Improved Post-Processing for GMM based Adaptive Background Modeling

Deniz Turdu, Hakan Erdogan

Faculty of Engineering and Natural Sciences
Sabanci University, Orhanlı Tuzla 34956, Istanbul, Turkey
denizturdu@su.sabanciuniv.edu, haerdogan@sabanciuniv.edu

Abstract

In this paper, we propose a new post-processing method for Gaussian mixture model (GMM) based adaptive background modeling which was proposed by Stauffer and Grimson. This is a ubiquitous and successful background modeling method. A drawback of this method is that it assumes independence of pixels and relies solely on the difference between current pixel value and its past values. This causes some errors within the foreground region and results in fragmentation of foreground objects detected. Our method uses relaxed-thresholding and adds foreground edge information in close proximity of detected foreground blobs. The close proximity is obtained as the union of convex hulls of close-by regions which we call the hysteresis region. Our results show that we can achieve increased recall rate with the proposed method without much decreasing the precision of the conventional method.

1. Introduction

In robust video surveillance applications, segmentation of foreground and background is a primary concern. Since the cameras are stationary in such applications, background modeling based foreground detection methods are widely used [4]. Such a method is composed of two main parts: modeling the background and detecting the foreground. In the first part, a background model is determined and in the second part, by comparing that background model to the current frame, the foreground objects are detected.

Background can be modeled using adaptive or non-adaptive techniques. Non-adaptive techniques consider a static frame as the background model. These techniques are expected to fail in situations where the background is subject to some changes such as changes in illumination due to the time of the day. These techniques also fail when there are some moving parts in the background, like leaves of a tree in the background observed on a windy day [3]. For this reason, robust foreground extraction methods require updating the background model continuously.

A simple adaptive background technique is to take a temporal average of the frames and to consider this average as the background model. This technique fails when the motion of foreground objects is relatively slow, since those objects will be integrated heavily into the averaged background. Also, when the number of moving foreground objects increase and the background becomes less visible, this technique is expected to forget the image of the background. In addition, this technique has a particularly long recovery time for the background. The Kalman filter approach in [6] deals with the sudden illumination changes in the background, but it is also subject to this slow background recovery problem. The method described in [5] does not have this problem; it uses a single Gaussian to represent the distribution of background pixels. The method in [5] is suboptimal as compared to the Stauffer-Grimson method used in [1] which uses a Gaussian mixture model to represent a background pixel's distribution.

The method proposed in this paper mainly depends on the Stauffer-Grimson method, which is successful on solving the problems aforementioned. The Stauffer-Grimson method has been one of the most successful methods in the algorithm competition of VSSN'06 conference [2]. In this method, more than one Gaussian mixture components for a pixel are used. First few of the highest weighted mixture components assigned to a pixel are taken to be the background distributions for that pixel. The sum of the weights for the background distribution components should be above a threshold. Thus, the Gaussian mixture component with the highest weight is always assumed to be a part of the background distribution. In addition to that, since Stauffer-Grimson method analyzes each pixel independently from the others and the values observed are only the color values, some of the single piece foreground objects are detected as many separate smaller objects. Especially when some parts of the moving foreground object has color values similar to the colors of the background behind the object, the foreground object will partially be considered

as background and as a result, this single piece foreground object will be fragmented. To overcome this problem confronted in Stauffer-Grimson method, a secondary thresholding with a relaxed threshold in a hysteresis search region is performed and foreground edge extraction is applied in our proposed method.

This paper is organized as follows: In section 2, the Stauffer-Grimson method [1] will be explained. Standard post-processing operations performed on the foreground images will be mentioned in section 3. Section 4 will consist of the way of finding a hysteresis search region on the foreground, applying relaxed thresholding on that region, and the use of foreground edge segmentation besides these. Experimental results are shown in section 5, and we are going to conclude in section 6 presenting our future plans on the method.

2. GMM for the Background

In Stauffer-Grimson method, a GMM per pixel is used to model the background, thus a pixel is assumed to be independent from its spatial neighbors. It is considered that background image may change due to different lighting conditions and small movements. The Gaussian mixture components for a pixel have normalized weights calculated from the past observations. For simplicity, the image retrieved is assumed to be a 3 channel image and all the channels are independent, having the same variance values within themselves. The likelihood that a pixel has a value of $\boldsymbol{X_t}$ is assumed to be:

$$P(X_t) = \sum_{i=1}^{K} w_{i,t}\ \eta(X_t, \mu_{i,t}, \Sigma_{i,t})\ . \qquad (1)$$

Here $\boldsymbol{w_{i,t}}$ is the weight of the $\mathbf{i^{th}}$ Gaussian distribution within all $\boldsymbol{K}$ distributions assigned for that pixel. $\boldsymbol{\mu_{i,t}}$ is the mean value of the $\mathbf{i^{th}}$ mixture component and $\boldsymbol{\Sigma_{i,t}}$ is the covariance matrix of the component at time $\boldsymbol{t}$. In this equation, $\boldsymbol{\eta}$ shows a multivariate Gaussian density function. Related to the assumption in the model that the color channels are independent with equal inner variances, the covariance matrix is a diagonal matrix in the form $\Sigma_{i,t} = \sigma_k^2 I$. Each diagonal element in this matrix is equal to the variance of a single channel which is $\boldsymbol{\sigma_k^2}$.

The parameters of the mixture components are updated with new frames. A retrieved pixel value is compared with all the components of the mixture assigned to that pixel to find out if there is a match. A match is said to happen when the retrieved pixel value is within 2.5 times standard deviation of a mixture component. The update procedure is different for the matching component and other components. The mean values and the covariance matrices are updated for only the matching component. The update formulas for the matching component are given below:

$$\begin{aligned} \mu_t &= (1-\rho)\mu_{t-1} + \rho X_t \\ \sigma_t^2 &= (1-\rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T (X_t - \mu_t) \end{aligned} \qquad (2)$$

For both matching and not matching components, the update for the weights is done by,

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

where $\boldsymbol{M_{k,t}}$ is 1 for the matching component and 0 for the others, thus weight of the matching component is relatively increased. After updating the weights for all components, weights are re-normalized.

In update formulas, $\boldsymbol{\alpha}$ and $\boldsymbol{\rho}$ are the learning parameters determined experimentally in this paper. These parameters can be altered appropriately for different purposes of use [1].

In case there is not a match between the current pixel value and the mixture components related to that pixel, the component having the smallest likelihood with respect to the current pixel value is discarded. A new Gaussian component is created in place of the discarded one, having a mean value equal to the current pixel value, and a variance equal to a predetermined constant. Therefore, random short term changes in the image cannot form a dominant component in the model.

While the retrieved frames are processed in a pixel wise manner, each pixel is checked whether it is a part of the background or not. The whole process work as follows:

for $i, j =$ all pixels

Update GMM model

Sort Gaussian components according to weights

$$w_1 > w_2 > \ldots > w_K$$

$$B = \arg\min_b \left(\sum_{i=1}^{b} w_i > T_1\right)$$

$m =$ index of matched component

$$F_1(i,j) = \begin{cases} 0 & \text{if } m \le B \\ 1 & \text{if } m > B \end{cases}$$

end for

where $\boldsymbol{T_1}$ is a threshold value for the sum of the weights of the background components. From all Gaussian components belonging to a pixel, first $\boldsymbol{B}$ components with the highest weights are considered as background

components for that pixel, and the sum of their weights should be greater then the threshold value T_I. If the retrieved pixel value is matching one of its background components, then that pixel is assumed to be a background pixel. Otherwise, it is assumed that there is a foreground object on that pixel. In the end, after all pixels are checked in this way, the background hence the foreground (F_I) for the current frame is determined.

We examined the Stauffer-Grimson GMM method which forms the foundation of the technique proposed in this paper. In this method, some single piece foreground objects are detected as smaller separate objects, and the mixture component with the highest weight is always considered to be a background component. Thus, the method could be improved in terms of these problems. In the following section, the standard post-processing methods applied to F_I are discussed.

3. Standard Post-Processing

After the foreground image F_I is retrieved, it is subject to some post processing operations:

3.1. Opening and Closing

In the foreground, there are usually many small holes in the detected objects and there are also some small objects detected which are not really objects, but occur due to the noise and the dynamic background. Since opening smoothes the contour of an object by eliminating protrusions and closing smoothes the contours by filling the gaps and holes on the contours [7], an opening and a closing operation are performed as standard post-processing morphological operations.

3.2. Connected Component Analysis

A connected component analysis is performed on F_I after the morphological operations. The connected regions are obtained and the contours $\{C_1, C_2, \ldots, C_K\}$ of these regions on the foreground image are found. In this work, we only take the external contours into consideration, it is assumed that the foreground objects will not have a hollow structure.

3.3. Minimum Area Filtering

Let A_{min} be the minimum area value determined experimentally. For an object with external contour C having a pixel wise area of A_C, the object will be filtered out of F_I if $A_C < A_{min}$. Thus very small pieces that were not eliminated through the morphological operations and that are too small to be foreground objects or parts of foreground objects are left out.

After the standard post-processing is finished, the set of contours surrounding the foreground regions is formed and an updated foreground image $\hat{F}_I$ is formed by union of interiors of these contours.

These post-processing steps are not able to prevent the fragmentation problem in Stauffer-Grimson method, thus a secondary relaxed thresholding in a specific hysteresis region is used together with the foreground edge information afterwards. These additional techniques will be explained in detail in the following section.

4. Hysteresis Thresholding

To reduce the fragmentation of foreground objects in Stauffer-Grimson method, we introduce an improved method that uses hysteresis thresholding and foreground edge detection.

4.1. Hysteresis Search Region

Let $d_{min}(C_n, C_m)$ be the minimum distance between two object contours C_n and C_m on $\hat{F}_I$. We iterate through all contour pairs and if the distance between two contours is less than the threshold D_{max}, we find the interior of the convex hull of those two contours, $\mathbf{H_{nm}}$. Then the union of such convex hull regions is taken to form a mask where the relaxed threshold will be applied. This union image operates as a mask combining pairs of regions that have a high probability for belonging to the same single region foreground object. Assuming there are N_R contours on $\hat{F}_I$, the process below is performed:

$H(i,j) = 0 \quad \text{for all } i, j.$
for $n = 1, \ldots, N_R$
  for $m = n+1, \ldots, N_R$
    $D = \text{distance}(C_n, C_m)$
    if $D < D_{\max}$
      $H_{nm} = \text{convexhull}(C_n \cup C_m)$
      $H = H \vee H_{nm}$
    end if
  end for m
end for n

At the end of this process, the hysteresis search region indicator image H is obtained. It can be considered as a foreground image with the union of convex hulls of close contours in $\hat{F}_I$. The relaxed thresholding applied in this mask is presented in the following section.

4.2. Relaxed Thresholding

For the pixels in H, a secondary thresholding with a lower threshold value than T_1 is applied. In Stauffer-Grimson method, the sum of the weights of the Gaussian mixture components are compared to the single threshold value T_1 and the primary foreground image F_1 is formed. This causes the highest weighted Gaussian component to be considered as background regardless of its weight. This may cause problems when there is a highly variable background and the current foreground pixel value is close to one of the background Gaussian components (which becomes more likely due to the variable background). Because of this, in relaxed thresholding, only the weight of the matching Gaussian component is taken into consideration. Let the weight of the matching Gaussian be w_m and T_2 be the relaxed threshold value, then the relaxed thresholding is applied as:

for $i, j = \text{all pixels}$

$m = \text{index of matched component}$

$$F_2(i,j) = \begin{cases} 1 & \text{if } w_m \le T_2 \\ 0 & \text{if } w_m > T_2 \end{cases}$$

end for

Standard post-processing is also applied to F_2 to obtain $\hat{F}_2$. There is an additional dilation step after the opening and closing operations[1]. This relaxed thresholded foreground mask $\hat{F}_2$ is then considered only inside H. This is done by a "logical and" operation between H and $\hat{F}_2$ ($U = H \wedge \hat{F}_2$)[2]. Thus, the background pixels positioned between two detected foreground objects in $\hat{F}_1$ with a relatively low weight are also marked as foreground in the mask image U. This relaxed thresholding reduces fragmentation in $\hat{F}_1$. After the improved foreground candidate U is obtained, the edge change information in the foreground is also used for further improvement as we explain in the next section.

4.3. Foreground Edge Extraction

A drawback of Stauffer-Grimson method is that the pixels are assumed to be independent and neighborhood information is not used in determining foreground regions. The gradient operator uses neighbors of a pixel to determine spatial derivatives of the intensity image. The information in change of gradients in the background versus the foreground should be complementary to the color change information used in Stauffer-Grimson method. Thus, in addition to relaxed thresholding, we also employ a foreground edge detection algorithm to determine foreground edge pixels inside the hysteresis search region. Foreground edges will mark only the edges of foreground objects. We consider only external contours to cover foreground objects, thus no hollow foreground objects are allowed. Because of this assumption, correctly detected foreground edge points will possibly aid in determining the whole foreground object as a single object.

Foreground edges are determined in the following fashion. Let the intensity image at current time t be P and consider that the gradient history images in the horizontal and vertical directions up to time t are $\Lambda_{x,t}$ and $\Lambda_{y,t}$ respectively. Then foreground edge detection is realized by the following algorithm:

apply Gaussian smoothing on P

$$G_x = \frac{\delta P}{\delta x},\ G_y = \frac{\delta P}{\delta y}$$

Update gradient histories as:

$$\Lambda_{x,t} = (1-\alpha_e)\Lambda_{x,(t-1)} + \alpha_e G_x$$
$$\Lambda_{y,t} = (1-\alpha_e)\Lambda_{y,(t-1)} + \alpha_e G_y$$

$$D_{edge} = \sqrt{\left(G_x - \Lambda_{x,t}\right)^2 + \left(G_y - \Lambda_{y,t}\right)^2}$$

for i, j = all pixels in E_{fg}

$$E_{fg}(i,j) = \begin{cases} 0 & \text{if } D_{edge}(i,j) < T_{edge} \\ 1 & \text{otherwise} \end{cases}$$

endfor

In this procedure, α_e is the edge learning parameter found experimentally, and T_{edge} is the edge threshold parameter. To find directional gradients on the current frame, a Sobel operator with a 3x3 kernel is used. For the x-derivative and for the y-derivative, kernels used are:

---

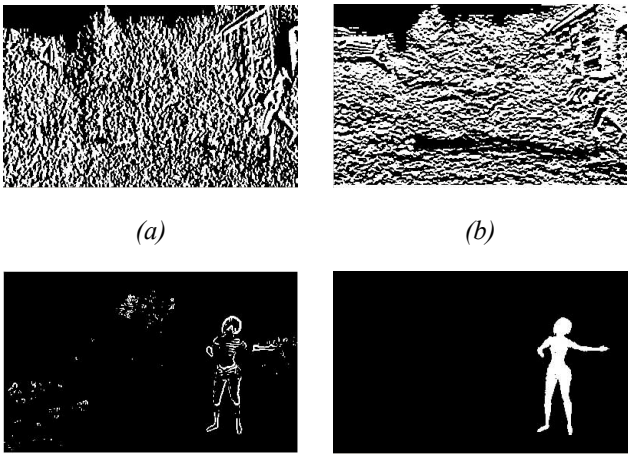
[1] Inside the hysteresis region, the probability of pixels being in the foreground is higher. Thus, we perform an additional dilation step. An additional dilation step added to the original method (in the entire image) would result in intolerably lower precision since there is no safety net of the hysteresis region.

[2] Note that, performing the relaxed thresholding inside hysteresis region H is mathematically equivalent to performing it for the entire image and then ANDing the result with H. Thus, we use two explanations interchangeably. The computational load of the two alternatives can be different but negligible.

$$\begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{pmatrix} \text{ respectively.}$$

We also experimented with the Laplacian operator instead of the gradient images; but our experiments show that Laplacian operator results in a slightly worse foreground change detection than the procedure above.

This foreground edge information is then used as follows. A pixel inside the hysteresis search region H is considered as a foreground pixel if it passes the relaxed threshold test or it is found as a foreground edge using the test above. In other words, we form a new foreground mask image by the following operation $\boldsymbol{U_2 = H \wedge (F_2 \vee E_{fg})}$. In Figure 1, edge images for a random frame taken from a test video can be seen.



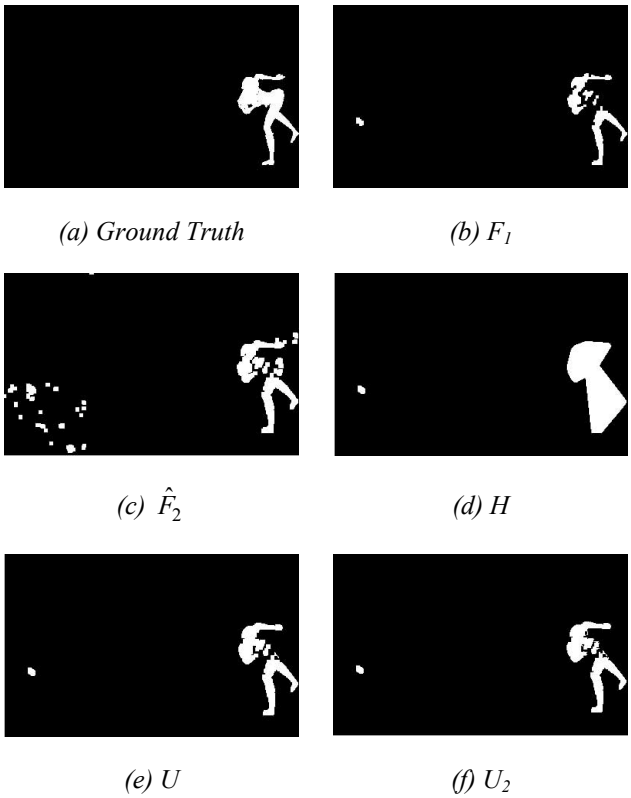
(a) *(b)*

(c) *(d)*

Figure 1: (a) G_x (b) G_y (c) Foreground edge mask E (d) Ground truth

In the Figure 1 (c), the edges of the foreground object are detected from the spatial gradients. Also, some of the background edges are included in the detected foreground edge in Figure 1 (c), because of the motion in the background. But when this resulting edge image is combined with the relaxed thresholding results, it enhances the foreground segmentation performance.

Also, as clearly seen in Figure 2 (b), the foreground object is detected in many smaller separated parts by Stauffer-Grimson method. This fragmentation problem is mainly reduced in the proposed method and the improved result is seen in Figure 2 (f).



(a) Ground Truth *(b) F_1*

(c) $\hat{F}_2$ *(d) H*

(e) U *(f) U_2*

Figure 2: (a) Ground truth (b) Primary foreground formed by Stauffer-Grimson method (c) Relaxed thresholded foreground (d) Union of convex hulls (e) U, the result of Logical AND operation on H and $\hat{F}_2$ (f) Result of logical OR operation on U and E_{fg}

5. Experimental Results

All experiments were performed using the application programming interface prepared for the algorithm competition in VSSN'06 (International Workshop on Video Surveillance & Sensor) by ACM.

During an initial background learning period, the parameters in the results were not taken into consideration. For video files Vid1 and Vid2 which have lengths less than 1 minute, the length of this learning period was 10 seconds. For the longer video in Vid3 file, this period was 25 seconds long. This rule has been taken from VSSN'06 competition rules.

Experiments were realized on a PC with Intel Celeron® 1.7 GHz Mobile Processor, 480 MB of RAM on Microsoft Windows® XP. The results are summarized in the tables below.

	$\hat{F}_1$	***H***	***U***	U_2
Precision (%)	45,43	28,07	45,35	45,44
Recall (%)	84,61	95,09	90,39	90,75
Avg.FalseAlarm (pel)	1219	3787	1571	1579
Avg. Miss (pel)	358	114	224	215
FPS	16	14	14	14

Table 1: Comparison of methods on video 1

	Stauffer-Grimson		Proposed Method	
	Vid2	Vid3	Vid2	Vid3
Precision (%)	48,43	33,17	47,00	32,94
Recall (%)	90,96	96,30	94,48	97,67
Avg.FalseAlarm	2746	3084	3011	3461
Avg. Miss	515	159	286	100

Table 2: Comparison of the base method with the improved one. Vid2 file consists of a video without an only background frame. Vid3 file consists of sudden illumination changes.

For the results above, all experimental threshold values used were:

$T_1 = 0.7$, $T_2 = 0.5$, $T_{edge} = 200$, $D_{max} = 10$, $A_{min} = 100$

In Table 1, the base method itself is compared with different combinations of additional techniques introduced in this paper. Video file of length 32 seconds (Vid1) including dynamic background and illumination changes in the background is used for this comparison. First column is the base method, Stauffer-Grimson only. Second column (marked ***H***) considers every pixel inside the hysteresis search region. Third column (marked ***U***) involves use of only relaxed thresholding in the search region, without the edge information. Last column (marked U_2) is the method in which both relaxed thresholding and foreground edge extraction are used inside the search region. Best performance in terms of precision and recall are achieved by incorporating relaxed thresholding and foreground edge information in hysteresis region.

Table 2 shows the results of comparing the performances of the base method and the improved method on videos of different characteristics. In these comparisons, although the precision is nearly the same, the improved method is better than the base method in terms of recall with a rate of %1-4.

6. Conclusions and Future Work

This paper shows that integrating a type of hysteresis thresholding on a mask of union of convex hulls and using foreground edge extraction enhances the segmentation of the foreground by reducing the foreground fragmentation and increasing the recall without a significant change in the precision.

In the future work, we are planning to integrate the foreground edge information to the model using another GMM.

References

[1] Stauffer C, Grimson W. E. L., "Adaptive background mixture models for real-time tracking," in Proceedings. 1999 *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (Cat. No PR00149). IEEE Comput. Soc. Part Vol. 2, 1999.

[2] http://mmc36.informatik.uni-augsburg.de/media-wiki/data/6/65/VSSN-Algo.pdf

[3] P. KaewTraKulPong and R. Bowden, "An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection," chapter 11, pages 135--144. *Video-Based Surveillance Systems*. Kluwer Academic Publishers, Boston, 2002.

[4] Wei Xu, Yue Zhou, Yihong Gong, and Hai Tao, "Background modeling using time dependent Markov random field with image pyramid," in *Proc. IEEE Motion'05*, January 2005.

[5] Christopher Richard Wren , Ali Azarbayejani , Trevor Darrell , Alex Paul Pentland, "Pfinder: Real-Time Tracking of the Human Body," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* v.19 n.7, p.780-785, July 1997

[6] C. Ridder, O. Munkelt & H. Kirchner, "Adaptive background estimation and foreground detection using Kalman filtering," in *Proc. ICAM*, 193--199, 1995.

[7] Gonzalez, Rafael C. "Digital image processing using MATLAB," chapter 9. Upper Saddle River, N. J. Pearson Prentice Hall, c2004.