

# An Algorithm of Detecting Moving Foreground Based on an Improved Gaussian Mixture Model

Mingjie Wang<sup>1</sup>, Jesse S. Jin<sup>1,2</sup>(✉), Xianfeng Han<sup>1</sup>, Wei Jiang<sup>1</sup>,  
Yifei Jing<sup>1</sup>, Lei Gao<sup>2</sup>, and Liping Xiao<sup>2</sup>

<sup>1</sup> School of Computer Software, Tianjin University, Tianjin 300000, China  
{wang\_major, hanxianf, jiangweitju}@163.com,  
jinsheng@tju.edu.cn, lovejing0306@foxmail.com

<sup>2</sup> Beijing Aerospace Institute of Automatic Control, Beijing 100085, China  
{thrstone, xlp027}@sina.com

**Abstract. Objective:** The technique of detecting moving regions has been playing an important role in computer vision and intelligent surveillance. Gaussian mixture models provide an advanced modeling approach for us. Although this method is very effective, it is not robust when there are some lighting changes and shadows in the scenes. This paper proposes a mixture model based on gradient images. **Method:** We firstly calculate gradient images of a video stream using the Scharr operator. We then mix RGB and gradient, and use a morphological approach to remove noise and connect moving regions. To further reduce false detection, we make an AND operation between two modeling results and result in final moving regions. **Result:** Finally, we use three video streams for analysis and comparison. Experiments show that this method has effectively avoided false detection regions resulting from lighting change and shadow, and improves the accuracy of detection. **Conclusion:** The approach demonstrates its promising characteristics and is more applicable in real-time detection.

**Keywords:** GMM (Gaussian mixture Model) · Gradient image · Scharr operator · Sobel · Morphological · AND operation

## 1 Foreword

In area of computer vision, the moving foreground detection is the foundation of intelligent surveillance, and its accuracy will influence the subsequent processes [1] including target tracking, object classification, etc. Moving foreground detection is about detecting the moving area in a video stream according to certain characteristics of video sequences. Only on this can the feature extraction of target and pattern recognition be conducted, which makes advanced image analysis and interpretation possible. Therefore, the study on foreground detection algorithms has significant importance [2].

The advanced background modeling methods mainly include average background model, nuclear density model, Gaussian mixture model (GMM) and so on. Nowadays, the GMM brought forward by Grimson and Stauffer in 1999 should be the most widely

used and most effective one. In the Gaussian mixture model, every state is demonstrated by a Gaussian process function [4] and is an expansion of single Gauss distribution. The single Gauss model regards each individual pixel in video sequences as an independent random variable and assumes they all follow Gauss distribution. The single Gauss model to a certain degree improves the deficits of traditional background modeling method like background subtraction method. However, in actual scene, pixels are usually in the form of multi-modal. Consequently, Grimson and other experts present that each pixel should follow a mixed Gauss distribution. The part with high weight in the Gauss distribution represents background area, while the part with low weight represents the moving area.

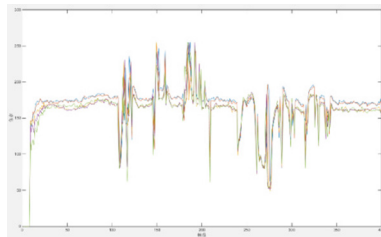
The GMM effectively improves deficits of low-level modeling methods and is robust with respect to repetitively moving background area, detection of slowly moving target, etc. However, under the influence of sudden change of illumination and shadows formed because objects are shaded, standard mixture model will mistake the background detection for foreground. In order to solve this problem, this article raises the Gaussian Mixture Model based on gradient images (G-GMM) which integrates with traditional Gaussian Mixture Model based on RGB (GMM) to form a new background model. This combines advantages of both by keeping the complete moving areas at the same time effectively eliminating disturbance of changes of illumination as well as shadows, which greatly enhances accuracy and stability of moving foreground detection.

## 2 Traditional GMM

Traditional GMM regards each pixel of video sequence as a random variable and assumes they all follow the mixed distribution consisting of  $K$  single Gauss distribution. Five fixed example pixels are taken from the experimental video LabText.avi, and the RGB value is sampled in the first 400 frames. Then the weighted sum of R, G, B for all samples is drawn in a coordinate graph, as in Fig. 1.

A typical mixed model assumes pixels are independent of each other and the whole video is a stable randomized procedure. The probability density function is as follows:

$$P(X_t) = \sum_{i=1}^K w_i^t \eta(X_t, \mu_i^t, \Sigma_i^t) \quad (1)$$



**Fig. 1.** The weighted sum of RGB for five example pixels in the video sequence (0.299R + 0.587G + 0.114B)

In which  $X_t$  is the pixel value of the corresponding pixel at time of  $t$ ;  $K$  is the number of single Gaussian distribution in the mixture model, and the bigger  $K$ 's value the better the classification effect; the amount of calculation will follow up and increase. Normally the value will be  $3 \sim 5$ .  $w_i^t$  is the weight estimation of  $i^{th}$  Gaussian distribution at the time of  $t$  in the mixture model;  $\mu_i^t$  is the mean value of  $i^{th}$  Gaussian distribution at the time of  $t$  in the mixture model;  $\Sigma_i^t$  is the co-variance matrix of  $i^{th}$  Gaussian distribution at the time of  $t$  in the mixture model; limited by the complication of calculation, the co-variance matrix usually is approximated to  $\sigma_i^2 I$  (that is, imagine three channels of pixels are independent of each other, and in which  $I$  is unit matrix),  $\eta$  is probability density function of single Gaussian distribution [5]

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

Based on Fig. 1 and formulas above, we can know the probability density function of mixture model is determined by weight, mean value and variance. Suppose the parameter vector of the model is  $\theta(w_1, \mu_1, \sigma_1, w_2, \mu_2, \sigma_2, \dots, w_K, \mu_K, \sigma_K)$ . The process of Gaussian mixture background modeling is actually the process of constantly estimating and updating parameter vector  $\theta$ .

In multivariate statistics, the common method used for parameter estimation is the Maximum Likelihood Estimation Method; however during the background modeling, pixel value  $X$  is the only observation data that can be used. The classification number of this pixel value cannot be observed, and the incomplete data when looking for the maximum value of Likelihood function will lead to a transcendental equation with plural roots. It is caused by the fuzziness between the observation data  $X$  and its matching distribution sequence number. As a result, after introducing the hidden data  $Y$  to make up a complete data pair  $(X, Y)$ , we use EM method of calculation to conduct parameter estimation and can get the updating formula of weight  $w_k^t$

$$w_k^t = (1 - \alpha)w_k^{t-1} + \alpha L_k^t \quad (3)$$

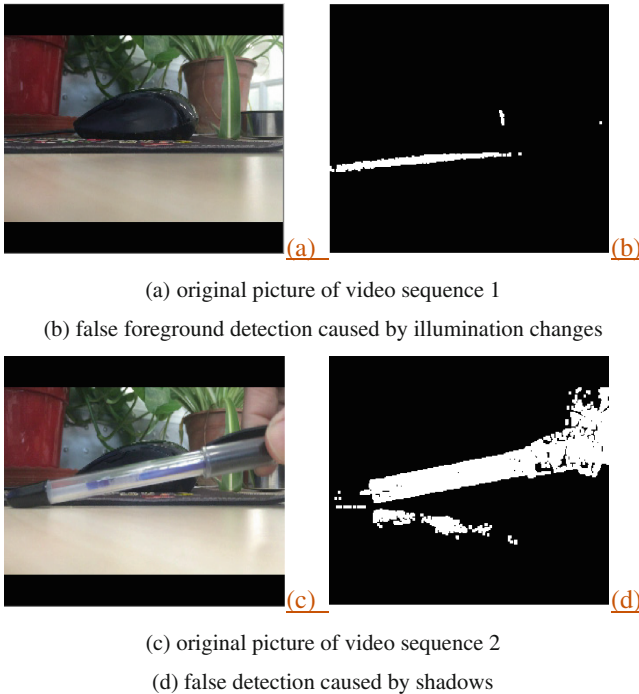
In which  $\alpha$  is the learning rate of the mixture model and determines the speed of background modeling. If it is set too big, the iteration will get over the optimal value making the algorithm fail to converge. If it is set too small, the convergence rate will be slowed down resulting in higher detection error rate. When new pixel values match with the  $K^{th}$  distribution  $L_k^t = 1$ , or  $L_k^t = 0$ ; the weight of  $K$  distribution is adjusted in the process of iteration by using formula (3) and new pixel values. When new pixel values match with the  $K^{th}$  distribution of the mixture model, the mean value and variance of this distribution will be updated according to formulas (4) and (5)

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho X_t \quad (4)$$

$$\sigma_t = (1 - \rho)\sigma_{t-1} + \rho(X_t - \mu_t)^T (X_t - \mu_t) \quad (5)$$

In which,  $\rho$  is another learning rate,  $\rho = \alpha\eta(X_t|\mu_k, \sigma_k)$  [5]. If there is no distribution matching with  $X_t$  in the present mixture model, the Gaussian distribution with the minimum weight will be replaced by new Gaussian distribution with an mean of  $X_t$  and at the same time one larger variance and smaller weight of its will be initialized. The rest mean value and variance of Gaussian distribution remain and only the weight is attenuated. Then each weight will go through normalization processing [6], and each Gaussian distribution will be sequenced from high to low according to priority  $\left(p_k^t = \frac{w_k^t}{\sigma_k^t}\right)$ , then a judgment threshold  $\lambda$  will be set. The weight of the sorted Gaussian distribution will be accumulated from the maximum value and when  $\sum_{i=1}^N w_i^t \geq \lambda$ , the accumulation stops. At that time, we believe the first  $N$  distributions feature background pixels, thereby constructing new background model [7].

The above process is the traditional Gaussian Mixture Model based on RGB. Compared to low-level modeling method such as background subtraction method, it has better performance in terms of foreground detection. This model can not only update background adaptively, but also allow repetitive moving areas to be existed in the background. Although the moving background will be identified as foreground when it enters the scene, staying for a longer time, it will be absorbed and turned into background pixel. Unfortunately, in practical application, there is always sudden illumination change and shading. The Grimson's GMM under this circumstance will result in wrong foreground detection, as illustrated in Fig. 2.



**Fig. 2.** False detection of the traditional GMM resulted from illumination changes and shadows

Figure 2 illustrates the deficiency of traditional GMM in practical application. In applications with strict requirements of detection accuracy, false detection caused by such as fire and smoke detection alarm system, illumination changes and shadows will lead to false alarm, and further unnecessary expenses.

### 3 Gradient Image

Gradient is a kind of feature descriptor which is identified as local discontinuity of pixels and is mainly used to feature sudden change of pixels, texture structure and so on. The gradient value of a pixel point in the image is the first-order differential of pixel function regarding coordinate of this point and is used to describe the speed of pixel value change. Now we only consider the special gradient in X and Y directions; the formula is in (6)

$$|\nabla I(x, y)| = \sqrt{\left[\frac{\partial I(x, y)}{\partial x}\right]^2 + \left[\frac{\partial I(x, y)}{\partial y}\right]^2} \quad (6)$$

In the image, the pixel function  $I(x, y)$  is a discrete non-differentiable function, so we can only use approximation methods to approach its first-order differential in the X and Y directions. Common first-order differential operators mainly include Roberts operators, Prewitt operators, Sobel operators, Scharr operators and so forth, among which, Sobel operators are the most popular. Its essence is convolution operation. It is easy to calculate this kind of operators and the speed of processing is fast; besides, the cost of time and internal storage is low. A regular Sobel operator (the size of its core is  $3 \times 3$ ) is defined as follows [9]:

$$\frac{\partial I(x, y)}{\partial x} = \begin{vmatrix} I(x+1, y+1) + 2I(x+1, y) + I(x+1, y-1) \\ -I(x-1, y+1) - 2I(x-1, y) - I(x-1, y-1) \end{vmatrix} \quad (7)$$

$$\frac{\partial I(x, y)}{\partial y} = \begin{vmatrix} I(x-1, y-1) + 2I(x, y-1) + I(x+1, y-1) \\ -I(x-1, y+1) - 2I(x, y+1) - I(x+1, y+1) \end{vmatrix} \quad (8)$$

According to the calculation of formulas (6), (7) and (8), we can get gradient values of each point in the image (single channel/three-channel). The size of the convolution kernel in a Sobel operator determines the accuracy of gradient calculation. Big kernels can better approach derivatives but the cost is high; small kernels are sensitive to noises [10]. In this article, we will analyze operators and how to choose the size of cores in the next section.

Creating image  $T$  with a similar size of original picture  $I$  and the same number of channels and making  $T = |\nabla I(x, y)|$ .  $T$  is called the gradient image of  $I$ , as illustrated in Fig. 3.



(a) original picture (b) gradient image based on Sobel operators (the size of core is 5\*5)

**Fig. 3.** Gradient image

Compared to colored images, gradient images have the feature of not being easily influenced by illumination change, quantization noises, shadows, mild distortion of images and affine transformation etc. [11]. For example, when the weather in the video scene changes from sunny to cloudy, the pixel values of colored images will also change greatly. On the contrary, pixel values of corresponding gradient images will not be changed. The reasons are as follows.

Suppose the pixel changing amplitude caused by illumination change or shadows in the convolution area is  $\Delta I$ . Here we assume the light evenly changes, then the pixel value in the areas changes to  $I(x, y) + \Delta I$ ; introducing it into formulas (7) and (8), we

can know 
$$\frac{\partial[I(x, y) + \Delta I]}{\partial x} = \left| \begin{array}{c} I(x+1, y+1) + \Delta I + 2I(x+1, y) + 2\Delta I + I(x+1, y-1) + \Delta I \\ -I(x-1, y+1) - \Delta I - 2I(x-1, y) - 2\Delta I - I(x-1, y-1) - \Delta I \end{array} \right|$$

$$= \frac{\partial I(x, y)}{\partial x}$$

In the same way,  $\frac{\partial[I(x, y) + \Delta I]}{\partial y} = \frac{\partial I(x, y)}{\partial y}$ . As a result, theoretically the gradient value in the image influenced by balanced illumination will not change.

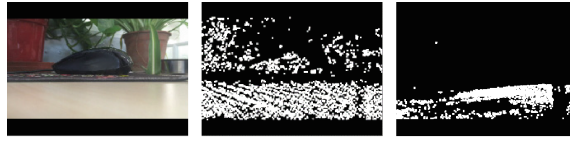
## 4 G-GMM Background Modeling

**GMM Based on Sobel Gradient Images.** The essence of Sobel operators is to conduct convolution operation at each pixel point. The common  $3 \times 3$  convolution

core is  $\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$ ,  $\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$ . The size of a convolution core determines

the accuracy of gradient calculation. The sizes Sobel operators support are  $3 \times 3$ ,  $5 \times 5$  and  $7 \times 7$ . We can separately use these three convolution operations to get gradient image sequence and conduct GMM and foreground detection on this basis. The result is shown in Fig. 4.

**GMM Based on Scharr Gradient Images.** In Sobel operators, the weights of convolution cores are small, which makes the gradient calculation is easy to be influenced by noises and leads to unsatisfying accuracy. Therefore, we use Scharr filter to sole the inaccuracy caused by approximation of  $3 \times 3$  Sobel operator gradients. The Scharr filter is as fast as Sobel filter, but the accuracy is higher [12]. The convolution cores of



(a1) original pictures; (a2) GMM based on  $7 \times 7$  core gradients; (a3) GMM based on RGB



(b1) original pictures; (b2) GMM based on  $5 \times 5$  core gradients; (b3) GMM based on RGB



(c1) original pictures; (c2) GMM based on  $3 \times 3$  core gradients; (c3) GMM based on RGB

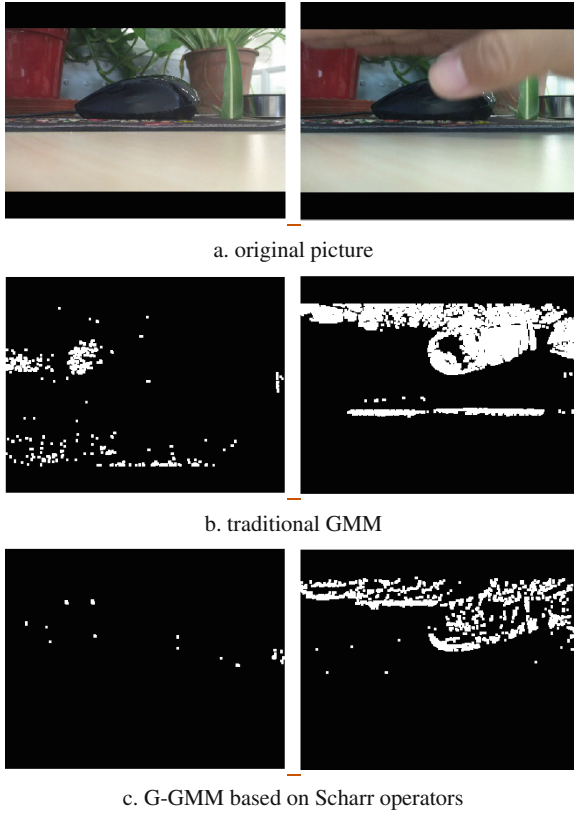
**Fig. 4.** Three GMMs based on Sobel operators having different core sizes

Scharr operators are  $\begin{bmatrix} -3 & 0 & 3 \\ -10 & 0 & 10 \\ -3 & 0 & 3 \end{bmatrix}$ ,  $\begin{bmatrix} 3 & 10 & 3 \\ 0 & 0 & 0 \\ -3 & -10 & -3 \end{bmatrix}$ . The gradient images in video sequences can be calculated by using Scharr operators, see formula (9), which can also be used to conduct GMM and detect moving foreground. The result is shown in Fig. 5.

$$|\nabla I(x, y)| = 0.5 \times \left[ \frac{\partial I(x, y)}{\partial x} \right] + 0.5 \times \left[ \frac{\partial I(x, y)}{\partial y} \right] \tag{9}$$

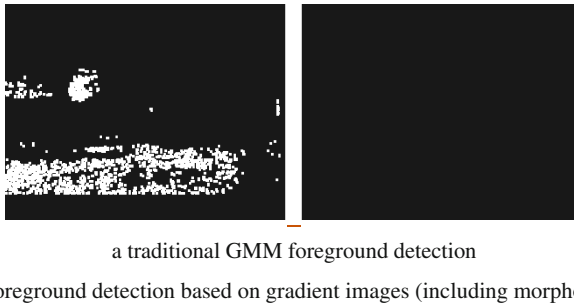
From Fig. 5, we can know, the foreground detection algorithm based on Scharr gradient images has the benefits of being resistant to illumination changes, shadows and so forth just like the foreground detection algorithm based on  $5 \times 5$  Sobel gradient images. In addition, the former has better noise-resistance than the latter, which leads to less false detection points. Therefore, in this experiment, we will use Scharr operators to calculate gradient image sequence.

**Morphological Filter AND Operation.** This article first carries out two consecutive erosion calculation on foreground detection results based on gradient images, eliminates most isolated false detection points, and then uses one dilation calculation to connect moving areas making the internal part of targets more substantial and well connected, which also facilitates further AND operation. The results of morphological filtering are shown in Fig. 6, after morphological filtering, foreground detection



**Fig. 5.** G-GMM

algorithm based on G-GMM can eliminate most isolated false detection points, meantime keep the moving connected region and enhance the accuracy of foreground detection.



**Fig. 6.** The effect of morphology filter



The morphology filter eliminates most noises in G-GMM detection results. When the video scene is relatively simple, it has good detection effect and nearly no false detection points. When the scene is complicated, some noises still exist due to limitation of gradient calculation accuracy, but it still has strong robustness to influence of illumination and shadows. In order to further eliminate false detection points, we combine the detection result of traditional GMM and G-GMM, keeping the high accuracy of foreground detection and eliminating shadow detection. Assume the foreground detection result of traditional GMM is  $T(x, y)$ , and the result of G-GMM is  $G(x, y)$ . The final result is  $F(x, y)$ . Combining two results in AND operation:  $F(x, y) = T(x, y) \cap G(x, y)$ , then we can get the final detection result.

**Improved GMM Flow Chart.** The improved Gaussian Mixture foreground detection flow is shown in Fig. 7 [14].

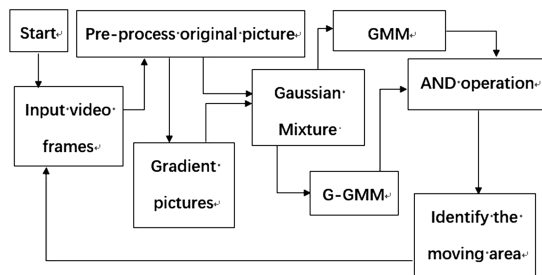


Fig. 7. The process of improved Gaussian Mixture Models

## 5 Experiment and Analysis

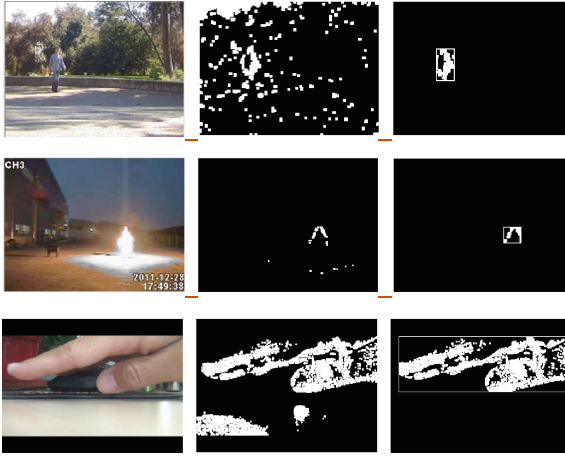
In this experiment, the initial value of learning speed in GMM is preset as 0.01 and Scharr operators are used in gradient image calculation. Three videos are compared in the following pictures and the video information can be found in Table 1.

Table 1. Information of experimental videos

Name of video	Source	Format	Resolution	Type
original_view_40	Documents [15]	avi	180*144	Outdoors
LabFire	VisiFire	avi	320*240	Indoors
LabText	Self-collected	avi	320*240	Indoors

Carrying out moving foreground detection on three videos in Table 1 by using traditional GMM and models in this article respectively. The results are shown in Fig. 8.

Figure 8b shows results of foreground detection of GMM based on RGB. There are lots of false detection points in both video 1 and video 3. The results of video 1 prove the algorithm in this article has higher stability because in a stable background video, the learning speed of gradient background is faster than RGB [16]. The traditional

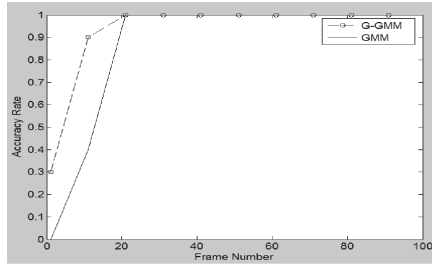


**Fig. 8.** Experimental comparison (a). Original video pictures; (b). Traditional GMM; (c). Method in this article

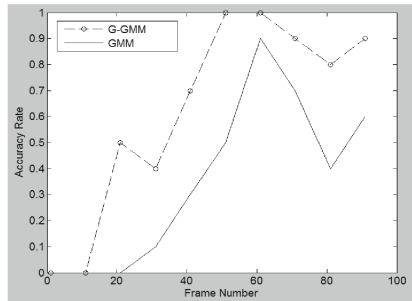
GMM used in video 2 and video 3 causes false detection because of illumination changes and shadows generated by shading of objects. This paper makes use of the characteristic that gradient images have robustness to changes of illumination to eliminate the above false detection effectively. At the same time, it makes use of morphological operation AND operation to guarantee the accuracy of foreground detection effectively, and has certain robustness indoors and outdoors. In order to explain the false detection rate of this algorithm more accurately, we run statistics on the first 100 frames in the above three videos and calculate relative detection rate and get the following result.

From the comparison in Fig. 9, we can see the method of this article represented by dotted lines all show excellent performance in foreground detection in three videos used in the experiment. It can both overcome the detection deficiency of traditional Gaussian Mixture Model and obtain high accuracy. Besides, it also has fast convergence rate [17].

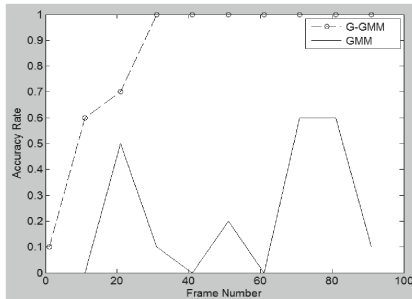
Although the method in this article solves the problems of traditional GMM when illumination changes and shadows occur and effectively reduces false detection rate of moving foreground, it has some disadvantages. First, the algorithm adopts gradient approximation Scharr operators when doing the gradient image calculation. When facing complicated scenes, noises still exist. Next, the time this algorithm cost is three times more than that of traditional algorithm. When the video pixels are high, the real-time is what we need to consider [18]; however, the real-time in the video used in the experiment can be guaranteed. In this day and age when hardware resources are relatively rich, the requirement of practical application for accuracy of foreground detection is higher in order to avoid huge waste caused by poor detection. To sum up, the detection model proposed by this paper is more suitable for real-time detection of moving foreground.



(a)



(b)



(c)

**Fig. 9.** The comparison of detecting accuracy rate (a). Video 1 (original\_view\_40); (b). Video 2 (LabFire); (c). Video 3(LabText)

## 6 Conclusion

Compared to some low-level foreground detection methods like background subtraction method, traditional Gaussian mixture model has higher detection efficiency; however, it still cannot solve the problems generated by illumination changes and shadows. In order to solve this problem, this article proposes a G-GMM based on gradient images. First we use Scharr filter operators to calculate gradient image sequences, then conduct Gaussian mixture modeling according to sequences of original pictures and sequences of gradient images and also use morphology filter to carry out filtering and connecting operation on the initial results. In order to maintain accuracy of detection and eliminate shadows, we compare detection results of two models and

finally obtain the detection results in the moving area. The experiment shows the method in this article can effectively reduce false foreground detection caused by changes of illumination and shadows. Compared to traditional GMM, this method has better detection efficiency and robustness, thereby being more suitable for application on actual moving foreground monitoring platforms.

## References

1. Chen, Z., Ellis, T.: A self-adaptive Gaussian mixture model. *Comput. Vis. Image Underst.* **122**, 35–46 (2014)
2. Yuan, C.F., Wang, C.X., Zhang, X.G., et al.: Video segmentation of illumination abrupt variation based On MOGs and gradient information. *J. Image Graph.* **12**(11), 2068–2070 (2007)
3. Basri, I., Achamad, A.: Gaussian mixture models optimization for counting the numbers of vehicle by adjusting the region of interest under heavy traffic condition. In: *International Seminar on Intelligent Technology and Its Applications, 2015 (ISITIA 2015)*, Surabaya (2015)
4. Qiao, S.J., Jin, K., Han, N., et al.: Trajectory prediction algorithm based on Gaussian mixture model. *J. Softw.* **26**(5), 1048–1049 (2015)
5. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: *Computer Vision and Pattern Recognition, 1999 (CVPR 1999)*, Fort Collins, CO (1999)
6. Cui, W.P., Shen, J.Z.: Moving target detection based on improved Gaussian mixture model. *Opto-Electron. Eng.* **37**(4), 119–121 (2010)
7. Chen, X.Y.: *The research and implementation of background modeling and updating algorithm based on mixture gaussian model*. Northeastern University, Shenyang (2009)
8. Lei, L.Z.: On the edge detection method of digital image. *Bull. Surv. Mapp.* **40**(3), 40–41 (2006)
9. Lu, Z.Q., Liang, C.: Edge thinning based on Sobel operator. *J. Image Graph.* **5**(6), 516–518 (2000)
10. Yu, S.Q., Liu, R.Z.: *Learning OpenCV*, pp. 169–171. Tsinghua University Press, BeiJing (2009)
11. Pan, X.H., Zhao, S.G., Liu, Z.P., et al.: Detection of video moving objects by combining grads-based frame difference and background subtraction. *Optoelectron. Technol.* **29**(1), 34–36 (2009)
12. Meng, Y.F., OuYang, N., Mo, J.W., et al.: A shadow removal algorithm with Gaussian mixture model. *Comput. Simul.* **27**(1), 210–212 (2010)
13. Ma, Y.D., Zhu, W.F., An, S.X., et al.: Improved moving objects detection method based on Gaussian mixture model. *Comput. Appl.* **27**(10), 2544–2546 (2007)
14. He, K., Ju, S.G., Lin, T., et al.: Image denoising on TV numerical computation. *J. Univ. Electron. Sci. Technol. China* **42**(3), 459–461 (2013)
15. Gorelick, L., Blank, M., Shechtman, E., et al.: Actions as space-time shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(12), 2247–2248 (2007)
16. Fan, W.C., Li, X.Y., Wei, K., et al.: Moving target detection based on improved Gaussian mixture model. *Comput. Sci.* **42**(5), 286–288 (2015)
17. Feng, W.H., Gong, S.R., Liu, C.P.: Foreground detection based on improved Gaussian mixture model. *Comput. Eng.* **37**(19), 179–182 (2011)
18. Sun, D., Liu, J.F., Tang, X.L.: Edge detection based on density gradient. *Chin. J. Comput.* **32**(2), 299–302 (2009)