Journal of Southeast University (English Edition) Vol. 29, No. 4, pp. 379 - 383 Dec. 2013 ISSN 1003—7985

Adaptive moving target detection algorithm based on Gaussian mixture model

Yang Xin[1,2,3] Liu Jia[1] Fei Shumin[2] Zhou Dake[1]

([1]College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)
([2]School of Automation, Southeast University, Nanjing 210096, China)
([3]Key Laboratory of Photoelectric Control Technology, Luoyang 471000, China)

Abstract: In order to enhance the reliability of the moving target detection, an adaptive moving target detection algorithm based on the Gaussian mixture model is proposed. This algorithm employs Gaussian mixture distributions in modeling the background of each pixel. As a result, the number of Gaussian distributions is not fixed but adaptively changes with the change of the pixel value frequency. The pixels of the difference image are divided into two parts according to their values. Then the two parts are separately segmented by the adaptive threshold, and finally the foreground image is obtained. The shadow elimination method based on morphological reconstruction is introduced to improve the performance of foreground image's segmentation. Experimental results show that the proposed algorithm can quickly and accurately build the background model and it is more robust in different real scenes.

Key words: moving target detection; Gaussian mixture model; background subtraction; adaptive method

doi: 10.3969/j.issn.1003 - 7985.2013.04.005

Moving target detection is the first step in information extraction in computer vision applications. At present, the common moving target detection algorithms are the optical flow method[1], the inter-frame difference method[2] and the background subtraction method[3].

The background subtraction method is to achieve the moving target detection through the difference between the current frame and the background image. The method is simple in principle and it is easy to implement. It can provide a more complete characterization data, but the background images are more sensitive to changes of light and external conditions.

At present, the commonly used background subtraction method is based on the Gaussian model. Pfinder[4] used the single Gaussian statistical model to extract the background. Stauffer et al.[5] modeled background by the Gaussian mixture model. This method can handle the interference such as light changes and background confusion. In order to reduce the error rate of detection in the complex environment, Zhang et al.[6] used the support vector machine to make further judgment for foreground pixels on the basis of the GMM. Lee et al.[7] increased the convergence speed by calculating the adaptive learning rate for each pixel. In Ref. [8], we used the global threshold method to segment the difference image, but the detection effect is not ideal when the target itself changes greatly.

This paper models the background by the GMM and adaptively estimates the number of Gaussian distributions. After obtaining the difference image, the pixels are adaptively divided. Then they are binaried respectively. The threshold here is also adaptive. After the above processes, this system can extract the foreground objects and it has strong robustness.

1 Gaussian Mixture Model

A particular pixel in the image sequence (x_0, y_0) and its values are assumed to be $(x_1, \ldots, x_i, \ldots, x_t)$ at time 1 to t. Each pixel value is modeled by a mixture of K Gaussian distributions. The probability of observing the current pixel value can be written as

$$p(x_t) = \sum_{i=1}^{k} w_{t,i}\eta(x_t, u_{t,i}, \Sigma_{t,i}) \tag{1}$$

where $w_{t,i}$ is an estimate of the weight of the i-th component at time t, and $\sum_{i=1}^{k} w_{t,i} = 1$; $u_{t,i}$ is the mean value of the i-th component at time t; $\Sigma_{t,i}$ is the covariance matrix of the i-th component at time t. And η is a Gaussian probability density function

$$\eta(x_t, u_{t,i}, \Sigma_{t,i}) = \frac{1}{(2\pi)^{n/2} \left| \Sigma_{t,i} \right|^{1/2}} \cdot \exp\left(-\frac{1}{2}(x_t - u_{t,i})^{\mathrm{T}} \sum_{t,i}^{-1} (x_t - u_{t,i}) \right) \tag{2}$$

where n is the dimensions of a Gaussian distribution function and $\Sigma_{t,i} = \sigma_i^2 I$.

When a new pixel arrives at time $t+1$, x_{t+1} needs to be

Received 2013-06-17.

Biography: Yang Xin (1978—), male, doctor, associate professor, yangxin@nuaa.edu.cn.

Foundation items: The National Natural Science Foundation of China (No. 61172135, 61101198), the Aeronautical Foundation of China (No. 20115152026).

Citation: Yang Xin, Liu Jia, Fei Shumin, et al. Adaptive moving target detection algorithm based on Gaussian mixture model[J]. Journal of Southeast University (English Edition), 2013, 29(4): 379 - 383. [doi: 10.3969/j.issn.1003 - 7985.2013.04.005]

compared with the mean values of K distributions. The first step is to choose the matched distribution by some rule. The rule is

$$|x_{t+1} - u_{t,i}| < c\sigma_{t,i}^2 \qquad i = 1, 2, \cdots, k \tag{3}$$

where c is the standard threshold, and usually $c = 2.5$ is appropriate. If the i-th distribution satisfies the rule, the current pixel value matches the i-th distribution. The parameters of the distribution which matches the new pixel are updated as[9]

$$\mu_{t,i} = (1-\rho)\mu_{t-1,i} + \rho x_t \tag{4}$$

$$\sigma_{t,i}^2 = (1-\rho)\sigma_{t,i}^2 + \rho(x_t - \mu_{t,i})^2 \tag{5}$$

$$\rho = \alpha w_{t,i} \tag{6}$$

where α is the learning rate, $0 < \alpha < 1$; ρ is the second learning rate.

The mean values and variances of the distributions which are not matched remains unchanged. The weights of K distributions are updated as

$$w_{t,i} = (1-\alpha) w_{t-1,i} + \alpha M \tag{7}$$

The value of M is denoted as 1 for the model which is matched and 0 for the remaining models. In order to improve the reliability of the background model, the weights need to be normalized, $\widetilde{w}_{t,i} = w_{t,i} / \sum_{i=1}^{k} w_{t,i}$. Then K distributions are ordered in a descending sort by the value of $\widetilde{w}_{t,i}/\sigma_{t,i}$. The first B distributions are used as the background model, and B is estimated as

$$B = \underset{b}{\operatorname{argmin}}\left(\sum_{i=1}^{b} \widetilde{w}_{t,i} > T_1\right) \tag{8}$$

where the threshold T_1 is a measure of the minimum fraction of the background model.

2 Improved Algorithm

2.1 Number of distributions

The background of the real scene is complex, and different pixel values with varying degrees of confusion keep changing all the time. If the number of the Gaussian distributions is fixed, the big number leads to the inefficiency of the algorithm, and the small number leads to the bad accuracy of the algorithm. So, a single Gaussian model is used for the pixels that remain almost unchanged. If the background pixel changes greatly, the number of Gaussian distributions will increase adaptively. Oppositely, the number of distributions will be decreased when the background pixel tends to be steady. The maximum number of distributions is $K_{\max}$. If the current k distributions are not matched with the target pixel, and $k < K_{\max}$, then $k = k + 1$, and a new distribution will be generated and added to the background model. However, if $k = K_{\max}$, the new distribution will not replace the one whose weight is the minimum immediately until its weight is greater than the minimum one. Before replacement, the mean value and variance of the new distribution will be updated as Eqs. (4) and (5), and the weight will be updated as Eq. (7). α is replaced by the sigmoid function[10] as

$$\alpha = \frac{1}{1 + e^{0.1(T_2 - j)}} \tag{9}$$

where constant T_2 is used to control the time that foreground pixels are incorporated in the background after remaining stable; variable j is used to count, which records the matching times between observation and the new Gaussian distribution. After the new distribution is added to the background, variable j is cleared, and the new k weights need to be normalized.

In the classical algorithm, the first successful match is used as the match result. In fact, the new pixel may be matched with several distributions, and the first match may not be the optimal one. In this paper, each distribution will be matched with the new pixel to find the optimal match. The optimal match can be obtained as

$$L = \operatorname{argmin}\left\{\frac{|x_t - \mu_{t,i}|}{\sigma_{t,i}}\right\} \tag{10}$$

If the optimal match of continuous 50 frames is the i-th distribution, and $k > 1$, then $k = k - 1$, and the distribution whose weight is the minimum will be removed directly.

Fig. 1 shows the adaptively chosen process for Gaussian distributions.

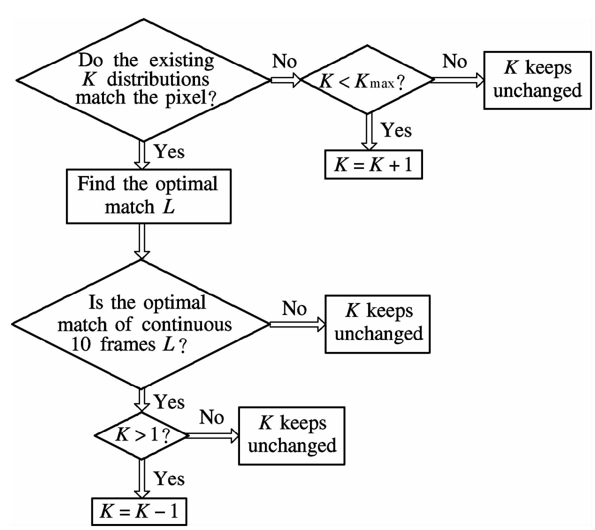

Fig. 1 Adaptively chosen process for K

2.2 Background update

2.2.1 Update of σ^2

The update of the variance parameter σ^2 in Eq. (5) is a decreasing function which converges to zero. This shows that the description of the random process for background tends to a steady state with the accumulation of time. But in reality, there exist noise, camera-shake, etc. The high stability easily leads to the detection misjudgment that occurs on the next time at the same position. So, some modifications have been made for the update of σ^2; that is, σ^2 will be kept remaining relatively stable when it converges to about 15% of the initial value. Fig. 2 shows the update process of σ^2.

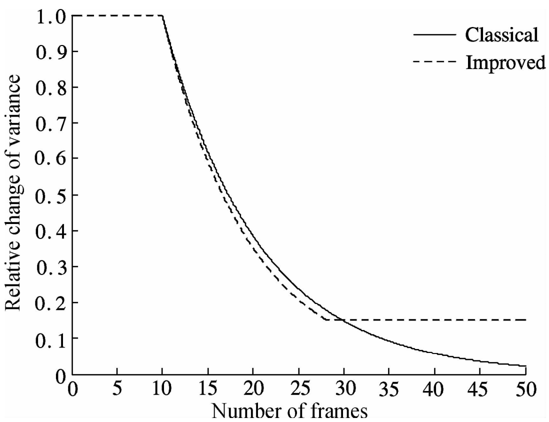

Fig. 2 Update process of σ^2

2.2.2 Background update for light mutation

When the light suddenly changes, most of the background points will be misinterpreted as foreground points by the classical algorithm. So, the proportion of pixels with a large change of the current frame in the total number of pixels δ is calculated. If δ is greater than threshold T_3 (70% is appropriate), it shows that the light has changed. Then the Gaussian distribution k_{b1} with the highest priority in all the pixels needs to be updated as follows:

$$\mu_{t,b1} = f_k(x,y),\ w_{t,b1} = \max(w_{t,i}),\ \sigma^2_{t,b1} = 30 \tag{11}$$

where $f_k(x,y)$ is the current frame, and $\max(w_{t,i})$ is the maximum weight.

2.3 Adaptively threshold binarization

With the establishment of the background model, the difference image can be obtained after the subtraction of the current frame and the background frame as

$$D_k(x,y) = |f_k(x,y) - f_b(x,y)| \tag{12}$$

where $D_k(x,y)$ is the difference image, and $f_b(x,y)$ is the background frame.

When the difference of grayscale between the foreground object and the background is great, there exists an obvious bimodal in the gray histogram. Then the global threshold method is a good way to obtain the segmentation result. But when the foreground target image is more complex, such as when the grayscale of the target object changes greatly, or it is close to the grayscale of the background, there will not be an obvious bimodal in the gray histogram; thus, the global threshold method will not work well. To solve the problem above mentioned, segmentation with an adaptive threshold is used in this paper. The basic idea is to classify the pixels of the difference image, and then different thresholds will be adopted in segmentation processing for different parts of the difference image.

In this paper, pixels are classified into two parts as follows:

$$D_k(x,y) \in D_{k1}(x,y) \qquad D_k(x,y) < T_4 \tag{13}$$

$$D_k(x,y) \in D_{k2}(x,y) \qquad D_k(x,y) \geqslant T_4 \tag{14}$$

$$T_4 = \frac{1}{4}(\mathrm{median}[D_k(x,y)] + \max[D_k(x,y)] + \min[D_k(x,y)]) \tag{15}$$

where T_4 is the adaptive threshold; $D_{k1}(x,y)$ represents the part of the image whose pixels are less than T_4; $D_{k2}(x,y)$ represents the part whose pixels are greater than T_4. Finally, the foreground image can be obtained by segmentation of classified images. As the two parts have different features, different thresholds need to be used. T_5 is assumed to be the threshold for $D_{k1}(x,y)$, and T_6 is for $D_{k2}(x,y)$. The process is

$$R_{k1}(x,y) = \begin{cases} 0 & D_{k1}(x,y) < T_5 \\ 1 & D_{k1}(x,y) \geqslant T_5 \end{cases} \tag{16}$$

$$R_{k2}(x,y) = \begin{cases} 0 & D_{k2}(x,y) < T_6 \\ 1 & D_{k2}(x,y) \geqslant T_6 \end{cases} \tag{17}$$

$$T_5 = |\mathrm{median}[D_{k1}(x,y)] - 0.5\max[D_{k1}(x,y)] - 0.5\min[D_{k1}(x,y)]| \tag{18}$$

$$T_6 = T_4 + 3 \times 1.4826 \times \mathrm{median}[D_{k2}(x,y) - \mathrm{median}[D_{k2}(x,y)]] \tag{19}$$

where $R_{k1}(x,y)$ is the foreground image of part $D_{k1}(x,y)$; $R_{k2}(x,y)$ is the foreground image of part $D_{k2}(x,y)$, and T_5, T_6 are adaptive thresholds.

2.4 Shadow elimination

The shadow is often mistaken as the foreground object. Its existence brings great difficulties in target detection. The HSV color space is an intuitive color model, and it can quantify the difference between the shadow and the moving target. The determination whether a pixel is a shadow pixel or not is as follows:

$$\mathrm{SP}(x,y) = \begin{cases} 1 & \alpha_s \leqslant \dfrac{I^{\mathrm{V}}(x,y)}{B^{\mathrm{V}}(x,y)} \leqslant \beta_s \cap (I^{\mathrm{S}}(x,y) - B^{\mathrm{S}}(x,y)) \leqslant \tau_{\mathrm{S}} \cap |I^{\mathrm{H}}(x,y) - B^{\mathrm{H}}(x,y)| \leqslant \tau_{\mathrm{H}} \\ 0 & \text{others} \end{cases} \tag{20}$$

where $I^{\mathrm{H}}(x,y)$, $I^{\mathrm{S}}(x,y)$ and $I^{\mathrm{V}}(x,y)$ are the HSV components of the pixel value $I(x,y)$ at coordinate (x,y) in the input image; $B^{H}(x,y)$, $B^{S}(x,y)$ and $B^{V}(x,y)$ are the HSV components in the background model. If $I(x,y)$ is determined as a shadow pixel, $SP(x,y)$ is set to be 1; otherwise 0. The two parameters α_S and β_S satisfy the equation, $0<\alpha_S<\beta_S<1$. The value of α_S depends on the power of the shadow; β_S is used to enhance the robustness to noise, which makes the brightness of the current frame not be too close to the background. The parameter τ_S is less than 0, and the choice of the parameters τ_S and τ_H is done empirically.

3 Experiments and Results

To verify the detection results of the proposed algorithm, experiments are performed by Matlab. In the experiments, the maximum number of distributions K_{max} is set to be 5; variance σ_0 is set to be 30; and weight ω_0 is set to be 0.2.

Fig. 3 shows a sequence of the outdoor scene. Fig. 3(a) shows one frame of the scene, and Fig. 3(b) shows the background frame. Fig. 3(c) is the foreground image obtained by the classical GMM. Fig. 3(d) is the foreground image obtained by the improved algorithm. We can find that there exist noise and holes in Fig. 3(c). So the improved algorithm can effectively suppress noise and cavitation.

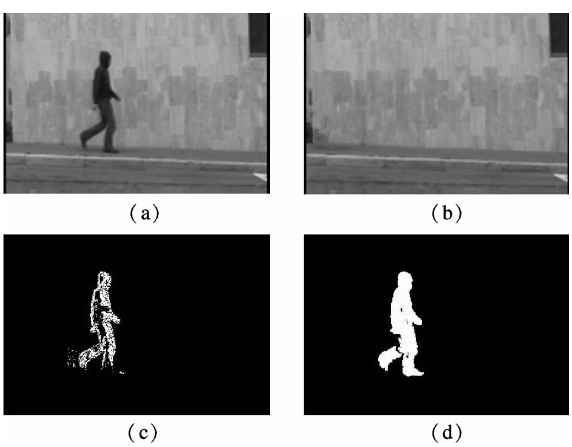

Fig. 3 Detection results of outdoor scene. (a) Original input image; (b) Background frame; (c) Result of classical GMM; (d) Result of improved algorithm

Fig. 4 shows a sequence of the indoor scene, and the part of the moving target's grayscale is close to the background. As the light is strong, there exists shadow, and the shadow underfoot is obvious. This experiment is to verify the ability to remove the shadow and the adaptivity of the algorithm proposed in this paper.

Fig. 4(c) is the result of the classical GMM, and Fig. (d) is the result of the improved algorithm. We can see that the shadow is detected as the part of the target. By the shadow elimination, the improved algorithm successfully removes the shadow. As the target's grayscale is close to the background, there are holes in detection results. The improved algorithm makes the holes disappear by a smaller segmentation threshold used in the region whose grayscale is close to the background.

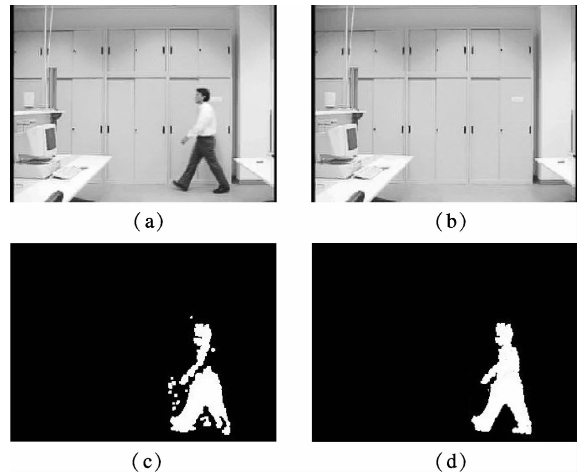

Fig. 4 Detection results of indoor scene. (a) Original input image; (b) Background frame; (c) Result of classical GMM; (d) Result of improved algorithm

Fig. 5 also shows a sequence of the indoor scene, and the difference is that the grayscale of the target itself is small and does not change a lot. In the classical algorithm, the lighter part of the target is recognized as background, causing the detection error. The improved algorithm adjusts the thresholds adaptively, and it can reduce the holes.

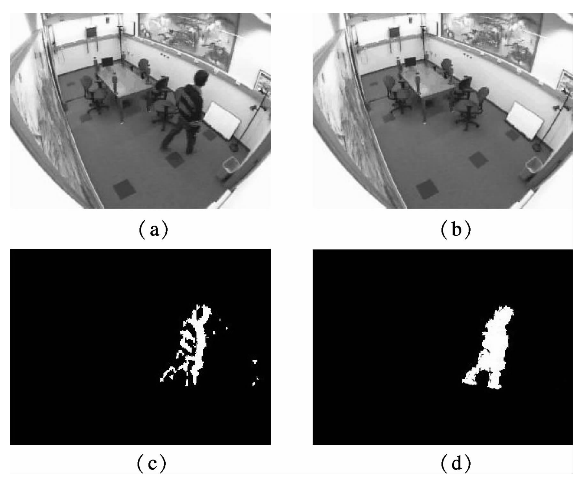

Fig. 5 Detection results of indoor scene. (a) Original input image; (b) Background frame; (c) Result of classical GMM; (d) Result of improved algorithm

4 Conclusion

In this paper, we present a detection algorithm. First, we model the background by the GMM, and the number of distributions here is estimated adaptively. Secondly, we classify the pixels of the difference image into two categories, and each category has its own segmentation threshold. Finally, we obtain the foreground image after shadow elimination.

There are several problems that should be resolved in future research works. In this paper, we apply three videos in the experiments, and the results are not bad. But we cannot guarantee whether another video can be detected well or not. So, we will test more videos to find a more accurate threshold formula. And the three videos tested in experiments are shot by static camera, but if the camera shakes, will the algorithm proposed in this paper work well? How to make the algorithm more extensively self-adaptive needs to be studied further.

References

[1] Zhang J Y, Barron J L. Optical flow at occlusion[C]// *The Ninth Conference on Computer and Robot Vision*. Toronto, Canada, 2012: 198 - 205.

[2] Tsai D M, Lai S C. Independent component analysis-based background subtraction for indoor surveillance [J]. *IEEE Transactions on Image Processing*, 2009, **18**(1): 158 - 160.

[3] Piccardi M. Background subtraction techniques: a review [C]// *IEEE International Conference on Systems, Man and Cybemeties*. Sydney, Australia, 2004, **4**: 3099 - 3104.

[4] Wren C R, Azarbayejani A, Darrell A, et al. Pfinder: real-time tracking of the human body [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 780 - 785.

[5] Stauffer C, Grimson W E L. Adaptive background mixture models for real-time tracking[C]// *IEEE Conference on Computer Vision and Pattern Recognition*. Fort Collins, USA, 1999: 246 - 250.

[6] Zhang J, Chen C H. Moving objects detection and segmentation in dynamic video backgrounds [C]// *IEEE Conference on Technologies for Homeland Security*. Woburn, MA, USA, 2007: 64 - 69.

[7] Lee D S. Effective Gaussians mixture learning for video background subtraction[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, **27**(5): 827 - 832.

[8] Li G, Zeng R L, Lin L. Moving target detection in video monitoring system[C]// *Proceedings of the Sixth World Congress on Intelligent Control and Automation*. Dalian, China, 2006: 9778 - 9781.

[9] Xia Y Q, Ning S H, Shen H. Moving targets detection algorithm based on background subtraction and frames subtraction [C]// *IEEE International Conference on Industrial Mechatronics and Automation*. Wuhan, China, 2010: 122 - 125.

[10] Aboueldahab T, Fakhreldin M. Adaptive control of dynamic nonlinear systems using sigmoid diagonal recurrent neural network [C]// *IEEE International Conference on Systems, Man and Cybernetics*. Istanbul, Turkey, 2010: 4341 - 4345.

一种自适应的基于混合高斯模型的运动目标检测算法

杨　欣[1,2,3]　刘　加[1]　费树岷[2]　周大可[1]

([1] 南京航空航天大学自动化学院, 南京 210016)
([2] 东南大学自动化学院, 南京 210096)
([3] 光电控制技术重点实验室, 洛阳 471000)

摘要:为提高运动目标检测的可靠性,提出了一种自适应的基于混合高斯模型的运动目标检测算法. 该算法利用混合高斯分布对每个背景像素建模,高斯分布的个数不是固定不变的,而是随着像素值的混乱程度自适应变化. 差分图像的像素按大小被分为 2 部分,然后对这 2 部分分别进行自适应阈值化分割,得到前景图像. 利用基于形态学重构的阴影消除方法来改善前景图像分割的性能. 不同实际场景的实验结果表明该算法能够快速准确地建立背景模型,且具有更强的鲁棒性.

关键词:运动目标检测;高斯混合模型;背景差分;自适应方法

中图分类号:TP391.41