM02 J value was 24.56, 28.46, 32.38, 36.30 and 40.26. The luminance of peak white point in each adjusted image was 68.84, 99.46, 137.72, 183 and 235.37(cd/m2). The image was adjusted to have all possible average CIECAM02 C and J combinations. Therefore total 20 test images were generated for the psychophysical experiment. including one reference image. The image having average CIECAM02 C value of 29.18 and J value of 32.21 was selected as the reference image,

The psychophysical experiment was done in a dark room. Each test image was shown on the wide-gamut LCD display along with the reference image on the black background. The size of both test and reference images was 5.63x9(cm). To evaluate the overall perceived brightness of the test images, a magnitude estimation method was used. Each participant evaluated the overall brightness of each image in a number comparing with the reference image assigned to have 50. Before starting the experiment, training was done to let the participants understand the concept of chroma and lightness by using a Munsell student's workbook. Then 5 participants evaluated all the test images and the order of showing images was randomized per each participant. To analyze the data, all participants' responses were averaged.

The results showed that as the average CIECAM02 J of the image increases, participants evaluated the brightness higher when the average CIECAM02 C value is the same. Also as the average CIECAM02 C of the image increases, participants tended to evaluate the brightness higher while the average CIECAM02 J value is the same proving the Helmholtz-Kohlrausch effect. For example, participants evaluated the image which has average CIECAM02 C of 48.61 looked brighter about 19% than the image which has average CIECAM02 C of 19.45. Also it is found that the Helmholtz-Kohlrausch effect becomes larger as the average Brightness of the image increases. Currently, further experiments are in progress with more test images. Final experimental results will be reported in the final paper.

9395-35, Session 6

Preferred tone curve characteristics of transparent display under various viewing conditions

Sooyeon Lee, Hye-young Ha, Youngshin Kwak, Ulsan National Institute of Science and Technology (Korea, Republic of); Hyosun Kim, Youngjun Seo, Byungchoon Yang, Samsung Display Co., Ltd. (Korea, Republic of)

Transparent display makes users to see the scene behind the display alongside the images on the display. See-Through characteristic of a display allows providing new application which other types of displays cannot have. However, those advantages also could give image quality degradation to us since the contrast of the reproduced images will be reduced because of the transmitted light. Understanding of the effect of transmitted lights on transparent display image quality will be helpful to make better use of the transparent display. In this study the OLED-type transparent display was simulated on LCD monitor by adding background scene to the reproduced images. The effect of the level of transmission and surround luminance on the preferred tone curve is investigated.

For transparent display simulation, it is assumed that the transparent displays have four different transparency levels, 17, 52, 70 and 87%, which were shown under four levels of surround luminance, dark (0 cd/m2), dim (28 cd/m2), average (70 cd/m2) and bright (231 cd/m2). The color gamuts of all the transparent displays were set to sRGB with 140 cd/m2 peak white. Tone curves of transparent displays were controlled by simulating various monitor gamma. Four test images, 3 pictorial images having skin, green grass, blue sky and various chromatic colors, and 1 captured web site, were selected for the experiment. 312 images having different transmission and surround condition were generated. 15 university students were participated in the psychophysical experiment. They passed Ishihara test and 100 FM hue

test. They had adaptation time under each surround condition for 5 minutes and were asked to answer the degree of preference of each image under each condition using 7-point Likert-scale (7 : the most preferred, 1 : the least preferred). The data was analyzed using Torgerson's Law of Categorical Judgment Condition D.

As the result, it is found that as the surround luminance and transmission increase, preferred images tend to have lower gamma. The surround luminance changes affected the preferred gamma more significant than that of transmission level. The captured web site has no significant difference for the different level of transmission and surround luminance. Those results suggest that image quality requirement for a transparent display is different from conventional opaque displays. Therefore, further studies on image quality of transparent displays are needed for the widespread of transparent display.

9395-36, Session 6

A method for estimating colors of scene illuminants under spatially non-uniform environments

Harumi Kawamura, Ayumi Matsumoto, Akira Kojima, Nippon Telegraph and Telephone Corp. (Japan)

In color image analysis, colors have rich information to process. However, they substantially affect the detecting of objects or the extracting of features from an image. This is because the colors in an image include those of the scene illuminants irradiating the objects in the image. Thus, estimating the colors of the scene illuminants and removing them is a useful technique in color imaging.

Most conventional methods of estimating scene illuminant colors appearing in an image assume that there is only one light source and that the illuminant spans a scene uniformly. However, the validity of this assumption is limited since in most cases there are several scene illuminants and their illumination rates vary depending on the distance and direction to the light source. Several methods have focused on the several-illuminants case; however, they assume that one light source exists in a small region of the image and cannot estimate illuminant colors when two different light sources are irradiating that region.

Our method makes it possible to estimate the mixed colors of scene illuminants comprising two different light sources, such as fluorescent light and daylight. The method estimates the illumination rates of two illuminants irradiating objects in a scene in cases where the rates vary from location to location. Generally, the colors of a region in an image are expressed as a product of the sensitivity function of a color sensor, the region's surface reflectance, and illuminants, which we formulate as the product of two illuminants and their rates. Therefore, the colors of the region are expressed as a function of illumination rates. In the estimation, the colors of two regions that have the same surface reflectance but in different locations are used. The method uses the property that the colors derived in the regions comprise the plane through the origin in a three-dimensional color space, meaning the plane is unique to the surface reflectance. By determining the normal of the plane, the surface reflectance common to each color region is estimated. Thus, using the estimated surface reflectance makes it possible to convert the two colors into the illumination rates and to use the rates to estimate the colors of the scene illuminants.

We conducted experiments by using numerical simulations; the rates of the two illuminants were estimated by using two simulated colors reflected on the surface of the same objects but in different locations. The two colors were calculated using the spectral distribution of fluorescent lamp (4000 K) and CIE daylight (6500 K) and 235 kinds of "typical" reflectance from the ISO-TR 16066 object colour spectra database. The illumination rates were fixed as 1.0:0.1 and 0.1:0.5 for each reflectance. The obtained results showed that for most cases in the dataset the illumination rates of the two illuminants were estimated to have values similar to those cited above. However, the estimations were less accurate in cases when the reflectance was almost flat in the visible wavelength range because of the validity of surface reflectance representation using the basis function.

9395-37, Session 7

How colorful! A feature it is, isn't it?

Fritz Lebowsky, STMicroelectronics (France)

A display's color subpixel geometry provides an intriguing opportunity for improving readability of text. True type fonts can be positioned at the precision of subpixel resolution. With such a constraint in mind, how does one need to design font characteristics? On the other hand, display manufactures try hard in addressing the color display's dilemma: smaller pixel pitch and larger display diagonals strongly increase the total number of pixels. Consequently, cost of column and row drivers as well a power consumption increase. Perceptual color subpixel rendering using color component subsampling may save about 1/3 of color subpixels (and reduce power dissipation). This talk will try to elaborate the following questions while considering several different layouts of subpixel matrices: Up to what level are display device constraints compatible with software specific ideas of rendering text? Can simplified models of human visual color perception be easily applied to text rendering on displays? How linear is human visual contrast perception at band limit of spatial resolution? How much does visual acuity vary at 20/20 vision? How to best consider preferred viewing distance for readability of text? How colorful does the rendered text appear on the screen? How much of color contrast will remain?

9395-38, Session 7

Why simulations of colour for CVD observers might not be what they seem

Phil Green, Gjøvik Univ. College (Norway)

No Abstract Available

9395-39, Session 8

An interactive app for color deficient viewers

Cheryl Lau, Nicolas Perdu, Ecole Polytechnique Fédérale de Lausanne (Switzerland); Carlos E. Rodriguez Pardo, Univ. of Rochester (United States); Sabine Süsstrunk, Ecole Polytechnique Fédérale de Lausanne (Switzerland); Gaurav Sharma, Univ. of Rochester (United States)

About 7% of the US male population has a red-green color deficiency [Montgomery 1995]. These individuals encounter some difficulties in everyday life such as telling when meat is cooked, detecting a sunburn, and picking out ripe bananas from unripe bananas [Flück 2010]. Tasks that involve matching colors, such as selecting matching clothes, are especially difficult. LED lights on electrical devices and charts and graphs in documents and presentations pose a problem to these individuals since products are often designed for individuals with normal color vision. Various smartphone apps are available to help color deficient individuals perform daily tasks and to enable others to visualize what the world looks like as a color deficient individual. As there are different types of color deficiencies, some of these apps would benefit from knowing what type of deficiency the user has in order to provide enhanced customization. We present an app to determine a user's color deficiency type.

Smartphones have become extremely popular and present a convenient way for accessing information. In 2013, about 65% of US mobile consumers owned smartphones [Nielsen 2014]. These personal devices can be customized and personalized to suit each owner's individual needs. As a small, portable device, a smartphone is easy to carry around for convenient access throughout the day. With direct finger input, the touchscreen provides an intuitive interface for the user to interact with many apps. Taking advantage of the prevalence of smartphones and their touchscreen inputs, we develop an app for testing for color deficiencies that can be

conveniently administered. Knowing the user's color deficiency type gives other apps the opportunity for further customization.

Color deficiencies can be categorized into three types, according to which of the three normal cone types is affected. The L cone is affected in the protan case, the M cone is affected in the deutan case, and the S cone is affected in the tritan case. The protan and deutan cases, comprising the red-green deficiencies, form the majority of color deficient viewers while the tritan case is very rare. Our app can distinguish between the protan, deutan, and tritan cases. The Ishihara plate test is the most widely used test for determining whether an individual has a red-green color deficiency or not. During this test, the observer looks at printed plates containing colored dots that form a pattern, and the observer has to identify the number in the image or trace the winding path in the image. As the test is mainly designed to distinguish between color-normal and red-green deficient observers, only six of the 38 plates are designed to help distinguish between protans and deutans, and four of the six involve reading numerals. In contrast, our test is designed to distinguish between protans, deutans, and tritans. During our test, the user matches colored squares. This task does not require a knowledge of language or numerals and with its right/wrong binary result, is easier to judge than a winding path trace.

We present an app with a simple test to distinguish between color deficiency types, especially between the common protan and deutan red-green types. There are multiple rounds in the test, and on each round the user is presented with four colored squares. Two of the four squares are extremely similar in color while the other two are very different, and the user must choose the two closest matching squares. We choose the four colors to be colors that would be hard to distinguish for a certain type of deficiency but possible to distinguish for the other deficiency types. For each round, we generate confusing colors for a particular deficiency type, and after all the rounds, the user is categorized as the deficiency type for which he got the least amount of correct answers.

Our method for selecting the colors of the four squares is based on dichromatic perceptual models. Dichromats are those that lack one of the three cone types. For the four squares, we choose three color samples along the viewer's confusion line and randomly choose one to duplicate. To avoid exact matches, we then add a small random perturbation vector to the duplicated color. We specifically choose confusion lines for one viewer type that have colors that the other viewer types can distinguish easily. For example, consider a test round of four squares designed for a protanope. All four squares will look similar to the protanope as the colors have been sampled from the protanope's confusion line, but the three dissimilar colors are possible to distinguish by deuteranopes and tritanopes enabling them to easily select the two matching squares. To ensure this principle, we test randomly generated protanopic confusion lines and keep only the ones whose extremes, bounded to the linear RGB gamut, have visible differences above a threshold for deuteranopic and tritanopic viewers. This is done for the other two viewer types to select confusion lines containing colors that are hard for the viewer to distinguish but possible for the other viewer types to distinguish. We conducted preliminary user tests and find that that test distinguishes between protanopic, deuteranopic, and tritanopic perception with high reliability.

D. Flück. 2010. Living with color blindness. http://www.color-blindness.com/2010/03/30/ living-with-color-blindness/

G. Montgomery. 1995. Seeing, Hearing, and Smelling The World: Breaking the Code of Color. A Report from the Howard Hughes Medical Institute.

Nielsen. 2014. The U.S. Digital Consumer Report. http://www.nielsen.com/us/en/insights/reports/2014/the-us-digital-consumer-report.html

9395-40, Session 8

Evaluating color deficiency simulation and daltonization methods through visual search and sample-to-match: SaMSEM and ViSDEM

Joschua Simon-Liedtke, Ivar Farup, Gjøvik Univ. College (Norway); Bruno Laeng, University of Oslo (Norway)

Color image quality in images can be significantly reduced for color deficient people since color contrast and/or perception of edges is reduced leading to problems in retrieving relevant information from the image or identifying objects in the image. Several daltonization methods have been proposed to regain color image quality and color deficiency simulation methods have been proposed to better under- stand difficulties of color deficient observers for normal sighted ob- servers. How well these methods perform, and how accurate they depict or enhance vision of color deficient observers is mainly left unknown. We propose methods that compare and rank different color deficiency daltonization and simulation methods by using behavioral experiments, in which we analyze respond times and accuracy of sample-to-match and visual search tasks.

Most commonly, daltonization methods are verified by applying them on Ishihara test plates, and checking whether color deficient people are able to retrieve the information or not. However, Ishihara test plates do not satisfactorily represent real-life images since they are graphical products and the image points are separated from each other by white background, something that rarely exist in real-life images. Thus, we propose methods that use real-life images for evaluation purposes instead.

The first evaluation method is a sample-2-match task to evaluate the performance of color deficiency simulation methods. In this two-step experiment, the observer is firstly shown either the original or the simulated version of an image for a couple of seconds, and then secondly he/she is shown both the simulated version and the original version. The task for the observers is to answer whether the previously shown image is now located on the left or on the right. We tested the method on 4 simulation methods – simulation methods by Viénot, and an adjusted Viénot method, Brettel and Kotera – using 44 different images.

The second evaluation method is a visual search task to evaluate the performance of color deficiency daltonization methods. We pre- pared different images associated with certain tasks that can only be answered correctly depending on the color hue of the content of the image – do berries have the same color as the background, do cloths have the same color etc. –, and colored the objects in the images with colors that are difficult for color deficient observers. The observers were shown the images, and a statement associated with the image is presented below the images. The task for the observers is to answer whether he/she agrees to the statement or not. We tested the method on 4 daltonization methods – daltonization methods by Anagnostopoulos, Kotera, Kuhn and Huan – using 23 different image tasks.

The experiment has been conducted with 23 observers, of which 14 were color deficient of different types and severenesses, and of which 9 were normal sighted. A detailed analysis of the result will be included in the final paper. Also, the ranking of the different color deficiency simulation and daltonization methods will be presented in the final paper.

9395-41, Session 8

Image color reduction method for color-defective observers using a color palette composed of 20 particular colors

Takashi Sakamoto, National Institute of Advanced Industrial Science and Technology (Japan)

This paper describes a color enhancement method using a color palette especially designed for protan and deutan defects (commonly known as red-green color blindness).

Color vision defects occur when the cells in the retina fail to function normally. In many cases, the dysfunction is caused by particular genetic factors and affects a significant number of people. In the USA, approximately 7% of males and 0.4% of females suffer from either protan or deutan color vision defects. Globally, there are more than 200 million people who suffer from some form of congenital color vision defect. Such people have difficulty distinguishing particular sets of colors such as red and black, red and green, pink and sky blue, and pea green and yellow.

Congenital color vision defects are caused genetically, and the condition is currently incurable. The most reasonable and effective way to assist people with color vision defects is to improve their visual environment. This means not only improving the color design of the surrounding environment (e.g., in traffic signs, direction boards, maps, and various types of print), but also replacing the colors used by imaging devices (e.g., in TVs, projectors, electronic bulletin boards, PC monitors, and smart-phones). Image color enhancement is the one effective way to replace particular sets of colors confused by color-defective observers.

More than a dozen color enhancement methods and software programs to assist people with color vision defects have been reported. These methods and software programs are based on mathematical backgrounds and calculation. Consequently, they require complicated and substantial processing. Such limitations may affect real-time processing when imaging devices do not have powerful CPUs or image processing engines.

The proposed method is based on a simple color mapping. It does not require complicated computation, and it can replace confusing colors with protan/deutan-safe (p/d-safe) colors. There are no reports or papers describing a simple color mapping for p/d-safe photographic enhancement. Color palettes that have been proposed are typically composed of few p/d-safe colors. Hence, the number of colors is insufficient for replacing the colors in photographs.

Fortunately a p/d-safe color palette composed of 20 particular colors has been proposed in June, 2009 [Ito et al., 2009]. The author found that this particular p/d-safe color palette could be applied to image color reduction in photographs. Thus, it provides a means to replace confusing colors in photographs for protan and deutan defects. The proposed method is unsuitable for high-definition color imaging, however, because of the image degradation caused by the color reduction. Yet, it can be executed without any processing delays by using a lookup table, and it can be implemented in low-performance information devices such as smart-phones.

This paper shows the results of the proposed color reduction in photographs that include typically p/d-confusing colors (viz., red and black). These confusing colors can be replaced and color-defective observers can distinguish them after the proposed color reduction process.

9395-42, Session 8

Adaptive colour rendering of maps for users with colour vision deficiencies

Anne Kristin Kvitle, Phil Green, Peter Nussbaum, Gjøvik Univ. College (Norway)

A map is an information design object for which canonical colours for the most common elements are well established. For a CVD observer, it may be difficult to discriminate between such elements - for example, for an anomalous trichromat it may be hard to distinguish a red road from a green landscape on the basis of colour alone. This problem becomes more intractable as more complex geographic information is overlaid, particularly on mobile devices where the calibration state and viewing conditions are unknown and therefore the precise colour stimulus cannot easily be predicted by the map designer.

In the Hypercept project, we address this through an adaptive colour schema in which the conspicuity of elements in a map to the individual user is maximised. The principal objective of this paper is to outline a method to perform adaptive colour rendering of map information for customised users with colour vision deficiencies.

The palette selection method is based on a pseudo-colour palette generation technique which constrains colours to those which lie on the boundary of a reference object colour gamut.

A user performs a colour vision discrimination task designed to evaluate the user's relative discrimination in different regions of colour space. Analogously to the Farnsworth-Munsell 100-hue test, users arrange colour patches by hue, and the hue difference between samples is progressively reduced to determine the users' threshold discrimination level under the device display and viewing conditions. No prior knowledge of the user's particular vision deficiency is needed to interpret the test results.

Based on the results of the test, a palette of colours is selected using the pseudo-colour palette generation method, which ensures that the perceived difference between palette elements is high but which retains the canonical colour of well-known elements as far as possible. Colours are transformed between CIE colour space and device encoding using ICC colour management. The extensibility of the palette selection method, allows larger palettes, to be added by re-computing the palette while continuing to retain conspicuity between elements.

We show examples of colour palettes computed for a selection of normal and CVD observers, together with maps rendered using these palettes. In subsequent phases of the project we will refine the palette selection method and perform tests with observers.

9395-43, Session 8

Spatial Intensity Channel Replacement Daltonization (SIChaRDa)

Joschua Simon-Liedtke, Ivar Farup, Gjøvik Univ. College (Norway)

Color image quality in images can be significantly reduced for color deficient people since color contrast and/or perception of edges is reduced, leading to problems in retrieving relevant information from the image or identifying objects in the image. Global daltonization methods have been proposed to regain color image quality by adding lost information from out-of-gamut colors to the original image globally.

We propose two methods that provide better solutions for color deficient observers – namely protanopes, deuteranopes, protanomal and deuteranomal observers – by increasing correlation between the layers of the original color image using spatial color-to-gray methods that integrate information from the original image channels into the lightness channel. We propose using spatial color-to-gray methods that are either capable of translating color contrasts into lightness contrasts or that are capable of translating color edges into lightness edges.

In the first method we compute a gray image from all three RGB channels and replace the original intensity channel in IPT space, as defined by Ebner et al., with the newly computed gray image. This truly leads to redundant information in the intensity channel, but makes it possible for color deficient observers to retrieve information that would otherwise be hidden for them in the color channels.

In the second method we enhance the previously proposed method by additionally adding information from the IPT image's red–green channel to the newly computed gray image. This puts additional focus on difficult red–green contrasts that might lead to confusion for color deficient observers like anomalous trichromats, respectively contrasts that would otherwise be invisible since they are hidden in the red–green channel for color deficient observers like dichromats.

We tested two implementations using color-to-gray methods that focus on preserving color contrast proposed by Ali Alsam et al. and/or color edges proposed by Øyvind Kola?s et al. on different types of images leading to moderate results depending on the content of the image. It can be shown that the spatial methods work best on real-life images were confusing colors are directly adjacent to each other, respectively where they are in close proximity. On the other hand, composed artificial images with borders of white space between colors – like for example in the Ishihara plates – lead only to unsatisfactory results, as expected.

9395-44, Session 8

Preferred memory color difference between the deuteranomalous and normal color vision

YeSeul Baek, Youngshin Kwak, Ulsan National Institute of Science and Technology (Korea, Republic of); Sungju Woo, Chongwook Park, KAIST (Korea, Republic of)

The goal of this study is to evaluate the difference of the preferred hues of familiar objects between the color deficient observer and the normal observer. A 24-inch LCD monitor was used and the color gamut of this monitor was set to sRGB. The maximum luminance of display was 245 cd/m2. Thirteen test color images were selected covering fruit colors, natural scene and human faces. The images contained red (apple, steak and tomato), yellow (orange, lemon and banana), green (kiwi and watermelon), blue (blueberry and sky), purple (cosmos) and skin color (Oriental people). The image manipulation tool was prepared using MATLAB program to control hue of the input image. To manipulate the test images, sYCC color space was used. The RGB values of the test images were converted to sYCC values. After hue adjustment, YCC was converted back to RGB color space. One color deficient observer (male) and two normal color vision observers (1 female and 1 male) were participated in the psychophysical experiment. For color deficient observer, type of color vision deficiency and deficiency severity were tested using the Ishihara plates and the Farnsworth–Munsell 100-hue. As a result, he was deuteranomal and moderate to severe with a total error score of 144.

The psychophysical experiment was conducted to find out the preferred hue angle. The observer freely changed the hue angle of the objects of the test images using the image manipulation tool and the hue-transformed image was immediately displayed on the monitor. They were asked to generate the most preferred image and the most natural image.

The selected images were analyzed using CIELAB values of each pixel. The images selected by color deficient observer were converted using the color deficiency simulation algorithm developed by Mochizuki et al.

Data analysis results showed that in the case of naturalness, both groups selected the similar hues for the most of image. It is assumed that color deficient observer choose the similar color with normal because of the experience for long periods of time. On the other hand, in the case of preference, the color deficient observer preferred more reddish images. In the case of red and yellow images, the preferred hue was rotated to a clockwise direction in CIELAB a*b* plane. It means the color deficient observer preferred more reddish (even purplish) color than normal vision observers. However in green and blue images the preferred hue was rotated to a counter-clockwise direction. Since the deuteranomalous observer has relatively week perception for red and green region, they may prefer more reddish or greenish color.

Conference 9396: Image Quality and System Performance XII

Tuesday - Thursday 10–12 February 2015

Part of Proceedings of SPIE Vol. 9396 Image Quality and System Performance XII

9396-1, Session 1

Advanced mechanisms for delivering high-quality digital content

Mikolaj I. Leszczuk, Lucjan Janowski, AGH Univ. of Science and Technology (Poland)

A large variety of multimedia data sources leads to a variation, not only in terms of content, but also in terms of the quality presented.

In terms of mechanisms for delivering high-quality digital content, a common feature of existing systems is the need to ensure an acceptable quality of streamed video sequences, regardless of the load transmission medium, the type of access network, and the terminal equipment employed by the end-user.

In this paper, we present research which, due to the above requirements, seeks to develop a system for ensuring adequate levels of quality of video sequences. The system consists of metrics to assess video quality and mechanisms for optimizing quality using information provided by the metric. The aim of the study was to develop ways of providing high-quality multimedia content. The use of different methods for the transcoding and adaptation of multimedia content was planned. Metrics and algorithms were designed to assess the quality of video sequences.

Modern and universal systems of databases containing vast numbers of videos will only be fully useful when it is possible to deliver them to users with an acceptable quality. Therefore, the aim of the research was to provide this. Consequently, the main innovation of the proposed solution is a comprehensive evaluation of the perceived quality of video, and optimization based on information provided by the proposed metrics. The complexity of the evaluation consists of taking into account both distortion typical of the video acquisition process, and that caused by lossy compression for streaming needs. By conducting quality tests and using advanced statistical methods for the analysis of the data received, it was possible to create models simulating the measured values of quality parameters on the resulting video quality, expressed as a scale intelligible to the user, i.e. the 5-step Mean Opinion Score scale. Another important issue was to optimize quality based on information supplied by metrics either stored in the form of meta-data or computed on demand. Optimizing quality takes into account not only the network parameters, but also the characteristics of video sequences, which is an important innovation.

2 Large Set of Coding Parameters

As the first step of this research, a large database of different coding parameters was created. For each sequence in the database, a set of full reference quality models were computed.

The considered full reference quality metrics are: Peak Signal-to-Noise Ratio – PSNR, Structural Similarity Index Metric – SSIM, Video Quality Metric – VQM, and Visual Information Fidelity – VIF.

The previously mentioned FR measurement algorithms were compared.

VQM has been shown to be the best FR metric, having a fit factor of subjective data higher than other metrics by 20% or above. Therefore the VQM metric was used for further analysis of the obtained results.

3 Application Implementation

A possible application area of this database is the development of methods for delivering high-quality multimedia content.

3.1 Controlling Compression

Having a quality metric makes it possible to describe the database of all subjective test results, so that the best can be selected from the source for each sequence. However, with such a large number of sequences, some differ only minimally in terms of quality. For example, the number of slices, as anticipated, has no impact on quality. However, due to the calculation rounding problem of the same metric, values obtained for the different sequences are almost always different. Therefore, as well as choosing the best sequence, it was necessary to choose sequences which differed slightly in terms of quality.

Clustering analysis of the results allowed the group to replace the most suitable examples of the sequences whose quality was not worse than the quality of the best sequences +0.1. A coefficient of 0.1 resulted from the analysis of clusters.

Analysis of small groups of the sequences aimed to select the most general compression parameters. Ideally, the parameters do not all come from the same sequence, as the dependence of the parameters of compression forces reference analysis of the same sequence, which greatly complicates the process. Of course, if it has a great impact on the quality of the obtained sequence, analysis of the reference cannot be ignored.

The final result of the analysis of the data received is a function of the parameters of the compression rate.

3.2 Streaming

Compliance with the five most popular web browsers has been demonstrated. The only browser which offers native support for all play modes in HTML5/H.264 is Google Chrome. For the other, popular web browsers, a solution temporarily circumventing the problem is available, consisting of a Windows Media Player plug-in for Windows and VideoLAN Client for Linux and OS X.

The Matroska container has been directed as a multimedia container file. This format, in addition to supporting the x264 encoder, also enables variable bit rate, variable image resolution and variable frame rate video streaming.

3.3 Integration

Selecting the encoder, and defining the compression parameters and the type of multimedia container, made it possible to create a script for video compression while maintaining the quality of the image.

An application written in the scripting language Bash processes the video file, selects the encoding parameters, and performs its compression.

3.4 Tests

As a test, the video transcoder was checked for correct operation for different: 1. Multimedia container file inputs

2. Encoding formats as input file

3. Quality models

4. Erroneous input data

The following factors were taken into account: output video, and the type of error information generated.

4 Conclusions

The application works correctly for a wide range of input file formats, different types of multimedia containers and types of compression. During the test, problems occurred with just one format. The transcoder correctly interprets changes in the video quality model definition and uses its parameters. In the event of errors, the user is informed of their occurrence, so long as the nature of the error can be identified.

9396-2, Session 1

Towards assessment of the image quality in the high-content screening

Yury Tsoy, Institut Pasteur Korea (Korea, Republic of)

High-Content Screening (HCS) [1] is a powerful technology for biological research, which relies heavily on the capabilities for processing and analysis of cell biology images. The quality of the quantification results, obtained by analysis of hundreds and thousands of images, is crucial for reliable decision-making and analysis of biological phenomena under

study. Traditionally, a quality control in the HCS refers to the preparation of biological assay, setting up instrumentation, and analysis of the obtained quantification results, thus skipping an important step of assessment of the quality of images to be analyzed.

So far, only few papers have been addressing this issue, but no 'standard' methodology yet exists, that would allow estimating imaging quality and pointing out potentially problem images. Manual quality control for images is apparently impossible due to the large number of images. Even though HCS readers and software are gradually improving, by using more advanced laser focusing, compensation of abberations, using special algorithms for noise reduction, deconvolution, etc. it is impossible to eliminate all the possible sources of corruption of the image quality.

In this research the importance of the image quality control for the HCS is emphasized. Possible advantages of detecting problem images are:

- Validation of the visual quality of the screening.

- Detection of the potentially problem images.

- More accurate setting of the processing parameters.

For the estimation of visual quality the Power Log-Log Slope (PLLS) [2] is applied. Previously by Bray et al. [3] it was shown that for the case of bimodal distribution of PLLS values, it is possible to detect out-of-focus images. Further discarding the images with bad quality allowed significant improvement of accuracy of the well-plate analysis results. However several important questions were not addressed by Bray et al., such as:

- possible loss of information due to gating of wells from the analysis;

- gating of wells for the plates having images with more than 1 band;

- analysis of images with large PLLS values;

- analysis of screens with unimodal distribution of PLLS values;

Resolving these drawbacks might improve the overall reliability of the image quality assessment and increase accuracy of the HCS data analysis. The following improvements are proposed:

- the wells with 'abnormal' PLLS values are not excluded from analysis, but marked as potential sources of outliers;

- the information about outliers is used for validation of the plate quality and HCS quantification (e.g. cell counts, infection ratio) and analysis results (e.g. DRC parameters, hit/target candidates) and making recommendations for elimination of outliers;

This approach greatly contributes to the through quality control as it covers large part of the HCS processing pipeline starting from assay optimization and ending with DRC parameters or target candidates. Preliminary research results show that such approach allows detection of unreliable estimation of the Dose-Response Curve parameters even when R2 value is above 0.8.

References:

1. Zanella F; Lorens JB & W., L. High content screening: seeing is believing Trends Biotechnol., 2010, 28, 237-245.

2. Field DJ, Brady N. Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. Vision Res. 1997;37:3367–3383

3. Bray, M-A, Fraser AN, Hasaka TP, Carpenter AE Workflow and metrics for image quality control in large-scale high-content screens. 2012, Journal of Biomolecular Screening 17(2):135-143

9396-4, Session 1

Information theoretic methods for image processing algorithm optimization

Sergey F. Prokushkin, DCG Systems Inc. (United States); Erez Galil, Univ. of California, Santa Cruz (United States)

Modern image processing pipelines (e.g., those used in digital cameras) are full of advanced, highly adaptive filters that often have a large number of tunable parameters (sometimes > 100). This makes the calibration procedure for these filters very complex, and the optimal results barely achievable in the manual calibration; thus an automated approach is a must and selection of an appropriate algorithm performance metric becomes a necessary and important task. Although the ultimate goal of the filter calibration is to achieve an optimal perceived image quality, or signal fidelity, we argue that an important pre-requisite step is optimization of the algorithm adaptive characteristics, i.e. its ability to differentiate between signal and noise. In other words, the total parameter calibration problem can be split into the two simpler sub-problems: first, finding a subspace of the full parameter space that corresponds to the maximal physical information transfer, and second, finding within this subspace a domain that provides the best subjective image quality. Existing methods for signal fidelity evaluation are based on full reference quality assessment (FR QA) where test images are compared against a reference high quality image, by measuring a 'distance' between a test and a reference image in perceptually meaningful way. We propose an approach in which the adaptive filter is benchmarked against a (non-adaptive) linear filter of the equivalent strength, and the adaptive filter parameters are adjusted to maximize its adaptive characteristics. A special calibration chart consisting of resolution and noise measurement regions is proposed to make the benchmarking simpler. Using this calibration target, an information theory based metric for evaluation of algorithm adaptive characteristics ('Adaptivity Criterion') is calculated that can be used in automated filter optimization. Specifically, we use a version of the integral Shannon-Fellgett-Linfoot formula for information capacity to extract the 'information transfer' through the filter. Many advanced filters are often placed closer to the end of an image processing pipeline where the noises in different channels become correlated and cross-dependent on the signal values. As a result, basic statistical characteristics become difficult to evaluate. In particular, different spectral components of the signal become statistically correlated and their information capacities cannot be simply added, due to a substantial mutual information present. However, one can argue that in many practical cases, the upper bound information capacity value (as in the original Shannon's theorem) can be sufficient to assess and optimize filter 'adaptivity'. We illustrate the method's application on the example of a noise reduction algorithm calibration.

The key characteristics of the method is that it allows to find a sort of 'an orthogonal decomposition' of the filter parameter space into 'filter adaptivity' and 'filter strength' directions. Since the Adaptivity Criterion is a measure of a physical "information restoration" rather than perceived image quality, it helps to reduce the set of the filter parameters to a smaller subset that is easier for a human operator to tune and achieve a better subjective image quality. With appropriate adjustments, the criterion can be used for assessment of the whole imaging system (sensor plus post-processing).

9396-5, Session 1

Forward and backward tone mapping of high dynamic range images based on sub band architecture

Ines Bouzidi, Azza Ouled Zaid, National Engineering School of Tunis (Tunisia)

1. PROBLEM STATEMENT

Despite the progress in digital image processing, most of today used images are still limited when compared to the real shot scene. This limitation is more pronounced when the concerned scene has a high degree of contrast. In this case, the highlights tend to wash out to white, and the darks become big black blobs. High Dynamic Range (HDR) images appear to be an effective solution to faithfully reproduce a realistic scene as they have the advantage to handle a wide range of tonal values. However, display devices such as CRTs and LCDs cannot display such images. In this context, different Tone Mapping (TM) algorithms1, 2 have been developed to reduce the dynamic range of the images, while maintaining their significant details. On the other hand, the emergence of new HDR displays including those of mobile devices raised the issue of enhancing Low Dynamic Range (LDR) content to a broader range of physical values that can be perceived by the human visual system. In the recent years, the production of HDR image from a series of LDR pictures with different exposure times, is being of practical use. However, the problem is more difficult when considering a single LDR image without any extra HDR information.

2. PROPOSED METHOD

In our work, we designed a novel tone mapping/inverse tone mapping system based on subband architecture. This system is inspired from the companding framework developed by Li et al.3 During the TM stage, the HDR image is firrstly decomposed in subbands using symmetrical analysis-synthesis filter bank. The subband coefficients are then rescaled by using a predefined gain map. The main advantage of our technique is that it allows range expansion. During the inverse Tone Mapping (iTM) stage, the LDR image is passed through the same subband architecture. But, instead of reducing the dynamic range, the LDR image is boosted to an HDR representation. Moreover, the developed iTMO results in high fidelity content compared to the original HDR image. During the LDR expansion, an optimization module has been used to select the gain map components that minimize the reconstruction error. As a result, our iTMO results in high fidelity content compared to the original HDR image. It is worth mentioning that the tuning parameters of our TM system are not image dependent. Consequently, our solution can be used in conjunction with any lossy or lossless LDR compression algorithm. Another benefit of the proposed iTM framework is that HDR content is constructed directly from the LDR representation without the use of any extra information.

The proposed HDR TM algorithm involves the following steps:

1. Luminance information is first extracted from the original HDR image.

2. TM process is achieved on luminance channel using subband architecture.

3. The obtained LDR image is expanded to construct a first HDR version.

4. A residual luminance is then computed by subtracting the recovered HDR image from the original one.

5. The residual information is then TM and added to the first LDR image estimation. The three last steps

are iterated till a specific condition is satisfied.

6. The corrected LDR image is finally saved in any LDR supported format (bmp, jpeg, png, ...)

3. RESULTS

To examine the effectiveness of the proposed TM/iTM scheme, we start by analyzing the quality of the obtained LDR images compared with those obtained by other TM methods. For this purpose we use two assessment metrics. The first one is a blind/no-reference metric named Q metric.4 The second one, named tone-mapped image quality index (TMQI),5 evaluates the quality of the tone mapped image using its corresponding HDR content as reference. The TMQI metric combines a signal fidelity measure with naturalness estimation. Finally, we evaluate the visual quality of the retrieved HDR images according to the original ones. Simulations have been performed on four different HDR images. We compare our results to state of the art TM operators and two Photoshop TM methods. The quality evaluation of the tone mapped images using the Q metric indicates that our method has better results compared to the reference methods. Based on the TMQI results, we can notice that, compared to the state of the art TM methods, our framework performs the best since it delivers LDR images which faithfully preserve the structural fidelity of the original HDR images.

REFERENCES

1. F. Drago, K. Myszkowski, T. Annen, and N. Chiba, Adaptive logarithmic mapping for displaying high

contrast scenes," Computer Graphics Forum 22, pp. 419{426, September 2003.

2. F. Durand and J. Dorsey, Fast bilateral _ltering for the display of high-dynamic-range images," ACM

Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 21, pp. 257{266, July 2002.

3. Y. Li, L. Sharan, and E. Adelson, Compressing and companding high dynamic range images with subband

architectures," ACM Transactions on Graphics 24(3), pp. 836{844, 2005.

4. X. Zhu and P. Milanfar, Automatic Parameter Selection for Denoising Algorithms using a No-Reference

Measure of Image Content," IEEE Transactions on Image Proessing 19, pp. 3116{3132, December 2010.

5. H. Yeganeh and Z. Wang, Objective Quality Assessment of Tone-Mapped Images," IEEE Transactions on Image Processing 22, pp. 657{667, February 2013.

9396-6, Session 1

Perceptual patch-based specular reflection removal for laparoscopic video enhancement

Bilel Sdiri, Univ. Paris 13 (France) and Gjovik Univ. College (Norway); Azeddine Beghdadi, Univ. Paris 13 (France); Faouzi Alaya Cheikh, Gjøvik Univ. College (Norway)

Removing specular reflection is an important preprocessing step that can significantly improve the outcome of subsequent processing steps related to laparoscopic images, such as classification, segmentation and registration. Therefore, it can have a direct effect on the surgeon's work and the patient's safety. Many techniques can be found in the literature proposing to resolve this problem from different points of view. Applying these approaches on endoscopic images, however, does not necessarily give satisfying results because of the particular characteristics of the endoscopic scenes and that of the human tissues. Many of the proposed highlights removal techniques do not take into consideration the scene content spatio-temporal correlation and the Human Visual System (HVS) sensitivity to the specularities. In this work, we will present a novel patch-based highlight removal technique that exploits the spatio-temporal correlation of the data by using the specular free patches to correct the corrupted pixels. In this process we propose to take into account the HVS sensitivity to the luminance/color variations in the area surrounding the corrected pixels to avoid any perceptual confusion of the surgeon.

9396-7, Session 2

Aberration characteristics of conicoidal conformal optical domes

Wang Zhang, Dongsheng Wang, Shouqian Chen, Zhigang Fan, Harbin Institute of Technology (China)

Most conformal optical system designs concentrate on the researches of ellipdoidal domes. However, ellipsoidal domes are not the best ones which possess excellent aerodynamic performance among conicoidal conformal domes. This paper investigates the aberration characteristics of conicoidal conformal domes. For the purpose of the study, different types of infrared conformal domes, including ellipsoidal domes, parabolic domes and hyperbolic domes, with the common fineness ratio of 1.0 and an index of refraction of 2.25 are modeled by mathematical computation. To investigate the dynamic characteristics of the on-axis aberrations, the curves of third order fringe Zernike polynomials of the domes across field of regard are obtained by decomposing the incident wavefront at the exit pupil. The Zernike aberrations across field of view are also achieved to investigate the varying characteristics of off-axis aberrations. To find the similarities and difference between spherical domes and conformal domes at both field of regard and field of view, also make the foundation for optical design, spherical domes and defocused spherical domes are modeled for reference researches. The changing rules and the features of the aberrations among different types of conicoidal conformal domes are drawn as conclusions at the end of the paper.

9396-8, Session 2

MTF evaluation of white pixel sensors

Albrecht J. Lindner, Kalin Atanassov, Jiafu Luo, Sergio R. Goma, Qualcomm Inc. (United States)

The RGB Bayer layout is well-established for color filter arrays in color image sensors. However, there are ongoing efforts to explore alternative designs with different spectral sensitivity curves and different sizes of the unit cell [1]. One possible alternative CFA design uses white pixels, i.e. pixels that collect photons across the entire visible spectrum (WRGB sensor) [2, 3, 4]. The benefit of white pixels is that they are more sensitive and thus have better performance in low-light conditions. The drawback is that white pixels are achromatic and thus do not provide any information related to color.

We design a test target with four quadrants. Two quadrants are achromatic where we only modulate the luma channel and the other two quadrants are chromatic, where the chroma channel varies, but the luma channel is constant. From the two achromatic (chromatic) quadrants, on is a Siemens star [5] and the other is a zone plate. We arrange the four quadrants into one image and place visual marker in the corners for automatic detection. The target is then printed on non-glossy paper and mounted on a foam board.

We use two sensors that only differ in the CFA layout, one with traditional Bayer and one with a novel WRGB layout. The distance to the target is chosen so that the target is imaged at 1000 pixels width and height where – by design – the highest frequency is half the sampling frequency f_s/2. We take images with each sensor at 10 different lux levels and adjust the sensor gain in order to have the same digital count on the sensor.

We automatically analyze the target photos to estimate the sensors' MTF curves for all lux levels. The analysis shows that the added white pixels improve the resolving power for the luma channel at the expense of resolving power in the chroma channels. For the chroma channels we find that at the spatial frequency where the MTF of the Bayer sensor falls off to half of its starting value, the MTF of the WRGB sensor is already at one fourth of its starting value.

[1] Z. Sadeghipoor Kermani, Y. Lu and S. Süsstrunk, Optimum Spectral Sensitivity Functions for Single Sensor Color Imaging, Proceedings of IS&T/SPIE EI: Digital Photography VIII, 2012.

[2] Hiroto Honda, Yoshinori Iida, Yoshitaka Egawa, and Hiromichi Seki, A Color CMOS Imager With 4 ? 4 White-RGB Color Filter Array for Increased Low-Illumination Signal-to-Noise Ratio, IEEE Transactions on electron devices, vol. 56, no. 11, 2009.

[3] Alexander Getman, Jinhak Kim, Tae-Chan Kim, Imaging system having White-RGB color filter array, IEEE International Conference on Image Processing, pp. 569 - 572 , 2010.

[4] Chang-shuai Wang, Jong-wha Chong, An Improved White-RGB Color Filter Array Based CMOS Imaging System for Cell Phones in Low-Light Environments, IEICE TRANSACTIONS on Information and Systems, vol. E97-D no.5, pp.1386-1389, 2014.

[5] Christian Loebich, Dietmar Wüller, Bruno Klingen, Anke Jäger, Digital Camera Resolution Measurement Using Sinusoidal Siemens Stars, SPIE-IS&T Electronic Imaging Digital Photography III, 2007.

9396-9, Session 2

Intrinsic camera resolution measurement

Peter D. Burns, Burns Digital Imaging (United States); Judit Martinez Bauza, Qualcomm Inc. (United States)

Objective evaluation of digital image quality usually includes analysis of spatial detail in captured images. There are several well-established methods for this purpose, including those based on test objects containing elements such as edges, lines, sine-waves and random patterns. The resulting measure of the system's ability to capture image detail is often expressed as a function of spatial frequency. While the origin of such measures is in linear-system analysis, it is understood that for practical imaging systems there is no unique Modulation Transfer Function (MTF). However, useful spatial frequency response (SFR) measures have been developed as part of imaging standards. Examples of these are the edge-SFR [1], and S-SFR [2] (based on a polar sine-wave feature) in the revised ISO 12233 standard [3]. A more recently-developed measure of image-texture capture is based on pseudo-random objects (e.g. circles) in a 'dead-leaves' pattern [4].

Although the previously-developed methods have found success in the evaluation of system performance, the systems in question usually include spatial image processing (e.g. sharpening or noise-reduction), and the results are influenced these operations. Our interest, however, is in the intrinsic resolution of the system. By intrinsic resolution we mean the performance primarily defined by the lens and imager, and not influenced by subsequent image processing steps that are invertible. Examples of such operations are brightness and contrast adjustments, and simple sharpening and blurring (setting aside image clipping and quantization). While these operations clearly modify image perception, they do not in general change the fundamental image information that is present. Evaluation of this fundamental performance is important, for example, when evaluating and comparing image-capture systems which employ different image processing. For the analysis and comparison of both conventional and computational camera systems, where it is not practical to disable the image processing, we need a method based on processed images.

In this paper we describe a method to measure an intrinsic SFR computed from test image(s) for which spatial operations may have been applied. As an example, consider image capture by an ordinary camera or scanner followed by a digital image sharpening operation. Using established SFR methods and comparing results for the same camera with and without filtering would yield different results. However, since the lens and imager have not changed and the filtering is invertible, we seek a method for which the two intrinsic SFR results are equivalent. In addition, for operations for which there is loss of spatial information, the desired method would indicate this reduction. In other words, the measure would 'see through' operations for which image detail is retrievable, but measure the loss of image resolution otherwise.

For this method we adopt a two-stage image capture model. The first stage comprises a locally-stable point-spread function (lens), the integration and sampling by the detector (imager), and the introduction of detector noise. The second stage comprises the spatial image processing. Since the image processing is not restricted to linear operations, this model is used with the understanding that its parameters will approximate the corresponding measures for an ideal model system.

The method estimates the effective signal and noise spectra, and these are used along with the two-stage model, to derive a measure of the intrinsic (signal) SFR. The method requires the acquisition of several replicate images of a test target such as a dead-leaves pattern. A mean noise spectrum is derived from the set of images. The signal transfer function of the end-to-end two-stage system model is obtained using a cross-spectral estimation method, similar to the recently described by Kirk, et al. [5]. Based on the estimated output noise-spectrum we derive an effective SFR due to the spatial image processing (second stage). Since this is modelled as being cascaded with the input lens-imager SFR (first stage), the intrinsic resolution is estimated by correcting the end-to-end signal SFR by that computed for the image processing alone. The computational elements of the proposed method fall within a well-established statistical estimation and signal processing framework.

The validation of this method was done using both simulation and actual camera evaluations. The following conclusions are derived from the results of the validation process:

- Reliable measurement using small areas (locality); this is useful for systems for which characteristics vary across the image field
- Stable and consistent results for both simulated and camera testing
- Robust (invariant) results for a range of camera ISO settings (200-1600)
- Stable results for different test-target texture patterns, due in part to the use of cross-spectral signal estimation
- For cases with no post processing, the intrinsic SFR compared favorably with edge-SFR for low-noise image capture
- For several levels of invertible operations (e.g. sharpening, contrast

stretching) the desired results were achieved, i.e. equivalent intrinsic (retrievable) resolution was reported

• Desired loss of intrinsic resolution was reported for several imaging paths where expected, e.g., non-linear filtering, image resampling

• Improved stability achieved via data de-trending, and automatic registration of the image regions of interest used for both cross-spectral signal and auto-spectral estimation

As with any measurement based on an approximate model, there are limitations to the method. We describe these along with suggestions for control of measurement variability and bias, including computational details and selection of test target characteristics.

REFERENCES

[1] Burns, P. D., "Slanted-Edge MTF for Digital Camera and Scanner Analysis," Proc. PICS Conf., IS&T, 135-138 (2000).

[2] Loebich, C., Wüller, D., Klingen, B., Jäger. A., "Digital Camera Resolution Measurement Using Sinusoidal Siemens Stars," Proc. SPIE 6502, 6502N (2007).

[3] ISO 12233:2014, Photography -- Electronic still picture imaging -- Resolution and spatial frequency responses, ISO, (2014).

[4] Cao, F., Guichard, F., and Hornung, H., "Measuring texture sharpness of a digital camera," Proc. SPIE 7250, 72500H (2009).

[5] Kirk, L., Herzer, P., Artmann, U. and Kunz, D., "Description of texture loss using the dead leaves target: current issues and a new intrinsic approach," Proc. SPIE 9014, (2014).

9396-10, Session 3

Mobile phone camera benchmarking in low light environment

Veli-Tapani Peltoketo, Sofica Ltd. (Finland)

High noise values and poor signal to noise ratio are traditionally associated to the low light imaging. Still, there are several other camera quality features which may suffer low light environment. For example, what happens to color accuracy and resolution in low light imaging and how the camera speed behaves in low light? Furthermore, how low light environment should be taken account in the camera benchmarking and which metrics are the critical ones in low luminance?

Definitely high noise values may decrease color accuracy metrics since noise causes wrong colored pixels to the image. Also resolution may decay when the noise will dim the sharp edges of the image. On the other hand, texture resolution measurement will be challenging in high noise environment because the measurement algorithms may interpret noise as a high frequency texture.

The low light environment will affect especially to the focus speed of the image capturing. Obviously, exposure time will increase also. However, it is interesting to see if the image pipeline functionalities like denoising and sharpening will increase the image capture time in low light environment and how the high noise will influence to the image compression time. In this work, low light images are captured without flash to concentrate to the low light characteristics of the sensor and camera module.

To compare mobile phone cameras in different environments easily, a single benchmarking is commonly used. The score is a combination value of measured camera metrics. However, quality and speed metrics will be affected differently in different luminance environment and thus it should be considered which ones are useful in each environment.

The work contains standard based image quality measurements including noise, color accuracy and resolution measurements in three different light environments: 1000 lux, 100 lux, and 30 lux which represent overcast day, general indoor lighting, and dim indoor lighting correspondingly. Moreover, camera speed measurements are done and important image capture factors like exposure time and ISO speed are recorded in each environment. Detailed measurement results of each quality and speed category are revealed and compared between light environments. Also a suitable benchmark algorithm is evaluated and corresponding score is calculated to find an appropriate metric which characterize the camera performance in different environments.

Measured mobile phones are selected from the different prize categories and operating systems. The measurements are done using software based and automatic test system which is executed towards application programming interface of different operating system. This approach gives comparable measurement values between different operating systems and removes the influence of mobile phone specific camera applications.

The work is a continuation to the previous papers of the author which concentrated to the generic mobile phone benchmarking metrics and noise characteristics of mobile phone cameras in different light environments.

The result of this work introduces detailed image quality and camera speed measurements of mobile phone camera systems in three different light environments. The paper concludes how different light environment influences to the metrics and which metrics should be measured in low light environment. Finally, a benchmarking score is calculated using measurement data of each environment and mobile phone cameras are compared correspondingly.

9396-11, Session 3

Luminance and gamma optimization for mobile display in low ambient conditions

Seonmee Lee, Taeyong Park, Junwoo Jang, Woongjin Seo, Taeuk Kim, Jongjin Park, Moojong Lim, Jongsang Baek, LG Display (Korea, Republic of)

Since mobile display has become a high-performance device, low power consumption technologies are required to complement a problem of battery usage. Recently, eye strain and visual discomfort are emerging as a major issue because people spend many hours staring at mobile displays in their lives. The solution for these issues, an automatic brightness control is a common function on the mobile display. This function can decide the screen brightness appropriately using a light sensor detecting of the ambient illumination. As a result, the brightness control has merits in both the power saving and the visual comfort. However, this function is not optimized and do not reflect that the requirements for displays viewed in bright and dark lighting are very different [1][2][3].

This study presents effective method to provide user's visual comfort and reduce power consumption through the optimization of display luminance and gamma curve using mobile display in low ambient conditions. The study was carried out in two steps. First, perceptual experiment was conducted to obtain appropriate luminance for visual comfort. Second, gamma curve was optimized to offer power saving effect without reducing image quality in lower luminance than in appropriate luminance.

Step 1. Luminance optimization

We conducted perceptual experiment to provide user's comfortable viewing experience by optimizing the display luminance in ambient light. To identify the optimal display luminance level for visual comfort, two luminance levels were defined. One is appropriate luminance, the other is minimum luminance. Appropriate luminance stands for a comfortable level that user can recognize object clearly, while minimum luminance is the threshold level that user can recognize object.

We set illumination condition 0lux(dark room), 50lux(family living room) and 100lux(very dark overcast day) which are illumination of general low ambient light environment [4]. The experiment was performed in the constructed indoor lighting system to reproduce low ambient light condition. Ten subjects participated in the experiment, and the paired comparison method was used to investigate optimal luminance. Subjects were presented two images of different luminance through each of the 5.5inch FHD LCD. And then, they were asked to answer following questions:

Figure 1 shows both appropriate luminance and minimum luminance tend to increase with the ambient illumination increases. It means that a gap between surrounding ambient light and the display luminance affect increasing of eye discomfort in dark. We obtained appropriate luminance and minimum luminance as shown in Table 1.

Step 2. Gamma optimization

We maximized the power saving effect through optimization of gamma curve without reducing image quality in lower luminance than in appropriate luminance. We defined s-curve gamma, and we optimized parameters which can provide equal image quality in 20 percent lower luminance compared with appropriate luminance. Then, the optimized parameters were verified through perceptual experiment. The concept of proposed method is presented in Figure 2 (a).

To compensate the decreased luminance, we defined s-curve gamma expressed by the Formula (1). Din is input image data, and Dout is output luminance data for setting of gamma curve. L1 is current luminance, and L2 is decreased luminance for power saving.

S-curve gamma is divided into concave curve defining low gray level(0≤Din≤?) area and convex curve defining high gray level(?≤Din≤255) area. Nodal point(Np) connects concave curve and convex curve. As can be seen in Figure 2 (b), we can separately control the low gray level and high gray level by setting parameter ? and ?.

Np is a point divide low gray level area and high gray level area. In case of lower Np, contrast of the image would be worse because low gray level area would be brighter. Thus, we set optimal Np through analyzing histogram of dark area of images commonly used in mobile display. After setting Np, we optimized parameter ? and ?. ? is parameter that affects contrast and low gray level expression, and ? is parameter that affects brightness and high gray level expression. ? value was obtained by considering contrast and the expression of low gray level area without reducing quality. Then perceptual experiment was performed to optimize ? value. We split ? value into five levels. Ten subjects participated in the experiment, and the paired comparison method was used for optimizing ? value. Subjects were presented two different ? level images, and they were asked to evaluate images according to the questions about brightness, high level gray expression and total image quality preference. The most preferred level chosen by the subjects was set as the optimal ? value. We conducted perceptual experiment to verify performance of s-curve gamma. Twenty subjects participated in the test consist of ten display image quality experts and ten non-experts. They were presented three different conditions of images through each of the 5.5inch FHD LCD, and asked to evaluate images to the following questions by 9 scales:

Condition 1. Appropriate display luminance + 2.2 gamma

Condition 2. 20 percent lower luminance compared with appropriate luminance + 2.2 gamma

Condition 3. 20 percent lower luminance compared with appropriate luminance + s-curve gamma

Question 1. How would you rate the preference of contrast(brightness, low level gradation, high level gradation)?

Question 2. How would you rate the overall image quality?

9 : Like extremely 8 : Live very much 7 : Like moderately 6 : Like slightly 5 : Neither like nor dislike

4 : Dislike slightly 3 : Dislike moderately 2 : Dislike very much 1 : Dislike extremely

In low gray area, we designed s-curve gamma similar to 2.2 gamma. Meanwhile, in high gray area, we designed s-curve gamma higher than 2.2 gamma. Thus, 'Low level gradation' of s-curve gamma was no significant difference between all conditions and 'High level gradation' was not rated higher than others. However, 'Contrast' and 'Brightness' were especially rated higher score than others although the luminance of s-curve gamma was 20% lower than 2.2 gamma. Consequently, subjects recognized an equal overall image quality between Condition 1 and Condition 3 as shown in Figure 3 (b). We have achieved power saving effect by applying s-curve gamma without reducing image quality in lower luminance than in appropriate luminance.

References

[1] R. M. Soneira, "Brightness Gate for the iPhone & Android Smartphones and HDTVs – Why existing brightness controls and light sensors are effectively useless", Display Mate Technologies, http://www.displaymate.com/AutoBrightness_Controls_2.htm, 2010

[2] Ma, TY. Et al, "Automatic Brightness Control of the Handheld Device Display with Low Illumination", Computer Science and Automation Engineering, pp. 382-385, 2010

[3] Rafal Mantiuk et al, "Display Considerations for Night and Low-Illumination Viewing", APGV 09, pp. 53-58, 2009

[4] Yu-Ting Lin et al, "Minimum ambient illumination requirement for legible electronic-paper display", Display, Volume 32, Issue1 pp.8-16, 2011

9396-12, Session Key

Print quality and image quality: kissing cousins or feuding in-laws? *(Keynote Presentation)*

Jan P. Allebach, Purdue Univ. (United States)

As digital imaging systems become an increasingly ubiquitous part of our daily lives, and are accompanied by ever-increasing demands to improve quality, add features, reduce cost, and shorten the time-window for product development, print quality and image quality have grown in importance. They provide tools that are essential throughout the product life cycle from early development through field deployment and production usage. Some may say that print quality and image quality are the same thing. Others may claim that print quality is a special case of the more general image quality framework. In this talk, I will explore what exactly, is meant by these two terms, what they have in common, and how they differ. I will argue that in fact, print quality encompasses a broader range of issues than image quality, and that it is used in a larger part of the imaging product life cycle than is image quality. I will then discuss some of the challenges and opportunities for researchers in this field, especially as they pertain to development of algorithms to assess print quality.

9396-13, Session 4

A new method to evaluate the perceptual resolution

Miho Uno, Shinji Sasahara, Fuji Xerox Co., Ltd. (Japan)

Today, various approaches to evaluate perceived resolution of the printed matter have been used, but all of them based on visual estimations. Meanwhile, Fogra proposed a method (Fogra L-Score evaluation) that can be evaluated objectively the resolution of the printed matter made by production printers. Feature of Fogra L-Score is that the score is calculated by using number of patterns that can be resolved.

Investigation of Fogra L-Score evaluation is performed on monochrome printed matter from 3 different marking technologies (offset, liquid electrophotograpy and electrophotograpy have been selected). As in the case of verifying Fogra L-Score, the RIT Contrast-Resolution Form (consists of a two dimensional array of patterns, each of which contains concentric circles of varying spacial frequency and contrast) was selected for this study. As a result of investigating the reliability of Fogra L-Score, even in samples of the same score, it was confirmed there is a difference in appearance between samples at higher frequency. This problem occurs because Fogra L-Score is calculated from the number of resolved pattern. Therefore Fogra L-Score becomes a discrete value and doesn't have enough fineness of unit.

To resolve this problem, we need the new metric which is correlated more than L-Score to the perceived resolution. As a method of achievement, we propose a novel approach to increase resolution by using the dimensional cross-correlation coefficient which allows finding the exact overlapping position of the scanned image and the reference. The underlying concept of proposed measurement method is based on an idea from Fogra L-Score evaluation.

In this work 6 printed samples listed in Digital Print Forum 2008/2009 of IPA (IDEAlliance) issue were used as evaluation image and a high quality scanner was used. The spatial filter is applied to the scanned image and the reference test pattern. The scanned image is compared with the reference test pattern by a two dimensional cross-correlation.

The core part of a new method is the calculation of the cross-correlation coefficient between every pattern in the reference test pattern and

corresponding pattern in the scanned image. Analysis of the cross-correlation coefficient of each pattern is started on top of every column in the vertical direction and left of every row in the horizontal direction. The pattern which has the first local minimal value of the cross-correlation coefficient is considered as a resolution limit.

Therefore, all following patterns must be marked as not resolvable and these correlation coefficients are corrected to zero.

A new score is obtained by using the cross-correlation coefficients. In order to confirm the effect of this method, visual evaluation test was performed. Using 6 printed samples attached to Digital Print Forum 2008/2009, 7 panelists evaluate perceived resolution by ranking method. Correlation between resolution score and perceived resolution with subjective evaluation were improved to 0.85 in the new score from 0.73 in L-Score. Furthermore, using a monochromatic landscape image printed in Digital Print Forum 2008/2009, visual evaluation test for perceived sharpness was performed. As a result, perceived sharpness could be expressed by the contrast resolution and a new score, the correlation coefficient between the subjective evaluation results was 0.99.

9396-14, Session 4

MFP scanner motion characterization using self-printed target

Minwoong Kim, Purdue Univ. (United States); Peter Bauer, Jerry K. Wagner, Hewlett-Packard Co. (United States); Jan P. Allebach, Purdue Univ. (United States)

Multifunctional printers (MFP) are products that combine the functions of a printer, scanner, and copier. Our goal is to help customers to be able to easily diagnose scanner or print quality issues with their products by developing an automated diagnostic system embedded in the product. We specifically focus on the characterization of scanner motions, which may be defective due to irregular movements of the scan-head. The novel design of our test page and two-stage diagnostic algorithm are described in this paper. The most challenging issue is to evaluate the scanner performance properly when both printer and scanner units contribute to the motion errors. In the first stage called the uncorrected-print-error-stage, aperiodic and periodic motion behaviors are characterized in both the spatial and frequency domains. Since it is not clear how much of the error is contributed by each unit, the scanned input is statistically analyzed in the second stage called the corrected-print-error-stage. Finally, the described diagnostic algorithms output the estimated scan error and print error separately as RMS values of the displacement of the scan and print lines, respectively, from their nominal positions in the scanner or printer motion direction. We validate our test page design and approaches by ground truth obtained from a high-precision, chrome-on-glass reticle manufactured using semiconductor chip fabrication technologies.

When scanning a document in flatbed mode, motion errors can result from gear backlash, runout, chain or belt stretch, and wobble of the scan-bar, as it is pulled along under the flatbed glass. When scanning in sheet-feed or automatic document feeder (ADF) mode, the same error sources are operative, except the last two listed above.

Our goal is to quantify the quality of scanner motion achieved when scanning in either of these two modes. Since it is intended that the diagnostic procedure be conducted by the customer, with the unit at the customer's facility, it is desirable that the customer be able to print any diagnostic

pages that are needed using the printer in the MFP unit itself, rather than relying on him or her having access to an expensive, precision target.

We have developed a special test page consisting of columns of scan-line marks (SLMs). By calculating the centroids of these SLMs in the scanner motion direction, we can estimate the position of each scan line in this direction. The columns of SLMs are displaced, and offset in the scanner motion direction, from column-to-column, to allow measurement of the displacement of each succeeding scan-line, even in the presence of printer dot gain that would cause SLMs that are adjacent in the direction of the scanner motion to fuse together, and prevent estimation of the centroid of each individual SLM. The columns of SLMs are arranged in three separate groups at the left side, center, and right side of the test page, in order to provide estimates of motion quality at three different horizontal positions across the scanned region. (For purposes of our discussion here, we assume that the scanner motion is in the vertical direction of the page being scanned.)

Unfortunately, the positions of the SLMs on the printed test page may also contain errors due imperfections in the motion provided by the print engine. These errors may be due to most of the same sources mentioned above, as well as errors in the fabrication and motion of the polygon mirror that is used to scan the laser beam across the organic photoconductor drum surface. In order to separate the errors in the positions of the SLMs as they are placed on the test page from the errors in the positions of the SLMs, as they are captured during the scanning process, we scan the printed test page twice – once from top to bottom, and once from bottom to top, and introduce a novel statistical analysis that cancels the RMS error in the SLM positions due to the print engine motion. This then allows us to separately estimate the RMS error in the SLM positions due to the scanner and the printer.

In addition to the high-frequency errors caused by all the sources discussed above, there are a number of possible sources of low-frequency errors during both the printing of the test page and subsequent scanning of it. These include paper-feed skew during printing, stretch of the paper after printing, skew during placement of the page for flat-bed scanning, and skew during scanning with the ADF unit. We print registration marks along the left and right margins of the page, and perform a homographic correction procedure to correct for these low-frequency error sources.

9396-15, Session 4

Autonomous detection of ISO fade point with color laser printers

Ni Yan, Purdue Univ. (United States); Eric Maggard, Roberta Fothergill, Renee J. Jessome, Hewlett-Packard Co. (United States); Jan P. Allebach, Purdue Univ. (United States)

Image quality is an area of growing importance as imaging systems become ever more pervasive in our daily lives. In this paper, we consider the problem of autonomously assessing the presence of fading in prints from home or office color printers. An accurate understanding of the capacity of a printer cartridge to generate prints without observable fading is very important to the manufacturers of the cartridges, as well as the manufacturers of the printers that use those cartridges.

1. PSYCHOPHYSICAL EXPERIMENT

The standard ISO/IEC 19798:2007(E) defines fading as a noticeably lighter area, 3 mm or greater in length, located anywhere in one or more of the bars along the right and bottom margin of the diagnostic page of a 5-page standard test suite that consists of 4 typical document pages followed by the special diagnostic page. Following the recommendations of ISO/IEC 19798:2007(E), we conducted a psychophysical experiment that gives us a sequence of training samples that comprises a subset of printed test pages and an ISO fade page. All samples before the ISO fade page in the sequence are presumed not to be faded; and the rest are presumed to be faded. This information comprises the ground truth for training our fade detection algorithm. In addition, we use the first sample page in the sequence as a master image.

2. FADE DETECTION ALGORITHM

For each color, we analyze the two saturated bars since imaging scientists/engineers only look at these regions in the psychophysical experiment. Each bar is divided into overlapping segments with length equal to 1/10 of the length of the bar and width equal to that of the bar. To eliminate boundary effects, the segments overlap by 1/2 of their length. Each segment is processed individually, we simply refer to a given segment as the region of interest (ROI).

For a given bar, two extreme color points PW and PS are extracted from the

master image, between which we assume all the color points of testing bar are distributed. We use the white point of the master image as point P W. It is calculated as the average CIE L*a*b* value of the white margin. We use the most saturated pixel of the entire color bar as point PS. The saturation of a given pixel is approximated by

S=sqrt(((a*)^2+(b*)^2))/L*

For a given ROI, we project the scattered color points onto the line PSPW. Since the extreme point PW is the white point and the extreme point PS is the dark point, color points that are closer to point PW are lighter than those that are closer to point PS.

From the master image, we calculate the overall average L*a*b* value of the bar and project it onto PSPW. Also, from the test image, we project the overall average L*a*b* value of the bar onto PSPW. We refer to these two points as the projected average master L*a*b* (L*a*b*PAM) and the projected average test L*a*b* (L*a*b*PAT). They are used in the iteration and machine learning procedure.

For each segment in a bar, we first apply the K-means algorithm to segment this ROI into 2 clusters – the faded (F) and the non-faded (N) regions. This provides our initial Fade Map that is intended to show faded regions. We then carry out the procedure to correct the Fade Map to better align it with the judgments of the expert observers. We start by comparing the average lightness L?F of the ROI with the threshold L?PAT + ?1. If L?F is less than this threshold, then the faded part of the ROI is too large, and contains pixels that are too dark. We reduce the size of the faded region by applying the K-means algorithm just to the faded region alone. Then, we merge the newly declared non-faded pixels with the previous non-faded region. This reduces the size of the faded region by removing the darker pixels that it previously contained, and thereby increases L?F. If L?F is still less than L?PAT + ?1, we repeat this procedure. Otherwise, we terminate the iteration, and output the final corrected Fade Map.

On the other hand, if for the initial Fade Map L?F > L?PAT + ?1, we check to see if L?F is greater than a second threshold L?PAT + ?2. If it is, then follow a procedure that is analogous to that described above, which is used to reduce the size of the faded region. The values for ?1 and ?2 are chosen empirically. For each bar at each pixel location, we compute the final Fade Map as the logical AND between the two Fade Maps for the segments that overlap at that location.

3. MACHINE LEARNING TOOL

We use a supervised learning model with a Support Vector Machine (SVM) as our machine learning tool. We train an SVM classifier with the training samples and test it on the printed diagnostic pages that were not chosen during the psychophysical experiment. Feeding the data into the machine learning trainer, we get an SVM classifier that is to be used as the prediction tool.

We compute 4 features: (1) fade percentage – the number of detected faded pixels over the total number of pixels ? 100; (2) ?E1: ?E between L*a*b*PAT and the faded pixel cluster centroid CF; (3) ?E2: ?E between L*a*b*PAT and the non-faded pixel cluster centroid CN; (4) ?E3: ?E between L*a*b*PAM and the bar average L*a*b*. Features (2)-(4) are calculated as averages over all segments in that bar. For testing, we apply the SVM classifier to each test sample to see whether that page is faded or not. A misclassification can be either a non-faded page is classified as faded or a faded page is classified as non-faded. The error rate is 7.14% for testing cyan bar, 9.68% for black bar, and 0 for blue bar, magenta bar and red bar.

9396-16, Session 5

Autonomous detection of text fade point with color laser printers

Yanling Ju, Purdue Univ. (United States); Eric Maggard, Renee J. Jessome, Hewlett-Packard Co. (United States); Jan P. Allebach, Purdue Univ. (United States)

Nowadays, image quality is an area of growing importance as imaging systems become more and more pervasive in our daily lives. Most of the image-analysis-based quality measures depend on the judgments of expert observers, based on mapping functions and/or machine-learning frameworks, according to all aspects of the image or some specific image attributes or defects [1–4].

Laser electro-photographic printers are complex systems which can produce many possible artifacts that are very different in terms of their appearance in the printed output. Fading due to depletion of toner is one of the issues of most critical concern for print quality degradation with color laser electro-photographic printers [5–7]. As the toner in a given cartridge becomes depleted, the customer may start to notice fading in certain areas of a printed page. An accurate understanding of the capacity of a printer cartridge to generate prints without observable fading is very important to the manufacturers of the cartridges, as well as the customer who intends to utilize every cartridge to the fullest extent possible, while providing good quality printed output. With the growing popularity of managed print services, accurate prediction of the capacity of a printer cartridge takes on an entirely new importance for the vendor. Thus, there is a great need for a means to accurately predict the fade point while printing pages with typical content.

ISO/IEC 19798:2007(E) specifies a process for determining the cartridge page yield for a given color electro-photographic printer model. Starting with a new cartridge, a suite of four typical office document pages and a diagnostic page consisting of solid color bars, is printed repeatedly until the cartridge is depleted, followed by visual examination of the sequence of printed diagnostic pages to determine where in the sequence fading first occurred, and set it as the printing fade point. In this test, fade is defined as a noticeably lighter area, 3 mm or greater located in the bars around the diagnostic page of the test suite. But this method is a very costly process since it involves visual examination of a large number of pages. And also the final decision is based on the visual examination of a specially designed diagnostic page, which is different than typical office document pages.

In this paper, we propose a new method to autonomously detect the text fading in prints from home or office color printers using a typical office document page instead of a specially designed diagnostic page. This task is particularly challenging because the fading behavior is location dependent: it varies from character to character. We hypothesize that the expert observer determines the fade point depending on the local character contrast in the faded test patches, as well as the global strength of the color characters relative to a reference page. In our method we scan and analyze the printed pages to predict where expert observers would judge fading to have occurred in the print sequence. Our approach based on a machine-learning framework in which features derived from image analysis are mapped to a fade point prediction. We train our predictor using a sequence of scanned printed text pages, and analyze them character-by-character to extract a set of novel features that are indicative of fade. Calculation of the features involves the following steps: color space conversion, binary threshold, morphological operation, connected components, common mask alignment, classification, and a machine-learning algorithm to develop the predictor for the text fade point. In this paper, we only present results for laser electro-photographic printers that use dry toner; but our approach should be applicable to other print technologies as well.

References

[1] X. Jing, S. Astling, R. Jessome, E. Maggard, T. Nelson, M. Q. Shaw, and J. P. Allebach, “A General Approach for Assessment of Print Quality,” Image Quality and System Performance X, SPIE Vol. 8653, P. D. Burns and S. Triantaphillidou, Eds. San Francisco, CA, 3-7 February 2013.

[2] J. Zhang, S. Astling, R. Jessome, E. Maggard, T. Nelson, M. Q. Shaw, and J. P. Allebach, “Assessment of Presence of Isolated Periodic and Aperiodic Bands in Laser Electrophotographic Printer Output,” Image Quality and System Performance X, SPIE Vol. 8653, P. D. Burns and S. Triantaphillidou, Eds. San Francisco, CA, 3-7 February 2013.

[3] M. Q. Nguyen, S. Astling, R. Jessome, E. Maggard, T. Nelson, M. Q. Shaw, and J. P. Allebach, “Perceptual Metrics and Visualization Tools for Evaluation of Page Uniformity,” Image Quality and System Performance XI, SPIE Vol. 9016, S. Triantaphillidou and M.-C. Larabi, Eds. San Francisco, CA, 3-5 February 2014.

[4] S. Hu, Z. Pizlo, and J. P. Allebach, “JPEG Ringing Artifact Visibility Evaluation,” Image Quality and System Performance XI, SPIE Vol. 9016, S. Triantaphillidou and M.-C. Larabi, Eds. San Francisco, CA, 3-5 February 2014.

[5] S. Kiatkamjornwong, K. Rattanakasamsuk, and H. Noguchi, “Evaluation of

some strategies to control fading of prints from dye-based ink jet," Journal of Imaging Science and Technology, vol. 47, no. 2, pp. 149–154, 2003.

[6] Y. Ye, "Study on the fading of color inkjet printing paper," Chung-kuo Tsao Chih/ China Pulp and Paper, vol. 23, no. 12, pp. 14 – 17, 2004.

[7] E. Hoarau and I. Tastl, "A novel sensor to visually track fading of printed material," in NIP & Digital Fabrication Conference. Society for Imaging Science and Technology, vol. 2010, pp. 423–425, 2010.

9396-17, Session 5

Photoconductor surface modeling for defect compensation based on printed images

Ahmed H. Eid, Brian E. Cooper, Lexmark International, Inc. (United States)

Manufacturing imperfections of photoconductor (PC) drums in electrophotographic printers cause low-frequency artifacts that can produce objectionable non-uniformities in the final printouts. These non-uniformities are most visible in low-density, large (constant) area, high-frequency halftoned regions. Artifacts result from variation in the thickness of the charge layer of the PC. This variation may occur during fabrication as well as during use (i.e., from wear). Environmental conditions can also influence the density variation. During printing, successive revolutions of the cylindrical PC drum will form a repeating pattern of over- and under-developed print areas down the page (i.e., in the process direction).

In this paper, we propose a technique to analyze the PC defects using the 2D details of the input images, assuming prior knowledge of the PC's circumference. We combine together and filter multiple complete cycles of the PC drum from the input image, using wavelet filtering. This pre-processing step removes the perpendicular defects from the image and eliminates the high frequency defects parallel to the PC non-uniformity. Using the expectation maximization (EM) algorithm for density estimation, we model the non-uniformities caused by the PC drum as a mixture of Gaussians. The model parameters then provide a measure to quantify the PC defects.

In addition, a 2D polynomial fitting approach characterizes the spatial artifacts of the drum, by analyzing multiple revolutions of the printed output from the previous wavelet filtering step. The approximation component (mainly the zero frequency component) of the filtered image is removed, so the output image forms three groups of values: positive values representing the over-developed areas, negative values representing under-developed areas, and zero values that represent the properly developed areas. A 2D polynomial fitting approach models this image compactly, reducing memory storage within the device. This model becomes the defect-compensating profile of the defective PC drum.

To compensate for PC drum artifacts, the printer must align the position of the defect-compensating profile with a known position on the PC drum. Then the printer can compensate for the over- and under-developed areas that correspond to the defective regions on the drum's surface.

Our preliminary results show a high correlation of the modeled artifacts from different revolutions of a drum using both the EM mixture of Gaussians and the polynomial fitting models. In addition, we apply the objective quality metric, derived from the mixture of Gaussians parameters, to print samples from a set of ten different printers. Our experiments show high correlation between the proposed metric and subjective assessment of print quality experts.

9396-18, Session 5

Controlling misses and false alarms in a machine learning framework for predicting uniformity of printed pages

Minh Q. Nguyen, Jan P. Allebach, Purdue Univ. (United States)

In our previous work* , we presented a block-based technique to analyze printed page uniformity both visually and metrically. The features learned from the models were then employed in a Support Vector Machine (SVM) framework to classify the pages into one of the two categories of acceptable and unacceptable quality.

In this paper, we introduce a new set of tools for machine learning in the assessment of printed page uniformity. This work is primarily targeted to the printing industry, specifically the ubiquitous laser, electrophotographic printer. We use features that are well-correlated with the rankings of expert observers to develop a novel machine learning framework that allows one to achieve the minimum "false alarm" rate, subject to a chosen "miss" rate. Surprisingly, most of the research that has been conducted on machine learning does not consider this framework.

During the process of developing a new product, test engineers will print hundreds of test pages, which can be scanned and then analyzed by an autonomous algorithm. Among these pages, most may be of acceptable quality. The objective is to find the ones that are not. These will provide critically important information to systems designers, regarding issues that need to be addressed in improving the printer design. A "miss" is defined to be a page that is not of acceptable quality to an expert observer that the prediction algorithm declares to be a "pass". Misses are a serious problem, since they represent problems that will not be seen by the systems designers. On the other hand, "false alarms" correspond to pages that an expert observer would declare to be of acceptable quality, but which are flagged by the prediction algorithm as "fails". In a typical printer testing and development scenario, such pages would be examined by an expert, and found to be of acceptable quality after all. "False alarm" pages result in extra pages to be examined by expert observers, which increases labor cost. But "false alarms" are not nearly as catastrophic as "misses", which represent potentially serious problems that are never seen by the systems developers. This scenario motivates us to develop a machine learning framework that will achieve the minimum "false alarm" rate subject to a specified "miss" rate. In order to construct such a set of receiver operating characteristic (ROC) curves, we develop various tools for the prediction, ranging from an exhaustive search over the space of the linear and nonlinear discriminants to a modified SVM framework. We then compare the curves gained from those methods. Our results show a significant improvement in miss and false alarm rates, compared to the usual SVM, and the promise for applying a standard framework to obtain a full ROC curve when it comes to tackling other machine learning problems in industry.

*:Minh Q. Nguyen, Renee Jessome, Stephen Astling, Eric Maggard, Terry Nelson, Mark Shaw, Jan P. Allebach, "Perceptual metrics and visualization tools for evaluation of page uniformity", in Image Quality and System Performance XI, Sophie Triantaphillidou; Mohamed-Chaker Larabi, Editors, Proceedings of SPIE Vol. 9016 (SPIE, Bellingham, WA 2014), 901608.

9396-19, Session 5

Estimation of repetitive interval of periodic bands in laser electrophotographic printer output

Jia Zhang, Jan P. Allebach, Purdue Univ. (United States)

In the printing industry, electrophotography (EP) is a commonly used technology in laser printers and copiers. In the EP printing process, there are many rotating components involved in the six major steps: charging, exposure, development, transfer, fusing, and cleaning. If there is any irregularity in one of the rotating components, repetitive defects, such as

isolated bands or spots, will occur on the output of the printer or copier. To troubleshoot these kind of repetitive defect issues, the repeating interval of these isolated bands or spots is an important clue to locate the irregular rotating component. In this paper, we will describe an algorithm to estimate the repetitive interval of periodic bands, when the data is corrupted by the presence of aperiodic bands, missing periodic bands, and noise.

Here are the steps illustrating how our algorithm works. In the first step, we use a fixed threshold to obtain the banding profile from the scanned print-out using the processing pipeline described in [1]. This banding profile is a 1-D signal containing the information about the bands, such as strength, lighter or darker than the background, and their postions along the process direction of the paper. In the second step, we choose a fixed target number of periodic bands, and then from the set of all candidate bands, choose that number of bands and assign them to be periodic bands, and with the remaining bands assigned to be aperiodic bands. We build a cost funtion, with the estimated periodic interval as our unknown variable, which calculates the normalized mean squared error (MSE) between the estimated periodic bands positions and our data. We show that it is possible to obtain a closed-form solution for the minimum MSE estimate of the periodic interval, given a fixed membership assignment. For this assignment, we calculate our cost function and the estimate of the periodic interval. Then, we repeat this for all possible periodic band membership assignments. We choose the membership assignment and resulting estimated periodic interval which achieves the minimum error among all possible periodic band membership assignments. This result will be kept for this fixed number of target periodic bands.

In the third step, we repeat the second step for all possible target numbers of periodic bands, ranging from three to the total number of candidate bands that we identified in the first step. For each target number of periodic bands, we obtain an estimated interval of periodic bands and a minimum value of the cost function which is achieved with this target number of periodic bands. These results constitute one column of our results table, which will be completed in the following step. In the fourth step, we repeat the above three steps for a set of decreasing values of the threshold, which is used in the first step to determine whether or not a band is present. For each threshold value, we obtain a separate column. These columns form our results table, which is used in the following step.

Finally in the fifth step, we observe certain trends across the data in the results table, and by fitting the data to low-order mathematical functions, we identify the break-points that correspond to the most reliable estimate of (a) the number of periodic bands that are present, (b) the set of these periodic bands from among all the candidate bands, and (c) the best estimate of the interval between the bands in the chosen set of periodic bands. After all these five steps, we will also check whether missing periodic bands are present, and update our results to improve accuracy.

To illustrate the effectiveness and robustness of our algorithm, we will provide example processing results of actual print-outs with isolated periodic bands, with aperiodic bands present and some periodic band missing.

[1] Zhang, J., Astling, S., Jessome, R., Maggard, E., Nelson, T., Shaw, M., and Allebach, J. P., “Assessment of presence of isolated periodic and aperiodic bands in laser electrophotographic printer output”, Image Quality and System Performance X, 86530N, February 2013.

9396-20, Session 6

Image quality optimization via application of contextual contrast sensitivity and discrimination functions

Edward W. S. Fry, Sophie Triantaphillidou, John Jarvis, Gaurav Gupta, Univ. of Westminster (United Kingdom)

Digital images are produced with the intention of being viewed by human observers. Therefore, to achieve optimised image quality, modelling and incorporating observer preference characteristics is essential in imaging system engineering. Providing a ‘one size fits all’ algorithm for the optimisation of image quality is particularly challenging, due to the observer quality estimation process being deeply complex and subjective, involving both low-level cortical processing and high-level interpretative processes of perception (1). Consequently, observer judgements naturally vary from image to image, affected by: image content, location of image artefacts, past observer experiences and the intended image application (2). This suggests, that if image quality enhancement algorithms are to be universally effective, they should be adaptive, at least taking into account the suprathreshold sensitivity variations of the HVS (human visual system), in reaction to different image stimuli.

Previous image quality optimisation research, has applied the CSF (contrast sensitivity function) as a weighting-function, to the luminance channel of ideally-filtered spatial frequency image octaves. This resulted in optimal sharpness, which is known to correlate with observed image quality (3,4). More recently, state-of-the-art CSFs and discrimination functions, have been directly measured from complex pictorial scenes, defining the HVS’s ability to detect and discriminate information in real images, within the context of complex masking information (5). These ‘contextual’ CSFs (cCSFs) and ‘contextual’ discrimination functions (cVPFs) should be more directly applicable to image quality optimisation, since observer quality estimation involves both the detection and discrimination of similarly masked signals. Nevertheless, it should not be directly assumed, that either cCSFs or cVPFs provide perfect band-weightings for image quality optimisation. In this paper, the hypothesis is that these newly measured functions, form more suitable weighting-functions than a common CSF.

Enhancement of microcontrast, by boosting contrast in the highest visible frequencies, is known to increase sharpness, clarity and three-dimensionality (4), which are central to image quality optimisation. Progressive boosting of mid to high frequency contrast, may also increase image quality, by compensating for the characteristic smooth decay in modulation with increased frequency, recognisable in the MTFs (modulation transfer functions) of common imaging systems. Therefore, this paper suggests that smooth amplification of contrast over higher visible frequencies, may optimise image quality further than a sudden amplification of the very highest visible frequencies. Interestingly, cCSFs and cVPFs follow this profile, naturally peaking at the higher visible frequencies, indicating that cortical processing prioritises these frequencies, when detecting/discriminating contextual information.

This paper also proposes, that all images of a recognisable nature, and perceived to be of high quality, should show some degree of naturalness. Image naturalness is assessed according to the observer’s own expectations, with respect to an internal reference or memory representation of the image, and with consideration to the perceived environmental conditions, under which the scene was captured (6). Naturalness should not be confused with fidelity of representation, since it has been proven, in an experiment involving adjustment of the chroma of images, that perceptually natural images are generally not exact reproductions of scenes (6). Results from this paper, confirm the above findings of Federovskaya et al. also apply to contrast adjustment within the frequency domain.

The experimental method for this paper, involves a number of controlled mutations being applied to each weighting-function once normalised, including the common CSF, cCSF, cVPF and a flat function. This process generates mutated curve shapes, to be applied as band-weightings, to ten single-octave log-cosine filtered image bands. Systematic mutation of the functions, clarifies whether the original functions can be optimised further, when applied as contrast weighting-functions, and provides alternative curve variations for investigation. Mutations are engineered to contain identical logarithmic areas as their corresponding original functions, thus preserving their spectral power. The image quality of all weighted images, is then assessed in a two-alternative forced choice psychophysical test, alongside images sharpened in Adobe Photoshop. Similar tests are undertaken with respect to sharpness and naturalness.

Results show that non-mutated cCSFs and cVPFs, are clearly outperforming non-mutated common CSF and flat functions, with greater consistency across all test images. The highest quality results from weighting-functions based upon cCSFs and cVPFs, that have been mutated to boost contrast smoothly, peaking in the higher visible frequencies. These band-weightings result in increased sharpness and quality, beyond the capabilities of Adobe Photoshop’s ‘Sharpen’ and ‘Sharpen-More’ filters, with greater naturalness and consistency across all test images. They also achieve peak image quality, with higher overall weighting than less successful mutations, suggesting that

if optimised further, they may be applied with increased intensity, resulting in further enhancement. Analysis of scatter plots investigating correlations between quality, sharpness and naturalness, suggests a strong correlation between quality and sharpness, and that both sharpness and naturalness are required to be preserved to achieve optimal quality.

Further investigation into quality optimisation, via the adjustment of contrast in the frequency domain, could form the basis for adaptive image quality optimisation algorithms. Research would benefit from statistical analysis of the test images, including analysis of their contrast spectra and phase information, as well as further examination of the relationships between observed quality, sharpness and naturalness. If the above relationships were understood, algorithms could potentially be engineered to account for observer preference, scene content, required output situation and capture system characteristics.

References:

1. Radun J. et al. Content and Quality: Interpretation-Based Estimation of Image Quality. ACM Transactions on Applied Perception (TAP). 2008 Jan; 4(4).

2. Ahmuda A, Null CH. Image Quality: A Multidimensional Problem. In Watson AB, editor. Digital Images and Human Vision. Cambridge: MIT Press; 1993. p. 141-148.

3. Bouzit S, Macdonald L. Does Sharpness Affect the Reproduction of Colour Images. Proc. SPIE 4421, 9th Congress of the International Colour Association, 902. 2002.

4. Bouzit S. Sharpness of Displayed Images. PhD thesis. University of Derby; 2004.

5. Triantaphillidou S, Jarvis J, Gupta G. Contrast Sensitivity and Discrimination of Complex Scenes. Proc. SPIE 8653, Image Quality & System Performance X, 86530C. 2013.

6. Federovskaya E, de Ridder H, Blommaert F. Chroma Variations, and Perceived Quality of Color Images of Natural Scenes. Color Research and Application. 1998; 22(2).

9396-21, Session 6

A study of slanted-edge MTF stability and repeatability

Jackson K. M. Roland, Imatest LLC (United States)

The slanted-edge method of measuring the modulation transfer function (MTF) has become a well known and widely used image quality testing method over the last 10 years. This method has been adopted by multiple international standards over the years including standards from ISO and IEEE. Nearly every commercially available image quality testing software includes the slanted-edge method and there are numerous open-source algorithms available. This method is, without question, one of the most important image quality algorithms in use today. The algorithm itself has remained relatively unchanged since it's original publication in ISO 12233:2000. Despite that stability in the latest 2014 revision of the ISO 12233 standard there was a major modification to the defined target. In the original 2000 edition of ISO 12233 the target was required to have a minimum edge contrast of 40:1. In the current revision the edge contrast is specified at 4:1. This change reflects a change in understanding about the slant edge measurement, with high contrast the measurement becomes unstable and so the contrast was lowered. This raises a question, how stable is the slanted-edge method? And under what conditions will it be most stable? There are a wide variety of test conditions are are know to impact the results of this method. Everything from system noise to low-light conditions and target contrast to target material can affect the results from this method.

The first part of this paper explores these test conditions and the impacts they have on the stability and precision of the slanted-edge method. The results of this experimentation are used to define ideal test conditions for the slanted-edge method. Another set of factors to consider are the inputs to the algorithm itself. The algorithm requires the use of OECF to linearize the input data but this is not always practical or possible. An alternative way to accomplish the same goal would be to assume an estimated gamma to apply. The significance of this estimation is studied and the efficacy of this is tested. Finally there is the algorithm itself to consider. The original method as described by ISO 12233 uses a first order fit to determine the line location and does not attempt to compensate for the noise present in the non-edge components of the tested region. Other algorithms created since the original have used second order fits and applied noise reduction to non-edge components of the region. These variations are tested to determine the affect on the overall stability and precision of the slanted- edge method. The results of these experiments are used to create a profile for the set of conditions and methods that provide the most stable and repeatable test possible. Future work that is being considered includes testing other widely used and adopted methods of measuring MTF to determine their own stability as well as how they compare to the slanted-edge method.

9396-22, Session 6

Comparative performance between human and automated face recognition systems, using CCTV imagery, different compression levels, and scene parameters

Anastasia Tsifouti, Home Office (United Kingdom) and Univ. of Westminster (United Kingdom); Sophie Triantaphillidou, Univ. of Westminster (United Kingdom); Mohamed-Chaker Larabi, Univ. de Poitiers (France); Efthimia Bilissi, Alexandra Psarrou, Univ. of Westminster (United Kingdom)

The police use both human (i.e. visual examinations) and automated (i.e. face recognition systems) identification systems for the completion of face identification tasks. Uncontrollable environmental conditions, such as variable lighting encountered by CCTV camera systems, create challenges on the usefulness of the reproduced imagery. The usefulness of the imagery is further compromised by compression, implemented to satisfy limited storage capacity, or transmission bandwidth.

Image usefulness is associated with image quality and relates to the suitability of the imagery to satisfy a specific task [1]. In this context, the specific identification task requires enough useful facial information to remain in the compressed image in order to allow human and automated identification systems to identify a person.

The aim of this investigation is to identify relationships between human and automated face identification systems with respect to compression. Further, to identify the most influential scene attributes on the performance of each identification system. The work includes testing of the systems with compressed footage consisting of quantified scene (footage) attributes. These include measures of camera to subject distance, angle of the face to camera plane, scene brightness, and spatio-temporal busyness. These attributes have been previously shown [2] to affect the human visibility of useful facial information, but no much work has been carried out to assess the influence they have on automated recognition systems.

Results and test material (i.e. 25 scenes, H.264/MPEG-4 AVC compression method, and bitrates - ranging between 300kbps to 2000kbps) previously obtained during an investigation with human observers [2] are also used here. In the investigation with humans, experienced civilian analysts and police staff, were polled to give their opinion on what they considered to be acceptable reduction of information from an uncompressed reference source for maintaining facial information for identification. The results were analyzed by fitting psychometric curves against the different levels of compression.

In the present investigation three basic automated systems [3] are assessed: Principal Component Analysis – PCA, Linear Discriminant Analysis - LDA, and Kernel Fisher Analysis – KFA. Their assessment is based on analysis of similarity scores. Similarity scores provide a distance measure (e.g. range between 0 – not match, and 1 – perfect match) of facial information between two images of faces, or biometric signatures [4,5]. There is a

variety of testing procedures to evaluate automated systems [6,7]. In this investigation, the automated systems were trained with a dataset of uncompressed images of faces (i.e. the enrolls or known faces). The gallery dataset included both compressed and uncompressed images (i.e. the probes, or unknown faces). Every single image in the gallery dataset was compared against every single image in the trained dataset and the produced similarity scores were used for further analysis.

Similarly to the human investigation, the automated systems were assessed based on the similarity score distance, between a compressed image from its uncompressed version. Non-linear modeling was used to analyze the results. The response of the automated systems to each different level of compression was different. PCA was not affected much by compression. LDA and KFA were affected by compression and they produced similar results.

Results between human and automated identification systems were compared and relevant correlations were drawn between the performance of each system and the selected scene attributes. Results show that the automated identification systems are more tolerant to compression than humans. For automated systems, mixed brightness scenes were the most affected (i.e. produced the lowest scores) and low brightness scenes were the least affected (i.e. produced the higher scores) by compression. In contrast for humans, low brightness scenes were the most affected and medium and mixed brightness scenes the least affected by compression. Adler and Dembinsky [8] have also shown that low brightness scenes have scored the lowest by humans, but not by automated algorithms.

The results of the present work have the potential to broaden the methods used for testing imaging systems for security applications. For example, in this investigation security systems are tested using controlled footage in terms of conveyed information, which allows a better understanding on how the systems perform. Also, often the term image quality for security applications is used similarly for both humans and automated systems [9] whilst this investigation proves that different scene attributes influence identification systems differently.

[1] Yendrikhovskij. S. N., “Image quality and colour characterization,” in: McDonald, L and Ronnier Luo, M., [Colour image science - Exploiting digital media], John Wiley and Sons, pp. 393- 420 (2002).

[2] Tsifouti, A., Triantaphillidou, S. , Bilissi, E., and Larabi, M.-C. ,“Acceptable bit rates for human face identification from CCTV imagery,” in Proc. SPIE 8653, Image quality and system performance X. San Francisco, USA. 865305 (2013).

[3] Struc, V., “The PhD face recognition toolbox,” University of Ljubljana, Faculty of Electrotechnical Engineering, Slovenia, (2012).

[4] Hu, Y., Wang, Z., “A Similarity Measure Based on Hausdorff Distance for Human Face Recognition”, in Proceedings of the 18th International Conference on Pattern Recognition, Volume 03, IEEE Computer Society, (2006).

[5] Mane, V.A., Manza, R.R., kale, K.V., “The role of similarity measures in face recognition”, International Journal of Computer Science and Application, ISSN 0974-0767, (2010).

[6] Phillips, P. J., Moon, H., Rizvi, S. A., Rauss, P. J., “The FERET Evaluation Methodology for Face Recognition Algorithms,”IEEE Trans. on pattern analysis and machine intelligence. 22(10), (2000).

[7] Delac, K., Grgic, S., Grgic, M.,:Image compression in face recognition – a Literature Survey,”in: Delac, K., Grgic, M., Bartlett, M. S., [Recent Advances in face recognition], IN-the, Croatia, pp. 1-14, (2008).

[8] A. Adler, T. Dembinsky, “Human vs automated measurements of biometric sample quality”, Canadian Conference on Electrical and Computer Engineering, pp. 2090-2093, (2006).

[9] Information technology-Biometric sample quality, BSi, PD ISO/IEC TR 29749 – 5:2010, 2010

9396-23, Session 6

A study of image exposure for the stereoscopic visualization of sparkling materials

Victor J. Medina, Mines ParisTech (France) and Peugeot Citroën Automobiles S.A. (France); Alexis Paljic, Mines ParisTech (France); Dominique Lafon-Pham, Mines Alès (France)

1- INTRODUCTION

The context of this work is the use of predictive rendering to obtain a perceptually correct computer-generated (CG) representation of sparkling paint materials using physically based rendering methods [1] and stereoscopic visualization [2]. We concentrate on the perceptual validation of the images generated by the rendering engine in comparison with physical references.

When we look at our reference plates, our perception of their texture depends largely on the flake density, the amount of flake layers, and the strength of the sparkles [3]. What we perceive is the result, as processed by our visual system, of the radiation emitted by the scene. In the case of CG images, the human visual system perceives something that has been preprocessed by the rendering engine and the visualization devices.

The imaging process converts a radiance signal –absolute scale– into an image – classic scale, from 0 to 255– via a series of optical transformations. A very important parameter for this conversion is the exposure, which measures how much of the energy radiated by each observed point is captured by the sensor during the imaging process.

In order to study the relationship between the overall image exposure and the perception of flake density, depth, and sparkling, we want to present observers with a series of stereoscopic photographs of the same plate, taken with a colorimetrically calibrated camera at varying exposure times, and ask them to select the photograph that they think is closest to the physical reference. This way, we expect to find a relationship between image exposure and the perceived aspect of the materials.

2- MATERIAL PROPERTIES:

The materials that we are working with are car paint coatings. They are formed by a layer of pigment containing micrometric-scale metallic flakes with specific optical properties, covered by a layer of transparent resin. The metallic flakes are deposited in the pigment following a pseudo-random distribution at different depths and orientations, which creates a sparkling and glittering effect when they interact with the light.

The visualization of materials such as ours is a complex process where we must take into account many factors, such as the effect of the light sources, material shape or flake composition. For example, since flakes can be found at different depths and orientations, it is possible that the sparkle created by large flakes makes them appear much bigger than they are [4, 5]. Similarly, shiny flakes located in deeper layers might appear the same size as duller ones that are closer to the surface. Although the paints can be applied over a number of different shapes, for now we are only focusing on the visualization of flat plates to reduce the number of factors that can influence the results.

3- EXPOSURE AND PERCEPTION:

As mentioned earlier, the imaging process –i.e., the conversion of the reflected light arriving onto the sensor into an image– entails the conversion of the light reflected by the objects in the scene –in our case, a paint plate– via an optical transformation. This conversion involves moving from a physical magnitude (radiance) to a classic notation between 0 and 255. Exposure is key in this transformation as it controls the correspondence between physical values and gray levels, which becomes especially important for highly illuminated points prone to over saturate the acquisition system (overexposure).

The amount of light illuminating the scene (image illuminance) is directly dependent on three factors: lens aperture, shutter speed and scene luminance. By keeping the scene luminance and lens aperture constant, we vary the shutter speed to obtain photographs at different exposure values.

Depending on the chosen exposure, the displayed photograph may vary from an image with a clearly perceptible depth and large overexposed areas, to a flatter image with smaller overexposed areas. Hence, we can see that choosing the right exposure is essential to determine the perceived flake sparkling throughout the image as well as texture depth.

In this experiment we want to create a series of stereoscopic photographs of the same paint plate, captured with different exposures –again, with a camera that has been colorimetrically calibrated under the same conditions. We will present these photographs to a group of observers, and ask them to select the one they perceive as being the most similar –in terms of flake density, depth, and sparkling– to a physical reference plate. As a result, we expect to be able to find an adequate exposure value as a function of those three parameters. In turn, we can use this function in the rendering engine to select the right exposure in each case, to obtain the desired material aspect.

REFERENCES:

[1] P. Shirley, R. K. Morley, P.-P. Sloan, and C. Wyman, "Basics of physically-based rendering," in SIGGRAPH Asia 2012 Courses, p. 2, ACM, 2012.

[2] F. da Graça, A. Paljic, D. Lafon-Pham, and P. Callet, "Stereoscopy for visual simulation of materials of complex appearance," in IS&T/SPIE Electronic Imaging, vol. 90110X, SPIE,

2014.

[3] N. Dekker, E. Kirchner, R. Super, G. van den Kieboom, and R. Gottenbos, "Total appearance differences for metallic and pearlescent materials: contributions from color and texture," Color Research & Application, vol. 36, no. 1, pp. 4-14, 2011.

[4] C. McCamy, "Observation and measurement of the appearance of metallic materials. Part i. macro appearance," Color Research & Application, vol. 21, no. 4, pp. 292-304, 1996.

[5] C. McCamy, "Observation and measurement of the appearance of metallic materials. Part ii. micro appearance," Color Research & Application, vol. 23, no. 6, pp. 362-373, 1998.

9396-24, Session 7

QuickEval: a web application for psychometric scaling experiments

Khai Van Ngo, Jehans Jr. Storvik, Christopher A. Dokkeberg, Ivar Farup, Marius Pedersen, Gjøvik Univ. College (Norway)

The perceived quality of images is very important for the end-user, industry, engineers and researchers. Subjective image quality assessment is a challenge for many in the field of image quality. The threshold for conducting subjective experiment can be high; they are time-consuming, they can be difficult to set-up and carry out, they might require special software. The most common methods for conducting such experiments include paired comparison, rank order and category judgement [1]. In paired comparison experiments observers judge quality based on a comparison of image pairs, and the observer is asked which image in the pair is the best according to a given criterion, for example which has the highest quality. For rank order experiments the observer is presented with a number of images, who is asked to rank them based on a given criterion. Rank order can be compared to doing a pair comparison of all images simultaneously. In category judgment the observer is instructed to judge an image according to a criterion, and the image is assigned to a category.

These types of experiments are also very often carried out in controlled laboratory settings with a limited group of observers. These often rely on software that has been designed to cover the functionality required for that specific experiment, and they software is very often not publicly available. Attempts have also been made to allow web-based image quality experiments. Qiu and Kheiri [2] presented "Social Image Quality" a web-based system for paired comparison experiments. In the TID2008 database experiments were also conducted online, showing little difference between controlled and uncontrolled experiments [3].

To the best of our knowledge a tool for conducting paired comparison, rank order, and category judgement does not exist in a single software. We have not been able to find software allowing for rank order experiments. In this paper we propose a web application, Quickeval, to create and carry out psychometric experiments both for controlled and uncontrolled environments. The type of psychometric experiments supported by the software is pair comparison, category judgement and rank order. It also support experiment on a large scale, enabling experiments over the internet.

Quickeval is an application where the researcher easily can set up experiments, invite users and extract the experimental results. The application supports two different modes; Observer mode and Scientist mode. The Scientist mode has many options when creating an experiment. First of all, the scientist has to upload pictures which are to be used in the experiment. The scientist has many options during creation, such as type of experiment (paired comparison, rank order or category judgement), random unique picture queues for every observer (the scientist can of course make their own custom picture queue as well), background color for the experiment. After an experiment has been successfully created, then the observers can take them.

Observers can choose to register as an observer with name, age, country etc., or they can choose to log in as "anonymous", although some experiments still require the observer to give information such as age or country. When logged in, observers will receive a list of all public experiments from every scientist registered with Quickeval. Experiments may be set as hidden if the scientist doesn't want anyone to carry out their experiment, or if only invited people are allowed to take the experiment. Some experiments may not be unlocked before the observer has undergone a simple Ishihara test to determine if the observer is colorblind or not or if they do not do a simple monitor calibration procedure.

Once an experiment has been started, the observer is taken to a screen with the pictures that are to be evaluated. This screen only contains the actual pictures. Great care has been taken in order not to put any other disturbing elements. The pictures shown will of course depend on the method of the experiment, and the picture queue. For example a pair comparison experiment will show only two pictures side by side, and the original in the middle if the scientist chose this option during experiment creation. The observer simply left clicks on the picture that is experienced as the best one. The picture is then highlighted, and the observer can proceed to the next image pair. Care was taken in order to make sure pictures weren't scaled. The observer also has the option of panning in the images if they do not fit the screen resolution, and the images can be shown in full screen. Results from the carried out experiment gets stored in a database for later use by the scientist.

The application enables the scientist to easily view the results from an experiment, or it can be downloaded in various formats such as .csv. This allows the scientist to parse the results in their own preferred way.

Quickeval also follows best practice and common guidelines subjective image quality evaluation [4].

Bibliography

[1] P. G. Engeldrum, Psychometric Scaling, a toolkit for imaging systems development, Imcotek Press Winchester USA, 2000.

[2] G. Qiu og A.Kheiri, «Social Image Quality,» in Image Quality and System Performance VIII, San Francisco, CA, 2011.

[3] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, J. Astola, M. Carli og F.Battisti, «TID2008 – A Database for Evaluation of Full Reference Visual Quality Assessment Metrics,» Advances of Modern Radioelectronics, vol. 10, pp. 30-45, 2009.

[4] CIE, Guidelines for the evaluation of gamut mapping algorithms, CIE TC8-03, 156:2004.

9396-25, Session 7

A database for spectral image quality

Steven Le Moan, Technische Univ. Darmstadt (Germany); Sony T. George, Marius Pedersen, Gjøvik Univ. College (Norway); Jana Blahová, Technische Univ. Darmstadt

(Germany); Jon Yngve Hardeberg, Gjøvik Univ. College (Norway)

Assessing the perceived quality of an image pertains to predicting how the human visual system reacts to variations in different features such as contrast, structure, chroma or hue. There is a tremendous amount of literature on this topic and many so-called metrics have been proposed to predict human perception of greyscale and color images, with or without reference data (i.e. an original undistorted image). There is however a need to measure image quality in higher dimensions with the recent advent of multispectral technologies, image appearance models and multi-channel printing. Spectral reproduction workflows have recently raised a considerable interest, particularly for applications where multiple viewing conditions are to be considered. Yet processing spectral data (i.e. for image quality assessment) is not without challenges due to the size, high-dimensionality, but also lack of inherent perceptual meaning of such data. //// Recent studies have formulated the problem of spectral image quality assessment (SIQA) as a multiple color image quality problem, in which spectral data is rendered for different illuminants and quality is then evaluated for each rendering [1]. The problem then pertains to the following question: how do image distortions actually change across illuminants? Preliminary results have demonstrated that even with chromatic adaptation, achromatic distortions (e.g. lightness-contrast, lightness-structure) are much more stable than e.g. chroma or hue for a variety of image reproduction algorithms (gamut mapping, dimensionality reduction, compression). There are however a few more problems to solve in SIQA and this paper intends to provide tools to address them. //// We introduce a new image database dedicated to SIQA. A total of 10 scenes representing different material surfaces were captured by means of a 160 band hyperspectral system with a spectral range between 410 and 1000nm (the VNIR-1600 hyperspectral line scanning camera designed by Norsk Elektro Optikk): cork, wool, tree leaves on top of paper, oranges skin, oil painting, CMYK halftones printed on fluorescent paper, light skin (top of the hand), dark skin (top of the hand), wood, as well as a part of a Gretag color checker. “Realistic” spectral distortions were then designed so as to create five reproductions for each original image, and thus a basis for image comparison: a spectral gamut mapping [2] optimized for two different illuminants (daylight and incandescent light), two kinds of spectral reconstruction from a lower-dimensional representation (a 6-channel filter wheel model and the LabPQR interim connection space) and JPEG2000-like compression. Note that each distortion was applied only to the visible channels of the images. Finally, a paired comparison psychophysical experiment was carried out. A total of 16 observers were asked to select the reproduction that they considered the closest to the original, when the scenes were rendered under daylight (CIED50) and incandescent light (CIEA). Note that the CCT of the display used in this experiment was adjusted differently for each illuminant (5000K for CIED50 and about 2850K for CIEA), in order to give the most accurate rendering possible. Different preferences were recorded for each illuminant, thus confirming the fact that image quality is an illuminant-dependent matter. Nevertheless, in accordance with the results obtained in [1], we observed that achromatic distortions such as the loss of spatial structure yielded by the JPEG2000-like compression were perceived as equally annoying under both lights. The database, which will be publicly available, is intended to be used in a variety of applications related to SIQA and material appearance modelling. In this paper, we demonstrate its usefulness in the evaluation of spectral image difference measures.

REFERENCES

[1] Le Moan, S. and Urban, P., “Image-difference prediction: From color to spectral”, IEEE Transactions on Image Processing 23(5), 2058-2068 (2014).

[2] Urban, P. and Berns R. S., “Paramer Mismatch-based Spectral Gamut Mapping”, IEEE Transactions on Image Processing 20(6), 1599–1610 (2011).

9396-26, Session 7

Alternative performance metrics and target values for the CID2013 database

Toni I. O. Virtanen, Mikko Nuutinen, Jenni E. Radun, Tuomas M. Leisti, Jukka P. Häkkinen, Univ. of Helsinki (Finland)

Amongst the top priorities of the image quality research is the creation of a computational model (in the form of an algorithm) that is capable of predicting the subjective visual quality of natural images. An established practice is to use publicly available databases when the performance of new image quality assessment (IQA) algorithms is tested or validated [1]. This paper proposes target values to evaluate the performance of (IQA) algorithms utilizing a newly published multiply distorted real camera image database (CID2013) [2]. With multiply distorted stimuli, there can be more disagreement between observers as the task is more preferential than that of distortion detection. One observer might prefer one sample, for example with less noise, while another prefers a different sample with better sharpness but more noise. Therefore it is important that especially with multiply distorted databases, reporting only the Mean Opinion Scores (MOS) is not sufficient. The CID2013 database consist the full data for each observer instead of just the averaged MOS scores for the images, allowing more thorough investigation of the properties of the subjective data. If the consensus between the observers in sorting the samples in their order of quality is low, the target performance level should be reduced. One such target level could be the point when a data statistic measure is saturated, e.g. when increase in sample size won’t change its value significantly. We have calculated Root-Mean-Square-Error (RMSE) value, as a data statistic measure, for the subjective data as a function of the number of observers, applied from [3]. Each observer and the averages of combinations of two, three, four, etc. subject sets are compared against the overall MOS of all subjects. The data shows that after 15 subjects the RMSE value saturates around the level of four, meaning that a target RMSE value for an IQA algorithm for CID2013 database is four. Investigating the full data of CID2013 opens up additional ways to evaluate the performance of IQA algorithms to complement the traditional Pearson linear correlation coefficient (LCC), Spearman rank ordered correlation coefficient (SROCC) and RMSE, performance measures. With full data we can calculate e.g. the number of statistical differences between every image pair, using Linear Mixed Models ANOVA with Heterogeneous Compound Symmetry (HCS) covariance matrix. We prefer the Linear Mixed Models because it can handle data more effectively than standard methods, such as simple ANOVA can [4]. Simple ANOVA makes the assumption that the MOS distributions are normally distributed and the variance is equal between variables, e.g. the images. Linear Mixed Models is a better fit to the data as with single images the distribution can be strongly skewed towards either ends of the scale, depending on its overall quality compared to other images in the test. Finally we also need to use Bonferroni correction to control the higher risk of Type I error that multiple comparisons introduce. If the quality difference between image samples is statistically significant and an algorithm can predict this quality order, it means that it has a good performance. This performance can be aggregated to a simple percentage figure on how many image pairs, out of every statistically significant image pair, an algorithm can correctly predict. If however the difference is not statistically significant in the subjective data, should we demand any better performance with IQA algorithms? Having access to the whole data also allows other type of information to be acquired. For example, we can calculate the point where the average standard deviation no longer decreases significantly with the addition of new subjects, meaning that an optimal amount of observers has been reached. This figure is useful when planning of future experiments with the same experimental setup in subjective tests

1. Dinesh Jayaraman, Anish Mittal, Anush K. Moorthy and Alan C. Bovik, Objective Quality Assessment of Multiply Distorted Images, Proceedings of Asilomar Conference on Signals, Systems and Computers, 2012.

2. Virtanen, T., Nuutinen, M, Vaahteranoksa, M, Häkkinen, J., Oittinen, P., “CID2013: a database for evaluating no-reference image quality assessment algorithms,” IEEE Transactions on Image Processing, accepted, to appear 2014.

3. Nuutinen, M., “Reduced-Reference Methods for Measuring Quality Attributes of Natural Images in Imaging Systems,” Aalto University, DOCTORAL DISSERTATIONS 157/2012, 184 pages

4. Nuutinen, M., Virtanen, T., Häkkinen, J., “CVD2014 - A database for evaluating no-reference video quality assessment algorithms,” submitted to review.

Conference 9396:
Image Quality and System Performance XII

9396-27, Session 7

Extending subjective experiments for image quality assessment with baseline adjustments

Ping Zhao, Marius Pedersen, Gjøvik Univ. College (Norway)

In a typical working cycle of subjective image quality assessment, it is common to invite a group of human observers to give perceptual ratings on multiple levels of image distortions. Due to the complexity of rating procedures and the number of image distortions involved, the stimuli might not be rated in a single experiment session. In some other cases, adopting a newly introduced image distortion to existing ones requires repeating the entire previous subjective experiment. The whole rating process is non-trivial and it consumes considerable time and human resources. One potential answer to this research challenge can be the baseline adjustments [1]. This method incorporates common stimuli to form a baseline for determining the comparability of ratings between different experiment sessions and may allow the computation of scale values expressed relative to responses for the baseline stimuli [2]. The effectiveness of a common baseline strongly associates with the type and number of stimuli included in the baseline, but the discussions regarding this topic were largely ignored in the previous researches. In this paper, we conduct an experimental study to verify and evaluate the baseline adjustment method regarding extending the existing subjective experiment results to a new experiment session. The first goal of this research is to verify that the baseline adjustment is an appropriate method for extending subjective experiment, and the second goal is to identify the type and number of stimuli that we should use in the common baseline in order to minimize the workload and complexity of subjective experiments.

Comparing to the conventional researches focusing on case studies of hypothetical data sets [1, 2], we focus on the data sets collected from a real subjective experiment which incorporates twenty human observers to evaluate the spatial uniformity of a projection display by observing a series of projected test images. We split the original stimuli into three groups. The stimuli in the first group are assumed to be used in the existing experiment session only, while the stimuli in the second groups are assumed to be used in the new experiment session. The rest of them are selected as the common baseline for scaling the raw ratings. A copy of the common baseline is distributed to the stimulus group with no modification, and another copy of it is distributed to the second stimulus group with randomly added noises following a standard Gaussian distribution. The purpose of adding noises is to simulate the differences across sessions in baseline ratings due to difference in judgment criteria of observers but not the difference in perception. Then the two stimulus groups sharing a common baseline are merged and scaled by mean Z-score method to produce the final perceptual ratings with respect to specific scaling procedures. Since the raw perceptual ratings are collected from a single experiment session, we should expect a high correlation between the median Z-scores of the raw perceptual ratings and the baseline adjusted version if the common stimuli selected are appropriate.

The results from a preliminary research indicates that the most optimal common stimuli to extend the subjective experiments should be the ones giving the smallest variance in perceptual ratings upon all human observers in the existing experiment session. The results from the upcoming research are expected to suggest that it is possible to reduce the number of common stimuli so the time and labor cost of extended subjective experiment can be largely reduced if we are able to tolerate a certain level of unreliability with respect specified statistic confidence interval (results and numbers to be presented in the manuscript). Although the conclusions from this research are derived within the theme of image quality assessment but they can be applied to all types of researches incorporating subjective experiment in general.

[1] T. C. Brown and T. C. Daniel, "Scaling of Ratings?: Concepts and Methods," Rocky Mountain Forest and Range Experiment Station, technical report, 1990.

[2] T. C. Brown, T. C. Daniel, H. W. Schroeder, and G. E. Brink, "Analysis of Ratings: A Guide to RMRATE," Rocky Mountain Forest and Range Experiment Station, technical report, 1990.

9396-28, Session 7

Subjective quality of video sequences rendered on LCD with local backlight dimming at different lighting conditions

Claire Mantel, Jari Korhonen, DTU Fotonik (Denmark); Jesper M. Pedersen, Søren Bech, Bang & Olufsen (Denmark); Jakob Dahl Andersen, Søren O. Forchhammer, DTU Fotonik (Denmark)

Local backlight dimming is a technology that aims at both saving energy and improving the quality of images rendered on Liquid Crystal Displays (LCDs). It consists in reducing the luminance of the display backlight in areas where the local image content is dark and does not require full intensity. Two types of defects can emerge from the intensity of the backlight: leakage, when the Liquid Crystal (LC) cells fail to block light completely, producing grayish black pixels; and clipping, when not enough light is provided to the LC cells to reach the intended luminance. This paper investigates the role of the ambient light level in a viewing room on the perception of videos rendered with local backlight dimming. As leakage can appear only in the dark areas, the elevation of the minimum black level due to reflections and the decrease in contrast due to ambient light adaptation particularly influence its perception.

We set-up a subjective test in which participants rated the quality of video sequences presented on an LCD platform with a controllable backlight at various ambient light levels. Two local backlight dimming algorithms were applied: the first algorithm represented a conventional LCD (Full backlight), i.e. all LED segments were set uniformly to their full intensity, and the second algorithm (the Gradient Descent algorithm), aimed at achieving best quality by adapting the backlight to the content according to a display model. Participants were located at a distance of three times the display height and the display used is a 46'' LCD which was rotated by 15 degrees in order to emphasize leakage perception. Each of the twenty participants rated every stimuli using a continuous scale and repeated the experiment twice. Sequences were shown in a room with dark walls at three different ambient light conditions: no light (approximately 0 lux), low ambient light (approximately 5 lux) and higher ambient light (60 lux). The peak white of the display was kept at the constant value of 490Cd/m? for all ambient light levels. For the test five sequences all containing dark areas that could show leakage defect were used. Participants were given time to adapt to each light condition before performing the required task.

An ANOVA analysis of the data shows that the used dimming method (Algorithm) and content (Sequence) have a significant effect on the ratings. The Gradient Descent algorithm is significantly preferred to full backlight under all light conditions. The ambient light (Light) does not have a significant influence alone, but the interaction between Light and Sequence, as well as Light, Sequence and Algorithm, are significant. Post hoc tests (Tukey $p<0.05$) show that there is a significant difference between the ratings attributed at the two lower ambient light levels and those attributed at the higher ambient light level. Our work hypothesis was that the subjective quality differences between full backlight and Gradient Descent algorithms would diminish when the ambient light level becomes higher. The results support this hypothesis as the differences between the two algorithms tested are less visible (i.e. the grades obtained are closer) at the higher ambient light. The main difference between the two algorithms is that the Gradient Descent algorithm shows globally less leakage but varies over time (on the contrary of the Full backlight). Therefore change in the algorithm ratings implies that the perception of leakage is less significant at higher ambient light.

We have also observed that the subjective preferences are highly dependent on the different contents. Indeed, depending on the sequence, the ambient light levels can cause differences in the subjective grades of more than 15% of the whole scale to no difference at all. In addition, some test subjects tend to prefer full backlight even in cases when the majority of test subjects show a clear preference towards the Gradient Descent algorithm. The content dependency can be explained by different temporal and spatial characteristics affecting the visibility of leakage, i.e. for some contents the leakage and its temporal variations have higher impact on perceived quality

than for other contents. Different individual preferences can be explained by personal tendencies of paying major attention on different aspects in visual content.

Further analysis of the results will include the ambient light computation in the display model to provide accurate input to quality metrics and evaluate how their performances are affected.

9396-29, Session 7

Study of the impact of transmission parameters on the QoE of video calling services over the LTE/4G network

Maty Ndiaye, Orange SA (France) and Univ. of Poitiers (France); Gwenael Le Lay, Orange SA (France); Hakim Saadane, Mohamed-Chaker Larabi, Univ. de Poitiers (France); Catherine Quinquis, Orange SA (France); Clency Perrine, Univ. de Poitiers (France)

Today, the deployment of LTE/4G networks and the technological advances in electronic components and handheld devices allowed to solve problems (device quality and network capacity) that have limited in past the ascension of the video calling services. Consequently, video calling services are seen afresh by telecom network operators as an attractive and financially interesting market. More, their accessibility to consumers increased significantly and they represent therefore a good alternative to conventional phone. Ensure the best quality for this service became thus a major issue for network operators and service providers in order to remain competitive in this growing market. However, managing and assessing end-user QoE is a challenging domain where intensive investigations are conducted. Indeed, Supply the best possible quality of experience (QoE) for this type of service depends on several parameters from device, encoding and transmission sides. The proposed work aims to assess the perceived quality of a video calling service performed on a real (not simulated) LTE/4G network. To reach this goal, subjective quality assessment tests are conducted to determine how QoS (Quality of Service) mechanisms affect the user's QoE. Preliminary results confirm that QoE is mainly impacted by packet loss, jitter and delay. However the impact of these three transmission parameters which vary over time and in case of user mobility seems to be different from that observed with simulated networks.

9396-30, Session 8

RGB-NIR image fusion: metric and psychophysical experiments

Alex E. Hayes, Graham D. Finlayson, Univ. of East Anglia (United Kingdom); Roberto Montagna, Spectral Edge Ltd. (United Kingdom)

RGB-NIR color image fusion has been proposed as a way of improving photography to better capture and portray details lost in normal image capture, including detail through haze and detail at distance. Our paper starts with the simple observation that while the goal of RGB-NIR fusion is laudable the produced pictures are often far from pleasing photos (at least in our opinion). To test this hypothesis the main contribution of this paper is to carry out a standard pairwise preference comparison of leading algorithms together with giving the observer the choice of choosing the original.

We selected four methods for image fusion of the visible RGB and near-infrared (NIR). Firstly, the Spectral Edge method[1], which extends the Socolinsky and Wolff (SW) image fusion method[7] in two ways. First, the SW method is directed toward mapping N channels to 1, whereas Spectral Edge provides a new theory for N to 3 (or generally N to M image fusion). Second, the SW method necessarily introduces artifacts into the processed images whereas the Spectral Edge method was designed so this does not happen. This said, the Spectral Edge method is not as 'local' as the SW processing and the produced image fusion results are much less dramatic. The second image method we consider is the dehazing method of Schaul et al.[3]. Here a luminance image is merged with NIR in a pyramidal scheme (inspired by Toet[10]) and then the new fused luminance image is combined with the chromaticities of the original color images. The pyramidal scheme uses edge-preserving filters[8] to avoid - to a large extent - artifacts such as bending and halos common in many fusion schemes. Third, the luminance channel and NIR channel are simply averaged and used as a luminance channel replacement, as suggested by Fredembach and Süsstrunk[9]. The final method fuses the luminance channel of the RGB image with the NIR image using a standard non-edge-preserving ratio of low-pass pyramid (ROLP) technique. All these methods are 4 to 3 channel, i.e., they produce a color RGB output image. As a control, the unprocessed RGB images are included in the measurement as well.

The psychophysical experiment is a pairwise comparison test (Thurstone's law case V), the standard technique for evaluating image processing algorithms both in academic research and in industry, as used by Connah et al.[6]. On each comparison they have to select the image they prefer according to personal taste (through forced-choice, i.e., there is no ``I don't know'' option). All pairs of images (for the different algorithms and the same scene) are presented twice (each pair is presented as a left right pair and as a right-left pair). 10 images from the EPFL RGB-NIR data set are used[13]. We also adopt ISO 3664:2009 recommendations for carrying out image preference experiments. The pairwise preferences for multiple observers are counted in a score matrix and then using Thurstone's method we convert the scores into algorithm ranks and then to preference scores. Significantly, the Thurstone method also returns confidence intervals (and so it is possible to conclude whether one algorithm is significantly better than another). Images are displayed on a calibrated monitor in a dark room, and each comparison is repeated twice in the course of a trial (swapping place between left and right images). As the number of comparisons is relatively high (200), each experiment is split into two sessions of 100 comparisons each, to reduce fatigue in volunteers.

Preliminary preference results, with 6 observers naive about the experiment, conclude that the Spectral Edge method is the leading algorithm, second is the dehazing method of Schaul et al., third is the ROLP fusion, fourth is the original RGB image, and the luminance channel and NIR average is in last place. The Spectral Edge method - which provides much less detail than the competing methods - is, we believe, preferred because of the lack of artifacts, because it is closer to what is expected (from a normal photographic diet), and because the color aspect of image fusion is integral to the method (it is not based on luminance fusion).

Given this preference result we build on - as a simple proof of concept test - existing image fusion metrics to predict our preference results. Our new metric extends the metric of Xydeas and Petrovi?[4] , which is based on measuring how much gradient information from the input images is transferred to the output image, and was recently rated as a top metric in a large-scale comparison of image fusion metrics[5].

Our metric extends Xydeas and Petrovi?'s method to explicitly quantify the 4 to 3 fusion (as opposed to the 2 to 1 image fusion originally considered). This represents the detail transfer component of the metric. As we are considering image fusion for photography, we also incorporate measures of colorfulness and contrast. Colorfulness, as measured in CIELUV chromaticity, has been correlated with observer preference, up to a certain point[12], as has root mean square (RMS) contrast[11], so these are included as weights in the metric. We take the output RGB image into CIELUV color space, and calculate its mean chroma from the u* and v* channels, and its RMS contrast from the L* channel. Our new metric is the sum of these three elements.

Our metric gives the same ranking of the methods, with the exception of the original RGB image, which it places higher. But, the original, in effect, represents a boundary for image fusion (it is an image where no fusion has taken place). Our metric is very much a work in progress and we will further develop before the conference (to amongst other things, consider these kinds of boundary cases).

The conclusion of our study is that observers do like to see NIR data in photographic images but only when the resulting images are similar to conventional photographic images. Or, seeing at distance or through haze is preferred only if it looks natural.

References:

[1] Connah, D., Drew, M. S., & Finlayson, G.D. (2014). Spectral Edge Image Fusion: Theory and Applications. Accepted for publication in Proc. IEEE ECCV2014.

[2] Finlayson, G. D., Connah, D., & Drew, M. S. (2011). Lookup-table-based gradient field reconstruction. Image Processing, IEEE Transactions on, 20(10), 2827-2836.

[3] Schaul, L., Fredembach, C., & Süsstrunk, S. (2009, November). Color image dehazing using the near-infrared. In ICIP (pp. 1629-1632).

[4] Xydeas, C. S., & Petrovi?, V. (2000). Objective image fusion performance measure. Electronics Letters, 36(4), 308-309.

[5] Liu, Z., Blasch, E., Xue, Z., Zhao, J., Laganiere, R., & Wu, W. (2012). Objective assessment of multiresolution image fusion algorithms for context enhancement in night vision: a comparative study. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34(1), 94-109.

[6] Connah, D., Finlayson, G. D., & Bloj, M. (2007, January). Seeing beyond luminance: A psychophysical comparison of techniques for converting colour images to greyscale. In Color and Imaging Conference (Vol. 2007, No. 1, pp. 336-341). Society for Imaging Science and Technology.

[7] Socolinsky, D. A., & Wolff, L. B. (2002). Multispectral image visualization through first-order fusion. Image Processing, IEEE Transactions on, 11(8), 923-931.

[8] Farbman, Z., Fattal, R., Lischinski, D., & Szeliski, R. (2008, August). Edge-preserving decompositions for multi-scale tone and detail manipulation. In ACM Transactions on Graphics (TOG) (Vol. 27, No. 3, p. 67). ACM.

[9] Fredembach, C., & Süsstrunk, S. (2008, January). Colouring the near-infrared. In Color and Imaging Conference (Vol. 2008, No. 1, pp. 176-182). Society for Imaging Science and Technology.

[10] Toet, A. (1990). Hierarchical image fusion. Machine Vision and Applications,3(1), 1-11.

[11] Calabria, A. J., & Fairchild, M. D. (2003). Perceived image contrast and observer preference I. The effects of lightness, chroma, and sharpness manipulations on contrast perception. Journal of imaging Science and Technology, 47(6), 479-493.

[12] Fedorovskaya, E. A., de Ridder, H., & Blommaert, F. J. (1997). Chroma variations and perceived quality of color images of natural scenes. Color Research & Application, 22(2), 96-110.

[13] Brown, M., & Susstrunk, S. (2011, June). Multi-spectral SIFT for scene category recognition. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on (pp. 177-184). IEEE.

[14] ISO 3664:2009 (2009). Graphic technology and photography - Viewing conditions, ISO, Geneva, Switzerland.

9396-31, Session 8

Non-reference quality assessment of infrared images reconstructed by compressive sensing

Juan Esteban Ospina, Hernan Dario Benitez Restrepo, Pontificia Univ. Javeriana, Cali (Colombia)

Recent works demonstrate that compressive sensing theory enables to reconstruct sparse or compressible images from fewer measurements than Nyquist-Shannon sampling theory. This novel theory exploits the sparsity property and the capability of the optimization algorithms to reconstruct an image. This scheme uses approximately 10% of the image discrete cosine transform (DCT) coefficients in the reconstruction process. This approach could support the implementation of low cost infrared (IR) sensors that would require less elements to attain reconstructed IR images with good quality. Generally, the peak signal-to-noise ratio (PSNR) or mean square error (MSE) are used to evaluate the reconstructed image. Latest works on perceptual quality image exhibit that image quality assessment (IQA) measures, which are based on visual system features, are better correlate with human judgements than traditional measures. Since the IR images are representations of the world and have natural features like images in the visible spectrum, it would be interesting to use these IQA measures to assess IR image quality perception. In this work, we compare the quality of a set of indoor and outdoor IR images reconstructed from measurement functions formed by linear combination of their pixels. The reconstruction methods are: linear discrete cosine transform (DCT) acquisition, DCT augmented with total variation minimization, and compressive sensing scheme. The PSNR, three full-reference (FR) IQA and four no-reference (NR) IQA measures computes the qualities of each reconstruction: multiscale structural similarity (MSSIM), visual information fidelity (VIF), information fidelity criterion (IFC), local phase coherence (LPC-SI), blind/referenceless image spatial quality evaluator (BRISQUE), naturalness image quality evaluator (NIQE) and gradient singular value decomposition (GSVD), respectively.

We perform a differential mean opinion score (DMOS) test to compare the measure predictions with human judgements. We observe that MSSIM is highly correlated with human judgements, but it needs the reference image. Among NR measures, GSVD has the highest Pearson linear (PLCC) and Spearman rank-order (SRCC) correlation coefficients, and the lowest root mean square error (RMSE). Nonetheless, FR measures outperforms GSVD. Based on these results, we use MSSIM to compare the reconstruction methods. We find that the compressive sensing scheme produces a good-quality IR image (PLCC=0.8626, SRCC=0.8563 and RMSE=5.6223), from 30000 random sub-samples and 1000 DCT coefficients (2%). However, linear DCT provides higher correlation coefficients than compressive sensing scheme by using all the pixels of the image and 31000 DCT (47%) coefficients. Finally, the comparison of running times for each reconstruction method shows that compressive sensing and linear DCT are the slowest and fastest techniques, respectively.

9396-32, Session 8

Study of the effects of video content on quality of experience

Pradip Paudyal, Federica Battisti, Marco Carli, Univ. degli Studi di Roma Tre (Italy)

Demand of multimedia services, including video streaming and Video on Demand (VoD)is increasing popularity due to the advancements in wireless mobile technologies including Long Term Evolution (LTE/LTE-Advanced). Moreover competition among service providers is also increasing. It is then of fundamental importance the ability to deliver the video content through an error prone network, with the quality required by a customer for a specific application.Video service is more prone to network transmission impairments so the ability of guaranteeing a desired level of perceived quality/Quality of Experience (QoE) is crucial.

Perceived quality can be measured by subjective and objective approaches. Subjective method is based on the collection of user opinion and it is measured in terms of Mean Opinion Score (MOS). This process is impractical, complex and costly in terms of time. Many objective models have been proposes to estimate perceived quality from QoS parameters: PLR, Jitter, Depaly, Throughput, and encoding artifacts. However, perceived video quality not only depends on QoS parameters and encoding artifacts but also on its content, context, motion, etc.

Our goal is to investigate the effect of video content on perceived quality and to determine the major parameters related to video/content itself and use that parameter to find the better mapping model. Here, we discuss the effect of video content which is characterized by Spatial Perceptual Information (SI), Temporal Perceptual Information (TI), total motion coefficient and data rate in perceived video quality metrics. For the analysis, SI is computed based on the Sobel filter and TI is based upon the motion difference feature. Whereas total motion coefficient is computed based on absolute differences between frames, and data rate is extracted directly from the source video sequences. For the analysis study we have exploited the ReTRiEVED video quality database.Test videos are generated by considering practical scenario using network emulator and source videos with heterogeneous content characterized by wide span of spatial-temporal plane, motion and data rate. Also, a subjective experiment has been performed for collecting subjective scores.

Analysis of the result shows that:

• perceived video quality does not only depend on key Quality of Service (QoS) parameters: delay, packet loss rate, jitter and throughput and coding and compression artifacts. Color, context, content, motion, may impact on the overall judgment.

• Perceived video quality is not so influenced by the source data rate.

• Perceived video quality is not so dependent on SI, however it depends strongly on TI and it decreases significantly for every increase in TI.

• There is a strong relation between video quality and total motion coefficient and it decreases significantly for every increase in motion coefficient.

From the analysis, it results that with network level QoS and encoding parameters, we also have to consider the content information (especially TI and total motion coefficient) during the design and testing of QoE/QoE mapping model for close performance.

9396-33, Session 8

The effects of scene content, compression, and frame rate on the performance of analytics systems

Anastasia Tsifouti, Home Office (United Kingdom) and Univ. of Westminster (United Kingdom); Sophie Triantaphillidou, Univ. of Westminster (United Kingdom); Mohamed-Chaker Larabi, Univ. of Poitiers (France); Graham Doré, Home Office Centre for Applied Science and Technology (United Kingdom); Efthimia Bilissi, Alexandra Psarrou, Univ. of Westminster (United Kingdom)

Video analytics (VA) are computerized autonomous systems that analyze events from camera views for applications, such as traffic monitoring and behavior recognition [1]. The Image Library for Intelligent Detection Systems (i-LIDS) is a set of video surveillance datasets, a UK Government standard for VA systems [2]. One of the i-LIDS scenarios is investigated in this paper: the Sterile Zone (SZ). The SZ dataset is segmented into shorter video clips of approximately 40 minutes duration. The SZ is a low complexity scenario, consisting of a fence (not to be trespassed) and an area with grass. The analytics system needs to alarm when there is an intruder entering the scene (an attack).

The aim of this investigation is to identify the effects of compression and reduction of frame rate to the performance of four analytics systems (labeled A, B, C, D). Furthermore, to identify the most influential scene features, affecting the performance of each VA system under investigation. The included four systems have obtained UK Government approval by been tested with "uncompressed" footage [2].

The work includes testing of the systems with D1 PAL resolution of uncompressed and compressed (7 levels of compression with H.264/MPEG-4 AVC at 25 and 5 frames per second) footage, consisting of quantified scene features/content. The scene content of 110 events was characterized and scenes were classified into groups. The characterization includes groups of scenes based on: scene contrast (contrast between main subject and background), camera to subject distance, spatio-temporal busyness, subject description (e.g. one person, two people), subject approach (e.g. run, walk), and subject orientation (e.g. perpendicular, diagonal). Additional footage, including only distractions (i.e. no attacks to be detected) is also investigated. Distractions are elements in the scene, such as abrupt illumination changes and birds that could be falsely recognized by the systems as, intruders.

To be able to measure the performance of the analytics systems, a method was developed to simultaneously play the video clips without degradation, record the alarm events raised by the analytics systems without time delays, and compare the raised alarms with ground truth data. The rules determining whether an a alarm event was true, or false, were defined as follows: if an alarm falls within the ground truth alarm period then a true match is recorded; if an alarm occurs outside of the ground truth period then that is noted as a false alarm. The obtained results have scores of 1 to the correctly detected attacks and 0 to the un-detected attacks. To estimate the consistency of experimental results, each clip was repeated 10 times.

There are some small variations on the results between the repeated times, which is due to the noise added to the video signal (i.e. as part of the output of footage to the detection systems), and/or the actual intrinsic parameters of the analytics systems (i.e. how it is tuned) and/or the properties of the events (i.e. variation is triggered by certain events). Each analytics system produced different results. For example, variation of results is increased with highly compressed clips at 25fps for system A, but this is not the same for system B (i.e. uncompressed clips produced more variation than highly compressed clips). This variation is increased for most systems when reducing the frame rate from 25fps to 5fps.

The results indicate that detection performance does not monotonically degrade as a function of compression. This phenomenon is similar to relevant findings from face recognition studies [3-5]. Every system has performed differently for each compression level, whilst compression has not adversely affected the performance of the systems. Overall, there is a small drop of system performance from 25fps to 5fps.

Most false alarms were triggered for a specific destruction clip (i.e. which includes abrupt changes of illumination due to the sunny day and small clouds) and at high compression levels.

Since the performance of the examined analytics systems is not affected by compression in the same manner, a factorial analysis for proportional data (i.e. to take into consideration the repeated measures) is applied to each individual level of compression and frame rate [6]. This will allow to identify correlations between scene features, compression and frame rate for each system. For example, two people were easier to be detected than one person in the scene (i.e. scene feature of subject description category) by most systems for most levels of compression and frame rate.

Little research has been done in the area of image compression and analytics systems as currently only few scenarios are capable for autonomous analysis (SZ is one of them). Nonetheless this area is receiving a large amount of research investment [1]. In a world of rapidly technological changes, analytics systems will need to be more flexible and be able to be used in post-event forensics and with limited transmission bandwidth (e.g. through an Internet Protocol network).

Understanding how the analytics systems are behaving with compression, frame rate reduction and scene content could contribute to the improvement in the development of such systems. For example, the developed compressed datasets could be used by manufacturers to improve the performance of their systems with compression and reduced frame rate.

[1] Regazzoni, C. S., Cavallaro, A., Wu, Y., Konrad, j., and Hampapur., Video Analytics for surveillance: Theory and Practice, Signal Processing Magazine, IEEE, 27(5), pp. 1617, (2010)

[2] Home Office CAST, Imagery Library for Intelligent Detection Systems: the i-LIDS user guide, v4.9, No10/11, (2011)

[3] Delac, K., Grgic, S., Grgic, M.,Image compression in face recognition – a Literature Survey,"in: Delac, K., Grgic, M., Bartlett, M. S., [Recent Advances in face recognition], IN-the, Croatia, pp. 1-14, (2008).

[4] Delac, K., Grgic, S., Grgic, M., Effects of JPEG and JPEG2000 Compression on Face Recognition. Pattern Recognition and Image Analysis, Springer Berlin. p. 136-145, (2005).

[5] Wat, K. and Srinivasan, S. H., Effect of compression on face recognition. in Proc. of the 5th International workshop on image analysis for multimedia interactive services, (2004).

[6] Crawlry, M.J., Proportion Data in Statistics: An introfuction using R, John Wiley & Sons Ltd, pp. 247-262, (2005).

9396-34, Session 8

How perception of ultra-high definition is modified by viewing distance and screen size

Amélie Lachat, Jean-Charles Gicquel, Jérôme Fournier, Orange SA (France)

Context

TV technology is in constant evolution. For example, High Definition television (HD-TV) has improved Standard television (SD-TV), and now the new Ultra High Definition television (UHD-TV) technology has been defined to ameliorate the existing HD-TV standards [1]. Among the improvements of UHD-TV is the number of pixels per image that has been increased twice. Consequently, the resolution has been changed from 1920x1080 (HD-TV) to 3840x2160 (UHD-TV). This change influences the optimal viewing distance in terms of visual performance or users' preference. The recommendation ITU-R BT.2246 defines 1.5H as the optimal or design viewing distance (DVD) for optimal perception of UHD-TV, where H is the height of the screen [2].

There are few papers discussing the impact of the viewing distance on the users' perception or preference in the literature. Some studies use the definition "optimal viewing distance", while others "preferred viewing distance" [2,3]. However, it is important to distinguish between them because the first one is defined by the size and resolution of the screen [2] whereas the second one considers other parameters e.g. the country, the individual preferences, the habits etc. [3-4]. This difference points out the various environmental and technical constraints to take into account while studying the impact of viewing distance. For instance, in France home TV screen diagonal has been increasing annually and in 2013 it has reached 34.4 inches in average [5]. However, such data are not always considered in experimental research and for example smaller screen sizes are used for experiments decreasing its reliability [3]. Besides, the optimal viewing distance might not be adapted to the viewing environment of users. For instance, in the USA the television viewing distance at home is around 3.3m [6], while in Great Britain it is 2.7m [7]. Taking this data into account, the influence of both the technical and environmental parameters should be considered.

Objective

As with all new technologies in the process of normalization it is important to understand the influence of new technical parameters on user's perception before giving recommendations. The main objective of this study is to explore the added value of Ultra High Definition Television in terms of user's perception considering the viewing conditions recommended by ITU-R BT.2246 [2] and representative of home environment. Therefore, this study aims answering the following questions: (a) Do users perceive video quality degradation/improvement varying the resolution, the screen size and viewing distance? (b) Do users' feelings (comfort and preferences) change with the viewing distance? To sum-up, the optimal visualization parameters for a viewer perception are investigated for various viewing distances, scene contents and screen sizes.

Method

Four different outdoor contents (3 documentaries and 1 sport content) with different complexity levels have been filmed using the UHD camera at 60Hz by the 4EVER [8] project. For each scene, three different resolution have been generated. Original UHD contents were downscaled to HD and SD resolutions using Lanczos filter. This was the t processing performed to the contents.

The psychophysical test set-up consisted of two 55" and 84" UHD displays, which were placed in two different rooms. All prepared 10 seconds video sequences were presented at 1.5H, 3H and 4.5H distances (optimal viewing distance for respectively UHD ,HD, SD), where H is the height of the screen. The subjective test was divided into six sessions. One session was defined by one screen size and one viewing distance; its average duration was 25 minutes. To avoid visual fatigue, each subject assessed 3 sessions per day.

Furthermore, 30 subjects evaluated the different scenes. After visualization of each test sequence, observers were asked to answer seven questions concerning feelings, viewing distance and scene quality. For each observer the session, scene and question orders were randomized. The analysis was accomplished computing Mean Opinion Scores and exploiting the answers of observers.

Conclusions

This study was launched to better understand and compare the perception of UHD format with lower resolution ones. Therefore, the influence of the viewing distance, screen size and scene content on quality perception, preferences and feelings of users was investigated. The votes of 30 observers were analyzed. Generally, the observers preferred shorter viewing distances such as 1.5H or 3H than 4.5H, where some of them wanted to move closer to the display. Hence, the viewing distance has a significant influence on the perceived quality, especially for SD. It was also established that visual annoyance and discomfort are rather evoked by SD whatever the viewing distance is. But at the distance of 1.5H nobody wanted to move backwards.

Moreover, the highest MOS was obtained at optimal viewing distance or DVD for each resolution. However, a small difference in terms of perceived quality was found between HD and UHD.

[1] Recommendation ITU-R BT.2020: Parameter values for ultra-high definition television systems for production and international programme exchange. (2012, August).

[2] Recommendation ITU-R BT.2246 : The present state of ultra-high definition television (2014, April)

[3] Lee, D.-S. (2012): Preferred viewing distance of liquid crystal high-definition television. Applied Ergonomics, 43(1), 151–156.

[4] Nathan, J. G., Anderson, D. R., Field, D. E., & Collins, P. (1985). Television Viewing at Home: Distances and Visual Angles of Children and Adults. Human Factors: The Journal of the Human Factors and Ergonomics Society, 27(4), 467–476.

[5] GFK. Marketing Research : Bilan 2013 des marchés des biens techniques en France. GFK, 2014.

[6] Nathan, J. G., Anderson, D. R., Field, D. E., & Collins, P. (1985). Television Viewing at Home: Distances and Visual Angles of Children and Adults. Human Factors: The Journal of the Human Factors and Ergonomics Society, 27(4), 467–476.

[7] Tanton, N.E., 2004. Results of a Survey on Television Viewing Distance. BBC R&D white paper WHP090, June.

[8] 4EVER project : http://www.4ever-project.com/

9396-35, Session 9

A no-reference video quality assessment metric based on ROI

Lixiu Jia, Xuefei Zhong, Yan Tu, Wenjuan Niu, Southeast Univ. (China)

A no reference video quality assessment metric based on the region of interest (ROI) was proposed in this paper. The objective video quality was evaluated by integrating the quality of the most two important compressed artifacts, i.e. blurring distortion and blocking distortion of the ROI. An objective bottom-up ROI extraction model combining center weighted temporal saliency model, center weighted color opponent model, center weighted luminance contrast model and center weighted frequency saliency model based on spectral residual was built. The center weighted matrix was used to simulate the preference of human eyes to the whole layout of the image. The temporal saliency model was constructed by computing the magnitude discrepancy of discrete wavelet transform between two consecutive frames. One low frequency information and three high frequency information obtained from consecutive two frames by using sym4 wavelet transform were subtracted from frame to frame. It not only avoided the phenomenon of the "holes" comparing with the two consecutive frames subtraction, but also had a low computation complexity and high detection accuracy. A mean fusion technique was used to obtain the final saliency maps. The no reference video quality assessment metric of the two artifacts was built after the computation of the objective blurring and blocking values of ROI of videos.

The corresponding subjective experiment was conducted to evaluate the performance of our no reference video quality assessment model and objective ROI extraction metric. The subjective experiment included three sessions, i.e. overall quality assessment, blurring assessment and blocking assessment. The 4 original videos were compressed to 5 different bitrates including 0.82Mbps, 1Mbps, 1.2Mbps, 1.5Mbps and 2 Mbps using H.264 coding method. Totally 24 videos were used in the experiment. The 18 videos of overall 24 videos were used to build objective video quality metric and the residual 6 videos were used to verify the model. All the videos were presented by the SMI Experiment Center and displayed on the 46 inch

HYUNDAI LCD TV with the resolution of 1920?1080. The single stimulus method was adopted in the experiment. The 2-Dimensional Gaussian kernel function was used to extract the human density maps of the overall videos from the subjective eye tracking data extracted from the SMI BeGaze software. The radius ? of the 2D Gaussian kernel function was 72 in this paper. The experimental results indicate that our objective ROI extraction metric has a higher AUC (the area under a receiver operating characteristic (ROC) curve) value. In order to verify the effects of ROI on objective video quality assessment metric, the strengths of blurring and blocking artifacts of the videos are calculated according to the different weight of ROI, which changes from 50% to 100% with the interval of 5%. The results indicate that the correlation between the subjective mean opinion score (MOS) and the objective quality scores improves with the increasing ROI weight for both blurring and blocking artifact metric. From the comparison of the average computational time of objective strength of blurring and blocking artifact of each frame of the video, we can conclude that the ROI metric obviously decreases computation complexity, especially for the objective metric of blurring artifact.

The R square of our objective video quality metric was 0.858. The Pearson Linear Correlation Coefficient (PLCC) between subjective quality MOS and objective quality scores was 0.947 and 0.934 for model set-up and verifying videos respectively.

Comparing with the conventional video quality assessment metrics, the metric proposed in this paper not only decreased the computation complexity, but also improved the correlation between subjective MOS and objective scores.

9396-36, Session 9

Comparison of no-reference image quality assessment machine learning-based algorithms on compressed images

Christophe M. Charrier, Univ. de Caen Basse-Normandie (France); Hakim Saadane, XLIM-SIC (France) and Univ. de Nantes (France); Christine Fernandez-Maloigne, Univ. de Poitiers (France)

Lossy image compression techniques such as JPEG2000 allow high compression rates, but only at the cost of some perceived degradation in image quality.

The way to evaluate the performance of any compression scheme is a crucial step, and more precisely available ways to measure the quality of compressed images. There is a very rich literature on image quality criteria, generally dedicated to specific applications (optics, detector, compression, restoration, etc.).

From several years, a number of researches have been conducted to design robust No-Reference Image Quality Assessment (NR-IQA) algorithms, claiming to have made headway in their respective domains. NR-IQA algorithms generally follow one of three trends. 1) Distortion-specific approaches: These employ a specific distortion model to drive an objective algorithm to predict a subjective quality score. These algorithms quantify one or more distortions such as blockiness, blur, or ringing and score the image accordingly. 2) Training-based approaches: these train a model to predict the image quality score based on a number of features extracted from the image. 3) Natural scene statistics (NSS) approaches: these rely on the hypothesis that images of the natural world (i.e., distortion free images) occupy a small subspace of the space of all possible images and seek to find a distance between the test image and the subspace of natural images.

In this paper, we present an extensive comparative study of well-known NR-IQA training-based algorithms.

The trial NR-IQA algorithms used are the following: 1) BIQI, 2) DIIVINE, 3) BLIINDS, 4) BLIINDS-2, 5) BRISQUE and 6) NIQE . The implementations of the algorithms were either publicly available on the Internet or obtained from the authors.

To provide comparison of NR-IQA algorithms, three publicly available databases are used: 1) LIVE database, 2) TID2008 database and 3) CSIQ image database.

The Spearman correlation coefficient is computed between the subjective values and the predicted scores obtained from trail NR-IQA algorithms.

In addition, to ascertain which differences between NR-IQA schemes performance are statistically significant, we applied an hypothesis test using the residuals between the DMOS values and the ratings provided by the trial IQA algorithms. This test is based on the t-test that determines whether two population means are equal or not. This test yields us to take a statistically-based conclusion of superiority (or not) of an NR-IQA algorithm.

As a first step, all trial algorithms have been compared without performing any new learning phase.

Aanalyzing and discussing about obtained results, one point of view is related to the fact that a same NR metric provided different values of correlation when one considers same distortions from different databases. In addition when considering no learnt degradation, NR metrics are not relevant (e.g., results obtained for Non eccentricity pattern noise and Mean shift (intensity shift) degradations). This behavior was somewhat expected since all metrics have been trained on LIVE subsets. In that case, NR metric seems to fail for quality prediction when no learnt degradation are used.

As a second step of this study, we investigate how the selection of the training set influences the results. To perform such an investigation, we proceed as follows : the training set is generated using different subsets randomly extracted from all databases. The remaining subsets are used to construct the test set. All trial NR-IQA algorithms are trained on this new training set. Finally, performances of the used NR metrics are estimated on the new test set.

A bootstrap of 999 replicates for the generation of the training set is used to obtain robust results.

All obtained results for the three trial databases will be discussed and analyzed in the final version of the paper.

9396-37, Session 9

Objective evaluation of slanted edge charts

Harvey H. Hornung, Marvell Semiconductor, Inc. (United States)

Camera objective characterization methodologies are widely used in the digital camera industry. Most objective characterization systems rely on a chart with specific patterns, a software algorithm measures a degradation or difference between the captured image and the chart itself.

The Spatial Frequency Response (SFR) method, which is part of the ISO 12233 standard, is now very commonly used in the imaging industry, it is a very convenient way to measure a camera Modulation transfer function (MTF). The SFR algorithm can measure frequencies beyond the Nyquist frequency thanks to super-resolution, so it does provide useful information on aliasing and can provide modulation for frequencies between half Nyquist and Nyquist on all color channels of a color sensor with a Bayer pattern. The measurement process relies on a chart that is simple to manufacture: a straight transition from a bright reflectance to a dark one (black and white for instance), while a sine chart requires handling precisely shades of gray which can also create all sort of issues with printers that rely on half-toning. However, no technology can create a perfect edge, so it is important to assess the quality of the chart and understand how it affects the accuracy of the measurement.

In this article, I describe a protocol to characterize the MTF of a slanted edge chart, using a high-resolution flatbed scanner. The main idea is to use the RAW output of the scanner as a high-resolution micro-densitometer, since the signal is linear it is suitable to measure the chart MTF using the SFR algorithm. The scanner needs to be calibrated in sharpness: the scanner MTF is measured with a calibrated sine chart and inverted to compensate for the modulation loss from the scanner. Then the true chart MTF is computed. This article compares measured MTF from commercial charts and charts printed on printers, and also compares how the contrast of the edge (using different shades of gray) can affect the chart MTF. Then considering a modulation

threshold for the chart (if the camera is measured in the RAW domain, then the chart MTF can also be inverted and the threshold can be lower), I define the range of frequencies that the chart can provide and then for a give camera resolution and focal length, for what distance range the chart can reliably measure the camera MTF.

9396-38, Session 9

Evaluating the multi-scale iCID metric

Steven Le Moan, Jens Preiss, Technische Univ. Darmstadt (Germany); Philipp Urban, Fraunhofer-Institut für Graphische Datenverarbeitung (Germany)

1. INTRODUCTION

Despite the fact that there is still much to understand about how the Human Visual System (HVS) interprets an image, recent approaches of Image Quality Assessment (IQA) have allowed to reach very high levels of correlation with human judgment [1, 2]. In this paper, we focus on so-called Full-Reference IQA metrics, which take as an input two digital images (reference and reproduction) and provide a score corresponding to the degree of distortion perceived by the HVS. Recently, Lissner et al. [3] introduced the Color-Image-Difference (CID) metric, inspired by the well-known SSIM index [4], and designed especially to improve the prediction of the latter on chromatic distortions. It was shown to be particularly effective on gamut-mapping distortions. Later, Preiss et al. [5] proposed an improved version - referred to as iCID - in order to account for specific artifacts such as chromatic ringing or chromatic edges. Though the purpose of this most recent publication leans towards optimizing gamut mapping rather than IQA, the iCID metric showed great potential on the 2013 version of the renowned Tampere Image Database (TID2013) [2]. Compared to the original CID, this improved version uses less parameters and takes into account a larger variety of artifacts. In this paper, we introduce a multi-scale version of the iCID and we demonstrate its efficiency on three renowned databases. Note also that we intend to make the Matlab code for the metric available with this publication.

2. MULTI-SCALE ICID

Let O (original) and R (reproduction) be two tri-chromatic images (defined e.g. in sRGB). In the iCID framework O and R are first normalized with an image-appearance model with respect to the viewing conditions (e.g. visual resolution, illuminant, luminance). Note that normalizing the images with respect to the visual resolution is particularly crucial in order to consider the differences between the chromatic and achromatic contrast sensitivities of the human visual system. This is done by filtering the input images with contrast sensitivity functions adapted from iCAM (see e.g. [6]). The images are then converted to the nearly perceptually uniform LAB2000HL color space [7], which is optimized for the CIED65 illuminant. Seven so-called Image-Difference Features (IDF) are then computed, by means of terms adapted from the SSIM index [4]: Lightness-Difference, Lightness-Contrast, Lightness-Structure, Chroma-Difference , Chroma-Contrast, Chroma-Structure and Hue-Difference. They are extracted in the form of IDF maps which depict their spatial organization, and which are then averaged so as to produce a single score for each IDF. We refer to the original paper by Preiss et al. [5] for further explanations. Although all maps are originally computed on a single scale, we propose to compute the contrast and structure terms on 5 different scales, as suggested in [8]. Each map is then averaged and the features are combined into the multi-scale iCID (MS-iCID) score. In addition, MS-iCID uses three parameters to balance the contribution of the seven IDFs. Note that unlike state-of-the-art metrics whose parameters are trained over a particular image quality database (e.g. [9]), these ones were selected by means of a visual inspection by three expert observers, so as to avoid artifacts when the metric is used as an objective function to optimize gamut mapping (see [5] Section III.F).

3. EXPERIMENTS

We compared the proposed MS-iCID metric with 6 state-of-the-art metrics: the original single-scale iCID [5], the feature similarity index (FSIMc) [9], the PSNR-HA [10], the SSIM [4], the MS-SSIM [8], and the visual information fidelity index (VIF) [11]. We compared them over the following databases, which are widely used in the image quality community: TID2013 [2], CSIQ [12], LIVE (release 2) [1]. In contrast to CSIQ and LIVE, TID2013 includes chromatic distortions and is, therefore, of particular interest in our study. Note that we assumed a visual resolution of 20 cycles/degree for TID2013 and LIVE, whereas the best results on CSIQ were achieved for a resolution of 16 cycles/degree. For comparison, we measured the Spearman Rank Order Correlation Coefficient (SROCC) with Mean Opinion Scores. Our results show that the multi-scale version of iCID performs better than the single-scale one on each database. We observe especially that the MS-iCID metric yields the best correlation with human judgment on TID2013 with a SROCC of 0.8610. This is particularly noteworthy given that, unlike the current best-performing metric (FSIMc), MS-iCID was not trained on any database in particular. On the same database, other metrics yield SROCC of 0.8510 (FSIMc), 0.7859 (MSSIM), 0.7792 (PSNR-HA), 0.6930 (single scale iCID) and lower values for SSIM and VIF. Because MS-iCID is designed principally for chromatic distortions, it does not perform significantly better than the state-of-the-art metrics on the LIVE database, which features distortions that are mostly achromatic. Nevertheless, we note that MS-iCID ranks first when it comes to measuring Gaussian blur, JPEG2000 compression and particularly JPEG transmission errors (with minimum improvements of 0.0327 SROCC over the state-of-the-art) in TID2013.

REFERENCES

[1] Sheikh, H. R., Sabir, M. F., and Bovik, A. C., “A statistical evaluation of recent full reference image quality assessment algorithms”, IEEE Transactions on Image Processing 15(11), 3440-3451 (2006).

[2] Ponomarenko, N., Ieremeiev, O., Lukin, V., Egiazarian, K., Jin, L., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., et al., “Color image database tid2013: Peculiarities and preliminary results”, in 4th European Workshop on Visual Information Processing EUVIP201, (2013).

[3] Lissner, I., Preiss, J., Urban, P., Scheller Lichtenauer, M., and Zolliker, P., “Image-difference prediction: From grayscale to color”, IEEE Transactions on Image Processing 22(2), 435-446 (2013).

[4] Wang, Z., Bovik, A., Sheikh, H., and Simoncelli, E., “Image quality assessment: From error visibility to structural similarity”, IEEE Transactions on Image Processing 13(4), 600-612 (2004).

[5] Preiss, J., Fernandes, F., and Urban, P., “Color-image quality assessment: From prediction to optimization”, IEEE Transactions on Image Processing 23(3), 1366-1378 (2013).

[6] Reinhard, E., Khan, E. A., Akyz, A. O., and Johnson, G. M., “Color imaging: fundamentals and applications”, AK Peters, Ltd. (2008).

[7] Lissner, I. and Urban, P., “Toward a uni fied color space for perception-based image processing”, IEEE Transactions on Image Processing 21(3), 1153-1168 (2012).

[8] Wang, Z., Simoncelli, E. P., and Bovik, A. C., “Multiscale structural similarity for image quality assessment”, in Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on, 2, 1398-1402, IEEE (2003).

[9] Zhang, L., Zhang, L., Mou, X., and Zhang, D., “Fsim: a feature similarity index for image quality assessment”, IEEE Transactions on Image Processing 20(8), 2378-2386 (2011).

[10] Ponomarenko, N., Ieremeiev, O., Lukin, V., Egiazarian, K., and Carli, M., “Modified image visual quality metrics for contrast change and mean shift accounting”, in CAD Systems in Microelectronics (CADSM), 2011 11th International Conference The Experience of Designing and Application of, 305-311, IEEE (2011).

[11] Sheikh, H. and Bovik, A., “Image information and visual quality,” IEEE Transactions on Image Processing 15(2), 430-444 (2006). ---------- [12] Larson, E. C. and Chandler, D. M., “Most apparent distortion: full-reference image quality assessment and the role of strategy,” Journal of Electronic Imaging 19(1), 011006 (2010).

9396-39, Session 10

Image quality evaluation of LCDs based on novel RGBW sub-pixel structure

Sungjin Kim, Dong-Woo Kang, Jinsang Lee, Jaekyeom Kim, Yongmin Park, Taeseong Han, Sooyeon Jung, Jang Jin Yoo, Moojong Lim, Jongsang Baek, LG Display (Korea, Republic of)

Introduction

Recently, many TV manufacturers have released various sizes and high resolution of TVs. In case of the higher resolution LCD TVs in limited size, improvement of panel transmittance is required to increase luminance and reduce power consumption.

Adding a white sub-pixel to RGB sub-pixel structure is one of well-known solutions to improve panel transmittance since there is no color filter on white sub-pixel [1]. Figure 1 shows sub-pixel structures of RGB and RGBW LCD with the same resolution [2]. Although there are many methods to realize RGBW sub-pixel structure, RGBW type3 is more suitable method because of higher backplane compatibility with RGB type. In WRGB LCD, the area of RGB sub-pixel decrease to 75%, but the remaining 25% is filled with white sub-pixel. Thus, RGBW LCD has 50% higher white luminance and 25% lower primary color luminance compared with RGB LCD [3].

There are several advantages using white sub-pixels in RGBW LCD. First, the RGBW LCD can be easily applied to HDR (High Dynamic Range) display or power saving mode because of high luminance. Second, it can express improved edges using only white sub-pixel, and it is superior to RGB LCD in terms of image quality [4]. Third, combination of RGBW panel and WCG (wide color gamut) BLU (backlight unit) can not only enhance color, but also maintain high luminance.

In this paper, we focused on image quality of RGBW and RGB LCD, and evaluated how strength and weakness of RGBW and RGB LCD affect image quality under TV viewing condition. In order to evaluate them, a reference video clip was prepared considering program frequency from TV broadcast videos in Korea. As a result of the reference video analysis, RGBW LCD is likely to improve image quality more because most of colors are located around white point. In addition, IEC-62087 was also analyzed in terms of color distribution, and most of colors are located around white point as can be seen in Figure 2.

Experimental Setup

Table 1 summarizes the optical performance of a reference RGB, RGBW and RGBW + WCG BLU LCD. Two sessions of visual experiment were conducted. First, image-quality attributes such as 'Brightness', 'Naturalness', 'Colorfulness' and 'Contrast' additional to overall image quality were evaluated using 24 test still images. Second, overall image quality using 17 video clips from the reference video was evaluated.

Table 1. Comparison of RGB, RGBW and RGBW + WCG BLU LCD

LCD RGB RGBW RGBW + WCG BLU

Resolution 3840 x 2160 (Ultra High Definition)

Color gamut

(compared to BT.709 gamut) 100% 100% 123%

Luminance 400 nits 600 nits 600 nits

Twenty observers participated in the subjective evaluation, and all of the observers were image quality engineers in display field aged 25 to 40. The viewing distance was set to 3H, where H refers to the height of the display. The illumination condition was about 200 lux which is common illumination of general building areas by ISO 9241-307 standard. Observers viewed and evaluated images of three different LCDs according to the following questions. The answer of the question was judged by 9 scales as follows.

Session 1. How would you rate the preference of brightness (naturalness, colorfulness, contrast)?

Session 2. How would you rate the overall image quality?

9:Like Extremely 8:Like very much 7:Like moderately 6:Like slightly 5:Neither like nor dislike

4:Dislike slightly 3:Dislike moderately 2:Dislike very much 1:Dislike extremely

Result and Analysis

RGBW LCD obtained higher scores than RGB LCD except for 'Colorfulness' as shown in Figure 3(a) and (b). Since RGBW LCD has higher luminance around achromatic area, 'Brightness' and 'Contrast' test could be rated higher. Furthermore, for 'Naturalness', natural colors such as skin, sky or grass are also located near to achromatic color. Thus, it was though that RGBW LCD obtains higher 'Naturalness' scores. Meanwhile, RGBW LCD's 'Colorfulness' was assessed not higher than RGB LCD because luminance values of RGB primary colors of RGBW LCD are 25% lower than RGB LCD. Consequently, overall image quality of RGBW LCD was rated higher than RGB LCD.

In order to enhance RGBW LCD's colorfulness compared to RGB LCD's, it can be recommended that backlight is replaced with a wide color gamut backlight (WCG BL). If the peak luminance value is the same, 'Colorfulness' could be improved although the luminance values of RGB primary colors do not increase. As can be seen in Figure 3 (a) and (b), RGBW LCD using WCG BL shows better performance than RGBW LCD. In detail, 'Colorfulness' rated worse in RGBW LCD was significantly improved additional to 'Brightness' and 'Contrast'.

9396-41, Session 10

Is there a preference for linearity when viewing natural images?

David Kane, Marcelo Bertamio, Universitat Pompeu Fabra (Spain)

No Abstract Available

Conference 9397: Visualization and Data Analysis 2015

Monday - Wednesday 9–11 February 2015

Part of Proceedings of SPIE Vol. 9397 Visualization and Data Analysis 2015

9397-1, Session 1

An evaluation-guided approach for effective data visualization on tablets

Peter S. Games, Boise State Univ. (United States); Alark Joshi, Boise State Univ. (United States) and Univ. of San Francisco (United States)

There is a rising trend of data analysis and visualization tasks being performed on a tablet device. Apps with interactive data visualization capabilities are available for a wide variety of domains. We investigate whether users grasp how to effectively interpret and interact with visualizations. We conducted a detailed user evaluation to study the abilities of individuals with respect to analyzing data on a tablet through an interactive visualization app.

Based upon the results of the user evaluation, we find that most subjects performed well at understanding and interacting with simple visualizations, specifically tables and line charts. A majority of the subjects struggled with identifying interactive widgets, recognizing interactive widgets with overloaded functionality, and understanding visualizations which do not display data for sorted attributes. Based on our study, we identify guidelines for designers and developers of mobile data visualization apps that include recommendations for effective data representation and interaction.

9397-2, Session 1

Plugin free remote visualization in the browser

Georg Tamm, Philipp Slusallek, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (Germany)

Today, users access information and rich media from anywhere using the web browser on their desktop computers, tablets or smartphones. But the web evolves beyond media delivery. Interactive graphics applications like visualization or gaming become feasible as browsers advance in the functionality they provide. However, to deliver large-scale visualization to thin clients like mobile devices, a dedicated server component is necessary. Ideally, the client runs directly within the browser the user is accustomed to, requiring no installation of a plugin or native application. In this paper, we present the state-of-the-art of technologies which enable plugin free remote rendering in the browser. Further, we describe a remote visualization system unifying these technologies. The system transfers rendering results to the client as images or as a video stream. We utilize the upcoming World Wide Web Consortium (W3C) conform Web Real-Time Communication (WebRTC) standard, and the Native Client (NaCl) technology built into Chrome, to deliver video with low-latency.

9397-3, Session 1

Ensemble visual analysis architecture with high mobility for large-scale critical infrastructure simulations

Todd Eaglin, Xiaoyu Wang, William Ribarsky, William J. Tolone, The Univ. of North Carolina at Charlotte (United States)

Nowhere is the need to understand large heterogeneous datasets more important than in disaster monitoring and emergency response, where critical decisions have to be made in a timely fashion and the discovery of important events requires an understanding of a collection of complex simulations. To gain enough insights for actionable knowledge, the development of models and analysis of modeling results usually requires that models be run many times so that all possibilities can be covered. Central to the goal of our research is, therefore, the use of ensemble visualization of a large scale simulation space to appropriately aid decision makers in reasoning about infrastructure behaviors and vulnerabilities in support of critical infrastructure analysis. This requires the bringing together of computing-driven simulation results with the human decision-making process via interactive visual analysis. We have developed a general critical infrastructure simulation and analysis system for situationally aware emergency response during natural disasters. Our system demonstrates a scalable visual analytics infrastructure with mobile interface for analysis, visualization and interaction with large-scale simulation results in order to better understand their inherent structure and predictive capabilities. To generalize the mobile aspect, we introduce mobility as a design consideration for the system. The utility and efficacy of this research has been evaluated by domain practitioners and disaster response managers.

9397-4, Session 2

OSNAP! Introducing the open semantic network analysis platform

Peter J. Radics, Nicholas F. Polys, Shawn P. Neuman, William H. Lund, Virginia Polytechnic Institute and State Univ. (United States)

Graph visualization continues to be a major challenge in the field of information visualization, meanwhile gaining importance due to the power of graph-based formulations across a wide variety of domains from knowledge representation to network flow, bioinformatics, and software optimization.

We present the Open Semantic Network Analysis Platform (OSNAP), an open-source visualization framework designed for the flexible composition of 2D and 3D graph layouts.

Analysts can filter and map a graph's attributes and structural properties to a variety of geometric forms including shape, color, and 3D position.

Using the Provider Model software engineering pattern, developers can extend the framework with additional mappings and layout algorithms.

We demonstrate the framework's flexibility by applying it to two separate domain ontologies and finally outline a research agenda to improve the value of semantic network visualization for human insight and analysis.

9397-5, Session 3

iGraph: a graph-based technique for visual analytics of image and text collections

Yi Gu, Chaoli Wang, Univ. of Notre Dame (United States); Jun Ma, Robert J. Nemiroff, Michigan Technological Univ. (United States); David L. Kao, NASA Ames Research Ctr. (United States)

In our daily lives, images and texts are among the most commonly found data which we need to handle. We present iGraph, a graph-based approach for visual analytics of large image and text collections. Given such a collection, we compute the similarity between images, the distance between texts, and the connection between image and text to construct iGraph, a compound graph representation which encodes the underlying relationships among these images and texts. To enable effective visual navigation and comprehension of iGraph with tens of thousands of nodes and hundreds of millions of edges, we present a progressive solution that offers collection

overview, node comparison, and visual recommendation. Our solution not only allows users to explore the entire collection with representative images and keywords, but also supports detail comparison for understanding and intuitive guidance for navigation. For performance speedup, multiple GPUs and CPUs are utilized for processing and visualization in parallel. We experiment with two image and text collections and leverage a cluster driving a display wall of nearly 50 million pixels. We show the effectiveness of our approach by demonstrating experimental results and conducting a user study.

9397-6, Session 3

Exploring hierarchical visualization designs using phylogenetic trees

Shaomeng Li, Univ. of Oregon (United States); R. Jordan Crouser, MIT Lincoln Lab. (United States); Garth Griffin, Tufts Univ. (United States); Connor Gramazio, Brown Univ. (United States); Hans-Joerg Schulz, Univ. Rostock (Germany); Hank Childs, Univ. of Oregon (United States); Remco Chang, Tufts Univ. (United States)

Ongoing research on information visualization has produced an ever-increasing number of visualization designs.

Despite this activity, limited progress has been made in categorizing this large number of information visualizations.

This makes understanding their common design features challenging, and obscures the yet unexplored areas of novel designs. With this work, we provide categorization from an evolutionary perspective, leveraging a computational model to represent evolutionary processes, the phylogenetic tree. The result — a phylogenetic tree of a visualization design corpus — enables better understanding of the various design features of information visualizations, and further illuminates the space in which the visualizations lie, through support for interactive clustering and novel design suggestions. We demonstrate these benefits with our software system, where a corpus of two-dimensional hierarchical visualization designs is constructed into a phylogenetic tree.

This software system supports visual interactive clustering and suggesting for novel designs; the latter capacity is also demonstrated via collaboration with an artist who sketched new designs using our system.

9397-7, Session Key

The Palomar transient factory *(Keynote Presentation)*

Peter E Nugent, Lawrence Berkeley National Lab. (United States) and University of California, Berkeley (United States); Yi Cao, Caltech (United States); Mansi Kasliwal, The Carnegie Observatories (United States)

Astrophysics is transforming from a data-starved to a data-swamped discipline, fundamentally changing the nature of scientific inquiry and discovery. New technologies are enabling the detection, transmission, and storage of data of hitherto unimaginable quantity and quality across the electromagnetic, gravity and particle spectra. The observational data obtained during this decade alone will supersede everything accumulated over the preceding four thousand years of astronomy. Currently there are 4 large-scale photometric and spectroscopic surveys underway, each generating and/or utilizing hundreds of terabytes of data per year. Some will focus on the static universe while others will greatly expand our knowledge of transient phenomena.

Maximizing the science from these programs requires integrating the processing pipeline with high-performance computing resources. These are coupled to large astrophysics databases while making use of machine learning algorithms with near real-time turnaround. Here I will present an overview of one of these programs, the Palomar Transient Factory (PTF). I will cover the processing and discovery pipeline we developed at LBNL and NERSC for it and several of the great discoveries made during the 4 years of observations with PTF.

9397-8, Session 4

Emotion-prints: interaction-driven emotion visualization on multi-touch interfaces

Daniel Cernea, Technische Univ. Kaiserslautern (Germany) and Linnaeus Univ. (Sweden); Christopher Weber, Achim Ebert, Technische Univ. Kaiserslautern (Germany); Andreas Kerren, Linnaeus Univ. (Sweden)

Emotions are one of the unique aspects of human nature, and sadly at the same time one of the elements that our technological world is failing to capture and consider due to their subtlety and inherent complexity. But with the current dawn of new technologies that enable the interpretation of emotional states based on techniques involving facial expressions, speech and intonation, electrodermal response (EDS) and brain-computer interfaces (BCIs), we are finally able to access real-time user emotions in various system interfaces. In this paper we introduce emotion-prints, an approach for visualizing user emotional valence and arousal in the context of multi-touch systems. Our goal is to offer a standardized technique for representing user affective states in the moment when and at the location where the interaction occurs in order to increase affective self-awareness, support awareness in collaborative and competitive scenarios, and offer a framework for aiding the evaluation of touch applications through emotion visualization. We show that emotion-prints are not only independent of the shape of the graphical objects on the touch display, but also that they can be applied regardless of the acquisition technique used for detecting and interpreting user emotions. Moreover, our representation can encode any affective information that can be decomposed or reduced to Russell's two-dimensional space of valence and arousal. Our approach is enforced by a BCI-based user study and a follow-up discussion of advantages and limitations.

9397-9, Session 5

GPU surface extraction using the closest point embedding

Mark Kim, Charles Hansen, The Univ. of Utah (United States) and Scientific Computing and Imaging Institute (United States)

Isosurface extraction is a fundamental technique used for both surface reconstruction and mesh generation. One method to extract well-formed isosurfaces is a particle system; unfortunately, particle systems can be slow. In this paper, we introduce an enhanced parallel particle system that uses the closest point embedding as the surface representation to speed-up the particle system for isosurface extraction. The closest point embedding is used in the Closest Point Method (CPM), a technique that uses a standard three dimensional numerical PDE solver on two dimensional embedded surfaces. To fully take advantage of the closest point embedding, it is coupled with a Barnes-Hut tree code on the GPU. This new technique produces well-formed, conformal unstructured triangular and tetrahedral meshes from labeled multi-material volume datasets. Further, this new parallel implementation of the particle system is faster than any known methods for conformal multi-material mesh extraction. The resulting speedups gained in this implementation can reduce the time from labeled data to mesh from hours to minutes and benefits users, such as bioengineers, who employ triangular and tetrahedral meshes.

9397-10, Session 5

Advanced texture filtering: a versatile framework for reconstructing multi-dimensional image data on heterogeneous architectures

Stefan Zellmann, Yvonne Percan, Ulrich Lang, Univ. zu Köln (Germany)

Reconstruction of 2-d image primitives or of 3-d volumetric primitives is one of the most common operations performed by the rendering components of modern visualization systems. Because this operation is often aided by GPUs, reconstruction is typically restricted to first-order interpolation. With the advent of in situ visualization, the assumption that rendering algorithms are in general executed on GPUs is however no longer adequate. We thus propose a framework that provides versatile texture filtering capabilities: up to third-order reconstruction using various types of cubic filtering and interpolation primitives; cache-optimized algorithms that integrate seamlessly with GPGPU rendering or with software rendering that was optimized for cache-friendly "Structure of Array" (SoA) access patterns; a memory management layer (MML) that gracefully hides the complexities of extra data copies necessary for memory access optimizations such as swizzling, for rendering on GPGPUs, or for reconstruction schemes that rely on pre-filtered data arrays. We prove the effectiveness of our software architecture by integrating it into and validating it using the open source direct volume rendering (DVR) software DeskVOX.

9397-11, Session 5

A client-server view-dependent isosurfacing approach with support for local view changes

Matthew Couch, Timothy S. Newman, The Univ. of Alabama in Huntsville (United States)

A new approach for view-dependent isosurfacing on volumetric data is described. The approach is designed for client-server environments where the client's computational capabilities are much more limited than those of the server and where the network between the two features bandwidth limits, for example 802.11b wireless. Regions of the dataset that contain no visible part of the isosurface are determined on the server, using an approximate isosurface silhouette and octree-driven processing. The visible regions of interest in the dataset are then transferred to the client for isosurfacing. The client also receives additional components of the volume to enable rapid generation of renderings at new viewpoints with limited additional data transfer. Experimental results for application of the approach to volumetric data are also presented.

9397-12, Session 6

Comparative visualization of protein conformations using large high resolution displays with gestures and body tracking

Matthew Marangoni, Thomas Wischgoll, Wright State Univ. (United States)

Automatically identifying protein conformations can yield multiple candidate structures. Potential candidates are examined further to cull false positives. Individual conformations and the collection are compared when seeking flaws. Desktop displays are ineffective due to limited size and resolution. Thus a user must sacrifice large scale content by viewing the micro level with high detail or view the macro level while forfeiting small details.

We address this ultimatum by utilizing multiple, high resolution displays. Using 27, 50", high resolution displays with active, stereoscopic 3D, and modified virtual environment software, each display presents a protein users can manipulate. Such an environment enables users to gain extensive insight both at the micro and macro levels when performing structural comparisons among the candidate structures. Integrating stereoscopic 3D improves the user's ability to judge conformations spatial relationships. In order to facilitate intuitive interaction, gesture recognition as well as body tracking are used. The user is able to look at the protein of interest, select a modality via a gesture, and the user's motions provide intuitive navigation functions such as panning, rotating, and zooming.

Using this approach, users are able to perform protein structure comparison through intuitive controls without sacrificing important visual details at any scale.

9397-13, Session 6

FuryExplorer: visual-interactive exploration of horse motion capture data

Nils Wilhelm, Anna Vögele, Univ. Bonn (Germany); Rebeka Zsoldos, Univ. für Bodenkultur Wien (Austria); Theresia Licka, Univ. für Bodenkultur Wien (Austria) and Veterinaermedizinische Univ. Wien (Austria); Björn Krüger, Univ. Bonn (Germany); Jürgen Bernard, Fraunhofer-Institut für Graphische Datenverarbeitung (Germany)

The analysis of equine motion has had a long tradition in the past of human mankind. Equine biomechanics aims at detecting characteristics of horses indicative of good performance. Especially, veterinary medicine gait analysis plays an important role in diagnostics and in the emerging research of long-term effects of athletic exercises. More recently, the incorporation of motion capture technology contributed to an easier and faster analysis, with a trend from mere observation of horses towards the analysis of multivariate time-oriented data. However, due to the novelty of this topic being raised within an interdisciplinary context, there is yet a lack of visual-interactive interfaces to facilitate time series data analysis and information discourse for the veterinary and biomechanics communities. In this design study, we bring visual analytics technology into the respective domains, which, to our best knowledge, was never approached before. Based on requirements developed in the domain characterization phase, we present a visual-interactive system for the exploration of horse motion data. The system provides multiple views which enable domain experts to explore frequent poses and motions, but also to drill down to interesting subsets, possibly containing unexpected patterns. We show the applicability of the system in two exploratory use cases, one on the comparison of different gait motions, and one on the analysis of lameness recovery. Finally, we present the results of a summative user study conducted in the environment of the domain experts. The overall outcome was a significant improvement in effectiveness and efficiency in the analytical workflow of the domain experts.

9397-14, Session Key

Some difficult visualization problems: big science, big computer systems, and big data *(Keynote Presentation)*

Kenneth I. Joy, Univ. of California, Davis (United States)

No Abstract Available

9397-15, Session 7

Weighted maps: treemap visualization of geolocated quantitative data

Mohammad Ghoniem, Maël Cornil, Bertjan Broeksema, Mickaël Stefas, Benoît Otjacques, Ctr. de Recherche Public - Gabriel Lippmann (Luxembourg)

A wealth of census data relative to hierarchical administrative subdivisions are now available. It is therefore desirablehierarchical data visualization techniques, to offer a spatially consistent representation of such data. This paper focusesa widely used technique for hierarchical data, namely treemaps. In particular on a specific family of treemaps, designed to take into account spatial constraints in the layout, called Spatially Dependent Treemap (SDT). The contributions ofpaper are threefold. First, present "Weighted Maps", a novel SDT layout algorithm and discuss the algorithmic differences with the other state-of-the-art SDT algorithms. Second, we present the quantitative results and analyses of a numbermetrics that where used to assess the quality of the resulting layouts. The analyses are illustrated with figures generated from various data sets. Third, we show that the Weighted Maps algorithm offers a significant advantage for the layoutlarge flat cartograms and multi-level hierarchies having a large branching factor.

9397-16, Session 8

Evaluating lossiness and fidelity in information visualization

Richard Brath, Ebad Banissi, London South Bank Univ. (United Kingdom)

We describe an approach to measure visualization fidelity for encoding of data to visual attributes based on the number of unique levels that can be perceived; and a summarization across multiple attributes to compare relative lossiness across visualization alternatives. These metrics can be assessed at design time in order to compare the lossiness of different visualizations; and examples are provided showing the application of these metrics to two different visualization design scenarios. Limitations and dependencies are noted along with recommendations for other metrics that can be used in conjunction with fidelity and lossiness to gauge effectiveness at design-time.

9397-22, Session PTues

Reactive data visualizations

Curran Kelleher, Haim Levkowitz, Univ. of Massachusetts Lowell (United States)

Constructing interactive visualizations is a complex task. The task becomes even more complex when multiple visualizations are presented and linked together through interactions. The issue at the core of interactive visualization and linked views is management of complex data flows and update patterns. Even with the wealth of visualization toolkits and libraries that exist today, there is a need for an abstraction that addresses these core issues. The Model View Controller paradigm can be combined with functional reactive programming to enable straightforward creation of reactive systems based on data flow graphs. Consider visualizations such as the bar chart, line chart, stacked area chart, parallel coordinates and choropleth map. These visualizations share many underlying primitives such as scales, axes, margins and labels. Interactive forms of these visualizations also share interaction techniques for selecting visual marks such as rectangular brushing, hovering, clicking, panning and zooming. These visualization primitives can be encapsulated as data dependency subgraphs. In this paper we demonstrate the effectiveness of our proposed approach in several visualization examples including multiple linked views. An open source proof of concept implementation of our reactive visualization approach is available on GitHub at https://github.com/curran/model and https://github.com/curran/reactivis .

9397-24, Session PTues

Visualization and classification of physiological failure modes in ensemble hemorrhage simulation

Song Zhang, Mississippi State Univ. (United States); William A. Pruett, Robert Hester, The Univ. of Mississippi Medical Ctr. (United States)

In an emergency situation such as hemorrhage, doctors need to predict which patients need immediate treatment and care. This task is difficult because of the diverse response to hemorrhage in human population. Ensemble physiological simulations provide a means to sample a diverse range of subjects and may have a better chance of containing the correct solution. However, to reveal the patterns and trends from the ensemble simulation results is a challenging task. We have developed a visualization framework for ensemble physiological simulations. The visualization helps users identify trends among ensemble members, classify ensemble member into subpopulations for analysis, and provide prediction to future events by matching a new patient's data to existing ensembles. We demonstrated the effectiveness of the visualization on simulated physiological data. The lessons learned here can be applied to clinically-collected physiological data in the future.

9397-25, Session PTues

Time-synchronized visualization of arbitrary data streams

Paul Kolano, NASA Ames Research Ctr. (United States)

Savors is a visualization framework that supports the ingestion of data streams created by arbitrary command pipelines. Multiple data streams can be shown synchronized by time in the same or different views, which can be arranged in any layout. These capabilities combined with a powerful parallelization mechanism and interaction models already familiar to administrators allows Savors to display complex visualizations of data streamed from many different systems with minimal effort. This paper presents the design and implementation of Savors and provides example use cases that illustrate many of the supported visualization types and the commands used to generate them.

9397-26, Session PTues

3D chromosome rendering from Hi-C data using virtual reality

Yixin Zhu, Siddarth Selvaraj, Philip Weber, Jennifer Fang, Jürgen P. Schulze, Bing Ren, Univ. of California, San Diego (United States)

No Abstract Available

9397-27, Session PTues

Visualizing uncertainty of river model ensembles

John van der Zwaag, Song Zhang, Robert J. Moorhead, Mississippi State Univ. (United States); David Welch, Lower Mississippi River Forecast Ctr. (United States); Jamie Dyer, Mississippi State Univ. (United States)

Ensembles are an important tool for researchers to provide accurate forecasts and proper validation of their models. To accurately analyze and understand the ensemble data, it is important that researchers clearly and efficiently visualize the uncertainty of their model output. In this paper, we present several methods for visualizing uncertainty in 1D river model ensembles. 2D and 3D inundation maps are generated by combining the 1D river model output with high-resolution digital elevation model data. We use the strengths of commonly used techniques for analyzing statistical data, and we apply them to the 2D and 3D visualizations of inundation maps. The resulting visualizations give researchers an easy method to quickly identify the areas of highest probability of inundation while also seeing the entire range of the ensemble output. It also allows forecasters to generate inundation maps to clearly show the general public the areas that are more likely to be in danger of flooding.

9397-28, Session PTues

Remote visualization system based on particle based volume rendering

Takuma Kawamura, Yasuhiro Idomura, Hiroko N. Miyamura, Hiroshi Takemiya, Japan Atomic Energy Agency (Japan); Naohisa Sakamoto, Koji Koyamada, Kyoto Univ. (Japan)

In this paper, we propose a novel remote visualization system based on particle-based volume rendering (PBVR) [Sakamoto et al, Computers&Graphics 34, 34 (2010)], which enables interactive analyses of extreme scale volume data located on remote computing systems.

The remote PBVR system consists of Server, which generates particles (rendering primitives), and Client, which processes volume rendering, and the particles are transferred from Server to Client.

The size of particle data is determined only by visualization parameters such as the image resolution and the transfer function, and becomes significantly smaller than the original volume data.

Depending on network bandwidth, the level of detail of images is flexibly controlled to attain high frame rates.

Server is highly parallelized on various parallel platforms with General Purpose Graphic Processing Units or multi-core CPUs using either a hybrid MPI-CUDA programing model or a hybrid MPI-OpenMP programing model.

The mapping process is accelerated by two orders of magnitudes compared with a single CPU, and structured and unstructured volume data with $sim$$10^8$ cells is processed within a few seconds.

Compared with commodity Client/Server visualization tools, the total processing cost of remote scientific visualization is dramatically reduced by using the remote PBVR system.

9397-17, Session 9

An image-space Morse decomposition for 2D vector fields

Guoning Chen, Univ. of Houston (United States); Shuyu Xu, Univ. of Houston (United States)

Morse decompositions have been proposed to compute and represent the topological structure of steady vector fields. Compared to the conventional differential topology, Morse decomposition and the resulting Morse Connection Graph (MCG) is numerically stable. However, the granularity of the original Morse decomposition is constrained by the resolution of the underlying spatial discretization, which typically results in non-smooth representation. In this work, an Image-Space Morse decomposition (ISMD) framework is proposed to address this issue. Compared to the original method, ISMD first projects the original vector field onto an image plane, then computes the Morse decomposition based on the projected field with pixels as the smallest elements. Thus, pixel-level accuracy can be achieved. This ISMD framework has been applied to a number of synthetic and real-world steady vector fields to demonstrate its utility. The performance of the ISMD is carefully studied and reported. Finally, with ISMD an ensemble Morse decomposition can be studied and visualized, which is shown useful for visualizing the stability of the Morse sets with respect to the error introduced in the numerical computation and the perturbation to the input vector fields.

9397-18, Session 9

Subsampling-based compression and flow visualization

Alexy Agranovsky, Univ. of California, Davis (United States) and Lawrence Berkeley National Lab. (United States); David Camp, Lawrence Berkeley National Lab. (United States); Kenneth I. Joy, Univ. of California, Davis (United States); Hank Childs, Univ. of Oregon (United States) and Lawrence Berkeley National Lab. (United States)

As computational capabilities increasingly outpace disk speeds on leading supercomputers, scientists will, in turn, be increasingly unable to save their simulation data at its native resolution. One solution to this problem is to compress these data sets as they are generated and visualize the compressed results afterwards. We explore this approach, specifically subsampling velocity data and the resulting errors for particle advection-based flow visualization. We compare three techniques: random selection of subsamples, selection at regular locations corresponding to multi-resolution reduction, and introduce a novel technique for informed selection of subsamples. Furthermore, we explore an adaptive system which exchanges the subsampling budget over parallel tasks, to ensure that subsampling occurs at the highest rate in the areas that need it most. We perform supercomputing runs to measure the effectiveness of the selection and adaptation techniques. Overall, we find that adaptation is very effective, and, among selection techniques, our informed selection provides the most accurate results, followed by the multi-resolution selection, and with the worst accuracy coming from random subsamples.

9397-19, Session 9

A multi-resolution interpolation scheme for pathline based Lagrangian flow representations

Alexy Agranovsky, Harald Obermaier, Univ. of California, Davis (United States); Christoph Garth, Technische Univ. Kaiserslautern (Germany); Kenneth I. Joy, Univ. of California, Davis (United States)

Where the computation of particle trajectories in classic vector field representations includes computationally involved numerical integration, a Lagrangian representation in the form of a flow map opens up new alternative ways of trajectory extraction through interpolation. In our paper, we present a novel re-organization of the Lagrangian representation by sub-sampling a pre-computed set of trajectories into multiple levels of resolution, maintaining a bound over the amount of memory mapped by the file system. We exemplify the advantages of replacing integration with interpolation for particle trajectory calculation through a real-time, low

memory cost, interactive exploration environment for the study of flow fields. Beginning with a base resolution, once an area of interest is located, additional trajectories from other levels of resolution are dynamically loaded, densely covering those regions of the flow field that are relevant for the extraction of the desired feature. We show that as more trajectories are loaded, the accuracy of the extracted features converges to the accuracy of the flow features extracted from numerical integration with the added benefit of real-time, non-iterative, multi-resolution path and time surface extraction.

9397-20, Session 10

Enhancing multidimensional data projection using density-based motion

Ronak Etemadpour, Oklahoma State Univ. (United States); Angus G. Forbes, Univ. of Illinois at Chicago (United States)

The density of points within multidimensional clusters can impact the effective representation of distances and groups when projecting data from higher dimensions onto a lower dimensional space. This paper examines the use of motion to retain an accurate representation of the point density of clusters that might otherwise be lost when a multidimensional dataset is projected into a 2D space. We investigate how users interpret motion in 2D scatterplots and whether or not they are able to effectively interpret the point density of the clusters through motion. Specifically, we consider different types of density-based motion, where the magnitude of the motion is directly related to the density of the clusters. We conducted a series of user studies with synthetic datasets to explore how motion can help users in various multidimensional data analyses, including cluster identification, similarity seeking, and cluster ranking tasks. In a first user study, we evaluated the motions in terms of task success, task completion times, and subject confidence. Our findings indicate that, for some tasks, motion outperforms the static scatterplots; circular path motions in particularly give significantly better results compared to the other motions. In a second user study, we found that users were easily able to distinguish clusters with different densities as long the magnitudes of motion were above a particular threshold. Our results indicate that it may be effective to incorporate motion into visualization systems that enable the exploration and analysis of multidimensional data.

9397-21, Session 10

A survey and task-based quality assessment of static 2D colormaps

Jürgen Bernard, Fraunhofer-Institut für Graphische Datenverarbeitung (Germany) and Technische Univ. Darmstadt (Germany); Martin Steiger, Fraunhofer-Institut für Graphische Datenverarbeitung (Germany); Sebastian Mittelstädt, Univ. Konstanz (Germany); Simon Thum, Fraunhofer-Institut für Graphische Datenverarbeitung (Germany); Daniel A. Keim, Univ. Konstanz (Germany); Jörn Kohlhammer, Fraunhofer-Institut für Graphische Datenverarbeitung (Germany) and Technische Univ. Darmstadt (Germany)

Color is one of the most important visual variables since it can be combined with any other visual mapping to encode more information without using any additional space on the display. Encoding one or two dimensions with color is widely explored and discussed in the field. Also mapping multidimensional data to color is applied in a vast number of applications, either to indicate similar, or to discriminate between different elements or (multidimensional) structures on the screen. A variety of 2D colormaps is present in literature, holding large variances with respect to different perceptual aspects. Likewise, many of them have a different focus on the underlying data structure as a consequence of the various application tasks existing for multivariate data. Thus, a large design space for 2D colormaps exists which makes the development and the use of 2D colormaps cumbersome. According to our literature research, 2D colormaps have not been subject of in-depth quality assessment. Therefore, we present a survey of 2D colormaps as applied for information visualization and related fields. Moreover, we map seven quality assessment metrics for 2D colormaps to seven important tasks for multivariate data analysis. Finally, we present the quality assessment results of the 2D colormaps with respect to every seven analysis tasks, and contribute guidelines on which colormaps to select or to create for each analysis task.

Conference 9398: Measuring, Modeling, and Reproducing Material Appearance 2015

Monday - Tuesday 9–10 February 2015

Part of Proceedings of SPIE Vol. 9398 Measuring, Modeling, and Reproducing Material Appearance 2015

9398-1, Session Key

Generative appearance models in the perception of materials and their properties *(Keynote Presentation)*

Roland W. Fleming, Justus-Liebig-Univ. Giessen (Germany)

No Abstract Available

9398-2, Session 1

Imaging artwork in a studio environment for computer graphics rendering

Brittany D. Cox, Roy S. Berns, Rochester Institute of Technology (United States)

INTRODUCTION

A frequently used technique for capturing surface information is photometric stereo, assuming a Lambertian surface. A common practice to determine surface normal is to employ a three-light photometric stereo technique. Research in this area has focused on increasing the number of lights to improve the accuracy of the resulting normal map by allowing pixels that contain highlights or shadows to be excluded from calculations.

Berns, et al. employed the use of polarizers with a 4LI, or four-light imaging, setup to remove specular highlights. Cross polarization was used to eliminate first surface reflectance including specular highlights. This technique addressed the problems associated with highlights but did not directly eliminate the presence of shadows. A shadow was assumed to occur in only one light direction. That reduced the number of light directions to three since the signal was null.

The Berns, et al. method was implemented in a photography studio by adding two Xenon strobes to the usual set up of a pair of strobes placed on each side of the painting. Unfortunately, cross polarization resulted in images that appeared blurry. This paper describes the results of using a selection technique to remove highlights from images captured using the 4LI system, but without using cross-polarization.

IMAGING

Both the 4LI polarization technique and the new technique employed four Xenon strobes placed at 45° from the surface of the painting and 90° from each neighboring light along an annulus with the camera placed perpendicular to the surface of the painting. The only difference in the physical setup between the Berns method and the new technique is that the Berns, et al. method placed linear polarizers in front of each light and camera while the new method did not.

The procedure outlined by Berns, et al. was performed and repeated for both methods. Images were taken using each light sequentially of a uniform white background for flat fielding, a glossy black cue ball to define each light direction, a Xrite ColorChecker Classic for color calibration, and two acrylic paintings, one varnished and one unvarnished.

IMAGE ANALYSIS, RESULTS, AND DISCUSSION

Two maps are required for computer graphics rendering: diffuse albedo and surface normal. For the surface normal maps, the computational removal of highlights was achieved by implementing a thresholding method. This method assumed that a highlight is only present in a pixel for one direction at a time. In the thresholding equation:

if: imgA(x,y) > avgImg(x,y) + avgImg(x,y)*k,

then: imgA(x,y) is a highlight

x and y denote a pixel location, imgA is a single image taken from one of the lights directions, avgImg is the calculated average without the maximumum value of all four light directions, and k is a constant, accounting for any skewing that may have occurred if a pixel was in shadow. The constant k was 0.75, selected by visual evaluation. The goal was to avoid any unnecessary removal of pixels, thus preserving as much surface normal information while still eliminating highlights. Pixels above the threshold were assumed to be highlights and removed, resulting in three light directions for calculating the surface normal.

Once the highlights were computationally removed, a normal map was calculated using information from three or four lights for the appropriate pixel. These calculations, like those done by the 4LI method, were based on the technique proposed by Woodham. A comparison of these two normal maps indicated that the threshold method created normal maps with more variation in the Z direction. This is attributed to the increase in contrast present in these images that made the boundaries of impasto more pronounced. There is more high frequency detail in these maps. However, the normal maps resulting from the threshold method exhibited more noise while the cross-polarized images generated normal maps that provided smoother transitions in areas with changes in topography and reduced artifacts introduced by pixels in highlighted regions.

The second map required for computer-graphics rendering was the diffuse albedo (color) map. The diffuse color of the paintings was calculated by taking an average of each pixel excluding the maximum. Excluding the maximum removes the influence of highlights from the diffuse color. Using a simple average of all four cross polarized images, as presented in the Berns, et al. method, resulted in images with a decrease in contrast. This decrease in black resulted in the appearance of a less crisp image by comparison.

The final component to rendering, gloss, is addressed by setting the appropriate properties in Maya®. The diffuse images and their corresponding normal maps were imported into Maya® and assigned properties of a matte plastic, a preset material of the software. Matte plastic was chosen because it was the closest preset material with characteristics similar to dry acrylic paint. The index of refraction of the unvarnished painting was set to 1.40 and the index of refraction of the varnished painting was set to 1.52, the approximate index of refraction values of dry acrylic paint and an average varnish, respectively. The reflective glossiness property for the varnished painting was also changed to 0.600, compared to 0.500 for the unvarnished painting. The Fresnel Reflection ray-tracing option was used along with the mental ray rendering software plugin to generate realistic images that mimicked light interaction with the surface of each painting simulating a D65 spotlight 45° to the right of the center of the painting and a camera, or observer 45° to the left of the painting.

Qualitatively, there is little discernable difference between images rendered using the discussed techniques. Imaging paintings without cross polarization produced a rendered image of the paintings that compared very well in image sharpness and surface texture detail to a rendering generated from cross-polarized images.

CONCLUSIONS

A new imaging approach has been developed and implemented for capturing input data for computer graphics rendering of paintings. The approach improved previous research by eliminating the need for cross-polarization, reducing studio setup time, and improving image sharpness. Different thresholding and averaging methods for producing surface normal and diffuse albedo maps might produce images that are equivalent to so-called “beauty shots” where the photographer combines diffuse and direct lighting to maximize the most important properties of the painting when viewing a static image. The most important result is the possibility of having a single studio setup when imaging materials with appreciable micro- and macro-structure. Accentuating or diminishing structure becomes a dynamic tool, reducing the need for reshooting for different purposes.

9398-3, Session 1

Predictive rendering of composite materials: a multi-scale approach

Thomas Muller, Patrick Callet, Alexis Paljic, Philippe Porral, Mines ParisTech (France)

Predictive rendering of material appearance means going deep into the understanding of the physical interaction between light and matter and how these interactions are perceived by the human eye. In this paper we describe a new approach to the predictive rendering by relying on the multi-scale nature of the involved phenomena. Referring to recent work on physical modeling of complex materials, we show how to predict the aspect of a material based on its composition and its morphology up to the scale of a few atoms.

Over the past years we have been working towards improving the rendering process both for industry and cultural heritage. Car paints are a fantastic use case, especially when dealing with metallic, pearlescent or flip-flop coating effects. The spectral simulation of polychromy appearance in medieval architecture and sculpture brings to light that humans have for a long time investigated making complex optical materials - for beauty. In both cases we can observe a wide range of effects in the interaction between light and matter: reflection, refraction, scattering, chromatic dispersion, interferences (thin layer), birefringence, polarization, de-polarization, etc. The first part of this paper shows and illustrates with practical use cases, that none of these effects are negligible when focusing on predictive rendering. In this respect we start relating the history of the research of the pearlescent effect as an example of the human obsession with the magic of light. We follow with a short survey of the predictive rendering of composite materials focusing on the materials whose morphological structures are multi-scale (from a few atoms to a few millimeters: nanoparticles, thin layer, photonic structures, flakes, etc.).

The second part of this paper report our practical experience: The problems we have had using believable rendering and the success we have had taking into account the intrinsic property of light and matter. So, we detail our multi-scale approach to the predictive rendering of composite materials. At a scale far below the considered wavelength (<<380 nm in the case of visible light) we focus on acquiring the intrinsic parameters of the materials as optical constants. Indeed, at this physical scale all the materials can be characterized by their optical constants such as the complex dielectric function or the complex index of refraction (including anisotropic property of crystals). In the case of composite materials (small and not percolated inclusions) we consider the optical constants of the corresponding effective media. These can be measured with a spectroscopic ellipsometer, or calculated analytically (in simple cases with pre-defined shapes of inclusions), or computed iteratively by a homogeneization process (general case). We shall extensively compare the three methods and describe some cross validation results. Far above the wavelength, on a perfectly flat surface, we can directly use these optical constants within the Fresnel equations and predict accurately the aspect of a given material. Our approach is built on recent work in modeling the multi-scale dispersion of nanoparticles in a hematite coating. We use as reference hematite cubic crystals at nanoscale specially produced for these experiments. The phase function associated to the scattered radiance depends on the geometrical state of pigments and on their nature. Extrinsic parameters as pigment shapes and sizes also modify the optical appearance of the resulting composite material. Nature (n,k) and structure (diameter, kind of shapes: spherical, acicular, flakes, platelets, etc.) are always linked at any scale. At this stage a local illumination model is enough to render a simple scene composed by a flat panel (hematite coating) and a point light. On a rough surface we have to account for the angular resolution and spectral sensitivity of the observing optical system (human eye, camera, etc.). Small scale interactions between light and matter can be then integrated into a bidirectional reflectance distribution function. As the spatial resolution of the observed radiance is dependent on the observer position, we propose a multi-scale, spectral and polarized BRDF representation and a practical implementation using spherical wavelets.

In conclusion, we provide a solution for the two main scale steps in predictive rendering of composite materials. The “nano” to “micro” step is a physical boundary related to the considered wavelength and the “meso” to “macro” step is a physiological boundary related to the observer. Finally we explore some of those boundary problems, especialy when the scale of morphological structure of the material is close to the wavelength, and we give some directions to go deeper by predicting the appearance of a material based on its atomic structure.

9398-4, Session 1

Estimating reflectance property from multi-focus images by light field camera and its application

Norimichi Tsumura, Kaori Baba, Chiba Univ. (Japan); Shoji Yamamoto, Tokyo Metropolitan College of Industrial Technology (Japan); Masao Sambongi, Olympus Corp. (Japan)

?In this paper, we propose a method to estimate the reflectance property from multi-focus images for light source reflected on the object. The blurred information of the light source on the surface is expected to be the practical method to estimate the reflectance property, even though various methods are proposed to estimate the reflectance property. However, the degree of the blurred information will be changed with the position of focus in the camera. Therefore, we introduce the light field camera which can change the position of the focus after the image are captured. In this research, we choose the image where the light source is focused on the object surface. Based on the blurred information of the focused light source, we estimated the reflectance property of the object. The estimated reflectance property is applied to inverse rendering for auto appearance valance.

9398-5, Session 1

Experiments with a low-cost system for computer graphics material model acquisition *(Invited Paper)*

Holly E. Rushmeier, Yitzhak Lockerman, Luke Cartwright, David Pitera, Yale Univ. (United States)

There are many computer graphics applications that require visual simulation of the physical world. Applications include animation, games, virtual reality training simulators and architectural and industrial design. In these applications varying levels of accuracy for material appearance models are required. In many cases there is much greater need for models with visual richness than for models with a high level of numerical accuracy in individual spatial and angular measurements. An inexpensive acquisition system that can produce approximate data for complex materials would often be more useful in an application than access to a highly accurate instrument for making high spatial and angular resolution light scattering measurements. In this work we present the design of and initial tests of components in an inexpensive system for acquiring computer graphics material appearance models.

Visually rich materials are characterized by spatial variations of light scattering properties, small scale geometric variations and subsurface scattering. The state of the art for acquiring the data to model such materials is to use multi-light, multi-camera position systems that record the light scattered from a surface patch of material. The resulting data is referred to as a bidirectional texture function BTFs, originally introduced by Dana et al. in 1999. While capturing visual richness, raw BTFs are not suitable for graphics applications. Computer graphics requires models that are compact to reduce memory requirements. The models must also facilitate importance sampling to incorporate into an efficient light simulation scheme for rendering. Finally, models are most useful if they can be edited easy to achieve particular visual effects. Some approaches

for fitting analytic models to BTF data have made progress towards these requirements (e.g. Wu et al. 2011), but compactness and editabilty in particular are generally limited. In the design of a new system we take the approach that more efficient and compact models can be produced by fitting models to components of material appearance, rather than attempting to model all effects bundled together. Furthermore, separating components of appearance allows for more efficient adaptive sampling of individual materials to capture the components that are most critical to model for that particular material.

Components of appearance can be separated optically using the high spatial frequency lighting patterns introduced by Nayar et al. in 2007. Nayar et al. demonstrated that the first scattering or direct light from surfaces can be separated from multiple scattering or global light by illuminating surfaces with pairs of light patterns. They also demonstrated that using the direct light only in techniques for estimating geometry from surfaces are more accurate. Our system design concept is to use these observations to acquire data for subsurface scattering from the global data, and surface structure and surface reflectance from the direct light. The image of global light can be used to generate a map of the different BSSRDF characteristics of points on the sample surface. This map then can be used to adaptively sample positions on the surface to estimate the BSSRDF functions. The image of the direct light can be obtained for several different camera positions. The direct light data can be used to estimate both surface BRDFs and a map of the surface normal.

Our system design includes a small number of cameras and projectors. For inexpensive easily controlled cameras we use Raspberry Pi computers and cameras. For inexpensive pattern generation, we use miniature (i.e. fit in the palm of your hand) DLP projectors. The light levels recorded by the system can be calibrated by imaging patterns shown on a spectralon sample. Images of an X-rite color passport chart can bemused to calibrate color for the system. Each camera/project pair is clamped together, so that the camera view is nearly coincident with the point of original of the projection pattern. Relative positions of the camera/projector pairs can be calibrated by classic computer vision camera calibration. The camera exposure times are limited by, on the low end, the DLP refresh rate and on the high end by the Raspberry Pi control software. Within these limits, the cameras are used to capture multiple exposures of each pattern to be combined into high dynamic range images.

We first consider recovering the global lighting for estimating material subsurface scattering. Global lighting includes both surface subscattering and multiple scattering events on the sample surface. We note that these two effects have different signatures in the light pattern images. Subsurface scattering is at a peak at the edge of each lit image, and decays with distance from the edge. Surface multiple scattering effects occur at some distance from the scattering site, and don't show this drop-off. Our initial experiments have found that the two different effects show peak light levels for light patterns of different scales. We can use these observations to segment the an image of global lighting into subsurface and surface events. This segmentation can be used as a diagnostic to determine whether there is significant subsurface scattering that needs to be further probed and models. When subsurface scattering is detected, we can use the results of multiple pattern scales to determine the maximum radius of significant subsurface scattering. Next, for the subsurface scattering image, can be segmented into regions of different levels and spectra of scatter. This segmented image and the estimated radius of scattering effects can then be used to specify a pattern of spatially sparse regions to estimate the BSSRDF at isolated points. A map of BSSRDF can then be made using the segmented images and spatially sparse measurements.

Maps of the surface normals and surface bidirectional reflectance distribution function (BRDF) variations can be generated from the direct light images. Images lit from three directions (i.e. using three camera/projector pairs) can be used to compute the normals using photometric stereo. Basic photometric stereo assumes a Lambertian surface. Large deviations from this assumption can be detected by masking out pixels where the observed reflection exceeds the reflected light from the white spectralon. These high specular values can also be used to estimate surface normals -- e.g. a high value for an image when the camera and projection centers are approximately the same indicates that the normal is in the direction of the pair.

At each step in the material model estimation, multiple estimates can be made using images from a single camera and multiple projection directions. Variations in results for the different camera views indicates the reliability of the assumption of diffuse reflection and multiple scattering.

The appropriate evaluation of models from our system is comparison of a computer graphics rendering of the material with the physical sample. Our plan is to obtain models that can be rendered with the physically accurate Mitsuba renderer. By comparing renderings with the physical sample we can evaluate the success of our system, and its suitability for various classes of materials.

9398-6, Session 1

BTF Potts compound texture model

Michal Haindl, Vaclav Remes, Vojtech Havlicek, Institute of Information Theory and Automation (Czech Republic)

This paper introduces a method for modeling mosaic-like textures using a multispectral parametric Bidirectional Texture Function (BTF) compound Markov random field model (CMRF).

The primary purpose of our synthetic texture approach is to reproduce, compress, and enlarge a given measured texture image so that ideally both natural and synthetic texture will be visually indiscernible, but the model can be easily applied for BFT material editing.

The compound Markov random field model consist of several sub-models each having different characteristics along with an underlying structure model which controls transitions between these sub models. The proposed model uses the Potts random field for distributing local texture models in the form of analytically solvable wide-sense BTF Markovian representation for single regions among the fields of a mosaic.

The control field of the BTF-CMRF is generated by the Potts random field model build on top of the adjacency graph of a measured mosaic. The Potts MRF parameter is estimated from the adjacency graph of the mosaic approximated by the Voronoi diagram.

The Voronoi diagram was chosen due to its suitable representation of the intended class of the modeled man made textures and simultaneously due to the simplicity of its estimation and synthesis.

The compound random field synthesis combines the modified fast Swendsen-Wang Markov Chain Monte Carlo sampling of the hierarchical Potts MRF part with the fast and analytical synthesis of single regional MRFs.

The local texture regions (not necessarily continuous) are represented by an analytical BTF model which consists of single factors modeled by the adaptive 3D causal auto-regressive (3DCAR) random field model. The 3DCAR model can be analytically estimated as well as synthesized and thus to avoid a time consuming Monte Carlo sampling.

The proposed BTF-CMRF^{PV3AR} model is well suited to model various types of man made surfaces such as floor, textile, or stained glass random mosaics.

We have successfully tested the method mainly on man made surfaces such as different types of linoleum, smoothed stones, stone walls or stained glass and on selected natural materials such as chipped nacre.

The visual quality of the resulting complex synthetic textures generally surpasses the outputs of the previously published simpler BTF-MRF models.

Textures on which the algorithm will exhibit poor quality are regular textures or textures with fixed patterns such as most textile or knitted wool textures.

The model allows for seamless multispectral texture synthesis and enlargement with an extremely high compression rate independent of the-size of the-desired resulting texture.

The data needed to be stored is comprised of only several dozens of parameters.

Using a simple modification of the method we can use it for texture editing (by changing the-local texture models for several indexes of the-control field), we can use it for modeling BTF textures or even the synthesis of new, unmeasured textures by manually assigning the model's parameters.

9398-7, Session 1

Statistical Analysis of Bidirectional Reflectance Distribution Functions

Carlos Zubiaga Peña, INRIA Bordeaux (France); Laurent Belcour, Univ. de Montréal (Canada); Carles Bosch, Univ. de Girona (Spain); Adolfo Muñoz, Univ. de Zaragoza (Spain); Pascal Barla, INRIA Bordeaux (France)

Bidirectional Reflectance Distribution Functions (BRDFs) are commonly employed in Computer Graphics and Computer Vision to model opaque materials. Analitical BRDF models have the advantage of providing a small set of editable parameters, either based on observation or physical considerations. An alternative is to rely on data-driven BRDFs, which come from measured real-world materials. Such BRDF data exhibit more complex reflectance functions than analytical models; on the other hand, they are not readily editable.

What difference does it make in practice though, when material, object shape and environment illumination interact to yield a complex appearance?

A BRDF is a complex 4D function of both light and view directions, which reduces to a 3D function in the case of the isotropic materials that we consider here. When computing the radiance reaching the eye from a surface point, the function of interest further becomes two-dimensional since the view direction is then held fixed. Such a 2D BRDF slice acts as a filter on the local environment lighting. As a result, the difference between two types of BRDF are likely to appear as differences in filter properties (e.g., blur size, intensity, etc).

The goal of our work is to understand the statistical properties of such a filter as a function of viewing elevation. To this end, we have conducted a study of measured isotropic BRDFs where we have computed statistical moments for each viewing angle. For a given BRDF, our approach is to characterize its corresponding filter properties for incident to grazing viewing directions.

Our study is based on the MERL database, which provides 100 measured isotropic materials of various kinds (metals, plastics, fabrics, paints, etc).

Note that most BRDFs may be conceived as the sum of at least a diffuse component and a glossy/specular component, which are blended in MERL materials. This is a problem for our analysis, since moments are good descriptors of the shape of a function as long as it is unimodal. The first step of our process thus consists in separating a BRDF into a sum of a diffuse component and a specular component using a simple heuristic. More complex separation routines could be devised and are likely to yield more than two components, but this is out of the scope of this work. In practice, we have found by visual inspection that our heuristic provides good diffuse/specular separations for 42 out of the 100 MERL BRDFs, which we use for our study.

Another issue is that the hemisphere of directions, when parametrized by elevation and azimuth angles, is a periodic domain due to the azimuthal dimension. This is problematic for the computation of odd-order moments, hence we present an alternative planar parametrization of the hemisphere, aligned with the view direction. Using this view-dependent parametrization, we are then able to compute 2D raw moments up to order 4 for each viewing elevation and for each color channel.

Raw moments may then be combined to yield classic cumulants up to order 4: mean, variance, skewness and kurtosis, which grow in dimensionality with increasing order. The final step of our study is to make sense of what we call cumulant profiles, which express a given cumulant as a function of viewing elevation. To this end we fit low-parameter functions to each type of cumulant profile; these provide key insights in the filter properties corresponding to a given measured BRDF.

The first observation we draw from our study is that many of the cumulant profiles are close to zero. This is due to our choice of parametrization that emphasizes the symmetry across the scattering plane, observed in most materials (models and data). As a consequence, the mean lies on the horizontal axis, and there are no co-variance, co-skewness or co-kurtosis terms, which significantly simplifies filter properties.

A second important observation is that hemispherical clamping has a significant effect on odd-order cumulants for small viewing elevations. In particular, the mean slightly moves away from the peak reflectance, and skewness is progressively introduced at grazing angles. These filter properties are likely to affect the appearance of a material close to its silhouette.

Third, cumulants are correlated across both dimensions of our parametrization and across cumulant orders. For instance, profiles of variances along both dimensions always start at the same value, since at incidence a BRDF slice is radially symmetric. Coming toward grazing angles, variances seem to follow somewhat opposite behaviors, so much that their average is close to constant across elevation angles. This suggests that the area of the filter will approximately remain the same across viewing elevation, or that the amount of blur will be roughly preserved. We have also found the average of variances to be correlated to the mean, which is in part due to hemispherical clampling since BRDF slices of large variances are affected early on. This suggests that filters with smaller support will yield more sharp and distorted reflections, an effect which is particularly noticeable along silhouettes of shiny objects.

As a last observation, we have found that the main differences in color lie in the energy of a BRDF slice as a function of viewing elevation. In particular, some variations in hue and saturation toward grazing viewing angles are quite striking, but to our knowledge no material model has dealt with such effects.

We believe that our statistical analysis of measured BRDFs puts analytical and measured BRDFs in a new perspective. It reveals their effects as view-dependent filters; in particular, it emphasizes which effects from BRDF data will be sacrificed by a given choice of analytic BRDF model.

Our study also opens the way to novel applications in Computer Graphics and Computer Vision. We envision the possibility to manipulate measured BRDFs based on cumulants to let artists create variations on real-world materials. It also has the potential to enable more direct material estimation and image-based material editing, by directly estimating and manipulating filter properties in images.

9398-8, Session 1

Principal component analysis for surface reflection components and structure in the facial image and synthesis of the facial image in various ages

Misa Hirose, Saori Toyota, Chiba Univ. (Japan); Nobutoshi Ojima, KAO Corp. (Japan); Keiko Ogawa-Ochiai, Kanazawa Univ. Hospital (Japan); Norimichi Tsumura, Chiba Univ. (Japan)

In this paper, principal component analysis is applied to pigmentation distributions, surface reflectance components and facial landmarks in the whole facial images to obtain feature values. Furthermore, the relationship between the obtained feature vectors and age is estimated by multiple regression analysis to modulate facial images in woman of ages 10 to 70. In our previous work, we analyzed only pigmentation distributions and the reproduced images looked younger than the reproduced age by the subjective evaluation. We considered that this happened because we did not modulate the facial structures and detailed surfaces such as wrinkles. By analyzing landmarks represented facial structures and surface reflectance components, we analyze the variation of facial structures and fine asperity distributions as well as pigmentation distributions in the whole face. As a result, our method modulate the appearance of a face by changing age more physically.

Conference 9398: Measuring, Modeling, and Reproducing Material Appearance 2015

9398-9, Session 1

Extrapolation of bidirectional texture functions using texture synthesis guided by photometric normals

Heinz C. Steinhausen, Rodrigo Martín, Dennis den Brok, Matthias B. Hullin, Reinhard Klein, Univ. Bonn (Germany)

Numerous applications in computer graphics and beyond benefit from accurate models for the visual appearance of real world materials. Their common representation as bidirectional texture functions (BTFs) subsumes variations in geometry and non-local lighting effects into an "apparent" bidirectional reflectance distribution function (aBRDF) for each position on the surface. Unlike BRDFs, aBRDFs do not adhere to Helmholtz reciprocity and energy conservation and have too many degrees of freedom to be reasonably represented by parametric models.

When it comes to measuring BTFs, a discrete sampling of the function is typically obtained using image-based setups like camera domes, gonioreflectometers or cameras attached to kaleidoscopic arrangements of mirrors. One conceptual drawback of such data-driven models is the limited sample size. To justify the common assumption of far-field illumination, one has to maintain a certain ratio between the distance of the sample to the cameras and lights on one side and the sample size on the other side. Most acquisition setups compared by Schwartz et al. [1] thus support sample sizes not larger than 10 cm by 10 cm. Obviously, materials like leather, structured wallpapers or wood contain structural elements on different scales. While BTFs obtained by the aforementioned methods are able to capture small- and medium-scale structure, their measurements naturally lack information on the unseen large-scale structure.

Direct acquisition of a BTF for the larger sample, e.g. by sequentially treating different regions of the material, is prohibitive with regards to acquisition and postprocessing time as well as memory requirements. We therefore propose a method that extends recent research presented by Steinhausen et al. [2] to extrapolate BTFs for large-scale material samples. The input is a measured and compressed BTF for a small region of the material sample, as well as a set of large-scale photographs that serve as additional constraints. We extend this work in several manners:

While the method by Steinhausen et al. only relies on rectified and resampled photographs of the full sample, we propose using normal maps as part of the constraints guiding the extrapolation process. Combined with further imagery and neighborhood information, these are used as distribution maps narrowing down the search space for suitable aBRDFs per texel to a large extent. In order to acquire normal maps for nearly flat materials of size in the range of 20 cm by 20 cm, we build upon an idea proposed by Pan and Skala [3] that makes use of an off-the-shelf flatbed scanner to scan the full sample in four different rotations. As flatbed scanners are available in a variety of sizes, this allows us to expand the range of covered scales at a moderate cost.

As another use case, we propose to utilize this method to obtain the same amount of detail with reduced measurement effort. This is achieved by reducing the size of the measured sample, enabling the acquisition of reflectance data for multiple material samples at once. The missing parts are then reconstructed using our method.

To provide optimal parameters of operation, we evaluate the effect of different samplings, i.e. positions from which the photographic constraint images are taken, on reconstruction quality for ground-truth data. We compare our results to Steinhausen's original method based on texture optimization, and demonstrate the improvement in computational efficiency that results from a pixel-based, as opposed to patch-based, texture synthesis scheme.

[1] Schwartz, C., Sarlette, R., Weinmann, M., Rump, M., Klein, R. (2014). Design and Implementation of Practical Bidirectional Texture Function Measurement Devices Focusing on the Developments at the University of Bonn. Sensors. 2014; 14(5):7753-7819.

[2] Steinhausen, H. C., den Brok, D., Hullin, M. B., & Klein, R. (2014, June). Acquiring Bidirectional Texture Functions for Large-Scale Material Samples. In International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision (WSCG 2014).

[3] Pan, R., & Skala, V. (2013, September). Normal Map Acquisition of Nearly Flat Objects Using a Flatbed Scanner. In Virtual Reality and Visualization (ICVRV), 2013 International Conference on (pp. 68-73). IEEE.

9398-10, Session 1

Using line-scan camera based structure from motion for high resolution 3D reconstruction

Pengchang Zhang, Tomoyuki Takeda, Jay Arre O. Toque, Ari Ide-Ektessabi, Kyoto Univ. (Japan)

3D geometrical shape acquisition based on structure from motion has been attracting increasing interests and demands in the realm of cultural heritage digitization. Although it has found many successful applications, high resolution still remains a challenging problem for researchers. Previous studies mostly focused on area camera based structure from motion for imaging medium or large size objects such as architecture, monuments, and sculptures. However, these methods demonstrated much lower resolutions. In this paper, focus is given to increase the resolving ability of reconstructed 3D points by using line scan cameras. This technique is capable of delivering much greater details of the geometry, texture, color, and material information which are quite beneficial to the scientific documentation, restoration, research and visualization of cultural heritages.

There are several critical problems that need to be addressed: (1)camera calibration; (2) image correspondence; and (3) shape extraction.

An area camera obtains the whole image at one shot with its optical center fixed while a line scan camera captures an image by shifting the optical center along one direction perpendicular to its linear image sensor. In this way, a line scan camera has a distinct geometric image formation model compared with an area camera. In this paper, the mathematical equation describing the relationship between the coordinate of a pixel in captured image and its corresponding 3D spatial coordinate is analyzed and derived as the basis for camera calibration.

To get the 360-degree view of an object, a rotating table is used in this study to sit the object on while keeping the camera position fixed. During the capture of the images, the table is rotated by a certain degree at each scan. Such intrinsic parameters as focus length remains constant during the whole image capturing process but changes in position and orientation of the camera need to be estimated as long as the table is rotated. Therefore, estimation of intrinsic and extrinsic parameters in this study is conducted separately. The intrinsic parameters only need to be estimated once, while camera positions and orientations need to be estimated for each viewpoint. A self-designed calibration rig is used with known 3D coordinate of some physical feature point in space to estimate the intrinsic parameters of the camera; on the other hand, extrinsic parameters are estimated without utilization of any calibration rig, called self-calibration.

There are two classes of method for self-calibration: (1) based on epi-polar equations; (2) factorization method. For area cameras, the calculation of epipolar equations is generally a linear problem, however, as we will show in this paper, epipolar equations for line scanners are non-linear in essence which pose significant complexity in solving the problem. Therefore, a factorization method is discussed in this paper to estimate both the camera position and pose at each viewpoint and the 3D shape of the object at the corresponding view.

Correspondence is a critical issue when recovering 3D shape in methods such as stereo vision and shape from motion. For high resolution reconstruction, dense correspondence is both a indispensable and extremely challenging task. In this paper, due to the viewpoint variation in the image capturing process, Affine-SIFT algorithm, which is a fully affine-invariant extension to SIFT(Scale-Invariant Feature Transform) feature detection algorithm, is employed to establish dense correspondence with high accuracy. To reduce the computational complexity of Affine-SIFT, a GPU version of the algorithm is implemented in this study.

In factorization method, shape extraction is achieved simultaneously with estimation of camera positions and orientations. To use the factorization method to image sequence acquired by line scan cameras, perspective

projection model is approximated to an affine projection at the cost of decreased accuracy. On one hand, bundle adjustment is used to minimize the difference between the estimated 3D coordinates and their corresponding actual physical coordinates, thus compensating for the accuracy loss greatly; on the other hand, the solution to the factorization of measurement matrix into motion matrix and shape matrix is not unique, and two sets of symmetrical solution are generated. The correct unique solution is found by taking the imaging process of the rotating table-based 3D scanner into account.

The proposed line-scan camera based shape from motion method is experimentally verified and evaluated by designing and implementing the ideal through a table-rotating 3D scanner. In this scanner, 360-degree view of the object is achieved by rotating the table where the object sits on by a certain degree at each scan and the corresponding image is captured by the line scan camera. To perform the factorization method, a minimum number of 3 adjacent images are used. All the partial 3D views of the object generated by a run of the factorization are integrated into a full 360-degree 3D view of the object in a common coordinate system. 3D artworks are used as the subjects for experiment. A laser scanner is used as the reference for comparison in order to evaluate the accuracy of the reconstructed result. Experimental result shows that the proposed method demonstrates great effectiveness and efficiency in reconstructing high resolution 3D shape of cultural heritage.

9398-11, Session 2

NIST appearance metrology program *(Invited Paper)*

Maria E. Nadal, National Institute of Standards and Technology (United States)

Appearance greatly influences a customer's judgment of the quality and acceptability of manufactured products. One of the primary missions of the National Institute of Standards and Technology (NIST) is to strengthen the U.S. economy by working with industry to develop and apply technology, measurements and standards. NIST has established a program for the appearance characterization of objects that includes calibration services for color and specular gloss standards and research in the characterization of novel coatings. These services are designed to meet demands for improved measurements and standards to enhance the acceptability of the final products.

For the calibration of color standards, a 0:45 reference spectral reflectometer has been developed at NIST. The instrument specifications were determined by a series of simulation of errors caused by the measuring instrument on the calibrated spectral reflectance of color standard and calculated color values. Particular attention was given to the inherent properties of the instrument such as stray light level, random noise, and wavelength uncertainty. The calibrations are performed for an influx angle of 0 degree and efflux angle of 45 degrees for wavelength from 380 nm to 780 nm in 5 nm. This instrument measures the spectral reflectance properties of samples, from which color quantities are calculated.

For the characterization of novel coatings such as gonioapparent materials, a five-axis gonio-spectrometer has been established at NIST. These novel coatings require exceptional processing conditions and characterization methods, which are different from the traditional single-geometry methods. The necessary set of measurement geometries is determined by the complexity of the scattering mechanisms present in these coatings. The NIST five-axis gonio-spectrometer measures the spectral reflectance of samples over a wide range of illumination and viewing angles for in-plane and out-of-plane geometries. Three detection and illumination systems have been implemented. The first system is the traditional single-element silicon diode combined with monochromatic illumination. For this set-up, the data acquisition sequence for one set of incident and viewing angle for the visible range of 380 nm to 780 nm takes about four hours. Therefore, the reflectance measurements of one gonioapparent sample are very time consuming. The second system is a fiber-coupled CCD array spectrometer with a white illumination source. Unlike the first system, the array spectrometer measures the entire spectrum simultaneously, dramatically decreasing the acquisition time. In this set-up the wavelength range was extended to 380 nm to 1050 nm. The measurement uncertainty was determined to be comparable to similar instruments operating in a single channel configuration. A comparison between the two setups shows a 0.4 % agreement. The third system consists of monochromatic illumination with the fiber-coupled CCD array spectrometer and is used for the characterization of fluorescing optically activate pigments.

9398-12, Session 2

Metrological issues related to BRDF measurements around the specular direction in the particular case of glossy surfaces

Gaël Obein, Conservatoire National des Arts et Metiers (France); Jan Audenaert, Katholieke Univ. Leuven (Belgium); Guillaume Ged, Conservatoire National des Arts et Metiers (France); Frédéric B. Leloup, Katholieke Univ. Leuven (Belgium)

Among the visual attributes implicated in the visual appearance of materials, gloss is known to be the second most important attribute beside color [1]. Within particular domains, such as the printing industry, the aspect of gloss is even gaining in importance, thanks to the introduction of and the recent developments achieved with 2,5D and 3D printers. While the surface relief of a print reproduction enhances the glossy effects, a visual mismatch of gloss on e.g. an art reproduction may yield a poor quality sensation. Today the control of the level of gloss is hence requested.

Yet, a good working knowledge requests appropriate measurement techniques. Unfortunately, the soft metrology [2] of surface gloss lags behind. Indeed, a glossmeter is not capable of quantifying the subtleties of the visual gloss sensation [3], and new measurement procedures must therefore be proposed. This scope is part of a European metrological project called "xDReflect" [4]. The results presented here have been obtained in the framework of this project.

The light reflected at the surface of a material can be fully characterized by the Bidirectional Reflectance Distribution Function (BRDF). Among the complete BRDF, visual gloss is principally related to physical reflection characteristics located around the specular reflection direction. This particular part of the BRDF is usually referred to as the specular peak. A good starting point for the physical description of gloss could be to measure this specular peak. Unfortunately, such a metrological characterisation is not trivial.

As a matter of fact, the width of the specular peak of a glossy surface may become very narrow, with full width at half maximum values below 1° in zenith. An appropriate measurement of such a peak requests BRDF measurement devices with an enhanced angular resolution, i.e., very small solid angles of detection, and these are achieved on just a few gonioreflectometers [5].

Moreover, the diminution of the solid angle of detection engenders very high BRDF values, which are strongly dependent on the instrument function of the gonioreflectometer used. In absence of any standardization regarding the optical design of BRDF measurement instrumentation, each measurement device has its own particular optical design (i.e., different dimensions of illumination and receptor apertures). In result, the direct comparison of measurement results obtained on different devices becomes problematic. This point is a drag on the development of specular peak measurements and thus, in more general, on the optical metrology of gloss.

In this paper we address this issue. By way of example, BRDF measurements around the specular peak of two white surfaces, one being matte and the other one being glossy, are described. The measurements were performed on two high level gonioreflectometers having a different optical design. Important discrepancies in the measurement results of the glossy sample, due to the convolution effect, are presented and discussed. Finally, luminance maps obtained from renderings with the acquired BRDF data on both devices, exemplify the large visual differences that might be obtained

for glossy samples of which the BRDF data, although both being correct, significantly differ.

This article points out the metrological challenges in the field of BRDF measurements of glossy surfaces. These difficulties must be known by metrologists and by users of BRDF measurement data. Their comprehension of parameters affecting the measurement results is an inevitable step towards progress in the metrology of surface gloss, and thus also towards a better metrology of appearance in general.

1. CHRISTIE, J., "Evaluation of the attribute of appearance called gloss", CIE-Journal, 5, 2, 41-56, (1986).

2. COMMISSION INTERNATIONALE DE L'ECLAIRAGE TECHNICAL REPORT 175:2006, "A framework for the measurement of visual appearance" (CIE, 2006).

3. LELOUP, F. B., POINTER, M., DUTRE, P., HANSELAER, P. « Geometry of illumination, luminance contrast, and gloss perception ». J Opt Soc Am A, 27, pp2046-2054. (2010)

4. HÖPE, A., KOO, A., VERDU, F., LELOUP, F. B., OBEIN, G., WÜBBELER, G., CAMPOS, J., IACOMUSSI, P., JAANSON, P., KÄLLBERG, S., ?MÍD, M., 2014, "Multidimensional Reflectometry for Industry" (xD-Reflect) an European research project, Measuring, Modeling, and Reproducing Material Appearance, Proceedings of SPIE Vol. EI108, 4-5 February, San Francisco, USA.

5. OBEIN, G., OUARETS, S., GED, G., 2014, Evaluation of the shape of the specular peak for high glossy surfaces (invited paper), Measuring, Modeling, and Reproducing Material Appearance, Proceedings of SPIE Vol. EI108, 4-5 February, San Francisco, USA.

9398-13, Session 2

Upgrade of goniospectrophtometer GEFE for near-field scattering and fluorescence radiance measurements

Berta Bernad, Alejandro Ferrero, Alicia A. Pons, María Luisa Hernanz, Joaquín Campos Acosta, Consejo Superior de Investigaciones Científicas (Spain)

In last years, the measurement of appearance of objects has gained increasing relevance because of their wide application in industry and basic research. Appearance, according to the International Commission on Illumination [Pointer 2006], is the visual sensation through which an object is perceived to have attributes such as size, shape, color, texture, gloss, transparency, opacity, etc. The specification of many correlates of visual attributes needs to be investigated yet, and this task requires the availability of spectrophotometric measurements of materials.

Spanish Council for Scientific Research's Optic Institute (IO-CSIC) designed and developed the gonio-spectrophotometer GEFE, that allows the spectral sBRDF (Spectral Bidirectional Reflectance Distribution Function) of surfaces to be measured at any illumination/detection geometry, including out-of-plane and actual retro-reflection geometries [Rabal 2012].

In order to extend its capabilities, so that it is possible to make a more complete description of the scattering of optical radiation by surfaces, the following implementations are under development within the activity of the project "Multidimensional reflectometry for industry" funded by the European Metrology Research Program (EMRP):

- Incorporation of a camera in the detection system that allows spatially-resolved measurements to be obtained, providing information of the radiation emitted from different points of the analyzed surface. This capability is very important in the study of the BSSRDF (Bidirectional Scattering-Surface Reflectance Distribution Function) [Nicodemus 1977], texture, uniformity and other effects, such as sparkle. In addition, spatial resolution is extremely useful to understand and estimate uncertainty sources, for example, the sample inhomogeneity, usually ignored in far-field measurements.

- Inclusion of a monochromator in front of the luminous source to provide the system with monochromatic irradiation. This will allow fluorescence measurements to be carried out. Furthermore, this kind of lighting offers the possibility to perform spectral measurements with any detector, and not only with those provided with spectral analysis capability.

- Development of the ability to perform spectral transmittance measurements, in order to determine the Bidirectional Transmittance Distribution Function (BTDF). This feature, together with the BRDF, represents the Bidirectional Scattering Distribution Function (BSDF) and, therefore, the complete far-field characterization of the scattering properties of a material.

As a result of this upgrade, GEFE should be able to perform the complete characterization of the scattering of flat surfaces and, therefore, provide relevant spectrophotometric measurements to study the appearance. Beyond the present ability to measure color and its variation with geometry, the system will be able to explore other perceptual categories as texture, translucency, fluorescence or sparkle.

The new design of the gonio-spectrophotometer, the camera-assisted characterization procedure and results about some fundamental instrument's properties, such as the solid angle of illumination or the uniformity in the plane defined by the sample, will be presented. This more thorough knowledge of GEFE will allow the radiance uncertainty budget estimation to be improved. In addition, preliminary results of spectrophotometric measurements of fluorescence and sparkle, from specially selected samples, will be presented.

References

• Pointer, M. (2006). A Framework for the Measurement of Visual Appearance.CIE Publication 175-2006: CIE TC1-65 Technical Report.

• Rabal, A. M., Ferrero, A., Campos, J., Fontecha, J. L., Pons, A., Rubiño, A. M., & Corróns, A. (2012). Automatic gonio-spectrophotometer for the absolute measurement of the spectral BRDF at in-and out-of-plane and retroreflection geometries. Metrologia, 49(3), 213.

• Nicodemus, F. E. (1977). Geometrical considerations and nomenclature for reflectance (Vol. 160). Washington, D. C: US Department of Commerce, National Bureau of Standards.

9398-14, Session 2

Rapid acquisition of bidirectional texture functions for materials

Dennis den Brok, Heinz C. Steinhausen, Matthias B. Hullin, Univ. Bonn (Germany); Reinhard Klein, Rheinische Friedrich-Wilhelms-Univ. Bonn (Germany)

Analytical or physically-based reflectance models such as the bidirectional reflectance distribution function (BRDF) and its spatially varying sibling (SVBRDF) generally fail to accurately represent the noticeable non-local effects such as interreflections and self-shadowing that can be observed on many common real-world materials.

The bidirectional texture function (BTF) accounts for these effects. Its image-based nature, however, makes material BTFs extremely cumbersome to acquire: in order to adequately sample high-frequency details, many thousands of images of a given material with different lighting and viewing positions have to be obtained. Existing BTF acquisition setups employ a "brute-force" approach where all desired images are acquired explicitely. Parallel setups such as camera domes and kaleidoscopes significantly accelerate this process, but they still suffer from long exposure times for HDR capture and the sheer number of samples required.

We overcome both of these problems by means of an efficient "bootstrapping" method that allows for sparse acquisition with low exposure times:

In the "bootstrapping" step we create a heterogeneous database of material BTFs with full angular resolution.

In order to mitigate the long exposure times we capture the individual materials' appearance lit by many lights at once ("multiplexed illumination"). Due to the approximate linearity of the superposition of light sources, the desired single-light images can then be obtained by solving an appropriate

linear system ("demultiplexing"). This process, however, decreases the ratio of signal to signal-dependent noise to an extent that is often intolerable when the number of light sources per illumination pattern is large. On the other hand, we are interested in using as many light sources as possible in order to be able to use very short exposure times. We show that many-light illumination patterns derived from Hadamard transform optics yield a level of noise that is still tractable through non-local filtering, exploiting the correlations and self-similarities in the material's appearance. Following this strategy alone, we are typically able to reduce acquisition times by about 90%.

Once the database has been created, we perform all further BTF measurements both under multiplexed illumination and sparsely; i.e. we acquire images only for a small subset of all illumination patterns and camera positions. We subsequently approximate the missing values using the database. To this end, we fit linear models to the database's per-texel BRDF-like data via truncated singular value decomposition. In principle, approximations of the missing values could then be obtained by using model parameters fit in the subspace determined by the sparse sampling. However, as the data is very high-dimensional, the number of samples required for this straight-forward approach to be stable is infeasible. We therefore apply Tihonov regularization, based on the assumption that our linear models generalize well to novel materials. In that case, the models' basis vectors corresponding to large singular values are likely to contribute the most also to representations of novel materials. We hence propose to use the singular values as regularization weights. That way, we are able to reconstruct BTFs with full angular resolution from very sparse measurements and decrease acquisition times even further. Moreover, we show that this approach also helps mitigating the demultiplexing noise.

We evaluate the proposed method on a number of real-world materials in a camera dome scenario. As we shall explain, the method readily applies to other common BTF acquisition setups as well.

9398-15, Session 2

An abridged goniometer for material appearance measurements

Adria Fores Herranz, Mark D. Fairchild, Rochester Institute of Technology (United States); Ingeborg Tastl, Hewlett-Packard Labs. (United States)

Man-made objects have a geometry that is easy to model with current CAD applications and 3D scanning techniques to be used in virtual environments. However, the capture of material appearance is a more challenging task. Goniospectrophotometers can be used to measure how the light is reflected from a material at different directions, but the time consuming measurement process and expensive nature of those devices limits its use.

The generation of realistic synthetic images relies on an accurate representation of the material appearance of objects, and improvements on its capture and representation would improve the realism of previews of to be printed 3D objects, regular 2D printing taking into account the appearance of the substrate and inks used, and the accurate representation of real materials in computer generated images.

In this paper a cost-effective, fast, and scalable solution to capture the material appearance is presented. The main idea behind this work is the use of simpler devices, commonly used for quality control applications, and to combine the measurements to represent the material appearance of an object. Devices used for quality control applications are used in this work as they are fast, as they need to measure multiple samples in a short period of time and to constrain their measurements to important perceptual properties.

The main idea behind the measurement technique presented in this paper is to split the information to capture to represent the material appearance into two different attributes: color and gloss. The same separation is used to represent BRDFs with analytical models, where a diffuse lobe is commonly used to represent the color of an object, and a specular lobe is used to represent the gloss appearance. The technique presented would generalize for any isotropic material, but the monochromatic capture of the specular lobe by the device used in this work limits the current applicability of this technique to dielectrics, since the color of metals is known to be represented in the specular lobe and it will not be captured.

The diffuse reflectance of a material will be acquired using the 45:0 measurement geometry (illumination:measurement directions), this geometry is commonly used in colorimetry to avoid capturing the specular component. The X-Rite i1 spectrophotometer will be used in this work to capture the spectral reflectance of a material, using a PTFE created from pressed teflon powder as standard.

The requirements for the gloss measurements are the following: high angular resolution, high dynamic range, and the measurement of multiple incident directions. The high angular resolution is required to correctly capture the width of the specular lobe. The high dynamic range is required to be able to capture materials from diffuse that have almost negligible specular lobes to high gloss. Finally, the ability to measure at different incident directions would allow to measure the Fresnel effect behavior on a specific material, which models the increase in reflectance when the incident direction goes towards grazing angles.

The Rhopoint IQ is a DOI-Gloss-Haze meter used for quality control applications. Its design is similar to a 20/60/85 gloss meter, but a linear sensor of 512 pixels covering +/-7.25 degrees around the mirror direction at 20 degrees is used instead of a single diode to provide an angular resolution of 0.028 degrees. The Rhopoint IQ reports Specular Gloss at the three different geometries, and by using the linear sensor the DOI, Haze, RSPEC (average gloss at the peak of the specular lobe), and the goniometric curve are also provided for the 20 degree geometry. The gloss attributes reported by the device are relative attributes computed against the standard used in calibration (black glass or mirror).

By default, the goniometric curve is reported in gloss units, and is expected to only be used as an aid to better understand the material angular reflectance of the sample measured. Access to the RAW sensor data was given to us by the manufacturer, thus allowing the use of the device to perform BRDF measurements.

The diffuse reflectance of a material is set directly to the 45:0 measurement, and the parameters of the analytical BRDF model used and the specular reflectance can be non-linearly optimized to approximate the BRDF data measured with the Rhopoint IQ.

The Rhopoint IQ characteristics limit the applicability of this technique to non-metallic and non-goniochromatic materials, and as only a single incident direction is captured with the linear sensor the increased reflectance towards grazing angles (Fresnel effect) behavior is not captured. Thus, the fact that different materials might have a different behavior when the incident direction goes towards grazing angles will not be measured.

The technique presented is later evaluated by comparing its results with the high accuracy measurements of a goniospectrophotometer, and the approximations obtained when the optimization process starts with the high accuracy measurements.

The Murakami GCMS-10x goniospectrophotometer, which measures the spectral reflectance factor as a function of incident and detection angles, was used as a reference instrument in order to verify the performance of the developed technique by measuring a set of 36 printed samples varying from high to low gloss with both devices.

The Rhopoint IQ was found to provide a good approximation of the specular lobe measurements obtained with the goniospectrophotometer. Two main differences are found between the measurements of both instruments: A small underestimation of the peak of the specular lobe for high glossy materials, which is due to the limited dynamic range of the Rhopoint IQ sensor when compared to the Murakami, and the limitation to capture the width of broad specular lobes related to low gloss materials, which is due to the lack of angular coverage beyond the +/-7.25 degrees of the mirror direction of the Rhopoint IQ.

To conclude, a novel cost-effective, fast, and scalable solution to measure material appearance is presented in this paper. The main contribution is to separate the measurement of different appearance properties of materials, color and gloss, while using simpler devices to perform each measurement. A good approximation was obtained when comparing the new technique to a goniospectrophotometer, except for a small underestimation of the peak of the specular lobe of high gloss materials and the limitation to capture the specular lobe width of broad specular lobes found on low gloss materials.

The use of the Rhopoint IQ enables the technique presented in this work, but at the same time its design constrains the usage of the presented technique to non-metallic materials, due the achromatic capture of the specular lobe, and the fresnel effect is not captured as only a single incident angle is measured.

A promising avenue of future work is the creation of a measurement device that builds upon the initial Rhopoint IQ design and addresses the limitations found in this paper. Having a RGB filter array in the linear sensor would allow to capture colorimetric information of the specular lobe while still preserving reasonable angular resolution. A change in the lens design would allow to capture beyond the current +/-7.25 degrees from the mirror direction, thus being able to correctly capture the specular lobe width of low gloss materials. Finally, the addition of a second linear array at another measurement geometry would allow to (1) capture the fresnel effect behavior and (2) capture goniochromatic effects when combined with the RGB filter. The improvement of the sensor's dynamic range would also lead to an overall accuracy improvement.

9398-16, Session 2

New generation of Fourier optics instruments for fast multispectral BRDF characterization

Pierre M. Boher, Thierry Leroux, Véronique Collomb-Patton, Thibault Bignon, ELDIM (France)

Spectral BRDF measurements are generally made with goniometric solutions which are generally quite expensive and time consuming when high angular resolutions are needed. The same type of problem was observed for the viewing angle characterization of emissive displays in the past. When ELDIM introduced commercially Fourier optics instruments for that purpose in 1992, the method was rapidly adopted by all the main display makers for its simplicity and rapidity. All the main features of Fourier optics instruments come from the patented optical configuration which allows controlling the angular aperture of the system independently of the measurement spot size, with a high collection efficiency. The light rays are collected by ultra-large angular aperture Fourier optics (up to ±88°) and refocused on a primary Fourier plane. This plane is reimaged on an imaging detector via field optics and imaging optics. The spot size is defined by a diaphragm at a location where we have a direct image of the measurement spot. Near this location color filters or band pass filters allow luminance, color or spectral measurements. Measurements on non-emissive surfaces requires an additional method to illuminate the sample surface. It is realized across the same optics using a beam splitter and an additional lens. Full diffused illumination and collimated illumination are realized using integration sphere and fiber optics on one illumination Fourier plane respectively. A multispectral version of this system has been introduced in 2008 for the characterization of LCD displays (1) and version with reflective option was applied to e-papers and color shifting paints (2-3). Evaluation of the display performances under external illumination has been performed with the same instrument (4-5).

In this paper we present a cost effective version of this instrument that allows full diffused reflectance and spectral BRDF measurements rapidly and without complex setup. The angular aperture (±60°) and the high angular resolution (< 0.1°) allows rapid measurement at different wavelengths. The collimated beam illumination is made across the same optics with white LEDs positioned on one illumination Fourier plane obtained using a beam splitter and an additional lens. The wavelength resolution is limited (~20nm) but sufficient to compute accurate surface aspects under any type of illumination. Experimental results obtained in different field of applications will be presented. For cosmetics the capacity to measure completely the backscattering is very interesting. Backscattering properties of light diffusers can be quantified and related the morphology. Their impact when included inside foundations can also be investigated. Other applications in the field of displays will be also presented. Electronic papers and small size displays when used outdoor can see their characteristics strongly degraded. The capacity of the instrument to measure both emissive properties under dark conditions and reflective properties allows to quantify precisely the contrast and color gamut of the displays under any illumination conditions (4). Display aspect for any kind of content and any position of the observer can even been computed (5).

(1) "New multispectral Fourier optics viewing angle instrument for full characterization of LCDs and their components", Pierre BOHER, Thierry LEROUX, Thibault BIGNON, David GLINEL, SID, Los Angeles, USA, May 18-23, P89 (2008)

(2) "Optical characterization of e-papers using multispectral Fourier optics instrument", Pierre BOHER, Thierry LEROUX, 2nd CIE Expert Symposium on Appearance, 8-10 September, Ghent, Belgium (2010)

(3) "Spectral BRDF of color shifting paints and e-papers using multispectral Fourier optics instrument", Pierre BOHER, Thierry LEROUX, Thibault BIGNON, Eurodisplay, 7.4 (2011

(4) "Optical measurements for comparison of displays under ambient illumination and simulation of physico-realistic rendering", Pierre BOHER, Thierry LEROUX, Thibault BIGNON, Vincent LEROUX, Dearborn, Michigan, October 20-21 (2011)

(5) "Physico-realistic simulation of displays", Pierre BOHER, Thierry LEROUX, Véronique COLLOMB PATTON, Vincent LEROUX, Vehicle Display Symposium, Dearborn, October 18-19 (2012)

9398-17, Session 2

Color calibration of an RGB digital camera for the microscopic observation of highly specular materials

Juan Martínez-García, Mathieu Hébert, Alain Trémeau, Univ. Jean Monnet Saint-Etienne (France)

Image acquisition systems provide their users with a practical and inexpensive way to record colored spatial information from a scene or an object. Nevertheless, the recorded colors might be different from the ones perceived by the human visual system and could also vary between different devices due to the fact that the spectral response of their RGB channels are device dependent; hence, unless the effective spectral responses of the imaging system are determined, colors can be compared from each other only if they are measured with the same imaging system. The color calibration of an imaging device consists in building a transformation law between the RGB colors recorded from a set of colored samples (learning samples), and the CIE-XYZ or CIE-L*a*b* values derived from spectral measurements of these samples, considered as the ground-truth, device-independent colors. The key point for accurate color calibration is that the color chart and the objects to measure must be observed under the same lighting conditions and with the same capture settings for the imaging device. As a consequence, to prevent saturation of the detector or too low light signal, color chart and objects need to have comparable reflectances in the selected illumination-observation geometry. Most color charts available on the market place, such as the GretagMacbeth ColorChecker® or ColorChecker Digital SG®, are Lambertian, or at least very diffusing. They are therefore not adapted to highly specular surfaces, whose reflectance in the specular direction is generally much higher than the reflectance of a diffuse material. Even when the channels are not completely saturated, the colors recorded by the calibrated device will appear clearly brighter than they would actually be perceived by the human vision system. Various illustrations will illustrate that point in our study.

The first difficulty to calibrate a device for observation of specular samples is that no specular color chart is available. Semi-glossy or glossy color charts are not adapted neither, because the light reflected by their surface, i.e. the gloss, is mostly achromatic. Moreover, none of the existing charts are homogenous enough at the microscopic scale to enable the calibration of a microscope, which is the aim of our study. In this paper, we thus propose four different methods to calibrate a microscopic imaging system that overcome the aforementioned limitations and show that one of them is an efficient alternative to classical calibration methods, with a comparable accuracy.

Methods:

In order to obtain a color chart with colored specular reflections to test the different methods, we used 50 colored Supergel ROSCO filters and placed them in front of a calibrated mirror. It provides an adequately regular surface since the microscope is focused on the mirror, which is very homogeneous at microscopic scale.

Several algorithms have been previously studied to calculate the relationship between device dependent and device independent spaces in the color calibration of imaging devices, such as: neural networks [1], three-dimensional lookup tables [2], and polynomial transformations. And it has been demonstrated that both polynomial transformation and neural networks have similar performance even though polynomial transformations are easier to implement and require a smaller number of calibration samples to achieve good results [3]. For that reason, we use a polynomial transformation similar as the one proposed in [4], by performing not only 3rd degree polynomial fitting but also 2nd and 4th degrees in order to study the effects of polynomial overfitting.

The first two studied methods are very simple. They consist in adjusting the imaging device calibrated with 48 Munsell matte color sheets to assess our highly specular samples. Since the specular samples reflect much more light than the diffuse color chart, we reduce the exposure time of the camera (first method) or the power supply of the lamp (second method) in order to prevent saturation of the sensors. The samples were measured with the device and also with the X-rite Color i7 spectrophotometer (based on the diffuse:8° geometry) to compute the ground-truth colors. In the two methods, with these two sets of measurements, the polynomial transformation is applied and the calibration parameters are obtained. The specular testing samples are measured in the same conditions, the colors recorded by the imaging system are transformed by using the calibrated transformation, and compared with the ground-truth colors issued from spectral measurement. The quality of the calibration is assessed by computing the CIELAB DeltaE 1994 value for each sample, and considering the average of the set of testing samples.

The third method relies on a 72 specular learning samples, i.e. a mirror successively covered by 72 LEE filters. The samples were measured both with the device and with the i7 spectrometer. The polynomial transformation parameters were obtained, and then applied to the colors recorded by the device for the same testing samples as for the first and second method.

Lastly, the fourth method is based on the same learning samples as the third method but the spectral reflectances were measured by using a different optical set-up that reproduces the illumination-observation geometry of the microscope, i.e. the 0°:0° geometry where the samples are illuminated and observed at the normal of their surface. The transformation parameters were subsequently obtained and applied to the colors recorded from the testing samples.

Results:

A complete comparison between the four methods has been carried out, permitting an interesting discussion on the influence of different parameters on the calibration accuracy of each method. Through our experiments, we conclude that the third method with 2nd degree polynomial fitting is the most accurate: by comparing the colors recorded by the device after transformation and those measured for our set of 50 specular testing samples, we obtained an average CIE-DeltaE 1994 color difference of 2.1 units, which is a good performance compared with other studies presented in the literature (for diffusing materials): 2.5 units in [4], 2.5-3.0 units in [5] and 2.2 in [6]. This method is already being used for color measurements of photochromic dyes created by laser insolation of clear plates.

[1] Artusi, J. Electron. Imaging, 2003

[2] Hung, Proc. SPIE 1448, 1991

[3] Cheung, Color. Technol., 2004

[4] Charrière, App. Opt., 2013

[5] Quintana, Comput. Med. Imaging Graph, 2011

[6] Jackman, Meat Sci., 2012

9398-18, Session 2

An Image-Based Multi-Directional Reflectance Measurement Setup for Flexible Objects

Aditya S. Sole, Ivar Farup, Gjøvik Univ. College (Norway); Shoji Tominaga, Chiba Univ. (Japan)

Appearance of an object material is measured with the goal of objectively describing and quantifying human visual impressions with measurement values. Measurements help us communicate the appearance of the material in numerical terms.

Material appearance can be described using surface reflectance properties. The most appropriate way to measure the surface reflectance is by performing physical measurements [1]. Physical reflectance measurements are performed (for example, in the packaging and car paint industry) to characterise and control reproduction of materials like printed packages and paints [2,3]. To completely capture the reflectance of opaque object materials, bidirectional reflectance distribution function (BRDF) can be calculated. BRDF completely describes the reflectance of an opaque surface at a given point [1]. BRDF is usually recorded using custom built high end measurement instruments known as gonio-spectrometers and multi-angle spectrophotometers. However, using a gonio-spectrometer to measure the BRDF of the material is a time consuming and an expensive process and is mainly used in research environments [4].

The principal objective of this paper is to outline an image based measurement setup to perform multi-directional reflectance measurements at relatively cheaper and faster way. These measurements can then further be used for BRDF estimation. The image based measurement setup records, the light reflected from the object material, at multiple incident and reflection angles. We use a point light source to illuminate and a RGB camera to record the reflected radiance from the flexible object that is wrapped around a cylinder of known radius. RGB cameras have been used for process control (in-line measurement, capturing information from micro control elements, etc.) in packaging industry. The recorded radiance is used to estimate the reflectance of the object material using a reference white sheet curved onto the same cylinder (similar in the way the colour sample is wrapped around the cylinder) and a reference calibration tile (spectralon) measured at normal to the camera. In the setup, the object is wrapped around a cylinder of known radius. The object curvature helps us collect information on the amount of light reflected from the object surface for different incident and reflection angles. Due to sample curvature, light reflected at different directions, can thus be recorded from the single image captured by the camera. The flexible object material (for example, paper) to be measured is assumed to be homogeneous.

A geometrical method, presented in previous studies (paper accepted at CIC, 2014) successfully estimate the incident (?i) and reflection (?r) angles for the given point (P) on the material surface, from the image captured (of the curved sample). The methods corrects for the geometrical distortion caused due to mapping of a curved surface to a flat camera sensor array.

We use a reference white sheet and the spectralon tile to calculate irradiance at the curved object surface based on the normalised camera sensor information. Spectralon tile with a reflectance of approximately 0.99 over the visible range is used (assuming it as a lambertain surface) along with the curved reference white sheet to estimate the reflectance of the sample object. Using lambert's cosine law and the spectralon tile, we calculate the reflectance at the points on the reference white sheet (which make same incident (?i) and reflection (?r) angles as the points (P) make on the object surface). Using this data we calculated the approximate irradiance at these points that are then used to estimate the reflectance at the given point (P) on the curved surface of the object. As the object surface is being assumed to be homogeneous, the reflectance estimated at the corresponding points (P) is the reflectance information that corresponds to the incident (?i) and reflection (?r) angles, the point (P) makes with its normal. This way we are able to record reflectance information at multiple incident and reflection angles from the captured image. Equations, to calculate the reflectance from the sample radiance captured using the camera and incident point light source, are derived and discussed in the paper.

Results obtained indicate that the image-based technique can be used to measure the multi-directional reflectance of flexible opaque object materials that can further be used to estimate the BRDF. The object material properties, sample curvature and camera resolution decides the number of incident and reflection angles at which the bi-directional reflectance can be calculated using this method. The reflectance calculated from the method proposed in this paper was compared with measurements made using a diffuse/0° (Specular Included) spectrophotometer and a Tele-spectro-radiometer (TSR) on the selected colour samples. CIEXYZ_Y value was compared between the measurements, which showed a good relation with the calculated reflectance. The results are presented in the paper.

The object material measured using the above method and the corresponding reflectance calculation presented in the paper outlines an image based multi-directional reflectance measurement method that can further be used for BRDF estimation (for example – of opaque packaging materials).

References

1. Stephen R. Marschner, Stephen H. Westin, Eric P. F. Lafortune, Kenneth E. Torrance, and Don- ald P. Greenberg. Image-based brdf measurement including human skin. In 10th Eurographics Workshop on Rendering, pages 139 – 152, 1999.

2. C. S. McCamy. Observation and measurement of the appearance of metallic materials. part 1. macro appearance. Journal, 1996.

3. K. Kehren. Optical Properties and Visual Appearance of Printed Special Effect Colors. PhD thesis, Technischen Universita ?t Darmstadt, Darmstadt, Germany, April 2013.

4. G. Baba. Gonio-spectrophotometry of metal-flake and pearl-mica pigmented paint surfaces. In Proceedings of the fourth Oxford conference on spectrophotometry, pages 79–86. SPIE, 2003.

9398-20, Session 3

Extended visual appearance texture features

Simon-Frédéric Désage, Gilles Pitard, Maurice Pillet, Hugues Favrelière, Fabrice Frelin, Univ. de Savoie (France); Serge Samper, Univ. de Rennes 1 (France); Gaetan Le Goic, Univ. de Bourgogne (France)

Nowadays, perceived quality of visual appearance is [still] an industrial matter. The CIE (International Commission on Illumination) provided a technical report CIE 175:2006 “A framework for the measurement of visual appearance”. This report defines four headings under which possible measures might be made: color, gloss, translucency and texture. Considering color and translucency as known and controlled by the industry, we have chosen to work on the gloss and texture to quantify the impact on the human perception. This paper is then an echo of a “soft metrology” approach [1], because this paper presents some measurements to compare with the visual human behavior. A known difficulty is the interdependence of these measurements because translucency can influence color, which may influence gloss, and texture is probably a combination of all three [2].

Our starting point is the exploitation of most advanced visual surface representation and image processing approach of texture classification.

Some recent works have shown different representations of visual surface as appearance measurement. We propose to use a similar method for gloss measurement with a recent visual texture representation called Bidimensional Texture Function (BTF) [3] [4]. We do a recall of BTF and Apparent Bidimensional Reflectance Distribution Function (ABRDF) definitions, a basic taxonomy of BTF and a typical device [4]. For editing the useful function, we use a pixel-wise modeling method called Polynomial Texture Maps (PTM). There are other methods identified and more efficient such as Hemispherical Harmonic (HSH), but we start with the simpler method [4] [5]. We propose a surface appearance framework from ABRDF slice in three parts: color, material and relief. The shape of ABRDF slice is a feature of material, respectively the power for the color, and the direction for the relief. We can extract three images for each information type from ABRDF slice. This framework can help to discuss about texture definition. There are surface texture and image texture. One image combines these three types of information whereas we can distinguish three sub-textures from ABRDF slice. Then, we must characterize each image – photography or numerical representation.

Some works have shown the useful of Haralick features for texture classification [6]. The difficulty to get a good classification was to have the “good” image of the product surface, i.e. an image with a uniform or well directed lighting and a uniform or well directed viewing. Of course, it is necessary to have a suitable resolution. However, a human controller scans different lighting and viewing positions to get the right image. The right image is the one with the anomalies if they exist. Hence, the idea is to stick to the controller behavior and collect all the useful images of surface.

The Haralick features used to describe a single image. We do a recall of the generic method to compute a gray-level co-occurrence matrix (GLCM) and applied Haralick features [6], while GLCM method is one of the most well-known and widely used texture features. GLCM measures the spatial dependency of two gray-levels, given a displacement vector. Several works are identified as using GLCM method for defect detection [7].

The co-occurrence matrix method is a statistical method for textural defect detection. There are other statistical methods such as histogram properties, local binary pattern (LPB) or autocorrelation. There are also other methods: structural methods, filter-based or model based methods [7] [8]. In this paper, we focus on the co-occurrence matrix method to ease the demonstration.

We present the (extension) adaptation of Haralick to BTF and some results of Bidimensional Haralick Functions (BHF).

We show a comparison between results of BHF and results of Haralick features from 3 ABRDF images. We study the computational time and performance between the two methods.

2 CONCLUSION AND PERSPECTIVES

There are three perspectives of this work for visual inspection.

The first perspective is when visual inspection is applied to aesthetic field, there are two challenges:

1 – Have a Repeatable and Reproducible (R&R) method [9].

2 – Have a human perception-like processing with visual attributes.

We can use some elements of this paper for establishing features of visual data. This data can be used for trying to understand the human visual behavior, because we can verify uniformity and reproducibility of visual exploration and evaluation data.

The second perspective is for automatic defect classification. In fact, some works have shown the useful of Haralick features for automatic defect classification from different co-occurrence matrix of one image [10] [11].

The third perspective is to use the different well-known methods for editing BTF-like functions, such as Polynomial extended Lafortune Reflectance Model (PLM RF), Reflectance field factorization (PCA RF) or HSH [4] [5].

3 ACKNOLEGDMENTS

We thank our SYMME partners in MESURA project as well as Savoie Mont Blanc Industries and Conseil General 74 to enable us to carry out this research by giving us resources.

4 REFERENCES

[1] Krynicki, Jean-Claude. “Introduction to Soft metrology” XVIII IMEKO World Congress. 2006

[2] Eugène, Christian. “Measurement of “total visual appearance”: a CIE challenge of soft metrology.” 12th IMEKO TC1 & TC7 Joint Symposium on Man, Science & Measurement. 2008.

[3] Filip, Jiri, and Michal Haindl. “Bidirectional texture function modeling: A state of the art survey.” Pattern Analysis and Machine Intelligence, IEEE Transactions on 31.11 (2009): 1921-1940.

[4] Haindl, Michal, and Jirí Filip. “Visual Texture.” Advances in Computer Vision and Pattern Recognition. Springer-Verlag London, London, January (2013).

[5] Wang, Oliver, et al. “Material classification using BRDF slices.” Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on. IEEE, 2009.

[6] Haralick, Robert M., Karthikeyan Shanmugam, and Its' Hak Dinstein. "Textural features for image classification." Systems, Man and Cybernetics, IEEE Transactions on 6 (1973): 610-621.

[7] Xie, Xianghua. "A review of recent advances in surface defect detection using texture analysis techniques." Electronic Letters on Computer Vision and Image Analysis 7.3 (2008): 1-22.

[8] Nixon, Mark, Mark S. Nixon, and Alberto S. Aguado. Feature extraction & image processing for computer vision. Academic Press, 2012.

[9] Maire, Jean-Luc, Maurice Pillet, and Nathalie Baudet. "Gage R2&E2: an effective tool to improve the visual control of products." International Journal of Quality & Reliability Management 30.2 (2013): 161-176.

[10] Porebski, Alice. Sélection d'attributs de texture couleur pour la classification d'images. Application à l'identification de défauts sur les décors verriers imprimés par sérigraphie. Diss. Université Lille 1, 2009.

[11] Porebski, Alice, Nicolas Vandenbroucke, and Ludovic Macaire. "Iterative feature selection for color texture classification." Image Processing, 2007. ICIP 2007. IEEE International Conference on. Vol. 3. IEEE, 2007.

9398-21, Session 3

Goniochromatic difference between effect coatings: Is the whole more than the sum of its parts?

Jana Blahová, Technische Univ. Darmstadt (Germany); Eric J. J. Kirchner, Niels Dekker, Akzo Nobel Coating B.V. (Netherlands); Marcel P. Lucassen, LUCASSEN Colour Research (Netherlands); Lan Njo, Ivo van der Lans, Akzo Nobel Coating B.V. (Netherlands); Philipp Urban, Fraunhofer-Institut für Graphische Datenverarbeitung (Germany); Rafael Huertas, Univ. de Granada (Spain)

Special effect coatings have been increasingly used in many industries (e.g. automotive industry) over the past two decades. The measurement of perceived color differences on such coatings cannot be done by means of traditional color-difference formulas (e.g. dEab, CIEDE2000) as they lack to consider distinct optical properties such as coarseness, glint and goniochromatism. However, there is a need to ensure quality and colorimetric accuracy when designing and processing special effect coatings. This paper is motivated by this need and introduces a visual experiment intended to provide the basis for future work on a new generation of color-difference formula(s) particularly designed for effect coatings.

In daily situations we come across 3D objects covered with special effect colors (e.g. an automobile) and our judgment is made in its entirety. Is it possible to predict the overall color difference of special effect colors by combining multiple formulas measuring the color-difference for each geometry? This paper pertains to the diversity of perceived difference between the judgment made for two geometries at the same time and made separately in two sessions.

Observers were asked to evaluate the perceived difference of each test pair under the mid specular angle 45:0 and far specular angle 45:65 referred by ASTM E 2194-03 as 45 and 110 geometry, respectively. We divided our experiments in two parts, which we will refer to as approach 1 and approach 2. In approach 1, observers wereshown the samples under the two geometries immediately one after the other, whereas in approach 2, two sessions were carried out with at least a day inbetween, each session being dedicated to a single geometry.

We conducted this experiment following the so-called Grey Scale method. A total of 66 metallic patches were prepared: 14 for reference and 52 for test. They were created by spraying various special effect coatings on a steel substrate with an automated HTE (High Throughput Experimentation) equipment [1]. The reference patches were prepared with grayscale coatings, in order to have 7 pairs with color differences ranging from about 0.8 to 3.7 CIEDE2000 units. As for the test patches, 26 pairs were coated based on multiple use. Detailed specifications will be given in the final paper.

A group of 7 observers (2 women, 5 man) with color-normal vision was asked to find the reference pair (grayscale) that best matches the perceived difference between each test pair. This strategy is referred to as the Gray Scale method. The appearance of metallic patches was viewed from a distance of 500mm for test and of 550mm for reference. A black mask in size of 100x110mm was placed on each test pair to specify a visual field of 11.5?. A similar mask was used also in case of reference pairs with a field of vision of 10?. Note, that each pair was placed horizontaly without space in between.

The experiment was conducted in a completely darkened room to minimize flare. Two similar light sources (with a CCT of 4700K each) were used to illuminate the test and reference pairs separately. The incident angle of the light on the test pairs was fixed throughout the experiment (45?). Observers were asked to look at the test pairs under two different angles, corresponding to the 45 and 110 geometries.

The results indicate that there is no significant change in the observers' ratings when the two different geometries are shown immediately after one another or in two different sessions. The mean difference between scores obtained when using approach 1 and 2 was 0.32 with standard deviation of 0.63. Detailed results will be given in the final paper.

These results will serve as a basis for future investigation on deriving a color difference equation, or modifications of existing equations, dedicated and specific for metallic coatings.

[1] E. d. Wolf, "HTE and screening techniques in paint development," 2011.

9398-22, Session 3

Visual comparison testing of automotive paint simulation

Gary W. Meyer, Univ. of Minnesota, Twin Cities (United States); Curtis Evey, Dassault Systemes (United States); Jan Meseth, Dassault Systemes (Germany); Ryan Schnackenberg, Dassault Systemes (United States); Charles Fan, General Motors Co. (United States); Chris Seubert, Ford Motor Co. (United States)

A research project is underway to determine the remaining sources of inaccuracy when commercially available software is used to simulate the color of car paint. Although compelling realism is easy to achieve today with computer graphic software, manual intervention (so-called tweaking) is often necessary to produce a final color result that is acceptable within the automotive industry for design, engineering, and marketing purposes. The goal of this project is to determine whether these ad hoc adjustments are necessary due to such factors as a flaw in the simulation algorithm or the calibration of the display device, or whether they are caused by something not yet accounted for. The solution of this problem is important in order to improve the quality of the final result and to decrease the amount of time required to achieve it.

The initial goal of the project is to determine whether computer simulations, produced from data acquired with industry standard measurement devices and displayed on calibrated monitors, can be successfully compared to samples observed under controlled lighting and viewing conditions. The paint sample properties to be compared are the color (as measured with either the X-Rite MA98 or the BYK-Gardner BYK-mac multi-angle spectrophotometers), the orange peel (as measured with the BYK-Gardner wave-scan), and the flake structure (as measured with the BYK-mac). The sample is observed in a booth (the BYK-Gardner byko-spectra effect) where the viewing angle relative to the light source can be carefully controlled. The light booth was rendered using CAD data provided by the manufacturer. The simulation (produced using RTT DeltaGen) was reproduced on a monitor (the Eizo CG275W) that has known colorimetric properties and is able to automatically calibrate itself.

Experimental Setup

The Byk Gardner byko-spectra effect light booth was employed to observe the real automotive paint panels. This device is specially manufactured for this purpose and contains two different types of light sources to illuminate

the panels. A fluorescent tube provides diffuse illumination for the panel at a constant incident angle of 45 degrees, and an arm on the side of the booth allows the sample to be rotated and viewed at a series of aspecular angles (-15°/15°/25°/45°/75°/110°). The booth also contains three projector bulbs that illuminate the paint sample from three different directions so that the flake sparkle can be seen. In the work reported in this paper we only used the diffuse fluorescent illumination.

To allow similar observations to be made from the computer monitor and the light booth it was necessary to slightly modify the usual operating procedure for the booth. The light level in the booth was decreased so that light reflected from a 100% reflecting surface in the booth was at the same level as the peak light output of the monitor (110 cd/m2). This was accomplished by placing neutral density theatrical lighting gels in front of the fluorescent light source. For most of our work the sample in the booth was viewed with the front door open instead of through the slit in the front door. This was done to avoid the construction of a lighting hood for the monitor that would have introduced an artificial viewing circumstance for the display.

The computer graphic simulation of the light booth was produced using RTT DeltaGen software. A geometric model of the booth was created from the CAD data provided by Byk Gardner, and appropriate textures and colors were applied to the booth's surfaces. Interior lights and illumination were modeled using discrete lights available in DeltaGen and by ray tracing an environment map to "bake in" certain lighting effects. Using D65 illumination, multi-angle measurements of the paint panels were made using both a Byk Gardner Byk-mac and an X-Rite multi-angle spectrophotometer. These measurements were read into the DeltaGen program via a computer interface, and a specially written shader was employed to generate the correct sRGB value on the screen. Because only the diffuse fluorescent light was used to illuminate the samples in the booth, visual inspection showed it was only necessary to set the "grain" parameter in the shader to correctly simulate the appearance of the metal flakes. A empirically derived linear correspondence was employed to set the value of the "grain" parameter from the Si Byk-mac measurement.

The simulation was displayed on a calibrated Eizo Color Edge Model CG 275W monitor. The monitor's auto calibration procedure was used to balance the monitor for a D65 white point, 2.2 gamma, and peak luminance of 110 cd/m2. Although spectral measurements of the light source in the booth were made using a StellarNet spectrometer, for the purpose of the colorimetric calculations it was found to be more effective to assume a D65 light source and to use the DeltaGen post process feature to adjust the final color temperature of the light and the scene. We speculate that this is necessary due to drift in the color temperature of the light source with time. Measurements were also made using a telephotometer to confirm that the distribution of light intensity in the simulation is consistent with the distribution in the actual scene.

Experimental Results

We have begun a series of tests to determine the level of equivalency between a simulation of the paint panels in the light booth and the actual color panels displayed in a real light booth. In the first test we asked subjects to offer opinions about the quality of the match between panels presented in a simulated light booth and panels exhibited in an actual light booth. The goal of this simple test was to determine whether the comparison was close enough to warrant a subsequent time consuming task based evaluation. In the second test subjects were asked to compare two paint panels and determine whether they did or did not match. They were asked to perform this task in two separate conditions: examining the panels in a simulation and evaluating them in the light booth.

In the first test subjects were requested to examine the simulated paint panel and the actual paint panel while observing them from the same point of view. Subjects were asked to stand between the simulation and the light booth, and they were told to look back and forth between the monitor and the booth. After inspecting both they were asked to grade the quality of the match by assigning a numeric rating between 0 and 5. Solid colors were presented at only the 45 degree viewing direction and the metallic colors were shown at two different viewing directions. There were a total of 14 different color panel and viewing angle combinations. Eighteen human evaluators participated in the test, eight were experts from industry and ten were non-experts.

The result of this simple comparison test is that most discrepancies between the simulation and the real panels were considered to be small. The average rating for all subjects was 3.7 and none of the panels was given an overall average rating below. The solid colors were given a generally higher rating than the metallic colors, and the presentations at 45 degrees were typically rated higher than those at other angles. The color experts were generally more critical than the non-color experts, but the difference is not great (0.3 on average) and their impressions are well correlated with those of the non-experts. The test also permitted the refinement of viewing conditions and experimental procedures for subsequent tests.

In the second test subjects were asked to perform a color paint comparison task that is representative of the type of judgment that is made every day in the automotive industry. For this test, fifteen color samples were submitted by car manufacturers and paint suppliers. For each sample, two identically sprayed standards were provided and one sample that was considered, by industry standards, to be a mismatch. Nine subjects (color normal) were shown pairs of the samples, as physical samples in the light booth and as computer graphic simulations on the monitor, and they were asked to decide whether the samples matched. At their discretion, the subjects were allowed to rotate the samples to different viewing angles, using a mouse button for the computer simulation and the light booth handle for the real panels. To make these judgments they were asked to position themselves so that they could see the samples through the viewing slot in the light booth (the door was subsequently opened for the tests) and so that they had a similar view of the samples displayed on the monitor.

The results of these color comparison tests show that there was little difference between the decisions made using the light booth and the computer monitor. Although the average error rate when using the computer simulation was slightly higher than the error rate achieved when making the comparison in the light booth, this difference is not statistically significant ($z = 3.673$, $p < .05$). A test of equivalency shows that if there is an average error rate difference of 1.5 or lower (10% for 15 samples) it is impossible to tell the difference between the results obtained using the computer simulation and those achieved using the light booth ($t(16) = 5.353$, $p < .05$ and $t(16) = -4.742$, $p < .05$). A significant finding is that it takes only half the amount of time to conduct the virtual tests as it does to perform the experiments that use the real panels.

Discussion

The results obtained to this point are encouraging and indicate that there is potential for making color critical decisions from a computer simulation. A simple comparison between the simulation and the light booth shows that the rendering makes sense and captures the important appearance properties of the paint. While cross media color comparisons are inherently problematic, it is important to confirm that the picture is consistent with what is seen in the booth. The color matching test shows that it is possible to use a simulation to duplicate the results for a typical industrial appearance test. This result is valuable because it demonstrates the potential to use computer graphic simulations to solve real color matching problems.

There are also limitations to the work that has been accomplished thus far. Decreasing the light level in the booth to match the peak monitor output undoubtedly increases the error rate when working with the real physical samples. In the future, this problem could be solved by employing high dynamic range monitors that output significantly more light. It is also difficult to put a single computer graphic image, synthesized from a fixed point of view, on the same footing with a light booth that is observed from a range of viewing positions and a continuous range of viewing directions (even though we tried to control for these variations). Tracked virtual reality systems may eventually overcome these limitations. Finally, we need to run our color matching experiment with a group of experts from industry. These results will tell us whether the work has relevance to these specialists.

Conclusions

These results suggest that standards for creating, displaying, viewing, and comparing automotive paint simulations can improve the results that are achieved when using these tools. To the extent that observations of the physical sample and the simulation can be put on an equal footing, computer simulations derived from standard industrial measurements look very similar to the actual paint samples. This result can at least be used as a starting point for evaluating color in synthetic automotive imagery before

considering the confounding effects of complex lighting and complicated inter-reflections. These standards could also play a role in developing virtual techniques that could be used for qualifying paint suppliers and for solving color harmony problems.

9398-23, Session 4

Goniometric properties of paints and varnish

Paola Iacomussi, Michela Radis, Giuseppe Rossi, Istituto Nazionale di Ricerca Metrologica (Italy)

Visual appearance is fundamental for our interaction with the real world and, especially in works of arts, colour perception is the most important one and involves at least three different aspects: the way the artist wants to give the visual message, the level of conservation of the work and the exhibition conditions. Restorers and curators must consider the first aspect setting up the correct actions for the other two: restoration (pictorial and varnish removal and addition), and exposition location (lighting set up, observers positions and level of adaptation). All these aspects play a relevant role and affect our ability to perceive and interpret the message and the colour of the work of art.

Painters know that the perception of their works is strongly influenced by the gloss: to enhance saturation and also to protect the artefact, artists are used to put some varnishes on. Varnishes are transparent and colourless substances that are applied, as final transparent coating, to paintings surface to improve the aesthetic and secondarily to protect them from abrasion and deposit. The painting surface usually has microscopic imperfections as, for example, pigment particles protruding from it, that cause a certain degree of diffuse reflection of light on the surface. The diffuse white light causes a de-saturation of colour.

Varnishes make uniform the microscopic surface imperfections of paint layer, causing a reduction of diffuse light reflection on the surface and increasing the colour saturation. Smooth painting surface has more saturated colours.

Protective varnishes can be glossy or matt, natural or synthetic, and their behaviour in reflecting light in space strongly affects the colour appearance of paints: increases chroma and brightness and intensifies the depth sense of picture. Great painting masters knew very well, by experience, how the varnish affects the paints appearance, but restorers not so well and, differently from the past, painting are now exposed to public in museum, and the appearance is related also to the geometrical conditions of exposition.

In order to investigate the appearance of colours of painting with protective varnish and to provide useful indication to restorers in choosing the protective varnish and to works of art exhibition designers, goniometric investigations on photometric and colorimetric properties of varnished tempera are necessary. This goniometric characterisation is not performed in the field of cultural heritage, because restoration labs are not equipped with a goniophotometer, but the relevance of this type of characterisation is clear, especially for materials used in restoration and integration.

The spectral reflection factor of different tempera coloured samples with four different protective varnishes (glossy and matt, synthetic and natural) is under analysis at INRIM using a Perkin Elmer Lambda 900 spectrometer and the INRIM goniometer for materials analysis.

The INRIM goniometer is based on height different mechanical (and computer controlled) movements in order to arrange the sample in all possible orientations in space respect a fixed lighting source (CIE standard illuminant A) and a detector able to move only in the horizontal plane. The source is kept fixed at about 3m from the goniometric centre, in order to reach higher stability and accuracy during the measuring time. The photometric measurements are performed using a relative technique based on CCD camera, while the spectral measurements for the tempera and varnish investigations are performed with a portable spectrophotometer PR650. The paper will describe in detail the goniometer and the measurement method.

The samples are wood based with a preparatory layer of cementite, 20% diluted with aqua regia, and two layers of gouche tempera, branded WINSOR&NEWTON. Seven different colours are considered and for each colour, two samples are made. Each sample is divided in 3 zones: on two of them a protective varnish, at two glossy levels, is applied, while the remaining part the sample is not covered by varnish. In this way is possible to compare goniometric appearance and colorimetric properties of varnished and unprotected tempera.

Natural (organic) and synthetic varnishes are under test. The organic resins used to produce the natural varnish, have a good optical quality, because they have very low molecular weight (about 400-700 a.m.u.) and a relatively high refractive index. Low-molecular-weight resins often produce coatings of low viscosity that make uniform the rough painted substrates and produce more colour saturation, but are very susceptible to photochemical degradation and to auto-oxidation reactions. Modern synthetic varnishes are generally more stable than natural ones and less sensitive to ageing: synthetic resins build high viscosity varnishes which reproduce the roughness of the underlying paint layers and, therefore, usually produce less colour saturation, but enhance the artist way of apply colours. The knowledge of the goniometric appearance of varnished paint helps restorers choosing the best varnishes not considering only the chemical and mechanical stability.

Each zone of the samples is measured using a double beam spectrophotometer Perkin Elmer Lambda 900, equipped with a 150 mm integrating sphere (measurement condition 8/d), in the range 250-2500 nm with 1 nm step. The goniometric investigations are performed using the INRIM goniometer for material analysis for two directions of incidence and sixteen direction of observation. The directions of incidence are typical of museum lighting set-up (30° and 60°) and the sixteen directions of observation simulate different observer positions in a museum exposition.

A good lighting set-up allows to enjoy a works of art in the best possible viewing conditions, considering both the environment around the painting and the correct perception of colour, especially regarding geometric metamerism, glare and saturation. The knowledge of spectral-goniometric properties of varnished painting will be helpful for restorers in order to chose the protective varnish regarding the final effect on colour perception depending on the condition of exposition and also to lighting designers to set up the best condition of exposition.

Preliminary investigations show that at 30° of incidence the luminance factor is, for all directions of observation, lower than at 60° of incidence and also colours are less altered. Glossy and matt varnishes affect differently the width of the cone of colour desaturation, but again the preliminary results show, again, a smaller width for the direction of 30° incidence. These results, even if are only raw and refers to very few samples, allow to say that in works of art exposition it should be better to prefer lighting set up with the direction of incidence on the painting of 30° or lower, in order to enjoy the best appearance of the colours.

This work is part of the European research project “Multidimensional Reflectometry for Industry” EMRP Project IND52 xD-Reflect. The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union.

9398-24, Session 4

Goniochromatic and sparkle properties of effect pigmented samples in multidimensional configuration

Andreas Hoepe, Kai-Olaf Hauer, Sven Teichert, Dirk Huenerhoff, Christian Strothkaemper, Physikalisch-Technische Bundesanstalt (Germany)

The Physikalisch-Technische Bundesanstalt (PTB), the National Metrology Institute of Germany, is operating two robot-based gonioreflectometers and an integrating sphere facility for the measurement of visual appearance-related quantities. A general aim of current research in reflectometry is to find an irreducible set of scaling parameters for goniometric effects like lightness-flop, color-flop or sparkle/graininess. The variety of partially new

effects generated by dedicated effect pigments created a complicated situation for metrology and the traceability of such materials. In the course of the European Metrology Research Program (EMRP) a joint research project entitled "Multidimensional Reflectometry for Industry" was formed in order to give industry support to overcome these challenges.

The EMRP is a metrology-focused program of coordinated Research & Development (R&D) funded by the European Commission and participating countries within the European Association of National Metrology Institutes (EURAMET). It supports and ensures research collaboration between the partners by launching and managing different types of project calls. The presented work is part of this project, with its short name xD-Reflect indicating the multidimensional aspects (x > 3) of modern reflectometry. The xD-Reflect project is conducted by a consortium of 8 National Metrology Institutes (NMIs) and 3 universities or research institutes, respectively. The general objective of xD-Reflect is to meet the demands from industry to describe the overall macroscopic appearance of modern surfaces by developing and improving methods for optical measurements which correlate with the visual sensation being evoked.

The expression "goniochromatic" denotes materials having a strong angular dependent reflection behavior. They show a color impression that depends on the spatial arrangement of illumination and observation relative to the surface of the artifact. The goniochromatic colors are produced based on the interference effect in which the incoming light is partly reflected from the particle surface and partly refracted through it. Depending on the coating thickness and variation of the angle of incidence of the applied radiation, a colorful appearance of the reflected light at various spatial directions is produced. However so far, goniochromatic pigments and paints are a challenge for metrology as their color, saturation and brightness all change with visual perspective and the ambient lighting conditions. This makes it necessary to overcome the fixed and standardized directional reflection geometries of the past like 45°:0° and 0°:45° recommended by the CIE (Commission Internationale de l´Eclairage).

In order to fulfill the outlined requirements, measurements of the reflection behavior of selected effect pigmented samples, from an entity of more than 50 samples in stock, were performed for angular configurations in three-dimensional space. For this purpose different pigments in different concentrations and different gradations applied to different backgrounds (black/white) were studied with respect to the described effects of goniochromatism and sparkling in directional reflection geometries.

The measurements were performed on black and/or white primed metal sheets coated with different effect pigments from the line-up of the MERCK pigment families of Iriodin®, Colorstream®, Miraval® and Xirallic®, as well as Cromaflair® pigments of JDS Uniphase Flex Products.

The investigations can be distinguished into two parts, so-called "in-plane" and "out-of-plane" measurements. The differentiation between both configurations is based on the orientation of the vectors of incident and reflected radiation as well as on the orientation of the sample normal. If all three vectors are in one plane this is the "in-plane" configuration. If they are not in a common plane, which is in fact valid for most of "real live" geometrical configurations, this is denoted as "out-of-plane".

Extensive "in-plane" measurements were performed at three different incident angles of 15°, 45° and 65°, according to the fact that these angles are to some extend realized in commercial multi-angle spectrophotometers. While the incident angle was fixed, the reflection angle was varied ranging from +80° to -80° relative to the surface normal in steps of 5°. These data-points are forming the so-called aspecular line. In addition, so called "15°-interference lines" in Cis- and Trans-configuration for varying incident angles 0° to 80° (step size also 5°) were determined. The notation Cis- and Trans- comes from chemistry and denotes the orientation relative to the specular peak. Here 15°-Cis means the measurement at an angle of 15° in the direction of the light source and 15°-Trans denotes an angle in the opposite direction, away from the source.

From the measured spectra CIELAB color coordinates L*a*b* for different illuminants and standard observers can be calculated. The data measured in the different geometrical configurations span a wide range within the a*b*-plane, covering thereby all four quadrants and showing the need for multi-geometry measurements in order to fully describe the goniochromatic behavior of effect pigments. In a further step, the three-dimensional reflection behavior of the samples for various incident angles was determined ('out-of-plane' measurements). An example for this approach is a fixed incident angle of 45° with the measurement of the spectral radiance factor at 141 positions spread over half of the hemisphere above the sample surface. The angular range is in this case 0° to 80° (angular spacing 10°) for the polar angle and 0° to 180° for the azimuth angle (with variable angular spacing).

The second aspect of the project is the study of sparkle properties of effect pigments. Sparkling can be recognized as many tiny but very intense light spots, like bright stars twinkling at the night sky. Sparkle is an obvious effect to human observers, but cannot be measured with prevalent spectrophotometers because of its small length scale. Furthermore, the same effect pigments, when measured and perceived in diffuse illumination, causes a new visual texture effect named graininess or coarseness. There is currently only one commercial instrument on the market measuring a quantity called "sparkle index" allocating thereby a scale to this effect for quantification. The drawback of the current situation is that the definition and basis of computation of this quantity is not revealed. Also the correlation of this index as a pure number with the sensation of the human eye is unclear. Driven by this matter of fact extensive research for an open source and traceable sparkle index is necessary which is performed in our lab with a combined Imaging Luminance and Imaging Color Measurement Device (ILMD/ICMD) having HDR capability (HDR: high dynamic range) and due to a special imaging optics a resolution of approx. 28 μm on the device under test.

Acknowledgement: This report was compiled within the EMRP IND52 Project xD-Reflect "Multidimensional reflectometry for industry". The EMRP is jointly funded by the EMRP participating countries within EURAMET and the European Union.

9398-25, Session 4

Anisotropic Materials Appearance Analysis using Ellipsoidal Mirror

Jiri Filip, Radomir Vavra, Institute of Information Theory and Automation (Czech Republic)

Digital material appearance relies on a number of representations. They range from well established reflectance representations such as BRDF to more complex texture and light scattering representations based on BTF, BSSRDF.

While the latter group captures material appearance in all its complexity; including shadowing, masking, inter-reflections, and subsurface scattering, the former is based on simplified assumptions of opaque and flat materials omitting any texture information. However, many appearance measurement and modeling approaches go even further and reduce their processing costs by sacrificing material anisotropy.

Generally, an anisotropy is the property of being directionally dependent, as opposed to isotropy, which implies identical properties in all directions. It is most often observed in chemistry for the evaluation of property in different crystallographic directions or in physics for analysis of the directional speed of light.

In computer graphics isotropy and anisotropy is used for the evaluation of directional material's appearance as well as for capabilities of related methods. When a material's reflectance is constant for fixed view and illumination irrespective of the rotation of the material around its normal the material is considered isotropic; otherwise it is considered anisotropic.

Anisotropic materials, due to their atypical expensive appearance, are often used for achieving of unique eye-catching look of many man-made products. Fabrics for fancy apparel clothing are probably the most typical example. Anisotropic finishing of plastics or metals is often used to create an attractive appearance of personal items or electronic gadgets. Genuine wooden materials are also very popular due to their specific anisotropic behavior.

Why is the capturing of visual anisotropy difficult? Mainly due to the fact that it enormously expands measurement state space when compared to isotropic material capture. In the case of BRDF, one has to capture four dimensions instead of three. Therefore, it is time-demanding to sample

this space uniformly while still maintaining results comparable to isotropic measurements.

In this paper we study materials' visual anisotropy and introduce an inexpensive and rapid detection technique based on only an ellipsoidal mirror and compact camera. Our technique allows us to determine the strength of anisotropy, the main anisotropy axes, and the width of anisotropic highlights.

Although the concept of BRDF measurement using parabolic or ellipsoidal mirrors has been already researched, it requires a relatively complex and calibrated setup of a dedicated camera and projector.

All these setups place the measured material into focal point of the mirror, share the optical and illumination axis using coaxial pair of camera and projector, and thus allow the recording of multiple illumination or view directions in a single image.

The main advantage of such an arrangement is the elimination of any mechanical component in the measurement setup. Principal disadvantages are: often a limited range of recorded elevation angles, variable reflectance attenuation across the elevations, or a low dynamic range of the measurements.

Most setups place the material into a mirror focal point in such a way that the material is aligned with the axis of the mirror. This is not ideal for fast anisotropy detection as it requires sample extraction and positioning inside the mirror.

Our anisotropy detection technique originates out of these setups; however, it is considerably simplified. It consist of an ellipsoidal reflector, wit an opening at the narrowest part. It is then attached to the measured material. This reflector is photographed by a compact camera having an optical axis aligned with the reflector axis also representing normal of the measured surface.

In contrast to other setups, we do not attempt to record the material's BRDF, but only its anisotropic behavior. Therefore, we do not sweep the mirror by a controlled ray of light from the projector, but use instead the flash from our camera. We do not use any special gantry and the image is taken from a distance 6 feet. This is done in order to capture the reflector (in diameter 4 inches) approximately symmetrically.

Such an omni-directional illumination lights the measured material from many azimuthal angles. As the material sample is not positioned in the mirror's focal point, illumination elevations angles vary. This allows for the capturing of azimuthal direction of anisotropic reflection. In the image of material at the opening (1 inch wide) taken by the camera is each location illuminated from two elevations at azimuths 180 degrees apart. Naturally, these directions are averaged in the captured photo. Mean illumination elevation angle across entire material plane is 50 degrees.

When an isotropic material is analyzed, the captured image contains a circularly shaped peak in the center. This is caused by the accumulation of reflected rays at that location. However, for an anisotropic material we additionally observe a couple of triangle shaped highlighted areas running symmetrically from the center to the edges of the opening. Their azimuthal orientation coincides with the direction of the main anisotropy axis/axes, while their width corresponds to width of the anisotropic highlight for the illumination elevation 50 degrees.

To validate our results, we used eight complicated anisotropic fabric materials and compared their microstructure scans with the detected anisotropy direction recorded by our method. Two of the materials exhibited distinct anisotropy highlights corresponding to two anisotropy axes. When detailed BRDF of the tested materials were recorded and rendered we could observe an exact coincidence of the detected and real anisotropy highlights and their directions.

Finally, we have shown on three dense BRDF measurements how detected information about anisotropy highlight direction and width can improve BRDF measurement strategy. We compared a uniform sampling with adaptive sampling using the detected anisotropy axis and highlight width. We placed samples along anisotropic highlights. Their contours were obtained for fixed azimuth of half angle direction in Rusinkiewicz's parameterization. This azimuth corresponds to the detected anisotropy axis. Additionally, we added two surrounding contours with the distance from the main one defined by detected highlight width. Finally, we applied radial basis function interpolation to reconstruct BRDF.

In this way we achieved better reconstruction than uniform sampling given the same number of samples (1000 reciprocal). We compared the reconstructed BRDFs and their renderings using computational measures (RMSE,PSNR,VDP2) and found a notable visual improvement when the detected information about material anisotropy was used for the generation of a sufficient sampling pattern.

The introduced method allows for a very convenient and fast detection of material's anisotropy strength, main anisotropy axes, and corresponding highlights shape. These features are beneficial for variety of tasks, ranging from material appearance acquisition (as shown in this paper) to its automatic classification or image-based material retrieval.

9398-26, Session 4

Changing the color of textiles with realistic visual rendering

Mathieu Hébert, Univ. Jean Monnet Saint-Etienne (France) and Institut d'Optique Graduate School (France); Lambert Henckens, Lembart S.A.S. (France); Justine Barbier, Lucie Leboulleux, Marine Page, Lucie Roujas, Institut d'Optique Graduate School (France); Anthony Cazier, Univ. Jean Monnet Saint-Etienne (France) and Institut d'Optique Graduate School (France)

Everyone knows the variety of appearances of fabrics that can be obtained according to the selected fibers, the structure of the yarns, their preparation by spinning, coating and dyeing, the weaving and the finishing... Infinite variations of color, shine, glaze, transparency and texture may be obtained. Their visual rendering is therefore difficult to predict from these numerous parameters, and only the production of samples allows visualizing the desired rendering. However, producing samples is often an expensive and time-consuming constraint for professionals. A realistic simulator of fabrics, where parameters could be virtually modified, would therefore represent a high benefit. This is especially true in applications where the fabric is integrated into a global multi-material design and thus needs to satisfy precise rendering constraints.

As a first attempt towards this objective, we studied the possibility to visualize on a computer display the appearance of a fabric when the colors of the warp and weft yarns are varied while keeping their respective gloss and shine properties. The proposed approach relies on images of one fabric sample where the two types of yarns can be distinguished thank to their different optical properties and localized in the image pixel by pixel. A simple optical model is then used to modify the color of each pixel.

We considered two types of fabrics where this approach seemed to be applicable. The first one, obtained by Jacquard weaving, was made of polyester yarns with different colors for the warp and the weft. By making a color picture of the fabric with a table scanner and performing a thresholding on the color channels, we could identify the respective positions of the colored yarns and modify the hue and chroma, while preserving their lightness in order to keep the shadows and the weak shine effect.

The second type of fabric was satin ribbons made of warp yarns in polyester and weft yarns in metallo-plastic material. Thresholding of a color image is not possibly anymore because the reflectance of the highly specular metallic yarns strongly varies due to the multiple orientations of their surface. However, we can use the fact that the polyester, a dielectric material, and the metallic yarns do not reflect polarized light in the same way, as a consequence of the Fresnel formulas for dielectrics and metals. We developed an angular-polarized imaging system where a collimated white light source rotates around the ribbon sample, and a black&white camera preceded by a telecentric imaging objective observes the sample in the normal direction, with a x5magnification. Polarizing filters were placed after the light source and before the camera. The gloss measurements were performed on a ribbon made of black polyester and achromatic metallic yarn. For each incident angle of light, we captured images in parallel- and cross-polarization modes. In cross-polarization mode, the black polyester is invisible because the polarized light reflected by its surface is fully blocked

by the polarizer; the captured image therefore shows only the specular reflections by the metallic yarns. In parallel-polarization mode, the captured image shows the specular reflections by both polyester and metallic yarns, and by subtraction of the two images, by assuming that the metallic yarns reflect the same amount of light in the two polarization modes, we obtain an image of the specular reflections by the only polyester yarns. This procedure was repeated for many incident angles of light from 5° to 80°.

For the simulation of piecedeyed ribbons with polyester warp and metallic weft, where the two types of yarns get different colors, we used a simple reflection model based on the following assumptions: the shine of the metallic yarn is not modified in radiance and angular distribution, but simply colored with constant saturation and hue; the gloss of the polyester yarns remains achromatic, but an angle-independent colored light component is added. Therefore, the images made at each incident angle where transformed in the following way: we multiplied the gloss image of the metallic yarns by distinct R, G and B values representative of the metallic color; we added a constant RGB value to the non-zero pixels in the gloss image of the polyester; we finally summed the two color images. The introduced RGB values may be measured for example on a solid color ribbons. In the resulting color images, we can see the color, the shadows, and the shine of the polyester yarns. We can also preview the twinkling effect of the ribbon by displaying successively the images corresponding to the different incidence angles of light.

After having explained the optical concepts underlying the proposed methods, we will discuss their limits. One limit is the fact that the colors are displayed on a computer screen. The perception of colors is therefore different from real surfaces, and the texture resolution is limited by the resolution of the screen. However, we can compare on the display the simulated images of ribbons with pictures made of real ribbons in the same illumination and viewing conditions. We could verify with yarn-dyed Jacquard fabrics and piece-dyed ribbons that the images simulated with the proposed methods reproduce real pictures of them with a visually acceptable accuracy.

9398-27, Session 5

3D printed glass: surface finish and bulk properties as a function of the printing process *(Invited Paper)*

Susanne Klein, Hewlett-Packard Co. (United Kingdom); Michael Avery, Ctr. for Functional Nanomaterials (United Kingdom); Robert Richardson, Paul Bartlett, Univ. of Bristol (United Kingdom); Regina Frei, Univ. of Portsmouth (United Kingdom); Steven J. Simske, Hewlett-Packard Co. (United States)

3D-printing, along with other additive manufacturing (AM) and rapid prototyping (RP) techniques, involves building up structures in a layer by layer fashion based upon a computer design file. Such techniques are well suited to the production of one-off, complex structures that would often be difficult to produce using traditional manufacturing methods. There has been rapid growth and interest in this field during recent years and a range of techniques are now available which make use of many common materials such as plastic, metal, wood and ceramic. However, relatively little has been done to develop AM using glass.

Most commonly today glass objects are made from a melt. A molten glass blob is taken from the melt and either blown into a mold or free blown. During the blowing process the glass solidifies and after further cooling the typical shiny, glassy, surface and an ideally bubble free bulk is achieved. The much older glass making technique is the so called kiln glass method already used by the Egyptians. Glass powder or frit is filled into a mold and fused in an oven, the so called kiln. In this process shaping takes place at room temperature and fusing is a second step. The smaller the frit size the finer the detail on the finished piece but the less transparent it is.

The kiln glass process lends itself to 3D printing since the printing, i.e. shaping, can be done at room temperature. Direct glass printing methods are powder bed methods or extrusion methods. A pattern or mold is generated during indirect glass printing methods.

Since glass is turning into a viscous fluid during the firing process any shape generated by 3D printing has to be supported during fusing. Depending on which 3D printing technique is used to produce the so called greenware and what support material surrounds the object during firing the surface finish and the bulk properties differ greatly. We will report different techniques and the physical properties (Young's modulus, opacity and density) of glass generated by them.

9398-28, Session 5

Color-managed 3D printing with highly translucent printing materials

Can Ates Arikan, Alan Brunton, Tejas Madan Tanksale, Philipp Urban, Fraunhofer-Institut für Graphische Datenverarbeitung (Germany)

Many 3D printing applications require the reproduction of an object's color in addition to its shape. This is of particular importance for design-prototypes or for 3D copies where a texture-mapped 3D scan of an object shall be reproduced in a color consistent way.

Today, only two commercially available 3D printing technologies allow full color reproduction: the powder-binder and the layer-laminated method. Both are based on inkjet technology to apply CMYK inks onto a base material (white powder or paper) of which the 3D shape is formed. These base materials possess only low translucency and thus have comparable optical properties as known from 2D printing on paper. To characterize these printers, spectral measurement instruments from the graphic arts industry (circular 45°/0° geometry) can be used [1]. Due to uneven surface topography of powder-binder printouts, Stanic et al. [2] recommend to use sphere spectrophotometers (d/8° geometry) for characterization. They observed that such spectrophotometers yield more reliable measurements than 45°/0° instruments.

Another 3D printing technology, called multi-jetting or poly-jetting, allows combining multiple materials into a single object. It uses also inkjet technology but to directly build the 3D shape from UV curing inks. This technology enables a higher resolution and improved surface finish compared to the previously mentioned techniques. Furthermore, it is able to print fully transparent materials as well as materials which have different tactile properties. This makes multi-jetting very interesting for graphical 3D printing.

We will first describe a color-managed 3D printing workflow consisting of:

1. Voxelizing the texture-mapped 3D model that serves as the input of the workflow.

2. Mapping texture colors to tonal values by ICC profiles (which include gamut mapping and separation).

3. Mapping tonal values to the surface of the 3D model's voxel representation.

4. Creating material halftones from the tonal values and assigning them to voxels.

5. Extracting slices from the voxel representation for layer-wise printing

This 3D printing workflow is an adaptation of the standard 2D printing workflow. In addition to the geometry processing of applying color onto curved surfaces (2D manifolds), a major difference to 2D printing consists in characterizing the printing system, i.e. creating the ICC profiles. This is because printing materials available for multi-jetting are highly translucent causing problems in measuring these materials with standard spectralphotometers as described below. Even though we expect that less translucent materials will be developed for such printers in future, employing translucent materials for 3D color reproduction might be advantageous, e.g. to reproduce material appearances such as skin, which consist of translucent layers and have a high degree of subsurface scattering. Therefore, we focus in this paper on the measuring aspects required to generate ICC profiles for multi-jet 3D printers employing highly translucent materials.

Such materials require another approach for colorimetrically characterizing these printers because spectral measurements obtained from instruments used in the graphic arts industry (circular 45°/0° geometry) do not correlate well with perceived colors viewed under diffuse illumination:

A spectrophotometer illuminates a spot on the material's surface, measures the response and relates the measured data to a reference for obtaining the reflectance factor. If the spatial intensity distribution of the light spot and its convolution with the light transport's point spread function (which depends on the printing material) are not equal over the measuring area, some light is transported within the material away from the measurement area and does not contribute to the obtained reflectance factor. We measured the reflectance factor of a white material using a spectroradiometer in a viewing booth (the white reference used was measured at the German metrology institute PTB, see [3] and with two common spectrophotometers used in the graphic arts industry (same backing was employed in all measurements). The results show almost similar reflectance factors for both spectrophotometers with a CIELAB value of approximately (L*, a*, b*)=(70, -6, -8) computed for CIED50 and the CIE 1931 observer. The spectroradiometer results were (L*, a*, b*)=(88.5, 0.9, 8.6), which agrees much better with the perceived color.

This indicates that commercially available spectrophotometers used in the graphic arts industry cannot be used to characterize 3D printing systems with highly translucent printing materials. In order to colorimetrically characterize the printer, a spectral or colorimetric camera may be used to capture the printing target under 45°/0° illumination (e.g. in a viewing both). We use the Canon 5D Mark III that has a Vora Value [4] of 0.9455 (i.e. it is almost colorimetric) and a filtered tungsten SoLux daylight lamp which is a spectrally-broadband CIED50 simulator. Because of subsurface light transport, the reflectance factor of each patch is biased by adjacent patches. Therefore, the patch-size must be large (we used squares with an edge length of 1.25 cm) and only pixels within the center of each patch have to be used for colorimetrically characterizing the printer. In the final paper, we present the accuracy of this characterization compared to spectroradiometric measurements and present 3D color-prints created on the Objet500 Connex3 printer from Stratasys.

[1] C. Parraman, P. Walters, B. Reid, and D. Huson, Specifying colour and maintaining colour accuracy for 3d printing," in SPIE/IS&T Electronic Imaging Conference, 6805, pp. 68050L-1-68050L-8, (San Jose), 2008.

[2] M. Stanic, B. Lozo, T. Muck, S. Jamnicki, and R. Kulcar, Color measurements of three-dimensional ink-jet prints," in NIP & Digital Fabrication Conference, 2008(2), pp. 623-626, Society for Imaging Science and Technology, 2008.

[3] K. Kehren, P. Urban, E. Dörsam, A. Höpe, and D. R. Wyble, Performance of multi-angle spectrophotometers," in Proceedings of the Midterm Meeting of the International Color Association, Zürich: AIC, pp. 473-476, 2011.

[4] P. L. Vora and H. J. Trussell, Measure of goodness of a set of color-scanning filters," Journal of the Optical Society of America A 10, pp. 1499-1508, 1993.

9398-29, Session 5

Towards gloss control in fine art reproduction

Teun Baar, Océ Print Logic Technologies (France) and Télécom ParisTech (France); Hans Brettel, Télécom ParisTech (France); Maria V. Ortiz Segovia, Océ Print Logic Technologies (France)

The studies regarding fine art reproduction mainly focus on the accuracy of colour and the recreation of surface texture properties. Since reflection properties other than colour are neglected, important details of the artwork are lost. For instance, gloss properties, often characteristic to painters and particular movements in the history of art, are not well reproduced. The inadequate reproduction of the different gloss levels of a piece of fine art leads to a specular reflection mismatch in printed copies with respect to the original works. Such a mismatch affects the perceptual quality of the printout and disrupts the actual intent of the artist. Currently, there is a lack of printing workflows that would allow us to control the gloss level of a printout in a per-area basis. The absence of standardised metrics to describe glossiness is an additional constraint. Our goal is to study the print workflow needed to control gloss locally in a printout. We propose to use different print parameters of a 3D high resolution printing setup to produce different gloss levels on a single printout. Our method can be used to control gloss automatically and in crucial applications such as fine art reproduction.

In our research we attempt to generate and control specular gloss locally because of its importance in fine art reproduction and special effects rendering for printing applications. When the appearance of gloss is considered in fine art, it is often related to the use of varnish applied by artists, either to protect the painting or to alter the look of the painting selectively [1]. Several studies that focus on measurements of varnish and gloss reflection properties have found remarkable differences between painters and movements in art history showing the importance of varnish in the perception of fine art [2]. For instance, painters like Picasso, Braque and Munch preferred matte appearances and would therefore deliberately not varnish some of their artwork [3]. On the other hand, there are painters that varied intentionally the shininess at a local level through the adjustment of the oiliness of the paint, by adding varnish or by other means [1]. As the spatially varying gloss level in fine art is an important indicator, not only of a specific movement and painter, but also the artist's intensions, it should not be omitted during reproduction. Recent work on the reproduction of fine art with 3D print systems exhibit the need for improvements regarding the recreation of the different gloss levels present in a piece of artwork [4].

Gloss also plays a major role in print quality assessment, where gloss variation is either applied intentionally or shown as a result of printing artefacts. Distinct printing applications that require local gloss differences include the fields of security printing [5] and material reproduction with known bidirectional reflection distribution functions [6].

We focus on controlling the appearance of specular gloss, through the variation of different print parameters. In our workflow, we use print parameters that correspond to a given intended gloss level resulting in a reproduction that shows local gloss variations consistent with the input gloss information. For example, we know that the power of the UV lamp used in UV-curing printers, affects the surface roughness and thereby the gloss level of the printout. For this particular parameter, our strategy is to determine the required intensity of the UV lamp that corresponds to a given gloss level.

We used two prototype printers, a wet-on-wet and a wet-on-dry UV-curable system, that have the ability of superimposing several layers of ink in the same printing area. We characterised the influence of ink coverage, (multi-layer) print mode, amount of UV light exposure and varnish on the resulting gloss level of different samples. Specular gloss values of the printed samples were then measured at an angle of 60° using a MG628-F2 multi-angle gloss meter.

Through the use of such a set of characterisation patches we are able to control the gloss level of new patches by finding the closest match to the intended ones. We tested our method by printing six patches with 100% coverage of black ink, increasing the gloss level from 10 GU to 35 GU with equal steps of 5 GU. A combination of print parameters was selected for each patch so that the resulting gloss value would closely match the intended gloss value. The results show that we are able to control the gloss level with an average deviation of 2.6 GU.

In this research we focussed on local control of the gloss appearance by adjusting several print parameters in a multi-layer printing system. Although we currently worked with 2D prints, we envision extending our approach to control the appearance of gloss in more intricate surfaces such as textured fine art. In future research we will work on integrating our research in a complete workflow to control the colour and gloss level of prints that would lead to more accurate and appealing reproductions of fine art.

[1] Kirsh, A., Levenson, R.S., Seeing through Paintings: Physical examination in art historical studies, Yale University Press, New Haven, (2000)

[2] Liang, H. et al, "En-face optical coherence tomography – a novel application of non-invasive imaging to art conservation", Optics Express 6133, (2005)

[3] "The Unvarnished Truth: Mattness, 'Primitivism' and Modernity in French Painting, c.1870-1907" The Burlington Magazine 136, 738-746 (1994).

[4] Elkhuizen, W.S., Zaman, T., Verhofstad, W., Jonker, P.P., Dik, J., Geraedts, J.M.P., "Topographical scanning and reproduction of near-planar surfaces of paintings" Electronic Imaging; Measuring, Modeling, and Reproducing Material Appearance (2014).

[5] Hodgson, A., "The use of Gloss Effects from Inkjet Printing for Brand Identification, Personalisation and Security," International Conference on Digital Production Printing and

[6] Malzbender, T., Samadani, R., Scher, S., Crume, A., Dunn, D. and Davis J., "Printing reflectance functions," ACM Trans. Graph., 31, 20:1–20:11, (2012).

9398-30, Session 5

Exploring the bronzing effect at the surface of ink layers

Mathieu Hébert, Univ. Jean Monnet Saint-Etienne (France) and Lab. Hubert Curien (France); Maxime Mallet, Alexis Deboos, Institut d'Optique Graduate School (France); Pierre H. Chavel, Lab. Charles Fabry (France); Deng-Feng Kuang, Institut d'Optique Graduate School (France); Jean-Paul Hugonin, Mondher Besbes, Lab. Charles Fabry (France); Anthony Cazier, Univ. Jean Monnet Saint-Etienne (France) and Institut d'Optique Graduate School (France)

Printing is based on coloring a white reflecting surface through continuous or discontinuous layers of inks. The inks are colored materials, generally non-scattering or weakly scattering, that play a role of spectral filtering. However, it sometimes occurs, especially in inkjet printing, that a colored shine appears in the specular direction whose hue is noticeably different from the color of the ink itself, a well-known phenomenon often called "bronzing" or "gloss differential" by professionals of photo printing, or amateur photographers who print their pictures in inkjet on photo-quality paper. Discussions including recommendations to avoid the effect can be found on the internet in forums specialized in photo printing, but a satisfactory physical explanation is elusive, as it also is in the scientific literature on printing sciences and technologies. Beyond our scientific interest for this physical phenomenon, our study was also motivated by its consequences on the accuracy of prediction models in the color reproduction domain.

By observing that the effect disappears when a coating, even a simple fingerprint, is deposited on the ink, it seems that the phenomenon is related to the ink-air interface, through either its relative refractive index or its topology. To our knowledge, the only web page where a sound physical explanation is proposed illustrates the phenomenon in the case of ball pen inks, for which the effect is often pronounced, and analyses the influence of the absorbance of the ink on the reflectance of the air-ink interface [1]. Indeed, the Fresnel formulas directly state that the reflectance of an interface with a difference in the complex relative refractive index is an increasing function of that difference. This applies to the imaginary part of the index, representative of the absorbance of the medium. It could also explain why the phenomenon disappears with an overlay, because the ink-air interface is replaced with the coating-air interface with a lower relative index difference. Our experiments clearly confirm that this phenomenon essentially explains the colored shine observed on the surface of the inkjet inks.

We also investigated a possible additional effect of the light scattering by the nano-roughness of the ink-air interface under collimated illumination, which is known to produce a small reddish shine when the standard deviation of the elevation profile is smaller than the wavelength of light. We could verify, with an atomic force microscope, that the roughness of the studied surfaces is on this order of magnitude. However, the spectral reflectances predicted by wave scattering models do not coincide with the measured ones. We thus concluded, as in Ref. [2], that light scattering cannot explain the observed shine colors and we focused on the Fresnel reflectance of the air-ink interface.

Most phenomenological models that predict the spectral reflectance of colored surfaces, such as the classical Williams-Clapper model for gelatin prints and Clapper-Yule model for halftone paper prints, describe the coloring medium as a homogenous effective medium with a real refractive index around 1.5. Absorption is characterized by the spectral transmittance of the coloring layer. However, this model is valid only for weakly absorbing media. Since the saturation of the colors is related to the capacity of the ink to strongly absorb light in a specific spectral band, the absorptance in this band may overpass the threshold of validity of this approximation. Ink should then be modelled as a homogenous effective medium with complex refractive index, where the imaginary part is directly related to the absorption coefficient of the medium, and the Fresnel reflectance of the interface should use that complex relative refractive index. The imaginary part increases with absorption, and so does the Fresnel reflectance. This yields the paradox that at the surface of red ink that absorbs green light, a green shine is observed. The shine and the diffuse light filtered by the ink have therefore complementary colors.

Since the spectral complex refractive index of the ink is needed to predict the shine color , we address the way to obtain it. Ellipsometry, the favorite technique for the measurement of refractive indices at strongly absorbing media, requires very flat surfaces: inks are too rough for accurate measurement. Instead, we propose to measure the reflectance and the transmittance at normal incidence of a transparency film coated with the ink, and to deduce the complex index from a model describing the flux transfers within the coated film. The Fresnel reflectances predicted with the obtained values agree with the measured spectrum of the sheen, except for the yellow ink where an important deviation is observed in a rather small spectral band, probably because the ink is fluorescing. Fluorescence is the first limitation of this method: an extended model that has not been developed yet, allowing the distinction between the light components reflected by the interfaces and the light components emitted by fluorescence, will be necessary. The second limitation arises in layers that absorb light completely, and thus have zero transmittance. This was the case with the black and blue BIC ball pen inks that we tested, for which dilution was necessary.

We also discuss the impact of bronzing effects on the spectral reflectance models for calibration procedures, including those intended for halftone prints. In the phenomenological models used to predict the spectral reflectance of halftone prints or overprinted surfaces, the transmittance of the primaries are generally calibrated from the measured reflectance of printed patches. If the shine is concentrated in a small solid angle, and if the measuring geometry prevents from the capture of the specularly reflected light, bronzing should have no impact on the calibration of the model. In contrast, if the surface of the print is rough, the shine may be observed through a large solid angle and part of it can be included into the reflectance measurement. Unless this part can be accurately predicted, there is a risk of error in the calibration of the model. Spectral Bi-directional Reflectance Distribution Function measurements, and an analysis of the geometry of the reflectance measurement device can help assessing this risk.

[1] http://www.phikwadraat.nl/specular_colors/

[2] NPIRI Color Measurement Task Force, "A method for : the identification and assessment of the presence of bronzing in printing ink films," American Ink Maker, oct. 2001, vol.79, n°10, p.99-104.

9398-31, Session 5

Controlling colour-printed gloss by varnish-halftones

Sepideh Samadzadegan, Technische Univ. Darmstadt (Germany); Teun Baar, Océ Print Logic Technologies (France) and Mines ParisTech (France) and Télécom ParisTech (France); Philipp Urban, Fraunhofer-Institut für Graphische Datenverarbeitung (Germany); Maria V. Ortiz Segovia, Océ Print Logic Technologies (France); Jana Blahová, Technische Univ. Darmstadt (Germany)

Controlling gloss in addition to colour is highly desirable in digital printing. Different parameters in a printing process such as the type of substrate

and inks, the print mode, drying time, etc. influence the glossiness of the printout. Since some of these factors cannot be spatially controlled, gloss artefacts, such as differential gloss [1]-[3], may reduce the print quality drastically. Therefore, varying the gloss level intentionally might be highly desirable to avoid gloss artefacts and for aesthetic purposes.

Although some experiments have been conducted to change the glossiness of black samples using spatially-varying varnish [4], in this paper we demonstrate how to control the glossiness of the final colour-printed samples independently of the underlying colour by varnish halftones. Moreover, a psychophysical experiment will be conducted in order to relate the gloss measurements to gloss perception and to obtain a perceptual gloss scale.

In our study we printed on a 140 gram outdoor paper as well as a rigid media using the Océ Arizona 480 GT printer. This printer allows multi-layer printing in a xy-resolution of 450 dpi and a layer thickness of 8 μm. From preliminary experiments combining different print parameters we found that varying halftone screens for varnish deposition enables us to create gloss levels in the range of "semi-matt" to almost "high gloss", according to the NCS Gloss Scale product description.

Varnish is often applied in a full-coverage fashion, where an area of the printout is either fully covered with a special ink, or not at all. In our attempt to obtain intermediate gloss levels between these two extremes, we printed different colour-samples where their glossiness varied by controlling the amount of varnish deposition by halftoning.

To investigate the colour-gloss gamut we utilized Direct Binary Search (DBS) halftoning with varnish coverage in the range of 0 to 60% by steps of 5%. Since the previous experiments showed that the gloss values of samples with varnish coverages above 60% were not considerably different, we decided to use 60% as the maximum level of coverage. In total, 13 different varnish levels were created.

The varnish halftone layer was applied onto C, M, Y, R=(M+Y), G=(C+Y), B=(C+M), W and C+M+Y ramps consisting of 8 different ink coverages in the range of 12.5 to 100% by steps of 12.5%. It should be mentioned that the Arizona printer has five inks which are C, M, Y, K, and W. In total, 13 ? 8 ? 9 = 936 samples were printed.

The glossiness of the printed test samples were measured using a MG628-F2 multi-angle gloss meter for ?20?^0, ?60?^0, and ?85?^0 of specular gloss. The measurement results show a monotonic relationship between the varnish coverage and the gloss value, i.e. the larger the varnish coverage the higher the glossiness independently of the underlying colour.

Furthermore, the colour of all samples was measured using a ?45?^0/0^0 spectrophotometer which shows the influence of increased varnish coverage on colour. This colour change might be the result of varnish layers reflection properties. In our experiment, we found only colour differences of maximum 1 ?ΔE?_ab^* around the average colour of each of the series of patches consisting of 13 different varnish coverages. This colour difference is close to the just noticeable difference (JND).

Looking at these results, printing colour samples with different glossiness in the range of "semi-matt" to almost "high gloss", according to NCS Gloss Scale range, is possible; i.e. using this approach we are able to cover a fairly wide range of specular gloss values.

Considering samples with 50% ink coverage and ?60?^0 specular gloss measurements, the first two varnish coverages (0 and 5%) lead to the gloss values in the range of 9 to 20 which indicate the "semi-matt" appearance. The second two varnish coverages (10% and 15%) result in the gloss values between 21 and 39 indicating the "satin-matt" samples. Applying the varnish coverages of 20% and 25% lead to the gloss values in the range of 40 and 59 representing the "semi-gloss" appearance. The rest of varnish coverages from 30% to 60% yield to the cases where the printed samples are called "glossy" and almost "high gloss" with the gloss values between 60 and 89.

Moreover, a psychophysical experiment using 15 observers with normal colour vision will be conducted in order to investigate the relationship between the gloss measurements and perception. In this experiment we consider the samples with 50% ink coverage. Among a set of 13 samples for each printed colour 6 of them will be selected. Three of them will be chosen from the "semi-matt" to "semi-gloss" range and the rest three will be selected out of the "glossy" and almost "high gloss" samples.

For the reference samples we will utilize 6 samples out of the available 28 samples of the NCS Gloss Scale. These reference samples are called: "matt", "semi-matt", "satin-matt", "semi-gloss", "glossy", and "high gloss" with ?60?^0 specular gloss values of 6, 12, 30, 50, 75, and 95 respectively. All of them have medium grey colour.

Both the reference and test samples will be attached to grey colour cylindrical shapes in order to provide different angles for gloss perception.

The experiment will be conducted in a dark room and inside a viewing booth with D50 illumination. The observers will be asked to hold the samples in their arms' length and inside the viewing booth comparing them with the fixed reference samples to give scales according to the reference gloss scales from 1 to 6. Assignments of intermediate gloss values will be allowed by the step of 0.5. The samples will be given to observers randomly with no specific colour order.

Although by using varnish-halftones we have some predictions about the glossiness range of the final print, more accurate varnish coverages resulting in samples with perceptually equal gloss steps is required in order to build a perceptually uniform four dimensional colour-gloss space. Colour-gloss gamut mapping shall be performed in the future based on this colour-gloss space.//

References:

[1] Zeise, E. and Burningham, N., "Standardization of Perceptually Based Image Quality for Printing System," IS&T NIP18, 698-702 (2002).

[2] Kuo, C., Ng, Y. and Wang, Y., "Differential Gloss Visual Threshold under Normal Viewing Conditions," International Conference on Digital Printing Technologies 18, 467-470 (2002).

[3] Baar, T., Samadzadegan, S., Brettel, H., Urban, P. and Ortiz Segovia, M. V., "Printing gloss effects in a 2.5D system," Electronic Imaging; Measuring, Modeling, and Reproducing Material Appearance (2014).

[4] Fores, A., Ferwerda, J., Tastl, I. and Recher, J., "Perceiving Gloss in Surfaces and Images," 21st Color and Imaging Conference Proc. IS&T, 44-51 (2013).

9398-32, Session 5

Reproducing oil paint gloss in print for the purpose of creating reproductions of Old Masters

Willemijn S. Elkhuizen, Boris A. J. Lenseigne, Technische Univ. Delft (Netherlands); Teun Baar, Océ Print Logic Technologies (France) and Institut Mines-Télécom (France); Wim Verhofstad, Océ Technologies B.V. (Netherlands); Erik Tempelman, Technische Univ. Delft (Netherlands); Jo M. P. Geraedts, Technische Univ. Delft (Netherlands) and Océ Technologies B.V. (Netherlands); Joris Dik, Technische Univ. Delft (Netherlands)

Recent advances in computer graphics, 2D- and 3D-printing have led to more and more realistic imitations of material appearance, either in the digital or physical domain. The development of a 3D scanner [1] in combination with High Resolution 3D printing technology of Océ Technologies is among these advances. It has already led to the possibility to capture and reproduce the colour and texture of historic oil paintings by e.g. Rembrandt and Van Gogh. However, expert evaluation have established that this does not create a totally realistic material appearance yet: the typical gloss and the translucency of the paints are missing and the colours and textures are not yet completely accurately reproduced [2].

Accurate reproduction of paintings (using a digital process instead of manually) would offer multiple opportunities for the field of cultural heritage, ranging from commercial and educational applications to supporting art historic research: for instance, it could be used to create historical reconstructions of aging effects, recreate missing pieces of artworks, or support the restoration practice by providing samples. Apart from these practical benefits, such reproductions also challenge the study of perception; furthermore, it seems reasonable to assume that the knowledge

can also be used for other material reproduction, e.g. ceramic heritage.The aim of this paper is elaborate on the results and challenges of capturing and reproducing oil paint gloss (and colour) using a scanning setup and a full colour printing system.

Handmade painted samples were created for the purpose of scanning and reproduction of the gloss. Although newly painted samples are likely different in gloss from old paintings, creating and using samples offers two advantages: a) the colour of the samples is uniform and can be controlled (therefore differences in appearance can be attributed to gloss) and b) (near) optically smooth samples can be created, which minimizes the effect texture on the gloss perception. Samples were made using high quality oil paint in thirteen colours, and varnished using two types of acrylic varnishes (matte and glossy acrylic varnish). To create optically smooth samples the oil paint is mixed with various auxiliary materials (Alkyd medium, Linseed oil and turpentine) and siccative is added to speed up 'drying' time of the paint; it can normally take up to one year for full oil polymerisation. One set of samples is made using a pallet knife, and a second set of samples is created using an automated blade coater.

In order to capture the spatially varying gloss of a painted surface, a setup was built consisting of a professional grade camera (40Mp sensor) and a linear array of LEDs placed vertically. The sample was illuminated from different angles and from a constant distance and series of multi-exposure HDR images were shot for each position. The RGB images are converted to recover the radiance of the scene and the variations of the radiance over the different illuminations are used to build a first order estimate of the local glossiness. This approach gives a first estimate of the gloss across the surface, although subtle texture elements (present in the first set of samples) show up as variations in the gloss map.

Digitally printed samples were made with the Océ Arizona 480 GT printer. In order to match the colour of the painted samples, measurements of the painted samples were made using a spectrophotometer. The painted colours are approximated using an absolute colorimetric colour profile. It was found that many of the colours of the painted samples, lie outside the gamut of printer for the case of a D65 light source.

Two different strategies were combined to create varying gloss in the print. First, to create a matte effect, the colour layers were printed using a multilevel half-toning algorithm, as described by Baar et al. [3] This generated the most matte effect. Secondly, on top of the matte colours, half-toned varnish (of the Océ Arizona 480GT printer) was added in increasing coverage to create increasing gloss levels. Varnish coverings between 0 to 60% were used (meaning 0 to 60% of the pixels are printed) using a specially developed half-toning algorithm. Varnish coverage above 60% did not lead to higher gloss levels. When viewed under the microscope, it can be observed that the 60% varnish coverage already forms a full layer on the surface, explaining this effect. At lower coverage, puddling of the varnish can be observed.

Three types of printed samples were created. One set of samples was printed with gloss patches (with increasing gloss coverage in 5% increments). These were be used to measure gloss and evaluate the uniformness of gloss within a patch. A second set of samples was creating with a gloss gradient, to evaluate the smoothness of transitions that can be created. A third set of samples was printed using the gloss map from the scan, to create a relatively matched gloss print, meaning the lowest gloss level of the scan is linked to the lowest gloss level of the printer, and highest scan level to highest printer gloss level.

On visual inspection the printed gloss patches show gradually increasing gloss. Also, the printed gloss gradients show smooth transitions and the printed gloss map shows the varying gloss characteristics of painted sample. However, differences can also be observed between the painted and printed gloss samples. When viewing the printed samples at close range (<50cm) the print resolution print is visible. This creates an orange peel effect, especially visible in the glossy samples, different from the smooth surface of the painted samples. Furthermore, measurements of the painted samples give a lowest gloss value of 1.2GU at 60° measurement (being unvarnished oil paint). The lowest printable value lies around 20GU (at 60°). This difference was also observed visually.

Measurements of the painted and printed samples are presented (i.e. tri-gloss measurements), as well as participants' visual evaluations.

[1] Zaman, T., "Development of a Topographic Imaging Device for the Near-Planer Surfaces of Paintings," Repository, Delft University of Technology (2013).

[2] Elkhuizen, W.S., Zaman, T., Verhofstad, W., Jonker, P.P., Dik, J., and Geraedts, J.M.P., "Topographical scanning and reproduction of near-planar surfaces of paintings," in Electron. Imaging 2014 9018, M. V. Ortiz Segovia, P. Urban, and J. P. Allebach, Eds., 901809 (2014).

[3] Baar, T., Samadzadegan, S., Ortiz Segovia, M. V., Urban, P., and Brettel, H., "Printing gloss effects in a 2.5D system," in Electron. Imaging Meas. Model. Reprod. Mater. Appear. (2014).

9398-33, Session 5

3D printing awareness: the future of making things

Fabrizio Valpreda, Politecnico di Torino (Italy)

The advent of 3D printing is giving us new production opportunities but is creating new economical and social assets. In the paper we will analyze the new conditions we will live in, first of all from the point of view of the industrial design, building a parallelism with the two previous "big design revolutions", and from the human being cultural values too, focusing the survey with an approach based on the Open Design and the new ways of sharing knowledge.

The first "big design revolution" humankind face off (interesting for our argumentation) was the "printing revolution". From the moment when the movable type printing appeared, any parameter related to culture, knowledge, and in general anything meant to be known and transmitted human by human, changed forever.

Hypothetically any person around the world was able to know anything thanks to sheet paper printed documents: cheap, easy to reproduce and move, collectable in books, to be archived in a whole new generation of libraries.

This new asset was so strong, solid and effective that the results of that revolution, more human behavioral than technological, are still visible in our everyday life.

There are several relevant aspects of this historical change that deserve to be underlined but here we will point our attention focus upon some of them only:

- knowledge was not anymore reserved to few, mostly of them rich, people because printed words readable by any human around were available to everybody

- one of the most important natural resources, trees, started to be used for a brand new, almost immediately and heavily diffused activity

- the diffusion of printed books took only a decade to stabilize and spread in the known world

- the legacy of the traditional books production was quite immediately reduced to a tiny percentage of the whole global book printing

- people were able to edit, print and share their own writing essay, this last aspect was obviously enhanced by internet and digital technologies.

Not all this happened at the same time, of course, and not all happened the simple way we are summarizing here. We can say, however, that all this happened, and most important it is going to happen again with the 3D printing technology, in a manner that we intend to understand, since we guess that this time it will be a lot more disrupting.

In order to better understand what is going to happen with the 3D printing, we have to remember another historical event that changed human being life forever: the industrial revolution.

Before the advent of the steam machines applied to the production of textiles and iron manufacturing, any object produced was made thanks to handicraft activities: the invention of the steam machine and the development of the use of more efficient water powered machines changed every human production activity.

And, like for “movable type printing revolution”, again the effects of the industrial revolution are still visible today and changed definitively our way to relate to objects: the consumerism was born thanks to these two “big design revolutions”.

The first one of this two revolutions, however, has already met some kind of transformation that not only few experts are defining as a no-way-back-to-past: digital text documents creation and editing, which use have been spread all over the world around the mid 80s, are commonly used by anybody. And more, any of us is able to print anything created, received via email or downloaded from the Internet. These possibilities open a new approach in sharing the knowledge: the so called “Open approach”.

All this is heavily transforming the asset of knowledge spreading via text documents: what once was wrote by writers, edited by professionals, printed by huge print services, distributed via book shops and archived in libraries can now be created, edited, printed, uploaded and distributed by any of us.

This is, again, a revolution, but what is extraordinary is that it is limited to 2 dimensions.

What if all this will happen to objects? What if we add the third dimension to all this processes? In short terms, what if any of us was able of printing, and sharing, not only his own book, but a bicycle too? Or a phone, or a wheel chair? Or a gun? Or food?

We can say that even if not steadily, even if not always seriously, all this is already possible.

And the experts are telling us that during the next decade all this will come to a very serious version of itself.

A simple logical reasoning can help us think about what is the scenario we are talking about.

If we think about “printing revolution” we can say that manuscript was overruled by movable type printing which was replaced by personal printing and the conversion to the use of digital data. This is widely considered the killer of of traditional printing and book production. E-books are the next version of books, but are they also their worst enemy? And if yes, is this what we wanted to happen? Shouldn’t we say the same about music, since the traditional version if its market is dead?

If we do the same logical trick with “industrial revolution” we can say that handicraft was killed by industry and the relationship with objects and their value changed for ever. But what will arrive with personal 3D printing and the use of digital 3D data exchange? Will it kill industrial production? Will it damage the social tissue?

The idea of reading a book on an e-book reader will migrate to objects? And how will be our relation and perception of such kind of objects? Will we enjoy objects virtually online and print them out if needed or will we print anything just to test it for a while and then trash it?

The makers movement and the spread of the FabLabs around the world is suggesting us a different story that carries some interesting issues. And the Open Design approach is very strong ally to this new strategic asset.

The situation, at last, involves of course all of us, the way we share knowledge, the way we use matter and energy, what we will decide to do with billion of people working in traditional factories and what we will decide to do with what we will be able to print out of our personal 3D printers.

The paper will analyze this scenario, with the target of reaching a more robust awareness regarding not only design but also social choices, related to the future of the things we will produce in the (near?) future.

Conference 9399: Image Processing: Algorithms and Systems XIII

Tuesday - Wednesday 10–11 February 2015

Part of Proceedings of SPIE Vol. 9399 Image Processing: Algorithms and Systems XIII

9399-1, Session 1

Links between binary classification and the assignment problem in ordered hypothesis machines

Reid B. Porter, Los Alamos National Lab. (United States); Beate G. Zimmer, Texas A&M Univ. Corpus Christi (United States)

Ordered Hypothesis Machines (OHM) are large margin classifiers that belong to the class of Generalized Stack Filters which were originally developed for non-linear signal processing. In previous work we showed how OHM classifiers are equivalent to a variation of Nearest Neighbor classifiers, with the advantage that training involves minimizing a loss function which includes a regularization parameter that controls class complexity.

In this paper we report a new connection between training OHM classifiers for binary classification and the Linear Assignment problem. The linear assignment problem is one of the most famous and well-studied optimization problems in the field of operations research. It focuses on how best to assign a set of N items (e.g. jobs or tasks) to a set of N other items (e.g. workers or machines). Binary classification is another combinatorial optimization problem that is perhaps as famous and as well-studied as the assignment problem. It focuses on how to best design a classifier from training examples, an important question in the field of machine learning. In its simplest form, binary classification involves choosing a function from a function class F (or hypothesis space) that minimizes the number of mistakes.

The motivation, definition and solution methods for binary classification and the Assignment problem differ in many ways. For example, a solution to the Assignment problem can be found in polynomial time by (among others) the Hungarian algorithm. But for binary classification, finding a function that minimizes the number of classification mistakes is known to be NP hard even for relatively simple function classes such as linear classifiers. In this paper we show that using a commonly used convex surrogate for the number of mistakes (based on the hinge loss) in combination with a particular choice of function class (the class of Generalized Stack Filters) leads to a situation where the binary classification and Assignment problems form a primal-dual pair.

In the paper we summarize the prior work required to formalize this observation and discuss the relationships between the two problems with the help of experiments with synthetic data. The duality sheds new light on the OHM training problem, and opens the door to new training methods. We present some initial work in this area and develop solution methods inspired by the Transportation problem (a generalization of the Assignment problem) to solve a change detection task in remotely sensed imagery. We conclude with a brief description of how future work exploring the connection between these two optimization problems may extend beyond classification and potentially benefit other applications in non-linear signal processing.

9399-2, Session 1

Optimized curve design for image analysis using localized geodesic distance transformations

Billy Braithwaite, Harri Niska, Irene Pöllänen, Tiia Ikonen, Keijo Haataja, Pekka J. Toivanen, Univ. of Eastern Finland (Finland); Teemu Tolonen, Univ. of Tampere (Finland)

We consider geodesic distance transformations in digital images. Given a N x M digital image, a distance mapping is produced by evaluating local pixel distances. Distance Transformation On Curved Space (DTOCS) evaluates shortest geodesics of a given pixel neighborhood by evaluating the height displacements between pixels via integer approximation. We propose an optimization framework, based on multi-objective genetic algorithms and artificial neural network (ANN)-based sensitivity analysis, for optimal coefficient and feature selection for DTOCS in a pattern recognition scheme. The proposed optimization framework yields promising results for e.g. image segmentation, region detection and general pattern recognition for image analysis. We exemplify initial experiments using complex breast cancer imagery. Furthermore, we will outline future research work which will complete the research done in this paper.

DTOCS evaluates the height displacements between pixels in a discrete 8-square grid via city block metric. The pixels are processed in a four time raster scanning iterating two times on the image to guarantee convergence. When DTOCS evaluates the pixel distances, a curvature coefficient is introduced to the evaluation, which governs the amount of curvature of an gray-scale image. The basic definition of DTOCS, i.e. setting the curvature coefficient to unit value, yields a global distance mapping. This will not give any meaningful information about the image, if the image has complex structures or textures. By setting the coefficient between zero and unit value, initial experiments suggests that the distance mapping produced by DTOCS becomes more localized.

Our optimization framework is formulated as a multi-objective minimization problem to determine the minimal amount of DTOCS distance images, generated with optimal coefficients, for maximum classification rate. We use multi-objective genetic algorithms and sensitivity analysis of an ANN classifier to determine a Pareto-optimal subset of feature images, generated by DTOCS with Pareto-optimal coefficients.

The overall optimization scheme can be divided into five phases. First we generated a selected number of DTOCS images with selected curvature coefficients. Secondly, we train the initial ANN classifier using all DTOCS images as training images. Thirdly, multi-objective genetic algorithms are used to search for a Pareto-optimal DTOCS features. Fourthly, the acquired Pareto-optimal features are validated using external images. The final step is the selection of the final Pareto-optimal DTOCS features and external validation.

In our experiments, we approached the optimum curvature design for DTOCS by means of ANN-based image segmentation of complex histopathological imagery. The imagery are tissue samples of breast cancer. Our goal is to segment the cancerous tissue, healthy (non-cancerous) tissue and fat deposits from the images.

For our experiments, we generated 20 different DTOCS images with curvature coefficients chosen in a logarithmic scale between zero and one. As for the ANN model, we chose the standard multi-layered perceptron (MLP) model. The configurations for the MLP model and the genetic algorithms were experimentally chosen via trial-and-error. 32 histopathological images were used in the experiments, which were randomly divided into three subsets: a subset of 5 training images, a subset of 5 images for validating the Pareto-optimal feature subset, and 22 images for validating the the final selected Pareto-optimal feature subset.

After doing cross validation to the final Pareto-optimal feature subset, we obtained a classification rate of 89% for the fat deposits, 27% for the healthy connecting tissue and 85% for the cancerous tissue.

The proposed optimized framework for geodesic distance transformation yields promising results for using standard ANN models for pattern recognition purposes in image analysis. The optimum curvature design for DTOCS offers a computationally efficient solution for producing localized distance mapping of the image without resulting to more expensive computational solutions e.g. sliding-window approach. Due to the small and subtle interclass variances between tissue types in breast cancer imagery, DTOCS cannot give an accurate region separation if the regions are too similar. However for homogeneous regions DTOCS is able to give an accurate region separation via localized distance mapping. A more in-

depth analysis is needed for the implications of the curvature effect w.r.t shortest geodesics in digital image, which will be a topic on its own. For future research, we plan to improve the optimum curve design for pattern recognition in image analysis, especially for breast cancer analysis for a broader class of tissue type segmentation or classification, and using a more robust ANN model for DTOCS -based pattern recognition.

9399-3, Session 1

Adaptive graph construction for Isomap manifold learning

Loc Tran, Old Dominion Univ. (United States); Zezhong Zheng, University of Electronic Science and Technology of China, (China); Guoqing Zhou, Guilin University of Technology (China); Jiang Li, Old Dominion Univ. (United States)

In manifold learning, the goal is to learn a linear representation of nonlinear structures present in quantitative data. The methods used for this often incorporate a simple neighborhood selection algorithm such as k-nearest neighbors. Rigid approaches such as this may inadequately reflect the intrinsic structure of nonlinear manifolds since the optimal number of neighbors selected may vary at different points on the manifold. We introduce a novel adaptive neighborhood selection approach for manifold graph construction and apply it to classical Isomap.

The Isomap algorithm consists of three steps. First, a neighborhood is created to form a neighborhood graph. Next, a geodesic distance matrix is calculated from the neighborhood graph. This matrix will consist of pairwise distances between every sample of the data set where the distances are calculated by paths from the neighborhood graph. Finally, dimensionality reduction is performed on the distance matrix to form the lower dimensionality space while preserving the geodesic distances found in the second step. A successful manifold using Isomap depends largely on the formation of the neighborhood graph. One of the major drawbacks of the Isomap algorithm is that it is susceptible to leaking or short circuiting. This can happen if neighbors are selected such that a short cut is created between two areas that do not follow the intrinsic structure of the manifold. As a result, large geodesic distances are misrepresented as short distances because of a poorly formed graph. The leaking problem is especially prevalent in noisy data sets. The graph construction is thus a very critical step for the Isomap method to create a successful manifold.

Because of the short circuiting problem, it is undesirable to have a large neighborhood. In the original graph construction method, the nearest neighbors are chosen from k-nearest neighbors using Euclidean distance. The selection of k is critical in creating a good graph. If k is too large the risk of short circuiting is higher. On the other hand, if k is too small, the graph may not be fully connected. Another problem that occurs is that a good k value for one location may be inappropriate for another location. In this paper, these problems are addressed with an adaptive neighbor selection method implemented with the L1-norm. In the literature, it has been shown that sparse representation has applications in a wide range of computer vision and pattern recognition including face and object recognition, compressive sensing, subspace learning, medical image analysis, and dictionary learning.

We propose an adaptive graph construction approach that is based upon the sparsity property of the L1 norm. The L1 enhanced graph construction method replaces k-nearest neighbors in the classical approach. The proposed algorithm is first tested on the data sets from the UCI data base repository which showed that the proposed approach performs better than the classical approach. Next, the proposed approach is applied to two image data sets and achieved improved performances over standard Isomap.

9399-4, Session 1

Colony image classification and segmentation by using genetic algorithm

Weixing Wang, Chang'an Univ. (China); Lishen Yang, Jianqing Guo, Zhanfu Meng, Henan Polytechnic University (China)

In the segmentation and classification of colony images, there are numbers of applications including food, dairy, hygiene, environment monitoring, water, toxicology, sterility testing, AMES testing, pharmaccuticals, paints, sterile fluids and fungal contamination.

Colony image segmentation plays a key role in automatic visual systems. The segmentation problem has been formulated from different perspectives. Image segmentation and classification of colony images plays a key role in automatic visual systems.

This paper describes a new algorithm using for segmentation and classification of colony images. It is based on a genetic approach that allow us to consider the segmentation problem as a global optimization, and the new classifier introduced here is based on fuzzy-integration schemes controlled by a genetic optimization procedure. Two different types of integration are proposed here, and are validated by experiments on real data sets for Machine. Results show the good performance and robustness of the integrated classifier strategies. More details are described as the follows.

In this study, the data set U is composed of 300 pre-classified images representing three different kinds of colonies: food colony (class A), water colony (class B), and bacteria (class C). The classification problem is hard because these classes have similar shape and texture. The training set, T, is composed of 180 images (90 for each class). The number of elements of T has been set equal for all classes because they are represented in U with the same abundance and their structural complexity is assumed comparable.

In the OAR-classification scheme, the training set for the two classes is built starting from T by assigning 90 elements to the kth class and 180-(k?90) for the superclass of the remaining. The genetic optimization parameters have been found after G=1000 iteration for each level of the tree, with probability of crossover pc=0.8 and probability of mutation pm= 0.02. The cardinality of the population is fixed to 90 chromosomes at each iteration. The number of neurons was 8 for the hidden layer and 2 for the output layer. In the AAA-classification scheme, the genetic optimization parameter have been found after G=3000 iteration with probability of crossover pc=0.8 and probability of mutation pm = 0.02. The cardinality of the population is fixed to 90 chromosomes at each iteration. The number of neurons was 10 for the hidden layer and 4 for the output layer.

9399-5, Session 1

Real-time affine invariant gesture recognition for LED smart lighting control

Xu Chen, Miao Liao, Xiao-Fan Feng, Sharp Labs. of America, Inc. (United States)

Gesture recognition has attracted extensive research interest in the field of human computer interaction. Realtime affine invariant gesture recognition is an important and challenging problem. This paper presents a robust affine view invariant gesture recognition system for realtime LED smart light control. As far as we know, this is the first time that gesture recognition has been applied for control LED smart light in realtime. Here we focus on the recognition of two types of gestures (a so-called ``gesture 8'' and a ``five finger gesture'').

In the proposed framework, the camera captures the image of the field of view of the downlight and performs the gesture recognition.

Once the hand gestures are robustly detected and tracked, they are employed to control the status of lighting. when light is off, the presence of "gesture 8" will turn it on. In order to improve system robustness, instead

of detection in a single frame, we require detection of "gesture 8" in 6 continuous frames to turn on the light, so is for turning light off. When light is on, user can slide "gesture 8" to turn up and turn down the light brightness. In particular, sliding towards the thumb direction turns down light brightness, and vice versa. When light is on, the presence of 5 finger gesture will turn it off. In addition to 6 continuous frame constraint, we also require the presence of 5 finger gesture of both hands to turn off the light, because users may unintentionally show 5 finger gesture with either hand but not both.

In the process of gesture recogniton, employing skin detection, hand blobs captured from a top view camera are first localized and aligned. Subsequently, SVM classifiers trained on HOG features and robust shape features are then utilized for gesture recognition. By accurately recognizing two types of gestures ("gesture 8" and a "5 finger gesture"), a user is enabled to toggle lighting on/off efficiently and control light intensity on a continuous scale. In each case, gesture recognition is rotation- and translation-invariant.

With evaluation on more than 40,000 frames of those different gestures of different orientations from different people, the proposed gesture recognition algorithm achieves very high accuracy, with both of true positive rate and true negative rate more than 93% on single images (higher accuracy is achieved on video). At the same time, our implementation runs in real-time, with a frame rate of 20 fps, on a PC.

9399-6, Session 1

Steganography in clustered-dot halftones using orientation modulation and modification of direct binary search

Yung-Yao Chen, Sheng-Yi Hong, Kai-Wen Chen, National Taipei Univ. of Technology (Taiwan)

This paper proposed a new approach to encode data messages in halftone images, based on orientation modulation and the modification of clustered-dot direct binary search [1]. The goal of this study was to produce encoded halftones with a clustered-dot halftone texture that is preferable for electrophotographic printers. Embedding information in halftone prints is widely used in numerous applications including tracing and labeling. Barcodes typically provide a simple solution for these uses [2]. However, barcodes require additional overt marks on a page which are unrelated to the original content of this page. Hiding information within a contextually relevant image, rather than using a barcode, is a more attractive solution.

For electrophotographic printers, clustered-dot halftones are preferable because of the homogeneous halftone texture. There are four categories for encoding information in a clustered-dot halftone [3]: 1) encoding by changing the orientation of dot clusters [4]; 2) encoding by changing the phase shift of dot clusters [5]; 3) encode by adding single-pixel shifts of dot-clusters [6]; and 4) encode by modulating dot size through laser intensity modulation [7]. The method proposed in this paper belongs to the first category. The most common technique for generating clustered-dot halftones is based on screening [8]; a two-dimensional threshold array is periodically tilled in the image plane. The spatial arrangement of the thresholds in the threshold array indicates the order in which pixels are turned on or white within a selected region (halftone cell). In [4], the threshold function is modified cell-by-cell (i.e., the arrangement of thresholds in the threshold array varies by cell), enabling users to generate elliptical dot clusters and to guide the orientation of the dot cluster within each halftone cell according to the embedded data. In contrast to [4], the proposed method changes only the orientation of dot cluster in selected cells, and the same threshold array is used throughout the image plane, to provide consistent dot-cluster shapes.

First, the traditional direct binary search (DBS) algorithm is presented; the method of incorporating the proposed encoding method into a modified DBS is then described. The DBS algorithm is an iterative method that minimizes the perceptual-error-based metric between the original grayscale image and the halftone image [9, 10]. This metric incorporates a model for the human visual system (HVS). This is accomplished by manipulating each pixel of an initial halftone until a local minimum of the metric is achieved; in the DBS, toggle and swap operations are used to create trial changes of halftone patterns (Fig. 1). Several data-embedding approaches have been developed based on the DBS [11, 12]. However, because of DBS characteristics, both methods produce stochastic dispersed-dot halftone textures which may not be preferable for electrophotographic printers.

The proposed approach is a cell-wise embedding method which comprises three steps: 1) selection of embeddable halftone cells; 2) message encoding using orientation modulation; and 3) modification of the DBS optimization framework. First, to generate a clustered-dot halftone, a traditional 45° clustered-dot screen was used (Fig. 2). Because the halftone screen employed in this study was diagonally symmetrical, the input image was divided into cells of size 4 ? 4 pixels. To select the embeddable halftone cells, the property of screened halftones called "dot profile function" was applied. When screening, the family of binary textures used to render each level of constant tone is called the dot profile function, and a one-to-one relationship exists between the dot profile and the threshold array. For a region of the input image with a nearly constant value, the halftone image consists of only the pattern from the dot profile function; for a region of the input image with high frequency detail, the pattern from the dot profile function does not appear. Variations in these shapes based on edge detail are referred to as "partial dots," and the shape is not predictable. An input image is examined to identify the cells with patterns from the dot profile function. Cells with symmetrical patterns (i.e., an M-by-M dot cluster) are filtered to perform orientation modulation. The preserved cells are called "carriers," and the filtered cells are called "non-carriers". The locations of both types of cells are recorded, respectively.

Second, because the input image and the screen type are known, the dot-cluster shape of each preserved cell is known and is guaranteed to be asymmetrical. To improve identifiability and enhance data-carrying capacity, variable-bit-length codes (1 bit and 2 bits) are embedded within carriers with varying rotation angles, as shown in Figs. 3(a) and 3(c). After encoding the codes in the selected cells by using orientation modulation, the resulting halftone is sent to the modified DBS framework.

Third, a modified block-based DBS search strategy was used to preserve the dot-cluster shape. The halftone being tested was divided into non-overlapping blocks (in this study, 4 ? 4 blocks because of the cell size), and the DBS optimization framework is performed block by block. For the blocks corresponding to non-carrier cells, the traditional DBS is performed, i.e., the effect of trial toggles and swaps are computed for each pixel in the block, and only the trial change corresponding to the largest reduction in the error metric was accepted. On the other hand, for blocks corresponding to carrier cells, only swaps were allowed, and dot clusters were set as the smallest unit for swaps, as shown in Figs. 3(b) and 3(d). In addition, the error metric from [1] was adopted which is proved, equivalently generates stochastic clustered-dot texture when it is minimized. For decoding, the embedding data were retrieved by scanning and extracting individual rotation angles of the dot clusters in recorded cells.

This study provided a novel orientation modulation encoding method which can be incorporated into a modified DBS-based optimization framework, improving image quality and resisting dot gain effect of the output halftone. There are two advantages: first, consistent (but varying rotation angles) dot-cluster shapes which come from same threshold array were used throughout the output halftone image, producing a homogeneous texture. Second, compared with the traditional DBS, the block-based search strategy reduced computational complexity significantly. It is because for the greedy search strategy in traditional DBS, if some portions of a halftone converge to its local minimum of the error metric more rapidly than other portions, the repeat computation required to process these areas is wasted.

Reference

[1] P. Goyal, M. Gupta, C. Staelin, M. Fischer, O. Shacham, and J.P. Allebach, "Clustered-dot halftoning with direct binary search," IEEE Trans. Image Processing, vol. 22, no. 2, pp. 473–487, Feb. 2013.

[2] R. Villán, S. Voloshynovskiy, O. Koval, and T. Pun, "Multilevel 2D bar codes: Towards high capacity storage modules for multimedia security and management," IEEE Trans. Inf. Forensics Security, vol. 1, no. 4, pp. 405–420, Dec. 2006.

[3] Y. Y. Chen, R. Ulichney, J. P. Allebach, M. Gaubatz, and S. Pollard, "Stegatone performance characterization," Proc. SPIE, Media Watermarking,

Security, and Forensics 2013, vol. 8665, pp. 54–66, Aug. 2013.

[4] O. Bulan, G. Sharma, and V. Monga, "Orientation modulation for data hiding in clustered-dot halftone prints," IEEE Trans. Image Processing, vol. 19, no. 8, pp. 2070–2084, Aug. 2010.

[5] G. Sharma and S. G. Wang, "Show-through watermarking of duplex printed documents," Proc. SPIE Security, Steganography, and Watermarking of Multimedia Contents VI, vol. 5306, pp. 19–22, Jan. 2004.

[6] R. Ulichney, M. Gaubatz, and S. Simske, "Encoding information in clustered-dot halftones," IS&T NIP26 (26th Int. Conf. on Digital Printing Technologies), pp. 602–605, Sep. 2010.

[7] P. J. Chiang, J. P. Allebach, G. T. C. Chiu, "Extrinsic signature embedding and detection in electrophotographic halftoned images through exposure modulation," IEEE Trans. Inf. Forensics Security, vol. 6, no. 3, pp. 946–959, Sep. 2011.

[8] R. Ulichney, Digital Halftoning, MIT Press, Cambridge, MA, USA, 1987.

[9] M. Analoui and J. P. Allebach, "Model-based halftoning using direct binary search," Proc. SPIE Human Vision, Visual Processing, and Digital Display III, vol. 1666, pp. 96–108, Mar. 1992.

[10] D. J. Lieberman and J. P. Allebach, "A dual interpretation for direct binary search and its implications for tone reproduction and texture quality," IEEE Trans. Image Processing, vol. 9, no. 11, pp. 1950–1963, Nov. 2000.

[11] D. Kacker and J.P. Allebach, "Joint halftoning and watermarking," IEEE Trans. Image Processing, vol. 22, no. 2, pp. 1054–1068, Apr. 2003.

[12] J. M. Guo, C. C. Su, Y. F. Liu, H. Lee, and J. Lee, "Oriented modulation for watermarking in direct binary search halftone images," IEEE Trans. Image Processing, vol. 51, no. 4, pp. 4117–4127, Sep. 2012.

9399-7, Session 2

Machine learning for adaptive bilateral filtering

Iuri Frosio, NVIDIA Corp. (United States); Karen O. Egiazarian, Tampere Univ. of Technology (Finland) and NVIDIA Corp. (United States); Kari A. Pulli, NVIDIA Corp. (United States)

We describe here a supervised learning procedure aimed at estimating the relation between a set of local image features and the local optimal parameters of an adaptive bilateral filter. A set of two entropy based features is used to represent the properties of the image at a local scale. Experimental results show that the adaptive bilateral filter developed with the proposed method outperforms other versions of the bilateral filter where parameter tuning is based on empirical rules. Beyond bilateral filter, the proposed learning procedure represents a general framework that can be used for the development of a wide class of adaptive filters.

9399-8, Session 2

Real-time 3D adaptive filtering for portable imaging systems

Olivier Bockenbach, TechGmbH.com (Germany); Murtaza Ali, Texas Instruments Inc. (United States); Ian Wainwright, ContextVision AB (Sweden); Mark Nadeski, Texas Instruments Inc. (United States)

1 INTRODUCTION

In the area of medical imaging, there is a growing interest for portable devices. Tablets are used in emergency rooms for a quick diagnostic that allows sending the patient into the appropriate section of the hospital. In those cases, it is not uncommon to visualize 3D volumes obtained from ultrasound, CT and MRI scans.

3D adaptive filtering has been demonstrated to be of significant clinical value when it comes to enhancing the acquired data allowing radiologists to establish fast and accurate diagnostics. However, 3D adaptive filtering is compute intensive and its achievable image quality has often been traded off because of issues with processing power and heat dissipation. However, progress in processor technology allows considering devices with ARM cores, DSP cores and FPGAs.

2 OVERVIEW

Adaptive filtering is based on the detection and the evaluation of the properties of structures found in the scanned data, usually in the form of orientation and magnitude.

Adaptive filtering can be used in 3D, operating on points and lines as in the 2D case but also on surfaces that span in all dimensions. Indeed the third dimension is used for adding consistency and confidence to the processing. For example, a point in a 2D plane can be the result of the intersection of that plane with a line or a point caused by noise. This distinction is difficult to make when operating only in the 2D plane. On the other hand, using multiple planes can detect the presence of a line and deterministically decide for the noise point or the line.

By adding the third dimension to the problem, 3D adaptive filtering indeed adds a significant amount or processing requirement. Section three is dedicated to demonstrating how to optimize 3D adaptive filtering. It also shows how to use high performance computing techniques to leverage specific acceleration features present in DSPs.

3 METHODS AND MATERIAL

3.1 Hardware

The TMS320C6674 is a high performance DSP with four C66x DSP cores with fixed and floating-point capabilities, equipped with a rich set of industry standard peripherals such as Giga bit Ethernet.

The C66x core can issue eight instructions in parallel operating on two sets of 32-bit registers. The core also allows data level parallelism through the use of SIMD instructions which can operate on up to 128-bit vectors.

The instruction scheduling is based on an "in order" approach which leaves a lot of the optimization burden on the talents of engineering and of the compiler but results in a very power efficient implementation.

Besides following a standard memory hierarchy, the TMS320C6674 device also includes 512 KB of L2 caches per core that is most efficiently used in combination with the chip's Enhanced Direct Memory Access (EDMA). The EDMA can perform double strided memory access, allowing desired scatter-gather operations.

3.2 Algorithm

Two main properties are combined in the filtering process: the image gradient and the local orientation of structures. The processing consists in evaluating those properties in a given 3D neighborhood in the images and, with that information, steer the way the filter operates in terms of noise reduction and feature enhancement. For establishing quantitative values for the local orientation and amplitude, our study implements the necessary infrastructure for analytical signal analysis through banks of quadrature filters.

Nevertheless, using plain three dimensional filter kernels for both the orientation and actual filtering represents a burden that is not compatible with the processing power that can be embedded in tablets or other portable devices. Instead, our implementation uses a set of 1D filters aimed at computing a polynomial expansion of the considered 3D filters, leading to reducing the required processing power by an order of magnitude.

3.3 Implementation

Designing a real time requires fitting all required data in internal memory. Our implementation is organized around producing one output line along the x dimension at a time. The convolution oriented nature of the processing requires loading a collection of lines around the one being processed and moving one step in the y dimension requires to load xz subplane whose height is compatible with the required extent of the filter kernels. The loading of the next subplane is overlapped with processing the previous y line making the processing time completely CPU bound.

The performance of a CPU bound implementation is dictated by the clock cycle count. We have reduced the cycle count by using the SIMD capabilities of the C66x core and reducing the size of the data items. All intermediate

results are on 16 signed integers.

4 RESULTS

4.1 Performance results

The performance was assessed against a volume size of 512x256x128 pixels sampled at 10 Mvoxels/s. The selected memory layout requires ~410KB to hold all required data, which is below the limit of the 1MB available for each core. To sustain the required throughput the algorithm must execute 1.93 Gcycles/s. Hence, two C66+ cores clocked at 1.3 GHz achieve the desired performance.

4.2 Image quality

The image quality was assessed against volume sequences acquired by true 3D probes processed with both the reference 2D adaptive filter and 3D adaptive filter. As anticipated, the current implementation shows significant improvements over the 2D reference notably in the homogeneity of the tissues. Taking the difference with the reference 3D, the current implementation shows minor deviations in areas with a string gradient and energy variation. Those differences are limited to +5/-5 on the greyscale.

5 CONCLUSION

We have investigated the implementation of a 3D adaptive filtering algorithm on a high performance DSP based platform. The problem of limited in chip memory can be solved with appropriate handling of neighborhoods. With the help of the compiler and the SIMD unit, our implementation is capable of processing a volume of 512x256x128 voxels on two C66x cores in less than one second.

We have run our implementation against a set of volumes acquired from Ultrasound scanners equipped with a true 3D probe. The selected fixed-point representation can deliver similar image quality as our reference floating point implementations.

9399-9, Session 2

Joint demosaicking and integer-ratio downsampling algorithm for color filter array image

Sangyoon Lee, Moon Gi Kang, Yonsei Univ. (Korea, Republic of)

Most digital color cameras use a single CCD or CMOS sensor. However, single sensor can perceive only the brightness of a pixel. Therefore, color filter array (CFA) is used to obtain three different primary color information at each pixel. The Bayer pattern is most widely used as a CFA pattern. In the image captured by CFA structure, each pixel contains only one of the three color components such as red, green and blue, and that kind of image is called the mosaic image. The other two missing colors should be estimated from neighboring pixel to produce full color image. This process is referred as color demosaicking, which is also called as color interpolation process.

A number of color demosaicking algorithms for single sensor with CFA structure have been developed since the late 80's. Color interpolation method using inter-channel correlation in an alternating projections was proposed by Gunturk et al. Lu and Tan proposed two steps demosaicking: an interpolation step to produce full color image and a post-processing step to reduce demosaicking artifact. An adaptive homogeneity-directed demosaicking algorithm was presented by Hirakawa .

Beside demosaicking process, image resizing is another important issue. In the previous work in resizing process is almost focused on interpolation process. As the size of imaging sensor becomes bigger, the resolution of image grows higher. Currently, multiple mega-pixel imaging sensor and captured image is widely used. However, the resolution of display device is not as high as image device. Moreover, portable digital devices such as mobile phone or display in digital camera have much lower resolution than actual image size. Therefore, image downsampling process is arising as an important issue. The conventional method to generate downsampled full color image from sensor with CFA follows two steps: color demosaicking first and downsampling later. However, this process requires huge memory to store the result of color demosaicking process and long time to compute whole process. For these reason, it is important to consider joint demosaicking and downsampling approach.

Using subpixel-based joint demosaicking and downsampling methods for Bayer image were proposed by Fang et al. Subpixel-based downsampling can be applied only in LCD display and downsampling ratio is fixed with 6:1.

In this paper, we present the joint demosaicking and integer-ratio downsampling algorithm for CFA image. The proposed method produce full color image from Bayer CFA image and also resized by integer ratio.

Bayer CFA pattern structure has two G components and one R, B component. Bayer pattern can be divided in 2x2 block. In 2x2 block, G components are placed at diagonal position, and the others at the rest position. The image with Bayer CFA consists of three subsampled primary color information. The subsampled color components can be represented as product of fully sampled color component and subsampling function. The Fourier transform of Bayer CFA image can be decomposed to three components: luma component on baseband, chrominance component on (pi,pi) and another chrominance component on (0,pi) and (pi,0). The forward and inverse transform matrices between R, G, B channels and L, C1, C2 channels can be obtained by transform matrices. With appropriate lowpass, highpass and bandpass filter, one luma and two modulated chrominance components can be obtained from Bayer CFA image. The estimated chrominance components are modulated to baseband, and then all components are transformed to R, G and B channel by transforming matrix. Unlike L and C1 components, C2 components appears on (0,pi) and (pi,0), let them be C2a and C2b. C2a and C2b is overlapped with L in v and u direction respectively, so C2 information can be obtained by weighted summation of C2a and C2b values. Moreover, anti-aliasing filter is designed to avoid aliasing that can be generated in downsampling process. It is important that designing filters precisely to estimate accurate color demosaicked image. To estimate the filters, true color images and Bayer CFA images are used. Using L, C1, C2 components and Bayer CFA image, the filter for each component is estimated by least square solution. To choose appropriate C2 component, there needs more filters that determine the weight w. The interference between L and C2 information is detected most at (0,+-3pi/8) and (+-3pi/8,0), the Gaussian bandpass filters centered at frequencies (0,+-3pi/8) and (+-3pi/8,0) are designed. The weight w is calculated as portion of the energies of filtered results. Moreover, there is chance that aliasing occurs in downsampling process, so the anti-aliasing filters should be applied to the estimated filters. The anti-aliasing filters should have relevant cutoff frequencies that satisfy Nyquist sampling frequency. Filtering in frequency domain requires a lot of hardware resources. Therefore, filters can be transformed to spatial domain and applied to reduce hardware resources. The point spread function (PSF) of filters can be calculated by the inverse Fourier transform. By experiments, the sizes of filters can be minimized. However, Gaussian bandpass filters act as difference to row and column directions and filters can be substituted by difference operators. Finally, the filters for joint demosaicking and downsampling are obtained by cascading the estimated filters and the anti-aliasing filters.

The proposed algorithm is tested on the Kodak test set. In the experimental results, the proposed method is compared with "Demosaicking first and downsampling later" method. The test image with 512x768 is downsampled to 256x364 and the results are shown in Figure 6. The proposed method produces the result as good as original image without aliasing artifacts. Moreover, the proposed method does not need any frame buffer because it can produce full color component simultaneously. And it does not need to produce full size demosaicked image because. These reduce the hardware resources and computational costs of the proposed method.

In this paper, we presented joint demosaicking and downsampling with integer-ratio algorithm. Compared with other available method, experimental results showed that proposed method performs reasonable. It prevented aliasing artifact with preserving edge and detail well. With proposed algorithm, full color components were obtained simultaneously and the demosaicking and the downsampling process were performed at the same time. Therefore, the hardware resources and the computational cost of the proposed method might be much less than those of the conventional algorithms.

9399-10, Session 2

Intermediate color interpolation for color filter array containing the white channel

Jonghyun Kim, Sang Wook Park, Moon Gi Kang, Yonsei Univ. (Korea, Republic of)

Most digital cameras use a color filter array (CFA) to reduce the cost and the size of equipment instead of three sensors and optical beam splitters. The Bayer CFA pattern is a representative pattern. It consists of primary colors such as red, green and blue (R, G and B) and it has twice as many green pixels as red or blue ones. The G channel that occupies half of the total number of pixels, in particular, has a quincuncial structure. This is a geometric pattern consisting of five points arranged in a cross and forming an "X" shape. Due to the subsampling of the color components, color interpolation is required in order to reconstruct a full color image. Also, color aliasing should not appear during the process of color interpolation. With the quincuncial structure of the Bayer CFA pattern, several color interpolation algorithms have been developed to prevent color aliasing and to improve spatial resolution.

Recently, CFA patterns other than the Bayer CFA pattern have been designed to reduce the amount of color aliasing from the luminance component and the chrominance components. Also, other CFA patterns have been presented to provide robustness to aliasing and noise. However, there are many advantages to the CFA pattern that contains the W channel. Since the W channel has broad spectral bands and a high light sensitivity compared with R, G and B channels, the W channel has a high signal-to-noise ratio (SNR) regardless of the color temperature of an illumination. Consequentially, it is possible to obtain faster shutter speed or shorter exposure time in low light conditions. In addition to its effect on exposure, it allows for the reduction of motion blur.

However, various color interpolation algorithms for the Bayer CFA pattern are inapplicable, because the CFA pattern that contains the W channel does not have a quincuncial structure. To utilize a color interpolation algorithm for the quincuncial structure, it is necessary to convert the CFA pattern containing the W channel to a quincuncial structure. A method in which the R, G and B channels are obtained by color interpolation with quincuncial patterns is studied. The positions of the G and W channels are only estimated by partially applying the linear color interpolation model through frequency analysis. Finally, any edge adaptive color interpolation method for the Bayer CFA pattern can be applied after generating two quincuncial patterns.

In this paper, intermediate color interpolation based on multi-scale gradients is proposed to obtain the quincuncial structure from the pattern containing the W channel. Two quincuncial patterns are obtained through intermediate color interpolation. Then, a full color image is reconstructed by using a color interpolation algorithm. By generating an intermediate quincuncial pattern, various color interpolation algorithms can be applied to it.

The proposed intermediate quincuncial interpolation method begins with edge directional estimates. After estimating the initial results, the final quincuncial patterns are obtained based on the multi-scale gradients. The color difference between the G and W channel is defined as the multi-scale gradients.

The first step of the algorithm is obtaining initial edge directional estimates for the G and W channels. Generally, the G and W pixel values correlate well with each other because the spectrum band of the G channel is located on the center of the spectrum band of the W channel. Additionally, a structure in which the G and W pixels are arranged in a cross and repeated diagonally is fully interpolated by using the gradients. And the edge direction for interpolation is estimated by using the gradients of neighboring pixels. Taken together, the weighted average of the gradients of the G pixel is added to the weighted average of the adjacent W pixel. Then, we obtain the color difference between the W pixels and the G pixels.

Subsequently, the color differences are combined to form the directional difference estimation which is divided by the size of the local window. According to the distance between the center pixel and the outermost pixel, the weighted difference estimation and the weights for each direction are obtained. Finally, the missing G channel is interpolated at the W channel. It is defined by the sum of the weighted directional difference estimation.

An approach based on multi-scale gradients gives more reliability in terms of edge directions and high frequency information than when only using initial edge directional interpolation. Estimations of W pixels are calculated in same manner. Finally, two quincuncial patterns with different color channels among the W and G channels are obtained by intermediate color interpolation.

The performance of the proposed method is evaluated on the Kodak PhotoCD image set. The pattern containing the W channel is generated by subsampling with the CFA pattern from a full color image. It is assumed that the level of the W channel is an average of the R, G and B channels. For comparison, the linear color interpolation method is utilized to generate intermediate images with a quincuncial pattern. By applying the proposed method, another quincuncial pattern is generated. Next, the color interpolation method proposed in multiscale gradients-based color filter array interpolation (MSG) is used for the quincuncial patterns equally. As shown in the peak signal-to-noise ratio (PSNR) comparison results, the proposed algorithm outperforms the conventional method. Its average PSNR is higher than the conventional method by 1.13-1.87dB. Moreover, the results of the proposed method show that visual artifacts such as grid effects and false colors are reduced.

In this paper, a method for generating quincuncial patterns based on multi-scale gradients was proposed. The developed method was carried out with a CFA sensor containing the W channel. First, the missing W and G channels at the G and B locations were obtained by initial edge directional interpolation respectively. Then, the initial estimates were used to generate the intermediate quincuncial patterns. This approach based on multi-scale gradients through initial interpolation provided information about the reliability of the edge direction as well as the high frequency information. The performances of the proposed method and the conventional method were compared in the experimental results section. According to the experimental results, the algorithm proposed in this paper outperforms the conventional method in terms of subjective visual quality as well as objective value.

9399-11, Session 3

The future of consumer cameras *(Invited Paper)*

Sebastiano Battiato, Marco Moltisanti, Univ. degli Studi di Catania (Italy)

In the last two decades multimedia, and in particular imaging devices (camcorders, tablets, mobile phones, etc.) have been dramatically diffused. Moreover the increasing of their computational performances, combined with an higher storage capability, allows them to process large amount of data. In this paper an overview of the current trends of consumer cameras market and technology will be given, providing also some details about the recent past (from Digital Still Camera (DSC) up today) and forthcoming key issues. We report just one example: the market trend of DSCs sales where is evident significant decrease in the last years (Fig.1). More data will be presented and discussed at conference time,

After a brief introduction to the fundamentals of image acquisition on consumer cameras [1, 2] with some details about the differences in terms of overall performances, cost and market success of various cases-study, we will introduce the current and mainly the next technologies, with related computational frameworks and applicative scenarios. Recent market forecasts already give ideas about novel and emerging sectors (with tons of related applications and services) such as wearable devices and depth cameras. The success of new imaging devices will be strictly connected with the consumer behaviour and needs of sharing in real time (e.g. by update the own status on Facebook, etc.) and/or add semantic tags (e.g., people, place, etc.) to the multimedia content itself for further processing and/or dedicated services (e.g., audience measurement, lifeblogging, summarization, etc.).

[1] Image Processing for Embedded Devices – Eds. S. Battiato, A.R. Bruna, G. Messina, G. Puglisi – Applied Digital Imaging ebook series ISBN: 978-1-60805-170-0 – Bentham Science publisher, 2010.

[2] S. Battiato - Single-Sensor Imaging Devices: an Overview - Chapter in Handbook of Digital Imaging, edited by Michael Kriss – John Wiley & Sons Ltd (2014 – In press);

9399-12, Session 3

Challenges towards a smart security camera *(Invited Paper)*

Lucien Meijer, Ildiko Suveg, Bosch Security Systems (Netherlands)

In this paper, we will present a smart security camera that is autonomous in operation and can provide good image quality for scene recognition, while offering low-bit rate and absence of artifacts. Security cameras are employed in many types of user scenarios, indoor and outdoor environments, and should work continuously 24 hours a day, seven days a week yielding the best possible image quality. Cameras are often employed without special adjustments made from the out-of-the-box factory defaults, while having unknown user scenario and variable scene lighting conditions. More complex algorithms for computer vision and pattern recognition, so called video analytics are gradually getting embedded into security cameras.

Currently, the term “Intelligent IP camera” is often understood as camera with embedded video analytics that is able to track and recognize objects in the scene, as well as automatically detect out-of-order acts or emergency situations. We extend the term of Intelligent IP camera by also considering the other components of a security camera: digital image processor and the encoder. In this sense, we are developing a smart camera concept, that aims to automatically and on-fly optimize the three key blocks of a camera system: image pipeline, encoder and intelligent video analytics (IVA) by adding intelligence to each of the blocks and also sharing information / data between the blocks. In this way, the system as a whole can benefit from the knowledge of individual components. We propose a hardware and software architecture that allows the tight integration of the components.

For instance, well exposed images can help the video analytics to detect objects of interest in the scene. Furthermore, having the objects of interest detected in the scene, image exposure can be optimized on them. In addition, the IVA block can transfer the regions of detected important objects in a priority order to the encoder and it can detect and classify important events. This information can provide a trigger signal to the encoder to increase the overall quality of encoding during the important detected event, and/or to increase locally (inter-frame) the quality of encoding of detected objects of interest, while reducing the quality of unimportant image regions.

Image quality (according to contextual or perceptive criteria) is determined not only by the quality of optics and sensor, but also to a significant degree by the intelligence built into the image pipeline. In order to achieve full camera automation, it is important to detect the type of the scene. For example, classification to indoor or outdoor scene can be of interest, since many image pipeline settings and algorithms should have different behaviour outdoor compared to indoor. One can think of different parameters for a colour constancy algorithm, as well as different parameters for sharpness, contrast and noise reduction algorithms. The image pipeline should also perform content-dependent enhancement and prepare the signal for optimal compression: strengthens useful information and removes noise. Intelligent dynamic noise reduction (iDNR) optimizes bandwidth by dynamically adjusting the degree of noise reduction based upon an analysis of motion detected in the scene. Motion detection and the estimation of the magnitude of motion are performed in IVA.

Another important direction of development is to improve the encoders, by introduction of adaptable control algorithms for encoding depending on the load of communication channels and content of the observed scene.

In this paper we describe the hardware and software architecture that allows the tight integration of the components: image processor, scene analysis and encoder and also the challenges in developing such a system.

9399-13, Session 3

Image quality based x-ray dose control in cardiac imaging

Andrew G. Davies, Stephen M. Kengyelics, Amber J. Gislason-Lee, Univ. of Leeds (United Kingdom)

Background

The control of radiographic settings in cardiac X-ray imaging systems is governed by a closed loop feedback system. This automatic dose control (ADC) is responsible for ensuring that the quality of images produced is of sufficient quality for the clinical task, that the radiation dose to the patient is acceptable, and that the operation limits of the system components are observed. The ADC must accommodate differences in patient body habitus, and respond to changes in imaging conditions dictated by the user, such as changes in the projection angle being used. A common method of ADC design is for the X-ray system to maintain a constant X-ray dose rate to the X-ray image receptor. This approach may not mean that the quality of images produced is specifically tailored to the clinical application in hand (context), and in some cases may mean that some patients are receiving more radiation dose than necessary as there is a link between image quality and patient dose.

Aims

This study aimed to investigate the effect on image quality and patient dose of an ADC design that sought to maintain a constant image quality rather than radiation dose to the receptor using computer simulation.

Material and Methods

A software simulation of a simplified X-ray imaging system was created in MATLAB (Mathworks, Natick MA). Patients were simulated as simple geometric blocks of polymethylmethacylate (PMMA) which has similar X-ray attenuation to soft tissue. The thickness of the PMMA phantom was varied to simulate patients of different size. The X-ray source was simulated as typical of cardiac X-ray system (a tungsten anode X-ray tube with an anode angle of 11 degrees, inherent filtration of 3.5 mm Al, constant potential 80 kW generator). X-ray output spectra and attenuation data were obtained from published sources. Only primary X-ray photons are considered in the simulation. The patient couch was simulated as an attenuator of 3.5 mm Al. The anti-scatter grid was simulated as absorbing 50% of the primary beam. The X-ray image receptor scintillator was simulated as a 0.55 mm CsI layer. The source to object distance was set to 70 cm, and the source to image distance adjusted to leave a 10 cm gap between the phantom exit and the receptor. An acquisition rate of 15 frames per second was assumed for dose rate calculations. A detail was included in the phantom to simulate a coronary artery filled with contrast medium.

The simulator predicted the following quantities: phantom entrance surface dose (ESD), and energy deposited in the image receptor’s scintillator in the background and in the shadow of the detail. From these the noise in the image background and in the area covered by the vessel detail and the contrast of the detail were calculated, and combined into a contrast to noise ratio, CNR, used to represent the quality of the image that would have been produced.

Two ADC methods were modelled using computer simulation. The first method (constant dose) set the X-ray tube potential difference (kV) to achieve a target dose per image at the image receptor. The tube current (mA) and pulse duration (ms) were derived from look up tables based on the kV and the relationships between kV and mA and kV and ms were programmed to be those found on a commercially available cardiac X-ray system (Allura Xper FD10, Philips Healthcare, The Netherlands). The second ADC model operated in a similar manner but aimed to achieve a target CNR of the vessel detail. The ADC designs were compared in terms of the dose delivered and predicted image quality for PMMA thicknesses of 15, 20, 25 and 30 cm. The target detector dose for the constant dose ADC was set to 90 nGy per frame, and two target CNRs for the constant quality ADC were investigated.

Results

For both ADC designs there was a non-linear relationship between ESD and phantom thickness. For the constant dose ADC the ESD rates were

1.0, 3.2, 89 and 15.6 mGy/s for the 15, 20, 25 and 30 cm thickness phantom respectively. This relationship was more extreme for the constant quality ADC. At the lower quality mode the corresponding ESD rates were 0.37, 1.4, 5.2 and 23.2 mGy/s, and for the higher quality mode 0.64, 2.4, 9 and 23.2 mGy/s.

Predicted image quality is inversely related to phantom thickness for the constant dose ADC, and was predicted to be 20.7 18.9, 16.2 and 10.8 for the 15, 20, 25 and 30 cm thickness phantom respectively. CNR for the constant quality ADC and was 12.2 for the lower quality setting and 16.3 for the higher quality setting except for the 30 cm phantom at the higher quality setting. For this combination the ADC was not able to maintain the quality of images as it had reached the maximum permitted kV, and produced a lower quality (12.2) and ESD than was expected.

Conclusion

X-ray imaging systems must deliver adequate image quality for the clinical task being undertaken. However, the image quality delivered by a constant receptor dose ADC decreases with patient size. An image quality normalising ADC could produce a more context aware operation by setting the quality of images to that required to the clinical task, tailored to the needs of the operator. By using such a system it is possible that considerable radiation dose savings could be made (reducing harm to the patient), especially for thinner patients. However, maintaining a constant dose for very large patients may result in an unacceptably high radiation dose rate being delivered.

9399-14, Session 3

Selecting stimuli parameters for video quality assessment studies based on quality similarity distances

Asli E. Kumcu, Ljiljana Platiša, Univ. Gent (Belgium); Heng Chen, Vrije Univ. Brussel (Belgium); Amber J. Gislason-Lee, Andrew G. Davies, Univ. of Leeds (United Kingdom); Peter Schelkens, Vrije Univ. Brussel (Belgium); Yves Taeymans, Univ. Ziekenhuis Gent (Belgium); Wilfried Philips, Univ. Gent (Belgium)

Purpose:

This work presents a methodology to optimize the selection of multiple parameter levels of an acquisition, degradation, or post-processing process applied to stimuli intended to be used in a subjective image or video quality assessment (QA) study. Parameter values are often chosen within the parameter space (e.g. a step size equal to some compression bit-rate or noise standard deviation value) or with technical measures (e.g. peak signal-to-noise ratio (PSNR)). However, it is known that processing parameter values and technical / objective measures of quality are usually nonlinearly related to human quality judgment. If they are used to select the parameter levels of a stimulus, the quality levels may be unevenly distributed within the considered quality range. This may skew the estimate of the quality model. For example, at extremely high or low quality levels, humans often perceive very little or no quality differences. If stimuli parameter levels are not well chosen, an objective measure may incorrectly estimate the levels at which subjective quality differences become perceptible. Similarly, the shape of the nonlinear relationship between subjective scores and objective measures at intermediate quality levels may also be incorrectly estimated.

To overcome this, we propose a method for modeling the relationship between parameter levels and perceptual quality distances using a paired comparison procedure where subjects judge the perceived similarity in quality. Our goal is to use the modeled relationship to select parameter values that are equidistant in the perceptual quality space; that is, the perceived quality of the stimuli should be distributed equally within the considered quality range. This approach is tested on two clinical applications: (1) selection of compression levels for laparoscopic surgery video, and (2) selection of dose levels for an interventional x-ray system.

Methods:

For the compression experiment, a 10-second scene from a High-Definition (1920x1080 pixels) laparoscopic surgery video was extracted and compressed by H.264/AVC at seven bit-rates (20, 8.5, 5, 3.5, 2.5, 1.85, and 1.5 Mbps) with parameters optimized for low-latency encoding and decoding, resulting in 8 stimuli. We consider two successive compression levels, e.g. 20 and 8.5 Mbps, as being parametrically adjacent. A total of twenty-eight pairs of sequences (paired combinations of the reference and seven compressed sequences) were presented to two observers.

For the x-ray dose experiment, a static anthropomorphic chest phantom containing contrast filled coronary arteries (Radiology Support Devices Alderson Phantoms, Long Beach, USA) was imaged at six dose levels on an Allura interventional x-ray system (Philips Healthcare, Best, The Netherlands). The sequences (890 x 890 pixels) were acquired at 15 frames per second in raw format without any proprietary image processing. We consider two successive dose levels as being parametrically adjacent. A total of fifteen pairs of sequences (paired combinations of six dose levels) were presented to four doctors from Ghent University Hospital specializing in interventional cardiology and/or cardiac electrophysiology.

For both experiments, subjects evaluated the similarity in quality of each video pair using a continuous scale from 0 (completely different) to 100% (exactly the same quality). Sequences were presented sequentially using the video presentation protocol from the Double-Stimulus Continuous Quality Scale (DSCQS) method defined in ITU-R recommendation BT.500. Presentation ordering was randomized. Sequences were displayed on a 24" surgical display (MDSC-2124, Barco, Kortrijk, Belgium).

The analysis procedure consisted of six steps. First, a mean dissimilarity matrix was computed from the raw similarity scores. Classical multidimensional scaling was used to reduce the dimensionality of the dissimilarity matrix, producing a distance matrix. The eigenvalues were used to determine the minimum number of dimensions that accurately reproduced the original distances. The Euclidean distance between parametrically adjacent stimuli were computed in the reduced dimension space and then fit to the perceptual distances, resulting in a model of the relationship between the parameter level and quality difference within the range of parameters tested. This function was used to select parameter levels that were perceptually equidistant.

Finally, the estimated quality differences were compared to the distribution of the quality scores obtained from a follow-up Single Stimulus Continuous Quality Evaluation (SSCQE) experiment for each application. In the follow-up x-ray dose experiment, subjects were also asked to evaluate the 'visibility of the coronary tree' on a continuous scale from 0 (Poor) to 100% (Excellent).

Results:

The results for both experiments indicated that two dimensions were sufficient for estimating the perceptual quality distance between parametrically adjacent stimuli. The relationship for the compression experiment was modeled with a power law function. The lowest compression level (20 Mbps) was first selected such that it was sufficiently similar in quality to the reference sequence, but with some visible compression artifacts. The remaining three bit-rates (5.6, 2.9, and 1.85 Mbps) were selected using the perceptual function such that the differences between the four bit-rates were approximately perceptually equidistant. Quality scores from the follow-up SSCQE experiment conducted with the chosen bit-rates were approximately evenly distributed, indicating that the similarity distance model correctly estimated the relationship between the subjective quality scores and the chosen parameter values.

A linear relationship between dose and quality differences was found for the x-ray dose parameter experiment. The quality scores from the follow-up SSCQE experiment showed poor correlation with the estimated quality differences. However, scores for the 'visibility of the coronary tree' were moderately (monotonically and nonlinearly) related to dose. In this clinical application, the similarity distances estimated the doctors' subjective opinion of the visibility of clinically relevant structures, but not the quality rating itself. This suggests that the appropriate subjective task must be chosen for some medical applications.

Conclusions: Our two pilot experiments indicate that a paired-comparison, quality similarity judgment task can assist the selection of optimal parameter levels for a subjective QA study. This methodology ensures that the perceived quality scores of the selected parameter levels (compression bit-rate and x-ray dose) are evenly distributed within the quality range of

interest. The type of subjective assessment task may have an impact on the estimated relationship with the parameter level. Future work will consider more robust statistical approaches to account for inter-subject variability in the use of the quality similarity scale.

9399-29, Session PTues

No-reference visual quality assessment for image inpainting

Viacheslav V. Voronin, Vladimir A. Frantc, Vladimir I. Marchuk, Alexander I. Sherstobitov, Don State Technical Univ. (Russian Federation); Karen O. Egiazarian, Tampere Univ. of Technology (Finland)

Inpainting has received a lot of attention in recent years and quality assessment is an important task to evaluate different image reconstruction approaches. In many cases inpainting methods introduce a blur in sharp transitions in image and image contours in the recovery of large areas with missing pixels and often fail to recover curvy boundary edges. Quantitative metrics of inpainting results currently do not exist and researchers use human comparisons to evaluate their methodologies and techniques. Most objective quality assessment methods rely on a reference image, which is often not available in inpainting applications. This paper focuses on a machine learning approach for no-reference visual quality assessment for image inpainting based on the human visual property. Our method is based on observation that when images are properly normalized or transferred to a transform domain, local descriptors can be modeled by some parametric distributions. The shapes of these distributions are different for non-inpainted and inpainted images. Next, we use a support vector regression learned on assessed by human images to predict perceived quality of inpainted images. We demonstrate how our predicted quality value repeatably correlate with qualitative opinion in a human observer study. We show that our approach outperforms known and widely used algorithms on a selected image dataset both in terms of Z-score and Pearson product-moment correlation coefficient. Results are shown on a human-scored dataset for different inpainting methods.

9399-30, Session PTues

Pentachromatic Colour Spaces

Alfredo Restrepo, Univ. de los Andes (Colombia)

We generalize results presented previously, for dimensions 3 and 4, to dimension 5. We exploit the geometric properties of the 5-hypercube [0, 1]5 in order to give a mathematical model for colour vision in the case of 5 photoreceptor types and for the corresponding additive colour combination with five primary lights. The paper is concluded with a gen- eralisation to dimension n ? N.

The interval [0,1] is meant to be the set of possible intensities of a primary light, in an additive colour combination, or of a photoreceptor response being the basis for colour detection or perception. We start then with the cubic colour space [0, 1]5 ? R5 of colour points (v, w, x, y, z) ? R5, sometimes denoted as (p0, p1, p2, p3, p4) as well.

The interior of the 5-hypercube is an open 5-cell and we characterise the position of its points with reference to the boundary of the hypercube. The boundary ? := ∂[0,1]5 consists of the points having at least one 0-valued coordinate or one 1-valued coordinate.

The points of the hypercube having a given coordinate pi equal to 1, or to 0, can be classified into 5 ? 2 = 10 4-cubes, their union is ?, which is a PL (piecewise linear) topological 4-sphere S4. In fact, in addition to these 10 4-cubes, a cellular decomposition of ? with 40 3-cubes, 80 squares, 80 edges and 32 points, results as follows: by fixing 2 of the coordinates pi and pj of points in the 5-cube with values in {0, 1}, you get the 40 3-cubes of ?; fixing 3 coordinates in {0, 1}, you get the 80 squares of ?; fixing 34 coordinates, 80 edges and, fixing all 5 coordinates, 25 = 32 points or vertices. 4 Out of the 40 3-cubes in ?, there are 20 that do not have neither the point s = (0 0 0 00)norw=(11111)asavertex. These 20 cubes for man"equatorial S3"of?,c alled ?; the points of ? are precisely those points of ? having at least one coordinate of value 0 and one coordinate of value 1. This equatorial ?, also called the chromatic sphere (of codimension 2 in R5) is used to define pentachromatic hue by giving coordinates to the points of [0, 1]5 on the basis of ?. There are other ways to define a pentachromatic hue as there are other closed 3-manifolds (unions of 3-cubes in ?) in the cell complex ?, but the equatorial 3-sphere is in a sense a canonical choice.

In fact, since each of the cubes in ? can be triangulated into 3! = 6 tetrahedra, each of which containing the points with a specific ordering of their coordinates (e.g. p0 = 0 ≤ p1 ≤p2 ≤p3 ≤p4 =1determinesatetra hedroninthecube{p0 =0,p4 =1}),yougeta triangulation of ? into 20 ? 6 = 120 tetrahedra, each of which containing the points with one of the possible orderings of the coordinates of a colour (each such tetrahedron precisely corresponds then to an element of the symmetric group S5). We say that the colour points on each of these tetrahedra belong to a given (1 out of 120 possible) hue family. All colours in a given hue family have the same ordering regarding the relative contributions from each of the photoreceptors (an analogy with the trichromatic RGB case would be for example the family of the oranges, where R ≥ G ≥ B.).

As each colour point (p0, p1, p2, p3, p4) not on the achromatic segment ? := {(?, ?, ?, ?, ?) ? R5 : ? ? [0, 1]} belongs to the unique triangle (called the chromatic triangle of the colour point) having as one side ? and as opposing vertex the point (q0, q1, q2, q3, q4) ? ? with by qi = pi?m, where m := min{pi : i ? 5} and ? := max{pi}?min{pi}, we say ? that the points of ? give the hue of the chromatic points of the hypercube; also, that each chomatic colour point belongs to at least one hue family (points with a hue on a boundary of a 3-cubes of ? belong to exactly two families).

Consider the graph having as nodes the tetrahedra in the chromatic sphere (the elements of S5) where two nodes are joined by an edge precisely when the two corresponding permu- tations differ by a transposition of two consecutive elements. When a colour changes from a hue family to another, the two corresponding permutations differing by a transposition of two consecutive elements, we say that a mild hue family change has occurred. In the graph, we find a Hamiltonian circuit; in this way, we give a cyclic order for the hue families; similar to the trichromatic case where you can give a cyclic order to all the hues.

For colour processing, it is better to transform the hypercube into a round "Runge" space, do the transformations on this Euclidean 5-ball before going back to coordinates in the piecewise linear hypercube. It is more intuitive and less prone to ending up with a "forbidden" colour. For each colour point p = (p0,p1,p2,p3,p4) ? [0,1]5, consider the midrange ? := 0.5[max{pi}+min{pi}] and the range ?. Each colour point can so be mapped to a point on the triangle of points with coordinates (??); it is an isosceles triangle with base {(?,?) : ? ? [0,1],? = 0} and height {(?,?) : ? = 0.5,? ? [0,1]}. We say that ? measures the luminance of the colour and that ? measures its chromatic saturation. Also, it is shown that each chromatic triangle maps bijectively to this ?-? (luminance-saturation) triangle. By considering a "solid of revolution" or, more precisely, by spinning the triangle with respect to a round S3 (a round version of ?) you get a double-cone type space. If the triangle is deformed into half a round disk, after spinning, you get a round 5-ball, called Runge 5-space.

Several algorithms for colour processing are given. Also, the routines for the transfor- mations between the hpercube, the double-cone type space and the round Runge space are given.

9399-31, Session PTues

A comparative study of two prediction models for brain tumor progression

Deqi Zhou, Princess Anne High School (United States); Loc Tran, Old Dominion Univ. (United States); Jihong Wang, The Univ. of Texas M.D. Anderson Cancer Ctr. (United States); Jiang Li, Old Dominion Univ. (United States)

MR diffusion tensor imaging (DTI) technique together with traditional T1 or T2 weighted MRI scans supplies rich information sources for brain cancer

diagnoses. These images form large-scale, high-dimensional data sets. Due to the fact that significant correlations exist among these images, we assume low-dimensional geometry data structures (manifolds) are embedded in the high-dimensional space. Those manifolds might be hidden from radiologists because it is challenging for human experts to interpret high-dimensional data. Identification of the manifold is a critical step for successfully analyzing multimodal MR images.

We have developed various manifold learning algorithms (Tran et al. 2011; Tran et al. 2013) for medical image analysis. This paper presents a comparative study of an incremental manifold learning scheme (Tran. et al. 2013) versus the deep learning model (Hinton et al. 2006) in the application of brain tumor progression prediction. The incremental manifold learning is a variant of manifold learning algorithm to handle large-scale datasets in which a representative subset of original data is sampled first to construct a manifold skeleton and remaining data points are then inserted into the skeleton by following their local geometry. The incremental manifold learning algorithm aims at mitigating the computational burden associated with traditional manifold learning methods for large-scale datasets. Deep learning is a recently developed multilayer perceptron model that has achieved start-of-the-art performances in many applications. A recent technique named "Dropout" can further boost the deep model by preventing weight coadaptation to avoid over-fitting (Hinton et al. 2012).

We applied the two models on multiple MRI scans from four brain tumor patients to predict tumor progression and compared the performances of the two models in terms of average prediction accuracy, sensitivity, specificity and precision. The quantitative performance metrics were calculated as average over the four patients. Experimental results show that both the manifold learning and deep neural network models produced better results compared to using raw data and principle component analysis (PCA), and the deep learning model is a better method than manifold learning on this data set. The averaged sensitivity and specificity by deep learning are comparable with these by the manifold learning approach while its precision is considerably higher. This means that the predicted abnormal points by deep learning are more likely to correspond to the actual progression region.

9399-32, Session PTues

Enhancement of galaxy images for improved classification

John A. Jenkinson, Artyom M. Grigoryan, Sos S. Agaian, The Univ. of Texas at San Antonio (United States)

In this paper, the classification accuracy of galaxy images is demonstrated to be improved by enhancing the galaxy images. Galaxy images often contain faint regions that are of similar intensity to stars and the image background, resulting in data loss during background subtraction and galaxy segmentation. Enhancement darkens these faint regions, enabling them to be distinguished from other objects in the image and the image background, relative to their original intensities. The heap transform is employed for the purpose of enhancement.

Segmentation then produces a galaxy image which closely resembles the structure of the original galaxy image, and one that is suitable for further processing and classification. 6 Morphological feature descriptors are applied to the segmented images after a preprocessing stage and used to extract the galaxy image structure for use in training the classifier. The support vector machine learning algorithm performs training and validation of the original and enhanced data, and a comparison between the classification accuracy of each data set is included. Principal component analysis is used to compress the data sets for the purpose of classification visualization and a comparison between the reduced and original feature spaces. Future directions for this research include galaxy image enhancement by various methods, and classification performed with the use of a sparse dictionary. Both future directions are introduced.

9399-33, Session PTues

Face retrieval in video sequences using Web images database

Marco Leo, RadioLabs (Italy); Federica Battisti, Marco Carli, Alessandro Neri, Univ. degli Studi di Roma Tre (Italy)

The automatic annotation of broadcasting news programs is a challenging task for multimedia press review. Usually in a video stream the head of a person moves continuously and changes in facial expressions, lighting conditions, and camera motion produce significant distortions in the image appearance that can largely affect recognition performances. In this paper a tool for automatic face identification in TV broadcasting programs is presented. The proposed approach is based on a joint use of Scale Invariant Feature Transform descriptor and Eigenfaces-based approach. The algorithm has been tested on video sequences using a database of images acquired starting from web search. Experimental results show that the joint use of these two approaches improves the recognition rate.

9399-34, Session PTues

Development and validation of an improved smartphone heart rate acquisition system

Gevorg Karapetyan, Rafayel Barseghyan, Hakob G. Sarukhanyan, Institute for Informatics and Automation Problems (Armenia); Sos S. Agaian, The Univ. of Texas at San Antonio (United States)

In this paper we propose an improved mobile phones/tablets application for robust heart rate (HR) measurement. The effectiveness and robustness over the existing approaches have been demonstrated under different distances from camera source and illumination conditions. Moreover, we compare the developed method with related commercial applications of remote heart rate measurements on mobile devices including results derived from mobile devices to electrocardiograph (ECG) and from Food and Drug Administration (FDA) approved sensors (devices). Furthermore, we present the comparison between Red (645 nm), Green (530 nm) wavelength light and Blue (470 nm) Light Reflection Photoplethysmography (PPG) for HR monitoring during motion under different conditions. We show that the green wavelength light, in general, has the potential to be a better method for monitoring HR during normal daily life. Finally, we have performed various experiments with different mobile phones and tablets. HRs were collected simultaneously from 10 subjects, ages 22 to 65, by using the 3 devices during 5-minute periods, while at rest, reading aloud under observation, and playing a video game or doing research under different distances from camera source and illumination conditions.

9399-35, Session PTues

New 2D discrete Fourier transforms in image processing

Artyom M. Grigoryan, Sos S. Agaian, The Univ. of Texas at San Antonio (United States)

There are many unitary transforms which are used effectively in image processing, such as the Fourier, Hadamard, and Haar transforms. The theory of the Fourier transforms and effective methods (or, fast algorithms) of the discrete Fourier transforms (DFT's) are well developed and used for solving different problems in the area of data processing, such as filtering, compression, restoration, etc. This transform has the simple physical meaning and transfers the data defined on the real space into the complex. Other transforms, for instance the Hadamard transform do not keep many useful properties which the Fourier transform has, for instance, the property

of the spectra multiplication for the linear convolution, the existence of the transform and fast algorithms for arbitrary order, etc. Therefore, this transform cannot contest with the Fourier transform for the 2^r-by-2^r case, although the Hadamard transform can be effectively performed using the well-known recurrent formula for computation.

Since the Fourier transforms are a fundamental tool in signal and image processing. There have been many attempts to generalize the commonly used Fourier transform, including the quaternion Fourier transform tailored to color images. In this paper, the concept of the two-dimensional Fourier transform is defined in the general case, when the form of relation between the spatial points (x,y) and frequency points (omega1,omega2). is defined in the exponential kernel of the transformation by a nonlinear form L(x,y;omega1,omega2). The traditional concept of the 2-D DFT is defined for the Diophanous form x(omega1)+y(omega2). This 2-D DFT is the particular case of the Fourier transforms described by such forms L(x,y;omega1,omega2). We describe a new class of the Fourier transforms with nonlinear forms describing the relation between the spatial and frequency points. We analyze in detail the one- and two-dimensional cases, and the multidimensional case is described similarly. For effective calculation of Fourier transforms of this class, we derive another complete system of functions from the Fourier transforms as a system which split the mathematical structure of these transforms. In other words, we describe the functions which can split directly the complex two-dimensional structure of the Fourier transform into a number of the separate 1-D transforms. The complete system of such functions is similar to the discrete paired functions which first were introduced for effective calculation of the 2-D Fourier, Hadamard, Hartley, and cosine transforms. Properties of such 2-D discrete Fourier transforms are described and examples of application in image processing are given. The special case of the N-by-N-point 2-D Fourier transforms, when N=2^r, r>1, is analyzed and effective representation of these transforms is proposed. Together with the traditional 2-D DFT, the proposed class of 2-D DFTs can be used in image processing in image filtration and image enhancement.

9399-36, Session PTues

Printed Arabic optical character segmentation

Khader Mohammad, Muna Ayyesh, Birzeit Univ. (Palestinian Territory, Occupied); Aziz Qaroush, Iyad Tumar, Birzeit University (Palestinian Territory, Occupied)

A considerable progress in recognition techniques for many non-Arabic characters has been achieved. In contrary, few efforts have been put on the research of Arabic characters. In any Optical Character Recognition (OCR) system the segmentation step is usually the essential stage in which an extensive portion of processing is devoted and a considerable share of recognition errors is attributed. In this research, a novel segmentation approach for machine Arabic printed text with diacritics is proposed. The proposed method reduces computation, errors, gives a clear description for the sub-word and has advantages over using the skeleton approach in which the data and information of the character can be lost. Both of initial evaluation and testing of the proposed method have been developed using MATLAB and shows 98.7% promising results.

9399-37, Session PTues

Highly accelerated dynamic contrast enhanced MRI using region of interest compressed sensing

Amaresha S. Konar, Rashmi Rao, Nithin N. Vajuvalli, C. K. Dharmendra Kumar, Divya Jain, Sairam Geethanath, Dayananda Sagar Institutions (India)

Introduction: Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) is one of the most preferred methods to understand and analyze tumor micro-environment of tissue by injecting the contrast agent (1). DCE-MRI can be used to acquire high contrast images to qualitatively analyze and to measure properties of the particular tissue voxels such as T1, T2, diffusion tensor, magnetisation transfer, metabolite concentration, Ktrans (contrast transfer), etc., which are very useful in clinical analysis of cancer tissues (2). One significant drawback of MRI is slow data acquisition and reconstruction compared to other imaging modalities such as Computed Tomography (CT). DCE-MRI needs accelerated technique to acquire more data during the post contrast. Compressed Sensing (CS) is a widely used acceleration technique in MRI for reconstruction of data from highly undersampled measurements (3). Our group has recently developed a novel CS technique to accelerate MRI called Region of Interest Compressed Sensing (ROICS) (4). ROICS allows for increased sparsity, which is one of the key criteria for CS reconstruction. It is based on the hypothesis that superior CS performance can be achieved by limiting the CS to Region of Interest (ROI). In DCE MRI, this ROI would be the suspected cancer region which is smaller in size compared to entire image. This would be localized typically using a T2 weighted image. This work aims at application of ROICS on DCE MRI data to achieve increased acceleration during post contrast.

Theory: Conventional CS can be represented by eq. [1]

min?m?(??F_u (m)-y??_2+????(m)??_1) [1]

where, m is the current estimate of the image to be obtained, Fu is the undersampled orthonormal Fourier operator: F(.)* Undersampling mask, y is the undersampled k-space measured by the acquisition process, ? is the regularization factor, determined by methods like Tikhonov regularization or L-curve optimization (5), ? is the sparsifying transform operator and ||.||k is the k-norm operator.

Data consistency term was evaluated in the spatial domain and ROICS eq. was derived by weighting the spatial data consistency term over a ROI and this results in eq. [2],

min?m?(??F^(-1) ?(F?_u (m)-y)*W??_2+????(m*W)??_1) [2]

where, F-1 is the inverse Fourier transform and W is the Ns*Ns diagonal matrix containing a spatial weighting that one can use to specify and evaluate a ROI, of the dimensions of the image. Data sparsity is the key criterion in CS and ROI mask inclusion enhances the data sparsity in the ROICS reconstruction and which results in better reconstruction compared to conventional CS.

Method: Breast DCE-MRI taken from the cancer imaging archive website (6) and they have acquired using Siemens 3T scanner. Full breast coverage was acquired with a 3D gradient echo-based Time-resolved angiography WIth Stochastic Trajectories (TWIST) with TE/TR 2.9/6.2 ms, 320x320 in-plane matrix size for 28 images and the contrast agent used for study was Gd (HP-DO3A). T2 map data was obtained for considering the suspected cancer region by selecting Region of Interest (ROI). CS and ROICS technique has been applied for the dataset at chosen acceleration factors 2x, 5x, 10x, 15x, and 20x. ROI selected mask was considered as a weighting function to perform ROICS by restricting the reconstruction to ROI, where as conventional CS reconstruction was performed for the entire image. Reconstruction error in ROI selected region for CS and ROICS images at chosen acceleration factors were evaluated by calculating the Root Mean Square Error (RMSE) by using the equation [3]

RMSE=√((∑_(i=1)^n??(x_p-x_r)?^2)/N) [3]

where, xp, xr are the CS/ROICS reconstructed and full k-space reconstructed image respectively and N is number of elements in xp or xr.

Results: Superior performance of ROICS can be observed over conventional CS qualitatively at chosen acceleration factors 2x, 5x, 10x, 15x and 20x. It can be observed that the ROICS outperforms conventional CS, as the acceleration increases the noise in the CS image increases where as in ROICS even at 20x it is able to retain information than CS. Reconstruction using 5 percent of the data (20x) shows the utility of our proposed ROICS reconstruction technique and it is quantified by RMSE value. RMSE value for both CS and ROICS technique were calculated at different chosen acceleration factors and the graph shows that ROICS is superior than CS. 5x acceleration onwards there is a significant increase in the RMSE value for the conventional CS technique where as there is a consistency in the RMSE value for ROICS method.

Conclusion: ROICS has been applied on breast DCE MRI data for the first

time and qualitatively and quantitatively ROICS performs better than the conventional CS. CS and ROICS reconstructed images will be used to compare the obtained pharmacokinetic maps from Tofts Model in time domain and in frequency domain (7).

References:

Jackson A, Buckley DL, Parker GJ. Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology. Springer, 2004

Tofts PS. Quantitative MRI of the brain: measuring changes caused by disease. John Wiley, 2003

Lustig, Michael, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. Magnetic resonance in medicine 58.6 (2007): 1182-1195

Amaresha Shridhar Konar, Steen Moeller, Julianna Czum, Barjor Gimi, Sairam Geethanath. Region of interest compressive sensing (ROICS). In Proceedings of the 21st Annual Meeting of ISMRM, Salt Lake City, USA, 2013. p. 3801

Hansen, Per Christian, and Dianne Prost O'Leary. The use of the L-curve in the regularization of discrete ill-posed problems. SIAM Journal on Scientific Computing 14, no. 6 (1993): 1487-1503

https://wiki.cancerimagingarchive.net/display/Public/QIN+Breast+DCE-MRI

Nithin N Vajuvalli, Krupa Nayak, Sairam Geethanath," Accelerated Pharmacokinetic Map Determination for Dynamic Contrast Enhanced MRI using Frequency-Domain Based Tofts Model ", EMBS, IEEE -2014

9399-38, Session PTues

Super Resolution Algorithm for CCTVs

Seiichi Gohshi, Kogakuin Univ. (Japan)

CCTV systems are installed in urban areas worldwide for safety and security purposes. The images and video from these systems function to suppress crime and play an essential part in our lives. The cost of security cameras is falling, and CCTV systems will have even more widespread applications in the near future. However, the images recorded by them do not always have sufficient quality.

Super resolution (SR) is a hot topic in the image and video field. It improves resolution. However, the signal processing it uses is intended for color images taken in daylight. CCTVs do not produce clear images in the poor lighting conditions after sunset; therefore, infrared cameras are used instead. It is not easy for the previously proposed SRs to improve the quality of infrared images.

Considering the purpose of CCTV systems, it is necessary to make a real-time signal processing for their video cameras. However, the previously proposed SR algorithms require iterations, and developing real-time hardware based on these algorithms is not a simple task. A fundamentally new method is necessary to improve the overall image quality of CCTV systems.

In this paper, we propose a novel signal processing method for infrared images. It uses nonlinear signal processing (NLSP) to create higher frequency elements not contained in the original image. It also increases the contrast and image quality. NLSP is a simple algorithm that can work in real time. Using this method, it is easy to develop real-time hardware for video and images. It increases the potential capabilities of CCTV systems operating at night.

9399-39, Session PTues

Intended motion estimation using fuzzy Kalman filtering for UAV image stabilization with large drifting

Tiantian Xin, Hongying Zhao, Sijie Liu, Lu Wang, Peking Univ. (China)

Videos from a small Unmanned Aerial Vehicle (UAV) are always unstable because of the wobble of the vehicle and the impacts of surroundings, leading to vertigo, visual fatigue and even misjudging. Electronic image stabilization aims at removing the unwanted wobble and obtaining the stable video. However, it is always difficult to obtain stable real-time intended motion to get close to real flying path and remain as much information of video images, especially with large drifting because of motion compensation. This paper presents an effective method to estimate the intended motion for UAV image stabilization with large drifting using fuzzy Kalman filtering.

Firstly, we analyze Kalman filtering process of global motion and the impacts of observation error covariance matrix R on the filtering results. In the filtering process the value of R is always changing and difficult to determine its precise value. It is observed that the value of R, which stands for the weight of current observation, can largely influence the performance of Kalman filter and the optimal intended motion estimation.

Then we choose fuzzy Kalman filtering to adjust the value of R adaptively to improve filtering effects. Fuzzy control can avoid precise mathematical modeling and build fuzzy systems to process those which cannot be abstracted into precise equations. It uses membership and fuzzy rules to control filtering process.

At last actual UAV data with large drifting is experimented with fuzzy Kalman filtering method proposed in this paper. The correction of R is the output of filter to adjust the value of R adaptively corresponding to different circumstances. We obtain not only smooth intended motion closer to real flying path but also stable UAV videos remaining much more information. Results show that fuzzy Kalman filtering method in this paper can effectively solve the intended motion estimation problems for UAV videos with large drifting.

9399-40, Session PTues

On-line rock fragment delineation on simple thresholding using fuzzy comprehensive evaluation

Zhongpu Jia, Henan Polytechnic Univ. (China)

The distributions of rock fragment size and shape are important indicators to evaluate the production quality. The traditional measurement is by manual and sieving, which is a time-consuming, limited sampling, inefficiency and hard laboring method. To overcome the disadvantages, the image processing technique has been used for the measurement and analysis of rock fragments on-line since 30 years ago. As a general on-line monitoring system, the image acquisition sub-system is setup with a CCD camera either over a moving conveyor belt or at the end of a conveyor belt. In the former situation, the fragments overlapping each other make image segmentation hard, and in the later situation, the fragments are in a gravitational falling stream and easily delineated. In the falling stream, the sky can be used as the image background, and the most fragments are separately falling in air, a simple thresholding algorithm is possibly used for the fragment detection in real time, but the main problem is to split the touching fragment clusters.

Anyhow, in order to meet the on-line requirement of the rock fragment measurement and analysis, a thresholding algorithm is the better choice for image segmentation, because it is a widely used image segmentation algorithm which divides an image into two classes of targets and background. Normally, a global thresholding algorithm has the good effect on the image in which the gray level difference between targets and background is obvious, but for the falling rock fragment images, due to the factors of the weather variation, dust, fragment touching, shadows and other noise, it is hard to do image segmentation completely. In order to overcome this drawback, this paper proposes an adaptive threshold algorithm based on fuzzy comprehensive evaluation. In the algorithm, it firstly segments a gray level image by using a global thresholding algorithm which is evaluated and selected based on five widely used thresholding algorithms, and then it runs the thresholding algorithm on the touching fragment regions separately, which meets the criterion or has abnormal size, shape, gray level difference and gradient magnitude, and it repeats the operations until no fragment touching object can be separated further.

The classification and measurement of rock fragments is very important in Mining and Construction Engineering. The quality evaluation can be achieved by monitoring fragments on-line. The monitoring system acquires and analyses the fragment images from a gravitational falling stream at the end of a moving conveyor belt, and the key function of the system is to construct an image segmentation algorithm which can work accurately in real time. To reach to this goal, an adaptive thresholding algorithm with Fuzzy Comprehensive Evaluation is proposed. Firstly a grabbed image is roughly segmented by using a global auto-thresholding algorithm. Then each of the objects is measured and analyzed if it includes the multiple fragments touching each other, based on the fuzzy comprehensive evaluation method in which the salient fragment features of area, perimeter, shape, gradient magnitude and gray-level flatness are extracted, and for each of the features, the membership function is constructed experimentally. Finally, each of the touching fragment regions, as one image, is auto-thresholded again, and this procedure is repeated until no region can be further separated. The experimental results show that comparing to Cluster analysis, Graph based, FCM image segmentation algorithms the new algorithm can make image segmentation well for the falling fragments on-line.

9399-41, Session PTues

A perceptual quality metric for high-definition stereoscopic 3D video

Federica Battisti, Marco Carli, Alessio Stramacci, Univ. degli Studi di Roma Tre (Italy); Atanas Boev, Atanas P Gotchev, Tampere Univ. of Technology (Finland)

The use of 3D video is increasing in many fields, such as entertainment, military simulations, medical application, etc., consequently also recording, transmission, and processing of 3D video is getting more common thus creating artifacts that can affect the perceived quality.

A challenging task is to find a metric able to predict the perceived quality with a low computational complexity so as to be used in real-time applications. To this aim, several metrics have been proposed for 2D videos, whereas for 3D videos research is still ongoing due to more complex influence of the stereoscopic cues. In our approach we believe that a combination of existing metrics can result in a new metric able to predict the MOS more accurately.

The proposed procedure can be described as follows:

1. The definition of three models to take into account different characteristics of a 3D transmission:

• Cyclopean view (CV): is given by the overlapping between left and right view.

Binocularrivalry(BR):occurs when the eyes try to focus on a single point in a scene as a result of two slightly different views. Even though occlusions are a natural source of artifacts, the major contribute to BR is given by the distorted views only. Reference view is not taken into account for this model.

• Binocular depth (DQ): takes into account the amount of depth in different stereoscopic videos.

2. Test of a set of state of the art metrics or candidates to each model. In particular, Mean squared error (MSE), Gradient Normalized Sum of Squared Difference (SSD), Peak Signal-to-Noise Ratio HVS (PSNR-HVS), PSNR-HVS without masking effect (reduction of masking and contrast operation), PSNR-HVS-M with masking effect, Fea- ture Similarity Index (FSIM), FSIM2, FSIM3, Structural Similarity Index (SSIM), SSIM applied on the luminance component, and SSIM applied to the chrominance components.

3. A subjective test has been performed to collect the MOS for validation purposes. The test set has been selected by considering high bit-rate, high quality and stable shooting condition. The Nantes-Madrid-3D-Stereoscopic-V1 (NAMA3DS1),1 database has been selected. The database provides high 3D visual quality sequences, included uncompressed and encoded videos with bit-rate of 25fps. The test set is composed by 10 original sequences. Each sequence has been affected by 9 different artifacts (blocking, down-sampling, edge enhancement, and combinations of these artifacts). The subject was asked to rank the perceived quality of the video in a range 1 (worst quality) to 5 (best quality).

4. Metric combination selection. The adopted procedure for defining the best metric combination is inspired by2 . In more details: to assess the effectiveness of a metric fj , the Spearman correlation between the MOS and the metric under test is computed. In table 1 the correlation with respect to the MOS is reported. As can be noticed the best result is given from f25 with 0.81. The last row of the table shows the correlation provided by the combination of all the metrics for each single model. In this case we notice that CV is the strongest model (it includes the first three columns) with a maximum correlation value of 0.91. However it includes just one model so that it can not provide general results.

The features combination is obtained through a linear regression. This operation needs as input the features we want to combine and the reference vector (in our case the MOS). The goal is to find the parameters that are able to minimize the error between our selected features and the MOS. To the aim of selecting the minimum number of features needed to design a new metric, the sequential feature correlation , has been applied.

An optimal combination of features has been found which results in a correlation value of 0.91 with MOS.

9399-42, Session PTues

Content-aware video quality assessment: predicting human perception of quality using peak signal to noise ratio and spatial/temporal activity

Benhur Ortiz-Jaramillo, Jorge Oswaldo Niño-Castaneda, Ljiljana Plati?a, Wilfried Philips, Univ. Gent (Belgium)

Video-based systems such as digital television, surveillance and tele-medicine, are becoming more common in everyday life. Quality control of such systems is very important for increasing the user satisfaction. Since the end-user is often a human observer, quality control mechanisms have to include measures which mimic the HPoQ. Particularly, quality assessment of videos has an important role in evaluating and improving the performance of video-based systems. Overall, quality assessment of videos can be grouped in two main categories: subjective and objective methods.

On the one hand, subjective assessment is performed by a group of human subjects who evaluate (processed or corrupted) videos according to certain well-defined criteria such as those defined in the related ITU standards. Often the result of such assessments is a Difference Mean Opinion Score (DMOS) per assessed video. When a sufficiently large group of human subjects is available, this methodology represents the most accurate technique of measuring HPoQ in video based systems. However, such techniques are in general complex, expensive, and time consuming. Therefore, they are unpractical for real time video processing and hard to incorporate into a system design process.

On the other hand, objective assessment tries to estimate HPoQ by using quantitative operations based on mathematical models and computer algorithms. Objective approaches have proven to be a desirable alternative for measuring video quality because they can be cheap, time efficient and easy to incorporate into video-based systems. Currently, there exists a large variety of objective VQMs which ranges from very simple models like PSNR, to more complex ones, involving artifact measures such as blockiness, blurring, and noise, among others. Recent developments in VQM are based on modeling the properties of the human visual system, e.g., modeling the contrast sensitivity, spatial/temporal masking effects, and color perception.

In general, current VQMs are either too computationally complex and/or not generic enough for a wide variety of video contents. It is well known that the latter problem is mainly due to the strong dependency of HPoQ on video content. Although such dependency is well known, only few existing quality metrics directly account the influence of content on HPoQ. Hence, most of the current VQM used for estimating HPoQ are very limited in performance.

Our aim is to advance the existing VQM by including video content features. In this paper we propose a methodology which involves off-line training of the parameters of nonlinear functions and its relationship with video content features. Video content in this work means spatial and temporal activity in the video sequence, i.e., the extent of details and motion. The proposed methodology is based on observations made from the change of different VQM in function of source content. Without loss of generality, in this work we use PSNR and gradient statistics with the purpose of illustrating the potential of the methodology.

We map PSNR to DMOS by using a nonlinear function which depends on parameters determined by linear combinations of spatial and temporal activities. The spatial and temporal activities are estimated by computing statistics on spatial and temporal gradients as well as spatial dependencies of pixel values. The coefficients of the linear combinations are estimated off-line with the purpose of controlling two parameters: the halfway point and the dispersion of PSNR values over that point. Thus, assuming DMOS values in the range [0, 1], the halfway point is taken as the PSNR value when DMOS is 0.5 for a given source video content. The dispersion parameter indicates how fast the quality drops as PSNR decreases, i.e., how big is the rate of change of DMOS with respect to PSNR for certain source video content.

In summary, this work investigates the influence of content on video quality assessment. By including content features, we develop a computationally simple but effective methodology which is able to perform well under multiple types of content. Importantly, the proposed methodology is generic enough to be used with different distortion metrics and/or content features. Experimental results over four different public video quality databases show that, even by using very simple metrics and content features, the proposed methodology is competitive with state-of-the-art metrics.

9399-43, Session PTues

Multi-volume mapping and tracking for real-time RGB-D sensing

Lingni Ma, Egor Bondarev, Peter H. N. de With, Technische Univ. Eindhoven (Netherlands)

A detailed 3D reconstruction of arbitrary scenes is of a great interest to many applications, including robotics, virtual/augmented reality, 3D printing, civil engineering, etc. In recent years, the low-cost RGBD cameras boost the research in RGBD mapping and tracking, which is widely used for 3D reconstruction. KinectFusion, the first real-time mapping and tracking algorithm for handheld RGBD sensor was introduced in 2011. The algorithm fuses depth images with truncated signed distance function (TSDF) into a volume for scanned scenes and uses the iterative closest point (ICP) algorithm for tracking. Based on KinectFusion, many succeeding systems are derived to improve the mapping and tracking performance. The following techniques are reported on the state-of-art systems: the rolling buffer technique for unlimited spatial modeling, the volumetric color fusion or texture mapping for colored maps, the RGBD odometry combined with ICP for robust tracking and SLAM techniques with trajectory correction after loop closure.

Despite this progress, 3D models obtained from the current systems are not of comparable visual quality as original RGBD color images and there is not much research reported for improvement. Due to the powerful parallelism of current GPUs, most systems perform volumetric data fusion on a GPU. A volume is a 3D voxel array, where each voxel corresponds to some spatial area. A voxel usually takes 32bits to store TSDF and color value. In order to cover enough spatial range for tracking and that the given memory usage grows cubically with voxel resolution, the commonly used volume is a 3m cube divided by 512x512x512 voxels. The individual 5.86mm cubic voxel is too large to keep full visual details in color images, as well as the fine geometrical details in depth images. To solve this problem, we propose the following contributions: 1) multi-resolution volumetric fusion on GPU for RGBD mapping; and 2) an optimal volume positioning and shifting strategy to reconstruct unbounded space. The experimental results reveal that the proposed system provides highly-detailed color maps and accurate tracking data. The system works in real time (30Hz), while maintaining almost the same requirement.

To solve the dilemma between a voxel dimension and a global volume dimension given limited GPU memory, we develop a solution using multi-resolution volume processing. The principal idea is to use small voxels at near field to obtain detailed 3D maps, meanwhile using large voxels at far field to cover sufficient range for tracking. To implement this principal, we deploy two cubic volumes on GPU: one small volume of 1m dimension with 512x512x512 voxel resolution and one large volume of 3m dimension with 256x256x256 voxel resolution. We refer the smaller one as the RGB-TSDF volume, and fuse both depth and color images into it. We refer the bigger one as the Helper volume, and fuse only depth images into it. While the RGB-TSDF volume provides detailed 3D maps, the Helper volume assists in tracking.

To make best use of two volumes, their positioning is optimized. The RGB-TSDF volume is positioned closer to the sensor for better visual capturing and fusion and the Helper volume is further from the sensor to cover wider range. The sensor is placed outside of the volumes, due to the fact that RGBD cameras normally do not provide data within very short distance. We define a viewing range parameter as the distance from the camera to the volume center along the camera view direction. When the camera moves through the space, both volumes are virtual shifted to include the appearing vision field. Upon volume shifting, we always reposition the volume to maintain an optimal viewing range, which is defined as half of the volume dimension plus the sensor occlusion distance.

For quantitative evaluation, we perform two sets of experiments on the TUM RGBD datasets. In the first experiment, we assess the mapping quality with respect to visual details. For this purpose, we render images of the 3D models from the same viewpoints as the camera. Taking the original color images as references, the peak signal to noise ratio (PSNR) and the structure similarity (SSIM) of rendered images are computed and compared among maps fused in volumes of 3m, 2m and 1m dimension. The average PSNR of the 3m, 2m and 1m volume is 20.439dB, 21.428dB and 22.448dB, respectively. The average SSIM of the 3m, 2m and 1m volume is 0.798, 0.823 and 0.865, respectively. The results show that the use of 1m volume greatly improves the coloring and geometry modeling quality.

In the second experiment, we evaluate the tracking performance by measuring the relative pose error (RPE) frame to frame with respect to the ground truth. The tracking error of results obtained with a single 3m volume is compared to the tracking error from our proposed 1m RGB-TSDF volume plus a 3m Helper volume. On average, the RPE of both scenario is about 0.002212m in translation and 0.001578deg in rotation. The results show that our method gives comparable tracking quality. In addition, we observe that the average frame processing time is under 30ms, despite fusion of two volumes. This is suitable for real-time processing of RGBD sensor at 30Hz frame rate. With respect to GPU memory, the contributing method requires 64MB extra space to store the Helper volume. This amount is marginal to the 1GB memory required by the 512x512x512 voxel resolution to fuse depth and color.

In conclusion, we have presented our multi-resolution volume fusion algorithm for RGBD mapping and tracking system. The experimental results show that our algorithm provides much detailed color maps without sacrificing the tracking performance, the processing frame rate and the memory requirement.

9399-44, Session PTues

Preserving natural lighting by strobe-lit video

Olli J. Suominen, Atanas P. Gotchev, Tampere Univ. of Technology (Finland)

Capturing images in low intensity, ambient lighting conditions poses significant problems in terms of achievable image quality. Either the sensitivity of the sensor must be high, filling the resulting image with noise, or the scene must be lit with artificial light, destroying the aesthetic quality of the image. While the issue has been previously tackled for still imagery using cross-bilateral filtering, the same problem exists in capturing video. We propose a method of illuminating the scene with a strobe light synchronized to every other frame captured by the camera, and merging

the information from consecutive frames alternating between high gain and high intensity lighting. The movement between the frames is compensated using block matching based motion estimation.

The uniform lighting conditions between every other frame make it possible to utilize conventional motion estimation methods, circumventing the image registration challenges faced in fusing flash/non-flash pairs from non-stationary images. Motion estimation is dependent on the texture of the image, but the subsequent motion vector based reconstruction is only used as a guide for edge-aware filtering. Therefore important areas (i.e.

edges and detail) are likely matched correctly, while areas with low texture have no impact on the filtering. The results of the proposed method are shown to closely resemble those computed using the same filter based on reference images captured at perfect camera alignment. The method can be applied starting from a simple set of three frames to video streams of arbitrary lengths with the only requirements being sufficiently accurate syncing between the imaging device and the lighting unit, and the capability to switch states (sensor gain high/low, illumination on/off) fast enough.

9399-23, Session 4

On detailed 3D reconstruction of large indoor environments *(Invited Paper)*

Egor Y Bondarev, Technische Univ. Eindhoven (Netherlands)

The advances in development of depth-based imaging sensors enable 3D digitalization of surrounding environments, where the common flat 2D representation is enriched with a geometry data of a captured scene. The resulting 3D textured models might represent captured reality in a detailed, accurate and photorealistic manner. Therefore, the 3D reconstruction technology has recently pulled an attention from various industrial domains, such as robotics, healthcare, surveillance, construction and military. However, the sensor- and algorithmic- maturity level of the state-of-the-art reconstruction technologies is not yet high enough to provide detailed, accurate and photorealistic 3D models. This problem especially renders for reconstruction of large-scale and arbitrarily-shaped indoor environments. The reasons for this phenomenon are portrayed in the following paragraphs.

An accurate reconstruction of textures and 3D structures of a large and geometrically complex scene poses two major requirements: sufficient geometry- and texture- detalization of individual objects and accuracy of object positioning in global dimensions. The state-of-the-art sensing and reconstruction approaches can be classified into the three following types. The first approach is based on a lengthy and continuous use of a close-range depth-sensor (e.g. Kinect or stereo camera) [1,2], where the second one yields combination of multiple individual scans form a far-range depth-sensor (e.g. LIDAR), and the third hybrid approach deploys multiple types of depth sensors. All approaches require multi-modal fusion of the obtained depth and texture data into a single textured 3D model.

Let us now outline the pro's and con's of these approaches with respect to the two main requirements. In the first approach, the sufficient detalization of individual objects is granted by low-range intrinsics and sensor path flexibility. However, the problem of severe localization drifts appears during lengthy capturing sequences. For instance, capturing a 10.000 m3 scene by a structured-light sensor would require at least 1 hour sensing time (100K depth images), resulting in a 1-5 meters sensor-localization drift. This significantly hinders the second requirement on intra-object positioning and global dimension consistency. The second approach does not suffer from intra-object and global positioning inconsistency, since it deploys far-range sensor scans where point clouds can be accurately aligned to a reference point cloud. However, tackling the first requirement on sufficient detalization of objects is extremely challenging for scenes with complex geometry, since this would require thousands of far-range scans from various positions and viewing angles. Taking into account that one laser scan returns millions of points, processing the total point set would require inadequate amount of hardware resources and time. The third approach based on the hybrid deployment of close- and far-range sensors helps addressing both requirements on detalization and positioning, while introduces two challenges on fusion of the data obtained from multiple sensors of intrinsically different types. These challenges include (a) the issue of synchronization and autonomous localization of different sensors in temporal and spatial domains, and (b) the problem of integration of the data captured from low-range sensors (e.g. depth map of TSDF grid [3]) into the data obtained by far-range sensors (e.g. point cloud).

To address the two vital requirements, we are experimenting with all three approach types. In this paper, we mainly focus on our algorithmic research on the low-range and hybrid approach types, where the last one yields the promising idea of combining the best of both sensor worlds.

The high detalization and accuracy of individual geometric shapes in the scene is addressed by a cluster of the following proposed methods. For the low-range sensor data, we introduce the real-time KinfuLargeScale algorithm able to reconstruct large environments [3]. It was recently extended with the following algorithmic blocks in the pipeline. To filter the depth noise, we propose a weight-based depth fusion algorithm, where each depth point at every frame is integrated into a global 3D model based on its distance to the sensor, distance to the line-of-sight and the angle between its normal and the line-of-sight [4]. For localization accuracy improvement, we propose an approach where localization results from multiple simultaneously running tracking algorithms (e.g. ICP [5] and RGB-D odometry) are weighted and combined based on the amount of specific features found for each tracking algorithm in an individual frame. For instance, should the captured scene contain low amount of texture data and sufficient amount of geometry data, the ICP localization output is weighted for the final localization computation higher than the RGB-D output, and vise versa. To achieve very dense 3D resolution (125 points/cm) and still keep the scanning range and localization accuracy, we have introduced a double voxel-grid. The primary high-resolution voxel grid is of a small size (1-2 m3) enabling refined 3D reconstruction at 125 points/cm density, where the secondary low-resolution voxel-grid is large (27-64 m3) and used for tracking and localization. To obtain photorealistic textures and then accurately map them on the corresponding 3D surfaces, we have proposed an advanced approach that accumulates color values for each valid voxel with a specific weighting strategies: (a) valid color-points allocated on edges of hypothetical surfaces are integrated with lower weight than the color-points allocated far from such edges, and (b) valid color-points featuring large angle between its normal and the line-of-sight are underweighted to the color-points with normals pointing to the sensor. The accumulated colors of the points are then interpolated, thereby generating a textel for each face of a resulting mesh [6]. Due to the introduced high resolution of the voxel grid and the weighting strategies, most of the textures generated in our experiments can be perceived as obtained from a photo camera.

Merging these local reconstruction methods with the far-range 3D data would bring local detalization and global consistency. In this paper we introduce our experimental results on algorithms for localization of low-range sensor in the global reference model from the far-range sensor and on fusion of colored dense voxel-grid data with colored sparse point-cloud data. Moreover, we present a unique 3D data decimation approach with a consequent octree-based rendering engine, enabling real-time visualization of massive (up to 500mln points) point clouds in contemporary online browsers.

9399-24, Session 4

Person re-identification by pose priors

Slawomir Bak, Filipe Martins de Melo, Francois Bremond, INRIA Sophia Antipolis - Méditerranée (France)

1. Introduction:

Person re-identification is a well known problem in computer vision community.

This task requires finding a target appearance in a network of cameras with non-overlapping fields of view.

The changes in person appearance together with inter-camera variations in lighting conditions, different color responses, different camera viewpoints and different camera parameters

make the appearance matching task extremely difficult. Current state of the art approaches very often concentrate on metric learning that

use training data to search for matching strategies minimizing the appearance changes (intra-class variations), while highlighting distinctive properties

(maximizing inter-class variation) of the target. Although, usually metric learning boosts the recognition accuracy, it requires large training data for each camera pair in a camera network.

This already can make these approaches inapplicable in small camera networks where acquisition of labeled data is unattainable.

We claim that using metric learning for handling all difficulties related to appearance changes is not not effective and unscalable.

In contrary, in this paper we divide the person re-identification problem into several sub-problems:

(1) registration (alignment);

(2) pose estimation;

(3) pose matching and

(4) invariant image representation.

We focus on the third problem, ie. pose matching strategies, employing metric learning. We learn several metrics for handling different pose changes, generating pool of metrics.

Once metric learning pool for matching different poses has been learned, we can apply it to any pair of camera. This makes our approach scalable to large camera networks.

2. The method:

We divide the person re-identification problem into four sub-tasks:

(1) appearance registration,

(2) pose estimation,

(3) pose matching and

(4) camera invariant image representation.

In the following sections we discuss issues and approaches related to these tasks, while focusing on the pose matching problem.

2.1. Appearance registration:

The changing viewpoint in a network of cameras is an important issue and might significantly distort the visual appearance of a person. This problem has a direct impact on the re-identification accuracy.

Eliminating perspective distortions in an image region containing the target is often called image rectification. We employ two state of the art techniques [1,2] and provide detailed evaluation.

Finally we select one that brings better performance (full version of the paper will contain elaboration of both techniques, discussing pros and cons).

2.2. Pose estimation:

Person pose is estimated using 3D scene information and motion of the target [1]. Employing this simple but effective method, we obtain sufficiently accurate pose information,

which is used for selecting a metric from a pool of metrics. The metric pool is learned based on pairs of images with particular poses (see section 2.3).

2.3. Pose matching:

In this section we learn the matching strategy of appearance extracted from different poses. We employ well known metric learning tools for matching given poses.

Let us assume that pose can be described by the angle between the motion vector of the target and the viewpoint vector of the camera (see details in [1], full paper will contain detailed figures).

Thus for each target appearance we can express the pose as the angle in the range of [0,360). We decide to to divide this range into n bins.

Given n bins of estimated poses, we learn how to match different poses corresponding to different bins. In the result, we learn n*(n+1)/2 metrics.

While learning metrics, we follow a well known scheme based on image pairs, containing two different poses of the same target (details on generating training data will be provided in the full version of the paper).

The learned metrics stand for the metric pool. This metric pool is learned offline and is not dependent on camera pair. In the result, once metric pool is learned, it can be used for any camera pair.

Given two images from different (or the same) camera, we first estimate the poses for each image. Having two poses, we select a corresponding metric from the metric pool.

The selected metric provides the strategy to compute similarity between two images.

2.4. Camera invariant image representation:

In this section we provide an overview of different descriptors that are usually employed for the re-identification task. In general, image is divided into an image patches (regions)

that are represented by image descriptors (e.g. hog, covariance, brownian, lbp, color histograms, etc). We discuss advantages and disadvantages of given descriptors, while providing their analysis.

Given descriptors are evaluated in section 3.

3. Experimental results:

This sections evaluates the re-identification performance on SAIVTSOFTBIO dataset that consists of 8 cameras and 152 subjects.

We provide detailed results w.r.t. each sub-task: (1) appearance registration, (2) pose estimation, (3) pose matching and (4) invariant image representation.

We illustrate the improvements at each step of the processing chain, discussing its importance. Finally, we show the impact of metic pool.

Our approach shows that dividing the re-identification problem into sub-tasks gives better understanding of this complex problem and allows to obtain increase in the recognition accuracy.

[1] Improving Person Re-identification by Viewpoint Cues, Bak et al, AVSS 2014.

[2] Human detection by searching in 3d space using camera and scene knowledge, Y. Li et al, ICPR 2008.

9399-25, Session 4

Fast planar segmentation of depth images

Hani Javan Hemmat, Arash Pourtaherian, Egor Bondarev, Peter H. N. de With, Technische Univ. Eindhoven (Netherlands)

1. INTRODUCTION

The advent of low-cost real-time depth sensors has attracted multiple studies in 3D perception and reconstruction for a wide range of applications, including robotics, health-care, and surveillance. One of the major challenges in such applications is the real-time execution of the algorithms. For indoor environments, perceiving the geometry of surrounding structures is very important. On the average, up to 95% of indoor-environment structures consist of planar surfaces [1]. Therefore, fast and accurate detection of such features has a crucial impact on quality and functionality of the 3D applications, i.e. decreasing model size (decimation), as well as enhancing localization, mapping, and semantic reconstruction. The available planar-segmentation algorithms based on surface normals and/or curvatures extraction [2, 3], linear fitting and Markov chain Monte Carlo [4], 3D Hough transform [5] and hybrid region growing [6] are computationally expensive and challenging for real-time performance. In these algorithms, a real-time segmentation of planes is normally achieved by sacrificing the image quality [7]. In this paper, we propose a fast planar segmentation algorithm solely based on identification of intersecting lines on a plane without processing any surface normal. The proposed method is capable of handling VGA-quality depth images at a high frame rate, upto 6 FPS, with a single-thread development.

2. SEGMENTATION

The proposed algorithm segments planar surfaces in a depth image by utilizing two intrinsic plane properties: (1) each planar surface is bounded between 3D edges in the depth image, and (2) each two intersecting lines

represent a planar surface. Based on these intrinsics, segmentation of the planar surfaces is performed as followings. The first phase deter- mines all 3D edges in the depth image. In the next phase, we search for lines between points on opposite edges. The final phase tries to merge the identified lines into a planar surface.

2.1. Edge detection

In order to detect 3D edges, the algorithm scans the depth image line-by-line in four directions (vertically, horizontally, left- and right-diagonally). An edge is detected in aspect of depth-value change. At each scan, 3 types of edges are detected: jump, corner, and curved edges. We formulate the detection criterion for each edge type:

Jump edge. The 3D points Pi and Pj on current line of the scan direction are on the opposite sides of a jump-edge, if the absolute difference of their depth values is greater than a pre-defined threshold.

Corner edge. The 3D point Pi on current line of the scan direction is a corner edge, if points before and after Pi (1) lay on a line, and (2) their line-slopes difference is greater than a pre-defined threshold.

Curved edge. The 3D point Pi on current line of the scan direction is a curved edge, if points on one side of Pi are laying on a line and points on the other side are not.

2.2. Line-based plane detection

At the second phase, to segment the planes, the algorithm tries to merge the points on each pair of intersecting detected-lines into a plane according to the second above-described intrinsic. It continues merging until it covers all the points on each line. The resulting planar segments need improvements (Section 2.3).

2.3. Plane Enhancement

We process the detected planes further in order to improve the segmentation outcome. The enhancement is performed in three stages:

Curvature validation: A valid plane should have a curvature less than the desired threshold. The plane candidates are converted to corresponding pointclouds and their eigenvalues are calculated by the Principal Component Analysis (PCA). A plane is considered valid if its curvature values in the direction of eigenvectors v0 and v1 are less than thresholds T0 and T1, respectively.

Merging separate-segments: Due to holes and occlusions in a depth image, different segments of a plane may be identified as separate planes. In order to merge the separate segments, we check two conditions for each pair of plane segments. First condition is true when the segments are parallel and second condition is true when they lay on the same planar surface.

Size validation: A valid plane is required to contain a certain minimum number of points. This improves segmentation quality by removing the small segments that could not be merged to any other larger planes.

3. RESULTS

We applied the proposed segmentation algorithm on several datasets, including both real and artificial VGA-quality depth images of indoor scenes. Our algorithm execution time is compared to various methods from the PCL library. The execution time of the presented method is compared to RANdom SAmple Consensus (RANSAC), Least MEDian of Squares (LMEDS) and M-Estimator SAmple Consensus (MSAC). The proposed algorithm features the smallest execution time for all datasets (mean: 0.23 s). An interesting finding is the large amount of standard deviation to mean ratio for all other methods. Moreover, since the presented method is dealing with depth images, it is less dependent on the image content and is faster in terms of execution performance.

4. CONCLUSION

In this paper, we have introduced a fast planar-segmentation method for depth images avoiding any normal estimation. The proposed algorithm searches for 3D edges in the depth image and finds the lines between these edges. Then, it merges all the points on each pair of intersecting lines into a plane. Finally, various enhancement stages are applied to improve segmentation quality. Furthermore, due to the multi-threaded design of the proposed algorithm, we expect to achieve a factor of 10 speedup by deploying a GPU version.

REFERENCES

[1] R. B. Rusu, PhD thesis, Technische Universitaet Muenchen, Germany (2009).

[2] R. B. Rusu et al., in [ICRA], 3212–3217 (2009).

[3] B. Oehler et al., in [ICIRA], 145–156 (2011).

[4] C. Erdogan et al., in [CRV], 32–39, (2012).

[5] R. Hulk et al., Journal of Visual Communication and Image Representation 25(1), 86–97 (2013).

[6] J. Xiao et al., Robotics and Autonomous Systems 61(12), 1641–1652 (2013).

[7] D. Holz et al., in [Proc. of RoboCup 2011], 306–317, (2012).

9399-26, Session 4

Machine vision image quality measurement in cardiac x-ray imaging

Stephen M. Kengyelics, Amber J. Gislason-Lee, Univ. of Leeds (United Kingdom); Claire Keeble, UNIVERSITY OF LEEDS (United Kingdom); Derek R. Magee, Andrew G. Davies, Univ. of Leeds (United Kingdom)

Background

Percutaneous coronary intervention (PCI) is an effective, minimally invasive, treatment for cardiovascular disease resulting from a narrowing of coronary arteries that works by mechanically improving the flow of blood to the heart. The critical dependence on x-ray imaging to visualize coronary arteries, made opaque by use of iodine contrast agents, during PCI procedures requires the use of ionizing radiation and its concomitant hazard to both patients and staff. Lowering radiation dose is advantageous, but any arbitrary reduction may result in deterioration of diagnostic visual information that may compromise patient care. Currently most cardiac x-ray imaging systems used during PCI procedures regulate their radiation output by adjusting a number of system parameters to seek to maintain a constant average output signal from the x-ray detector. This automatic dose control (ADC) scheme is somewhat limited in that it principally only accounts for the x-ray attenuation properties of the patient and not the requirements of the clinical imaging task.

Aims

The purpose of this work is to report on a machine vision approach to the automated measurement of x-ray image contrast of coronary arteries filled with iodine contrast media during interventional cardiac procedures. The aim is to investigate a means of providing real-time context sensitive information to augment the ADC of the x-ray imaging system such that the quality of the displayed image is guided by clinically relevant image information.

Methods

A machine vision algorithm was developed in MATLAB (Mathworks, Natick MA). The algorithm creates a binary mask of the principal vessels of the coronary artery tree by thresholding a standard deviation map of the direction image of the cardiac scene derived using a Frangi filter. Vessel centerlines are determined using skeletonization of the mask and the average contrast of the vessel is calculated by fitting a parametric Gaussian profile at a number of points along and orthogonal to the vessel over the extent of the width of the binary mask and averaging the results. Profile data are normalized to their maximum value and inverted by subtracting this value from unity and thus the derived contrast values may be considered as local to the vessel. Profiles resulting in minimum mean square errors of less than 0.9 are rejected from the analysis. The average width of the vessel is determined by measuring the width of the binary mask orthogonal to the centreline at points along the vessel coincident with the contrast measurements and the results averaged.

The algorithm was applied to sections of single frames from 30 left and 30 right coronary artery image sequences from different patients. The images were unprocessed and were captured using a custom data-capture device fitted to an ALLURA FD10 cardiac imaging system (Philips Healthcare) situated in the Leeds General Infirmary, United Kingdom.

For validation the algorithm was constrained to operate on small identifiable sections of the iodine filled vessel. The average contrast and width of the vessel was calculated and recorded for a total of 20 profiles. Five manual measurements of the contrast and width of the vessel were also performed over the region of interest using a standard distance measurement and profile tool. The five results were then averaged. A Bland-Altman assessment for agreement was used to compare the two methods for the measurement of contrast and vessel width. A range of agreement was defined as the mean bias ± 2 ? standard deviation of the differences between the two methods.

Results

For the 60 vessels the range of contrast was 0.26-0.74 and the range of vessel widths was 1.07-5.4 mm considering results from both methods. The maximum standard deviation for single averaged measurements from the algorithm using 20 samples was 0.05 for average contrast and 0.3 mm for average width. For the manual measurements the maximum standard deviation for single averaged measurements from the manual method using 5 samples was 0.05 for average contrast and 0.5 mm for average width. The Bland-Altman analysis indicates that the 95% limits of agreement between the two methods ranged from -0.05 to +0.04 with a mean bias of 0.004 for the measurement of contrast and -1.7 to -0.13 mm with a mean bias of -0.9 mm for the measurement of vessel width, with the auto-method measuring lower than the manual method.

Conclusions

The machine vision algorithm produced average contrast measurements with a narrow range of 95% limits of agreement of -0.05 to +0.04 compared with manual measurements. This is considered to be acceptable for the purposes of automatic measurement of contrast in the context of image quality measurement in cardiac imaging. For the average vessel width measurements the algorithm had a mean bias of -0.9 mm and a range of -0.7 mm to -0.13 mm. The bias is thought to be associated with morphological operations used to clean the binary mask image prior to analysis which results in a thinning of its outline. The machine vision algorithm has the potential of providing real-time context sensitive information to guide the image quality of cardiac x-ray imaging systems.

9399-27, Session 4

Multiview image sequence enhancement

Ljubomir Jovanov, Hiêp Q. Luong, Tijana Ruzic, Wilfried Philips, Univ. Gent (Belgium)

In order to completely describe the scene, it would be necessary to capture the light distribution function in each point of the scene, for each azimuth and elevation. Based on such a 5D representation of the scene it would be possible to select a subset of viewpoints corresponding to the viewing positions of the multiview display. At the current stage of development it is possible only to sample such a function in the case of static scene. In order to capture a dynamic scene choice is limited to multi camera systems, since the range of 3D video capture sensors (e.g. time-of-flight depth cameras) is not sufficient to cover the whole scene.

In this paper we describe our research on preprocessing techniques for multiview video sequences. Due to the different characteristics/settings of the cameras in the multivew setup and varying photometric characteristics of the objects in the scene, the same object may have different appearance in the sequences acquired by the different cameras in the setup. Images representing views recorded using different cameras in practice have different local noise, color and sharpness characteristics.

In order to provide a viewer with a complete immersion, mutliview displays must be able to show a large number of views and to provide seamless transition between them. Since it is physically impossible to place such a large number of cameras sufficiently close to each other, views have to be synthesized based on the images from the existing cameras. The use of such images for view synthesis without a proper pre-processing creates disturbing artifacts in synthetized views, which manifest in color, noise and blur variations.

View synthesis algorithms introduce artifacts due to errors in disparity estimation/bad occlusion handling or due to erroneous warping function estimation. If the input multivew images are not of sufficient quality and have mismatching color and sharpness characteristics, these artifacts may become even more disturbing. The main goal of our method is to simultaneously perform multiview image sequence denoising, color matching and the improvement of sharpness in the slightly blurred regions. By performing this preprocessing step the amount of artifacts in the interpolated views can be significantly reduced.

In the proposed approach we first perform computationally efficient pilot denoising step, in order to achieve more reliable estimates of temporal and multiview correspondences. After finding these correspondences we estimate noise-free frame by employing best matching segments from previous frames and optimizing objective visual quality measures. After denoising step we detect de-focused textured regions in all camera views and transfer the edges and the texture from the views in which they appear sharp. The improvement of the blurred regions is especially important for microstereopsis cameras where these artifacts often occur due to the characteristics of the optical system of such camera. Finally we perform local color correction by mapping colors of all cameras into the desired color palette using the previously estimated correspondences.

In order to evaluate the visual quality we apply the proposed method on a number of publicly available sequences and sequences recorded by ourselves. We evaluate the visual quality of the proposed method by employing the stereoscopic display coupled with the head tracker in order to emulate the behavior of a multiview display. The results of the study confirm that the proposed method significantly reduces the amount of the artifacts in the synthesized views.

9399-28, Session 4

How much image noise can be added in cardiac x-ray imaging without loss in perceived image quality?

Amber J. Gislason-Lee, Univ. of Leeds (United Kingdom); Asli E. Kumcu, Univ. Gent (Belgium); Stephen M. Kengyelics, Laura A. Rhodes, Andrew G. Davies, Univ. of Leeds (United Kingdom)

Purpose/Aims: In the treatment of coronary heart disease, dynamic X-ray imaging systems are used during percutaneous coronary interventional (PCI) procedures to see anatomy and clinical devices inside the human body. X-ray system settings are controlled automatically by specially-designed X-ray dose control mechanisms whose role is to ensure an adequate level of image quality is maintained with an acceptable radiation dose to the patient. If image quality is set too high, unnecessarily high levels of X-ray dose are used. X-ray exposure is harmful to humans; it may cause damaging short term effects such as skin burns and long term genetic effects such as cancer.

The utility of cardiac X-ray images is in their interpretation by a cardiologist during an interventional procedure, therefore image perception by a cardiologist should be considered to ascertain the required level of image quality to set the dose control. However, it is not well understood how changes in X-ray settings made by the dose control are related to a clinical professional's perception of image quality. For example, image degradation from an increase in image noise may not be perceived at all. With the long term goal of devising an image quality metric for an intelligent dose control system, we aim to determine the amount of noise which can be added to a patient image without making any difference to the perceived quality of the image. Noise is directly related to radiation dose, therefore results may demonstrate potential for a reduction in radiation dose used for PCI procedures.

Methods: Angiograms were selected from 5 different PCI patients in our local cardiac catheter laboratory. Images were selected to represent adult cardiac patient sizes (body mass index 23 to 44 kg m-2) and to include angular cardiac views commonly used in clinical practice. Images were obtained on an Allura Xper FD10 interventional X-ray system (Philips Healthcare, The Netherlands), modified by the manufacturer to allow for

image capture without computer image enhancement processing. Quantum noise was added to patient images; the Allura X-ray system had been used to perform calibrations and technical measurements of image quality required for the image synthesis software used to add image noise. For each patient image, a collection of that image representing various levels of degradation (i.e. dose levels) was created.

Observers (cardiologists) viewed image sequences on a medical grade monitor with appropriately dimmed ambient lighting and placement of the monitor to simulate the cardiac catheter laboratory. The software used to run the viewing sessions and estimate the observers' perception of image degradation was written in MATLAB specifically for this task, executing a "staircase" or "transformed up/down" psychophysics experiment.

The original and degraded images of the same patient were shown side by side as an image pair; left and right placement of the images in the pair were randomized and recorded in a keyfile on the host computer. Observers were asked to focus their attention on the clarity with which the coronary arteries of the heart were shown, answering the question "which side [left or right, in the image pair] shows the vessels more clearly?" in a two alternative forced choice (2AFC) test. The level of degradation was set high in the first few image pairs, making the difference between the left and right images apparent; these easy decisions allowed for a period of observer training, and results from training images were not used for data analysis.

The 1 up / 3 down rule was used following training: when an observer chose the original image three consecutive times (three "correct" responses), the level of degradation was reduced in the next image pair; when the observer chose the degraded image one time (an "incorrect" answer), the level of degradation was increased in the next image pair. A level of degradation was eventually reached where the observer had difficulty deciding between the original and degraded image, indicating that the degradation was no longer perceptible. Several reversals in direction (up and down) around a certain degradation level represented the observers' inability to consistently make correct decisions at that level. The mean of the reversal points (excluding training data) was the estimated level of degradation no longer perceived by the observer, known as the point of subjective equality (PSE). Each staircase terminated when the precision of the PSE estimate was below a threshold or the observer had viewed 50 image pairs. No time limit was imposed.

Results: The staircase experiment was found to be a time efficient method of measuring medical image perception. The point of subjective equality from preliminary results was approximately 20% dose reduction.

Conclusions: Results indicate that there is scope to increase the noise of cardiac X-ray images by up to 20% before it is noticeable by clinical professionals. This has implications in automatic dose control design, indicating that clinical observers are not as sensitive to changes in noise as would have been assumed by imaging physicists prior to this study. In addition, results indicate a potential for 20% radiation dose reduction during PCI procedures. The impact of such a result put into practice would affect both patient radiation dose and daily exposure incurred by catheter lab personnel.

9399-15, Session 5

Metamerism in the context of aperture sampling reconstruction

Alfredo Restrepo, Julian D. Garzon, Univ. de los Andes (Colombia)

Perceptually, but without considering contrast effects, two light beams are seen as hav- ing the same colour when the photoreceptor responses are the same: the photoreceptor responses are the only source of information to the human visual system (also the pho- tosensitive ganglion cells play a role). If the colours are the same but the lights have different spectra, the lights are said to be metameric. In computer vision, metamerism is defined as well: two pixels are said to have the same colour if their RGB values are the same.

The spectral responses of photoreceptors are more or less bell shaped, although more than one mode in the corresponding curves is also a possibility. A photoreceptor sensitive to only one wavelength is pretty much useless. Metamerism is then the result of aperture- sampling the spectra of lights, below a required sampling rate. In the case of human trichromacy, spectrum signals defined in the nanometer wavelength interval of [400, 700] are sampled using only three aperture samples.

When the natural domain parameter (as opposed to the Fourier fequency domain pa- rameter) of a signal s(?) is transformed with a homeomorphism h producing the signal s(?), ? = h(?), its bandwidth properties can change, in fact, it can go from not being band limited to being so and viceversa so, perhaps, it is possible that in some domain, different from wavelength, the spectral signals be band limited and that a small number of aperture samples be enough to reconstruct the spectrum signal. For example, sin(t) and sin(t3) have different bandwidth properties. Also, the composition g ? s of a signal s(?) with an (e.g. even and nondecreasing on the nonnegative semiaxis) function changes its spectrum; for example, sin(t) is band limited with zero bandwidth, | sin(t)| is not band limited and sin2(t) is again band limited with bandwidth of 2. So, it is plausible that under a certain transformation, 3 samples be approximately enough to characterise the spectra in the natural world.

The problem of aperture-sampling reconstruction has been considered for example in the context of scatterometer image reconstruction [Reconstruction form aperture-filtered samples with application to scatterometer image reconstruction; IEEE trans. on Geoscience and Remote Sensing, vol. 49, no. 5, May 2011, pp. 1663-1675]. Their aperture sampling functions An(x) are here the spectral response functions of the photoreceptors, and their natural domain variable x is here the wavelength variable ?.

The problem of the reconstruction of a band limited signal on the basis of a distorted version of it has been considered in the context of audio, of SSB modulation and Hilbert transform.

Mathematically, the problem of metamerism is traditionally stated in terms of the kernel of a linear transformation, here, we consider it also in the light of underdetermined aperture sample reconstruction, where the pseudo inverse of a certain full-rank matrix is considered.

We consider in particular the cases of dichromacy, trichromacy, tetrachromacy and pen- tachromacy; that is where 2, 3, 4 or 5 aperture samples for the wavelength nanometer interval, say [300, 800].

Surely Schr ?odinger's optimal colours, also related to the uniques as in blue-lowhpass red- stopband, green-bandpass, yellow-highpass (filter terminology regarding wavelength rather than electromagnetic frequency) are an effective classification of the spectra of lights in the natural world. Also: achromatic-allpass, so also Hering's basic colours.

9399-16, Session 5

Tensor representation of color images and fast 2D quaternion discrete Fourier transform

Artyom M. Grigoryan, Sos S. Agaian, The Univ. of Texas at San Antonio (United States)

Since the introduction of the fast Fourier transform (FFT) by Cooley-Tukey (1967), the Fourier analysis has become one of the most frequently used tools in signal and image processing. For color images, the traditional approach of processing images in the frequency domain is reduced to processing each color channel separately. The Fourier transform-based method is also one of these methods, which in many cases is applied to each color plane-component of the color image separately. In other words, the color image is considered as a triplet of separate 2-D gray scale images and each of these images represents red, green, or blue component of the color. Quaternion numbers of Hamilton's was used in Ell's works (1993) where the new concept of the quaternion Fourier transform was introduced, and after that time much attention was given to the transformation of the color components to the imaginary subspace of the quaternion numbers, "imaginary part" of which consists of three components. It becomes clear that the discrete quaternion Fourier transform is well-suited for color image processing applications, since it processes all three color components (R,G,B) simultaneously, it captures the inherent correlation between the

components, it does not generate color artifacts or blending, and finally it does not need an additional color restoration process. The quaternion Fourier transform was effectively used for enhancing color images in Grigoryan and Agaian works (2014) by generalizing the method of alpha-rooting in the quaternion algebra.

In this paper, we describe and analyze different methods of calculation of the 2-D QDFT and present a new approach in processing the color images in the frequency domain. Between two different spatial and frequency spaces exists an intermediate space, the so-called frequency-and-time space, or representation of the image in the form of one-dimensional signals which are generated by a specific set of frequencies. We introduce one of such representation for color images, which is called the color tensor representation when three components of the image in the RGB space are described by one dimensional signal in the quaternion algebra. The concept of the splitting-signals of images can also be used in other color model of images. The tensor representation is effective, since it allows us to process the color image by 1-D quaternion signals which can be processed separately. This representation also splits the algebraic structure of the 2-D QDFT, since each of the quaternion signals defines the 2-D QDFT in the corresponding subset of frequency-points. These subsets cover the entire Cartesian lattice of frequency-points on which the 2-D QDFT is defined. Therefore the 1-D quaternion signals are called the splitting-signals of the color image in tensor representation. The tensor transform-based algorithm of the 2-D QDFT is effective and simple, and uses less number of multiplications than the known methods. For instance, the tensor algorithm for the 2^r?2^r-point 2-D QDFT uses 18N^2 less multiplications than the well-known method of calculation based on the symplectic decomposition. The saving of 18N^2 multiplications has place also for other cases of N, because the tensor representation does not required multiplications. For the N=2^r case, we also can use a modified tensor representation which is similar to the paired representation for the gray scale images. The proposed algorithm is simple to apply and design, which makes it very practical in color image processing in the frequency domain.

9399-17, Session 5

Algorithms of the q2^r? q2^r-point 2D discrete Fourier transform

Artyom M. Grigoryan, Sos S. Agaian, The Univ. of Texas at San Antonio (United States)

Fast unitary transformations are well-known and used widely in image processing, namely in image compression, restoration and enhancement, linear filtration and image reconstruction. Other unitary transform are also used in image processing. The suitability of unitary transforms in each of the above applications depends on the properties of basic functions of transforms as well as on the existence of fast algorithms, including parallel ones. One of the most important transformations in image processing is the Fourier transformation with different fast algorithms for one- and two-dimensional cases. We here mention the general algorithm of splitting, namely the manageable split algorithm for computing unitary transforms of arbitrary orders by using different partitioning in the frequency domain that yield new representations of images. These types of partitioning have been used effectively for multidimensional transforms and can be also used for the one-dimensional case. The number of partitions depends on the type of the unitary transforms and their orders. The greater the order of the transform, the more opportunities to obtain such partitioning that may result in an effective decomposition of the transform. For instance, the split algorithm for the 2^r-point discrete Fourier transform requires a minimum number 2^{r-1}(r-3)+2 of operations of multiplications. The same splitting can be used for the Hadamard transform, and as a result, the computation of the 2^r-point discrete Hadamard transform uses on the average no more than six operations of additions per sample.

In this paper, we use the concept of partitions revealing transforms for computing the 2-D DFT of order q2^r?q2^r, where r>1 and q is a positive odd number. By means of such partitions, the 2-D transform can be split into a number of short transforms, or 1-D M-point DFTs where M< q2^r. In the one dimensional case, the partitions determine fast transformations that split the q2^r-point Fourier transform into a set of Nk-point transforms, where k=1:n and N1+... +Nn=q2^r, and minimizes the computational complexity of the q2^r-point discrete Fourier transform. The 2-D q2^r?q2^r-point DFT can be calculated by the column-row method with 2(q2^r) 1-D DFTs, each of which can be split by the short transforms, by means of the 1-D paired transforms. Another and more effective algorithm of calculation of the 2-D q2^r?q2^r-point DFT is based on the splitting by the 2-D tensor or paired transform which lead to the calculation with a minimum number of 1-D DFTs.

9399-18, Session 5

A method for predicting DCT-based denoising efficiency for grayscale images corrupted by AWGN and additive spatially correlated noise

Aleksey S. Rubel, Vladimir V. Lukin, National Aerospace Univ. (Ukraine); Karen O. Egiazarian, Tampere Univ. of Technology (Finland)

Noise is one of prevailing factors that affect quality (including visual) of images. Quality of images acquired by conventional optical systems and devices is usually decreased by additive noise. To increase quality of images, some denoising operations could be carried out. However, it is known that such operations are unable to always increase image quality; sometimes denoising can even lead to quality decreasing. Thus, it is desirable to have some preliminary information about denoising efficiency before applying denoising.

Among modern and recently developed filters, the best performance is usually provided by denoising techniques that belong to two groups: methods based on orthogonal transforms and the non-local denoising methods. It is often necessary to pay attention to visual quality of denoised images and to perform analysis of filtering efficiency using modern visual quality metrics. Keeping this in mind, a practical task is to predict filtering efficiency before starting to denoise a given image. If prediction is reliable (accurate enough) and much faster than filtering itself, then it becomes possible to undertake (including automatic) decision is it worth carrying out filtering or no. This can save processing time in automatic or automated chains (procedures) of image processing for different applications.

The main idea of denoising efficiency prediction is the following. Assume that there is a parameter that is commonly accepted as the one able to adequately characterize filtering efficiency. An example of such a parameter is the aforementioned IPSNR (larger this parameter is more efficient denoising is observed). Assume also that there is a statistical parameter able to jointly characterize image complexity and noise level (in fact, complexity of noise removal). Examples of such parameters are averaged (for all image blocks) probabilities1 that amplitudes of DCT coefficients in blocks are larger (or smaller) than a certain threshold (that is connected with additive noise standard deviation assumed a priori known or pre-estimated with appropriate accuracy). Suppose now that there is a rather strict connection between IPSNR (or other parameter characterizing filtering efficiency) and aforementioned statistical parameter where this connection (dependence) is described analytically. Then, by estimating the considered statistical parameter and inserting it into the established dependence one gets a predicted IPSNR that allows characterizing denoising efficiency and decision undertaking.

Clearly, there are many questions that arise as, e.g., what is the best parameter describing filtering efficiency, what statistical parameter to use, how to get the aforementioned dependence and for what filters it is strict (good) enough, for what types of noise the efficiency prediction can be performed, how fast prediction can be and so on. Some of these questions have been already fully or partly answered. In particular, the following has been already demonstrated. First, the threshold 2? (? denotes additive noise standard deviation) is a quite good threshold for AWGN case1 (although this choice can be not the best). Second, for obtaining the desired dependences one needs to carry out preliminary (off-line) study with considering many test images and noise levels, obtaining scatter-plots (in the coordinate system filtering efficiency vs statistical parameter), and curve fitting into these scatter-plots for obtaining the desired dependences. Third, quite good

(strict) dependences have been obtained for IPSNR for the conventional DCT filter and BM3D, slightly worse dependences have been got for improvement of visual quality metric PSNR-HVS-M. Fourth, it has been shown that prediction can be achieved for white signal-dependent noise of known type under condition that this dependence of local variance on local mean is a priori known or pre-estimated.

This means that the prediction approach is becoming more and more universal. However, only the cases of white noise (signal-independent and signal dependent) have been analyzed yet. Meanwhile, the case of spatially correlated noise is important as well since it happens in practice quite often and spatially correlated noise degrades visual quality more than AWGN with the same variance. Because of this, in this paper we concentrate on considering the case of spatially correlated noise assuming that its 2D DCT spectrum for 8x8 pixel blocks is a priori known or pre-estimated with high accuracy. Here we have studied three cases of additive noise. The first one is AWGN that can be generated easily by embedded means in various modelling programs. Other two cases relate to spatially correlated noise type. We have called them middle and strong correlation. To obtain additive spatially correlated noise, 2D zero mean AWGN realization is needed. Then, 2D FIR filters with certain matrices of weights are applied.

The basic principle of the proposed prediction method is the use of some parameter that can be estimated from some noisy image with priori known noise characteristics. As such parameter, amount (percentage) of spectral components that will be removed (or preserved) by filter can be used. Thus, the algorithm of predicting will be, in some sense, similar to the conventional DCT-based denoising procedure.

9399-19, Session 6

Cost Volume Refinement Filter for Post Filtering of Visual Corresponding

Shu Fujita, Takuya Matsuo, Norishige Fukushima, Yutaka Ishibashi, Nagoya Institute of Technology (Japan)

Basic image processing for depth maps, optical flows or segment images is important. The problem to estimate these corresponding maps can be defined by a discrete label-based problem. There are a lot of methods for the problem, such as Markov random field optimization, convex optimization, and non-local filtering method. After the processing, usually edge-preserving filter is applied as post filtering. For such visual correspondence problems, joint filtering such as joint bilateral filtering, which uses corresponding maps and also uses RGB images, is a suitable for refining detail of object boundaries. Filtering the corresponding maps directly make blur; thus filtering some transformed domain is preferred.

In this paper, we propose a cost volume refinement filter (CVRF), which a generalized filtering in a range-transformation for refinement of the corresponding maps. In general, there are three main steps in the filtering with range-transformation. They are (a) building a cost volume, (b) refining cost slices, (c) merging the cost slices. The cost volume consists of cost slices, which have splatted range-values of a corresponding map by using some functions to a higher dimension. Then, the slices are refined by some filtering methods. After that, we obtain an output image by merging the cost slice. These processes are the general flow in the cost volume filter.

revious works which are the basis of the CVRF is individually formulated, for example, metrics for building a cost volume, and kind of filters for refining the cost slices. These methods are not well conducted and not evaluated for various applications. The contributions of the paper are:

(1)Generalizing each step (a) to (c),

(2)Adding range and domain resizing characteristics to this for the more generalized format,

(3)Comparing with various methods with three applications; depth map, optical flow and segmentation estimation.

Now, we review each step (a) to (c);

(a) Building a cost volume: The cost volume consists of the cost slices, and the number of the slices is the number of candidates of labels. The slices are defined by the difference between the pixel and label value. The metrics of distance should be a monotonically increasing function, then we adopt L1, L2 norm and exponential function as the representative.

(b) Refining cost slices: The cost slices are filtered by the effect of noise reduction method, but a recommended type of the filter is edge-preserving filtering (e.g., bilateral filtering, joint bilateral filtering and guided filtering). Moreover, weighted filtering, whose weight around object boundaries is low, have higher performance. In our work, we use the weight of the trilateral filter as the weight map.

(c) Merging the cost slices: The final output pixel value is simply selected by using winner-takes-all strategy for the minimum cost value or selected by sorting for weighted median output.

In this way, we can can obtain a refined map through the CVRF.

In our experiment, we refine depth maps as a representative of single channel maps and optical flows as a representative of multi-channel maps. The depth maps are estimated by block matching and semi-global matching method. The methods for the estimating optical flows are F-TV-L1 and Farneback method.

We evaluate the performance of post filtering of the CVRF applied by various methods. The methods of building a cost volume are L1, L2 norm and exponentially decreasing function. In addition, we up/down-sample the dynamic range by converting the number of the cost slices to 128 or 512 from 256 as resizing of the range format. For refining the cost slices, we use the Gaussian filter, the guided filter, the joint bilateral filter. We also compare the performance between using the trilateral weight map and not. Moreover, we evaluate the performance of CVRF about noise reduction for Gaussian noise and up-sampling as the resizing of the domain format. Since we assume that depth maps obtained by a depth sensor usually include noises, we also up-sample the depth map including noises. At this time, we add Gaussian noise to the down-sampled depth map.

Our experimental results show what format is appropriate for the CVRF. For refining the cost slice, the amount of improvement is bigger when we use the edge preserving filters. Especially, the performance of the joint bilateral filter is better than the guided filter in terms of the edge preserving filters. Then, the accuracy becomes higher when we use the weight map. For building the cost volume, L2 norm or exponential function is valid because L1 norm is not robust for noises.

Moreover, in the dynamic range up/down-sampling, we should not down-sample the dynamic range if we want sub-pixel accuracy. In the result of resolution up-sampling, the CVRF with the edge preserving filter has good performance if the target image does not include noises. By contrast, if the target image includes noises, the edge non-preserving smoothing filter is better for refining cost slices. The reason is that the noises have been extended by up-sampling, and hence the edge preserving filter cannot reduce the noises because of the edge preserving effect. We can also confirm that the optical flow refinement by the CVRF is effective, but the edge non-preserving smoothing filter becomes better when the quality of input flow is low. This reason is almost same as the noisy image up-sampling.

In this paper, we generalized a cost volume refinement filter (CVRF) and evaluated the performance of CVRF applied by various methods. Experimental results showed that the CVRF is effective for various applications. Moreover, we demonstrated the appropriate methods for the cost volume filter by comparing each method.

9399-20, Session 6

Depth remapping using seam carving for depth image based rendering

Ikuko Tsubaki, Kenichi Iwauchi, Sharp Corp. (Japan)

For stereoscopic displays, depth range of a stereo image pair, which is a range from the minimum to the maximum depth (disparity) value, is one of important factors to control the three-dimensional appearance. If a displayed stereo image has too large depth, viewers are unable to fuse binocular images and therefore perceive double images. Depth remapping is a technique to control the depth range of stereo images, and utilized in stereoscopic content production and stereoscopic displays. The depth range is decreased by depth remapping if the range is wider than a target range.

There has been much research concerning depth remapping, such as linear and nonlinear transformation. If the target depth range is much smaller than the input range, a stereo image synthesized from the corrected depth map by conventional methods may give much smaller sense of depth to viewers. Especially, the details of depth structure are lost and not perceived.

Seam carving is an intelligent image resizing algorithm without distorting the important objects, unlike simple scaling methods. A seam is defined as a path across the image from one side to another with the minimum cost based on the gradient information, and pixels on the seam are removed. Dynamic programming has been originally used to find the appropriate seam. Graph cut based seam carving has been introduced later for resizing video sequence, where video images are treated as a space-time volume, and a seam is extended from 1D path on a 2D image, to 2D manifold in a 3D volume. The 2D manifold is called as a seam surface, and it must be monotonic, including only one pixel in each row, and connected.

A depth remapping method which preserves the details of depth structure is proposed in this paper. We apply seam carving, which is an effective technique for image retargeting, to depth remapping. An extended depth map is defined as a space-depth volume, and a seam surface is defined as a manifold in the 3D volume. The seam surface is monotonic, including only one pixel in each coordinate, and connected. We use the seam surface as a cutting surface for reducing depth range. The depth range is reduced by removing depth values on the seam surface from the space-depth volume in our concept. The appropriate seam surface is selected using graph cut algorithm. After the corrected depth map is created, a stereo image pair is synthesized from the corrected depth map and an input image by conventional depth image based rendering (DIBR).

The process of depth remapping is explained below. We define an extended depth map which has a depth histogram in neighboring region at each pixel. A seam surface is defined as a manifold which connects depth values with small values in the extended depth map at each pixel, and derived by selecting the small values under the condition that depth value is continuous between adjacent pixels. The optimal seam can be found using the graph cut algorithm by minimizing the cost. A grid-like graph is constructed from the extended depth map in which every node represents a voxel, and connects to its neighboring voxels. Two virtual terminal nodes, source and sink, are created. Every internal node is connected with 6 arcs. Four of them are diagonal so as to create a connected seam surface. We assign the weight based on the value of the extended depth map to the arc.

A partitioning of the graph to source and sink is performed by max-flow min-cut algorithm. The seam surface is obtained as a cutting surface by partitioning. The arc which has small value is easy to select as the minimum cut.

Converting is executed by decreasing depth values by 1 at pixels which have the depth value larger than the seam surface. The meaning of this process is that the cuboid of the extended depth map is divided into two parts by the seam surface, one part is pruned on the surface, and the two parts are again bonded to each other. As a result, the length of depth direction of the cuboid becomes shorter by 1. This process is iterated until the depth range becomes the target range. After the corrected map is obtained, a stereo image pair is synthesized by DIBR using conventional image warping technology.

The proposed algorithm was tested with two depth maps, “Cones” and “Art”. We reduced the range of depth from 159 to 32 for Cones, from 147 to 32 for Art. The depth range of the corrected map is much smaller than the originals, but these stereo images give a natural sense of depth to viewers. We confirmed that the corrected maps have the detail depth structure. For comparison, the conventional depth range reduction based on linear transformation and histogram equalization is performed. The corrected depth maps by conventional methods have less detail information.

9399-21, Session 6

Depth map occlusion filling and scene reconstruction using modified exemplar-based inpainting

Viacheslav V. Voronin, Vladimir I. Marchuk, Alexander V. Fisunov, Svetlana V. Tokareva, Don State Technical Univ. (Russian Federation); Karen O. Egiazarian, Tampere Univ. of Technology (Finland)

RGB-D sensors are relatively inexpensive and are commercially available off-the-shelf. However, owing to their low complexity, there are several artifacts that one encounters in the depth map like holes, mis-alignment between the depth and color image and lack of sharp object boundaries in the depth map. Depth map generated by Kinect cameras also contain a significant amount of missing pixels and strong noise, limiting their usability in many computer vision applications. In this paper we present an efficient hole filling and damaged region restoration method that improves the quality of the depth maps obtained with the Microsoft Kinect device. The proposed approach based on a modified exemplar-based inpainting and LPA-ICI filtering by exploiting the correlation between color and depth values in local image neighborhoods. The edges of the objects are sharpened and aligned with the objects in the color image using such approach. Several examples considered in this paper show the effectiveness of the proposed approach for large holes removal as well as recovery of small regions on several test images of depth maps. We perform a comparative study and show that statistically, the new algorithm deliver good quality results compared to existing algorithms.

9399-22, Session 6

Real-time depth image-based rendering with layered dis-occlusion compensation and aliasing-free composition

Sergey Smirnov, Atanas P Gotchev, Tampere Univ. of Technology (Finland)

Depth Image-based Rendering (DIBR) is a popular view synthesis technique which utilizes the RGB+D image format, also referred to as view-plus-depth scene representation. Classical DIBR is prone to dis-occlusion artefacts, caused by the lack of information in areas behind foreground objects, which appear visible in the synthesized images. A number of recently proposed compensation techniques have addressed the problem of hole filling. However, their computational complexity does not allow for real-time view synthesis and may require additional user input. In this work, we propose a hole-compensation technique, which works fully automatically and in a perceptually-correct manner. The proposed technique applies a two-layer model of the given depth map, which is specifically tailored for rendering with free viewpoint selection. The main two components of the proposed technique are an adaptive layering of depth into relative ’foreground’ and ’background’ layers to be rendered separately and an additional blending filtering aimed at creating a blending function for aliasing cancellation during the process of view composition. The proposed real-time implementation turns ordinary view-plus-depth images to true 3D scene representations, which allow visualization in the fly-around manner.

Conference 9400: Real-Time Image and Video Processing 2015

Tuesday 10 February 2015

Part of Proceedings of SPIE Vol. 9400 Real-Time Image and Video Processing 2015

9400-1, Session 1

Customized Nios II multi-cycle instructions to accelerate block-matching techniques

Diego González, Guillermo Botella, Carlos Garcia, Univ. Complutense de Madrid (Spain); Uwe Meyer-Baese, Anke Meyer-Baese, Florida State Univ. (United States); Manuel Prieto-Matias, Univ. Complutense de Madrid (Spain)

In this contribution, we describe how to accelerate our sensor application thanks to the inclusion of a new customized instruction in the Nios II processor Instruction Set Architecture (ISA). This instruction helps to reduce the time leak discovered with the profiling information in the aforementioned motion estimation techniques. Our system has been provided with two customized instructions, which best fit to our application. This is an extension of the work addressed in [1] regarding monocycle customized instruction and it constitutes an alternative in terms of acceleration regarding using C2H compiler [2], so constructing a new enhanced multi-cycle instruction based system is achieved as a 450 KPPS performance in the best case, equivalent to a SoC which processes 50?50 @ 180 fps.

Moreover, in this work will be explained in detail every memory type available in our Altera FPGA platform. Extracting the most powerful memory types and doing an exhaustive testing plan with the selected memories, we have found another way to improve the system. Finally, combining the memory system design with the customized instructions explained, an additional improvement is achieved.

Regarding multi-cycle custom instruction, the average performance got is about 45% for the full set of parameters: window and macroblock sizes, algorithms and processor architecture used. The maximum throughput using this design is an improvement about 75% (window and macroblock sizes of 32, FST, Nios II/e processor). With the optimization of using the memory types available in the design, an improvement of 60% was achieved in the execution time. Considering, finally the combination of both techniques, an improvement of 80% was reached on average and a 90% for the optimum case.

Now, commenting the achieved results, we can observe that looking to the FST technique we almost achieve 6 KPPS in the best case, despite of the used sequence, due to the high number of operations requested by this exhaustive technique with every processed frame. The best case is achieved when using a macroblock size of 64 executing under the Nios II/s processor, due to the exploitation of the instruction cache that this processor provides and the low number of iterations needed when using this macroblock size. Due to all the reasons explained previously along this work, we can accomplish a small sensor of 50?50 @ 2.5 fps.

Focusing on the 2DLOG technique, we achieve nearly 350 KPPS in the best case when executing the “Carphone” sequence, due to the fact that this algorithm needs less number of operations to process a frame than the FST technique. The best case is achieved when using a macroblock size of 32 executing under the Nios II/f processor, despite of the used sequence, due to the use of the instruction cache that this processor provides, and the exploitation of the data cache, that only this processor provides, when using this macroblock size. The final performance obtained suggests that a SoC working with a little sensor of 50?50 @ 130fps would be fully functional.

Regarding the TSST algorithm, we achieve nearly 450 KPPS in the best case when executing the “Carphone” sequence, due to the fact that this algorithm needs even less number of operations to process a frame than the 2DLOG technique. The best case is achieved when using a macroblock size of 32 executing under the Nios II/f processor, despite of the used sequence, due to the use of the instruction cache that this processor provides, and the exploitation of the data cache, that only this processor provides, when using this macroblock size. In this situation we can construct a SoC which processes 50?50 @ 180 fps and 170 fps respectively for either multi-cycle or monocycle approaches.

9400-2, Session 1

Hardware design to accelerate PNG encoder for binary mask compression on FPGA

Rostom Kachouri, ESIEE Paris (France); Mohamed Akil, Ecole Supérieure d’Ingénieurs en Electronique et Electrotechnique (France)

Within the framework of the Demat+ project, we aim to propose a complete solution for storage and retrieval of scanned documents. In this context, the intended paperless application to achieve have to transmit the scanned documents with a low-bandwidth network to a computer cloud. The overall idea consist to partition the source document into three layers: a foreground layer, a background layer, and a binary mask, and then to use different compression strategies for the same document. In literature, the often used taxonomy distinguishes two categories of image compression format: the lossless compression formats and the lossy ones. The lossless compression formats perform compression on the image matrix. It is worth noting that the transformation between a raw format and the lossless compression one is bijective. Which means that when decompressing a lossless compressed image, the original image is restored, and it is a 100% identical copy of the original. On the other side, the lossy compression formats achieve better compression rate at the cost of image degradation. A quantification stage is applied in this case on the frequency transform of the image. We note that the foreground and background layers which contain, respectively, the color information of the text and the original background of the image can be compressed via a lossy compression format like JPEG for example. By against, the binary mask, where the text, and possibly pieces of thin strokes when they exist, are located is necessarily subjected to a nondestructive compression. One of the most interesting lossless compression formats is PNG (Portable Network Graphics). In fact, it provides a portable, legally unencumbered, wellcompressed, well-specified standard for lossless bitmapped image files.

To respond to real-time constraints, aimed to be respected by the SagemCom company, we expect in this paper to accelerate the employed PNG encoder for binary mask compression through a hardware implementation. Indeed, while compare with software encoding, parallel processing is the most significant feature of high efficiency. Front of a custom development, based on ASIC technology, the capacity and performance of current FPGAs are such that they present a much more realistic alternative than they have been in the past. Effectively, FPGAs allow a rapid soft reconfiguration of on chip hardware. Moreover, with a sufficient number of parallel operations, FPGAs can offer better performance-price and power dissipation than state of the art microprocessors or DSPs. Many implementations of lossy and lossless image compression encoders and decoders were performed on reconfigurable FPGA circuits.

In this paper, we discuss a hardware accelerated implementation of the PNG encoder for binary mask compression on FPGA. An optimized architecture is proposed as part of an hybrid software and hardware co-operating system. For its evaluation, the new designed PNG IP has been implemented on the ALTERA “Arria II GX EP2AGX125EF35” FPGA. The experimental results show a good match between the achieved compression ratio, the computational cost and the used hardware resources. The paper is organized as follows: first we present briefly the PNG encoder in section II. The performed PNG encoder optimizations, in order to well-ensure binary mask compression, are provided in section III. Section IV describe the proposed hardware design to accelerate the optimized PNG encoder on FPGA. Then, the obtained experimental results are discussed in section V. Finally, section VI concludes the discussion.

Conference 9400:
Real-Time Image and Video Processing 2015

9400-3, Session 1

Real-time algorithms enabling high dynamic range imaging and high frame rate exploitation for custom CMOS image sensor system implemented by FPGA with co-processor

Blake C. Jacquot, Nathan G. Johnson-Williams, The Aerospace Corp. (United States)

Prior generations of space systems typically used CCDs for visible imaging. As the CCD industrial base declines, CMOS imagers are becoming the dominant visible image sensor for space-based applications. The slow CMOS adoption relative to other industries is driven largely by CCDs having characteristics that fit well with space surveillance needs. Such capabilities include time delay and integrate (TDI), low fixed pattern noise (FPN), and high dynamic range. The generally lower dynamic range of CMOS imagers can be partly addressed by techniques and algorithms for High/ Wide Dynamic Range CMOS (HDR/WDR CMOS). While algorithmic HDR imaging helps bridge the gap between CCDs and CMOS imagers, it is either computationally intensive or requires high bandwidth data transmission.

Real-time algorithms help address computation and bandwidth issues for CMOS image sensor systems, but implementing real-time algorithms still requires significant computing power. The system must capture, analyze, and store every pixel according to algorithm constraints without dropping frames.

This paper documents the implementation of a real-time multiple samples algorithm to extend dynamic range for a custom-built camera module with commercial CMOS image sensor. Though used in ground-based applications, this algorithm is new to space systems, which typically use well adjustment for dynamic range extension. The multiple samples technique (in model form) allows arbitrarily large dynamic range with appropriately scaled computational demands. The algorithm is implemented by FPGA to relieve burdens on host processor.

The multiple samples algorithm samples and stores each pixel at higher rates than intended for the final HDR image and processes these samples according to the algorithm constraints. At evenly spaced intervals in time, each pixel value is recorded. At the end of the frame cycle, each pixel is analyzed to determine if it has reached full well and is saturated or not. If not saturated, the last recorded value is stored for the HDR image. However, if saturated, the algorithm backtracks along the pixel samples until it finds the last sample before saturation (LSBS). With this LSBS sample and record of the origin, the algorithm extrapolates based on a linear fit to what value the pixel would have if it did not reach full well at frame end. This extrapolated pixel value is stored for the HDR frame. After the frame time, all pixels are reset and the process repeats.

The CMOS image sensor used in this demonstration has 1280 ? 1024 format with 10-bits per pixel and frame rates of 0.5 kHz to 100 kHz. Crucial to demonstrating the multiple samples method is non-destructive readout (NDR) and global shutter.

These real-time algorithms support applications such as motion tracking in day/night vision environments and help close the gap between CCD and CMOS imagers as well as sensor data generation rates and available bandwidth for satellite imaging.

9400-4, Session 1

Fast semivariogram computation using FPGA architectures

Yamuna Lagadapati, Mukul V. Shirvaikar, Xuanliang Dong, The Univ. of Texas at Tyler (United States)

The semivariogram is a statistical measure of the spatial distribution of data based on Markov Random Fields (MRFs). It is similar to other measures of texture like spatial covariance and co-occurrence matrices.

Semivariogram analysis is a computationally intensive algorithm that has typically seen applications in the geosciences and remote sensing areas. Recently applications in the area of medical imaging are being investigated, resulting in the need for efficient real time implementation of the algorithm. The semivariogram is a plot of semivariances for different lag distances between pixels. A semi-variance, ?(h), is defined as the half of the expected squared differences of pixel values between any two data locations with a lag distance of h. The semivariance value typically increases with the lag distance converging to a constant limit constant called the “sill”. The value increases rapidly at low lags and then progresses linearly. A simple mathematical function can be used to describe the underlying stochastic process of experimental variograms. Such a mathematical model, known as an authorized model, must satisfy the following conditions: an intercept on the ordinate (axis of semi-variance), a monotonically increasing section and conditional negative semi-definite. Only functions that ensure non-zero variances may be used for variograms. Examples of simple authorized models are exponential, Gaussian and spherical models. The semivariogram graph is typically fitted to an exponential model using least squares approximation. Due to the need to examine each pair of pixels in the image or sub-mage being processed, the base algorithm complexity for an image window with n pixels isO(n^3).

This paper presents a technique for the fast computation of the semivariogram using a custom FPGA architecture. The design consists of several modules dedicated to the constituent computational tasks. The squared difference module computes the squares of the gray-level differences between pixel pairs. The distance module computes the distances between pixel pairs under consideration. A counter module keeps track of the number of pixel pairs for each lag distance. A correlated accumulator module aggregates the data and computes the result. Image data is fed to the modules through a pipeline mechanism from memory. A modular architecture approach is chosen to allow for replication of processing units. This allows for high throughput due to concurrent processing of pixel pairs. The current implementation is focused on anisotropic semivariogram computations only. Directional semivariogram implementation is anticipated to be an extension of the current architecture, ostensibly based on refinements to the current modules. The algorithm is benchmarked using VHDL on a Xilinx Spartan-3E FPGA for varying image sizes. Medical image data from MRI scans is utilized for the experiments. Computational speedup is measured with respect to Matlab implementation on personal computer with an Intel i7 multi-core processor. Preliminary simulation results indicate that a significant advantage in speed can be attained by the architecture making the algorithm viable for implementation in medical devices.

9400-5, Session 1

2D to 3D conversion implemented in different hardware

Eduardo Ramos-Diaz, Univ. Autónoma de la Ciudad de México (Mexico); Victor Gonzalez-Huitron, Volodymyr Ponomaryov, Instituto Politécnico Nacional (Mexico); Araceli Hernandez-Fragoso, Colegio de Estudios de Posgrado de la Ciudad de México (Mexico)

Conversion of available 2D data for release in 3D content is a hot topic for providers and for success of the 3D video applications, in general. It naturally completely relies on virtual view synthesis of a second view given by original 2D video. Disparity map estimation is a central task in 3D content generation but still follows a very difficult problem for rendering novel images precisely. DM estimation opens a wide variety of interesting new research topics and applications, such as virtual view synthesis, high performance imaging, remote sensing, image/video segmentation, 3DTV channels, object tracking and recognition, environmental surveillance, and other applications.

There exist different approaches in DM reconstruction, among them manually and semiautomatic methods that can produce high quality DMs but they demonstrate hard time consuming and are computationally expensive. The key of success in performance of a reliable embedded real

time capable stereovision system is the careful design of the core algorithm. The tradeoff between execution time and quality of the matching should be handled with care and is a difficult task. However, for extracting dense and reliable 3D information from observed scene, stereo matching algorithms are computationally intensive. To enable both accurate and efficient real time stereovision in embedded systems, we propose several hardware implementations of two designed frameworks for an automatic 3D color video generation based on 2D real video sequence.

The analyzed frameworks include together processing of neighboring frames using the following blocks: the RGB to CIELa*b* color space conversion, wavelet transform (WT) with edge detection via soft thresholding procedure proposed by Donoho using HF wavelet sub-bands (HF, LH and HH) or pyramidal scheme, color segmentation via k-means on a*b* color plane, up-sampling in wavelet case, disparity map (DM) estimation using stereo matching between left and right images (or neighboring frames in a video) based on criterion SSD, adaptive post-filtering, and finally, the anaglyph 3D scene generation. The SSIM and QBP criteria are applied in order to compare the performance of the proposed frameworks against other 3D computation techniques.

During numerous simulation experiments using wavelet families (Coiflets, Symlets, Haar, Daubechies, Biorthogonal, and wavelet atomic functions (up, upn, fupN, gn), we have selected the Haar wavelet in the final hardware implementation because of its simplicity and computational low cost.

The designed techniques has been implemented on DSP TMS320DM648, Matlab's Simulink module over a PC with Windows 7, and using graphic card (NVIDIA Quadro K2000) demonstrating that the proposed approach can be applied in real-time processing mode.

The time values needed, mean SSIM and BP values for different hardware implementations (GPU, Single CPU, Multicore CPU and DSP) are exposed in this paper. The data for time consumption and quality metrics in case of stereo pair Tsukuba (384x288), set of images from Middlebury set (27 stereo pairs, size of 1390 x 1110) and complete Tsukuba video sequence (1800 frames) are presented and discussed in this work.

In case of GPU implementation, when the stereo matching stage has been implemented in parallel only, we can obtain sufficient time reduction for large processing windows. The DSP and CPU implementations run slower in comparison with GPU and Multicore CPU. The presented processing times are consistent with the devices speeds; this is because the CPU at 3.4 GHz is faster than the DSP processor (900 MHz), but on other hand, the DSP module has in more than 30 times less power consumption.

9400-6, Session 1

A real-time GPU implementation of the SIFT algorithm for large-scale video analysis tasks

Hannes Fassold, JOANNEUM RESEARCH Forschungsgesellschaft mbH (Austria); Jakub Rosner, Silesian Univ. of Technology (Poland)

The SIFT (Scale-Invariant Feature Transform) algorithm is one of the most popular methods for extracting and describing local features in an image. Due to its robustness against scale/rotation changes and partial occlusions, it is widely used for tasks like action detection, panoramic image stitching and autonomous vehicle navigation. Furthermore, it is used very successfully in all sort of video analysis tasks like object recognition, instance search, duplicate and near-duplicate detection and clustering by visual similarity.

Despite the practical importance of the SIFT algorithm, so far only one CUDA implementation ("SiftGPU") exists which is somewhat outdated and does not give the same reliable results as e.g. the high-quality SIFT CPU implementation in the HessSIFT library, as not all robustness-enhancing steps within the algorithm have been ported to the GPU.

In this work, we present an efficient GPU implementation of the SIFT descriptor extraction algorithm using CUDA for NVIDIA GPUs with Fermi architecture or later. The major steps of the algorithm (extrema detection in the DoG scale-space, key point candidate refinement & filtering, key point descriptor calculation) are presented. For each step we describe how to efficiently parallelize it massively, how to take advantage of the unique capabilities of the GPU like shared memory / texture memory and how to avoid or minimize common GPU performance pitfalls like scattered memory accesses and branch divergence.

We compare the GPU implementation with the reference CPU implementation from the HessSIFT library in terms of runtime and quality. The results show a significant speedup factor of approximately 3 - 5 for SD resolution and 5 - 6 for Full HD resolution with respect to the multi-threaded CPU implementation (Geforce GTX 480 GPU vs. Quad-Core Xeon 3.0 GHz CPU). This allows us to extract 1,000 SIFT descriptors for each frame of an SD video in real-time. Even for video in FullHD resolution, we achieve real-time processing when calculating up to 5,000 SIFT descriptors at 10 fps, which is usually enough for video analysis tasks. Furthermore, quality tests show that the GPU implementation gives nearly identical results than the CPU HessSIFT routine.

Efficient extraction of SIFT descriptors is not only relevant for traditional SIFT matching, but also for compressed feature representations such as Vector of Locally Aggregated Descriptors (VLAD) or Vectors of Locally Aggregated Tensors (VLAT). While these methods are highly optimized for matching, descriptor extraction is a performance bottleneck, which is addressed by our proposed approach.

We further describe the benefits of GPU-accelerated SIFT descriptors for applications such as near-duplicate video detection, which aims at detecting duplicates almost identical video segments in large video data sets, or linking video segments by shooting location or salient object.

9400-7, Session 2

Real-time deblurring of handshake blurred images on smartphones

Reza Pourreza-Shahri, Chih-Hsiang Chang, Nasser Kehtarnavaz, The Univ. of Texas at Dallas (United States)

Many images captured by smartphones appear blurred due to handshakes. One way this problem is alleviated is by recreating the scene of interest and recapturing it while keeping the hand steady. However, there are situations when the moment of interest cannot be repeated. This paper discusses a camera app or a menu option on smartphones to deblur such blurred images.

The problem of image deblurring has been extensively studied in the image processing literature, e.g. [1-3], particularly within the field of computational photography. Conventional techniques such as those that estimate the point-spread-function (PSF) and then utilize a deconvolution filter are computationally demanding and face real-time implementation challenges on mobile devices such as smartphones because of their limited computational resources and memory compared to desktop computers. An attempt was made in [4, 5] to estimate the PSF in real-time by using an inertial sensor attached to the camera platform. However, it was seen that the calibration between the inertial sensor and the camera severely affected the deblurring performance.

In [6, 7], a computationally efficient approach to image deblurring was introduced by capturing a short-exposure image immediately after or during the time the blurred image got captured. Although the short-exposure image appeared darker, it did not suffer from blurring due to its short exposure duration. Then, an adaptive tonal correction (ATC) algorithm was used to enhance the short-exposure image such that its brightness and contrast matched the brightness and contrast of the blurred image. In [7], it was shown that the ATC algorithm was able to produce deblurred images with lower MSE and higher SSIM than those produced by the PSF approach in [4].

Inspired by the ATC algorithm, a new image deblurring algorithm is introduced in this paper that is computationally more efficient than the ATC algorithm and is thus more suited for implementation on smartphone platforms. The first step of this algorithm involves performing a low rank matrix approximation [8]. This is achieved by applying SVD (Singular Value Decomposition) to extract a low rank approximated image from the

blurred image. This approximated image does not suffer from blurring while incorporating the image brightness and contrast information. The second step of this algorithm involves combining the eigenvalues extracted from the low rank approximated image with those from the short-exposure image. In contrast to the ATC algorithm, this new algorithm does not require any initialization and also does not require any iterative search. The results that will be reported in the full length paper will show that this new algorithm runs about 70% faster than the ATC algorithm. In addition, the results of an actual real-time implementation of this algorithm on an Android smartphone will be reported.

A sample experimental result is shown in Fig. 1. Figs. 1(a) and 1(b) denote a short-exposure and a regular or auto-exposure image which appears blurred because of handshakes when the image was taken. The short-exposure image appears not blurred but darker. Figs. 1(c) and 1(d) illustrate the deblurring outcomes by using the ATC and the introduced algorithm, respectively. As shown in Fig. 2, the cumulative histogram of the introduced algorithm has a similar shape to that of the blurred image.

[1] Fergus, R., Singh, B., Hertzmann. A., Roweis, S. and Freeman W., "Removing camera shake from a single photograph," ACM Trans. Graph, 25(1), 787-794, (2006).

[2] Portz, T., Zhang, L., and Jiang, H., "High-quality video denoising for motion-based exposure control," Proc. of IEEE International Conference on Computer Vision Workshop, 9-16, (2011).

[3] Ben-Ezra, M., and Nayar, S., "Motion-based motion deblurring," IEEE Trans. on Pattern Analysis and Machine Intell., 6(26), 689-698, (2004).

[4] ?indelá?, O., and ?roubek, F., "Image deblurring in smartphone device using built-in inertial measurement sensors," Journal of Electronic Imaging, 22 (1), (2013).

[5] Joshi, N., Zitnick, K., and Szeliski, R., "Image deblurring using inertial measurement sensors," ACM SIGGRAPH, 29(4), 1-9, (2010).

[6] Razligh, Q., and Kehtarnavaz, N., "Image blur reduction for cell-phone cameras via adaptive tonal correction," Proc. of IEEE International Conference on Image Processing, (113-116), (2007).

[7] Chang, C-H., Parthasarthy, S., and Kehtarnavaz, N., "Comparison of two computationally feasible image de-blurring techniques for smartphones", Proc. of IEEE Global Conference on Signal and Information,743-746, (2013).

[8] Shen, H., and Huang, J., "Sparse principal component analysis via regularized low rank matrix approximation," J. Multivar. Anal., 99 (6), 1015–1034, (2008).

9400-8, Session 2

Real-time object tracking for moving target auto-focus in digital camera

Haike Guan, Norikatsu Niinami, Ricoh Co., Ltd. (Japan);
Tong Liu, Ricoh Software Research Ctr. (China)

Focusing at a quickly moving object accurately is difficult and important for taking photo of the target successfully in digital camera.

Because the object often moves randomly and changes its shape frequently, position and distance of the target should be estimated at real-time so as to focus at the objet precisely. We often fail to take the photo because focus is set at background or other object.

We propose a new method of real time object tracking to do auto-focus for moving target in digital camera. Video stream, which is used for monitoring, is applied for the moving target tracking.

Particle filter tracking is used to deal with problem of the target object's random movement and shape change. Color and edge features are used to measure the object's states. Parallel processing algorithm is developed to realize real-time particle filter object tracking easily in digital camera. Hardware is used efficiently to realize parallel real-time processing. Features of multiple particles are calculated parallel.

Movement prediction algorithm is also proposed to remove focus error caused by position difference between tracking results and target object's real position when the photo is taken. Because it takes some time to release shutter after we get tracking results, the target object moves to other place during the time lag. We use tracking results of former frames to estimate the target object movement by least square estimation.

Experiment results in digital camera demonstrate effectiveness of the proposed method.

We designed randomly moving target object to evaluate performance of our proposed real-time tracking method. The target object for tracking is set on a shaved stage. The stage is moved periodically by known frequency. Target moving speed, lighting condition and intensity change, target size, contrast change is also tested to show effectiveness of our method. Random motion object tracking of nature scene is also made to demonstrate robustness of our proposed method.

We embedded the parallel object tracking algorithm in digital camera. Position and distance of the moving target object is obtained accurately from the video stream. SIMD processor is applied to enforce parallel real-time processing. The parallel hardware can be used efficiently by our algorithm. Features of multiple particles are calculated at same time parallel. Limited small cache memory is also used effectively by our algorithm to increase processing speed. Several look-up tables are designed to make calculation of particle filter tracking more quickly without losing tracking accuracy. Processing time less than 60ms for each frame is obtained in the digital camera with CPU of only 162MHz.

9400-9, Session 2

Embedded wavelet-based face recognition under variable position

Pascal Cotret, Stéphane Chevobbe, Mehdi Darouich, Commissariat à l'Énergie Atomique (France)

Key challenge in embedded systems is to decrease the computing power and power consumption needed to execute complex algorithms in keeping the appropriate quality for the targeted applications. In image processing, and especially in face recognition using subspace learning technic such as eigenfaces [1], the computing power is directly proportional to the resolution of input images. Thus, by limiting the memory access and the memory size, the power consumption will decrease in a significant manner. We show in [2] that an efficient way to decrease the input image resolution is to apply a wavelet-based transformation such as the LeGall 5/3 wavelet on the input image. This method is robust against illumination variation. The quantity of input data is reduced by a factor 22K, where K is the level of the wavelet decomposition.

This work studies wavelet-based eigenfaces robustness under several level of decomposition against face position shifting in both horizontal and vertical directions; and in scale. In the literature, many methods have been studied to overcome the misalignment issue by reconstructing an aligned Region Of Interest (ROI) [3] or using correlation filters to propose robust face recognition method [4]. However, in a real-time embedded context, the limited processing power prevent us to consider such complex methods.

This work studies the impact of face position in relation to the snapshot used for learning purposes while Yale face database allows us to measure robustness against light variation as well.

For face position, a tolerance of +/- 10% of the ROI size is obtained with satisfying recognition rates (over 75%) and a high level of wavelet decomposition (data amount divided by 64).

In this context, this work shows implementation results of a wavelet-based face recognition method on a standard computer.

For the Yale face database B, using wavelet pre-processing gives a decrease of 90% in terms of computation time without damaging face recognition rates on both standard computers and embedded boards (providing less computational power).

For instance, recognition task on a set of 32 faces with light variations is processed in 211 seconds for a wavelet decomposition level K = 1 while this task is performed in around 22 seconds for K = 3 on a recent x86-based computer. These results allow us to implement a face recognition

application with smooth video feedback and a larger face database. These results also confirms the interest of wavelet transform for face recognition on ultra-embedded vision systems. It also allows us to consider the use of a low complexity face detection stage compliant with the limited processing power available on such systems.

REFERENCES

[1] M. Turk and A. Pentland, "Eigenfaces for recognition," J. Cognitive Neuroscience, vol. 3, no. 1, pp. 71–86, Jan. 1991. [Online]. Available: http://dx.doi.org/10.1162/jocn.1991.3.1.71

[2] S. Courroux, S. Chevobbe, M. Darouich, and M. Paindavoine, "Use of wavelet for image processing in smart cameras with low hardware resources," Journal of Systems Architecture - Embedded Systems Design, vol. 59, no. 10-A, pp. 826–832, 2013.

[3] S. Yan, H. Wang, J. Liu, X. Tang, and T. Huang, "Misalignment-robust face recognition," Image Processing, IEEE Transactions on, vol. 19, no. 4, pp. 1087–1096, April 2010.

[4] M. Sawides, B. Kumar, and P. Khosla, ""corefaces" - robust shift invariant pca based correlation filter for illumination tolerant face recognition," in Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, vol. 2, June 2004, pp. II–834–II–841 Vol.2.

9400-10, Session 2

Embedded application specific signal processor (ASSP) for real-time computation of 3D depth maps

Rajashekar Reddy Merugu, Ushakiran Thoyyeti, Alekhya Darsi, Venu Kandadai, Manjeera Digital Systems Pvt. Ltd. (India)

Need for implementation of several complex image and video processing algorithms on embedded mobile devices is steadily increasing. Examples include using facial recognition to unlock the Smartphone and other devices, auto-stitching of multiple images (panorama stitching), moving object detection, gesture recognition, 3D-depth mapping and so on. Most of these algorithms are computationally intensive and required massive data processing. Existing embedded hardware solutions such as GPUs, multi-core CPUs, coprocessor extensions such as SIMD and VLIW do not have adequate performance required to process such complex algorithms in real time.

Reconstructing a detailed depth-map from a pair of images of stereo cameras is of considerable interest due to the increasing number of applications, both in graphics and in vision. Depth map estimation algorithms based on multi camera acquisitions are in fact, time consuming, and need heavy compute resources. For example performing 3D- depth map computation for a 640?480 image in real time requires higher end graphics processor such as NVIDIA GEFORCE 8800 GTX with 128 cores. No solution on embedded devices is available as yet.

In this paper we present a middle stratum operation (MSO) based application specific signal processor (ASSP) based on data-plane architecture called Universal Multifunction Accelerator (UMA). MSOs include the mathematical complexity as well as data access in a higher level operation. MSOs consist of a combination of basic mathematical operations. UMA is vector processing engine of MSOs. UMA operates on local memory with a novel interface to memory such that vector operands for all types of MSOs are available. Since UMA is a pure data plane processor, it needs a control flow engine (CFE). Any simple microcontroller is adequate as a CFE. A multimedia ASSP can now be designed with UMA core as the computing engine. Such an ASSP, called Multimedia Data-plane Processor (MDP), includes multiple UMA cores, local memory (64KB per core) and a CFE. Any multimedia algorithm can be accelerated using MDP by porting its C-code on UMA. This provides very high acceleration and ultra low power consumption.

We have implemented 3D-depth map algorithm on MDP. MDP with eight UMA cores can perform 50 level disparity map computation (with a 7?7 window size) on 640?480 image with 8.73 Mega Cycles. Therefore MDP operating at 500 MHz can compute up to 57 fps at 640?480. GEFORCE 8800GTX can only perform under 48 fps. In addition MDP consumes only 248 mW for 30fps operation a fraction of the power 8800 GTX consumes. Details of MSO based computing, and implementation of 3D disparity map computation on UMA will be presented in this paper.

9400-11, Session 3

FIR filters for hardware-based real-time multi-band image blending

Vladan Popovic, Yusuf Leblebici, Ecole Polytechnique Fédérale de Lausanne (Switzerland)

The limited angle of view of modern cameras has created the need to combine two or more images into a single one, to increase the effective angle of view. The creation of panoramas or image mosaics has been a popular research topic over the past years. Instead of using a wide angle camera lens, e.g. fish-eye lens, which introduces visually perceivable distortions, it is preferable to use post-processing algorithms to create a photo-mosaic after taking the photographs.

A major issue in creating photo-mosaics resides in the fact that the original images do not have identical brightness levels. The problem manifests itself by the appearance of a visible seam at the position where the images overlap. The blending algorithms based on a weighted average between pixels in every image can reduce or even completely remove the seams. However, the drawback of a weighted average lies in high frequency blurring in the presence of any small image alignment error. A possible solution to this issue consists of using a multi-band blending (MBB) algorithm [1] where the high frequencies are combined separately, thus avoiding blurring.

The MBB requires Laplacian Pyramid (LP) decomposition and separate blending of the pyramid's levels. The main disadvantage of the above mentioned algorithms is a high required precision of the filter coefficients and long processing time. This results in a reduced processing frame rate, which is unsuitable for real-time processing systems. Recently, a systolic implementation of MBB using binomial filters [2] allowed real-time blending of very high resolution images, which was not possible with previous implementations [3].

In this paper, we create a set of FIR filters consisting of several custom-designed Butterworth and Chebyshev filters, 5/3 and 9/7 wavelets and already proposed Gaussian filters [1], [2]. We apply these filters to LP decomposition of the benchmark dataset, and analyze their influence on image quality and hardware complexity. The resulting photo-mosaics are compared using three different no-reference quality measures: blur [4], edge quality [5] and naturalness [6]. We show that 5/3 wavelet filter pair results in the best image quality, especially when edge quality is concerned.

An FPGA system implementing real-time 2D non-separable systolic filtering scheme is proposed, with timing results comparable to [2], i.e. approximately 95 fps for 1080p frame resolution. The system is implemented on Virtex-7 FPGA. Thanks to fully pipelined processing, the output frame rate is solely dependent on the achieved processing frequency and not on input load of the system. Additionally, the resulting image quality is independent of the system where the algorithm is being implemented on. Hence, the presented analysis and FPGA implementation is easily portable to other processing platforms, without any loss in the image quality.

The work presented in this paper allows further research and development in real-time image processing areas such as panoramic multi-camera systems, surveillance, object recognition and tracking, which are currently limited by either low image resolution or low frame rate.

REFERENCES

[1] M. Brown and D. Lowe, "Automatic panoramic image stitching using invariant features", IJCV, 74(1), 2007.

[2] V. Popovic et al., "Real-time hardware implementation of multi-resolution image blending", ICASSP, 2013.

[3] Y. Song et al., "Implementation of real-time Laplacian pyramid image fusion processing based on FPGA", Proceedings of SPIE, vol. 6833, 2007.

[4] P. Marziliano et al., "A no-reference perceptual blur metric", ICIP, 2002.

[5] C.S. Xydeas and V. Petrovic, "Objective image fusion performance measure," Electronics Letters, 36(4), 2000.

[6] A. Mittal, R. Soundararajan and A. C. Bovik, "Making a completely blind image quality analyzer", IEEE Signal processing letters, 22(3), 2013.

9400-12, Session 3

Iris unwrapping Using the Bresenham Circle Algorithm for Real-Time Iris Recognition

Matthew T. Carothers, Hau T. Ngo, Ryan N. Rakvic, Randy Broussard, U.S. Naval Academy (United States)

Iris recognition is one of the best biometric-based authentication methods available today and is mainly being implemented as a security measure. No two irises are equivalent, even in twins, making authentication of this unique data the keyhole and iris recognition the efficient key. The first step in an iris recognition system is to find and extract human irises in a static image or video frame. Most modern iris recognition systems segment the captured iris and unwrap its omnidirectional data into a matrix template more suitable for logical analysis. The template is then compared to others in a database to determine a match.

This paper presents a design and implementation of a parallel computational architecture for an efficient iris recognition algorithm. The proposed design is implemented with high performance Field Programmable Gate Array (FPGA) technology. The United States Naval Academy's Ridge-Energy Detection (RED) algorithm will be used in this design. The segmentation step of our iris recognition system has already been designed and implemented with FPGA technology. Once the iris boundaries and center have been found, the template creation and template matching steps will be carried out to determine a match. One of the most commonly used methods to unwrap an iris image into a matrix template is based on the Coordinate Rotation Digital Computer (CORDIC) technique. CORDIC is a method of calculating various trigonometric and transcendental functions by rotating vectors. One of many useful applications of the CORDIC algorithm is coordinate transformation. Specifically, polar coordinates are used to label the omnidirectional segmented iris image, which are then converted to Cartesian representations for processing. To do this, the CORDIC algorithm performs a series of simple shift-and-add operations. Although these operations are simple, the large number of required iterations for each transformation creates a drawback. This drawback is increased processing time and the need for a larger power supply. The computational requirement due to the complexity of the coordinate transformation step is too high to support real-time applications in a processor-based system. In this work, an unrolled pipelined and parallel architecture is designed to implement the CORDIC-based coordinate transformation module in order to increase the throughput of the system. This design also includes data enhancement through histogram equalization as well as interpolation to account for the uneven circumferences of the iris's inner and outer boundaries. A parallel XOR tree is designed to rapidly compute the hamming distance in the template matching module. The proposed design is targeted for a low-cost FPGA prototype board from Altera for a cost effective iris recognition system. The large databases that will be used to test the entire system contain pictures of human eyes captured in the near infrared spectrum. The entire system is implemented using VHDL and Altera development software tools. High performance computational modules are an FPGA-based embedded system using Altera's Avalon Streaming protocol. With an FPGA as the main integration platform our expectations are to set system processing time and overall power consumption to levels more efficient than current iris recognition systems released today.

9400-13, Session 3

Real-time joint deflickering and denoising for digital video

Zhimin Xu, Fan Zhang, Sijie Ren, Lenovo Group Ltd. (Hong Kong, China); Edmund Y. Lam, The Univ. of Hong Kong (Hong Kong, China)

Recent developments in image sensors have made cameras in smartphones approaching the abilities of human vision in terms of resolution and sensitivities. However, the miniaturization of image sensors causes tiny pixel size which eventually produces severe noise. On the other side, the use of rolling shutter in these inexpensive devices further degrades final imaging qualities, such as artifacts caused by flickering AC lighting fixtures. In this paper, we present a real-time solution to automatically detect the flicker type and noise level, and alleviate both flicker and sensor noise adaptively. In addition, the computational cost of our algorithm is relatively low in order to reduce its power consumption when implemented on smart devices.

Flicker and sensor noise are both assumed as additive artifacts, yet show different statistical correlations in spatio-temporal domain. Thus we propose an adaptive spatio-temporal filter to eliminate both of them. When the noise level is above a certain threshold and no flicker occurs, we will use a fast bilateral filter to smooth the pixels. Image registration is unnecessary in this case, since the photometric similarity term of bilateral filter can avoid the ghost effect in motion pictures. To accelerate the filter, sparse neighborhood pixels are sampled and the repaired frames are recursively cached, so that fewer neighbor pixels are involved in the computation.

When flicker happens, our approach merges the current frame with two consecutive frames after an efficient registration process via gradient feature matching. The best weighting for frame merging is achieved by analyzing the flicker type and its frequency. The registration is a relatively time-consuming and computationally intensive process, it means more power consuming compared to our denoising process. So we only use such operations for deflickering, although it can deal with noise to some degree.

To test our method, degraded videos are generated from a set of reference video clips. The sensor noise is simulated as an additive independent distributed Gaussian noise, while the flicker is approximated by sine wave horizontal stripes. Denoising methods are evaluated by the quality gain of the denoised video over the corresponding generated video. Our method achieves comparable PSNR gain and better SSIM gain, in comparison with BM3D.

To demonstrate the low computational expense of our approach, we have implemented our method with OpenGL ES on a Lenovo K900 smartphone, which is equipped with the Intel Atom Z2580 processor. We are able to display a denoised and deflickered viewfinder directly on the phone's screen. Our implementation runs at 30 fps (i.e., the camera's maximum frame rate).

9400-14, Session 3

Real-time object tracking using robust subspace learning in particle filter

Wen Lu, Institute of Optics and Electronics (China) and Univ. of the Chinese Academy of Sciences (China); Yuxing Wei, Institute of Optics and Electronics (China)

The main challenge in designing a robust object tracking algorithm is the real-time problem and the inevitable variation in the images of the tracked object's appearance over time. Various factors can be responsible for such variation, e.g., changes in the view point, changes in illumination, changes to the shape (deformations, articulations) or reflectance of the object, or partial occlusion of the target. Most existing algorithms are able to track objects, either previously viewed or not, in short durations and in well controlled environments. However these algorithms usually fail to observe the object motion or have significant drift after some period of time, due to drastic change in the tracked object's appearance or large lighting variation in its surroundings. Therefore, an important theme in object tracking research is the design of a flexible model or representation which can adapt to appearance changes. This paper proposes a robust and adaptive appearance model for tracking complex natural objects based on the subspace technique in real-time using DSP and FPGA hardware.

Observing that low-dimensional linear subspaces are able to effectively

model image variation caused by illumination and pose change. On the one hand, most related tracking algorithms using a pre-trained view-based eigenspace representation, it is imperative to collect a large set of training images covering the range of possible appearance variation, and this representation, once learned, is not updated. On the other hand, the conventional subspace learning, in the sense of least-squares approximation, is susceptible to outlying measurements. To address these two important issues, we present a robust subspace learning method that incrementally learns a low-dimensional subspace representation, which can efficiently adapt online to changes in the appearance of the target. Tracking then is led by the state inference within the framework in which a particle filter is used for propagating sample distributions over the time.

Conventional methods of computing eigenspace are singular value decomposition (SVD) and eigenvalue decomposition (EVD). But they are all performed in batch mode which computes the eigenspace using all the observations simultaneously. And the computational complexity is O (m*m*m) where m is the minimum value between the data dimension and the number of training images. To address this problem, in this paper, an incremental algorithm is used to continually update the subspace model. Incrementally updating the subspace does not use the offline learning phase required by other eigentrackers, allowing one to track objects for which a database of training images is not even available.

Besides, the methods of EVD and SVD, in the sense of least mean squared error minimization, are susceptible to outlying measurements. To build an eigenspace model which is robust to “outliers”, we wish the pixels of the acquired images receive different treatment and all the observations have different influence on the estimation of subspace. To enable a selective influence of individual images and pixels, in this paper, the method of EVD is generalized into a weighted approach, which considers individual pixels and images diversely, depending on the corresponding weights.

Experimental results on tracking visual objects in video sequences where the targets undergo large changes in pose, occlusion and illumination, demonstrate the real-time and robustness of our tracking algorithm.

9400-15, Session PTues

Efficient fast thumbnail extraction algorithm for HEVC

Wonjin Lee, Hanyang Univ. (Korea, Republic of); Gwanggil Jeon, Univ. of Incheon (Korea, Republic of); Jechang Jeong, Hanyang Univ. (Korea, Republic of)

< Proposed algorithm>

The proposed algorithm produces an average value per 4?4 size for thumbnail extraction in HEVC. The average value for one thumbnail pixel per 4?4 size is generated with the use of reference pixels. Thus, the average value is defined as summation of sample values which are residual DC value of DCT domain of TU (Transform Unit) and average value of prediction block of PU. The residual DC value can be directly extracted because the residual DC value is located to (0, 0) position of DCT domain. The prediction block is generated by using the upper and left reference pixels which are boundary pixels of previously reconstructed blocks according to intra prediction modes. The average value of prediction block can be calculated by using only reference pixels. The proposed algorithm uses the predefined equations according intra prediction mode.

The current intra prediction reconstructs all pixels in a block even if whole pixels in a reconstructed block do not need to generate the thumbnail image. Therefore, in order to efficiently reconstruct pixels needed for generating the prediction block, we propose the partial intra prediction which reconstructs only 4?4 boundaries.

The residual block of TU is calculated by IDCT. The computational complexity of IDCT also reduced by calculating only 4?4 boundaries. For example, when TU and PU (Prediction Unit) are 8?8, the partial intra prediction generates only 4?4 boundary samples (28 samples) and the partial IDCT is performed by using the two partial DCT basis which are 8?1 matrices of 4th and 8th row of the 8x8 DCT basis.

< Experimental results>

The proposed algorithm was compared to the conventional method which is sub-sampled by 1/4 of image reconstructed by the HEVC reference decoder. The proposed method was tested on HEVC video reference software (HM 13.0). All test sequences for HEVC (Class A: 2560?1600, Class B: 1920?1080, Class C: 832?480, Class D: 416?240, Class E: 1280?720, Class F: 832?480, 1024?768, 1280?720) were used in the experiments. The proposed algorithm evaluated the consumed thumbnail extraction time compared to conventional method. The thumbnail extraction time of proposed method is faster than conventional one in all sequences from 17.28% to 30.10%.

Subjective qualities of both thumbnail images were very similar although thumbnail extraction time was significantly reduced.

9400-17, Session PTues

Parallel hybrid algorithm for solution in electrical impedance equation

Volodymyr Ponomaryov, Marco Robles-Gonzalez, Ari Bucio-Ramirez, Marco Ramirez-Tachiquin, Instituto Politécnico Nacional (Mexico); Eduardo Ramos-Diaz, Univ. Autónoma de la Ciudad de Mexico (Mexico)

This study is inspired in medical imaging technique that it can be useful to control and monitor diseases that impacted in human’s health, such like tumors, chronic and degenerative ills and cancer. This technique is known as Electrical Impedance Tomography procedure that is not considered as a medical imaging analysis due to the equation that governs this problem.

A novel framework that can obtain the solution for Electrical Impedance Equation (EIT) is presented. Many experts in the field considered the EIT as ill-posed problem because of the stability in solution of this equation, in which the result depends upon the initial data used; thus, any variation in the initial data affects the approximation inside the domain with boundary.

In this paper, we present a novel scheme that includes the analysis of the inverse problem using the Finite Element Method (FEM) in order to obtain the conductivity within a unit circle domain, following the estimated conductivity is introduced in the algorithm via a technique of the Pseudoanalytic Function Theory where the Taylor series in formal powers method is used, performing the solution in EIT for electric potential in the boundary.

The convergence and stability of this ill posed problem is also analyzed in this work using proposed hybrid procedure based on regularization to increase the convergence performance. Numerous experimental results for estimated electric potential by means of resolving together inverse and direct EIT problems for different types of conductivity distributions are exposed in this study justifying good performance of novel method. These results are performed using the FEM and numerical approximation of the Taylor series in formal powers approach developed in the structural programming paradigm in C++. An analysis of the Taylor series in formal powers employing a parallel programming paradigm in C++ is included with the objective to modify the algorithm and to develop a possible real-time mode in computing the solution with proposed method. This is performed with help of hardware based on multi-core processors and GPU platform.

9400-18, Session PTues

Fast-coding robust motion estimation model in a GPU

Carlos Garcia, Guillermo Botella, Univ. Complutense de Madrid (Spain); Francisco de Sande, Univ. de Las Palmas de Gran Canaria (Spain); Manuel Prieto-Matias, Univ. Complutense de Madrid (Spain)

The results observed using the convolution operation as a case study determined that the use of accelerators is an efficient solution not only in terms of performance but also in power consumption (watts/fps). The

use of this type of technology will loosely satisfy real-time requirements, since the system is able to process up to 190 frames per second for a resolution of 1024x1024. Taken into account the promising results observed with the convolution in a graphics accelerator, our proposal deal with an implementation of a neuromorphic motion estimation algorithm to estimate motion using OpenACC programming paradigm. We hope that performance results observed would move to full model because it's based on multiple convolution operations in temporal and spatial filtering or steering stages. A deep study and evaluation of a full model would be done using this new programming paradigm for GPUs if the paper is finally accepted.

Next items summarized the main tasks to be performed:

• Implementation of the full model using OpenACC paradigm.

• Speedups evaluation in all the stages involved in the robust motion estimation model using CAPS and PGI compilers.

• Extended evaluation into other architectures that support OpenACC as Intel's Xeon-Phi or AMD-APU technology.

9400-19, Session PTues

Real-time single-exposure ROI-driven HDR adaptation based on focal-plane reconfiguration

Jorge Fernández-Berni, Ricardo A. Carmona-Galán, Rocío del Río, Instituto de Microelectrónica de Sevilla (Spain); Richard Kleihorst, Wilfried Philips, Univ. Gent (Belgium); Ángel B. Rodríguez-Vázquez, Instituto de Microelectrónica de Sevilla (Spain)

When it comes to extracting meaningful information from a scene, vision algorithms have to cope with varying illumination conditions taking place at both intra-frame and inter-frame levels. Without a strategy to address this issue, important details can be missed due to saturation in over-exposed regions or noise and lack of contrast in under-exposed regions. The most usual technique to prevent this from happening consists of taking multiple captures per frame with different exposure periods. The resulting images are combined into a single one featuring a much wider intra-frame dynamic range. Inter-frame changes of illumination are accommodated by adapting the range of exposure periods correspondingly. Unfortunately, this technique creates artifacts if motion happens to occur during multi-exposure. These artifacts can in turn mislead the scene analysis performed by the vision algorithm. Other approaches at sensor chip level have been proposed that do not require multiple captures: well capacity adjustment, time-to-saturation measurement, logarithmic pixel response etc. Generally speaking, all these reported techniques targeting High Dynamic Range (HDR) deal globally with the image content. In other words, there is no special consideration for specific regions in the process of adjusting the capture according to the illumination conditions. However, vision algorithms usually focus their attention in particular so-called Regions Of Interest (ROI). Once a certain ROI is spotted, the algorithm tracks it across the scene while carrying out prescribed analytics. This tracking and corresponding analytics should not be affected by variations in the illumination over the ROI. Indeed, the priority should be to adapt the capture for that ROI while ensuring that new ROIs can still be detected in case they enter the scene.

All in all, this paper describes the operation of a prototype QVGA smart imager capable of adjusting the photo-integration time of multiple ROIs concurrently, automatically and asynchronously with a single exposure period. Each pixel of this prototype incorporates two photo-diodes. One of them senses the pixel value itself whereas the other, in collaboration with its counterparts in a particular ROI, senses the mean illumination of that ROI. Additional circuitry interconnecting both photo-diodes enables the automatic and asynchronous adjustment of the integration time for each ROI. Since the pixel matrix works in Single-Instruction Multiple-Data (SIMD) mode, the operation takes place in parallel for all the ROIs. The sensor can be reconfigured on-the-fly according to the requirements of a vision algorithm. To this end, the array of pixels is surrounded by peripheral registers that can be modified on a frame basis. These registers encode the pixel interconnection patterns that set the different ROIs. An illustrating sequence composed of images directly extracted from the chip can be downloaded from http://www.imse-cnm.csic.es/mondego/RTIVP15/. In this example, the integration time is initially determined by the global mean illumination of the scene. This leads to a noisy capture of poorly illuminated regions as well as to saturated pixels in regions featuring very high illumination. At a certain point, the ROI-driven HDR adaptation is activated. The algorithm conducting the focal-plane reconfiguration is the Viola-Jones face detector provided by OpenCV.

9400-20, Session PTues

Edge pattern analysis on GPU

Bo Jiang, Guangzhou Institute of Biomedicine and Health (China)

Edge detection serves as the basic transformation of signals into symbols and it enables and heavily influences the performance of subsequent higher level pattern processing. Edge detection is defined as the process of detecting and representing the presence of and locations of image of discontinuities in the two-dimensional image signal distributions. Depending upon the final application, different algorithms are proposed to detect edges in images. Those algorithms are designed differently for their own specific purpose. But, in general, the edge detection process has two main steps: filtering, and detection and localization.

The most commonly used methods in the second step of edge detection compute edges from the derivative of the intensity values. This localizes edges at pixels where intensity transitions occur, such as different order derivative operators, e.g. Roberts, Sobel, Prewitt, Laplacian, etc. However, an edge is formed by a group of pixels, whose characteristics are not solely determined by a single pixel value. But, generally, one specific detector such as step, ramp, or stair edge detector suited for one type of edge patterns is not applicable in other edge patterns. In this paper, a novel edge detection and localization algorithm based on the analysis of basic edge patterns is developed to achieve this goal. The ramp, the impulse, the step and the sigmoid (RISS) edge patterns are chosen as basic edge profiles, for they can represent other edge profiles in ideal or non-ideal images. Experimental results on real images support that the proposed edge detection algorithm with edge pattern analysis is effective in enhancing the accuracy of edge detection and localization, even in noisy conditions.

As the edge pattern expands from step pattern to a series of edge pattern, it becomes a generic algorithm successfully used for a wide variety of imagery. Therefore, its real-time application has far-reaching implications. However, as the potential utility and complexity of the algorithm expands, so do the computational requirements of the algorithm. The Graphics Processor Unit (GPU) is one high performance computing platform, which offers highly-parallel computation and a flexible, programmable environment with relatively low cost, and has recently been applied to many applications. In this paper, we implement the proposed algorithm on a CUDA-enabled GPU. For the various combinations of configurations in our test, the GPU accelerated RISS algorithm shows a scalable speedup as the resolution of an image increases. We achieve 20 frames per second in real-time processing (640X480), which is the acceptable frame rate for many diverse applications. Also, to achieve better performance on the GPU, the execution configuration optimizations are demonstrated to discuss the trade-off among its features associated to its compute capability.

9400-21, Session PTues

Task-oriented quality assessment and adaptation in real-time mission-critical video streaming applications

James M. Nightingale, Qi Wang, Christos Grecos, Univ. of the West of Scotland (United Kingdom)

In recent years video traffic has become the dominant application on the

Internet with global year-on-year increases in video-oriented consumer services. Driven by improved bandwidth in both mobile and fixed networks, steadily reducing hardware costs and the development of new technologies, many existing and new classes of commercial and industrial video applications are now being upgraded or emerging. Some of the use cases for these applications include areas such as public and private security monitoring for loss prevention or intruder detection, industrial process monitoring and critical infrastructure monitoring. The use of video is becoming commonplace in defence, security, commercial, industrial, educational and health contexts.

Towards optimal performances, the design or optimisation in each of these applications should be context aware and task oriented with the characteristics of the video stream (frame rate, spatial resolution, bandwidth etc.) chosen to match the use case requirements. For example, in the security domain, a task-oriented consideration may be that higher resolution video would be required to identify an intruder than to simply detect his presence. Whilst in the same case, contextual factors such as the requirement to transmit over a resource-limited wireless link, may impose constraints on the selection of optimum task-oriented parameters.

This paper presents a novel, conceptually simple and easily implemented method of assessing video quality relative to its suitability for a particular task and dynamically adapting videos streams during transmission to ensure that the task can be successfully completed. Firstly we defined two principle classes of tasks: recognition tasks and event detection tasks. These task classes are further subdivided into a set of task-related profiles, each of which is associated with a set of task-oriented attributes (minimum spatial resolution, minimum frame rate etc.). For example, in the detection class, profiles for intruder detection will require different temporal characteristics (frame rate) from those used for detection of high motion objects such as vehicles or aircrafts. We also define a set of contextual attributes that are associated with each instance of a running application that include resource constraints imposed by the transmission system employed and the hardware platforms used as source and destination of the video stream. Empirical results are presented and analysed to demonstrate the advantages of the proposed schemes.

9400-22, Session PTues

A simulator tool set for evaluating HEVC/SHVC streaming

James M. Nightingale, Tawfik A. Al Hadhrami, Qi Wang, Christos Grecos, Univ. of the West of Scotland (United Kingdom); Nasser Kehtarnavaz, The Univ. of Texas at Dallas (United States)

Video streaming and other multimedia applications account for an ever-increasing proportion of all network traffic. The recent adoption of High Efficiency Video Coding (HEVC) as the H.265 standard provides many opportunities for new and improved multimedia services and applications in the consumer domain. Since the delivery of version one of H.265, the Joint Collaborative Team on Video Coding have been working towards standardisation of a scalable extension (SHVC) to the H.265 standard and a series of range extensions and new profiles. As these enhancements are added to the standard the range of potential applications and research opportunities will expand. For example the use of video is also growing rapidly in other sectors such as safety, security, defence and health with real-time high quality video transmission playing an important role in areas like critical infrastructure monitoring and disaster management, each of which may benefit from the application of enhanced HEVC/H.265 and SHVC capabilities.

The majority of existing research into HEVC/H.265 transmission has focussed on the consumer domain addressing issues such as broadcast transmission and delivery to mobile devices with the lack of freely available tools widely cited as an obstacle to conducting this type of research. In this paper we present a toolset that facilitates the transmission and evaluation of HEVC/H.265 and SHVC encoded video on an open source emulator. Our toolset provides researchers with a modular, easy to use platform for evaluating video transmission and adaptation proposals on large-scale wired, wireless and hybrid architectures. The toolset consists of pre-processing, transmission, SHVC adaptation and post-processing tools to gather and analyse statistics. It has been implemented using HM15 and SHM5, the latest versions of the HEVC and SHVC reference software implementations to ensure that currently adopted proposals for scalable and range extensions to the standard can be investigated.

We demonstrate the effectiveness and usability of our toolset by evaluating SHVC streaming and adaptation to meet terminal constraints and network conditions in a range of wired, wireless, and wireless mesh network scenarios. Our results are compared with those for H264/SVC, the scalable extension to the existing H.264/AVC advanced video coding standard. The proposed toolset would significantly facilitate further research in delivering adaptive video streaming based on the latest video coding standard over wireless mesh and other ad hoc networks.

9400-23, Session PTues

Dynamic resource allocation engine for cloud-based real-time video transcoding in mobile cloud computing environments

Adedayo A. Bada, Jose Alcaraz-Calero, Qi Wang, Christos Grecos, Univ. of the West of Scotland (United Kingdom)

The recent explosion in video-related Internet traffic has been driven by the widespread use of smart mobile devices, particularly smartphones with advanced cameras able to record high-quality videos. Although many of these devices offer the facility to record videos at different spatial and temporal resolutions, primarily with local storage considerations in mind, most users only ever use the highest quality settings. The vast majority of these devices are optimised for compressing the acquired video using a single built-in codec and have neither the computational resources nor battery reserves to transcode the video to alternative formats. Given the number of potential distribution channels for video content (e.g. satellite transmission, social media, adaptive Internet streaming, broadcast SD and HD TV, etc.) and the varying specifications of the end users' devices, there is a clear need to be able to dynamically provide the same content in a number of different formats for adaptation purposes.

These observed limitations of mobile devices can be overcome by offloading computationally intensive tasks to cloud-based services. However, real-time intelligent cloud resource allocation schemes for different transcoding jobs with varying resource requirements is challenging and require further research and development. As an on-demand service, the utilization of the vast resources available in the cloud comes at minimal cost to the user, and dynamically allocating resources as per requirement will go a long way in reducing cost incurred by the mobile devices. It not only reduces battery consumption but also relieves the device of any computational intensive task.

This paper proposes a new low-complexity dynamic resource allocation engine for cloud-based video transcoding services that are both scalable and capable of being delivered in real-time. Firstly, through extensive experimentation, we establish resource requirement benchmarks for a wide range of transcoding tasks. The set of tasks investigated covers the most widely used input formats (encoder type, resolution, amount of motion and frame rate) associated with mobile devices and the most popular output formats derived from a comprehensive set of use cases, e.g. a mobile news reporter directly transmitting to the TV audience, with minimal usage of resources both at the reporter's end and of the cloud infrastructure. These benchmark results are then exploited to create a reference database with predefined combinations of inputs and outputs, which is a product of rigorous experiments in defining the conversion process at an optimized level. The proposed cloud-based video transcoding process is platform-independent and codec-agnostic; it is a function of the input and expected video output, dynamically determining the most efficient method for its transcoding based on the reference database. Moreover, this database is automatically extensible by learning the requirements of new types of input tasks, and thus the resource allocation engine is able to adapt to the emergence of a broader range of transcoding tasks. This paper presents and implements the required resource-efficient schemes that provide the most

effective usage of the cloud with respect to real-time transcoding tasks, and demonstrates the effectiveness of the proposed schemes with experimental results

9400-24, Session PTues

Subjective evaluation of H.265/HEVC based dynamic adaptive video streaming over HTTP (HEVC-DASH)

Iheanyi C. Irondi, Qi Wang, Christos Grecos, Univ. of the West of Scotland (United Kingdom)

With the surge in Internet video traffic, real-time HTTP streaming of video has become increasingly popular. Especially, applications based on the MPEG Dynamic Adaptive Streaming over HTTP standard (DASH) are emerging for adaptive Internet steaming in response to the unstable network conditions. Integration of DASH streaming technique with the new video coding standard H.265/HEVC is a promising area of research in light of the new codec's promise of substantially reducing the bandwidth requirement. The performance of such HEVC-DASH systems has been recently evaluated using objective metrics such as PSNR by the authors and a few other researchers. Such objective evaluation is mainly focused on the spatial fidelity of the pictures whilst the impact of temporal impairments incurred by the nature of reliable TCP communications is also noted. Meanwhile, subjective evaluation of the video quality in the HEVC-DASH streaming system is able to capture the perceived video quality of end users, and is a new area when compared with the counterpart subjective studies for current streaming systems based on H.264-DASH. Such subjective evaluation results will shed more light on the Quality of Experience (QoE) of users and overall performance of the system. Moreover, any correlation between the QoE results and objective performance metrics will help designers in optimizing system performance.

This paper presents a subjective evaluation of the QoE of a HEVC-DASH system implemented in a hardware testbed. Previous studies in this area have focused on using the current H.264/AVC or SVC codecs and have hardly considered the effect of Wide Area Network (WAN) characteristics. Moreover, there is no established standard test procedure for the subjective evaluation of DASH adaptive streaming. In this paper, we define a test plan for HEVC-DASH with a carefully justified data set taking into account longer video sequences that would be sufficient to demonstrate the bitrate switching operations in response to various network condition patterns. The testbed consists of real-world servers (web server and HEVC-DASH server), a WAN emulator and a real-world HEVC-DASH client. We evaluate the QoE by investigating the perceived impact of various network conditions such as different packet loss rates and fluctuating bandwidth, and the perceived quality of using different DASH video stream segment sizes on a video streaming session and using different video content types. Furthermore, we demonstrate the temporal structure and impairments as identified by previous objective quality metrics and capture how they are perceived by the subjects. The Mean Opinion Score (MOS) is employed and a beyond MOS evaluation method is designed based on a questionnaire that gives more insight into the performance of the system and the expectation of the users. Finally, we explore the correlation between the MOS and the objective metrics and hence establish optimal HEVC-DASH operating conditions for different video streaming scenarios under various network conditions.

9400-25, Session PTues

Impact of different cloud deployments on real-time video applications for mobile video cloud users

Kashif A. Khan, Qi Wang, Chunbo Luo, Xinheng Wang, Christos Grecos, Univ. of the West of Scotland (United Kingdom)

The latest trend to access mobile cloud services through wireless network connectivity has amplified globally among both entrepreneurs and home end users. Although existing public cloud service vendors such as Google, Microsoft Azure etc. are providing on-demand cloud services with affordable cost for mobile users, there are still a number of challenges to achieve high-quality mobile cloud based video applications, especially due to the bandwidth-constrained and error-prone mobile network connectivity, which is the communication bottleneck for end-to-end video delivery. In addition, existing accessible clouds networking architectures are different in term of their implementation, services, resources, storage, pricing, support and so on, and these differences have varied impact on the performance of cloud-based real-time video applications. Nevertheless, these challenges and impacts have not been thoroughly investigated in the literature.

In our previous work, we have implemented a mobile cloud network model that integrates localized and decentralized cloudlets (mini-clouds) and wireless mesh networks. In this paper, we deploy a real-time framework consisting of various existing Internet cloud networking architectures (Google Cloud, Microsoft Azure and Eucalyptus Cloud) and a cloudlet based on Ubuntu Enterprise Cloud over wireless mesh networking technology for mobile cloud end users. It is noted that the increasing trend to access real-time video streaming over HTTP/HTTPS is gaining popularity among both research and industrial communities to leverage the existing web services and HTTP infrastructure in the Internet. To study the performance under different deployments using different public and private cloud service providers, we employ real-time video streaming over the HTTP/HTTPS standard, and conduct experimental evaluation and in-depth comparative analysis of the impact of different deployments on the quality of service for mobile video cloud users. Empirical results are presented and discussed to quantify and explain the different impacts resulted from various cloud deployments, video application and wireless/mobile network setting, and user mobility. Additionally, this paper analyses the advantages, disadvantages, limitations and optimization techniques in various cloud networking deployments, in particular the cloudlet approach compared with the Internet cloud approach, with recommendations of optimized deployments highlighted. Finally, federated clouds and inter-cloud collaboration challenges and opportunities are discussed in the context of supporting real-time video applications for mobile users.

9400-26, Session PTues

Improving wavelet denoising based on an in-depth analysis of the camera color processing

Tamara N. Seybold, Arnold & Richter Cine Technik GmbH & Co. Betriebs KG (Germany); Mathias Plichta, Walter Stechele, Technische Univ. München (Germany)

While Denoising is an extensively studied task in signal processing research, most denoising methods are designed and evaluated using readily processed image data, e. g. the well-known Kodak data set. The noise model is usually additive white Gaussian noise (AWGN). This kind of test data does not correspond to nowadays real-world image data taken with a digital camera. Using such unrealistic data to test, optimize and compare denoising algorithms may lead to incorrect parameter tuning or suboptimal choices in research on real-time camera denoising algorithms.

To understand the difference, let's review the color image capture via a digital camera, which is the usual way of image capture nowadays. One pixel captures the light intensity, thus the sensor data corresponds linearly to the lightness at the pixel position. To capture color data, a color filter array (CFA) is used, which covers the pixels with a filter layer. Thus the output of the sensor is a value that represents the light intensity for one color band at one pixel position. This data cannot be displayed before further steps are applied. These steps are the white balance, the demosaicking, which leads to a full color image, and the color transformations. These color transformations adapt the linear data, which is linear to the brightness, to displayable monitor data, which is adapted to the monitor gamma and color space. The transformation is usually highly nonlinear.

These steps lead to a noise characteristic that is fundamentally different

from the usually assumed AWGN: through demosaicking it is spatially and chromatically correlated and through the nonlinear color transformations the noise distribution is unknown. As most denoising algorithms – eg. wavelet hard thresholding – include the noise variance to adjust the denoising strength, the complicated camera noise characteristic poses a problem when standard algorithms shall be used in real camera data processing.

To adapt the standard methods for real camera noise, the noise characteristics of the camera need to be included. In this paper we therefore first derive a precise analysis of the noise characteristics for the different steps in the color processing. Based on real camera noise measurements and on the simulation of the camera processing steps, we obtain a good approximation for the noise variance. We show how this approximation can be used in standard wavelet denoising methods: We first investigate hard thresholding, with a threshold that is calculated based on our approximation, and additionally show how the bivariate thresholding formula can be improved. The computational complexity of our method is very low, as we include the approximation results using a look-up-table (LUT). This LUT is only generated once and then can be used for denoising the complete image sequence. Our implementation can run real-time for HD video sequences in an FPGA.

Our results show the both the hard thresholding and the bivariate wavelet denoising method can be improved. Using real camera data as well as simulated data we show the advantage in visual quality and additionally we calculate image quality metrics for both the standard and the improved methods. We conclude that our noise analysis improves the denoising methods significantly while the computational complexity is almost equal to the standard method.

9400-27, Session PTues

Impulsive noise suppression in color images based on the geodesic digital paths

Bogdan Smolka, Silesian Univ. of Technology (Poland); Boguslaw Cyganek, AGH Univ. of Science and Technology (Poland)

Noise reduction in color digital images is, despite many years of research, still a current and active topic of low-level image processing. The proliferation of inexpensive image-capturing devices, combined with the miniaturization of the optical systems and the increase of sensors' resolution, causes that the need for fast and efficient denoising algorithms is still not satisfied.

In the paper a novel filtering design based on the concept of exploration of the pixel neighborhood by digital paths is presented. The paths start from the boundary of a filtering window and reach its central element. The cost or penalty of a transition between adjacent pixels is defined in the hybrid spatial-color space. Thus, an optimal path of minimum total cost, leading from pixels of the window's boundary to its center can be determined. To decrease the computational complexity, the two-pass chamfer distance algorithm has been applied to calculate the total cost values.

The cost of an optimal path determines the degree of similarity of the central pixel to the samples from the local processing window. Obviously, if the central pixel of the sliding window is an outlier, then all the paths leading from the boundary, will have high costs and the minimal one will also be high. In the case of a cluster of a few similar pixels injected into the image by the impulsive noise process, the minimal cost will be high as well, as the pixels will form a small 'island' with steep 'cliffs', which has to be climbed from the surrounding 'sea' of pixels.

The straightforward scheme for the removal of the noisy pixels, would be to construct a switching filter based on the minimal cost value. However, such a design is very sensitive to the thresholding parameter and therefore more robust soft-switching design has been implemented, which substantially reduces the artifacts caused by the erroneous classification of pixels as noisy or not disturbed. In this way, the filter output is calculated as a weighted mean of the central pixel and an estimate constructed using the information on the minimal costs assigned to each image pixel. So, first the cost of an optimal path is used to build an interpolated, smoothed image and in the second step the minimal cost is utilized for constructing the weight of the soft-switching design. Thus, two parameters are needed, however the experiments revealed that they are not sensitive to image structures and noise intensity. Therefore, no additional, usually computationally expensive, adaptive designs are needed.

The experiments performed on a set of standard color images, prove that the efficiency of the proposed algorithm is superior to the state-of-the-art denoising techniques in terms of the objective restoration quality measures, both for low and high noise contamination ratios. The proposed filter, due to its low computational complexity can be applied for real time image denoising and also for the enhancement of video streams.

9400-28, Session PTues

Optimal camera exposure for video surveillance systems by predictive control of shutter speed, aperture, and gain

Juan Torres, Jose Manuel Menendez, Univ. Politécnica de Madrid (Spain)

This paper establishes a real–time auto-exposure method to guarantee that surveillance cameras in uncontrolled light conditions take advantage of their whole dynamic range while provide neither under nor overexposed images. State–of–the–art auto-exposure methods base their control on the brightness of the image measured in a limited region where the foreground objects are mostly located. Unlike these methods, the proposed algorithm establishes a set of indicators based on the image histogram that defines its shape and position. Furthermore, the location of the objects to be inspected is likely unknown in surveillance applications. Thus, the whole image is monitored in this approach. To control the camera settings, we defined a parameters function (F) that linearly depends on the exposure time and the electronic gain; and it is inversely proportional to the square of the lens aperture diameter. When the current acquired image is not overexposed, our algorithm computes the value of F that would move the histogram to the maximum value that does not overexpose the capture. When the current acquired image is overexposed, it computes the value of F that would move the histogram to a value that does not underexpose the capture and remains close to the overexposed region. If the image is under and overexposed, the whole dynamic range of the camera is therefore used, and a default value of the F that does not overexpose the capture is selected. This decision follows the idea that to get underexposed images is better than to get overexposed ones, because the noise produced in the lower regions of the histogram can be removed in a post-processing step while the saturated pixels of the higher regions cannot be recovered.

The proposed algorithm was tested in a video surveillance camera placed at an outdoor parking lot surrounded by buildings and trees which produce moving shadows in the ground. During the daytime of ten days, the algorithm was running alternatively together with a representative auto-exposure algorithm in the recent literature. Besides the sunrises and the nightfalls, multiple weather conditions occurred, which produced light changes in the scene: sunny hours that produced sharpen shadows and highlights; cloud coverages that softened the shadows; and cloudy and rainy hours that dimmed the scene. Several indicators were used to measure the performance of the algorithms. They provided the objective quality as regards: the time that the algorithms recover from an under or over exposure, the brightness stability, and the change related to the optimal exposure. The results demonstrated that our algorithm reacts faster to all the light changes than the selected state–of–the–art algorithm. It is also capable of acquiring well exposed images and maintaining the brightness stable during more time. Summing up the results, we concluded that the proposed algorithm provides a fast and stable auto-exposure method that maintains an optimal exposure for video surveillance applications. Future work will involve the evaluation of this algorithm in robotics.

Conference 9400:
Real-Time Image and Video Processing 2015

9400-29, Session PTues

Real-time object recognition in multidimensional images based on joined extended structural tensor and higher-order tensor decomposition methods

Boguslaw Cyganek, AGH Univ. of Science and Technology (Poland); Bogdan Smolka, Silesian Univ. of Technology (Poland)

In this paper we propose and analyze properties of the real-time object recognition system operating with the multidimensional vision signals. The main novelty is a proposition of joining the extended structural tensor with the Higher-Order Singular Value Decomposition framework for object recognition in tensor subspaces [3-4]. Object recognition is done by pattern projection into the tensor subspaces obtained from the factorization of the signal tensors representing the input video signal [1]. However, instead of taking only the intensity signal the novelty of this paper is to first build the extended structural tensor representation from the image signal. The extended structural tensor extracts information on dominating structures in the local areas around each pixel based on color values, their derivatives, as well as their mutual products [5]. This way the higher order pattern tensors are built from the training samples that convey more discriminative information on image content. These are then decomposed with the Higher-Order Singular Value Decomposition into the pattern tensor subspaces [8]. Finally, recognition relies on measurements of the distance of the test pattern projected into the tensor subspaces obtained from the training tensors. Due to high-dimensionality of the input data, tensor based methods require high memory and computational resources. However, recent achievements in the technology of the multi-core microprocessors and graphic cards make possible real-time operation on the multidimensional signals [7]. Proposition of implementations which allow real-time processing of multidimensional signals constitute the second novelty of this paper. It is shown and analyzed in this paper with examples of action recognition in the video signals [6].

1. Cyganek, B. Object Detection in Digital Images. Theory and Practice, Wiley (2013)

2. Kolda, T.G., Bader, B.W. Tensor Decompositions and Applications. SIAM Review, Vol. 51, No. 3, 455-500 (2008)

3. Lathauwer de, L. Signal Processing Based on Multilinear Algebra. PhD dissertation, Katholieke Universiteit Leuve (1997)

4. Lathauwer de, L., Moor de, B., Vandewalle, J. A Multilinear Singular Value Decomposition. SIAM Journal of Matrix Analysis and Applications, Vol. 21, No. 4, 1253-1278 (2000)

5. Luis-Garc´ıa R., Deriche R., Rousson M., Alberola-L´opez C.: Tensor Processing for Texture and Colour Segmentation. Lecture Notes in Computer Science, Vol. 3540, pp. 1117–1127, 2005.

6. Nefian, A. V.: Embedded Bayesian networks for face recognition, IEEE International Conference on Multimedia and Expo, August (2002)

7. OpenMP. www.openmp.org (2013)

8. Savas, B., Eldén, L. Handwritten Digit Classification Using Higher Order Singular Value. Pattern Recognition, Vol. 40, No. 3, 993-1003 (2007)

9400-30, Session PTues

A near infrared real-time video retrieval projection system based on Da Vinci platform and DMD

Aly Ahmed A. Khalifa, Hussein A. Aly, Military Technical College (Egypt)

Design and implementation of a near-infrared video projection system based on Davinci Digital Signal Processor (DSP) platform and Digital Micromirror Device (DMD) is introduced in this paper. And also describes the block diagram, hardware structure and video data processing algorithms of real-time video retrieval in detail. This system take two numbers (start frame, end frame) via serial port then it decode a part of pre-captured video indexed by start frame and end frame using DM6446EVM kit then up-sampling this video by THS8200 daughter card then project it by D4100 kit with DMD in infrared region. We also show the incompatibility between the DM6446EVM kit and the D4100 kit and how we develop a solution for that incompatibility.

9400-31, Session PTues

Near real-time operation of public image database for ground vehicle navigation

Ehsan A. Ali, Samuel Kozaitis, Florida Institute of Technology (United States)

A successful night navigation system can help drivers of ground vehicles see clearly in a degraded visual environment due to darkness. Such a system would increase safety and would also be useful for security. An important consideration for such a system is that it operates in real-time.

Our system was designed for terrestrial vehicle navigation in dark conditions and used color information from a public database of images to assign color information to an intermediate image. Initially, an image was obtained by combining two spectral bands of images, thermal and visible, in an effort to enhance night vision imagery. However, the fused image gave an unnatural color appearance. Therefore, a color transfer based on look-up table (LUT) was used to replace the false color appearance with a colormap derived from a daytime reference image. We obtained the reference image from a public database using the GPS coordinates of the vehicle. Using this approach, we were able to produce imagery acquired at night that appeared as if in the daylight.

We found that database images used for colorization had to be supplied using an automated procedure often on the order of seconds or minutes depending on how much the environment changes to allow the system to operate in real-time.

Other than the public database, the entire system can be implemented in a stand-alone fashion. The system has four inputs and two outputs. Two of the inputs are from cameras, thermal and visible, one input is from a GPS sensor, and the other input receives images from a public database. The two outputs consist of a connection to a public database that sends requests for images and the display of the final result of the night scene. Such an architecture can form the basis of other systems related to vehicle navigation with enhanced imagery.

9400-32, Session PTues

Efficient FPGA-based design of hexagonal search algorithm for motion estimation

Baishik Biswas, Rohan Mukherjee, Indrajit Chakrabarti, Pranab K. Dutta, Indian Institute of Technology Kharagpur (India)

Motion estimation plays a fundamental role in real-time video coding application encompassing a broad domain including digitally compressed video, ranging from low bit-rate streaming to HDTV broadcast. This paper proposes an efficient VLSI architecture for Hexagonal search algorithm for real-time video processing. The proposed architecture has demonstrated appreciable performance as compared to majority of the existing motion estimation architectures. The FPGA-based design proposes a novel addressing mechanism for accessing the search pixels in the reference frame in order to find the best match for the current macro-block whose motion vector has to be determined. The work concentrates on reducing the chip area drastically with respect to existing architecture without compromising the speed requirement of real-time processing. Major

contribution of the VLSI design lies in accommodating the entire search through a single processing element rather than systolic array based structures with multiple processing elements and on-chip buffers. Again, since the number of search points is not fixed and depend on the motion content of the frame, the proposed architecture adopts a combinatorial logic based design to address the iterative search pattern and access the stored pixels required for estimating the motion. This handling of datawith simplified control also helps lower the critical path length.The proposed memory addressing scheme with simplified memory organizationis based on extending the current and reference frame size in nth power of 2, (say 512?512, in case of standard CIF format of 352?288), while storing them in block RAMs. Thus the 18-bit pixel addresses can only be generated by counter logic. Generation of the memory address offsets for various search points with respect to a particular position can also be done by applying some combinatorial logic on the output of a counter. This simplifies the area to a large extent. A 10-bit variable is also introduced which takes care of variation of position of the blocks with respect to the minimum distortion position. Efficient generation of this variable to locate the position having the best match poses a design challenge. Sequential processing of each block is efficiently organized by four major modules of the design, namely Address Generation Unit, SAD module, Motion Vector module, Control Unit. The implementation shows that the avoidance of parallel processing and data reuseschemes of few existing designs do not create any barrierin achieving high frame rate. The proposed design when implemented using Verilog HDL and synthesized using Xilinx ISE on Virtex-5 technologyhas demonstrated a maximum operating frequency of 320MHz.Requiring 1700 clock cycles on an average to find the best match for a given 8?8 macro-block, the architecture can process 118 frames of standard CIF format working at a frequency of 320 MHz. Area requirement of data-path with control unit is calculated in a fashion to make the measure independent of technology and comes out as 3.4K gate equivalent.

Conference 9401: Computational Imaging XIII

Tuesday - Wednesday 10–11 February 2015

Part of Proceedings of SPIE Vol. 9401 Computational Imaging XIII

9401-16, Session PTues

ISAR for concealed objects imaging

Andrey Zhuravlev, Vladimir Razevig, Igor A. Vasiliev, Sergey I. Ivashov, Bauman Moscow State Technical Univ. (Russian Federation); Viacheslav V. Voronin, Don State Technical Univ. (Russian Federation)

The problem of concealed threat detection under clothing on ground transport hubs and other crowded places remains actual as such places remain targets of terrorist attacks. Unlike in the airports where thorough inspection involving microwave body scanners and manual search is possible, bringing the same level of security to other transport systems remains a challenge. The paper addresses the issue of transport and infrastructure security by applying the concept of inverse aperture synthesis to such a target like a walking person. The proposed ISAR system consists of a vertical linear array of microwave elements and a synchronous time-of-flight video sensor. The vertical resolution in this system is achieved by the vertically distributed microwave antenna array while horizontal resolution is obtained due to subject's walking in the perpendicular direction. The coherent processing of the acquired radar signal results in a synthetic radar image from non-stationary target due to synchronous video processing that gives required trajectories of subject's body parts and clothes. After acquiring this dual channel record during the time of subject's intersecting the inspection area, a synthetic radar image can be obtained for an arbitrary moment by processing the accumulated radar signal and the video sequence. The described concept is illustrated by computer simulation and experiments conducted with a setup consisting of a vertical scanner with movable continuous wave radar and a video sensor. The signal acquisition is performed by the stop motion technique when the subject moves incrementally and remains stationary during each scan and video frame capture. The resulting synthetic radar images are presented. The requirements to an electronically switched system are given to make signal acquisition and processing in real time.

9401-17, Session PTues

Three-dimensional gas temperature measurements by computed tomography with incident angle variable interferometer

Satoshi Tomioka, Shusuke Nishiyama, Samia Heshmat, Yasuhiro Hashimoto, Kodai Kurita, Hokkaido Univ. (Japan)

This paper presents a method to measure a three-dimensional gas temperature distribution without inserting probes to the gas using a couple of techniques of computed tomography and an optical interferometer.

The computed tomography is a well-known technique to determine three-dimensional internal distributions of the attenuation coefficient non-destructively. In conventional computed tomography, X-rays are employed as the probe beams, and two-dimensional line integral distributions of the attenuation coefficient along the probe beams with various incident angles are observed as a set of projection data; then the internal distribution of the attenuation coefficients is reconstructed from the projection data set.

When the object is gas, the object is transparent to the X-ray or to other optical waves. However, an optical difference can be applied to measure the temperature distribution, since the temperature depends on the refractive index, and the optical difference is equal to an integral of the refractive index distribution along the optical path. The temperature distribution can be easily determined by replacing the intensity measurement system with the interferometer system that can measure the optical difference. However, there are two major difficulties; one is the limit of a measurement system, and the other is found in an evaluation of the line integral.

The difficulty of measurement system is that the angle of the incident beam is limited in a certain range. If the object could be rotated, the projection data from all direction of the incident beam would be obtainable. However, the gas cannot be rotated to avoid a convection flow. In this situation, the incident angle must be controlled by moving and rotating the mirrors equipped with the interferometer. However, the incident angles are limited because of the restriction of mirror arrangements.

The second drawback is the complexity of data processing of fringe patterns observed by the interferometer. The integral of the refractive index is observed as a two-dimensional phase distribution; however, the phase is expressed by a multi-valued function of the integral; i.e. the phase is limited in [-pi, +pi], which is called wrapped phase. Therefore, a phase unwrapping technique is required. In addition, the background fringe analysis for each fringe pattern with different incident directions is required, since the background fringe pattern depends on the difference of the angle between the probe beam and the reference beam of the interferometer which changes as the incident angle changes. Furthermore, a filtering is required to obtain the wrapped phase. Since this filtering by traditional half-plane filters in Fourier domain sometimes fails, an improved filtering method is required.

To solve the first problem, we have added a few additional projection data to the set of the limited angle projection data. To solve the second problem, we developed several new algorithms such as a phase unwrapping algorithm, a carrier frequency detection, a noise reduction, and a carrier peak isolation.

The validity of these improvements is shown through the experimental measurements in the temperature distributions around candle flames.

9401-18, Session PTues

An MRI myocarditis index defined by a PCA-based object recognition algorithm

Rocco Romano, Univ. degli Studi di Salerno (Italy); Igino De Giorgi, Azienda Ospedaliera Univ. (Italy); Fausto Acernese, Gerardo Giordano, Univ. degli Studi di Salerno (Italy); Antonio Orientale, Giovanni Babino, Azienda Ospedaliera Univ. (Italy); Fabrizio Barone, Univ. degli Studi di Salerno (Italy)

Magnetic Resonance Imaging (MRI) has shown promising results in diagnosing myocarditis that can be qualitatively observed as enhanced pixels on the cardiac muscles images.

In order to select enhanced pixels, a new PCA-based recognition algorithm [1] was used. In particular, radiologist delimits the myocardial region on an image and selects a small region of enhanced pixels. The algorithm looks at the selected enhanced pixels and defines, for each pixel, a neighborhood of the pixel itself. The neighborhoods are used as image vectors and the eigenvectors of their covariance matrix are extracted. For each pixel in the selected myocardial region a neighborhood and an image vector is then obtained. The image vector, projected in the eigenvector basis, is then compared to the image vectors of the selected enhanced pixels. If it can be considered an enhanced pixel, is counted. The ratio of enhanced pixels and the total pixels in the delimited myocardial region represents the myocarditis index, which should give a quantitative measure of the cardiac muscle inflammation. The algorithm was implemented in Matlab. A group of 10 patients, referred to MRI with presumptive, clinical diagnosis of myocarditis, was analyzed. The exam-analyzed images were obtained on a 1.5 – T Intera CV MRI Unit (Philips Medical Systems) with an inversion recovery contrast enhanced MRI, performed after intravenous injection of gadobutrolo (Bayer) 10 ml, using a three-dimensional T1-weighted turbo-field echo technique in the cardiac short axis and long axis planes (TR 4.16 ms, TE 1.39 ms, flip angle 15°, slice thickness 8 mm, matrix 130 x 252, FOV 440 mm). The inversion time was adjusted for optimal suppression

of normal myocardial signal (inversion time approximately 250 ms), and the images were obtained within 15 min after injection of gadobutrolo. To assess intra- and interobserver variability, two observers blindly analyzed data related to the 10 patients by delimiting myocardial region and selecting enhanced pixels.

The myocarditis index values ranged from 30% to 60%, for different patients of different myocarditis seriousness. In order to assess myocarditis seriousness, two radiologists blindly assigned, with respect to the patient case history and images, a value from 1 to 5 (1 less serious, 5 very serious). There was a significant correlation ($P<0.001$; $r = 0.94$) between myocarditis index and the assigned values. There was a good intra- and interobserver reproducibility, in fact a maximum CoV of 3% (Coefficient of Variation, defined as the SD of the differences between the two separate measurements divided by their means, and expressed as the percentage) for intraobserver and interobserver reproducibility was obtained. The mean difference between the two observer measurements was 1.6% and the 95% limits of agreement on the Bland-Altman plot were -14.6 to 17.8 %.

[1] Matthew T and Pentland A, Eigenfaces for Recognition, Journal of Cognitive Neuroscience, 1991; 3, 1, 71 - 86.

[2] Bland JM, Altman DG, Measuring agreement in method comparison studies, Statistical Methods 1999; 8, 135 – 160.

9401-19, Session PTues

Inverse lighting for non-homogeneous objects from color and depth image using wavelet representation

Junsuk Choe, Hyunjung Shim, Yonsei Univ. (Korea, Republic of)

Creating a computer-generated model of virtual object has been one of the most important topics in a 3D imaging technology. Realistic 3D models are utilized in various applications such as games, movie special effects, and a virtual training system. The computer-generated model consists of three scene attributes and they include a bidirectional reflectance distribution function (BRDF), geometry and lighting.

In this paper, we aim to extract the lighting attribute from images, which is known as the inverse lighting problem. Existing methods for inverse lighting can be classified into two groups depending on the lighting model, either the basis model or point light model. For the basis model, the lighting is represented by a weighted sum of the basis functions that present a locality property in spatial frequency domain. Among basis models, spherical harmonics basis functions (SHBF) are known to be optimal for a convex Lambertian object. The point light model represents the lighting by a set of point lights and it has a locality property in spatial domain. Consequently, the basis model is effective for low frequency lighting while the point model is suitable for high frequency lighting. Lately, a hybrid method suggests combining the SHBF basis and point light model for representing all frequency lighting. However, this method integrates two separate models without considering their redundancy. As a result, it is possible to yield the errors in the sequential estimation process.

This paper proposes a new unified framework to employ Haar wavelet basis functions for estimating all frequency lighting using a pair of color and depth image. The Haar wavelet basis function holds the locality property in spatial and spatial frequency domain simultaneously. In addition, utilizing the fact that there is no redundancy in orthogonal basis functions, we can formulate the inverse lighting problem by linear estimation. Among all possible wavelets, we need to select a set of the important basis functions because estimating all wavelet coefficients are intractable given input images. We use an existing method, the hierarchical refinement algorithm, to decide the importance of each wavelets and obtain the optimal bases. Given selected basis functions, we compute their coefficients to best fit the input image by a least square method and then reconstruct the lighting attributes by a sum of weighted bases.

We perform the preliminary experiments and verify that wavelets can represent all frequency lighting effectively. We synthesize the Lambertian diffuse reflection by a single cosine lobe and reconstruct this by our wavelet based light model. Then, we simulate the specular reflection using a narrow band Gaussian lobe and perform the reconstruction using the proposed model. Using 40 basis functions, we obtain the reconstruction rate of 96.50% for the diffuse reflection and that of 99.29% for specular reflection. These experiments show that the Haar wavelet basis function is suitable for all frequency inverse lighting. In the final manuscript, we apply our unified framework onto both synthetic and real data for the performance evaluation. By conducting various experimental studies, we show the effectiveness of proposed method.

9401-20, Session PTues

A quantum algorithm for multipath time-delay detection and estimation

John J. Tran, Information Sciences Institute (United States); Kevin J. Scully, Darren L. Semmen, The Aerospace Corp. (United States); Robert F. Lucas, Information Sciences Institute (United States)

Resolving closely spaced objects (CSO) below the Rayleigh limit on a two-dimensional focal plane is an important part of automated target recognition and tracking. A brute force algorithm requires time and/or space that scales exponentially with the maximum number of objects the algorithm will resolve; to avoid being trapped by local minima of the cost function, this approach is often taken. Recently, we presented a theoretical translation of this approach for an adiabatic quantum computer (Tran, el al 2014) with resource requirements that do not depend on the number of objects. However, we could not implement it with the restrictions on problem-size and connectivity imposed by current D-Wave machines, potential adiabatic quantum computers.

Here we present a one-dimensional CSO resolution problem which requires fewer resources: the multipath time-delay (MPTD) problem, i.e., resolution of multiple copies of a single radio signal that reach the receiver via slightly different paths. We also present experimental results of the algorithm's performance on the D-Wave architecture. This research effort provides evidence that a real-world problem can be solved using a quantum annealer.

In summary, this presentation is a continuation of the theoretical talk we presented at the 2014 SPIE conference. Here, we present numerical results from our experimentation with the CSO algorithms. These results demonstrate that a class of intractable problems can be solved on an adiabatic optimization device.

9401-21, Session PTues

A no-reference perceptual blurriness metric based fast super-resolution of still pictures using sparse representation

Jae-Seok Choi, Sung-Ho Bae, Munchurl Kim, KAIST (Korea, Republic of)

In recent years, perceptually-driven super resolution (SR) via sparse representation techniques has been proposed to lower the computational complexity. Also, super resolution using sparse representation is known to produce competitive high resolution images with relatively lower computational costs compared to other SR techniques. Nevertheless, it is still difficult to be implemented with substantially low processing power for real-time applications. In order to speed up the processing time of SR, much effort has been made with efficient methods which selectively incorporate elaborate computation algorithms for perceptually sensitive image regions based on Just Noticeable Distortion (JND). However, JND is not suitable to be used for super resolution in perspective of perceptual visual quality because:

1) Most of the SR techniques start off by upscaling lower resolution (LR) images with bi-cubic interpolation. So, the edges tend to be blurred, which deteriorates perceptual visual qualities for the resulting super resolution

images. Therefore, such edge or texture regions must be carefully treated in visual quality perception of Human Visual System (HVS). However, JND thresholds are often obtained with relatively higher values in edge or texture regions.

2) Typical JND models that consider luminance contrast and spatial masking effects often yield relatively lower threshold values in plane areas compared to edge or texture areas. JND-based SR may generate surplus active pixels in plane areas which are perceptually less sensitive to blur distortion, so unnecessarily increasing computation time.

To overcome these problems, we first propose a fast super resolution method with sparse representation, which incorporates a no-reference Just Noticeable Blur (JNB) metric. That is, the proposed fast super resolution method efficiently generates super resolution images by selectively applying a sparse representation method for perceptually sensitive image areas which are detected based on the JNB metric. The JNB metric in our proposed fast SR method allows us to discover the amount of blurriness in blurred edges that can be perceived by human. Low blurriness values indicate that the edges are not strongly blurred, thus being seldom noticeable to human perception. Therefore, a simple bi-cubic interpolation technic can only be applied to the plane, edge or texture areas with low blur distortions. On the other hand, for the edge or texture areas with relatively higher blur distortions, sparse representation is used after the simple bi-cubic interpolation to carefully construct SR images. By doing so, SR images can be fast obtained by selectively applying sparse representation with perceptual quality reasonably maintained. The experimental results show our JNB-based fast super resolution method is about 4 times faster than a non-perceptual SR for 256X256 test LR images. Compared to a JND-based SR method, our JNB-based fast super resolution method is about 3 times faster, with approximately 0.1 dB higher PSNR and slightly higher SSIM value in average. This result indicates that our proposed perceptual JNB-based SR approach improves the subjective quality and has better performance.

9401-22, Session PTues

Efficient capacitive touch sensing using structured matrices

Humza Akhtar, Ramakrishna Kakarala, Nanyang Technological Univ. (Singapore)

Capacitive touch sensing technology, both in consumer electronics and in wall mount displays seem to be omnipresent in the future. Capacitive sensor electrodes are arranged on a touch panel in grid formation. In our previous paper [1], we talked about capacitive touch panel grid pattern design. However in this paper, we extend our research scope to the capacitance sampling issue faced by large scale touch panel designers and propose structured driving voltage matrices along with simplified signal reconstruction algorithms for capacitive touch sensing. The issue of efficient sampling is now more relevant than ever because research has been started for designing large scale capacitive touch panels.

In a capacitive touch panel, the electrodes can be arranged in a single or double layer formation, with each electrode representing a touch coordinate. The electrode array rows (driving lines) are driven and the capacitance readings are sensed through the columns (sensing lines). Increase in scanning time is major issue while constructing a large scale touch screen. Increasing the electrode density (using multi-resolution analysis) allows the touch controllers to capture a detailed picture of the touching object but will increase the scanning time as the controller will have to measure capacitance at every single intersection of driving and sensing electrode in the grid. This issue may be resolved using compressive sensing technique. The idea behind compressive sensing is that super-resolved signals and images can be reconstructed using far fewer data (or measurements) than what is usually considered necessary (Nyquist theorem) [2]. In [3] the authors used a random sampling matrix for their compressive sampling framework, however in order to save limited resources available in touch controllers we propose structured binary circulant matrices which will not only require less memory but also simplify the recovery algorithm.

Much prior work has been done in the study of structured matrices for compressive sensing [4],[5] and [6]. All of these matrices satisfy the Restricted Isometry Property (RIP), which is a necessary condition for a matrix to be able to recover the sparse signal from fewer measurements [2]. We are primarily interested in the binary structured matrices, for example the toeplitz matrix and circulant matrix, due to their easy construction on the fly and low storage requirements. In this paper, we construct binary matrices from algorithms given in [5] and [6] and try to recover our sparse signal using various reconstruction algorithms. During our research we measure the degradation in performance of various binary matrices with decrease in SNR.

For recovery algorithm to work with least error, the unknown signal must be sparse, which is true in the case of capacitive touch panels. The sparsity k is number of simultaneous touches on the touch panel. We exploit the fact that our signal is also clustered as well as sparse to our advantage and this clustering of signal can help in further simplifying the recovery algorithm. We have experimented with various recovery algorithms during our research and finally decided on CoSaMP algorithm [7] due to its simplicity, ease of implementation and robustness to noise. We propose a simpler version of CoSaMP algorithm for binary sensing matrices. Detailed explanation with example scenarios is presented along with the corresponding reconstruction algorithm.

References:

[1] Akhtar, H.; Kakarala, R., “A Methodology for Evaluating Accuracy of Capacitive Touch Sensing Grid Patterns”, Journal of Display Technology, vol.10, no.8, pp. 672,682, Aug. 2014.

[2] E. J. Candes, “Compressive sampling”, Proceedings on the International Congress of Mathematicians, Madrid, August, pp. 22-30, 2006.

[3] Chenchi Luo; Borkar, M.A.; Redfern, A.J.; McClellan, J.H., “Compressive Sensing for Sparse Touch Detection on Capacitive Touch Screens”, Emerging and Selected Topics in Circuits and Systems, IEEE Journal on, vol. 2, no. 3, Sept. 2012, pp. 639-648.

[4] Bajwa, W.U.; Haupt, J.D.; Raz, Gil M.; Wright, S.J.; Nowak, R.D., “Toeplitz-structured compressed sensing matrices.” Statistical Signal Processing, SSP’07. IEEE/SP 14th Workshop on. IEEE, 2007.

[5] DeVore, R.A., “Deterministic constructions of compressed sensing matrices.” Journal of Complexity 23.4 (2007), pp. 918-925.

[6] Amini, A. and Farokh M., “Deterministic construction of binary, bipolar,and ternary compressed sensing matrices.” Information Theory, IEEE Transactions on 57.4 (2011), pp. 2360-2370.

[7] Needell, D., and Tropp, J.A., “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples.” Applied and Computational Harmonic Analysis 26.3 (2009), pp. 301-321.

9401-1, Session 1

Motion compensated content adaptive mesh reconstruction of 4D cardiac SPECT data

Francesc Massanes, Jovan G. Brankov, Illinois Institute of Technology (United States)

In the field of tomographic imaging, it is customary to sample volumetric data using a uniform pattern and voxel bases functions. In [1] we proposed a content-adaptive mesh model (CAMM) reconstruction where the volumetric data is sampled using a non-uniform pattern.

In [2]–[5] we parallelized this algorithm so that it could become practical in terms of needed computation time. We have reported parallel implementation results for 3D volumes [4] and non-circular cameras trajectories [5]. In this paper we take it one step forward and develop a fully mesh-based reconstruction for 4D data as obtained in cardiac gated emission tomography.

In this paper we use the incremental/decremental (ID) algorithm [8] to generate the 3D mesh model from a filtered back projection (FBP) pre-reconstructed images, this is a change from our previous work [4,5] where we used the emission diffusion (ED) [1] algorithm. In this case we only use ED as a starting point for ID algorithm.

Since we are going to be dealing with 4D (3D x-y-z space +1D time due to cardiac gating) data, we have a set of mesh model nodes (spatial samples) in every gated frame. Furthermore, the nodes from one gated frame are not independent of the nodes in other gated frames so we can relate them with motion vectors.

Due to the space constraint of this summary, we will simply reference the methodology used to estimate the motion as can be found in [12].

At this point we can use our previous algorithm [2]-[5] designed in 3D to build an operator for each gate and find the nodal values at each gate. Ideally, if motion is accurately estimated, there should be only one set of nodal values.

For our validation step, we simulate a Pricker Prism3000 SPECT system with a low-energy high-resolution (LEHR) collimator and a TC99m labeled sestamibi as the imaging agent. The simulation is performed using the SIMIND [10] system. The emision and attenuation images used in the simulation are generated using the 4D NURBS-based cardiac-torso (NCAT) [11] 2.0 phantom.

The mesh and its motion were estimated using the actual's NCAT phantom, by the time of the conference we will be performing this on the filtered back projection data instead of the NCAT's phantom. The generated geometry has 33698 nodes and 206076 elements (tetrahedrons). The acceleration on this paper is achieved by using parallel computation in OpenCL which ran on a Intel(R) Xeon(R) CPU E5520 at 2.27GHz with four quad-cores.

The execution time to build the 16 system matrices and the ML-EM reconstruction in under 15 min. At this time we are not accounting for the motion estimation step but it will be evaluated by the time of the conference. The ML-EM reconstruction for the pixel-based reconstruction lasted 48 min.

The generalization of our previous three-dimensional work to account for temporal gating has promising results. It shows good reconstruction performance and it is requires a shorter execution time than the classical pixel based method. By the final manuscript submission we expect to significantly refine our results.

REFERENCES

[1] Y. Yang, M. N. Wernick, J. G. Brankov, A fast approach for accurate content-adaptative mesh generation, IEEE Trans. on Image Processing, vol. 12, no. 8, 2008, pp. 866-881.

[2] F. Massanes, J. G. Brankov, GPU Based Calculations of a SPECT Projection Operator for Content Adaptative Mesh Model, IEEE NSS/MIC 2011.

[3] F. Massanes, J. G. Brankov, Calculations of a SPECT Projection Operator on a Graphical Processing Unit, SPIE in Medical Imaging 2012: Physics of Medical Imaging 2012.

[4] F. Massanes, J. G. Brankov, Parallel computation of a SPECT projection operator for a content adaptative mesh model, 9th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2012.

[5] F. Massanes, J. G. Brankov, OpenCL-Accelerated Computation of a 3D SPECT Projection Operator for Content Adaptive Mesh Model, 12th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine, 2013

[6] R. Boutchko, A. Sitek, G. T. Gullberg, Practical implementation of tetrahedral mesh reconstruction in emission tomography, Phys. Med. Biol. 58, 2013.

[7] M. D. Adams, A Flexible Content-Adaptive Mesh-Generation Strategy for Image Representation, IEEE Trans Image Process., vol. 20, no. 9, 2011.

[8] M. D. Adams, A Highly Incremental/Decremental Delaunay Mesh-Generation Strategy for Image Representation, Signal Processing, vol. 93, no. 4, Apr. 2013, pp. 749-764.

[9] Y. Yang, M. N. Wernick, J. G. Brankov, A fast approach for accurate content-adaptative mesh generation, IEEE Trans. on Image PRocessing, vol. 12, no. 8, 2008, pp. 866-881.

[10] M. Ljungberg, S. E. Strand, A Monte Carlo program for the simulation of scintillation camera characteristics, Computer Methods and Programs in Biomedicine, vol. 29, no. 4, pp. 257272, Aug. 1989.

[11] W. P. Segars, Development and Application of the New Dynamic NURBS-based Cardiac-torso (NCAT) Phantom, University of North Carolina at Chapel Hill, 2001.

[12] T. Marin, J. G. Brankov, Deformable left-ventricle mesh model for motion-compensated filtering in cardiac gated SPECT, Med. Phys, vol. 37, no. 10, Oct, 2010.

[13] J. E. Stone, D. Gohara,G. Shi, OpenCL: A Parallel Programming Standard for Heterogeneous Computing Systems, IEEE Des. Test, vol 12. n. 3, May 2010.

9401-2, Session 1

Image reconstruction in the presence of non-linear mixtures utilizing wavelet variable-dependency modeling in compressed sensing algorithms

Lynn M. Keuthan, The George Washington Univ. (United States); Jefferson M. Willey, U.S. Naval Research Lab. (United States); Robert J. Harrington, The George Washington Univ. (United States)

In the past decade, compressed sensing theory has shown that signals with sparse representations can be reconstructed from far fewer measurements than the rate given by the Nyquist Sampling Frequency. Compressed sensing shows that data and signals in many real-world problems (composed of non-linear mixtures) can be accurately represented by sparse representations. Indeed, the success of compressed sensing in imaging, video, and audio applications can be attributed to the fact that most images and video and audio recordings of practical interest have sparse representation in a transform domain.

Given the non-linear nature of compressed/sparse sensing and the ability to recover information from a non-linear mixture, compressed/sparse sensing is a promising area to explore addressing variable-dependencies and optimization of performance based on accounting for variable-dependencies. Initial compressed sensing recovery techniques assumed that the sparsity transform coefficients were independently distributed and did not exploit dependencies between transform coefficients to improve recovery performance. More recently some algorithms have been proposed to exploit some variable dependencies for improved compressed sensing recovery. To efficiently exploit variable dependencies, the nature and characteristics of the dependencies need to be accurately accounted for. This paper proposes to apply wavelets and their theoretical principles, and the structural statistical modeling of dependencies, in the reconstruction algorithm to improve feature optimization in the presence of non-linear mixtures.

Compressed Sensing reconstruction algorithms are usually iterative. They use the previous signal estimate in consequent reconstruction to identify significant coefficients. They implicitly or explicitly introduce weights, which are often based on the magnitude of the former signal estimate. They typically do NOT exploit signal dependencies between coefficients. They typically set:

$$wk(i) = 1/|xk\text{-}1(i)|$$

where wk(i) denotes a weight associated w/the wavelet coefficient xk(i) and xk-1(i) denotes the wavelet coefficient at iteration k-1.

Conventional Compressed Sensing algorithms, as shown above, utilize the magnitudes of the sparse domain coefficients in an independent manner while determining the weights to be used in the next iteration. Therefore, statistical dependencies between sparse domain coefficients are not exploited. More recent Compressed Sensing reconstruction algorithms have limitations in regard to accounting for variable-dependencies, and, therefore, in regard to optimization of performance for many real world applications, including image processing, particularly in regard to reconstruction of natural images with random-appearing/complex dependencies.

This paper seeks to enhance Compressed Sensing reconstruction based on feature dependencies obtained by exploiting Wavelet Theory and the modeling of structural statistical dependencies. Our approach to optimization is to incorporate dependency characteristics into the Compressed Sensing reconstruction algorithms to optimize reconstruction

performance. We focus on Sparsifying Transforms and the Bayes Least Squares-Gaussian Scale Mixtures (BLS-GSM) model to model dependencies and optimize reconstruction of images. Our paper provides a technical review of pertinent algorithms for Compressed Sensing and then incorporates dependency characteristics into the reconstruction algorithms by using wavelets and structural statistical dependencies to model and construct dependency characteristics during the formulation of coefficient weights during successive iterations. Both qualitative image reconstruction performance improvement and quantitative performance improvement based on minimization of reconstruction error are shown.

9401-3, Session 1

Machine learning deconvolution filter kernels for image restoration

Pradip Mainali, Rimmert Wittebrood, TP Vision (Belgium)

This paper presents a novel algorithm to restore a sharp image from its blurry form. To restore the sharpness, de-convolution/inverse filtering is required, which is in general an ill-posed problem. The system remains ill-posed even when the degradation model is known. Therefore, regularization is extensively studied in the literature to alleviate it. In this paper, we propose a novel approach of using the machine learning technique to regularize the inverse filtering. The pixel class based de-convolution filters are optimized offline. The image is restored by classifying the pixel and the filter optimized for the corresponding pixel class.

The proposed algorithm consists of offline learning and online filtering stages. The offline learning/training requires a database of huge number of images (e.g. 8 billions of pixels in our database) as the training samples. The algorithm first learns the dictionary/codebook of the pixel classes that encodes various local characteristic of the pixel within a local patch of size e.g. 3x3. The dictionary is learnt by extracting the local pixel features over all the pixels in the database and by applying the clustering algorithm such as k-means clustering to group the pixel features that are similar. The pixel features are grouped into K compact pixel classes, achieving a small error while pixel classification. Once the dictionary is learnt, the pixels in the database are assigned a class label. Later, the de-convolution filter coefficients for each pixel class are optimized using a huge number of source and the degraded image pairs in the training database. The Least Mean Squares (LMS) criteria is used to optimize the filter coefficients for each pixel class using the source and the degraded pair of the images. The degradation model that needs to be applied during the training stage is inverse filtering that is observed during the filtering. Therefore, for the sharpness restoration, we generate the degraded image by applying the Gaussian blur on the source image. Consequently, the offline stage effectively generates the filters optimized per pixel class to perform an inverse filtering.

In an online stage, the algorithm performs pixel classification based on its local characteristics. The image pixels in the window of size e.g. 3x3 are used to construct the feature vector exactly as done during the offline learning. The nearest matching pixel class is searched in the dictionary. Depending on the pixel class to which the pixel belongs, the corresponding filter is selected and the image pixel is filtered with the optimized filter coefficients for that pixel class. Moreover, the pixel can be classified into a multiple class. The filtered results from the multiple filters are combined by weighting with their class distance.

To restore the images at different blur impacts, the filters are optimized for few discrete blur levels. The blur impact of the input image is measured and the filters optimized for the corresponding blur impact is used to restore the image. The proposed algorithm is evaluated over the various images, which shows the superior performance over the state-of-art methods.

9401-4, Session 1

Sparsity based noise removal from low dose scanning electron microscopy images

Alina Lazar, Youngstown State Univ. (United States); Petru S. Fodor, Cleveland State Univ. (United States)

Among the various surface characterization techniques used in material science, microbiology, and nanotechnology, electron microscopy has emerged as a central player when high spatial resolution topography or compositional information is required. While this type of microscopy is extremely versatile and provides extreme spatial resolution from micrometer to sub-nanometer scales, random fluctuations in both the emission of electrons from the electron gun, as well as the production of secondary electrons upon their interaction process with the surface, drastically reduce the signal-to-noise ratio (SNR). The noise profile dominated by this Poisson distributed shot noise is further complicated by the typical electric and thermal additive Gaussian noise associated with the signal processing units.

The standard approach in increasing the SNR in the context of scanning electron microscopies is to increase the dwell time, i.e. the time the imaging beam spends at each location on the sample, as the signal detected is directly related with the beam interaction time with the surface. Besides the obvious increase in the scan times required, the main disadvantage of this approach is the increased potential for beam induced damage in the sample. Most organic samples and some inorganic ones such as zeolites are highly susceptible to the thermal damage associated with the high-energy primary beam electrons bombardment. Thus, in these applications, a low dose associated with short scan times is required to prevent sample damage.

In this work we present an alternative methodology for obtaining high quality electron microscopy data using small dwell times, i.e. low dose, based on offline image processing. The techniques used take advantage of the inherent sparsity of electron microscopy images. An analysis of more than 450 images acquired from a broad range of biological, material and device samples, shows that in a typical SEM acquired data the image energy is concentrated in less than 20% of the DCT transform coefficients, enabling us to develop sparsity based algorithms for denoising. To this end we develop a noise model for the system and determine the instrument specific white Gaussian noise from measurements of the statistical fluctuations in the signal. With the knowledge of the white noise level, a variance stabilization technique is applied to the raw data followed by a patch-based denoising algorithm. Successful denoising results are presented both for images with known levels of mixed Poisson-Gaussian noise, as well as for raw images. The quality of the image reconstruction is assessed using the standard peak signal-to-noise-ratio (PSNR) as well as based on measures specific to the application of the data collected. These include accurate identification of objects of interest and structural similarity. For natural images high-quality results are recovered from noisy observations, allowing for example high rates for feature identification (better than 80%), while decreasing the scanning time tenfold. The techniques developed can be easily adapted to other scanning imaging methods affected by mixed noise environments including confocal, fluorescence and near field optical microscopy.

9401-5, Session 1

Recovery of quantized compressed sensing measurements

Grigorios Tsagkatakis, Foundation for Research and Technology-Hellas (Greece); Panagiotis Tsakalides, Foundation for Research and Technology-Hellas (Greece) and Univ. of Crete (Greece)

The mathematical theory of Compressed Sensing has been applied to various engineering areas ranging from one-pixel cameras, to range imaging and medical ultrasound imaging, to name a few. The theory developed within CS suggests that one can achieve perfect reconstruction of a signal from a small number of random measurements, far below the

typical Shannon-Nyquist sampling limit. Recovery from the compressed measurements is possible by exploiting the sparsity of the signal, when expressed in an appropriate dictionary.

Despite the recent explosion of CS sampling architectures, the majority of literature studies the scenario where a sufficiently large number of randomly encoded measurements are available at the decoder for signal reconstruction. Although signal sampling and reconstruction are crucial, systems that implement CS must use a finite number of distinct symbols via a quantization process in order to efficiently store and transmit the random linear measurements.

The necessity of quantization for storage and transmission has motivated the investigation of the effects of quantization on the recovery capabilities of CS. The challenges associated with analyzing and overcoming the limitations of quantization are directly related to the non-linear nature of the process. The problem has been studied for different quantization schemes, ranging form uniform to sigma-delta to the extreme case of 1-bit CS where the recovery algorithm is presented merely with the signs of the random measurements.

In this work, we propose a novel formulation of CS recovery where the effects of quantization are taken into account during reconstruction. Formally, we consider a non-linear mapping function of the sparse signal encoding the contribution of quantization and enforcing consistancy with the estimated measurements. Recovery in the presence of the non-linear function can be achieve by solving a modified greedy minimization problem, the Orthogonal Matching Pursuit (OMP), with additional constraints on the consistency of the recovered signal with respect to its quantized counterpart.

The merits of introducing the non-linear quantization during the process of sparse minimization were investigated under various conditions in software simulations. The results indicate that the proposed minimization is significantly better in estimating compressable signals in both moderate (8 bits per measurement) and severily limited (4 bits per measurement) situations. Especially in the low sampling rates reigim, the proposed scheme achieves a reduction by half of the estimation error. We should also point out that the additional performance gains are accompanied with a marginal increase in decoding complexity. The performance gains combined with the encoding efficiency can be extremely valuable in resource limited imaging scenarios.

9401-6, Session 1

Mobile image based color correction using deblurring

Yu Wang, Purdue Univ. (United States); Chang Xu, Qualcomm Inc. (United States); Carol J. Boushey, Univ. of Hawai'i Cancer Ctr. (United States) and Purdue Univ. (United States); Edward J. Delp III, Purdue Univ. (United States)

Dietary intake, the process of determining what someone eats during the course of a day, provides valuable insights for mounting intervention programs for prevention of many chronic diseases such as obesity and cancer. Accurate methods and tools to assess food and nutrient intake are essential for epidemiological and clinical research on the association between diet and health. The goals of the Technology Assisted Dietary Assessment (TADA) System, developed at Purdue University, is to automatically identify and quantify foods and beverages consumed by utilizing one image of a users food acquired with a mobile device (mobile telephone). The current system consists of four main steps: preprocessing, segmentation, classification and volume estimation. Every step is essential in the sense that a relatively small error may result in misclassification of the food or errors in the volume estimation. Color serves as one of the key features in food image processing. It is important that we acquire accurate color images from the mobile device. In our previous work an image quality measurement technique as well as a linear color correction model using the CIELAB color space was described.

In this paper, we propose an image quality enhancement technique by combining image de-blurring and color correction. The idea is based on the assumption that users include a specifically designed color checkerboard (i.e. a fiducial marker) in their food images. This particular type of fiducial marker provides a reference for the scale and pose of the objects in the scene and to provide color calibration information for the camera and scene lighting. In the image de-blurring step, a quality metric is used to detect blurriness in the food image. Once blur is detected, a blur kernel is estimated for the most salient patch. The estimation process is done in a coarse-to-fine approach in order to avoid local minima. Then, a de-convolution technique is used to recover the unblurred image. In the color correction step, we combine the Gray World model with a revised linear color correction model using the LMS color space to achieve faster color adaption. The Macbeth color checker, which is a calibrated 24-color reference chart, is used to measure the Root Mean Square Error (RMSE) between the reference image acquired with D65 lighting and the corrected image. The results show that our method can lower the error by approximately 10 percent compared with our previous work. This implies that even though the CIELAB color space is more uniform with respect to the Human Visual System (HSV), it is not necessarily the best choice when it comes to the linear color correction model, since the model expects each channel to be correlated. Overall, the proposed image enhancement technique improves the blurred image quality significantly in the sense that it provides more reliable image analysis results.

9401-7, Session 2

Spectral x-ray diffraction using a 6 megapixel photon counting array detector

Ryan D. Muir, Nicholas R. Pogranichniy, Purdue Univ. (United States); J. Lewis Muir, Argonne National Lab. (United States); Shane Z. Sullivan, Purdue Univ. (United States); Kevin Battaile, Anne M. Mulichak, Argonne National Lab. (United States); Scott J. Toth, Purdue Univ. (United States); Lisa J. Keefe, Argonne National Lab. (United States); Garth J. Simpson, Purdue Univ. (United States)

A statistical approach was developed for unmixing polychromatic X-ray scattering patterns acquired on a photon counting array detector, opening up new strategies for improved diffraction analysis. Statistical models of the detector response as a function of photon energy were developed and fit to the measured responses of each pixel in a 6 megapixel photon counting array detector. A comparator circuit for photon counting and a digital counter is integrated into each pixel in the array. Absorption of an X-ray photon produces a current transient, which in turn is converted to a voltage transient by a resistor. If the peak of the voltage transient exceeds the threshold voltage of the comparator, a count is recorded. Following calibration, the probability density function (pdf) for detecting a count as a function of photon energy was determined for each pixel in the array. Armed with this knowledge, the measured counts acquired at multiple threshold values allowed separation of dual color composite patterns into two monochromatic scattering patterns. This capability can improve the speed and accuracy of dual-wavelength measurements to aid in addressing phase uncertainty in diffraction by simultaneously providing diffraction patterns at photon energies producing both normal and anomalous dispersion. In addition, scanning the sample through a narrow 5-10 micrometer diameter X-ray beam allows diffraction microscopy measurements for structure-specific imaging.

9401-8, Session 2

Anomaly detection of microstructural defects in continuous fiber reinforced composites

Stephen E. Bricker, Univ. of Dayton Research Institute

(United States); Jeffrey P. Simmons, Craig Przybyla, Air Force Research Lab. (United States); Russell C. Hardie, Univ. of Dayton (United States)

Ceramic matrix composites (CMCs) with continuous fiber reinforcements have the potential to enable the next generation of high speed hypersonic vehicles and/or significant improvements in gas turbine engine performance due to their exhibited toughness when subjected to high mechanical loads at extreme temperatures (2200F+). Continuous fiber reinforcements provide significant fracture toughness, crack growth resistance, and strength. However, CMCs are quite complex and little is known about how stochastic variation and defects in the as processed material impact material properties. Of particular interest is the ability to characterize nominal fiber reinforcement, both on the individual fiber scale, and at the tow or meso-scale. The objective here is to determine unusual characteristics of the manufactured CMC in an automated fashion and correlate these characteristics to material properties to help improve the microstructure design, performance prediction, simulation, and fabrication.

Characterization of CMC microstructure is possible by means of automated serial sectioning and imaging. Serial images are acquired with an optical microscope following each polishing of the material. However, the resulting stack of images can be very large with tens of thousands of features of interest and manual analysis of the images is impractical for all but the smallest volumes. In this work, we present a local-entropy based fiber segmentation algorithm for 3D imagery of reinforced CMCs. Segmentation of fibers by simple thresholding is often problematic in CMCs due to low contrast between fiber and matrix phases. Due to a more uniform microstructure, the fibers tend to have a much more homogenous value than the surrounding matrix. This homogeneity is exploited for fiber segmentation by using a local entropy filter, in conjunction with a series of 3D connectivity and morphological operations to obtain robust 3D fiber segmentation. The details of the segmentation algorithm are presented here along with an accuracy quantification based on a comparison to ground truth data. Once the fibers are identified and segmented, a reduced mathematical description of the fiber geometry can be obtained by fitting ellipses to cross sections along its length. These simplified geometrical descriptions are useful for generating digital microstructure volumes that are meshed and simulated using the FEM to predict response at the scale of the individual fibers.

After fiber segmentation, numerous features are computed to describe the orientation, shape, flow, and consistency of the fibers. For example, the fiber velocity field and gradient are estimated to characterize fiber path behavior, where velocity refers to the change in a fiber's position along the fiber axis direction. Fiber coatings are also segmented and automatically characterized. This processing allows us to validate the fabrication parameters and characterize the real microstructure.

Statistical anomaly detection is also employed to automatically detect fibers whose orientation and path are inconsistent relative to the other fibers in the image based on the computed features described previously. In particular, an overall probability density function (pdf) for the computed features is estimated using a multivariate Gaussian mixture model (GMM). This model captures the aggregate behavior of the fibers and coatings. The areas of anomalous fiber patterns are automatically detected based on the likelihood value provided from the GMM pdf. Anomalous behavior in the material is quantified and visualized for multiple features. The detection and characterization of anomalous microstructure is a critical step in accurate material simulation and property prediction for the integrated computation materials engineering (ICME) of RFC based components.

9401-9, Session 2

Phase retrieval in arbitrarily-shaped aperture with the transport-of-intensity equation

Lei Huang, Brookhaven National Laboratory (United States); Chao Zuo, Nanjing University of Science and Technology (China); Mourad Idir, Brookhaven National Laboratory (United States); Weijuan Qu, Ngee Ann Polytechnic (Singapore); Anand Asundi, Nanyang Technological University (Singapore)

Phase information is not easy to detect directly since the energy-based optical sensors detect the intensity only. But sometimes it contains what we really want. As one of the phase retrieval techniques, the famous Transport-of-Intensity Equation (TIE) is continuously researched after proposed by Teague in 1983. Due to its simple setup and easy implementation, the TIE has been widely used in many applications. In fact, the TIE is well studied under homogeneous Neumann boundary condition that the phase derivatives in the normal directions at boundary of the field of view (FOV) equal to zero. A few applications belong to this condition, for instance, the observation of a cell that is center-positioned in a "flat wavefront". Nowadays, the widely used Fast Fourier transform (FFT) based TIE solver works well for phase retrieval in this case. Nevertheless, in many other applications, such as wavefront sensing, it is impossible to have the optical wave always satisfy the homogeneous Neumann boundary condition. In this case, the energy inside the FOV is not conserved, because the energy exchanges at the boundary during the recording distance is being changed. The FFT-based TIE solver fails to retrieve the phase under this condition, and to solve this problem, Zuo et al. suggested adding a rectangular hard aperture to limit the optical field under test and then to keep the energy conservation. However, in practice it is quite challenging to add an aperture whose shape is exactly a rectangle, due to the difficulty in aperture fabrication and system alignment, or the other existing pupils (e.g. telescopes) obstructing the aperture shape to be a rectangle. Consequently, the performance of Zuo's method is limited in real applications. In this work, we present a new method to solve the TIE under general boundary condition by adding an arbitrarily-shaped hard aperture. Since the aperture shape can be arbitrary in our proposed method, the difficulty of aperture fabrication and system alignment can be significantly reduced. In our method, the TIE is solved by using iterative discrete cosine transforms (DCT) method, which contains a phase compensation mechanism through iterations to improve the retrieval results. The proposed method is verified in simulation with an arbitrary phase, an arbitrarily-shaped aperture, and non-uniform intensity distribution. Real experiment is also carried out to check its feasibility. Comparing to the classical FFT based TIE solver, the proposed method is accurate and applicable for any phase distribution. Compared with Zuo's DCT method, the added aperture in hardware can be in an arbitrary shape which results in a low requirement on aperture fabrication and alignment, and data processing procedure is extremely automatic and easy to use. These features of the proposed method significantly enhance the flexibility of TIE measurement with hard aperture in real applications. It is very easy and extremely straightforward to use in a practical measurement as a flexible phase retrieval tool.

9401-10, Session 2

Acceleration of iterative image reconstruction for x-ray imaging for security applications

David G. Politte, Washington Univ. in St. Louis (United States); Soysal Degirmenci, Washington Univ. in St Louis (United States); Carl M. Bosch, Nawfel Tricha, SureScan Corp. (United States); Joseph A. O'Sullivan, Washington Univ. in St. Louis (United States)

Three-dimensional image reconstruction for scanning baggage in security applications is becoming increasingly important. Compared to medical x-ray imaging, security imaging systems must be designed for a greater variety of objects. There is a lot of variation in attenuation and nearly every bag scanned has metal present, potentially yielding significant artifacts. Statistical iterative reconstruction algorithms are known to reduce metal artifacts and yield quantitatively more accurate estimates of attenuation than linear methods.

For iterative image reconstruction algorithms to be deployed at security

checkpoints, the images must be quantitatively accurate and the convergence speed must be increased dramatically. There are many approaches for increasing convergence; two approaches are described in detail in this paper. Each approach is implemented

on real data from a SureScanTM x1000 Explosive Detection System? and compared to straightforward implementations of the alternating minimization (AM) algorithm of O'Sullivan and Benac [1], both unregularized and regularized (using a Huber-type edge-preserving penalty originally proposed by Lange [2]).

Ordered subsets [3, 4] is a well known range decomposition technique for accelerating iterative algorithms, but does not guarantee convergence. Our analysis starts with a novel reformulation of convergent ordered subsets that was originally introduced by Ahn, Fessler et al. [5]. The derivation of the AM algorithm is modified to use new surrogate functions. Contrary to Ahn and Fessler's approach, the surrogate functions are not quadratic so this method yields a straightforward modification of the ordered subsets AM algorithm with guaranteed convergence.

The first acceleration technique (beyond the convergent ordered subsets algorithm) that has been implemented is based on scheduling the number of subsets used in a sequence of ordered subsets iterations. Using a large number of subsets in our ordered subset implementation yields a final cost that is higher than that achievable without ordered subsets or with convergent ordered subsets. Scheduling the number of subsets to decrease slowly as the iterations proceed yields lower final value of the cost than using any fixed number of subsets.

The second acceleration technique varies the step size for the updates in the AM algorithm. The AM algorithm yields additive updates for the attenuation values with a multiplicative factor that is chosen to guarantee convergence. This guarantee can yield smaller step sizes that are more conservative than necessary. Larger step sizes may yield lower values of the cost, but may not guarantee convergence. Step sizes may be selected in a number of ways including line-search, an optimization method, and a fixed schedule. Line-search and optimization methods have been investigated by Kaufman using the EM algorithm for PET imaging [6]. The line-search and optimization methods require multiple computations of the cost (or its derivative) at each step, but these computations are typically significantly less than the cost of either a forward or a backward projection that are required for a full iteration. If the line-search also takes into account the nonnegativity of attenuation, then thresholding of the updates would require additional forward projections. Furthermore, the fixed schedule requires no additional computations but it is not guaranteed to converge for every schedule.

?SureScanTM is a trademark of the SureScan Corporation.

[1] O'Sullivan, J. A. and Benac, J., "Alternating minimization algorithms for transmission tomography," Medical Imaging, IEEE Transactions on 26(3), 283–297 (2007).

[2] Lange, K., "Convergence of EM image reconstruction algorithms with Gibbs smoothing," Medical Imaging, IEEE Transactions on 9, 439–446 (Dec 1990).

[3] Kamphuis, C. and Beekman, F., "Accelerated iterative transmission CT reconstruction using an ordered subsets convex algorithm," Medical Imaging, IEEE Transactions on 17, 1101–1105 (Dec 1998).

[4] Erdogan, H. and Fessler, J. A., "Ordered subsets algorithms for transmission tomography," Physics in Medicine and Biology 44(11), 2835 (1999).

[5] Ahn, S., Fessler, J., Blatt, D., and Hero, A., "Convergent incremental optimization transfer algorithms: application to tomography," Medical Imaging, IEEE Transactions on 25, 283–296 (March 2006).

[6] Kaufman, L., "Implementing and accelerating the em algorithm for positron emission tomography," Medical Imaging, IEEE Transactions on 6, 37–51 (March 1987).

9401-23, Session 2

Rotationally-invariant non-local means for image denoising and tomography

Suhas Sreehari, Singanallur Venkatakrishnan, Purdue Univ. (United States); Lawrence F. Drummy, Jeffrey P. Simmons, Air Force Research Lab. (United States); Charles A. Bouman, Purdue Univ. (United States)

No Abstract Available

9401-11, Session 3

Restoration of Images for the Granular Mirror Space Telescope

Xiaopeng Peng, Alexandra Artusio-glimpse, Garreth J Ruane, Grover A Swartzlander Jr., Rochester Institute of Technology (United States)

Modern space-based telescope mirrors are limited in size, which constrains the light collecting power and resolution. A space telescope with large effective mirror may be achieved by swarms of micro mirrors. In this paper, we present an analysis of such a Mirror Swarm Space Telescope (MSST) system.

The basic idea of MSST is to combine a swarm of mirror segments to achieve resolution comparable to that of a single mirror with equivalent size to the entire mirror swarm distribution. Spatially coherent light from distant stars focused onto a single detector array by the mirrors. The captured images contain high resolution information, but are significantly affected by aberration and speckle due to the random distribution and constant motion of the micro-mirrors. Multi-frame deconvolution techniques are employed to restore a high resolution image.

In this paper, we demonstrate using MSST to recover the image of binary stars using both numerical and laboratory simulations. Two distant light sources are used to represent binary stars. An aperture mask with randomly distributed holes is placed in front of a lens to represent the mirror swarm. By changing the distribution of holes, the temporal variations of the mirror positions are simulated. A CCD detector is placed at the focal plane to capture the formed images. Light sources of both narrow-band and broad-band are used to mimic the star light. Phase masks are used in laboratory experiment to simulate the phase aberration, and different levels of piston and tilt are examined numerically. Detector noise with different signal to noise ratio(SNR) is also investigated.

We implement three types of multi-frame deconvolution methods to reconstruct a high-quality image from the captured data: non-blind, blind, and hybrid deconvolution. Non-blind deconvolution assumes the point spread function (PSF) is known. Therefore, an approximate PSF is measured and Wiener Filter is applied to recover the binary stars. Blind deconvolution recovers the image and PSF simultaneously. We implement S. Harmeling's iterative blind deconvolution (IBD) method, which efficiently solves the multi-frame deconvolution. Finally, we combine the Wiener filter and IBD by using the Wiener filter results as image prior and PSF prior to propagate the IBD.

The results indicate that the hybrid method is robust to both phase aberration and detector noise. Specifically, binary stars with angular separation 1.22 ?/D may be distinguished with 15 randomly placed micro-mirrors, each with a random amount of piston ranging from 0 to 12 ϖ radians and tilt ranging from -11 ? /D to 11 ?/D. Here, ? is the wavelength of the light source and D is the maximum allowable distance between the mirrors. In addition, the SNR could be as low as 17.25 dB. The Wiener filter alone is robust to phase aberration, but sensitive to noise. On the other hand, the IBD method alone is robust to detector noise, but sensitive to the phase aberration.

In summary, the swarm mirror space telescope we proposed is capable of high resolution image formation while potentially allowing a much larger collecting area, more flexibility of control, and lower cost.

9401-12, Session 3

Regularized image registration with line search optimization

Lin Gan, Gady Agam, Illinois Institute of Technology (United States)

Image registration is normally solved as an optimization problem where an objective function is minimized based on a selected similarity metric. A major factor contributing to the complexity of image registration is the inclusion of special constraints on the solution in the optimization process. For example, in medical image registration a smoothness constraint is imposed to adhere to physical tissue characteristics.

The line search procedure is commonly employed in unconstrained nonlinear optimization. At each iteration step the procedure computes a step size that achieves adequate reduction in the objective function at minimal cost. The procedure uses termination conditions that guarantee global convergence. The Wolfe conditions are among the most widely used termination conditions. They consist of a sufficient condition and a curvature condition. The sufficient condition (also known as the Armijo condition) requires a sufficient decrease in the objective function given the selected step size. The curvature condition rules out unacceptably short steps. In addition to a well chosen step size, a well chosen descent direction is also required to guarantee global convergence.

In this paper we extend the constrained line search procedure with different regularization terms so as to improve convergence. The extension is addressed in the context of constrained optimization to solve a regularized image registration problem. Specifically, the displacement field between the registered image pair is modeled as the sum of weighted Discrete Cosine Transform basis functions. A Taylor series expansion is applied to the objective function for deriving a Gauss-Newton solution. We consider two regularization terms added to the objective function. A Tikhonov regularization term constrains the magnitude of the solution and a bending energy term constrains the bending energy of the deformation field. We modify both the sufficient and curvature conditions of the Wolfe conditions to accommodate the additional regularization terms.

To evaluate the performance of the proposed extension, we generate known deformation fields and warp images according to them so as to have a test collection with known deformation fields. In this manner, by comparing estimated deformation fields to the known ones, we obtain an objective evaluation criterion that is not affected by image appearance. In addition to evaluating accuracy we also evaluate the computational cost. We compare the solutions obtained from nine cases: a) without regularization term and line search; b) without regularization term and with a backtracking line search; c) without regularization term and with a Wolfe condition line search. d) with Tikhonov regularization and without line search; e) with Tikhonov regularization and a backtracking line search; f) with Tikhonov regularization and a Wolfe condition line search; g) with bending energy regularization and without line search; h) with bending energy regularization and a backtracking line search; h) with bending energy regularization and a Wolfe condition line search.

The experimental evaluation results show that a solution obtained with bending energy regularization and Wolfe condition line search achieves the smallest mean deformation field error among 100 registration pairs. This solution shows in addition an improvement in overcoming local minima.

9401-13, Session 3

Rectangular approximation of buildings from single satellite image using shadow analysis

Gurshamnjot Singh, Mark Jouppi, Avideh Zakhor, Univ. of California, Berkeley (United States)

Automatic building extraction from remotely sensed data is an important problem with many applications such as urban planning, population estimation, disaster management, and assessing human impact on the environment. There exists numerous building detection methods using aerial LiDAR data, as well as stereo processing of satellite imagery. The advantage of satellite over LiDAR is larger coverage, higher spatial resolution, and lower cost, while LiDAR data can be more accurate as it directly measures range or distance. In this paper, we develop a methodology for rectangular approximation of buildings and their heights using a single satellite image by exploiting the shadows. We start with a pan sharpened satellite image by combining the panchromatic and multi-spectral imagery. We then detect straight line segments in the pan sharpened satellite image. In addition, we find shadow pixels by analyzing the intensity histograms of the pan sharpened satellite image, and eliminating pixels corresponding to vegetation using the NVDI index. We then select line segments that could potentially correspond to building-shadow boundaries by taking into account the direction of the sun for the satellite image, and the previously detected shadow pixels. In doing so, it is possible for adjacent buildings with unequal heights to share the same building-shadow line segment. As such, we decompose such building-shadow line segments into multiple segments, each corresponding to a building with a different height.

To assign a rectangle to each building-shadow line segment, we first use texture to segment the pan sharpened satellite image. Next we assign the "best" segment to each building-shadow line segment depending on a number of factors such as (a) Morphological Building Index (MBI) for that segment, (b) the vegetation content of that segment as measured by NVDI, (c) the geometric shape of the segment and its spatial relationship with respect to the position and orientation of the line segment under consideration. Finally, we construct a rectangular approximation for that building-shadow segment by exploiting the shape and size of the segment associated with it, and assuming near Manhattan geometry for the building.

While in rural areas with low building density, the above approach results in distinct mostly non-overlapping rectangles, in urban areas, the resulting rectangular approximations heavily overlap with each other. To "untangle" the overlapping rectangles in the latter case, we construct a graph where each node corresponds to a rectangle and the edges between the nodes are assigned a pair of weights that reflect the degree of overlap between the two rectangles. For each connected component of the graph, we then prune the rectangles with high overlap, and remove the edge between rectangles with low overlap, and repeat the connected component analysis until there are no "edges" between the nodes, and all rectangular approximations have been de-tangled.

We demonstrate the effectiveness of our proposed scheme in both rural and urban satellite images of Jordan. We achieve detection rate of 90% (68%) and false alarm rate of 9% (58%) for rural (urban) areas.

9401-14, Session 3

Webcam classification using simple features

Thitiporn Pramoun, King Mongkut's Univ. of Technology Thonburi (Thailand); Jeehyun Choe, He Li, Qingshuang Chen, Purdue Univ. (United States); Thumrongrat Amornraksa, King Mongkut's Institute of Technology Thonburi (Thailand); Yung-Hsiang Lu, Edward J. Delp III, Purdue Univ. (United States)

Thousands of sensors are connected to the Internet. The ``Internet of Things" will contain many ``things" that are image sensors. This vast network of distributed cameras (i.e. web cams and surveillance systems) will continue to exponentially grow. We are interested in how these image sensors can be used to sense their environment.

How can this ever increasing amount of imagery be interpreted to extract valuable information related to weather forecast, traffic control, environment management, and location identification? Interpreting and learning from such images pose many challenges. Indoor-outdoor scene classification is a key step in the interpretation process. In this paper, we propose a method to classify images into indoor or outdoor scenes using a set of simple visual features and classification.

Our approach relies on multiple visual features. We consider four different types of features: color and edge information, line orientation and texture information. We investigate eight features to classify a camera as "looking" at an indoor or outdoor scene. These features are: two color features, the Scalable Color Descriptor (SCD) and the Dominant Color Descriptor (DCD), the number of edge pixels, the number of vertical and horizontal lines from the Hough transform, and texture features from gray-level co-occurrence matrix (GLCM), in particular the contrast, correlation and entropy. Classification is based on the use of nearest neighbor using the L1 norm.

For our experiments we used the MIT Scene Understanding (SUN) database. Five hundred images were used for training (250 indoor images and 250 outdoor images). Fifty images are used for testing. We conducted two experiments to test the impact of color features on the classification process. The first experiment used all the features while the second experiment excluded color information. We achieve 68% accuracy using all features and 74% excluding the color information.

In the full paper we will show more details of how we weight the features in the classifier. We will show accuracy in the terms of precision and recall also we will compare our approach with other web cam classification methods.

9401-15, Session 3

Flatbed scanner simulation to analyze the effect of detector's size on color artifacts

Mohammed Yousefhussien, Roger L. Easton Jr., Raymond Ptucha, Rochester Institute of Technology (United States); Mark Q. Shaw, Brent Bradburn, Jerry K. Wagner, David Larson, Hewlett-Packard Co. (United States); Eli Saber, Rochester Institute of Technology (United States)

Flatbed scanners have been widely used over the past several decades for consumer, industrial, and scientific purposes. Existing literature on flatbed scanner design generally addresses individualistic issues such as optical resolving power, color moiré, motion errors, etc., or investigates applications such as the use of a color scanner as a color measurement device. Accurate simulations of flatbed scanners can shorten the development cycle of new designs, increase image quality, and lower manufacturing costs. Despite the ubiquitous nature of these devices, there are relatively few publications to aid the document scanner designer in simulating various lens designs, sampling strategies, and sensor configurations. Similar challenges face the remote sensing community with numerous publications devoted to assist the engineers in designing the desired systems with minimum time and effort. Leveraging concepts from the remote sensing community, a flatbed document scanner simulation system has been developed. The process in flatbed scanners follows the general approach of the imaging chain model that is a well-understood area in the remote sensing community. The main differences, among others, between the flatbed scanning process and a remote sensing imaging system are: 1) the target of interest; 2) the geometry of the target; and 3) the source of illumination. Targets for flatbed scanners are typically printed documents or photos with a 2D geometry and illuminated with artificial light sources. On the other hand, Remote sensing targets can be of any real-world objects with various geometrical shapes in 3D-space and illuminated with sunlight. Such differences determine many factors in the scanning system such as the necessary resolution, the required spectral sensitivities, the lens and the detector type. The image chain that represents the scanning process consists of several components. First, the light source illuminates a printed document which is characterized by a certain reflectance value. The product of the light and the reflectance is then propagated through the optical element and passed to the sensor. After digitization, different image reconstruction and correction algorithms are used to produce an optimized output. In this paper, we present an end-to-end flatbed document scanner simulator. To demonstrate the efficacy of this system, we illustrate the effect of the sensor height on color artifacts created by a flashing RGB illuminant. The proposed simulation takes into consideration variables such as the intensity and duration of the illuminant, the scanning rate, the sensor aperture, the detector MTF, and finally the motion blur created by the movement of the sensor during the scanning process. These variables are modeled mathematically using Fourier analysis, functions that model the physical components, convolutions, sampling theorems, gamma corrections, and proper motion algorithms in order to produce accurate representations of practical illustrations. The input to the simulation system is a digital representation of a continuous image. The digital input image can be a single frequency pattern, a radial chirp-like pattern, or a high resolution scanned document. We compared the results with side-by-side actual scans for validation. The simulation has been found to model the scanning process effectively both on a theoretical and experimental level.

Conference 9402: Document Recognition and Retrieval XXII

Wednesday - Thursday 11–12 February 2015

Part of Proceedings of SPIE Vol. 9402 Document Recognition and Retrieval XXII

9402-1, Session Key

Printing presses and polyphonic pianos: unsupervised transcription for documents and music *(Keynote Presentation)*

Dan Klein, Univ. of California, Berkeley (United States)

No Abstract Available

9402-2, Session 1

Ground truth model, tool, and dataset for layout analysis of historical documents

Kai Chen, Ecole d'ingénieurs et d'architectes de Fribourg (Switzerland); Mathias Seuret, Hao Wei, Univ. de Fribourg (Switzerland); Marcus Liwicki, Univ. de Fribourg (Switzerland) and Technische Univ. Kaiserslautern (Germany); Jean Hennebert, Univ. de Fribourg (Switzerland) and Haute Ecole Spécialisée de Suisse occidentale (Switzerland); Rolf Ingold, Univ. de Fribourg (Switzerland)

In this paper, we propose a new dataset and a ground-truthing methodology for layout analysis of historical documents with complex layouts. The dataset is based on a generic model for ground-truth presentation of the complex layout structure of historical documents. For the purpose of extracting uniformly the document contents, our model defines five types of regions of interest: page , text, block , text line , decoration , and comment. Unconstrained polygons are used to outline the regions. A performance metric is proposed in order to evaluate various page segmentation methods based on this model. We have analysed four state-of-the-art ground-truthing tools: TRUVIZ, GEDI, WebGT, and Aletheia. From this analysis, we conceptualized and developed Divadia , a new tool that overcomes some of the drawbacks of these tools, targeting the simplicity and the efficiency of the layout ground truthing process on historical document images. With Divadia, we have created a new public dataset. This dataset contains 120 pages from three historical document image collections of different styles and is made freely available to the scientific community for historical document layout analysis research.

9402-3, Session 1

Use of SLIC superpixels for ancient document image enhancement and segmentation

Maroua M. Mehri, Univ. de La Rochelle (France); Nabil Sliti, Univ. de Sousse (Tunisia); Pierre Héroux, Univ. de Rouen (France); Petra Gomez-Krämer, Univ. de La Rochelle (France); Najoua Essoukri Ben Amara, Univ. de Sousse (Tunisia); Rémy Mullot, Univ. de La Rochelle (France)

Designing reliable and fast segmentation algorithms of ancient documents has been a topic of major interest for many libraries and the prime issue of research in the document analysis community. Thus, we propose in this article a fast ancient document enhancement and segmentation algorithm based on using Simple Linear Iterative Clustering (SLIC) superpixels and Gabor descriptors in a multi-scale approach. Firstly, in order to obtain enhanced backgrounds of noisy ancient documents, a novel foreground/background segmentation algorithm based on SLIC superpixels, is introduced. Once, the SLIC technique is carried out, the background and foreground superpixels are classified. Then, an enhanced and non-noisy background is achieved after processing the background superpixels. Subsequently, Gabor descriptors are only extracted from the selected foreground superpixels of the enhanced gray-level ancient book document images by adopting a multi-scale approach. Finally, for ancient document image segmentation, a foreground superpixel clustering task is performed by partitioning Gabor-based feature sets into compact and well-separated clusters in the feature space. The proposed algorithm does not assume any a priori information regarding document image content and structure and provides interesting results on a large corpus of ancient documents. Qualitative and numerical experiments are given to demonstrate the enhancement and segmentation quality.

9402-4, Session 1

Software workflow for the automatic tagging of medieval manuscript images (SWATI)

Swati Chandna, Danah Tonne, Thomas Jejkal, Rainer Stotzka, Karlsruher Institut für Technologie (Germany); Celia Krause, Technische Univ. Darmstadt (Germany); Philipp Vanscheidt, Hannah Busch, Univ. Trier (Germany); Ajinkya Prabhune, Karlsruher Institut für Technologie (Germany)

Digital methods, tools and algorithms are gaining in importance for the analysis of digitized manuscript collections in the arts and humanities. One example is the BMBF funded research project "eCodicology" which aims to design, evaluate and optimize algorithms for the automatic identification of macro- and micro-structural layout features of the medieval manuscripts. The main goal of this research project is to provide better insights into high dimensional datasets of medieval manuscripts for humanities scholars. The heterogeneous nature of the large humanities data and the need to create a database of automatically extracted reproducible features for better statistical and visual analysis, are the main challenges in designing a workflow for the arts and humanities.

As a solution, this paper presents a concept of a workflow for the automatic tagging of medieval manuscripts. As a starting point, the workflow uses medieval manuscripts digitized within the scope of the project "Virtual Scriptorium St. Matthias". Firstly, these digitized manuscripts are ingested into a data repository. Secondly, specific algorithms are adapted or designed for the identification of macro- and micro-structural layout elements like page size, writing space, number of lines etc. And lastly, a statistical analysis and scientific evaluation of the manuscripts groups are performed. The workflow is designed generically to process large amounts of data automatically with any desired algorithm for feature extraction. As a result, a database of objectified, reproducible features is created which helps to analyze and visualize hidden relationships of around 170,000 pages. The workflow shows the potential of automatic image analysis by enabling the processing of a single page in less than a minute. Furthermore, the accuracy tests of our workflow show that automatic and manual analysis are comparable. The usage of a computer cluster will allow the highly performant processing of large amounts of data. The software framework itself will be integrated as a service into the DARIAH infrastructure to make it adaptable for wider range of communities.

Conference 9402:
Document Recognition and Retrieval XXII

9402-5, Session 2

Math expression retrieval using an inverted index over symbol pairs

David Stalnaker, Richard Zanibbi, Rochester Institute of Technology (United States)

We introduce a new method for indexing and retrieving mathematical expressions, and a new protocol for evaluating math formula retrieval systems. The Tangent search engine uses an inverted index over pairs of symbols in math expressions. Each key in the index is a pair of symbols along with their relative distance and vertical displacement within an expression. Matching expressions are ranked by the percentage of symbol pairs matched in the query, and the percentage of symbol pairs matched in the candidate expression (using the harmonic mean). We have found that our method is fast enough for use in real time and finds partial matches well, such as when subexpressions are re-arranged (e.g. expressions moved from the left to the right of an equals sign) or when individual symbols (e.g. variables) differ from a query expression. In an experiment using expressions from English Wikipedia, student and faculty participants (N=20) found expressions returned by Tangent significantly more similar than those returned by an existing text-based retrieval system (Lucene) adapted for mathematical expressions. Participants provided similarity ratings using a 5-point Likert scale, evaluating expressions from both algorithms one-at-a-time in a randomized order to avoid bias from the position of hits in search result lists. For the Lucene-based system, precision for the top 1 and 10 hits averaged 60% and 39% across queries respectively, while for Tangent mean precision at 1 and 10 were 99% and 60%. Both an online demonstration and source code for Tangent are available.

9402-6, Session 2

Min-cut segmentation of cursive handwriting in tabular documents

Brian L. Davis, William A. Barrett, Scott D. Swingle, Brigham Young Univ. (United States)

Handwritten tabular documents, such as census, birth, death and marriage records, contain a wealth of information vital to genealogical and related research. Much work has been done in segmenting freeform handwriting, however, segmentation of cursive handwriting in tabular documents is still an unsolved problem. Tabular documents present unique segmentation challenges caused by handwriting overlapping cell-boundaries and other words, both horizontally and vertically, as “ascenders” and “descenders” overlap into adjacent cells. This paper presents a method for segmenting handwriting in tabular documents using a min-cut/max-flow algorithm on a graph formed from a distance map and connected components of handwriting. Specifically, we focus on line, word and first letter segmentation. Additionally, we include the angles of strokes of the handwriting as a third dimension to our graph to enable the segmentation of overlapping letters. Word segmentation accuracy is 89.5% evaluating lines of the data set used in the ICDAR2013 Handwriting Segmentation Contest. Accuracy is 98% for a specific application of segmenting names in a noisy census record where the names are not touching. Accuracy decreases depending on the degree of overlap with results still above 75% for word and line segmentation of tabular documents containing highly convoluted, overlapping handwriting.

9402-7, Session 2

Cross-reference identification within a PDF document

Sida Li, Liangcai Gao, Zhi Tang, Yinyan Yu, Peking Univ. (China)

Cross-references, such like footnotes, endnotes, figure/table captions, references, are a common and useful type of page elements to further explain their corresponding entities in the target document, however, the researches on this topic are quite sparse. In this paper, we address the task, cross-reference identification in a PDF document, and present a robust method as a case study of identifying footnotes, and figure references. The proposed method first extracts footnotes and figure captions and then matches them with their corresponding references within a document. A number of novel features within a PDF documents, i.e., page layout, font information, lexical and linguistic features of cross-references, are utilized for the task. And clustering is also adopted to handle the features that are stable in one document but varied in different kinds of documents so that the process of identification is adaptive with document types. In addition, this method leverages results from the matching process to provide feedback to the identification process and further improve the algorithm accuracy. The primary experiments in real document sets show that the proposed method is promising to identify cross-reference in a PDF document.

9402-8, Session 2

Intelligent indexing: a semi-automated, trainable system for field labeling

Robert Clawson, William A. Barrett, Brigham Young Univ. (United States)

We present Intelligent Indexing: a general, scalable, collaborative approach to indexing and transcription of non-machine-readable documents that exploits visual consensus and group labeling while harnessing human recognition and domain expertise. In our system, indexers work directly on the page, and with minimal context switching can navigate the page, enter labels, and interact with the recognition engine. Interaction with the recognition engine occurs through preview windows that allow the indexer to quickly verify and correct recommendations. This interaction is far superior to conventional, tedious, inefficient post-correction and editing. Intelligent Indexing is a trainable system that improves over time and can provide benefit even without prior knowledge. A user study was performed to compare Intelligent Indexing to a basic, manual indexing system. Volunteers report that using Intelligent Indexing is less mentally fatiguing and more enjoyable than the manual indexing system. Their results also show that it reduces significantly (30.2%) the time required to index census records, while maintaining comparable accuracy. (a video demonstration is available at http://youtube.com/gqdVzEPnBEw)

9402-9, Session 3

Re-typograph phase I: a proof-of-concept for typeface parameter extraction from historical documents

Bart Lamiroy, Univ. de Lorraine (France); Thomas Bouville, Atelier National de Recherche Typographique (France); Blégean Julien, Hongliu Cao, Salah Ghamizi, Univ. de Lorraine (France); Romain Houpin, Univ. de Lorraine (France); Matthias Lloyd, Univ. de Lorraine (France)

This paper reports on the fist phase of an attempt to create a full retro-engineering pipeline that aims to construct a complete set of coherent typographic parameters defining the typefaces used in a printed homogenous text. It should be stressed that this process cannot reasonably be expected to be fully automatic and that it is designed to include human interaction. Although font design is governed by a set of quite robust and formal geometric rule sets, it still heavily relies on subjective human interpretation. Furthermore, different parameters, applied to the generic rule sets may actually result in quite similar and visually difficult to distinguish typefaces, making the retro-engineering an inverse problem that is ill conditioned once shape distortions (related to the printing and/or scanning

process) come into play.

This work is the first phase of a long iterative process, in which we will progressively study and assess the techniques from the state-of-the-art that are most suited to our problem and investigate new directions when they prove to not quite adequate. As a first step, this is more of a feasibility proof-of-concept, that will allow us to clearly pinpoint the items that will require more in-depth research over the next iterations.

9402-10, Session 3

Clustering of Farsi sub-word images for whole-book recognition

Mohammad Reza Soheili, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (Germany); Ehsanollah Kabir, Tarbiat Modares Univ. (Iran, Islamic Republic of); Didier Stricker, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (Germany)

Redundancy of word and sub-word occurrences in large documents can be effectively utilized in an OCR system to improve recognition results. Most OCR systems employ language modeling techniques as a post-processing step; however these techniques do not use important pictorial information that exist in the text image. In case of large-scale recognition of degraded documents, this information is even more valuable. In our previous work, we proposed a sub-word image clustering method for the applications dealing with large printed documents. In our clustering method, the ideal case is when all equivalent sub-word images lie in one cluster. To overcome the issues of low print quality, the clustering method uses an image matching algorithm for measuring the distance between two sub-word images. The measured distance with a set of simple shape features were used to cluster all sub-word images. In this paper, we analyze the effects of adding more shape features on processing time, purity of clustering, and the final recognition rate. Previously published experiments have shown the efficiency of our method on a book. Here we present extended experimental results and evaluate our method on another book with totally different font face. Also we show that the number of the new created clusters in a page can be used as a criteria for assessing the quality of print and evaluating preprocessing phases.

9402-11, Session 3

Gaussian process style transfer mapping for historical Chinese character recognition

Jixiong Feng, Liangrui Peng, Tsinghua Univ. (China); Franck Lebourgeois, Institut National des Sciences Appliquées de Lyon (France)

Historical Chinese character recognition is very important in larger scale historical documents digitalization, but it's a very challenging problem due to lack of labeled training samples. This paper proposes a novel non-linear transfer learning method, namely Gaussian Process based Style Transfer Mapping (GP-STM). The GP-STM extends traditional linear Style Transfer Mapping (STM) by using Gaussian process and kernel methods. With GP-STM, the knowledge of printed Chinese character samples is extracted to help the recognition of historical Chinese character samples. To demonstrate our framework, we first compare different feature extraction methods; then build a modified quadratic discriminant function (MQDF) based classifier with printed Chinese character samples; finally we perform our GP-STM model on Dunhuang historical documents. Different kernels and their parameters are explored. The performance with different proportion of training samples is evaluated. The visualization examples of STM and GP-STM with different parameters are given. Experimental results show that, there is a remarkable increase of accuracy rate by nearly 15 percentage points (from 42.78% to 57.51%) after GP-STM is adopted. Compared with STM, GP-STM also has an improvement by over 8 percentage points (from 49.18% to 57.51%). These results prove the effectiveness of our model.

9402-12, Session 3

Boost OCR accuracy using iVector based system combination approach

Xujun Peng, Huaigu Cao, Raytheon BBN Technologies (United States); Premkumar Natarajan, The Univ. of Southern California, Marina del Rey (United States)

Optical character recognition (OCR) is a challenging task because most existing preprocessing approaches are sensitive to writing style, writing material, noises and image resolution. Thus, a single recognition system cannot address all factors of real document images. In this paper, we describe an approach to combine diverse recognition systems by using iVector based features, which is a newly developed method in the field of speaker verification.

Prior to system combination, document images are preprocessed and text line images are extracted with different approaches for each system, where iVector is transformed from a high-dimensional supervector of each text line and is used to predict the accuracy of OCR. We combine hypotheses from multiple recognition systems according to the overlap ratio and the predicted OCR score of text line images. We present evaluation results on an Arabic document database where the proposed method is compared against the single best OCR system using word error rate (WER) metric.

9402-13, Session 4

Exploring multiple feature combination strategies with a recurrent neural network architecture for off-line handwriting recognition

Luc Mioulet, Univ. de Rouen (France) and Airbus Defence and Space (France); Gautier Bideault, Univ. de Rouen (France); Clément Chatelain, Institut National des Sciences Appliquées de Rouen (France); Thierry Paquet, Univ. de Rouen (France); Stephan Brunessaux, Airbus Defence and Space (France)

The BLSTM-CTC is a novel recurrent neural network architecture that has outperformed previous state of the art algorithms in tasks such as speech recognition or handwriting recognition. It has the ability to process long term dependencies in temporal signals in order to label unsegmented data. This paper describes different ways of combining features using a BLSTM-CTC architecture. Not only do we explore the low level combination (feature space combination) but we also explore high level combination (decoding combination) and mid-level (internal system representation combination). The results are compared on the RIMES word database. Our results show that the low level combination works best, thanks to the powerful data modeling of the LSTM neurons.

9402-14, Session 4

Spotting handwritten words and REGEX using a two stage BLSTM-HMM architecture

Gautier Bideault, Luc Mioulet, Univ. de Rouen (France); Clément Chatelain, Institut National des Sciences Appliquées de Rouen (France); Thierry Paquet, Univ. de Rouen (France)

In this article, we propose an hybrid model for spotting word and regular expressions (REGEX) in handwritten documents. The model is made of the state-of-the-art BLSTM (Bidirectional Long Short Time Memory) neural

network for recognizing and segmenting characters, coupled with an HMM to build line models able to spot the desired sequences. Experiments on the Rimes database show very interesting results.

9402-15, Session 4

A comparison of 1D and 2D LSTM architectures for the recognition of handwritten Arabic

Mohammad Reza Yousefi, Mohammad Reza Soheili, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (Germany); Thomas M. Breuel, Technische Univ. Kaiserslautern (Germany); Didier Stricker, Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (Germany)

In this paper, we present an Arabic handwriting recognition method based on recurrent neural network. We use the Long Short Term Memory (LSTM) architecture, that have proven successful in different printed and handwritten OCR tasks.

Applications of LSTM for handwriting recognition employ the two-dimensional architecture to deal with the variations in both vertical and horizontal axis.

However, we show that using a simple pre-processing step that normalizes the position and baseline of letters, we can make use of 1D LSTM, which is faster in learning and convergence, and yet achieve superior performance.

In a series of experiments on IFN/ENIT database for Arabic handwriting recognition, we demonstrate that our proposed pipeline can outperform 2D LSTM networks.

Furthermore, we provide comparisons with 1D LSTM networks trained with manually crafted features to show that the automatically learned features in a globally trained 1D LSTM network with our normalization step can even outperform such systems.

9402-16, Session 4

Aligning transcript of historical documents using dynamic programming

Irina Rabaev, Rafi Cohen, Jihad A. El-Sana, Klara Kedem, Ben-Gurion Univ. of the Negev (Israel)

We present a simple and accurate approach for aligning historical documents with their corresponding transcription. First, a representative of each letter in the historical document is cropped. Then, the transcription is transformed to synthetic word images by representing the letters in the transcription by the cropped letters. These synthetic word images are aligned to groups of connected components in the original text, along each line, using dynamic programming. For matching we use two different feature groups: profile-based features and gradient structural and concavity (GSC) features. The method was tested on several datasets and it provides excellent results.

9402-17, Session 4

Offline handwritten word recognition using MQDF-HMMs

Sitaram N. Ramachandrula, Hewlett-Packard Labs. India (India); Mangesh Hambarde, Hewlett-Packard India Sales Pvt Ltd. (India); Ajay Patial, Hewlett Packard India Sales Pvt Ltd. (India); Dushyant Sahoo, Indian Institute of Technology Delhi (India); Shaivi Kochar, Jamia Millia Islamia Univ. (India)

We propose an improved HMM formulation for offline handwriting recognition (HWR). The main contribution of this work is using modified quadratic discriminant function (MQDF) [1] within HMM framework. In an MQDF-HMM the state observation likelihood is calculated by a weighted combination of MQDF likelihoods of individual Gaussians of GMM (Gaussian Mixture Model). The quadratic discriminant function (QDF) of a multivariate Gaussian can be re-written by avoiding the inverse of covariance matrix by using the Eigen values and Eigen vectors of it. The MQDF is derived from QDF by substituting few of badly estimated lower-most Eigen values by an appropriate constant. The estimation errors of non-dominant Eigen vectors and Eigen values of covariance matrix for which the training data is insufficient can be controlled by this approach. MQDF has been successfully shown to improve the character recognition performance [1]. The usage of MQDF in HMM improves the computation, storage and modeling power of HMM when there is limited training data. We have got encouraging results on offline handwritten character (NIST database) and word recognition in English using MQDF HMMs.

9402-18, Session Key

The internet archive: challenges and solutions for large scale document repositories *(Keynote Presentation)*

Brewster O. Kahle, Internet Archive (United States)

No Abstract Available

9402-19, Session 5

Separation of text and background regions for high performance document image compression

Wei Fan, Jun Sun, Satoshi Naoi, Fujitsu Research and Development Center Co., Ltd. (China)

We describe a document image segmentation algorithm to classify a scanned document into different regions such as text/line drawings, pictures, and smooth background. The proposed scheme is relatively independent of variations in text font style, size, intensity polarity and of string orientation. It is intended for use in an adaptive system for document image compression. The principal parts of the algorithm are the generations of the foreground and background layers and the application of hierarchical singular value decomposition (SVD) in order to smoothly fill the blank regions of both layers so that the high compression ratio can be achieved. The performance of the algorithm, both in terms of its effectiveness and computational efficiency, was evaluated using several test images and showed superior performance compared to other techniques.

9402-20, Session 5

Metric-based no-reference quality assessment of heterogeneous document images

Nibal Nayef, Jean-Marc Ogier, Univ. de La Rochelle (France)

No-reference image quality assessment (NR-IQA) aims at computing an image quality score that best correlates with human perceived image quality and/or an objective quality measure, without any prior knowledge of reference images. Although leaning-based NR-IQA methods have achieved the best state-of-the-art results so far, those methods perform well only

on the datasets on which they were trained. The datasets usually contain homogeneous documents, whereas in reality, document images come from different sources. Documents can be scanned or captured by different cameras, they have different image characteristics and varying layouts. It is unrealistic to collect training samples of images from every possible capturing device and every document type.

Hence, we argue that a non-learning IQA method is more suitable for heterogeneous documents. We propose a NR-IQA method for the objective of OCR accuracy. The method combines distortion-specific quality metrics.

The final quality score is calculated taking into account the proportions of, and the dependency among different distortions. Experimental results show that the method achieves competitive results with learning-based NR-IQA methods on standard datasets, and performs better on heterogeneous documents.

9402-21, Session 6

Clustering header categories extracted from web tables

George Nagy, Rensselaer Polytechnic Institute (United States); David W. Embley, Brigham Young Univ. (United States); Mukkai Krishnamoorthy, Rensselaer Polytechnic Institute (United States); Sharad Seth, Univ. of Nebraska-Lincoln (United States)

Revealing related content among heterogeneous web tables is part of our long term objective of formulating queries over multiple sources of information. Two hundred HTML tables from institutional web sites are segmented and each table cell is classified according to the fundamental indexing property of row and column headers. The categories that correspond to the multi-dimensional data cube view of a table are extracted by factoring the (often multi-row/column) headers. To reveal commonalities between tables from diverse sources, the Jaccard distances between pairs of category headers (and also table titles) are computed. We show how about one third of our heterogeneous collection can be clustered into a dozen groups that exhibit table-title and header similarities that can be exploited for queries.

9402-22, Session 6

A diagram retrieval method with multi-label learning

Songping Fu, Xiaoqing Lu, Peking Univ. (China); Lu Liu, Jingwei Qu, Institute of Computer Science & Technology, Peking University (China); Zhi Tang, Institute of Computer Science & Technology, Peking University (China) and State Key Laboratory of Digital Publishing Technology, Beijing (China)

In recent years, the retrieval of plane geometry figures (PGFs) has attracted increasing attention in the fields of mathematics education and computer science. However, the high cost of matching complex PGF features leads to the low efficiency of most retrieval systems. This paper proposes an indirect classification method based on multi-label learning, which improves retrieval efficiency by reducing the scope of compare operation from the whole database to small candidate groups. Label correlations among PGFs are taken into account for the multi-label classification task. The primitive feature selection for multi-label learning and the feature description of visual geometric elements are conducted individually to match similar PGFs. The experiment results show the competitive performance of the proposed method compared with existing PGF retrieval methods in terms of both time consumption and retrieval quality.

9402-24, Session 6

Detection of electrical circuit elements from documents images

Paramita De, Sekhar Mandal, Amit Kumar Das, Indian Institute of Engineering & Technology, Shibpur (India); Bhabatosh Chanda, Indian Statistical Institute, Kolkata (India)

In this paper a method to detect the electrical circuit elements from the scanned images of electrical drawings is proposed. The method,

based on histogram analysis and mathematical morphology, detects the circuit elements, for example, circuit components, wires, and generates a connectivity matrix which may be used to find similar, but spatially different looking circuit using graph isomorphism. The work may also be used for vectorization of the circuit drawings utilising the information on the segmented circuit elements and corresponding connectivity matrix.

The novelty of the method lies in its simplicity and adaptability to work with a tolerable skewed image and the capability to segment symbols irrespective of their orientation.

The proposed method is tested over a data-set containing more than one hundred scanned images of a variety of electrical drawings.

Some of the results are presented in this paper to show the efficacy and robustness of the proposed method.

9402-25, Session 7

Missing value imputation: with application to handwriting data

Zhen Xu, Sargur N. Srihari, Univ. at Buffalo (United States)

Missing values make pattern analysis difficult, particular with limited available data. In longitudinal research, missing values accumulate, thereby aggravating the problem. Here we consider how to deal with temporal data with missing values in handwriting analysis. In the task of studying development of individuality of handwriting, we encountered the fact that feature values are missing for several individuals at several time instances. Six algorithms, i.e., random imputation, mean imputation, most likely value imputation, and three methods based on Bayesian network (static Bayesian network, parameter EM, and structural EM), are compared with children's handwriting data. We evaluate the accuracy and robustness of the algorithms under different ratios of missing data and missing values, and useful conclusions are given. Specifically, static Bayesian network is used for our data which contain around 5% missing data to provide adequate accuracy and low computational cost.

Conference 9403: Image Sensors and Imaging Systems 2015

Monday - Tuesday 9–10 February 2015

Part of Proceedings of SPIE Vol. 9403 Image Sensors and Imaging Systems 2015

9403-1, Session 1

2.2um BSI CMOS image sensor with two layer photo-detector

Hiroki Sasaki, Ai Mochizuki, Yuki Sugiura, Ryoji Hasumi, Kentaro Eda, Yoshitaka Egawa, Hirofumi Yamashita, Kenji Honda, Tatsuya Ohguro, Hisayo S Momose, Hiroshi Ootani, Toshiba Corp. (Japan); Yoshiaki Toyoshima, Toshiba Materials Co., Ltd. (Japan); Tetsuya Asami, Toshiba Corp. (Japan)

Multi-layer photo-detector pixel in CMOS imagers is one of the alternative color acquisition solutions to conventional single-layer photo-detector (1LPD) solution. Several reports have already been reported and low color aliasing for high spatial frequency images have been attained [1][2][3][4]. This paper reports on a study of two-layer photo-detector 2LPD structure implemented in BSI architecture to see a feasibility of a small pixel 2LPD BSI as a possible solution for better low light SNR. For recent small pixels, low light SNR is dominated by photon shot noise and increasing the number of photons that photo-detector can receive is the only way to improve low light SNR. Theoretical investigation of luminance SNR (YSNR) improvement in two-layer photo-detectors for small size pixel suggests that Magenta-green CFA checker pattern is one of the possible solutions to improve low light SNR [5]. In this study, we focused on color signal variation in 2LPD that is enhanced by BSI process, especially backside thinning, and crosstalk from adjacent pixels.

A 2.2um BSI CMOS imager with a 2LPD test pixel array with a Magenta-Green CF array has been fabricated and evaluated. Magenta pixel (2.2um x 4.4um) has Blue/Red 2LPDs and vertical charge transfer (VCT) path for Blue photo-detectors implemented on back-side Si surface. Green pixel (2.2um x 2.2um) has 1LPD. 2LPDs and VCTs were implemented by high-energy ion implantation from front side. Measured spectral response curves from 2LPD fitted well with those estimated based on light-absorption theory for Silicon detectors. The number of photons absorbed between depth t1 and depth t2 at wavelength L, Nabs(t1,t2,L) is described by the formula: Nabs(t1,t2,L) = N0(L)(exp(-a (L)t1)- exp(-a(L)t2)), where N0(L) is the number of photons entering the silicon surface and a(L) is the wavelength dependent absorption coefficient. N0(L) is calculated by measured and calculated data from 1LPD. We fabricated different thickness of 2LPD. To measure the thickness of 2LPD, these structures were observed by Scanning Spreading Resistance Microscopy (SSRM). Compared 2.5um to 2.8um, normalized sensitivity of Green in 1LPD decreased 96%, whereas that of Blue in 2LPD decreased 78% and that of Red in 2LPD increased 115%. This result shows that a variation of backside thinning in BSI process is caused a great color signal variation.

We fabricated test patterns of the VCT which is isolated from a backside photo-detector and evaluated a crosstalk to the VCT path from adjacent pixels. The crosstalk increases with metal shield edge even from coverage 0.2um to 0.1um. When 2LPD is implemented in a smaller pixel than 2.2um, light shielding metal coverage should be smaller otherwise optical aperture of Magenta pixel decreases and hence QE for magenta pixel gets lower. Therefore, to achieve a low crosstalk structure for vertical charge transfer path is one of the key issues for 2LPD in BSI pixel.

Our measurement results show that the keys to realize 2LPD in BSI pixel are; (1) reduction of crosstalk to the vertical charge transfer path from adjacent pixels and (2) controlling Si photo-detector thickness variance to reduce color signal variation.

[1] Merrill, R., “Color Separation in an Active Pixel Cell Imaging Array Using a Triplewell Structure,” U S. Patent 5,965,875, 1999.

[2] Lyon, R., Hubel, P., “Eyeing the Camera: Into the Next Century”, IS&T/TSID 10th Color Imaging Conference Proceedings, Scottsdale, AZ, USA; 2002 pp. 349-355

[3] Findlater, K.M., Renshaw, D., Hurwitz, J.E.D., Henderson, R.K., Purcell, M.D., Smith, S.G. & Bailey, T.E.R. 2003, “A CMOS image sensor with a double-junction active pixel”, IEEE Transactions on Electron Devices, vol. 50, no. 1, pp. 32-42.

[4] Tweet, D.J., Lee, J.-., Speigle, J.M. & Tamburrino, D. 2009, “2PFCTM Image Sensors: Better Image Quality at Lower Cost”, Proceedings of SPIE - the International Society for Optical Engineering, 7250, art. no. 725007

[5] Fossum, E.R. 2011, “Investigation of Two-Layer Photodetectors for YNSR10 Improvement in Submicron Pixels”, Proceedings of International Image Sensor Workshop 2011

9403-2, Session 1

A compact THz imaging system

Aleksander Sešek, Andrej Švigelj, Janez Trontelj, Univ. of Ljubljana (Slovenia)

The objective of this paper is the development of a compact low cost imaging THz system, usable for observation of the objects near to the system and also for stand-off detection. The performance of the system remains at the high standard of more expensive and bulkiest system on the market. It is easy to operate as it is not dependent on any fine mechanical adjustments. As it is compact and it consumes low power, also a portable system was developed for stand-off detection of concealed objects under textile or inside packages. The requirements listed rule out all optical systems like Time Domain Spectroscopy systems which need fine optical component positioning and requires a large amount of time to perform scan and the image capture pixel-by-pixel. They are also almost unusable for stand-off detection due to low output power.

The selected principle of operation is the use of a solid state THz source which is compact and portable but it remains the most expensive building block. The THz source core is VDI’s 13GHz source and 24x multiplication chain with highest output power up to 4mW or alternatively the same setup with Spacek source and highest output power of 300μW. The THz source illuminates the target with a pulse modulated signal with 100% modulation index and with a modulation frequency high enough for fast image frame rate. The modulation frequency ranges from 1 kHz to 3 kHz.

The key parameter of imaging system is the number of image pixels. This determines the scanning speed in the case of single sensor or the sensor array dimension for the staring system. The described system is a combination of both approaches using 2 lines of linear sensor array in the x direction of the image and a scanning technique with pivoting mirror for the y direction. The linear array uses both lines of sensors with the equal central operating frequency, but it also allows the usage of sensors with different operating frequencies. In that case so called “multicolor” THz imaging is possible, especially usable in drug continence detection and material properties determination.

With this approach the repetition rate of the image is up to 10 frames per second, allowing near real time operation. The system operates in reflection mode and with some modifications also in transmission mode. Both modes are described in the paper and some imaging results are shown.

For the sensor array, both the response time and the noise equivalent power are critical parameters to obtain acceptable range and resolution of the THz camera. The response time of the chosen microbolometer sensor is less than 1μs and its noise equivalent power was measured to be down to 5pW/√Hz.

Using a single lens, 10m range is obtained for a 1024 pixel THz image. The illuminated area at the 5m stand-off distance is about 1m2 but the scanned area is about 0,2m x 0,2m to ensure uniform illumination.

Signal processing hardware and software are discussed in the paper and the proprietary THz sensor is presented and described.

9403-3, Session 1

Signal conditioning circuits for 3D-integrated burst image sensors with on-chip A/D conversion

Rémi Bonnard, Josep Segura Puchades, Fabrice Guellec, CEA-LETI (France); Wilfried Uhring, Institut de Physique et Chimie des Matériaux de Strasbourg (France)

Ultra high speed (UHS) image sensors are a cutting edge field of imaging. UHS imaging is mainly used for study high speed phenomenon (detonics, plasma forming and laser ablation). In the last few years, important works have been made on this topic both in CCD [1] and CMOS [2] technology. At ultra high speed, the data read-out is a bottleneck in term of speed. Therefore the sensor works in burst mode. It acquires a burst of images and stores it in an on-chip memory. Then the burst is read-out of the memory at low speed. In the past few years, 3D integration appeared as a key technology for imaging. It is especially true for UHS imaging as this technology allows highly parallel data acquisition to increase the frame rate. 3D integration permits a higher surface dedicated to electronic without impacting the fill-factor increasing both memory depth (number of frame per burst) and sensitivity.

We proposed in [3] a 3D-integrated burst image sensor with on-chip A/D conversion. Thanks to 3D integration the conversion is performed during the burst acquisition. On the top layer a cluster of pixels multiplexes the signal to the ADC on the middle layer. The ADC feeds digital memories on the bottom layer with the burst of images. As 3D integration allows heterogeneous integration, an analog technology node is used for the sensing circuits and an advanced technology node is used for the memories. This sensor frame rate is 5 Mfps with more than 1000 frames per burst and for a pixel pitch of 50µm. As far as we know, there is no such UHS image sensor performing the A/D conversion and it exceeds the state of the art by a factor 5 in term of memory depth.

To implement this image sensor, it is necessary to design an analog front-end that performs the current integration, the global shutter acquisition and multiplexes the pixels to the ADC. We describe and compare here three structures of analog front-ends. The first architecture based on source follower buffer shows a dynamic range allowing a 8 bits A/D conversion and a sensitivity of 700nA (1400 W.m-2). The second based on unbiased transistors reaches the same dynamic and sensitivity for a power consumption divided by 5. The third architecture based on direct injection performs the integration on a chosen capacitance thus reaching a sensitivity of 2 nA (4 W.m-2) for the same dynamic. For those three structures the read-out noise is below the half LSB. The results are based on post layout simulation of those structures implemented in a standard cmos 0.18µm technology.

[1] Etoh, T.G et al., A 16 Mfps 165kpixel backside-illuminated CCD, Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2011 IEEE International, pp.406,408, (2011)

[2] Tochigi, Y et al., A global-shutter CMOS image sensor with readout speed of 1Tpixel/s burst and 780Mpixel/s continuous, ISSCC, 2012 IEEE International, pp. 382 384, (2012)

[3] Bonnard, R., Guellec, F., Segura, J., Dupret, A., Uhring, W., New 3D-Integrated Burst Image Sensor Architectures with in-situ A/D conversion, Proc. DASIP 2013, (2013)

9403-4, Session 1

A 4MP high-dynamic-range, low-noise CMOS image sensor

Cheng Ma, Jilin Univ. (China) and Gpixel Inc. (China); Yang Liu, Jing Li, Quan Zhou, Gpixel Inc. (China); Yuchun Chang, Jilin Univ. (China); Xinyang Wang, Gpixel Inc. (China)

In this paper we present a 4M-pixel (2048 x 2048) high dynamic range, low noise and low dark current CMOS image sensor which is very suitable for scientific imaging and high-end surveillance applications. The pixel is designed based on a 4-T PPD pixel. During the readout of the array, pixel signals are first amplified and then feed to a low power column parallel ADC array which is already presented in [1]. Measurement results show that the sensor achieves a full well of 120ke- and a dark read noise of less than 1.7e- at 24fps speed.

In a 4-T PPD based CMOS image sensor, normally the dark read noise is dominated by the in-pixel source follower noise and the analog signal processing noise if there is no other specific noise (such as KTC noise of in-pixel sampling cap) induced. In this design, we try to increase the conversion gain of the floating diffusion and amplification factor in the analog processing to lower the total read noise. While this has a limitation of full well as in this case, the swing in the floating diffusion is the limiting factor. In order to readout the full well, we tried the same method as in [2] by a second charge transfer and high FD capacitance.

Unlike the conventional design which stores the charges on the gate of a MOS cap, we store the charges in the MOS channel so one transistor is saved and this gives us more fill factor in the pixel. The pixel is schematically shown in Figure 1. Figure 2 shows the sensor architecture. Two analog readout chains are used for each pixel to both output the high gain and low gain pixel value, the high gain chain is optimized for low readout noise and the low gain chain is optimized for high full well capacity.

In this design, we use a low power column parallel ADC array [1] to digitize all the amplified pixel analog signals. The conversion algorithm is shown in Figure 3. Basically two clocks are used to count the time interval between t1 and t7. Most of the time, the column counter is counting with the low frequency clock. Only within one CLOCK LOW period after the comparator flipping, the column counter is counting with high frequency clock. So a lot of power can be saved and at the same time the substrate is much “quiet” compared to the conventional counting methods as in [3] and [4].

In below sections, a detailed description of the sensor design and pixel analysis will first be presented, followed by the chip characterization results and analysis.

9403-5, Session 1

Multi-camera synchronization core implemented on USB3 based FPGA platform

Ricardo M. Sousa, Univ. da Madeira (Portugal); Martin Wäny, Pedro Santos, AWAIBA Lda. (Portugal); Morgado Dias, Univ. da Madeira (Portugal) and Madeira Interactive Technologies Institute (Portugal)

Centered on Awaiba’s NanEye CMOS image sensor family and an FPGA platform with USB3 interface, the aim of this paper is to demonstrate a new technique to synchronize up to 8 individual self-timed cameras with minimal error. Small form factor self-timed camera modules of 1 mm x 1 mm or smaller do not normally allow external synchronization. However, for stereo vision or 3D reconstruction with multiple cameras as well as for applications requiring pulsed illumination it is required to synchronize multiple cameras. In this work, the challenge of synchronizing multiple self-timed cameras with only 4 wire interface has been solved by adaptively regulating the power supply for each of the cameras. To that effect, a control core was created to constantly monitor the operating frequency of each camera by measuring the line period in each frame based on a well-defined sampling signal. The frequency is adjusted by varying the voltage level applied to the sensor based on the error between the measured line period and the desired line period. To ensure phase synchronization between frames, a Master-Slave interface was implemented. A single camera is defined as the Master, with its operating frequency being controlled directly through a PC based interface. The remaining cameras are setup in Slave mode and are interfaced directly with the Master camera control module. This enables the remaining cameras to monitor its line and frame period and adjust their own to achieve phase and frequency synchronization.

The control core was tested in a laboratory environment with up to 4

cameras. Results indicate it is able to maintain perfect synchronization in changing ambient temperatures. Tests show that even in the presence of temperature gradients of 60 ºC between cameras, phase and frequency synchronization is still preserved. The temperature range (tested from -2 ºC to 60 ºC) is only limited by the self-imposed voltage limits to the camera supply. Furthermore, the average phase error between frames was measured to be 3.77 μs. Taking into account that the nominal frame rate of a NanEye camera is 44 fps this means the average error is approximately 0.017%. The control module has also shown it is independent from cable length and NanEye camera version. The result of this work will allow the realization of smaller than 3mm diameter 3D stereo vision and 3D meteorology equipment in medical endoscopic context, such as endoscopic surgical robotic or micro invasive surgery.

9403-6, Session 2

Compressed hyperspectral sensing

Grigorios Tsagkatakis, Foundation for Research and Technology-Hellas (Greece); Panagiotis Tsakalides, Foundation for Research and Technology-Hellas (Greece) and Univ. of Crete (Greece)

A fundamental problem that HyperSpectral Imaging sensors must address is how to collect the three dimensional HSI data, along the two spatial and the one spectral dimensions, using either a single, a 1D array, or a 2D array of detectors. The discrepancy between the required and the available dimensionality of detectors has sparked different philosophies in HSI acquisition system designs, each one with its own specific capabilities in terms of spatial, temporal, and spectral resolution. A key shortcoming shared by all current methods lies in the high scanning repetition rates required for generating the complete 3D hyperspectral datacube. In the case of spatial/spectral scanning, multiple lines/pixels have to be scanned, while for 2D frame scanning systems, multiple frames have to be acquired in order to obtain the complete spectral profile of the scene. This limitation is responsible for a number of issues that hinder HSI performance, including slow acquisition time and motion artifacts. To address these limitations, Snapshot Spectral Imaging (SSI) systems acquire the complete spatio-spectral cube from a single or a few captured frames, i.e., during a single or a few integration periods, without the need for successive frame acquisition.

In this work, we proposed a novel SSI-HSI architecture that can achieve high quality reconstruction of the hypercube from a limited number of frames, without resorting to moving parts, by exploiting the theory of Compressed Sensing (CS). According to the CS framework, perfect reconstruction of a signal is possible from a small number of random measurements far below the typical Shannon-Nyquist sampling limit, provided the signal can be sparsely represented in a collection of elementary examples

The proposed system is composed of the following elements: (i) a coding mask that can allow or block the incoming light according to a specific dynamically changing sampling pattern. Such a sampling mechanism can be implemented using a Digital Micromirror Device (DMD); (ii) an array of optical filters that filter the incoming light allowing only a specific set of spectral bands to propagate; and (iii) an array of lenses, also called a Lenslet array, that focus the filtered light onto the imaging sensor, such as a CCD or a CMOS device. During each sampling instance, the imaging sensor captures light from the entire scene, however the recorded values correspond to a mixture of spectral bands, defined by the specific coding. Recovery of the spectral profile at a specific location is then formulated as l1 minimization problem.

Software simulations were carried out to understand the recovery performance of AVIRIS HSI data using the proposed imaging architecture, where a binary sampling pattern of 10 active elements per frame and a dictionary consisting of 280 training hyperspectral profiles were employed. While traditional imaging with the proposed architecture would require the acquisition of 220 spectral frames, the proposed architecture exhibited exceptional performance from a few as 10 frames. The results suggest that a high quality reconstruction of the full hypercube is possible from a significantly smaller number of frames thanks to the proposed efficient encoding and decoding procedures.

9403-8, Session 2

Hyperspectral atmospheric CO2 imaging system based on planar reflective grating

Xueqian Zhu, Lei Ding, Xinhua Niu, Shanghai Institute of Technical Physics (China)

In order to measure the high precision absorption spectrum of CO2, fit the research requirements of the source and sink of CO2, a kind of hyperspectral imaging system is designed. It includes three channels, the 1.6μm waveband can get a clearly absorption spectrum of CO2, but affected by many other factors, the precision can hardly meet the requirements. So the others two channels are indispensable. The 0.76μm waveband can show the atmospheric optical path accurately, which can get rid of the affection of the scatter caused by clouds and aerosols, and the 2.06 waveband can detect the atmospheric pressure and hygral change in the upper air. The information collected by all the channels can provide good supplement and perfection for the inversion data, and finally realize the hyperspectral measurement of the source and sink of CO2.

The optical system includes two parts, the fore-optics and the tri-channel grating spectrometer system. The fore-optics consists of two symmetric Cassegrain telescopes, which can reduce the effective aperture and the size of the optical elements behind. The three channels of the grating spectrometer system work on a same slip angle, so every channel has a similar structure in the theoretical design, contains a large planar reflective grating . The mechanical structure matched is in process now.

It already has a demonstration system to verify the performance and indices of the 1.6μm waveband. The InGaAs Infrared Focal Plane Array (FPA) with a low noise acquisition system is served to provide low noise digital signals. This FPA has a pixel array of 640x512, 25μm pixel pitch, is sensitive to 0.9μm to 1.7μm short wave infrared (SWIR) band. The detector pixel operability is 99.91%. With the experiments processed in the library and external field, we can achieve the results such as signal to noise ratio (SNR) and modulation transfer function (MTF). Analyzing all these data can help to estimate the system's property and demonstrate the whole system meets the requirements of hyperspectral measurement.

9403-9, Session 2

Design, fabrication and characterization of a polarization-sensitive focal plane array

Dmitry Vorobiev, Zoran Ninkov, Rochester Institute of Technology (United States)

We present the design, fabrication and characterization of the Rochester Institute of Technology Polarization Imaging Camera (RITPIC), a snapshot polarimeter for visible and near-infrared remote sensing applications. RITPIC is a compact, light-weight and mechanically robust imaging polarimeter that is deployable on terrestrial, naval, airborne and space-based platforms. RITPIC is developed using commercially available components and is capable of fast cadence imaging polarimetry of a wide variety of scenes, over a broad spectral range. To derive the polarization properties of a scene, we employ a variation of the division-of-focal plane modulation strategy. RITPIC is fabricated by hybridizing a MOXTEK, Inc. micropolarizer array (MPA) with a Truesense KAF-1603 CCD. The result is a "general purpose" polarization-sensitive imaging sensor, which can be placed at the focal plane of a wide number of imaging systems (and even spectrographs). We present our efforts to date in developing this technology and examine the factors that fundamentally limit the performance of these devices. Finally, we identify some applications for which these devices appear ideally suited.

9403-10, Session 2

High dynamic, spectral, and polarized natural light environment acquisition

Philippe Porral, Patrick Callet, Philippe Fuchs, Thomas Muller, Mines ParisTech (France); Etienne Sandré-Chardonnal, Eclat Digital Recherche (France)

In the field of the image synthesis, the simulation of the material's appearance requires the rigorous resolution of the light transport equation. This implies to take into account all the elements that may have an influence on the spectral radiance, and that are perceived by the human eye. Obviously, the reflectance properties of the materials have a major impact in the calculations, but other significant properties of light such as the spectral distribution and the polarization must be taken into account, in order to expect correct results. The image rendering under natural light is very dependent of a rigorous characterization of the source, unfortunately real maps of the polarized or spectral environment corresponding to a real sky do not exist, except only for a few simplistic parametric models. Therefore, it seemed necessary to focus our work on capturing such data, in order to have a system to quantify all the properties and capable of powering our future simulations in a renderer.

In this work, we develop and we characterize a device designed to capture the entire light environment, by taking into account both the dynamic range of the spectral distribution and the polarization states, in a measurement time of less than two minutes. Sky light is collected by a fisheye lens through a series of polarizing and bandpass filters mounted on two motorized wheels. Each sequence of (5 x 21 = 105) images is recorded by a "Wide Dynamic Range" CMOS sensor. Thereafter, the images are processed and transcribed in a data format, inspired by polarimetric imaging and fitted for a spectral rendering engine, which exploits the "Stokes-Mueller formalism."

We are assuming that the proper consideration of this new information will improve:

- The accuracy of color and aspect reproduction

- Efficiency of the simulation for several specific effects (such as absorption, dispersion, diffraction, interference, etc.),

- Anticipate the metamerism phenomena.

9403-11, Session 2

A high-sensitivity 2x2 multi-aperture color camera based on selective averaging

Bo Zhang, Keiichiro Kagawa, Taishi Takasawa, Min-Woong Seo, Keita Yasutomi, Shoji Kawahito, Shizuoka Univ. (Japan)

To demonstrate the low-noise performance of the multi-aperture imaging system, an ultra-high-sensitivity multi-aperture color camera with 2?2 apertures is being developed. In low-light conditions, random telegraph signal (RTS) noise and dark current white defects become visible that greatly degrades the quality of the image. RTS noise is generated by capturing and emission of carriers in the channel of MOSFET randomly by the traps near the silicon-silicon dioxide interface. The RTS noise in CMOS image sensors (CISs) is a large issue especially in low-light applications. In addition, as the transistor scales down, RTS noise will be larger.

To reduce the RTS noise and dark current white defects in low-light conditions as well as to increase the number of incident photon, the redundancy of the multi-aperture imaging system is exploited. In the multi-aperture imaging system, an array of a lens and a CIS is utilized. One lens and one sensor constitute an aperture like a traditional single-aperture camera. The same sensors are used in every aperture and all of the apertures capture the same scene at the same time.

In the multi-aperture imaging system, multiple images can be captured simultaneously. However, the noise levels of the corresponding pixels for a subjective point are different. Here, we combine the multiple images to reproduce the single final image with lower noise by the noise reduction method called a selective averaging. This method is operated pixel by pixel. A pixel in the reproduced image has multiple sub-pixels in every aperture. In preparation, we calculate the variance of sub-pixel value in the dark condition. Then, during image capturing, we sort those variances from the minimum to the maximum. After that, a combination variance is calculated based on the synthetic variance, that is, the equation V_m=(1/m^2)sum{v_i} (i=1, ..., m). V_m is the combination variance; v_i is the sorted variance; m is the number of apertures, which changes from 1 to the number of all apertures. We find out the minimum combination variance and define the sub-pixels which were used to calculate the minimum combination variance as the selected sub-pixels for the virtual pixel. The pixel value of the final image is calculated by averaging the pixel values of the selected sub-pixels only. A large noise causes a large variance. If the variance is relatively large, the combination variance for m apertures can be larger than the combination variance for m-1 apertures. Thus, the sub-pixels with large noise are automatically excluded. In simulation, the effective noise in the peak of noise histogram is reduced from 1.38e- to 0.48 e- in a 3?3-aperture system. The RTS noise and dark current white defects has been successfully removed.

In this work, low-noise color sensors with 1280?1024 pixels fabricated in 0.18um CIS technology are used. The pixel pitch is 7.1?m?7.1?m. The noise of the sensor is around 1e- based on the folding-integration and cyclic column ADCs. The low voltage differential signaling (LVDS) is applied to improve the noise immunity. A synthetic F-number of 0.6 will be achieved with 2?2 apertures in the prototype.

9403-29, Session 2

Acousto-optic imaging with a smart-pixels sensor

Kinia Barjean, Univ. Paris 13 (France); Kevin Contreras, Jean-Baptiste Laudereau, Institut Langevin (France); Eric Tinet, Dominique Ettori, Univ. Paris 13 (France); François Ramaz, Institut Langevin (France); Jean-Michel Tualle, Univ. Paris 13 (France)

Acousto-optic imaging (AOI) is an emerging technique in the field of biomedical optics which combines the contrast allowed by diffuse optical tomography with the resolution of ultrasound (US) imaging . The US wave modulates both the refractive index and the scattering particles position. This creates sidebands around the laser frequency with a shift equal to the US wave frequency . By filtering the resulting frequency-shifted photons, the so-called tagged photons, one can recover the local light irradiance within the turbid medium. The frequency shifts here considered, in the range of a few MHz, are very small compared to the light frequency. Although high finesse spectral filters, using spectral hole-burning at cryogenic temperatures, have proven their ability to discriminate the tagged photons , a lot of groups are currently working on interferometric detection schemes, which seem to be more accessible for medical applications .

Interferometry presents the advantage of moving the frequency shift into more accessible regions. However one has to deal with the limited spatial coherence of diffuse light, which appears as a speckle pattern. The short correlation time of this speckle pattern encountered with biological tissue, in the submillisecond range, is also an issue. Moreover, the signal flux is very low. Another difficulty concerns the axial resolution, along the US beam propagation axis: the use of US pulses is not always the best solution as the tagged signal is really weak in that case. In this work we have implemented a CMOS smart-pixels sensor dedicated to the real-time analysis of speckle patterns . A highly sensitive lock-in detection we implemented in each pixel allows us to extract the tagged photons after an appropriate in-pixel post-processing. Axial resolution is obtained through a new method, Fourier-Transform AOI, based on a sine modulation of the acoustic beam. In comparison to existing techniques used in this field, this method should allow to get a consequent improvement of the signal to noise ratio. With this system we can acquire images in scattering samples with a spatial resolution in the 2mm range, with an integration time compatible with the dynamic of living biological tissue.

9403-12, Session 3

Simulation analysis of a backside illuminated multi-collection gate image sensor

Vu Truong Son Dao, Takeharu Goji Etoh, Ritsumeikan Univ. (Japan); Edoardo Charbon, Zhang Chao, Technische Univ. Delft (Netherlands); Yoshinari Kamakura, Osaka Univ. (Japan)

We have proposed a backside-illuminated multi-collection gate (BSI MCG) image sensor. Each pixel has a group of collection gates (CG) located around its center. The image sensor is divided into groups of pixels; each group is vertically connected to a ring oscillator (RO) unit mounted on a different wafer.

We designed a test chip including an imaging device and a RO driver, though, at this moment; these devices were not stacked. The imaging device consists of: (a) 32x48 pixels driven by a conventional driver; (b) 1x2 pixels driven by the test RO driver from a separate dice. Each pixel is a hexagonal MCG BSI one that stores 5 consecutive frames.

We obtained the following simulation results:

(1) In each pixel, mean and standard deviation of the electron travel time were 0.62ns and 0.17ns, respectively. The maximum travel time was from 0.6ns to 1.4ns if a generation site was near the center to near the edge of a pixel.

(2) The RO driver can achieve a pulse width of 1.4ns with a voltage swing of 4.2V and 20% overlapping pulses. The minimum pulse width reduced to 0.77ns with a decreased voltage swing of 2.6V and 25% overlapping pulses.

Therefore, we can confirm that the proposed test chip can achieve the target.

9403-13, Session 3

Analysis of pixel gain and linearity of CMOS image sensor using floating capacitor load readout operation

Shunichi Wakashima, Fumiaki Kusuhara, Rihito Kuroda, Shigetoshi Sugawa, Tohoku Univ. (Japan)

Small and low power image sensors are used in various fields such as smartphones, medical endoscopes and cameras for industrial robot arms. The demand for low power image sensors will increase with the development of communication technologies in the near future. To achieve lower power CMOS image sensors, the pixel signal readout operation without column current sources has been reported [1][2]. In [1], we reported a small, low power and low noise CMOS image sensor using floating capacitor load readout operation. However, distinctive characteristics of the pixel gain and linearity range of a CMOS image sensor using floating capacitor load readout operation were not mentioned in detail. These characteristics are important when decreasing the power supply voltage for the lower power consumption. In this paper, we demonstrate that the floating capacitor load readout operation has higher readout gain and wider linearity range than conventional pixel readout operation, and report the reason.

In conventional CMOS image sensor, the in-pixel driver transistor drives constant current when pixel signal read out. The readout gain is determined by the transconductance, the backgate transconductance and the output resistance of the in-pixel driver transistor and the load resistance. In floating capacitor load readout operation, the in-pixel driver transistor is connected to a sample/hold capacitor which has been reset to 0V and set floating, then in-pixel driver transistor charges up the sample/hold capacitor. Since there is no current source and the load is the sample/hold capacitor only, the load resistance approaches infinity. Therefore readout gain is larger than that of conventional readout operation. Moreover, since current flow is almost nothing at the end of signal readout, floating capacitor load readout operation suppresses decreasing of readout gain due to the voltage drop in pixel select transistor. By these effects, floating capacitor load readout operation has lager readout gain than conventional readout operation.

In conventional readout operation, output voltage is lower than input voltage by the gate-to-source voltage to drive the constant current and by the voltage drop in pixel select transistor and parasitic resistance of pixel output vertical signal line. And the output voltage needs to be higher than the drain-to-source voltage of current source which drives in saturation region. In floating capacitor load readout operation, the gate-to-source voltage is lower and the voltage drop is almost nothing because the current flow is almost nothing at the end of signal readout. Therefore the linearity range is enlarged for both high and low voltage limits in comparison to the conventional readout operation.

The effect of linearity range enlargement becomes more advantageous when decreasing the power supply voltage for the lower power consumption. To confirm these effects, we fabricated a prototype chip using 0.18um 1-Poly 3-Metal CMOS process technology with pinned PD. The chip size is 2.5mm(H)x3.4mm(V), the pixel size is 2.8um(H)x2.8um(V), and the number of pixels is 1140(H)x768(V). As a result, we confirmed that floating capacitor load readout operation increases both readout gain and linearity range. This operation is effective for lower power image sensors.

[1]S. Wakashima et al., "A CMOS Image Sensor using Floating Capacitor Load Readout Operation," Proc. Int. IS&T/SPIE Electronic Imaging, vol. 8659, 86590I-1-86590I-9 (2013).

[2]B. Cremers, et al., "A 5Megapixel, 1000fps CMOS Image Sensor with High Dynamic Range and 14-bit A/D Converters," Proc. Int. IISW, pp.381-383 (2013).

9403-14, Session 3

Addressing challenges of modulation transfer function measurement with fisheye lens cameras

Brian M. Deegan, Patrick E. Denny, Vladimir Zlokolica, Barry Dever, Laura Russell, Valeo Vision Systems (Ireland)

Modulation transfer function (MTF) is a well defined and accepted method of measuring image sharpness. The slanted edge test, as defined in ISO12233 is a standard method of calculating MTF, and is widely used for lens alignment and auto-focus algorithm verification. However, there are a number of challenges which should be considered when measuring MTF in cameras with fisheye lenses. Due to trade-offs related Petzval curvature, planarity of the optical plane is difficult to achieve in fisheye lenses. It is therefore critical to have the ability to accurately measure sharpness throughout the entire image, particularly for lens alignment. One challenge for fisheye lenses is that, because of the radial distortion, the slanted edges will have different angles, depending on the location within the image and on the distortion profile of the lens. Previous work in the literature indicates that MTF measurements are robust for angles between 2 and 10 degrees. Outside of this range, MTF measurements become unreliable. Also, the slanted edge itself will be curved by the lens distortion, causing further measurement problems. This study summarises the difficulties in the use of MTF for sharpness measurement in fisheye lens cameras, and proposes mitigations and alternative methods.

9403-15, Session 3

Designing a simulation tool for smart image sensors

Michel Paindavoine, Univ. de Bourgogne (France); Laurent Soulier, Stéphane Chevobbe, CEA LIST (France); Pierre Bouchain, Univ. de Bourgogne (France)

Thanks to CMOS technology, it is possible now to design smart image

sensors involving CMOS photo-detectors with image processing functions implemented at the focal plane level. These processing functions can be for example contrast increasing (high dynamic range), image filtering using convolutions, morphological operations or transforms for image compression. Several approaches are used to implement such treatments: pixel-level, at a group of pixels or at the foot of column imagers. These functions can also be implemented in analog or digital or even with mixed solutions.

Depending on the application, it is therefore desirable to explore different architectural solutions prior to retain an optimal solution in terms of processing speed, minimum complexity and low power consumption. For this, various powerful simulation tools have already been developed and presented in the literature but in general they are limited to study the photo-detectors without taking into account the associated image processing functions.

So our goal was to provide a tool for simulation of smart image sensors taking into account the complete chain from the image acquisition to the image processing in the focal plane.

Our tool, developed in C ++, allows to take into account the different possible scenarios. For image acquisition, like the existing simulators, photoelectric characteristics of photo-detectors can be considered depending on the chosen integration technology. Therefore it is possible to study a priori the performance of the imagers in terms of sensitivity, dynamic range and noise. The originality of our simulator is based on the fact that snapshot and rolling shutter acquisition modes can be simulated simultaneously with the photoelectric characteristics. It is thus possible to take into account the speed of the objects to be detected in the image in order to study the optimum image acquisition mode in relation with the physical parameters of the application. Similarly, the binning function can be simulated and again depending on the physical parameters (brightness level, for example), the choice of binning size can be established a priori.

In connection with the image acquisition functions, the image processing functions can also be simulated. Therefore, it is very easy to compare different treatment approaches such as those that pool several pixels (macro-pixels) for the same processing unit. In this case, the size and format of the macro-pixel can be explored in order to retain the optimal topology for the intended application. Similarly, in relation to architectural constraints, the choice between analog, digital or mixed solutions can be studied at this level.

In this article, we will present the different principles of our simulator and through some sample applications, such as detection of straight lines in natural images, we will show how to simulate an intelligent imager.

9403-16, Session 3

An ASIC for speckle patterns statistical analysis

Jean-Michel Tualle, Kinia Barjean, Eric Tinet, Univ. Paris 13 (France); Dominique Ettori, Univ. Paris-Nord (France); Antoine Dupret, Commissariat à l'Énergie Atomique (France); Marius Vasiliu, Univ. Paris-Sud 11 (France)

The real-time statistical analysis of speckle patterns is a generic technological bottleneck for many biomedical applications, when diffuse light is under consideration. The statistics of speckle fluctuations can be mainly related to the microscopic movements, what is called Diffuse Correlation Spectroscopy (DCS). The DCS introduces a new kind of contrast factor, linked to tissue perfusion, which reveals to be highly promising for medical diagnosis [1,2]. Furthermore, the speckle fluctuations associated to a wavelength modulation of the light source can be exploited to perform time-resolved measurements of diffuse light propagation in a low-cost way [3]. This last method is currently the only way to perform time-resolved DCS measurements in thick media [4]. To finish with, interaction of diffuse light with an acoustic wave can also induce speckle fluctuations: from the discrimination of such fluctuations, it is possible to select photons "tagged" by the acoustic wave. This is the sketch of acousto-optic imaging [5] that allows mixing of the good spatial resolution of acoustic imaging and of the pertinent contrast of optical imaging.

Of course, speckle analysis suffers from the low spatial coherence of speckle patterns: there is no incentive to take detectors bigger than a coherence area (the speckle "grain" size) for recording speckle fluctuations. Multi-pixels detectors are in fact the good tool for such a task, although they face difficulties: the intensity level of the diffuse light is quite low, leading to a very weak signal at the pixel level that can be far lower than the photon level for one frame. There is therefore a need of a setup with high sensitivity, capable of outputting a signal from noise through averaging on a high number of pixels. Furthermore, such a processing has to be done at a very high acquisition rate to follow speckle fluctuations, which are in the sub-millisecond range. We present a bi-dimensional pixel CMOS detector array specially designed for this task, with parallel in-pixel demodulation and time-resolved correlation computation. Optical signal can be processed at a rate higher than 10,000 samples per second with demodulation frequencies in the MHz range. The performances of this new design will be discussed compared to previous versions of this ASIC [6]. The implications of those achieved improvements will be illustrated through examples in the field of biomedical imaging.

[1] DURDURAN T., CHOE R., YU G., ZHOU C., TCHOU J.C., CZERNIECKI B.J., and YODH A.G., Diffuse optical measurement of blood flow in breast tumors, Opt. Lett. 30 (21), pp. 2915-2917 (2005).

[2] LI J., JAILLON F., DIETSCHE G., MARET G., and GISLER T., Pulsation-resolved deep tissue dynamics measured with diffusing-wave spectroscopy, Opt. Ex. 14 (17), pp. 7841-7851 (2006).

[3] TUALLE J.-M., TINET E. and AVRILLIER S., A new and easy way to perform time-resolved measurements of the light scattered by a turbid medium, Opt. Comm. 189 (4-6), 211-220 (2001).

[4] J.-M. TUALLE, H.L. NGHIEM, M. CHEIKH, D. ETTORI, E. TINET and S. AVRILLIER, Time-Resolved Diffusing Wave Spectroscopy for selected photon paths beyond 300 transport mean free paths, Journal of the Optical Society of America A 23 (6), 1452-1457 (2006).

[5] FARAHI S., BENOIT E., GRABAR A.A., HUIGNARD J.-P., and RAMAZ F., Time resolved three-dimensional acousto-optic imaging of thick scattering media, Optics Letters 37 (13), pp. 2754-2756 (2012).

[6] TUALLE J.-M., DUPRET A., VASILIU M., Ultra-compact sensor for diffuse correlation spectroscopy, Electronics Letters 46 (12), 819-820 (2010).

9403-17, Session 4

A SPAD-based 3D imager with in-pixel TDC for 145ps-accuracy ToF measurement

Ion Vornicu, Ricardo A. Carmona-Galán, Ángel B. Rodríguez-Vázquez, Instituto de Microelectrónica de Sevilla (Spain)

Single-Photon Avalanche Diodes (SPADs) can be employed to detect the arrival of a reflected pulse of light, thus emerging as a feasible alternative for generating a depth map of the scene. SPADs arranged in a bi-dimensional array can effectively associate an estimation of the time-of-flight (ToF) of the pulsed light reflected to each point in the image. Apart from this, ToF measurement can be also applied to positron emission tomography (PET) and to other biomedical techniques using a faint light source like fluorescence lifetime imaging (FLIM).

Amongst the major limitations for ToF estimation based on SPADs are background illumination, and the time resolution of the detection. The former problem could be addressed with spatial correlation methods or using specially designed illumination schemes, especially in the case of outdoor operation. In this paper, we will deal with the latter problem, i. e. the exact time stamping of the detection event. For this to be achieved, we have incorporated an 11b resolution time-to-digital converter (TDC) to each pixel of a 64 ? 64-SPAD array. This array can implement a double functionality: ToF estimation and photon counting. The sensor chip has been designed in a 0.18?m standard CMOS technology, achieving a pixel pitch of 64?m. The complete sensor array fits in 4.1 ? 4.1mm2. The rest of the area up to the 5 ? 5mm2 contains analog I/O buffers, fast signal distribution trees,

row decoder, fast data serializer and a programmable phase locked-loop (PLL). Each pixel contains a SPAD, an active quenching and recharge circuit with adjustable dead time —down to 4ns— to reduce afterpulsing, a ripple 8b counter, a compact pseudo-differential voltage-controlled ring oscillator (VCRO), an encoder and the 11b memory. The counter generates the coarser bits of the time stamp, while the finer 3b are determined by the phases of the VCRO properly encoded. By interpolation of the 8 phases of the VCRO, a minimum time bin of 145ps can be detected, for a power consumption of only 9?W per TDC, amongst the smallest reported and certainly the best time resolution/power consumption trade-off reported in this technology, to the best of our knowledge. The in-pixel TDC occupies 1740μm2 which is smaller than the state-of-the-art. The PLL provided on-chip tunes the reference voltage for the array VCROs, in order to overcome global drift of process parameter and temperature variations. The measured standard deviation of the TDCs across the array is 19 codes, i. e. 1%. This figure is evaluated without applying any pixel-to-pixel calibration. The FWHM jitter of the TDC is 133ps (or 0.92LSB). The last two measurements have been performed at 90% of the full dynamic range (or 270ns).

9403-18, Session 4

Neuro inspired smart image sensor: analog Hmax implementation

Michel Paindavoine, Univ. de Bourgogne (France);
Jerome Dubois, Univ. de Picardie Jules Verne (France);
Purnawarman Musa, Univ. of Gunadarma (Indonesia)

Visual recognition of a familiar object - such as the image of the Eiffel Tower in its natural environment - is obtained easily by a human subject. The execution of the same task on a “classical” computer requires complex and costly algorithms in terms of computing power. Thus, “Neuro-Inspired” approach, based on models from biology, allows to reduce the computational complexity. During the last 20 years many vision models have been proposed such as those from T.Poggio [1] , Y.Lecun [2] and S.Thorpe [3]. The Hmax model proposed by T.Serre et al [4], which builds on the work of Poggio, shows that the recognition of an object in the visual cortex mobilizes V1, V2 and V4 areas. From the computational point of view, V1 corresponds to the area of the directional filters (for example Sobel filters, Gabor filters or wavelet filters). This information is then processed in the area V2 in order to obtain local maxima. This new information is then sent to an artificial neural network. This neural processing module corresponds to area V4 of the visual cortex and is intended to categorize objects present in the scene.

In order to realize autonomous vision systems (consumption of a few milliwatts) with such treatments inside, we studied and realized in 0.35?m CMOS technology prototypes of two image sensors. These prototypes integrate in the focal plane analog functions capable to achieve spatial filters (V1 processing) and local maxima (V2 processing). The first analog retina - consisting of 64x64 pixels - has allowed us to show that it is possible to achieve analog filters such as Sobel filter in a time less than 0.2 ms for the whole image [4]. The second analog retina - also composed of 64x64 pixels - carried out in parallel local maxima on basic neighborhoods of 2x2 pixels and this in less than 1ms [5]. Combining the outputs of local maxima on these basic neighborhoods allows us to make in a few milliseconds maxima on neighborhoods 8x8 pixels. Thus, the coupling of the two retinas allows us to achieve the V1 and V2 layers of Hmax model.

From the description of the models of neuro-inspired vision, we present in this paper our analog image sensors and their coupling in order to achieve the V1 and V2 processing of Hmax model. Finally, we present the results obtained in the context of an application of face detection.

Bibliography :

[1] Hierarchical models of object recognition in cortex. Maximilian Riesenhuber and Tomaso Poggio, Nature, 1999

[2] Convolutional Networks and Applications in Vision Yann LeCun, Koray Kavukcuoglu and Clément Farabet, IEEE ISCAS 2010

[3] Suggestions for a Biologically Inspired Spiking Retina using Order-based Coding. Simon J. Thorpe, Adrien Brilhault, José-Antonio Perez-Carrasco, IEEE ISCAS 2010

[4] Robust Object Recognition with Cortex-like Mechanisms. Thomas Serre, Lior Wolf, Stanley Bileschi, Maximilian Riesenhuber, and Tomaso Poggio, IEEE PAMI vol. 29,n°. 3, march 2007

[5] A 10 000 fps CMOS sensor with massively parallel image processing. Dubois, J., Ginhac, D., Paindavoine, M., & Heyrman, B. (2008). IEEE Journal of Solid-State Circuits, 43(3), 706-717.

[6] Design and implementation of non-linear image processing functions for CMOS image sensor. Purnawarman Musa, Paindavoine, M. SPIE Asia Photonics Conference, Beijing, China, 5-7 November 2012.

9403-19, Session 4

A 12-bit 500KSPS cyclic ADC for CMOS image sensor

Zhaohan Li, GengYun Wang, Leli Peng, Cheng Ma, Yuchun Chang, Jilin Univ. (China)

In this paper we present a 12-bit, 500KSPS Cyclic ADC which is very suitable for CMOS image sensor array column-level readout circuit applications. The cyclic ADC can achieve higher resolution by using a lower frequency system clock than traditional ramp ADCs. The ADC we presented is made by 0.18?m 1P6M CMOS process and only occupies a chip area of 8?m?374?m. Post simulation results show that Signal-to-Noise-and-Distortion-Ratio (SNDR) and Efficient Number of Bit (ENOB) reach 63.7dB and 10.3bit, respectively.

As shown in figure 1, the converter includes switch capacitor network, a differential amplifier, a sub-ADC and a digital correction circuit. Firstly, the pixel signal is sampled by different sampling signals and passes through the sub-ADC to get a 1.5-bit digital code which will be transmitted to digital correction. Secondly, the 1.5-bit digital code is converted to an analog signal by a DAC and then subtracted by the input signal to obtain a residue signal which is gained by 2 times and sampled for the next period. Thirdly, the previous 2 steps are repeated till we get the resolution needed.

Figure 2 shows a timing diagram. The converting flows of the cyclic ADC is described as following.1) Signal sampling stage. The exposure signal from pixel is sampled to C1. 2) Reset sampling stage. The reset signal is sampled to C2. 3) Capacitors reverse state. Signals collected in above two steps are sent to the op-amp output terminal and hold. Meanwhile the signals are converted to the first group of digital codes by the sub-ADC. 4) Amplification stage. Op-amp given different output voltages in three different situations. 5) Resample state. The system resampled output signals in stage 4 to C1 and C2. The second group digital codes are obtained by the sub-ADC at the same time. Then the circle of stage 4 and 5 is repeated till the final digital codes are obtained. This convert flow is shown in figure 3.

Figure 4 shows a micro photograph of proposed cyclic ADC with a timing generator. It is designed for test separately. The chip is made by 0.18?m 1P6M CMOS process. The size of ADC core designed for one or two columns pixels application is 8?m wide and 374?m length. Post simulation results show that SNDR and ENOB reaches 63.7dB and 10.3bit, respectively.

9403-20, Session 4

14bit pipeline-SAR ADC for image sensor readout circuits

GengYun Wang, Can Peng, Tianzhao Liu, Cheng Ma, Ning Ding, Yuchun Chang, Jilin Univ. (China)

A two stage 14bit pipeline-SAR analog-to-digital converter includes a 5.5bit zero-crossing MDAC and a 9bit self-timing SAR ADC for image sensor readout circuits [1][2] built in 0.18um CMOS process is described with low power dissipation as well as small chip area. In this design, we employ comparators instead of high gain and high bandwidth amplifier, which consumes as low as 20mW of power to achieve sampling rate of 40MSps and 14bit resolution.

As number of pixels in image sensor becoming more and more, demand for high speed and high resolution ADC is increasing. This paper presents a 14bit 40MSps pipeline-SAR ADC for image sensor readout circuits. Conventional pipeline ADC is an analog-to-digital converter structure consists of some MDACs contain residue voltage amplifier, sub-ADC and sub-DAC. The switch capacitor MDAC needs a high gain and high bandwidth amplifier to improve settling accuracy. As the feature size of integrated circuits process developing shorter and shorter, intrinsic gain and output resistance of a transistor with a short channel becomes smaller and smaller. Therefore, it is more and more difficult to design the high performance amplifier with the continue technology scaling. Recently, some papers present several methods to replace the high gain and high bandwidth amplifier with several comparators and current sources [3][4].

High bit MDAC is not usually applied in amplifier based switch capacitor circuits, because high feedback factor induce high bandwidth amplifier essentially. But this requirement disappears in zero-crossing comparator switch capacitor circuits. In zero-crossing MDAC topology, integrator settles the residue voltage with charge injecting into feedback capacitors. MDAC output voltage export to next stage and input value of the comparator approaches the system common voltage at the same time. Because comparator has offset voltage, current source can not be shut down at the accurate time. To solve this problem, 5.5bit zero-crossing MDAC is divided into a coarse step and a fine step. During the coarse step, charge is injected into feedback capacitors from a huge current source, and the input value approximately approaches to the common value which is regarded as zero value very quickly. At the fine step, the input value approaches the common value very slowly driven by a small current source.

The second stage is a 9bit successive approximation register ADC[5] which is based on the self-timing algorithm. A conventional N-bit SAR ADC requires N cycles to complete one conversion. However, it is complicated to generate a such high frequency clock from an on-chip PLL block. Therefore, asynchronous self-timing topology is adopted to generate high speed clock by SAR ADC internal logic.

According asynchronous algorithm, after low speed non-overlapping time clock CLKL turns down, the comparator yields output code and the output digital logic generate signal Logic which is feed back into the clock system to settle next cycle state and set a fast clock CLKH at the same time. After nine clock periods of CLKH, all the output codes are generated successfully.

9403-21, Session 5

Power noise rejection and device noise analysis at the reference level of ramp ADC

Peter Ahn, JiYong Um, EunJung Choi, HyunMook Park, JaSeung Gou, SK Hynix, Inc. (Korea, Republic of); KwangJun Cho, KangBong Seo, SangDong Yoo, SK hynix (Korea, Republic of)

Today's state-of-the-art CMOS Imagers require high resolution display at high frame rate with low noise performance. In order to achieve high resolution format at high frame rate, a ramp ADC, which is capable of converting large array of analog signals into digital bits at once, has been widely accepted as the ideal candidate for generating digital representation of analog signals from each pixel. However, in order to realize low noise performance, it is crucial to generate as little noise as possible at the reference level from which the ramp signal of ADC is generated. Therefore, the paper pinpoints the sources of noise seen at the ramp ADC's reference level and concludes with the means to suppress the identified noise sources to realize low noise state-of-the-art CMOS imagers.

Noise sources that are investigated in the paper are divided into two groups - power noise and device noise. In order to suppress the reference level corruption from identified noise sources, noise appearing at the reference level is quantified in terms of design variables (bias current, device dimension, load capacitance et cetera). By quantifying the amount of generated noise in terms of design variables, the paper works as a solid design reference which would guide a design engineer through a low noise CMOS imager design by showing which variables to be modified for noise reduction.

Regarding the power noise, Power Supply Rejection (PSR) of Band-Gap Reference Generator (BGR) and Current Bias Generator (IGEN) are investigated. A ramp generator, which also takes considerable amount in reference level generation, is ignored in PSR analysis for its superb PSR characteristic induced from the current mirroring architecture. PSR of BGR and IGEN are quantified by solving the equivalent small signal circuit of each block. Small signal circuit analysis on PSR of BGR and IGEN suggests that while both building blocks require an error amplifier with high voltage gain for strong power rejection, each amplifier in corresponding building block must have a contrasting PSR characteristic for outstanding power noise rejection. To be more specific, power noise from BGR is well suppressed when BGR's error amplifier has poor Power Supply Rejection while power noise from IGEN is well suppressed when its error amplifier has superb Power Supply Rejection.

Regarding the device noise, noise generated from each device (BJTs, MOSFETs and Resistors) in the circuit is expressed in terms of its equivalent noise current or noise voltage. Each of the noise current and noise voltage is given in terms of design variables and is multiplied by proper signal gain to estimate the signal corruption at the reference level. Moreover, the total output noise generated from each device at the reference level is obtained by integrating each of the gain-multiplied noise current/voltage through its signal bandwidth.

The noise quantification in terms of design variables are validated by comparing the noise level obtained from the analysis and that obtained from HSPICE simulation.

9403-22, Session 5

The effect of photodiode shape on dark current for MOS imagers

Steven Taylor, DTS, Inc. (United States); Bruce Dunne, Heidi Jiao, Grand Valley State Univ. (United States)

The effect of photodiode (PD) shape was studied in the attempt to reduce the dark current in MOS imagers. In such imaging systems, each pixel ideally produces a voltage directly proportional to the intensity of light incident on the PD. However, due to various non-idealities, the PD performance is compromised by the presence of dark current. Dark current, also known as leakage current, is the undesirable current that flows through the PD under no illumination. The result is that dark current is the most significant source of noise degrading overall image quality, particularly for low light environments. Unfortunately, due to the statistical variability of dark current, it is not possible to simply correct the readout voltage error via subtraction. To reduce this effect, recent research suggests that PD shape and features have an influence on dark current levels [1][2]. We consider PDs with different corners while maintaining high fill-factor rates. Furthermore, we test both rectangular and triangular shapes to exploit charge transfer characteristics. In all, five PD geometries were built to test the influence of PD shape on the dark current signal. These geometries include a traditional square shape (sized at 300μm x 250μm and serving as the control), two square shapes with increasingly rounded corners (135 and 150 degrees), a triangular design with sharp corners and finally, a triangular design with 120 degree corners. The fabrication process is a nonstandard single diffusion, metal gate, two-metal process with a minimum feature size of 10μm, utilizing spin-coating at several stages. Results indicate the PDs with square shape and 90o corners exhibit the lowest dark current and highest readout voltage. Furthermore, the triangular shape suggests improved charge transfer characteristics; however, this improvement appears to be negated by an increase in dark current response. Therefore, our findings indicate that the traditional PD square shape is the preferred design.

[1] I. Shcherback, A. Belenky and O. Yadid-Pecht, "Empirical Dark Current Modeling for Complementary Metal Oxide Semiconductor Active Pixel Sensor," Optical Engineering, vol. 41, no. 6, pp. 1216-1219, December 2001.

[2] B. Shin, S. Park and H. Shin, "The Effect of Photodiode Shape on Charge Transfer in CMOS Image Sensors," Solid State Electronics, vol. 54, pp. 1416-1420, 2010.

Conference 9403:
Image Sensors and Imaging Systems 2015

9403-23, Session 5

High-speed binary CMOS image sensor using a high-responsivity MOSFET-type photodetector

Byoung-Soo Choi, Sung-Hyun Jo, Myunghan Bae, Pyung Choi, Jang-Kyoo Shin, Kyungpook National Univ. (Korea, Republic of)

In this paper, we propose a novel high-speed binary complementary metal-oxide semiconductor (CMOS) image sensor using a high-responsivity gate/body-tied (GBT) MOSFET-type photodetector. The proposed binary CMOS image sensor was designed using standard CMOS 2-poly 4-metal 0.18-??m technology and does not require any special additional process. The high-responsivity GBT MOSFET-type photodetector is based on the p-type MOSFET with the gate connected to the body of the MOSFET. The incident light through the gate affects the body potential of the GBT photodetector. Then the output signal of the GBT photodetector is amplified because of the transistor characteristics. The photodetectors used in the conventional image sensors are pinned photodiodes, silicon-on-insulator (SOI) photodetectors, hole accumulation diodes (HADs) and bipolar junction transistor (BJT) photodetectors. The GBT photodetector with a high responsivity is suitable for a pixel of the binary CMOS image sensor. In addition, it is possible to increase the resolution of image sensor because the GBT active pixel sensor (APS) occupies a small area.

Generally, the CMOS image sensor has multi-bit contrast information, but the binary CMOS image sensor has only 1-bit digital information. It means that it has only black and white pixels of the image. The proposed binary image sensor does not require analog to digital converter (ADC), thus the proposed image sensor can be designed for low power consumption and high speed operation. In CMOS image sensors (CISs), power efficiency is essential because the battery power of the portable devices is limited. High-speed operation of the image sensor also becomes important as the resolution increases. The binary image sensor has been used for texture recognition, edge detection, bar-coding, target tracking and motion recognition.

The operating principle of the proposed image sensor is as follows. The output signal of the GBT APS is applied to the input of the comparator. Under high illumination, the GBT APS output voltage is larger than the reference voltage, and the output signal of the comparator is high. However, the output signal of the comparator is low under low illumination. This reference voltage of the comparator can be adjusted from outside the chip and define the number of pixels of black and white. This output signal of the comparator represents 1-bit digital information of the binary image sensor. The proposed image sensor is composed of the GBT pixel array (176 ? 144), scanners, comparators and memories. The pixel size of the GBT APS is 5.6 ??m ? 5.6 ??m, which was determined in consideration of the column parallel binary processing circuit and the optics. Size of the entire image sensor chip is 1.3mm ?1.5 mm. We expect the image sensor to operate at a speed of higher than 200 frames per second. Operation of the proposed binary CMOS image sensor can be changed from the binary mode to the analog mode and vice versa by simple switching in a single chip.

9403-24, Session 5

Design considerations for low noise CMOS image sensors

Ángel B. Rodríguez-Vázquez, Univ. de Sevilla (Spain); Fernando Medeiro, Rafael Dominguez-Castro, Anafocus (Spain)

No Abstract Available

9403-25, Session PTues

An improved Sobel edge detection algorithm based on the idea of median filter

Shuang Cui, GengYun Wang, Teng Chen, Yuchun Chang, Zhen Huang, Jilin Univ. (China)

With the development of society, the digital image processing technology has been more and more widely used in people's daily life. It is significant to study image segmentation which is an important part of digital image processing. Edge detection, as the basis of profile-based image segmentation technology[1], is the main part studied by researchers. As a simple image pre-processing technology, while median filter eliminates the noise, it also causes the image fuzzy, weakening the edge information of image.

In this paper, by using the weighted thoughts and changing square template into circular template to reduce the influence of pixels around images on the template center's pixels and eliminate noises at the same time, it will well retain the edge information of the image and improve the image clarity after filtering. Besides, it will improve the two defects that the traditional Sobel operator[2] can not accurately detect real edge points with weak response in the dark areas and the pseudo edge points with strong response are detect out in bright areas and combine the normalized ideas with Sobel operator to put forward an improved algorithm.

The new algorithm not only improves the noise immunity of the traditional Sobel operator, but also improves the accuracy of edge detection. Through the experiment, the results show that the improved algorithm can well detect the real edge in dark area and prevent the pseudo edge points in the bright area so as to improve the accuracy of edge detection. Compared with the traditional Sobel operator, the improved algorithm can make the extracted edge delicate, continuous, pseudo edge points fewer and have high accuracy.

Figure 1 is the iris vascular's traditional median filter and weighted average filter comparison figure after adding Gaussian noise. As shown in both Figure 2 and Figure 3, the effect of noise on the image can be well eliminated. But Figure 3 is clearer than Figure 2. Especially the capillary portion in Figure 3 is very clear and its details are completely kept.

Figure 4 is the red blood cells' edge detection results comparison figure conducted by the traditional Sobel operator and the improved Sobel operator respectively. As shown in Figure 5, there are more pseudo edge points in the inner contour of erythrocytes; while Figure 6 has a good figure of the red blood cells and has fewer pseudo edges.

Figure 8 and Figure 9 are the traditional Sobel operator and the improved Sobel operator' edge detection results comparison figure after adding Gaussian noise. It can be seen that due to poor noise immunity of the traditional Sobel operator, its image has a lot of pseudo edge points; while the improved Sobel operator well suppresses the influence of Gaussian noise on edge detection and has high accuracy.

9403-26, Session PTues

Short wave infrared hyperspectral imaging for recovered postconsumer single and mixed polymers characterization

Giuseppe Bonifazi, Roberta Palmieri, Silvia Serranti, Univ. degli Studi di Roma La Sapienza (Italy)

Post-consumer plastics resulting from packing and packaging represent about the 60% of the total plastic wastes (i.e. 23 million of tons) produced in Europe. The EU Directive (2014/12/EC) fixes as target that the 60%, by weight, of packaging waste has to be recovered, or thermally valorized. When recovered, the same directive established that packaging waste has to be recycled in a percentage ranging between 55% (minimum) and 60% (maximum). The non-respect of these rules can produce that large

quantities of end-of-life plastic products, specifically those utilized for packaging, are disposed-off, with a strong environmental impact. The application of recycling strategies, finalized to polymer recovery, can represent an opportunity to reduce: i) not renewable raw materials (i.e. oil) utilization, ii) carbon dioxide emissions and iii) amount of plastic waste disposed-off. Aim of this work is to perform a full characterization of different end-of-life polymers based products, constituted not only by single polymers but also of mixtures, in order to realize their identification for quality control and/or certification assessment of the different recovered products as resulting from a recycling plant where classical processing flow-sheets, based on milling, classification and separation, are applied. To reach this goal, an innovative sensing technique, based on the utilization of a HyperSpectral Imaging (HSI) device working in the SWIR region (1000-2500 nm), is proposed. HSI is an innovative, fast and non-destructive technique able to collect both spectral and spatial information from an object. The recorded information is contained in a "hypercube", a 3D dataset characterized by spatial data (i.e. x and y axis, representing the pixel coordinates) and spectral data (i.e. z axis, representing the wavelengths). HSI represents an attractive solution for quality control in several industrial applications. Following this strategy single polymers and/or mixed polymers recovered can be determined. The main advantage of the proposed approach is linked to the possibility to perform "on-line" analyses, that is directly on the different material flow streams, as resulting from processing, without any physical sampling and classical laboratory "off-line" determination.

9403-27, Session PTues

Designing and construction of a prototype of (GEM) detector for 2D medical imaging application

Abdulrahman S. Alghamdi, Mohammed S. AlAnazi, Abdullah F. Aldosary, King Abdulaziz City for Science and Technology (Saudi Arabia)

Due to the limited resolution and accuracy of several technologies that are able to get a digital X-ray image with a good performance in the very high rates, micro-pattern technology can achieve these features by using the most effective example of which is gas electron multiplier (GEM). The main objective of this project is to develop a two dimensions imaging that can be used in medical imaging purposes. The project consists of the theoretical parts of the process, including simulating the best detector dimensions, geometry, and the best energy range of the applied radiation. Furthermore, constructing a large active area of triple GEM detector, and preparing the necessary setup parts for medical imaging system assumed. This paper presents the designing and construction of a prototype of triple-GEM detector (10cm x10cm) that can achieve the goals of the project as a first step toward attaining the goals of this project. In addition, the preliminary results from X-ray and some gamma sources as a testing of the prototype detector includes the discussions of outlined tasks and achievements will be presented. The future plan of the whole project and more details about the next stages will be presented in this paper as well.

9403-28, Session PTues

Enhanced correction methods for high density hot pixel defects in digital imagers

Rahul Thomas, Glenn H. Chapman, Rohit Thomas, Simon Fraser Univ. (Canada); Israel Koren, Zahava Koren, Univ. of Massachusetts Amherst (United States)

Previous research has shown that "Hot Pixels" is the most common type of defects in modern digital imagers, with their number in a given imager increasing over time. They are believed to be caused by cosmic rays, and shielding cannot fully prevent their manifestation and increasing numbers. In our recent studies we have developed an empirical formula to project the growth of hot pixel defects in terms of defects/year/mm2, and discovered that hot pixel densities will grow via a power law of inverse of the pixel size raised to a power of about 3. This formula indicates that the defect rate will increase drastically when the pixel size falls below 2 microns, and can potentially reach a density of 12.5 defects/year/mm2 at ISO 25,600. This could cause a significant deterioration in the image quality, making defect correction of images vital.

This paper presents a correction algorithm that uses a weighted combination of two terms: one is traditional interpolation of neighboring values, and the other is a correction of the hot reading based on the estimated parameters of the specific hot pixel. The hot pixel is modeled with an offset value plus a coefficient of growth with exposure time, both of which increase with the sensitivity but remain nearly constant after formation. The weights can vary from one pixel to the other, and depend on defect severity, ISO, exposure time and complexity of the pixel neighborhood.

During our analysis of hot pixels, we discovered that contrary to common belief, illumination does affect the hot pixel parameters. At low illumination, the classic dark hot pixel model applies, while at higher illuminations both the offset and the dark current are amplified creating a different behavior during exposure. Above some threshold illumination, the offset and dark current enhancement nearly saturates. We used this new model as the second component in the correction algorithm discussed in the paper.

To test the accuracy of our correction algorithm, we found an innovative way of obtaining the real value (defect-free) of a defective location. Our technique involves a simple translation of the image sensor in order to recover the true value that was originally covered by the defective pixel. To quantify the error of this translational experimental method we collected data for 50 good pixels before and after translation, resulting in an average error of ±6.1%, which is within the expected noise error.

We tested our algorithm on two sets of images, using cameras with known hot pixel defects. The first set was nearly uniform backgrounds ranging from dark fields to light gray fields. For these uniform pictures, we found that interpolation correction dominates. We then tested the algorithm on images that were more complex in nature, and found that our weighted algorithm is more effective than interpolation alone because it makes use of the specific hot pixel parameters when interpolation alone falls short.

Conference 9404: Digital Photography and Mobile Imaging XI

Monday - Tuesday 9–10 February 2015

Part of Proceedings of SPIE Vol. 9404 Digital Photography XI

9404-1, Session 1

Multimode plenoptic imaging

Andrew Lumsdaine, Indiana Univ. (United States); Todor G. Georgiev, Qualcomm Inc. (United States)

No Abstract Available

9404-2, Session 1

Automatically designing an image processing pipeline for a five-band camera prototype using the local, linear, learned (L3) method

Qiyuan Tian, Henryk Blasinski, Stanford Univ. (United States); Steven P. Lansel, Olympus America Inc. (United States); Haomiao Jiang, Stanford Univ. (United States); Munenori Fukunishi, Olympus America Inc. (United States); Joyce E. Farrell, Brian A. Wandell, Stanford Univ. (United States)

Implementation of new camera designs is slowed by the need to develop novel image processing algorithms that are tuned for the new design. To speed camera development, we developed an algorithm (Local, Linear, Learned or L3) that automatically creates an image processing pipeline for a given design. The L3 method is a data driven algorithm combining machine learning and camera simulation. We used the L3 algorithm to implement a pipeline for a prototype camera with five color channels. We describe how we modeled the prototype, tested the accuracy of the calibration, used the L3 algorithm to create an image processing pipeline and accelerated the pipeline using graphics processing units (GPUs).

9404-3, Session 1

Efficient illuminant correction in the local, linear, learned (L3) method

Francois G. Germain, Iretiayo A. Akinola, Qiyuan Tian, Stanford Univ. (United States); Steven P. Lansel, Olympus America Inc. (United States); Brian A. Wandell, Stanford Univ. (United States)

1. Introduction.

The L3 algorithm automatically generates a standard image processing pipeline (sensor correction, demosaicking, illuminant correction, noise reduction) for novel camera architectures [1-4]. The algorithm selects optimal parameters for the pipeline for a specific sensor and camera architecture. The standard image processing pipeline comprises two parts. First, the pipeline classifies each pixel (based on a statistical analysis of the responses at the pixel and its neighborhood) into one of many possible classes [1]. Second, the processing pipeline linearly transforms the data of the pixel and its neighborhood to the target output space.

The classes are defined by the user and the linear filters for each class are selected through a training process. The selection of the filters rely on an accurate camera software simulation; we used the Image Systems Engineering Toolbox (ISET) simulation software [5, 6]. The simulation creates many samples of how natural scenes are converted into sensor data. The output data, i.e. CIE XYZ values for consumer photography are computed from the known scene radiance, and the L3 algorithm learns linear filters that convert the sensor data in each class to the CIE XYZ values.

2. Motivation: Illuminant correction.

Illuminant correction is an essential part of the image processing pipeline that is necessary to account for the perceptual phenomenon of color constancy [7]. The visual system adapts to changes in the ambient light level and color, and this adaptation has the general effect of preserving the color appearance of an object (e.g. a white shirt) across conditions (sunlight to tungsten light). Because the human visual system adapts to the illuminant, to preserve color appearance the linear filters in L3 method used for rendering must adapt, too.

In the original algorithm, we anticipated selecting classes and learning a table of linear filters for a large number of illumination conditions (Figure 1a). Each of these illuminant-dependent tables would be stored and applied for rendering under the appropriate illuminant. This approach pays a significant cost in computation and storage [1]. Hence, we explore an alternative approach that separates illuminant correction from the rest of the pipeline. In this approach, we train an illuminant-independent table for only the within illuminant conditions: data acquired under one illuminant are transformed to a target output space, with demosaicking, denoising and sensor correction, but without illuminant correction (Figure 1b). We then apply an illuminant-dependent color correction linear transform to the output in the target space. This architecture requires training and storing only one table of L3 filters and one color correction matrix for each illuminant of interest.

3. Methods.

We used ISET [5,6] to simulate digital camera processing. The ISET camera model includes the simulation of scene radiance, optics, sensor electronics and image processing pipeline (here the L3 pipeline). In this abstract, we model the camera as an f/4 lens with diffraction limited optics and focal length of 3mm. The sensor uses a RGB/W color filter array (CFA) layout described in [1], with parameters values as in Figure 2. The table of L3 filters is learned from simulated data of natural scenes following the process described in [1].

Within-illuminant (WI) tables of L3 filters are learned by simulating the sensor data and the target display XYZ data under the same illuminant. Here, we learn a table for a D65 (WI_D65) illuminant and a different table for tungsten illuminant (WI_Tun) . A cross-illuminant (XI) table is learned by simulating the sensor data under a tungsten illuminant and the target display XYZ data under a D65 illuminant. Finally, a 3x3 linear transform T from tungsten to D65 is learned by linear regression on display XYZ data rendered ideally under tungsten and D65 illuminants.

Color reproduction error was evaluated using a custom Natural-100 test chart (Figure 2) comprising of natural surface colors and a gray strip. The error was measured in CIE2000 ?? units [8] by comparing the ideal rendered XYZ display data under a D65 illuminant with the rendered data using the different L3 tables.

4. Results.

First, we compared the WI_D65 and WI_Tun tables in order to assess the suitability of WI_D65 as illuminant-independent within-illuminant table (as presented in Figure 1b). The optimized filter weights differ slightly between the illuminant conditions. For example, under tungsten illuminant conditions the blue pixel data are generally noisier than the corresponding classes with D65 illumination. Consequently, the optimal filters weights sum more broadly over the blue pixels in the tungsten filters than in the D65 filters.

Even though the derived tables differ, for color reproduction error the rendering differences are rather small. This can be seen by comparing Figure 3b and 3c which show the same camera data rendered with the WI_D65 and WI_Tun tables respectively derived from D65 and tungsten illumination. We use WI_D65 as reference illuminant-independent within-illuminant table in the rest of the paper.

Second, we compared renderings using the illuminant-independent

table (here WI_D65) followed by a linear transformation T for illuminant correction in XYZ space (WI_D65+T) with renderings using the cross-illuminant table (XI) between a tungsten illuminant and D65 target output (Figure 4).

We quantified the color rendering differences between the XI, WI_Tun+T and WI_D65+T pipelines using the Natural-100 test chart by comparing the XYZ values with those of the ideal XYZ values (these are known because the data are simulated). The color reproduction errors in the XI condition are identical to those in the WI_Tun+T condition, while they are slightly smaller than those in the WI_D65+T condition, indicating that color reproduction accuracy is mostly affected by the replacement of WI_Tun by the illuminant-independent table (here WI_D65). However, the differences do not appear to be very significant for most consumer applications (Table 1).

5. Conclusion.

Using a single illuminant-independent L3 table followed by a linear color correction transformation is efficient without significantly reducing color reproduction accuracy. Additional analyses need to be performed to understand the consequences of this approach for spatial resolution and noise.

In the full paper, we will present a wide variety of rendered images comparing the accuracy of color reproduction for different CFAs and illuminants. We will provide an extended analysis of the method trade-offs, with detailed specifications regarding the measure and its limits, as well as the calibration and simulation used in our experiments.

6. References.

[1] Tian, Q., Lansel, S., Farell, J., and Wandell B. "Automating the design of image processing pipelines for novel color filter arrays: local, linear, learned (L3) method," in IS&T/SPIE Electronic Imaging, International Society for Optics and Photonics (2014).

[2] Lansel, S. and Wandell, B., "Local linear learned image processing pipeline," in Imaging Systems and Applications, Optical Society of America (2011).

[3] Lansel, S. P., Local Linear Learned Method for Image and Reflectance Estimation, PhD thesis, Stanford University (2011).

[4] Lansel, S., Wandell, B., et al., "Learning of image processing pipeline for digital imaging devices," (Dec. 7 2012). WO Patent 2,012,166,840.

[5] Farrell, J. E., Xiao, F., Catrysse, P. B., and Wandell, B. A., "A simulation tool for evaluating digital camera image quality," in Electronic Imaging 2004, 124–131, International Society for Optics and Photonics (2003).

[6] Farrell, J. E., Catrysse, P. B., and Wandell, B. A., "Digital camera simulation," in Applied Optics 51(4), A80–A90 (2012).

[7] Wandell B., Foundations of Vision, Sinauer Associates (1995).

[8] Sharma, G., Wu, W., and Dalal E., "The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations," in Color Research & Application 30(1), 21–30 (2005).

9404-4, Session 2

Reflection removal in smart devices using a prior assisted independent components analysis

Pramati Kalwad, National Institute of Technology, Karnataka (India); Divya Prakash, Indian Institute of Technology, Kanpur (India); Phanish H. Srinivasa Rao, Samsung R&D Institute India - Bangalore (India)

When photographs are taken through a glass or any other semi-reflecting transparent surface, in museums, shops, aquariums etc., we encounter undesired reflection. Reflection Removal is an ill-posed problem and is caused by superposition of two layers namely the scene in front of camera and the scene behind the camera getting reflected because of the semi-reflective surface. Modern day hand held Smart Devices (smartphones, tablets, Fablets, etc) are typically used for capturing scenes as they are equipped with good camera sensors and processing capabilities and we can expect image quality to be similar to a professional camera. In this direction, we propose a novel method to reduce reflection in images, which is an extension of Independent Component Analysis (ICA) approach, by making use of two cameras present - a back camera (capturing actual scene) and a front facing camera. When compared to the original ICA implementation, our method gives on an average of 10% improvement on the peak signal to noise ratio of the image.

9404-5, Session 2

Measurement and analysis of the point spread function with regard to straylight correction

Julian Achatzi, Gregor Fischer, Fachhochschule Köln (Germany); Volker Zimmer, Leica Camera AG (Germany); Dietrich W. Paulus, Univ. Koblenz-Landau (Germany); Gerhard Bonnet, Spheron-VR AG (Germany)

Introduction

Stray light is the part of an image that is formed by misdirected light. I.e. an ideal optic would map a point of the scene onto a point of the image. With real optics however, some parts of the light gets misdirected. This is due to effects like scattering at edges, Fresnel reflections at optical surfaces, scattering at parts of the housing, scattering from dust and imperfections — on and inside of the lenses — and further reasons. These effects lead to errors in color-measurements using spectral radiometers and other systems like scanners. Stray light is further limiting the dynamic range that is achievable with High-Dynamic-Range-Technologies (HDR) and can lead to the rejection of cameras due to quality considerations.

Therefore it is of interest, to measure, quantify and correct these effects. Our work aims at measuring the stray light point spread function (stray light PSF) of a system which is composed of a lens and an imaging sensor. We seek to use this PSF in order to qualify and quantify the system's stray light, and to find new ways of correcting for stray light.

State of the Art

In order to quantify a system's amount of stray light, DIN 58186 proposes to measure the brightness of the image of a black spot, which is surrounded by a bright white area. The system's stray light measure is then defined as the ratio of light in the black spot image over light in the bright surround.

DIN 58186 further proposes to measure the point spread function in order to visualize the distribution of stray light over the image field. A model for the stray light PSF has been proposed in, which is calibrated by using an extent, uniform light source and two exposure times, the latter because of the high dynamic range of the stray light PSF. This model is then used in order to correct images for stray light. Other research concentrates on correcting for stray light in spectral radiometers and scanners. These methods commonly use a calibrated stray light correction matrix. This is a feasible approach for systems with only a few pixels, however this is not the case for high resolution systems like, e.g. digital cameras.

Despite all of this work, the question how to distinguish stray light from other effects that misdirect light -- e.g. aberration, diffraction, defocus etc. -- stays unanswered. In this work we propose a new method for measuring the stray light PSF and we seek to answer some questions concerning stray light quantification, qualification and correction.

Stray light measurement

The PSF is the image of a point light source (PLS). I.e. it is the 2D correspondent of the impulse response function. The PSF of an optical system however is not shift-invariant, i.e. it depends on the position of the point light source in the object-plane. Optical systems are commonly characterized by means of the modulation-transfer-function (MTF), which is the absolute value of the Fourier transformed PSF. The MTF is generally measured by using a slanted edge or a Siemens star. While being commonly used and reasonable for most applications, these methods are not practicable for measuring the stray light PSF. That is because firstly the targets of these methods do have a spatial extent, which is problematically because of the PSF's shift-variant nature. Secondly these methods consider

the PSF to be "relatively compact and localized". Stray light however influences the broad wings of the PSF, which are furthermore at a much lower signal-level than the PSF's peak, such that the stray light PSF does have a very high dynamic range, which thirdly, is not considered by the regular methods.

The results, of classical approaches to quantify straylight like DIN 58186, depend on the size of the black-spot (or light-trap). However the black-spots' size isn't varied, the influence therefore not measured. These classical measurements could be simulated with varying spot-size, using a model of the straylight-PSF, resulting in better qualifications of the systems straylight-behaviour.

Hence we suppose a new method which measures the stray light PSF by the means of an approximate point light source and HDR-Techniques. The point light source consists of a high-power LED and a small aperture. This light source, and the camera, are inside of a dark room which has been specially designed for stray light measurements. I.e. this room does have black walls, black carpet and a black ceiling, in order to reduce light scatterings from the room itself.\

We use HDR-Techniques in order to capture the very high dynamic range of the stray light PSF. That is we capture a series of images with varying exposure time, in order to capture the peak of the PSF and its wide wings. While ordinary HDR-Techniques only use such exposure-series, to increase the dynamic range of the camera system, we further adapt to our unique scene by varying the brightness of the PLS. When doing so, one has to take further considerations about the LED's emission-spectrum, as it changes with junction-temperature. We use a modified version of the maximum-likelihood-estimator approach from in order to fuse these images into one HDR-image. Hereby we are able to further improve the signal quality at the wings of the stray light PSF.

In this work we will use our novel measurement-technique in order to calculate classical characteristics, like the MTF and straylight-measurements like DIN 58186. The results will be compared with the classical methods. We will further develop some new measures for the qualification of an optical system's straylight behaviour. We further seek to find new ways of correcting for straylight.

9404-6, Session 2

Advances in image restoration: from theory to practice *(Keynote Presentation)*

Filip Sroubek, Institute of Information Theory and Automation (Czech Republic)

We rely on images with ever growing emphasis. Our perception of the world is however limited by imperfect measuring conditions and devices used to acquire images. By image restoration, we understand mathematical procedures removing degradation from images. Two prominent topics of image restoration that has evolved considerably in the last 10 years are blind deconvolution and superresolution. Deconvolution by itself is an ill-posed inverse problem and one of the fundamental topics of image processing. The blind case, when the blur kernel is also unknown, is even more challenging and requires special optimization approaches to converge to the correct solution. Superresolution extends blind deconvolution by recovering lost spatial resolution of images. In this talk we will cover the recent advances in both topics that pave the way from theory to practice. Various real acquisition scenarios will be discussed together with proposed solutions for both blind deconvolution and superresolution and efficient numerical optimization methods, which allow fast implementation. Examples with real data will illustrate performance of the proposed solutions.

9404-7, Session 3

From Maxwells equations to efficient filter flow with applications in blind image deconvolution *(Invited Paper)*

Michael Hirsch, Max-Planck-Institut für biologische Kybernetik (Germany)

No Abstract Available

9404-8, Session 3

Parameterized modeling and estimation of spatially varying optical blur *(Invited Paper)*

Jonathan D. Simpkins, Robert L. Stevenson, Univ. of Notre Dame (United States)

No Abstract Available.

9404-9, Session 3

Making single image deblurring practical *(Invited Paper)*

Jue Wang, Adobe Systems (United States)

No Abstract Available

9404-10, Session 4

Blind deconvolution of images with model discrepancies

Jan Kotera, Filip Sroubek, Institute of Information Theory and Automation (Czech Republic)

Single channel blind deconvolution is the inverse problem of estimating latent image u from a single observed blurred image g satisfying the convolutional degradation model

g = u*h+n,

where h, called point spread function (PSF), is unknown and n is random additive noise. Since we have only one observation and no knowledge of the PSF, the problem is extremely ill-posed and hence requires careful choice of regularizers and optimization procedure. Simultaneous recovery of u and h can be expressed as a standard MAP (Maximum A Posteriori) estimation

P(u,h|g) ~ P(g|u,h)P(u,h) = P(g|u,h)P(u)P(h)

where P(g|u,h) is the noise distribution and P(u), P(h) are the prior distributions on the latent image and blur kernel, respectively. One of the key ideas of new algorithms is to address the ill-posedness of blind deconvolution by characterizing the prior P(u) using natural image statistics, because numerically convenient (convex) priors tend to favor blurred images and the estimation gets stuck in the no-blur solution h=delta. We use the heavy-tailed distribution of image gradients

P(u) = C prod_i exp(-alpha||D_i u||^p),

where D_i u is the gradient at i-th pixel, and 0<=p<=1. Natural images have gradient distribution with roughly p=0.5. In the equivalent energy minimization formulation, this prior corresponds to L^p norm of image gradients and we are able to use even L^0 or nearly-L^0 (with continuous approximation around zero) norm, which produces nearly flat images with pronounced salient edges and this has been shown to aid the blur estimation.

For the numerical minimization, we use alternating direction method of multipliers, estimating the latent image u and the PSF h in an alternating manner. By using Fourier transform for direct matrix inversion or pixel-wise (thresholding-like) operations in the update equations, the algorithm is very fast, therefore applicable to large images and blurs. It has now become a standard to equip any MAP estimation of u,h with additional ingredients like multiscale approach, energy terms reweighing, selective PSF thresholding, image region reweighing, and more because without these steps, simultaneous u,h estimation struggles to work reliably. It has been pointed out that such steps are in fact essential and yet seldom given enough attention. We take care to carefully document, explain and analyze the whole estimation pipeline.

Often silently neglected is the problem of boundary artifacts in blind deconvolution. The information about pixel intensities outside the image support is missing and this produces heavy artifacts (ringing) in deconvolution if not properly treated, plus it also hinders the PSF estimation. Recently, several almost identical methods have been proposed which deal with boundary pixels and allow for fast implementation in FT. We further extend this method to deal with saturated pixels, which are automatically detected during the estimation process: we remove from the energy term all pixels, where the estimated value in u is outside the image intensity range, therefore saturated pixels are not penalized and do not harm the PSF estimation.

Lastly, given the large amount of blind blur estimation and deconvolution methods just in the last decade, there is an increasing need to compare the performance of a particular method with others. Currently, there is no standard widely used way of performance evaluation (result to ground-truth comparison). For blur estimation evaluation, people traditionally use either mean squared error (MSE) of the PSFs or MSE of the estimated sharp image to the original. The problem with PSF MSE is that while PSF estimation is a necessary intermediate step, the ultimate goal is the sharp image estimation. Therefore, PSF error should be measured as it would propagate to image estimation. Two PSFs with the same MSE to the common ground-truth can produce vastly different results (both in terms of MSE and human perception) when used for deconvolution. However, the problem with using simply the MSE of the deconvolved images for the PSF estimation evaluation is that the result is highly dependent on the used non-blind deconvolution method. Some non-blind methods are more forgiving to PSF error than others, some explicitly deal with border problems, some do not and the introduced error may overshadow the PSF error, all methods require tuning of usually several parameters. As a result, we can get completely different orderings (best to worst) of different PSF estimations just by using different non-blind methods or one method with different parameter settings, meaning that such process is useless as a performance evaluation tool across different author teams, where test conditions differ.

As a solution we propose a method which can accurately calculate the MSE introduced by the PSF to the image estimation just by one formula, without the need to perform the actual deconvolution. More specifically, the error measure we propose has the form

err = E[MSE(u_h, u_h')],

where u_h and u_h' are sharp images reconstructed with the correct PSF h and the estimated PSF h', respectively, and E is probabilistic expectation over images and noise, which are treated as random variables. The error measure reflects how the estimation error affects the image reconstruction. For example, an image with horizontal translation symmetry will be unaffected by horizontal blur, therefore if h and h' for such image differ only by a horizontal blur b (i.e., h = h'*b or vice versa), then E=0, which corresponds to the fact that the reconstructed images are identical. Plus, such 'one-formula' measure can very well serve as a standard because test results are easily reproducible, therefore meaningful.

To summarize, the main contributions of our work are detailed description and freely available implementation of fast blind blur estimation and deconvolution method base on heavy-tailed non-convex priors and incorporating other latest trends, extension of masking to automatically detect and handle saturated regions, and a new proposed method (plus implementation) for meaningful comparison of blur estimation results between different methods.

9404-11, Session 4

Motion deblurring based on graph Laplacian regularization

Amin Kheradmand, Univ. of California, Santa Cruz (United States); Peyman Milanfar, Univ. of California, Santa Cruz (United States) and Google (United States)

Graph-based image representation is an effective tool for describing the underlying structure of an image by encoding the similarity relationships between different pixels of the image in kernel similarity and associated Laplacian matrices. In this paper, we develop a regularization framework for image deblurring based on a new definition of the normalized graph Laplacian. We apply a fast scaling algorithm to the kernel similarity matrix to derive the symmetric, doubly stochastic filtering matrix from which the normalized Laplacian matrix is built. We use this new definition of the Laplacian to construct a cost function consisting of data fidelity and regularization terms for regularizing the ill-posed motion deblurring problem. The final deblurring estimate is obtained by minimizing the resulting cost function in an iterative manner. Furthermore, the spectral properties of the Laplacian matrix equip us with the required tools for spectral analysis of the proposed method. We verify the effectiveness of our iterative algorithm via synthetic and real examples for motion deblurring.

9404-12, Session 4

A system for estimating optics blur PSFs from test chart images

Radka Tezaur, Nikon Research Corp. of America (United States); Tetsuji Kamata, Nikon Corp. (Japan); Li Hong, Stephen D. Slonaker, Nikon Research Corp. of America (United States)

Photographic lenses suffer from optical aberrations that cause blur in captured images. Modern cameras featuring high resolution sensors can make this blur clearly visible even when expensive, high quality lenses are used. Also, even if the lens has aberrations well controlled, images can still exhibit blur that is due to the diffraction of light passing through a small aperture. It is possible to reduce or completely remove this blur by applying deconvolution as one of the steps in the image processing pipeline. In order to be able to do that, though, point spread functions (PSFs) that sufficiently accurately characterize the optics blur induced by the lens are needed. The optics blur PSFs are different for each color channel, and they also depend on the lens settings (focal length, focus distance, and f-number). Most importantly, though, the optics blur PSFs are also strongly spatially variant. In the case of ideal lens, perfectly corresponding to the lens design, the number of PSFs that are needed can be reduced by exploiting the rotational symmetry of the lens and PSFs need to be obtained only for a dense enough set of different image heights (distances from the center of the frame). However, in practice, lenses can exhibit significant amount of decentering, which results in PSFs that are not completely symmetric. To accurately characterize blur produced by a real lens, PSFs are needed for a sufficiently dense two-dimensional grid of points covering the entire image area. Obtaining so many PSFs in a reasonable amount of time is a big challenge. One possible way is estimating them using captured images of a special test chart (see, for example, the paper E. Kee et al.: Modeling and Removing Spatially-Varying Optical Blur, IEEE Int. Conf. Computational Photography, Anchorage, Alaska , 2011). We have developed a new test chart, which is convenient to use, and software for automatic processing of the captured test chart images, allowing us to obtain optics blur PSFs for an entire image frame in an efficient manner.

Our test chart design takes into consideration both the accuracy of PSF estimation and the ease of use. The chart consists of square black tiles with white random circle patterns that are separated by a white grid. The binary black and white image is easy to print and does not require any grayscale calibration. The black tile background reduces the problems with the cropped image boundaries during the PSF estimation and the random circle

pattern has been selected for its spectral properties, helping us to achieve good estimated PSF accuracy. The white grid separating the tiles makes it easier to properly align a camera to the chart. Also, it allows us to quickly and reliably segment the tiles, even in the presence of strong geometric distortion and vignetting, which can occur especially with wide angle lenses.

We estimate the PSFs at 805 different positions covering the entire frame, corresponding to the intersections of 23 horizontal and 35 vertical white lines in our test chart. At each position, the four surrounding black tiles are cropped out and used for PSF estimation. The corners of the tiles are used to estimate the parameters of an affine transformation that locally models the geometric distortion present at that part of the image. An ideal sharp image needed for non-blind PSF estimation is artificially synthesized by our software. It matches the correct pattern in each of the surrounding four tiles as well as the geometric distortion observed at that location. To estimate the PSF for each color channel we use a cost function comprising a least squares fidelity term and a total variation regularization term. To minimize the cost function we apply variable splitting that leads to an iterative algorithm in which the cost function is minimized in alternating fashion over the PSF variable and the auxiliary variables introduced by variable splitting. The convergence of the iterative algorithm is very fast. Typically, no more than 10 iterations are needed. We experimented also with some other methods for estimating the PSFs and we selected the one described above because it helps to reduce the noise in the estimated PSFs and yields PSFs that cause fewer artifacts in restored images, when the estimated PSFs are used for optics blur removal.

On a PC equipped with Intel Xeon W3530 CPU (2.80) GHz and 24 GB of memory, running Matlab version R2014a under 64-bit Windows 7 operating system, processing of a 40-megapixel captured image and estimating all 805 PSFs for each of the three color channels takes less than 3 minutes. We have tested the developed system with several different lenses and successfully used the estimated PSFs to remove optics blur from images of the test chart that were used for PSF estimation as well as from other types of images, capturing different scenes. The main limitation of the proposed system is that it is difficult to obtain optics blur PSFs for different magnifications or, in other words, for different focus distances, as this requires using test charts of a different size. Large focus distances are especially difficult, both because of the required size of the test chart and the space with controlled conditions needed to capture the test chart images. However, our experiments have shown that the restoration process possesses a fair amount of robustness with respect to the different focus distance and, with PSFs obtained from our test chart images, we were able to successfully reduce optics blur even in many images for which the focus distance was considerable larger, including infinity.

9404-13, Session 5

Computational photography and state of the art in image processing *(Keynote Presentation)*

Peyman Milanfar, Google (United States)

Modern image processing is the enabler for a new paradigm in consumer photography. It is the art, science, and engineering of producing a great shot (moving or still) from small form factor, mobile cameras. It does so by changing the rules of image capture -- recording information in space, time, and across other degrees of freedom -- while relying heavily on sophisticated and robust algorithms to produce a final result. Coupled with the ubiquity of devices, and recent algorithmic and hardware advances, open platforms for imaging will inevitably lead to an explosion of technical and economic activity, as was the case with other types of mobile applications. In this landscape, clever and robust algorithms, take center stage and enable unprecedented imaging capabilities in the user's hands. These recent approaches to image processing have brought together several powerful data-adaptive methods from Graphics, Computer Vision, Signal Processing, Machine Learning, and Statistics. These approaches are deeply intellectually connected. In this talk, I will present a framework for understanding some common underpinnings of these methods, leading to new insights and a broader understanding of how these diverse methods result in the state of the art.

9404-14, Session 5

Gradient-based correction of chromatic aberration in the joint acquisition of color and near-infrared images

Zahra Sadeghipoor Kermani, Ecole Polytechnique Fédérale de Lausanne (Switzerland); Yue M. Lu, Harvard Univ. (United States); Sabine Süsstrunk, Ecole Polytechnique Fédérale de Lausanne (Switzerland)

No Abstract Available

9404-15, Session 5

Visible and near-infrared image fusion based on visually salient area selection

Takashi Shibata, NEC Corp. (Japan) and Tokyo Institute of Technology (Japan); Masayuki Tanaka, Masatoshi Okutomi, Tokyo Institute of Technology (Japan)

A typical digital camera with a silicon-based sensor is capable of capturing near-infrared (NIR) light by removing a hot mirror, which is in front of the sensor to block NIR. Recent advances in computational photography techniques (e.g. designing color filter array or beam-splitter) provided an easy way to acquire a NIR image and a visible image simultaneously. This paper presents a novel image fusion algorithm for a visible image and a NIR image.

For the proposed fusion, the image is selected pixel-by-pixel based on local saliency. In this paper, the local saliency is measured by a local contrast. Then, the gradient information is fused and the output image is constructed by a Poisson image editing preserving the gradient information of both images. The proposed framework provides various applications including denoising, dehazing, and image enhancement. Experimental results demonstrate that the proposed algorithm has comparable or even superior performance to existing methods, which are proposed for each specific application, in terms of image quality.

9404-16, Session 5

Fast HDR image upscaling using locally adapted linear filters

Hossein Talebi, Guan-Ming Su, Peng Yin, Dolby Labs., Inc. (United States)

A new method for upscaling high dynamic range (HDR) images is introduced in this paper. Overshooting artifact is the common problem when using linear filters such as bicubic interpolation. This problem is visually more noticeable while working on HDR images where there exist more transitions from dark to bright. Our proposed method is capable of handling these artifacts by computing a simple gradient map which enables the filter to be locally adapted to the image content. This adaptation consists of first, clustering pixels into regions with similar edge structures and second, learning the shape and length of our symmetric linear filter for each of these pixel groups. This new filter can be implemented in a separable fashion which perfectly fits hardware implementations. Our experimental results show that training our filter with HDR images can effectively reduce the overshooting artifacts and improve upon the visual quality of the existing linear upscaling approaches.

9404-17, Session 5

Cinematic camera emulation using two-dimensional color transforms

Jon S. McElvain, Walter C. Gish, Dolby Labs., Inc. (United States)

No Abstract Available

9404-18, Session 6

Image quality assessment using the dead leaves target: experience with the latest approach and further investigations

Uwe Artmann, Image Engineering GmbH & Co. KG (Germany)

The so-called „texture loss“ is a critical parameter in the objective image quality assessment of todays cameras. Especially cameras build in mobile phones show significant loss of low contrast, fine details which are hard to describe using standard resolution measurement procedures. The combination of very small form factor and high pixel count leads to a high demand of noise reduction in the signal-processing pipeline of these cameras. Different work groups within ISO and IEEE are investigating methods to describe the texture loss with an objective method. The so-called dead leaves pattern has been used for quite a while in this context.

Image Engineering could present a new intrinsic approach at the Electronic Imaging Conference 2014, which promises to solve the open issue of the original approach, which could be influenced by noise and artifacts. In this paper, we present our experience with the new approach for a large set of different imaging devices.

We show, that some sharpening algorithm found in todays cameras can significantly influence the Spatial Frequency Response based on the Dead Leaves structure (SFR_DeadLeaves) results and therefore make an objective evaluation of the perceived image quality even harder. For an objective comparison of cameras, the resulting SFR needs to be reduced to a small set of numbers, ideally a single number. The observed sharpening algorithms lead to much better numerical results, while the image quality already degrades due to strong sharpening. So the measured, high SFR_DeadLeaves result is not wrong, as it reflects the artificially enhanced SFR, but the numerical result cannot be used as the only number to describe the image quality. We propose to combine the SFR_DeadLeaves measurement with other SFR measurement procedures as described in ISO12233:2014. Based on the three different SFR functions using the dead leaves pattern, sinusoidal Siemens Stars and slanted edges, it is possible to obtain a much better description if the perceived image quality. We propose a combination of SFR_DeadLeaves, SFR_Edge and SFR_Siemens measurements for an in-depth test of cameras and present our experience based on todays cameras.

9404-19, Session 7

An ISO standard for measuring low light performance

Dietmar Wüller, Image Engineering GmbH & Co. KG (Germany)

To measure the low light performance of todays cameras has become a challenge. The increasing quality for noise reduction algorithms and other steps of the image pipe make it necessary to investigate the balance of image quality aspects.

The first step to define a measurement procedure is to capture images under low light conditions using a huge variety of cameras and review the images as well as the metadata of these images.

Image quality parameter that are known to be affected by low light levels are noise, resolution, texture reproduction, color fidelity, and exposure. For each of the parameters thresholds below which the images get unacceptable need to be defined. Although this may later on require a real psychophysical study to increase the precision of the thresholds the current project tries to find out wether each parameter can be viewed as an independent one or if multiple parameters need to be grouped to differentiate acceptable images from unacceptable ones.

Another important aspect is the definition of camera settings? For example the longest acceptable exposure time and how this is affected by image stabilization. Cameras on a tripod may produce excellent images with multi second exposures.

After this ongoing analysis the question is how the light level gets reported? Is it the illuminance of the scene, the luminance of a certain area in the scene or the exposure? All these aspects are currently investigated and the results will be presented in the paper as well as the strategy ISO 19093 will follow to define the measurement of low light performance of cameras.

9404-20, Session 7

ISO-less?

Henry G. Dietz, Univ. of Kentucky (United States); Paul Eberhart, University of Kentucky (United States)

The concept of specifying the sensitivity of photographic emulsions to light by a single number dates at least to Warnerke’s Sensitometer in 1880 and has continued to evolve through standards such as DIN 4512, ASA Z38.2.1-1943, and ISO 12232:20076. Various procedures are defined for measuring the light sensitivity of different films, and there are several methods permitted for determining ISO speed of a digital camera. The catch is that most digital cameras allow the ISO setting to be changed -- but what really changes is not clear. It probably is not the quantum efficiency (QE) of the sensor, but more likely an analog gain applied before analog-to-digtal conversion. Given a converter that covers the full dynamic range of the sensor with sufficient resolution, would gain adjustments have a significant impact on the image quality?

The claim has been made that many digital cameras are effectively “ISO-less,” with little change in image quality no matter what the ISO setting. For example, nearly identical JPEG images can be produced from raw captures using a Sony A7 for a well-exposed image at ISO 1600 and a 4-Ev underexposed image at ISO 100. In contrast, the Canon 5D III image quality is dependent on ISO setting, as the Magic Lantern (ML) “dual ISO” support clearly demonstrates.

This paper evaluates the “ISO-less” properties of a variety of consumer digital cameras using three methods. The first involves interpretation of sensor dynamic range measurements published by DxOMark. The second directly compares raw image captures using different ISOs with all other parameters held constant. However, both those methods can be biased by differences between the actual and reported (so called “market”) camera settings, and they also fail to account for the impact of processing inside the camera to create the JPEG files most photographers use. Thus, the third method involves using the Canon Hack Development Kit (CHDK) and ML to put custom code inside various Canon cameras so that ISO setting and exposure parameters can be precisely controlled and the camera’s native JPEG processing can be applied.

Having characterized the behavior of many cameras in response to ISO changes, it becomes clear that most cameras are not entirely ISO-less, but that ISO setting changes do not have the expected effect with respect to changes of other exposure parameters. Various researchers have studied how ISO setting choices can be optimized for multi-shot high dynamic range (HDR) imaging, but understanding how ISO settings can be decoupled from other exposure parameters can lead to better image quality even for single captures. Thus, a new approach to selecting ISO settings and other exposure parameters based on observing the scene dynamic range is proposed. This approach is implemented and experimentally evaluated inside Canon PowerShot cameras using CHDK to install custom code.

9404-21, Session PTues

Overcoming the blooming effect on autofocus by fringe detection

Shao-Kang Huang, Dong-Chen Tsai, Homer H. Chen, National Taiwan Univ. (Taiwan)

In the presence of light bloom or glow, multiple peaks may appear in the focus profile and mislead the autofocus system of a digital camera to an incorrect in-focus decision. We present a novel method to overcome the blooming effect. The key idea behind the method is based on the observation that multiple peaks are generated due to the presence of false features in the captured image, which, in turn, are due to the presence of fringe (or feather) of light extending from the border of the bright image area. By detecting the fringe area and excluding it from focus measurement, the blooming effect can be reduced. Experimental results show that the proposed anti-blooming method can indeed improve the performance of an autofocus system.

An autofocus system typically contains two basic operations: measuring image sharpness (focus measurement) [1]?[5] and searching for in-focus lens position [6]?[8]. The search performance of an autofocus system relies in part on the accuracy of focus measurement. When pointing a camera at a scene, the focus values of the images captured by the camera as the lens moves along the optical axis form a focus profile. The focus profile normally has a single peak that corresponds to the in-focus lens position, at which the sharpest image can be obtained.

However, when taking photos of a scene with intensely bright light shining into the camera, the photo sensors receiving the bright light may obtain excessive photons that diffuse into neighboring photo sensors. If the bright part of the image where the photo sensors are overly exposed is adjacent to a relatively dark part, the fringe of light would extend from the bright area to the dark area and create the so-called blooming effect. As a result, false features that do not exist in the original scene are generated in the fringe region, raising the focus value of the image improperly and resulting in multiple peaks in the focus profile. In a circumstance like this, the competition between the true features and the false features affects the performance of autofocus. If the overall strength of the true features is stronger than that of the false features, the in-focus lens position can still be detected. Otherwise, the existence of multiple peaks may fool the autofocus system and make it capture a blurry image corresponding to a false peak. As an illustration, a three-peak focus profile and the captured images corresponding to the three peaks are shown in Fig. 1. In general, the likelihood of the focus profile being multi-peaked is higher when the camera has a shallower depth of field because the captured image has a larger out-of-focus area (and hence weaker true features).

In short, multiple peaks may appear in a focus profile due to the blooming effect and mislead the autofocus system to an incorrect in-focus decision. Our goal is to prevent the blooming from affecting the in-focus decision for a digital camera during the autofocus process.

Most previous anti-blooming approaches are sensor centric in that they seek to build a drain structure into the photo sensors [9]?[11] to prevent the occurrence of blooming. The resulting image may appear too dim for dark regions and requires additional tone mapping operation or its variants to enhance the image appearance. In contrast, our work aims at resolving the impact of blooming, if occurs, on autofocus at the image processing pipeline level of a digital camera. Specifically, we detect the fringe regions and exclude them from the focus measurement process to reduce the impact of blooming on autofocus.

The idea of using multiple focus windows at different locations of the image has been proposed to improve the accuracy of autofocus [12]. However, this approach does not work for the problem considered here. When blooming appears, some of the focus windows will be attacked by the blooming, and the result of in-focus decision for these focus windows will be incorrect, as discussed above. The result of the in-focus decision for the other focus windows, although remaining intact, will be usable only if these focus windows are on the object of interest. In practice, this is rarely the case unless the object of interest is a big planar surface covering most of the image and parallel to the image plane.

Specifically, our method reduces the impact of blooming on autofocus by detecting the fringe region in the captured image and excluding it from the computation of focus value. The proposed method contains two main steps: 1) fringe detection and 2) focus value computation. As described earlier, the fringe region extends from the border of a bright region. Since the sensors in the bright region must have saturated before the photons start to overflow into sensors in the fringe region, we locate the saturate region in the image before detecting the fringe region. All the operations are performed on the image area within the small focus window from which the focus value of the image is computed. The extraction of saturate and fringe regions is performed on the luminance component only; the chrominance components are discarded.

For each pixel of the focus window, luminance and gradient criteria are applied to determine candidate saturate pixels. Then the morphological opening operator is applied to the image to filter out incorrect candidates caused by noise. Next, we detect the fringe region. All pixels external to the saturate region are candidate pixels of the fringe region. Because the fringe of light appearing in the image is the result of excessive photons spilling over the saturated photo sensors, the fringe region is typically a ring around the saturate region. Finally, we compute the focus value. Only the gradients of the pixels outside the saturate and fringe regions are considered in the focus value computation.

The proposed method is tested on a RED Epic-M digital cinema camera equipped with the Mysterium-X sensor [13] and a Canon 70-300mm F4-5.6L IS USM telephoto lens. The results of three typical test scenes in our experiments are shown in Fig. 2. We can see that all focus profiles generated by our anti-blooming method have one sharp peak and that the peak of each focus profile is located at the correct in-focus lens position. In contrast, the focus profiles obtained without anti-blooming have multiple peaks, and the largest peak is not always located at the true in-focus lens position. Take Fig. 2(d) as an example, an autofocus system can easily take lens position 39 to be the in-focus position since it gives the highest focus value. With our method, the autofocus system can avoid such misjudgment. A more detailed version of Figs. 2(a) and 2(d) is provided in Figs. 3 and 4, respectively. The middle row of Fig. 3 shows the feature maps of the sequence of images filtered by a difference-of-Gaussian kernel. As shown in Fig. 4, the presence of false features improperly increases the focus value, particularly when the lens is farther away from the in-focus position. Eventually, it dwarfs the true peak. By eliminating the contribution of such false features to the final focus value, we can keep the focus profile in proper shape with a single peak at the true in-focus lens position. Fig. 2(b) shows a case where the size of the light beam occupies a good portion of the focus window. In this case, our method would accordingly increase the excluding region. But the amount of features in the focus window is already small, even for the in-focus image. A high threshold would evidently result in a decrease of the focus value in our method. Despite of the decrease, we can see from Fig. 2(e) that the anti-bloomed focus profile is nicely peaked. This appealing feature enables an autofocus system to accurately locate the in-focus lens position. In contrast, the focus profile obtained without anti-blooming has multiple peaks and is likely to fail an autofocus system.

Fig. 2(c) shows a case more complicated than the ones in Figs. 2(a) and 2(b). While the native features are strong enough to sustain the attack of false features under the blooming effect and maintain a sharp peak at the in-focus lens position, a local peak appears at around lens position 19. Therefore, an autofocus system trapped into the local peak would generate a blurry image. In contrast, the anti-bloomed focus profile generated by our method is clean and has one single peak.

In this work, we develop a novel anti-blooming method for digital autofocus. It reshapes a multi-peak focus profile into a single-peak focus profile as desired without any prior knowledge of the optical characteristics of the camera. Furthermore, the in-focus lens position remains intact in the reshaping process so that the sharpest image is captured. Since only simple image processing operations are involved, the computational overhead introduced by the proposed method is confined to a small range without affecting the real-time performance of an autofocus system.

The proposed method has been developed under the assumption that that the saturate area and the fringe area due to blooming only partially occupy the focus window used for autofocus. It is our plan to remove this assumption by adaptive window resizing and reselection in the future.

9404-22, Session PTues

Stable image acquisition for mobile image processing applications

Kai-Fabian Henning, Alexander Fritze, Eugen Gillich, Uwe Mönks, Volker Lohweg, Ostwestfalen-Lippe Univ. of Applied Sciences (Germany)

Motivation

In today's world, mobile devices in form of smartphones and tablets are widespread and of high importance for their users. They not only enable communication, but also allow for daily planning (calendar, timer, notices, etc.) as well as use of multimedia contents (music, images, etc.) or games. People rely on their devices and immerse them into their daily life. Hence, the market of mobile devices is still growing. Over time, versatility and performance of such devices increase. Accordingly, the number and quality of embedded sensors increase. This leads to the opportunity of using mobile devices for more specific tasks like image processing for mobile health [1] or in an industrial context, e.g. banknote authentication [2].

For the analysis of images, meeting certain requirements is crucial, i.e. the image quality (blur, contrast, illumination, etc.) as well as a defined relative position of device and object to be inspected. In order to fulfill these requirements using a mobile device, some obstacles have to be taken into account. First, the image sensors are low-cost and prone to a certain amount of image noise. Second, contrary to conventional image processing applications, a mobile device is handheld and not fixed in its position. This leads to motion blur artifacts. Third, mobile devices are used in constantly changing environments implying that different ambient illuminations have to be considered. These obstacles have to be overcome to enable a proper image acquisition in the context of image processing on mobile devices.

Approach

We present a new approach for handheld stabilized image acquisition of an arbitrary planar object. Therefore, object detection methods are combined with sensor fusion concepts.

Our goal is to guide the user moving the device to a defined position and to automate the image capture process. The latter is triggered depending on the alignment of the device and the object. Furthermore, an image acquisition state that depends on the current motion and environment of the device is considered for triggering.

The approach comprises object detection, motion estimation, and sensor fusion.

Object Detection In the first step the object has to be localized and its position relative to the mobile device has to be estimated. We refer to this as pose estimation. Basically our method relies on a marker-less feature detection approach. The approaches presented in [3] and [4] are analyzed. Both comprise methods to detect, describe and match feature points of an image, so that the pose estimation steps are the same.

Initially, the feature points in an image of a reference object are detected and described. During runtime, feature points are detected and described in the current camera frame. Next, the descriptions of the camera frame are matched with the reference descriptions to get a set of corresponding feature points.

If a certain amount of corresponding point pairs is matched, the approach presented in [5] is used to estimate a 3?3 rotation matrix and 3?1 translation vector. Rotation and translation define a three-dimensional (3D) transformation of the reference object to the real object position relative to the device.

We apply the transformation to the reference object boundaries to receive a 3D-contour representing the real object position. This contour is projected into the two-dimensional (2D) camera frame to give the user a visual feedback.

Note that the object is detected even if object and background have similar texture and contrast. Therefore, our approach is applicable to a wide range of planar objects on arbitrary backgrounds.

Motion Estimation Embedded sensors of a mobile device provide information about its current motion, orientation, etc. In combination with the estimated pose we compute the relative movement between the device and the object. We use this to predict the object's position and a region-of-interest (ROI) for feature detection in the subsequent camera frames. This accelerates the pose estimation process.

Sensor Fusion The sensors' information is also used to determine an image acquisition state that supports the triggering for image acquisition. For example, if the accelerometer delivers continuously and rapidly changing values this implies that a capture without motion blur is not possible. This kind of information is integrated into the application by sensor fusion concepts as follows:

In order to compute a score value for the image acquisition state, we map sensor data to fuzzy membership functions [6, 7, 8]. A membership function describes the degree of membership of a feature, in this case single sensor values, to a defined class. For image acquisition only two classes are required. Either an image should be captured or not.

To be more robust in the detection of an appropriate image acquisition state, we include the system introduced in [6] and refined in [9] and [10]. Besides sensor data, image quality measures are used to create membership functions. By aggregation of all membership values an overall score value for the image acquisition state is computed.

Finally, image acquisition is triggered if the following conditions are met: First, the object has to be in the right position defined by the tracking results. Second, the image capturing state has to be above a certain threshold.

Results

The proposed approach provides a reliable user support to move the device to a predefined position relative to the object to be inspected. Furthermore, the automated stabilized image acquisition process leads to significantly improved image quality in terms of image processing requirements.

Conclusion and Outlook

This new approach results in stable handheld image capture (contrast, illumination, motion blur, etc.) of arbitrary objects in a predefined position. This is a fundamental task for object inspection applications on mobile devices. Thus, it is applicable to a broad field of applications. In future work the performance will be improved and investigations for adapting our approach to 3D objects will be executed.

References

[1] Perera, C.; Chakrabarti, R.: The utility of mHealth in Medical Imaging. In: Journal MTM 2(3), 2013;

[2] Lohweg, V.; Dörksen, H.; Hoffmann, J. L.; Hildebrand, R.; Gillich, E.; Hofmann, J.; Schaede, J.: Banknote authentication with mobile devices. In: IS&T/SPIE Electronic Imaging, Media Watermarking, Security, and Forensics, 2013.

[3] Bay, H.; Tuytelaars, T.; Gool, L. V.: SURF: Speeded Up Robust Features. In: ECCV 2006, 9th European Conference on Computer Vision, Bd. 3951, Springer, 2006.

[4] Taylor, S.; Rosten, E.; Drummond, T.: Robust feature matching in 2.3µs. In: IEEE CVPR Workshop on Feature Detectors and Descriptors: The State of the Art and Beyond, 2009.

[5] Lepetit, V.; Moreno-Noguer, F.; Fua, P.: E PnP: An Accurate O(n) Solution to the PnP Problem. In: International Journal Computer Vision, vol. 81, no. 2, 2009.

[6] Zadeh, L. A.: Fuzzy Sets, Information and Control, vol. 8, no. 3, pp. 338–353, 1965.

[7] Mönks, U.; Voth, K.; Lohweg, V.: An Extended Perspective on Evidential Aggregation Rules in Machine Conditioning. In: IEEE CIP 2012, 3rd International Workshop on Cognitive Information Processing, pp. 1-6, 2012.

[8] Bocklisch, S. F.; Priber, U.: A parametric fuzzy classification concept. In: Proc. International workshop on Fuzzy Sets Applications, Akademie-Verlag, Eisenach, Germany, 1986.

[9] Mönks, U.; Lohweg, V.: Machine Conditioning by Importance Controlled Information Fusion. In: IEEE ETFA 2013, 18th International Conference on Emerging Technologies and Factory Automation, pp. 1-8, 2013.

[10] Mönks, U.; Lohweg, V.: Fast Evidence-based Information Fusion. In: IEEE CIP 2014, 4th Internation Workshop on Cognitive Information Processing, pp. 1-6, 2014.

9404-23, Session PTues

Near constant-time optimal piecewise LDR to HDR inverse tone mapping

Qian Chen, Guan-Ming Su, Dolby Labs., Inc. (United States); Peng Yin, Dolby Labs Inc (United States)

In a backward compatible HDR image/video compression, it is a general approach to reconstruct HDR from compressed LDR as a prediction to original HDR, which is referred to as inverse tone mapping. Experimental results show that 2-piecewise 2nd order polynomial has the best mapping accuracy than 1 piece high order or 2-piecewise linear, but it is also the most time-consuming method because to find the optimal pivot point to split LDR range to 2 pieces requires exhaustive search. In this paper, we propose a fast algorithm that completes optimal 2-piecewise 2nd order polynomial inverse tone mapping in near constant time without quality degradation. We observe that in least square solution, each entry in the intermediate matrix can be written as the sum of some basic terms, which can be pre-calculated into look-up tables. Since solving the matrix becomes looking up values in tables, computation time barely differs regardless of the number of points searched. Hence, we can carry out the most thorough pivot point search to find the optimal pivot that minimizes MSE in near constant time. Experiment shows that our proposed method achieves slight better PSNR, while saving 3 times computation time than traditional exhaustive search in 2-piecewise 2nd order polynomial inverse tone mapping with continuous constraint.

9404-24, Session PTues

Face super-resolution using coherency sensitive hashing

Anustup Choudhury, Andrew Segall, Sharp Labs. of America, Inc. (United States)

No Abstract Available

9404-25, Session PTues

An evaluation of the effect of JPEG, JPEG2000, and H.264/AVC on CQR codes decoding process

Max E. Vizcarra Melgar, Mylène C. Q. Farias, Alexandre Zaghetto, Univ. de Brasília (Brazil)

QR Codes are two-dimensional structures used to transmit information through a print-scan channel. They were proposed in 1994 by the Japanese company Denso Wave Incorporated. In a previous publication, we proposed a technique to generate colored QR Codes. Up to our knowledge, this was the first technique used to generate Colored QR (CQR) Codes which had the purpose of increasing the stored data capacity without using black modules in the Encoding Region. Given that the capture, storage and transmission of a QR Code may involve its compression, it is necessary to evaluate the effect of lossy compression on the decoding process. In this paper, our goal is to evaluate the effect that degradation inserted by common image compression algorithms has on the decoding process. At the same time, we verify the maximum compression rate that a CQR Code image can be compressed without affecting the decoding process. For this study, we have tested three popular compression algorithms: the JPEG, JPEG2000, and H.264/AVC. JPEG is still the most popular compression standard, while JPEG2000 is its successor. H.264/AVC, on the other hand, is a video compression standard that has recently been used for image compression, surpassing the JPEG2000 performance. The proposed Colored QR (CQR) Codes are made up of 49 ? 49 modules. The colors red (255,0,0), green (0,255,0), blue (0,0,255) and white (255,255,255) are used to represent information and redundancy bits, while the colors black (0,0,0) and white (255,255,255) are used as Function Patterns. These set of colors were chosen because they are all maximally equidistant on the RGB color space, what makes them less prone to errors in the decoding stage. Modules are distributed over two distinct regions: Function Patterns and Encoding Region. The modules of the Function Patterns have the sole purpose of detecting the presence of a CQR Code in the image. From the 2401 (49 ? 49) modules of the CQR Code, 192 are inside the Function Patterns and 2209 are inside the Encoding Region. Since each module inside the Encoding Region represents 2 bits, there are 4418 bits available to carry data.

Our tests were performed with CQR Codes with seven modules of Quiet Zone on each side. The original CQR Code images (1 module per pixel) were resized to make them ten times larger, resulting in a CQR Code of 630 ? 630 modules, 490 pixels of Encoding Regions, and 140 pixels of Quiet Zone in the vertical or horizontal directions. The input images are in 8 bits/pixel bitmap format (RGB). In this paper, we evaluate the distortions from the compression using the algorithms JPEG, JPEG2000, and H.264/AVC.

In the case of JPEG compression, CQR Codes can be decoded from rates greater than 0.3877 bpp (average value for set), which corresponds to the highest compression rate (level 100 in JPEG compression). At this rate, the CQR Code obtains (on average) 74 corrupted symbols, which translates into an average percentage of 26.81% of symbols. This percentage of corrupted symbols can be corrected by the Reed-Solomon algorithm. For JPEG2000 compression, CQR Codes can be decoded with compression rates greater than 0.1093 bpp (average value for set), which corresponds to level 212 in JPEG2000 compression. For this rate, an average of 97.4 corrupted symbols were found, which corresponds to an average percentage of 35.30% of symbols. Rates like 0.1001 or 0.0810 cause the CQR Code to be decoded with a greater percentage of corrupted symbols than what is allowed. Finally, for H.264/AVC compression, CQR Codes can be decoded with rates higher than 0.3808 bpp (average value for set). This rate is also the highest rate possible, which corresponds to level 51 in H.264/AVC. The average percentage of corrupted symbols for this rate is 0.18%. It was verified that the CQR Code is successfully decoded for compression rates higher than 0.3877 bpp, 0.1093 bpp and 0.3808 bpp for JPEG, JPEG2000 and H.264/AVC, respectively. The algorithm that presented the best performance was H.264/AVC, followed by JPEG2000, and, finally, by JPEG.

9404-26, Session PTues

Stitching algorithm of the images acquired from different points of fixation

Evgeny Semenishchev, Vacheslav Voronin, Vladimir Marchuk, Maria Pismenskova, Don State Technical Univ. (Russian Federation)

I. INTRODUCTION

There are various factors that cause difficulties in the problem of united image obtaining, including the formation of image composition. Also it includes problems with combining images acquired from different points of fixation. Figure 1 shows an example of fixation frame, obtained of the free point of view. Points A and B is subject of fixation camera. When it moves from point A to point B, the follow characteristics could be changed: illumination, viewing angle, brightness, focal length, etc. Also camera destabilization could exist and it causes blur or rotation of fixation point relatively first camera. These parameters affect the further steps of stitching images.

Image obtained by combining frames from different points of fixation of one object depends on the choice of conjugation points. The location of these points with respect to the fixation point may give different results when obtaining combined image. Obtain of final image is also affected by: the increased time between frames by the movement of the camera, the focal length, the nature of relationship between the subject and its background.

In the paper [2] considered the main shortcomings of the algorithms for combining images. The most computationally simple algorithms which are used to combine images have these disadvantages. Algorithms with the minimal or without artifacts are computationally complex. In [3, 4], the presented algorithms are based on the method SURF. These algorithms

automatically stitch panoramic image, and they may produce the artifacts when combine frames of different scales. Another computationally complex example uses methods based on the neural network [5 and other]. Also, examples of algorithms allowing the combination of images can be found in the papers Davis, J, He Panli, Gao G, Zhaoxia Fu, Guangyu Bian, Yonggi Wu et al. Most of the algorithms are computationally complex and don't give good results always in the case of the conventional stitching images obtained from different points of fixation.

II. MODEL STITCHING IMAGES

Figure 2 shows the model of the stitching image.

A simplified mathematical model of the stitching image is represented as:

, (1)

where: - first image; - second image; - region that stitch the images together; - the first image with the region that stitch the images together; - the second image with the region that stitch the images together.

Substituting the expressions for and in (1) and introducing the stitching parameter, we obtain the following form of the mathematical model:

, (2)

where: ? – coefficient of the area transformations.

III. ALGORITM STITCH THE IMAGES TOGETHER

Most of the standard algorithms creating a mosaic image do not take account the difference in scale, the possibility of combining images taken from different fixation, the intensity change of illumination. To complement the presented opportunities and improve processing speed in this paper we propose an algorithm that stitch the images together acquired from different points of fixation (Figure 2).

Analysis of the effectiveness will be evaluated with a couple of test images with a slight different scale, angles and obtained from different points of fixation.

This algorithm is implemented as follows (basic steps). On the first step is carried out loading images. In the second stage of the algorithm the detection of the boundaries is produced. An analysis of the literature reveals that the Canny detector shows the best visual and quantitative results when works with clean pictures as well as with images subject to various distortions.

In the next step, analysis of detailed objects is produced for the selected boundaries. Found detailed objects allow you to split the image into the informative area and background. Combining of uninformative areas do not require complex changes with the image. As a method of analysis is used the method of "density", because it does not require operator intervention.

When looking for detalisation objects by "density" on the first step the general factor of detail is considered on the whole image, which is defined by the formula (3)

, (3)

where: - value of the pixel with coordinates and ; – rows; – columns; – base coefficient of detalisation.

On the next step, similarly to expression (3) calculation of the density is produced in each sliding window.

(4)

where - general coefficient of detalisation; 0,1- averaging factor associated with the automatic selection of the window size equal to 10% of the total image.

Next, and are compared and a decision is made about of detail in this window. In the next step, the window shifts and is made similar calculations for it (4).

In the next step, each detalisation object is separated from the others and key points are searched in this object. For these reference points is carried out coordinate binding images.

Subsequent processing is performed in the each of identified areas. In base of the analysis is method of search for correspondences, shown that the method of SURF is one of the most efficient and fast algorithms. This approach identifies the specific points on the image and creates their descriptors that are invariant to scale and rotation. This means that the description of the key points will be the same, even if the sample size would be changed and rotated by an angle.

Point found in the previous step, in most cases, contains erroneous compliance. At the next stage the search of the false compliance is produced. To eliminate false correspondences analysis of their cross-correlation is used. Selection a block of 5?5 around each point of compliance for all detalisation objects in all images. The correlation coefficient is calculated between the blocks of the corresponding points of the corresponding detalisation objects. In case of exceeding the threshold sigma blocks are considered coincident. If it is less or equal to, the point is considered to be false. As a result analysis of all area discards false leaving only the true.

Since the basic steps are made for the areas with hight-detalisation, this improves performance when combining high-resolution images. The next step is to analyze the key points and determination of the scaling factor, contrast coefficient and ? coefficient of area transformations.

Combining images produced with the linewise bonding by means of the analysis of edges in the combining area (optimal seam searching result). Averaged result of analysis shows an increasing of processing speed with high-resolution images up to 40% (The table comparing the processing speed from size of images will be built). Examples of combining are presented in Figure 4.

IV. CONCLUSION AND FUTURE WORK

REFERENCES

2. Williams Don, Burns Peter D, Image Stitching: Exploring Practices, Software, and Performance. Archiving Conference, Archiving 2013 Final Program and Proceedings, pp. 126-131(6).

3. Jun Zhu, Mingwu Ren. Image Mosaic Method Based on SIFT Features of Line Segment. Computational and Mathematical Methods in Medicine. Volume 2014 (2014), Article ID 926312, 11 pages.

4. Lin Zeng, Shengping Zhang, Jun Zhang, Yunlu Zhang. Dynamic image mosaic via SIFT and dynamic programming. Machine Vision and Applications (2014) 25:1271–1282. DOI 10.1007/s00138-013-0551-8.

5. Zhiyuan Li, Weiting Kong, Yongzhao Zhan, and Junlei Bi. A Novel Image Mosaicking Algorithm for Wireless Multimedia Sensor Networks. Hindawi Publishing Corporation International Journal of Distributed Sensor Networks Volume 2013, Article ID 719640, 9 pages.

Conference 9405: Image Processing: Machine Vision Applications VIII

Tuesday - Wednesday 10–11 February 2015

Part of Proceedings of SPIE Vol. 9405 Image Processing: Machine Vision Applications VIII

9405-1, Session 1

Multiple object detection in hyperspectral imagery using spectral fringe-adjusted joint transform correlator

Paheding Sidike, Vijayan K. Asari, Univ. of Dayton (United States); Mohammad S. Alam, Univ. of South Alabama (United States)

Hyperspectral imaging (HSI) sensors provide plenty of spectral information to uniquely identify materials by their reflectance spectra, and this information has been effectively used for target detection and identification applications. HSI detection algorithm can be generally classified into stochastic and deterministic approaches. Deterministic approaches are comparatively simple to apply since it does not require the statistical information of the targets and background classes. Some popular deterministic algorithms such as spectral angle mapper (SAM) and spectral fringe-adjusted joint transform correlation (SFJTC) have been used for hyperspectral image processing. Specifically, the decision criterion of SAM is based on the similarity of the angle between two spectral vectors, while the SFJTC determines a desired target by analyzing the correlation peak intensity between an unknown spectral signature and a known reference spectrum. Compared to SAM, the SFJTC technique is able to accommodate noise and variations of the spectral signatures. However, both SAM and SFJTC techniques were designed to detect only similar patterns in constant time using hyperspectral information. Although the JTC based dissimilar pattern detection algorithms from two-dimensional image have been introduced in some literatures, but HSI based dissimilar target detection using JTC is yet to be done. Thus, in this paper, a new HSI deterministic approach is proposed to perform multiple dissimilar target detection in hyperspectral imagery by use of class-associative filter design. In this technique, input spectral signatures from a given hyperspectral image data cube are correlated with the multiple reference signatures using class-associative technique. To achieve better correlation output, the concept of SFJTC and the modified Fourier-plane image subtraction technique are incorporated in the multiple target detection process. The output of this technique provides sharp and high correlation peaks for a match and negligible or no correlation peaks for a mismatch. In detail, if there are dissimilar patterns present in the scene, the proposed algorithm yields equal-high correlation peaks for each target simultaneously without losing inherent advantage of the SFJTC and class-associative filtering. Similar to some deterministic target detection approaches, it also does not require any training step in whole detection process, whereas in many statistical machine learning techniques, the reference signature and none-target information are required before performing target recognition process. Furthermore, the decision matrix such as peak-to-clutter mean is also integrated in the algorithm which makes the performance of the proposed technique is highly based on the signature (shape) of the target but not the amplitude. This makes the proposed algorithm intensity invariant since the reflectance information of a material is usually preserved in hyperspectral imagery whereas the intensity may change due to environmental conditions. The feasibility of the proposed technique has been tested using real-life hyperspectral imagery. Computer simulation results show that the proposed algorithm can successfully detect multiple dissimilar targets while satisfying the equal correlation peak criteria by adjusting parameters in the class-associative JTC filter formulation, such that it will be an excellent candidate of a pattern recognition technique in HSI for near-to real-time applications.

9405-2, Session 1

Dynamic hierarchical algorithm for accelerated microfossil identification

Cindy M. Wong, Dileepan Joseph, Univ. of Alberta (Canada)

MOTIVATION:

Object recognition is a difficult problem within computer vision. For some applications, human-based computation may be the best approach, whereby a computer algorithm outsources steps to humans. We believe microfossil identification is such an application, one of sufficient importance to engage human volunteers. Marine microfossils provide a useful record of the Earth's resources and prehistory via biostratigraphy. In particular, to study hydrocarbon reservoirs and prehistoric climate, geoscientists identify the species of foram tests, i.e., fossilized remains of the phylum foraminifera, found in core samples. Hundreds of miles of core are available in public and private repositories. Only a tiny fraction has been used for biostratigraphy because microfossil identification is labour intensive.

STATE OF THE ART:

Automated microfossil identification has been investigated since the 1980s, initially with rule-based systems. These systems assisted knowledgeable users with identification through the refinement of a list of possible species. Unfortunately, users still had to examine each specimen under a microscope. While artificial neural network (ANN) systems have shown more promise for reducing expert labour, they have not displaced manual identification. One system relied on expensive scanning electron microscope images that rendered specimens unsuitable for geochemical analysis. Another system had a high correct rate for one species using optical images, but also a high incorrect rate all species considered. A third optical system did not have these issues. However, its "fat" ANN architecture and relatively small dataset meant that its generalization ability was questionable.

METHODOLOGY:

In our human-based computation approach, the most difficult step, namely feature, genus, or species identification, is outsourced via a frontend website to human volunteers, who are given a tutorial. A backend algorithm, called dynamic hierarchical identification (DHI), uses unsupervised, supervised, and dynamic learning to accelerate microfossil identification. The unsupervised learning clusters specimens so that volunteers need not identify every specimen during supervised learning. With agglomerative hierarchical clustering, specimens are organized into a tree by visual similarity. Direct identifications of specimens use the tree to propagate information, i.e., indirect identifications and confidence levels, to visually-similar specimens. Human inputs and computation outputs are also in a feedback loop, providing dynamic learning. In particular, indirectly-identified specimens are prioritized for human analysis by computing the impact of their direct identifications on total confidence level.

RESULTS:

The DHI algorithm proved effective at accelerating microfossil identification. Using a dataset of foram tests that were directly identified by an expert, we evaluated correct and incorrect genus and species rates versus "time", represented by the number of specimens directly identified. In Monte Carlo trials, we randomly varied the sequence of direct identifications and used the k-nearest neighbour (kNN) method to obtain indirect identifications. With this approach, correct and incorrect rates versus time showed substantial dependence on sequence. At each moment, we took the highest correct and lowest incorrect rates, across trials, as the best benchmark results. The lowest correct and highest incorrect rates represented the worst benchmark results. The DHI algorithm achieved comparable rates to the best benchmark results, much better than the, equally likely, worst benchmark results.

Conference 9405:
Image Processing: Machine Vision Applications VIII

9405-3, Session 1

Deep convolutional neural network (CNN) for landmark recognition

Lin Sun, Cong Zhao, Chang Yuan, Lenovo (Hong Kong) Ltd. (Hong Kong, China)

Landmark recognition is always the pain point for the customers whatever during the tourism or photo album organization since the Global Positioning System (GPS) can not be helpful for this precise location service. Landmark recognition not only can provide the customers more convenient trips experience but more flexible album management. In this paper we want to introduce our recent work on recognizing landmark automatically without the help of computer vision algorithm and deep learning techniques. Deep learning has become the dominant in machine learning and computer vision areas. Many challenging task has achieved the remarkable performance using deep learning techniques. The most famous one is: when the deep learning is applied in the IMAGENET challenge, the performance is boosted a lot compared with the traditional computer vision algorithms. We will also apply the deep learning algorithm, particularly the popular and effective Convolutional Neural Network (CNN) in this task. We build a 72 landmark database of China, containing 12817 images in total. We randomly extract 10 images from each landmark category as the testing images and the remaining as training. We build an 8-layer CNN with 5 convolutional layer and 3 fully connected layer to handle this problem. Since this complex neural network can not be trained sufficiently just using our landmark data at hand, we use the imagenet database with 1281167 of 1000 categories to obtain the initial parameters for the CNN and fine-tune the whole network using our collected landmark data. During the testing we find that the first convolutional layer and final fully connected layer have a great affect on the final performance. Therefore, we transfer the parameters trained from the imagenet data except the first convolutional layer and final fully connected layer. We fine-tune the CNN to adapt the new parameters from the first layer and fully connected layer to the new landmark data. During the fine-tuning we also investigate the affect of the number of feature map and the kernel size to the final performance. We find that the small number of feature map and small kernel size is better for this task since it will reduce the over fitting problem. The whole accuracy on the newly collected landmark dataset can achieve 92.1%. In this paper we also propose some methods to reduce the false alarm, such as increasing more similar data or adjust the recognizing threshold.

9405-4, Session 1

Monitoring Arctic landscape variation by pole and kite mounted cameras

Rusen Oktem, Univ. of California, Berkeley (United States); Baptiste Dafflon, John E. Peterson, Susan S. Hubbard, Lawrence Berkeley National Lab. (United States)

Optic surveillance is an important part of monitoring environmental changes in various ecological settings. Although remote sensing provides extensive data, its resolution is yet not sufficient for scientific research focusing on small spatial scale landscape changes. We are interested in exploiting high resolution image data to observe and investigate the landscape variations at a small spatial scale arctic corridor in Barrow, AK, as part of the DOE Next-Generation Ecosystem Experiments (NGEE-Arctic). A 500x40 m corridor at this site is occasionally imaged by a low-altitude kite mounted consumer grade (RGB) camera and a 35 m transect is continously imaged by two separate - one capturing in NIR and the other capturing in visible range - pole mounted consumer grade stationary cameras. Surface and subsurface features along this 35 m transect are also sampled by electrical resistivity tomography (ERT), temperature loggers and water content reflectometers .

Images acquired 50 m above the ground by the kite-mounted camera are processed by a proprietary software to generate a 2D orthomosaic image and a DEM (Digital Elevation Map) of the terrain, at less than 5 cm and 10 cm resolutions, respectively. A segmentation algorithm working with HSV and HSL color planes and at multiple resolution levels is developed and implemented for identifying different terrain features. A basic K-means clustering is applied in the lowest resolution scale to generate initial clusters. The algorithm uses an expectation maximization approach in the upper scales, to assign clusters into expected features. The algorithm is found succesful in identifying areas of different types of vegetation and dry regions, and will be combined with DEM to quantify statistics belonging to (naturally formed) polygonal regions of the monitored terrain.

The two stationary cameras are mounted on the same pole, with slightly differing field of views. Images from the two are required to be registered through a pixel position mapping, so that data corresponding to the areas of interest from both can be combined together. Use of vegetation index is one way of monitoring the landscape variation after the melting of the snow, however change of image capturing parameters due to change in environmental light direction and light intensity calls for a compensation process. This is achieved by use of reference objects, where the compensation transform parameters are optimized in the least squares sense with the objective function defined over the intensity response of the reference objects. A polynomial model for the RGB image, and a nonlinear exponential function model for the NIR image are used for the intensity compensation.

Monitoring arctic landscape variation and soil characteristics helps to understand the arctic ecosystem feedbacks to the climate. The results of this study helps to achieve this purpose by investigating new measurement techniques in a small spatial scale test environment through ground based and low-altitude kite based high resolution time lapse images.

9405-5, Session 1

Hyperspectral imaging using a color camera and its application for pathogen detection

Seung-Chul Yoon, Tae-Sung Shin, Gerald W. Heitschmidt, Kurt C. Lawrence, Bosoon Park, Gary Gamble, Agricultural Research Service (United States)

Rapid detection and identification of pathogenic bacteria in complex food matrices, such as meat products, are important for both the food industry and regulatory agencies. The public demand for safe and high-quality foods has increased the need for research to develop rapid methods for accurate and reliable detection of foodborne pathogens, e.g. Shiga-toxin producing Escherichia coli (STEC), Campylobacter, Salmonella, and Listeria. Although commercially-viable rapid methods for pathogen detection have been around for many years, traditional culture-based direct plating methods are still the “gold standard” for presumptive-positive pathogen screening in many microbiology laboratories, where agar media are routinely used for isolation, enumeration, and detection of pathogenic bacteria. In practice, highly skilled technicians visually screen and manually select presumptive-positive colonies by trial and error for microscopic, biochemical, serological and molecular confirmation tests whose results may or may not be obtained rapidly. For this reason, the culture methods are labor intensive and prone to human subjective errors. Another challenge with direct plating is that competitive microflora often grow together with target microorganisms on agar media and can appear phenotypically similar.

Recently, researchers at the Agricultural Research Service (ARS) of the U.S. Department of Agriculture (USDA) have developed visible and near-infrared hyperspectral image classification algorithms in an attempt to reduce human error and increase the screening throughput for detection and identification of pathogenic colonies on solid agar media, such as Campylobacter, Salmonella and STEC. However, transferring the new hyperspectral imaging technique to analysts of a regulatory agency, such as USDA Food Safety and Inspection Service performing high-throughput routine testing of food samples, is expensive and time-consuming until the effects and effectiveness of the technique become proven in real conditions. Therefore, there was an immediate need for research to develop a cost-effective imaging solution using a regular digital color camera and the developed hyperspectral image processing algorithm, for which a hyperspectral reflectance reconstruction technique using color images can

be a viable solution.

This paper reports development of the hyperspectral reflectance reconstruction technique based on polynomial multiple regression from either RGB color bands or RGB and near-infrared bands. The target application was to detect and classify "Big Six" non-O157 STEC serogroups that are responsible for approximately 113,000 illnesses and 300 hospitalizations annually in the United States. A study was conducted to compare the performance of the spectral recovery technique between RGB color data obtained by a DSLR camera (Nikon D700) and composite color data obtained by two hyperspectral imaging spectrometers (Specim Imspector, V10E and V10M). The optimal polynomial order was studied in terms of R squared value. The preliminary data analysis found that data standardization was effective for numerical stability especially with higher order polynomial regressions. Adding a near-infrared band to a regression model improved the reconstruction power in the spectral range between 700 and 1000 nm, compared to a regression model using only three RGB color channels. The expected outcome of the research will be an imaging system using a DSLR camera and a diffused light illuminator for minimizing glares and glints.

9405-6, Session 2

Fast face recognition by using an inverted index

Christian Herrmann, Jürgen Beyerer, Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (Germany)

Searching for faces in large video datasets is becoming a topic of increasing practical interest. While face recognition appears to be solved for smaller and controlled scenarios, it remains a challenge for uncontrolled large scale applications like searching for persons in surveillance footage or internet videos. For example, in the context of forensic analysis of surveillance footage, a typical challenge is to find all appearances of a specific person, probably a criminal, in the given data.

Technically on the one hand, current productive systems focus on the best shot approach where only one representative frame from a given face track is selected by some heuristic and the remaining frames are discarded. This approach sacrifices recognition performance to achieve high recognition speed by ignoring the available information in the discarded frames.

On the other hand, systems achieving state-of-the-art recognition performance, like DeepFace which was recently published by Facebook, do not mind a slow recognition speed and use pairwise frame comparison approaches for video data which makes them impractical for large scale applications.

We suggest a set of measures to address the problem of achieving a high recognition speed combined with a decent recognition performance. First, extending the extracted local image features with the feature location and head pose allows collecting all features from one face track together in a single feature set. The extension enforces that only features describing the same face region and head pose can be successfully compared.

Secondly, the inverted index approach, which became popular in the area of image retrieval, is applied to that feature set. A face track is thus described by a set of indexed visual words and for each word a reference to this track is stored in the database index. Searching the database for a given person represented by a face track requires only an index-lookup for the respective visual words. This is very fast, because the index data structure of the dataset can be pre-computed for lookups.

Through these measures, we collect the information from all frames of a face track which allows better recognition performance than best shot approaches. In addition, the inverted index makes the recognition speed independent of the database size which allows constantly high recognition speeds.

Evaluation on a dataset of several thousand videos shows the validity of the proposed approach. In comparison to the best-shot approach, it shows comparable recognition speed and increases the recognition performance towards the state-of-the-art systems.

9405-7, Session 2

Advanced colour processing for mobile devices

Eugen Gillich, Helene Dörksen, Volker Lohweg, Ostwestfalen-Lippe Univ. of Applied Sciences (Germany)

Motivation

Recent applications of mobile devices (smartphones, tablets, etc.) are reaching the era to be used professionally for image processing tasks (e.g. m-health applications [1] or banknote authentication [2]). This is in stark contrast to the fact that mobile devices were not designed for such applications, especially in terms of image processing requirements like stability and robustness [3].

In the framework of our contribution, we concentrate on color processing problems of mobile devices. A simple example that represents the problem are inhomogeneous appearances of images under differently colored illumination of various environments. The problem might arise from diverse factors, e.g. failure of the white balance of the device, critical pixel size of the sensor for reddish colors, or low sensor dynamic range [4-5].

In our paper we propose a technique that contributes to the solution of the presented color processing problem. We demonstrate the performance of the approach by an example of a pattern recognition task.

Methods and Results

Based on the image processing method from [3], we present an extension of the approach for handling complex tasks of generating low-noise and sharp images without filtering. For an application of the approach regarding pattern recognition tasks, we show that a combination of our method together with an optimization of Machine Learning fundamentals leads to significantly more stable results and higher performance.

Color Space Image Pre-Processing: Our method is based on the fact that we analyze spectral and saturation distributions of color channels. Furthermore, the RGB space is transformed into a more convenient space, a particular HIS space (Hue, Saturation, Intensity)—named de-noised HIS-model (dHSI-model) [3], defined below. As we are interested in grayscale images, we generate the grayscale by a control procedure that takes into account the color channels and their noise behavior as well as chromatic aberration of (simple) optics of a mobile device [4]. This results in an adaptive color mixing model under the constraint of noise reduction. In the conclusion we receive a white balance under the above mentioned constraint. A noise reduced grayscale image for image post-processing is generated.

Feature Adaption and Classifier Refinement: Color processing for mobile devices for pattern recognition tasks might be hindered due to the spread of the measurements in the feature space [3], caused by the above-mentioned effects in color pre-processing. The spread of the measurements could lead to the false classification. Since it is not possible to construct a training dataset having information about the complete variety of color and light sources, the probability that an object is classified correct or wrong will be equal. For improvement of classification, the sensitivity of individual features is statistically analyzed. Features with lower sensitivity are adapted for classification. For classification technique we are interested in a simple classifier with a low number of parameters. The simplicity of the classifier is important here, since the dataset for the training is incomplete. Due to the incompleteness, a classifier with a higher number of parameters might tend to over-fit. To overcome the failures, we apply the ComRef method [6] for construction of a simple and trustful classifier.

References:

[1] Perera, C.; Chakrabarti, R.: The utility of mHealth in Medical Imaging. In: Journal MTM 2(3), 2013.

[2] Lohweg, V.; Dörksen, H.; Hoffmann, J. L.; Hildebrand, R.; Gillich, E.; Schaede, J.; Hofmann, J.: Banknote authentication with mobile devices. In: Media Watermarking, Security, and Forensics 2013 (03-07.02.2013) IS&T/ SPIE Electronic Imaging 2013, San Francisco, USA, Feb 2013.

[3] Gillich, E.; Dörksen, H.; Lohweg, V.: Generation of robust optical paths – Color Processing for Mobile Devices. In: Optical Document Security - The Conference on Optical Security and Counterfeit Detection IV, San Francisco, CA, USA Jan 2014.

[4] Schöberl, M., Brückner, A., Foessel, S., and Kaup, A., "Photometric limits for digital camera systems: SPIE," Journal of Electronic Imaging 21(2), 020501–1–020501–3 (2012).

[5] Goodman, J. W., [Introduction to Fourier optics], Roberts & Co. Publishers, Englewood and Colo, 3 ed. (2005).

[6] Dörksen, H.; Lohweg, V.: Combinatorial Refinement of Feature Weighting for Linear Classification. In: 19th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA 2014) Barcelona, Spain, Sep 2014.

9405-8, Session 2

A rotation invariant descriptor using Fourier coefficients for object detection

Quamrul H. Mallik, Abelardo Gonzalez, Pablo H. Lopez, Maja Sliskovic, SEW-EURODRIVE GmbH & Co. KG (Germany)

Rotational invariance in a feature descriptor is a desirable property for many computer vision applications. Examples include the general areas of object detection, texture or pattern recognition and parts-based object detection. Specifically for mobile applications such as hand-held devices, or industrial assistance systems; a feature descriptor that is intrinsically rotation invariant can be quite useful. At the same time, the size of the descriptor plays a significant role for embedded real-time applications. A successful descriptor for the stated application areas, thus, has to find a balance among rotation invariance, size and discriminative power.

In presented work, we study the question of achieving rotational invariance by using formal Mathematical Analysis. We concentrate on developing a descriptor that may be directly used in applicative areas of computer vision; such as human detection or face detection on a mobile platform; where the orientation of the imager cannot be guaranteed.

Historically, analytical rotational invariance has proven to be difficult to achieve in a reasonably sized descriptor with acceptable discriminative power. One standard approach is to normalize orientation to the dominant orientation of a detected keypoint. Intuitively, the accuracy of such an approach greatly depends on correctly assigning an invariant orientation to the keypoints. Furthermore, in applications where dense-scanning of the whole image becomes necessary (i.e. a scanning-window), a keypoint-based approach may not be the best choice. Other methods such as structured learning of artificially rotated samples of the object may not be practical enough due to prohibitively high memory requirements and complexity of the following non-linear classifier. Conversely, Analytical rotational invariance is often achieved at a cost of reduction in discriminative power and increasing size of the final descriptor. Previous approaches utilizing the Fourier Transform and/or Radon Transform illustrate these issues.

In presented work, we develop descriptors that guarantee invariance to arbitrary rotations to the image by using Formal Analysis. By an over-complete description of the two dimensional gradient field in the Fourier domain, we achieve performance that is superior or on-par with the state-of-the-art in both (a) rotation invariance and (b) gradient-based shape description. We base our descriptor on the well-known SIFT/HOG descriptor and its Fourier derivatives. Unlike previous work, our descriptor is reasonably sized (approximately the same size as conventional HOG), is over-complete; and achieves globally-true rotational invariance by calculating a single spatially-centered descriptor. Our experiments demonstrate that these factors are crucial in order to achieve competitive discriminative performance.

Through a systematic study of the effects of changing parameters such as convolution kernel sizes, sampling radii and sampling frequency we develop an optimal descriptor. This optimal descriptor is combined with a non-linear classifier (AdaBoost and SVM) and post-processing steps such as Non-Maximal Suppression to create a complete detection system. The performance of the classifier is experimentally validated using well established datasets in the areas of face detection and human detection. Using publicly available datasets for human detection and face detection, we achieve detection rates of 90% and 99%; at 1E-4 FPPW and 1E-2 FPPW respectively. These preliminary results indicate the superiority of our approach. Finally, we present an optimized implementation using vector processing to implement convolutional routines on an embedded arm processor; enabling mobile applications.

9405-9, Session 2

Image calibration and registration in cone-beam computed tomogram for measuring the accuracy of computer-aided implant surgery

Walter Lam, The Univ. of Hong Kong (Hong Kong, China); Henry Y. T. Ngan, Hong Kong Baptist Univ. (Hong Kong, China); Peter Wat, Henry Luk, Tazuko Goto, Edmond Pow, The Univ. of Hong Kong (Hong Kong, China)

This paper presents a robust image calibration and registration method for the CBCT. To link the virtual (imaging) to the reality (surface), a guidance with RF markers (usually multiple metal/ceramic balls) is fitted during a tomographic image acquisition. These RF markers are regarded as "floating" above the surface and registered at a 3D position guidance in the imaging. Therefore, this guidance link up the surface (clinical reality) and the internal structures visualized in the imaging if the guidance can fit the reproducibly onto the surface such as any hard tissues like bone and teeth. By the reference of a cube's corner as a RF marker, it defines mathematical convertible Carterian (x, y, z)-coordinates between clinical reality and imaging. This corner point where three surfaces met defines the origin (o) and a corner line where any two surfaces met defines the x-, y- and z-axis of the coordinates. In this way, every vowel (composition unit) of the imaging and its reality counterparts will be assigned a corresponding x-, y- or z- coordinates. The physical cubic corner (real physical domain) and the imaging cubic corner (computerized virtual domain) are then matched, calibrated and registered robustly. The imaging is a coordinated 3D virtual model of the human body showing both the external (surface) and internal structures. Herein, surgical procedures can be planned virtually and the virtual surgical plan might be transferred to the (radiographic) guidance or a new surgical guidance might be made straight away w.r.t. the (x, y, z)-coordinates. For example, in an oral implant placement, guiding sleeve with a depth control can be determined on the guidance with or without the help of a robot according to the virtual planning. The oral implant can then be placed in the designed position by a surgeon in the patient's mouth. Moreover, the cube's corner can also be easily calibrated at a chair-side (in a comparison with multiple balls) to allow a real time surgical navigation with the aid of the imaging. The navigation screen will further inform the surgeon whether the direction of bone drilling (to create a site for the implant placement) is following to the right planning with a reference to the imaging. This allows a certain degree of freedom during a surgery comparing with the traditional guidance approach. Accuracies of various guidance methods, including the use of guides (static) and the navigation system (dynamic), can be compared to a gold standard: the planning in the imaging w.r.t. their Carterian (x, y, z)-coordinates. A case who requested replacement of a missing front tooth by an oral implant using static guide system was illustrated. The x-, y- and z- coordinates of the implant apex w.r.t. the RF marker were measured in the implant planning (x=0.27mm, y=-21.90mm, z=-6.87mm) as well as the actual implant placed in the after-surgery's CBCT (x=0.39mm, y=-21.63 mm, z=-7.06mm). Implant apex was measured since it was the deepest implant part placed in a human body and most deviated from the planning and causing damage.

9405-10, Session 3

An video saliency detection method based on spacial and motion information

Kang Xue, Xiying Wang, Weiming Li, Gengyu Ma, Haitao Wang, Samsung Advanced Institute of Technology (China)

The human visual system has a remarkable ability to quickly grasp salient regions in static and dynamic scenes without training. Based on such ability, human can understand the scenes easily. Saliency detection method may reveal the human's attention mechanism, as well as model their fixation selection behavior. However, computationally identifying such salient regions that match the human's attention is a very challenging task. To obtain automatic, effective and accurate saliency extraction methods, many researchers pay lots of attention in this area in the last decade. In our paper, we propose a novel bottom-up video saliency extraction method, which includes two parts: first is static saliency detection and second is dynamic saliency detection. In static saliency detection step, we consider the static saliency detection as a classification problem: a scene can be divided to two classes – saliency class and non-saliency class. To solve such problem, we introduce a Canonical Correlation Analysis (CCA)-based classification method. Unlike some current methods using super pixel segmentation, we propose a multi-cue strategy (using different features, such as Lab color, Luv color and coordinate) to describe each pixel with lower time consuming. Firstly, a frequency analysis, which permits the full use of global information, is used to initialize the scene to coarse saliency region and non-saliency region. Then we samples from the saliency and non-saliency region as the training samples to train the CCA projection matrix. Finally, the scene can be classified as saliency and non-saliency region. In the dynamic detection step, we combine motion feature with the spatial feature to represent object's dynamic saliency in temporal domain. The system is initialized by extracting static saliency in the 1st frame in a video sequence. Then, background is modeled using non-saliency region and all foregrounds can be detected in a new frame. Considering each foreground's motion contrast, we combine such motion information with static saliency information together to determine each foreground's dynamic saliency. Note that because human eyes are more sensitive to moving object, we give higher bias on motion information in this fusion. We test our saliency detection method in two ways. First, to objectively evaluate our new static saliency detection method, we compare our results with other state-of-the-art methods for object or interest segmentation. We chose the large accurate dataset (MSRA 1000) as the testing data. In such comparison, our method outperforms other methods in both accuracy and effectiveness, and the time-consuming is 0.35 second per frame on Matlab. Second, we test our dynamic saliency detection method in some video sequence. The results illustrate that our algorithm can represent the difference of human's visual attention when they face a scene with and without motion objects. In our paper, we propose an effective static and dynamic saliency detection method by introducing different kinds of feature extracted from frequency, spatial and temporal domain. Our method shows its high performance in the testing on both images and video.

9405-11, Session 3

Depth-Map and Albedo Estimation with Superior Information-Theoretic Performance

Adam P. Harrison, Dileepan Joseph, Univ. of Alberta (Canada)

Motivation

Lambertian photometric stereo, where surface normals and albedos are estimated from sets of images, is a seminal computer vision method. Nonetheless, using a depth map in the image formation model, instead of surface normals, is preferred from an information-theoretic perspective. It reduces parameters by a third, thereby decreasing susceptibility to over fitting. The Akaike information criterion (AIC) quantifies this trade-off between goodness of fit and over fitting. Obtaining superior AIC values requires a robust and efficient means to produce maximum-likelihood (ML) estimates of the depth map and albedo. This paper presents such a method.

State of the Art

This work differs from approaches in the literature, which either do not attempt to produce an ML estimate, offer only an approximate ML estimate, and/or decouple depth-map and albedo estimation. Popular methods use two linear steps. They first execute photometric stereo and then estimate the depth map. However, these linear approaches transform noise between steps, making it difficult to realize an ML estimate. For this reason, popular methods can struggle with real-world image noise. Recently, we published a linear ML method that accounts for noise transformations. It exhibits significantly greater ability to handle image noise. While effective, the method relies on approximations of noise distributions that limit the accuracy of the ML depth-map estimate. As well, like all linear methods, the depth map and albedo are not estimated jointly, further limiting accuracy.

Methodology

Accuracy may be improved by abandoning the linear two-step approach in favour of a nonlinear one-step approach that operates directly on images. Yet, practical inconveniences stemming from a nonlinear generative model have so far stalled this direct approach. To overcome these obstacles, this paper kick starts the nonlinear approach with the linear one. Our linear ML method provides a robust initial solution, which is then used as input to a nonlinear refinement process using direct image observations. A nonlinear separable least-squares reformulation reduces the size of the refinement problem by half. Where possible, sub-problems are tackled independently for each pixel, further reducing computational demands. These innovative approaches enable an efficient nonlinear estimation method.

Results

Efficacy of the direct refinement method is demonstrated with comprehensive experiments that highlight the benefits of a more parsimonious model. These include experiments that demonstrate the visual inaccuracy of over fitting under noisy conditions, validating our information-theoretic approach. One manifestation of this inaccuracy is poor image reconstruction under light directions not used in the estimation step. With over fitting, depth maps and albedos also prove more sensitive to image noise. These qualitative results are supported by an extensive statistical analysis using multiple instances of images corrupted by increasing levels of noise. Refined depth maps and albedos produce superior AIC metrics than with photometric stereo. Moreover, improvement in reconstruction error compared to the linear ML approach is demonstrated. These results indicate that the direct refinement method offers a practical and effective means to generate accurate ML depth-map and albedo estimates that are superior in an information-theoretic sense.

9405-12, Session 3

Shot boundary detection and label propagation for spatio-temporal video segmentation

Sankaranaryanan Piramanayagam, Eli Saber, Nathan D. Cahill, David W. Messinger, Rochester Institute of Technology (United States)

This paper proposes a two-stage algorithm for spatio-temporal video segmentation. In the first stage, shot boundaries are detected in the video by comparing dissimilarity between 2-D segmentations of each frame. In the second stage, the 2-D segments are propagated across frames within individual shots.

Videos are essentially a collection of scenes that are subdivided into shots, or temporally unbroken video events. Each shot in turn is comprised of many spatio-temporal regions. Detecting shots and segmenting videos into spatio-temporal regions are important pre-processing steps in many video processing applications such as tracking, indexing, summarization, and retrieval. As opposed to previous techniques for shot detection that are block-based, the shot-detection stage in our algorithm uses natural 2-D segmentations of each frame not only to detect shot boundaries, but to provide input for the second stage of propagating 2-D segments within shots to identify natural spatio-temporal regions in the video.

The method developed for our shot detection stage consists of three major steps. In the first step, each individual video frame is segmented into homogeneous regions by a gradient based region growing (GSEG) algorithm. In the second step, dissimilarity between two adjacent frames is computed. Here, we overlay two frames segmentation map and assume both frames have the superimposed segmentation. For both frames, color

histograms of the individual segmented regions are found. Chi-square distance between first frame histograms and second frame histograms are computed. Then, average of these distances weighted by number of pixels each region occupies is calculated. The weighted average represents the dissimilarity between the two adjacent frames. In the final step, the entire sequence of pairwise frame dissimilarities is analyzed to detect cut or gradual transition. Once shot boundaries are found, the second stage of our algorithm proceeds by propagating the 2-D region labels from individual frames within each shot to generate spatio-temporal segments. The label propagation step is based on a multivariate analysis of variance (MANOVA).

We tested our segmentation based shot detection method on the TRECVID 2007 video dataset and compared it with block-based method. For the block-based shot detection method, we divided each frame into 32/16/12/1 blocks. We found that using only two adjacent frames yielded unsatisfactory results for both our approach and the block-based approach for shot detection. Hence, we adopted a moving window approach to find cuts and gradual transitions. In the moving window approach, each frame is compared against multiple preceding and succeeding frames. To quantitatively compare algorithms we computed precision, recall and F-measures for transitions. Preliminary cut detection results on the TRECVID 2007 dataset indicate that our algorithm (Recall: .9329 Precision: .9581) has comparable shot-detection results to the best of the block-based approach (Recall: .9278 Precision: .9565). We plan to investigate the outcome with other state-of-the-art segmentation algorithms in literature.

In addition to the quantitative validation of our shot-detection stage, we show qualitative results of the resulting spatio-temporal video segmentation after our second stage of label propagation is performed.

9405-26, Session PTues

Context-based handover of persons in crowd and riot scenarios

Jürgen Metzler, Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (Germany)

In order to control riots in crowds, it is helpful to get the ringleader under control. A great support to achieve this task is the capability of observing them automatically by using one or several video cameras. Especially, the automatic tracking of individual persons in a video sensor that is monitoring a crowd is helpful for security personell in order to get ringleader and aggressors under control. Also, machine vision can be applied for the conservation of evidence so that legal proceedings can be initiated. In both cases, it is important that automatically tracked persons do not get lost in the camera system.

In this paper, we propose a context-based approach for handover of persons between different camera's fields of view. It uses an appearance-based re-identification method for matching persons as well as groups between images of various camera perspectives. The approach can be applied for overlapping as well as non-overlapping fields of view, so that a robust capturing of global trajectories of individual persons in camera networks is possible.

At the first step, a set of persons is detected in a single image that is very well distinguishable from the background and from other persons in the scene, and hence, can be tracked robustly. In many cases these are persons who are wearing loud colored clothes that can be easily detected by color analysis. Then, a covariance descriptor is calculated for each image region of a detected person representing the variances and covariances of gradient, color and spatial information. These descriptors are introduced by Porikli et al. and are also used in our application for tracking individual persons in crowds. A previous evaluation of several tracking methods resulted that the tracking approach based on covariance descriptors are the most appropriate one for crowd and riot scenarios. After the descriptors of these person-related image regions are calculated, they are matched with descriptors extracted from other images. If there are overlapping camera's fields of view, the number of comparisons can be limited. Also, available calibration information of the cameras can reduce them further. If the fields of view are non-overlapping, we only use geographical information to reduce them.

Our main objective is the robust handover of persons between different camera's fields of view in crowded situations. On the one hand, there are generally a lot of persons, and so a high probability to detect image regions that can be matched easily. However, on the other hand, there are a lot of persons who looks similar to each other in such situations which makes the matching more difficult. In order to improve the matching step, we only re-identify persons wearing clothes that are distinguishable from the background and other persons in the crowd. All other persons are handed over by considering contextual information in a second step.

The approach is evaluated on a data set which we collected during a Crowd and Riot Control (CRC) training of the German armed forces. Four cameras were installed in 25 meter high on an observation platform and recorded data from nadir view. The data set consists of three different levels of escalation. First the crowd started with a peaceful demonstration. Later there were violent protests, and third, escalations of the riot where offenders bumped into the chain of guards.

The evaluation of our approach shows a significant improvement of the handover of persons if context information is used. Furthermore, it shows that it can be also improve other re-identification approaches. In summary, the result is a robust method for handover of persons between different camera's fields of view in crowd and riot scenarios which can improve the situational awareness and the conservation of evidence in such scenarios.

9405-27, Session PTues

3D motion artifact compenstation in CT image with depth camera

Young Jun Ko, Jongduk Baek, Hyunjung Shim, Yonsei Univ. (Korea, Republic of)

In medical imaging systems, achieving high image quality is very important for detection of early stage cancers. To improve the detectability, it is desirable to have high resolution, less noisy, and artifacts free images. However, in daily clinical scans, patient motion is not avoidable and may introduce blurs, streaks or discontinuities in reconstructed images.

Motion artifacts are common problems in computed tomography (CT) systems and several correction methods have been developed. Existing approaches are roughly divided into two categories; either reducing a scan time or compensating the corrupted measurements by estimating the human motion. Although their methods are useful to reduce motion artifacts, both categories have inherent limitations. In the first category, a short scan time yields a low signal-to-noise ratio because the amounts of incoming X-ray photons are proportional to the scan time. On the other hand, the performance of second category is bounded by the accuracy of motion estimation because they estimate the motion using a generic human motion model. While our method belongs to the second category, we alleviate its weakness by extracting accurate patient motions. In this paper, we propose a new motion-free CT system using depth cameras. Because the depth camera is effective for motion capture, motion estimation can be performed effectively, and thus minimize the motion artifacts in the reconstructed CT images.

Our CT system includes multiple depth cameras to capture the 3D motion of patient body. Depth cameras are positioned around the patient and record the depth video of patient in approximately 30 fps with few millions of pixels. To extract the motion from the depth video, we compare two consecutive depth frames. By assuming rigid motion, we can parameterize the 3D motion by rotation and translation. We evaluate the score of motion parameters based on the number of points, RMSE between two frames and regularization constraints. By repeating this for all possible rotation and translation, we optimize the most probable motion parameters corresponding to the best score.

To compensate motion artifacts, we modify the raw measurements (referred to as sinogram) from CT scans using the motion data and then reconstruct the CT images from the modified sinogram. For example, in the presence of the translation motion, a subset of sinogram data is shifted by several pixels. Because we can convert the distance unit into the sinogram pixel unit upon the CT specification, we predict the pixel shift in sinogram from the motion. Consequently, we restore the sinogram by shifting pixels back.

We expect to achieve the performance improvement because our framework compensates the motion in system level by modifying sinogram data while previous methods perform image deconvolution on reconstructed CT images. Moreover, capturing patient motion over the entire surface overcomes the accuracy limitation of motion estimation by generic human motion model. Based on the simulation, we succeed to identify rigid motion parameters in arbitrary 3D point sets and compensate rigid motion artifacts in simulated CT scans. In the final manuscript, we extend our method to non-rigid motion artifacts correction.

9405-29, Session PTues

Robust detection for object under occlusions

Yong Li, Chunxiao Fan, Yue Ming, Beijing Univ. of Posts and Telecommunications (China)

Motivation for this work:

Past years witnessed a great progress in the field of object detection. Most existing methods are based on HOG and DPM. However, the detection of objects under partial or heavy occlusion remains to be a challenge. To address this challenge, some current algorithms tend to treat occlusions as an unstructured source of noise and explicitly establish models based on specific datasets, e.g., synthetically occluded training data. Other algorithms try to combine detection with segmentation, but can only be applied to a certain specific types of detection models or scenarios. The goal of this work is to develop detection techniques that are not limited to particular scenarios and datasets. To this end, we propose detecting objects under occlusions through combining the HOG descriptor and E2LSH.

Insight and Methods:

Our insight comes from the simple fact that human vision can easily detect objects regardless of the occluders appearance. In other words, for human vision there is no strong relationship between occluder and occludee. In an ideal situation, we expect that the object under occlusions can still be detected as if the occlusion did not exist.

This work adopts the HOG descriptor. Since an object's HOG feature under occlusion contains both the object part and the occluder part, we hope to reserve the object part and discard the occlude part. For the object part, E2LSH is used to search similar elements in dataset. If the number of similar elements (denoted by x) is greater than a threshold, then there is a potential object in this window with the confidence of x.

The similarity is calculated based on Euclidean distance. If there are two HOG features (denoted by A and B, respectively), the distance between them is d. Then, the distance between a part of A and a part of B should be less than d. Therefore, the similarity between two parts should be greater than the similarity between A and B.

This suggests that if an area is recognized as an object, its part feature should be detected with higher confidence.

Experimental results:

We compared our algorithm with DPM. DPM was trained on the dataset VOC 2012 and then tested on some images from Internet. For this setup, most objects under heavy occlusion were detected. Nevertheless, when trained on INRIA, the detection performance degraded rapidly under even small partial occlusion. This phenomenal difference originates from the fact that the VOC 2012 itself encapsulates an abundance of occlusion information and patterns, and hence the detection algorithm trained with it still performs very well. The robustness of DPM against occlusions is affected by the dataset on which the training is conducted. To show the proposed method performs better than DPM, we trained both DPM and our method on INRIA, which contains less information about occlusions than the VOC 2012.

Dollar's results in 2012 show that most occluded parts lie either in the lower area or in the side of an object. Motivated by this, we remove the entries from the HOG feature corresponding to these areas and reserve a part feature of 217 dimensions. The '217' is an empirically determined number from a great deal of experiments.

When there was no occlusion, our approach had a lower detection precision than DPM. However, all pedestrians can be captured although the detected bounding boxes do not precisely enclose the pedestrians. When there was partial occlusion, DPM missed detecting pedestrians in some test images, while our approach was not affected by the partial occlusion. When heavy occlusion existed DPM missed most objects, but our approach was still unaffected.

9405-31, Session PTues

Human action classification using procrustes shape theory

Wanhyun Cho, Chonnam National Univ. (Korea, Republic of); Sangkyoon Kim, Soon-Young Park, Mokpo National Univ. (Korea, Republic of); Myungeun Lee, Seoul National Univ. (Korea, Republic of)

The key contribution of our work is to provide a novel approach for human activity recognition by applying Product manifold theory and Procrustes shape technique for 2D landmark shape sequences. The main idea is to compute the full Procrustes mean shape vector for a sample of configuration matrices of landmarks belongs to product manifold, and compute the partial Procrustes tangent space using a pole as the full Procrustes mean. To measure the distance between two different human actions represented as two volume images, we have shown that the geodesic distance between two volume images belongs to product manifold can be approximate by the Euclidean distance between two vectors in tangent product space corresponding to these volumes. And then, we are looking for a human action in which can minimize this distance. In order to implement these theories, we consider the mathematical properties of product manifolds and the content of Pocrustes shape analysis. First, we factorize each volume images belonging to training dataset as a time ordering sequence of images, and we extract pre-shape configuration vector of landmarks from each frames consisting a image sequence. Then, we have obtained a random sample of pre-shape configuration vectors from all videos stored in tanning database by using similar procedure, and we compute mean shape vectors for random sample of extracted shape vectors. In the second step, in order to recognize the query human action video, we derive a sequence of the pre-shape configuration vectors from given query video, and we project each shape vector on the tangent space with respect to the pole taking on a sequence of the mean shape vectors corresponding with a target video. We recognize a query video as target video that can minimize the distance between two sequence of the pre-shape vectors and the mean shape vectors. We assess the proposed method using Weizmann human action data sets. Experimental results reveal that the proposed method performs very well on these data sets.

9405-32, Session PTues

Sub-pixel estimation error over real-world data for correlation-based image registration

Pablo S. Morales Chavez, Ramakrishna Kakarala, Nanyang Technological Univ. (Singapore)

Image registration is a rather simple and commonly used method to describe the motion or geometric deformation occurred between two consecutive frames in an image sequence. Correlation, or window-matching, based image registration detects the differences in the images by comparing the patterns present in a search window centred at the potential matching pixels. One of the mayor drawbacks of the correlation based techniques is the integer-limited precision of the computed results. See for example [5], for an optical navigation application that requires sub-pixel accurate data. Methods for refining the pixel-accurate results are used as a post-processing step once two corresponding pixels have been selected.

In this work, we are interested in the evaluation of the accuracy of the sub-pixel generating methods. In particular, we evaluate the results of the

sub-pixel refining process of the optic flow detected using a correlation algorithm in close range imagery. The contribution of this paper is two fold. First, we design and provide a data set of engineered close range image sequences with available sub-pixel optic flow ground truth. The motion described in the sequences represent the two-dimensional motion of a rigid planar object with different textures. We use chequerboard patterns (binary and with multiple intensities), straight lines and single-intensity objects. For each of the patterns, we record a sequence with either a circular or linear motion or describing a square. There are no independent moving objects and all the pixels describe the same motion, starting and finalizing in the same position.

The motion sequences are generated using a Newport VP-25XL two-dimensional motion stage platform with a claimed bidirectional repeatability of 0.14 micrometers, see [7]. The optical flow sub-pixel ground truth is calculated using the data reported by the motion stage and the known parameters of the used optical system.

The usual test beds for evaluating the sub-pixel accuracy, in the context of close range imagery, are computer generated patterns describing a linear movement across the sequence. The public data sets presented in [1] and [4], contain image sequences suitable for the evaluation of optic flow algorithms when used in outdoor oriented applications. They are well known and are considered as valuable reference data. See [6] for a recent study involving the two data sets. However, this data is not adequate for our purposes, as it contains independent moving and non-rigid objects. We are interested into filling this lack of reference data with the presented data set.

And second, we use the presented data set to evaluate an extension of the approached introduced in [2]. In that paper, the authors suggested a method designed to cancel the bias induced by the normalized cross correlation when used as a cost function in a correlation based motion detection algorithm. For computing the sub-pixel measurements, the authors restricted their study to the parabola fitting method. We extend such approach in the sense that the sub-pixel measurements can also be computed using other commonly used methods, for example using the techniques analysed in [3]. We compare the performance of both sub- pixel estimation approaches and corroborate the results presented in [2] using a broader data set.

References

1. Baker, S., Scharstein, D., Lewis, J. P., Roth, S. and Black, M. J. and Szeliski, R. A Database and Evaluation Methodology for Optical Flow. In Proc. ICCV, p. 1–8, 2007.

2. Cheng, P. and Menq, C.-H. Cancelling bias induced by correlation coefficient interpolation for sub-pixel image registration. In Measurement Science and Technology, 24(3):321–334, 2013.

3. R. Fisher and D. Naidu. A comparison of algorithms for subpixel peak detection. In Image Technology, Advances in Image Processing, Multimedia and Machine Vision,Springer-Verlag, 1996, pp. 385–404

4. Geiger, A., Lenz, P. and Urtasun, R. Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite. In Proc. CVPR, p. 3354–3361, 2012.

5. HP Designjet Z6100-series Printers: Optical Media Advance Sensor. http://h10088.www1.hp.com/gap/Data/en/Z6100_optical.pdf, retrieved 2014 [online].

6. Vogel, C., Roth, S., and Schindler, K.. An Evaluation of Data Costs for Optical Flow. Pattern Recognition, p. 343–353, 2013.

7. VP-25XL, High Precision Compact Linear Stage. User Manual. http://assets.newport.com/webDocuments-EN/images/VP-25XL_User_Manual.pdf, retrieved 2014 [online].

9405-33, Session PTues

Understanding video transmission decisions in cloud base computer vision services

Rony Ferzli, Nijad Anabtawi, Arizona State Univ. (United States)

Cloud based computing has gained ground in recent years due to several advantages such as scalability, redundancy, abundance of computational power, and elimination of server maintenance and configuration. Due to the fact that most computer vision algorithms are computationally demanding, cloud based solutions are ideal for such scenarios and as such becoming the focus of recent research. Commercial cloud solution to be used in computer vision are emerging such as Microsoft Azure [1] or CloudCV [2].

In such scenarios, bandwidth consumption is highly critical, increased bandwidth will increase the cost, add complexity to the cloud platform adding the need to load balance, as well as adding additional computational power. To reduce the bandwidth consumption the following optimization can be done at the camera source: reduce resolution, reduce frame rate, and reduce quality by compressing at a higher bitrate or any combination of the above.

Reduction of cost andresources needed in the cloud through the adoption ofof the measures outlined above may, however, have negative consequence on the computer vision algorithms. For example, reducing the resolution may lead to a reduced recognition rate. To understand better the implications, this paper aims at conducting a study performing a sensitivity analysis to reach some conclusion about optimal transmission parameters that will reduce bandwidth usage while concurrently ensuring that the CV algorithms results are marginally affected. The approach to be used is described hereafter:

a) Pick a Compute Vision (CV) scenario

b) For each scenario, pick the raw input raw source video

c) Run experiment using the original raw video, note the results

d) Modify the source video by adding resizing, compression , frame rate deduction as well as combination of the three

e) Feed each video back to the CV system and note the results

f) Gather the data and draw conclusion and recommendations

This study will pick several known computer vision algorithms/scenarios important in numerous computer vision applications including activity recognition, automotive safety, and surveillance and perform the sensitivity analysis noted above, to name few:

a) Face detection and tracking: many techniques for face detection and tracking exist, the most widely used and adopted in this paper are CAMshift [3-4] and KLT algorithm [Ref].

b) Motion-Based Multiple Object Tracking: such as detecting moving cars or people. These are performed using foreground detector based on Gaussian mixture models (GMMs) [5]

Preliminary results are obtained using MATLAB, openCV , and Intel Media SDK. The Intel SDK contains a highly optimized H.264 encoder to be used to compress the sequences. MATLAB/OpenCV will be used to implement the computer vision algorithms. Video Sequence of size 640x360 is used and compressed at different bitrates ranging from 50 to 600 kbps. The compressed streams are then decoded and fed to the computer vision system to check for the results. Results of the motion based tracking system when a) the original uncompressed video b) the compressed video at 100 kbps are compared. It can be seen that compression artifacts affect heavily the decision , from the experiments conducted the algorithm will start producing false positive as we go below 200 kbps for the specified resolution. In the full paper more video sequences will be generated with different resolution/frame rate and compression ratios (as well as more scenarios) to understand the limitation and draw a set of recommendations that will be very useful.

References

[1] Microsoft Azure, http://azure.microsoft.com/en-us/

[2] CloudCV: Large-Scale Parallel Computer Vision on the Cloud. http://cloudcv.org/objdetect/

[3] G.R. Bradski “Real Time Face and Object Tracking as a Component of a Perceptual User Interface”, Proceedings of the 4th IEEE Workshop on Applications of Computer Vision, 1998.

[4] Viola, Paul A. and Jones, Michael J. “Rapid Object Detection using a Boosted Cascade of Simple Features”, IEEE CVPR, 2001.

[5] Zivkovic, Z., “Improved adaptive Gaussian mixture model for background subtraction,” Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., vol.2, no., pp.28,31 Vol.2, 23-26 Aug. 2004.

9405-34, Session PTues

An auto focus framework for computer vision systems

Rony Ferzli, Nijad Anabtawi, Arizona State Univ. (United States)

An out of focused captured images are blurry in nature, so if the level of blurriness can be predicted accurately then it can detect the images that are out of focus captured from the camera, triggering the drop of these images from the computer vision pipeline and at the same time providing a controlled feedback channel to adjust the camera focal length to the correct level. The framework communicates with the camera though its standard API interface using known media frameworks for a given operating system (such as Media Foundation Framework for Windows). Thus the proposed framework consists of the following

1) Camera out of focus detection

2) Control Feedback loop

3) Camera interface

At the heart of the out of focus detection is the blur metric, being able to accurately assess blur in an image is crucial. The blurriness of an image is a function of its spectral density. Narrow spectrum implies the image is more blurry. Thus the image blurriness can be measured by measuring the shape of its spectrum. Bivariate kurtosis can be used to measure the shape and shoulder of a two dimensional probability distribution. It is known that the low frequencies correspond to the slowly changing components of an image and high frequencies correspond to faster gray level changes in the image, which gives information about the finer details such as edges. When an image is in focus, the high frequency components are maximized to define the edges sharply. Thus kurtosis, which measures the width of the shoulder of the probability distribution, corresponding to the high frequencies, can be used to measure the sharpness (i.e the inverse of blurriness). The higher the kurtosis the blurrier is the image. Note that block based motion estimation and compensation is performed between frames to exclude any error in blur assessment due to motion.

Next the kurtosis will be calculated over a sliding window of 1/3 of the frame rate (for 30 fps video that corresponds to 10 frames). For each window, the median is calculated, if the median is above a threshold 'T' then the frame at the middle of the window is flagged as out of focus.

Once a frame is flagged as out of focus, based on the kurtosis value a proportional controller is used to feed the right signal to the camera interface to adjust the focal length accordingly.

Simulation results are conducted using the 'akiyo' CIF clip and simulating out of focus blur by running a Gaussian window of different sizes. Then the blurry sequence is fed to the proposed framework to get the kurtosis value and based on the controller feedback the image is sharpened (simulating the focal length adjustment by API). Preliminary results obtained are encouraging whereas in the full paper detailed analysis will be shown about the accuracy of the proposed framework and how this can be implemented efficiently in hardware.

9405-35, Session PTues

Innovative hyperspectral imaging (HSI) based techniques applied to end-of-life concrete drill core characterization for optimal dismantling and materials recovery

Giuseppe Bonifazi, Nicoletta Picone, Silvia Serranti, Univ. degli Studi di Roma La Sapienza (Italy)

1. Problem statement and motivation for the work

The possibility to develop an efficient recovery and reuse of concrete materials resulting from end-of-life (EOL) buildings, and/or civil constructions dismantling/recycling, represents one of the main targets in the Construction and Demolition Waste (C&DW) sector. EOL concrete materials recycling is strongly increasing in these last years, the main reasons being linked to: i) the decrease of steady supplies of good quality natural aggregates, ii) the need to secure ample supplies of concrete aggregates and, finally, iii) the environmental constraints more and more limiting the C&DW wastes disposal, especially in urban regions. The fulfilment of the previous mentioned goals can produce several benefits, that is: i) a reduction of new non-renewable resources exploitation, ii) a strong reduction of the costs linked to transport and energy production, iii) the possibility to utilize materials that otherwise should be lost (i.e. land filled), iv) land preservation in respect of future urban development and, finally, v) the reduction of the impact of new exploitation activities on the environment. The possibility to utilize efficient, reliable and low cost analytical tools able to perform detection/control actions finalized to assess: i) concrete characteristics, before demolition, and ii) physical chemical attributes of the resulting products (i.e. particles) after demolition and processing, represents one of key issue to develop innovative process/control actions/procedures inside the C&DW sector. C&DW recycled granulate products show, for their nature, larger variations in quality (e.g. particles characterized by different degree of mixing between cement, sand, gravel) compared to natural aggregate and limestone. The identification of a suitable on-line sensor technology for quality measurement and control of recycled streams along the entire chain, from demolition to "new" cement and/or mortar production, could dramatically help to detect DW resulting products fluctuations and to deal with them accordingly. Furthermore the possibility to implement on-line control strategies allows performing a certification of the products resulting from demolition.

Reflectance Spectroscopy (RS) across the visible, near and short infrared spectral region (400-2500 nm) has been recently proposed as a tool to assess the status of the concrete in situ . The fundamental vibrations of most building materials generate spectral information in the mid-infrared region (2500-14000 nm), and overtones and combination modes in the near and shortwave infrared region (900-2500 nm). Such an approach, even if quite powerful, only allows collecting information in a specific region of the investigated materials and/or product. The possibility to map the different materials constituting a concrete structure to dismantle and/or a system of particles resulting from comminution is almost impossible to reach adopting this approach. On the other hand the possibility to perform a topological assessment of the different materials constituting a concrete and/or concrete-milled-derived-products is what really need to design innovative recycling strategies in DW sector. The fulfilment of this goal can be reached adopting an innovative characterization strategy, based on HyperSpectral Imaging (HSI) working in shortwave infrared region (SWIR), able to realize a low-impact-real-time collection of the information concerning dismantled materials.

2. Methods

Hyperspectral images were acquired using the SISUChema XL™ Chemical Imaging Workstation (Specim, Finland), embedding an ImSpector™ N25E imaging spectrograph (Specim, Finland) working in the range from 1000 to 2500 nm, with a spectral sampling/pixel of 6.3 nm (active pixel 320 (spatial) x 240 (spectral) pixels), coupled with a MCT camera with pixel resolution of 14 bits. The device is controlled by a PC unit equipped with the ChemaDAQ data acquisition software (Specim, Finland). The images were acquired with a 31 mm lens and a field of view of 100 mm.

Spectral data analysis was performed using the PLS_Toolbox™ (Version 7.3, Eigenvector Research, Inc., Wenatchee, WA) under Matlab® environment (Version 7.5 (R2007b), The Mathworks, Inc., Natik, MA).

Analyses were carried out in three steps: i) spectral preprocessing, to eliminate or minimize variability of spectral signals unrelated to the property of interest (like those from multiplicative scatter effects or from base-line drifts); ii) Principal Component Analysis (PCA), to perform an exploratory data analysis; iii) Partial Least-Squares Discriminant Analysis (PLS-DA), to build the models and validate them. The built models were thus applied on concrete drill core samples, produced by portland and furnace slag cements, in order to perform their constituents classification.

3. Experimental results

The utilisation of an HSI based platform to demonstrate the potentialities of this approach was developed with reference to concrete drill cores coming

from a building to dismantle in the province of Groningen (Netherlands). In order to analyse a representative set of samples, drill cores were collected in different building locations (i.e. floor, room/corridor, wall/floor/façade/pillar, etc.). Analyses have been carried out in two steps, that is: i) identification/characterisation of the different materials (i.e. mortar and aggregates) constituting the drill core slices and ii) HSI based drill core slices mapping according to materials detected spectra. Results obtained by the adoption of the HSI approach showed as this technology can be successfully applied to analyse quality and characteristics of C&DW before dismantling and as it results as final product to re-utilise after demolition-milling-classification actions. The proposed technique and the related recognition logics, through the spectral signature detection of finite physical domains (i.e. concrete slice and/or particle) of different nature and composition, allows; i) to develop characterization procedures able to quantitatively assess end-of-life concrete compositional/textural characteristics and ii) to set up innovative sorting strategies to qualify the different materials constituting drill core samples.

4. Conclusions

The potentiality offered by an HSI based approach to perform a full characterisation of solid waste streams was described and analysed. After a general presentation on the main characteristics of a typical SW&HW integrated architecture able to implement this kind of the analysis both off- (i.e. laboratory scale) and on-line (i.e. industrial scale), an example referred to end-of-life concrete was described and analysed. The approach can be "easily" extended to other secondary raw materials (i.e. cullets, fluff, metals scraps, electric and electronic waste, compost, etc.), where expensive and sophisticated control architectures cannot be often adopted both for technical (e.g. particles of different size, shape and composition) and economic reasons (i.e. high analytical costs, long time delay in respect of sample collection, strong environmental impact, etc.).

9405-36, Session PTues

Localizing people in crosswalks with a moving handheld camera: proof of concept

Marc Lalonde, Claude Chapdelaine, Samuel Foucher, CRIM (Canada)

The present work is a small contribution to a larger project that aims at measuring the trajectory of blind subjects who are asked to cross a street intersection with the aid of various cuckoo-chirp type signals. Some signals may be better designed and more efficient in guiding a blind person to align himself/herself with the crosswalk. One key component of the project is to measure the person's deviation with respect to the center of the crosswalk. Although competing techniques are generally efficient and readily available (e.g. GPS), they are not quite appropriate in this context due to their bulkiness and limited spatial resolution (15 cm at most is required). One approach that is being explored is based on the semi-automatic analysis of the video footage taken during the experiments. Due to logistical as well as optical reasons, positioning a single fixed camera in such a way that the whole crossing event is recorded at good resolution is impossible. So the strategy under analysis is to use a moving, handheld camera to record the subject's displacement and then, in offline mode, to track and localize his/her feet with respect to markings painted on the ground for spatial referencing.

The positioning is relative to some origin arbitrarily set to be the intersection between sidewalk edge and the center of the crossing. This relative positioning is possible if a maximum of spatial cues are detected and tracked in the video shot, and then properly mapped to the physical markings (painted gridlines) on the street. Such mapping is nonambiguous for lines parallel to the crosswalk (type V) but in the case of lines perpendicular to the crosswalk (type H), line counting is necessary in order to figure out which portion of the crosswalk is being observed by the moving camera. Each video frame is analyzed to detect situations where the subject's feet are found in a parallelogram made of pairs of V and H lines, in which case the mapping information can be used to deduce his/her relative position.

Tracking lines proceeds by asking the user to supply pairs of points that model all lines visible in the first video frame. At each subsequent frame, points are moved in a systematic fashion inside a small neighborhood and a line likelihood is evaluated. The procedure can be seen as a kind of adjustment of the position of the line model so that a fit criterion is maximized at each frame. In case the fit is inadequate (likelihood value is too low), tracking for this paticular line halts. Lines are tracked independently.

Accurate spatial positioning of the subject with respect to the crosswalk requires detection and tracking of his or her feet throughout the video shot. The easiest and most robust approach for performing tracking involves the use of a distinctive marker (yellow tape) affixed to the right foot of the subject. Tracking of the yellow sticker is based on the selection of the optimal combination of YCbCr color channels that best discriminates between the sticker and background (asphalt, clothing, road markings, etc.). The selection is guided by the maximization of the separability between the probability distributions of sticker and background colors. Once the image coordinates of the centroid P of the yellow patch are found, a verification is made as to whether or not P lies in a parallelogram defined by some of the line markings being tracked. In case of a positive verification, the relative location of the point P inside the parallelogram can be reprojected by bilinear interpolation to its real-world counterpart.

In order to test the application and validate the whole process, we analyzed a video sequence lasting about 27 seconds, which is the average time taken by the blind subjects to cross the street. The images were captured using a consumer handheld camcorder held by an observer about 2-3 meters behind the subject. In addition to video acquisition, the subject's position was recorded using a field computer equipped with a GNSS receiver. Because the GNSS data are tainted with localization errors, it appears that there is greater agreement between the vision-based data and what can be observed in the video sequence than with the GNSS data. Many more video sequences are being analyzed in order to validate this claim.

9405-37, Session PTues

Fused methods for visual saliency estimation

Amanda S. Danko, Siwei Lyu, Univ. at Albany (United States)

Our visual system processes an extreme amount of data about the world around us every single second. The complex cognitive mechanisms involved are as of yet not fully understood, and therefore of particular interest to a variety of research fields. In the computer vision community, modeling these processes efficiently on large amounts of visual data has become a well-known problem. Finding prominent objects and regions within an image requires powerful reduction strategies, yet provides for optimal results for image retrieval[5], compression, object detection[6,7], and many other common machine vision problems [4].

While there exist many saliency models with different perceptual justifications and mathematical formulations, we argue for the fusion of the latest techniques to improve upon the independent methods significantly. Bottom-up pre-attentive saliency models are categorized as they are stimulus-driven rather than guided by a specific task. Context-aware models generally incorporate some other sensory information into an otherwise stimulus-driven method. Having examined the state-of-the-art in these two categories, we find that the combination of two (or more) methods between them yields a saliency map closer to the ground truth.

We put this theory to the test in two experiments. Primarily, we evaluate the saliency maps generated by a set of ten state-of-the-art methods against those generated by our model, across the MSRA[1] and NUSEF[2] datasets. We find that on average our model yields definitive improvements upon recall and f-measure metrics with comparable precisions. The next test of our method involves content-based image-retrieval, with the use of open-source library LIRE[3]. Over 8,500 images from the MSRA and NUSEF datasets were indexed to a searchable dataset. We then employ the saliency maps from the first experiment to create crops of 2,000 original images, containing only the most salient part(s). This is done for each model, and

the crops are used as query images to the dataset. LIRE assigns a score to all returned results, and we use this score as an evaluation metric, in combination with the position which the correct result showed in the result list. We find that all searches using our fused method returned more correct images and additionally ranked them higher than the searches using the original methods alone.

The simple fusion method of visual saliency proposed here highlights the abilities of state-of-the-art models while compensating for their deficiencies. We demonstrate the effectiveness of our fusion method with experiments on visual saliency results as well as applications in content-based image retrieval tasks, demonstrating improved performance. In the future, we will apply this method to other computer vision problems including object detection and image summarization, and expand the experiments to other datasets for a variety of challenges.

References

[1] T. Liu, J. Sun, N-N. Zheng, X. Tang, H-Y. Shum. Learning to detect a salient object. In CVPR, 2007.

[2] R. Subramanian, H. Katti, N. Sebe, M. Kankanhalli, T-S. Chua. An eye fixation database for saliency detection in images. In ECCV, 2010.

[3] L. Mathias, S. Chatzichristofis. Lire: Lucene image retrieval an extensible java cbir library. In ACM ICM, 2008.

[4] L. Itti. Automatic foveation for video compressing using a neurobiological model of visual attention. IEEE Transactions on Image Processing, 2004.

[5] R. Datta, D. Joshi, J. Li, J. Wang. Image retrieval: Ideas, influences, and trends of the new age. ACM Computing Surveys, 1:1–60, 2008.

[6] E. Erdem, A. Erdem. Visual saliency estimation by nonlinearly integrating features using region covariances. Journal of Vision, (4), 2013.

[7] L. Zhang, Z. Gu, H. Li. Sdsp: A novel saliency detection method by combining simple priors. In ICIP, 2013.

9405-38, Session PTues

Classification of hyperspectral images based on conditional random fields

Yang Hu, Eli Saber, Sildomar Monteiro, Nathan D. Cahill, David W. Messinger, Rochester Institute of Technology (United States)

Rapid developments in satellite and sensor technologies have led to a significant increase in the availability of high resolution remotely sensed images. The ability to collect images remotely is expected to far exceed the capacity to analyze these images manually. Consequently, image analysis techniques are urgently needed to support analysts in generating results in an efficient and timely manner. Hyperspectral image classification, especially land-over classification, is one of the most promising areas. The classification results can be employed in various areas, such as object tracking in national defense system and agricultural field change detection.

The conventional probabilistic frameworks, such as Markov Random field (MRF), have the drawback of incorporating occasional inaccurate assumptions. Conditional Random Fields (CRFs) are advantageous over traditional models due to its following characteristics: no requirement of the independence assumption for observations, flexibility in defining local and pairwise potentials, and an independence between the modules of feature selection and parameter learning procedure. On the other hand, hyperspectral images that are correlated spatially and spectrally can provide abundant information across the bands to improve classification results. Therefore, CRFs possess the capability to generate accurate classification results in remotely sensed images.

CRFs have demonstrated better performance in classificationfrom previous work. Multinomial logistic regression (MLR) and Ising/Potts models categorize images by generating local and pairwise potentials, followed by utilizing graph cut or loopy belief propagation (LBP) as the inference method. In order to improve the performance, we propose to extract more meaningful features and apply the tree-reweighted belief propagation as the approximate inference method.

The framework of the algorithm contains the following modules:

Module 1 selects the training and testing datasets in a reasonable way by fixed size block sampling and stratified sampling based on the population of each class. The latter method preserves the class distribution information of a whole dataset and assures all classes to appear in most of the training and testing subsets.

Module 2 serves to determine the CRF graphical model by defining the class of each pixel based on observations and its first order clique with its four neighbors. Local and pairwise potentials, including local intensity, spatial information, histogram of gradient and initial classified results, are taken into consideration in this step.

Module 3 performs the pseudo likelihood learning procedure to estimate the parameters for local and pairwise potentials, followed by the tree-reweighted belief propagation inference method

The results of the above proposed algorithms using two standard publicly accessible hyperspectral image datasets, AVIRIS Indian Pine and HYDICE DC Mall, are presented in this paper. Comparisons between state-of-the-art classification methods based on Support Vector Machines and CRFs are also illustrated.

9405-39, Session PTues

Pro and con of using Gen<i>cam based standard interfaces (GEV, U3V, CXP and CLHS) in a camera or image processing design

Werner Feith, Sensor to Image GmbH (Germany)

Based on more than 100 interface core deliveries worldwide and our work in the standard committees I will sum up the technical possibilities of the Genicam software standard as well as its hardware standards of GEV, U3V, CXP and CLHS and the potential resources and limitations in the FPGA and PC world of standards implementation.

As most applications are driven by individual and restricting custom specifications, I will include samples of a few typical customer as standards driven interface specification phases and conclude on a technical reasonable implementation as potential cost of the implementation of the customers product.

The talk shall conclude in an advice on when to use standards, or parts of a standard, or total user defined implementations to give orientation to the intended audience of camera as embedded image processing designers for their future work on interfaces.

The content of the talk will be adjusted and reviewed by the heads of the technical committees of EMVA for Genicam, AIA for GEV, U3V and CLHS and JIIA for CXP to keep the talk as neutral as possible. A good change for this coordinating action would be early October 2014 as the Genicam meeting is taking place from 6-10 October in Yokohama this fall.

9405-40, Session PTues

Geological applications of machine learning on hyperspectral remote sensing data

Kevin Tse, Y. Li, Edmund Y. Lam, The Univ. of Hong Kong (Hong Kong, China)

No Abstract Available

9405-13, Session 4

An edge-from-focus approach to 3D inspection and metrology

Fuqin Deng, Jia Chen, Harbin Institute of Technology (China); Jianyang Liu, Southwest Jiaotong Univ. (China); Zhijun Zhang, Jiangwen Deng, Kenneth S. M. Fung, ASM Pacific Technology Ltd. (Hong Kong, China); Edmund Y. Lam, The Univ. of Hong Kong (Hong Kong, China)

A high-resolution 3-D imaging system is often required in many machine vision applications especially in automatic inspection of semiconductors. Since the edges of an inspected object are the most important clues in such applications, we build a motorized high-resolution focusing system to capture these edges and propose an edge from focus approach to extract the 3-D data effectively and efficiently.

In our focusing system, we use a coaxial setup both in the illumination and the imaging paths to avoid the occlusion in other 3-D imaging systems such as stereo and laser triangulation systems. Equipped with a large numerical aperture objective lens, the proposed system can capture the high-resolution edges of the inspected object within a narrow depth of field. Moreover, by motorizing the focusing system along the optical axis, we can extend the measuring range based on the captured image sequence.

However, due to the limited depth of field in this high-resolution system, using it for our semiconductor inspection and metrology applications such as tracing the height of the wire loop along its edges, we have to capture a large number of images in order to cover the whole wire loop. Then, most of the data within the captured images are out-of-focus, which distract us from extracting the useful data for further applications. Hence, we present an effective and efficient edge from focus approach to extract the 3-D edges for an inspected object. First, by averaging these captured high-resolution images including both in-focus and out-of-focus data, we synthesize a uniformly smoothed image to remove most of the noise within the image sequence. Second, based on this single synthesized image, we apply an edge detection technique to locate the positions of all the 2-D edges of the whole inspected object. With these two image processing operations, we do not need to detect the trivial edges within each image. Furthermore, we do not need to spend time and effort on merging these trivial edges and removing the false edges of the inspected object either. Therefore, our approach can locate the true edges of the object efficiently. Third, since the contrast of the edge changes gradually when it moves from the in-focus plane to out-of-focus planes, we can estimate the depths of the edges by analyzing the contrast variation within the image sequence. Hence, at each location of the previously located edge, we formulate the depth estimation as an optimization problem and solve for the depth by locating the peak of the focus curves along the relative position of this edge to the imaging system. Finally, we obtain the sparse 3-D edges of the inspected object, which can be used for high-level inspection and metrology applications. For comparison, in the conventional depth from focus method, one can reconstruct a full-field but noisy depth image, which is difficult and inconvenient to be used for extracting the valid data for further applications. Both simulation and real experiments demonstrate the effectiveness of the imaging system and the proposed approach.

9405-14, Session 4

Improved metrology of implant lines on static images of textured silicon wafers using line integral method

Kuldeep Shah, Eli Saber, Rochester Institute of Technology (United States); Kevin Verrier, Varian Semiconductor Equipment Associates, Inc. (United States)

In semiconductor wafer manufacturing processes, the measurement of mask wearing over time is important to maintain the quality of the wafers and the overall yield. Mask wearing can be estimated by measuring the width of lines implanted by it on the substrate. An automatic defect detection system can increase accuracy of the implantation process and reduce waste. Previous methods proposed image analysis algorithms to detect and measure these lines. They have been shown to perform well on polished wafers. Although it is relatively easier to capture images of textured wafers, the contrast between foreground and background is extremely low. In this paper, we propose an improved technique for detecting and measuring implant line widths on textured semiconductor wafers. A rectangular region-of-interest is selected from the pseudo-square shape of the wafers. Due to repeated patterns of textured lines in the image, a fast non – local means method is used as a pre – processing step to denoise the image. Following image enhancement, the previously proposed Line Integral technique can be used to extract the position of each line in the image. Full – Width One –Third Maximum approximation is then used to calculate line widths in pixel units. The line widths thus measured are converted into real – world metric units using a photogrammetric approach involving the Sampling Distance. The proposed technique is evaluated for line detection using a dataset of synthetic and real life images of textured wafers. Since ground truth for line width measurement is not available, we calculate the histogram of line widths for each wafer image. A large qualitative improvement can be seen in lines detected by the proposed technique. The variance of computed line widths for every wafer is approximately 0.0029, which is lower than the state-of-the-art. The robustness to noise of the proposed technique is evaluated using the same synthetic images and procedure that were used for evaluation in the state-of-the-art. These synthetic images are generated with varying amounts of noise. Precision, recall and F-measure values are calculated to benchmark the proposed algorithm. We found the proposed technique to be more robust to noise, with critical SNR value reduced by 10dB in comparison to the existing method.

9405-16, Session 4

Multispectral imaging: an application to density measurement of photographic paper in the manufacturing process control

Raju Shrestha, Jon Yngve Hardeberg, Gjøvik Univ. College (Norway)

Multispectral imaging, which extends the number of imaging channels beyond the conventional three, has shown to be beneficial for a wide range of applications. This paper presents yet another practical industrial application of multispectral imaging in photographic paper manufacturing process control.

Photographic paper is a paper coated with light sensitive chemical layer (called emulsion), which when exposed to light captures a latent image that is then developed to form a visible image. The quality of photographic paper is measured in different aspects such as color range, archival properties, instant dry-to-touch etc. and all these quality factors are determined by the emulsion layer. Since printing technology is based on subtractive color mixing of four colors (CMYK) [1], small changes in the ink (colorant) thickness of a colorant have a much greater effect on the resulting color. In order to assure acceptable levels of product quality and consistent output, it is important to monitor samples and accordingly control the manufacturing process. Thickness of an ink determines how much it absorbs portion of the spectrum. Density, which is defined as the fraction of radiation (light) absorbed at specific wavelengths, is used to measure the ink thickness. Reflectance density of a sample material is expressed as negative logarithm to the base 10 of the reflectance factor [2]. Reflectance factors for the four colorants are calculated from the spectral reflectance of the surface and four standard complementary filters (Red, Green, Blue, and a wide band filter called Visual are used). Density can be used to derive other values such as dot gain, print contrast, ink trapping, grayness etc. Reflection densitometer is widely used to measure densities of colorants. But being a spot measurement device, it is not useful to monitor the density of the colorants in a larger area of the photographic paper during the manufacturing process. In order to process control in the manufacturing of photographic paper, it is essential to do measurements in a larger area for maintaining

the quality and consistency of the ink densities, and hence the appearance of the photographic prints. We propose here a LED illumination based multispectral imaging system (LEDMSI) that uses an RGB camera and a number of optimal LEDs, for fast and accurate estimation of the reflectance image of a scene [3], from which density images for the four colorants are obtained.

We designed and developed a 9-band LEDMSI constructed with a Nikon D600 camera and an iQ-LED module from Image Engineering [4], in collaboration with the FUJIFULM Manufacturing, Netherlands. 9-band images of a scene are acquired in three exposures, by turning three different types of LEDs in each exposure. Optimal combinations of LEDs are selected from 22 different LEDs, based on optimization of spectral estimation. We conduct measurements of a number of printed strips containing color charts of different ink densities that are used as references in the manufacturing process. From the spectral reflectance image, densities of the four colorants in every color patch are computed. We analyze the results and compare the measurement accuracy with the reference values. Preliminary experimental results confirm the effectiveness of the multispectral imaging device. We believe that the proposed system would be used in the process control of the manufacturing of the photographic papers soon. We think that this work shows a very good example of industrial application of multispectral imaging.

References

[1] Berns, R. S., Billmeyer and Saltzman's Principles of Color Technology, 3rd edn., Wiley, 2000.

[2] ISO 5, Photography and graphic technology - Density measurements, ISO, 2009.

[3] Shrestha, R. & Hardeberg, J. Y., Multispectral imaging using LED illumination and an RGB camera, in Proceedings of 'The 21st Color and Imaging Conference (CIC)', IS&T, pp.8–13, 2013.

[4] Image Engineering, 'iQ-LED Technology', http://image-engineering-shop.de/shop/article_iQ-LED/IQ-LED.html, 2014.

9405-17, Session 4

Self-calibration of monocular vision system based on planar points

Yu Zhao, Lichao Xu, Univ. of Science and Technology of China (China)

This paper proposes a method of self-calibration of monocular vision system which is based on planar points. Using the method proposed in this paper we can get the initial value of the three-dimensional(3D) coordinations of the feature points in the scene easily, although there is a nonzero factor between the initial value and the real value of the 3D coordinates of the feature points. From different viewpoints, we can shoot different pictures, and calculate the initial value of external parameters of these pictures. Finally, through the overall optimization, we can get all the parameters including the internal parameters, the distortion parameters, the external parameters of each picture and the 3D coordinates of the feature points. According to the experimental results, in about 400mm?400mm field of view, the mean error and the variance of 3D coordinates of the feature points is less than 20um.

9405-18, Session 5

A comparative study of outlier detection for large-scale traffic data by one-class SVM and kernel density estimation

Henry Y. T. Ngan, Hong Kong Baptist Univ. (Hong Kong, China); Nelson H. Yung, Anthony G. Yeh, The Univ. of Hong Kong (Hong Kong, China)

Traffic data captured from large city network is generally enormous in data size. Apart from the fact that a large variety and number of sensors such as inductive loops, ultrasound, radar, laser and surveillance cameras are being deployed on all types of roads over the years, and more importantly, these sensors work 24x7 non-stop to generate vehicle counts, density, flow rate, etc as a basis for management and control. Presently, human operators monitor most of the traffic situations, especially for abnormal ones such as congestion, diversion, road work and accidents. Obviously, the tasks of monitoring and analysis would have to depend on machines in the future. Moreover, as data size increases, there is always errors due to all sorts of reasons inherent in the data set itself. For example, error may come from sensors or when it is transmitted to the TCSS. Errors are to be removed or corrected, while data representing abnormal situations need to be identified effectively. Therefore, this research described in this paper aims at detecting outliers (abnormal events, data errors) for large-scale traffic data. The assumption is that outliers are assumed to be the minority in the data while inliers are the majority. A traffic database with 764,027 vehicles collected from one of the busiest 4-armed junctions (19 traffic directions, 23 sessions each direction) at Hong Kong is utilized for evaluation. The traffic data is originally in a format of spatial-temporal (ST) signal and there are a high similarity among ST signals, no matter with outliers or not. A ST signal is in term of vehicle volume in one direction (spatial axis) versus traffic cycle (temporal axis) and with high dimension of traffic cycles from each 3-hour session. The dimension of every ST signal is reduced using principal component analysis (PCA). Herein, the first two coefficients of each ST signal are representative enough, hence they are extracted and represented as a pair of (x,y)-coordinates. Then, each traffic signal has 23 pairs of (x,y)-coordinates, for which are assumed to conform as a Gaussian mixture model (GMM). Two typical outlier detection (OD) methods, namely one-class SVM and KDE, are used because a likelihood on GMM can be utilized to as threshold to detect any outlier. One-class SVM, as a common machine learning tool, labels training data as one-class and classifies whether input data is the same class as the training class or not. The same class means an inlier, otherwise it is an outlier. KDE is a non-parametric method to estimate a random variable's probability density function (pdf). By estimating the pdf on the KDE, a threshold can be employed on a certain level of pdf as likelihood to classify a datum is an outlier or not. Any datum outside a thresholded likelihood is classified as an outlier. In the performance evaluation, the one-class SVM and KDE achieved average detection success rate of 59.27% and 95.20%. It is believed this research can be beneficial to researchers and practitioners of traffic incident management and intelligent transportation systems.

9405-19, Session 5

Image-based dynamic deformation monitoring of civil engineering structures from long ranges

Matthias Ehrhart, Werner Lienhart, Technische Univ. Graz (Austria)

MOTIVATION

Today, many large-scale civil engineering structures are permanently monitored to provide early warnings and to initiate counter actions from structural failure. Total station measurements are commonly used to determine 3D movements with measurement intervals of several minutes or hours. However, these measurements do not provide information on the vibration behavior of the structures. For this purpose other sensors like accelerometers have to be installed on the object. In this paper we present a new approach to utilize existing state of the art total stations for the vibration monitoring from long distances using passive targets.

METHODS

Modern total stations are equipped with cameras integrated in the telescope. Currently, these cameras are used for documentation purposes only. The telescope cameras are promising sensors for accurate deformation measurements from long distances since the spatial resolution of the resulting image data benefits from the 30x optical magnification of the telescope.

With the instrument used in this investigation, it is possible to capture video streams in real time with 10fps and a spatial resolution of 1.7''/pixel which corresponds to 1.2mm at a distance of 150m. Furthermore, the instrument is equipped with a laser distance measurement unit. With the known distance it is possible to automatically set the camera's focus position and to relate the angular quantities gained from image processing to units of length.

To measure the vibrations of civil engineering structures, it is necessary to have prominent features in the camera's video stream. To achieve accurate results, we use circular target markings rigidly attached to the object. After detecting the contour of the marker, we estimate its center based on the contour coordinates. Herby, we employ a least squares adjustment according to the Gauss-Helmert model from which the parameters and their standard deviations are derived. The knowledge of the standard deviations is important in monitoring applications in order to evaluate the statistical significance of the measured deformations.

EXPERIMENTS & RESULTS

The proposed method was tested in the laboratory and on a real structure. To verify the accuracy of the complete system, we mounted a marker on a motorized translation stage. The movements of the used translation stage can be controlled with an accuracy of better than 0.01mm. Afterwards, the target was automatically moved to different positions and a video was taken at each position. Compared to the reference movements given by the translation stage, the computed movements from single frames showed deviations of less than 0.2mm. The movements resulting from the averaged videos at each position showed deviations of less than 0.05mm.

To proof the feasibility of the measurement system, a field test on a footbridge was carried out. Reference vibration values were determined using accelerometer measurements. The results were compared with the vibrations derived by image processing using the videos collected with the total station.

CONCLUSION

In his paper we demonstrate that existing total station hardware can be used to detect vibrations of civil engineering structures with a high accuracy from distances of more than 100m. Compared to conventional image-based systems, the orientation and position of the camera can be determined in a global coordinate system with high accuracy. The stability of the camera's viewing direction is monitored with an integrated high precision tilt sensor. The approach is based on robust hardware which can withstand harsh environmental conditions.

9405-20, Session 5

Building and road detection from large aerial imagery

Shunta Saito, Yoshimitsu Aoki, Keio Univ. (Japan)

(1) A problem statement and motivation for this work: Building and road detection from aerial imagery has many applications in a wide range of areas including urban design, real-estate management, and disaster relief. The extraction of regions of buildings and roads from aerial imagery has been performed by human experts manually, so that it is very costly and time-consuming process. Our goal is to develop a system for automatically detecting buildings and roads directly from aerial imagery. Many attempts at automatic aerial imagery interpretation have been proposed in remote sensing literature. Much of early works use local features to classify each pixel or segment to an object label, but this kind of approach needs some prior knowledge on object appearance or class-conditional distribution of pixel values. Some of these methods also need a segmentation step as a pre-process. There, we use Convolutional Neural Networks (CNNs) to learn mapping from raw pixel values in aerial imagery to object labels (e.g., buildings, roads, and otherwise.) It is equivalent to generate three-channel maps from aerial imagery. (2) Methods: Firstly we divide aerial imagery into 64 x 64 image patches. Each patch is three-channel (RGB) color image and normalized by subtracting mean value in the patch and divided by a standard deviation calculated from all patches. We prepared more than three hundred thousand image and label patches for training CNNs. The objective function is defined as softmax over probabilities of three types of objects, building, road, and otherwise. Each output patch of CNNs is three-channel mask image and each channel represents probability for each object, so that the summation over channels at a pixel is always one. We train CNNs using stochastic gradient descent. Finally we evaluate our system on a large-scale road and building detection datasets proposed by Mnih [1]. This dataset includes aerial imagery, corresponding building labels, and road labels. Mnih has also proposed an automatic system for labeling aerial imagery [1], but it provides a single channel output (building or road). Considering the trade-off between building and road likelihood while training CNNs, the predicted map patches can be more accurate. Because a pixel belonging to building label is always not road, vise versa. (3) Experimental results: The most common metrics for evaluating building or road detection systems are precision and recall. Precision and recall are known as correctness and completeness respectively, in remote sensing literature. The precision is the fraction of predicted building or road pixels that are true buildings or roads, while the recall of a set of predictions is the fraction of true building or road pixels that are correctly detected. We use these metrics to evaluate our system. We train the CNNs using training dataset, and then use test dataset to calculate precision and recall. The experimental results show state-of-the-art or better performance on the same dataset as the methods proposed by Mnih [1]. (4) Conclusion: We proposed an automatic aerial imagery labeling system using Convolutional Neural Networks (CNNs). It was trained using a large-scale building and road detection dataset proposed by Mnih [1]. We considered the trade-off between building and road likelihood while learning by using softmax over probabilities of all object types as the objective function of CNNs. Then we presented state-of-the-art or better performance of our system on the dataset [1]. (5) Reference: [1] Volodymyr Mnih, "Machine Learning for Aerial Image Labeling", Ph.D Thesis, University of Toronto, 2013

9405-21, Session 5

Interactive image segmentation tools in quantitative analysis of microscopy images

Reid B. Porter, Christy Ruggiero, Los Alamos National Lab. (United States)

In material science and bio-medical domains there is great value in automatically detecting, delineating and quantifying particles, grains, cells, neurons and other functional "objects" within digital microscopy images. These are challenging problems for image processing because of the variability in object appearance that inevitably arises in real world image acquisition and analysis. Even within the same image, consistent and reliable segmentation of objects across the field of view is difficult, and as we look to apply methods across multiple object types, multiple images, multiple instruments and instrument settings, off-the-shelf segmentation methods become increasingly ineffective.

One of the most promising (and practical) approaches that can address some of these challenges is interactive image segmentation. These algorithms are designed to incorporate input from a human operator to tailor the segmentation method to the image at hand. Interactive image segmentation is now a key tool in a wide range of applications within microscopy as well as more general applications, such as background removal tools in consumer photo editing.

Historically, interactive image segmentation methods have tailored segmentation on an image-by-image basis, and information derived from operator input is not transferred between images, e.g. Interactive graph cuts, marker based segmentation and active contour methods. However more recent advances in machine learning, and more specifically structured output prediction, have opened the door to a new class of methods that accumulate and learn from the operator input over longer periods of time. Unlike previous methods these new "learning" methods reduce the need for operator input over time, and can potentially provide a more dynamic balance between customization and automation for different applications.

In this paper we provide a unified review of historical and more recent interactive image segmentation methods and describe how different methods provide solutions to inference and learning problems in different ways. We then develop and evaluate some of the new "learning" methods

with respect to a number of particle and grain segmentation tasks found in microscopy images of metals and particulate materials. We report on the current performance with respect to the real-world application requirements, such as quantitative microscopy analysis for forensic evidence requirements, and outline the remaining challenges to improve the reliability and scalability of these methods in the future.

9405-22, Session 6

Camera-based forecasting of insolation for solar systems

Daniel Manger, Frank Pagel, Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (Germany)

With the energy transition towards renewable energies, electricity suppliers are faced with huge challenges. Especially the increasing integration of solar power systems into the grid gets more and more complicated because of their dynamic feed-in capacity. While flexible energy storages for short-term surpluses are still a research topic, there are also initiatives to control the feed-in capacity. For example, in Germany, new solar power systems have to be equipped with a remote shut-down functionality so that grid operators can limit or even shut down solar power systems in case of grid overloads. Since according to the law, the owners of the solar power systems have to be reimbursed for the lost feed-in compensation, this leads to the absurd situation that ultimately, the end-consumers are paying for the unused power, too. To solve the problem by dynamic energy storages and for the proper stabilization of the grid, the feed-in capacity of a solar power system within the next seconds, minutes and hours should be known in advance. Coarse predictions are possible with the information of the weather forecast or from satellite images. However, both the temporal and the spatial resolution of satellite images are not sufficient for an accurate short-term forecast for one system at a certain location.

In this work, we present a camera-based system for forecasting the feed-in capacity of a solar system. To this end, the camera is targeted at the sky and clouds are segmented, detected and tracked. For the computation of dense optical flow information, an automatic adjustment of the image contrast is performed. The unsupervised segmentation of clouds was evaluated with different color spaces. The temporal aspect of the forecast is then obtained from the tracking information of the clouds. The quantitative prediction on the other hand is performed with learning based algorithms. Using a small solar panel, a resistor and measuring devices, feed-in capacity training data for the respective images was collected. With this training data, we build a model which tries to associate the future feed-in capacity with the visual appearance of the clouds approaching the sun. Because clouds look differently when approaching the sun, the offline training stage includes a back-tracking of the clouds in front of the sun. Attention is also paid for the localization of the sun in the image as well as for the handling of blooming and smear effects in the image caused by the sun. The system is evaluated with several sequences of different types of clouds and for different points in time in the future showing the applicability of the camera-based short-term forecasting of insolation for solar systems.

9405-23, Session 6

3D barcodes: theoretical aspects and practical implementation

David Gladstein, Cogswell Polytechnical College (United States); Ramakrishna Kakarala, Nanyang Technological Univ. (Singapore); Zachi I. Baharav, Cogswell Polytechnical College (United States)

Barcodes have served for many years as a very robust and efficient way to connect the printed physical world with computers. From the one-dimensional barcodes printed on most products for easy price and inventory management, to the two-dimensional barcodes used on packages by the post office, and in recent years to the plethora of QR codes used to enable scanning by consumer mobile devices.

In this paper we introduce the idea of 3D barcodes as shown in Fig.~ref{fig:3D}. The barcode is composed of an array of 3D cells, called modules, and each can be either filled or empty, corresponding to two possible values of a bit. These barcodes have great theoretical promise thanks to their very large information capacity, which grows as the cube of the linear size of the barcode, and in addition are becoming practically manufacturable thanks to the ubiquitous use of 3D printers.

In order to make these 3D barcodes practical for consumers, it is important to keep the decoding simple using commonly available means like smartphones.

We therefore limit ourselves to decoding mechanisms based only on three projections of the barcode, which imply specific constraints on the barcode itself. The three projections produce the marginal sums of the 3D cube, which are the counts of filled-in modules along each Cartesian axis.

The flow of the work is as follows.

We start by introducing the novel concept of 3D barcodes, and relating it to previous work on barcodes. We then move to the theoretical side, as we analyze and show the applicability of the work done on the reconstruction of 2D binary matrices based on row and column projections. We discuss both some of the original papers (e.g., [1]) and many more recent ones [2-5]. The latter work suggests ways to determine the uniqueness of the matrix given the marginal sums, the cardinality of the non-unique cases, and the cases where marginal sums do not correspond to any feasible matrix. We build upon this work to give upper bounds on the information content of a 3D barcode, given the constraint of decoding using marginal sums. Furthermore, we present simulations to illustrate the applicability and relevance of these upper bounds to practical cases.

To demonstrate practicability, we describe the construction of a 3D code in a piecewise manner. This construction enables the decoding the cube in slices. This is important since the complexity of decoding a large 3D code as a whole becomes intractable very quickly. For example, using a brute force dictionary look-up method, even a small cube of 5x5x5 elements on side produces 2^125 possibilities. Moreover, decoding the cube piecewise affords robustness and makes decoding feasible on consumer devices instantly. The practical application combines both computer vision and image processing techniques, which we will describe.

We close with the demonstration of various printed 3D barcodes, and describe practical aspects of their manufacturing.

9405-24, Session 6

Still-to-video face recognition in unconstrained environments

Haoyu Wang, Changsong Liu, Xiaoqing Ding, Tsinghua Univ. (China)

In this paper, we introduce a novel and robust solution for still-to-video face recognition in unconstrained environments. Face recognition has been an active research topic for many years, owing to its broad range of applications and the exponential increase in computing power. A number of methods have been proposed to deal with the conventional face recognition problem of identifying a face from a single image. However, such methods may fail to work in video-based face recognition tasks. Although videos provide much more information than images, it is hard to exploiting and utilizing it effectively. Different video sequences of the same person may contain several kinds of variations, including pose, facial expression, illumination, image resolution and occlusion. Untagged and unpredictable variations lead to mistakes in recognition. Motion blur and compression artifacts also deteriorate recognition performance. How to recognize faces from videos captured in unconstrained environments has become a new challenge for researchers.

Moreover, in real world applications such as law enforcement, video surveillance and e-passport identification, only a single still image per person is available. It is infeasible to use covariance matrix to describe variations in each category. Consequently, traditional subspace-based

methods such as Linear Discriminant Analysis (LDA) cannot be applied in this scenario. Meanwhile, differing from the gallery set consisting of a single still image per person, probe sets are usually collected on the spot in the form of video sequences. It is crucial to match multiple probe images with the single gallery image correctly without being affected by possible outliers.

We incorporate a regularized least squares regression approach in the framework to tackle the multi-modality problem. Compared with those in the corresponding still images of the same subject, faces in video frames have very different appearances. It is inappropriate to directly calculate distances or similarities between these two modalities as usual. We assume that face images of the same person are identical in the identity space, and multi-modal or heterogeneous images can be obtained from identity vectors via linear transformations. By learning the projection matrix from each sample space to the identity space, images from various modalities can be transformed to the same space where the matching procedure is performed. The learning method is to solve a least squares regression problem with several regularization terms, which are based on prior knowledge and heuristic assumptions to avoid overfitting.

When it comes to the single still image per person problem, the traditional method is to match each probe image with the single gallery image and then combine results by some mathematical operations. However, it is more helpful to use synthesized virtual samples based on the single gallery image and face variations learned from the training set. As a result, set-to-set metric learning, instead of point-to-point one, can be utilized to make the matching results more robust and accurate. We inherit from the affine/convex hull-based approach to find nearest points as representatives of two point sets, and use improved collaborative representation with regularizations as the learning and matching algorithm.

Extensive experiments on COX-S2V, ChokePoint and YouTube Faces datasets using unconstrained video sequences demonstrate that the proposed framework performs better than many video-base face recognition methods with considerable improvements in recognition accuracy.

9405-25, Session 6

Realistic texture extraction for 3D face models robust to self-occlusion

Chengchao Qu, Eduardo Monari, Tobias Schuchert, Jürgen Beyerer, Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (Germany)

In the context of face recognition, 3D representation is shown to add useful information to the overall performance. Probably the most well-known approach to represent 3D faces is the 3D Morphable Model (3DMM) proposed by Blanz and Vetter. Separate linear subspaces for shape and texture are learned by the Principal Component Analysis (PCA) for compact representation. When 3DMM is fitted to a 2D image, not only the shape coefficients, but also the texture and illumination parameters are simultaneously estimated, resulting in a huge parameter space and slow runtime. Alternatively, the 3D shape can be inferred solely based on a few dozens of sparse 2D feature points, allowing for real-time reconstruction. However, since the texture parameters are completely left out, they must be extracted from the image afterwards. This paper addresses the possible problems in texture extraction caused by self-occluded facial region.

The complete 3DMM fitting takes account of the learned PCA texture model and Phong illumination to minimize the fitting error of shape and texture. Later, more rendering features, e.g., specular highlight and edge constraint, are added by Romdhani to avoid local minima. These methods can reconstruct the texture parameters including the occluded facial part according to the 3D dataset and the estimated illumination at the cost of computational time. However, if only the 3D shape is reconstructed for fast online applications, or real facial texture is needed, e.g., for forensic analysis, an extra texture extraction step is necessary.

This approach assumes the 3D shape of the face is already recovered. After registering the shape back to the 2D image, small displacements caused by limited 3DMM subspace are dealt with by affine warping and interpolation.

After determining the visibility of the vertices, color information of them can be extracted by linear non-uniform interpolation. To overcome the self-occlusion problem, some approaches leverage the symmetric property of the face by mirroring the visible part of the face and provide some good-looking results on well illuminated studio images. This naïve approach has two drawbacks. Even minor illumination difference between the mirrored parts will result in severe inhomogeneous intensity and poor visual effect. On the other hand, some facial regions, e.g., between the chin and neck, are often invisible in both face halves. In this work, the “bad” half is first determined. Starting from the cheek near the nose area, a virtual texture map for the homogenous area is generated iteratively by averaging the color of neighbored visible vertices until the whole face is “filled”. Although this step creates unrealistic, overly smoothed texture, illumination stays constant between the real and virtual texture. In the second pass, the mirrored texture is gradually blended with increasing weight. At places where the original vertices are visible, the real texture is used, otherwise, the generated texture is employed for blending. This scheme ensures a gentle handling of illumination and yet yields realistic texture. Because the blending area only relates to non-informative area, main facial features, i.e., eyes and mouth, still has unique appearance in different face halves, which is proven to be crucial to face recognition.

We evaluate our approach on several “in the wild” image and video datasets containing diverse pose and illumination variations. Experimental results reveal realistic rendering in novel poses robust to unconstrained conditions and small registration error.

Conference 9406: Intelligent Robots and Computer Vision XXXII: Algorithms and Techniques

Monday - Tuesday 9–10 February 2015

Part of Proceedings of SPIE Vol. 9406 Intelligent Robots and Computer Vision XXXII: Algorithms and Techniques

9406-1, Session 1

Synchronization of mobile robot's actuated wheels

Ville Pitkänen, Antti Tikanmäki, Juha Röning, Univ. of Oulu (Finland)

Historically, trapezoidal velocity profiles have been widely used to control engines. Nevertheless, the evolution of robots and their uses has led to the need of using smoother profiles, due to the demand of high precision and delicate movements. It has been shown that this can be achieved by minimizing the change of acceleration and using s-curve profiles. Moreover, to provide a good control of the movement of a robot, it is necessary to ensure that it will meet the desired velocity profile. Therefore, a way to prevent how the wheels will react on the soil becomes highly useful, in order to adapt the supplied torque.

This paper suggests a model to define an appropriate s-curve velocity profile given the desired starting and ending kinematic states for a mobile robot. The study is then focused on a one-wheel system to define the interaction between the soil and a wheel. This interaction is modelled and extended in order to calculate the required torque, drawbar pull and power needed to fulfil the desired s-curve velocity profile. Finally, an introduction to unicycle robots is given as an example of how the proposed models could be applied in the motion planning of a mobile robot.

9406-2, Session 1

Moving object detection from a mobile robot using basis image matching

Du-Ming Tsai, Yuan Ze Univ. (Taiwan); Wei-Yao Chiu, Industrial Technology Research Institute (Taiwan); Tzu-Hsun Tseng, Yuan Ze Univ. (Taiwan)

Mobile robots become very important for many potential applications such as navigation and surveillance. In this paper, we propose an image processing scheme for moving object detection from a mobile robot with a single camera. It especially aims at intruder detection for the security robot on either smooth or uneven ground surfaces. The proposed scheme uses the template matching with basis image reconstruction for the alignment between two consecutive images in the video sequence. The most representative template patches in one image are first automatically selected based on the gradient energies in the patches. The chosen templates then form a basis image matrix. A windowed subimage is constructed by the linear combination of the basis images, and the instances of the templates in the subsequent image are matched by evaluating their reconstruction error from the basis image matrix. For two well aligned images, a simple and fast temporal difference can thus be applied to identify moving objects from the background.

The proposed template matching can tolerate -10~+10 degree in rotation and -10~+10% in scaling. By adding templates with larger rotational angles in the basis image matrixes, the proposed method can be further extended for the match of images from severe camera vibrations. Experimental results of video sequences from a non-stationary camera have shown that the proposed scheme can reliably detect moving objects from the scenes with either minor or severe geometric transformation changes. The proposed scheme can achieve a fast processing rate of 32 frames per second for images of size 160x120 pixels.

9406-3, Session 1

Dealing with bad data in automated decision systems *(Invited Paper)*

Charles A. McPherson, Draper Lab. (United States)

The robotics and automation community has recently demonstrated the feasibility of level 4 automation in passenger vehicles. The DARPA Urban Challenge has lead to the Google Driverless Automobile pursuit. In addition, even the automated safety systems available in the BMW 3 Series or as recently introduced by Mercedes in their S-Class sedan are providing level 2 automation on the current market. According to a recent advertisement on the Mercedes Benz website, "Today's S-Class literally looks ahead, and 360 degrees around, to spot hazards in your path. A team of standard and optional systems can alert the driver, assist in braking, and even respond autonomously to help avoid collisions with other vehicles and pedestrians." With the increase in reliance upon automation to perform safety functions to fully autonomous operation, one reality poses a constant concern to developers. Sensors do, and will, fail. Furthermore, how will the overall system respond to component failures.

This paper examines a variety of methods of performing analysis of failures in system components with regard to how they effect the decision made by the system. We present the results of a recent study that investigated the characteristics of different methods of data filtering upon common features used in feature-based classifiers/detectors. We also present results of observations on handling bad data conditions and comparisons between the effects of bad data on a variety of classifiers such as Support Vector Machines, Maximum Likelihood, and Dempster-Shafer. Finally, we discuss advantages and disadvantages of a range of solutions to failure analysis ranging from discrete event simulations to pure statistical analysis which suggests that a mix of the two approaches may better provide understanding to system sensitivities.

9406-4, Session 2

Thorough exploration of complex environments with a space-based potential field

Alina Kenealy, Nicholas Primiano, Alex O. Keyes, Damian M. Lyons, Fordham Univ. (United States)

Robotic exploration, for the purposes of search and rescue or explosive device detection, can be improved by using a team of multiple robots. However, searching, exploration and mapping with a team of robots is still an active research area. Potential field navigation methods offer natural and efficient distributed exploration algorithms, mutually repelling team members to cover the area efficiently, they also suffer from field minima issues. Liu and Lyons proposed a Space-Based Potential Field (SBPF) algorithm that dispreses robots efficiently and also ensures they are driven in a distributed fashion to cover complex geometry.

In this paper, the approach is modified to handle two problems with the orginal SBPF method: fast exploration of enclosed spaces, and fast navigation of convex obstacles. A "gate-sensing" function was implemented which draws the robot to narrow openings, such as doors or corridors that it might otherwise pass by, to ensure every room can be explored. Secondly, an improved obstacle field vortexing function was developed which allows the robot to avoid walls and barriers while using their surface as a motion guide to avoid being trapped. Finally, the SBPF was integrated with the DP-SLAM algorithm.

Simulation results, where the modified SPBF program controls the MobileSim Pioneer 3-AT simulator program, are presented for a selection of maps that capture difficult to explore geometries. Physical robot results are also presented, where a team of Pioneer 3-AT robots is controlled by the modified SBPF program. Data collected prior to the improvements, new simulation results, and robot experiments, are presented as evidence of performance improvements.

9406-5, Session 2

Localization using omnivision-based manifold particle filters

Adelia Wong, Mohammed Yousefhussien, Raymond Ptucha, Rochester Institute of Technology (United States)

Developing precise, low-cost, and fast-converging spatial localization algorithms is an essential step in creating real-time, fully autonomous navigation systems. An agent must be equipped with a sensory input capable of procuring meaningful data from its surroundings. Data collection must be of sufficient detail and breadth to distinguish unique locations, yet coarse enough to enable real-time collection and processing. The issue is complicated by the broad range of applications and environments in which autonomous robots are expected to perform. Outdoor systems rely on global positioning systems and magnetometers as primary inputs. When combined with road maps and wheel encoders, accurate localization can be achieved. Indoor systems must rely primarily on proximity sensors. Active proximity sensors such as sonar and rangefinders have been used for localization, but sonar sensors are generally noisy and rangefinders are generally expensive. Passive sensors such as video cameras are low cost and feature-rich, but suffer from high dimensions and excessive bandwidth.

This paper presents a novel approach to robot localization using a video camera to collect precise sensory data for indoor robot navigation. To enable an agent to see 360o, an omnidirectional sensory system is utilized. The sensor consists of a low cost camera pointing upwards into a spherical mirror, enabling the collection of a continuous field of view in a single capture. Captured images undergo unwarping and normalizing to condition data for upstream processing. Unwarping extracts slices along a radial line to create a panoramic view of the agent's immediate surroundings. Normalization orients the image such that the dominant interior wall points upward. These conditioning steps not only map spherical images to a canonical representation more suitable for processing, but when combined with the final localization estimate, serves to accurately estimate the pose. Training images along with indoor maps are fed into a semi-supervised linear extension of graph embedding manifold learning algorithm to learn a low dimensional surface which represents the interior of a building. Manifold mapping reduces the complexity of the high-dimensional images while preserving their defining characteristics. The resulting representation is not only more accurate, but more computationally efficient.

An adaptive particle filter processor uses the low dimensional manifold surface descriptor as a semantic signature for particle filter localization. Test frames are conditioned, mapped to a low dimensional surface, and then localized via a particle filter algorithm. These particles are temporally filtered for the final localization estimate. Our method, which we call omnivision-based manifold particle filters (OMPF) improves upon existing localization algorithms. Its feature rich input, along with robust temporal elasticity and trajectory persistence, reduce convergence lag and increase overall efficiency. Indoor agents based on our OMPF navigation system are able to converge upon the most probable robot location in real time as the robot wanders about its environment. A comparison with other computer vision based approaches such as bag of words feature descriptors demonstrates the advantages of our OMPF method.

9406-6, Session 2

An online visual loop closure detection method for indoor robotic navigation

Can Erhan, Istanbul Teknik Üniv. (Turkey); Evangelos Sariyanidi, Queen Mary, Univ. of London (United Kingdom); Onur Sencan, Hakan Temeltas, Istanbul Teknik Üniv. (Turkey)

Visual loop closure detection problem is an active area especially for the indoor environments, where the global position information is missing and the localization information is highly dependent on odometry sensors. Although there exist a range of techniques that depend on pre-located markers or painted lines, we perform an algorithm that detects the natural landmarks called visual landmarks on the environment dynamically. Specifically, we adopt visual place recognition to close loops that is useful for the process of correctly identifying a previously visited location.

In this paper, we present an enhanced loop closure method based on unsupervised visual landmark extraction with saliency detection technique. Image frames are represented sparsely through these landmarks, which are ultimately used to determine the similarity between two images and detect loop-closing events. The place recognition technique that we implement is based on finding saliency regions on the image that is taken from the moving depth camera mounted on an autonomous ground vehicle, which we built before. Saliency regions are used to refer to the certain distinctive areas on the image patches. They are also suitable to represent locations in a sparse manner. Unsupervised extraction of visual landmarks is not a simple task for several reasons. Firstly, a saliency criterion is needed to measure the saliency of a given image patch. Secondly, an efficient search algorithm is needed to test this saliency criterion on a complete image sample and extract the most salient regions. Therefore, the saliency detection problem is formulated as an optimization problem, and an energy function which describes the distinctiveness of a given image patch is defined. To find the global optimum of the energy function efficiently, a Branch & Bound based search technique is employed. The utilized saliency detection technique is able to detect patches that are suitable to be used as visual landmarks, and it performs with very high efficiency.

The first step is extracting the most salient regions on the image. Once the landmarks are extracted, they are described and later re-identified using well-established ferns classifiers. After that, a similarity function is used to measure the similarity between two images through the landmarks identified in each image. In order to find out the similarity function, a straightforward function is defined that depend on the detection confidences and the psychical 3D coordinates of the landmarks on each image. The detection confidence of a single landmark is calculated using the number of overlapped detections around the landmark. Recognition of the previously visited or unvisited locations is determined in accordance with the similarity function.

The tests are done in the halls of the faculty with a known trajectory. Exemplary results and the practical implementation of the method are also given with the data gathered on the testbed with a Kinect mounted differential drive autonomous ground vehicle and it has been shown that our application achieves quite promising results.

9406-7, Session 2

Improved obstacle avoidance and navigation for an autonomous ground vehicle

Binod Giri, Hyunsu Cho, Benjamin C Williams, Hokchhay Tann, Shakya Bicky, Vishal Bharam, David J. Ahlgren, Trinity College (United States)

This paper presents improvements made to the intelligence algorithms employed on Q, an autonomous ground vehicle, for the 2014 Intelligent Ground Vehicle Competition (IGVC). In 2012, the IGVC committee

combined the formerly separate autonomous and navigation challenges into a single AUT-NAV challenge. In this new challenge, the vehicle is required to navigate through a grassy obstacle course and stay within the course boundaries (two white painted lines) that guide it toward a given starting GPS waypoint. Once the vehicle reaches this waypoint, it enters an open course where it is required to navigate to 7 other GPS waypoints while avoiding obstacles. After reaching the final waypoint, the vehicle is required to traverse through another obstacle course before completing the run. Q uses a modular parallel software architecture in which image processing, navigation and sensor control algorithms run concurrently. A tuned VFH algorithm allows Q to smoothly decelerate upon encountering obstacle fields and traverse them with relative ease. Majority of revisions occurred in the vision system, which detects white lines and barrels from raw images and informs the navigation component. A difficulty arose as barrels came in multiple colors; a simple color plane extraction would not suffice. To overcome this difficulty, laser range sensor data was overlaid on top of visual data. In addition, a significant update was made to our JAUS component: the component now accepted a greater variety of messages and showed better compliance to the JAUS technical standard. With these improvements, Q was able to secure the second place in the JAUS track.

Since its inception, Q underwent several iterations of change. Each year Trinity College Robotics Study Team pull together ideas to improve the vehicle's design. Once the new design is in place, a few members are tasked with implementation in hardware and software. This year five members saw through improvements in design and engineering, and participated in the IGVC.

9406-8, Session 3

Statistical approach for supervised codeword selection

Kihong Park, Seungchul Ryu, Seungryong Kim, Kwanghoon Sohn, Yonsei Univ. (Korea, Republic of)

Context

Over the last decades object categorization has received a lot of attention as a key element in computer vision field. The bag-of-words (BoW), one of the most successful methods, represents an image as an orderless collection of quantized local features called codewords. In BoW, how to select distinctive codewords is one of the most important issues.

Objective

Compared to its importance, codeword selection area didn't be developed enough. Many BoW algorithms just select codewords uniformly regardless of considering their importance even though some codewords which are significantly influence to improving the categorization performance, while other codewords disturb the satisfactory classification. This paper analyzes that which factors within codewords influence to a categorization performance, which enables us to propose distinctive codewords selection method based on statistical analysis of within- and cross-category codewords in a supervised manner.

Method

During a codeword generation process, statistics of local features with respect to category class are analyzed. Based on these statistics, we measure the confidence for determining distinctive codewords. Specifically, we define two types of confidences: within-category confidence and cross-category confidence. The within-category confidence of a codeword is measured by the appearance consistency across different images of same category based on the assumption that a consistently presented codeword is more significant for discriminating the category from others. In addition, if a codeword is generated from only a specific category of images, the codeword can be a discriminative indicator of the category. Based on this fact, the cross-category confidence is computed by the ratio of dominant category from which the codeword is generated. The combination of within- and cross-category confidences is used to select distinctive codewords. Unlike the conventional methods, only distinctive codewords are used to construct bag-of-words in the proposed method.

Result

The performance of the proposed method is evaluated for Caltech-101 dataset on Intel Core i7-2600 CPU 2.6GHz and 4G memory PC. The proposed method was compared with the conventional codewords selection method with three base algorithms: the BoW, sparse coding (SC), and locally-constrained linear coding (LLC). Note that the proposed distinctive codeword selection method can be applicable to any other base algorithms. The experimental results showed that the proposed method improves base algorithms with about 2~5% error rate reduction. Moreover, the proposed method reduced the number of codewords about 50%, and consequently overall complexity.

Novelty

First, unlike the conventional uniform codeword selection methods, a distinctive codewords selection method is proposed to reduce the redundant part of codewords and improve the discriminative of codewords. Second, within- and cross-category confidence measures are proposed based on the statistics of category labels of local features during codeword generation process. Third, the proposed method can be combined with any other base algorithms, i.e., the proposed method is very compatible to other methods and has flexible structure.

9406-9, Session 3

Multi-polarimetric textural distinctiveness for outdoor robotic saliency detection

Shahid Haider, Christian Scharfenberger, Farnoud Kazemzadeh, Alexander Wong, D. A. Clausi, Univ. of Waterloo (Canada)

Saliency detection is utilized in applications where distinguishing unique items in a scene is important.

One such application is the area of mobile robotics, where robots that rely on vision, while navigating outdoors to detect and identify objects, utilize saliency approaches to identify a set of potential candidates to recognize. The state of art in saliency detection for mobile robotics often rely upon visible light imaging using conventional camera setups aiming to distinguish an object against its surroundings based on factors such as feature compactness, heterogeneity and/or homogeneity. These methods limit themselves only to what can be captured using conventional camera setups, which can be hampered by image saturation seen on sunny days, as well as detector insensitivity to slight differences in colour. To address some of these issues, neutral density filters have been placed on cameras for mobile robotics to remove bright specular highlights, but require longer exposure times and do increase the sensitivity to slight colour differences.

To remedy these issues for mobile robotics, one is motivated to incorporate different optical modes to capture additional useful information about the scene. In this work, we propose a novel saliency detection method that not only incorporates an additional mode for saliency detection, but is also not well-explored in literature: visible light multi-polarimetric imaging. The incorporation of multi-polarimetric imaging for saliency detection is motivated by the optical property of materials known as Fresnel reflections. By observing the scenes split in reflected intensity between multiple polarization states, we can infer the distribution of the refractive index and rely upon that in determining object saliency.

In the proposed multi-polarimetric saliency detection approach, we captured a visible light image along with multiple polarization states of a scene. Rotational-invariant multi-polarimetric textural representations are extracted from the captured multi-polarimetric imaging data and a high-dimensional sparse texture model is learned from the representations. The multi-polarimetric texture distinctiveness of the scene is characterized using a fully-connected graphical model based on the sparse texture model, which is then used to determine the saliency at each pixel of the scene along with general visual attentive constraints.

To evaluate the efficacy of the proposed multi-polarimetric texture distinctiveness approach for the purpose of mobile robotics saliency detection, images were captured of stationary objects with similar colour intensities as their surroundings under strong natural ambient light, which

is considered difficult for existing saliency detection approaches. Based on the captured images, the proposed approach was then compared to existing state-of-the-art saliency detection approaches. It was observed that existing saliency detection algorithms struggled with determining the saliency of the objects due to color intensity similarities between the objects and their surroundings. On the other hand, the proposed multi-polarimetric texture distinctiveness approach, by utilizing polarimetric information in its saliency detection framework, was able to provide noticeably improved saliency maps. As such, the proposed approach shows considerable promise in significantly improving the detection of salient objects under difficult scenarios often encountered in mobile robotics, and merits further investigation.

9406-10, Session 3

Semantic video segmentation using both appearance and geometric information

Jihwan Woo, Samsung Electronics Co., Ltd. (Korea, Republic of); Kris Kitani, Carnegie Mellon Univ. (United States); Sehoon Kim, Hantak Kwak, Woosung Shim, Samsung Electronics Co., Ltd. (Korea, Republic of)

Segmenting objects in the video plays a very important role in many computer vision tasks. Tracking or recognizing the object of interest or analyzing the video requires segmentation as the first step. Therefore, there have been many researches in the video segmentation. The current state-of-the-art researches use color, texture, depth, or motion as feature for segmentation. However, the objects in video are composed of multiple colors, textures, and depths. Also, motions of the object add more complication. These facts make the semantic video segmentation a complex and hard problem. Many researchers only considered a single feature for the video segmentation. For example, Super-pixel or K-means clustering techniques extract the color feature such as 3-dimensional RGB vector for the every pixel and group them with similar feature. With these algorithms, a multi-color painted wall will be segmented with several parts since it is composed of multiple colors or textures.

On the other hand, human will successfully segment the wall into a one coherent object very easily even if there are many colors. The human brain considers not only appearance such as colors or textures but also geometric features like motions and structures of the object at the same time. We propose the video segmentation algorithm which is motivated by the human visual system. Our method provides a unified framework of the appearance and geometric features for the video segmentation. The proposed framework is composed of 3 steps: 1) super-pixel extraction, 2) optical flow tracking for the super-pixel and 3) homography estimation between super-pixels in the consecutive frames.

In the first stage, we use Simple Linear Iterative Clustering (SLIC) super-pixel algorithm as a starting point. We extract the color histogram in every super-pixel as an appearance feature. Secondly, we track every pixel in the super-pixel from (i)th frame to (i+1)th frame with the Lucas-Kanade optical flow method. In this stage, super-pixel correspondences are made between two consecutive frames. Thirdly, we estimate the homography matrix between super-pixel correspondences. Every super-pixel in the video has its own homography matrix. We use the homography matrix as a geometric feature. If two super-pixels are in the same segment, they exhibit not only similar color histograms but also similar homography. Furthermore, we intentionally group super-pixels together when two neighboring super-pixels have similar homography matrixes even if there is some difference in the two color histogram. The weighted sum of color histogram distance and homography transform distance in the two adjacent super-pixels are used to build the affinity matrix. Finally, we apply the spectral clustering method to merge initial super-pixels for semantic video segmentation.

We have benchmarked proposed algorithm with other state of the art super-pixel based segmentation algorithms. The proposed segment shows better semantic segmentation result compared to the previous super-pixel based method. The proposed segmentation method will serve as a basis for better high-level tasks such as recognition, tracking and video understanding.

9406-11, Session 3

Feature matching method study for uncorrected fish-eye lens image

Baofeng Zhang, Yanhui Jia, Tianjin Univ. of Technology (China); Juha Röning, Univ. of Oulu (Finland); Weijia Feng, Tianjin Normal Univ. (China)

Fish-eye lens is a kind of short focal length lens. The filed of view (FOV) of fish-eye lens can reach or even exceed 185° in horizontal and vertical directions. From the viewpoints of mathematics, in spite of the serve distortion with fish-eye lens image, it still maintained the one-to-one mapping relationship between the object space and the image space. A large literature suggests that more and more researchers utilize two fish-eye lenses to expand the stereo FOV of traditional stereo vision in recent years.

An accurate feature matching is one of the key processes of 3D reconstruction and panoramic image generation. The further from the center of image the lower resolution and the severe non-linear distortion are the characteristics of uncorrected fish-eye lens image. To a large extent, these great difficulties restrict the applications of fish-eye lens in stereo vision and panoramic vision. The method generally adopted with the stereo vision system built by fish-eye lenses is to rectify the fish-eye lens image into perspective projection image, and then to make the epipolar lines of the corrected fish-eye images parallel, after that to match feature points through some local feature extraction algorithms like SIFT. However, some issues have been founded in previous study of the authors of this article:

1. Perspective projection model is not suitable for treating the rectification of fish-eye lens image. If the view angle is close to 180°, according to the perspective projection model, the size of the corrected image will be infinite. It is inevitable that some image information might be lost during the conversion of fish-eye lens image to perspective image;

2. In essence image rectification is a kind of space transform. The image coordinates is usually not an integer vector after the transformation. For the non-integer coordinates, interpolation algorithm needs to be used;

In conclusion, the approach which first corrects distortion, then matches feature points will destroy the geometrical relationship of stereo vision system and affect the authenticity of 3D reconstruction. It remains to make further research to demonstrate the scientificity and correctness of the general method. Therefore, this paper will explore a special feature matching algorithm for uncorrected fish-eye lens image to avoid the rectification processing.

Local binary pattern (LBP) is a texture measurement based on grayscale. It is demonstrated in literatures that LBP has high capability of grayscale invariance and rotation invariance. These characters make it possible to use for the fish-eye lens images as local feature descriptor. In this paper, the problem is solved by combining SIFT and LBP for feature detection and feature description. Firstly, detect and extract interest points by SIFT from the pair of fish-eye lens images. Following that establish the interest point set and the corresponding interest regions. Thereafter, LBP is used for describing the interest region. The similarities of these regions are evaluated by chi-square distance. Finally, the only pairs of feature points will be found. It is shown that the feature matching approach proposed achieves a satisfying matching performance in uncorrected fish-eye lens image.

9406-12, Session 4

Shape simplification through polygonal approximation in the Fourier domain

Mark Andrews, The Univ. of Auckland (New Zealand); Ramakrishna Kakarala, Nanyang Technological Univ. (Singapore)

As many papers have demonstrated, the bounding contour of an object suffices to detect, classify, or recognize the object within standard databases such as the MPEG-7 or ETH-Z. Current research focuses on shape descriptors derived from the contour, such as IDSC [1], SSC [2], or

broken contour Fourier descriptors [3]. However, present work does not provide an understanding of the relationship between Fourier descriptors of an entire contour and those of a simplified contour obtained through polygonal approximation. In this paper, we derive the theory and explore the applications of simplifying shapes into polygons through the Fourier domain. Our contributions and ongoing work are summarized below.

The main theoretical result of our paper connects the Fourier series coefficients of a contour with those of a polygonal approximation. Suppose a closed, simple contour is represented by a complex-valued function x(s), with s?[0,L) representing arc-length. Since x(s) is periodic, it has a Fourier series expansion

x(s)=∑_c_m exp?(j2ϖms/L).

If we sample x(s) at N evenly-spaced (in arc-length) points, and connect those points to form a polygon, then the polygon's contour may also be represented by a periodic complex-valued function z(t) with Fourier series

z(t)=∑_ d_k exp?(j2ϖkt)

In this paper, we prove the following result connecting the polygon's coefficients d_k with c_m, those of the curve:

d_k=sinc^2?(k/N) ∑__(m?k mod N) c_m . (sum over m = k modulo N)

This result clearly shows the frequency domain connection between the two shape representations. In particular, it is apparent that the curve's Fourier coefficients c_m are encoded and obscured in the polygon's coefficients d_k via the selective indexing of the summation. For bandlimited shapes the condition m?k mod N is solved approximately by k?m for large N, corresponding to polygons of high fidelity. For small values of N the condition m?k mod N results in many coefficients c_m being combined in a complicated form of frequency selective blurring, resulting in a low-fidelity polygonal rendition of x(s). Nevertheless, it provides a useful and unexpected means of assessing polygonal shape representations of more complex curves.

In ongoing work, we apply the polygonal approximation to simplify shapes, and examine the effects of simplification on shape classification accuracy. We are particularly interested in the connection between the order of the polygon N, the effective Fourier bandwidth of the curve, and the resulting recognition performance. Our results are compared to the state of the art using various descriptors.

References

[1] H. Ling and D. Jacobs, "Inner distance using the shape context", IEEE Trans Pattern Analysis and Machine Intelligence, Vol. 29, No. 2, pp 286-299, 2007.

[2] V. Premachandran and R. Kakarala, "Perceptually motivated shape context which uses shape interiors", Pattern recognition , Vol 46, No. 5, pp 2092-2102 2013.

[3] C. Dalitz, C. Brandt, S. Goebbels, D. Kolanus, "Fourier descriptors for broken shapes", EURASIP Journal on Advances in Signal Processing, Vol. 2013, No. 161 (18 October 2013)

9406-13, Session 4

Graph optimized Laplacian eigenmaps for face recognition

Fadi Dornaika, Univ. del País Vasco (Spain); Ammar Assoun, Lebanese Univ. (Lebanon); Yassine Ruichek, IRTES SeT UTBM (France)

Introduction:

In recent years, a variety of nonlinear dimensionality reduction techniques (NLDR) have been proposed in the literature. They aim to address the limitations of traditional techniques such as PCA and classical scaling. Most of these techniques assume that the data of interest lie on an embedded non-linear manifold within the higher-dimensional space. They provide a mapping from the high-dimensional space to the low-dimensional embedding and may be viewed, in the context of machine learning, as a preliminary feature extraction step, after which pattern recognition algorithms are applied.

Laplacian Eigenmaps (LE) is a nonlinear graph-based dimensionality reduction method. It has been successfully applied in many practical problems such as face recognition. However the construction of LE graph suffers, similarly to other graph-based DR techniques from the following issues: (1) the neighborhood graph is artificially defined in advance, and thus does not necessary benefit the desired DR task; (2) the graph is built using the nearest neighbor criterion which tends to work poorly due to the high-dimensionality of original space; and (3) its computation depends on two parameters whose values are generally uneasy to assign, the neighborhood size and the heat kernel parameter.

To address the above-mentioned problems, for the particular case of the LPP method (a linear version of LE), L. Zhang et al. [1] have developed a novel DR algorithm whose idea is to integrate graph construction with specific DR process into a unified framework. This algorithm results in an optimized graph rather than a predefined one.

Proposed approach:

In this paper we propose to modify the traditional LE NLDR technique by adding an optimized graph construction paradigm in a way similar to the one developed in [1] around the LPP method. The obtained method, called GoLE, is based on a learning iterative process during which the graph is gradually updated taking into account the transformed data. Thus, our model potentially reduces reliance on the parameter k of the nearest neighbor in the space of the original data. Figure 1 illustrates the proposed GoLE algorithm. The criterion to be minimized is the LE criterion to which a regularization term on the graph is added.

Inputs: X --- Data matrix organized as follows X=[x_1,x_2,...,x_n] with x_i?R^L ;

S_0 --- Initial weight matrix in original space;

H --- Regularization parameter;

? --- Iterative stop threshold

Outputs: Matrix Y such that: x_i ? (y_1 (i),y_2 (i),...,y_m (i))

Where {y_i } are the eigenvectors needed for embedding in the m-dimension subspace

Procedure:

Project the data X onto a PCA transformed subspace

For k=1,2,...,MaxIter

Calculate the optimal projection matrix Y using the classical eigenvalues problem:

L z=? z with L=D-S and Y=Z^T

Update adjacency weight matrix S

Calculate the objective function value of J_k=J(Y,S)

If |J_(k+1)-J_k |<?

Break and Return Y;

EndIf

EndFor

Fig. 1. GoLE algorithm

Experimental results:

We evaluate the performance of our proposed method by applying it to the problem of face recognition. Four face databases were tested: ORL, UMIST, Extended Yale B and PIE. Recognition experiments are conducted on these face databases using LE and GoLE methods respectively. Test faces are classified using the Nearest Neighbor classifier in the new space. The percentage of training data was set to 30%.

The average recognition rates (%) obtained with four face databases are 90.9, 92.7., 67.8, and 40.3 when LE is used. These rates become 94.2, 95.9, 73.1, and 41.5 when the proposed method is used.

[1] L. Zhang, L. Qiao et S. Chen, Graph-optimized locality preserving projections, Pattern Recognition, Vol. 43, Issue 6, June 2010, Pages 1993–2002

9406-14, Session 4

A superfast algorithm for self-grouping of multiple objects in image plane

Chialun John Hu, SunnyFuture Software (United States)

If we apply our developed LPED method to a binary image (e.g., binary IR image, binary color contrast image and binary color distribution image in visible wavelength range), we can obtain the boundary points of all objects embedded in the more randomly distributed noise background in sub-milli-second time. Then we can apply our newly developed grouping or clustering algorithm to separate the boundary points for all objects into individual groups with each group representing the boundary points of one object only. If it is followed by our fast identification-and-tracking software, we can automatically identify each object by its unique geometry shape and automatically track its movement simultaneously for N objects like we did before for two objects.

This paper will concentrate at the criteria and the algorithm design of this super-fast grouping technique. It is not like the classical combinatorial clustering algorithm in which the computation time increases exponentially with the number of points to be clustered. It is an algebra (or linear) time which increases only linearly with the number of the total points to be grouped. The total time for automatic grouping of 80-100 boundary points into separated object boundary points is only a few mili-seconds or less.

9406-15, Session 4

Research on the feature set construction method for spherical stereo vision

Junchao Zhu, Li Wan, Tianjin Univ. of Technology (China); Juha Röning, Univ. of Oulu (Finland); Weijia Feng, Tianjin Normal Univ. (China)

Stereo vision can obtain the depth information of scene by calculated the parallax of the same point in observing space, which is simultaneously observed by both cameras. Hence, a blind area exists, and the overlapped region will be more limited. Fish-eye lens usually have a wider-than-hemispherical FOV, it is an efficient way to extend the Stereo Field of View (SFOV). Spherical stereo vision is a kind of stereo vision system built by fish-eye lenses, which discussing the stereo algorithms conform to the spherical model. Recently, spherical stereo vision is more and more common in computer vision applications.

Accurate feature detection is one of the key processes of feature matching, three-dimensional reconstruction and panorama image generation. The great difficulties which restrict the applications of fish-eye lens include the further from the center of image the lower resolution and the severe non-linear distortion. At present, the commonly method for the applications of Spherical Stereo Vision is to rectify the fish-eye lens image into perspective projection image, and then to make the epipolar lines of the corrected fish-eye images parallel, after that to match feature points through some local feature extraction algorithms like SIFT. However, this solution has the risk of breaking the original geometry of stereo vision system and affecting the accuracy of three-dimensional reconstruction, which remains to make further research to demonstrate the scientificity and correctness. Therefore, the core aim of this article is to explore a special method of feature set construction avoiding the fish-eye image distortion correction and protecting the original geometric constraints between the images.

Maximally Stable Extremal Region (MSER) utilizes grayscale as independent variables, and uses the local extremum of the area variation as the testing results. From the aspects of repeatability, identification and detection speed, MSER, SIFT and Harris are the high-performance local feature detector in the report of Tuytelaars. It has been proved by various researches that MSER is characterized by affine invariance, repeatability and stability, and has well robustness under the situation of perspective transformation, partial occlusion and illumination changes. MSER fits almost all detection of shapes and fields, and the algorithm is easy to achieve. The complexity of the detection approach is nearly linear. It is demonstrated in literatures that MSER is only depending on the gray variations of images, and not relating with local structural characteristics and resolution of image.

The epipolar in uncorrected fish-eye image will not be a line but an arc which intersects at the poles. It is polar curve. The theory of nonlinear epipolar geometric will be explored, and the method of nonlinear epipolar rectification will be proposed to eliminate the vertical parallax between two fish-eye images. Next, MSER will be combined with the nonlinear epipolar rectification method. The intersection of the rectified epipolar and the corresponding MSER region is determined as the feature set of spherical stereo vision. Experiments show that this study achieved the expected results. A novel feature set construction method for uncorrected fish-eye images has been studied unlike the conventional.

9406-16, Session 5

Development of autonomous grasping and navigating robot

Hiroyuki Kudoh, The Univ. of Electro-Communications (Japan); Keisuke Fujimoto, Hitachi, Ltd. (Japan); Yasuichi Nakayama, The Univ. of Electro-Communications (Japan)

Cost reduction is one of the most important business challenges in warehouse management. The majority of operations in warehouses are related to warehousing and removing; thus, reduction of personnel expenses for those tasks has a great deal of meaning. In recent years, much effort has been made to improve efficiency, accuracy and labor productivity. For example, directed picking, which is a sub system of the warehouse management system (WMS) and automatic warehouse, moves shelves or containers with full of goods to the place where workers will pack them. These practices require an introduction of new equipment, and the latter might need rebuilding a warehouse; hence, the rebuilding cost is huge, and it is hard to install new equipment in existing facilities. We decided to solve these problems by replacing workers with robots that move automatically in warehouses, pick up goods, place them on a cart, and carry them to destination. We dealt this challenge with the picking up operation. The picking up operation in a warehouse is composed of locating the requested items, moving to where the items are located, finding the items on the shelf, picking up the items from the shelf, placing them on the cart, and carrying them to where they will be packed. To achieve these operations, we designed three functions of the robot as follows. The first function is an autonomous moving system. The robot is equipped with a 2D laser rangefinder and four wheels to implement this. The robot makes a map of the warehouse and estimates its location with distance information from the 2D laser rangefinder by using a simultaneous localization and mapping (SLAM) algorithm with normal distribution transform (NDT). Using this generated map and the estimated location, the robot finds a route to the shelf and follows it. The second function is a position recognition system. The robot is equipped with a 3D distance and image sensor to recognize the positions of items on the shelf. It obtains the rough positions from image data by using an image recognition algorithm with Haar-like features trained with an AdaBoost algorithm and obtains the accurate positions by matching the shape data of items to the 3D distance information using an iterative closest point (ICP) algorithm. The last function is the collecting function. The robot uses its hand and the recognized positions of items to plan a path to pick up the items from above or front and place them on a cart. We tested this robot in an experimental environment that simulates the warehouse. It achieves a series of operations: moving to a destination, recognizing the positions of items on a shelf, picking up the items, placing them on the cart by using its hand, and returning to the starting location. These operations are the routine works in warehouse; hence, these operations are required for robots to replace workers in warehouses. The results of this experiment show the possibility of replacing workers with robots.

Conference 9406: Intelligent Robots and Computer Vision XXXII: Algorithms and Techniques

9406-17, Session 5

Fine grained recognition of masonry walls for built heritage assessment

Noelia Oses, Zain Foundation (Spain); Fadi Dornaika, Univ. of Alberta (Canada); Abdelmalik Moujahid, Univ. del País Vasco (Spain)

INTRODUCTION

Our objective is to develop image-based tools that help automate the application of protocols for built heritage protection. In this work, we address the image-based extraction of arrangement of the masonry for automatic classification (See Figure 1). At first glance, one may think of applying automatic image-based granulometry techniques in order to delineate the individual stones and then infer the category from the geometric description of the delineated individual blocks. In our case, the fully automatic, image-based delineation of individual stones can be very challenging mainly due to the characteristics of the objects to be delineated (delineation, in the context of our work, refers to the extraction of the outline of the construction blocks). These objects (built heritage) are exposed to the elements and, as a result, they suffer discolouring, cracking, erosion, and can even have vegetation growing in them. The image capturing is performed in an uncontrolled environment with lighting conditions possibly changing for different captures. Images can, therefore, have bright and dark areas, and shadows depending on the sun's position. These undesired effects in the images generate intensity gradients that are, often, completely unrelated to the physical delineation of stone and, as a consequence, conventional edge-based methods are not useful since the signal-to-noise ratio is very low. This motivates the development of a delineation method that does not rely on conventional edge detection methods. Our proposed framework for the semi-automatic delineation and classification of built heritage has two phases. In the first phase, the test image undergoes a series of processing steps in order to extract a set of straight line segments from which a statistical signature is inferred. This set of straight line segments constitutes a partial delineation of blocks. In the second phase the statistical signature is classified using machine learning tools.

PROPOSED APPROACH

As stated earlier, the protocols will analyse many different features of each built heritage object type before reaching a conclusion. One of the features that can be characterised geometrically, and, thus, is a good candidate to be analysed automatically through digital image processing, is the type of masonry arrangement. This is the feature we have chosen to prove that the information obtained with our proposed automatic delineation framework is meaningful and can be used successfully in the automatic classification of this type of features. Three classes of masonry arrangement have been defined in these protocols: the first class are blocks (usually irregular) not arranged in rows (Figure 1.(a)), the second are irregular blocks arranged in rows (figure1.(b)), and, third, regular (rectangular) blocks arranged in rows (figure1.(c)). The masonry walls used in this experiment are located at different sites in the Basque Country: Durango and Urdaibai. The data set contains 86 wall images; 33 belong to Class 1, 15 belong to Class 2, and the remaining 38 belong to Class 3. Each raw image undergoes the proposed delineation steps that provide a set of 2D straight segments. This is carried out via local processing and mode based region extraction. Several statistics are extracted from this set of segments and are used as the predictor variables in the classifier. Twenty eight statistics are extracted from the set of straight lines.

EXPERIMENTAL RESULTS

We evaluated five classifiers adopting the feature selection by a wrapper method. Compared to the scheme that uses all features (28 features), the accuracy of all classifiers was improved. For example, the recognition rate of the 1-NN classifier increased by 12.8 % with respect to the result with 28 features. The rate of success classification (in %) of NN1, NN-3, Naive Bayes, Trees and SVM are 84.9, 83.7, 83.7, 86.0 and 87.2, respectively. We can observe that, following automatic feature selection, the SVM classifier provided the best performance.

9406-18, Session 5

Visual based navigation for power line inspection by using virtual environments

Alexander Ceron-Correa, Univ. Militar Nueva Granada (Colombia) and Univ. Nacional de Colombia (Colombia); Iván Fernando Mondragón Bernal, Pontificia Univ. Javeriana Bogotá (Colombia); Flavio A. Prieto, Univ. Nacional de Colombia Sede Medellín (Colombia)

Several countries spent efforts and resources in the development and implementation of technologies for power line inspection. It is important to make efforts in developing methods for improving different processes of electrical infrastructure inspection and maintenance.

Power line detection is an important task in the inspection of electrical infrastructure for maintenance. For this reason, it is valuable to develop methods that reduce costs, risks and the logistic problems of processes that involve human manipulation, including manned flights by using UAVs. UAVs can be used for capturing images from different views that could be processed in order to navigate autonomously and record images for the detection of failures in the electrical infrastructure. It is good to mention that there are companies that offer different services and products for electrical inspection, including UAVs, but the use of autonomous visual navigation is not yet provided.

For this reason, visual based navigation strategies for UAV power line inspection are presented; a virtual environment for real time simulation was developed and a set of line detection methods were integrated and validated within the virtual environment. The first strategy is related with the obtaining of the initial pose of the UAV with respect to the power lines. The second strategy is for navigating over the power lines. The navigation is performed by using the information extracted from a virtual camera in a visual control scheme.

The development of a UAV based power line inspection system requires to perform many UAV navigation tests. It is necessary to find alternatives to avoid damages to the UAV platform by doing several tests in environments that have similarities with the real ones. This can be done by developing virtual environments for simulating under different conditions.

For the development of this project, we decided to use a ROS gazebo simulator called Tumsimulator, that was developed at the Technical University of Munich, which permits to simulate different kinds of robots and its environments. This simulator can represent important aspects of the UAVs such as the sensor information and specially the camera, since it allows to develop and validate computer vision techniques for visual servoing schemes.

A visual feature based navigation for a multi-rotor UAV is proposed, implemented and tested by using simulation in virtual environments. Two strategies for multi-rotor navigation were proposed and they can be improved by using object detection for cluttered environments. Additionally, they can be extended for other areas.

A three dimensional model of environments with the power lines was built in order to generate synthetic images of power lines with different points of view. Different configurations of the scene were created for validating computer vision techniques successfully.

9406-20, Session 6

PanDAR: a wide-area, frame-rate, and full color lidar with foveated region using backfilling interpolation upsampling

Terrell N. Mundhenk, Kyungnam Kim, Yuri Owechko, HRL Labs., LLC (United States)

LIDAR devices for on-vehicle use need a wide field of view and good fidelity. For instance, a LIDAR for avoidance of landing collisions by a helicopter needs to see a wide field of view and show reasonable details of the

area. The same is true for an online LIDAR scanning device placed on an automobile. In this paper, we describe a LIDAR system with full color and enhanced resolution that has an effective vertical scanning range of 60 degrees with a central 20 degree fovea. The extended range with fovea is achieved by using two standard Velodyne 32-HDL LIDARs placed head to head and counter rotating. The HDL LIDARS each scan 40 degrees vertical and a full 360 degrees horizontal with an outdoor effective range of 100 meters. By positioning them head to head, they overlap by 20 degrees. This creates a double density fovea. The LIDAR returns from the two Velodyne sensors do not natively contain color. In order to add color, a Point Grey LadyBug panoramic camera is used to gather color data of the scene. In the first stage of our system, the two LIDAR point clouds and the LadyBug video are fused in real time at a frame rate of 10 Hz. A second stage is used to intelligently interpolate the point cloud and increase its resolution by approximately four times while maintaining accuracy with respect to the 3D scene. By using GPGPU programming, we can compute this at 10 Hz. Our backfilling interpolation methods works by first computing local linear approximations from the perspective of the LIDAR depth map. The color features from the image are used to select point cloud support points that are the best points in a local group for building the local linear approximations. This makes the colored point cloud more detailed while maintaining fidelity to the 3D scene. Our system also makes objects appearing in the PanDAR display easier to recognize for a human operator.

9406-21, Session 6

3D object recognition based on local descriptors

Marek Jakab, Wanda Benesova, Slovenska Technicka Univ. (Slovakia); Marek Racev, Slovak University of Technology (Slovakia)

Object recognition is one of the most challenging task in computer vision in the last years.

Methods using local descriptors became widespread and a lot of further improvements and modifications of known methods are in the focus of researchers now.

Our final task is to develop a fast and robust method of 3D object recognition which could be used for visual controlled self-checkout system in a shop; hence, our goal is to build a 3D object recognition system working in the real-time. For this purpose we extend the RGB image of the object with the 3D information of the object derived from the depth map. 3D information contributes to the improvement of the recognition accuracy and reducing the computational time, too.

In this paper, we present an enhanced method of 3D object description and recognition based on local descriptors using RGB image and depth map information of the object.

In the pre-processing step, RGB image and depth image acquired by the sensor Kinect are aligned and artifacts which may occur during the process are morphologically removed. The object is then segmented using the method of growing region in the depth map. To select a seed point for this method, we assume that the object which has been recognized, is the nearest object from the device. The segmentation is quite simple, but well working and brings noticeable performance improvement.

In the next step, we can start the SIFT features detector taking into account only the features which are not located on the border of the segmentation mask. Our real-time system includes already published implementation of the SIFT detector and SIFT descriptor calculated on the graphical processor unit (GPU). A saved template, which is corresponding to one of the learned objects, includes information composed from different viewpoints.

Our main contribution is focused on the extension of the SIFT feature vector by the 3D information derived from the depth mask and we also propose a novel local 3D descriptor which includes a 3D description of the key point neighborhood. As so defined, the 3D descriptor can then enter into the decision-making process.

Two different approaches:

- classification using an extension of the SIFT feature vector by the depth local information, for example: absolute value of the difference between the depth minimum and depth maximum in the local area, standard deviation of the depth value in the local area,

- classification using the original SIFT descriptor in combination with our novel proposed 3D descriptor have been evaluated, each by two strategies for the final decision.

Our dataset used in training and evaluation process is composed of 3D objects like toys, books etc.

First results show improvements of the recognition accuracy if we compare the recognition system which includes the 3D local description and the same system without any 3D local description. Our experimental system of object recognition is working near real-time.

9406-23, Session 6

The study of calibration and epipolar geometry for the stereo vision system built by fisheye lenses

Baofeng Zhang, Chunfang Lu, Tianjin Univ. of Technology (China); Juha Röning, Univ. of Oulu (Finland); Weijia Feng, Tianjin Normal Univ. (China)

Fish-eye lens is a short focal distance (f=6~16mm) camera. The field of view (FOV) of fish-eye lens is near or even exceeded 180?180 degrees. It can capture hemispherical image information of the observation spaced by one time shooting. A large number of literatures have proved that the multiple view geometry system built by fish-eye lens will get larger stereo field than traditional stereo vision system which based on the pair of perspective projection images. There are massive calibration research achievements on the stereo vision conformed by the pinhole imaging model. However, it has been rarely reported the calibration study for the special stereo vision based on fish-eye lens.

This paper focuses on discussing the method of the internal and external parameters calibration for fish-eye stereo vision. A geometric model for the fish-eye stereo vision will be constructed firstly. Then the relationships between the various physical parameters will be described by the universal fisheye mathematical model. Based on that, a specialized calibration method named Digitalized Fish-eye Stereo Calibration (DFSC) will be proposed to calibrate the parameters of fish-eye stereo vision.

At the experiment segment, two different situations have been tested separately to evaluate the performance of DFSC. One case is paralleled optical axis with two fish-eye lenses. The other is perpendicular. The relationships between the various physical parameters are described by the generic fish-eye mathematical model which can generalize the four projection models of fish-eye lens: equidistant projection, solid angle projection, stereoscopic projection, orthogonal projection. A planar calibration pattern with features circle has been designed for calibration. DFSC assumes that the camera internal parameters didn't change with the movement of the vision system and the target. Firstly, the features circles will be extracted from the images of planar calibration pattern. The images are taken by the fisheye lens stereo vision system in different locations. Next, establish the mathematical relationship between each feature circle on the planar calibration pattern and the point in the images. Finally, calculate relevant parameters. The experimental results verify that DSFC has obtained satisfactory calibration results to the applications of the multiple view geometry system built by fish-eye lenses.

The remainder of this paper is organized as follows: First, in Section 2, we introduce fish-eye lens and the multiple view geometry system built by fish-eye lens. Thereafter, the geometric models and the mathematical models for two different situations of fish-eye stereo vision are proposed. Section 4, the calibration method, DFSC, is explored to calibrate the internal and external parameters of fish-eye stereo vision. Finally, the experimental results and conclusions are presented in the last part, respectively.

Conference 9406: Intelligent Robots and Computer Vision XXXII: Algorithms and Techniques

9406-24, Session PTues

Intermediate view synthesis for eye-gazing

Eu-Ttuem Baek, Yo-Sung Ho, Gwangju Institute of Science and Technology (Korea, Republic of)

Video conferencing is a communication technology to connect users anywhere in the world as if they were in the same room. The technology has become very affordable and cut down on travel-related expenses. Interest of video conferencing and telepresence system tends to increase in the international companies in recent years. Therefore, the demand for video conferencing is expected to grow steadily.

The lack of eye contact seems to be the most difficult one among the problems in video conferencing. Mutual gaze is difficult due to the disparity between the position of the camera and the position of the eyes on the screen. It results in unapproachable and even unnatural interactions. In order to overcome the problems, previous approaches have tried to enhance mutual gaze. There have approaches to remove the unpleasant interactions using some methods such as stereo matching and 3D modeling. However, a stereo matching and 3D modeling are too slow, and real time methods are difficult due to the heavy computation required for matching algorithms. we do not use a dense stereo matching technique in order to run in real-time, because it takes long time to do a dense stereo matching, and holes around objects are created due to the limitation on camera positions.

In this paper, we propose the eye gazing correction system targeted at the enterprise video conferencing system that requires two cameras and a display. Our setup is composed of the 55 inch full HD television and the two cameras. The two cameras installed vertically, one on the top and the other beneath the bottom of the television. The camera is a CCD camera. The distance between the screen and a user is about 2m, and the distance between the two cameras is about 79cm. We separate our system into three main steps. The outline of the algorithm is as follows: The first step is a preprocessing. We capture two images from the two cameras simultaneously. We fine the camera calibration matrix using a chessboard. In order to improve performance, we extract a person from an image using background difference. The second step is the view morphing. We use the two original camera parameters to find the virtual camera configuration. The extracted two images are rendered with 2D warping at the virtual center position. We detect the facial feature points using Haar feature-based cascade classifiers and the proportion of head, and construct a Delaunay triangulation. We apply 2D-to-2D inverse affine transformation in order to eliminate holes and overlaps. Finally, we synthesize the morphed face and the background. To be shown seamlessly, the contour around the morphed face is blended. The result of the synthesized image show that eye gazing is corrected and the image was synthesized seamlessly.

9406-25, Session PTues

Increasing signal-to-noise ratio of reconstructed digital holograms by using light spatial noise portrait of camera's photosensor

Pavel A. Cheremkhin, Nikolay N. Evtikhiev, Vitaly V. Krasnov, Vladislav G. Rodin, Sergey N. Starikov, National Research Nuclear Univ. MEPhI (Russian Federation)

Digital holography is technique which includes interference pattern recording with digital photosensor, processing of obtained holographic data and reconstruction of object wavefront. Increase of signal-to-noise ratio (SNR) of reconstructed digital holograms is especially important in such fields as image encryption, pattern recognition, static and dynamic displaying of 3D scenes, and etc. The method of increasing of SNR of reconstructed digital holograms by using light spatial noise portrait (LSNP) of camera's photosensor is presented.

LSNP is array of photosensor pixels photo response non-uniformities. It is mainly used for camera identification, determination of images origin and any post-processing done. Typical spatial noise value is about 0.5 % of camera signal value that is 2-4 times less than temporal noise. So use of the proposed method is effective after application of others methods of SNR increasing that suppress temporal noise. We investigated application of the LSNP compensation method in conjunction with popular averaging over frames method. Earlier the method of registered image SNR increasing by spatial noise suppression using compensation of photosensor LSNP was proposed. In this paper compensation of photosensor LSNP for increase of SNR of reconstructed digital holograms was proposed.

To verify the proposed method, numerical experiments with computer generated Fresnel holograms with resolution equal to 512x512 elements were performed. Propagation of light field from object plane to hologram plane and from hologram plane to reconstruction plane was calculated using direct calculation of Fresnel diffraction. The method of simulation of shots registration with digital camera is described. It takes into account measured noise and radiometric parameters of digital camera. Simulation of shots registration with digital camera Canon EOS 400D was performed.

It is shown that solo use of the averaging over frames method allows to increase SNR only up to 2 times, and further increase of SNR is limited by spatial noise. Application of the LSNP compensation method in conjunction with the averaging over frames method allows for 10 times SNR increase. This value was obtained for LSNP measured with 15-20 % error. In case of using more accurate LSNP, SNR can be increased up to 50 times.

9406-26, Session PTues

Camera calibration based on parallel lines

Weimin Li, Yuhai Zhang, Yu Zhao, Univ. of Science and Technology of China (China)

No Abstract Available

Conference 9407: Video Surveillance and Transportation Imaging Applications 2015

Tuesday - Thursday 10–12 February 2015

Part of Proceedings of SPIE Vol. 9407 Video Surveillance and Transportation Imaging Applications 2015

9407-33, Session PTues

Person re-identification in UAV videos using relevance feedback

Arne Schumann, Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (Germany)

With cheaper and cheaper recording hardware the amount of available surveillance video data is rapidly growing. The limiting factor in making use of this data is usually the not only the personell cost for human operators but also their limited attention span when reviewing the data. It is thus important to find ways to support operators by solving relevant tasks in an automatic or semi-automatic way. One important task in the context of surveillance is that of person re-identification. If a certain person is of interest to the operator, computer vision may help find other occurrences of that same person elsewhere in the data more quickly.

Person re-identification is well known to be a challenging problem. In the typical surveillance scenario inside a camera network approaches have to deal with challenges such as varying color footprints of cameras, lighting effects, image noise and varying viewpoints. In this work we investigate methods to approach the problem for aerial video data. In this even more challenging setting, the difference in viewing angles is much larger and may change over time (due to motion of the camera and the persons), camera motion has to be dealt with and, depepending on the position of the recording unmanned aerial vehicle (UAV), scale of persons may change significantly. While the latter two of these problems mainly impact the task of person detection, the former is a significant problem that needs to be addressed during re-identification.

As a baseline we investigate the performance of a number of image features in the aerial setting that are known to perform well for person re-identification in static surveillance camera networks. Such features usually focus on either color or texture information. Our feature pool includes color histograms, color structure descriptors, sobel and gabor filters, local binary patterns and discrete cosine transform features. We use these features in combination with established feature matching strategies to establish an automatic baseline approach.

We extend this automatic baseline by incorporating the operator in an iterative manner. Given a query person, the approach first generates a list of possible matches automatically using features from the pool. The operator then marks retireved matches as correct or incorrect and this feedback is used to generate a new list of matches that more closely resembles the operator's imressions. This semi-automatic procedure can be repeated for multiple iterations until a satisfying result is reached.

We use this feedback in multiple ways. First, we limit selection of features from the pool to only those features that most reliably produced results that were marked by the operator as correct. The selection of features is thus adapted to focus on those visual cues that the operator deems most discriminative. Second, we train a simple appearance classifier using the operator feedback which learns to distinguish the query person from others. Finally we use the feedback from previous queries (of other persons) to further restrict the search space and generate more hard negative examples which in turn can further improve the classifier and feature selection. We are thus able to improve performance not only over multiple feedback iterations of the same query but also over multiple uses of the system.

While traditionally the topmost returned matches are labelled by the operator, those may not be the best choice to imrpove accuracy in the next iteration. We also investigate active learning strategies tell the operator which results should be labelled instead of letting him make that choice. Such strategies include selecting those matches that the approach is most uncertain about, committee-votine where those matches are selected that the features and the classifier least agree upon or selecting those matches that cause the largest variance in the output.

To the best of our knowledge there is no video dataset that is recorded by a UAV and allows for person re-identification at the same time. We thus recorded our own dataset. The data contains a group of volunteers walking through a scene recorded by a low-flying UAV. No specific instructions for movement patterns were given. The data contains small groups of people as well as single persons. Each person appears multiple times to allow for the evaluation of re-identification approaches.

We evaluate the described baseline approach as well as the various improvements and demonstrate significant improvements in accuracy over multiple iterations of operator feedback. The dataset also contains stationary cameras on the ground which allows us to further evaluate re-identification between ground and aerial views. We also compare to the performance of a previous approach and show that re-identification performance was improved.

In conclusion, we present a person re-identifcation approach can handle the added challenges of aerial video data and integrates operator feedback in multiple ways. We demonstrate the performance of the described approach on a newly recorded dataset and show improvements over the baseline as well as a previous approach.

9407-34, Session PTues

Aerial surveillance based on hierarchical object classification for ground target detection

Alberto Vazquez-Cervantes, Juan Manuel García-Huerta, Teresa Hernández-Díaz, J. A. Soto-Cajiga, Hugo Jiménez-Hernández, Ctr. de Ingenieria y Desarrollo Industrial (Mexico)

Day to day, surveillance systems has become more common and they are considered as a cheaper way to supervise, monitor and take decisions over certain scenarios. This situation makes reliable the use of cameras as cheaper sensors in other devices. This is the case of unmanned aerial devices, such as drones. Drones have become more versatile to develop automatic and cheaper surveillance system due to the level of integration and wireless communication interfaces. In this sense, one common task in surveillance and monitoring systems is the location of objects in open scenarios. Localization of objects is a tough task because there are several non-controlled situations such as camera jittering, luminance pollution, occlusions, or geometry distortions due to camera projection. Regarding to these situations several approaches in the state of art, might not be robust in all scenarios. For instance, tracking and detection algorithms assume several scenario restrictions; others like, assumes fixed camera or previously knowledge of scenario. On the other hand, robust approaches like works well, at the cost of an increment on computational complexity. However, in aerial drones, both situations are not adequate. This is, if the algorithm is too simple, it will not locate objects under several affectations; instead if it is robust then the computational complexity will grow, dealing with low the battery lifetime, which affects directly to the drones autonomy.

Situations commented above, show the need to develop approaches with low complexity, but high reliability. This paper proposes a new algorithm to detection-by-example. In first place we provided a target objective to be located, after a sampling itinerary performed by unmanned drone with HD frontal camera. Visual evidence is acquired, analyzing sequences of frames. Likewise, the drone take as reference its internal GPS. Drone vehicle has an embedded system based on A8 Cortex with 4 cores. The detection algorithm is based in a dynamical parallel process, in which, most of the time behaves as a low complexity algorithm; i.e. algorithm make a hierarchical process of analysis. The top process has low complexity, which increase as it turns into the bottom process. The top process is used in general analysis and selecting candidate zones as possible target objects. Inferior hierarchical level uses more features in order to discriminate

the candidate from the target object. Deepness hierarchical analysis represents with high certainty candidate objects' related with the search objective. A top-down process is performed dynamically, resulting in a variable complexity, avoiding over-use of embedded resources. In terms of application, the algorithm complexity is expressed as an optimization of hardware architecture with the aim to an energy-efficiency.

The hierarchical approach uses a set of random features extracted from the target object. A random feature is the result of encoding as a binary string a set of features uniformly sampled from the target. Encoding process represents distinctive zones as binary strings. Distinctive zones are small-window-features under a transformation F, such that it is invariant to several image affectations; for our purposes, a tensor matrix from gradient of image is used. Binary representation have several properties to develop classifiers in huge dimensionalities; i.e. concepts like orthogonally, string mixing, or string superposing, under L1 metric and X-or rings, allow to develop a sparse associative field.

Associative field might define a process to make associations between encoding features with the target object in a hierarchical structure. The hierarchical structure is implemented as a similarity order relation between binary strings related with target features. The process of choice when evidence belongs to similar target is appreciated as a path into the tree. Tree position after test candidate evidence defines the level of similitude between target and a candidate. An acceptance criterion is used to define how deeper will be an adequate similitude among candidates and the target.

Finally, this approach is tested in controlled/not-controlled scenarios with different target objects. Controlled scenarios such as searching in a laboratory. Not-controlled scenarios as it is a basketball court. Besides, the search targets include textured geometrical objects. Additionally, tables of recognition ratio are showed and the analyses of resources used trough time are presented.

9407-35, Session PTues

A novel synchronous multi-intensity IR illuminator hardware implementation for nighttime surveillance

Wen Chih Teng, Jen-Hui Chuang, National Chiao Tung Univ. (Taiwan)

Abstract—Vision-based surveillance systems have gained increasing popularity. However, their functionality is substantially limited under nighttime conditions due to the poor visibility and improper illumination. With fixed-intensity, traditional infrared (IR) illuminators only work for a certain distance, resulting in undesirable imaging effects of overexposure / underexposure when an object is too close to (too far from) the camera. To overcome such a limitation, a novel multi-intensity IR illuminator (MIIR) is employed to extend the effective depth of field (DOF) for foreground detection in nighttime surveillance videos. By splitting a video into different channels (for different illumination levels), well-exposed images for far or near objects can then be selected among these channels. Accordingly, an effective algorithm of channel selection is proposed in this paper, which takes into account (i) stability of the foreground bonding box and (ii) degree of saturation of foreground image. Experiments on real nighttime surveillance videos show that the effective DOF can indeed be extended compared with the traditional infrared (IR) illuminators if both (i) and (ii) are considered. The usage of MIIR is expected to significantly enhance background/foreground detection in existing nighttime surveillance systems, and consequently promote the public security.

9407-1, Session 1

Road user tracker based on robust regression with GNC and preconditioning

Andreas Leich, Marek Junghans, Karsten Kozempel, Hagen Saul, Deutsches Zentrum für Luft- und Raumfahrt e.V. (Germany)

A novel tracking algorithm is introduced and tested on real world video image data of road users.

The algorithm relies on the evaluation of image intensities of all pixels in some predefined image region (region of interest – ROI) over consecutive images. A motion model is being fitted to the pixel data so that a cost function is in its minimum, when the model parameters fit the dominant motion in the image region. This approach is known in literature as the robust regression based approach to motion estimation. In contrast with least squares regression based motion estimation, robust regression promises improved dealing with outliers in the data.

In the case of robust regression, the surface of the cost function is multimodal in general, especially, when the ROI contains more than one moving objects. In order to address this issue, in contrast with widely adopted approaches that rely on random sampling (RANSAC), in this paper we investigate the graduated non convexity (GNC) approach, introduced by Blake and Zisserman in 1987. The idea of GNC is to do a scale space analysis of the parameter space where the cost function lives. Numerous attempts to apply GNC for finding the minimum of the cost function of a robust regression problem have been reported in literature without being able to present satisfactory results.

In this paper, we present an analysis of the reasons why prior approaches failed. Besides some minor traps and pitfalls that occur when designing a robust regression GNC algorithm, there is one main problem. We call this problem the condition problem. The condition problem occurs when a local minimum of the cost function is better conditioned than the global one. The condition problem occurs in almost every practical relevant dataset and causes the GNC algorithm to get stuck in a local minimum. We present a preconditioning method as solution for the condition problem. Finally we introduce an algorithm which improves the effectiveness of robust regression based GNC.

We analyze the effectiveness of the method both on synthetic data with randomly sampled outliers and in real world data within an experimental real time tracking system. The results on synthetic data show, that the proposed method stably works when the number of uncorrelated outliers exceeds 90%. For a data set with 90% outliers it is hard to see the ground truth even for a human observer.

The experimental setup is designed as follows: When the road user passes some entry ROI, then this loop emits a presence and classification signal using the state of the art Histogram of oriented Gradients algorithm (Hog). This signal triggers the road user tracker, which begins tracking the object until it leaves the image. In our implementation, a four-parametric motion model is fit to the data. The model describes the translation and zoom parameters of the target. The algorithm can be extended for an affine motion model. The algorithm precisely tracks the target in the road scene in real time. It continues tracking over some hundreds of meters.

9407-2, Session 1

Vehicle detection for traffic flow analysis

Vimal Varsani, Soodamani Ramalingam, Univ. of Hertfordshire (United Kingdom)

As technology advances and high quality video cameras become inexpensive, there has been a rise in automated video analysis through video processing in recent years. The increase in the processing power of computers in conjunction with their memory, has led to a growth of automated systems for video analysis that comprise of object tracking, which is an important field in computer vision. There has also been increased research in object tracking systems for military use. Combination of both these methods has led to numerous applications such as surveillance, teleconferencing, traffic monitoring etc.

Object tracking has been a major research topic in the past years with its applications stretching from basic statistics to defence. This report looks at some of the algorithms that can be used for tracking cars in particular for statistical analysis. The main methods for tracking discussed and

implemented are blob analysis, optical flow and foreground detection. A further analysis is also done testing two of the techniques using a number of video sequences that include different difficulties. The main of the project was to come up with an algorithm that performs tracking effectively. Object tracking is the process of automatically locating an object from one frame to another, in a video sequence or in an image. There are however some problems involved with tracking objects, such as; difficult object shapes that need to be tracked, objects moving randomly, noise in the video that needs to be filtered etc. A few of the applications for object tracking include;

• Automated video surveillance: This involves identifying certain moving objects in an area under surveillance and detecting anything suspicious. The system also needs to differentiate between different objects including animate and inanimate objects in the background.

• Monitoring traffic: Traffic is continuously monitored, and vehicles breaking certain traffic rules are tracked down and some advanced systems can also track road accidents, pedestrian activity etc.

• Human computer interaction: Tracking the eye movement of a person, or recognizing a face or other human gestures.

9407-3, Session 1

Vehicle speed estimation using a monocular camera

Wencheng Wu, Vladimir Kozitsky, Martin Hoover, Xerox Corp. (United States); Robert P. Loce, PARC, A Xerox Co. (United States); D. M. Todd Jackson, Xerox Corp. (United States)

Vehicle speed is a key traffic measurement required in an Intelligent Transport System (ITS). It is relevant to traffic flow and can be used for accident prediction or accident prevention etc. For example, instantaneous measurement of vehicle speeds at an intersection may be used to alter the duration of traffic signals temporarily to prevent accidents from occurring. Common sensors for vehicle speed measurement include inductive loops, radar, lidar, and stereo video cameras. There are several advantages that a monocular vision system can provide over alternatives. It is more cost effective, monocular imaging systems are widely available, are easier to install and maintain, and can serve other purposes such as surveillance. At a glance, it may appear quite simple to provide some measure of speed of an object using video cameras if the object of interest is properly detected/identified, and tracked. Indeed, much work has been done in this area but most focus on measuring average speed of vehicles. The issue is the achievable accuracy and precision of speed measurement of an individual vehicle when applying computer vision techniques using a monocular camera. We discuss causes of these issues in this paper.

The objective of this paper is to describe a speed estimation method for individual vehicles using a stationary monocular camera. The system includes the following: (1) object detection, which detects an object of interest based on a combination of motion detection and object classification and initializes tracking of the object if detected, (2) object tracking, which tracks the object over time based on template matching and reports its displacement frame to frame in pixels, (3) speed estimation, which estimates vehicle speed by converting pixel displacements to actual distances traveled along the road, (4) object height estimation, which estimates the distance from tracked point(s) of the object to the road plane, and (5) speed estimation correction, which adjusts previously estimated vehicle speed based on estimated object and camera heights

More specifically, our method starts by detecting an object of interest (i.e., license plate) within the camera field of view and then initializes the tracking accordingly. Although there exist many image-based license plate detection and/or recognition (LPR) algorithms, it is not necessary for this application to decipher the license plate information at this stage. For efficiency, we defer LPR to a later stage. Instead, we utilize a combination of motion detection and object classification to identify possible license plates. We first apply double-frame-differencing on frames at reduced spatial resolution to detect regions of interest (ROIs) indicating objects in motion. A pre-trained classifier is applied to each ROI to determine whether this ROI represents a license plate. If so, tracking of the detected plate is initiated.

Once a license plate is detected, our method tracks top left and right corners of the plate via template matching until the plate exits the scene. The initial templates are extracted when the plate was first detected. The templates are updated frame to frame in the tracking step to account for pose changes of the tracked vehicle. This is important for maintaining the tracking and for speed estimation. There are also interactions between tracking and object detection since not every object detected in a frame is a new object and not every object tracked is always detected in a frame. We will discuss in this paper how our method deals with these interactions in an efficient manner.

Next, an approximate speed of the tracked vehicle is estimated. Since the output of the object detection and tracking steps is a trajectory of license plate in pixel units, a conversion of units is required for computing vehicle speed. This conversion is often referred to as camera geometric transformation, a well-known art but has its drawback when applied to real world transportation settings. A rough speed estimate can be calculated based on the conversion of pixel to real-world coordinates and the frame-rate. We will discuss in the paper why this is not sufficient and how our method addresses that with a process to estimate height of each tracked plate. We thus develop a novel camera characterization method to meet this need.

Finally, we perform speed estimation correction, which adjusts previously estimated individual vehicle speed based on estimated plate height and camera height. The camera height is estimated once through our proposed camera characterization method. The height for the plate of each individual vehicle is estimated at run-time based on a priori knowledge of plate dimension and measured spatial characteristics of the camera. This height represents the distance between the tracked object/feature to the road plane, which is a critical measurement needed for accurate speed measurement using a monocular camera. It is the key to recovering missing information due to 3D-to-2D imaging. This is why we choose to track license plate rather than other vehicle features. Once the height of tracked feature(s) is estimated, the correction of speed measurement is simply a fractional correction based on the ratio of plate height to camera height. We demonstrate the effectiveness of our algorithm on 30/60 fps videos of three hundred vehicles travelling at speeds ranging from 30 to 60 mph. The 95-percentile speed estimation error of our method across the test set was within ?3% when compared to a lidar-based reference instrument.

In summary, our method relies on the detection and tracking of license plate, the estimation of the height of the plate above the road plane, and the knowledge of the camera height to achieve high accuracy of individual vehicle speed estimation. Key contributions of our method include the developments of (1) tracking a specific set of feature points of a vehicle to ensure consistent measure of vehicle speed, (2) a high accuracy camera characterization method, which does not interrupt regular traffic of the site, and (3) a plate and camera height estimation method for improving the accuracy of individual vehicle speed estimation. Additionally, we examine the impact of spatial resolution to the accuracy of speed estimation and utilize that knowledge to improve the computation efficiency of our algorithm. For example, we use relatively low spatial resolution for motion detection while using high spatial resolution for object classification and tracking. We also improve the accuracy and efficiency of our tracking over standard methods via dynamic update of templates and predictive local search.

9407-4, Session 1

Detecting and extracting identifiable information from vehicles in videos

Siddharth Roheda, Nirma Univ. (India); Hari Kalva, Florida Atlantic Univ. (United States); Mehul Naik, Nirma University Institute of Technology (India)

Traffic monitoring using cameras has become common in transportation and security applications. License plate recognition has been studied and a number of papers have been published on this [1, 2, 3]. Vehicles can be identified not only by a license plate but other information such as make,

model, color, and other identifiable traits such as bumper stickers.

License plate segmentation and recognition has been studied and a number of papers have been published. A key task of these systems is the separation of the license plate from the image. Some of the papers, [1] and [2], talk about segmentation of the plate using boundary/edge information. Others, [3], have used the texture features of the license plate to locate and segment it out. That is, they would look for a significant change in the gray level to detect the presence of characters. It also results in a high edge density area due to color transition. The approach used in [4] is to horizontally scan the image, looking for repeating contrast changes on a scale of 15 pixels or more. It assumes that the contrast between the characters and the background is sufficiently good and there are at least three to four characters whose minimum vertical size is 15 pixels. In [5], a contour detection algorithm is applied on the binary image to detect connected objects. The connected objects that have the same geometrical features as the plate are chosen to be candidates. This algorithm can fail in the case of bad quality images, which results in distorted contours.

We are developing a system to automatically describe a vehicle as a person would; e.g., red, Toyota tundra, with a bumper sticker "peace", and license tag 123-ABC. Such a system can help surveillance, improve analytics, and also allow encrypting these identifiable regions to ensure privacy.

The first step of the process is segmenting multiple cars in a video frame. Once the cars have been separated from the frame, we try to detect and segment out the areas which are needed in order to have a detailed description of the car. This means, all the areas that have text in them need to be segmented out. These would be the areas of license plate, bumper stickers, the make and model of the car or the logo of the company. To detect these parts we look for connected areas in the form of horizontal rectangles. Further, the gradient of the detected parts is checked so that the parts without texts can be eliminated.

The cropped regions are processed further to identify the license plate, make-model information, and bumper sticker if any. In order to do this, we use the fact that every license plate has information about the state of registration. If we are able to extract the information from one of the cropped image, we can classify this regions as a potential license plate. To further confirm that it is a license plate, we look for vertical rectangles instead of horizontal ones as before. Now, we check for the height to width ratio of the characters of the number plate. Each state has a certain standard of character sizes and if the ratios fall in a pre-determined range, the image is classified as a license plate. Once we have the license plate image, we use character recognition in order to obtain the license plate number. We are currently working on methods to detect logos using SIFT descriptors. This descriptor based matching can also be used to detect make and model information that is difficult to process with OCR techniques.

The system work with videos capturing the rear of the car and with all identifiable information clearly visible. The proposed solutions has very low complexity as it uses structure of the license plate and relative text size to simplify detection. The method tends to fail in situations where there are too many stickers, overlapping stickers, or when the resolution of the camera is too low. In such a case, though the license plate may be detected, it may not be possible to detect the license number due to bad resolution.

A total of 115 images were processed to evaluate the performance of this method. These were images that were taken from stationary traffic cameras, dashboard cameras and handheld digital cameras. The dataset had a mixture of very low resolution to high resolution pictures and also having varying clarity.

Out of these 115 images, in 92.17 % cases, license plate was successfully segmented out of the image. 82.92 % of times, the make and the model of the car could be segmented out and in 92.5% cases the stickers if present on the car could be segmented out. Once the segmentation of texts from the image was complete, the system was able to successfully able to extract the state information from the license plate in 87.34 % cases.

9407-5, Session 1

Electronic number plate generation for performance evaluation

Soodamani Ramalingam, William E. Martin, Talib A. A. S. Alukaidey, Univ. of Hertfordshire (United Kingdom)

Work on Automatic Number Plate Recognition (ANPR) as part of road safety by detecting and deterring a range of illegal road users is currently a key research work undertaken by the School of Engineering and Technology in collaboration with the UK Home Office and Hertfordshire Constabulary. A key research problem identified is the lack of an objective and independent assessment process for benchmarking ANPR systems in the UK. With this is mind, a simulation process was recently set up to generate car number plate images. As a first step, such plates show variability in character spacing for assessing ANPR systems which demonstrate the principles for benchmarking. This process avoids the need to manufacture a large number of physical licence plates, and the need for carrying out any resource intensive field trials by Law Enforcement Agencies which is currently the case. As an extension to the above work, it is proposed that an electronic licence plate be developed simulating a transparent display with a retro-reflective film fixed behind it. High transparency and high contrast were key requirements to be met. Further, the paper shows how these images may be used to determining SNR ratios for purpose of benchmarking ANPR algorithms.

The proposed algorithm simulates the physical characteristics of a number plate electronically and meets the requirements as specified by the British Standards BS AU 145d now in the process of being revised to BS AU 145e. This includes the physical dimensions of the plate and the characters that constitute the vehicle registration mark. It also follows the character set compliance as well the group combinations within it in accordance with the UK Driver and Vehicle Licensing Agency (DVLA) requirements. The reflectance properties of the plate are modelled by the opacity of the synthetic image and so are other conditions of weather, and speed of the car and the impact that has on the image capture.

The simulation process generates data sets for benchmarking commercial ANPR systems. Many tests in the performance assessment standard require the use of a large number of licence plates having different registration numbers and different synthesized defects (such as fixing bolts, de-lamination etc). When the electronic plate is viewed with an ANPR camera (visible or infra-red), the captured image should be, as far as possible, indistinguishable from the image of a physical licence plate bearing the same registration number and synthesized defects when viewed by the same camera. A metric is formulated to verify the quality of the plate generated.

As a final step, the images are subjected to ANPR algorithms to determine a benchmarking performance evaluation measure.

9407-6, Session 1

Efficient integration of spectral features for vehicle tracking utilizing an adaptive sensor

Burak Uzkent, Matthew J. Hoffman, Anthony Vodacek, Rochester Institute of Technology (United States)

There have been an extensive number of studies related to the air-to-ground object tracking problem, most of which rely on only kinematic information to accomplish target identification. In challenging cases this can lead to misidentification. With recent advancements in sensor technology, it is possible to incorporate discriminative spectral target features in addition to the kinematic information. Due to the high cost of spectral data, we cannot collect and process spectral data of the full scene. Therefore, a target tracking system can highly benefit from an adaptive sensor such as Rochester Institute of Technology Multi-object Spectrometer (RITMOS) that collects spectral data at only desired locations. This way

we can avoid collecting and processing a large amount of redundant data. RITMOS is also capable of acquiring a panchromatic image of a full scene. It employs a micromirror array to reflect light from individual pixels to either spectroscopy or imaging sensors. It cannot collect data in different modalities simultaneously; however, the switch from panchromatic to spectral for a pixel is achieved quickly due to the speed of micromirror arrays.

The first fundamental step in the tracking system is detection, as objects of interest must be observed to update the prior estimates. We use panchromatic images to perform detection and change detection, a well-known background subtraction method, is implemented for this purpose. Once moving objects are detected, spectral data is used for target identification to determine which detected object best matches the target of interest. Here, we note that the similarity score of the best match must be higher than a pre-defined threshold or the target is assumed to be lost in the frame. Comparison of the spectral features collected at the current time step with the spectral features of the known targets is achieved using a spectral distance metric.

The adaptive sensor is tasked to collect spectral data based on the location and covariance information provided by the filtering algorithm and its forecasting system. The sensor must be tasked to take spectral observations before targets are detected in a particular time step, so the usefulness of collected data depends on the performance of the tracker and the predictions. One will end up with large amount of background data when the tracker performs poorly. This is especially true in the intersections, where vehicles often perform motions that have nonlinear probability distribution functions (pdf). To avoid background data collection, we propose the incorporation of intersection masks that are generated from an external source to adaptively tune the filter to improve tracking in intersections. Some background data will still be observed with the target, so we propose a background data elimination method to minimize the effect of these spectra on the identification process. This method collects spectral features from the target's predicted future location to build a background database for that track. At the future time step, we compare the new features with the previously-built background database using a difference metric. Finally, we remove the spectral features that match the background database while keeping the remaining spectra.

In the filtering section, we apply a Gaussian Sum Filter (GSF) that employs the Extended Kalman Filter (EKF) to propagate the Gaussian components. The GSF approximates the target's pdf using a mixture of normal distributions. The GSF was chosen over other conventional methods due to its capability of accommodating large number of motion models and its low complexity. The EKF performs two updates on each Gaussian component: prediction and measurement. In the prediction update, a pre-defined motion model is applied to each component. Here we apply one of four different motion models to each component: constant velocity (CV), coordinated turn (CT), constant acceleration (CA) and stop model. Sticking with a fixed forecasting multi-modal set which contains same models regardless of the environment the target is interacting with in different scenarios can degrade tracking and identification accuracy. To improve this, we propose an adaptive multi-model forecasting framework that incorporates context based information on intersection geometry that is extracted from an online source.

Ground vehicle movement is usually heavily constrained by the surrounding environment. Vehicles travel on roads so in an intersection a vehicle typically has only several possible paths. This contextual information is useful because a vehicle changes its motion dynamics dramatically in order to execute a turn which results in a complex movement. Since we may not have prior information on which direction the target is turning, we might not have the best motion model set for this scenario. By incorporating an intersection mask of the area, we can adaptively switch to a better representative forecasting model set for a particular intersection. By doing so, we can collect more useful spectral data in the intersection and increase the tracker's performance. For instance, assume that in a T-intersection a target can only turn left or right. Once the target enters this intersection, we switch to a model set that is more strongly weighted towards CT models while reducing the number of the CV and CA models used for forecasting the components. The intersection mask is extracted by using data from OpenStreetMap which is a free, crowd-sourced project aiming to build a geographic database of the world.

We test the proposed tracking system by generating realistic and challenging scenarios using the Digital Imaging and Remote Sensing Image Generation (DIRSIG) model. In addition, the Simulation of Urban Mobility (SUMO) platform is used to generate realistic traffic simulations on the DIRSIG generated frames. To estimate the tracking accuracy, we use the Root Mean Square Error (RMSE) metric and the Track Purity metric is used to measure the efficiency of the target identification process. Overall, we have tested several scenarios with different types of intersections and varying level of challenge. The results demonstrate that we achieve persistent tracking with high Track Purity and RMSE accuracy by only using %1.5 of the data volume of a full hyperspectral camera.

9407-7, Session 2

Detection and recognition of road markings in panoramic images

Cheng Li, Ivo M. Creusen, Lykele Hazelhoff, CycloMedia Technology B.V. (Netherlands) and Technische Univ. Eindhoven (Netherlands); Peter H. N. de With, Technische Univ. Eindhoven (Netherlands) and Cyclomedia Technology B.V. (Netherlands)

INTRODUCTION

The government is responsible for the maintenance of road markings, since deterioration of road markings can lead to a decreased road traffic safety. Ideally, the condition of road markings should be periodically monitored. Performing this monitoring task manually is too labor intensive and too costly, therefore an automatic or semi-automatic solution would be preferable.

Most previous work on road marking detection has been performed in the context of Advanced Driver Assistance Systems (ADAS) and Autonomous Vehicles. For instance, a detection and recognition method based on the Hough transform is described in [1] [2]. More recently, in [3], the author proposes a practical system based on matching a detected region with template images. However, the previously described algorithms work on images from low-resolution video. Our designed system processes high-resolution panoramic images with a very wide field-of-view instead. Moreover, by comparing the results of images captured over different years, the deterioration can be tracked over time. In this paper, we describe the first stage of such a system, which is the detection and recognition of road markings in single highway panoramic images.

OUR APPROACH

Our proposed system for road marking detection consists of three main steps. First, Inverse Perspective Mapping (IPM) is performed to generate a top-view image. Second, road marking elements are extracted by a segmentation algorithm. Third, from each segment a feature vector is extracted, and a Support Vector Machine (SVM) [4] classifier is employed to distinguish segments based on their geometric features. Since some background segments can have similar shapes as road markings, the lanes that appear in the image are modeled using RANSAC and a Catmull-Rom spline [5] with the lane marking candidates classified by the SVM. Many non-road marking segments can be rejected based on the lane positions and the orientation of the segments (the orientation should be consistent with the orientation of the lane markings).?

SYSTEM OVERVIEW

The original spherical panoramic images are stored in an equirectangular format. We transform the panoramic image to a top-view by applying an Inverse Perspective Mapping, similar to [1][3]. The mapping matrix between spherical panoramic images and the top-view image only depends on the camera height of the ground plane and the azimuth angle of the camera with respect to the ground plane.

Road markings are designed to be easily observed by drivers, so pixels of road markings are brighter than the surrounding pixels. Based on this observation, we first compare the value of a pixel with the average of the surrounding pixels bounded using a window, and make a subtraction. If the residual pixel value of this pixel is larger than a manually defined threshold,

then this pixel is considered to be a candidate pixel of road markings. The size of the window is determined empirically, where it was found that a value of a 105?105 pixels provides the best performance for our data. The high-intensity pixels usually form connected segments. The segments that are too small to be considered as road marking regions are removed.

Each of the detected regions is classified into two categories based on their size: Long segments whose thickness corresponds to the width of a real lane marking are assumed to be lane marking candidates, and split into smaller segments for further classification. Prior to extracting the feature vector of each segment, the centroid location, scaling and rotation should be normalized. The features are then extracted by calculating the distance from the center to the boundary of each segment at certain angles. In our experiment, we have chosen these angles from to with a step size of . The value of this step size was chosen empirically and based on numerous experiments. To distinguish the road markings from other regions, we apply a Support Vector Machine (SVM) [4] algorithm to classify the shapes. In our experiment, a non-linear SVM with a radial basis function kernel is used.

LANE MODELING

In order to decrease the number of false detections by the SVM, we utilize our prior knowledge that the road is bounded by solid lane markings, and any segments found outside cannot belong to road markings. For this purpose, we model lanes based on lane marking candidates. In our model, we apply RANSAC and Catmull-Rom spline [5] to interpolate straight lanes and curved lanes respectively. Position and orientation of road marking candidates relative to the lane markings are used to reject false detections.

EXPERIMENTS AND RESULTS

Several experiments are performed to test the performance of our system. The performance is individually measured for the different kinds of road markings, such as lane markings, stripes, blocks and arrows.

Dataset: the dataset consists of panoramic images captured every 5m by a camera mounted on a driving vehicle. A panoramic image contains a road scene with a resolution of 4,800 ?2,400 pixels. We use 910 highway panoramic images under various lighting conditions, and we have manually annotated all road markings that are completely visible on the top-view images Evaluation metrics: we divide the detections into two categories, road markings such as line segments, blocks and arrows and lane markings (long uninterrupted lines). Detections of road markings and lane markings are compared with the corresponding manually annotated top-view image. A detection of a road marking is considered as True Positive (TP), if at least 90% of the pixels of the detection are located within the corresponding shape in the annotated top-view image. If a detection of road markings is not annotated in the top-view ground-truth image, it is counted as a False Positive (FP). Since lane markings are long and usually have a variable length in each image, lane markings annotated in the ground-truth image are first split into small blocks. Similar to detections of road markings, if 90% of the pixels of a detection of a lane marking block overlap with a corresponding annotated lane marking block in the ground-truth image, it is counted as a TP, otherwise it is counted as a FP. If 80% of the pixels of the lane marking blocks belonging to one lane marking are detected correctly, the lane marking is considered to be correctly detected. The performance is evaluated with precision and recall.

Experiments: we have tested our algorithm with the previously described dataset. In our experiment, we have selected 50 panoramic images as a training set for the Support Vector Machine, which contain road markings of line segments, blocks, arrows and lane markings. Our algorithm achieves a recall of 93.1%, 95% and 91.1% for line segments, blocks and arrows, respectively, with a corresponding precision of 90.4%, 95.2% and 91.7%. Additionally, it achieves a recall of 96.4% for lane marking blocks at a precision of 96.2%. If considering whole lane markings, it achieves a recall of 94.3% at a precision of 94.6%. The performance for blocks and lane markings is therefore quite promising.

FUTURE WORK

Since most false detections are caused by vehicles on the road that contain visually similar regions to road markings in the top-view images, as well as imperfections in the road surface. Therefore, most of the false positives can be further reduced by combining the results from the nearby panoramic images. Additionally, we will also extend our system towards additional road markings.

REFERENCES

[1] Rebut, J., Bensrhair, A., Toulminet, G.: Image segmentation and pattern recognition for road marking analysis. In: Industrial Electronics, IEEE Int. Symp. On. Volume 1. (May 2004) 727-732

[2] Maeda, T., Hu, Z., Wang, C., Uchimura, K.: High-speed lane detection for road geometry estimation and vehicle localization. In: SICE Annual Conference, 2008. (Aug 2008) 860-865

[3] Wu, T., Ranganathan, A.: A practical system for road marking detection and recognition. In: intelligent Vehicles Symp. (?), IEEE. (June 2012) 25-30

[4] Vapnik, V.N.: The nature of statistical learning theory (1995)

[5] Wang, Y., Shen, D., Teoh, E.K.: Lane detection using spline model. Pattern Recognition Letters 21(8) (2000) 677-689

9407-8, Session 2

Topview stereo: combining vehicle-mounted wide-angle cameras to a distance sensor array

Sebastian Houben, Ruhr-Univ. Bochum (Germany)

We present an application of stereo vision with wide-angle cameras in the context of Advanced Driver Assistance Systems.

Today, we find modern vehicles equipped with an abundance of sensors for all kinds of tasks that assist or protect

the driver and other traffic participants. The number of sensors and processing units has become a substantial factor in automobile manufacturing cost.

Middle and upper-class vehicles have recently also been deployed with an array of wide-angle or fisheye cameras whose output can be combined to a surround view around the vehicle. These cameras are often mounted at the front bumper, both side mirrors and the rear trunk lid.

However, these potent sensors are mostly used to show the surrounding to the driver in cumbersome manoeuvring situations, thus, neglecting their aptitude for automated driver assistance tasks.

We introduce a stereo image processing pipeline for these cameras' field of overlap which can be found in the region in front right and in front left of the vehicle (overlap of front and side mirror cameras), as well as in the blind spots behind the vehicle (overlap of rear and side mirror cameras). Thus, the four cameras can be combined in four stereo pairs.

Since we do not rely on additional hardware, we map out a cost-efficient way to extract distance information from image data which can be used in all kinds of driver assistance scenarios.

Furthermore, to allow for a simple adaptation to - or even a reutilization of - processing hardware we follow the pipeline of well-established stereo vision algorithms which are found in today's vehicle-mounted frontally aligned stereo camera pairs. Although the problem of this "classic" stereo and the proposed fisheye stereo vision may seem to be very similar, the large baseline resulting in a significantly different perspective as well as non-Lambertian reflections and large occluded image regions pose some challenges.

We consider the common pipeline to be laid out as follows:

1) Apply a fixed transformation to the input image pair to receive a rectified stereo image pair. This allows us to search correspondences only in the epipolar line of the other rectified image.

2) Apply a correspondence measure for each pixel pair of each row and store that information regardless of the actual computed similarity. The data structure containing the correspondence values is called the disparity matrix containing a correspondence value for each pixel position and every possible disparity.

3) Process the disparity matrix to obtain the disparity image containing the most plausible disparity (which can be mapped one-to-one to a physical distance) for each pixel. This step implements some smoothness assumptions while processing the disparity matrix. In the context of vehicle-based stereo the Semi-Global Matching (SGM) approach presented by

Hirschmueller [1] is one of the most prominent methods due to its trade-off between optimally estimating a smooth disparity map and its low runtime requirements.

Regarding 1) we build on the rectification scheme for wide-angle sensors based on [2]. A formulation of the epipolar constraint in three dimensions enables us to find epipolar curves in both input camera images containing possibly corresponding pixels. We establish a processing scheme which automatically determines the required rectification with a maximum field of overlap only using the intrinsic and extrinsic calibration of the involved cameras. In contrast to a classical rectification procedure, we found it useful to project the images on a common cylinder segment, not on a common plane as with classical stereo, in order to better account for the strong lens distortion of the participating fisheye cameras.

Regarding 2) we find the rectified image pair suffering from strong perspective distortions as well as contrast and intensity differences. Thus, we apply a set of local correspondence measures which are commonly deployed in the context of vehicle-based stereo vision: Sum-of-Absolute-Differences (SAD), Census-Transform, Rank-Difference each for different window sizes.

The capability of all measures to deal with the peculiarities of the rectified wide-angle image pair is evaluated by pixelwisely assigning the disparity with maximum similarity. Inspection of the derived disparity image allows us to quickly evaluate the measure.

Regarding 3) we provide a thorough investigation for the choice of the smoothness terms for gradual and discontinuous disparity changes. It is shown that the first have to be strongly preferred to the latter in order to adequately handle the strong occlusions and the partly poor similarity measure.

The theoretical limits of the proposed stereo setup are remarkable: due to the large baseline of 2.0m (front-side) and the large area of overlap of 80° (front-side) we obtain a maximum depth sensing of about 480m on average for a disparity of one pixel at a reasonable angular resolution of 5 pixels / degree. The minimum depth due to a maximum disparity of 150 pixels is estimated to be at 4.0m. However, the strong smoothness constraints during the disparity image's computation strongly limit these characteristic numbers. We therefore will provide an evaluation on real-world data as well.

Since it is cumbersome to obtain reliable ground-truth data for this new type of stereo system, we evaluate the entire processing pipeline on vehicle-safety relevant obstacles like parked cars or plants whose dimensions and distances can be obtained via manual segmentation and triangulation. This yields distances for the first and last detection of scenario-relevant events which provides a much more lifelike assessment of the proposed distance sensor.

We strongly believe that the combination of already deployed cameras and hardware to obtain a distance sensor is highly attractive due to its minimal costs. The characteristics are convincing and unveil an image-based mid-range distance sensor providing additional information about the vehicle surrounding that can straightforwardly be used for many vehicle-safety applications if combined with other vehicle-deployed sensors.

[1] Hirshmueller, H., "Accurate and efficient stereo processing by semi-global matching and mutual information",

Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2005, pp. 807-814

[2] S. Abraham, W. Foerstner, "Fish-Eye-Stereo Calibration and Epipolar Rectification", Journal of Photogrammetry and Remote Sensing, 59 (5), 2005, pp. 278–288

9407-9, Session 2

A machine learning approach for detecting cell phone usage by a driver

Beilei Xu, Palo Alto Research Center, Inc. (United States); Robert P. Loce, Peter Paul, PARC, A Xerox Co. (United States)

Cell phone usage while driving is common, but widely considered dangerous due to distraction to the driver. Because of the high number of accidents related to cell phone usage while driving, several states have enacted regulations that prohibit driver cell phone usage while driving. However, to enforce the regulation, current practice requires dispatching law enforcement officers at road side to visually examine incoming cars or having human operators manually examine image/video records to identify violators. Both of these practices are expensive, difficult, and ultimately ineffective. Therefore, there is a need for a semi-automatic or automatic solution to detect driver cell phone usage while driving. In this paper, we demonstrate a machine learning based method for detecting driver cell phone usage using a camera system directed at the vehicle's front windshield. The developed method consists of two stages: first, the frontal windshield region localization using the deformable part model (DPM), next, we utilize Fisher vectors (FV) representation to classify the driver's side of the windshield into cell phone usage violation and non-violation classes. The proposed method achieved about 95% accuracy with a data set of more than 100 images with drivers in different poses with or without cell phones.

9407-10, Session 2

Driver alertness detection using Google glasses

Chung-Lin Huang, Asia Univ. (Taiwan); Kuang-Yu Liu, National Tsin-Hua University (Taiwan)

Inspired by the so-called first person vision (FPV) research, we propose a driver alertness detection system by using Google glasses. It consists of two modules: the scene classification and the driver viewing angle estimation. First, we use bag of words (BoWs) image patch classification approach. Second, we establish the vocabulary tree to encode an input image as a feature vector. Third, we apply SVM to classify driver's view into interior/exterior scene. Finally, by using Google glasses and the windshield-mounted camera, we may estimate the driver gaze direction. Here, we apply two-view geometry to estimate the coordinate transformation between two camera systems based on a set of corresponding points in a pair of images.

The previous research used histograms of gradient (HoG) and support vector regression (SVR) for head pose estimation. The others tried to detect the driver cognitive distraction by fusing eye movement and driving performance or fused the vision-based detections of gaze patterns and environmental motion saliency maps. Here, we apply a SVM scene classifier to determine the image is the vehicle interior or exterior scene. Then, we apply corresponding points matching to estimate driver gaze direction. For SVM classifier training, we collect lots of images from google glasses to extract the image feature vectors. Then, we perform hierarchical k-mean clustering to build vocabulary tree which can be used to convert the images into feature vectors for SVM classifier. In the testing process, the feature vector extracted from the Google glasses is used to identify the interior or exterior scene of the vehicle.

Here, we find the interest points based on features accelerated segment test (FAST) and use BRIEF feature descriptor for a corner detection using decision tree algorithm. After corner detector training, we apply the corner detector to generate effective corner response function, and use the non-maximal suppression to find the corner features. Then, we compute the score function for each detected corner, and apply non-maximum suppression to retain the corner with higher score function. We measure the distance between two BRIEF vectors by Hamming distance. We find that it outperforms other feature descriptor approached such as SIFT and SURF on two images corresponded points matching.

To identify the vehicle interior/exterior scene of driver's view, we build the BoW model by collecting a set of features from images of Google glasses. The vocabulary tree approach represents the hierarchical BoWs model. First, we apply k-means clustering to training sets to generate k cluster centers. Then, we partition the feature sets into k groups so that each descriptor vector is assigned to the closet center of cluster. We apply the same process to each group of features recursively to build vocabulary tree level by level. The process terminates whenever the level of tree is up to maximum number of levels. For each input feature vector, we compare it with k cluster centers to choose the closest one as the word assignment. Here, we perform

hierarchical k-modes clustering for binary features to build vocabulary tree, and minimize the within-cluster sum of Hamming distance. Each image is represented by a histogram of word frequencies. To assign different weight to each word, we apply the TF-IDF for our vocabulary tree model. Each word is given a weight according to its relevance in the training sets.

The SVM is a nonlinear classifier created by fitting the maximum-margin hyperplane in a transformed feature space. In SVM, we need to find so-called kernel function to map the feature vector into a higher dimensional space. We choose the radial basis function (RBF) as the kernel function. To build a SVM classifier, the penalty parameter C of the error term and kernel parameters ? must be chosen. We apply the cross-validation and grid-search methods to determine (C,?) so that the classifier can predict testing data as more accurately as possible. In ?-fold cross-validation, we divide the training set into ? subsets of equal size. Then, we train a classifier with ?-1 subsets and use the remaining one subset to test. We do the process iteratively for ? times and obtain the cross-validation accuracy by averaging. Various pairs of parameter (C,?) are tried for cross-validation and the one with the best cross-validation accuracy is chosen.

We test 4500 pairs of image to calculate the average number of correspondences. To build the vocabulary tree and SVM classifier, we prepare 3162 vehicle interior scene images and 3083 vehicle exterior scene images. We select 5?106 feature points from the datasets to build vocabulary tree denoted by VT(k,L) with parameters (k,L) = (2, 12) or (3, 8), where k is branch factor, and L is maximum level of depth. We train the SVM classifier by searching for two parameters in RBF kernel by using 5-fold cross-validation accuracy. The training accuracy are 0.994 and 0.995 for VT(2,12) and VT(3,8) respectively. Finally, we use 2300 vehicle interior scene images and 5000 exterior scene images as testing samples.

VT(2,12) outperforms VT(3,8). It is because the numbers of visual words in VT(2,12) and VT(3,8) are 4096 and 6561 respectively. The features belong to vehicle-interior are the discriminative terms for distinguishing interior/exterior scenes. The features of exterior scene are changing due to the vehicle is moving. Hence, we should not separate the interior features into too many visual words. Non-hierarchical model VT(4096,1) performs almost as well as VT(2,12), but the computation time for quantization is about 165 times of the hierarchical BoW model because the number of distance computations for searching in VT(2,12) and VT(4096,1) are 24 and 4096 respectively. The execution time of the system is 85 msec/frame, in which the feature point matching is the most time-consuming. The precision for VT(2,12), VT(3,8), and VT(4096,1) are 98.30%, 99.77%, and 99.30 %, whereas the recall for VT(2,12), VT(3,8), and VT(4096,1) are 83.70%, 78.58%, and 83.03 %.

9407-11, Session 3

Close to real-time robust pedestrian detection and tracking

Yuriy Lipetski, Gernot Loibner, Oliver Sidla, SLR Engineering GmbH (Austria)

Introduction

Visual based pedestrian tracking has numerous practical applications. It can be used to collect various statistics – counting, people flow analysis etc. Those statistics can help to optimize routes, waiting times, and handling of peak times for public transportation points. Another application is video surveillance, e.g. for detection of individuals moving in a prohibited direction. Still, the state-of-the-art of the pedestrian tracking task is not good enough for many applications yet – especially for security area, where every false positive or missed detection is critical.

We present a tracking approach that is aimed to be useful for practical applications:

- Close to real-time (or even real-time for a fast computer) computation;
- Tested with many hours of video data;
- Tested on many different scenarios.

The main contribution of our work is:

- combination HOG/CNN which is novel;
- Extensive analysis of features for tracking and re-identification.

Method and implementation

The core part of our tracking approach is tracking-by-detection, which can be divided into the following parts:

- Detection of individuals in a still frame;
- Tracking of the detected individuals over the scene:
- Projection into the next frame using either KLT motion vectors or linear interpolation;
- Re-identification and local position adjustment using re-detection and appearance matching.

Our experience is that the quality of a pedestrian detector determines the overall tracking quality. We briefly describe our state-of-the art Histogram of Gradients (HOG) based people detector. Several HOG cascades are used which parameters (blocks and cells positioning) are optimized by our genetic optimization algorithm. The cascades are configured with an increased complexity. The detections found by the HOG detector are verified by the Convolutional Neural Network (CNN) based classifier afterwards. The remaining detections serve as input to the tracker.

The proposed HOG+CNN detector is very fast to compute, because the computationally intensive CNN detector is applied for only a few amounts of detection candidates. The CNN itself consists of five layers, having an input layer (which is sample image itself as a normalized grayscale image), two convolutional layers, one full-connected layer and the output layer.

Generally, two detector configurations can be applied depending on a scene – full body or upper body detector. Full body detector yields better quality while upper body detector is more suitable for situations with a higher people density, where stronger occlusions are more likely.

The tracker projects detections found by the HOG+CNN detector into the successive frame using motion information which is given by tracking points of the Kanade-Lucas-Tomasi (KLT) algorithm. To increase robustness of a projected position and to cope with partial occlusions, we take only those points which are valid for a predefined amount of last N frames (N is about 7..11). Moreover, the tracking points must be located inside of a detection bounding box during all of the last N frames. This method helps greatly against partial occlusions where tracking points belonging to the neighboring trajectory could affect the predicted position. If no valid tracking points are found, linear interpolation is applied. To speed up calculation time, KLT tracking is performed on a reduced image size.

The projected detections are verified by the:

- Detections of the HOG+CNN detector for the successive frame (detection matching);
- Appearance matching approach (feature matching).

Feature matching uses different kinds of appearance information. In the course of tracker development we did an extensive study on the usability of features for re-identification of pedestrians. We examined single features as well as their combinations:

- DCT
- Color Hist (RGB, HSV, lab color spaces)
- HOG
- LBP
- Brief
- Fast
- DCT + Color Hist
- HOG + Color Hist
- DCT + HOG + Color Hist

We show feature vector distance distributions of above mentioned features using the annotated TownCentre sequence. As a conclusion, following feature combination is chosen as the best representative:

- HOG features to include gradients statistics;
- Color histograms in the HSV color space room to include color statistics.

To cope with a possible KLT inaccuracy (due to the obstacles, noise, occlusions etc.) feature matching is done in a small neighborhood of the predicted position. The best feature match serves for the position correction.

Using the two verification methods described above we make a conclusion about successful tracking. If either re-detection check or feature matching fails, the corresponding trajectory is marked as potentially occluded or lost. Still, we continue the tracking process for that trajectory. If the trajectory could not be recovered during the last K successive frames (K is about 40..60), we close it.

While using the modern state-of-the art algorithms, the described tracking approach is fast to compute. To further speed up calculation time, simple background modeling is applied. It serves as a mask to switch off detection and tracking process in completely static regions. We achieve about 4..12 fps on a standard PC depending on a video resolution and people density.

In the evaluation section we present our tracking results. We show handling of occlusions, tracking functionality in a crowd, etc. Also we compare tracking quality against other approaches on a Town Centre sequence using CLEAR MOT metrics. Finally, conclusions and next work are described.

9407-12, Session 3

Development of a portable bicycle/ pedestrian monitoring system for safety enhancement

Colin T. Usher, Wayne D. Daley, Georgia Tech Research Institute (United States)

Pedestrians involved in roadway accidents account for nearly 12 percent of all traffic fatalities and 59,000 injuries each year. Most injuries occur when pedestrians attempt to cross roads, and there have been noted differences in accident rates midblock vs. at intersections. Additionally, vehicle-pedestrian collisions at midblock also tend to be more deadly for pedestrians. This is of significant concern to the Georgia Department of Transportation (GDOT) which is being proactive in exploring various approaches to increasing pedestrian safety.

Collecting data on pedestrian behavior is a time consuming manual process that is prone to error. This leads to a lack of quality information to guide the proper design of lane markings and traffic signals to enhance pedestrian safety. The problem stems from observations that "...When convenient and manageable crossing points are not identified, most pedestrians cross at random, unpredictable locations. In making random crossings, they create confusion and add risk to themselves and drivers..." (Turner, Sandt, Toole, Benz, & Patten, 2006).

Researchers at the Georgia Tech Research Institute are developing and testing an automated system that can be rapidly deployed for data collection to support the analysis of pedestrian behavior at intersections and midblock crossings with and without traffic signals. This system will collect and analyze video data to automatically identify and characterize the number of pedestrians and their behavior. It consists of a mobile trailer with four high definition pan-tilt cameras for data collection along with software to conduct the analysis. The software is custom designed and uses state of the art commercial pedestrian detection algorithms provided by SLR Engineering.

The software consists of two main components; a data processing and a reporting component. The processing component handles all tasks associated with loading and processing videos while the reporting component handles presentation of the post analyzed data and report generation. This paper will describe the design and operation of these software elements.

Current state of the art pedestrian detection systems work primarily with limited fields of view for detection at crosswalks or in indoor areas such as airports. For detection at mid-blocks, a system must be able to identify and track pedestrians both close to the system and far away (up to 800 to 1000 feet). This requires the use of high definition data, which increases the requirements in processing time and affects pedestrian detection accuracy. The research team has developed and tested algorithms to handle these unique scenarios.

Preliminary testing shows a detection accuracy of greater than 90% for pedestrian trajectory tracking. However, false detections on vehicles remain a problem. Approaches to filter out false trajectories based on their orientation leads to a vast reduction in these types of false detections. This paper will describe the testing methodology and present results on detection and tracking algorithm accuracy along with approaches to reducing the computational requirements.

It is anticipated that the successful implementation of the system will support the development of models for input into future intersection and traffic signal/lane marking design. As Atlanta becomes a more pedestrian-friendly city, this type of data will be crucial to the proper design and development of future crosswalks and midblock signals/markings.

9407-13, Session 3

Real-time pedestrian detection, tracking, and counting using stereo camera and range sensor

Santiago Olivera, Bo Ling, Migma Systems, Inc. (United States); David R. P. Gibson, Federal Highway Administration (United States); Paul Burton, City of Tucson (United States)

This paper describes a number of mechanisms developed to integrate information from a stereo camera and a scanning laser range finder in order to detect, track, and count multiple pedestrians in real time. The system processes up to ten video frames per second and pedestrian detection and counting accuracy is over 95%.

Pedestrian detection, counting and tracking technologies are considered the key components in ITS. Current pedestrian detectors are insufficient of detecting and tracking pedestrians to support automated counting for traffic surveys. Algorithms and techniques presented in this paper represent the most recent enhancements that include a reliable, stable, and high performance system that can be easily deployed and configured to work in a wide variety of environments. It can be installed up to 30 feet from the desired detection zone on a crosswalk at a height between 10 and 15 feet. Cameras and laser share the same line of view and cover a 60-degree field of view.

Stereo image pairs from the cameras and range information from the scanning laser are initially processed independently. Image acquisition follows a specific timing pattern that allows to process two consecutive stereo pairs at the same time. Consecutive images are used to identify and extract moving targets (i.e., pedestrians) in a process that uses color opponency to highlight salient features on the image, while binocular data are used to create a disparity map that will be used to identify potential pedestrians on the scene. Scanning laser data are filtered to remove noise and background clutter. Potential laser detections are further evaluated in order to determine whether they represent one or more pedestrians or false detections.

Potential targets from the range data are mapped onto the stereo images. A geometrical model was created to automatically map potential detections in laser coordinates (angle and distance) tostereo image pixel coordinates (rows and columns). The mapping model has been extensively evaluated and has been found to map correctly the laser coordinates irrespectively of camera height, distance to target, or camera angles. In this way, only detections that exist both on stereo images and range data are validated.

Validated potential targets are used to create tracks that are referenced across multiple frames. A track is created for every potential pedestrian. Each track contains information on image coordinates, ground coordinates, speed, walking direction, and acceleration of a particular potential target. Tracks also include estimated values for these variables at every frame which are obtained through estimation and used to facilitate the target-to-track matching process. A tracking control mechanism analyzes tracks and detections frame to frame in order to match tracks with detections, create new tracks, identify tracks for pedestrians that may have already left the crosswalk, etc.

Tracks are further analyzed and evaluated before declaring a true detection. Consistency of variables such as walking distance, walking angle, number

of detections, and detection pattern over time is considered in order to validate a true pedestrian detection. Information included on each track can also be used to estimate the intersection performance measures such as pedestrian walking speed, delay or pause times, walking patterns, etc, which can help traffic engineers in better planning and design at pedestrian facilities.

This real time system can also interact with a traffic controller in order to get the status of traffic signals or store pedestrian counting data and other measures for later review by traffic engineers.

A prototype of this platform has been fielded at Tucson, AZ for more than a year. In the paper, we will present the detection results and performance evaluation from this field trial system.

9407-14, Session 3

Active gated imaging for automotive safety applications

Yoav Grauer, Ezri Sonn, BrightWay Vision Ltd. (Israel)

The paper shall cover Active Gated Imaging System (AGIS) as related to technical description of the sensing and projecting elements in the automotive environment. The sensing element uses a Gated CMOS Imager Sensor Imager (GCMOS) which is comprised of a multiple fast shuttering pixel array in a single image readout process. In addition, this unique pixel design operates in a wide spectral spectrum with high sensitivity (QEXFF), high anti-blooming and low noise. The projecting element uses a semiconductor array laser. The laser array is a high power pulsed Near Infra-Red source with a low duty cycle. Both elements are synchronized to provide a unified active gated image at night time in clear and harsh weather such as rain or snow.

The paper shall review the system electro-optics performance in the lab and at the field versus simulations including; noise sources, SNR etc.

Laser safety approach shall also be described as related to operational scenarios of the system in automotive environment.

We shall present our approach of a Driver Assistance System (DAS) named BrightEye, which makes use of the AGIS technology in the automotive field.

Website: http://www.brightwayvision.com/

9407-15, Session 4

Arbitrary object localization and tracking via multiple-camera surveillance system embedded in a parking garage

Andre Ibisch, Sebastian Houben, Matthias Michael, Ruhr-Univ. Bochum (Germany); Robert Kesten, GIGATRONIK Ingolstadt GmbH (Germany); Florian Schuller, AUDI AG (Germany)

We illustrate a multiple-camera surveillance system installed in a parking garage to detect arbitrary moving objects. Our system is real-time capable and computes precise and reliable object positions. These objects are tracked to warn of collisions, e.g. vehicles with pedestrians or other vehicles. The proposed system is based on multiple grayscale cameras connected by a local area network with a shared field of view with the other cameras to handle occlusions and to enable multiple-view vision. A majority of parking garages is already equipped with such cameras, which our system economically benefits. The system's pipeline starts with the synchronized image capturing process separately for each camera. In the next step, moving objects are selected by a foreground segmentation approach. Subsequently, the foreground objects from a single camera are transformed into view rays inside a common world coordinate system and are joined to receive plausible object hypotheses. An initial intrinsic and extrinsic calibration is required once beforehand. Afterwards, these view rays are filtered temporally to arrive at continuous object tracks.

At first, to distinguish between static background and moving objects in the foreground, we implemented an adaptive background representation, based on an exponentially time-smoothed mean-image, and subtracted it from the original image. To detect the objects edges we choose a fast learning setup. Segmentation errors caused by shadows are reduced by the normalized cross-correlation method to identify image regions that only change due to varying illumination. Overexposed regions caused by strong light sources (e.g., vehicle spot lights) are minimized by separating, dilating, and extracting them from the original image.

Remaining segmented image regions are subdivided into identical blocks which are clustered with adjoining blocks to allow for a coarse connectivity of single objects. With the help of an image-world transformation we generate view rays emerging from the respective camera center to the center of a single block. We used this partitioning into blocks to create multiple view rays for a single object rather than regarding only rays through the region of interest's corners of the entire object. This method helps to make the following procedure more robust by averaging over the huge number of view rays. Afterwards, we generate intersection points, i.e the lowest distance between the view rays, from at least two cameras. Since the object assignment between two cameras can be ambiguous if more than one object is present, we use the foreground segmentation about the knowledge of the connected blocks:

To decide which block clusters belong together, we defne a quality criterion, (i.e.) namely the average error of a back projected intersection pair. In the next step, the intersection points that fullfill this criterion are clustered with the help of a distance function using either the euclidean distance or the density of a cylinder, a circle, or a convex hull. In order to track these clustered intersection point clouds, we combined a normal distribution tracking, based on the euclidean distance and the similarity of the distribution and a alpha-beta-filter to follow the course of the position.

In our experiments we used a precise LIDAR-based reference system to evaluate and quantify our system's precision: An array of distributed LIDAR sensors is installed a few centimeters above the ground to detect objects. Each LIDAR sensor distinguish between static and dynamic measurments. These dynamic measurements of all LIDAR sensors are mereged and analyzed afterwards. Both systems are calibrated to the same common world coordinate system to ensure a valid comparison.

The system's mean positioning error in a sequence containing several difficult situations (e.g. moving pedestrian, moving vehicle, or the vehicle switches on its lighting etc.) is 0.24m. Compared to the LIDAR-based reference system with an error of 0.19m this leads to the conclusion that our system is precise enough to locate objects for applications of object detection and collision warning.

In conclusion, the system provides good results if objects are separated in the images. In the presence of occlusion it is in many cases not possible to disjoin them with the presented method. In the future, we want to classify segmented image regions with a detector to provide more information on matching view ray pairs, thus, accelerating and robustifying their intersection.

9407-16, Session 4

Unsupervised classification and visual representation of situations in surveillance videos using slow feature analysis for situation retrieval applications

Frank Pagel, Fraunhofer-Institut für Optronik, Systemtechnik und Bildauswertung (Germany)

Today, video-based surveillance systems produce thousands of terabytes. This source of information can be very valuable, as it contains information about abnormal or similar events, periodic patterns as well as time and space dependent activities. Such information significantly helps for a proper assessment of current or past situations monitored by the video system. But if this data is unstructured, a search in a large-scale video footage can be exhausting or even pointless. Searching surveillance video footage is extremely difficult due to the apparent similarity of situations, especially for

human observers.

In order to keep this amount manageable and hence usable, this work aims at clustering situations regarding their visual content as well as motion patterns. The objective is to support situation retrieval in surveillance footage by clustering similar situations. In this paper, a "situation" is characterized by its spatial and temporal information. A spatial dependency of the image regions is simply reached by dividing the image into cells and by processing each cell independently (future work will aim at connecting these regions and finding relations between the cells). This cell-wise implementation already enables the definition of simple "and" queries, like "Show all situations, where this situation happened in cell A and this situation in cell B". Here, local features are textures of certain image cells that describe the visual appearance of an event in the image. This information is important for a human analyst as it helps navigating intuitively through a video. In this paper we used standard HOG features as local image descriptors, but also other descriptors like color histograms would be possible (but were not considered in this paper).

But events are also characterized by significant temporal dependencies. I.e. one important question to be answered here is "where did objects come from, that passed this image cell" and "where did object go to after passing that cell". Therefore, we introduced novel descriptors, called Franklets, which explicitly encode motion patterns for certain image regions. Franklet descriptors represent motion patterns for a certain image region and also encode that the motion information visually. They are calculated by an accumulation of flow fields over a history of time. As a result, Franklets can show either where (from which cells) objects are going to after, or where objects are coming from before they enter an image cell.

Both, HOG and Franklets are high dimensional features. Features extracted directly from the image data are here called "First level features". As a next step, we want to cluster and visualize the textural and temporal features in lower dimensional feature spaces. Therefore, Slow feature analysis (SFA) will be performed for dimension of the first level features based on the temporal variance of the features. Slow feature analysis recently became popular in clustering and classifying motion-dependent features, like activities or video contents. The idea is that temporally adjacent features (between two or more consecutive frames) are likely to belong to the same class. So in dependence of the features temporal covariance matrices, SFA determines the temporally most stable features. By choosing only the n most stable dimensions, lower dimensional features with higher discriminative properties can be achieved.

By reducing the dimension with SFA, we gain a higher feature discrimination compared to standard principal component analysis (PCA) dimension reduction approaches. However, PCA is still a useful intermediate step for noise reduction. Furthermore, due to the drastic dimension reduction by SFA, situations can be visualized according to their similarity. This can be done by clustering them fast and efficient in an unsupervised manner with self organizing maps (SOMs). SOMs can map higher dimensional features according to their similarity onto a 2-dimensional map, and hence an intuitive representation for human assessors. Here, each node in the SOM is linked with a representative thumbnail (like a cell screen shot) that is visualizing the "scene snippet". In practice, a human operator can select one (ore several) cells, select one out of several feature modes (e.g. HOG, Color Histogram, Franklet Forward/Backward, ...) and then select a situation descriptor in the SOM. This approach can be ideally used for supporting semi-automatic search tasks and navigation in large-scale video surveillance footage.

The online algorithm for situation classification and representation can be summarized as follows:

For all image cells do:

1) Calculation of first level features (HOGs, Franklets)

2) Calculation of PCA (noise reduction)

3) Calculation of SFA features

4) Map into SOM and visualize

For developing, training and evaluating the algorithms, a data set from the Hamburg Harbor Anniversary 2014 (1.5 million visitors) was used, where we had the possibility to capture video data for three full days from a high-rise building.

In this paper, we investigate the effects of dimension reduction via SFA and compared it to classical PCA dimension reduction. Also, we show clustering results with different features (namely HOG feature descriptors and novel visual motion descriptors, Franklets). We could show that by using SFA a significant improvement of the clustering results could be achieved compared to our baseline evaluation data set based on standard PCA techniques. Also, we could demonstrate the practical usefulness of Franklet descriptors for describing temporal events in order to support browsing applications in surveillance video footage. The feature clustering algorithms were integrated into an approach that visually clusters SFA-dimension-reduced features based on self-organizing maps (SOMs).

9407-17, Session 4

An intelligent crowdsourcing system for forensic analysis of surveillance video

Khalid Tahboub, Neeraj J. Gadgil, Javier Ribera, Blanca Delgado, Edward J. Delp III, Purdue Univ. (United States)

Video surveillance has been an active research area recently due to its importance in security and law enforcement. Video surveillance systems are widely deployed with the number of surveillance cameras increasing exponentially. The enormous amount of video data coming out of such systems makes it practically impossible to continuously monitor all the incoming video feeds. In the past few years, many automatic methods have been developed to detect abnormal events and threats. Such automatic methods face significant challenges due to occlusions, illumination changes and computational requirements. Therefore, the identification of threats and anomalies are not always achievable in real-time and videos are usually archived for forensic purposes.

Search and rescue missions and "after-the-event" investigations are examples of time-critical forensic tasks which require careful analysis of tens of hours of video. Annually, hundreds of search and rescue missions are carried out in North America. The search for the lost Malaysian airplane (MH370) and the April 15, 2014 Boston bombing investigation effort are specific examples of time-critical missions. Crowdsourcing has been proposed as a solution to provide the manpower required for forensic video analysis. Crowdsourcing was defined by J. Howe in 2006 as "taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call". It's also referred to the collective intelligence, the wisdom of the crowd or human computation. Crowdsourcing is often considered as an effective solution to problems that involve cognitive tasks Crowdsourcing platforms such as Amazon Mechanical Turk (MTurk), Freelancer and Mob4hire aim to use the collective intelligence of crowds to do tasks that machines find very difficult.

Law enforcement authorities have used some forms of crowdsourcing to reach out to volunteers and ask them to search for objects, suspicious events or missing people in video. The search for the Malaysian airplane had a publicly available crowdsourcing web-based platform. Other search and rescue missions use YouTube and ask volunteers to view videos online.

However, existing crowdsourcing platforms suffer some serious drawbacks. For example, they are open to the public hence, rising concerns about the privacy of the video data. Also, they typically, do not differentiate between experienced and non-experienced volunteers. In some of the crowdsourcing platforms, the number of views per video varies significantly which might lead to some of the videos not receiving adequate analysis. Machine learning techniques are often not incorporated in the crowdsourcing model and the processing resources available at the clients' side (members of the crowd) are not significantly utilized to aid the annotation process. The annotation process is usually tedious and the lack of interactivity makes the process less attractive to volunteers.

Our goal of this paper is to present an intelligent crowdsourcing system for forensic analysis of surveillance video. These videos include video recorded as part of search and rescue missions or large scale investigation tasks. We previously built a web-based video annotation system as a stand-alone crowdsourcing platform to help law enforcement officers. In this paper, we extend it by incorporating object detection and tracking methods. We also present a hierarchical pyramid model to distinguish the ability, experience and performance of crowd members.

The proposed system enables administrators to upload a set of videos as one investigation task or a search and rescue mission. Once the videos are uploaded, the administrator can specify the type of task under investigation. This can be one of the typical offenses or events such as “assault”, “battery”, “missing person” or “abandoned baggage”. Our proposed system operates in an autonomous way to produce the final result of the crowdsourcing analysis in the form of a set of video segments specifying the events of interest as one storyline. This process will help the law enforcement authorities to optimize and expedite their investigation.

The system consists of the following main components: client-side web-based enhanced annotation platform, backend object detection and identification module, training and teaching module, and an adaptive pyramid crowd model. The web-based annotation tool is built using JavaScript and HTML5 elements. This tool enables members of the crowd to view assigned videos and annotate them with pre-defined labels. Each annotation consists of a time interval and a spatial annotation represented by a rectangle encapsulating the object of interest. To facilitate the annotation process, a continuous adaptive meanshift (Camshift) object tracking algorithm runs on the client (member of the crowd) browser and tracks the specified object. Color and edge features are used for tracking and are selected for low computational overhead. To enhance the process with object detection methods, a backend process is initiated when a video is annotated. The goal of this detection is to find video segments containing the same object and display them to the member of the crowd instantly to validate the match.

Using training and teaching module, the administrators can define specific video labels. Video labels are used by members of the crowd to annotate videos. When a label is defined, the administrator uploads training and teaching videos. The teaching video demos the annotation process and explains what constitutes a specific offense or event. The training video gives members of the crowd a chance to practice with an actual assignment.

Our adaptive pyramid crowd model consists of several layers based on members experience and performance. Higher layers in the pyramid validate the output from lower layers. Assignment of crowd members into pyramid layers depends on the crowd size and performance history. To keep track of performance few factors are considered: training module performance, tasks validated by higher layers, number of tasks executed and how the member results compare against the crowd average. We assume the crowd averaged results are reasonably accurate as the theory “wisdom of the crowd” indicates.

By designing this system, forensic video analysis is made more efficient. This leads to a faster intervention by the law enforcement officers or by the search and rescue personnel. This system, being managed by the officers, is safer and more reliable than other crowdsourcing platforms.

9407-18, Session 4

Trusted framework for cloud based computer vision surveillance platforms

Rony Ferzli, Nijad Anabtawi, Arizona State Univ. (United States)

Cloud based computing has been gaining ground in recent years due to several advantages such as the scalability, redundancy, abundance of computational power, and elimination of server maintenance and configuration Due to the fact that most computer vision algorithms are computationally demanding, cloud based solutions are ideal for such scenarios and as such have become the focus of recent research. One such platform is CloudCV supporting known algorithms such as face detection, image stitching, classification and feature extraction [1].

One caveat of adopting cloud based computer vision platforms is securing the end point (i.e camera sensor). In on premise solutions, the camera is available locally in the same vicinity and as such physically secured. In contrast, in cloud based CV solutions, the camera is remotely located and transmitting video information to the server. To avoid eye dropping and tempering the data across the communication channel, well known protocols are widely used such as Secure Socket Layer (SSL) and Transport Layer Security (TLS); these protocols rely on certificates and asymmetric cryptography to authenticate the counterparty with whom they are communicating, and to exchange a symmetric key. This session key is then used to encrypt data flowing between the server and the camera.

But should the server trust the camera itself? The data at the source could be tampered with or another camera could be swapped for the original. Direct Anonymous Attestation (DAA) is a scheme that enables the remote authentication of a Trusted Platform Module (TPM) while preserving the user’s privacy. A TPM can prove to a remote party that it is a valid TPM without revealing its identity and without linkability [2].

One limitation of DAA is that TPM can be revoked only if the DAA private key in the hardware has been extracted and published widely so that verifiers obtain the corrupted private key. To address this issue an Enhanced Privacy ID (EPID) scheme was proposed by [3-5] providing a method to revoke a TPM even if the TPM private key is unknown.

In this paper, a framework is described where EPID can be used for building the trust between the camera and the cloud platform securing end to end path and opening the door for various applications related to surveillance, forensics and health care.

The overall framework consists of the following:

a) Camera’s Manufacturer: has the issuer role where it creates public key and issues a unique EPID private key to each camera member.

b) Camera: stores the private key into a trusted storage area at manufacturing time and performs signing when asked for.

c) Cloud platforms: acts as the verifier requesting the public key from a public server maintained by the manufacturer. The cloud platform sends a random message to be signed by the camera using its private key, the platform is responsible for the verification of correctness of the signature since it knows the corresponding public key.

Note that the prover (camera) proves the knowledge of some secret information to a verifier (cloud server) such that (1) the verifier is convinced of the proof and yet (2) the proof does not leak any information about the secret to the verifier.

To generate the keys, bilinear maps are used. Let G1 and G2 be two multiplicative cyclic groups of prime order p. Let g1 be a generator of G1, and g2 be a generator of G2. We say e: G1 x G2 ? GT is an admissible bilinear map function, if it satisfies the following properties:

For all u ? G1, v ? G2, and for all integers a, b, equation e(ua, vb) = e(u, v) ab holds. The result of e(g1g2) is a generator of GT. There exists an efficient algorithm for computing e(u, v) for any u ? G1, v ? G2.

The EPID scheme is derived from Boneh, Boyen, and Shacham’s group signatures scheme [6] and perform the following to generate the keys:

a) Chooses G1 and G2 of prime order p and a bilinear map function e : G1 x G2 ? GT.

b) Chooses a group G3 of prime order p with generator g3.

c) Chooses at random g1 h1, h2 ? G1 and g2 ? G2.

d) Chooses a random r ? [1, p-1] and computes w = g2r. The public key is (g1, g2, g3, h1, h2, w) and the issuing private key is r.

Once the cloud platform finishes attestation successfully, it can signal the camera to start transmitting video and process it in the cloud to perform computer vision algorithms such as face detection.

To be able to test the framework described above, an experimental Setup is built to test whether the trusted framework is working as expected, the following building blocks are used:

- Tablet 1: emulating trusted camera sensor usingan Intel based tablet to emulate the camera sensor. Intel has many hardware features related to trusted platform in particular the use of the EPID scheme [7]

- Tablet 2: with no trusted platform emulating malicious camera sensor.

- Cloud Computing Platform: use of CloudCV platform to perform face detection using clusters in the cloud.

- Cloud Attestation platform: acting as verifier trusted platform and signaling the cloud computing platform that the sensor is trusted and thus can receive video for analysis (i.e face detection)

Interaction between the cloud server and tablet is performed using https (i.e SSL) and RESTFUL protocol for communication. Preliminary experiments

show the validity of the proposed framework where Tablet 2 will not connect to the CloudCV while Tablet 1 will.

References

[1] CloudCV: Large-Scale Parallel Computer Vision on the Cloud. http://cloudcv.org/objdetect/

[2] Trusted Computing Group. "TCG TPM Specification 1.2," 2003, http://www.trustedcomputinggroup.org

[3] E. Brickell, L. Chen, and J. Li. "A New Direct Anonymous Attestation Scheme from Bilinear Maps." In Proceedings of 1st International Conference on Trusted Computing, Volume 4968 of Lecture Notes in Computer Science, pages 166–178, 2008.

[4] E. Brickell and J. Li. "Enhanced Privacy ID from Bilinear Pairing." Cryptology ePrint Archive, Report 2009/095, 2009.

[5] E. Brickell and J. Li. "Enhanced Privacy ID: a Direct Anonymous Attestation Scheme with Enhanced Revocation Capabilities." In Proceedings of the 6th ACM Workshop on Privacy in the Electronic Society, pages 21–30, 2007.

[6] D. Boneh, X. Boyen, and H. Shacham. "Short group signatures." In Advances in Cryptology -- Crypto, Volume 3152 of Lecture Notes in Computer Science, pages 41–55, 2004.

[7] Enhanced Data Protection with Hardware Assisted Security, Intel Corp, http://www.intel.com/content/www/us/en/data-security/security-overview-general-technology.html

9407-19, Session 4

Hierarchical video surveillance architecture: a chassis for video big data analytics and exploration

Sola O. Ajiboye, Philip M. Birch, Christopher R. Chatwin, Rupert C. Young, Univ. of Sussex (United Kingdom)

There is increasing reliance on video surveillance system for systematic derivation, analysis and interpretation of data needed for predicting, planning, evaluating and implementing public safety. This is evident in massive number of surveillance cameras deployed across public locations. For example, in July 2013, the British Security Industry Association (BSIA) reported that over 4 million CCTV cameras are installed in Britain alone – the BSIA also reveal that only 1.5% of these are state owned. This paper propose a framework that allows access to data from more cameras, with the aim of improving the accuracy of security alerts, public safety planning, and decision support systems that are based on state-owned video surveillance systems. It suggests an approach to designing video surveillance system as a unified communication system.

The accuracy of result obtained by public safety departments would improve if privately owned surveillance systems 'expose' certain video-generated metadata events such as triggered alerts. Subsequently, a police officer, for example, with appropriate level of system permission can query unified video systems across large geographical area such as a city or a country to predict the location of an interesting entity, such as a pedestrian or a vehicle. This becomes possible with our proposed novel hierarchical architecture, the Fused Video Surveillance Architecture (FVSA). At high level, FVSA comprises of a hardware framework that is supported by a multi-layer abstraction software interface.

Summary

Communication infrastructures, services and their operations are increasingly being unified to improve access to information. Some benefits of unified communication include informed decision on health, social, security, financial and economic outcomes and the possibility of further analytics and exploration of the generated data through cross-referencing. An example implementation of unified communication is smart cities where various services are being unified across various industries in cities, using big data technologies.

Despite the integration of systems across many industries, video surveillance systems are chiefly configured independently, mainly because of inherent complexities such as data protection requirements and absence of defined integration design and communication interface. With current systems, video data are sourced from various surveillance cameras in large volumes where data from cameras possess a level of personality whereas current technology does not provide a means to collectively harness latent information that are embedded in them.

Earlier researchers have published systematic approaches to generating event-based metadata from video surveillance systems – metadata persists abstracted structures and content that users can query to retrieve meaningful information on the associated video data including event detection an autonomous object tracking. Our work is established on the reality of metadata – with appropriate authorization implemented, surveillance systems can expose limited and controlled data, where the data can provide useful information beyond the political and economic boundaries of the system owners.

Owners of surveillance systems are responsible for securing video generated by cameras to protect peoples' privacy, secure their data, and protect their economic interest. It is therefore imperative to independently managed and protected video surveillance system from external access. Nonetheless, metadata generated from the systems can provide meaningful information without revealing the full content of the video.

We explore the need to share information – high-level components within any instance of our system architecture includes a metadata framework, which implement an event-based object detection and object tracking module; Intelligent Network Video Recorders (iNVRs), and secure APIs through Service Oriented Interface. With our novel architecture, each deployed video surveillance system is capable of achieving the following requirements:

- Autonomous and continuous identification, tracking and investigation of objects from any camera on the network.

- Persistence of the tracking data.

- Apply a level of authorisation and authentication on the data to prevent fraudulent access.

- Perform high compression on the video data so they are cheaper to store for a reasonable.

- Persist suspicious data (such as intrusion alerts) for future review by authorised personnel.

- Generate reports, statistical information and evidence of behaviour and events for informed decision-making.

Because each independent system provides an interface, through which its metadata can be queried, it is potentially a component of the larger network. The framework of our solution is compatible with the hierarchical structure of computer networks where for example, routers and switches operate at different logical layer or scope. We suggest a hierarchical design and a high-level configuration for video surveillance devices and services, making it possible to approach video networks in layers such as internal system (local) or external system (global). As in traditional networks, a video network is provided an interface through which it receives and shares data with other systems. In addition to becoming a unified system our design aims to integrate video surveillance systems with any communication-aware systems such as smart city.

9407-36, Session 4

A multi-objective optimization algorithm for efficient background subtraction in image processing

Ramesh Rajagopalan, Univ. of St. Thomas (United States)

Tracking moving objects is an important task in several computer vision applications. Background subtraction is the very first step in image processing. The background subtraction algorithm should be able to accurately extract the foreground pixels corresponding to the moving object. Several background subtraction techniques have been proposed in the literature. Among these, the mixture of Gaussians model is widely used due to its robustness to variations in lighting and higher accuracy.

However, the performance of the mixture of Gaussian model strongly depends on parameters such as learning rate, background ratio, number of Gaussians, and number of training frames. Manually fine tuning these parameters is a challenging task. In this work, we propose a multi-objective optimization algorithm to determine the optimal values of the learning rate, background ratio, and the number of training frames. The performance of the background subtraction algorithm is assessed using metrics such as precision and recall. Maximization of precision reduces the percentage of false positives while maximizing the recall reduces the number of false negatives. The multi-objective optimization algorithm addresses this challenge by obtaining a set of mutually non-comparable solutions called the Pareto-front which characterizes the tradeoff between precision and recall. Experiments based on the Wallflower test images demonstrate the superior performance of the proposed multi-objective optimization algorithm when compared to recently proposed approaches such as weight based genetic algorithm and particle swarm optimization.

9407-20, Session 5

Gender classification in low-resolution surveillance video: in-depth comparison of random forests and SVMs

Christopher D. Geelen, Rob G. J. Wijnhoven, ViNotion B.V. (Netherlands); Gijs Dubbelman, Peter H. N. de With, Technische Univ. Eindhoven (Netherlands)

1. Introduction

The growing number of surveillance cameras results in an increasing demand for automatic and intelligent video content analysis systems. Such a system enables more efficient monitoring, by only presenting interesting footage to the security personnel. This is achieved by characterizing objects by their attributes, such as gender, age and clothing.

The aim of this research is to develop a gender classification system, following a pedestrian detection system. Typical challenges in video surveillance involve low-resolution images, varying illumination and occlusions. Gender classification provides additional challenges. Although humans make distinct classifications of gender, the decision boundary is not well-defined, due to the large intraclass variation and interclass correlation.

Three distinctive fields in gender classification appear in literature: classification using human faces, gait and full body (see survey paper of Ng et al.[1]). The first two methods are not applicable to the surveillance domain, since they require either the requirement of high-resolution, frontal, up-close face images or a clean capture of the full-body periodic movement. Full-body gender classification tackles these issues in return for a lower classification score. Several features are used in literature, such as HOG (Cao et al.[2]), HOG+HSV (Collins et al.[3]) and BIF (Guo et al.[4]). Another promising feature not yet used in full-body gender classification is Local Binary Patterns (LBP)[5].

Our research proposes a gender classification system for low-resolution surveillance video, and evaluates three different feature descriptors (HOG, HSV and LBP). Furthermore, an in-depth evaluation is performed on two different classification systems (SVM and RF).

2. Approach

We use an approach where the location of each person is given by the preceding pedestrian detection system. These images are described by feature descriptors to extract relevant information. Next, these features are classified into males and females.

We consider three different feature descriptors. First, we have chosen Histogram of Oriented Gradients (HOG) as the shape descriptor, which uses image gradients to model shape information (Dalal and Triggs[6]). Next, Local Binary Patterns (LBP) are employed, which are broadly adopted[5] as an attractive, preprocessing stage in classification problems. LBP stores the difference of pixel intensities, resulting in high-dimensional texture descriptions, but which are still efficiently computable. Third, we use the Hue-Saturation-Value (HSV) color description, since it is scale-invariant and shift-invariant with respect to light intensity[7].

We also evaluate the work between two popular classification techniques that obtain state-of-the-art performance: Support Vector Machines (SVMs) and Random Forests (RFs). The main differences are that SVMs use all feature dimensions to construct the decision boundary, while RFs are ensembles of decision trees, which only use small subsets of feature dimensions. Furthermore, SVM is a deterministic classifier, while RF works with randomization. From literature, it is known that RFs require strong individual features and many training samples, while SVMs can handle small training sets or noisy features more robustly.

3. Evaluation

We use two performance criteria, the overall accuracy and the mean accuracy. Evaluation is performed on three datasets: the public MIT CBCL dataset and two novel datasets A and B, that better resemble typical surveillance scenes. These datasets contain more occlusions and viewpoints and approximately eight times more samples.

To provide a baseline of gender classification, a small experiment is performed to determine the human classification accuracy. Using a human test group of 15 different persons, they each classify 100 images of the novel dataset A. The result of this experiment is a 92.6 plus-minus 2.8% classification accuracy, and can be regarded as the maximum achievable classification performance for our problem.

System performance of several feature descriptor combinations is evaluated on the MIT CBCL dataset, using a linear SVM classifier. We outperform the state-of-the-art, obtaining an overall accuracy of 80.9 plus-minus 2.4% and a mean accuracy of 76.6 plus-minus 0.9% using both front- and back-view dataset images from the MIT CBCL dataset.

Classifier comparison

A linear SVM and RBF kernel SVM are evaluated with respect to the regularization parameter and kernel size. Results show a lower performance when using the RBF kernel SVM. Highest classification accuracy is obtained using the combination of LBP and HSV features, resulting in the highest classification accuracy of 89.3 plus-minus 0.2%. The gender classification problem is well separable using a linear hyperplane.

RF analysis determined the best forest structure with 50 trees and a depth of 14. When evaluating the degree of randomization, the highest accuracy is obtained when using the square root of total number of dimensions for individual feature descriptors. However, when combining descriptors, for a high classification accuracy the required degree of randomization is much higher. The highest accuracy is obtained with HOG and HSV features, obtaining an accuracy of 88.2 plus-minus 0.8%.

This high degree of randomization leads to the question whether RF can cope with high-dimensional features. Therefore, we now evaluate multiple feature dimensions during node splitting, and propose a hybrid RF-SVM structure, where in each RF node an SVM is trained on a subset of randomly chosen (LBP-HSV) feature dimensions. This outperforms both individual classifiers and results in an accuracy of 89.9 plus-minus 0.2%.

4. Discussion & Conclusion

The SVM finds a good generalization of the decision boundary, but ignores the inherently complex boundary of gender classification. RF can construct complex boundaries, but has difficulties exploiting the relations between feature dimensions and over-trains on specific dimensions. By using a hybrid RF-SVM classifier, the strengths of both classifiers are combined and results in the highest system performance in our research.

However, experiments indicate that the number of training samples (i.e. 5,800 samples) is too limited to fully extract all relevant information from the feature space. First, the accuracy of RF converges when increasing the maximum depth of the trees, in contrast to results from literature. Second, using an RBF kernel in SVM also results in a low accuracy. This indicates that not all information is extracted from the training samples. The number of samples is sufficient for a linear SVM to find a generalized decision boundary, but in order to construct the correct complex decision boundary with RF, more information, samples, about non-stereotypical pedestrians are needed.

Our proposed gender classification system uses a combination of LBP and HSV features and a hybrid RF-SVM classifier structure. We show that it outperforms the current state-of-the-art on the public MIT CBCL dataset. On our novel and realistic, low-resolution surveillance datasets, the system obtains an accuracy of 89.9 plus-minus 0.2% and approaches the (optimal)

human baseline of 92.6%.

[1] C. B. Ng, Y. H. Tay, and B. M. Goi, "Vision-based human gender recognition: A survey," arXiv preprint arXiv:1204.1611, 2012.

[2] L. Cao, M. Dikmen, Y. Fu, and T. S. Huang, "Gender recognition from body," in Proceedings of the 16th ACM international conference on Multimedia. ACM, 2008, pp. 725–728.

[3] M. Collins, J. Zhang, P. Miller, and H. Wang, "Full body image feature representations for gender profiling," in Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on. IEEE, 2009, pp. 1235–1242.

[4] G. Guo, G. Mu, and Y. Fu, "Gender from body: A biologically-inspired approach with manifold learning," in Computer Vision–ACCV 2009. Springer, 2010, pp. 236–245.

[5] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 24, no. 7, pp. 971–987, 2002.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, vol. 1. IEEE, 2005, pp. 886–893.

[7] K. E. Van De Sande, T. Gevers, and C. G. Snoek, "Evaluating color descriptors for object and scene recognition," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 9, pp. 1582–1596, 2010.

9407-21, Session 5

Detection and handling of occlusion in a object detection system

Ron M. G. op het Veld, Rob G. J. Wijnhoven, ViNotion B.V. (Netherlands); Egor Bondarev, Peter H. N. de With, Technische Univ. Eindhoven (Netherlands)

1 Introduction

This paper focuses on the application of object detection in video surveillance, where the occurrence of objects has a large variation in appearance and occlusions are often part of the scene. Due to these variations in appearance (e.g. different scales and orientation) and the unknown position in the image, detecting objects is a challenging task. While detection quality is constantly improving, state-of-the-art techniques struggle to detect objects that are in unusual poses or occluded [1]. Occlusions are situations in a scene where an object of interest is partly covered by another object, so that the object of interest is not completely visible in the actual scene view.

The objective of our research is to improve the robustness of object detection in situations with clutter and strong occlusions. The baseline system is a real-time sliding-window object detection system that uses Histogram of Oriented Gradients (HOG) [2] features and linear classification. Instead of explicitly detecting occlusions, our approach focuses on the detection of non-occluded object parts using multiple classifiers in parallel, each dealing with different partial object views. In this study we show results for detecting humans, but the proposed techniques are generic and can be applied also to other object classes (e.g. cars).

2 Approach

The adopted approach adds occlusion handling to our existing sliding-window-based detection system. We propose a novel classification system which combines multiple classifiers, where each classifier focuses on a different non-occluded region of the object of interest and the occluded region is discarded from the decision making. Therefore, each classifier is trained in a prior offline stage for a certain occluded region with variations in size and position. During the detection stage, each classifier is operated on all sliding windows covering the image. In case of occlusion, there will be a classifier in the total set that matches to the type of occluded region in the image and ignores that occluded part.

Once multiple classifiers are designed and selected, their results have to be merged in an efficient way to optimize classification performance. Since each classifier is trained independently, the margins for the linear classifiers are different.

Classifiers covering a larger visible region of the object have an inherently better classification performance and have better-defined margins. The merging of classifier results can be controlled by one or more metrics and the input results of the merging require calibration of the individual classifiers. To calibrate each individual classifier, classifiers can be normalized using the maximum achievable detection score as described in [3], which is very dataset dependent.

Instead, we propose to normalize the linear classifier by scaling the weight-vector (and bias) to unity energy. Apart from individual normalization, when using multiple classifiers, the influence of each individual classifier to the combined result can be tuned. This can be implemented by imposing a weight opposite to the occlusion level [3] for each classifier.

Hereby, it is assumed that a classifier covering a smaller area performs always worse than a classifier covering a larger area. Furthermore, we calibrate the decision threshold for each individual classifier using a fixed false-positive rate, to statistically equalize the number of false detections.

3 Evaluation

Based on statistics, we have manually designed 29 different classifiers with different occlusion masks, covering 95% of all types of occlusion [4]. For all experiments, we have trained our classifiers on the INRIA Person training set. To demonstrate the importance of the classifier size, we compare 9 different classifiers, ranging from smaller to larger bottom-to-top occlusions. Overall, we notice that the larger the occlusion covered by the classifier, the higher the miss-rate, which is mainly caused by the higher number of false detections. We have obtained comparable results for right-to-left and left-to-right occlusions. By combining multiple classifiers, there are more detections of the same object, resulting in higher scores after merging, thereby reducing the number of false detections.

Let us now consider the technique for merging the classifier results. We first calibrate the decision threshold (tclass) for each individual classifier, using a fixed false-positive rate. Thus, the performance of each classifier is fixed, so that the contribution to the final classification is determined. For calibration, we use DET-curves [2] at fixed false-positive rates. We have concluded that overall, the decision threshold is optimally set at reference point 1 ? 10?4 false postives per window, which is used for the system. This outperforms non-calibrated classifiers (tclass = 0) by 4%.

Although this advanced occlusion handling increases performance, computational cost increases as well. To reduce this cost, we have designed a cascaded implementation of classifiers. The largest-area classifier is evaluated first and only if the score is above an initial performance level, all other classifiers are executed. In this way, many negative samples are already discarded in the first step by the largest-area classifier. A suitable trade-off between performance and cost leads to the use of 17 classifiers, resulting in an improvement of 8% in detection performance for only 3.4% higher computation cost. Using 7 classifiers, the computational cost grows with 1.3%, while increasing detection performance by 7.6%. Another significant improvement is obtained by applying a different merging technique adopted from [3], which increases the total performance gain up to 19.9%. This will be elaborated further in the full paper.

4 Discussion & Conclusion

We have proposed a novel system for occlusion handling and integrated this in a sliding-window detection framework using HOG features and linear classification. Occlusion handling is obtained by the combination of multiple classifiers, each covering a different level of occlusion. For real-time detection, our approach with 17 classifiers obtains an increase of up to 19.9% detection performance. This gain is caused by both a different merging technique (NMS+Merging [3]) and the novel concept of applying multiple classifiers with normalization and calibration. We have proposed a cascaded implementation that increases computational cost by only 3.4%. The presented results are based on pedestrian detection, but our approach is not limited to this object class. Fixing the types of occlusions prior to training, creates the advantage that we do not need an additional dataset for training, which covers all possible types of occlusions.

References

[1] D. Hoiem, Y. Chodpathumwan, and Q. Dai, "Diagnosing error in object detectors," in Computer Vision–ECCV 2012. Springer, 2012, pp. 340–353.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Computer Vision and Pattern Recognition, CVPR 2005, pp. 886–893

[3] M. Mathias, R. Benenson, R. Timofte, and L. Van Gool, "Handling occlusions with franken-classifiers," Submitted for publication, 2013.

[4] P. Doll ?r, C. Wojek, B. Schiele, and P. Perona, "Pedestrian detection: An evaluation of the state of the art," Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 34, no. 4, pp. 743–761, 2012.

9407-22, Session 5

Spatio-Temporal Action Localization For Human Action Recognition in Large Dataset

Sameh Megrhi, Univ. Paris 13 (France); Marwa Jmal, Ecole Polytechnique de Tunisie (Tunisia); Azeddine Beghdadi, Univ. Paris 13 (France); Wided Souidene, Ecole Polytechnique de Tunisie (Tunisia) and Univ. Paris 13 (France)

Action recognition is an active research field in computer vision. It includes a wide range of applications such as video surveillance, physical security, gesture interpretation, behavior recognition, robotic vision, transport, video search/retrieval and human-machine interaction etc. Recognizing human actions in videos is a challenging task. In fact, videos may contain complex actions with large intra-class variability, poor quality and camera motion.

In the literature, videos are described in different ways. Several methods rely on encoding the entire video sequence. This obviously leads to a huge number of descriptors. Most these features do not describe the action since they focus on non-moving humans/objects in the scenes. In other works video sequences are described by a fixed number of frames leading to a poor interpretation of the observed scene. But, this technique shows good results when exploited in a video dataset where only one person is performing one action with static background and discarding camera motion (KTH dataset). However for realistic videos (YouTube), moving objects disappear or reappear in some sequences due to occlusions or changes in viewpoints. Moreover, actions can be continuous, not-continuous, superposed (jumping to avoid an obstacle while walking, or stop to drink while walking), etc. In these cases a discriminative video segmentation needs to be addressed carefully. For that reason, many approaches are based on the visual segmentation to rely on a significant video sequence rather than encoding the entire video or a randomly fixed frame number.

In this context, this paper attempts to address the problem of human action recognition in realistic videos captured by moving cameras by segmenting human motion, investigating the optimal sufficient frame number to perform action recognition, and also by introducing a video description scheme based on the trajectory of the Speed up Robust Feature (SURF).

To perform action recognition, moving humans/objects in videos need to be first detected and segmented. Here, we seek for reducing the amount of data involved in motion analysis while preserving the most important structural features. We first detects image edges using the canny edge detector. As follows, all the steps of the motion segmentation process will be applied on the edge frames. Interest points are densely detected and extracted based on dense SURF points with a temporal step of N frames. Then, optical flows of the detected key points between two frames are computed by the iterative Lucas and Kanade optical flow using pyramids. Since we are dealing with scenes captured by moving cameras, objects' motion involves necessarily the background and/or the camera motion. Hence, we propose to compensate the camera motion. To do so, we first cluster optical flow vectors using KNN clustering algorithm in order to assess, under the assumption that camera motion exists if most points move in the same direction. If it does, we compensate it by applying the affine transformation on each frame in which camera motion is detected using as input parameters the camera flow magnitude and deviation. Finally, after camera motion compensation, moving objects are segmented using temporal differentiating and a bounding box is drawn around each detected moving object.

Thereafter, the discriminative video segmentation is performed based on the extracted bounding boxes. First, a maximum of the spatial information contained in those video bounding boxes is captured by performing dense SURF extraction. Second, selective video snippets describing the detected action are extracted using a tracking process. Then, a motion angle is empirically settled to track significant motion and in the same time reduce camera motion effect by ignoring small horizontal displacements. This technique allows investigating a sufficient frame number to recognize significant human small actions called "actionlets". The combination of ordered actionlets leads to describe an entire human action. Our proposed technique offers various advantages on the computational complexity and time consumption. First, it is based on a limited number of relevant frames, so all non-significant ones are discarded. Furthermore, the proposed technique is an optimal temporal action detection. Moreover, it allows tracking linear vectors of displacement to avoid additional trajectory shape feature computation. The combination of ordered actionlets permits to describe an entire human action. To perform action recognition many existing approaches are based on spatio-temporal interest points. Almost all of the proposed methods used for detecting spatio-temporal descriptors are based on the extension of a 2D interest point (IP) to the temporal domain (1D). The limitation of these methods is that they handle spatial and temporal information in a common 3D space. However, they have different characteristics and associating them differently in a new scheme deserve to be more investigated. To overcome these problems, we suggest tracking SURF descriptors upon video segmented sequences in order to detect spatio-temporal features. For this purpose, we introduce a new displacement descriptor called DD. We also propose a robust video description based on a late fusion process of five trajectory descriptors. The employed descriptors are the following: first, spatio-tempral SURF (ST-SURF) is used for its spatial and temporal description aptitude. Second, motion boundary histogram (MBH) is employed for its ability to reduce camera motion observed in realistic videos. Third, the histogram of motion trajectory orientation descriptor (HMTO) based on the SURF region, position and scale is used to track local patches. In fact, for each detected SURF in a given bounding box, we define the interest point neighborhood size related to the SURF scale. For the detected patch, we extract dense displacement field based on optical flow algorithm. Motion trajectories orientations are then generated for every pixel by exploiting horizontal and vertical optical flow components. Finally, the displacement distance is a new descriptor introduced to describe the trajectory of interest points in every segmented video sequence. The descriptor extraction task is followed by the generation of the final video description. These features are used in order to represent video objects as a dictionary of visual words. In this paper we exploit the K-mean clustering algorithm to extract a codebook. The extracted visual words are then classified based on an X2 Kernel Support vector machine (SVM). The nonlinear SVM is a fast and efficient algorithm that maps histograms in a higher dimensional space. SVM demonstrated high classification results under challenging realistic conditions, including intra-class variations and background clutter.

We conduct the proposed action recognition framework on a big and realistic dataset called UCF 101. It is collected from YouTube, having 13320 videos from 101 action categories. The videos of 101 action categories are grouped into 25 groups, where each group consists in 4-7 videos of an action. The videos from the same group may share some common features, such as similar background, similar viewpoint, etc. Our results outperform several baseline results achieving (83,151%) of accuracy in the UCF101 dataset. The obtained performances are encouraging and prove the viability of the proposed video description.

9407-23, Session 5

Person identification from streaming surveillance video using mid-level features from joint action-pose distribution

Binu M. Nair, Vijayan K. Asari, Univ. of Dayton (United States)

We propose a real time person recognition algorithm for surveillance based scenarios from low-resolution streaming video, based on mid-level features extracted from the joint distribution of various types of human actions and human poses. The proposed algorithm uses the combination of a novel action recognition framework which produces per-frame probability estimates of the action being performed and a pose recognition framework which gives per-frame body part locations. The main focus in this manuscript is to effectively combine these per-frame action probability estimates and pose estimates to provide appropriate weights to optical flow trajectories from each body part across a 15 frame window. These weighted trajectories can then be used to provide unique signatures corresponding to an individual and can be used to identify individuals under specific surveillance areas. The entire proposed algorithm is divided into the 3 parts : Per-frame human action recognition algorithm ; phase based articulated body part models for human pose estimation in low resolution imagery; and motion flow trajectory selection and learning mechanism which combines the action probability estimates and human pose estimates in an kernel based svm. The human action recognition framework extracts local flow based motion and shape descriptors and computes the underlying action distribution through regression-based modeling. The action models are designed to be independent of time so as to incorporate action sequence length normalization, motion speed invariance and non-initialization of action states. We formulate an action descriptor based on the fusion of novel motion features ,Hierarchical Histogram of Oriented Flow (HHOF) and the Local Binary Flow Pattern (LBFP) computed from dense Lucas Kanade optical flow, and shape features known as the R-Transform. A time independent orthogonal action basis using Empirical Orthogonal Functional Analysis is computed for each action class where projections on it vary with time. Due to the one-to-one mapping between action feature space and the basis projections, the time series is modeled by computing the mapping using Generalized Regression Neural Net. Classification label and subsequent probability estimates of the input streaming sequence at each frame is provided by the GRNN action model which estimates the true projections. For the human pose estimation algorithm, the state of the art articulated parts model are modified to work on phase features obtained from a monogenic signal representation rather than the well-known HOG descriptors. Computing the monogenic signal from a two dimensional body part region enables us to separate out the local phase information (structural details) from the local energy (contrast) thereby achieving illumination invariance. Therefore, at every frame, we get phase-based human pose estimates and the action probability estimates. Then, across a 15 frame window, we compute dense trajectories and the corresponding trajectory descriptors in the scene and localize them based on the individual detected by the well-known DPM detector. Using the probability action estimates as additional constraints in a multiple kernel support vector machine(MKL-SVM) , an automatic allocation of kernel weights is possible with each kernel corresponding to the trajectory descriptors extracted from a specific detected body part. This approach containing a combination of multiple algorithms at play, has been tested on some low-resolution public action datasets such as the Weizmann and KTH and on a private dataset provided by AFIT, where not only annotations for actions are provided but also for the actors in the scene. The challenges in these datasets involve the variation of viewpoint and low resolution which is perfect for testing algorithms meant for surveillance video feed.

9407-24, Session 5

Scene projection by non-linear transforms to a geo-referenced map for situational awareness

Kevin Krucki, Vijayan K. Asari, Univ. of Dayton (United States)

There are many transportation and surveillance cameras currently in use in major cities that are close to the ground and show scenes from a perspective point of view. It can be difficult to follow an object of interest across multiple cameras if many of these cameras are in the same area due to the different orientations of these cameras. This is especially true when compared to wide area aerial surveillance (WAAS). To correct this problem, this research provides a method to non-linearly transform current camera perspective views into real world coordinates that can be placed on a map in real-time. Using a directed homography and perspective and affine transformation matrices, perspective views are transformed into approximate WAAS views, called the approximate wide area view image (AWAVI). The AWAVI is then placed into the cameras real world field of view (RWFOV), giving us the map view image (MVI). All images are then on the same plane, allowing a user to follow an object of interest across several cameras on a map. While these transformed images will not fit every feature of the real world as WAAS images would, the most important aspects of a scene (i.e. roads, cars, people, sidewalks etc.) are accurate enough to give the user situational awareness. Our algorithm is proven to be successful when tested on a number of different cameras from the downtown area of Dayton, Ohio.

The image transform takes place in multiple steps and on two markedly different types of images. The first type of image does not look above the horizon; the camera is looking only at the ground. The second type of image looks above the horizon and are much more challenging to accurately fit a map compared to the first type of image. For the algorithm to determine which type of image we are looking at, certain information from the camera must be known. Knowledge of the tilt parameter allows us to determine if the camera is looking above the horizon at the sky. Obviously, these parts of the image cannot be represented as the ground on a 2D map and therefore are removed before transforming. If parts of the image are above the horizon, then horizon estimation is used to identify where the horizon is and all pixels above this location are discarded. The horizon estimation technique finds the intersections among all nearly parallel lines, also known as vanishing points, in the image and uses an average to obtain an approximate location that is accurate enough for the algorithm.

After the sky is removed from the second type of image, all images are treated the same. To start, the scene is transformed using a directed homography, which allows us to transform the scene as if we are looking down on it from the sky. This homography process is based on an orthogonal set of vectors in camera space that define a specific plane. For our purpose, that specific plane is the ground, and our specified vectors are pointing up in 3D space. Once these vectors are defined, they are placed into a matrix, specifically a 3x3 rotation matrix. This matrix is then multiplied by an affine transformation matrix that takes into account the height and width of the image and the focal length of the camera. This resulting matrix is then multiplied by a translation matrix and another affine transformation to center the image and make it the appropriate resolution and size. This gives us the final transform matrix, which when multiplied with the image performs a perspective transform to give the AWAVI.

The AWAVI image needs to then be placed into its real world coordinates. The matrix needed for the transform to real world coordinates is attained by using the camera's RWFOV, which can be calculated by knowing the height, pan, tilt, zoom, and GPS coordinates of the camera and basic trigonometry. The camera's field of view in angular form, entirely different from its RWFOV, is the first parameter calculated from the equation ?=2 tan^(-1)?(d/2f), where d is the sensor size and f the focal length of the camera. By knowing this angle ?, the tilt of the camera, and how high it is off the ground, we can calculate the length, base and top width of the RWFOV using trigonometric relationships. Once the physical lengths and widths of the RWFOV are known, the distance the RWFOV is from the camera can be calculated by comparing the RWFOV length to the camera's base length. A camera's base length is the length of a camera's RWFOV if it is pointing straight down. Once base length is known, the four corners of the RWFOV are placed around the camera's GPS location. The AWAVI is then mapped into the four corners of the RWFOV to give the MVI.

The dataset used to test this algorithm was collected from cameras in downtown Dayton, Ohio controlled by The Institute for Development and Commercialization of Advanced Sensor Technology (IDCAST). It can be difficult to easily monitor all cameras they have access to. In particular, they were having a problem following a moving vehicle across several camera views while at the same time knowing what roads they were on. Therefore, IDCAST is putting all cameras and crime data into one easy to use app and want to integrate this research into their app. After analysis of different images from IDCAST cameras, it became apparent that this approach was not perfect for all images. If the camera had a large RWFOV or was aimed far above the horizon, a large amount of distortion occurred.

Therefore, different methods, such as arbitrarily widening the base or top of the camera field of view in the transform are discussed to remedy this distortion. A discussion of which type of transform to use for certain images and camera positions is included. Future work in this research area will include tracking cars and people in these images using automatic person and car re-identification techniques to give their specific GPS locations.

9407-25, Session 6

A vision-based approach for tramway rail extraction

Matthijs H. Zwemer, ViNotion B.V. (Netherlands); Dennis W. J. M. van de Wouw, ViNotion B.V. (Netherlands) and Technische Univ. Eindhoven (Netherlands); Egbert G. T. Jaspers, ViNotion B.V. (Netherlands); Svitlana Zinger, Peter H. N. de With, Technische Univ. Eindhoven (Netherlands)

INTRODUCTION

With a growing number of urban traffic participants, there is an increasing need for automatic collision assessment systems to support drivers. Trams need relatively long stopping distance and lack maneuvering possibilities. Additionally, the tram driver is overloaded by traffic information due to the busy crossing of traffic participants and the people entering/leaving the tram. Evidently, this leads to occasional dangerous situations or even collisions. Despite the progress in automatic traffic detection, the existing detection techniques for train tracks and car lanes cannot be applied due to the challenging conditions.

We propose a vision-based approach for tram-track detection to support a collision avoidance system, which alerts the driver about possible danger. Our aim is to robustly extract the track under different conditions and cope with partially occluded tracks, e.g. pedestrians at cross-walks. Furthermore, tram track configurations are more complex than train tracks, because of branching, sharp curves and busy surroundings with clutter.

Several studies are carried out on train track detection, a recent method proposed by Ross[1] can reliably detect branches, where a track splits. However, the method cannot detect far ahead and cannot handle occlusions. Gschwandtner et al.[2] uses comparable techniques as our implementation and is able to detect train tracks up to 40 meters distance. However, this system cannot handle occlusions and branches in the track. For our system, we adopt the concept of track segments from Gschwandtner et al.[2] and concentrate on an improved track reconstruction algorithm, in which detected track segments are used to reconstruct the track, while handling occlusions.

OUR APPROACH: TRACK SEGMENT DETECTION AND TRACK RECONSTRUCTION

TRACK DETECTION - Perspective distortion is removed by transforming the camera image to a bird's-eye view by implementing Inverse Perspective Mapping based on Muad et al.[3]. This allows our system to exploit the geometric properties of the track during detection. The transformed image is split into 16 equally-spaced horizontal parts such that the maximum possible curvature of the rails can be represented by a set of straight line pieces (See Fig. 1a). To increase the detection accuracy of the rails, a dedicated rail filter tuned to the width of the rail is used to enhance the rails. The filter response is normalized to the signal energy and a threshold is determined using Otsu's method.

Rail candidates are found in the binary image by detecting straight lines using a RANSAC algorithm adopted from Farin[4]. In our implementation, the RANSAC algorithm creates a hypothesis by selecting two random pixels p(x1, y1), q(x2, y2) from the set of rail pixels P, where the pixel positions satisfy a line-model constraint x = ay+b according to

a = (x2 - x1)/(y2 - y1),

b = (x1*y2 - y1*x2)/(y2-y1),

a <= tan(alpha),

where tan(alpha) represents the slope deviation from the straight forward direction. Multiple hypothesis are tested and the hypothesis which has the most rail pixels in its neighborhood is selected as rail candidate. The pixels supporting this hypothesis are removed and the algorithm starts from the beginning to find another line model until sufficient models are found.

Next, track candidates are constructed by employing a line-pair constraint, which involves the rail parallelism and track width, and applies to all combinations of rail candidates. The generated track candidates are used in our track reconstruction (Fig. 1b).

TRACK RECONSTRUCTION - We propose a novel method that treats track reconstruction as a generic graph problem, in which weights are assigned to edges (connections) between vertices (track candidates). Edges between track candidates are only defined if a path between the track candidates does not exceed the maximum curvature of the rails. Possible occlusions are handled by also connecting track candidates which lie further apart. The weight of an edge between two track candidates is computed using the angle and position of both track candidates.

Since each existing path in the graph represents a feasible combination of track candidates, it is not trivial to directly extract a single track of interest (see Fig. 1c). Many duplicate paths exist in which track candidates are skipped. These duplicate paths are removed by creating a directed acyclic graph without duplicate paths to any vertex (a max cost arborescence graph, see Fig. 1d) using Chu-Liu/Edmonds' Algorithm[5].

Finally, the correct path is chosen by the path likelihood based on its detection scores, and some prior and temporal information about the location of the path.

EXPERIMENTS AND RESULTS

ANNOTATIONS AND DATASETS - Multiple videos are acquired using a camera mounted on top of a tram driving through a crowded city center. Two recordings containing over 3,600 frames, where each frame is manually annotated, enable an extensive evaluation of the system. The frames are divided into datasets containing straight, curved and occluded tracks, allowing for automatic validation of each situation. A track is evaluated as Completely Found (CF), Partially Found (PF) or Not Found (NF).

EXPERIMENTAL RESULTS - Table 1 presents the track detection rates found on a frame-by-frame basis. The system shows promising results, where over 90 % of the annotated straight tracks are completely found and 7 % partially. Only in 2 % of the images, the wrong track is chosen, e.g. a parallel track. Curved tracks still pose challenges, where 60 % of the curved track segments is completely or partially found. Occlusion handling is useful but limited to straight tracks and only small occlusions (e.g. pedestrians/cyclists).

In the full paper we will disclose more experimental results including a comparison with the system of Ross[1] and an evaluation of the computational performance. We conclude that the proposed method for tram track detection offers an attractive solution for use in an urban environment, but its robustness needs further improvement, i.e. by exploiting the tram-network layout.

REFERENCES

[1] Ross, R., "Vision-based track estimation and turnout detection using recursive estimation", Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on, 1330-1335 (2010).

[2] Gschwandtner, M., Pree, W., and Uhl, A., "Track detection for autonomous trains", Proceedings of the 6th international conference on Advances in visual computing - Volume Part III , 19-28 (2010).

[3] Muad, A., Hussain, A., Samad, S., Mustaffa, M., and Majlis, B., "Implementation of inverse perspective mapping algorithm for the development of an automatic lane tracking system", TENCON 2004. 2004 IEEE Region 10 Conference, A, 207-210 Vol. 1 (2004).

[4] Farin, D., "Automatic video segmentation employing object/camera modeling techniques", 401-403 (PhD Thesis, Eindhoven University of Technology, the Netherlands, 2005).

[5] Tarjan, R. E., "Finding optimum branchings", Networks 7(1), 25-35 (1977).

9407-26, Session 6

Exploration towards the modeling of gable-roofed buildings using a combination of aerial and street-level imagery

Lykele Hazelhoff, Ivo M. Creusen, CycloMedia Technology B.V. (Netherlands); Peter H. N. de With, Technische Univ. Eindhoven (Netherlands)

Introduction

Buildings are one of the most commonly occurring and prominent man-made structures and represent a considerable value in society. Extraction of building properties is thus valuable for several stakeholders. For example, the orientation w.r.t. the south and the corresponding roof area are interesting for companies in the solar panel industry, as these parties could approach real-estate owners with propositions for solar panel placement. Besides this, the building volume and dimensions are interesting parameters for e.g. insurance companies and governmental bodies tasked with real-estate taxes. Furthermore, utility companies are interested in measuring e.g. the density of solar panels in a town, or in finding specific buildings suitable for placing communication devices.

These building properties can be extracted manually from widely available remote sensing data, such as aerial or street-level image databases, e.g. by browsing all images and manually annotating the properties of interest, or by searching for buildings that satisfy certain criteria. Such an analysis is error prone and time consuming, so that automated analysis of these databases is desirable. This work extends our previous study on building detection [1] and explores the design of an automated system for modeling of buildings, a task widely studied in literature (see e.g. [2]). We focus on the modeling of standard buildings with a gable roof (the most prominent house type in Western Europe), and thereby extracting important parameters such as size, volume and orientation.

Our approach

This work explores modeling of buildings based on a combination of aerial stereo imagery and street-level panoramic images, where we aim to combine the strong points of both image sources, as used for geo-localization in [3]. The aerial images typically provide an accurate overview of the building footprint, but estimation of the building height is often only accurate up to 0.5 m. This height can be measured more accurately (up to 10-cm accuracy) from the street-level panoramic images, but these images only show a front or side view of the building, where information about the building footprint cannot be retrieved. As a consequence, exploiting the combination of both image sources results in a complete and at the same time accurate estimate of the building properties.

We follow a model-based approach as the building appearance varies enormously. These variations result from (1) variations in buildings themselves (aspect ratio/size/type), (2) customized building extensions (verandas, garages, dormers, chimneys, etc), and (3) capturing conditions (lighting conditions, contrast, occlusions, etc). The use of a building model provides robustness towards these deviations, while still capturing the most important building properties, and allowing for the easy extraction of meaningful information from the model parameters. This approach has the advantage that the extracted models can be re-used for a variety of applications, e.g. detection of objects on each roof side, such as solar panels, dormers and roof windows, but also building value estimation and even can aid 3D city reconstruction.

System overview

The building modeling system consists of two consecutive stages. The first stage analyzes aerial stereo images to retrieve both a footprint of the building and an initial estimate of the height of both the roof ridge and roof gutters. The second stage refines these height measurements based on the analysis of nearby street-level panoramic images, thereby enabling more accurate estimations of the model parameters.

The first stage is initialized with a GPS position of the building, which is extracted from a database of address locations (or may be inserted manually). Based on this position, all nearby aerial stereo images are selected, and from each image containing the coordinate, a bounding-box window around the position is extracted. Within each bounding box, straight line segments are retrieved, and the longest line segment closest to the input coordinate is used as roof ridge, resulting in an estimate of the building length and orientation. Next, the roof gutter locations are estimated, as well as the gutter end-points. The ridge and gutter corner end-points are then used to estimate the height of both the ridge and gutters using triangulation. This results in an initial building model, containing the footprint, orientation and height of both the roof ridge and gutters.

The second stage refines the estimated heights based on nearby street-level images. In several frontal images, the roof and gutter lines are identified, as well as the side(s) of the building. Based on the building orientation extracted from the aerial images, we estimate the vanishing point of the roof/gutter lines in the street-level image. We apply a custom edge detector which boosts lines that intersect the horizon near the computed vanishing point, to locate the roof and gutter lines, and locate these lines using the model from the aerial image as an initial estimate. Next, the end-points of these lines are determined, and by combining the detected end-points from nearby street-level images, their 3D position can be triangulated. These confirmed and detected 3D points allow us to refine the building model with more accurate building height and dimensions. This results in a better model, and allows for a more accurate estimation of the important building parameters.

Preliminary results and future work

We currently have developed a prototype of the above-described system, and are aiming at a more complete framework for large-scale performance evaluation. The house street-level analysis is still ongoing work and is reaching a first prototype.

The performance evaluation focuses at three different aspects: the building footprint (i.e. area), the building orientation and the accuracy of the full building model.

This process involves the generation of ground truth to numerically assess the quality of the found parameters. However, accurate estimation of all model parameters is rather difficult, as this involves manual measurements of the corresponding parameters using the same images as used for processing. In case the parameters of interest are not clearly visible, inaccurate manual measurements are expected, and inaccuracies in the source data themselves propagate into the ground-truth parameters. This accuracy analysis is also further studied.

Initial results on about 20 different houses show that the building length and width can be accurately estimated from the aerial images, while the corresponding building heights are estimated within a margin of about 50 cm w.r.t. manual measurements performed in the street-level images. The building heights can be estimated more accurately from the street-level images, where a deviation reduction over a factor of two is expected. But since the building appearances vary considerably, a bigger test set is necessary to accurately estimate the performance in real-world circumstances. Therefore, extensive results on a larger test set will be included in the final paper.

References

[1] Hazelhoff, L. & With, P.H.N. de (2011). Localization of buildings with a gable roof in very high-resolution aerial images. Proceedings of Visual Information Processing and Communication II , 22-24 January, 2011, San Francisco, California, USA, (Proceedings of SPIE, 7882, pp. 788208).

[2] Jaynes, C et al. (2003). Recognition and reconstruction of buildings from multiple aerial images. Computer Vision and Image Understanding, vol. 90, no. 1, pp 68-98.

[3] Bansal, M. et al. (2011) Geo-Localization of Street Views with Aerial Image Databases. Proceedings of the 19th ACM international conference on Multimedia, pp. 1125-1128.

9407-27, Session 6

On improving IED object detection by exploiting scene geometry using stereo processing

Dennis W. J. M. van de Wouw, Technische Univ. Eindhoven (Netherlands) and ViNotion B.V. (Netherlands); Gijs Dubbelman, Peter H. N. de With, Technische Univ. Eindhoven (Netherlands)

1. INTRODUCTION AND PROBLEM STATEMENT

In order to reduce casualties by IEDs, reliable detection of changes along roads is of primary importance. Change detection on aerial and satellite images is widely researched and already applied to Countering IEDs. However, analysis of aerial footage has its limitations, such as resolution because of the large capturing distance and the uniform viewing point. Therefore, the interest for mobile change detection systems for ground-based surveillance is growing. This is a challenging task [1], where images from different viewpoints and time instants, have to be accurately registered and compared in real time. This paper aims at improving the robustness of such a change detection system using scene geometry acquired by a stereo camera. The proposed solution addresses the main challenges of the monocular system and aims at a more reliable system, especially in urban environments, and a higher robustness to strong shadows.

2. RELATED WORK AND CONTRIBUTIONS

Ground-based change detection is a challenging task. Some challenges, such as inaccurate image registration in poorly textured road scenes and false positives due to local shadows, are difficult to solve with a monocular setup [1]. Haberdar and Shah tackle the first problem by applying depth sensing [2]. They assume changes are located on the ground surface and register the ground planes instead of the entire image. In their work, the ground plane is found by segmenting the disparity map. Wathen et al. employ a LIDAR sensor combined with a color camera to populate a point cloud [3]. Change detection is performed point-by-point on the 3D-point cloud, hence reducing the effect of shadows which do not affect the depth.

This paper extends the work of van de Wouw et al. [1] with depth sensing through the addition of a second camera, yielding a fixed stereo setup. Similar to Haberdar and Shah [2], we register the ground plane of the live and historic recording. However, instead of using image segmentation, the ground plane is extracted by fitting a spline through the 3D point cloud. This allows for multi-planar warping in the case of non-planar ground planes. Our approach improves image registration, which results in more accurate change detection. Next, we exploit depth information to improve the change detection accuracy. We propose a post-filtering metric to reduce sensitivity to shadows in the scene, by validating the detected changes from the preceding stage. Localized 3D point clouds are used to distinguish real objects from shadows, exploiting the knowledge that shadows do not affect the depth of the scene. The resulting system is extensively tested on manually annotated real-world videos, similar to the earlier referred approach [1].

3. DEPTH EXTENSIONS

We extend the original change detection system with depth sensing, where the most significant changes w.r.t. the monocular system are now described.

3.1 STEREO CAPTURING

Images are captured by a stereo camera consisting of two state-of-the-art cameras with an adjustable baseline of up to 150 cm. At full-HD resolution, an object of 10x10x10cm will have a disparity difference of 5.7 and 3.8 pixels w.r.t. its background at a distance of 40 and 60 m, respectively. Taking into account that the disparity map has a sub-pixel resolution of 1/16th pixel, this is more than sufficient to accurately distinguish objects from their background.

3.2 IMAGE REGISTRATION

The matched features between historic and live images are the basis for image registration. In this case, only those features residing on the ground plane are used. This ground plane can be accurately determined by fitting a spline through the y-z histogram of the 3D point cloud obtained from the stereo camera [4]. The ground plane is then obtained by finding all triangulated points whose (y, z)-values satisfy the spline. This method easily extends to multi-planar warping in the case of non-planar ground surfaces, by finding piece-wise linear segments in the spline.

Instead of merely selecting all features that lie on the ground plane, new features are extracted within the ground plane region. Together with feature selection techniques, such as Adaptive Non Maximal Suppression [5], the feature distribution over the ground plane is significantly improved. We have found that this positively affects the robustness of the image registration.

3.3 DEPTH FILTERING

To reduce the sensitivity to shadows, an additional change characterization module is added. For each change blob found by the monocular change detection system, the depth map is (locally) compared to the depth map of the historic scene. A change blob representing a shadow will not lead to any differences in the depth map, while a physical change, e.g. an object, will show clear changes and discontinuities. The challenge lies in coping with different viewpoints (caused by driving a different path), which result in a difference in the relative depth between the live and historic scene. Therefore, the localized 3D point clouds are first co-registered and then compared. If the resulting clouds differ significantly, the change blob is accepted as a true change.

4. EXPERIMENTS AND RESULTS

Our extended change detection system is extensively evaluated similar to that of the earlier work [1]. The system is mounted on a vehicle and videos are acquired while driving through urban and rural environments. After the first recording, wooden test objects of 10x10x10cm in multiple colors are placed in the environment at predefined locations. These objects are manually annotated and assigned the correct GPS location, after which the system can be automatically validated. After placement of the objects, a second video is captured by driving the same route now containing the manually placed test objects. This process will be explained in more detail in the final paper.

The proposed depth-sensing solutions shows promising results, where false positives are reduced by more than 30% and that reliability improves in urban scenes. An extensive quantitative evaluation will be presented in the final paper.

5. ACKNOWLEDGEMENT

The research leading to these results has received funding from the D-SenS project in the European Union's FP7 program, managed by REA - Research Executive Agency under grant agreement "FP7-SME-2012".

9407-28, Session 6

Visual analysis of trash bin processing on garbage trucks in low resolution video

Oliver Sidla, Gernot Loibner, SLR Engineering GmbH (Austria)

Introduction

The recycling of garbage is standard practice in Europe and is becoming increasingly important in more countries world wide. Normally the waste is sorted manually into several categories (paper, glass, metal, organic waste, and non-recyclable waste) and then deposited into according bins. These garbage bins are shared amongst households in apartment buildings, houses in less dense populated areas keep their own bins, which are provided and maintained by the municipality.

Once a week garbage collecting trucks roam through their designated area on predefined paths to pick up and empty the bins. On each truck tour several hundred bins are thus handled and processed. Although there is a specified path which is followed by the trucks, and the addresses visited are generally known, information about the exact number of trash bins collected and their precise location is incomplete.

This work aims at improving knowledge about the location and number of garbage bins by monitoring a truck over the period of a whole day and automatically detecting and counting all processed bins. We present a vision based system which is able to detect, track and count the garbage bins

even in adverse imaging situations. These results of can provide valuable information for both the operating companies and municipalities, e.g. they can be used to optimize tour planning.

Vision System Overview

The purpose of our work was to design a vision prototype which can

• Detect garbage bins (1-2) as they are mounted on the truck for processing.

• Differentiate between 2 bin sizes.

• Count each type of bins.

• Detect the location (left, right, center) relative to the truck at which they are put back on their resting position on the street.

Trash bin detection is achieved using videos from monitoring cameras, which are mounted on the back side on top of the garbage collecting trucks. There is generally no illumination present to support the cameras, and due to limitations of the environment, technology and budgets, the image quality is very poor. Especially in dark situations (winter time, early morning) the videos contain huge amounts of noise and distortion due to the resulting compression artifacts.

A working prototype has been successfully implemented and it demonstrates that our approach for detection and tracking of the trash bins can deal with all challenges in a satisfying way. Our vision pipeline has been tested on more than 15 hours of video in which all detections have been checked and verified manually.

Method and implementation

The challenges at hand can be classified into the following sub-problems for which we present solutions in this work:

1) Detect different types of trash bins in low resolution and very noisy video sequences.

2) Analyze the processing sequence for a bin

a) mounting of the bin

b) clearing of its content into the truck

c) dismounting of the bin from the truck

3) Re-location to storage area – detection of final location relative to truck.

We have implemented a combination of detector, tracker and basic image analysis methods to cope with problems 1)-3) above:

Trash bin detection

The shape of the trash bins is basically rectangular, but due to the very wide field of view of the camera and the large amounts of noise, this shape can be distorted. Therefore direct modeling and detection of the bin top as a rectangle was disregarded. Instead we trained and used a HOG detector for the bins when mounted (and thus generally well aligned) on the truck.

Bootstrapping the detector with false positives improved its robustness significantly, this process also allows for much faster training because only a limited initial number of samples need to be provided manually. Typically about only 20 samples are sufficient to bootstrap for this type of rigid objects. After final training, we can achieve a verified recall rate of >98% at an accuracy of >99% with the proposed detector.

Analysis of bin processing on the truck

The detector is tuned to localize trash bins as they are mounted on the truck prior to clearing them in the waste compartment. This situation is visually present for he monitoring camera at least at two points in time, namely when it is mounted and then dismounted from truck. Sometimes the situation is ambiguous because a bin can be pushed into the truck's waste compartment more than once if it has not been completely emptied in the first try. It is therefore necessary to split this clearing process into its phases so that a bin is counted only once.

The monitoring camera on the back of the truck views the bin from top, and during the clearing process it is moved up towards the camera. We first tried to detect and analyze the apparent motion of the bin, respective the resulting optical flow, to detect this lifting motion. Unfortunately, due to the large amounts of noise and limited structure on the bin top, the motion information is not reliable enough. Instead, we resorted to a simpler solution, which uses change detection around the area of the truck compartment where the bin is pushed through:

1. Just prior after first detection record a reference image region Ir through which every bin must pass during the clearing operation

2. Observe Ir until a substantial change has occurred – mark the bin as 'cleared'

3. Wait for next detection of the bin in lowered position (street level)

4. Increase the bin counter and record a timestamp

In order cope with failing detections in all stages of the pipeline described so far, we have added timeouts so that the detector can re-initialize itself to back into a valid working state without getting confused.

Bin tracking after dismounting

After the bin has been lowered back to street level it is re-detected again (Step 4 above) with HOG. At this stage it is counted, and in order to find its final resting position on the street side, we try to follow its path as it is being carried away by the operator.

The colors of bins differ significantly from the color of asphalt and the steel gray of the truck. We can therefore use camshift, color meanshift tracking, to track the color blob of the trash bin top through the camera field of view. This approach works well in many cases, and a success rate of >90% for directional tracking can be achieved, which is sufficient for our application.

After the blob disappears at the image border, the complete bin clearing analysis cycle is finished. Confusing situations like operators keeping bins a long time on the truck, bins falling, etc. are handled using the built in timeouts.

Summary

We present a framework for pure vision based trash bin detection, counting and tracking. The prototype system works with good reliability and it is robust enough to be used even in very noisy and adverse imaging situations. The robustness of our framework is demonstrated using 15 hrs of real-world video data and comparison to manually extracted ground truth.

9407-29, Session 7

Toward to creation of interaction models: simple objects-interaction approach

Teresa Hernández-Díaz, Juan Manuel García-Huerta, Alberto Vazquez-Cervantes, Hugo Jiménez-Hernández, Ctr. de Ingenieria y Desarrollo Industrial (Mexico); Ana M. Herrera-Navarro, Univ. Autónoma de Querétaro (Mexico)

Day to day, surveillance systems become more common and they are considered as a cheaper way to supervise, monitor and take decisions over certain scenarios. However the usefulness of them are dismissed with the extra work related and the data analysis that they involve. The amount of sensors installed generates huge volumes of information, which in optimal conditions are useful when user attention are focused in relevant events. In other situations, the information is used as log post-relevant situations happen. This is, after the consequences of relevant situation involves in video, the majority of sources are used as data log of past events. In this situation, automatic event detection has become more important. Today, technical and theoretical infrastructure allow to develop efficient approaches in well-delimited scenarios. Several approaches as [2-7] show capable approaches to detect and infer actions in controlled situations. However, only a few of them [8-10], offer approximation to model and detect more complicated action in which more than one object is involved. Such as the case of two person displacing a table, or two person shaking their hands, or persons changing a package, etc. This level of descriptions is more complicated because involve the semantics of the scenario and the plane-projection of the current camera. Against to the difficulty to be characterized, these kind of actions are more common to be performed and more informative in decisions' taking process.

Technically this kind of actions are more complicated because they turn to be invariant to the spatiality that could be implicit in the camera and the set of features involved are not-fixed or unrepeatable. This proposal establishes the basis to develop a framework to model interaction with objects in motions. The aim of this work consists on dynamically follow a set of features and model the behavior of them along time. These features

are associated to objects in motions and aspects as merge, division or occlusion are considered as primitives to develop causal relation among them. Interactions to be covered include those in which two or more distinguishable objects affects directly the motion of others; i.e. a distant salute, or a cooperative action to be performed (change a place other objects, in which there are three objects in interaction). This proposal consists on elaborate a graph model based on causal dependencies of individual clusters discovered in all video frames and its affectation respect to the others.

The proposal firstly select an efficient way to represent objects in motions. Literature says that the problem of tracking is open because there are several variables non-controllable in a scenario which warrant a repeatability in the estimation process [7]. Then, this work deals the motion representation as a selection of dynamic features that results distinguishable and invariant to certain image affectation over time. The dynamic selection allows to discard and add new features over time. Being specific, these features are represented as the Harris features, (please note that other general feature approach can be suitable too) [11-14].

Locally for a current frame of video, a set of characteristic are dynamically detected, the object motion labeling consists on generated clusters with all characteristic detected. Clustering approach is used as the way to group locally the features and describe them as cumulus without the needed to make a registration process of each characteristic. Even each frame, the characteristics that represents each cluster change, centroids are hardly affected. This allows to associate for each pair of frames I_t and I_(t+1) a centroid of a particular cluster (c_t^i,c_t^j). Note that the number of cluster should be dismissed or augmented. This mean that objects are merged or separated. All centroid can be characterized in follows situations:

When they become stable in several time stamps.

When they become merged with two or more cumulus.

When they are separated.

First situation stands for the stability of the cumulus and represents the local historical time-stamp, where we can use as local behavior. Second situation implies that a new cumulus is generated, then the historical can be added to the new cumulus; but we must be careful because there may exist n historical options that correspond to the number of cumulus merged. A simple criterion is based on the bigger cumulus, this is, those which have more features. Finally the third situation implies that now two separated cumulus exist and they will have the same historical information.

This way to describe cumulus can be used as evidence for testing causal dependencies. Each cumulus generated in time is considered as a variable. Using causal theory, dependencies of changes in spatiality or in behavior (mixing or separating) are located. These dependencies reflect the affectation of cumulus which is related with most distinguishable information in each frame that correspond to objects in motion; i.e. this give an approach to a graph dependency of interaction in simple objects.

Finally, this approach is tested in controlled/not-controlled scenarios. Controlled scenarios include actions to be performed in close work environment (office, lab, and classmate). Not-controlled scenarios include open access between buildings in a university. Finally, to warrant the reliability in general scenarios, we tested with several sequences taken from PETS database.

Bibliography:

[1] Pearl, J. [Causality: models, reasoning, and inference], Cambridge university press, (2000)

[2] Poppe, R. and Poel, M., “Discriminative human action recognition using pairwise CSP classifiers”, Proc. FGR), 1-6 (2008).

[3] Wang, Y., Jiang, H., Drew, M., Ze-Nian, L. and Mori, G.,“Unsupervised discovery of action classes”, Proc. CVPR, 2, 1654-1661 (2006).

[4] Yang, C., Guo, Y., Sawhney, H. and Kumar, R., “Learning actions using robust string kernels”, Proc. HUMO, 4814, 313-327 (2007).

[5] Ke, S.-R., Thuc, H.L.U., Lee, Y.-J., Hwang, J.-N., Yoo, J.-H. and Choi, K.-H., “A Review on Video-Based Human Activity Recognition”, Computers, 2(2), 88-131 (2013).

[6] Aggarwal, J.K. and Ryoo, M.S.,”Human activity analysis: A review”, ACM Comput. Surv. 43(3), 16:1–16:43 (2011).

[7] Poppe, R.,”A survey on vision-based human action recognition “, Image and Vision Computing, 28(6), 976-990 (2010).

[8] Oliver, N.M., Rosario, B. and Pentland, A., “A Bayesian computer vision system for modeling human interactions”, IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(8), 831-843 (2000).

[9] Hu, H., Tan, T., Wang, L. and Maybank, S., “A survey on visual surveillance of object motion and behaviors”, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 34(3), 334-352 (2004).

[10] Ogale, A.S, Karapurkar, A. and Aloimonos, Y., “View-invariant modeling and recognition of human actions using grammars”, WDV, 4358, 115-126 (2007).

[11] Shi, J. and Tomasi, C., “Good Features to Track”, In IEEE Conf. CVPR, (1994).

[12] Zhou, H., Yuan, Y. and Shi, C., “Object tracking using SIFT features and mean shift”, Comput. Vis. Image Underst., 113(3), 345-352 (2009)

[13] Schmid, C., Mohr, R. and Bauckhage, C., “Evaluation of interest point detectors”, Int. J. Comput. Vision, 37(2), 151-172 (2000)

[14] Tuytelaars, T. andMikolajczyk, K., “Local invariant feature detectors: a survey”, Found. Trends. Comput. Graph. Vis., 3(3), 177-280 (2008)

9407-30, Session 7

Compressive sensing based video object compression schemes for surveillance systems

Sathiya N. Sekar, Anamitra Makur, Nanyang Technological Univ. (Singapore)

1. INTRODUCTION

Most video compression techniques consider video frame as a random signal and achieves compression by exploiting its stochastic properties. In applications such as video telephony, video surveillance and medical imaging, foreground objects can certainly be segmented and coded efficiently. Particularly in the case of surveillance videos, if the foreground moving objects are segmented from the background, they can be coded independently requiring far fewer bits compared to frame-based coding.

2. RELATED WORKS

An object based video compression framework was proposed in which the objects in the video frame are detected, tracked and then coded using the coefficients of the most significant principle components obtained via incremental principle component analysis [1]. Compressive Sensing (CS) theory ensured the recovery of a sparse signal using a small number of linear observations. Exact reconstruction was achieved using convex relaxation technique (e.g. Basis Pursuit) or greedy algorithms (e.g. Orthogonal Matching Pursuit). Motivated by CS, Huang et al proposed a Video Object Error Coding method (CS-VOEC) in which the foreground moving objects are segmented and object-based motion compensated from the previous frame, and then the sparse object error is coded using CS principle [2].

3. CONTRIBUTIONS

In the aforementioned techniques, motion estimation at the encoder might be computationally intensive. Motivation of CS on video is to have a simple encoder. Encoder can be kept simple by pushing the motion estimation from encoder to decoder. Therefore,we first propose a novel CS based Video Object Compression (CS-VOC) technique having a simple encoder and an appropriate decoding scheme. For scenarios where the object segmentation requires more computations (depending upon the number and size of objects), we propose a Distributed Compressive Video Sensing based Video Object Compression (DCVS-VOC) framework wherein the object segmentation is required only for certain frames.

3.1 CS BASED VIDEO OBJECT COMPRESSION (CS-VOC)

Encoder: At first, the foreground objects are segmented from the background using an edge-based object segmentation scheme (as in CS-VOEC). Object block is obtained for each segmented object by creating a

rectangular bounding box around the object using the object mask. Then the object block is sensed using a random CS matrix. Location,shape, dimension and CS measurement vector of the objects are transmitted to the decoder.

Decoding Scheme using Motion Estimation at the decoder: At the decoder, we consider the problem of reconstructing the object error block given its measurement vector and the previous reconstructed frame. First step is a novel object motion estimation (using CS measurements). Upon obtaining the object motion vector, motion compensation is performed. The motion compensated object error is sparse, and therefore, it can be efficiently reconstructed using CS recovery algorithms like Basis Pursuit (BP). Since we have additional information in the form of object mask, we propose to use Modified-CS (MOD-CS) for solving the CS problem. Video object block is recovered by adding the obtained object error to the motion compensated object block. The video frame can then be recovered using the reconstructed object blocks, object mask and the available background.

3.2 DCVS BASED VIDEO OBJECT COMPRESSION (DCVS-VOC)

Encoder: As in [3], video frames are classified into key frames (I-frames) and non-key frames (CS-frames). For key frames, objection segmentation is performed. These frames are then coded using conventional video compression. Key frames are transmitted periodically after a certain number of CS frames. This is similar to the Group-of-Pictures (GOP) in video coders. For CS frames, a tentative object mask is obtained by dilating the object mask of the preceding key frame. A disk-shaped structuring element (of radius = GOP size) is used for dilation. CS frames are then coded as in the case of CS-VOC by using the tentative mask.

Decoder: At the decoder, key frames are first reconstructed via conventional decoding. For recovering CS frames, a decoding procedure similar to that of CS-VOC is applied. Motion estimation in DCS-VOC decoder is performed with respect to both preceding key frame and succeeding key frame separately. Therefore, two motion vectors are obtained. The motion vector indicating the smallest motion and the corresponding key frame are considered for motion compensation. Reconstruction of sparse object error and the recovery of object block are done exactly in the same manner as in CS-VOC.

3.3 COMPLEXITY ANALYSIS

Our aim was to design a simple encoder and then to propose an appropriate decoding scheme. In CS-VOEC encoder, motion estimation and compensation requires O(nS) computations (S is number of search points for motion estimation and n is the signal length). Therefore, compared to existing encoders, encoders of CS-VOC and DCVS-VOC have a computational savings of at least O(nS). Decoder complexity of CS-VOEC, CS-VOC and DCVS-VOC will be of the order of computations required by the CS reconstruction algorithm.

4. EXPERIMENTAL RESULTS

We present the decoding performances in terms of object-PSNR (PSNR considering object pixels only). We used 80 video frames (size 288 X 384) from the 'walk' sequence available in CAVIAR.

Experiment-1: In this experiment, BP is used for CS reconstruction. Performance of CS-VOC is much closer compared to that of CS-VOEC. Also, DCVS-VOC gives the best performance compared to CS-VOEC and CS-VOC. For an under-sampling ratio (number of CS measurements/signal length) of 0.3, if the GOP size is fixed as 4, DCVS-VOC gives an average object PSNR of 26.84 dB whereas CS-VOEC and CS-VOC gives 23.16 dB and 22.71 dB respectively. Time taken by DCVS-VOC encoder for coding those 80 frames is 24.94 seconds whereas the times taken by CS-VOEC and CS-VOC encoders are 84.12 seconds and 81.95 seconds respectively.

Experiment-2: This experiment is to verify the improvement in CS reconstruction performance due to MOD-CS. Two CS reconstruction algorithms were applied: BP and MOD-CS. MOD-CS (using the object mask information) gives an improved performance compared to BP. CS-VOC using BP results in an object PSNR of 22.71 dB for an under-sampling ratio of 0.3 whereas CS-VOC using MOD-CS results in 23.40 dB for a ratio of 0.2. Therefore, effective use of mask information reduces the number of measurements required to sense the object block.

5. CONCLUSION

We proposed a novel CS based video object coding (CS-VOC) technique having a simple encoder. Compared to CS-VOEC, proposed CS-VOC not only has simple encoding but also gives a comparable decoding performance. In order to simplify the encoder further, we proposed a distributed framework (DCVS-VOC) wherein the object segmentation is done only for the key frames. DCVS-VOC gives better reconstruction performance compared to CS-VOC and CS-VOEC. We have shown that the CS reconstruction can be improved by applying MOD-CS in the place of traditional reconstruction algorithm (BP) thereby making effective use of the mask information available.

REFERENCES

[1] Asaad Hakeem, KhurramShafique, Mubarak Shah, ``An Object based Video Coding Framework for Video Sequences Obtained From Static Cameras," MULTIMEDIA '05, Proceedings of the 13th annual ACM international conference on Multimedia, Nov. 2005.

[2] H. Huang, A. Makur, and D. Venkatraman, ``Video object error coding method based on compressive sensing," Control, Automation, Robotics and Vision, 2008. ICARCV 2008. 10th International Conference on, vol., no., pp.1287-1291, Dec. 2008.

[3] T.T. Do, Yi Chen, D.T. Nguyen, N. Nguyen, Lu Gan, T.D. Tran, ``Distributed compressed video sensing," Image Processing (ICIP), 2009. 16th IEEE International Conference on, vol., no., pp.1393-1396, 7-10 Nov. 2009.

9407-31, Session 7

Improved colorization for night vision system based on image splitting

Ehsan A. Ali, Samuel Kozaitis, Florida Institute of Technology (United States)

The success of a color night navigation system often depends on the accuracy of the colors in the resulting image. We presented a method to improve the color accuracy of a night navigation system by initially splitting a fused image into two distinct sections before colorization.

Operating in a degraded visual environment due to darkness can pose a threat to navigation safety. Systems have been developed to navigate in darkness that depend upon differences between objects such as temperature or reflectivity at various wavelengths. Such systems use a combination of passive and active sensors to discriminate between objects. However, adding sensors increases the complexity of a system by adding multiple components that may create problems with alignment and calibration. An approach is needed that is passive and simple for widespread acceptance.

Our system was designed for terrestrial vehicle navigation in dark conditions and used color information from a public database of images to assign color information to an intermediate image. This image was obtained by combining two spectral bands of images, thermal and visible, in an effort to enhance night vision imagery. However, the fused image gave an unnatural color appearance. Therefore, a color transfer based on look-up table (LUT) was used to replace the false color appearance with a colormap derived from a daytime reference image. The reference image was obtained from a public database using the GPS coordinates of the vehicle. Using this approach, we were able to produce imagery acquired at night that appeared as if in the daylight.

One problem is that the colors in relatively large regions of the fused image such as roads and sky regions appear similar and therefore will be assigned similar colors in the final image.

We considered an approach that depends to some degree on what is expected. For example, when driving a ground vehicle, a road is typically at the bottom of a scene, and the sky is at the top. We used this concept to aid the development of a passive system to improve safety while driving a ground vehicle at night. We split the fused image into two sections, generally road and sky regions, before colorization and processed them separately to obtain improved color accuracy of each region. It was not necessary to separate the regions precisely, only to separate the dominate region in terms of area. Therefore, separating the image could be done quickly, which is suitable for real-time operation. Colorizing each region created noticeable errors where the image was separated so we "shuffled" the two colormaps into a single colormap before colorizing the image. This

approach eliminated problems at the boundary where the two images were separated.

Another problem in the original system was due to the fused and database images not being registered perfectly. Highly accurate registration was difficult because of the different sensors, conditions, times, and positions used. Therefore, small objects were sometimes completely missed because the colormaping process was generally dominated by two large groups of colors. Since the new approach increased the accuracy of the colors, some objects were more visible.

9407-32, Session 7

Evaluation of maritime object detection methods for full-motion video applications using the PASCAL VOC challenge framework

Shibin Parameswaran, Space and Naval Warfare Systems Ctr. Pacific (United States); Martin Jaszewski, Space and Naval Warfare Systems Command (United States); Eric Hallenborg, Bryan Bagnall, Space and Naval Warfare Systems Ctr. Pacific (United States)

Objectives: We present the results of our efforts to build an anomaly detection performance evaluation system for our RAPid Image Exploitation Resource (RAPIER)® Full Motion Video (FMV) system. We evaluate a number of algorithms for real-time maritime anomaly and target detection in full motion video from fixed and moving video sources. We take a pragmatic approach to the evaluation; using appropriate performance criteria, we address tradeoffs between detection accuracy and computational speed/ throughput.

System Overview: RAPIER® FMV is a system designed to detect and track maritime objects from a variety of fixed and moving video sources. It is intended to reduce the workload of video analysts by automatically cueing the analyst to targets or anomalous objects and providing basic information about the location and status of the detected object. RAPIER® FMV is designed to work either as a forensic tool for previously recorded video or as an alerting tool for real-time video streams. Its effectiveness, therefore, hinges on the selection of a detection algorithm that balances accuracy and speed.

Conference 9408: Imaging and Multimedia Analytics in a Web and Mobile World 2015

Wednesday - Thursday 11–12 February 2015

Part of Proceedings of SPIE Vol. 9408 Imaging and Multimedia Analytics in a Web and Mobile World 2015

9408-1, Session 1

Recent progress in wide-area surveillance: protecting our pipeline infrastructure *(Keynote Presentation)*

Vijayan K. Asari, Univ. of Dayton (United States)

The pipeline industry has millions of miles of pipes buried along the length and breadth of the country. Since none of the areas through which pipelines run are to be used for other activities, these areas need to be monitored so as to know whether the right-of-way of the pipeline is encroached upon at any point in time. Rapid advances made in the area of sensor technology have enabled the use of high end video acquisition systems to monitor the right-of-way of pipelines. The images captured by aerial data acquisition systems are affected by a host of factors that include light sources, camera characteristics, geometric positions and environmental conditions. We present a multistage framework for the analysis of aerial imagery for automatic detection and identification of machinery threats along the pipeline right of way, which would be capable of taking into account the constraints that come with aerial imagery such as low resolution, lower frame rate, large variations in illumination, motion blurs, etc. The complexity of large variations in the appearance of the object and the background in a typical image causes the performance degradation of detection algorithms. Our novel preprocessing technique improves the performance of automatic detection and identification of objects in an image captured in extremely complex lighting conditions. A background elimination method employing a relative variance and local entropy based analysis has been developed and it is found to be very effective in reducing the search regions in the aerial imagery for threat detection. Our object detection algorithm can automatically detect and identify machinery threats such as construction vehicles and equipment in the regions designated as the pipeline right-of-way. Our detection algorithm makes use of monogenic signal representation to extract local phase information. A novel classifier using a matching criterion along with a threshold for minimum distance is used to filter out false detections. The algorithm has been successfully tested on the aerial imagery containing different classes of construction equipment.

9408-2, Session 1

Alignment of low resolution face images based on a 3D facial model

Lu Zhang, Jan Allebach, Purdue Univ. (United States); Xianwang Wang, Qian Lin, Hewlett-Packard Co. (United States)

Faces often appear very small in surveillance videos since there is typically a large distance between the camera and the scene. Because of the importance of face recognition in many applications, face hallucination or super resolution of face images has become a thriving research field. Most learning based image synthesis models for face hallucination require alignment between test image and the training dataset. So one of the greatest challenges of face problems is the difficulty of aligning faces in low resolution images even when using the same face detection system.

Liu et al. proposed a robust warping algorithm to align low resolution faces [1]. The alignment approach finds an affine transform to warp the input face image to a template to maximize the probability of the aligned low-resolution face image. Unlike many other learning-based face alignment methods, it doesn't need to extract any local facial features that normally requires a larger number of face pixels. Since the affine transform only works for 2D face images, the limitation of this method is the requirement of frontal face images. Since, in real images the faces are normally not perfectly frontal; the frontal face must be estimated.

Therefore, in this work, we propose an improved 3D face alignment approach for 2D images based on an active shape model, for low-resolution images. We train a 3D face shape model with different view-based models from a 3D face database, then generate corresponding 2D face images with -90 to 90 degree rotation angles, in both horizontal and vertical directions. Our approach first determines the rotation angles using a search algorithm that minimizes the average error between the test image and the rotated template. Then, our method is able to synthesize a corresponding frontal image. Last but not least, to maximize the probability of low-resolution aligned face image, we use an eigenspace representation, similar to Liu's method. To make it robust, the algorithm also explores multiple starting points to achieve the best alignment result.

The effectiveness of this method is shown by experimental results with aligned low-resolution face images, generated high quality hallucinated faces, and thus an improved face recognition score.

[1] Liu, Ce, Heung-Yeung Shum, and William T. Freeman. "Face hallucination: Theory and practice." International Journal of Computer Vision 75.1 (2007): 115-134.

9408-3, Session 2

Piecewise linear dimension reduction for nonnegative data

Bin Shen, Purdue University (United States); Qifan Wang, Jan Allebach, Purdue Univ. (United States)

Dimension reduction is playing an increasingly important role in many computer vision and pattern recognition tasks due to the rapidly growing large scale data with high dimensionality.

It reduces the number of random variables under consideration, which not only saves computational and storage resources but also helps overcome the curse of dimensionality.

To explore the hidden structures of the original high dimensional data samples, ${x^i = (x^i_1, dots, x^i_m)^{top} | i = 1,dots, n}$, dimension reduction techniques find low dimensional representation, ${y^i = (y^i_1, dots, y^i_l)^{top} | i = 1,dots, n}$, where $l < m$, according to some certain criterion to capture the content in original data. The algorithms may be either linear or nonlinear. Typical linear methods, including principal component analysis (PCA), independent component analysis (ICA), nonnegative matrix factorization (NMF) and etc., are usually simple and efficient, while nonlinear dimensionality reduction techniques, such as locally linear embedding (LLE), Isomap, kernel tricks and Laplacian eigenmaps, allow nonlinearity during the dimension reduction to capture more property of data, for instance, underlying manifold structure, and thus to overcome the limitation of linear models. However, nonlinear ones are usually more computationally expensive, sometimes prohibitively.

NMF has been widely used in applications such as classification and clustering due to its easy theoretical interpretation and {desired practical performance}. It aims to approximate nonnegative high dimensional data by product of a low-rank basis matrix and another low-rank coefficient matrix, both of which are nonnegative. Variants of NMF algorithms are proposed to adapt to different situations. Sparse solutions of NMF are gained by adding extra sparseness regularization. Discriminative NMF algorithm maximizes the between-class distance and minimizes the within-class distance while learning the low dimensional representation. Research works have proposed NMF variants on manifolds by adding constraints on local structures, since high dimensional data of many applications are on low dimensional manifold.

However, all these methods are restricted to linear dimension reduction, and thus unable to capture complex nonlinear properties. Moreover,

many real world data require nonlinearity in dimension reduction due to their distribution, for example, handwritten digits form own manifolds in the feature space. Unfortunately, typical nonlinear dimension reduction techniques are prohibitively expensive when facing large data. For example, Isomap has to compute the pairwise distances between sample to find geodesic distance, kernel trick has to apply the kernel function to all pairs of samples, etc.

To allow nonlinearity in the dimension reduction and to keep the computational efficiency, a piecewise linear dimension reduction technique is proposed in this paper. Specifically, piecewise linear nonnegative matrix factorization (PLNMF) is proposed. The assumption is that though dimension reduction is nonlinear globally, it is linear locally, i.e. linear dimension reduction method is applicable when data are restricted to a region. Specifically, the entire space is divided into k regions, within each of which a local model is learned. A global nonlinear dimension reduction is composed of a collection of linear local dimension reduction models. The proposed piecewise linear algorithm has been evaluated on eight real world datasets, and it has been shown to generate more accurate results than the linear model, NMF, in the scenario of clustering.

9408-4, Session 2

Boundary fitting based segmentation of fluorescence microscopy images

Soonam Lee, Purdue Univ. (United States); Paul Salama, Kenneth W Dunn, Indiana Univ. (United States); Edward J Delp, Purdue Univ. (United States)

Advances in fluorescence microscopy have enabled biologists to image deeper into tissue than previously achievable. Of increasing interest is the quantitative analysis of acquired 3D image stacks. However, the volume of data makes manual quantification, especially of each cell or tissue, tedious and error-prone. Thus, automated processing methods, especially image segmentation, become vital.

Segmentation of fluorescence microscopy images remains challenging due to the significant light scatter in biological tissues, which has the effect of reducing both contrast and resolution. For this reason, segmentation approaches that are successful at distinguishing objects at the surface of tissue samples increasingly fail at depth. In addition, boundaries of biological objects, which are non rigid and tend to vary in shape and orientation, are not always clearly and completely delineated by fluorescent probes.

This paper proposes a 2D segmentation technique that combines thresholding and boundary fitting to delineate tubular objects in microscopy volumes. More specifically, due to inhomogeneous background, we use combination of adaptive and global thresholding as a preprocessing. After that, perform 2D connected component labeling followed by branch pruning to clean unnecessary end points. Lastly, reconnect entire boundary using end point matching and curve fitting. To fit curve, we find “shortest paths” based on geodesic distance between two matched end points and find an ellipse that best (in the least square sense) fits the shortest path.

To verify our proposed scheme, we apply the proposed technique to two data sets. The first set of images (WSM) was comprised of 512 images each 512 ? 512 pixels in size, while the second set (Lectin) consisted of 821 images each 640 ? 640 pixels. Both data sets were imaged using multiphoton fluorescence excitation microscopy especially from a 3D volume of rat kidney. In particular, the WSM data set is labeled with Hoechest 33342 (blue) and fluorescent phalloidin (red) dyes and the Lectin data set is labeled with Hoechst 33342 (blue) and a fluorescent lectin (red) dye. In both cases, the structures of interest reside mostly in the red (R) component of the data. Compared to a region-based active contour technique, the proposed scheme was more successful at segmenting tubular structure in microscopy images.

In the full paper, we will explain our proposed technique with more details with various examples. In addition, we will add block diagram of our proposed technique and show the result images from every step. Also, we will provide detail equations of ellipse fitting including objective function and its constraints. Finally, we will compare the segmentation results from our proposed scheme to other methods.

9408-5, Session 2

Robust textural features for real time face recognition

Chen Cui, Vijayan K. Asari, Andrew D. Braun, Univ. of Dayton (United States)

Automatic face recognition in real life environment is challenged by various issues such as the object motion, lighting conditions, poses and expressions. The Enhanced Local Binary Pattern (ELBP) is able to represent textural features of a face image in different lighting conditions. Instead of comparing the intensity of every neighborhood pixel with the center pixel’s intensity value directly as in LBP, ELBP description compares the total positive distance and the total absolute distance between the neighborhood pixels and the center pixel. It then counts the number of ones in the 8-bit code of the thresholded neighborhood values to be the representation of the pixel under consideration. The range of the intensity level of an input image is reduced from 0~255 to 0~8, if a 3 by 3 neighborhood is assumed. To improve the performance by reducing the false positives, we propose a face recognition system that uses a textural description methodology based on a refined ELBP feature set and a Support Vector Machine (SVM) for classification. The counting strategy of ELBP may merge the details of the neighborhood textural information. For instance, the labels of 11000000 and 00000011 within the ELBP 8-bit codes are both 2. To retain the most specific features of input face images, we modified the ELBP algorithm by replacing the counting procedure of ELBP by converting the samplings to a binary image after obtaining the 8-bit code. The proposed system is currently trained with several people’s face images obtained from video sequences captured by an input camera in a real life environment. It is tested with two image sets: one contains the disjoint face images of the trained people’s faces to determine the recognition rate; the other one includes face images of several non-trained people’s faces to investigate the percentage of the false positives caused by the non-trained faces. Single SVM creates high dimensional margins to classify the high dimensionality data into two groups. So the proposed system has multiple single SVM structures to achieve the multi-class SVM classification. Currently, the proposed system recognizes the trained faces frame by frame continuously. The recognition rate among 300 images of 10 trained faces is around 85%, and the percentage of the false positives in the second test with 5000 images of 30 non-trained faces is around 8%. In order to improve the system confidence due to the presence of any false recognitions, a “matching score” is assigned next to the recognized faces to help a human observer to do a reconfirmation. The proposed system is programmed and validated in MATLAB, and it is being migrated to C++ for improving its efficiency. Research work is progressing to tackle the partially occluded faces as well. It is envisaged that an appropriate weighting strategy applied to different parts of the face area would lead a good performance in this situation.

9408-6, Session 2

Autonomous color theme extraction from images using saliency

Ali Jahanian, S. V. N. Vishwanathan, Jan P. Allebach, Purdue Univ. (United States)

A Color theme (palette) is a collection of salient color swatches which can represent or describe the choices of colors in an artwork or a design such as a graphic, fashion item, or interior. For instance, when designers are asked to design a piece such a magazine cover, poster, or webpage, they often start by choosing a good image and then extracting the color palette from the image. The color palette is then used consistently through the design process. Designers usually choose a 3-color, 5-color, or occasionally a 7-color palette so that their designs are clean and sophisticated as opposed to busy and cluttered. In today’s “Instagram” world, with the enormous

number of digital images that can be downloaded from the web, and design examples, an automatic mechanism for extracting color palettes form images may facilitate the inspiration and creativity of designers. Applications of such an automatic mechanism, however, may also extend to other avenues of research in color quantization, transfer of the color mood of images, image retrieval based on colors, and image similarity metrics based on colors.

Traditionally, color palette extraction is performed by either utilizing a suite of well-known clustering techniques such as k-means and fuzzy c-means, or by quantizing color histograms of images. These tasks are mainly defined as an optimization problem which solves for minimizing an error. Recently, Lin and Hanrahan [1] introduced a different solution; a regression model trained on 1600 collected color palettes for 40 images from 160 human participants. Their model includes six types of features among which saliency is reported to be the main feature. Although their underlying images include several different painting styles and images, this approach, in general, may suffer from lack of spanning all the colors in a color space, as well as averaging and overfitting dilemmas. Hence, in this paper, we suggest a more autonomous mechanism which is also based on the saliency map of a given image.

To extract K color swatches from a given image, our algorithm takes the image as an input, and computes its saliency map histogram using a superposition of two saliency algorithms from [2] and [3]. The next step is to find the K choices of saliency pixels from this histogram, as follows: First, we find the maximum amplitude (most frequently occurring histogram value) and minimum amplitude (least frequently occurring histogram value) points as the first and second choices. Then, for the k-th choice, we find a saliency pixel with the furthest distance (Euclidean distance) to the k-1 choice. In this way, we obtain a set of K saliency pixels. Our approach captures the dominant background color, the color corresponding to a highly localized, yet important feature, and a well-spaced distribution of colors between these limit points. The corresponding sRGB pixels of this set represent the choice of color swatches for the input image. This method is not trained on a set of images, and does not depend on any results from psychophysical experiments.

We compare our results with the ones from Lin and Hanrahan [1], as well as the two current frequently used online tools, Adobe Kuler [4], and ColourLovers [5]. Our evaluations based on the theme extraction evaluation method presented in [1], suggest that our algorithm produces more acceptable results than the others. Nevertheless, conducting a psychophysical experiment to verify the results is our next step.

References

[1] S. Lin, and P. Hanrahan, “Modeling how people extract color themes from images,” In ACM Human Factors in Computing Systems (CHI), 2013.

[2] J. Harel, C. Koch, and P. Perona, “Graph-based visual saliency,” Advances in neural information processing systems, vol. 19, pp. 545, 2007.

[3] R. Achanta, S. Hemami, F. Estrada, and S. Susstrunk, “Frequency-tuned salient region detection,” In Computer Vision and Pattern Recognition (CVPR), IEEE Conference on, 2009.

[4] Adobe Kuler. http://kuler.adobe.com/.

[5] ColourLovers. http://www.colourlovers.com/.

9408-7, Session 3

m-BIRCH: an online clustering approach for multimedia and computer vision applications

Siddharth K. Madan, Kristin J. Dana, Rutgers, The State Univ. of New Jersey (United States)

In modern multimedia and computer vision applications datasets are large and updates with new data are ongoing. Methods of online clustering are extremely important for clustering large and time-varying datasets.

Traditional clustering algorithms process all points at once which is very inefficient, require memory equal to the entire dataset size which is very large, and cannot incrementally update the clustering decisions when new data comes in. Online clustering algorithms incrementally and efficiently cluster the data points, use a fraction of the dataset memory, and update the clustering decisions when new data comes in. In this paper we adapt a classic online clustering algorithm called Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) to incrementally cluster large datasets of features. We call the adapted version modified-BIRCH (m-BIRCH).

BIRCH was originally developed by the database management community, but has not been used in multi-media and computer vision. BIRCH has a number of strengths which makes it a useful algorithm for multimedia and computer vision applications. The incremental update step of BIRCH can be used with different batch clustering algorithms, resulting in incremental versions of multiple batch clustering algorithms. We use BIRCH with spectral, Gaussian mixture model (GMM), and K-means based clustering. Adapting GMM and spectral clustering to be used with BIRCH is a novel contribution of the paper. BIRCH makes final clustering decisions on a concise summary of the whole dataset and not on subsets of points; therefore the final clusters reflect the point distribution in the entire dataset. BIRCH can generate correct clusters in presence of outliers. Prior work has not demonstrated BIRCHs utility in non-convex clustering. In this paper, we show that BIRCH can be used to discover non-convex clusters. Modifications made in m-BIRCH enable data driven parameter selection and effectively handle varying density regions in the feature space. BIRCH obtains a concise representation of points close to each other in the feature space, and the threshold parameter controls the maximum average inter-point distance in the concise representation. The ideal method for data driven selection of the threshold parameter is computationally infeasible. The m-BIRCH algorithm systematically approximates the computationally infeasible ideal approach. The m-BIRCH algorithm separately processes regions of different density in the feature space.

Separate processing ensures that regions of different densities contribute in generating the final cluster centers.

We demonstrate the ability of m-BIRCH to, a) incrementally cluster large datasets using a dataset of 840K SIFT descriptors, b) handle outliers using a dataset of 60K outlier corrupted grayscale patches, c) detect non- convex clusters using datasets with challenging clustering patterns. The m-BIRCH algorithm required just 10 to 20% of the dataset memory to perform the clustering. We evaluate the vocabularies by classifying separate test sets. The experimental results clearly show the ability of m-BIRCH to incrementally and effectively cluster large datasets of features using just fraction of the dataset memory. We have made our implementation of the algorithm publicly available, which provides a useful clustering tool.

9408-8, Session 3

Enhanced features for supervised lecture video segmentation and indexing

Di Ma, Gady Agam, Illinois Institute of Technology (United States)

Lecture videos are common and increase rapidly. Consequently, automatically and efficiently indexing such videos is an important task. Video segmentation is a crucial step of video indexing that directly affects the indexing quality. We are developing a system for automated video indexing and in this paper discuss our approach for video segmentation and classification of video segments. The novel contributions in this paper are two fold. First we develop a dynamic Gabor filter and use it to extract features for video frame classification. Second, we propose a recursive video segmentation algorithm that is capable of clustering video frames into video segments. We then use these to classify and index the video segments. By indexing video content, we can support both topic indexing and semantic querying of multimedia documents.

Frames of lecture videos can generally be divided into two categories. The first category consists of text frames which include slides, handwritten board notes, and computer screen projections. The second category consists of frames with views of the presenter and/or other frames that do not contain presentation specific material. We are interested in the first group as it contains specific text information that can be used for indexing the video. The main problem with existing methods for lecture video

segmentation is their accuracy in locating the starting frame of text shots in different kinds of videos and their accuracy in classifying text shots. We address these problems in the proposed approach using enhanced features and a supervised learning algorithm.

The enhanced features we propose are based on dynamic Gabor filters. Since different lecture videos have different kinds of text size, font, orientation and layout, a single Gabor filter with fixed parameters cannot effectively detect text areas in all videos. By changing the filter parameters, we iteratively apply different Gabor filters on video frames until a candidate text area is detected. Post processing operations on the filter outputs are then applied to detect candidate text locations which are then used to compute features for text classification. The proposed features are based on intensity distributions as well as projection histograms, connected component distributions, and geometric entities such as line segments. Overall, we extract 143 features in 5 categories. The cross-validated accuracy of the frame classification based on over 40000 frames in this stage is over 95%. Subsequently, a recursive clustering algorithm is proposed to group frames into relevant shots and an integrated shot is used to produce an index frame. The OCR content of the integrated shot of each video segment is the used to index it. The proposed approach is evaluated on a test collection of 20 different types of lecture videos and compared to a commercial video lecture indexing system.

The proposed approach results in a similar True Positive Rate(TPR) 92.5% and lower False Discovery Rate(FDR) 1.4% compared with the commercial system(TPR= 94.8%, FDR=40.0%) demonstrate that the performance is significantly improved by using enhanced features.

9408-9, Session 3

Characterizing the uncertainty of classification methods and its impact on the performance of crowdsourcing

Javier Ribera, Khalid Tahboub, Edward J. Delp, Purdue Univ. (United States)

Video surveillance systems are widely deployed with the number of surveillance cameras increasing exponentially. The enormous amount of video data coming out of such systems makes it practically impossible to continuously monitor all the feeds. Automatic video analytics provide a scalable solution to extract useful information or identity threats.

One type of analysis is “crowd flow estimation”. Crowd flow refers to the number of people crossing a specific region in a given period of time. Automatic crowd flow estimation suffers from drawbacks due to varying levels of crowdedness or due to degraded video quality. Crowdsourcing is shown to help automatic video analysis perform better. Our previous work used crowdsourcing to enhance the performance of the automatic crowd flow estimation when the automatic method is uncertain about making a particular decision.

In this paper, we study several approaches to characterize the uncertainty of the classifier. We conduct an experimental evaluation using publicly available datasets and by introducing video quality degradations and packet losses.

9408-10, Session 3

Object tracking on mobile devices using binary descriptors *(Invited Paper)*

Andreas E. Savakis, Mohammad Faiz Quraishi, Breton Minnehan, Rochester Institute of Technology (United States)

As mobile devices increase in ubiquity, so has the demand that they perform more complex tasks. Computer vision tasks, such as tracking, present an interesting opportunity for mobile devices because of potential for many interesting applications including augmented reality. In this paper, we present mobile implementations of a robust and efficient object tracker.

Our tracker utilizes a dictionary of templates consisting of static and dynamic templates. Dynamic templates are continuously updated and capture object variations due to pose, illumination, and partial occlusion. Static templates represent high confidence instances of the object, for example when an object detector is used, and are used to prevent drift or recover from occlusions.

Dictionary elements are image patches that are represented by binary descriptors, such as Binary Robust Independent Features (BRIEF), and Binary Robust Invariant Scalable Keypoints (BRISK). For each new frame, a search grid of candidate locations is generated centered at the previous object location. The candidate location that yields the best match with a dictionary element is selected as the location of the tracked object. Binary descriptors offer fast computation, low memory requirements and easy matching using Hamming distance.

We present two mobile implementations using Apple’s iOS and Google’s Android operating system. We utilize Android’s Native Development Kit (NDK), which gives the performance benefits of using native code as well as access to legacy libraries, while Apple iOS offers easy integration and strong real time performance.

9408-11, Session 4

Comparing humans to automation in rating photographic aesthetics *(Invited Paper)*

Ramakrishna Kakarala, Abhishek Agrawal, Sandino Morales, Nanyang Technological Univ (Singapore)

No Abstract Available

9408-12, Session 4

Service-oriented workflow to efficiently and automatically fulfill products in a highly individualized web and mobile environment

Mu Qiao, Shutterfly Inc. (United States)

Service Oriented Architecture (SOA) is widely used in building flexible and scalable web sites and services. In most of the web or mobile photo book and gift business space, the products ordered are highly variable without a standard template that one can substitute texts or images from similar to that of commercial variable data printing. In this paper, the author describes a SOA workflow in a multi-sites, multi-product lines fulfillment system where three major challenges are addressed: utilization of manufacturing equipment, highly automation with fault recovery, and highly scalable and flexible with order volume fluctuation.

9408-13, Session 4

An interactive web-based system using cloud for large-scale visual analytics

Ahmed S. Kaseb, Everett Berry, Erik Rozolis, Kyle McNulty, Seth Bontrager, Youngsol Koh, Yung-Hsiang Lu, Edward J. Delp III, Purdue Univ. (United States)

What are the traffic conditions of the streets in New York? Is it raining at the Eiffel Tower now? Has the snow in a national park melted? Is a shopping mall crowded? These questions are frequently asked by city planners, meteorologists, and the general public. To answer such questions, we have constructed a system that can retrieve and analyze the live visual data from thousands of worldwide distributed cameras.

This paper presents CAM2 (Continuous Analysis of Many CAMeras) as a system that addresses the following problems: (i) It takes significant effort to analyze images from thousands of cameras simultaneously. CAM2 reduces that effort by providing an Application Programming Interface (API) that requires only slight changes to existing analysis methods. (ii) Analyzing the data from thousands of cameras simultaneously requires significant amounts of resources. CAM2 allocates cloud resources to meet the computation and storage requirements. (iii) Cameras are heterogeneous, i.e. they have different brands, resolutions, and methods for retrieving data. CAM2 hides this heterogeneity so that the same methods can analyze the data from different cameras.

This paper focuses on how to use the CAM2 's website and API. Users can submit, execute, and download the results of their analysis methods using CAM2 's website (cam2.ecn. purdue.edu) by following this procedure: (i) Users can view an interactive world map with nearly 40,000 geo-tagged cameras along with their recent snapshots. (ii) Users can select cameras for analysis using the cameras' locations (e.g. country, state, city, and timezone) (iii) Users can specify the desired analysis parameters, including the frame rates and durations. (iv) Users can upload their analysis methods that use CAM2 's API. The API is event-driven: when new frames arrive, the analysis methods are invoked. This event-driven API significantly simplifies the analysis methods because they do not have to communicate with heterogeneous cameras directly. (v) Users can download their analysis results.

Our experiments demonstrate that CAM2 can be used for a variety of image analysis methods (e.g. motion analysis and human detection), and is capable of analyzing 2.7 million images (141 GB images at 107 Mbps) from 1274 cameras over 3 hours using 15 Amazon EC2 cloud instances. The average resolution of the cameras is 0.44 Mega Pixels (MP), and they are deployed in North America, West Europe, and East Asia.

This paper has the following contributions: (i) To our knowledge, this is the first web-based system that enables users to execute methods analyzing the data from thousands of cameras simultaneously. (ii) The system provides an API that makes it easy to migrate existing analysis methods with only slight changes. The system can be used for a wide range of applications. (iii) The system provides access to nearly 40,000 geographically distributed cameras (more are being added) that we discovered worldwide. (iv) This system allows users to specify parameters for selecting cameras, the frame rates, and the durations for executing the analysis methods. (v) CAM2 allocates cloud resources to meet the computation and storage requirements of different analysis methods.

9408-14, Session 4

Proposed color workflow solution from mobile and website to printing

Mu Qiao, Terry Wyse, Shutterfly Inc. (United States)

With the recent introduction of mobile devices and development in client side application technologies, there is an explosion of the parameter matrix for color management: hardware platform (computer vs. mobile), operating system (Windows, Mac OS, Android, iOS), client application (Flesh, IE, Firefox, Safari, Chrome), and file format (JPEG, TIFF, PDF of various versions). In a modern digital print shop, multiple print solutions are used: digital presses, wide format inkjet, dye sublimation inkjet are used to produce a wide variety of customizable products from photo book, personalized greeting card, canvas, mobile phone case and more. In this paper, we outline a strategy spans from client side application, print file construction, to color setup on printer to manage consistency and also achieve what-you-see-is-what-you-get for customers who are using a wide variety of technologies in viewing and ordering product.

9408-15, Session 4

On-line content creation for photo products: understanding what the user wants

Reiner Fageth, CeWe Color AG & Co. OHG (Germany)

No Abstract Available

9408-16, Session 5

Driving into the future: how imaging technology is shaping the cars of future *(Invited Paper)*

Buyue Zhang, NVIDIA Corp. (United States)

A few years back, who could have imagined automotive being the new driving force for imaging and computer vision technology? But that is exactly what is happening today. Fueled by the development of Advanced Driver Assistance System (ADAS) and Autonomous Vehicles, as well as the proliferation of cameras and sensors, automotive has become a rich domain for imaging innovations. In this talk, I will talk about the imaging and vision problems we are facing in today's automotive system, the unique challenges, and what it takes to make the car "see". I will give some example solutions and point out the opportunities in solving imaging problems in the automotive space.

9408-17, Session 5

Worldview and route planning using live public cameras

Ahmed S. Kaseb, Wenyi Chen, Ganesh Gingade, Yung-Hsiang Lu, Purdue Univ. (United States)

How does the Times Square look like right now? Are the streets still snowy on the route to Chicago? Is the shopping mall crowded now? These questions and many more about tourist attractions, traffic, and weather, are being frequently asked. To find the answers, some people search on-line but can find only snapshots taken long time ago. To check the current traffic or weather information, people usually use text-based methods which provide no visual information. Visual information from live feeds can provide timely updates. Moreover, the visual information helps seeing the world, planning routes, and many more applications. In recent years, thousands of public cameras are being installed. However, few mobile applications can provide access to these live views.

This paper presents an Android mobile application with the following features: (i) Users can watch the live feeds from thousands of public cameras. We have constructed a system that includes an interactive map with nearly 40,000 geotagged public cameras around the world. Users can select cameras to watch the live views. (ii) Users can plan their routes based on the live video feeds from public cameras. After a user specifies a starting point and a destination, this application will show the live views from the cameras along a suggested route.

This system consists of two main components: (i) A server maintains a database with the cameras information, including their locations and the methods to retrieve data from these cameras. (ii) The Android application installed on mobile systems retrieves cameras' information from the server and renders a map with superimposed camera markers. When a user clicks a marker, the mobile application directly communicates with the camera and show the live views. The visual data are transmitted directly from the cameras to the mobile systems, without going through the server. This reduces latency and enhances scalability. The application uses Google Directions API to find the route between a starting point and a destination.

To evaluate the application, we compare it with similar applications in

terms of the total number of cameras, the cameras world coverage, and the number of cameras on various routes. Our experiments show that this application provides the largest number of cameras with a better coverage than any other applications. The experiments also show that our application yields the most number of cameras for various selected routes. We also consider the following performance metrics: (i) The response time loading the world map with superimposed cameras markers depends on the number of cameras. Our experiments show that this response time is 2 seconds for 5,000 cameras, and 11 seconds for 40,000 cameras. (ii) The response time of finding the best route from a starting point to a destination along with the cameras on the route depends on both the length and the topology of the route. Our experiments show that this response time does not exceed 1.6 seconds for 1,000-mile routes. (iii) The response time of retrieving a single frame from various cameras around the world is 32 millisecond.

This paper has the following contributions: (i) It allows mobile users to watch the live feeds from 40,000 public cameras. Users can watch tourist attractions, streets, malls, etc. (ii) It handles the heterogeneity of the cameras, i.e. they are of different brands (Axis, Panasonic, etc.), resolutions, frame rates, etc. Heterogeneity is a key challenge since the mobile application has to handle the different methods to communicate with cameras of various brands. (iii) It allows travellers to plan their routes by showing the live feeds from the cameras along the route between a starting point and a destination. (iv) It enhances scalability by eliminating the need for an intermediate server between the mobile devices and cameras.

9408-18, Session 5

Musical examination to bridge audio data and sheet music

Xunyu Pan, Timothy J. Cross, Liangliang Xiao, Xiali Hei, Frostburg State Univ. (United States)

The digitalization of audio is commonly implemented for the purpose of convenient storage and transmission of music and songs in today's digital age. Analyzing digital audio for insightful look at a specific musical characteristic, however, can be quite challenging for various types of applications. Many existing musical analysis techniques can examine a particular piece of audio data. For example, the frequency of digital sound can be easily read and identified at a specific section in an audio file. Based on this information, we could determine the musical note being played at that instant, but what if you want to see a list of all the notes played in a song? While most existing methods help to provide information about a single piece of the audio data at a time, few of them can analyze the available audio file on a larger scale. These same concepts apply to the creation or regeneration of new digital audio as well. The research conducted in this work considers how to further utilize the examination and generation of audio file by extracting and storing more information from the original audio file. In practice, we develop a novel musical analysis system Musicians Aid to process musical examination and generation of audio data. Musicians Aid solves the previous problem by storing and analyzing the information as it reads it rather than tossing it aside. By gathering and analyzing extra information from an audio source, the system can provide professional musicians with an insightful look at the music they created and advance their understanding of their work. Amateur musicians could also benefit from using it solely for the purpose of getting feedback about a song they were attempting to play. By comparing the information provided by our system with traditional sheet music, they could ensure what they played was correct. Finally, the similar music or songs can be regenerated using sound effects library available on most computing devices by using the musical characteristics collected from the original audio data, which has the potential for music information retrieval on local computing devices or across computing networks. In addition, the application could be extended over the Internet to allow users to play music with one another and then review the audio data they produced. This would be particularly useful for teaching music lessons on the web. The developed system is evaluated with songs played with guitar, piano, drum, and other popular musical instruments. The Musicians Aid system is successful at both analyzing and generating audio file and it is powerful in assisting individuals interested in learning and understanding music either locally or over the Internet.

9408-19, Session 5

Innovating instant image recognition on mobile devices: delectable the social wine app

Wiley H Wang, Cassio Paes-Leme, Derick Kang, Kevin Farrell, Jevon Wild, Delectable (United States)

No Abstract Available

9408-20, Session 6

Document image detection for mobile capturing

Zhigang Fan, SKR Labs (United States)

No Abstract Available

9408-22, Session 6

A scheme for automatic text rectification in real scene images

Baokang Wang, Changsong Liu, Xiaoqing Ding, Tsinghua Univ. (China)

Traditionally, text information is mainly acquired by scanning documents and manuscripts using flat-bed scanner. However, digital cameras, with its non-contact imaging style, excellent portability and high imaging quality, are more and more widely used to capture text in real scenes and becoming the most important access to text information digitization, thus creating urgent demand for camera-based text understanding. Unfortunately, arbitrary position and angle related to text area in real scene can frequently cause perspective distortion which most OCR systems at present could not manage. Dealing with perspective distortion is a major challenge in scene text recognition and understanding, while most previous studies focused on rectification of document images, seldom research put forward detailed analysis and scheme to settle distortion problem of text in real scenes. In this paper, a hierarchical scheme for automatic text rectification in natural scene images is proposed. We rely on geometric information extracted from characters themselves as well as their surrounding information, forming a overall-to-internal framework. The final goal is to estimate the horizontal and vertical vanishing point(VP)s and then a 3-step recovery procedure is performed rectifying horizontal foreshortening, vertical foreshortening and shearing respectively. We notice the fact that linear segments around text area (signboard borders or edges of buildings typically) as well as edges of character stroke themselves usually contain direction parallel or perpendicular to text alignment, which could provide rich information for VP-estimation. For the first step, linear segments are extracted from interested region of input image, then a J-Linkage based clustering is performed to separate line segments into several groups corresponding to a specific VP. To eliminate the affect of noise data, we perform a customized EM iteration to refine clustering results. Finally two dominant VPs are selected as estimated horizontal and vertical VPs (HVP and VVP) which then used for primary rectification. If the confidence(mainly examined by OCR results) of primary rectification result is not satisfying, secondary rectification procedure would start which mainly utilizes internal structure and strokes of characters. Considering the difference existed between Chinese and English character, we trained a simple language classifier to tell Chinese texts from English. HVP and VVP are estimated in two stages respectively. In the first stage, the whole text area is rotated according to the principal component direction of text pixels thus reducing the deformation level. Then RANSAC line fitting and some geometric operations are performed followed by 2-D projection profile scanning to generate and refine top and bottom line position. Notice that details of operations may be different according to language type. HVP is then acquired followed

by horizontal partial rectification. In the second stage, we use the partial rectified image as input and try to figure out VVP position. Shear angles of “valuable” connected component and slopes of “valuable” strokes are selected and used to implement a linear regression, which would show out the shearing trend along x axis and finally obtain VVP position. Experiments on synthesis text samples as well as real scene images in MSRA-TD500 dataset demonstrate the increase of recognition rate and improvement compared with some related algorithms.

Conference 9409: Media Watermarking, Security, and Forensics 2015

Monday - Wednesday 9–11 February 2015

Part of Proceedings of SPIE Vol. 9409 Media Watermarking, Security, and Forensics 2015

9409-1, Session 1

Exposing photo manipulation from user-guided 3D lighting analysis

Tiago J. de Carvalho, Cemaden (Brazil); Hany Farid, Dartmouth College (United States); Eric R. Kee, Columbia Univ. (United States)

1. INTRODUCTION

Within many different and complementary approaches to analyzing a photo for evidence of manipulation, physically-based methods are particularly attractive because they are applicable in low quality and low reso- lution images, and can be hard to counter attack since the physical measurements being made are the result of interactions in the 3-D physical world being projected into the 2-D image. Earlier lighting-based forensic techniques focused on estimating 2-D properties of lighting. This is because estimating the full 3-D lighting requires knowledge of the 3-D structure of the scene which is, of course, generally not readily available.

We describe a new 3-D lighting-based technique that overcomes this limitation. This technique leverages the fact that with minimal training, an analyst can often provide fairly reliable estimates of local 3-D scene structure, from which 3-D lighting can be estimated. We describe an easy way to use user-interface for obtaining 3-D shape estimates, how 3-D lighting can be estimated from these estimates, a perturbation analysis that contends with errors or biases in the user-specified 3-D shape, and a probabilistic technique for combining multiple lighting estimates to determine if they are physically consistent with a single light source.

2. METHODS

The projection of a 3-D scene onto a 2-D image sensor results in a loss of information. Recovering 3-D shape from a single 2-D image is at best a difficult problem, and at worst it is under-constrained. There is, however, evidence from the human perception literature that human observers are fairly good at estimating local 3-D shape from a variety of cues including, foreshortening, shading, and familiarity.1, 2 To this end, we ask an analyst to specify the local 3-D shape of surfaces.

An analyst estimates the local 3-D shape at different locations on an object by adjusting the orientation of a small 3-D probe. The probe consists of a circular base and a small vector (the stem) orthogonal to the base. An analyst orients a virtual 3-D probe so that when the probe is projected into the image, the stem appears to be orthogonal to the object surface, Figure 1. With the click of a mouse, an analyst can place a probe at any point p in the image. This initial mouse click specifies the location of the probe's base. As the analyst drags their mouse, they control the orientation of the probe by way of the 2-D vector v from the probe's base to the mouse location. This vector is restricted by the interface to have a maximum value of ? pixels.

Probes are displayed to the analyst by constructing them in 3-D, and projecting them into the image. The 3-D probe is constructed in a coordinate system that is local to the object, Figure 2, defined by three mutually orthogonal vectors:

??p?c?? ??v??

b1= f b2= 1v·(p?c) b3=b1?b2 , (1)

f

where p is the location of the probe's base in the image, and f and c are a focal length and principal point (described below). The 3-D probe is constructed by first initializing it into a default orientation in which its stem, a unit vector, is coincident with b1, and the circular base lies in the plane spanned by b2 and b3, Figure 2. The 3-D probe is then adjusted to correspond with the analyst's desired orientation which is uniquely defined by their 2-D mouse position v. The 3-D probe is parameterized by a slant and tilt, Figure 2. The length of the vector v specifies a slant rotation, ? = sin?1(?v?/?), of the probe around b3. The tilt, ? = tan?1(vy/vx), is embodied in the definition of the coordinate system, Equation (1).

The construction of the 3-D probe requires the specification of a focal length f and principal point c, Equation (1). There are, however, two imaging systems to be considered. The first is that of the observer relative to the display.3 This imaging system dictates the appearance of the probe in the image plane. In that case, we assume an orthographic projection with c = 0, as in.1,2 The second imaging system is that of the camera which recorded the image. This imaging system dictates how the surface normal n is constructed to estimate the lighting. In this case, if the focal length f and principal point c are unknown then f can be assigned a typical mid-range value and c = 0.

With user-assisted 3-D surface normals in hand, we proceed with estimating 3-D lighting. We begin with the standard assumptions that a scene is illuminated by a single distant point light source (e.g., the sun) and that an illuminated surface is Lambertian and of constant reflectance. Under these assumptions, lighting is estimated using a standard least-squares estimation.4

In practice there will be errors in the estimated light direction due to errors in the user-specified 3-D surface normals, deviations of the imaging model from our assumptions, signal-to-noise ratio in the image, etc. To con- tend with such errors, we perform a perturbation analysis yielding a probabilistic measure of the light direction. For simplicity, we assume that the dominant source of error is the analyst's estimate of the 3-D normals. A model for these errors is generated from a large-scale psychophysical study in which observers were presented with one of twelve different 3-D models, and asked to orient probes, like those used here, to specify the object's shape.2 Theobjectswereshadedwithasimpleoutdoorlightingenvironment. UsingAmazon'sMechanicalTurk a total of 45, 241 probe settings from 560 observers were collected.

The probe setting data is used to construct a probability distribution over the ground truth slant and tilt conditioned on the analyst's slant and tilt. This slant/tilt distribution is then used to build a distribution over the 3-D light direction. Specifically, for each of the analyst's probes, a ground truth slant/ tilt is randomly drawn. The light direction is then estimated, and contributes a small Gaussian density (over the 3-D light direction) that is projected into azimuth/elevation space. This density estimation procedure is repeated for 20, 000 random perturbations of the analyst's probes. The result is a kernel-density estimate of the distribution over light directions, given the analyst's estimate of the object geometry.

Multiple lighting distributions are constructed by an analyst, each distribution from a particular object in the image. A confidence region is computed in each distribution, identifying a potentially non-contiguous area in which the light source must lie. The physical consistency of an image is determined by intersecting these confidence regions. Forgery is detected if the intersection is empty, up to a specified confidence threshold.

3. PRELIMINARY RESULTS

We rendered ten objects under six different lighting conditions. Sixteen untrained users were each instructed to place probes on ten objects. Shown in the left column of Figure 3 are four representative objects with the user- selected probes. Shown in the right column are the estimated light positions specified as confidence intervals. The small black dot in each figure corresponds to the actual light position. On average, users were able to estimate the azimuth and elevation with an average accuracy of 11.1 and 20.6 degrees with a standard deviation of 9.4 and 13.3 degrees, respectively. On average, a user placed 12 probes on an object in 2.4 minutes.

Because this technique is based on sound physical models, we expect (and will show) that this analysis is equally effective when applied to real-world images and forgeries.

REFERENCES

[1] Koenderink, J. J., Van Doorn, A., and Kappers, A., "Surface perception in pictures," Attention, Perception, & Psychophysics 52, 487–496 (1992). 10.3758/BF03206710.

[2] Cole, F., Sanik, K., DeCarlo, D., Finkelstein, A., Funkhouser, T., Rusinkiewicz, S., and Singh, M., "How well do line drawings depict shape?," in [ACM Transactions on Graphics (Proc. SIGGRAPH)], 28 (Aug. 2009).
[3] Cooper, E. A., Piazza, E. A., and Banks, M. S., "The perceptual basis of common photographic practice," Journal of Vision 12(5) (2012).

[4] Johnson, M. K. and Farid, H., "Exposing digital forgeries by detecting inconsistencies in lighting," in [Proceedings of the 7th workshop on Multimedia and security], MM&Sec '05, 1–10, ACM (2005).

9409-2, Session 1

Thinking beyond the block: block matching for copy-move forgery detection revisited

Matthias Kirchner, Pascal Schoettle, Westfälische Wilhelms-Univ. Münster (Germany); Christian Riess, Stanford School of Medicine (United States)

We describe an efficient approach for finding duplicate patterns of a given size in integer-valued input data. By design, we focus on the spatial relation of potentially duplicated elements. This allows us to locate copy-move forgeries via bit-wise operations, without expensive block comparisons in the feature space. Experimental results suggest performance boosts by an order of magnitude and promise high accuracy.

9409-3, Session 1

The Krusty the Clown attack on model-based speaker recognition systems

Scott A. Craver, Alireza Farrokh Baroughi, Binghamton Univ. (United States)

No Abstract Available

9409-4, Session 2

Automation and workflow considerations for embedding Digimarc barcodes at scale

Tony F. Rodriguez, Don L. Haaga Jr., Sean Calhoon, Digimarc Corp. (United States)

The Digimarc Barcode is a digital watermark applied to packages and variable data labels that carries GS1 standard GTIN-14 data traditionally carried by a 1-D barcode. The Digimarc Barcode can be read with smartphones and imaging based barcode readers commonly used in grocery and retail environments. Using smartphones, consumers can engage with products (ingredients, dosage information for pharmaceuticals, etc.) and retailers can materially increase the speed of check-out, increasing store margins and providing a better experience for shoppers. Internal testing has shown an average of 53% increase in scanning throughput, enabling 100's of millions of dollars in cost savings for retailers when deployed at scale. To get to scale, the process of embedding a digital watermark must be automated and tightly integrated within existing workflows. Creating the automation tools and processes that enable supermarkets, with a combined total of over a $1 trillion in sales, to encode their packages as part of their production process, is a new challenge for the watermarking community and one that requires increased attention to measures of watermark robustness, reliability and constrained color spaces in print.

Image and audio watermarks have been deployed at significant scale with exceptional robustness within tightly controlled vertical markets, such as in television for audience measurement, or in travel documents to deter counterfeiting. Unlike these examples, the retail channel is one defined by dynamism, where packaged goods are sourced from a wide variety of suppliers (refer to Figure 1). Each of these suppliers in turn have their own prepress and printing vendors with whom they work with to deliver Consumer Packaged Goods. The typical supermarket in the United States has over 50,000 unique National Brand items for sale at any given time. Adding to the complexity is the recent growth in Private Brands that are unique to the retailer, adding to the diversity of packaging in the larger retail ecosystem.

To efficiently engage the retail supply chain and automate the embedding of Digimarc Barcodes, techniques of mass customization are utilized that focus on embedding the watermark during the proof printing step of the production process. Doing so requires a workflow that can be integrated into existing proofing steps while maintaining throughput and eliminating the need for additional human inspection (refer to Figure 2).

Package Artwork is typically represented by a collection of files in a variety of different formats. Bitmaps (*.tiff, *psd, etc.), vector imagery (*.ps, *.ai, etc.) and fonts (*.abf, *.ttf, etc.) are the most common file types that comprise a package. A final rendered packaged can be "built" using the aforementioned files using a variety of different strategies, from a 1 layer bitmap, to numerous layers of vector and bitmap imagery utilizing multiple fonts.

The files are typically delivered in a directory structure that is compressed. Parsing the manifest of files and ensuring that they produce a correctly rendered packaged when assembled is a critical first step of the embedding process. This step is typically referred to as "pre-flight". By way of comparison, one can think of this step as "compiling" the package similar to compiling source code, enabling the equivalent of Static Program Analysis.

Once complete, the embedding process can begin, including the validation of the Global Trade Identifier Number (GTIN) that will be embedded by the watermark. This is completed in partnership with the GS1 organization that assigns and manages the GTIN globally.

Embedding a watermark in a digital image represented in memory as RGB triples, is a well understood process. Over the last decade, ensuring that the resultant watermark can survive the print process has been explored at length as well. Package artwork is more complex and introduces a number of challenges for embedding, these include vector representations, different color models (CMYK, LAB, etc.), wide ink gamuts, multiple concurrent screens and finally print induced constraints (total ink, registration, etc.).

Once embedded a proof print or simulation thereof is created to facilitate visual inspection and testing. The creation of an accurate proof requires that relevant factors that impact print quality at the final production printing step are modeled. These include modeling color profiles, dot gain, and how inks behave on differing substrates.

One-dimensional barcodes are graded using a standard system that provides grades "A" thru "F". The grades are created based on numerous measurements regarding adherence to standards. A barcode that does not meet requirements as defined by trading partners in the supply chain can result in large monetary penalties in the form of "chargebacks", some as large at £40,000 per incident.

Finally, the watermark is tested using both manual and automated methods to produce a similar expression of expected robustness. At the writing of this abstract, the grading system is still under development. A large number of packages have been scanned using multiple robots to yield a data set to set baseline metrics and identify sources of variance that will be accounted for in the final grading implementation.

The Digimarc Barcode, built on a platform of image watermarking technologies adopted by magazines, provides consumers an easy and accurate way to engage with Consumer Packaged Goods. Consumers have clearly expressed a desire to do so as witnessed by the demand for applications such as Red-Laser, etc. Beyond enabling consumer engagement however, the Digimarc Barcode benefits the retailer and brand as well. It does so by materially increasing throughput at the front of store, reducing labor costs and shrinkage. This provides an additional benefit to the consumer as their wait times at front store are reduced. To achieve these benefits, a material percentage the packages that make up the typical basket need to be encoded with the Digimarc Barcode. This number can be as few as 2,000-4,000 for a typical supermarket. To deploy image watermarking at scale, new tools and workflows are being created.

The completed paper will describe specific workflow improvements and the

underlying tools that are increasing throughput and quality. Metrics used to quantify these improvements will be described along with the statistical process control steps used to capture and track gains. Particular focus will be on the impact of algorithmic improvements to the embedding and grading steps of the workflow. Additional lessons learned during the startup of a large scale studio focused on the embedding of image watermarks will be provided as well.

9409-5, Session 2

Watermarking spot colors in packaging

Alastair M. Reed, Tomas Filler, Kristyn R. Falkenstern, Yang Bai, Digimarc Corp. (United States)

No Abstract Available

9409-6, Session 2

Scanning-time evaluation of Digimarc barcode

Becky Gerlach, Daniel T. Pinard, Matthew Weaver, Adnan M. Alattar, Digimarc Corp. (United States)

This paper presents a speed comparison between the use of Digimarc Barcodes and the Universal Product Code (UPC) for customer checkout at point of sale (POS). The recently introduced Digimarc Barcode promises to increase the speed of scanning packaged goods at POS. When this increase is exploited by workforce optimization systems, the retail industry could potentially save billions of dollars. The Digimarc Barcode is based on Digimarc's watermarking technology, and it is imperceptible, very robust, and does not require any special ink, material, or printing processes. A checker can quickly scan Consumer Packaged Goods (CPG) embedded with the Digimarc Barcode using an image-based scanner without the need to reorient the packages with respect to the scanner. Faster scanning of packages saves money and enhances customer satisfaction. It reduces the length of the queues at checkout, reduces the cost of cashier labor, and makes self-checkout more convenient. This paper quantifies the increase in POS scanning rates resulting from the use of the Digimarc Barcode versus the traditional UPC. It explains the testing methodology, describes the experimental setup, and analyzes the obtained results. It concludes that the Digimarc Barcode provides at least 50% enhancement in number of items per minute (IPM) scanned over traditional UPC.

9409-7, Session 2

Digimarc Discover on Google Glass

Eliot Rogers, Tony F. Rodriguez, John Lord, Adnan M. Alattar, Digimarc Corp. (United States)

This paper reports on our implementation of Digimarc's Discover on Google Glass. Digimarc's Barcode technology allows a consumer packaged goods to be embedded with an imperceptible, unique, and robust identity that can be recovered using Digimarc's Discover application. Digimarc's Barcode is embedded all over the package and can be read at any orientation or reasonable distance. Therefore, a checker at a point of sale can scan the package faster and easier than using traditional barcodes. Similarly, a consumer can use his smart mobile device to interact with the package as he walks between the isles of a store without touching the package on the shelves. Digmarc's Discover can also decode embedded information from the ambient audio playing from the overhead speakers in a retail store. This information is also embedded in the audio using Digimarc's technology. Discover can augment/fuse this information with the information it decodes from the package on the shelf. Discover has been successfully ported to the iPad, the iPhone, and many Android based devices, but not yet to smart glasses such as Google Glass. Porting discover to smart glasses is a challenging task, but has the advantages of freeing the customer's hands, reducing the time between intention and action, and presenting relevant information directly on the customer's view. This paper identifies the challenges we met in implementing Digiamrc's Discover on Google Glass and reports on the ways we overcame them. It describes the system architecture and the audio, video and graphical interfaces. It also reports on our evaluation of the system performance including the achieved robustness, processing speed, and power consumption. Finally, it analyses the system bottlenecks and recommends ways to enhance the Glass hardware and software

9409-30, Session Key

Piracy conversion: the role of content protection and forensics *(Keynote Presentation)*

Richard Atkinson, Adobe Systems (United States)

In this session, Richard Atkinson (Adobe's chief strategist and leader of their piracy conversion effort) will walk us through their unconventional yet very logical approach to viewing pirate users as customers and how they are responding to win their business. Included will be aspects around how they view the areas of Content Protection and Forensics in terms of user-experience, operational value, and direction.

9409-8, Session 3

Benford's law based detection of latent fingerprint forgeries on the example of artificial sweat printed fingerprints captured by confocal laser scanning microscopes

Mario Hildebrandt, Otto-von-Guericke-Univ. Magdeburg (Germany); Jana Dittmann, Otto-von-Guericke-Univ. Magdeburg (Germany) and The Univ. of Buckingham (United Kingdom)

The general possibility of fingerprint forgery at crime scenes by transferring latent fingerprints to other objects is known for a long time (see e.g. in Harper, Fingerprint forgery - transferred latent fingerprints, 1937). The same paper also states that the presence of two or more latent fingerprints which are identical in all respects should raise suspicion during the analysis process. Such identical fingerprints are caused by transferring a latent print to one or more objects. However, ink-jet printers (see. e.g. in Schwarz, An amino acid model for latent fingerprints on porous surfaces, 2009) and dispensing techniques (see e.g. Staymates et al., Evaluation of a drop-on-demand micro-dispensing system for development of artificial fingerprints, 2013]) allow for printing latent fingerprints to arbitrary surfaces for evaluation purposes. The overall possibility of misusing such techniques for tampering with crime scenes is recognized as a challenge for biometrics in forensics by Kiltz et al. (Printed fingerprints: a framework and first results towards detection of artificially printed latent fingerprints for forensics, 2011). Furthermore, it is very easy to modify the printing template and thus, creating sufficient variations of the fingerprints to avoid raising any suspicion. First subjective detection properties are proposed by Kiltz et al. for high-resolution data from chromatic white light sensors. An automated pattern recognition based detection approach for printed fingerprints is proposed by Hildebrandt et al. (Printed fingerprints at crime scenes: a faster detection of malicious traces using scans of confocal microscopes, 2013). This approach uses circle-based and edge-based features for detecting printed fingerprints as a countermeasure for such an attack on biometrics in forensics. Dittmann et al. (Context analysis of artificial sweat printed fingerprint forgeries: Assessment of properties for forgery detection, 2014) studies the attack chain in general. Moreover, potential context properties from the attack, which help to describe the attack, are described more

detailed. Furthermore, potential detection and context anomaly properties are derived and suggested as an enabler for a proper detection of such forgeries for forensic experts during forensic investigation and interpretation of traces. Our work addresses this application context and suggests a new detection property derived from the so-called Sample Production Properties. Here, our idea is to use Benford's Law (Benford, The law of anomalous numbers, 1938) to solve the overall challenge to reliably detect printed fingerprints during the analysis process.

Benford's Law describes the probability distribution for the most significant digits (1st digit) within natural data. The application of this law in the domain of image processing is described by Pérez-González et al. (Benford's law in image processing, 2007) in the spatial (pixel) domain and for DCT coefficients. The authors show that the luminance in the pixel domain does not follow Benford's Law. However, the block-based DCT coefficients are shown to follow the distribution of Benford's Law. This is used by Pérez-González et al. for performing stegoanalysis on the example of image watermarking. In image forensics, Benford's Law has been also applied to detect multiple JPEG compressions or for estimating JPEG2000 compression rates.

Despite the observations for arbitrary images in Pérez-González et al. we adopt this approach to analyze data from a Keyence VK-x110 confocal laser scanning microscope (CLSM) in the spatial domain towards the distribution of the most significant digits d of Benford's Law.

In particular, we use the laser intensity and topography from the CLSM for our experiments. We perform a pre-processing of the topography data by subtracting a plane determined using the least squares method to compensate a slight tilt of the sample and slightly different measurement distances. Especially the topography data can be considered as natural data because it basically describes the height of the residue on a particular substrate. The feature space is formed by the difference of the probability of each most significant digit to the distribution within Benford's Law. Afterward, we use the WEKA data mining software to train classifiers using supervised learning in order to build a model that is suitable for detecting printed fingerprints.

The results of the average distribution of the most significant digits confirm the observations of Pérez-González et al. for intensity images. However, the distribution for topography data is very similar to Benford's Law with the exception of the probability for the digit 1. Furthermore, we can observe that the distributions for real and for printed fingerprints are different.

Our experimental setup consists of 3000 printed and 3000 real fingerprint samples from four test subjects acquired using a Keyence VK-x110 CLSM equipped with a 10x magnification lens. Each sample covers an area of approximately 1.3 by 1 millimeters. The resulting image size is 1024x768 pixels. Our test goals are three fold: a) we evaluate the detection performance within a 10-fold stratified cross-validation for all 6000 samples, b) we evaluate the detection performance with separate test and training sets, each with 3000 samples, c) we evaluate the detection performance after simulating various distortions and influence factors using StirTrace (Hildebrandt et al., From StirMark to StirTrace: Benchmarking pattern recognition based printed fingerprint detection, 2014) to determine the robustness of the trained models from b).

The best performance for test goal a) is achieved using the RotationForest classifier with 99.07%. However, even the worst performance in this evaluation using the SMO classifier achieves a detection performance of 94.52%. Hence, the Benford's Law based features seem to be suitable to distinguish between real and printed fingerprints.

The performance for test goal b) is significantly lower. Here, the best performance of 98.03% is achieved using the Bagging classifier. The other classifiers achieve a performance between 84.67% and 92.2%. The differences in the performance are reasonable because the size of the training set is smaller. Furthermore, the scans used for training and testing are completely independent, which might increase the errors due to slight differences in the parametrization of the measurement device and environmental conditions.

In the full paper the benchmarking for test goal c) is performed with StirTrace using the same parametrizations as used in Hildebrandt et. al. (From StirMark to StirTrace: Benchmarking pattern recognition based printed fingerprint detection, 2014) to evaluate the robustness of the detection using Benford's Law. Furthermore, we motivate and evaluate a fusion of the Benford's Law based approach with the existing features space from Hildebrandt et al.

9409-9, Session 3

Capturing latent fingerprints from metallic painted surfaces using UV-VIS spectroscope

Andrey Makrushin, Tobias Scheidat, Claus Vielhauer, Fachhochschule Brandenburg (Germany)

Each contact of a fingertip with an object results in depositing an impression of friction ridges on the object's surface. Therefore, fingerprint detection and analysis is currently one of the most important means of personal identification used in crime scene investigations. Moreover, it is asserted that more crimes have been solved using fingerprint evidence than due to any other reason.

The visibility and persistency of a fingerprint impression primarily depends on the type of a substrate or, to be more precise, on its soaking and reflection properties. Forensic practitioners classify substrates in three categories regarding their ability to absorb fingerprint residues, namely porous, non-porous and semi-porous substrates. Porous substrates require invasive development approaches to render latent fingerprints visible e.g. applying powder, ninhydrin, fuming with cyanoacrylate, or vacuum metal deposition. Non-porous substrates do not soak fingerprint residues enabling optical non-invasive approaches for detection and lifting fingerprints. Semi-porous substrates may or may not absorb fingerprint residues depending on its viscosity as well as on the soaking properties of a substrate. Development approaches should be specified here for each particular case individually. Reflection is another characteristic of a substrate influencing a fingerprint's visibility. Based on our experience, optical fingerprint detection and lifting can be successfully applied only for smooth non-textured substrates also referred to as cooperative substrates. Structured, textured and semi-transparent surfaces immensely complicate or almost disable optical sensing.

Our focus is on metallic painted surfaces that are non-porous and "sparkling". We investigate the feasibility of contactless non-destructive optical sensors to detect, acquire and digitally process latent fingerprints which is considered the initial step prior to chemical development in digital crime scene forensics. Chemical/physical development is beyond the scope of this work.

Metallic paint is applied on a wide scale as a car body finish. Since vehicles are involved in many criminal cases, a metallic painted surface is highly relevant for forensic investigators as a potential carrier of fingerprints. Although a fingerprint on this substrate often can be seen with the naked eye, a conventional digital camera as well as an optical microscope is not able to take a photograph which exposes a fingerprint pattern clearly enough for further forensic investigations. The surface is semi-transparent and a beam of visible light penetrates it and is reflected off of the metallic flakes randomly disposed in the paint so that fingerprint residues impede the light beam in an insignificant manner. As a result, the fingerprint is invisible to sensors which make use of visible light. So for the majority of optical acquisition approaches, metallic paint poses a specific problem.

In our previous work, two optical microscopes: a chromatic white-light sensor (CWL) and a UV-VIS spectroscope (FTR) are compared regarding their ability to capture fingerprints from ten substrates most frequently occurring at a crime scene. Both sensors scan a surface point by point. The lateral distance between adjacent points is specified to be 50 ?m resulting in output images with a resolution of approximately 500 ppi. A metallic painted surface has been recognized as the substrate with a high potential for applying FTR. For other non-porous substrates, CWL produces fingerprint images of comparable quality. Since CWL is significantly faster, FTR is superfluous for these substrates. For instance, the scan time for a region of 10x15 mm is around ten minutes with CWL and around five hours with FTR considering the standard parameterization of the scanning process. Nonetheless, the previous work lacks quantitative analysis demonstrating only first trends based on experiments with three specimens.

To complement this, the extended empirical study in this paper comprises scanning 100 regions almost completely filled out with one fingerprint and 100 empty regions to achieve statistically more significant results. Here, we show that when scanning metallic painted surfaces, CWL is incapable of providing satisfactory fingerprint images and FTR delivers clear fingerprint images within the UV range of the electromagnetic spectrum.

A fingerprint's visibility within a scan is assessed automatically using the streakiness score, a new measure suggested recently. The streakiness score is designed as an objective fingerprint visibility measure, representing the ratio of a ridge-valley pattern within an image. A value resides within the interval [0,1], where 0 indicates an absence of a ridge-valley pattern and a value of around 0.8 indicates an almost perfect ridge-valley pattern. The computation of the streakiness score comprises three steps: (1) global filtering in the Fourier domain and binarization; (2) local enhancement by means of Gabor-filtering; (3) calculation of pixels in friction ridges. It was demonstrated that the streakiness score based visibility assessment highly correlates with human perception and therefore can be used for automatic distinguishing between empty and fingerprint regions.

The experiments to be presented in the final paper show that the average streakiness score for CWL scans of fingerprints yields 0.0657 and for CWL scans of empty regions 0.0295. These low values correlate with human perception and the negligible difference between them indicates the general invisibility of fingerprints and inappropriateness of CWL. In contrast, the average streakiness score for fingerprints in FTR scans exceeds 0.5 within the range of 205-385 nm just as for empty FTR scans it fluctuates at around 0.1 within the same range, enabling automatic distinguishing between fingerprint and empty regions solely based on streakiness scores.

Considering automatic distinguishing between empty and fingerprint regions to be a two-class problem, the classification performance ought to be measured in terms of equal error rate (EER). In our experiments, the EER values do not exceed 8.75% for any isolated wavelength within the range of 205-385 nm. The lowest EER of 1.71% is observed on margins with the wavelengths of 208.5 and 381 nm. However, this result should be considered optimistic because these wavelengths bear noisy images containing no texture of the outer layer and only a faint shape of the ridge-valley pattern. The realistic discrimination performance in terms of EER fluctuates at around 7.00% and can be observed with exemplary wavelengths of 254 and 324 nm. Here, EER yields 6.67% and 7.00% respectively. These experimental results allow us to declare a UV-VIS spectroscope to be indispensable for the optical examination of surfaces painted with metallic paint.

The final paper will be completed with the extended overview of related works, operating principles of CWL/FTR, details on calculation of the streakiness score, thorough description of the experimental setup and discussion of the experimental results including diverse CWL/FTR fingerprint images and diagrams reflecting the relationship between an FTR wavelength and the streakiness score.

9409-10, Session 3

Comparative study of minutiae selection algorithms for ISO fingerprint templates

Benoît Vibert, Jean-Marie Le Bars, ENSICAEN (France); Christophe M. Charrier, Univ. de Caen Basse-Normandie (France) and ENSICAEN (France); Christophe C. Rosenberger, ENSICAEN (France)

Nowadays, electronic transactions are part of our daily life (e-commerce, smartphones, physical access control . . .). In order to guarantee the security of authentication, biometrics is often used. Many real applications benefit from this technology such as for user control and e-payment. Nevertheless, a biometric data is very sensitive and cannot be revoked in general (like access password). In order to ensure its security and privacy, a biometric data is usually stored in a Secure Element (SE). The Secure Element could be a SmartCard, with an on Card comparison algorithm inside. Two steps are necessary when we using a biometric system, 1) the enrolment and 2) the verification. The On card comparison algorithm permits to compute a matching score between a captured biometric template with the reference one. It is a common, the biometric template stored in the SE follows the ISO Compact Card standard. 1 This template is composed of a set of minutiae represented by 3 octets (1 octet for the X location, 1 octet for the Y location, 1 octet for the Angle and the type). A SE has hardware and software constraints as for example the size of memory, the number of data we can send with an APDU command 2 (ISO 7816 standard for the communication with a SE). These limitations have an impact on the embedded algorithm and the size of the fingerprint template. Generally, in an operational application, a fingerprint template is limited to 48 minutiae when stored in a SE. When the sensor extracts more minutiae than the On-Card-Comparison(OCC) is able to process, we have to reduce the size of the template by selecting the most appropriate minutiae. Some automatic methods have been proposed in the state-of-the-art, 3 but it is tedious to determine if a method is better than another one. The aim of this paper is to determine how the way to select the minutiae modifies the performance of the biometric authentication. For this purpose, wepropose several selection methods. In our setting, we consider only methods which require no knowledge about the image where the minutiae template comes from and the algorithm used for the minutiae extraction.

9409-11, Session 3

Detection of latent fingerprints using high resolution 3D confocal microscopy in non-planar acquisition scenarios

Stefan Kirst, Otto-von-Guericke-Univ. Magdeburg (Germany) and Fachhochschule Brandenburg (Germany); Claus Vielhauer, Fachhochschule Brandenburg (Germany)

(i) Motivation and Application Context

From forensics to biometrics, the impact of fingerprints is very import. Nevertheless, the detection of traces is always crucial for further analysis. Common lifting methods in criminalistic forensics like sticky tape and powder may alter the trace. Using contactless non-invasive repeatable 2D and 3D surface scanning methods like confocal microscopy is an upcoming solution. Especially for large or distributed crime scenes, the digital process of analyzing traces may gain importance, due to the potential of (semi-) automated trace acquisition in future. However, in this concept, several new challenges arise, of which the detection of traces is obligatory. Whenever traces are acquired from surfaces of complex texture, it is hard to detect fingerprints, i.e. to distinguish between the trace and the background by means of pattern recognition methods. This becomes even more challenging when the contactless acquisition is not possible with a perpendicular sensor perspective due to the substrate's shape or environmental constraints.

A profile-based benchmarking of forensic methods for digitized forensics has been introduced, where one particular profile is based on the concept of a coarse scan [3]. This coarse scan typically covers a large area of a surface using considerably lower resolutions than the sensor is capable of, to realise the investigation target of determining parameters, materials and regions of interest [1]. Finding traces on challenging surfaces with limited resolutions is still quite difficult when it comes to fingerprints [2,6]. Using high resolutional data instead, a detection-by-segmentation approach like proposed in locksmith forensic scenarios [9], improves detection rates. This could already be shown in [8] with an exemplary test set, feature space and pre-processed downscaled raw data.

The used keyence sensory [7] in our work provides perfectly aligned data streams (intenstity, topography, color) with an unblured focus within the acquisition parameter.

By finding angle-based and substrate-independent classification models, an optimal acquisition setup per substrate can be determined, increasing the detection rate from the starting point during the acquisition process. In addition to that, knowing that a trace is not ascertainable prior to the actual capturing will also save time and effort. By using an inhomogeneous test set of surfaces for the determination of general classification models for the detection of fingerprints we exemplarily address the fact that substrate with fingerprint traces on crime scenes are not always the same.

(ii) Addressed Technical Challenge

Our approach for the detection of fingerprints aims for the determination of distinctive differences between the substrate and the trace in the field of non-planar scan scenarios on challenging surfaces. The challenge here is to distinguish between traces and different surface textures using a block-based classification approach and the determination of critical angles for all used surfaces, in which such a distinction is not reliable any more. Finding correlations between different surfaces with latent fingerprints using a block-based classification approach with a suitable feature set results in classification models for both, substrate-dependent and -independent detection. Especially higher acquisition angles at the limit of the sensor's numeric aperture and beyond, introduce a great amount of noise and outliers. This makes the detection more challenging, but also enables additional information. So when acquiring homogeneous and highly reflective substrates like a hard disc drive platter beyond the sensor's numeric aperture, the only reflecting light is originated from the diffuse reflection of the fingerprint residue. Such behavior is valuable and addressed by the proposed test set. Not only the design and the actual acquisition of the test set is a time-consuming task, also the determination of the ground truth for the two-class problem for the classification is an exhausting challenge. Since morphological- or other enhancement methods are not sufficient, the masking of traces and background is done manually. This is mandatory since the human perception is the established tool of fingerprint identification experts in crime scene forensics.

(iii) Conceptual Approach

In the presented approach we follow the process model provided in figure 1. We utilize a test set of eight different surfaces consisting of: hard disc drive platter, matte aluminium foil, oak furniture veneer, white furniture veneer, green car paint, glossy aluminium foil, brushed metal and glossy black plastic. Using white furniture, brushed metal and non-glossy car body finish, the final test set includes the same representative materials like used in [2,6]. Furthermore, we extend this test set using a HDD-platter, representing an ideal surface, and substrates like matt aluminium foil, representing additional cooperative surfaces. To work with different sweat compositions we introduced a inter-person- and intra-person variance by using 2 fingerprints of 10 donors resulting in 20 fingerprints for each surface for each acquisition angle. To represent the differences in non-planar acquisition we capture all substrates in five different angles using ten degree steps (0°..40° w.r.t the scanning surface perpendicular).

We apply our feature extraction on the perfectly aligned intensity, topography and color data from the utilized 3D CLSM. After the manual masking process we determine classification models for all angles as well as overall models using a total of 974400 instances. To distinguish between traces and surface textures even in non-planar scans we therefore combine a texture recognition approach [4] using statistical features on gray-level-co-occurrence matrices [10] with topography based surface texture techniques [5] utilizing a selection of roughness features with a variety of additional features in a blockwise manner. The approach in [5] was used to classify surfaces, whereupon we use a selection of those roughness features to distinguish between the trace and surfaces. Instead of a CWL-device like used in [2,6], in this work we use a 3D confocal laser microscope, since it is more useful in non-planar acquisition scenarios due to its confocal principal of measurement. Like in [8], we also adapt the detection-by-segmentation approach proposed in [9]. We furthermore utilize additional features like Tamura- [11], CEED- [12], FCTH- [13] and PHOG- [14] features to finalize the feature set to 1635 features. To identify a possible overfitting we use classifiers of different classes to build the resulting classification models.

(iv) Preliminary Results

First results on general classification models using scans from HDD-platters in angles from 0° to 40° and white veneer from 0° to 10° show up to 92.4% (kappa 0,84) correctly classified instances. These results are fairly the same for both, a percentage split (66% training set, 34% evaluation set) and a 10-fold crossvalidation. The linear dependency created by a crossvalidation will perform better when the classification models are build using the whole test set due to a greater variance in fingerprint residues from different donors. This will be monitored by using both testing options in combination with a variety of classifier to detect a possible overfitting.

The final results will be presented and discussed in detail for all substrate/angle settings. In the matter of ascertainability of traces, the results will indicate the maximum possible acquisition angles for fingerprint detection on every surface material within the test set using this approach. We will also be able to suggest optimal acquisition angles within the given scenarios for the trace detection on specific surfaces. We also provide classification models for substrate-independent detection of latent fingerprints. The results also outline the future work in regards to improvement of the approach, feature selection as well as the integration of new relevant features.

[1] Hildebrandt, Mario; Kiltz, Stefan; Grossmann, Ina; Vielhauer, Claus, "Convergence of digital and traditional forensic disciplines: a first exemplary study for digital dactyloscopy", in Proceedings of the thirteenth ACM multimedia workshop on Multimedia and security (MM&Sec '11), pp. 1-8, DOI: 10.1145/2037252.2037254, 2011

[2] A. Makrushin, T. Kiertscher, R. Fischer, S. Gruhn, C. Vielhauer, and J. Dittmann, "Computer-aided contact-less localization of latent fingerprints in low-resolution cwl scans," in Communications and Multimedia Security, ser. Lecture Notes in Computer Science, B. Decker and D. Chadwick, Eds. Springer Berlin Heidelberg, 2012, vol. 7394, pp. 89–98.

[3] Hildebrandt, Mario ; Dittmann, Jana ; Pocs, Matthias ; Ulrich, Michael ; Merkel, Ronny ; Fries, Thomas ; Vielhauer, Claus (Bearb.) ; Dittmann, Jana (Bearb.) ; Drygajlo, Andrzej (Bearb.) ; Juul, Niels Christian (Bearb.) ; Fairhurst, Michael C. (Bearb.): Privacy Preserving Challenges: New Design Aspects for Latent Fingerprint Detection Systems with Contact-Less Sensors for Future Preventive Applications in Airport Luggage Handling.. 6583. In: BIOID : Springer, 2011 (Lecture Notes in Computer Science). - ISBN 978-3-642-19529-7, S. 286-298

[4] E. Clausing, C. Kraetzer, J. Dittmann, and C. Vielhauer, "A first approach for the contactless acquisition and automated detection of toolmarks on pins of locking cylinders using 3d confocal microscopy," in Proceedings of the on Multimedia and security, ser. MM&Sec '12. New York, NY, USA: ACM, 2012, pp. 47–56.

[5] S. Gruhn and C. Vielhauer, "Surface classification and detection of latent fingerprints: Novel approach based on surface texture parameters," in Image and Signal Processing and Analysis (ISPA), 2011 7th International Symposium on, sept. 2011, pp. 678 – 683.

[6] A. Makrushin, M. Hildebrandt, R. Fischer, T. Kiertscher, J. Dittmann, and C. Vielhauer, "Advanced techniques for latent fingerprint detection and validation using a cwl device," pp. 84 360V–84 360V–12, 2012.

[7] Keyence Corporation, "Vk-x100/x200 series 3d laser scanning microscope," [Online] available: http://www.keyence.com/products/microscope/microscope/vkx100200/vkx100 200 specifications 1.php, last checked 20/07/2014, 2013.

[8] S. Kirst, "Digitized Forensics: Segmentation of Fingerprint Traces on Non-Planar Surfaces Using 3D CLSM", in Magdeburger Informatiktage 2014, pp 17-22, ISBN 978-3-944722-12-2

[9] E. Clausing and C. Vielhauer, "Digitized locksmith forensics: automated detection and segmentation of toolmarks on highly structured surfaces," pp. 90 280W–90 280W–13, 2014.

[10] R. Haralick, K. Shanmugam, and I.Dinstein, "Textural features for image classification," in IEEE Transactions on Systems, Man, and Cybernetics SMC , vol. 3 (6), 1973, pp. 610–621.

[11] H. Tamura, S. Mori, T. Yamawaki. Textural Features Corresponding to Visual Perception. IEEE Transaction on Systems, Man, and Cybernetcs, Vol. SMC-8, No. 6, pp. 460–472, June 1978.

[12] S. Chatzichristofis, Y. Boutalis, CEDD: Color and Edge-directivity Descriptor: A Compact Descriptor for image Indexing and Retrival, in Computer Vision Systems, Lecture Notes in Computer Science Volume 5008, 2008, pp 312-322

[13] S. Chatzichristofis, Y. Boutalis, FCTH: Fuzzy Color and Texture Histogram - A Low Level Feature for Accurate Image Retrieval, in WIAMIS '08 Proceedings of the 2008 Ninth International Workshop on Image Analysis for Multimedia Interactive Services Pages 191-196, ISBN: 978-0-7695-3130-4

[14] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in Proceedings of the ACM International Conference on Image and Video Retrieval, pp.401-408, ISBN: 978-1-59593-733-9, 2007.

9409-12, Session 3

Benchmarking contactless acquisition sensor reproducibility for latent fingerprint trace evidence

Mario Hildebrandt, Otto-von-Guericke-Univ. Magdeburg (Germany); Jana Dittmann, Otto-von-Guericke-Univ. Magdeburg (Germany) and The Univ. of Buckingham (United Kingdom)

The reproducibility of sensor signals is an important topic in biometrics and forensics. Recently, optical, nano-meter range, contactless, non-destructive acquisition techniques are investigated and evaluated to be applied in crime scene trace forensics allowing for a detailed, full digitization of the trace without altering it. In result such techniques enable parallel investigations of different trace types without affecting each other. Particular examples are introduced e.g. in Leich et al. (Non-destructive forensic latent fingerprint acquisition with chromatic white light sensors, 2011) for fingerprint acquisition, for fiber traces in Arndt et al. (First approach for a computer-aided textile fiber type determination based on template matching using a 3d laser scanning microscope, 2012) or for locksmith forensics in Clausing et al. (A first approach for the contactless acquisition and automated detection of toolmarks on pins of locking cylinders using 3d confocal microscopy, 2012). Such techniques can be game-changing for forensic investigations by allowing for a more detailed, non-destructive investigation of a multitude of different traces. However, in order to achieve forensic acceptance of an acquisition and processing technique, particular testing criteria (expert witness testimony) exist. In the US, a Daubert challenge is often required for assessing scientific evidence in court. In a Daubert hearing multiple questions can be assessed: whether the theory or technique behind can be or is tested (1), is reviewed and published (2), shows a particular or potential error rate (3), its operation is in accordance to a standard procedure (4) and is accepted in the relevant community (5).

Especially error rate criterion (3) requires a detailed evaluation before a new technology can be accepted and used by forensic experts. Therefore, such new contactless acquisition techniques need to ensure appropriate quality and reproducibility of results as a foundation for further investigation steps. In this respect we find in Kiltz et al. (Revised benchmarking of contact-less fingerprint scanners for forensic fingerprint detection: challenges and results for chromatic white light scanners (CWL), 2011) a benchmarking approach suggesting six properties describing legal requirements, application-related properties, surface material properties, technical properties, input sensory technology and processing methods to allow for comparing such new acquisition sensory for particular forensic use cases. Particular use cases belong to a certain acquisition and sample pre-processing context and offer the possibility to describe relevant properties, which need to be tested in accordance to (3).

Within this work our goal is to test the sensor reproducibility for the latent fingerprint acquisition by comparing four sensing approaches (chromatic white light sensor FRT CWL 600 and FRT CWL 1mm, confocal laser scanning microscope (CLSM) Keyence VK-X110 and NIR/VIS/UV spectrometer FRT FTR). For a privacy-conform testing we use the latent fingerprint test set from Hildebrandt et al. (Creation of a public corpus of contact-less acquired latent fingerprints without privacy implications, 2013), which uses SFinGe for generating fingerprint patterns combined with artificial sweat printing as suggested by Schwarz (An amino acid model for latent fingerprints on porous surfaces, 2009) to create artificial latent fingerprint samples. Previous work in Sturm et al. (High quality training materials to detect printed fingerprints: Benchmarking three different aquisition sensors producing printing templates, 2013) compared three different acquisition sensors for producing printing templates for high quality training material for forensic experts to learn the detection of fingerprint forgeries by artificial sweat printing on the example of printed, real, latent fingerprints.

Our approach is to compare two scans by using a simple, known feature space set derived from image processing: pseudo-correlation and absolute difference of average gray values as used in Sturm et al., Pearson product moments correlation, cross-correlation and mean squared error in spatial and frequency domain (resulting in 10 features). The sensor reproducibility is measured by calculating the ten features between two scan images (intensity data). The resulting feature space is elaborated with six classifiers from the WEKA data mining software to see which classifier is suitable for handling reproducibility testing by considering all of the ten features and classifying scan data into small differences (reproducible) and large differences (non-reproducible).

Beside signal processing features, we employ additionally a biometric reproducibility score by using a NIST NBIS to evaluate if the scan data for a particular fingerprint achieves successful biometric matching rates. The successful matches are interpreted as scan reproducible. For plausibility checking the average matching scores are additionally compared with the results from the ten dimensional feature space. For the classification as well as for the biometric matching, a separate plausibility check is performed based on different fingerprint patterns by studying erroneous small difference classifications and correct large difference classifications for the ten features and false correct matches for the biometric matching.

In our experiments we use the privacy conform 24 different computer generated fingerprint patterns (Sorigin): original samples, printing templates), printed with artificial sweat (Canon PIXMA iP4600) creating physical trace samples (Strace), printed samples) on an overhead foil from Hildebrandt et al.. The test set of scans from a CWL600 in 500 and 1000ppi resolution for the classification tests of intra-sensor reproducibility (the 48 scans from Hildebrandt et al. is extended by acquiring it a second time resulting in 96 images (scan samples Sscan). Furthermore, for inter-sensor reproducibility additionally the first three printed computer generated fingerprints from Hildebrandt et al. are used and consecutively scanned two times using the four sensors with different acquisition resolutions ranging from 500ppi to approx. 4000ppi.

In summary, we perform four different evaluations: (1) classifier training and cross-validation of 500 and 1000ppi CWL600 scans in comparison to the printing template, (2) intra-sensor evaluation for the 96 CWL600 scans by comparing the scan samples with each other, (3) inter-sensor evaluation by comparing scan samples from different sensors with each other, (4) intra-sensor evaluation of consecutive scans.

In order to reduce the amount of comparisons with the spectrometer data which captures 2048 distinct wavelength images, we chose 8 arbitrary wavelengths from the NIR, VIS and UV band (250, 350, 450, 500, 550, 600, 650, 800nm). Our contributions in comparison to Sturm et al. can be divided into: a) Comparison and benchmarking of different sensing approaches in respect to the ability of reproducing a particular trace during acquisition: we define the Acquisition Context with details about sample acquisition by considering sensor properties, sensor parameters, environment, physical trace sample properties, acquisition noise and additionally define the Printed Fingerprint Source Context with details about original sample source (printing template with properties for original sample, original sample pre-processing) and sample printing (with printer & printer settings, ink properties, printing surface, environment and printing defects & artifacts). Furthermore, the Sample Pre-Processing Context is described with all pre-processing methods performed (before feature extraction of 10 dimensional feature space and calculation of the biometric reproducibility score). This allows for a detailed summary of testing influence factors in later assessments and can be seen as enhancement to the state of the art towards a more precise description of the conditions and assumptions for determining the error rates in the evaluation. b) We extend the objective feature space by further known, statistical features applied to the sensor data in spatial and frequency domain, resulting in overall 10 features, which are classified into reproducible and non-reproducible samples by using known classifiers. Furthermore, we introduce an automated sample alignment as prerequisite for the feature space calculation. c) We benchmark the reproducibility each sensor in an intra-sensor evaluation of consecutive scans. We show that the spatial and frequency feature space from a) is suitable in most cases for being used with Bagging classifiers, which achieve consistent results with the biometric verification rates (biometric reproducibility score).

Our results indicate that the CWL600 and spectrum sensing show in summary the best reproducibility. Here, for the intra-sensor evaluation of consecutive scans from a CWL600 sensor all samples are recognized correctly as small difference and the samples are successfully matched with an average matching score of 60.

Conference 9409: Media Watermarking, Security, and Forensics 2015

9409-31, Session Key

Steganography: the past ten years *(Keynote Presentation)*

Jessica Fridrich, Binghamton Univ. (United States)

In my talk, I will contrast the state of the art in 2004 and today highlighting the main achievements in the field, including information-theoretical analysis, scaling laws, quantitative steganalysis, coding, content-adaptive steganography, rich media models, optimal detectors derived using hypothesis testing, game-theoretical formulation, and algorithms aimed at large scale deployment of steganalysis in the real world. I will also look back and identify the main bottlenecks and areas where we are likely to see major breakthroughs in the next ten years.

9409-13, Session 4

Design of a steganographic virtual operating system

Elan Ashendorf, Scott A. Craver, Binghamton Univ. (United States)

A steganographic file system is a secure file system whose very existence on a disk is concealed. Customarily, these systems hide an encrypted volume within unused disk blocks, slack space, or atop conventional encrypted volumes. These file systems are far from undetectable, however: aside from their ciphertext footprint, they require a software or driver installation whose presence can attract attention and then targeted surveillance.

We describe a new steganographic operating environment that requires no visible software installation, launch- ing instead from a concealed bootstrap program that can be extracted and invoked with a chain of common Unix commands. Our system conceals its payload within innocuous files that typically contain high-entropy data, producing a footprint that is far less conspicuous than existing methods. The system uses a local web server to provide a file system, user interface and applications through an web architecture.

9409-14, Session 4

Content-Adaptive Pentary Steganography Using the Multivariate Generalized Gaussian Cover Model

Vahid Sedighi, Jessica Fridrich, Binghamton Univ. (United States); Remi Cogranne, Univ. de Technologie Troyes (France)

The aim of this paper is to present a significant improvement in the recent adaptive steganographic scheme based on the Multivariate Gaussian (MVG) model. This methodology is focused on finding the change rates that minimize the overall Fisher Information of embedding process. In practice, variance of pixels are unknown and have to be estimated. In this paper, it is specially shown that the performance of an adaptive scheme based on such methodology heavily relies on the variance estimator. Preliminary numerical results using different variance estimators also show that the original ensuing adaptive scheme can be greatly improved.

9409-15, Session 4

Towards dependable steganalysis

Tomas Pevny, Czech Technical Univ. in Prague (Czech Republic); Andrew D. Ker, Univ. of Oxford (United Kingdom)

No Abstract Available

9409-16, Session 4

Deep learning for steganalysis via convolutional neural networks

Yinlong Qian, Univ. of Science and Technology of China (China); Jing Dong, Wei Wang, Tieniu Tan, Institute of Automation (China)

Steganalysis is usually formulated as a binary classification problem to distinguish cover and stego objects. Existing methods build steganalyzers in two steps: feature extraction and classification. The success of steganalyzers mainly depends on the feature design. Currently, the most successful way to design feature is to firstly compute a family of noise residuals, then assemble multiple weak feature sets from different residuals to form a powerful high-dimensional set. By this way, the dimensionality of the feature set has risen from several hundreds to tens of thousands (e.g. 30,000 or more) in the past few years. But it is rarely known which features are important for steganalysis. Hence, feature design by assembling a large amount of weak features to form a high-dimensional feature set becomes more and more experiential and difficult.

This paper introduces deep learning for steganalysis, and we consider feature learning for steganalysis as a brand new paradigm. The idea of Deep Learning (DL) is considered to offer a better way to represent low-level information to high-level one. Deep learning models can learn hierarchical representations automatically by using architectures composed of multiple levels of non-linear operations. These models can solve the classification problem all the way from pixels to classifier automatically. It has theoretically as well as practically proved to be a more powerful learning scheme for many research areas such as computer vision and speech recognition.

In our paper, we conduct our deep learning based steganalyzer via a designed Convolutional Neural Network (CNN). CNNs, as one of the most representative deep learning models, automatically learn features directly from the raw input. They have been successfully applied to computer vision (CV) tasks. A typical CNN is composed of two kinds of layers: convolutional layer and fully connected layer. The output of each convolutional layer is a set of arrays called feature maps. At a convolutional layer, three kinds of operations, filtering, non-linearity, and pooling, are usually applied sequentially. For filtering, each output feature map usually combines convolutions with multiple inputs. It can capture local dependencies among neighboring elements. The element-wise non-linearity suppresses the output between fixed bounds. Then, pooling reduces the spatial resolution of each feature map and translates local information into more global one. Average pooling and max pooling are two typical pooling methods. With multiple convolutional layers, higher-level features are progressively extracted from lower-level ones. Then the learned features are passed to several fully connected layers for classification.

However, the steganalysis task is quite different from CV ones. Thus the feature representations of steganalysis-designed CNN should be different from existing CNNs. Actually, we have tested the existing CNN models developed for CV tasks as steganalysis models. But it turns out to be a failure, which means those CNN models are hard to capture the statistical properties that are important for steganalysis. Therefore, in the framework of deep learning, we propose a customized CNN model called Gaussian-Neuron CNN (GNCNN) for steganalysis purpose.

The proposed model takes an image as input. The first layer is a preprocessing layer that contains only filtering operation. We use the

widely used KV kernel for initialization. Followed by this layer are several convolutional layers for feature representation. The filtering has two important effects. First, it can be considered as a kind of linear projection. Hence, with different learnable kernels, more dependencies can be captured. Second, the projected data will have the distribution that concentrated at zero if the kernels are high-pass. The choice of non-linearity is key here, and it is designed as Gaussian function. By using it, the values around zero are strengthened. In steganalysis, a good model should capture the robust statistical characteristics of images, and such values are usually robust and useful because of its high probability of occurrence. Therefore, by Gaussian non-linear regularization, we enforce our model to learn the desired high pass filters. To our best knowledge, it is the first time that Gaussian function is used as non-linearity operation in deep CNNs, and we call this kind of CNN as Gaussian-Neuron CNN (GNCNN). In our work, we find that average pooling has a much better performance compared to max pooling. The reason should be that, compared with the mean value, the max value of the local region is unstable since the GNCNN works on some kind of noise. Then the learned features at the top convolutional layer are passed to several fully connected layers for classification. Feature learning and classification steps are unified under a single architecture. This means all the parameters in these two steps are optimized automatically and jointly.

The novel contributions of the proposed model are outlined as follows: 1) We introduce deep learning methods for steganalysis to learn hierarchical features automatically. 2) We propose a customized CNN model called GNCNN which considers some special traces caused by steganography. It can directly learns feature representations from raw pixels.

We verified our idea on deep learning for steganalysis in the experiments. The standardized database, BOSSbase 1.01 that contains 10,000 images, is used. We resized all those images into the size of 256*256. This processing of resizing is only due to the limitation of our computational capabilities. All the experiments were carried out on the same database. We built our GNCNN with one preprocessing layer, five convolutional layers, and three fully connected layers. We learned 256 features which are the output of the fifth convolutional layer. We compared our model with Spatial Rich Model (SRM) feature set implemented with Ensemble Classifiers on HUGO using three payloads, 0.3, 0.4, and 0.5 bpp, respectively. The detection error of our model is just about 3%-5% higher than that of SRM depending on the payload. The final manuscript will also include the detection error of our model against SRM on other two state-of-the-art embedding methods (WOW and S-UNIWARD). Additionally, we will collect images from Internet to construct a large and realistic data set to do more experiments. Finally, the computational complexity of the proposed scheme will be discussed and compared with existing methods.

9409-32, Session Key

Ultra-high definition, watermark detection, mobile video, and much more: a status report on ATSC 3.0 *(Keynote Presentation)*

Jerry Whitaker D.V.M., Madeleine Noland, Advanced Television Systems Committee (United States)

ATSC 3.0 is the next step in the evolution of television broadcasting. Expected to include Ultra High-Definition video, Internet compatibility, enhanced interactivity, on-demand service options, improved audience measurement tools, new audio and video functions with advanced codecs, and robust mobility, experts from around the work are working to develop next-generation services under the "ATSC 3.0" banner. This presentation will detail the sweeping overhaul of the broadcast TV standard now underway, including overviews of the Physical Layer, Protocols and Management Layer, and Applications and Presentation Layer. Special emphasis will be given to the video feature set expected from the developing standard, and efforts currently underway to evaluate watermark technologies for use in the new system.

9409-27, Session 5

Counter-forensics in machine learning based forgery detection

Francesco Marra, Giovanni Poggi, Univ. degli Studi di Napoli Federico II (Italy); Fabio Roli, Univ. degli Studi di Cagliari (Italy); Carlo Sansone, Luisa Verdoliva, Univ. degli Studi di Napoli Federico II (Italy)

With the powerful image editing tools available today, it is very easy to create forgeries without leaving visible traces.

Boundaries between host image and forgery can be concealed, illumination changed, and so on, in a naive form of counter-forensics.

In fact, most modern techniques for forgery detection rely on the statistical distribution of micro-patterns, enhanced through high-level filtering, and summarized in some image descriptor used for the final classification.

In this work we propose a strategy to modify the forgery at the level of micro-patterns to fool a state-of-the-art forgery detector.

Then, we investigate on the effectiveness of the proposed strategy as a function of the level of knowledge on the forgery detection algorithm.

Experiments show this approach to be quite effective provided a good prior knowledge on the detector is available.

9409-28, Session 5

Anti-forensics of chromatic aberration

Matthew C. Stamm, Drexel Univ. (United States)

Over the past decade, a number of information forensic techniques have been developed to identify digital image manipulation and falsifications. However, recent research has shown that an intelligent forger can use anti-forensic countermeasures to disguise their forgeries. In this paper, an anti-forensic technique is proposed to falsify the chromatic aberration present in a digital image. Chromatic aberration corresponds to the relative contraction or expansion between an image's color layers that occurs due to a lens's inability to focus all wavelengths of light on the same point. Previous work has used localized inconsistencies in an image's chromatic aberration to expose cut-and-paste image forgeries. The anti-forensic technique presented in this paper operates by estimating the expected lateral chromatic aberration at an image location, then removing deviations from this estimate caused by tampering or falsification. Experimental results are presented that demonstrate that this anti-forensic technique can be used to effectively disguise evidence of an image forgery.

9409-29, Session 5

An overview of methods for countering PRNU based source attribution and beyond

Ahmet E. Dirik, Uludag Univ. (Turkey); Husrev Taha Sencar, TOBB Univ. of Economics and Technology (Turkey); Nasir D. Memon, Polytechnic Institute of New York Univ. (United States)

Photo response noise uniformity (PRNU) based source attribution has proven to be a powerful technique in multimedia forensics. The increasing prominence of this technique, combined with its introduction as evidence in the court, brought with it the need for it to withstand anti-forensics.

Although robustness under common signal processing operations and geometrical transformations have been considered as potential attacks on this technique, new adversarial settings that curtail the performance of this technique are constantly being introduced. Starting with an overview of

proposed approaches to counter PRNU based source attribution, this work introduces seam-carving based image resizing and photographic panoramas as two such approaches and discusses how to defend against them.

9409-18, Session 6

Disparity estimation and disparity-coherent watermarking

Hasan Sheikh Faridul, Technicolor (France); Gwenaël Doërr, Séverine Baudry, Technicolor S.A. (France)

In the context of stereo video, disparity-coherent watermarking has been introduced to provide superior robustness against virtual view synthesis, as well as improving perceived fidelity. Still, a number of practical considerations have been overlooked and in particular the role of the underlying depth estimation tool on performances. In this article, we explore the interplay between various stereo video processing primitives and highlight a few take away lessons that should be accounted for to improve performances of future disparity-coherent watermarking systems. In particular, we highlight how inconsistencies between left and right disparity maps impact performances, thereby calling for innovative designs.

9409-19, Session 6

Estimating synchronization signal phase

Robert C. Lyons, John Lord, Digimarc Corp. (United States)

To read a watermark from printed images requires that the watermarking system read correctly after affine distortions. One way to recover from affine distortions is to add a synchronization signal in the Fourier frequency domain and use this synchronization signal to estimate the applied affine distortion. Using the Fourier Magnitudes one can estimate the linear portion of the affine distortion. To estimate the translation one must first estimate the phase of the synchronization signal and then use phase correlation to estimate the translation. In this paper we provide a new method to measure the phase of the synchronization signal using only the data from the complex Fourier domain. This data is used to compute the linear portion, so it is quite convenient to estimate the phase without further data manipulation. The phase estimation proposed in this paper is computationally simple and provides a significant computational advantage over previous methods while maintaining similar accuracy. In addition, the phase estimation formula gives a general way to interpolate images in the complex frequency domain.

9409-20, Session 6

Mobile visual object identification: from SIFT-BoF-RANSAC to Sketchprint

Sviatoslav V. Voloshynovskiy, Maurits Diephuis, Taras Holotyak, Univ. de Genève (Switzerland)

Discriminative and robust content representation plays a central role in digital fingerprinting and content identification. Complex post-processing methods are used to compress descriptors and their geometrical information, aggregate them into more compact and discriminative representations and finally re-rank the results based on the similarity geometries of descriptors. In addition, the security and privacy of content representation has become a hot research topic in multimedia and security communities. In particular, it was demonstrated that the joint storage of descriptors and their spatial geometries makes it possible to roughly reconstruct the original images constituting a serious privacy leak.

In this paper, we introduce a new framework for semi-local content representation based on Sketchprint descriptors. It extends the properties of local descriptors to a more informative and discriminative, yet geometrically invariant content representation. In particular it allows images to be compactly represented by several sketch descriptors without being fully dependent on re-ranking methods.

We consider several use cases, applying Sketchprint descriptors to natural images, text documents, packages and micro-structures where traditional local descriptors demonstrate poor performance.

9409-21, Session 6

Analysis of optically variable devices using a photometric light-field approach

Daniel Soukup, AIT Austrian Institute of Technology GmbH (Austria); Svorad ?tolc, AIT Austrian Institute of Technology GmbH (Austria) and Institute of Measurement Science (Slovakia); Reinhold Huber-Mörk, AIT Austrian Institute of Technology GmbH (Austria)

Diffractive Optically Variable Image Devices (DOVIDs), sometimes loosely referred to as holograms, are popular security features for protecting banknotes, ID cards or other security documents. The main idea behind holography is to capture a full 4-D light field of a scene on a 2-D analog recording medium. In this context, most DOVIDs are not “proper” holograms as they only exploit specific diffraction grating patterns in order to generate desired color-changing effects depending on both the viewing as well as illumination angle. Nonetheless, for the sake of precise examination of the DOVID security features (e.g., for the purpose of authentication or quality inspection), one has to consider methods and tools that are capable of capturing a broad range of spatial illumination and response configurations for any kind of hologram.

Existing equipment for forensic hologram/DOVID analysis is based either on microscopic analysis of the grating structure or sparse point-wise projection and recording of diffraction patterns. The state-of-the-art tool for hologram verification used in forensic investigations is the Universal Hologram Scanner (UHS). The UHS records diffraction patterns at discrete steps over the hologram area and performs an orientation vs. frequency analysis on the recorded data. For non-forensic analysis method for hologram verification using mobile devices was suggested recently. Unfortunately, no details or results on hologram verification performance were provided, the authors instead evaluated user performance w.r.t. guided navigation to image capturing positions.

From the acquisition point of view, light-field-based methods provide a promising option for the inspection of DOVIDs. Both plenoptic area-scan as well as multi-line-scan cameras acquire a part of the light-field function, which describes the intensity (radiance) of light passing through every point in space coming from every direction. Even though most practical light-field recordings do not cover all light directions, it is still possible to infer depth information from them. On the other hand, a narrow range of angles comprised in practical light-fields is typically not sufficient for revealing the full optical variability of the DOVID features.

Furthermore, as the illumination angle also plays a vital role in revealing the DOVID diffractive behavior, we suggest to use photometric illumination arrangements (i.e., changing the observer vs. illumination angle by solely varying the light source position) in order to broaden the range of directional stimuli at the grating structure. Indeed, a pure photometric setup would not require a light-field camera. The motivation for using a light-field camera in combination with a photometric stereo is to cover directional variation at two scales: (i) fine variations in the observer angle delivered by the light-field captured by the camera and (ii) coarse variations in the illumination angle in the photometric stereo. This allows for a detailed sampling the Bidirectional Reflectance Distribution Function (BRDF) of the DOVID.

In this paper, we propose a hybrid photometric light-field approach suitable for a precise analysis of hologram/DOVID security features through capturing their BRDF. The method is demonstrated on a practical task of the automated classification of different DOVID types. For this purpose, we propose a tailored feature descriptor which is robust against several

expected sources of inaccuracy but still specific enough for the given task. The suggested hybrid approach is analyzed from both theoretical and practical viewpoints and its advantages w.r.t. both methods (i.e., photometric stereo and light-fields) alone are discussed in detail. We show that the combination of both methods provides a reliable and robust tool for inspecting DOVID microstructures as well as overall DOVID behavior.

9409-33, Session Key

Do wearables really change anything? *(Keynote Presentation)*

Brian J. Hernacki, Intel Corp. (United States)

Security, privacy and wearables are all hot topics in the tech industry today but have we really thought through where these interests meet? Do wearables present a large change in how we do things? Do they present new security or privacy problems? Or are they just the same problems we've been dealing with in a new package? This presentation examine both the threats to security and privacy in the wearable space and explores how the market and form factors differ (or don't).

9409-22, Session 7

Phase-aware projection model for steganalysis of JPEG images

Vojtech Holub, Binghamton Univ. (United States) and Digimarc Corp. (United States); Jessica Fridrich, Binghamton Univ. (United States)

Statistical features used for JPEG steganalysis can be divided into two types. The first forms the features from quantized DCT coefficients, typically as co-occurrences of various intra and inter-block coefficient neighbors. An example of this approach is the JPEG Rich Model (JRM). In the second approach, the features are extracted from the spatial domain representation of the JPEG image. One may directly use one of the spatial rich models, such as the SRM, the Projection Spatial Rich Model (PSRM) or its version specifically adapted for detection in the JPEG domain by increasing the quantization step (PSRMQ3).

The features extracted directly from the quantized DCT coefficients appear slightly more effective against older JPEG steganographic algorithms that introduce characteristic artifacts into the distribution of JPEG coefficients, such as nsF5. Spatial domain features appear far superior against modern JPEG steganographic schemes, such as J-UNIWARD, which seems to preserve the dependencies among JPEG coefficients captured by the JRM rather well. Currently, the most successful feature set against J-UNIWARD is PSRMQ3, even though the projection form of the spatial rich model was originally designed for spatial domain steganalysis.

In this abstract, we introduce a new projection-type feature set inspired by the PSRM and the newly introduced DCTR that offers better detection performance than the DCTR at the fraction of the computational cost of the PSRMQ3. We call the feature set the Phase-Aware Projection Rich Model (PAPRM).

9409-23, Session 7

JPEG quantization table mismatched steganalysis via robust discriminative feature transformation

Likai Zeng, Xiangwei Kong, Ming Li, Yanqing Guo, Dalian Univ. of Technology (China)

The success of traditional steganalysis relies on the assumption that both training and testing samples are from the same source with similar statistical properties and feature distributions. However, this assumption is usually not true in practice, which may result in the mismatch problem and the drop in detection accuracy. In this paper, we present a novel method to mitigate the problem of JPEG quantization table mismatch, named as Robust Discriminative Feature Transformation (RDFT). RDFT transforms original features to new features based on non-linear transformation matrix. It can improve the statistical consistency of the training samples and testing samples and learn new matched feature representations from original features by minimizing feature distribution difference whilst preserving the classification ability of training data. The comparison to prior arts reveals that the detection accuracy of the proposed RDFT algorithm can significantly outperform traditional steganalyzers under mismatched conditions and it is close to that of matched scenario. RDFT has several appealing advantages: 1) it can improve the statistical consistency of the training and testing data; 2) it can reduce the distribution difference between the training features and testing features; 3) it can preserve the classification ability of the training data; 4) it is robust to parameters and can achieve the optimal performance under wide range of parameter values.

9409-24, Session 7

CFA-aware features for steganalysis of color images

Miroslav Goljan, Jessica Fridrich, Binghamton Univ. (United States)

It is rather surprising that color image steganalysis that uses the more complex structure of color images has largely been neglected by the research community. Recently, the authors of this abstract proposed a spatio-color rich model consisting of two parts – the spatial rich model (SRM) computed from the union of all three channels and the color rich model (CRM) formed by 3D co-occurrences of residuals taken across the color channels rather than in the spatial direction.

We propose a further extension of the CRM that is aware of the underlying spatial alignment of the Bayer color filter array (CFA). In order to apply the new feature set, the training as well as the testing needs to be carried out on images that were spatially synchronized w.r.t. the CFA position. Since the position of the Bayer CFA can be rather reliably detected in a single image using existing techniques, it is entirely feasible to assume that the steganalyst can spatially synchronize the training database as well as the tested image. The new CFA-aware features can boost the steganography detection in color images that exhibit traces of color interpolation, and this holds true for detection of both non-adaptive and adaptive steganography.

9409-25, Session 7

Segmentation based steganalysis of spatial images using local linear transform

Ran Wang, Xijian Ping, Tao Zhang, Zhengzhou Information Science and Technology Institute (China)

As an important issue in multimedia security, image steganalysis has attracted extensive attention. Although various steganalyzers have been presented, there are still many challenges in the field of steganalysis attacking the spatial domain steganography. The steganalysis features are affected by image content, and the impact is more serious than embedding. However, most existed steganalysis methods are carried out based on the statistical changes of the image data caused by secret messages embedding without considering the inherent features of image contents, which makes the steganalysis performance closely tied to the image content and quality of the experimental database. Especially, when the training and testing images are from different databases with dissimilar statistical features, the detection results are obviously affected, which makes it difficult to move the research of steganalysis from laboratory into practical application.

In this paper, a new image content-based blind steganalysis method for spatial image is proposed. The image characteristics change differently

according to the image content complexity after embedding, while the steganalysis features of the image regions with the same content complexity are similar. As a result, segmenting the given image to several sub-images assorted by the texture complexity will make the steganalysis features of each category of sub-images more centralized, and it will be easier to distinguish the cover images from the stego ones. The proposed algorithm is composed by segmentation, steganalysis extraction, training and testing. We have combined the segmentation based steganalysis framework with JPEG Rich Model feature set in paper [1] and obtained good performance. In this paper, we extend the framework to spatial domain.

When segmenting, we divide the image into several nonoverlapping blocks with the same size, and extract the features which can measure the texture complexity from each block. The blocks are classified according to the texture complexity features, and each category of blocks is seemed as one sub-image. Thus the texture complexity of the same category of sub-image from different images will be the same, and the effects of image content to the steganalysis result can be eliminated.

After segmenting, steganalysis feature set based on local linear transform (LLT) is extracted from each sub-image. LLT is a general computational framework in which an image is convolved with a bank of masks with relatively small sizes for all local neighborhoods in a sliding window fashion. The statistical changes caused by embedding can be regarded as the variety of image texture, while LLT can capture the slight changes of the local stochastic textures. The LLT masks we used are sensitive to image details such as edges, lines, and isolated points. We compute the k-th order LLT residuals by convolving the image for k times, quantize the LLT residuals to curb the dynamic range and extract features from the most significant part. The high order LLT residuals are modeled as Markov processes, and the third order adjacent occurrence matrices in three directions are extracted as the steganalysis features. In the proposed algorithm, we segment the image into 3 sub-images. 20 different LLT masks are selected and the truncated threshold is set to 1. As a result, the total feature dimension is 7380.

The LLT based steganalysis features extracted from each category of sub-images are trained separately to form a classifier. In the testing phase, the steganalysis feature of each segmented sub-image is sent to the corresponding classifier and the final decision is made using a weighted fusing process. The ensemble classifier proposed in [2] is employed.

We ran the experiments on several image databases including BOSSBase, BOWS2, UCID and a combined database. We generate stego images by embedding data with different message lengths and different embedding algorithms. The five steganography embedding methods we considered are: LSBM, AELSB, EA, WOW and HUGO. In order to meet the needs of practical application, we test the algorithms in different conditions, including when the training and testing database are the same and different. When the training and testing images are from the same database, compared with the 12753-dimensional SRM feature [3], the proposed feature set can obtain almost the same or better results when detecting LSBM and AELSB. While for EA, WOW and HUGO, the proposed method is a little weaker than SRM. The minimal total error of the proposed method is about 3 percent higher than SRM. Although the performance is not as good as SRM, the proposed algorithm has smaller feature dimension and the time for extracting the steganalysis features are only about one quarter of SRM. When the training and testing images are from different image databases, the detection results of the proposed method in most circumstances are better than SRM method for about 1 to 2 percent. This indicates that unlike the traditional steganalysis algorithms, the proposed method is not much affected when the training and testing database are mismatched. This is because the proposed algorithm is based on the image content and has sufficient consideration of the characteristics of the image itself. With this advantage, the proposed content based method can work better when there is a considerable diversity in image sources and contents, which is quite useful in the practical application.

In this paper, a steganalysis method of spatial images based on segmentation and local linear transform is proposed. The proposed method has sufficiently considered the influence of the content difference of images to the steganalysis result and can obtain good performance.

REFERENCES

[1] R. Wang, et al, “Steganalysis of JPEG images using block texture based rich models,” J. Electron. Imaging, vol. 22, no. 4, pp. 043033, 2013.

[2] J. Fridrich, J. Kodovsk?, M. Goljan, V. Holub, “Steganalysis of content-adaptive steganography in spatial domain,” in: Proc. 13th Int. Workshop. Inf. Hiding, Prague, Czech Republic, May 2011, pp. 102-117.

[3] J. Fridrich, J. Kodovsk?, “Rich Models for Steganalysis of Digital Images,” IEEE Transaction on Information Forensics and Security, vol. 7, no.23, pp. 868-882, 2012.

9409-26, Session 7

Steganalysis of overlapping images

James M. Whitaker, Andrew D. Ker, Univ. of Oxford (United Kingdom)

No Abstract Available

Conference 9410: Visual Information Processing and Communication VI

Tuesday - Thursday 10–12 February 2015

Part of Proceedings of SPIE Vol. 9410 Visual Information Processing and Communication VI

9410-10, Session PTues

Adaptive motion compensation without blocking artifacts

Timothy B. Terriberry, Mozilla (United States)

The Block Matching Algorithms used in most popular video codec standards introduce blocking artifacts which must be removed via residual coding or deblocking filters. Alternative transform stages that do not cause blocking artifacts, such as lapped transforms or wavelets, require motion compensation methods that do not contain blocking artifacts, since they are expensive to remove. We design a new Overlapped Block Motion Compensation (OBMC) scheme that avoids these artifacts while allowing adaptive blending window sizes and reducing over-smoothing and ghosting artifacts.

This has the potential to show significant visual quality improvements over traditional OBMC.

For more details, see the uploaded PDF.

9410-22, Session PTues

Spatial resampling of IDR frames for low bitrate video coding with HEVC

Brett Hosking, Dimitris Agrafiotis, David R. Bull, Univ. of Bristol (United Kingdom); Nick Easton, BAE Systems (United Kingdom)

While the demand for higher quality/higher resolution video increases, many applications fail to meet such demands due to low bandwidth restrictions. Frequent coding of IDR frames is essential for error resilience in order to prevent the occurrence of error propagation. Each IDR frame, however, consumes a huge portion of the available bitrate resulting in high levels of compression for future coded frames. In this paper we show that spatial resampling of IDR frames can increase the rate distortion performance by providing a higher and more consistent level of video quality at low bitrates.

9410-23, Session PTues

Speed-up keypoint mapping technique by multi-resolution and global information

Wei Qiao, Yong Li, Beijing Univ. of Posts and Telecommunications (China); Hongbin Jin, Beijing University of Posts and Telecommunications (China); ZhiGang Wen, Beijing Univ. of Posts and Telecommunications (China)

No Abstract Available

9410-24, Session PTues

Building reliable keypoint matches by a cascade of classifiers with resurrection mechanism

Jing Jing, Yong Li, Chunxiao Fan, Wei Qiao, Hongbin Jin, Beijing Univ. of Posts and Telecommunications (China)

No Abstract Available

9410-25, Session PTues

Automatic coloring to freehand line drawings in online

Saori Kurata, Fubito Toyama, Hiroshi Mori, Kenji Shoji, Utsunomiya Univ. (Japan)

Freehand line drawings are used effectively for telling our ideas, and easy to edit and redraw. When making a picture, line drawings are one of the most important elements to determine the impression of the picture. Further, by coloring the line drawing image, it becomes more impressive. But, coloring pictures is a time-consuming job and not easy. In this work, we make the time and effort consuming colorization process automatic for freehand line drawings being easy to draw.

On colorization of grayscale images such as old photographs, a method using given example color images was proposed. Their example-based method is based on the principle where similar texture regions in grayscale have same coloring. Their method, however, cannot be used in colorization of line drawings, since almost every region except on the lines has no texture.

As an image filtering approach, an example-based rendering method called “image analogies” is well known. The method synthesizes a filtered image B’ from a source image B by referring to an example pair of source image A and filtered one A’. By selecting an example pair, for example, a photograph and its oil painted version, a user can get an oil painting like image corresponding to an arbitrary source photograph given by the user. According to the usage, the user only selects an example pair and a source photograph.

On assisting in drawing of freehand sketches, a system called “ShadowDraw” was proposed. However its paradigm is not applied yet to the colorization of line drawings.

For a freehand line image drawn on a PC screen where a user-selected reference image, e.g., a color photograph, as a model is foggily displayed with low contrast, we proposed a method for automatic coloring with a constrained Delaunay triangulation that divides the image into small triangles. Using a prototype system based on the proposed method, users can complete an impressive picture by only drawing lines.

Our coloring method begins with the triangulation for the set of sampling points on the drawn lines, followed by sampling of color in each triangle on the reference image, smoothing of color among neighboring triangles, and painting of each triangle with the smoothed color. The result of the triangulation is modified such that it satisfies the constraint where its divided lines should not cross over the drawn lines not to mix colors beyond the drawn line. Our prototype system can display the result of coloring to the current drawings immediately for convenience. So, the user can check the effect for coloring against a newly drawn line at any time. As the result of the coloring depends on how the user draws freehand line drawings, it can be seen as an art work with the individuality of each user’s drawings.

9410-26, Session PTues

Frameless representation and manipulation of image data

Henry G. Dietz, Univ. of Kentucky (United States)

No Abstract Available

Conference 9410: Visual Information Processing and Communication VI

9410-1, Session 1

A new robust method for two-dimensional inverse filtering

Megan Fuller, Jae S. Lim, Massachusetts Institute of Technology (United States)

We propose a new method of inverse filtering a blurry image by subtracting the phase of the point spread function (PSF) from the phase of the degraded image and then using the corrected phase to reconstruct the image. We show that this method is much more robust to noise than simple inverse filtering techniques, and can even compete with certain basic image restoration systems (which rely on accurate knowledge of the statistics of the noise) at low noise levels. For symmetric blurring filters, the phase of the PSF is zero or ϖ. For these filters, the locations of the phase transitions are much more dependent on filter size than taper, we find that knowing only the size of the blurring filter is sufficient for phase-based deconvolution to produce a reasonable estimate of the original image.

9410-2, Session 1

Semi-blind deblurring images captured with an electronic rolling shutter mechanism

Ruiwen Zhen, Robert L. Stevenson, Univ. of Notre Dame (United States)

The electronic rolling shutter mechanism found in many digital cameras may result in spatially-varying blur kernels if camera motion occurs during an imaging exposure. However, existing deblurring algorithms cannot remove the blurs in this case since the blurred image doesn't typically meet the assumptions embedded in these algorithms. This paper attempts to address the problem of modeling and correcting nonuniform image blurs caused by the rolling shutter effect. We introduce a new operator and a mask matrix into the projective motion blur model to describe the blurring process of each row in the image. Based on this modified geometric model, an objective function is formulated and optimized in an alternating scheme. In addition, noisy accelerometer data along x and y directions is incorporated as a regularization term to constrain the solution. The effectiveness of this approach is demonstrated by experimental results on synthesized images.

9410-3, Session 1

Predicting chroma from luma with frequency domain intra prediction

Nathan E. Egge, Mozilla (United States); Jean-Marc Valin, Mozilla (United States) and Xiph.Org Foundation (United States)

This paper describes a technique for performing intra prediction of the chroma planes based on the reconstructed luma plane in the frequency domain. This prediction exploits the fact that while RGB to YUV color conversion has the property that it decorrelates the color planes globally across an image, there is still some correllation locally at the block level. Previous proposals compute a linear model of the spatial relationship between the luma plane (Y) and the two chroma planes (U and V). In codecs that use lapped transforms this is not possible since transform support extends across the block boundaries. We design a frequency domain intra predictor for chroma that exploits the same local correlation with lower complexity than the spatial predictor and which works with lapped transforms.

9410-4, Session 1

Restoration of block-transform compressed images via homotopic regularized sparse reconstruction

Jeffrey Glaister, Shahid Haider, Alexander Wong, David A. Clausi, Univ. of Waterloo (Canada)

No Abstract Available

9410-5, Session 1

Rain detection and removal algorithm using motion-compensated non-local mean filter

Byung Cheol Song, Seung Ji Seo, Inha Univ. (Korea, Republic of)

This paper proposes a novel rain detection and removal algorithm robust against camera motions. It is very difficult to detect and remove rain in video with camera motion. So, most previous works assume that camera is fixed. However, these methods are not useful for application. The proposed algorithm initially detects possible rain streaks by using spatial properties such as luminance and structure of rain streaks. Then, the rain streak candidates are selected based on Gaussian distribution model. Next, a non-rain block matching algorithm is performed between adjacent frames to find similar blocks to each block including rain pixels. If the similar blocks to the block are obtained, the rain region of the block is reconstructed by non-local mean (NLM) filtering using the similar neighbors. Experimental results show that the proposed method outperforms previous works in terms of objective and subjective visual quality.

9410-6, Session 1

Exploiting perceptual redundancy in images

Hongyi Liu, Wuhan University (China) and Wuhan University (China); Zhenzhong Chen, Wuhan Univ. (China)

Exploiting perceptual redundancy plays an important role in image processing. Conventional JND models describe the visibility of the minimally perceptible difference by assuming that the visual acuity is consistent over the whole image. Some earlier work considers the space-variant properties of HVSbased on the non-uniform density of photoreceptor cells. In this paper, we aim to exploit the relationship between the masking effects and the foveation properties of HVS. We design the psychophysical experiments which are conducted to model the foveation properties in response to the masking effects.

The experiment examines the reduction of visual sensitivity in HVS due to the increased retinal eccentricity. Based on these experiments, the developed Foveated JND model measures the perceptible difference of images according to masking effects therefore provides the information to quantify the perceptual redundancy in the images. Subjective evaluations validate the proposed FJND model.

9410-7, Session 2

Video pre-processing with JND-based Gaussian filtering of superpixels

Lei Ding, Ge Li, Ronggang Wang, Peking Univ. (China); Wenmin Wang, Peking University Shenzhen Graduate

School (China)

In this paper, we proposed a new method of video pre-processing based on region-adaptive Gaussian filtering. Firstly, the video frame is segmented into perceptually meaningful atomic regions—super-pixels. Secondly, for each super-pixel, a just-noticeable-distortion (JND) threshold is calculated by weighted averaging of luminance differences around each pixel in the super-pixel. Finally, we set the strength of Gaussian filter for each super-pixel according to its JND threshold. Experimental results show that the bit-rate of video can be reduced up to 29% without loss in visual quality.

9410-8, Session 2

Perceptual vector quantization for video coding

Jean-Marc Valin, Mozilla (United States); Timothy B Terriberry, Mozilla (United States) and Xiph.Org Foundation (United States)

No Abstract Available

9410-9, Session 2

Adaptive residual DPCM for lossless intra coding

Xun Cai, Jae S. Lim, Massachusetts Institute of Technology (United States)

In the Differential Pulse-code Modulation (DPCM) image coding, the intensity of a pixel is predicted as a linear combination of a set of surrounding pixels and the prediction error is encoded. DPCM-based approaches are used in many lossless image and video coding systems.

In this paper, we propose the adaptive residual DPCM (ARDPCM) for intra lossless coding. In the ARDPCM, intra residual samples are predicted using adaptive DPCM weights. The weights are estimated by minimizing the Mean Squared Error (MSE) of coded data and they are synchronized at the encoder and the decoder. The DPCM weights are estimated in a mode-dependent manner.

The proposed method is implemented on the High Efficiency Video Coding (HEVC) 12.0 reference software. Experimental results show that the ARDPCM significantly outperforms the HEVC lossless coding and the HEVC with the DPCM. Specifically, the coding gain of the ARDPCM relative to the HEVC lossless coding reaches as high as 12.9% for 720p sequences. In addition, the proposed adaptive strategy consistently results in significantly better performance compared to the DPCM with reasonable sets of fixed weights. The proposed method is also computationally efficient and HEVC bitstream compliant.

9410-11, Session 2

Arithmetic coding with constrained carry operations

Abo-Talib Mahfoodh, Michigan State Univ. (United States); Amir Said, LG Electronics MobileComm U.S.A., Inc. (United States)

We propose a new technique for constraining the carry propagation that can arise during the arithmetic coding process. Although the probability of long carry propagation is small, practical implementations of arithmetic coding must be designed to work with any set of input data and symbol probabilities, including those that could produce long carry propagation. The proposed solution is based on the fact that the encoder and decoder can keep track of base and interval length, and both can identify the situations when it desired to limit carry propagation by adjusting the interval. Our experimental results show small loss in compression.

9410-12, Session 3

Quality optimization of H.264/AVC video transmission over noisy environments using a sparse regression framework

Katerina Pandremmenou, Nikolaos Tziortziotis, Univ. of Ioannina (Greece); Seethal Paluri, Weiyu Q. Zhang, San Diego State Univ. (United States); Konstantinos Blekas, Lisimachos P. Kondi, Univ. of Ioannina (Greece); Sunil Kumar, San Diego State Univ. (United States)

No Abstract Available

9410-13, Session 3

Game theoretic wireless resource allocation for H.264 MGS video transmission over cognitive radio networks

Alexandros Fragkoulis, Lisimachos P. Kondi, Konstantinos E. Parsopoulos, Univ. of Ioannina (Greece)

No Abstract Available

9410-14, Session 3

Secure content delivery using DASH and open eeb standards

Hari Kalva, Florida Atlantic Univ. (United States); Vishnu Vardhan Chinta, Manipal Univ. (India)

Content protection is essential for many type of content delivery. On-demand video delivery today is mostly based on HTTP streaming. MPEG's Dynamic Adaptive Streaming over HTTP (DASH) has received a lot of interest and many implementations exist. While HTTP streaming does not natively support content protection, current systems rely on proprietary solutions that require browser plugins or proprietary media players. To improve interoperability of HTTP streaming solutions and eliminate the need for plugins, W3C has been working on extensions to the HTML5 standard. Media source extensions (MSE) are being standardized specified to enable playback of content delivered using HTTP streaming and Encrypted Media Extensions (EME) are being specified to enable content protection and digital rights management. These extensions offer a standards based solution to secure content delivery.

This paper presents a system for delivering secure content using web standards. We discuss the use of DASH, MSE, and EME and present an architecture for secure content delivery.

9410-15, Session 3

A method for ultra fast searching within traffic filtering tables in networking hardwire

Sergey V. Makov, Vladimir I. Marchuk, Don State Technical Univ. (Russian Federation); Viacheslav V Voronin, Don state technical university (Russian Federation); Vladimir A. Frantc, Don State Technical Univ. (Russian Federation);

Victor A Obukhovets, Taganrog Institute of Technology Southern Federal University of Russia (Russian Federation)

Visual information streaming requires high data transfer rates between the nodes of the network. That's why design of effective network devices is an actual problem.

The network usually consists of network bridges and switches to provide the transfer of data frames to the specified nodes. The main function of the bridge is to filter local traffic.

The process of filtering is based on a self-learning filtering table in the bridges and switches. Algorithms of filtering tables and the logic of the decision are defined by the standard IEEE 802.1D.

One of the filtering table efficiency criteria is the time needed to search associative information in lookup table. This information allows you to make a decision to filter or to retranslate the frame through the switch.

For switches with store-and-forward architecture search time should be less than time passed from end of destination address of shortest frame to start of the next frame. For 1GBit/s systems this time equals to 560ns.

Another criterion is the amount of required memory is important for the filtering table. This size can be calculated by multiplying the record length by the number of records in the table. Minimal required length of the record in most systems is the length of the searching key plus the length of associated information. The number of records depends on the maximum number of nodes. It should be noted that the filtering table design has influence on the size of the memory required.

The most popular method of table design is hashing tables. There are many ways to reduce the probability of collision in hashing tables. The most popular method to resolve collisions is the block method. This method does not eliminate the probability of collision, but provides a reduction to an acceptable value.

In summary, it should be acknowledge that these three criteria are relevant to compare the efficacy of filtering table design.

To eliminate disadvantages of well known methods we decided not to store the searching key in the tables. Instead we store associated information only: port attachment flag of the frame test block. The frame testing block calculates t hash values applying different laws by using the source and destination addresses, furthermore the port attachment bits are written in the table.

Hashes calculated for destination addresses are used to look up data in the table. Finally the search result is a conjunction of the port attachment bits read from all parallel tables and the collisions are detected if both summary port attachment bits are set to ones.

The decision on the need to retranslate frame is made when: a collision is detected; the frame source port is differ of summary port attachment bit value; when both summary port attachment bits are zeros.

With this modification the amount of memory required decreases down to 64 times given the fact that we do not store the search key.

During research of relation collision probability of filtering table designed by our proposed method we found an optimal number of parallel tables for a fixed total memory size

Our results suggest an optimal number of parallel tables for every fixed total memory size to allocate filtering tables.

Our proposed method of filtering table design provides a significant decrease of memory usage and simultaneously keeps acceptable level of collision probability. At the same time the lookup operation require only one read/write memory operation.

To reduce probability of collision it is possible to use adaptive hashing combine with proposed method. One of the parallel tables should have variable law of the hash calculating. When the collision appears the hash calculating law for one of parallel tables should be changed to another one. This allows to significantly reducing the collision probability without increasing memory usage and optimizing searching time.

9410-16, Session 4

A novel framework for automatic trimap generation using the Gestalt laws of grouping

Ahmad F. Al-Kabbany, Eric Dubois, Univ. of Ottawa (Canada)

No Abstract Available

9410-17, Session 4

Efficient graph-cut tattoo segmentation

Joonsoo Kim, Albert Parra, He Li, Edward J. Delp III, Purdue Univ. (United States)

Law enforcement is interested in exploiting tattoos on as an information source to identify, track and prevent gang-related crimes. Many tattoo image retrieval systems, that retrieve similar tattoo images from a tattoo image database, have been described. In a retrieval system tattoo segmentation is an important step for image retrieval accuracy by removing the background in a tattoo image. Existing segmentation methods do not extract the tattoo very well when the background includes textures and color similar to skin tones. In this paper we describe a tattoo segmentation approach by determining skin pixels around a tattoo. Regions near image edges are the only ones considered for segmentation. In these regions graph-cut segmentation using a skin color model and a visual saliency map is used to find skin pixels. After segmentation we determine which set of skin pixels are connected with each other that form a closed contour including a tattoo. The regions surrounded by the closed contours are considered tattoo regions. Our method segments tattoo well when the background includes textures and color similar to skin.

9410-18, Session 4

Contourlet transform based human object tracking

Manish Khare, Om Prakash, Rajneesh K. Srivastava, Ashish Khare, Univ. of Allahabad (India)

Human object tracking in video sequence is a crucial problem in computer vision applications. Tracking can be defined as a problem of estimating the trajectory of an object in the image plane as it moves around the scene. Object tracking is useful in many applications like target detection and interpretation, surveillance and security, traffic monitoring etc. [1]. Object tracking requires the segmentation [2,3] of object from scene followed by tracking. A good tracking algorithms should have ability to perform operation in real time environment, deal with non-rigid object, deal with noisy data, and deal with presence of occlusion.

In this paper we present our work on single object tracking. Various tracking algorithms are given in Yilmaz et al. [1], in which region based tracking [4] and feature based tracking [5] are very popular. Main shortcoming of region based tracking is that they requires several parameters such as object size, color, shape, velocity etc. Feature based tracking method can avoid these shortcomings by selecting suitable features using some heuristics. A lot of work have been proposed in different literatures, which uses concept of feature based tracking.

In recent past, transformation-based methods are becoming popular for object tracking. By transforming images from one domain to other, some information which might be difficult to obtain in one domain can easily and efficiently obtained in other domain. In transform domain processing, wavelet transform is very promising tool for object tracking purpose. Several methods exists for object tracking using wavelet transform. However, the wavelet transform does not handle curve discontinuity, and discontinuities

across a single curve affect all wavelet coefficients at the curve. For avoiding this problem, the ridgelet transform was introduced. The ridgelet transform provides a good representation for line singularities in 2-D space. Xiao et al. [6] presented an object tracking method based on the ridgelet transform, and proved to be an alternative of wavelet representation of image data. The ridgelet transform alone cannot represent curve discontinuous efficiently. The curvelet transform are capable of handling curve discontinuities efficiently. Nigam and Khare [7] uses properties of curvelet transform and proposed object tracking method based on curvelet transform. Curvelet transform does not work well when object shape is in form of some contour, because curvelet transform work well only at curve discontinuities. Contourlet transform [8] having advantages over wavelet transform and curvelet transform, as it has high directionality and represents salient features of images such as edges, curves and contours in a better way. Motivated by these facts and work of Khare and Tiwary [9], we have proposed a new method for object tracking based on contourlet transform. The proposed method is compared with the state-of-the-art method proposed by Ning et al. [5]. Qualitative performance is not enough to judge the quality of any method. Therefore we have performed quantitative performance of the proposed method in terms of two quantitative performance measures - Euclidean distance and Mahalanobis distance.

The contourlet transform holds various properties [8], in which better edge representation in form of smooth contour, better directionality and translation invariant properties of contourlet transform are more useful in contourlet transform based human object tracking.

A video contains a sequence of consecutive frames. Each frame can be considered as an image. If the algorithm can track moving object between two consecutive frames then it will be able to track object in video sequence. The proposed method consists of two step: (i). Segmentation algorithm, and (ii). Tracking algorithm.

Segmentation algorithm

The main objective of segmentation is to retrieve object of interest in first frame of video sequence. The segmentation is done in contourlet domain. For segmentation of object in first frame of video, we have used our earlier developed method [3]. The segmentation approach, as Khare et al. [3], consists of following steps-

(a) Contourlet decomposition of sequence of frames.

(b) Application of single change detection method on contourlet coefficients.

(c) Application of soft thresholding to remove noise.

(d) Application of canny edge detector to detect strong edges in contourlet domain.

(e) Detection of strong edges after inverse contourlet transform.

We have skipped morphological processing step of Khare et al. [3], because here we do not need accuracy in segmentation. Result of segmentation algorithm is shown in Figure 1, for frame no.75 of Hall monitor video sequence. It is clear that segmentation algorithms works well.

Tracking algorithm

Contourlet transform uses a directional multi-resolution analysis framework to represent images and can deal effectively with piecewise smooth image with smooth contours. Contourlet transform have property of better edge representation in form of smooth contour, and have better directionality. The translation invariant property of contourlet transform allows us to shift attention to the contourlet coefficients domain. Thus magnitude and energy of contourlet coefficients remain approximately invariant by translating the object in different frames of a video. The proposed tracking algorithm exploits these properties.

The proposed tracking algorithm does not require any other parameter except contourlet coefficients. Complete tracking algorithm is as follows-

Step 1:

if frame_no=1

Segment the first frame, by the method described in Khare et al. [3], and re summarized in subsection 3.1 of this paper.

Make the bounding box around the object with centroid at (C1, C2) and compute the energy of contourlet coefficients of the bounding box, say E

where, are the Contourlet coefficients at (i,j)th points.

Step 2:

for frame_no=2 to last_frame do

Compute the contourlet coefficients of the frame, say

search_region=3

if frame_no > 4

Predict the centroid (C1, C2) of the object in current frame with help of centroids of objects of previous four frames and basic equations of straight line motion.

end if

for i = - search_region to + search_region do

for j = - search_region to + search_region do

Cnew_1 = C1+i; Cnew_2 = C2+j;

Make a bounding_box with centroid (Cnew_1, Cnew_2)

Compute the difference of energy of contourlet coefficients of the bounding_box, with E, say di, j

for end

for end

find minimum{di, j} and its index, say (index_x, index_y)

C1 = C1+index_x; C2 = C2+index_y

Make the object in current frame with bounding_box with centroid (C1, C2) and energy of bounding_box, E as

end for

The proposed method for human object tracking has been tested on several video sequences. Here results are being presented for one representative video sequence – child video sequence. The result obtained by the proposed method has been compared with the method proposed by Ning et al. [5]. The child video sequence contains 458 frames of frame size 352 x 288, but we have shown results for frame with a difference of 50 frames. Experimental results for child video sequence are shown.

In Child video, the child object is changing its position and motion abruptly. The abrupt motion of the child is difficult to track. From figure 2, we can see that in the second frame of this video, the bounding box fully covers the object. The movement and direction of movement changes in frame 3 and continuing this the object stops in frame 100. Further we observe that from frame 150 the object reverses it direction of movement and comes at rest in frame 200. The proposed method tracked the object in all these frames accurately. After frame 200, the child object shows some different poses such as slow motion, fast motion, walking pose, bending pose etc. and the proposed method still keeps on tracking the object accurately. On the other hand, from Figure 3, one can see that in the tracking method proposed by Ning et al. [5], the bounding box does not move correctly with child object when it changes its direction of motion, a clear miss track results can be seen in frame 50, 100 and 150. From Figure 2 and 3, it is quite clear that the proposed tracking method performs well in comparison to method proposed by Ning et al. [5].

Figure 3(a) shows plot of Euclidean distance for the proposed method and method proposed by Ning et al. [5]. From Figure 3(a) it is clear that the proposed method has the least Euclidean distance between centroid of tracked bounding box and ground truth centroid in comparison to method proposed by Ning et al. [5]. From Figure 3(b) it is clear that the values of dissimilarity measure are small in case of proposed method, i.e. the ground truth centroids and computed centroids of the proposed method are almost same. From figure 3(b), it is also clear that the Mahalanobis distance remain constant in the proposed method as compared to the method proposed by Ning et al. [5].

In this paper, we have developed and demonstrated a new algorithm for tracking of human object in video based on contourlet transform. Contourlet transform has higher directionality, translation invariant in nature and it captures smooth contours. The proposed method does not depend on many properties of objects such as object size, shape, color etc. the proposed method needs no manual intervention and allows user to easily and quickly track the object in video. Experiments are performed on a number of video

sequences and results for one representative video sequence (child video sequence) have been presented and analyzed. Visual representation of experimental results indicates that the proposed tracking perform well. We have also compared with the proposed method by using well known quantitative performance metrics viz. Euclidean distance and Mahalanobis distance. These quantitative comparisons also prove that the proposed tracking method using contourlet transform is better than other discussed method.

References:

1. A. Yilmaz, O. Javed, M. Shah, "Object Tracking: a survey", ACM Computing Surveys, Vol. 38, No. 4, pp.1-45, 2006.

2. M. Khare, R. K. Srivastava, A. Khare, "Single Change Detection based Moving object Segmentation by using Daubechies Complex Wavelet Transform", IET Image Processing, doi: 10.1049/iet-ipr.2012.0428, 2013.

3. M. Khare, S. Nigam, R. K. Srivastava, A. Khare, "Contourlet Transform based Moving Object Segmentation", in proc. of IEEE International Conference on Information and Communication Technologies (ICT 2013), India, pp. 782-787, 2013.

4. T. L. Liu, H. T. Chen, "Real time tracking using trust region methods", IEEE Transaction on Pattern Analysis and Machine Intelligence, Vol. 26, No. 3, pp. 397-401, 2004.

5. J. Ning, L. Zhang, D. Zhang, C. Wu, "Robust object tracking using joint color texture histogram", International Journal of Pattern Recognition and Artificial Intelligence, Vol. 23, No. 7, pp. 1245-1263, 2009.

6. L. Xiao, H. Z. Wu, Z. H. Wei, Y. Bao, "Research and applications of a new computational model of human vision system based on Ridgelet transform, in proc. of International Conference on Machine Learning and Cybernetics, China, Vol. 8, pp. 5170-5175.

7. S. Nigam, A. Khare, "Curvelet transform based technique for tracking of moving objects", IET Computer Vision, Vol. 6, No. 3, pp. 231-251, 2012.

8. M. N. Do, M. Vetterli, "Contourlets". In: "Beyond Wavelets" Stoeckler, J., Welland, G.V. (eds.), pp. 1–27. Academic Press, New York, 2002.

9. A. Khare, U. S. Tiwary: "Daubechies complex wavelet transform based moving object tracking", IEEE Symposium on Computational Intelligence in Image and Signal Processing, Honolulu, HI, pp. 36-40, 2007.

9410-19, Session 4

Saliency-based artificial object detection for satellite images

Shidong Ke, Xiaoying Ding, Wuhan Univ. (China); Daiqin Yang, Wuhan University (China); Zhenzhong Chen, Wuhan Univ. (China); Yuming Fang, Jiangxi Univ. of Finance and Economics (China)

In this paper, we introduce a computational model of top-down saliency based on multiscale orientation information for artificial object detection for satellite images. Further more, the top-down saliency is integrated with bottom-up saliency to obtain the saliency map in satellite images. We compare our method to several state-of-the-arts saliency detection models and demonstrate the superior performance in artificial object detection for satellite images.

9410-20, Session 4

Quantitative analysis on lossy compression in remote sensing image classification

Yatong Xia, Wuhan Univ. (China); Zimeng Li, Wuhan University (China); Zhenzhong Chen, Daiqin Yang, Wuhan Univ. (China)

In this paper, we propose to use a quantitative approach based on LS-SVM to perform estimation of the impact of lossy compression on remote sensing image compression. Kernel function selection and the model parameters computation are studied for remote sensing image classification when LS-SVM analysis model is establish.

The experiments show that our LS-SVM model achieves a good performance in remote sensing image compression analysis. Classification accuracy variation according to compression ratio scales are summarized based on our experiments.

9410-21, Session 4

Image completion using image skimming

Ahmad F Al-Kabbany, Univ of Ottawa (Canada); Eric Dubois, Univ. of Ottawa (Canada)

No Abstract Available

Conference 9411: Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2015

Tuesday - Wednesday 10–11 February 2015

Part of Proceedings of SPIE Vol. 9411 Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2015

9411-16, Session PTues

Enterprise Mobility Management (EMM) - a way to increase the security of mobile devices

Jenny Knackmuss, Reiner Creutzburg, Fachhochschule Brandenburg (Germany)

Today's organizations and companies need a secure solution that helps them with tools to proactively monitor, control and protect the enterprise from end to end – across various devices, apps, data and the network. While enterprise mobility brings opportunity for users and organization, it can also imply risk.

The aim of this paper is to describe the security improvements in mobile device usage by Enterprise Mobile Management systems. Some case studies, forensic investigation, and ongoing security research problems are explained in detail.

9411-17, Session PTues

Security aspects of mobile medical devices: the case of insulin pumps

Jenny Knackmuss, Fachhochschule Brandenburg (Germany); Wilfried Pommerien, Städtisches Klinikum Brandenburg (Germany); Reiner Creutzburg, Fachhochschule Brandenburg (Germany)

Nowadays, wearable and implantable medical devices are being increasingly deployed to improve diagnosis, monitoring, and therapy for various medical conditions. Compared to other types of electronics and computing systems, security attacks on these medical devices have extreme consequences and must be carefully analyzed and prevented with strongest efforts. Often, the security vulnerabilities of such systems are not well understood or underestimated.

The aim of this paper is to demonstrate security attacks that can easily be done in the laboratory on a popular glucose monitoring and insulin delivery system available on the market, and also propose defenses against such attacks.

9411-18, Session PTues

Semi-automatic generation of multilingual lecture notes: Wikipedia books for algorithms and data structure courses in various languages

Jenny Knackmuss, Reiner Creutzburg, Fachhochschule Brandenburg (Germany)

The aim of this paper is to describe the process of semi-automatic generation of multilingual lecture notes in form of Wikipedia books.

In particular, we study the case of the generation of lecture notes for undergraduate courses on "Algorithms and Data Structures" in different languages: German, English,

The benefit and support of Wikipedia books and multimedia-based teaching in a course on Algorithms and Data Structures was described in an earlier paper.

Furthermore, we explain the advantage of Wikipedia books to support the blended learning process using modern mobile devices in multicultural and multilingual environments.

9411-19, Session PTues

Platform-dependent optimization considerations for mHealth applications

Sahak I. Kaghyan, Institute for Informatics and Automation Problems (Armenia); David Akopian, The Univ. of Texas at San Antonio (United States); Hakob G. Sarukhanyan, Institute for Informatics and Automation Problems (Armenia)

Integrated sensors in modern mobile devices provide multitude of opportunities in such fields and areas of research like mobile health (mHealth), sport and military systems development, etc. All mentioned areas have a direct relation to a human, to interaction with him along with identification and analysis of his physical activity. Thus, it is necessary to highlight the range of problems that come out when we deal with activity monitoring tasks, and particularly the analysis of the state of the mobile device that person cares and respective software application that runs on it.

In this paper the specific emphasis is on inertial sensors, accelerometers and gyroscopes, integrated GPS receivers and WLAN-based positioning capabilities of modern smartphones. Each of these sensing data sources has its own characteristics such as own data providing format, signal retrieving frequency/rate, etc., which impact energy consumption. It is necessary to mention, that energy usage measure and energy consumption measure is an important factor in assessing the efficiency of related mobile applications. Energy consumption significantly varies as sensor data acquisition is followed by data analysis with complex data transformations and signal processing algorithms. Such applications might also perform primitive and complex (i.e., representing a chain/sequence of primitives) physical movement activity recognition and classification. Therefore, the assessment of energy consumption is very important. This paper will address energy consumption assessment of activity recognition applications and related optimization problems.

9411-20, Session PTues

Stroboscopic image in smart phone camera using real time video analysis

Somnath Mukherjee, Kritikal Solutions Pvt. Ltd. (India); Soumyajit Ganguly, International Institute of Information Technology (India)

Motion capturing from a real time video and there by segmentation of the motion of any moving object from a sequence of continuous images or a video is not an exceptional task in computer vision area. Smart-phone camera application is an added integration for the development of such tasks and it also provides for a smooth testing A new approach has been proposed for segmenting out the foreground moving object from the background and then masking the sequential motion with the static background which is commonly known as stroboscopic image. Primarily a stroboscope is an instrument which is used to project any moving object to be slowly moving or stationary or it may be define as an instrument for observing moving bodies by making them visible intermittently and thereby

giving them the optical illusion of being stationary. Stroboscopes are used to measure the speed of rotation or frequency of vibration of a mechanical part or system

Traditional stroboscopic techniques in the photographic as well as filmy world generally have concentrated on simply opening the shutter at the beginning of the action and closing it at the end and recording the moving subject during the process. It becomes quickly evident that this approach has limits in terms of the length of time during which any given can be recorded because if the time is extended too far, too many images will superimpose on each other and it becomes impossible to determine the development or sequence of the action being investigated or visualized. To overcome this problem this approach can be applied instead of camera controlling where the video of the moving object or the cyclically moving in a static background is required to acquire the strobes of the sequential motion. Addition or removing object from a static background images can be estimated only by the help of a statistical model of the background scene where any intruding any object can be detected and estimated by changing of image pixel which are completely different with the statistical model of the scene, this particular functionality is known as background subtraction

The background subtraction technique has been properly estimated here and number of sequential motion have also been calculated with the correlation between the motion of the object and its time of occurrence. The background model scene has a fixed probability density function(pdf) over each image pixel. This pixels have a correlation in between each pdf and it's description to the background model. This background scene describes the static nature when there is no abrupt changes of he image that can be only happened due to intruding new object into the scene. The pixel intensity values and it's corresponding variance shows a significant change during this time and it can be described as simple as Gaussian model and also described as complex nature using Gaussian Mixture Model as well as it's improved model This can be a very effective application that can replace the traditional stroboscopic system using high end SLR cameras, tripod stand, shutter speed control camera control and position etc.

9411-21, Session PTues

Video quality assessment via gradient magnitude similarity deviation of spatial and spatiotemporal slices

Peng Yan, Xuanqin Mou, Xi'an Jiaotong Univ. (China); Wufeng Xue, Xi'an Jiaotong Univ (China)

Now more and more video processing applications need objective video quality assessment (VQA), such as compression, communication, printing, displaying, analysis, restoration and so on, so developing objective video quality measurement techniques that can predict perceived video quality automatically has been an increasing need. Because VQA needs considering the temporal distortion and motion effects, the methods that use the general image quality assessment (IQA) frame by frame directly to predict the video quality cannot get highly correlation with the video subjective scores. In this paper, we proposed a full-reference (FR) VQA methods that is non-distortion specific. The approach relies on a FR IQA algorithms called Gradient Magnitude Similarity Deviation (GMSD), which predicted the image quality from the inconsistency of local quality across the whole spatial locations. GMSD is an excellent and fast IQA algorithm. We use GMSD to analyse not only the spatial distortion but also the spatiotemporal dissimilarity of reference and distorted videos, and then we combine this dissimilarity into a VQA distortion index to estimate the perceived quality of distorted videos.

GMSD is an excellent and fast IQA algorithm, it predicts the image quality from the inconsistency of local quality across the whole spatial locations. In this paper GMSD is used to compute the dissimilarities. Our VQA algorithm is applied on the luminance component of videos only.

Firstly, we extracted the luminance component of videos. Secondly, for the spatial distortion of videos, we applied GMSD to detect the dissimilarities between the luminance components of reference video and distorted video frame by frame, and then average these IQA index to get one quality distortion index, which is called the Spatial GMSD index. Thirdly, in order to detect the spatiotemporal distortion, we first made spatiotemporal slices (STS) images from the luminance components of videos. STS images have two kinds: horizontal STS images (images constructed by the same row of every frame of the luminance components of a video) and vertical STS images (images constructed by the same column of every frame of the luminance components of a video). In our methods, we also used GMSD to detect the STS images dissimilarities. we first applied GMSD to the vertical STS images between the luminance components of the reference and test video to get the vertical spatiotemporal dissimilarities, which is called V-Slice GMSD index, and then applied the same procedure to the horizontal STS images to get the horizontal spatiotemporal dissimilarities, which is called H-Slice GMSD index. At last, we combined the spatial and spatiotemporal dissimilarities of a video into one video distortion index, which is called the Spatial and Spatiotemporal GMSD index (SSTS-GMSD), by multiplying the Spatial-GMSD index, the H-Slice GMSD index and the V-Slice GMSD index.

Testing on the LIVE Video Quality Databases and EPFL-PoliMI Subjective Video Quality Assessment Database demonstrates that our algorithm performs better than the state-of-art FR VQA algorithms.

9411-22, Session PTues

Fast heap transform-based QR-decomposition of real and complex matrices: algorithms and codes

Artyom M. Grigoryan, The Univ. of Texas at San Antonio (United States)

Methods of QR-decomposition (or factorization) of a nonsingular matrix into a unitary (or orthogonal in the case of real matrices) matrix and a triangular matrix are well known in mathematics. QR-decomposition is used in many applications in computing and data analysis. For example, the QR decomposition is an important task for many MIMO signal detection schemes. However, as was mentioned in many works (see for instance the work of Patel, Shabany, and Gulak, IEEE 2009), that the decomposition of complex MIMO channel matrices with large dimensions may leads to high computational complexity, which is important to consider, since for mobile communication applications with fast-varying channels, it is required to perform QR decomposition with low processing latency. QR decomposition is the problem of the solution of the linear system of equations, written in matrix form as Ax=y. The solution x can be found after the factorization of the matrix A=QR, where Q is an orthogonal matrix and R is a right triangular matrix, in the case when the dimensions of the known vector y and unknown x are equal. This QR decomposition is unique if the diagonal coefficients of the matrix R are positive. There are several methods for computing the QR-decomposition, such as the Gramm-Schmidt process and method of Cholesky factorization. We also mention two other methods: the Householder transformations (known also as Householder reflections) and the Givens rotations. In Given rotations, each rotation zeros one element in the subdiagonal of the matrix. Therefore, a sequence of N(N-1)/2 plane rotations are required for reduction of a square matrix (N-by-N to triangular form. Givens rotations require a large number of arithmetical operations, including multiplications and N(N-1)/2 square roots. The method of Householder transforms is the most applied method for QR-decomposition, which reduces the number of square roots to at most 2(N-1) and uses approximately 4N^3/3 multiplications.

In this paper, we describe a new look on the application of Givens rotations to the QR-decomposition problem, which is similar to the method of Householder transformations.

We apply the concept of the discrete heap transform, which has been introduced in digital signal processing to generate the signal-induced unitary transforms that was introduced by Grigoryan (2001). Both cases of real and complex matrices are considered and examples of performing QR-decomposition of matrices are given. We also illustrate the importance of the path of the heap transform in such decomposition. The traditional way of performing the rotations of data in order 1,2,3,... is not the best way, or path,

in computing the QR-decomposition. We describe briefly other more effective paths in the QR-decomposition by the heap transforms and give a comparison with the known method of the Householder transformations. The proposed method of QR decomposition for the complex matrix is novel and differs from the known method of complex Givens rotation (A. Malsev, ..., IEEE 2006) and based on the analytical equations for QR heap transforms. Many examples illustrated the proposed heap transform method of QR decomposition are given, algorithms are described in detail, and MATLAB-based codes are included.

9411-23, Session PTues

Design and development of a prototypical software for semi-automatic generation of test methodologies and security checklists for IT vulnerability assessment in small- and medium-sized enterprises

Thomas Möller, ASSECOR GmbH (Germany); Knut Kröger, Reiner Creutzburg, Fachhochschule Brandenburg (Germany)

The aim of this paper is to show the recent progress in the design and prototypic development of a software for semi-automatic generation of test methodologies and security checklists for IT vulnerability assessment in small and medium-sized enterprises.

In particular we focus on the important case of mobile Android devices in the BYOD environment.

9411-24, Session PTues

Optimal color image restoration: Wiener filter and quaternion Fourier transform

Artyom M. Grigoryan, Sos S. Agaian, The Univ. of Texas at San Antonio (United States)

The linear filtration of a degraded image is the well-known procedure in signal and image restoration. Methods of optimal restoration, including the Wiener filtration are used widely in signal and image processing and number of applications continues to grow. These methods have been developed for denoising medical images in ultrasound imaging, in parameter estimation for linear stochastic systems, in electron microscopy, Gaussian noise filtering from ECG, for acoustic signal processing in a pilot's cabin of aircraft, in reducing the acoustic noise in optimal multichannel linear filtering, in speech enhancement to suppress the interference of car noise and residual, or musical noise, in virtual sound images in headphone equalization, for the enhancement of small transients in gear vibration signals, and for many other applications.

In this paper we analyze the methods of optimal restoration of color images from the degradations that are defined by the additive noise and blurring effects. The Wiener filters are well-known and they are the best linear filters in the considered blurred-image-plus-noise model. The simple approach for color image restoration is described by applying the well-known Wiener filter in the frequency domain to each color plane-component of the color image separately. In other words, if the color image is processed for instance in the RGB color space, the color image is considered as a triplet of separate 2-D gray scale images and each of these images represents red, green, or blue component of the color. We apply a new approach for processing color image, which is similar to our work in image enhancement in the frequency domain, when the color images are considered in the quaternion algebra. Quaternion numbers of Hamilton's was used in Ell's works, and after that time much attention was given to the transformation of the color components to the imaginary subspace of the quaternion numbers, "imaginary part" of which consists of three components. The application of the optimal filtration is based on the concept of the two-dimensional quaternion discrete Fourier transform (2-D QDFT) and results in high quality of filtered images. The use of the 2-D DQFT in optimal filtration adapted to the case of color images is much promised. The color images are considered in the RGB format, and three color components (Red,Green,Blue) are placing into the tree imaginary parts (i,j,k) of the quaternion number space. Since in the 2-D DQFT, any given spatial variation of a color component (R,G,B) is separated into different real-and-imaginary parts of the spectral point, this transform may separate the information of colors in the spectral domain. Our preliminary results show that the application of the 2-D DQFT plus one Wiener filter can be effectively used for restoration of color images. The optimal Wiener filtration is described in detail, together with the simple inverse filtration in quaternion algebra. Examples of application of the proposed method on different color images and comparison with the traditional method when the 2-D DFT based optimal filtration is applied for each color plane separately are given.

9411-25, Session PTues

Fourier transforms with rotations on circles or ellipses in signal and image processing

Artyom M. Grigoryan, The Univ. of Texas at San Antonio (United States)

Many fast unitary transforms such as the Fourier, Hadamard, Haar, and cosine transforms, are widely used in different areas of engineering and science. The theory of the Fourier transform is well developed and the Fourier transform-based methods are effective for solving many problems in signal and image processing, communication, and biomedical imaging. In the discrete case of signals and systems, the transformation of the data from the time domain or image plane to the frequency domain can be extended by introducing a more general concept, than the Fourier transform. The N-point discrete Fourier transform (DFT) represents a beautiful system which rotates the data on the real plane around N circles. The data are referred to as the 1-D signals, or images in the 2-D case. Indeed, the traditional N-point DFT is defined as the decomposition of the signal by N roots of the unit, which are on the unit circle. On different stages of the DFT, the data of the signal or image are multiplied by the exponential factors, or rotated around the circles.

In this paper, a periodic rotation of the data around ellipses is described by the presented concept of the elliptic DFT. For that, the DFT is described in the real space, not complex, and we consider the block-wise representation which is effective and can be generalized to obtain new methods in spectral analysis. We describe the N-point elliptic DFT (EDFT) with basic 2x2 transformations that are not the Givens transformations, but rather rotations around ellipses. The elliptic transformation is therefore defined by different Nth roots of the identity matrix 2x2, the matrix whose groups of motion move a point around the ellipses. The properties of the EDFT are described and examples are provided and compared with the traditional DFT. The EDFT preserves main properties of the DFT, such as shifting and energy preservation. The EDFT distinguishes well from the carrying frequencies of the signal or image in both real and imaginary parts. The EDFT is parameterized and includes the DFT as a partial case when ellipses are circles. Our preliminary results show that by using different parameters, the EDFT can be used effectively for solving many problems in signal and image processing field, which includes problems such as image enhancement, filtration, encryption and many others. Examples of application of the N-block EDFT in signal and image processing are given.

9411-26, Session PTues

Indoor positioning system using WLAN channel estimates as fingerprints for mobile devices

Erick Schmidt, David Akopian, The Univ. of Texas at San Antonio (United States)

Conference 9411: Mobile Devices and Multimedia: Enabling Technologies, Algorithms, and Applications 2015

The Global Positioning System (GPS) has eventually become a common positioning technology. While GPS enabled many applications, satellite signals have yet to overcome many obstacles to enable indoor positioning. Meanwhile, due to the wide deployment of Wireless Local Area Networks (WLAN) in recent years, WLAN positioning algorithms have become popular for mobile device positioning in indoor environments. The most accurate WLAN positioning algorithms exploit the so-called fingerprinting concept which consists of two stages. In the offline stage (training), the Received Signal Strength Indicator (RSSI) from a set of available Access Points (AP) is measured for a number of reference locations and stored in a database often called a radio map. Due to the availability of many APs and the complex structure of indoor environments, this information is distinctive for each reference location and thus is called a position fingerprint. In the online stage (testing), a mobile device receives RSSIs from the APs and their fingerprint is compared to others that are stored in the radio map for the best possible match. WLAN fingerprinting can provide high accuracy in indoor environments when there are many APs which ensure fingerprint uniqueness. This paper provides a novel approach to use WLAN channel estimates as possible fingerprints for positioning algorithms for applications with lower number of access points. The paper provides an initial study.

9411-27, Session PTues

A health messaging system with privacy protection

Lakshmi Aaleswara, Anthony Chronopoulos, The Univ. of Texas at San Antonio (United States)

In this paper, we propose a new software system that employs features that help the organization to comply with USA HIPAA regulations. The system uses SMS as the primary way of communication to transfer information. Lack of knowledge about some diseases is still a major reason for some harmful diseases spreading. The developed system includes different features that may help to communicate amongst low income people who don't even have access to the internet. Since the software system deals with Personal Health Information (PHI) it is equipped with an access control authentication system mechanism to protect privacy. The system is analyzed for performance to identify how much overhead the privacy rules impose.

9411-28, Session PTues

Presentation of a web service for video identification based on Videntifier techniques

Silas Luttenberger, Reiner Creutzburg, Fachhochschule Brandenburg (Germany); Björn Jónsson, Reykjavik Univ. (Iceland)

This paper describes a web service for video identification. The main aspect of the study is to analyze the requirements and implementation for such a web service in combination with the Videntifier techniques. Possible fields of use are discussed as well as difficulties that might be important to consider for a successful implementation.

9411-29, Session PTues

An efficient contents-adaptive backlight control method for mobile devices

Qiao Song Chen, Ya Xing Yan, Xiao Mou Zhang, Hua Cai, Xin Deng, Jin Wang, Chongqing Univ. of Posts and Telecommunications (China)

For most of mobile devices with a large screen, image quality and power consumption are both of the major factors affecting the consumers' preference. Contents-adaptive backlight control (CABC) method can be utilized to adjust the backlight and promote the performance of mobile devices. Unlike the previous works mostly focusing on the reduction of power consumption, both of image quality and power consumption are taken into account in proposed method. Firstly, region of interest (ROI) are detected to divide image into two parts: ROI and non-ROI. Then, three attributes including entropy, luminance, and saturation information in ROI are calculated. To achieve high perceived image quality in mobile devices, optimal value of backlight can be calculated by a linear combination of the aforementioned attributes. Coefficients of the linear combination are determined by applying the linear regression to the subjective scores of human visual experiments and objective values of the attributes. Based on the optimal value of backlight, displayed image data are processed brightly and backlight is darkened to reduce the power consumption of backlight later. Here, the ratios of increasing image data and decreasing backlight functionally depend on the luminance information of displayed image. Also, the proposed method is hardware implemented. Experimental results indicate that the proposed technique exhibits better performance compared to the conventional methods.

9411-30, Session PTues

Local adaptive tone mapping for video enhancement

Vladimir Lachine, Qualcomm Inc. (Canada); Min Dai, Qualcomm Inc. (United States)

As new technologies like High Dynamic Range cameras, AMOLED and high resolution/big screen displays emerge on consumer electronics market, it becomes very important to deliver the best picture quality for mobile devices. Tone Mapping (TM) is one popular technique to enhance visual quality. However, the traditional implementation of Tone Mapping procedure is limited by pixel's value to value mapping, and the performance is restricted in terms of local sharpness and colorfulness. To overcome the drawbacks of traditional TM, we propose a spatial-frequency based framework in this paper.

In the proposed solution, intensity component of an input video/image signal is split on low pass filtered (LPF) and high pass filtered (HPF) bands. Tone Mapping (TM) function is applied to LPF band to improve the global contrast/brightness and HPF band is added back afterwards to keep the local contrast. The HPF band may be adjusted by a coring function to avoid noise boosting and signal overshooting. Furthermore, the HPF band may be adjusted by another function that may be derived from the Tone Mapping function to restore or enhance local sharpness. In order to preserve colorfulness of an original image, chroma components are corrected by means of saturation function that may be derived from the Tone Mapping function as well. Localized content adaptation is further improved by dividing an image to a set of non-overlapped regions and modifying each region individually.

The suggested framework allows users to implement a wide range of tone mapping applications with perceptional local sharpness and colorfulness preserved or enhanced. Corresponding hardware circuit may be integrated in camera, video or display pipeline with minimal hardware budget.

9411-1, Session 1

Practical Usefulness of Structure from Motion (SfM) Point Clouds Obtained from Different Consumer Cameras

Patrick Ingwer, Fabian Gassen, Stefan Püst, Melanie Duhn, Marten Schälicke, Katja Müller, Heiko Ruhm, Josephin Rettig, Eberhard Hasche, Arno Fischer, Reiner Creutzburg, Fachhochschule Brandenburg (Germany)

This paper deals with the usefulness and accuracy of point clouds obtained by different consumer cameras out of structure from motion (SFM) algorithms.

Key technologies are moving a lot in areas like medicine, structural engineering, car development and other.

There is yet another field where modern technologies are significantly affecting workflows and results - the entertainment sector. In this field particular technologies are essential, without those no profitable results would be possible. For instance camera tracking, chroma keying, image based lighting with its photorealistic lighting and rendering, motion capturing, fur / smoke / fire simulation, and so on.

One very new technology is helping the new content creation. This is modelling with the help of point clouds and also point cloud supported camera tracking. One way to create a usable point cloud is the utilization of Lidar scanning equipment like those by the Faro company. Another way is the calculation of a point cloud with the structure from motion (SFM) method out of a series of pictures from small objects or large areas. Lidar scanning is highly accurate but has a big problem, the scanner always needs a solid ground and sometimes a stable positioning of the scanner is impossible. Here, the SFM method is a good solution, maybe in conjunction with cranes, air drones, etc. with high mobility. One question is now - how accurate and precise are point clouds out of those SFM calculations to make use in high quality entertainment productions or other applications like structural engineering. Then the most important question is, which kind of consumer cameras are most reasonable to get best and useful results when using cranes or air drones.

This paper deals with the research of the practical usage of SFM method in applications where highly accurate point clouds are needed.

9411-2, Session 1

Sensor data formats supporting energy consumption assessments for smartphone-based M-health applications

Rodrigo Escobar, David Akopian, Rajendra Boppana, The Univ. of Texas at San Antonio (United States)

Advances in state-of-the-art sensors and communications technology make feasible longitudinal health monitoring for accurate diagnosis of various health conditions, smart condition-adaptive data collection and rapid responses to prevent emergency conditions.

Continuous health monitoring typically involves three stages: 1) health signals are sensed using wireless body and other sensors, 2) sensor readings are communicated to local gateways (e.g. a smartphone) for preliminary processing, data aggregation, temporary storage or transmission to remote servers, and 3) the data are is further processed by remote medical servers and archived in long-term storage devices for longitudinal tracking of health conditions by health-providers' servers. Data collection using wireless body sensor networks (WBSN) has been extensively investigated. In addition, electronic data record formats to specify the popular health condition formats in a manner suitable for use by final consumers of health data have also been standardized. However, the intermediate chain of longitudinal data collection management that requires managing battery-powered sensors and other devices is generally overlooked.

Most sensors and gateway (smartphone) devices are energy-constrained and any realistic scenario of longitudinal observation should take into account major limitations on data collection durations and should be able to manage the process based on available battery resources.

In this paper we propose a data format that facilitates multiple sensor data aggregation, communication and control in energy-constrained environments. The data format leverages three important aspects: 1) efficient sensor data aggregation and communication between gateways and processing servers; 2) incorporation of energy consumption rates and statuses of participant energy-constrained devices in communication protocols so that the system management can assess energy consumptions and plan realistic monitoring scenarios; and 3) assessment of energy consumption with various rate estimates such as vendor provided energy consumption specifications or more accurate calibrated data.

The data format proposed in this paper is suitable for collection of multiple sensor data and defines signals' parameters such as sampling rates and signal durations. It outperforms other known data formats for communication of health-related data in terms of readability, flexibility, interoperability and validation of compliant documents, and enables energy assessment capability for realistic data collection scenarios.

9411-3, Session 2

User-aware video streaming

Louis Kerofsky, Yuriy A. Reznik, Abhijith Jagannath, InterDigital, Inc. (United States)

Bandwidth demand due to video streaming is growing at a high rate. In mobile applications, increased network capability is being explored by 5G wireless development while improved compression design of MPEG-HEVC offers a 50% reduction in bitrate. Despite these improvements in network infrastructure and compression efficiency, further reduction in video bandwidth is desired. Traditional video delivery systems assume a conservative viewing setup and control the resolution and quality of video based on this conservative assumption. Modern smartphones and tablets offer an array of sensors capable of providing information about the video playback environment. Similarly, mobile devices are used in a variety of usage modes creating different viewing distances and ambient light levels. This combination of a variety of viewing conditions and the ability to dynamically sense them enables a complementary solution to address the growth in video bandwidth needs.

We propose adaptation to individual users' viewing conditions to reduce the bitrate needed for delivery of video content. Resolution which cannot be seen under the present viewing conditions is not transmitted. As a result, user experience is not impacted while a reduction in bandwidth is realized. A visual model is used to determine sufficient resolution needed under various viewing conditions. The visual model uses characteristics of the display device as well as properties of the viewing conditions, i.e. viewing distance. Sensors on a mobile device dynamically estimate properties of the viewing conditions, particularly the distance to the viewer. We propose a method to estimate viewing distance without a camera by relying on other sensors typical of a modern smartphone or tablet, such as the accelerometer and/or gyroscope. To implement this user adaptive video streaming, we leverage the framework of existing Adaptive Bit-Rate (ABR) streaming systems, i.e. HLS, Smooth Streaming or MPEG-DASH. Traditional ABR clients dynamically select different representations of video from a list provided by the server using estimates of the network conditions and client playback buffer. These representations differ in both resolution and/or bitrate. Our proposed method provides a hint to the stream selection logic in the form of the sufficient resolution based on the visual model and estimated viewing conditions. The stream selection logic at the client is modified slightly to remove representations from consideration when a lower resolution stream meets or exceeds the current sufficient resolution. Conceptually, representations are removed from consideration when a lower resolution representation has sufficient resolution for the current viewing conditions. It is possible to improve the system performance by including custom representations produced for various viewing distances but having identical resolution expressed in pixel dimensions.

Achievable bitrate savings relative to non-user aware video clients depend upon the display characteristics (resolution and physical size), the viewing distance, and the set of candidate ABR representations. We simulate the bitrate savings using samples of these parameters. Example smartphones and tablets were considered. Typical viewing distance distributions for tablets and smartphones were taken from the literature. Sample ABR representations were used to produce these simulation results.

9411-4, Session 2

Mobile-based text recognition from water quality devices

Shanti Dhakal, Maryam Rahnemoonfar, Texas A&M Univ. Corpus Christi (United States)

Measuring water quality of bays, estuaries, and gulfs is a complicated and time consuming process. YSI Sonde is an instrument used in research fields to measure water quality parameters such as pH, temperature, salinity, and dissolved oxygen, etc. Usually, this instrument is taken to different parts of a water body in a boat trip and researchers frequently note down different parameters displayed by the instrument's display monitor. In this project, a novel algorithm and a mobile application is developed for Android platform using Java Native interface to call C++ and OpenCv functions from Java program. The application allows a user to take a picture of the YSI Sonde monitor, extract text from the captured image and store it on a file. The file can be retrieved later for analysis. The main idea behind the application is to help the user get information quickly from the image and reduce human error while typing in the information to a sheet of paper. Also, since the extracted information is stored on a file, there is no fear of loosing the information, as opposed to writing down the information on a piece of paper. The image captured by the application is first processed to identify the rectangular region of interest from the image of the YSI Sonde display monitor. For this purpose, first canny edge detection algorithm in combination with mathematical morphology operation (dilation) is used. Next, a perspective transformation matrix is obtained from the corner points of the source image and the corner points of the destination image. To obtain corner points, first, probabilistic Hough line transform is used to identify lines in the image. The corner of the image is then obtained by determining the intersection of the detected horizontal and vertical lines in the image. Finally, the image is warped using the matrix, hence, removing the perspective distortion from the image. In the next step, piecewise linear transform is performed on the image to adjust the contrast of the image. Mathematical morphology operation black-hat with a rectangular structuring element is used to correct the shading of the image. The black-hat operation produces the image with the objects that are darker than the surrounding objects. Since the characters in the image are darker than the background, this operation is very desirable. Finally, the Otsu's binarization technique is used to binarize the enhanced image. The binarized image is then passed to the Optical Character Recognition (OCR) software. Tess-two library, a part of Tessaract engine is used for optical character recognition. Java Regular Expression (Java Regex) is used to recognize the digits and some special characters from the extracted text. It is also used to eliminate unwanted and insignificant characters and symbols. The extracted information is stored in a file on the memory card of the phone so that it can be downloaded to a computer. The algorithm was tested on 60 different images of YSI Sonde with different perspective features and shading. Experimental results, in comparison to ground-truth results, demonstrate the effectiveness of the proposed method.

9411-5, Session 2

Depth enhanced and content aware video stabilization

Albrecht J. Lindner, Kalin Atanassov, Sergio R. Goma, Qualcomm Inc. (United States)

Video stabilization is an important topic due to the omnipresence of mobile phones that enable amateur users to spontaneously capture videos without stabilization equipment such as a Steadicam. The small form factor of phones and the competitive pricing make software solutions an interesting option.

Video stabilization algorithms usually have three stages. First, the movement of the camera with respect to the scene is estimated with a temporal tracking of the scene content. Then the trajectory is processed in order to reduce jitter, but retain desired motions such as smooth panning and rotation. Finally the video sequence is re-rendered for the new smoothed trajectory. Each of these stages has challenges to overcome, but in our work we focus on the first stage. We do this by exploiting depth information which is becoming more and more available on mobile phones equipped with stereo cameras.

In order to track camera movement, a mapping function has to be learned that projects image points from one frame to another. In an ideal case it would be desirable to do a full 3D reconstruction of the scene and camera position, but this is computationally expensive and prone to errors leading to visual artefacts. Therefore, 2D alternatives such as homographies or similarity transforms are often used, which are easier to compute and more robust [1]. This advantage comes at the price that homographies are only valid for points that lie on a planar surface and similarity transforms are even more restrictive. Thus, scenes with objects at different depths are not well modeled.

We propose to use depth information in order to compute a homography only for points that have the same depth. With this technique the user can for example choose to stabilize a person in the foreground independent of the potentially moving background. Alternatively the user can stabilize on the background independent of objects in the foreground. This has two main advantages: the homography estimation is more robust (points with different depth that cannot be modeled are rejected) and the user can choose which part of the scene has to be stabilized.

It is not necessary that the user has to explicitly state which part of the scene has to be stabilized. It is for example reasonable to assume that those areas where the user taps to focus are also important to be stabilized. Alternatively, the algorithm can default on the closest object, which can be determined from the depth histogram.

The most related work to ours is from Liu et al. who also propose to use added depth information for an augmented scene understanding. The difference is that they aim at recovering the full 3D information of the camera trajectory which is computationally more expensive and less robust than a simpler 2D alternative.

9411-6, Session 2

Mobile micro-colorimeter and micro-spectrometer modules as enablers for the replacement of subjective quality checks of optically clear colored liquids by objective quality assurance with smartpads in-field

Dietrich Hofmann, Technology and Innovation Park Jena (Germany); Paul-Gerald Dittrich, Technology and Innovation Park Jena (Germany); Fred Grunert, MAZeT GmbH (Germany); Jörg Ehehalt, Mathias Reichl, RGB Lasersysteme GmbH (Germany)

In chemical, pharmaceutical and cosmetic industries the quality standards of liquids such as solvents, oils, fatty acids and fuels are growing. An important feature for quality of optically clear colored liquids is their color. Therefore objective mobile color quality assurance systems for liquids becoming increasingly important. This actual situation is influenced by growing quality claims for liquids as well as innovative possibilities for in-field measurements via smartpads in combination with micro-colorimeter and micro-spectrometer modules. Enablers to overcome subjective comparison methods of liquids with colored liquid or glass standards are objective tri-stimulus, quasi-spectral and spectral measuring methods. The paper focuses on the investigation of micro photonic rgb, true-color, multi-spectral and hyper-spectral sensor modules (hardware apps) in their combination with smartpads and specialized imaging software (software apps) for convenient, reliable and affordable quality assurance systems.

Aim of the paper is to describe the performance characteristics of the above mentioned sensor modules for the objective characterization of optically clear colored liquids under different working conditions. Typical influences

are different illuminations, optical dimensions of the sample cells and different liquid or glass calibration standards. Also the calculation of the XYZ tri-stimulus values from the different sensor types and their conversion into the Lab color space is influential.

In the paper is shown that and how:

- LED lighting is applicable as mobile illumination module,

- Adjustable LED lighting intensities and high sensitive sensors enable the application of unified sample cells,

- Micro-colorimeter modules with true-color-sensors have the same color accuracy as multi-spectral and hyper-spectral sensor modules in terms of colorimetric characterizations of optically clear colored liquids,

- Micro colorimeter modules with true-color and multi-spectral sensors can be calibrated by any kind of linear independent liquid color solutions,

- The Lab color space is the most suitable color space for liquid color measurements,

- Highly accurate colorimetric characterization of liquids according to DIN EN 1557 can be seamless applied to micro-colorimeter modules with tri-stimulus sensors,

- Smartpads are convenient, reliable and affordable mobile quality assurance systems in combination with micro-colorimeter and micro-spectrometer modules and specialized image processing software,

- Imaging software apps can be unified for the numerical and graphical representation of color spectra and color coordinates in the color spaces

In the paper is shown that and how innovative convenient, reliable and affordable measuring instruments with micro sensor modules, smartpads and software apps enable objective colorimetric characterizations of optically clear colored liquids at the point of interest.

9411-7, Session 2

Concept for practical exercises for studying autonomous flying robots in a university environment: part II

Nils Gageik, Erik Dilger, Sergio Montenegro, Julius-Maximilians-Univ. Würzburg (Germany); Stefan Schön, Fachhochschule Brandenburg (Germany); Rico Wildenhein, Brandenburg University of Applied Sciences, Department of Informatics and Media (Germany); Reiner Creutzburg, Arno Fischer, Fachhochschule Brandenburg (Germany)

The present paper demonstrates the application of quadrotors as educational material for students in aerospace computer science, as it is already in usage today.

The work with quadrotors teaches students theoretical and practical knowledge in the fields of robotics, control theory, aerospace and electrical engineering as well as embedded programming and computer science.

For this the material, concept, realisation and future view of such an course is discussed in this paper.

Besides that, the paper gives an brief overview of student research projects following the course, which are related to the research and development of fully autonomous quadrotors.

9411-8, Session 3

Smartphone-based secure authenticated session sharing in internet of personal things *(Invited Paper)*

Ram Krishnan, Jiwan Ninglekhu, The Univ. of Texas at San Antonio (United States)

Human ability to be more productive has always been limited by their limited ability to remember. In the context of password-based authentication, only a few usernames and passwords can be memorized which is further limited by the character length of passwords. With the rise of the Internet, social media and web services, it has almost become impossible to handle the big pool of personal usernames and passwords. Using simple, same or similar passwords can easily be stolen or hacked by password cracking tools and social engineering attacks. Therefore, a robust and painless technique to manage sign-in information for websites is necessary. In this paper, a novel technique of user authentication-credentials management via a smart mobile device such as a Smartphone, in a wireless environment is proposed. We present a secure user credential management scheme in which user's username and password linked with website's uniform resource locator (URL) or domain name is saved into the mobile device database via an interactive interface app. This user information is to be shared with the web browser on user's computer, required for specific website. We develop a customized browser extension tool and use it to import user's credentials: username, password and domain name information, linked with the required website from the mobile device either through Wi-Fi or Bluetooth network connection. The extension identifies the target objects in the webpage and auto-fills the form with the required authentication credentials ready for the user to execute. An Android platform application is developed as an interface to user input to store the required user information into the mobile device, and a Google Chrome extension is developed to import the user information from the mobile device into the browser. We integrate this scheme with encrypted network communication using standard encryption technique.

9411-9, Session 3

Door and window image-based measurement using a mobile device

Gady Agam, Guangyao Ma, Manishankar Janakaraj, Illinois Institute of Technology (United States)

We present a system for door and window image-based measurement using an Android mobile device. In this system a user takes an image of a door or window that needs to be measured and using interaction measures specific dimensions of the object. The existing object is removed from the image and a 3D model of a replacement is rendered onto the image. The visualization provides a 3D model with which the user can interact. When tested on a mobile Android platform with an 8MP camera we obtain an average measurement error of roughly 0.5%. This error rate is stable across a range of view angles, distances from the object, and image resolutions. The main advantages of our mobile device application for image measurement include measuring objects for which physical access is not readily available, documenting in a precise manner the locations in the scene where the measurements were taken, and visualizing a new object with custom selections inside the original view.

Our measurement process starts by taking an image of the object in which a calibration target is present. The purpose of the calibration target is to facilitate absolute length measurements. The user then clicks on the four corners of the object. Accurate selection of corners is crucial for correct measurements. Hence, we developed a process by which a user interface provides magnification of the area of interest and a dynamic motion rate facilitates accurate selection close to the final selection location. Once selected, the points are refined using a subpixel corner detection algorithm.

Given the selected corners of the object we compute a 2D projective map to rectify the image so that perspective distortion is removed from the object. In this rectification it is not possible to have a precise scale since the calibration target has not yet been detected. Given the rectified image we automatically detect the calibration target, extract corners from it, and map the initial user corner selections onto the rectified image. We then compute a second 2D projective map based on the corners of the calibration target for which we have absolute measurements. After this second rectification transformation we have correct scale and can measure the object.

Given the location of the object we remove the image part corresponding to it and use the resulting image as a texture map for a 3D wall model with

an opening that has the same aspect ratio as the object. We support two rendering modes. In the first the image is rectified so that the object is rectangular and so the 3D object can be inserted directly into the opening. In the second we distort the object and insert the object into the original non-rectangular opening.

The full paper contains a detailed description of the algorithms we used for measurement and rendering, and provides accuracy and time performance results as a function of various image parameters such as image size, distance from the object, and view angle.

9411-10, Session 3

Communication target object recognition for D2D connection with feature size limit

Jiheon Ok, Yonsei Univ. (Korea, Republic of); Soochang Kim, Young-Hoon Kim, Electronics and Telecommunications Research Institute (Korea, Republic of); Chulhee Lee, Yonsei Univ. (Korea, Republic of)

Recently, a new concept of device-to-device (D2D) communication, which is called "point-and-link communication" has attracted great attentions due to its intuitive and simple operation. With this technology, the user selects a target image, which is taken by the device camera and shown in the user's display. Hence, the user can connect to the target device intuitively without inputting any pre-identification information such as SSIDs and MAC addresses.

In this paper, we propose an efficient object matching algorithm that can be applied to look-and-link communications for mobile services. We consider a scenario that the devices in the vicinity of the user broadcast the descriptors under 300 bytes. In the scenario, the user selects a target object from the display to establish a connection between the user's terminal and the target object. The user selects the target object by a simple cropping operation. From the selected object image, features are extracted to generate the descriptor, which is then compared with received descriptors sent by nearby devices. Finally, the connection between the user's device and the target object is established.

It is desirable to reduce the descriptor size for the limited memory and low computational power of mobile terminals. Therefore, a fast and robust feature extraction method is required. The proposed descriptor is constructed by combining HSV color histograms, selected SIFT features and object aspect ratios. Since key point descriptors are high dimensional and a few thousand key points are usually extracted from an image, the SIFT feature is not appropriate for the look-and-link communication due to the large data size. In this paper, two techniques are used to reduce the SIFT feature size. First, the number of key points is cut down to reduce redundancy among key points using by calculating degree of fidelity at each key points using bilateral filtering. Secondly, the standard SIFT descriptor is converted to binary SIFT descriptors by quantizing feature vectors. The color feature is one of the most intuitive features to distinguish objects. In order to generate the HSV histogram, the HSV space is quantized into several bins. Considering the human vision model, the HSV space is quantized into 36 non-uniform bins. Finally, the binary SIFT descriptor, the HSV color histogram and the object size information (width and height) are combined to generate an object descriptor. The target descriptor extracted from a target object image selected by the user is compared with all the descriptors broadcast from nearby devices. Then the device with the highest matching score is selected and a communication link is established. The proposed algorithm achieves the accuracy of 87.8% in real time (about 21ms per one comparison) on 368 object images including signboard and device. The proposed algorithm can be used in combination with beamforming to further improve performance.

9411-11, Session 3

Photogrammetric 3D reconstruction using mobile devices

Dieter Fritsch, Miguel Syll, Univ. Stuttgart (Germany)

Photogrammetric 3D reconstruction has undergone a renaissance with the invention of Semi-Global Matching by Hirschmueller (2005). Using homologues parts of imagery of all kinds a pixel-by-pixel matching can be performed delivering very dense point clouds of excellent geometric quality. At the Institute for Photogrammetry the software SURE (Surface Reconstruction from Imagery) has been developed and is widely used in academia and industry (Rothermel et al, 2012).

Today's mobile devices are equipped with a variety of onboard sensors (GPS, camera, accelerometer, magnetic compass, air pressure, temperature, proximity, humidity, light, etc.) and can be linked with external sensors using WIFI, bluetooth or other short communication transfers. In our paper we demonstrate the development of an Android Application for Photogrammetric 3D Reconstruction that works on smartphones and tablets likewise. The photos are taken with mobile devices, and can thereafter directly be calibrated using standard calibration algorithms of photogrammetry and computer vision, on that device.

Three devices where evaluated with the objective to measure the time consumption for each device performing detection and matching tasks over 4 images of 1536 x 2048 pixels (6 combinations) . To do this it was developed a short program using OpenCV to detect SIFT features and compute its matchings. At the moment of this evaluation the detectors SIFT and SURF were not available in the Android OpenCV version, however, were implemented patching OpenCV. Because of lack of memory of mobile devices we concluded that two scenarios have to be investigated:

• Stand-alone applications that will run on the mobile device and performs all the processes to generate 3D point clouds

• The data collection system is linked to a Server application. The mobile device collects and calibrates the photos, and finds features. This information is send to a server for processing and sending the colored 3D point cloud back to the device for visualization. This option is programmed using Asynchronous Remote Procedure Calls (Asynchronous RPC) (Ananda, 1992)

At present, a Client-Server processing pipeline has been established, using Dropbox folders for the handshake. The following benefits were realized:

• Processing on a Server better hardware configurations can be used, for example: more processors, higher RAM, GPU and CUDA. Then the application would have (indirectly) the performance of a desktop application.

• Results can be delivered faster and also it is possible to process higher amount of pictures and higher resolutions. Of course, the speed will also depend on hardware capabilities and the software used on the Server.

• The reconstruction can be implemented using other software tools like VisualSFM or Bundler, in addition with dense reconstruction software like PMVS or SURE. Then the information would be processed using the best current algorithms.

• Time of development will be shorter and at the first version the results would be to the height of existing PC applications. When the PC reconstruction applications are upgraded then the functionality will also be beneficial.

9411-12, Session 4

Toward energy-aware balancing of mobile graphics

Efstathios Stavrakis, The Cyprus Institute (Cyprus); Marios Polychronis, Univ. of Cyprus (Cyprus); Nectarios Pelekanos, A.R.M.E.S. Ltd. (Cyprus); Alessandro Artusi, The Cyprus Institute (Cyprus) and Univ. de Girona (Spain);

Panayiotis Hadjichristodoulou, A.R.M.E.S. Ltd. (Cyprus);
Yiorgos Chrysanthou, Univ. of Cyprus (Cyprus)

In the area of computer graphics the design of hardware and software has primarily been driven by the need to achieve maximum performance. Energy efficiency was usually neglected, especially when designing algorithms and software applications, assuming that a stable, always-on power source was available. This is supported by the vast scientific literature in computer graphics that usually reports the performance of algorithms, as well as the respective image quality achieved.

However, the advent of the mobile era has brought into question these ideas and designs in computer graphics since mobile devices are both limited by their computational capabilities and their energy sources. The scientific community is now increasingly calling for the inclusion of energy efficiency analysis, along with performance and image quality [1].

Aligned to this emerging need in computer graphics for energy efficiency analysis we have setup a software framework to obtain power measurements using off-the-shelf hardware that allows for sampling the energy consumption over the power rails of the CPU and GPU at 10Hz. Measurements are obtained from a strategically designed 3D scene with varying levels of geometric complexity, texture resolution and partitioning of the CPU/GPU workload.

We analyzed the power consumption behavior of mobile devices when performing typical computer graphics operations in games, such as loading textures and dynamically modifying the geometric complexity of the scene using Levels of Detail (LOD). In addition, we have measured the energy consumption of the CPU in contrast to that of the GPU when applying the same image filter on the two processors. Aligned with previous work [2,3], some of these initial measurements provide evidence that 20-30% energy can be preserved using these techniques independently.

The goal of this work, which is currently in progress, is to combine the knowledge obtained from these initial measurements into a prototype energy-aware balancer of processing resources. The balancer dynamically selects the rendering parameters (e.g. LOD, texture resolution, etc.) and the most suitable processing unit (i.e. CPU or GPU) for selected rendering tasks, while trading off image quality. Dynamic selection of parameters is driven by simple heuristics, e.g. viewing distance or user-defined application settings; however more sophisticated automatic parameter-selection approaches, such as visual attention [4], may be used with the balancer.

[1] T. Akenine-Möller and B. Johnsson, “Performance per What?,” Journal of Computer Graphics Techniques (JCGT), Vol. 1, No. 1, pp. 37–41, Oct. 2012.

[2] J. M. Vatjus-Anttila, T. Koskela, and S. Hickey, “Power Consumption Model of a Mobile GPU Based on Rendering Complexity,” Sept. 2013, pp. 210–215, IEEE.

[3] M. Hosseini, A. Fedorova, J. Peters, and S. Shirmohammadi, “Energy-aware Adaptations in Mobile 3D Graphics,” in Proceedings of the 20th ACM International Conference on Multimedia, 2012, MM ’12, p. 1017–1020, ACM.

[4] C-LOD: Context-aware Material Level-Of-Detail applied to Mobile Graphics, G. A. Koulieris, G. Drettakis, D. Cunningham, K. Mania, Computer Graphics Forum, Vol. 33, (4) – 2014.

9411-13, Session 4

Optimized large-capacity content addressable memory (CAM) for mobile devices

Khader Mohammad, Birzeit Univ. (Palestinian Territory, Occupied)

A content addressable memory system includes CAM cells, each having a compare circuit and a memory bit cell that stores complementary bits. The main CAM design challenge is to reduce power consumption associated with large amount of parallel switching circuitry, without sacrificing speed or density. In this paper, we present a new technique to eliminate crowbar current during bit-cell write operation (saving 0.0114mA per cell in 22nm process), reduce average current consumption during cam operation and eliminate the need for routing the complementary data to every cam cell, saving routing track in smaller node technology where wire cap is dominant.

9411-14, Session 4

Fast retinex for color image enhancement: methods and codes

Artyom M. Grigoryan, Analysa M. Gonzales, The Univ. of Texas at San Antonio (United States)

In this paper, we consider the class of complex retinex methods of color image enhancement, which use filtering with different Gaussian kernels and additional post-processing stages for adjusting colors. The retinex as model of lightness and color perception of the human vision was proposed by Land (1977) and different implementations exist of the retinex algorithms which were first developed by Land and Mcann (2000). We mention the multi-scale retinex (MSR) and multi-scale retinex with color restoration which are effective methods for color image enhancement that proved color constantly and dynamic range compression. Different methods of retinex have their various advantages and limitation, some of them are far more useful than others, and they vary mostly in terms of how the illumination in a color image is estimated. It is important to note that there is no clear understanding of the best way of selecting many parameters of the algorithms, but follow the suggestions of the developers. Finally, the direct implementation of the multi-scale algorithm is a very slow procedure, for instance in MATLAB, when comparing with other powerful transform-based methods of image enhancement, for instance the method of alpha-rooting in quaternion algebra developed Grigoryan and Agaian (2014) and fast method of heap transforms (Grigoryan 2012).

We propose the implementation of the multi-scale retinex in the frequency domain, since one of the main part of the retinex enhancement is in calculating a few two-dimensional convolutions by the Gaussian functions when processing each color component of the image. The implementation of this part can be accomplished in a very fast way if we use the fast Fourier transform. The RGB color space is considered for the image, but other color model can be used as well. The methods of Fourier transform allows for using many other smoothing filters instead of the Gaussian function and analyze their performance in color image enhancement. The examples of color image enhancement are shown and comparison with existing methods of retinex enhancement is described, and simple MATLAB-based codes are given.

9411-15, Session 4

Cross-standard user description in mobile, medical oriented virtual collaborative environments

Rama Rao Ganji, Mihai Mitrea, Bojan Joveski, Afef Chammem, Télécom SudParis (France)

Defining, designing, implementing and deploying a virtual collaborative environment devoted to e-healthcare remain a challenging research task, which should face three-folded technical constraints. First, the various actors involved in the collaboration (general physicians, specialists, paramedical, first aid intervention units, etc.) should access the medical data in a personalized manner and under strict authentication/confidentiality constraints. Secondly, the medical data should be presented to the user according to well-established practices (e.g. lossless compression and colorimetric calibration for a large class of medical images). Thirdly, the medical applications are generally connected to intensive computation and memory resources consumption, which is a priori a blocker for their deployment on mobile environments.

The present paper advances a novel architecture for mobile, medical oriented virtual collaborative environments. In this respect, four different open standards belonging to the ISO/IEC JTC1/SC29 WG11 (a.k.a. MPEG) and W3C families complement each other so as to meet all the above-

mentioned constraints.

The data characterizing the various users (their identities, roles, rights, etc.) are represented according to the emerging MPEG-UD (MPEG user description) standard. The security issues (privacy, authentication) are solved by deploying webID principles. This choice ensures the potential interoperability among different healthcare systems, from the user management point of view.

Irrespective of their elementary types (text, image, video, 3D, ...), the medical data are aggregated into hierarchical, interactive multimedia scenes which are alternatively represented into MPEG-4 BiFS (Binary Formats for Scenes) or HTML5 standards. This way, each type of content can be optimally encoded according to its particular constraints (semantic, medical practice, network conditions, etc.). Such resource intensive tasks are performed on remote servers (in a private or public cloud).

The mobile device should ensure only the displaying of the content (inside an MPEG player or an HTML5 browser) and the capturing of the user interaction. The scalability with respect to the number of collaborators is ensured by specifying novel management rules for the collaborative messages.

The overall architecture is implemented and tested under the framework of the MEDUSA European project, in partnership with medical institutions.

The testbed considers a server emulated by a PC and 10 mobile users, featured with heterogeneous devices (tablets, smartphones, laptops) running under iOS, Android and Windows operating systems. The connection between the users and the server is alternatively ensured by WiFi and 3G/4G networks.

Both objective and subjective evaluations have been carried out. The objective evaluation considers the quality of the content displayed on the mobile device, the bandwidth consumption, the network round trip time as well as the maximal memory/CPU resources requested on the mobile device. The subjective evaluation relates to the quality of experience perceived by the user.

About the Symposium Organizers

IS&T, the Society for Imaging Science and Technology, is an international non-profit dedicated to keeping members and others apprised of the latest developments in fields related to imaging science through conferences, educational programs, publications, and its website. IS&T encompasses all aspects of imaging, with particular emphasis on digital printing, electronic imaging, color science, photofinishing, image preservation, silver halide, pre-press technology, and hybrid imaging systems.

IS&T offers members:

- Free, downloadable access to more than 16,000 papers from IS&T conference proceedings via www.imaging.org
- Complimentary online subscriptions to the *Journal of Imaging Science & Technology* or the *Journal of Electronic Imaging*
- Reduced rates on IS&T and other publications, including books, conference proceedings, and a second journal subscription.
- Reduced registration fees at all IS&T sponsored or co-sponsored conferences—a value equal to the difference between member and non-member rates alone—as well as on conference short courses
- Access to the IS&T member directory
- Networking opportunities through active participation in chapter activities and conference, program, and other committees
- Subscription to the IS&T *The Reporter*, a bi-monthly newsletter
- An honors and awards program

Contact IS&T for more information on these and other benefits.

IS&T
7003 Kilworth Lane
Springfield, VA 22151
703/642-9090; 703/642-9094 fax
info@imaging.org
www.imaging.org

SPIE is an international society advancing an interdisciplinary approach to the science and application of light. SPIE advances the goals of its Members, and the broader scientific community, in a variety of ways:

- SPIE acts as a catalyst for collaboration among technical disciplines, for information exchange, continuing education, publishing opportunities, patent precedent, and career and professional growth.
- SPIE is the largest organizer and sponsor of international conferences, educational programs, and technical exhibitions on optics, photonics and imaging technologies. SPIE manages 25 to 30 events in North America, Europe, Asia, and the South Pacific annually; over 40,000 researchers, product developers, and industry representatives participate in presenting, publishing, speaking, learning and networking opportunities.
- The Society spends more than $3.2 million annually in scholarships, grants, and financial support. With more than 200 Student Chapters around the world, SPIE is expanding opportunities for students to develop professional skills and utilize career opportunities, supporting the next generation of scientists and engineers.
- SPIE publishes ten scholarly journals and a variety of print media publications. The SPIE Digital Library also publishes the latest research—close to 20,000 proceedings papers each year.

SPIE International Headquarters
1000 20th St., Bellingham, WA 98225-6705 USA
Tel: +1 360 676 3290
Fax: +1 360 647 1445
help@spie.org • www.SPIE.org

2015 Electronic Imaging

SCIENCE AND TECHNOLOGY

Thank you for your participation!

www.electronicimaging.org

IS&T imaging.org

SPIE.