1 Introduction

Robots and computers have become a prominent aspect of our lives, and their presence will enable unique technologies. There are many difficulties to overcome before robots can interact fluidly with human beings. Imagine a robotic psychiatrist that can interpret a patient's emotions and formulate an appropriate response. Emotion and expression recognition are intuitive but extremely complicated tasks, and they lie at the base of human–computer interactions (HCIs) ranging from mobile computing and gaming to health monitoring and robotics. It is a challenging problem because of the inherent variability of expressions that may carry the same meaning, resulting in overfitting issues because of the limited size of the available training set, and also the presence of factors such as viewpoint, illumination, and occlusions. Humans combine both visual and audio information to recognize an emotion such as anger. If robots and computers are to interact with humans effectively (e.g., a robot psychiatrist), they need to be able to process both visual and audio information to produce a single output.

Computer recognition of human emotion and activity from video is an increasingly important topic of computer vision research.[1–3] Coverbal cues, such as facial expressions (including gaze), head, arm, finger, and body gestures, combined with speech prosody, signal how speakers feel about what they say.[4] An important body of research has been performed on facial expression recognition (FER) and emotional speech recognition.[5–9] Several databases are freely accessible to test systems, and challenges have been organized for emotion or activity recognition using each of these modalities (emotion and activities). Few cross-modality challenges have been performed and almost none trying to combine voice, face, and body gestures. The proposed project intends to develop a robust and accurate multi-modal expression recognition procedure by combining acoustic and face cues with body gestures, such as of the shoulders, arms, and hands, which would lead to enhanced HCI.

93% of human communication is non-verbal (body language, facial expression, and tone of the voice), and it imparts most of the information.[10,11] The following examples show the great potential of reliable and automated emotion recognition:

- Sensitivity to the emotional state of a student could have an important role in computer-aided learning; for instance, the computer can encourage a student when it detects that the student is excited.
- Researchers are working on smart rooms where people are assisted in their everyday activities by distributed intelligence in their living places. This concept is especially appealing for societies such as the European Union that have a rapidly aging population and where semi-automatic assistance to non-autonomous elderly people is becoming a subject of interest.
- The interaction between humans and virtual characters in computer gaming opens up the possibility of a new generation of game controllers and education.

Emotion recognition involves three main steps, namely, visual-based emotion recognition, audio-based emotion recognition, and finally a fusion of the decisions of the first two steps to produce the final emotion recognition. Each of these stages needs to be investigated in detail, and this Spotlight will focus on developing solutions for them.

One of the most obvious ways in which emotion is demonstrated is facial expression.[12] Computerized FER requires feature extraction, analysis of facial expression information, and analysis and understanding of human emotions.[13] In this Spotlight, a comprehensive study of face recognition, emotion recognition, and pain recognition is provided. A detailed survey of available well-known databases is also provided.

2 Face Recognition

2.1 Introduction

Face recognition is emerging as an active research area spanning several research branches, such as image processing, pattern recognition, computer vision, and neural networks.[14,15] Humans often use faces to recognize individuals. These systems have shown significant improvements in computing capability over the last few decades, and today they enable us to simulate similar recognitions automatically. In addition, face-recognition technology has many commercial applications. These applications range from static matching of controlled format photographs, such as photos on identification cards, to real-time matching of video sequences. Although humans seem to recognize faces easily, and faces are more easily remembered by humans than other objects when presented in an upright orientation, machine recognition is a complicated task. A general issue in this field: for a given image or video sequence, identify one or more people in the scene using a training database of faces. The solution to this issue involves segmentation of faces from non-face background objects, extraction of meaningful features from the face region, identification, and matching.

2.1.1 History of face recognition

During the early 1970s, typical pattern-classification techniques that used measured attributes among manually detected features in faces were applied. Looking at the history of face recognition since the early 1990s, research interest in face recognition has grown significantly, which is due to several reasons such as an increase in emphasis on civilian and commercial research projects, usage of neural network models and their implementations of real-time hardware systems, the increasing need for surveillance-related applications due to terrorist activities, and so on. During the mid-1990s, researchers focused more on problems such as face localization and segmentation, and extraction of features, such as eye, nose, mouth, and so on. In addition, one should not disregard the