Real-time Abnormal Motion Detection in Surveillance Video

Nahum Kiryati, Tammy Riklin Raviv†, Yan Ivanchenko, Shay Rochel
Tel Aviv University, † *Massachusetts Institute of Technology*
nk@eng.tau.ac.il, tammy@csail.mit.edu, yan@eng.tau.ac.il, shayrochel@hotmail.com

Abstract

Video surveillance systems produce huge amounts of data for storage and display. Long-term human monitoring of the acquired video is impractical and ineffective. Automatic abnormal motion detection system which can effectively attract operator attention and trigger recording is therefore the key to successful video surveillance in dynamic scenes, such as airport terminals. This paper presents a novel solution for real-time abnormal motion detection. The proposed method is well-suited for modern video-surveillance architectures, where limited computing power is available near the camera for compression and communication. The algorithm uses the macroblock motion vectors that are generated in any case as part of the video compression process. Motion features are derived from the motion vectors. The statistical distribution of these features during normal activity is estimated by training. At the operational stage, improbable-motion feature values indicate abnormal motion. Experimental results demonstrate reliable real-time operation.

1. Introduction

A video surveillance system covering a large office building or a busy airport can apply hundreds and even thousands of cameras. To avoid communication bottlenecks, the acquired video is often compressed by a local processor within the camera, or at a nearby video-server. The compressed video is then transmitted to a central facility for storage and display.

Abnormal motion detection is the key to effective and economical video surveillance. The detection of an abnormal motion can trigger video transmission and recording, and can be used to attract the attention of a human observer to a particular video channel. The problem is characterized by three related challenges. One is the reliability requirement, meaning that irregular events should be consistently detected, while the false-alarm rate should be sufficiently low. The second is effective characterization of normal motion, allowing discrimination between normal and abnormal activity. Third, abnormal motion detection should be accomplished using the limited computational power available at or near the camera.

This paper presents a novel real-time abnormal motion detection scheme. The algorithm uses the macroblock motion vectors that are generated anyway as part of standard video compression methods [3]. Motion features are derived from the motion vectors. Normal activity is characterized by the joint statistical distribution of the motion features, estimated during a training phase at the inspected site. During online operation, improbable-motion feature values indicate abnormal motion. Relying on motion vectors rather than on pixel data reduces the input data rate by about two orders of magnitude, and allows real-time operation on limited computational platforms.

Previous works that rely on segmentation, grouping or tracking have been reported in [7, 2, 13, 20, 16, 14, 6, 10]. Steps towards liberation from segmentation and tracking in activity analysis have been taken by [11, 18, 15, 4, 9]. Activity analysis relying on anticipated characteristics of human motion, such as periodicity, gait or gestures, can be found in [1, 12, 19, 11]. In [11], principal component analysis of the macroblock motion vectors was used to match the detected activity in a video stream to known human activities (walking, running, kicking), and for selective access of details from the uncompressed domain. Novelty or activity detection in video using pixel-level motion analysis has been reported by [8, 5].

Unlike most previous methods for video analysis, the suggested approach completely avoids segmentation and tracking. Taken together with the reliance on macroblock motion vectors and the lack of a-priori presumptions regarding normal motion, these design decisions distinguish our work from most of the available literature.

2 Method

2.1 From video to motion vectors

Common video compression schemes exploit both the spatial and the temporal (frame-to-frame) redundancy present in the image sequence [3]. A frame is either an *intra-frame* that is compressed as a full still image, eliminating its spatial redundancy, or an *inter-frame* represented by macro-block displacement vectors relative to (say) the previous frame, and an error image. Intra-frames are generated at constant intervals, to allow random-access to the content, and to reduce accumulated errors. An intra-frame is also provided when there is a significant change in the scene (e.g., an editing cut), so that representation of the current frame in terms of the previous one is inefficient.

A motion vector $V_{i,j} = \{V_{x_{i,j}}, V_{y_{i,j}}\}$ is associated with each $M_h \times M_w$ macro-block (i, j) in an inter-frame. In the current implementation $M_h = M_w = 16$ pixels. Generally, $i \in \{1, \ldots, i_{max}\}$, $j \in \{1, \ldots, j_{max}\}$. The motion vector points to the location of the most similar $M_h \times M_w$ block in the previous frame. Inter-frame l is then represented by a set of $n = i_{max} \times j_{max}$ motion vectors $V^l = \{V^l_{i,j}, i = 1, \ldots, i_{max}, j = 1, \ldots, j_{max}\}$ associated with its macroblocks, or by their $2n$ components $\{V^l_{x_{i,j}}, V^l_{y_{i,j}}, i = 1, \ldots, i_{max}, j = 1, \ldots, j_{max}\}$.

The difference between the current block and the reference block in the previous frame is compressed as part of the error image, and used for reconstruction. When the match between the current macro-block and the reference block is poor, the current block is compressed by itself and is referred to as an intra-block.

2.2 From motion-vectors to motion features

A small set F^l of $m << n$ features is derived from the set of motion vectors V^l. Its regular probability distribution is estimated during training. In the course of online operation, the feature vector F^l of the incoming frame is compared to the statistical model. If its probability is low, it is declared abnormal.

The surveillance domain knowledge allows "manual" selection of the m features in F^l. The advantage of human-designed features with respect to blindly generated ones is their clear conceptual meaning, providing insight and promoting testability and maintainability. Let

$$|V^l_{i,j}| = \sqrt{(V^l_{x_{i,j}})^2 + (V^l_{y_{i,j}})^2},$$

$$\Phi^l_{i,j} = \arctan(V^l_{y_{i,j}} / V^l_{x_{i,j}})$$

respectively denote the magnitude and direction of the motion vector $V^l_{i,j}$. The current implementation uses the following $m = 5$ features.

2.2.1 Total absolute motion

$$F^l_{TAM} = \sum_{i,j} |V_{i,j}| \tag{1}$$

This feature corresponds to the total motion in the scene. No distinction is made between the motion of 'objects' and the motion of, say, tree branches on a windy day.

2.2.2 Regional information

Dividing the frame into K rectangular sub-frames A_k, the *area of dominant motion* is obtained by:

$$F^l_{ADM} = k^* = \arg\max_k (\sum_{i,j \in A_k} |V^l_{i,j}|) \tag{2}$$

This feature is the index of the sub-area of frame l with the largest sum of absolute values of motion vectors. Informally, this is the part of the frame with the largest absolute motion. In the current implementation, $K = 9$.

The ratio between the total absolute motion in the dominant area A_{k^*} of frame l and the total absolute motion F^l_{TAM} is an indicator of *motion homogeneity* within the frame. Formally,

$$F^l_{MH} = \max_k \frac{\sum_{i,j \in A_k} |V^l_{i,j}|}{F^l_{TAM} + \epsilon} \tag{3}$$

The addition of the small positive constant ϵ to the denominator prevents division by 0 in static frames.

2.2.3 Directional information

The range of motion directions $\{-\pi, \pi\}$ is divided into R equal fractions of size $\Delta\varphi = 2\pi/R$. Let $r = 0 \ldots R-1$ be the angular fraction index. The *principal motion direction* is defined as the index of the most popular angular fraction:

$$F^l_{PMD} = r^* = \arg\max_r \sum_{i,j} (|\Phi^l_{i,j} - r\Delta\varphi| < \frac{\Delta\varphi}{2}) \tag{4}$$

where the sum is incremented if the arithmetic condition is satisfied.

A measure for the dominance of the principal motion direction is obtained by the ratio of the total motion in the principal motion direction and the total absolute motion in the frame:

$$F^l_{DPM} = \frac{\sum_{i,j} |V^l_{i,j}| (|\Phi^l_{i,j} - r^*\Delta\varphi| < \frac{\Delta\varphi}{2})}{F^l_{TAM} + \epsilon} \tag{5}$$

2.3 Training and online detection

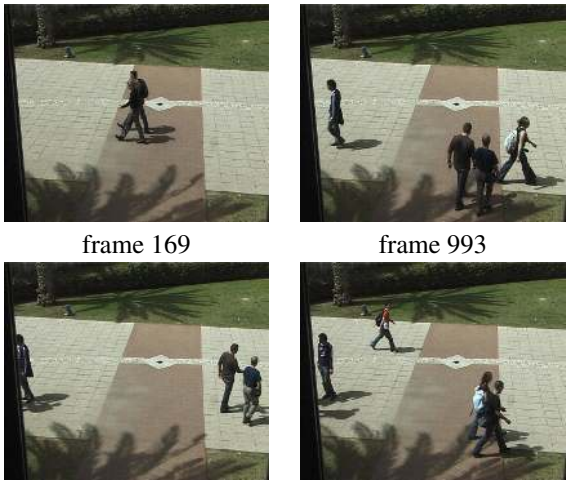
The feature vector F^l corresponding to frame l is represented by a point in an m-dimensional feature space. The essence of the training phase is estimation or modeling of the probability density function of feature vectors during normal conditions. Having an estimate of the probability density function allows, in the operational stage, to associate with each incoming frame the probability density of its feature vector under the normal motion hypothesis. The requirement of real-time computation at the full video rate supports the selection of a histogram that holds a discrete approximation of the m-dimensional probability density function of the feature vectors obtained during the training stage. In the detection phase, the feature vector associated with each incoming frame is computed. When the probabilities of the occurrence of k features vectors associated with k consecutive frames are below a threshold T, the k-est frame is declared abnormal.

3 Experimental results

The suggested abnormal motion detection algorithm was successfully tested at outdoor location. The algorithm was implemented in C++. The simplicity of the computations and the well-defined dynamic ranges allow fixed-point numerical representation. The code runs on a Pentium 4 2.8GHz PC with a Windows C++ graphical user interface at a rate of 75 frames per second, without optimization. This is three times faster than the video rate. In this experiment, the camera captured a pedestrian pathway from a nearby building. The complete movie, with analysis by our system, can be found at http://abn-motion.axspace.com. Abnormal/normal motion frames are framed with red/green respectively.

Roughly 50 minutes of video were acquired. About 41 minutes of normal pedestrian traffic were used for training. The 9-minute long test sequence contained normal and abnormal activity. The movie was captured using a SONY TRV900E PAL (25fps) digital video camera. It was then transformed to a computer in DV format and coded to MPEG-1 format using the generic MPEG-2 codec from the MPEG group website `http://www.mpeg.org/MPEG/MSSG/#source`.

Several representative frames from the video sequence are provided. Examples of frames showing normal behavior are presented in Fig. 1. A few frames detected as abnormal are presented in Fig 2. The frames shown belong to a jumping episode, to a running and grass-crossing episode and to service vehicle episode. Note that the semantic descriptions (jumping, running, grass-crossing) are provided merely for clarity. The operation of the algorithm is based on global motion features, without segmentation, tracking or any other attempt for semantic interpretation. These events are abnormal simply in the sense that similar motion patterns had not been observed (generally, have only rarely been observed) during the training session.



frame 169 frame 993

frame 1044 frame 1382

Figure 1: Examples of normal behavior.

4 Discussion

We presented a computationally efficient and reliable method for abnormal motion detection in compressed video streams. The input to the algorithm is the set of macro-block motion vectors (as well as intra-frame and intra-block flags) that are produced anyway by the compression process - an essential part of many modern video surveillance systems.

In the context of video analysis, 'normal' and 'abnormal' are fundamentally hard to define. The best current way to evaluate an abnormal motion detector is by learning the patterns of normal activity. Since the learning is based global motion features, no attempt is made to associate the motion abnormality detected with any 'object' in the scene. From a practical point of view, since the algorithm is used mainly for triggering video recording for later human analysis, or for triggering transmission to a human observer, detecting the 'object' that generated the abnormal motion is much less important than the detection of the abnormality itself. Fundamentally, the algorithm can detect abnormal motion that cannot be associated with any specific object. For example, panic in a crowd of people at a subway

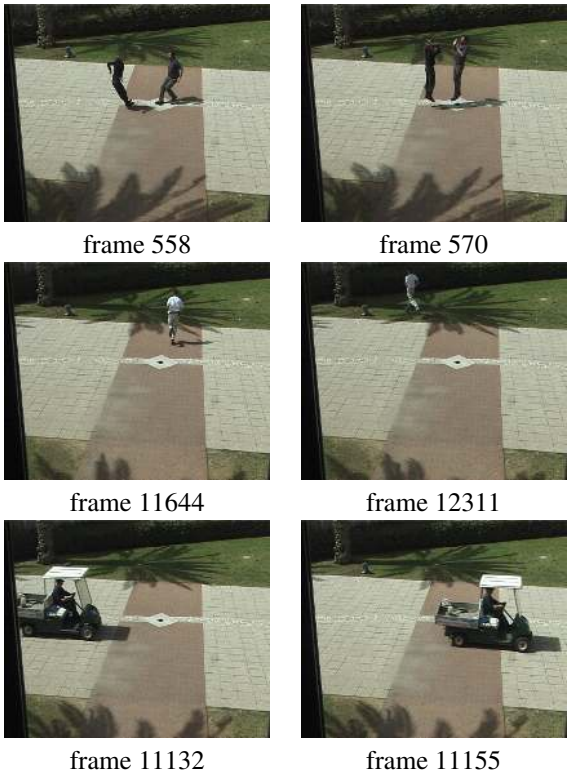


Figure 2: Frames taken from a jumping episode (first row), from a running and grass-crossing episode (second row) and from a service vehicle episode (third row) are detected as abnormal.
The complete movie, with analysis by our system, can be found at http://abn-motion.axspace.com.

station, or an unexpected tsunami wave at the seafront, are alarming situations characterized by abnormal motion, even though no particular 'object' in the scene can be associated with it.

The algorithm is modular, in the sense that different feature vectors can be suggested and alternative probability density estimation or modeling methods can be used. Further extension given long training sequences would be to learn normal pattern of activity from short frame sequences using for example Markov chains.

References

[1] J.K. Aggarwal and Q. Cai, Human motion analysis: a review, CVIU, Vol. 73, No. 3, pp. 428-440, 1999.

[2] C. Beleznai, B. Frühstück and H. Bischof, Tracking multiple humans using fast mean shift mode seeking, Int. Workshop on Visual Surveillance, pp. 25-32, 2005.

[3] V. Bhaskaran and K. Konstantinides, *Image and Video Compression Standards: Algorithms and Architectures*, Kluwer, 1997.

[4] O. Boiman and M. Irani, Detecting irregularities in images and in video, ICCV,2005.

[5] A. A. Efros, A. C. Berg, G. Mori and J. Malik, Recognizing action at a distance, ICCV, 2003.

[6] R. Hamid, Y. Huang and I. Essa, ARGMode - Activity recognition using graphical models, CVPR, Vol. 4, pp. 38-44, 2003.

[7] I. Haritaoglu, D. Harwood and L.S. Davis, W4: Real-time surveillance of people and their activities, PAMI, Vol. 22, No. 8, pp. 809-830, 2000.

[8] R.S. Gaborski, V.S. Vaingankar, V.S. Chaoji, A.M. Teredesai and A. Tentler, VENUS: A system for novelty detection in video streams with learning, FLAIRS Conference, 2004.

[9] C. Kaas, J. Luettin, R. Mattone and K. Zahn, Evaluation of a self-learning event detector, In Video-Based Surveillance Systems: Computer Vision and Distributed Processing, Kluwer, 2002.

[10] G. Medioni, I. Cohen, F. Bremond, S. Hongeng and R. Nevatia, Event detection and analysis from video streams, PAMI, Vol. 23, No. 8, pp. 873-889, 2001.

[11] B. Ozer, W. Wolf and A.N. Akansu, Human activity detection in MPEG sequences, Workshop on Human Motion pp. 61, 2000.

[12] R. Polana and R. Nelson, Detection and recognition of periodic non-rigid motion, IJCV, Vol. 23, No. 3, pp. 261-282, 1997.

[13] S. Rao and P.S. Sastry, Abnormal activity detection in video sequences using learned probability densities, TENCON, 2003, Vol. 1, pp. 369-372.

[14] C. Rao, A. Yilmaz and M. Shah, View-invariant representation and recognition of actions, IJCV, Vol. 50, pp. 203-226, 2002.

[15] J. Sherrah and S. Gong, VIGOUR: A system for tracking and recognition of multiple people and their activities, ICPR, pp. 179-182, 2000.

[16] C. Stauffer and W.E.L. Grimson, Learning patterns of activity using realtime tracking, PAMI, Vol. 22, No. 8, pp. 747-757, 2000.

[17] A. Veeraraghavan, A.R. Chowdhury and R. Chellappa, Role of shape and kinematics in human movement analysis, CVPR, Vol. I, pp. 730-737, 2004.

[18] T. Xiang and S. Gong, Beyond tracking: modelling activity and Understanding behavior, IJCV, Vol. 67, No. 1, pp. 21-51, 2006.

[19] Y. Yacoob and M.J. Black, Parametrized modeling and recognition of activities, ICCV,1998.

[20] H. Zhong, J. Shi and M. Visontai, Detecting unusual activity in video, CVPR, Vol. 2, pp. 819-826, 2004.