2018 IEEE 7th Data Driven Control and Learning Systems Conference
May 25-27, 2018, Enshi, Hubei Province, China

Moving Object Real-time Detection and Tracking Method Based on Improved Gaussian Mixture Model

Shanliang Zhu[1,3], Xin Gao[1], Haoyu Wang[1], Guangwei Xu[1], Qiuling Xie[2], Shuguo Yang[3*]

1. Research Center for Mathematical Modeling, School of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao 266061, China
E-mail: zhushanliang@qust.edu.cn

2. Finance Office, Qingdao University of Science and Technology, Qingdao 266061, China
E-mail: qustmaths@126.com

3. Institute of Intelligence Science & Data Technology, School of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao 266061, China
*Corresponding author, E-mail: ysg_2005@163.com

Abstract: In order to improve the reliability of moving objects detection and tracking, this paper presents a method for moving object real-time detection and tracking based on Vibe and Gaussian mixture model (GMM). This method uses the "Virtual" background model that is trained by video sequence instead of the first frame image for background modeling. And then the foreground object is extracted based on the pixel classification. Finally, according to the morphological method, the clearer moving targets are conducted to realize the real-time detection and tracking. The experimental results show that, in comparison with the current mainstream background subtraction techniques, our approach effectively works on a wide range of complex scenarios, with faster detection speed and more reliable detection results.

Key Words: moving object detection, Vibe, Gaussian mixture model, pixel classification

1 Introduction

Moving object detection has increased strikingly over the last decade. In order to detect and track objection automatically in videos, several algorithms are provided, which include background difference method [1-2], frame difference method [3], etc. Frame difference method is the basic principle of background subtraction, which compares a static background frame with the current frame of a video scene pixel by pixel. Background difference method is the most widely used due to simplicity of its principle. However, algorithms mentioned above can be extremely easily affected by the noise.

In this case, Barnich and Van Drogenbroek proposed a novel target detection algorithm based on background subtraction Vibe algorithm [4]. Compared with the traditional target extraction algorithm, it has the advantages of strong robustness, high accuracy and low hardware requirement. However, the algorithm always remains ghosting in the background in case static objects become moving.

Chen L. detected the ghosting according to the difference between ghosting and background [5]. Hu Z. H. improved the update probability and accelerated the rate of eliminating ghosting [6]. Through the analysis of foreground histogram and pixel variation, Li X. J. could determine the similarity of moving object in order to distinguish objects between moving one and ghosting [7]. Zhang K. used current frame to maintain background model [8], and Stauffer C. modified wrong background model to eliminate ghosting [9].

However, methods mentioned above have not been considered as fundamental method of eliminating ghosting, which is the main motivation of this paper.

In this technical note, we present a universal method for background subtraction. As the basis of ghosting origin, we exploit distribution randomness principle of Gaussian mixture model (GMM) to train video sequences, and obtain the initial background model based on probability distribution [10]. Then, based on dynamic threshold and dynamic update probability, the foreground object can be extracted by pixel classification method. The experiment result is given to illustrate that the method used in the paper is effective to eliminate ghosting in terms of both computation speed and detection rate.

This paper is organized as follows. In Section II, this method has been briefly described. Section III, we have implemented some methods to compare them with our model. We have described this technique and details our major innovations. Finally, Section IV is the conclusion of this paper.

2 Detection of a Universal Background Subtraction Technique

MS Word Authors: please try to use the paragraph styles contained in this document.

2.1 The Basic Principles of Vibe

Vibe is a machine recognition algorithm based on single frame video sequences, and is also the first algorithm to introduce stochastic theory and neighborhood correlation to video recognition. Vibe mainly consists of three parts:

*This work is supported by Shandong Province Key Research and Development Planning Project (2015GGX101020), the Research Project of Teaching Reform in Undergraduate Colleges and Universities in Shandong Province (Z2016Z005), and the Project of Shandong Province Education Science in 12th Five-Year Plan (YBS15014).

978-1-5386-2618-4/18/$31.00 ©2018 IEEE

background model initialization, foreground object extraction, and background template update.

Vibe is essentially a pixel classification algorithm that divides pixels into foreground targets and background targets.

The algorithm initializes the background model using the first frame in the video sequence. Then this algorithm calculating the Euclidean distances of the pixels in the two-dimensional space as the similarity between the new pixel and the sample set. Finally, according to the similarity to determine whether the pixel belongs to the foreground target. During the background update, when a pixel is judged as the foreground object, the sample set of the point need not to be updated. Otherwise, the sample set needs to be updated. The update mechanism follows the memory-less background update strategy.

2.2 The Establishment of the Background Model

The traditional Vibe algorithm uses the first frame for background modeling. When the first frame contains moving objects, it will left a hole behind in the background referred to as a ghost, resulting in Misjudgment of moving object [11-12]. To solve the problems of the classical Vibe algorithm, such as ghost, shadow and noise interference in moving object detection and tracking, this paper proposes an improved Vibe algorithm which combines GMM based on the probability of random multi-frame video sequence for background modeling.

The probability of the its pixel $x(i,t)$ in each frame of the video sequence can be expressed by a mixed Gaussian distribution, and the probability density function is as follows

$$p(X_{i,t}) = \sum_{k=1}^{K} \omega_{i,k,t} \frac{1}{(2\pi)^{n/2} |\sum_{i,k,t}|^{1/2}} e^{-\frac{1}{2}(X_{i,k}-\mu_{i,k,t})} \times \sum_{i,k,t}^{-1} (X_{i,k}-\mu_{i,k,t}), \quad (1)$$

where $\mu_{i,k,t}$ are the estimates of the means and $\sum_{i,k,t}$ are the estimates of the variances that describe the Gaussian components. $\omega_{i,k,t}$ is the weight parameter of the Kth Gaussian component. K is the number of the GMM, and is 3~5 [13].

Assuming that the three colors in RGB space are independent of each other, the covariance can be expressed as

$$\sum_{i,k,t} = \sigma^2_{i,k,t} I. \quad (2)$$

We estimate the background model by the first B largest clusters. Supposing that the components are sorted to have descending weights $\frac{\omega_k}{\sigma_k}$, we have

$$B = \arg\min(\sum_{k=1}^{K} \frac{\omega_k}{\sigma_k} > T), \quad (3)$$

where T is a measure of the minimum potion of the data that belongs to the background.

2.3 Updating the Background Model Over Time

Firstly, during the initialization of the model, we select the first frame of the corresponding video sequence, and initialize with the pixel values of each point in RGB space as the mean of K Gaussian components.

$$\omega_{i,k,t} = \frac{1}{k} \quad (4)$$

The initialized model is an inaccurate model. In order to represent the background of the video sequence, the initialized model needs to be continually update its model parameters. Judgment rules are as follows

$$|X_i - \mu_{k,j}| < \lambda \sigma_{k,j}, \quad (5)$$

where λ is a constant, usually takes 2.5~3 [13]. $\sigma_{k,j}$ is the standard deviation.

If the current pixel points match the K Gaussian component, the parameter do not need to be updated. Otherwise, the parameter will be updated by the following equations

$$\begin{aligned} \omega_k^{n+1} &= \omega_k^n + \frac{1}{1+n}(p(\omega_k \mid x_{n+1}) - \omega_k^n), \\ \mu_k^{n+1} &= \mu_k^n + \frac{p(\omega_k \mid x_{n+1})}{\sum_{i=1}^{n+1} p(\omega_k \mid x_{n+1})}(\mathrm{x}_{n+1} - \mu_k^n), \\ \sum\nolimits_k^{n+1} &= \sum\nolimits_k^n + \frac{p(\omega_k \mid x_{n+1})}{\sum_{i=1}^{n+1} p(\omega_k \mid x_{n+1})} \\ &\times((\mathrm{x}_{n+1} - \mu_k^n)(\mathrm{x}_{n+1} - \mu_k^n)^T - \sum\nolimits_k^n). \end{aligned} \quad (6)$$

For pixels that do not match any of the K Gaussian components, a new Gaussian component will be introduced instead of the smallest ω_k of the original K Gaussian components.

2.4 Pixel Model and Classification Process

1) The establishment of the sample set

Step1 Select one pixel (x,y) in the X-background which is trained by video frame and the pixel N surrounding the pixel together form the sample set $M(x,y)$ of the point. The elements in the sample set are all derived from the field of the pixel. The sample set corresponding to pixel $i(x,y)$ is

$$M_i(x,y) = \{i_1(x,y), i_2(x,y), \cdots i_n(x,y)\}.$$

Step2 Define a circular area $\mathrm{S}_R(i(x,y))$ with $i(x,y)$ as its center and R as its radius. We calculate the Euclidean distance between pixels and each element in the sample set, and when the distance is less than R, we add an approximate sample point. The total number of final approximate sample points is $\#P$ .When $\#P$ is greater than the given threshold value $\#_{\min} P$, the pixel is regarded as the background. When it is less than the given threshold, the pixel is considered as the foreground object.

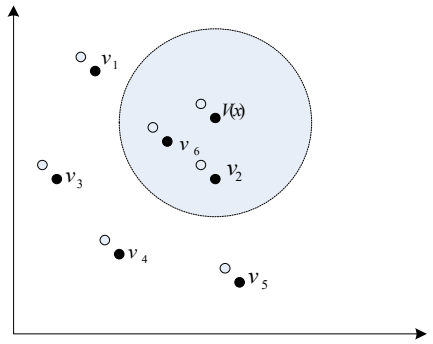

Fig. 1: Comparison of a pixel value with a set of samples

2) Sample set update

The sample set needs to be continuously updated to adapt to the changing of the video scene. In this paper, a more conservative updating strategy is adopted. For unmatched pixels, the pixel values will never be updated into the sample set. For the pixels that match successfully, the sample set is updated in three aspects [14].

- Memoryless update strategy

Each time it is determined that the background model of the pixel needs to be updated, and a sample value of the pixel sample set is randomly replaced with a new pixel value.

- Spatial neighborhood update strategy

For a pixel that needs to be updated, a background model of the neighborhood of the pixel is randomly selected, and the selected background model is updated with a new pixel.

- Time subsampling

Instead of processing a single frame of data, we need to update the process. But we need to update the background model at a certain update rate. When a pixel is judged as a background, it has a 1 / rate probability of updating the background model. The rate is the time sampling factor, and the general value is 16.

3) Based on the dynamic threshold of the background update

Background needs constantly to be updated to adapt to changes in the video scenes. In order to avoid a sample to stay in the sample set for a long time, a random update mechanism is introduced. When a certain pixel is judged as a background, there is an updated sample set of $\frac{1}{\delta(x,y)}$ probabilities. Considering that the video mostly is dynamic background, the model will inevitably appear a lot of false positives when the $\frac{1}{\delta(x,y)}$ is a fixed value. Therefore, this paper uses dynamic update and dynamic threshold to deal with the problem. The dynamic rules of update probability are as follows

$$\frac{1}{\delta(x,y)}=\begin{cases}\frac{1}{\delta(x,y)}+\sigma(x,y)\times\beta, & \delta(x,y)<\sigma(x,y)\times\alpha,\\ \frac{1}{\delta(x,y)}-\sigma(x,y)\times\beta, & \delta(x,y)>\sigma(x,y)\times\alpha,\end{cases} \tag{7}$$

where $\sigma(x,y)$ indicates acceptable threshold. α , β are the coefficient of variation. The update function for the threshold is similar to the above equation and will not be described.

4) Morphological de-noising of the target of the future

After the above algorithm is processed, the generated foreground objects contain a lot of noise, and there are some holes. Morphological methods are used to remove white spaces in binary images and fill holes in foreground targets. Based on the morphological theory, we perform morphological opening operations to eliminate the glitches on the isolated day shift and the edge of the foreground object. Then we expand the operation to fill the gap of the foreground object and make the edge of the object smoother.

3 Experimental Results

The test hardware platform for the AMD A6, 4GB RAM, software development environment for Windows 7, Matlab 2016a. The experiment uses 30 sample values as the sample set for each pixel, and the distance threshold is initially set to 30, and the initial value of the background model update rate is set as 1/20. The selected video sequence set is the D problem of China Post-Graduate Mathematical Contest in Modeling. All the tests of Gaussian mixture model and Vibe algorithm are completed by using the video sequence and compared with the mainstream algorithms.

3.1 Ghost Elimination Effect Analysis

In order to verify that the G-Vibe algorithm can solve the ghosting problem caused by the first frame containing the foreground object, the experimental results of G-Vibe is compared with Vibe, Siltp and frame difference method. The experimental results are shown in Figure 2.

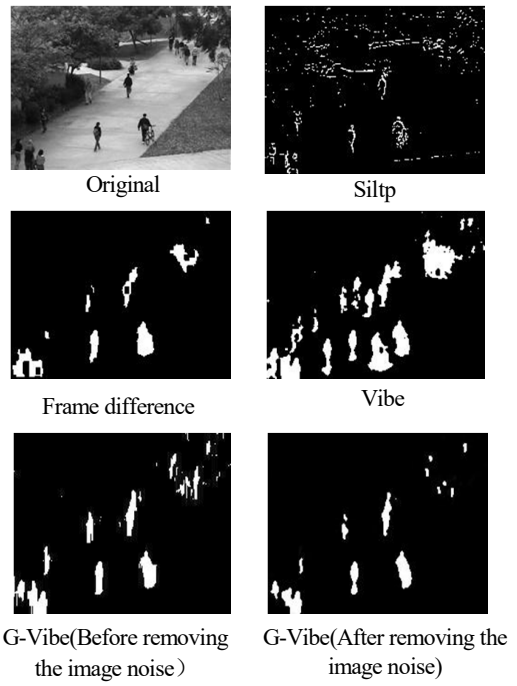

Fig. 2: Comparison of algorithms

Experimental results show that only the Vibe algorithm appears ghosting, while the rest of the algorithms are not

present. The result of frame difference method is well, but there are many voids in the foreground and the details are not in place. The results of Siltp are obviously affected by the dynamic background, and the foreground objects contain a lot of false positives. The details of G-Vibe algorithm results are not well-disposed and contain some misjudgment points before removing the noise. However, the foreground target is complete and has fewer misjudgment points after de-noising.

3.2 Dynamic Background Processing Effect Analysis

In this paper, we use water ripple as an example to verify the simulation. The experimental results are as follows.

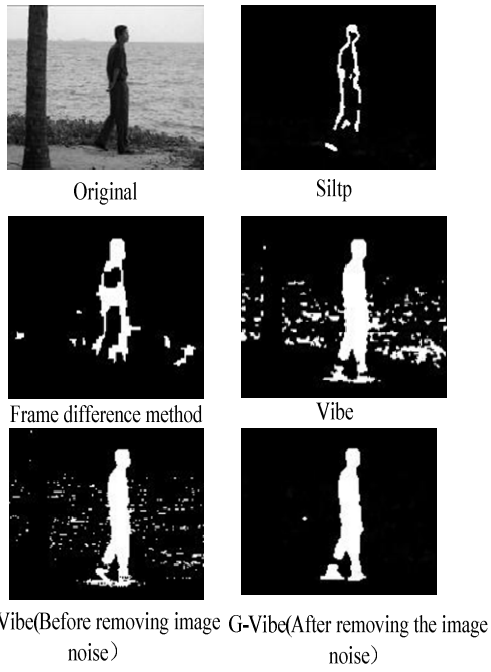

Fig. 3: Comparison of algorithms

Experimental results show that the Vibe algorithm contains a large number of false positives. The results of frame difference method to deal with the prospects contain a lot of holes and obvious white spots. Though water wave has a little effect on the foreground, the results of Siltp have obvious holes. The results of Vibe algorithm have a large number of white spots. The results of G-Vibe algorithm before de-noising are similar to that of Vibe. The results are satisfactory after de-noising. The human figure is complete while the white spots disappear, and the shadow is eliminated.

3.3 Algorithm Performance Analysis

Many metrics can be used to assess the output of a background subtraction algorithm. These metrics usually involve the following quantities [15]: precision(p), recall(R), specificity (SP), false positive rate (FPR), false negative rate (FNR) and percentage of wrong classifications (PWC). The paper compares and analyzes the performance of four algorithms: Siltp, frame difference method, Vibe and G-Vibe.

Table 1: Algorithm Performance Evaluation Index

Method	P	R	SP
Siltp	0.7156	0.6547	0.9358
Frame difference	0.7369	0.6644	0.9360
Vibe	0.8032	0.6931	0.9576
G-Vibe	0.8562	0.7213	0.9875
Method	FPR	FNR	PWC
Siltp	0.0168	0.0182	3.3318
Frame difference	0.0097	0.0163	2.9736
Vibe	0.0079	0.0105	2.1807
G-Vibe	0.0064	0.0095	2.0548

Table 1 shows that P, R of G-Vibe algorithm were significantly higher than the other three algorithms, FPR of G-Vibe was significantly lower than the other three algorithms. Thus G-Vibe algorithm has high reliability.

3.4 Algorithm Time Complexity Analysis

In order to verify the computing efficiency of G-Vibe algorithm, this paper compares the average time spent on processing each frame of the four algorithms. The specific results are shown in the following table.

Table 2: Algorithm Time Complexity

Method	Multi-view video target running time	Single foreground target video run time
Siltp	0.04430	0.04451
Frame difference	0.04964	0.05103
Vibe	0.01785	0.01854
G-Vibe	0.07856	0.07598

It can be seen from the Table 2 that the G-Vibe algorithm is slightly slower than other algorithms. But the G-Vibe algorithm is totally acceptable because it has a good effect of suppressing the ghosting.

4 Conclusions

In this paper, we present an effective learning algorithm named G-Vibe. The algorithm uses a continuous multi-frame image to train a Gaussian mixture model. The corresponding background of the image is obtained. Finally, the foreground object is extracted according to the pixel classification algorithm. On the basis of the above results, we use the morphological theory to deal with the prospect and makes it more perfect. Experimental results show that the performance of G-Vibe algorithm is obviously superior to other algorithms, and the effect of ghost suppression is obvious. The computational efficiency is equal to that of the other three algorithms. However, it is obvious that the image recognition time of the first frame is longer, mainly because of the high complexity of the training process.

References

[1] C. Chiu, M. Ku, and W. Liang, Robust object segmentation system using a probability based background extraction algorithm, *IEEE Trans. on Circuits and Systems for Video Technology*, 20(1): 5185-5192, 2010.

[2] N. Goyette, P. Jodoin, and F. Porikli, et al., Change detection. net: A new change detection benchmark dataset, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 1-8, 2012.
[3] L. Zhang, L. M. Cheng, and W. He, Application of improved frame difference method based on video in traffic flow measurement, *Journal of Chongqing University*, 27(5): 32-33, 2004.
[4] O. Barnich, and D. Van, Vibe: A universal background subtraction algorithm for video sequences, *IEEE Transaction on Image Processing*, 20(6): 1709-1724, 2011.
[5] L. Cheng, X. Z. Cheng, and Z. T. Fan, Ghost suppression algorithm based on Vibe, *Journal of China Jiliang University*, 24(4): 425-429, 2013.
[6] Z. H. Hu, W. X. Zhang, and Y. Wang, Moving target detection algorithm based on improved Vibe, *Electronic Technology Applications*, 43(4): 129-132, 2017.
[7] X. J. Li, J. T. Li, and Y. F. He, Ghost suppression of target similarity measure, *Electronic Technology Applications*, 31(3): 926-928, 2014.
[8] K. Zhang, L. Zhang, and M. H. Yang, Real-time compressive tracking, *Proceedings of the 12th European Conference on Computer Vision*, Berlin, 864-877, 2012.
[9] C. Stauffer, and W. Grimson, Adaptive background mixture models for realtime tracking, In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1999: 246-252.
[10] W. Zhou, Y. Liu, and W. Zhang, Dynamic background subtraction using spatial-color binary patterns, *International Conference on Image & Graphics--IEEE Computer Society*, 314-319, 2011.
[11] B. Shoushtarian, and H. E. Bez, A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking, *Pattern Recognition Letters*, 26(1): 5-26, 2005.
[12] Y. W. Li, K. Cao, and J. Wang, An improved Vibe ghost suppression algorithm, *Journal of Guangxi University (Natural Science Edition)*, 42(2): 712-719, 2017.
[13] Z. Zivkovic, Improved adaptive Gaussian mixture model for background subtraction, Pattern Recognition, *International Conference on IEEE Computer Society*, 28-31, 2004.
[14] P. Kaewtrakulpong, and R. Bowden, An improved adaptive background mixture model for real-time tracking with shadow detection, In *Proc. 2nd European Workshop on Advanced Video Based Surveillance Systems,*2002, 135-144.
[15] X. Liu, J. Yao, and X. Hong, et al., Background subtraction using Spatio-Temporal group sparsity recovery, *IEEE Transactions on Circuits & Systems for Video Technology*, 1-5, 2017.