Improved Moving Target Detection Based on Multi-Model Mean Model

IOP Conf. Series: Earth and Environmental Science **252** (2019) 052134 doi:10.1088/1755-1315/252/5/052134

Improved Moving Target Detection Based on Multi-Model Mean Model

Weiwei Wang[1, *], Deyong Gao[1, a], Yangping Wang[2, 3, b] and Decheng Gao[4, c]

[1]School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou, China
[2]Gansu Provincial Engineering Research Center for Artificial Intelligence and Graphics & Image Processing, Lanzhou, China
[3]Gansu Provincial Key Lab of System Dynamics and Reliability of Rail Transport Equipment, Lanzhou, China
[4]Gansu Institute of Metrology, Lanzhou, China

*Corresponding author e-mail: 1319144981@qq.com, [a]258680916@qq.com, [b]13519311970@qq.com, [c]1451703192@qq.coom

Abstract. Aiming at the problem of low detection accuracy of multi-mode mean model in complex scenarios, an improved detection method of moving target based on multi-mode mean model is proposed.Firstly, the background model is constructed using the multi-mode mean value model.According to different scene information, different thresholds are set and adjusted adaptively.The foreground image obtained by background difference method is detected by frame difference method, and the experiment is compared and analyzed.The detection rate and the error rate are reduced, and the detection accuracy is improved. Finally, the simulation results of three-segment video verify the effectiveness of the proposed method.

1. Introduction

In computer vision, detection of moving objects is the premise to obtain image sequence information and motion recognition, and also an important part of intelligent monitoring system [1].

The detection methods of moving targets can be roughly divided into the following three types: optical flow method [2-4], inter-frame difference method [5] and background difference method [6-7].Optical flow method is computationally intensive and highly demanding hardware equipment, which is not suitable for real-time detection.Inter-frame difference method has omission detection and error detection, while light and shadow have a great impact on the detection effect [8].Background difference method is the most widely studied and applied method at present. It is to identify moving objects according to the difference image [9].

To solve the above problems, this paper proposes an improved moving target detection algorithm based on multi-mode mean model, which uses multi-mode mean model to improve the problem of large computation and slow speed, and sets adaptive threshold to reduce the impact of noise on detection results. The interframe difference method verifies the results of the background difference method, which provides an accuracy guarantee for the detection of foreground targets and effectively reduces the influence of illumination changes.This algorithm combines the advantages of frame difference method and background difference method, which can make the detection of moving target more accurate.Experimental results show that compared with hybrid gaussian model [10] and VIBE algorithm [11], the improved algorithm has better performance in real-time performance and detection accuracy.

2. Multi-mode mean model

The multi-mode mean model is constructed by calculating the average of the multi-channel color time series of the video frame image. Its basic idea is to capture successive N-frame video frame images and calculate the multi-mode distribution mean of the pixels in each color channel. The K mean model describes the background model (the value of K is usually taken as 3-7), and the background model is described by equation (1) [12]:

$$B_{i,t} = \{S_{i,R}; S_{i,G}; S_{i,B}; C_{i,t}; r_{i,t}; p_{i,t}\} \quad i = 1,2,.....K \tag{1}$$

$B_{i,t}$ is the background pixel value of the i-th model at time t, $S_{i,R}, S_{i,G}, S_{i,B}$ represent the sum of the values of the three RGB color channels in the i-th model at time t, $C_{i,t}$ represents the sum of the number of matches between the current pixel value of time t and the i-th background model. $r_{i,t}, p_{i,t}$ represent the correspondence between the history frame and the pixels in the i-th background model, and is used to represent the level of the background model.

An image is represented by a multi-channel image, assuming that the number of channels is x, equation (2) can be used to describe the average value of each color component of each pixel point when the time is t:

$$u_{i,t,x} = S_{i,t,x} / C_{i,t} \tag{2}$$

$S_{i,t,x}$ represents the sum of the values of each color component of model k at time t, $C_{i,t}$ represents the accumulated number of pixel values matching the mode k in the t-frame corresponding to the mode. This method constructs a background model using RGB three-channel images, so x=3.

3. Improved moving target detection algorithm

3.1. Threshold selection and adjustment

In the process of background initialization, a time series of all pixels is obtained. For one pixel, median filtering is used to calculate the average value to obtain the expected value of this pixel. Then, all values are compared with the expected value to obtain the maximum difference, and the resulting maximum difference is regarded as the basic value of the threshold.The final threshold of the image is set to 2-3 times the base value, and its mathematical description is as in equation (3) and $f(x,y)$ is not equal to $Max1, Max2, Min1, Min2$.

$$TH_1(x,y) = 2Max(|f(x,y) - E_i(x,y)|) \quad \mathrm{i} = 1,2...10 \tag{3}$$

In formula (3): $Max1, Max2, Min1, Min2$, indicate the maximum value, the second largest value, the minimum value, and the second minimum value in the pixel set, TH_1 represents the pixel value when the difference value of the corresponding image position graph is binarized, which can also be called threshold image.The advantage of this method is that the appropriate threshold is selected in different areas of the scene according to the different monitoring scene information [14].

Figure 1. Background image

Figure 2. Threshold image

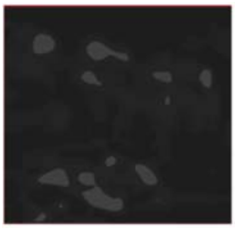

Figure 3. Left threshold image

Figure 4. Intermediate threshold image

As shown in Figure 1, there is a tree growth on the left side of the image. This area usually does not have a moving target appearing or moving. A relatively large threshold is set, corresponding to the relatively bright area on the left side of the image 2, and the bright area marked red on the left side of the image 2 is enlarged, as shown in figure 3.In the portion of the middle road of the image 1, there is a movement of the target, and a relatively small threshold value must be set, corresponding to the darker region in the image 2, and the dark region of the middle red is enlarged in figure 2, corresponding to figure4.

3.2. *Extraction of foreground target*

The main idea of foreground moving target extraction [15] is to compare the current video frame image pixel by pixel with the already constructed background model to determine whether it is a foreground pixel or a background pixel. When the formula (4) is met, the pixel is seen as the foreground pixel, set to 255, otherwise the pixel is the background pixel and the pixel size is set to 0.

$$(c_{i,t-1} > TH) \wedge (\bigcap_x | I_{t,x} - u_{i,t-1,x} | \le E_x) \tag{4}$$

3.3. *Moving target detection process*

The basic idea is to use the adaptive threshold image as the threshold of the background difference method in the detection stage, and use the background difference method to detect the moving target. Once an abnormality is detected, the differential operation of three consecutive frames is obtained to obtain more accurate information of the moving target. The detection flow chart in this article can be described in figure 5.

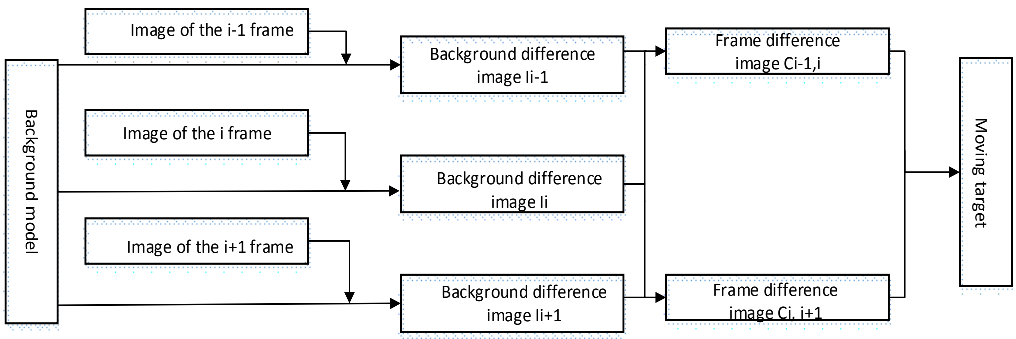

Figure 5. Flow chart of moving target detection

In the improved algorithm, the following two conditions are to be taken:

Judgment condition 1 (background difference method):

$$\begin{aligned} & b_k = \left| f(x, y, t_k) - B(x, y, t_k) \right| \\ & X = \left\{ Y \mid b_k \geq TH_1, Y \in A \right\} \\ & N_X > N_1 \end{aligned} \tag{5}$$

In formula (5): $f(x, y, t_k)$ represents the gray value of the image (x, y) at the time t_k; $B(x, y, t_k)$ represents the gray value of the background pixel (x, y) at the time t_k; TH_1 is the gray threshold of the image, and the detection area of the image is used to indicate A.In X, the number of the minimum elements that meet the early-warning condition is represented by N_1; in X, the number of elements satisfying the internal judgment condition is represented by N_X [15].(Under initialization conditions, k= i-1)

$$result1 = \begin{cases} 1 & if\ N_x \geq N_1 \\ 0 & else \end{cases} \tag{6}$$

If $result1 = 0$, it represents a target or object that has not been moved in the current monitoring scene, it is not necessary to perform deeper processing on the currently captured image, and the judgment is completed. If $result1 = 1$, represents that compared with the previous frame, the image of the first frame has changed, which may be caused by a change in illumination or a movement of the object into the monitoring area. Then, the image of the next frame is captured as the second frame image, let $k = i$ and the judgment of the first condition is continued. If $result1 = 0$, it means that the change of the first frame image is not caused by the former , but by the intrusion of moving objects into the monitoring area,an alert notification is issued; if $result1 = 1$, it indicates that the image of the second frame has also changed, it is not certain what the cause is. Just enter the judgment of the second condition.

Judgment condition 2(inter-frame difference method):

$$\begin{aligned} & C_{i-1,1} = \left| b_{i-1} - b_i \right| \\ & X = \left\{ Y \mid C_{i-1,1} \geq TH_2, Y \in A \right\} \\ & N_Y > N_2 \end{aligned} \tag{7}$$

In formula (7): b_i represents the gray value of the image difference image at the time t_i of the background; TH_2 is the gray threshold of the image sample; the number of the lowest elements in X that meet the pre-warning condition is denoted by N_2.The number of elements satisfying the internal judgment condition in X is represented by N_Y [16].

$$result2 = \begin{cases} 1 & if\ N_Y \geq N_2 \\ 0 & else \end{cases} \tag{8}$$

On the basis of the first judgment condition, if $result2 = 1$, it represents a moving object is entered into the monitoring area, an alarm notification is initiated. If $result2 = 0$, the change in the first frame image and the second frame image is likely to be caused by a change in illumination, the next frame image is captured as the image of the third frame, let $k = i + 1$, and continue to judge the first condition. If $result1 = 0$, on behalf of the first frame image and the second frame image changes are moving objects into the monitoring scene, an alarm notification is issued; if $result1 = 1$, so, the second condition

IOP Conf. Series: Earth and Environmental Science **252** (2019) 052134 doi:10.1088/1755-1315/252/5/052134

is determined. If $result2=1$, it means that the moving object has entered the monitoring area, an alarm notification is issued. If $result2=0$, take the image of the third frame obtained as the background and update the background.

4. Experimental results and analysis

Simulation experiments were carried out in opencv environment, and two video sequences were simulated.Their parameters are as follows: the frame rate of video sequence 1 and 2 is 25f/s, and their total frame number is 440 and 1200 respectively.

In video sequence 1, the moving target (vehicle) keeps moving rapidly, and the images at frames 52, 104 and 208 are respectively captured for simulation experiment.Figure 6 is the result of simulation experiment.

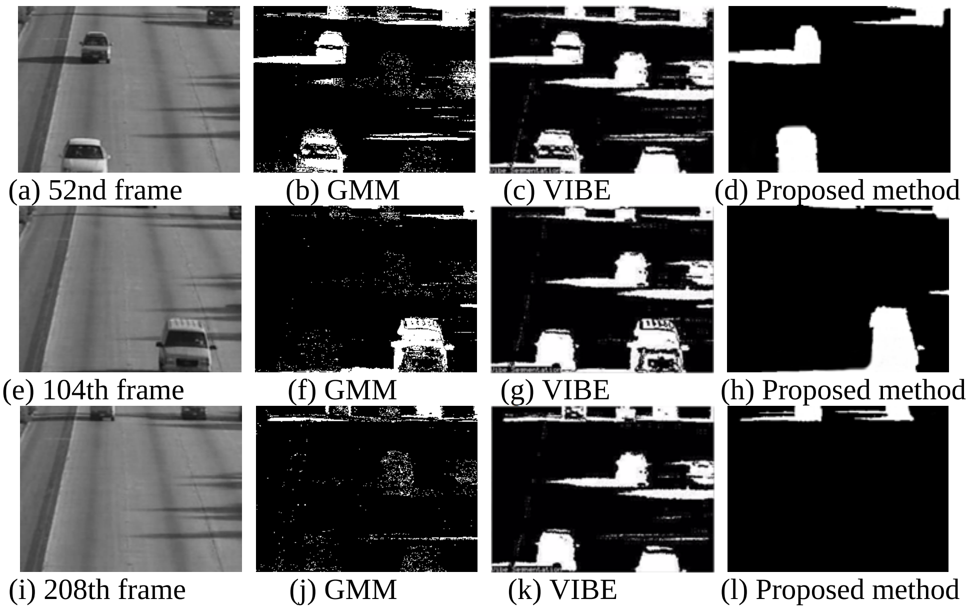

Figure 6. Comparison of Video Sequence 1

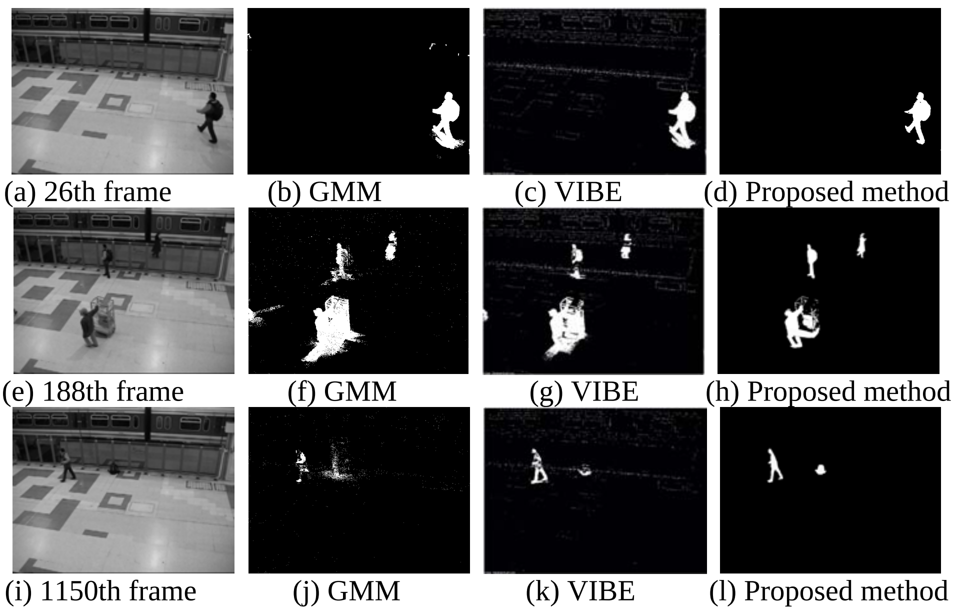

Figure 7. Comparison of Video Sequence 2

As can be seen from figure 6, the foreground image obtained by VIBE algorithm has relatively little noise but a lot of shadow, while the hybrid gaussian algorithm has less shadow but a lot of noise.The moving target extracted by this algorithm is relatively complete and the noise is reduced.

In video sequence 2, pedestrian video monitored indoors is a complex scene with multiple moving objects.The images at frames 26, 188 and 1150 were captured and simulated respectively.Figure 7 shows the results of the experiment.

As can be seen from figure 7, the moving target detected by the hybrid gaussian algorithm has a void phenomenon and a lot of noise.The operational target detected by VIBE is relatively complete, but part of the background railing is also detected as the foreground.The foreground image obtained by the algorithm in this paper is relatively complete and can suppress noise well.

The following performance evaluation indexes are used to analyze the performance of several algorithms [17].Pr stands for Precision, and FPS stands for Frames per Second. FPR (False Positive Rate) represents False alarm Rate, FNR (False Negative Rate) represents missed detection Rate, PWC represents wrong score Rate and is a comprehensive evaluation index.The experimental results of video sequences 1 and 2 are shown in table 1.

Table 1. Algorithmic Performance Evaluation Indicators

Performance	Method	FPR	FNR	Pr	PWC (%)	FPS (fps)
	GMM	0.0123	0.304	0.7577	2.9343	19
video 1	VIBE	0.0418	0.1521	0.5306	3.8105	51
	Proposed method	0.0113	0.1027	0.8904	1.3968	41
	GMM	0.0146	0.3160	0.7491	2.8151	20
video 2	VIBE	0.0237	0.1540	0.6739	2.8324	49
	Proposed method	0.0095	0.0872	0.9647	1.3780	40

According to the data in table 1, the frame rate of the algorithm in this paper remains above 30fps with a certain real-time performance.As can be seen from the accuracy, VIBE algorithm has a large amount of drag and shadow in the dynamic background due to fixed threshold and update rate, while hybrid gaussian algorithm has a large amount of noise and cavity phenomenon in the detected target, with the highest leakage rate, and the accuracy of this algorithm is relatively optimal.

5. Conclusion

In this paper, an improved moving target detection algorithm based on the multi-mode mean value model is proposed. Taking the multi-mode mean value model as the background model, the computational complexity is reduced and the detection rate of moving targets in video is improved.At the same time, an adaptive threshold is adopted for noise, and the inter-frame difference method is applied. Based on the background difference method, the detection accuracy is improved.In terms of detection accuracy and real-time performance, this algorithm has strong interference ability and is easy to implement.

Acknowledgments

This work was financially supported by National Natural Science Foundation of China (61162016 and 61562057), by the Science and Technology Project of Gansu Colleges and Universities (2017D-08), by Science and Technology Plan Project of Gansu (18JR3RA104 and 1504FKCA038).

References

[1] SU Z, WANG W,XU C.Optical correlation detection technology of moving target under low contrast environment [J]. Chinese Journal of Scientific Instrument, 2013, 34 (2): 319-325.
[2] Sun D, Roth S, Black M J. Secrets of optical flow estimation and their principles [C] IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2010: 2432-2439.
[3] Zhang Z P, Jiang W S, Zhang J. Vehicle borne panoramic image matching method based on

optical flow feature clustering [J]. Acta Geodaetica et Cartographica Sinica,2014, 43 (12): 1266-1273.
[4] Yang Y D. Theory and application of optical flow method in moving target recognition field [J]. Electronicn DesignEngineering, Electronic Design Engineering, 2013, 21 (5): 24-26.
[5] Yin Y, Wang Y F. Moving Object Detection Based on Spatial LocalCorrelation [J]. Opto-Electronic Engineering, 2009, 36 (2): 1-5.
[6] Yang H, Wang C, Jiang W T, et al. Target detection algorithm based on random background modeling [J]. Journal of Applied Optics, 2015, 36 (6): 880-887.
[7] SU S Z, Li S Z, Chen S Y, et al. Survey of pedestrian detection technology [J]. Electronic Journal, 2012, 40 (4): 814-820.
[8] Yang Y, Tang H M. Video-based pedestrian vehicle detection and classification [J]. computer engineering, 2014, 40 (11): 135-138.
[9] Huang X S, Huang P, Cao Y Q. An Improved Moving Target Detection Algorithm Based on K-SVD Dictionary Learning [J]. Microelectronics and Computer, 2014, 31 (3): 5-8.
[10] STAUFFER C, GRIMSON W E L. Adaptive Background Mixture Models for Real-Time Tracking [C] // IEEE Computer Society. 1999; 22-46.
[11] Barnich O, Van Droogenbroeck M. ViBe: a universal background subtraction algorithm for video sequences [J]. IEEE Transactions on Image Processing, 2011, 20 (6): 1709-1724.
[12] Zhao G P, Bo Y M.Target fusion detection method based on multimodal mean time space model [J]. Journal of Image and Graphics, 2010, 15 (8): 1254-1259.
[13] Wang H, Li A H, Cui Z G, et al. A foreground detection algorithm for improving visual background extraction in complex background [J]. Application Research of Computers, 2017, 34 (4): 1261-1264.
[14] Deng J G. Research on real-time video front and back separation and synthesis technology [D]. Xidian University of Electronic Technology, 2012.
[15] Nan Y X.Research on Key Technologies of Residual Detection in Video Surveillance [D]. Wuhan University of Technology, 2014.
[16] Li Z F, Zhu L Y. Moving target detection based on improved background difference method [J]. Instrument technology, 2012, 01 (1): 34-36.
[17] Goyette N, Jodoin P M, Porikli F, et al. Changedetection.net: A new change detection benchmark dataset [C] // Computer Vision and Pattern Recognition Workshops. IEEE, 2012: 1-8.