Precision of Some Motion Detection Methods Using Background Subtraction in Traffic Surveillance Video

Boris Nikolov[1] and Nikolay Kostov[2]

Abstract – **In these days video surveillance is used for many purposes such as security, traffic control, special measurement systems etc. In most of these systems the video stream is processed by motion detection algorithms. The goal of proposed algorithms is not only to detect motion, but to correctly subtract the foreground moving objects and to separate the foreground motion from the background motion.**

In this paper, some methods for background subtraction in video surveillance are investigated. An experiment and the results of estimating the footage of real traffic video is proposed. Conclusions about precision and computational cost for each method are given.

Keywords – **Motion detection, Background subtraction, Video surveillance.**

I. INTRODUCTION

In modern video surveillance systems motion detection is used to estimate the motion of moving objects. This is used for counting the moving objects, to recognize specific forms and to alarm if there is crossing the security border. The goal of this article is to investigate the precision of the methods used for motion detection. This is an important problem because the decision in motion detection algorithm must be correct and must avoid false detection caused by moving objects that belong to the background of the scene.

II. BACKGROUND SUBTRACTION METHODS

Background subtraction methods are used for identifying motion in video sequences using algorithm which includes creating a background model to represent the scene with no moving objects. Each pixel in the frame $I_{t(x,y)}$ is compared with the estimated background model $B_{t(x,y)}$. The pixels in the current frame that differs from the estimated background are classified as foreground. The moving objects in the scene are represented by areas of foreground pixels. To prevent incorrect classification of moving objects in the background like tree branches waving by the wind, the correctly estimation of the background model is important.

[1]Boris Nikolov is with the Technical University of Varna, Faculty of Electronics, Department of Radio Engineering, Bulgaria, Varna 9010, 1 Studentska Str.
E-mail: boris.nikolov@yahoo.com
[2]Nikolay Kostov is with the Technical University of Varna, Faculty of Electronics, Department of Radio Engineering, Bulgaria, Varna 9010, 1 Studentska Str.
E-mail: n_kostov@mail.bg

A. Temporal Averaging

Temporal averaging is a method used for background subtraction in video sequences. This method estimates background B_t by calculating the median value of the previous frames, [1]. The background is given by

$$B_{t+1} = \alpha I_t + (1-\alpha) B_t, \tag{1}$$

where α is a learning rate and is used for applying the adaptive mean.

The difference between the current frame and the background is given by

$$D_t = \left| I_t - B_t \right|. \tag{2}$$

The algorithm classifies the pixels as foreground by the rule

$$M_t(x,y) = \begin{cases} 0, & D_t(x,y) \le T \\ 1, & D_t(x,y) > T \end{cases}, \tag{3}$$

where T is a threshold preventing camera noise.

B. $\Sigma-\Delta$ background modelling

$\Sigma-\Delta$ (Sigma Delta) background modelling method is called $\Sigma-\Delta$, because the similarity to analog to digital conversion of a time varying signal using $\Sigma-\Delta$ modulation as it is interpreted in [2].

A current pixel is classified as foreground if the absolute pixel difference is greater or less than the estimated $\Sigma-\Delta$ variance. The $\Sigma-\Delta$ variance is different to the standard mathematical definition of variance. As it is defined in [2] it is a measure of the variation of the colours of each pixel over time. Depending on the difference between the current pixel value and the background the $\Sigma-\Delta$ variance are updated at each time step by incrementing or decrementing them by one.

The current background is given by

$$B_{t+1}(x,y) = \begin{cases} B_t(x,y) - 1, & D_t(x,y) < 0 \\ B_t, & D_t(x,y) = 0 \\ B_t(x,y) + 1, & D_t(x,y) > 0 \end{cases}, \tag{4}$$

where D_t is the difference, which is

$$D_t = I_t - B_t. \tag{5}$$

As it is proposed in [2], to adapt for different conditions at different areas in the image, a per pixel $\Sigma-\Delta$ variance, $V_{t(x,y)}$, is used as a threshold for each pixel. To update this variance $V_{t(x,y)}$ we implement the equation

$$V_{t+1}(x,y)=\begin{cases}V_t(x,y)-1, & C.\left|D_t(x,y)\right|<V_t(x,y)\\ V_t(x,y), & \left|D_t(x,y)\right|=0\\ V_t(x,y)+1, & C.\left|D_t(x,y)\right|>V_t(x,y)\end{cases}, \quad (6)$$

where C is a user set parameter that determines how large the difference must be, before the variance is updated. The foreground mask is given by

$$M_t(x,y)=\begin{cases}0, & \left|D_t(x,y)\right|\le V_t(x,y)\\ 1, & \left|D_t(x,y)\right|>V_t(x,y)\end{cases}. \quad (7)$$

High $\Sigma-\Delta$ variance is typical for pixels of repeating backgrounds, for example, leaves waving by the wind and wavy surface of water. Those pixels would not be classified as foreground because the difference is minor than the variance.

C. Mixture of Gaussians

The Mixture of Gaussians method describes each pixel in the frame by using multiple Gaussian distributions. Each pixel is represented by a distribution with its associated variance, weight and mean, [3].

The probability of observing the current pixel value *x* at time *t* at a particular pixel location is given by

$$P(x)=\sum_{i=1}^{K}\omega_{i,t}\eta(x;\mu_{i,t},\Sigma_{i,t}), \quad (8)$$

where K is the number of Gaussians distributions representing each pixel, $\omega_{i,t}$ is the weight of the i^{th} Gaussian at time t, η is the Gaussian probability density function with parameters: x is the current pixel, $\mu_{i,t}$ is the mean of the i^{th} distribution at time t, and $\Sigma_{i,t}$ is the covariance of the i^{th} distribution at time t.

The Gaussian probability density function η is given by

$$\eta(x;\mu,\Sigma)=\frac{1}{(2\pi)^{\frac{n}{2}}\left|\Sigma\right|^{\frac{1}{2}}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}. \quad (9)$$

A particular value *x* being observed at a pixel location has high probability if it is close to the mean of Gaussian distribution with high weight with a low variance. So, this is the Gaussian distributions that best describe each pixel. To update each distribution we use an adaptive learning rate described in [8]. This learning rate depends on the strength of the match between the current pixel value and the i^{th} distribution, q_i.

The weight of the i^{th} Gaussian at time *t* is given by

$$\omega_{i,t}=(1-\alpha)\omega_{i,t-1}+\alpha q_i. \quad (10)$$

In case there is no match it is created a new Gaussian distribution with a mean equal to the current pixel value, a low weight and a high variance. The new distribution replaces the distribution with the lowest weight and highest variance. It is assumed that the background is represented by Gaussian distributions with the highest weight and lowest variance. To estimate the background, the distributions are first sorted in order of decreasing ω/σ. The pixels that belong to the background are the first C distributions and C is given by

$$C=\arg\min_c\left(\sum_{k=1}^{c}\omega_k>T\right). \quad (11)$$

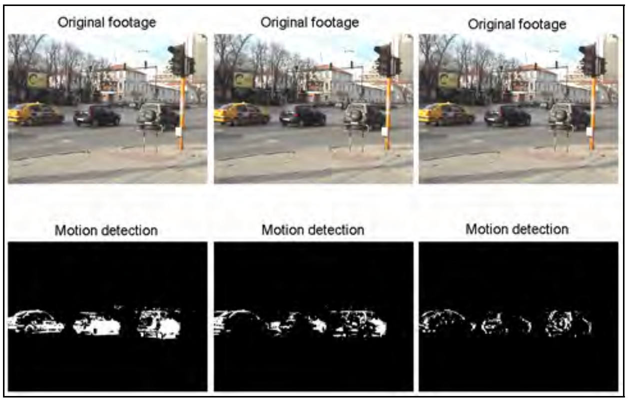
III. Experiment and Results

The proposed three methods for motion detection using background subtraction are implemented in Matlab and are shown. The processed video is captured at a city crossway where are many moving objects like cars and people. The shooting camera is stationary. The lighting of the scene is equal for all experiments. The methods, Temporal Averaging and Sigma-Delta background subtraction are ran for the original frame rate 25fps and resolution of 720x576 pixels. The Mixture of Gaussians method requires a lot of computational power and because of that it is ran for the same footage at 15fps and QVGA resolution of 320x240 pixels.

A. Temporal Averaging

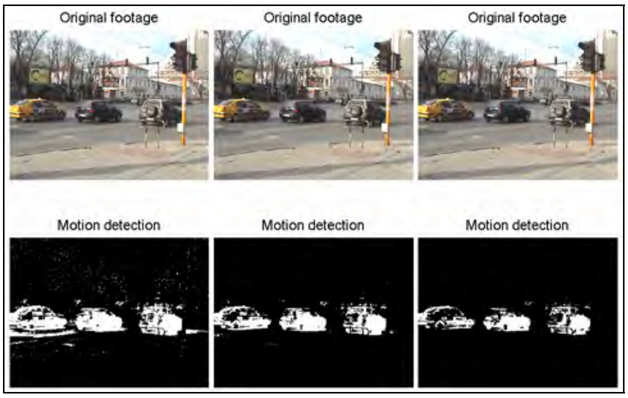
The main feature of this method is calculating historical background of the scene, Eq. (1). The learning rate α determines how fast the background is updated, so it should be between 0 to 1. As the learning rate is close to 1, the background is updating faster. This is not suitable for estimating relatively slow moving objects which are classified as a background. To estimate correctly slow moving objects α should be close to 0.1. If α is to much close to zero the natural changes in the background should not be updated in time and there will be incorrect foreground classification. In Fig. 1 are shown the original footage and the motion detection image for same sequence estimated for three different values of α. The threshold from equation (3) is equal for the three images. In Fig. 1a, there is a maximum similarity in motion detection, $\alpha=0.01$. In the next case, Fig. 1b, $\alpha=0.1$ there is an obvious trace after the moving objects, which is caused by incorrect background classification. The moving objects are assigned to the background too fast and when the object passes away the real background is estimated like a foreground. When α is to close to 1 like Fig. 1c $\alpha=0.8$, only the edges of the moving object will be detected.

The threshold T in Eq. (3) determines the resistance of noise in the image and flickering backgrounds. In Fig. 2, three screenshots of three amounts T are shown. The dimension of T is the number of the *255-th* levels of the *8 bit* depth luminance of the image in the algorithm. As the threshold is rising, the noise is more reduced, but small moving objects are hard to detect.



(a) *(b)* *(c)*

Fig. 1.Temporal Averaging method executed for three amounts of α , $\alpha = 0.01$ (a); $\alpha = 0.1$ (b); $\alpha = 0.8$ (c)

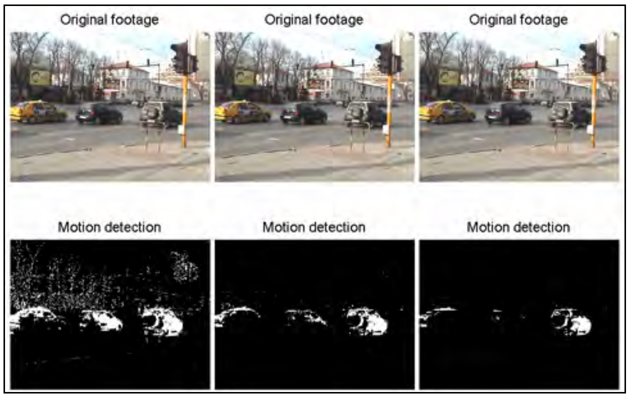


(a) *(b)* *(c)*

Fig. 2. Temporal Averaging method executed for three amounts of T , $T = 25$ (a); $T = 40$ (b); $T = 55$ (c)

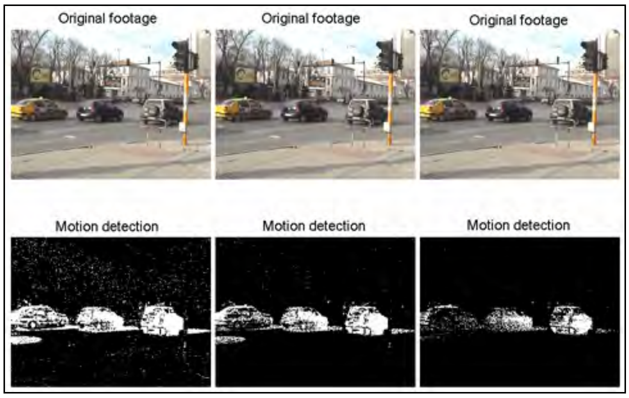
B. $\Sigma - \Delta$ background modelling

In Fig. 3 are shown the results of running the $\Sigma - \Delta$ background modeling method for the same footage as the previous method. The algorithm is executed for three amounts of the parameter C, Eq. (6). This parameter determines the speed of increasing the $\Sigma - \Delta$ variation and as consequence of that the resistance against noise and flickering background. In the original footage in Fig. 3 there are three cars moving from left to right. The first car is moving in area with low $\Sigma - \Delta$ variance and it is relatively equal represented in the three cases of motion detection, respectively Fig. 3a $C = 1$; Fig. 3b $C = 5$; Fig. 3c $C = 20$. As C goes up the second and third car are getting less visible in the motion detection image, Fig. 3b and Fig. 3c. This is determined that they are moving in the trace of the first car where $\Sigma - \Delta$ variance is high. The positive in this case is the high resistance against noise and flickering backgrounds. So, $\Sigma - \Delta$ background modeling method is not suitable for heavy traffic video estimating. But there is high precision and noise resistance when the moving objects passes in relatively long interval.



(a) *(b)* *(c)*

Fig. 3. $\Sigma - \Delta$ background modeling executed for three amounts of C , $C = 1$ (a); $C = 5$ (b); $C = 20$ (c)



(a) *(b)* *(c)*

Fig. 4. Mixture of Gaussians background subtraction method executed for three amounts of T , $T = 0.25$ (a); $T = 0.35$ (b); $T = 0.5$ (c)

C. Mixture of Gaussians

Mixture of Gaussian Background subtraction method is very precise, even the shadows of the moving objects are detected. This method is very complex to implement and configure and has relatively long running time. At every time step new distribution is created when the new observed pixel does not match previous Gaussian distribution. If this pixel belongs to a moving object which then goes away, the mean and variance of the previous distribution, which represented the true background, would not be changed. The algorithm is executed for three values of the threshold T , Eq. (11) and the results are shown in Fig. 4. High levels of the threshold determine high resistance against noise without serious corrupting of the moving objects estimation in the output motion detection image. In Fig. 4b and Fig.4c the noise is quite attenuated but the shapes of the foreground objects are still correctly represented.

IV. CONCLUSION

Three methods for motion detection using background subtraction were presented. The results of estimating real traffic surveillance video footage were proposed.

The conclusions about the precision of each method is analyzed according the purpose of implementation in practice.

The Temporal Averaging method is simple and do not require high computational power for maintains. Despite the simple algorithm, this method is surprisingly correct and resistant to noisy background, only if the learning rate α is correctly chosen. High levels of α are appropriate for scenes of fast moving objects. But in traffic surveillance sequences where most of the foreground objects are moving relatively slow, the amount of α should be near to 0.1.

The $\Sigma-\Delta$ background modeling method is a little bit complicated than the Temporal Averaging method, but it is not flexible enough when it is used for estimating heavy traffic video. Increasing the amount of the parameter C reduces the noise in the background but also creates a trace after the first moving object. This trace has high $\Sigma-\Delta$ variance, which determines low sensibility for motion in those pixels for the next consecutive frames. As a result there is a serious problem in estimating heavy traffic where most of the cars will be invisible.

Mixture of Gaussians is complex method for background subtraction. In our experiment this method shows highest precision. Even the shadows of the moving objects are detected. To achieve an optimal result using the mixture of Gaussians method there are many parameters to be adjusted and this is a disadvantage of this method. Another disadvantage is that it is computationally expensive and requires relatively large amount of memory to store, update and sort multiple Gaussian distributions.

The choice of the appropriate motion detection method depends on available computational power, specific surveillance conditions and requirements for processing time. Temporal Averaging and $\Sigma-\Delta$ background modeling are useful in real time systems, where is no need of high precision motion detection. For best precise results Mixture of Gaussians is a suitable decision but requires huge computational power.

ACKNOWLEDGEMENT

This paper is funded by National Ministry of Education and Science of Bulgaria under project "Investigation of Algorithms for Motion Detection in Motion Pictures", 2011, Technical University of Varna.

REFERENCES

[1] Janne Heikkilä and Olli Silvén. "A real-time system for monitoring of cyclists and pedestrians" Image and Vision Computing, Vol. 22, Issue 7, Visual Surveillance 1 July 2004 Pages: 563–570.

[2] Antoine Manzanera and Julien C. Richefeu. "A new motion detection algorithm based on sigma-delta background estimation". Pattern Recognition Letters In Advances in Visual information Processing, Vol.28 No.3 Pages: 320–328, 1 Feb 2007.

[3] C. Stauffer and W. E. L. Grimson. "Adaptive background mixture models for real-time tracking". In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol 2, page 252, 1999.

[4] Vesna Zeljkovic, Dragoljub Pokrajac. "Improved Illumination Independent Moving Object Detection Algorithm in Infrared Video Sequences" In Proceedings of ICEST Conference 2005, Nis, Serbia and Montenegro, pages 342-345, 2005.

[5] Cheung Sen-Ching and Chandrika Kamath. "Robust techniques for background subtraction in urban traffic video". Proceedings of Electronic Imaging: Visual Communications and Image Processing 2004 (Part One), January 20-22 2004, San Jose, California. Bellingham WA:SPIE, 5308, Pages: 881–892.

[6] Andrea Prati, Ivana Mikic, Costantino Grana, and Mohan M. Trivedi. "Shadow detection algorithms for traffic flow analysis: a comparative study". In Proceedings of IEEE Intelligent Transportation Systems Conference , pages 340–345, Aug 2001.

[7] Ahmed Elgammal, Ramani Duraiswami, David Harwood, Larry S. Davis, R. Duraiswami, and D. Harwood. "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance" In Proceedings of the IEEE, VOL. 90, No. 7, July 2002 Pages 1151–1163, 2002.

[8] Dar S. Lee. "Effective Gaussian mixture learning for video background subtraction", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27 Issue 5, Pages: 827–832, 2005.

[9] M. Vargas, S. L. Toral, F. Barrero, J.M. Milla. "An Enhanced Background Estimation Algorithm for Vehicle Detection in Urban Traffic Video", Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems Beijing, China, October 12-15, 2008, Pages 784 – 790.

[10] Nan Lu, Jihong Wang, Q.H. Wu and Li Yang. "An Improved Motion Detection Method for Real-Time Surveillance", Proceedings of International Journal of Computer Science, IAENG, 35:1, IJCS_35_1_16, January 2008.