An improved mixed gaussian-based background modelling method for fast gesture segmentation of mobile terminals

Chengfeng Jian, Tao Lu, Xiaoyu Xiang, Meiyu Zhang*

Computer Science and Technology College, Zhejiang University of Technology, Hangzhou 310023, China

zmy@zjut.edu.cn

ABSTRACT. *The gesture segmentation of mobile terminals faces two major problems: the segmentation effect is constrained by complex background, and the timeliness is dampened by the limited resources of mobile devices. To solve these problems, this paper puts forward a detection method based on improved mixed Gaussian background modelling. The first step is to improve the distribution of the mixed Gaussian background model. The number of Gaussian distributions was controlled adaptively to reduce system computing load and storage. Then, the learning rate was controlled in light of special scene change rate, aiming to enhance the adaptability of gesture segmentation to environmental changes. The experimental results show that the proposed method can rapidly eliminate environmental interferences, achieve effective hand segmentation and realize good computing performance, despite the massive changes to the background scene.*

RÉSUMÉ. *La segmentation des gestes des terminaux mobiles se heurte à deux problèmes majeurs: l'effet de la segmentation est limité par un arrière-plan complexe et la rapidité est freinée par les ressources limitées des terminaux mobiles. Pour résoudre ces problèmes, le présent document propose une méthode de détection basée sur une modélisation améliorée d'arrière-plan gaussien mixte. La première étape consiste à améliorer la distribution du modèle d'arrière-plan gaussien mixte. Le nombre de distributions gaussiennes a été contrôlé de manière adaptative afin de réduire la charge de calcul et le stockage du système. Ensuite, le taux d'apprentissage a été contrôlé à la lumière d'un taux de changement de scène particulier, dans le but d'améliorer la capacité d'adaptation de la segmentation des gestes aux changements environnementaux. Les résultats expérimentaux montrent que la méthode proposée permet d'éliminer rapidement les interférences environnementales, de réaliser une segmentation gestuelles des mains efficace et d'atteindre de bonnes performances informatiques, malgré les modifications massives apportées à la scène d'arrière-plan.*

KEYWORDS: *mixed gaussian model, background modelling, learning rate, gesture segmentation.*

MOTS-CLÉS: *modèle gaussien mixte, modélisation d'arrière-plan, taux d'apprentissage, segmentation gestuelle.*

DOI:10.3166/TS.35.243-252

1. Introduction

The precision of gesture recognition depends heavily on accurate gesture segmentation (Chen *et al.*, 2017), which in turn calls for precise detection of hand movement. Currently, the most popular detection methods for moving objects include time difference method, optical flow method and background subtraction method (Plyer *et al.*, 2016; Gui *et al.*, 2017). Among them, only the background subtraction method applies to the object detection of mobile terminals. Background subtraction mainly falls into mean background modelling, single Gaussian background modelling, and mixed Gaussian background modelling (Zheng *et al.*, 2016; Goyal and Singhai, 2017; Mangal and Kumar, 2016). The mixed Gaussian background modelling, which represents the pixels with multiple Gaussian distributions, stands out for its ability to accurately simulate the background of multi-peak distribution in real-time.

For mobile terminal applications, the gesture segmentation based on mixed Gaussian background modelling still has the following disadvantages: (1) The processing is slow due to the small computing and storage amounts of the mobile terminal; (2) The detection effect is affected by the virtual shadows easily produced at the dynamic changes of the background scene. To solve the defects, Katsarakis *et al.*, (2016) accelerated the elimination of virtual shadows by controlling the spatial change of learning rate, which pushed up the mean learning rate per frame and increased the computing load. Azzam *et al.*, (2016) put forward a spatial global mixed Gaussian model based on RGB pixels. Huang *et al.*, (2015) presented a new moving object detection method based on pixel-based time and space information. However, all these methods require a high system storage.

In light of the above, this paper attempts to improve the mixed Gaussian background modelling method from two aspects. On the one hand, the number of Gaussian distributions was selected adaptively for each pixel, so as to reduce the consumption of mobile computing resources without sacrificing the detection effect; on the other hand, a special scene change rate was defined to control the learning rate and accelerate the elimination of the virtual shadows produced at massive changes of the background.

2. Improved mixed gaussian background model

2.1. Construction of background model

Let $X=\{X_1, X_2, X_3,\ldots,X_t\}$ be the pixels of video frames collected in chorological order, with X_t being the pixel sample of the video frame collected at time t.

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} \, \mathrm{H}(X_t, \mu_{i,t}, \phi_{i,t}^2) \quad (1)$$

where K is the number of Gaussian distributions in the mixed Gaussian model; $\omega_{i,t}$ is the weight of the i-th Gaussian distribution in the mixed Gaussian model at time t;

$$H(X_t, \mu_{i,t}, \phi_{i,t}^2) = \frac{1}{\sqrt{2\pi^n}\,|\phi_{i,t}|} e^{-\frac{1}{2}(X_t-\mu_{i,t})^T \phi_{i,t}^{-2}(X_t-\mu_{i,t})} \tag{2}$$

$$\begin{aligned} -\frac{1}{2}(X_t-\mu_{i,t})^T \phi_{i,t}^{-2}(X_t-\mu_{i,t}) &= -\frac{(X_t-\mu_{i,t})^T(X_t-\mu_{i,t})}{2\phi_{i,t}^2} \\ &= -\frac{|(X_t-\mu_{i,t})^T\,||(X_t-\mu_{i,t})|}{2\phi_{i,t}^2} \\ &= -\frac{|(X_t-\mu_{i,t})\,||(X_t-\mu_{i,t})|}{2\phi_{i,t}^2} \\ &= -\frac{(X_t-\mu_{i,t})^2}{2\phi_{i,t}^2} \end{aligned} \tag{3}$$

According to (2) and (3), the probability density function of each mixed Gaussian distribution can be expressed as:

$$H(X_t, \mu_{i,t}, \phi_{i,t}^2) = \frac{1}{\sqrt{2\pi^n}\,|\phi_{i,t}|} e^{-\frac{(X_t-\mu_{i,t})^2}{2\phi_{i,t}^2}} \tag{4}$$

where $\mu_{i,t}$ and $\omega_{i,t}$ are the mean and the covariance matrix of the i-th Gaussian distribution at time t, respectively; $\omega_{i,t}$ is the weight representing the similarity ratio between the sample value of the current distribution and the mixed model of image X.

Each new pixel X_t should be compared with the first K Gaussian distributions according to equation (5) below:

$$|X_t - \mu_{i,t-1}| \le 2.5 * \sigma_{i,t-1} \tag{5}$$

If the mean deviation of the Gaussian distribution is within 2.5σ, the new pixel matches the Gaussian distribution. If the Gaussian distribution satisfies the background requirement, the new pixel belongs to the background; otherwise it belongs to the foreground.

2.2. Update to the background model

The traditional mixed Gaussian background modelling often mistakes the massive changes of the background for the foreground. The resulting virtual shadows will dampen the modelling accuracy and the effect of foreground extraction.

If the new pixel X_t matches the i-th Gaussian distribution at time t-1, the weight,

mean and covariance matrix should be updated by the following equations (Jian and Wang, 2014):

$$\omega_{i,t} = (1-\alpha)\omega_{i,t-1} + \alpha \tag{6}$$

$$\mu_{i,t} = (1-\beta)\mu_{i,t-1} + \beta X_{i,t} \tag{7}$$

$$\sigma_{i,t}^2 = (1-\beta)\sigma_{i,t-1}^2 + \beta(X_{i,t} - \mu_{i,t})^T (X_{i,t} - \mu_{i,t}) \tag{8}$$

where α is the learning rate; β is the ratio of the learning rate to the weight.

If the new pixel X_t fails to match any Gaussian distribution, the original distribution with the smallest weight should be replaced without changing the mean and variance. The mean and covariance matrix should be updated as above, while the weight should be updated by the following equation:

$$\omega_{i,t} = (1-\alpha)\omega_{i,t-1} \tag{9}$$

After each update, all weights should be normalized such that the total weight equals 1, and all Gaussian distributions should be ranked again by priority $\rho_{i,t}$. Here, $\rho_{i,t}$ refers to the priority of the *i*-th Gaussian distribution in the mixed Gaussian background model at time t.

$$\rho_{i,t} = \frac{\omega_{\mathrm{i,t}}}{\sigma_{\mathrm{i}}} \tag{10}$$

Many Gaussian distributions are generated during the computing process. Considering the limited resources of mobile terminals, the redundant Gaussian distributions should be removed. Thus, this paper proposes to scan all Gaussian distributions of each pixel at the interval of f frames. If a pixel has more than 3 Gaussian distributions, the distribution with the lowest priority should be deleted. Then, the weight of each Gaussian distribution should be checked. If the current weight is smaller than the initial weight and the priority is lower than the initial priority, the Gaussian distribution should be determined as redundant and deleted.

2.3. Foreground segmentation

After ranking all Gaussian distributions by priority $\rho_{i,t}$ at time t, and the top B Gaussian distributions should be selected for further analysis. Then, the matching relationship between each pixel value X_t and the top B Gaussian distributions at time t. If the pixel value X_t matches one of the top B Gaussian distributions, then this pixel is a background point; if the pixel value fails to match any of the top B Gaussian distributions, then this pixel is a foreground point, i.e. a moving object. B should satisfy the following equations:

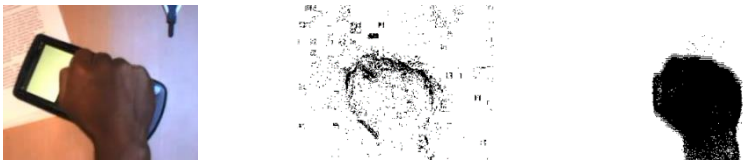
$$B = \arg\left(\min_b \left(\sum_{i=1}^{b} \omega_{i,t} > T \right) \right) \tag{11}$$

where T is the proportion of the background.

When the complex background undergoes massive change, the system may mistake part of the background as the foreground and extract it as a moving object. The resulting virtual shadows will dampen the effect of gesture segmentation (Lopez-Rubio *et al.*, 2015). The virtual shadows can be removed rapidly by increasing the learning rate. However, if the learning rate is fixed at a large value α_1, the moving object will be treated as part of the background if it does not move in a short time.

In view of this, a special scene change rate γ is designed here. If the scene change rate γ_t is above the threshold U, the learning rate should be increased to the large value α_1; if the scene change rate γ_t is below the threshold U, the learning rate should be reduced to the smaller value α_0.

Figure 1 shows the experimental results of the original and improved foreground extraction algorithms. Figure 1(a) presents a partial image of the video frame of the FG-Net database. It can be seen that the hand was recognized as part of the background when it did not move in a short time. Figure 1(b) illustrates the extracted foreground when the learning rate was fixed at a large value α_1. It is clear that the hand almost disappeared from the foreground. Figure 1(c) is the foreground extracted by the improved method.



(a) Original video frame; (b) Results of original algorithm; (c) Results of improved algorithm

Figure 1. Experimental results of foreground extraction.

The great variation in the mean greyscale of the image reveals violent changes of the scene of the video frame at time t, while the small variation in skin colour of the image shows the limited movement of the moving object, i.e. the hand, in the foreground. Hence, there must be great changes to the background of the frame image. Thus, the scene change rate can be calculated as:

$$\gamma_t = \frac{R_t}{S_t} \tag{12}$$

where R_t and S_t are respectively the mean change rate of grayscale and skin colour of the image. The mean grayscale change rate R_t can be obtained as:

$$R_t = \frac{|h_t - h_{t-1}|}{h_{t-1}} \tag{13}$$

where h_t is the mean greyscale of the image at time t. The mean skin colour change rate S_t can be obtained as:

$$S_t = \frac{|H_t - \delta| - |H_{t-1} - \delta|}{|H_{t-1} - \delta|} \tag{14}$$

where δ is the parameter value; H_t is the mean H (hue) of the image at time t. The HSV (hue, saturation, value) is a colour space reflecting the intuitive features of colour. It is widely used for skin colour detection. The component H (hue) depicts the colour information of the image, which reacts relatively slowly to the change in illumination. Besides, the H of skin colour generally falls between 22 and 28 (Figure 2).

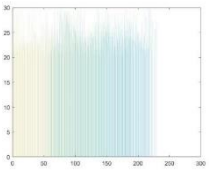


Figure 2. Concentrated area of the H of skin colour in HSV colour space.

3. Algorithm implementation

Firstly, the video was inputted to collect the frame images. After that, these images were pre-processed, and the background was simulated by the improved mixed Gaussian background modelling method. The modelling was followed by rapid and accurate extraction of the foreground, and the segmentation of the gesture. The algorithm is implemented as follows:

4. Experimental verification

4.1. Dataset collection

Two types of experimental data were collected from two sources, namely, a video from the FG-Net database and a video taken by the author with a fixed camera. In the former video, the hand colour of the model was similar to the desk colour, and the

objects on the desk were changed constantly (i.e. the video background is constantly changing). In the latter video, the background was complex and underwent massive changes, which can reflect the improvement effect.

4.2. Adaptive selection of the number of gaussian distribution

Table 1 lists the mean number of frames in the two videos processed per second by the original algorithm and the improved algorithm. It is clear that the improved algorithm was much faster than the original one and reduced the computing load of the mobile terminals.

Table 1. Processing speeds of the original and improved algorithms (fps/s).

Method	Original algorithm	Improved algorithm
FG-Net video	2.36	8.20
The author's video	3.13	10.42

4.3. Controlling the learning rate based on special scene change rate

(a) Original video frame; (b) Results of original algorithm; (c) Results of improved algorithm

Figure 3. Adaptabilities of the original and improved algorithms.

This subsection further compares the foreground extraction effect of the original algorithm and the improved algorithm. The original frame video is shown in Figure 3(a), and the effects of the original and improved algorithms are displayed in Figures 3(b) and 3(c), respectively. As shown in Figure 3, when another person suddenly entered the background, the original algorithm incorrectly extracted the person to the foreground, while the improved algorithm eliminated the virtual shadows rapidly. Hence, the original algorithm cannot adapt to the background changes, while the improved algorithm has strong adaptability to the massive changes of the background.

The original and improved algorithms were further contrasted by three parameters: precision, recall and F1-measure. The precision and recall rate evaluate the quality of the extraction results, while the F1-measure is the harmonic mean of the precision and recall, reflecting the overall performance of the extraction.

High precision indicates the correct rate of the detection is high, and high recall means a high proportion of gestures are correctly detected (i.e. few gestures are not detected). The FI-measure cannot reach a high value unless these parameters are at a high level. As shown in Table 2 below, the proposed algorithm enjoys a high F1-measure.

Table 2. Precisions, recalls and F1-measures of the original and improved algorithms.

Source	Parameter	Original algorithm	Improved algorithm
FG-Net video	Precision (%)	76.65	89.39
	Recall	30.62	75.20
	F1-Measure	44.91	81.33
The author's video	Precision	98.03	97.25
	Recall	48.62	76.13
	F1-Measure	63.86	85.33

5. Conclusions

For rapid gesture segmentation of mobile terminals, this paper proposes a fast, improved mixed Gaussian background modelling method. The proposed method was modified from the traditional mixed Gaussian background modelling method through the optimization of the learning rate and the number of Gaussian distributions. Through experiment, it is proved that the improved algorithm is adaptable to massive changes of the background and works effectively in gesture segmentation.

Acknowledgements

This work was supported in part by National Natural Science Foundation of China under Grant No. 61672461 and No. 61672463.

References

Azzam R., Kemouche M. S., Aouf N., Richardson M. (2016). Efficient visual object detection with spatially global Gaussian mixture models and uncertainties. *Journal of Visual Communication and Image Representation*, Vol. 36, No. 1, pp. 90-106. https://doi.org/10.1016/j.jvcir.2015.11.009

Chen D., Li G., Sun Y. (2017). An Interactive Image Segmentation Method in Hand Gesture Recognition. *Sensors*, Vol. 17, No. 2, pp. 539-550. https://doi.org/10.3390/s17020253

Cui Z. G., Wang H., Li A. H. (2017). Moving object detection based on optical flow field analysis in dynamic scenes. *Acta Physica Sinica*, Vol. 66, No. 8, pp. 56-63. https://doi.org/10.7498/aps.66.084203

Goyal K., Singhai J. (2017). Review of background subtraction methods using Gaussian mixture model for video surveillance systems. *Artificial Intelligence Review*, Vol. 2, No. 1, pp. 246-252. https://doi.org/10.1007/s10462-017-9542-x

Huang W., Liu L., Yue C. (2015). The moving target detection algorithm based on the improved visual background extraction. *Infrared Physics and Technology*, Vol. 71, No. 1, pp. 518-525. https://doi.org/10.1016/j.infrared.2015.06.011

Jian C. F., Wang Y. (2014). Batch Task Scheduling-oriented Optimization Modelling and Simulation in Cloud Manufacturing. *International Journal of Simulation Modelling*, Vol. 13 No. 1, pp. 93-101.

Katsarakis N., Pnevmatikakis A., Tan Z. H. (2016). Improved Gaussian Mixture Models for Adaptive Foreground Segmentation. *Wireless Personal Communications an International Journal*, Vol. 87, No. 3, pp. 1-15. https://doi.org/10.1007/s11277-015-2628-3

Lopez-Rubio F. J., Dominguez E., Palomo E. J., López-Rubio E., Baena R. M. L. (2015). Selecting the Color Space for Self-Organizing Map Based Foreground Detection in Video. *Neural Processing Letters*, Vol. 43, No. 2, pp. 345-361. https://doi.org/10.1007/s11063-015-9431-8

Mangal S., Kumar A. (2016). Real time moving object detection for video surveillance based on improved GMM. *International Journal of Advanced Technology & Engineering Exploration*, Vol. 4, No. 26, pp. 17-22.

Plyer A., Besnerais G. L., Champagnat F. (2016). Massively parallel Lucas Kanade optical flow for real-time video processing applications. *Journal of Real-Time Image Processing*, Vol. 11, No. 4, pp. 713-730. https://doi.org/10.1007/s11554-014-0423-0

Zheng A., Zhang L., Zhang W. (2016). Local-to-global background modeling for moving object detection from non-static cameras. *Multimedia Tools and Applications*, Vol. 76, No. 8, pp. 11003-11019. https://doi.org/10.1007/s11042-016-3565-1