Automated recognition of human activities in video streams in real time

Sebastiaan van den Broek, Johan-Martijn ten Hove, Richard den Hollander, and Gertjan Burghouts

A system using standard surveillance cameras senses human activities relevant for early detection of military threats.

Early detection of human activities that indicate a possible threat is needed to protect military bases or other important infrastructure. Currently, human observers are much better than computers in detecting human activities in videos. However, in many cases human operators have limitations. For example, many cameras often cover an area, and an operator can only watch one of them at a time. Also, fatigue may limit the time in which an operator can effectively perform. In military situations, resources are limited, and a full-time operator may not be available at all. For these reasons, it is desirable that computers assist in such surveillance in the future. But for that to become reality, the computer system must be able to detect people in the scene, track them, and recognize their activities.

Automated recognition of human activities is a true challenge because activities occur in many ways. There are activities that are performed by one person (running), by two people (fighting), one person with an item (pickup), two people with an item (exchange), and one person interacting with the environment (digging). Recognition of a wide range of human activities requires that the system be able to represent all of these elements. To identify the focus of attention, people must be distinguished from other parts of the scene. To capture walking patterns and associate multiple observations over time, for example, people must be tracked. To analyze their activity, their movement and appearance must be described. Then, using all of this information, we must determine what the people are doing.

State-of-the-art research has focused on person detection, tracking, motion features, and classification. But human activity recognition remains a challenge. To advance activity recognition technologies, the US Defense Advanced Research Projects Agency (DARPA) initiated the Mind’s Eye program. Several teams developed algorithms that were tested on thousands of video recordings of 48 selected activities, such as walking, running, pickup, exchange, and digging. The system achieved prominent results in DARPA’s benchmark. Our key innovations are identifying characteristic motion patterns for each activity and representing interactions between two people and between a person and an item or the environment. We integrated these new algorithms into a system that uses the best available person detection and tracking algorithms, optimizing the parameters to recognize human activities.

A military system for compound security requires detecting particular human activities. In 2013, we participated in a US Army event called Adaptive Red Team/Technical Support Operational Analysis (ART-TSOA) 13-4 to experiment with our system at a forward operating base (FOB) at Camp Roberts, California. This FOB is very large and situated in a wide, dry, and more or less empty area. We developed detectors for climbing a fence (see Figure 1) and leaving an item behind on the ground next to the fence.

Continued on next page
The next requirement is that the system operate in real time. Several adaptations we made resulted in real-time performance. As a preprocessing step the resolution is reduced and the user sets a region of interest. The motion features are implemented on the graphics processing unit (GPU), we use real-time person detection and tracking, and activity recognition is performed for only a limited set of human activities.

The real-time system was trained to detect a person who is digging, placing or picking up an object, walking, or running. The system was demonstrated at the ART/TSOA 14-2 event in Stennis, Mississippi, a setting very different from Camp Roberts. The site at Stennis is a small, watery area with more activity around the smaller FOB. But the goal was the same: to detect human activity of interest at an early stage, such that the Army can take countermeasures when there is a potential threat. An essential asset of a military system is showing an alarm on the common operational picture (COP), which is a situational awareness system displaying information from different sources geographically on a map and providing a combined picture to different users. We connected our system to the COP to display alarms when a person was digging in the field of view of a surveillance camera. The alarm showed a snapshot of detected human activity and its location on a map (see Figure 2).

Figures 3 and 4 show action recognition, where a rectangle indicates the track and circles show the locations of significant motion. Detected human activity is indicated above the video stream (see Figure 3) and communicated to the US Army’s COP by sending a message with an activity location and a snapshot of the video image, which is then displayed on a geographical view (see Figure 2).

In summary, we have developed a real-time system for recognizing a set of human activities in video streams. The system, demonstrated during a field trial organized by the US Army, has successfully detected activities such as a person who is digging, picking up items, or placing items in a scene. Demonstration videos are available online. The system was connected to a COP to show alarms and snapshots of the detected activity.

Our next step is to recognize scenarios, i.e., compounds or sequences of human activities. For instance, the placement of an improvised explosive device involves a sequence of actions. We are exploring methods to do this in a follow-up project sponsored by DARPA. The first promising results have been achieved in the European Union’s Seventh Framework Programme Security Research Project ARENA, in which cargo theft was

Continued on next page
recognized by analyzing longer-term behaviors.\textsuperscript{7} Recently, we
started a project in collaboration with the Royal Marechaussee
and Qubit Visual Intelligence at Amsterdam’s Schiphol Airport,
where we intend to detect activities such as people falling on the
ground and theft of bags amid large crowds.

\textbf{This work is supported by DARPA’s Mind’s Eye program. The content
of the information does not necessarily reflect the position or policy of
the US government, and no official endorsement should be inferred.}

\textbf{Author Information}

Sebastiaan van den Broek, Johan-Martijn ten Hove,
Richard den Hollander, and Gertjan Burghouts
Netherlands Organisation for Applied Scientific Research (TNO)
The Hague, The Netherlands

Sebastiaan van den Broek, PhD, is a research scientist in the Intel-
ligent Imaging research group. His interests in the field of image
processing and information fusion range from detection to clas-
sification, including tracking, multisensor fusion, and situation
assessment.

Johan-Martijn ten Hove is a research scientist at the Intelligent
Imaging research group at TNO. He studied applied physics at
the University of Twente (1999–2005). He is a software engineer
in the CORTEX project within DARPA’s Mind’s Eye program
about recognition of events and behaviors.

Richard den Hollander received his PhD in electrical engineer-
ing from Delft University of Technology in 2007. Since 2006 he
has been a research scientist at TNO, where he develops image
processing algorithms for various vision tasks. His research in-
terests include computer vision and pattern recognition.

Gertjan Burghouts, PhD, is a senior research scientist in visual
pattern recognition at the Intelligent Imaging research group
at TNO. He is the principal investigator for DARPA’s CORTEX
project. He has written papers in internationally renowned jour-
nals and has over 900 citations.

\textbf{References}

1. \url{http://www.darpa.mil/Our_Work/I2O/Programs/Minds_Eye.aspx}
   DARPA Mind’s Eye program website. Accessed 22 May 2014.
2. \url{http://www.visint.org} Website containing the DARPA Mind’s Eye video data
doi:10.1016/j.patrec.2013.01.024
4. H. Bouma, G. Burghouts, L. de Penning, P. Hanckmann, J.-M. ten Hove,
   S. Korzec, M. Kruithof, \textit{et al.}, \textit{Recognition and localization of relevant human behavior in
   of negative samples and two-stage combination of multiple features for action de-
doi:10.1007/s00138-013-0514-0
6. \url{http://www.youtube.com/intelligentimaging} TNO channel containing videos
   showing examples of action recognition. Accessed 22 May 2014.
7. G. Sanromà, L. Patino, G. J. Burghouts, K. Schutte, and J. Ferryman, \textit{A unified
   approach to the recognition of complex actions from sequences of zone-crossings}, \textit{Image