Journal of Information Hiding and Multimedia Signal Processing ©2016 ISSN 2073-4212
Ubiquitous International Volume **7**, Number **1**, January **2016**

Multi-Channel Adaptive Mixture Background Model for Real-time Tracking

Da-Jie Guo, Zhe-Ming Lu* and Hao Luo

School of Aeronautics and Astronautics
Zhejiang University
Hangzhou, 310027, P. R. China
zheminglu@zju.edu.cn

Received August, 2015; revised October, 2015
(Communicated by Zhe-Ming Lu)

Abstract. *Extracting the segmentation of moving regions in image sequences in real-time is a preliminary stage for many computer vision tasks such as video camera surveillance and human-machine interface communication. A typical method for real-time moving region detection is background subtraction. The first step is to construct a background model. Numerous background models have been introduced to solve this problem for different requirements. The most efficient one is Gaussian mixture model, but it still exists some problems. This paper discusses old modeling methods and proposes a new method base on chromatic channels to construct a background model. By reviewing the existing background update equations, we find some inappropriate points and take some modifications. This makes the method adaptive more accurately to a changing environment without consuming too much time. The result reveals that our improved method is efficient and well-performed.*
Keywords: Smart Surveillance, Video Content Analysis, Foreground Extraction

1. **Introduction.** A fixed position camera is a typical application in many visual surveillance systems. In the past, it is limited by computer calculating ability to process video sequence. As a result, many systems are slow and can not process images in real time or only available in a restricted condition. But now, the performance of computers, as the prediction of Moore's law, is growing fast. This makes it possible to design a more complex, efficient, robust and universal background model for analyzing video streams in real time.

The simplest way to gain foreground moving areas is to store a reference image which is generated by calculating the mean of former N frames which contain no moving objects in the scene first. Then subtract the reference image from the new coming frame to get a set of differential values. Finally, set a threshold value to separate the background and foreground pixels. The result is presented as a binary image mask, in which *0* represents the background and *1* represents the foreground. This method encounters many problems such as the difficulty of getting definitely clean reference images and the omission of new background objects. The background objects moving after the training period and the foreground motionless objects during the training period would be considered as permanent foreground objects. In addition, the frequently changing illumination caused by daylight or clouds also impacts the accuracy [1]. Ridder et al. [2] used the Kalman filtering theory to build a background model which recorded the illumination changes of

each pixel. It is a recursive method, so they took many parameter corrections to make the system robust enough for the light changing in the environment. While this method employed the automatical threshold adaption, it still recovered too slowly to handle the multimodal background model. Elgammal et al. [3] processed each pixel by a kernel estimator. Kernel exemplars were extracted from a moving window. They also introduced a method to reduce the result of small motions by bringing in a spatial coherence. It was done by comparing simply connected components with the background model of its circular neighbourhood. Grimson et al. [4, 5, 6] employed an adaptive nonparametric Gaussian mixture model to solve these problems. Their model could also suppress the effect of small repetitive motions such as moving branches or leaves of trees and small camera movements. P.KaewTraKulPong et al [1] proposed an improved Gaussian mixture model. They began to estimate the Gaussian mixture model by expected sufficient statistics update equations. Then they switched to L-recent window version update equations when the first L recent samples were processed. It raised the speed of adaption in the means and the covariance matrices. They also introduced a method to detect moving shadows. Zivkovic [7] also employed an improved method to cancel the influence of the prior video frames. His algorithm could adjust the distribution number as needed. In this way, the processing time of per frame is reduced.

Our background model is based on Grimson et al.'s [4, 5, 6], P. KaewTraKulPongetal et al.'s [1] and Zivkovic's methods[7], so in Section 2.1 and Section 2.2, we review their methods respectively, while our method is presented in Section 2.3. In Section 3, we present some experiments and compare results. In Section 4, we draw the final conclusion.

2. **Modelling Methods.** In this Section, we will discuss Stauffer et al.'s work [4, 5, 6] and KaewTraKulPong et al.'s work[1] and Zivkovic's work[7] below. They all introduced a modelling method for backgrounds based on mixture Gaussian models. The model builds K Gaussian distributions for each pixel and each Gaussian segment represents a possible pixel value distribution. For distinguishing the significance of each Gaussian segment, they designed a weight parameter for every segment of each pixel which reflects the probability of that segment appearing. Then they selected the highest B probable segments as background distributions which order by fitness value. The pixel value stays longer, the more possible it belongs to the background. A pixel belonging to static objects in the scene tents to generate some tight distributions, on the contrary, a pixel of moving object lean to generate some loose distributions due to the different reflecting surfaces during the movement. To ensure the method worked continuously and adapted to the light changing or small repetitive motion, they applied an update strategy. When a new frame is coming, all pixels values are compared with their respective existing Gaussian segments. All segments will be updated and then judge whether the pixel is foreground or not. If no match segment is found, a new Gaussian segment will be created using the unmatched value as mean and two initial values as variance and weight parameter respectively. Then this pixel will be regarded as foreground at that time.

2.1. **Gaussian Mixture Background Model.** Every pixel in a frame is modelled by a mixture of K Gaussian distributions. At any time t, for a pixel at (x, y), which represents x row and y column [4], its history values set is:

$$\{X_1, ..., X_t\} = \{I(x, y, i), 1 \leq i \leq t\} \tag{1}$$

Where I is the image sequence. The probability of this pixel value appearing in time t is:

$$P(X_t) = \sum_{i=1}^{K} w_i \eta(X_t, \mu_i, \sigma_i^2) \tag{2}$$

Where w_i is the estimate of weight of the $i-th$ Gaussian segment. $\eta(X, \mu_i, \sigma_i^2)$ is the Gaussian Distribution probability density function of $i-th$ segment:

$$\eta(X_t, \mu_i, \sigma_i^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(X-\mu_i)^2}{2\sigma^2}} \tag{3}$$

Where μ_i is the mean and σ_i is the variance of $i-th$ segment.

All distributions are ordered by the parameter called fitness value which equals to $\frac{w_i}{\sigma_i}$. It is an effective order reference, because the value increases when the weight increased and the variance decreases [7]. The first B segments are chosen as the background model, where B is:

$$B = arg\min_b(\sum_{k=1}^{b} w_k > T) \tag{4}$$

Where T is a threshold of the minimum part of the data that should be regarded as background model component. Background subtraction is executed by comparing every pixel with each distribution. If one pixel value is in 2.5 standard deviations of any of the B distributions, that pixel is estimated as background. Otherwise, it will be considered as the foreground of the scene.

If current pixel value matches one of existing distributions, all segments will be update by the equations:

$$\widehat{w}_{i,t+1} = (1-\alpha)\widehat{w}_{i,t} + \alpha O_{i,t+1} \tag{5}$$

$$\widehat{\mu}_{i,t+1} = (1-\rho)\widehat{\mu}_{i,t} + \rho X_{t+1} \tag{6}$$

$$\widehat{\sigma}^2_{i,t+1} = (1-\rho)\widehat{\sigma}^2_{i,t} + \rho(X_{t+1} - \widehat{\mu}_{i,t})^2 \tag{7}$$

$$\rho = \alpha\eta(X_{t+1}, \widehat{\mu}_{i,t}, \widehat{\sigma}_{i,t}) \tag{8}$$

$$O_{i,t} = \begin{cases} 1 & \text{if } w_{i,t} \text{ is the first matched distribution} \\ 0 & \text{other} \end{cases} \tag{9}$$

where $\widehat{w}_{i,t}$, $\widehat{\mu}_{i,t}$ and $\widehat{\sigma}_{i,t}$ is the estimation weight ,mean and variance of *i-th* Gaussian segment respectively. α is a constant which reflect the learning rate of this background model.

When a value matches none of existing distributions, it will substitute the last distribution for a new one which composes of a mean equaling that value, a default initial variance and an initial weight. As an unsupervised learning algorithm, only 2 parameters, T and α, need to be preset before processing the video data. For example, if a new object invades into the scene, pixel values in invaded regions change violently, which causes that pixels match small weight segment or even none, so these regions are marked as a part of foreground. If an object maintains in static state long enough, its weight exceed $(1-T)$ and then it can be regarded as a piece of the background. Reviewing update equations, we can figure out that the object must stay in static state for approximate $\log_{1-\alpha} T$ frames. For example, for $T = 0.7$ and $\alpha = 0.005$ we get 71 frames.

2.2. **Improved Gaussian Mixture Model.** Though the Gaussian Model is robust as how it is explained to be in papers [4, 5], it exists a vast promotion roomage. P. KaewTraKulPong et al introduced an improved model in [1].Their estimating of mixture model started by expected sufficient statistics update equations. Then it switched to L-recent windows version after the first L frames had been processed. The initial estimate improves both the accuracy of the estimation and the performance of the tracker allowing fast convergence to a stable state. The L-recent step gives priority over recent data therefore the background model can adapt to changes of scene quickly.

Their expected sufficient statistics update equations(first three) and the L-recent window equations(next three) are shown below:

$$\widehat{w}_{i,t+1} = \widehat{w}_{i,t} + \frac{1}{N+1}(O_{i,t+1} - \widehat{w}_{i,t}) \tag{10}$$

$$\widehat{\mu}_{i,t+1} = \widehat{\mu}_{i,t} + \frac{O_{i,t+1}}{\sum_{k=1}^{t+1} O_{i,t}}(X_{t+1} - \widehat{\mu}_{i,t}) \tag{11}$$

$$\widehat{\sigma}^2_{i,t+1} = \widehat{\sigma}^2_{i,t} + \frac{O_{i,t+1}}{\sum_{k=1}^{t+1} O_{i,t}}((X_{t+1} - \widehat{\mu}_{i,t})^2 - \widehat{\sigma}^2_{i,t}) \tag{12}$$

$$\widehat{w}_{i,t+1} = \widehat{w}_{i,t} + \frac{1}{L}(O_{i,t+1} - \widehat{w}_{i,t}) \tag{13}$$

$$\widehat{\mu}_{i,t+1} = \widehat{\mu}_{i,t} + \frac{1}{L}(\frac{O_{i,t+1}X_{t+1}}{\widehat{w}_{i,t+1}} - \widehat{\mu}_{i,t}) \tag{14}$$

$$\widehat{\sigma}^2_{i,t+1} = \widehat{\sigma}^2_{i,t} + \frac{1}{L}(\frac{O_{i,t+1}(X_{t+1} - \widehat{\mu}_{i,t})^2}{\widehat{w}_{i,t+1}} - \widehat{\sigma}^2_{i,t}) \tag{15}$$

Zoran Zivkovic proposed another improved Gaussian model [7]. He added a parameter c_T which was defined as a prior evidence support coefficient. They started with one component centered on the first sample and new components were added as required. For example, for a chosen $\alpha = 1/T$ you could require that at least $c = 0.01 * T$ samples support a component and you get $c_T = 0.01$. His recursive update equations are:

$$\widehat{w}_{i,t+1} = \widehat{w}_{i,t} + \alpha(O_{i,t+1} - \widehat{w}_{i,t}) - \alpha c_T \tag{16}$$

$$\widehat{\mu}_{i,t+1} = \widehat{\mu}_{i,t} + \frac{\alpha O_{i,t+1}}{\widehat{w}_{i,t+1}}(X_{t+1} - \widehat{\mu}_{i,t}) \tag{17}$$

$$\widehat{\sigma}^2_{i,t+1} = \widehat{\sigma}^2_{i,t} + \frac{\alpha O_{i,t+1}}{\widehat{w}_{i,t+1}}((X_{t+1} - \widehat{\mu}_{i,t})^2 - \widehat{\sigma}^2_{i,t}) \tag{18}$$

2.3. **Our Multi-Channel Mix Background Model.** As we can see, the previous update equations of background model contain some complex components which decrease the calculating speed. For example, $\rho = \alpha\eta(X_t, \mu_i, \sigma_i^2)$ consumes much time, so we will replace ρ with α because of its weak impact. Another weakness is that the previous models all constructed the Gaussian distributions based on only one channel value, but most mainstream surveillance systems provide chromatic video streams. So taking advantage of rest channels will enhance the performance of the algorithm.

In our model, only the matched one is updated at one time and other segments maintain the last state. The multi-channel update equations are shown below:

$$\widehat{w}_{i,t+1} = \widehat{w}_{i,t} + \alpha(1 - \widehat{w}_{i,t}) \tag{19}$$

$$\widehat{\mu}_{i,t+1} = \widehat{\mu}_{i,t} + \alpha(X_{t+1} - \widehat{\mu}_{i,t}) \tag{20}$$

$$\widehat{\sigma}^2_{i,t+1} = \widehat{\sigma}^2_{i,t} + \alpha((X_{t+1} - \widehat{\mu}_{i,t})^2 - \widehat{\sigma}^2_{i,t}) \tag{21}$$

For a pixel, if it contain R, G, B three channels, $\mathbf{X}_t = (x_{t,1}, x_{t,2}, x_{t,3})^T$, our method will record and update mean and variance of all channels using equations above but only one weight for each pixel. When a new sample frame comes, the judging criteria is:

$$\sum_{c=1}^{C}(x_{t+1,c} - \widehat{\mu}_{t,c})^2 < T_{var}\sum_{c=1}^{C}\widehat{\sigma}^2_{t,c} \tag{22}$$

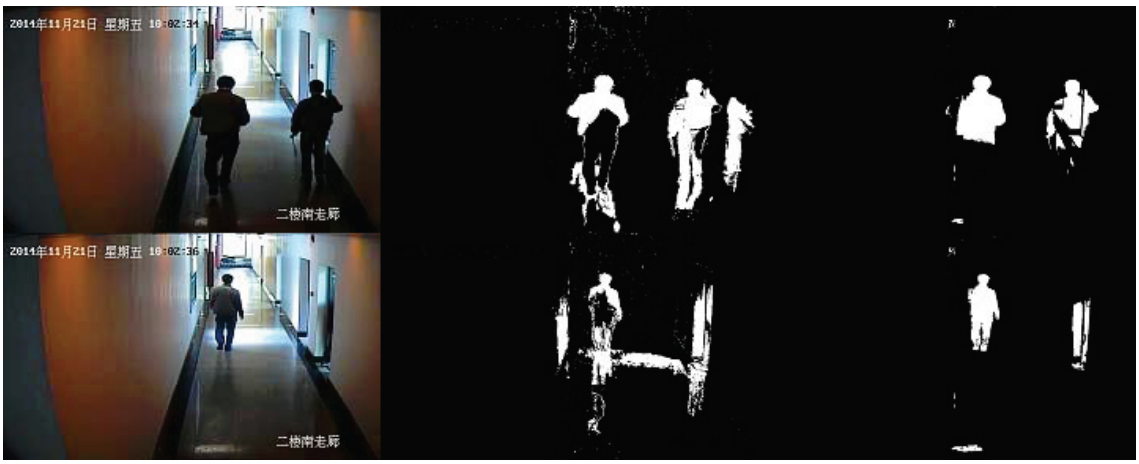

FIGURE 1. The first column presents origin images at 50th, 100th frame respectively, the second column is the result using Zoran Zivkovic's method and the last column is the result of ours

and the fitness value is calculated by:

$$fitness = \frac{w}{\sqrt{\sum_{c=1}^{C} \widehat{\sigma}_c^2}} \tag{23}$$

where C is the total channel number of the video stream and T_{var} is the variance threshold constant(In ours algorithm we choose 3^2). After the update procedure is completed, the sum of weights may not equal to one. To deal with it, our algorithm do a normalization.

3. **Experiments.** In this section, we analyze the performance of our algorithm. The maximum number of Gaussian segments N is configured to 5, the update learning rate α is set to 0.005 and the background weight threshold T is set to 0.7.

We select a fragment of real surveillance indoor scene image sequences to test our model. This scene contains the variable light casting from the window at the end of the corridor and many pedestrians walking from one side to another side or going into the door at right. Compared with the algorithm of Zoran Zivkovic [3], our model detects the foreground objects more accurately and Zoran Zivkovics exists more noise pixels. Four image samples are presented in Figure 1 and Figure 2, two pedestrians, one pedestrians at left, no people, one pedestrians at right respectively. The resolution of this video is 960x576 and the running environment is a 2.93GHz Intel E7500 PC. Every frame consumes about 50 milliseconds and Zoran Zivkovic's consumes about 26 milliseconds.

4. **Conclusion.** In this paper, we propose a new background model method for multi-channel image sequences based on Stauffer et al and rebuild a real-time update scheme for tracking the moving things. Our method only need two preset parameters, background learning rate α and threshold T, as same as many other background subtraction algorithms. We compare ours with another improved Gaussian mixture model proposed by Zivkovic. The results show that ours perform better. Though the consumption of time is higher, about 20 frames per second, it is enough for real-time tracking.

Acknowledgments. This work is supported by the National Natural Science Foundation of China under grant 61171150, and also supported by Zhejiang Provincial Natural Science Foundation of China under grant R1110006.

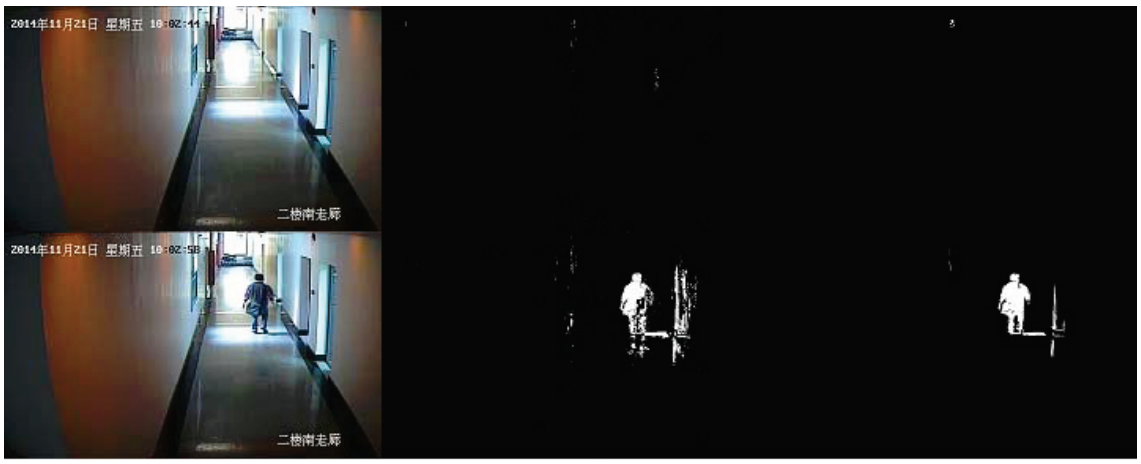

FIGURE 2. The first column presents origin images at 300th, 650th frame respectively, the second column is the result using Zoran Zivkovic's method and the last column is the result of ours

REFERENCES

[1] P. KaewTraKulPong, and R. Bowden, An improved adaptive background mixture model for realtime tracking with shadow detection, *Proceedings of the 2nd European Workshop on Advanced Video-Based Surveillance Systems*, 2001.

[2] C. Ridder, O. Munkelt, and H. Kirchner, Adaptive background estimation and foreground detection using Kalman-filtering, *Proceedings of International Conference on Recent Advances in Mechatronics*,ICRAM95, UNESCO Chair on Mechatronics, pp. 193-199, 1995.

[3] A. Elgammal, D. Harwood, and L. Davis, Non-parametric model for background subtraction, *IEEE IEEE International Conference on Computer Vision FRAME-RATE WORKSHOP*, 1999.

[4] C. Stauffer, and W. E. L. Grimson, Adaptive background mixture models for real-time tracking, *Proceedings of 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol.2, 1999.

[5] W. E. L. Grimson, C. Stauffer, R. Romano, and L. Lee, Using adaptive tracking to classify and monitor activities in a site, *Proceedings of 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Barbara, CA, June, pp. 22-29, 1998.

[6] C. Stauffer, and W. E. L. Grimson, Learning patterns of activity using real-time tracking, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.22, no.8, pp. 747-757, 2000.

[7] Z. Zivkovic, Improved adaptive gaussian mixture model for background subtraction, *Proceedings of the 17th International Conference on Pattern Recognition*, Cambridge, England, pp.28-31, 2004.

[8] P. Withagen, K. Schutte, and F. Groen, Likelihood-based object tracking using color histograms and EM, *Proceedings of IEEE International Conference on Image Processing*, Rochester, NY, USA, pp. 589-592, 2002.

[9] A. Prati, I. Mikic, M. M. Trivedi, and R. Cucchiara, Detecting moving shadows: algorithms and evaluation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 7, pp. 918-923, 2003.

[10] E.Hayman, and J. Eklundh, Statistical background subtraction for a mobile observer, *Proceedings of the 9th IEEE International Conference on Computer Vision*, Nice, France, pp. 67-74, 2003.

[11] A. Monnet, A. Mittal, N. Paragios, and V. Ramesh, Background modeling and subtraction of dynamic scenes, *Proceedings of the 9th IEEE International Conference on Computer Vision*, Nice, France, pp.1305-1312, 2003.