Reducing the Foreground Aperture Problem in Mixture of Gaussians Based Motion Detection

Á. Utasi and L. Czúni
Department of Image Processing and Neurocomputing
University of Pannonia
Egyetem u. 10, Veszprém, H-8200, Hungary
Phone: (+36) 88 624800 Fax: (+36) 88 624606 E-mail: czuni@almos.vein.hu

Keywords: motion detection, Mixture of Gaussians, foreground aperture problem

Abstract – Separating the moving image parts from the static background is an important phase in video surveillance applications. The method based on Mixture of Gaussians (MOG) is an often used and robust approach to learn the background automatically and adaptively. Known MOG methods often suffer from the phenomena called the foreground aperture problem, when parts of large moving homogenous regions become part of the background instead of being selected as moving pixels. This article introduces a new method to eliminate this problem.

1. INTRODUCTION

In recent years background modeling approaches became widely applied in different video processing problems even in real-time applications thank to the increase of processing power of computers and the relative simplicity of these algorithms. In these methods the background is modeled by the recent values of image pixels and the pixels of the new incoming frames are compared to these model parameters. The different approaches apply different model creation and decision process to designate moving areas.

In the Pfinder algorithm [1] Gauss distribution is used to model the pixels of the background. This approach works well for static backgrounds but fails in other cases such as waving trees, flickering objects, or water. Grimson and Stauffer applied a Mixture of Gaussians (MOG) technique first where several (typically 3 to 5) Gaussian functions were modeling such backgrounds [2]. The clear advantage of MOG is the ability of modeling multimodal backgrounds. In the recent years some extensions increased the performance of this technique [3].

Toyama et al. built a system (called Wallflower) of three components for background-foreground separation:

- The color of background is estimated from the previous frame by Wiener filtering on the pixel-level.
- The inner area of moving objects is filled by region-level image processing.
- Sudden global changes (like illumination changes due to switching the light on in a room) is handled by frame-level processing.

In this paper we show that with the extension of the original MOG model we can minimize the effect of the foreground aperture problem with only pixel-level steps.

2. MIXTURE OF GAUSSIANS FOR BACKGROUND MODELING

MOG is based on the assumption that the observation of pixels in a video can be well modeled by several Gaussian distributions. Consider K Gaussian distributions and let x be the value of a pixel at time t. The probability of x is:

$$P(x_t) = \sum_{i=1}^{K} \omega_{i,t}\eta(x_t, \mu_{i,t}, \Sigma_{i,t}) \qquad (1)$$

where $\omega_{i,t}$ is the weight, $\mu_{i,t}$ is the expected value, and $\Sigma_{i,t}$ is the covariance of Gaussian distribution η:

$$\eta(x_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x_t-\mu)^T\Sigma^{-1}(x_t-\mu)} \qquad (2)$$

To achieve high computing speed the color channels are considered as independent so the covariance matrix is reduced to $\Sigma_{i,t} = \sigma_k^2 I$ (where k stands for the RGB color channels).

The algorithm has to decide if a new coming pixel belongs to the background or to the foreground. According to [2] if the value of a pixel is not within the range of $2.5\sigma_{avr}$ from the expected value of any distribution then the pixel is set as foreground. (σ_{avr} stands for the average of σ_k of the three channels.) Otherwise, we have to investigate how probable the matching distribution is described by weights. First we update $\omega_{i,t}$ for all distributions such as:

$$\omega_{i,t} = (1-\alpha)\omega_{i,t-1} + \alpha M_{i,t} \qquad (3)$$

where α is a learning factor and $M = \begin{cases} 1 & \text{if distribution matches the current pixel} \\ 0 & \text{otherwise} \end{cases}$.

Also $\mu_{i,t}$ and σ_k are updated such as:

$$\mu_t = (1-\rho)\mu_{t-1} + \rho x_t \qquad (4)$$

$$\sigma_t^2 = (1-\rho)\sigma_{t-1}^2 + \rho(x_t-\mu_t)^T(x_t-\mu_t) \qquad (5)$$

where $\rho = \alpha\eta(x_t, \mu_t, \Sigma_t)$. It is common to give ρ a constant value to spare computation time [4]. The background model consists of the first B distributions of the ordered list of distributions where the position is determined by the value of ω/σ_{avr}. B is calculated according to the following expression:

$$B = \arg\min_{b}\left(\sum_{i=1}^{b}\omega_i > T\right) \tag{6}$$

where threshold T defines the number of distributions contributing to the background model. For example setting T to a relatively small value can result in an unimodal background model.

Now, we have to see if the matching distribution belongs to any of the first B distributions. If it does not then the pixel is set to foreground otherwise it is considered as a background pixel.

For more details of the basic MOG algorithm see [2].

3. PROBLEMS WITH BACKGROUND MODELING

In [5] Toyama et al. investigate the practical problems of MOG based foreground-background separation during the development of Wallflower. Depending on the application the following problems can occur:

- intensity changes (e.g. clouds appear);
- switching light on/off (in closed places);
- shaking trees;
- learning time;
- foreground aperture: moving large homogenous areas become part of the background;
- shadow detection.

In their tests they compared their own method with the original algorithms [2]. They apply region-level histogram based processing to solve the aperture problem without giving the required computation time in their paper.

H. Wang and D. Suter [3] complemented the original MOG model and tested 3 distinct algorithms to investigate the problems described in Wallflower. Their results show improvement to several problems but they did not deal specifically with the foreground aperture problem.

4. MODIFIED MIXTURE OF GAUSSIANS MODELING

In our paper we introduce a new method to minimize the foreground aperture problem. The purpose during the design of our approach was to keep the robustness of the original MOG method, to avoid region-level image processing, and to use only pixel-level steps similar to the original algorithm. This way we made an extension where not only the background but also the foreground has its own model.

4.1 Modeling background and foreground

The modeling of the background is basically identical with the original approach. The only difference is when the deviation of the background models decreases below a threshold.

In these places when we find a foreground pixel we create a foreground model for it. We have only one Gaussian distribution $F(\mu_t, \sigma_t)$, where μ_t denotes the expected value and σ_t stands for the deviation at time *t*.

Initially, we set the expected value of the foreground to the color of the pixel and the deviance to a relatively large value (f. e. 20). A foreground pixel can turn to be a background one but only in that case if its variance becomes small. Modification of μ_t and σ_t is similar to the background model (see eq. 4 and 5) from frame to frame. However, if the color of a foreground pixel fits the background then its deviance should be reduced drastically by applying a large ρ value. In this process when the foreground pixel turns to be part of the background at time *t* it is initialized as given below:

$$B_j(\omega_t, \mu_t, \sigma_t) = B_j(\omega_{ini}, F(\mu_{t-1}, \sigma_{t-1})) \tag{7}$$

where B_j is the background component of B with the smallest weight.

4.2 Neighborhood effect among foreground pixels

In the previous section we introduced a Gaussian distribution to represent homogenous foreground areas. Now, we have to ensure that these foreground pixels do not become part of the background due to the foreground aperture effect. To ensure this we have to investigate if pixels of the foreground model have foreground neighbors with similar expected value and smaller deviation. (Note that this refers to large homogenous moving areas: as a large plain region moves its margins at the side of the direction of motion have relatively high deviation but it decreases towards to other parts of the plain region.) If the variance is within a range of the investigated pixel then we have to increase the neighbor's deviation. Define the process of deviance flooding the following way:

$$flood\left(F_{i,j}(\mu_t, \delta_t)\right)$$

where $F_{i,j}(\mu_t, \delta_t)$ is the foreground model with expected value $F_{i,j}(\mu_t)$ and deviation $F_{i,j}(\delta_t)$ at position (i, j) and at time t. Define the set of foreground pixels satisfying the above requirements in radius $\Delta = 1$ by:

$$NF_{i,j}^{\Delta} = \left\{ F_{x,y}(\mu_t, \delta_t) : \begin{array}{r} \left|F_{i,j}(\mu_t) - F_{x,y}(\mu_t)\right| < M_{\max}; \\ 0 < F_{i,j}(\delta_t) - F_{x,y}(\delta_t) < D_{\max} \end{array} \right\} \tag{8}$$

where $i - \Delta \le x \le i + \Delta$ and $j - \Delta \le y \le j + \Delta$.

Then for $\forall F_{x,y}(\mu_t, \delta_t) \in NF_{i,j}^{\Delta}$ let $F_{x,y}(\mu_t, \delta_t) = (F_{x,y}(\mu_t), F_{i,j}(\delta_t))$ and call $flood(F_{x,y}(\mu_t, \delta_t))$ recursively for neighboring foreground pixels. This means that the foreground model's deviation, over margins of homogenous regions, spreads to the inner areas of objects. D_{max} blocks this spreading at disoccluded parts of the image and M_{max} modifies the sensitivity of this process.

5. TEST RESULTS AND ANALYSIS

We extended the original method of [2] with the above algorithm and used several different test videos to investigate the performance. We expected that at large moving homogenous regions less pixels were learnt by the background model thus foreground objects were less ragged.

Besides our own test videos we have chosen some, with the permission of authors, from those used by Toyama et al. when testing Wallflower [5]. In *Fig.1.* we can see one input frame, the output of the original MOG method [2], the output frame after morphological closing and the result of the proposed technique. Other results can be downloaded from http://193.6.41.149:8888/mog/

As can be seen the foreground area remained more connected and kept the shape of the objects. It clearly outperforms the often used morphological post-processing and besides keeping the objects integrity, partly or slowly moving regions remains detected (see the chair in the image).

In *Fig. 2.* three other frames are shown where numerical comparison was also carried out. *Table 1.* summarizes the performance of the original [2] and the proposed method by counting the misclassified pixel (either false positive or false negative).

Original MOG	Proposed MOG
4350	1934
3840	1778
2815	1447

Table 1. Misclassified pixels in columns 3 and 4 of *Fig. 2.*

To compare computational complexity we implemented the original and improved algorithm in the same Windows framework. Processing images with the resolution of 320x240 on a 3GHz Intel CPU PC the frame rate dropped from 12FPS to approximately 8FPS. In our experiments we used three background models and used no optimized *C* code. Here we also note that our approach does not need iterative optimization such as Markov Random Fields or similar techniques [6,7].

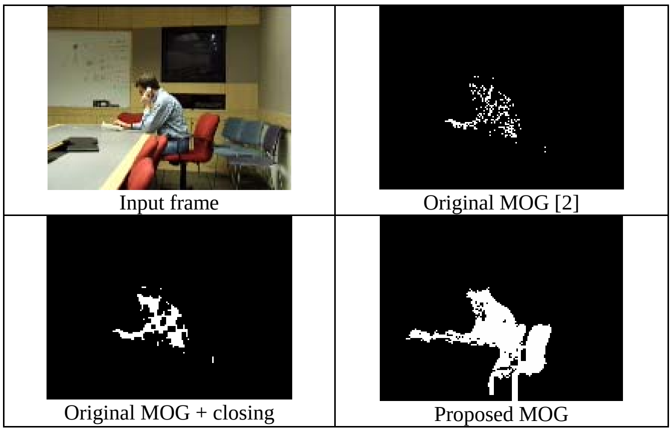


Fig. 1. Visual comparison of one frame from Wallflower [3].

6. CONCLUSION

In our paper we discussed the basics of MOG based motion detection. According to our experiences and to other papers the foreground aperture problem still seemed to be an important problem in motion detection thus we developed a new method to minimize this problem.

Compared to other techniques our approach reduces the undetected areas drastically, preserves the shape of the objects, detects weakly moving objects but decreases the frame rate with approximately 30%. This is achieved with an improvement of the original pixel-level MOG model, and without region-level image processing or iterative optimization. To illustrate our results we made several test videos, made numerical comparison and placed test videos on the internet.

7. ACKNOWLEDGEMENT

The authors would like to thank the help of Attila Licsár and the support of the GVOP-3.1.1.-2004-05-0388/3.0 national project.

REFERENCES

[1] C. R. Wren, et. al., "Pfinder: Real-Time Tracking of the Human Body", IEEE Trans. on PAMI, vol 19, no 7, p. 780-785, 1997.

[2] C. Stauffer and W.E.L. Grimson, "Adaptive Background Mixture Models for Real-time Tracking", CVPR, p. 246-252, 1999.

[3] H. Wang and D. Suter, "A Re-Evaluation of Mixture-of-Gaussian Background Modeling", ICASSP, p. 1017-1020, 2005.

[4] C. Stauffer and W.E.L. Grimson, "Learning Patterns of Activity Using Real-time Tracking", IEEE Trans. on PAMI, 22(8): 747-757, 2000.

[5] K. Toyoma, et. al., "Wallflower: Principles and Practice of Background Maintenance", ICCV, p. 255-261, 1999.

[6] Y. Wang and Q. Ji, "A Dynamic Conditional Random Field Model for Object Segmentation in Image Sequences", CVPR, vol. 1, p. 264-270, 2005.

[7] Cs. Benedek, T. Szirányi, "Markovian Framework for Foreground-Background-Shadow Separation of Real World Video Scenes", ACCVC, Lecture Notes in Computer Science, 2006.

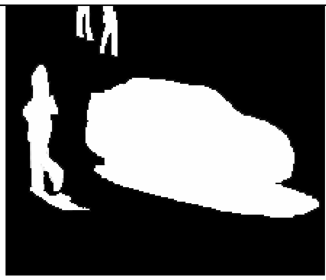
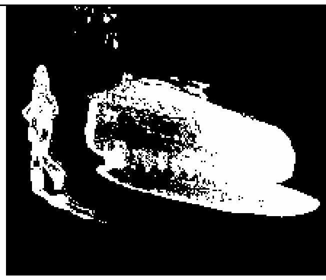
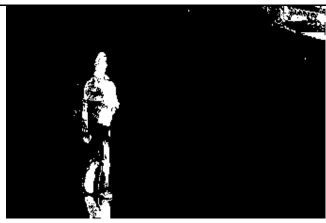
Original input	Ground truth (hand segmented)	Original MOG [2]	Proposed MOG
			
			
			

Fig. 2. Comparison of classical MOG and the proposed method.