Spatial-Temporal Algorithm for Moving Objects Detection in Infrared Video Sequences

Dragoljub Pokrajac[1], Vesna Zeljkovic[2], Longin Jan Latecki[3]

Abstract – Performance of moving objects detection algorithm based on spatiotemporal blocks on infrared videos is discussed. The algorithm decomposes spatiotemporal blocks using dimensionality reduction technique to obtain a compact vector representation of each block and to suppress the influence of noise. The proposed method is evaluated on monochrome and multispectral IR videos.

Keywords – Video surveillance; Motion detection; Infrared videos.

1. INTRODUCTION

In this paper, we evaluate the performance of motion detection algorithm introduced in [1]. Our main goal is to demonstrate that this novel technique is capable of successfully detecting moving objects in infrared videos.

Existing approaches for moving objects detection are mainly pixel based [2,3]. E.g., [4] is based on adaptive Gaussian mixture model of the color values distribution over time at a given pixel location. We adopted this approach in [1] but with a major difference that our computation is based on the spatiotemporal blocks. The novelty of our approach is the fact that we combine the pixel and region levels to a single level texture representation with 3D blocks. More precisely, we decompose a given video into spatiotemporal blocks, e.g., 8x8x3 blocks and then apply a dimensionality reduction technique to obtain a compact representation of color or gray level values of each block as vector of just a few numbers. The block vectors provide a joint representation of texture and motion patterns in videos.

We go away from the standard input of pixel values that are known to be noisy and the main cause of instability of video analysis algorithms. In contrast, the application of principal components instead of original vectors is expected to retain useful information while suppressing successfully the destructive effects of noise [5]. Hence, we have anticipated that the proposed technique will provide motion detection robust to various types of noise that may be present in infrared video sequences. This paper shows the practical approval of this theoretically asserted claim on a set of test videos.

[1]Dragoljub Pokrajac is with Delaware State University, CIS Dept and Applied Mathematics Research Center, 1200 N DuPont Hwy, Dover, DE 19901, USA. E-mail: dpokraja@desu.edu

[2]Vesna Zeljkovic is with Administration for Mutual Services of the Republic Entities, Nemanjina 24, 11000 Beograd, Serbia and Montenegro. E-mail: vesnaz@uzzpro.sr.gov.yu

[3]Longin Jan Latecki is with Temple University, CIS Dept., 1805 N Broad Street, Philadelphia, PA 19122, USA. E-mail: latecki@temple.edu

2. METHODOLOGY

The technique for moving object detection we use consists of dimensionality reduction by spatiotemporal blocks and detection of moving blocks. We treat a given video as three-dimensional (3D) array of gray pixels with two spatial dimensions and one temporal dimension. We use spatiotemporal (3D) blocks represented by N-dimensional vectors $\mathbf{b}_{I,J,t}$, where a block spans $(2T+1)$ frames and contains N_{BLOCK} pixels in each spatial direction per frame ($N=(2T+1)\times N_{BLOCK}\times N_{BLOCK}$). To reduce dimensionality of $\mathbf{b}_{I,J,t}$ while preserving information to the maximal possible extent, we use principal component analysis [5]. We estimate sample mean and covariance matrix of representative sample of block vectors corresponding to the considered types of movies and use the first N'=3 s eigenvectors of the covariance matrix **S** (corresponding to the largest eigenvalues) to create the $N\times N'$ projection matrix used for dimensionality reduction.The resulting transformed block vectors provide a joint representation of texture and motion patterns in videos.

The proposed algorithm for detecting moving blocks is an extension of the incremental EM algorithm for estimating the Gaussian mixtures [4]. As a generalization of the distance criterion proposed in [4], at each time instant t (corresponding to a frame number) we compute the Mahalanobis distances of the block vector with respect to the distribution components of the mixture. If the minimal distance (to one of distributions) is above a pre-specified threshold, the block is considered as outlier see Fig. 1, ("reset" mechanism), and labeled as 'moving'. Subsequently, the distribution component with the smallest estimated prior is replaced by a new Gaussian distribution. If the minimal Mahalanobis distance is smaller than a threshold, the block is considered as moving ("hold" mechanism) if: 1) the closest distributional component has relatively large variance but small prior probability and 2) an outlier has been detected within H frames preceding the current frame at the considered block position, see Fig 2. As demonstrated in [6], the proposed algorithm reduces false alarm rate and substantially outperforms the original Stauffer-Grimson algorithm. Details of the algorithm are provided in [1].

3. RESULTS

We demonstrate the performance of the proposed approach on two infrared video sequences. The first sequence, *RoofCam*, is obtained from Ohio State University Thermal Pedestrian Database [7]. Video was captured using a Raytheon 300D thermal sensor core with 75 mm lens. Camera was mounted on an 8-story building overlooking a pedestrian intersection on the OSU campus. Image size is 360x240 and was captured at varying sampling rates. The second sequence, *RocketLaunch*[1], is false color thermal infrared sequence of Spitzer Telescope launch taken from 3km distance. This 45s video shows the rocket passing through a cloud and includes cooling of the rocket plume after the rocket flies out of flame. In our experiments we use T=1 and N_{BLOCK} = 4 (for *RoofCam*) and N_{BLOCK} = 8 (for *RocketLaunch*). Overall experimental procedure followed the one described in [1]. Processed video-sequences are available on our website: http://ist.temple.edu/~pokie/data/TELSIKS005/.

Fig.3 contains moving objects detection for *RoofCam* video on three characteristic frames (38, 192, and 423). The identified moving blocks (right part of the image) are denoted by green and blue, depending on the mechanism which identified a moving block (reset-green and hold-blue). In the frame 38, a pedestrian appears in the "cyan" block in addition to three pedestrian near the fence on the left side of the scene. The algorithm is able to identify all four objects. Since the algorithm still learns the background, there are some "ghosts" in the vicinity of two pedestrians. In frame 192, a pedestrian walks through the "magenta" rectangular block. The four pedestrians on the frame are all properly identified by the algorithm. Since the rightmost pedestrian moves relatively fast, the majority of corresponding moving blocks are identified by the "reset" mechanism. The frame 423 contains two pedestrians, each of them correctly identified by the proposed algorithm, Observe that the proposed algorithm worked with a small number of false alarms in spite of the relatively high level of noise in the considered IR video.

Fig. 4. contains two frames (350 and 450) of the *RocketLaunch* video sequence along with the result on our moving objects detection algorithm. The frames show two characteristic phases of rocket launch. In the frame 350, the clouds are reflecting the bright infrared light of the hot rocket engines below so that clouds appear to light up and come down to meet the rocket. In this frame, this reflection is still merely visible (left) but is clearly identified by our algorithm (blue-colored ellipsoid above the rocket top). The algorithm is also able to identify the base of the rocket flame (green). In the frame 450, the base of the jet appears through the cloud. Our algorithm clearly identifies it using the "reset" mechanism.

The proposed algorithm can identify movement in particular region of interest. To accomplish this, we first define rectangular spatial windows corresponding to the regions of interest and compute the following spatial-windows based evaluation statistics. We count the number of identified moving block within the spatial window and normalize it with the window size.

In Fig.5, we show the computed statistics for *RoofCam* sequence on two rectangular blocks [120:140, 110:130] and [175:200, 280:300], annotated on Fig. 1. Visual inspection of the video sequence indicates the existence of two moving objects in the first block: one in frames 6-65 and another in the frames 242-284. The values of motion statistics correspond to these two intervals, indicating two peaks of motion activity. E.g., the motion statistics value in the frame 38 is large, corresponding to the moving objects actually appearing within the observed rectangular block in this frame (see Fig. 1). In the second block, [175:200, 280:300], a moving object appears in frames 171-210. This corresponds to a peak in computed motion statistics (reaching 24% for frame 191). Observe that due to the inertia of the proposed method, moving blocks continue to be identified in the observed rectangular regions for a while after the actual motion stops. However, by thresholding the motion statistics, it is still possible to get clear indication of the factual presence of the moving object in the rectangular region.

4. CONCLUSION

We have demonstrated that our moving object detection algorithm based on spatiotemporal blocks and linear variance-preserving dimensionality reduction can perform successful detection of moving objects in infrared videos. As a performance measure we, in addition to a visual evaluation, used spatial-windows based evaluation statistics and hand-labeled ground truth moving objects detection. The proposed moving object detection can provide low false positive error rates and successful identification of the front edge of the moving object. Currently, a practical real-time video-surveillance system using the proposed technology is under development. Our work in progress is concentrated on reducing the algorithm inertia and testing and improving its performance when background illumination changes.

[1]Video courtesy NASA/JPL-Caltech. Available at http://www.spitzer.caltech.edu/picturegallery/ir_launch.shtml

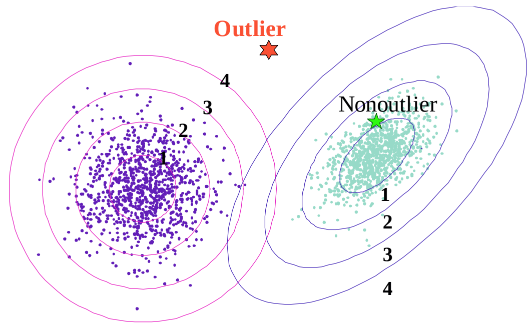

Fig. 1. Two Gaussian distributions and Mahalanobis distances from their centroids. An outlier (**left**) is identified as sample with minimal Mahalanobis distance larger than the *threshold=4*. A non-outlier (**right**) has minimal Mahalanobis distance of about 1.

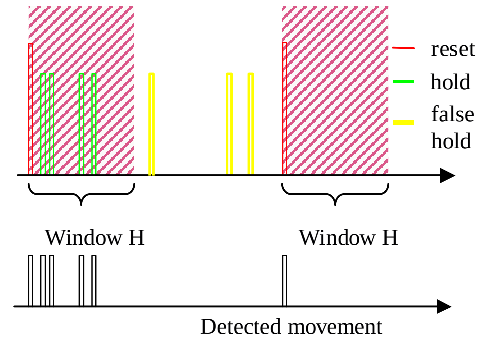

Fig. 2. Proposed mechanism for reducing false alarms; movement is detected by *hold* mechanism only within H frames after detection by the *reset* mechanism.

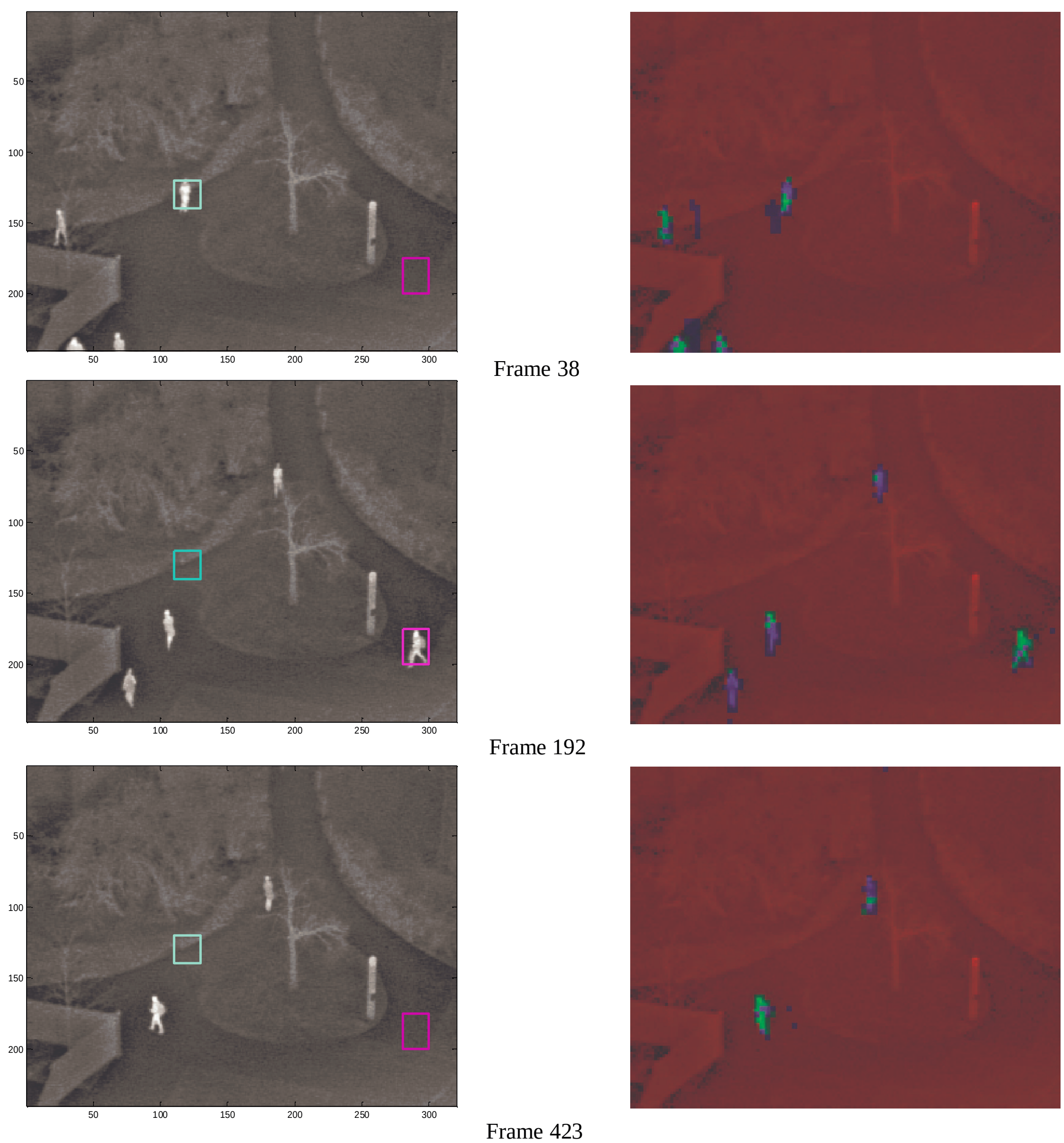

Fig. 3. Original frames 38, 192, and 423 of *RoofCam* video with Spatial blocks [175:200, 280:300] and [120:140, 110:130] denoted respectively by magenta and cyan (left) and result of moving objects detection (moving objects—green: reset, blue: hold mechanism; red—background), right

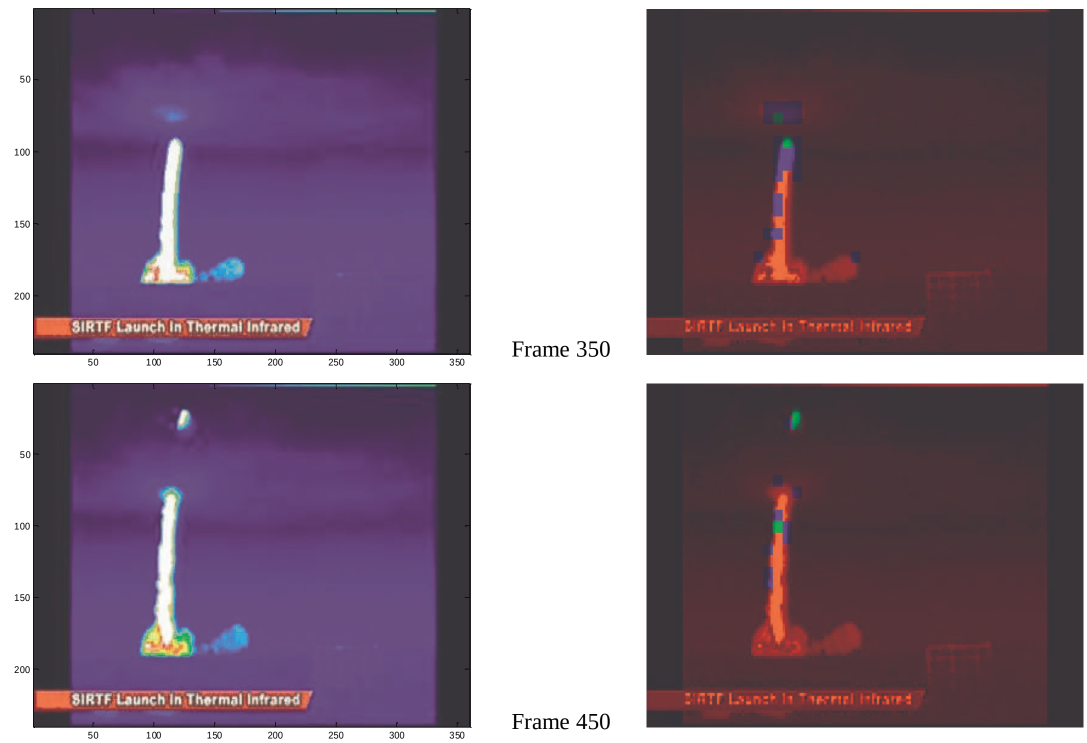

Fig. 4. Original frames 350 and 450 of *RocketLaunch* video, left and result of moving objects detection (moving objects—green: reset, blue: hold mechanism; red—background), right

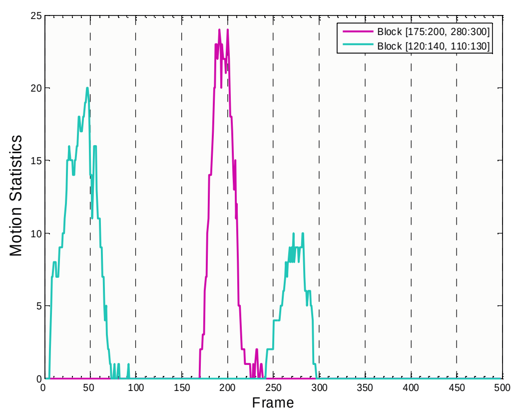

Fig. 5. Motion statistics—Percentage of identified moving objects—at spatial blocks [175:200, 280:300] and [120:140, 110:130] calculated for the *RoofCam* sequence

ACKNOWLEDGEMENT

D. Pokrajac and V. Zeljkovic have been partially supported by DoD HBCU/MI Infrastructure Support Program (45395-MA-ISP Department of Army). In addition D. Pokrajac has been partially supported by NIH-funded Delaware IDeA Network of Biomedical Research Excellence (INBRE) Grant, DoD HBCU/MI Infrastructure Support Program (45395-MA-ISP Department of Army), and the National Science Foundation (awards # 0320991, #HRD-0310163).

REFERENCES

[1] Pokrajac, D., Latecki, L. J. “Spatiotemporal Blocks-Based Moving Objects Identification and Tracking”, *In Proc. IEEE Int. W. Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS)*, Nice, France, 2003.

[2] Jain, R., Militzer, D., and Nagel, H. “Separating Nonstationary from Stationary Scene Components in a Sequence of Real World TV Images”, *In Proc. Int. Joint Conf. on Artificial Intelligence IJCAI 77 (Cambridge, MA, 1977)*, 612–618.

[3] Remagnino, P., G. A. Jones, N. Paragios, and C. S. Regazzoni, eds., *Video-Based Surveillance Systems*, Kluwer Academic Publishers, 2002.

[4] C. Stauffer, and W. E. L. Grimson, “Learning patterns of activity using real-time tracking”, *IEEE Trans. PAMI* 22(8) (2000), pp. 747–757.

[5] Jolliffe, I. T, *Principal Component Analysis*, 2nd edn., Springer Verlag, 2002.

[6] L. J. Latecki, R. Miezianko, V. Megalooikonomou, D. Pokrajac, “Using Spatiotemporal Blocks to Reduce the Uncertainty in Detecting and Tracking Moving Objects in Video,” *Int. J. Intelligent Systems Technologies and Applications*, in press.

[7] J. Davis and M. Keck, “A two-stage approach to person detection in thermal imagery”, IEEE OTCBVS WS Series Bench: In *Proc. Workshop on Applications of Computer Vision*, January 2005.