Noise-Resilient Detection of Moving Objects Based on Spatial-Temporal Blocks

Dragoljub Pokrajac[1], Vesna Zeljkovic[1], Longin Jan Latecki[2]
[1]Delaware State University, CIS Dept and Applied Mathematics Research Center., Dover,
[2]Temple University, CIS Dept., Philadelphia
E:mail: *latecki@temple.edu*

Abstract - *In this paper we discuss the resilience of moving objects detection algorithm based on spatiotemporal blocks on various types of additive and multiplicative noise. After a given video is decomposed into the spatiotemporal blocks, the algorithm uses dimensionality reduction technique to obtain a compact vector representation of each block and to suppress the influence of noise. We evaluate the algorithm performance by comparing "ground truth" (hand-labeled moving objects) to properly defined spatial-windows based evaluation statistics. Our results on a PETS repository video show that detection and tracking of moving objects is substantially improved in presence of Gaussian, speckle, multiplicative and Poisson noise.*

Keywords – *Video Surveillance, Motion Detection, Principal Component Analysis, Spatial-Temporal, Noise*

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 resistant to influence of various types of noise and to augment the reasons for such desirable behavior.

A common feature of the existing approaches for moving objects detection is the fact that they are pixel based [2,3,4,5,6]. One of the most successful of these approaches [7] 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 based on 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.

Observe that 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 [8]. Hence, we have anticipated that the proposed technique will provide motion detection robust to various types of noise that may be present in video sequence. This paper shows the practical approval of this theoretically asserted claim on a test video from PETS repository[1].

[1]Available at ftp://pets.rdg.ac.uk/.

2. METHODOLOGY

The technique for moving object detection we use consists of two major phases: 1) dimensionality reduction by spatiotemporal blocks; and 2) detection of moving blocks using incremental learning of Gaussian distributions and outlier detection.

We treat a given video as three-dimensional (3D) array of gray pixels $p_{i,j,z}$, i=1,…,X; j=1,…,Y; z=1,…,Z with two spatial dimensions X, Y and one temporal dimension Z. 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)× N_{BLOCK}× N_{BLOCK}). To reduce dimensionality of $\mathbf{b}_{I,J,t}$ while preserving information to the maximal possible extent, we use principal component analysis [8]. The resulting transformed block vectors $\mathbf{b}^{*}_{I,J,t}$ provide a joint representation of texture and motion patterns in videos.

For principal component analysis, 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 proposed algorithm for detection of moving blocks is a variant of the incremental EM algorithm for estimating the Gaussian mixtures in Stauffer und Grimson [7] extended by additional mechanism for detecting blocks corresponding to moving objects. The mixture consists of K components, and each component is specified by its estimated mean vector, a diagonal covariance matrix, and a distributional prior. As a generalization of the distance criterion proposed in [7], at each time instant t (corresponding to a frame number) we compute the squared Mahalanobis distances [9] of the block vector with respect to the distribution

components the mixture estimated for all blocks that appeared at the same position at previous time instants. If the minimal squared distance is above a pre-specified threshold, the block is considered as outlier and labeled as 'moving'. Subsequently, the distribution component that has the smallest estimated prior probability at the moment is replaced by a new Gaussian distribution. If the minimal squared Mahalanobis distance to one of distribution, the block still may belong to a moving object. Therefore, we employ the second criterion to detect moving blocks. First, we check whether an outlier has been detected within H frames preceding the current frame at the considered block position. If there were no outliers within the H previous frames, the block at the current frame is labeled as background. Otherwise, we label the block as moving if the closest distributional component has relatively large variance but small prior probability. Details of the algorithm are provided in [1].

3. RESULTS

We have demonstrated the performance of the proposed approach on sequences from the Performance Evaluation of Tracking and Surveillance (PETS) repository[2]. Processed video-sequences are available on our web site: http://tesla.desu.edu/~pokie/ELMAR2005/. Here, we present results on a video sequence from PETS2001[3] (here referred to as the *Outdoor video* sequence).

Since the original sequences contained RGB colors, we converted RGB to grayscale (PAL luminance). In addition, we reduced the size of the videos twice such that the frame size for the *Outdoor video* sequence is X=288, Y=384 (in contrast to the 576×768 pixel frames of the original video). In our experiments we use T=1 and N_{BLOCK} = 8, thus the length of a block vector $\mathbf{b}_{I,J,t}$ is N = 192 = 8×8×3. Overall experimental procedure followed the one described in [1].

3.1 Moving objects identification

We experimented with additive Gaussian, Salt&Pepper, multiplicative ("speckle") and Poisson noise. The additive Gaussian noise was zero mean, with variance ranging from 0.01 to 0.5. The Salt&Pepper noise densities varied from 0.05 to 0.2. The variance of the speckle noise ranged from 0.01 to 0.5. The result of the proposed approach on the *Outdoor video* sequence is illustrated in Fig. 1, for four different types of noise on Frame 2500 (blocks identified as moving are marked blue and green depending on the applied criterion). As we can see, the proposed technique is able to successfully and precisely detect moving objects even in case of relatively strong noise influence. The artifacts present in the Fig. 1 are mainly one-block dimensional, so they can be easily removed with median filter.

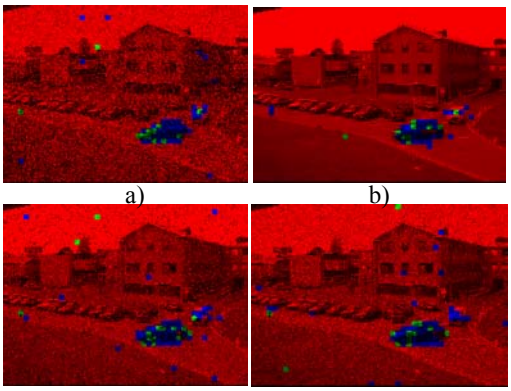

Fig. 1. Detection of moving objects on frame 2500 of *Outdoor video under* a) Gaussian zero-mean noise with variance 0.1 b) Poisson noise b) Salt and paper noise with density 0.1; d) Speckle noise with variance 0.1.

3.2. Performance evaluation

To demonstrate the influence of varying noise levels on the performance of our algorithm, we computed spatial-windows based evaluation statistics. We counted the number of identified moving block within a pre-specified spatial window and normalized it with the number of spatial blocks in the same window. We hand-labeled the observed spatial window by denoting time intervals when a moving object is present in the window in order to compare the result of automatic detection of moving objects with "ground truth".

In Fig. 2, we show the computed statistics for sequence without noise, different levels of Gaussian noise (0.01, 0.05, 0.1, 0.5) as well as ground truth moving objects detection in rectangular region (350, 510; 500, 600). It can easily be observed that, in spite of increased levels of noise, it is still possible to detect a moving object in a window by properly thresholding the observed statistics. Observe that such identification agrees with the ground truth.

3.3 Comparison to pixel-based approaches

Recall that we performed the detection of moving objects using the first three PCA components of each 3D block vector. Hence, we can observe each particular block in time through visualizing its trajectory in the feature space of the PCA components. As observed in [1], the method applied in this paper detects moving blocks that belong to "orbits" in the feature space. Fig. 3 presents the trajectories for block location (24, 28) in the absence of noise, and the trajectories for additive Gaussian noise with 0.01 variance. We can observe that, in spite of noise, the "orbits" are still clearly visible and separable from the two major clusters (that

[2]Available at ftp://pets.rdg.ac.uk/.
[3]ftp://pets.rdg.ac.uk/PETS2001/DATASET1/TESTING/CAMERA1_JPEGS/

correspond to the background frames at the beginning and at the end of movie).

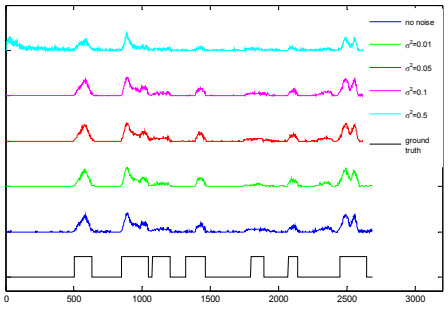

Fig. 2. Percentage of identified moving objects at spatial window (350,510; 500,600) calculated for original *Outdoor* video (no noise present), various levels of additive Gaussian noise, compared with hand-labeled ground truth (presence of moving object in the video as observed by a human).

In comparison to any pixel-based approach (e.g., Stauffer and Grimson [7]), our technique performs better since it reduces noise in background and can extract information about temporal change of texture (since it is based on spatiotemporal texture representation of 3D blocks instead of pixels). To demonstrate this, in Fig. 4 we plot loci of RGB color values that occur at the pixel location (185, 217), from the block (24, 28). In absence of noise (Fig. 4a) a pixel-based approach from [7] will be able to identify distributional components and identify moving object pixels as distributional outliers. However, in the presence of noise (additive Gaussian, variance 0.1), Fig. 4b, the distributions degenerate into one cluster with outliers corresponding not to actual moving objects but to high intensity noise. Hence, a pixel-based approach [7] will cease to properly identify moving objects and will result with high failure rate.

4. CONCLUSION

In this paper we have demonstrated that our moving object detection algorithm based on spatiotemporal blocks and linear variance-preserving dimensionality reduction is resistant on the influence of various types of noise.

We evaluated performance of the applied algorithm on benchmark videos from Performance Evaluation of Tracking and Surveillance (PETS) repository. 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 results indicate that a proper detection is still possible in spite of significant levels of additive or multiplicative noise. As we experimentally shown this can be explained by inherent capability of employed dimension reduction techniques to extract useful information from the signal (a vector representing a spatial-temporal block) while efficiently suppressing a noisy component. In contrast, pixel-based moving objects detections method become overwhelmed with the amount of noise present and cease to be useful.

Our work in progress is concentrated on applying incremental techniques to estimate feature projection matrix and on applying the proposed moving detection technique for more efficient tracking and trajectory clustering.

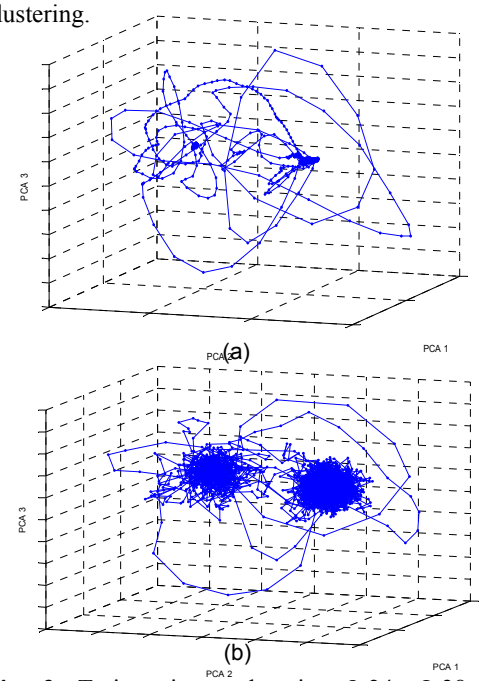

Fig. 3. Trajectories at location *I*=24, *J*=28 of the *Outdoor video* in feature space of first three block vector PCA components; a) Original video b) In presence of zero-mean Gaussian noise with variance 0.1.

ACKNOWLEDGEMENT

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), National Science Foundation (NSF) Infrastructure Grant (award # 0320991) and NSF grant "Seeds of Success: A Comprehensive Program for the Retention, Quality Training, and Advancement of STEM Student" (award #HRD-0310163). V. Zeljkovic 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).

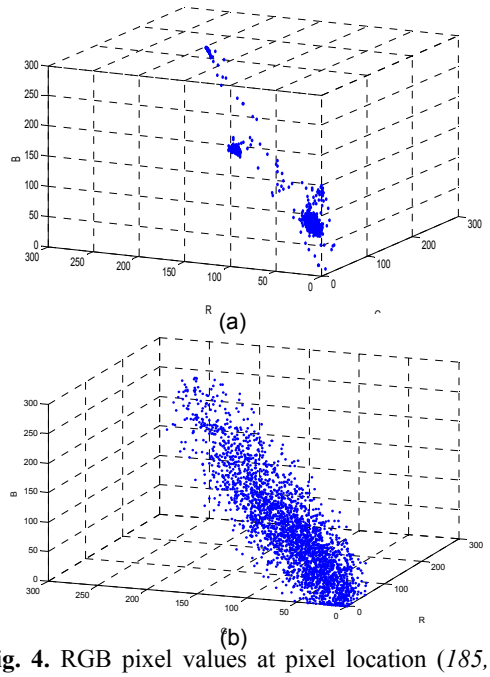

Fig. 4. RGB pixel values at pixel location (*185, 217*) (inside block 24, 28). a) Original video; b) Additive zero-mean Gaussian noise with variance 0.1.

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. International Joint Conference on Artificial Intelligence IJCAI 77 (Cambridge, MA, 1977)*, 612–618.

[3] I. Haritaoglu, D. Harwood, and L. Davis, "W4: Real-Time Surveillance of People and Their Activities", *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)* 22(8) (2000), pp. 809–830.

[4] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian Computer Vision System for Modeling Human Interactions", *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)* 22(8) (2000), pp. 831–843.

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

[6] C. Wren, A. Azarbayejani, T. Darrell, and A.P. Pentland, "Pfinder: Real-time Tracking of the Human Body", *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)* 19(7) (1997), pp. 780–785

[7] C. Stauffer, and W. E. L. Grimson, "Learning patterns of activity using real-time tracking", *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)* 22(8) (2000), pp. 747–757.

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

[9] Duda, R., P. Hart, and D. Stork, *Pattern Classification,* 2nd edn., John Wiley & Sons, 2001.