MVA2009 IAPR Conference on Machine Vision Applications, May 20-22, 2009, Yokohama, JAPAN

Robust Background Segmentation using Background Models for Surveillance Application

Tianci Huang, Jingbang Qiu, Takahiro Sakayori, Takeshi Ikenaga
Graduate School of Information, Production and System, Waseda University
N355, 2-7, Hibikino, Wakamatsu, Kitakyushu, 808-0135, Japan
E-mail: bond0060@ruri.waseda.jp

Abstract

Gaussian Mixture Models (GMM) is a very typical method for background subtraction because it possesses a strong resistibility to repetitive background motion. However when it comes to complex environment, some unexpected situations occur, e.g., when illumination changes, gradually or quickly, segmentation is generated with a poor result. Moreover, this method is not capable of distinguishing shadows of moving objects. In this paper features of intensity and texture information are utilized to eliminate the shadow of moving objects. Integrated with modified Gaussian mixture models by redefining the update criterion, proposed algorithm is adapted to the flexible illumination environment. To validate that the proposed algorithm is robust to apply on surveillance system, we provide a metric with set of variables for evaluation, a comparison had been made between proposal and original GMM, results show the accuracy improvement of models using our updated algorithm. Averagely at least of 34.8% decrease of false alarm rate proves the quality of segmentation has been significantly enhanced and proposal is more competent and stable for outdoor surveillance applications.

1. Introduction

Efficient and reliable detection of moving objects in video steams is significant issue for surveillance systems. Background subtraction is a conventional and effective solution to segment the moving objects from the stationary background. But in an actual scene, the complex background such as snowy or windy conditions, makes the conventional algorithm unfit for the real surveillance systems. Stauffer and Grimson [3,6] modeled each pixel as a mixture of Gaussians and an online EM Algorithm by P. KaewTraKulPong et al. [7] to update the model. Even through 3 to 5 Gaussian distributions are capable of modeling a multimodal background, there is a fact that this kind of pixel-based background modeling is sensitive to noise and illumination change, a lot of efforts made to modify the model or integrate other works with the model to make the GMM suitable for complex scene.

In [4], Javed et al. presented a number of important problems when using background subtraction algorithms, they proposed a solution using pixel, region and frame level processing, their algorithm is able to deal with quick illumination changes, but their technique is based on a complex gradients-based algorithm. Huwer et al. [2] proposed a method of combining a temporal difference method with an adaptive background model subtraction scheme to deal with lighting changes. Cucchiara et al. [5] analyzed the foreground as moving objects, shadow, and ghost by combining the motion information. Whereas the work is effective to solve the problems they mentioned, the computation cost is relatively expensive for real-time video surveillance systems because of the computation of optical flow.

In this paper, we propose to utilize a training parameter to cooperate with the GMM foreground classification as well as update criterions. The intervention of this parameter could prevent the process of gradual illumination changes from learning into the Gaussian models. Texture and color information based on block region are utilized to generate features which are capable of adjusting to quick light changes and removing shadow of moving objects.

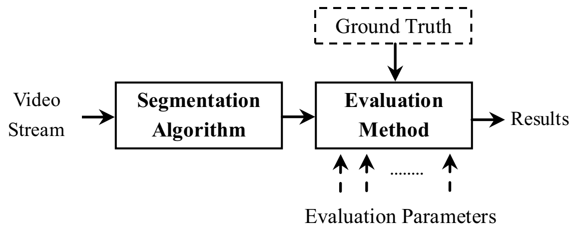


Figure 1. Typical Segmentation Evaluation Systems

A typical approach, shown as Fig. 1, to evaluate performance of motion detection is to introduce ground truth to provide independent and objective data, such as location and size, which can be related to the observations extracted from the video sequences. We evaluate the Tracking Detection Rate (TRDR) based on object metric compared with ground truth, also we compute the number of pixels for False Alarm Rate (FAR).

The remaining part of the paper is arranged as follows. Conventional GMM is briefly introduced in Section 2. Our proposal is specified in detail in Section 3. And in Section 4 performance evaluation metric is presented and experimental results are shown as well. Finally the conclusion is drawn in Section 5.

2. Gaussian mixture models

Stauffer and Grimson [3,6] introduced a mixture of K Gaussians (K from 3 to 5) for background subtraction. For a pixel X at time t, the probability of the pixel can be written as (1):

$$p(X_t) = \sum_{i=1}^{K} w_{i,t}\, \eta(X_t, \vec{u}_{i,t}, U_{i,t}) \quad (1)$$

in which

$$\eta(X_t; \vec{u}, U) = \frac{1}{(2\pi)^{\frac{n}{2}}|U|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t-\vec{u})^T U^{-1}(X_t-\vec{u})} \quad (2)$$

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

where $\vec{u}_{i,t}$ is the mean, α is the learning rate and $M_{k,t}$ is 1 for the model which matched and 0 for the remaining models. By assuming the red, green, and blue pixel values are independent and have the same variances. After the Gaussians are ordered by the value of w/σ, the first B distributions are chosen as the background model.

$$B = arg\ min\left(\sum_{j=1}^{b} w_k > T\right) \quad (4)$$

where T is a threshold that represents the minimum prior probability that the background is in the scene. In the paper, these parameters are set as α = 0.003, K = 3 and T = 0.4.

3. Proposed gaussian model

As the Row.2 in Fig. 2 shows, the mixture of Gaussian models is efficient to get rid of the periodic motions from a cluttered background, slow moving objects, long term scene change, and camera noises, but the method could not adjust to gradual light change for long time like sunset time, or quick light change, and it could not handle shadows either. We will discuss our solutions for the mentioned problems in this section.

3.1. Texture and intensity features

The GMM foreground mask shows as the right one in Row.2 of Fig. 2, a large areas of false positive foreground when there are quick light changes. GMM is not able to deal with this situation is the way it process on RGB color space, the model it utilizes can not handle the variability caused by illumination.

As to adjust the GMM to the quick light change enviroment for surveillance systems, we integrate the texture information to the foreground mask to remove this kind of false positive area. For the reason why GMM is not capable of removing shadow is that no heuristic exists to label Gaussian component as shadow. As the Row.2 of Fig. 3 shows, the shadow of moving objest exists, our solution is to utilize the intensity information to make the classification .

During the process of Gaussian model update process, current pixels in frame are used to training the correstponding pixels in background mask thus we can maintain a average background.

$$B_{pixel,i} = (1-\alpha) * B_{pixel,i-1} + \alpha * Cur_{pixel,i} \quad (5)$$

Where $B_{pixel,i}$ stands for the background pixel value at frame i, and $Cur_{pixel,i}$ is the corresponding pixel va-

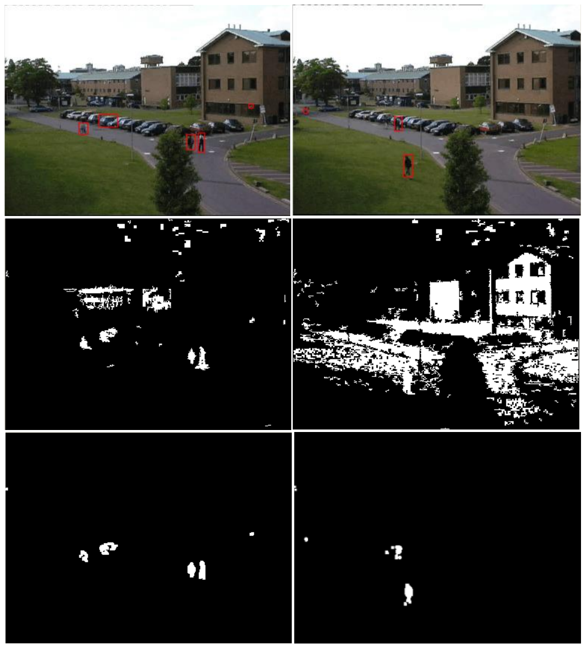


Figure 2. Row1: Original frame and detection results in red rectangles. Row2: Segmentation mask by GMM. Row3: Segmentation mask by Proposal.

lue at frame i, α is the same value in GMM.

Here we define two features as below :

$$F_1(X) = \frac{\sum_{i\in K} 2\cdot\|G_{Cur}(i)\|\cdot\|G_b(i)\|\cos\theta}{\sum_{i\in K}(\|G_{cur}(i)\|^2+\|G_b(i)\|^2)} \quad (6)$$

$$F_2(X) = \frac{\sum_{i\in K} I_{cur}(i)\cdot I_b(i) - \frac{1}{MN}\sum_{i\in K} I_{cur}(i)\cdot\sum_{i\in K} I_b(i)}{\sqrt{S(cur)\cdot S(b)}} \quad (7)$$

in which

$$S(x) = \sum_{i\in K} I^2{}_x(i) - \frac{1}{\mathrm{MN}}\left(\sum_{i\in K} I_x(i)\right)^2 \quad (8)$$

where K denotes the MxN neighborhood centered at the pixel X, $G_{Cur}(i)$ and $G_b(i)$ denotes gradient vector of current frame and averaging background mask, $I_{cur}(i)$ and $I_b(i)$ stands for the intensity at pixel i of current frame and background mask respectively, θ denotes the angle between $G_{Cur}(i)$ and $G_b(i)$. Gradient vector is obtained by laplace operator. When illumination change occurs in scene, texture of between current frame and background nearly keep same. In this way, we define the area of texture similariy at pixel X, as $F_1(X)$, if $F_1(X) > T_1$, the foreground block will be classified as background. In the proposed skeme similarity threshold T_1 is set as 0.7.

And in (7), we apply the intensity information to intervene foreground classification, the normalized cross-correlation $F_2(X)$ of intensity is calculated at pixels of foregroud area between current frame and averaging background mask. The foreground pixel is the shadow if $F_2(X) > T_2$, and T_2 are set as 0.7.

3.2. Improved gaussian mixture model

To prevent the process of gradual illumination change,

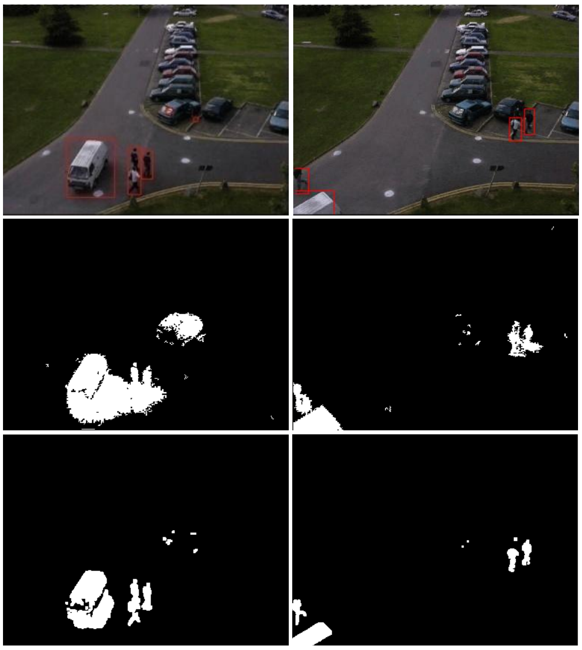


Figure 3. Row1: Original frame and detection results in red rectangles. Row2: Segmentation mask by GMM. Row3: Segmentation mask by Proposal.

e.g., sunset or sunrise scene, from being learning into the gaussian models, we introduce a training variable as $t_{pixel,j}$.

$$t_{pixel,j} = (1-\alpha) * t_{pixel,j-1} + \alpha * M * cur_{pixel} \quad (9)$$

$$\mathrm{M} = \begin{cases} 0, & if\ cur_{pixel} - pre_{pixel} > T \\ 1, & if\ cur_{pixel} - pre_{pixel} \le T \end{cases} \quad (10)$$

$t_{pixel,j}$ stands for the training pixel of each Gaussian model in frame j, and α is the learning rate. cur_{pixel} and pre_{pixel} respectively stands for the values of current pixel and previous pixel. Through this way, it results in a better background maintenance.

As Fig. 4 shows, in matching step the current pixel should belong to foreground or background, the criterion is not only by GMM, F_1 and F_2 as well as the $t_{pixel,j}$ intervene the classification criteria during model update.

4. Performance evaluation metric

4.1. Ground Truth

Evaluation based on Ground Truth (GT) offers a framework for objective comparison of performance of alternate surveillance algorithm. This kind of evaluation techniques compare the output of the algorithm with the GT obtained manually by drawing bounding boxes around objects, or marking up the pixel boundary of objects, or labeling objects of interest in original video stream.

We utilize the Pets sequence (including 4 sequences for outdoor) provided by IBM Research with 30 fps and a resolution of 384x288, to validate our proposed algorithm. Meanwhile, a comparison with the mask of conventional GMM is presented to show our approach significantly de-

```
While ( each pixel x in current frame) do
    If( cur_pixel – t_pixel < Tc ) then
        T_pixel_match =background;
    Else T_pixel_match = foreground;

    Model_Match =GMM_Match_Test ( cur_pixel )

    If ( Model_Match = No_match
        && T_pixel_match =foreground
        && F1(x)<T1 && F2(x)<T2 ) then
        Cur_Pixel = foreground;
        Update_NoMatch ( Model_Pixel, T_pixel_match );

    Else
        Cur_Pixel = background;
        Update_Match ( Model_Pixel, T_pixel_match);
end
```

Figure 4. Proposed Background Modeling Update Criterion In Pseudocode

creases the false motion detection and improve the segmentation quality.

To show the performance of improved background model, high level processes such as noise cleaning or connected component analysis algorithms were not introduced to the results of background subtractions.

4.2. Evaluation metric

Improved background modeling method is evaluated by the metric proposed by Black et al. in [1]. Because we only comparing the performance of detections, we only utilize a subset of the evaluation measures for this reason.

For the GT provided with dataset is XML format files and it describes the boundary boxes of certain several frames. Since there is the information of the size of boundary boxes and detailed locations in the frame. Thus we have to define the value of overlap between the foreground motion area and boundary boxes in dataset.

$$P_{object} = \frac{A_{obj} \cap A_{tru}}{A_{obj} \cup A_{tru}} \quad (11)$$

$$A(x,y,w,h) = \left[x - \frac{w}{2}, x + \frac{w}{2}\right] \times \left[y - \frac{h}{2}, y + \frac{h}{2}\right] \quad (12)$$

A_{obj} denotes the area of corresponding object which is generated by proposal, and A_{tru} denotes the foreground rectangle box of GT in dataset.

4.3. Evaluation parameters

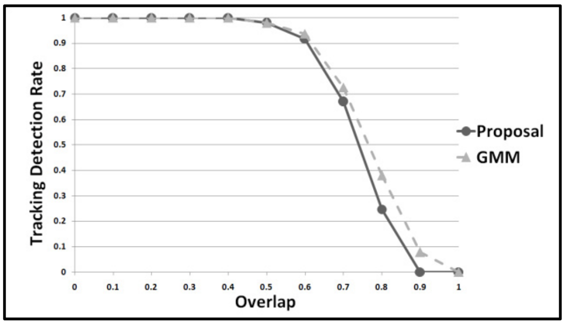
$$FAP = FP_{pixels} / (TP_{pixels} + FP_{pixels}) \quad (13)$$

$$TRDR = TP_{objects} / GT_{objects} \quad (14)$$

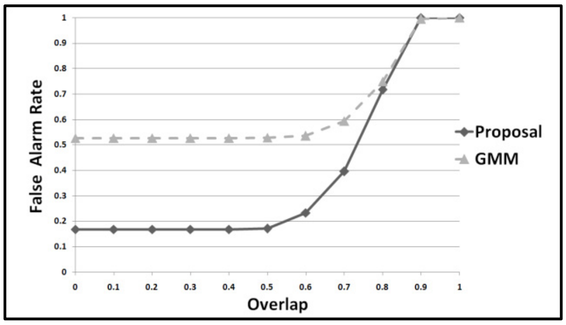
TP denotes True Positive, FP denotes False Positive, and FAP is the False Alarm Rate when TRDR means the Tracking Detection Rate. Once the percentage of object, P_{object} is bigger or equal than defined value of overlap, the detected object belongs to $TP_{objects}$. As for FAP, we count the total pixels of FP and TP for every frame, these

Table 1. Tracking Detection Rate and False Alarm Rate Of Pets2001Video Sequences

Overlap			0.1	0.3	**0.4**	**0.5**	0.6	0.8
PetsD1TeC1	TRDR	GMM	0.984	0.963	0.952	0.941	0.877	0.452
		Pro.	0.967	0.919	0.894	0.878	0.846	0.277
	FAR	GMM	0.504	0.508	0.509	0.509	0.519	0.646
		Pro.	0.144	0.149	0.151	0.160	0.170	0.553
PetsD1TeC2	TRDR	GMM	0.982	0.974	0.974	0.931	0.862	0.336
		Pro.	0.948	0.905	0.862	0.801	0.655	0.181
	FAR	GMM	0.364	0.365	0.365	0.375	0.384	0.598
		Pro.	0.095	0.098	0.124	0.139	0.220	0.580
PetsD2TeC1	TRDR	GMM	0.982	0.974	0.974	0.931	0.862	0.336
		Pro.	0.948	0.905	0.862	0.801	0.655	0.181
	FAR	GMM	0.364	0.365	0.365	0.375	0.384	0.598
		Pro.	0.095	0.098	0.124	0.139	0.220	0.580
PetsD2TeC2	TRDR	GMM	0.935	0.926	0.916	0.886	0.738	0.142
		Pro.	0.911	0.901	0.891	0.842	0.605	0.108
	FAR	GMM	0.521	0.522	0.522	0.527	0.547	0.817
		Pro.	0.102	0.103	0.105	0.130	0.221	0.813



(a)



(b)

Figure 4. (a) TRDR for PetsD2TeC2 Video Sequences (b) FAR for Pets D2TeC2 Video Sequences

datas are shown in Table.1, we present the FAP and TRDR for each outdoor video sequence with a few defined overlap, and experiments proved that when the overlap is between 0.4 and 0.5, this is the most ordinary percentage for actual foreground object occupied the boundary box in GT, which accounts for the data of these two column is most convinced.

5. Conclusion

This paper presents an updated scheme for object detection based on a mixture of Gaussian models. By utilizing intensity and texture information, and integrated with modified Gaussian mixture models by redefining the update criterion, proposed algorithm is robust for surveillance applications. Experimental results validate significant improvements compared with the standard scheme, and proposal enhances the quality of segmentation by averagely more than 34.8% FAP decrease without influencing TRDR.

6. Acknowledgment

This work was supported by fund from MEXT via Kitakyushu innovative cluster projects and CREST, JST.

References

[1] J. Black, T. Ellis, and P. Rosin, "A novel Method For Video Tracking Performance Evaluation," *International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*, pp. 125–132, 2003.

[2] S. Huwer and H. Niemann, "Adaptive Change Detection for Real-time Surveillance Applications," *Proceedings of 3rd IEEE Workshop On Visual Surveillance*, pp.37-45, 2000.

[3] C. Stauffer and W.E.L. Grimson, "Adaptive Background mixture Models for Real-time Tracking," *Proceedings of IEEE Int'l Conference on Computer Vision and Pattern Recognition*, pp. 246-252, 1999.

[4] O. Javed, K. Shafique, and M. Shah, "A Hierarchical Approach To Robust Background Subtraction Using Color And Gradient Information," *Proceedings of the Workshop on Motion and Video Computing*, pp. 22-27, 2000.

[5] R. Cucchiar, C. Grana, M. Piccardi, and A. Prati, "Detecting Moving Objects, Ghosts, and Shadows in Video Streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.25, Issue.10, pp.1337-1342, Oct.2003.

[6] C. Stauffer and W.E.L. Grimson, "Learning Patterns of Activity Using Real-time Tracking", *IEEE Transactions on Pattern Analysis & Machine Intelligence*, Vol.22, Issue.8, pp. 747-757, Aug 2000.

[7] P. KaewTraKulPong and R. Bowden, "An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection," *Proc.2nd European Workshop on Advanced Video Based Surveillance Systems*, AVBS01, Sept 2001.