29 June – 1 July 2006, Sofia, Bulgaria

Improved Spatial-Temporal Moving Areas Detection Method Resistant to Noise

Vesna Zeljkovic[1] and Dragoljub Pokrajac[2]

Abstract — **We discuss the resilience of the spatial temporal moving objects detection algorithm on various types of additive and multiplicative noise. Video is decomposed into the spatio-temporal blocks and dimensionality reduction is used to suppress the influence of noise. The moving object detection algorithm based on spatial temporal analysis is subsequently applied.**

Keywords — **Video analysis, Motion Detection, Principal Component Analysis, Noise Reduction.**

I. INTRODUCTION

We evaluate the performance of the improved spatial-temporal moving object detection method, resistant to noise. 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 [1-5]. General drawback of these methods is their sensitivity on various types of noise that may exist in video frames due to influence of insufficient illumination, video amplifier, lens system, electromagnetic interference, etc. Recently, a moving object detection algorithm based on principal components analysis is proposed [6]. It is demonstrated [7, 8] that the application of principal component analysis can contribute to significant reduction of the number of false positives (background objects falsely labeled as moving).

In this paper, we propose to apply principal component projection as preprocessing step of our pixel-based technique for detection of moving objects [9]. With the proposed preprocessing step, we combine the pixel and region texture information. More precisely, we decompose a given video into overlapping spatiotemporal blocks, e.g., 5x5x3 blocks, and then apply a dimensionality reduction technique to obtain a compact scalar representation of color or gray level values of each block. Subsequently, the illumination robust moving object detection algorithm that detects and tracks moving objects is applied.

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 [10].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. In our earlier paper [11] we demonstrated that the proposed technique is resilient to various levels of Gaussian noise. This paper demonstrates the robustness of the proposed technique to other noise types that may exist in video frames on a test video sequence from PETS repository', (available at ftp://pets.rdg.ac.uk/)

II. METHODOLOGY

The technique for moving object detection we use consists of the following major phases: extraction of the 3D filter coefficients with the PCA analysis; dimensionality reduction by spatiotemporal blocks; image filtering of a current frame with the noise removal filter; and detection of moving objects applying the pixel based method for moving object detection resistant to illumination changes.

We treat a given video as three-dimensional (3D) array of gray pixels $p_{i,j,t}$, $i=1,\ldots,X$; $j=1,\ldots,Y$; $t=1,\ldots,Z$ 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 I, J 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 represent the block vector $\mathbf{b}_{I,J,t}$ by a scalar while preserving information to the maximal possible extent, we use principal component analysis [10] (see Fig. 1 for illustration).

In 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 eigenvector of the covariance matrix **S** (corresponding to the largest eigenvalue). This eigenvector represents the coefficients of the 3D filter used for dimensionality reduction (that suppresses the noise). In practical realizations, the 3D filter can be emulated by three 2D filters applied on three subsequent frames.

After the dimensionality reduction, we apply the following pixel based algorithm for moving object detection and tracking [9]. Consider image sequence S consisting of N video frames. The sliding mask A_t is applied on every frame t.

We calculate the pixel variance in order to estimate the potential movement in the observed area, as follows:

$$\sigma_t^2(I)=\frac{1}{card\{A_t\}}\sum_{m\in A_t}\left(\frac{B_m}{C_m}K_t-median\{A_t\}\right)^2, \qquad (1)$$

[1]Vesna Zeljkovic is with Applied Mathematics Research Center, Delaware State University, 1200 N DuPont Hwy, Dover DE 19901, USA, E-mail: vzeljkovic@desu.edu

[2]Dragoljub Pokrajac is with Computer and Information Science Department and with Graduate Department of Applied Mathematics and Theoretical Physics, Delaware State University, 1200 N DuPont Hwy, Dover DE 19901, USA, E-mail: dpokraja@desu.edu

where pixel intensities within mask A_t are denoted with B_m for a reference—background frame that does not contain changing regions and with C_m for a current frame (where we are identifying moving objects). The estimated mean of the pixel intensity ratio within A_t is denoted with μ_{A_t} .

The illumination compensation coefficient is defined as

$$K_t = \frac{\sum_{m\in A_t} C_m}{\sum_{m\in A_t} C_{1m}} = \frac{\mu_t}{\mu_1}, \qquad (2)$$

where C_{1m} is pixel intensity for the first frame in the sequence.

The algorithm performs the analysis in time and space domains simultaneously, contributing to its resistance to the illumination changes and reducing the false detection, i.e. artifacts. We average estimated pixel variances for three successive pairs of frames and threshold this average to determine the presence of moving objects. This represents temporal aspect of analysis. The ratio of pixel intensities in A_t between two frames is used to estimate the pixel variance $\sigma_t^2(I)$ for the following three pairs of successive frames: frames *t*-3 and *t*-2, *t*-2 and *t*-1 and *t*-1 and *t*, where *t* is the current frame. Thus, we obtain three pixel variances for three successive and corresponding variance pair $\sigma_{t-2}^2(I)$, $\sigma_{t-1}^2(I)$ and $\sigma_t^2(I)$. Subsequently, we compute the average value of these three variances as follows:

$$\bar{\sigma}^2(I) = (\sigma_{t-2}^2(I) + \sigma_{t-1}^2(I) + \sigma_t^2(I))\,/\,3, \qquad (3)$$

After that, the mean value is subtracted from the pixel variance of the current and previous frame:

$$\sigma_t^2 * (I) = \sigma_t^2(I) - \bar{\sigma}^2(I) \qquad (4)$$

If $\sigma_t^2 * (I) \geq \varepsilon$ (a suitable threshold), the center of A_t is marked as changing region, i.e. as a moving area.

The proposed algorithm performs analysis in three dimensions, two spatial and one temporal. The time analysis, additionally to space analysis, helps with correct moving object detection and augments the precision of the algorithm. This algorithm, it should be pointed out, exploits local intensity and does not require color information. Hence, the method can easily be used for gray movies, i.e. BW movies or infra-red monochromatic movies.

III. Results

We have demonstrated the performance of the proposed approach on sequences from the Performance Evaluation of Tracking and Surveillance (PETS) repository[1]. Processed video-sequences are available on our web site: http://ist.temple.edu/~pokie/data/ACIVS2005. Here, we present results on a video sequence from PETS2001[1] (here referred to as the *Outdoor video* sequence). Video consisted of 2688 frames PAL standard (25 frames per second, 576×768 pixels per frame). In our experiments we use $N_{BLOCK} = 5$, thus the length of a block vector $\mathbf{b}_{I,J,t}$ is $N = 75 = 5\times5\times3$.

We experimented with additive Gaussian, Salt&Pepper, multiplicative ("speckle") and Poisson noise [12]. The additive Gaussian noise was zero mean, with variance ranging from 0.1 to 0.5. The Salt&Pepper noise densities varied from 0.05 to 0.2. The variance of the speckle noise ranged from 0.1 to 0.5. In Fig. 2, we demonstrate the effects of selected levels of each noise type on frame 2500 of the *Outdoor video* sequence.

The result of the proposed algorithm on Frame 2500 of the *Outdoor video* sequence is illustrated in Fig. 3, for the same types and intensities of noise.

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 moving car in foreground and slow-moving white van are identified as well as the pedestrian on the left part of the scene. There are no false moving objects identification and no artifacts are introduced.

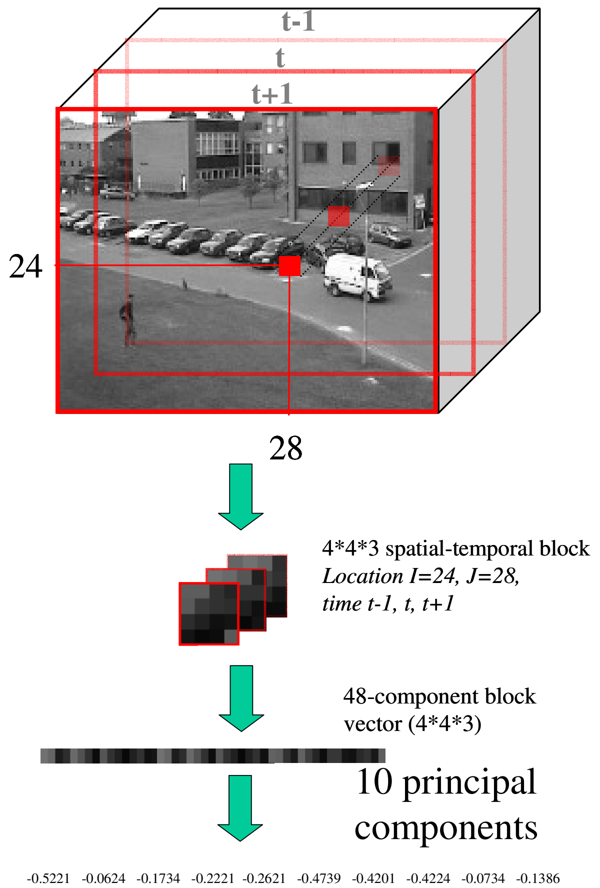
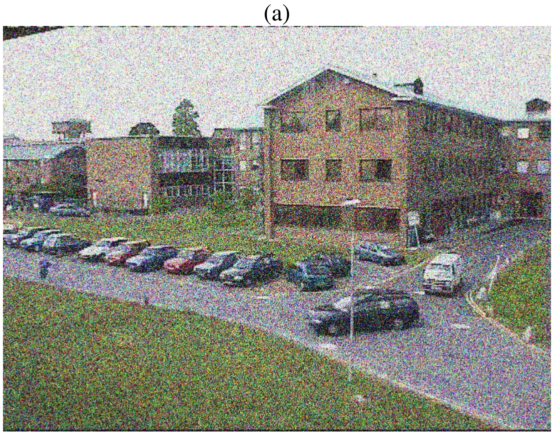
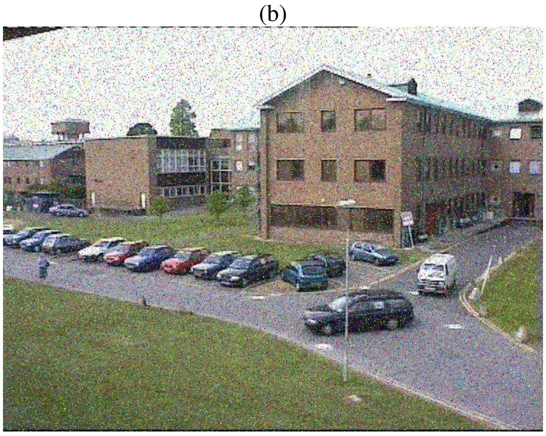


Fig. 1. Dimensionality reduction using spatial-temporal blocks.

[1]ftp://pets.rdg.ac.uk/PETS2001/DATASET1/TESTING/CAMERA1_JPEGS/


(a)


(b)


(c)

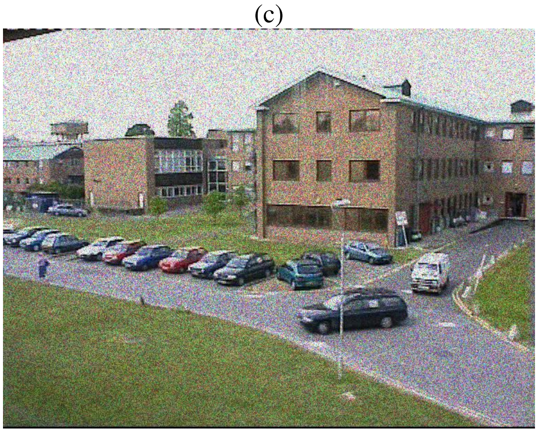
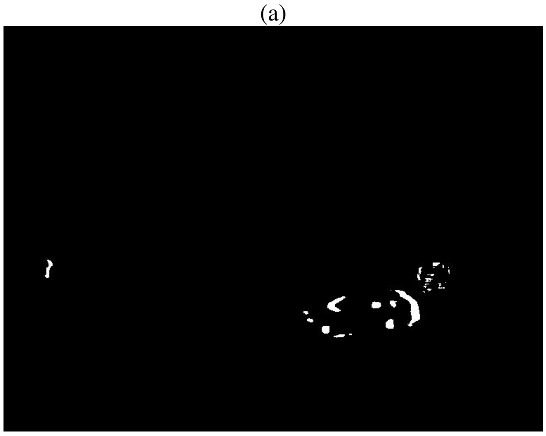
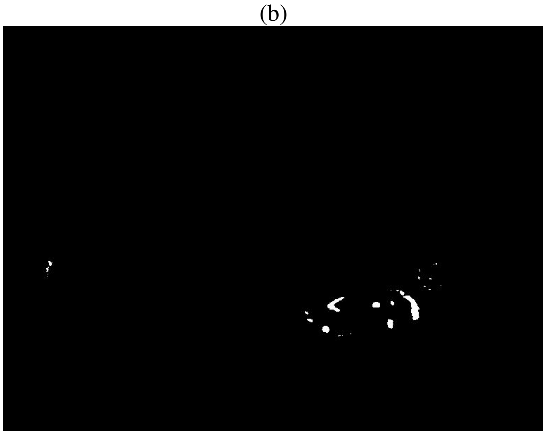

(d)

Fig. 2. The 2500th frame of *Outdoor video* under a) Gaussian zero-mean noise with variance 0.1; b) Salt and paper noise with density 0.1; c) Poisson noise; d) Speckle noise with variance 0.1.


(a)


(b)


(c)

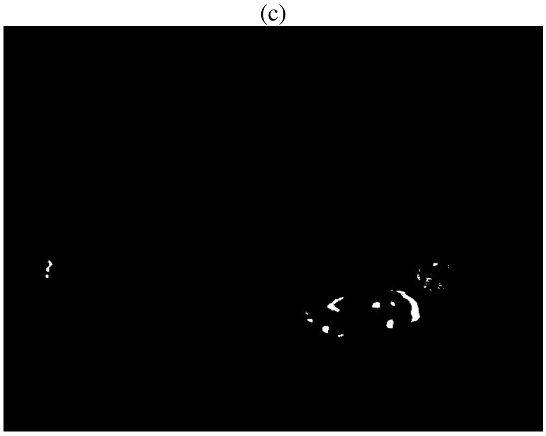

(d)

Fig. 3. Detection of moving objects on frame 2500 of *Outdoor video under* a) Gaussian zero-mean noise with variance 0.1; b) Salt and paper noise with density 0.1; c) Poisson noise; d) Speckle noise with variance 0.1

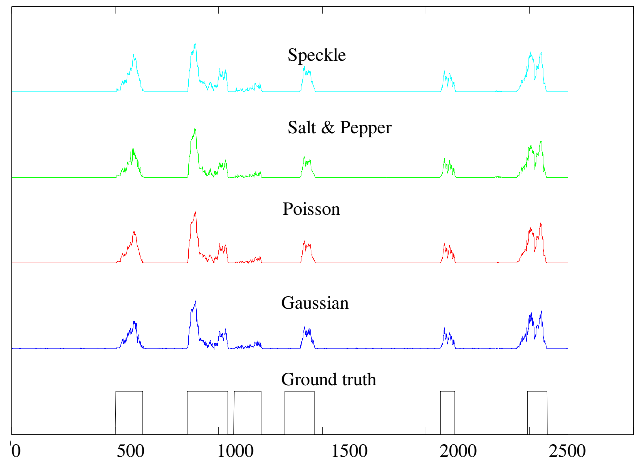


Fig. 4. Percentage of identified moving objects at spatial window (350,510; 500,600) calculated for *Outdoor* video with Gaussian zero-mean noise with variance 0.1, Poisson noise, Salt and paper noise with density 0.1, Speckle noise with variance 0.1, compared with hand-labeled ground truth (presence of moving object in the video as observed by a human)

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. 4, we show the computed statistics for sequence with Gaussian zero-mean noise with variance 0.1, Poisson noise, Salt and paper noise with density 0.1, and Speckle noise with variance 0.1, as well as ground truth moving objects detection in rectangular region (350, 510; 500, 600).

It can easily be observed that, in spite of various types of noise, it is still possible to detect a moving object in a window by properly thresholding the observed statistics. Observe also that such identification agrees with the ground truth.

IV. Conclusions

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 compared to human-identified 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 could be explained by inherent capability of employed dimension reduction techniques to efficiently suppress noisy component. Our work in progress is oriented towards theoretical justification of such behavior. Also, we would like to point out that we have not experimented with the influence of moving object velocity. That is also the part of our work in progress.

Acknowledgement

V. Zeljkovic has been partially supported by DoD HBCU/MI Infrastructure Support Program (45395-MA-ISP Department of Army). 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).

References

[1] 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.

[2] 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.

[3] 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.

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

[5] 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

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

[7] V. Zeljković, D. Pokrajac, L.J. Latecki, "Noise Robust Spatial-Temporal Algorithm for Moving Objects Detection", proc. 49th ETRAN Conf., 2005, in press.

[8] D. Pokrajac, V. Zeljković, L.J. Latecki, "Noise-Resilient Detection of Moving Objects Based on Spatial-Temporal Blocks", proc. 47th ELMAR Conf., 2005, in press.

[9] V. Zeljkovic, D. Pokrajac, A. Dorado, and E. Izquierdo, "Application of the Improved Illumination Independent Moving Object Detection Algorithm on the Real Video Sequence", In Proc. 6th International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2005.

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

[11] D. Pokrajac, V. Zeljković, "Influence of the Gaussian Noise Spatial-Temporal Method for Moving Objects Detection", submitted to 11th International Conference on Computer Analysis of Images and Patterns (CAIP) 2005.

[12] Gonzalez, R., Woods, R., Digital Image Processing, Prentice-Hall, 2002.