Background Modeling Algorithm for Multi-feature Fusion

Zhicheng Guo*, Jianwu Dang, Yangping Wang, Jing Jin
School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou, China
E-mail: lzjdgr@163.com Tel:+86-13919300072
Gansu Provincial Engineering Research Center for Artificial Intelligence and Graphics & Image Processing, Lanzhou, China
Gansu Provincial Key Lab of System Dynamics and Reliability of Rail Transport Equipment, Lanzhou, China
Lanzhou Bocai Technology Co., Ltd. Lanzhou, China

***Abstract*— In order to improve the accuracy of foreground target detection and establish a stable background model, this paper proposes a multi-feature fusion background modeling algorithm, which initializes the background model with the spatial correlation between the first frame pixel and the domain pixel, and quickly establishes the background. model. A multi-feature sample set consisting of pixel values, update frequency, update time, and adaptive dynamic coefficients is updated with temporal correlation of subsequent intra-pixels. According to the multi-feature sample set, the background complexity is adjusted to adjust the update speed of the model in different regions, which effectively improves the ghost phenomenon of the foreground target and reduces the false holes in the target and the false foreground in the background. The test results of multiple sets of data sets show that the proposed algorithm improves the adaptability and robustness of foreground target detection in scenarios with high dynamic changes.**

***Keywords*—machinevision; Multi-feature fusion; Target Detection; Background model**

I. INTRODUCTION

The background modeling of moving targets is one of the research hotspots and difficulties of machine vision and intelligent video processing. Its goal is to extract the changed regions from the video sequence, and effectively detect the moving targets for subsequent research work such as object tracking [1,2] Application understanding of target classification, behavior analysis, behavioral understanding, etc. plays an important role. At present, the commonly used detection methods include frame difference method [3], optical flow method [4], and background difference method [5]. The frame difference method uses two frames or several frames to perform differential operations to obtain moving targets. The extraction effect is related to the speed and frame rate of the moving target, and the calculation speed is fast, but it is not suitable for target detection with static or slow speed. The optical flow method is derived from the optical flow field. The moving target is obtained according to the continuous multi-frame video sequence and the difference of the motion of each pixel. The detection effect is good, but the algorithm complexity is high and the calculation amount is large, which is difficult to meet the real-time requirement. Compared with the former two background difference methods, the input image and the background model are used to obtain the moving target. The algorithm has the advantages of low overhead, high speed, high precision, accurate extraction target, etc. It has become the most commonly used method for moving target detection.

The detection performance of the background difference method depends mainly on the robust background model. The background model establishment and update algorithm directly affects the detection effect of the final target. Many scholars at home and abroad have carried out in-depth research on background modeling technology [6-10], which is mainly divided into two types: parameterization method and non-parameterization method. Parametric methods such as single Gaussian background model (SGM), mixed Gaussian background model (MOG) [11, 12], etc. Among them, the single Gaussian background model establishes a model represented by a single Gaussian distribution for each pixel value distribution, which is suitable for single-mode background. When the background is more complicated or the background pixels are multi-peak distribution, the model adaptive difference and detection rate incomplete. The mixed Gaussian background model has multiple Gaussian distributions, and the effect is better when the background pixels are multi-peak distribution, but the parameters are difficult to adjust, the computational complexity is high, and the real-time performance is not strong. Nonparametric methods such as CodeBook [13], ViBe [14], PBAS [15], etc. CodeBook establishes a time series model for each pixel in the video sequence, which can handle time fluctuations well, and the calculation amount is small, but the memory consumption is large, which is easy to cause codeword update error and poor dynamic adaptability. ViBe initializes the background model with a single frame video according to the spatial distribution characteristics of adjacent pixels, and updates the background model with a random replacement strategy. The modeling speed is fast and robust to noise and brightness changes. PBAS combines the advantages of both SACON and VIBE algorithms, but the judgment threshold and update rate are computationally complex and have poor real-time performance.

In this paper, a background modeling method based on pixel feature pixel value, frequency, update time and adaptive dynamic coefficient is proposed. The initial background

model is established by using pixel neighborhood correlation in single frame image, and the background model is updated. In the process, according to the multiple features of the video sequence, the ghost region that may be generated in the initial background model is quickly eliminated, and the threshold information is adaptively adjusted by using the feedback information generated by the actual background in the dynamic change, and the moving target is extracted in the complex scene.

II. BACKGROUND MODELING ALGORITHM FOR MULTI-FEATURE FUSION

The algorithm initializes the background model with neighborhood pixels, updates the background model by newly acquiring pixels, compares the background model and the current input pixel values to detect the foreground target, and is divided into three aspects: the working principle, initialization and update strategy of the model.[1]

A. Model working principle

In background modeling, commonly used feature information includes colors, shapes, textures, etc. For example, Nummiafo [7] et al. use the color information of the target to track the target, and Vignon [6] et al. use the shape information of the target to track the target. Although these algorithms consider the adaptability of the background model to the scene change, the adaptability of the model update strategy when the feature is degraded is neglected, and the features sensitive to the scene change cannot be updated in time, and the resolvability of the model is continuously reduced, which restricts the performance of the algorithm. This paper proposes a multi-feature background modeling method that combines sample pixel values, frequency, update time and adaptive dynamic coefficients.

The observations of the pixels at the same position in the video sequence image on the time axis may constitute a sequence $V(x_i)=\{x_1,x_2\ldots x_t\}$ where x_i represents a certain pixel of a frame in the image, t and i represent time, and $V(x_i)$ represents a pixel value at point x_i. A quaternary multi-feature sample set $M(x_i)=\{P(x_i),F(x_i),T(x_i),S(x_i),\}$ is created for each pixel sequence in a frame.

$P(x_i)=\{p_1,p_2\ldots p_N\}$ Is the background pixel value sample set at x_i, where N is the sample set size;

$F(x_i)=\{f_1,f_2\ldots f_N\}$ The frequency of each sample in the background sample set at x_i;

$T(x_i)=\{t_1,t_2\ldots t_N\}$ The latest update time for the background sample set at x_i;

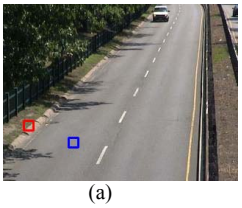
$S(x_i)=\{s\}$ Adaptive dynamic coefficient for x_i pixels;

The working principle of the model is to assume that all background pixel values are distributed inside the multi-feature sample set $M(x_i)$, while the foreground pixel values are distributed outside, $M(x_i)$ constitutes the boundary of the foreground background pixel, and the input $V(x_t)$ at time t is judged by the formula (1). Whether the pixel value is a background pixel, where G_F is the foreground pixel and G_B is the background pixel.

$$V(x_i)=\begin{cases}G_F,\{V(x_t)\cap M(x_i)\}<S(x_i)\\ G_B,\{V(x_t)\cap M(x_i)\}\geq S(x_i)\end{cases} \quad (1)$$

B. Model initialization

The traditional background modeling method to initialize the background model usually needs to initialize the sample model for image sampling of more than 10 frames. Not only the large amount of calculation will result in the inability to find moving targets in the first few frames. The algorithm takes the first frame of the video sequence to complete the initialization of the model and reduces the complexity of sampling. One frame of video, the pixels in different areas have strong consistency and correlation in position and gray level. As shown in Figure 1(a), select the area A in the first frame of the video (blue box) The gray value of the 5*5 neighborhood pixel of the mark, center coordinates 188, 96) and area B (red square mark, center coordinates 165, 35) is shown in Fig. 1(b)(c), and area A is The smooth center pixel has strong correlation with the neighborhood pixel, and the edge appears in the area B. The surrounding pixel and the center pixel have a large difference, but some pixels have strong correlation with the center pixel, and the algorithm utilizes the pixel. Neighborhood correlation initializes the background model sample set.



143	144	143	143	144
142	144	142	143	144
142	143	142	143	144
140	141	142	142	143
139	139	141	142	143

109	106	94	85	98
113	104	100	107	130
102	110	133	162	179
109	137	167	187	191
138	172	183	174	164

(a) (b) (c)

Fig.1 Pixel spatial correlation. (a) First frame image. (b) A area 5*5 pixel value. (c) B area 5*5 pixel value.

The background model is initialized with the first frame image, and the pixel value satisfying the formula (2) in the 5×5 window centered on the pixel x_i is initialized $P(x_i)$, $T(x_i)=\{t_i=1\,|\,i=1\ldots N\}$, and the number of x_i elements satisfying the formula (3) is the sample set size, and the dynamic index satisfies (4) Formula.

$$P(x_i)=\{V(x_i),V_j'(x_i)\,|\,j=1\ldots N\} \quad (2)$$

$$V_j'(x_i)\in\{|V_k'(x_i)-V(x_i)|<\varepsilon,k=1\ldots24\} \quad (3)$$

1 Project supported by: Project supported by the Natural Science Foundation of China(Grant Nos. 61661026; 61841303);Gansu Project supported by the Provincial Education Department of Gansu Province,China(Grant No.2017D-08);Project supported by the Natural Science Foundation of Gansu Province,China(Grant No.18JR3RA104);Project supported by the Young Scholars Science Foundation of Lanzhou Jiaotong University(Grant No.2016005); Lanzhou Talent Innovation and Entrepreneurship Project (2015-RC-86);

2 About the author: Guo Zhicheng (1981-), male, doctoral student, lecturer, machine vision, artificial intelligence, image processing. E-mail: lzjdgr@163.com.

$$S(s_i)=\frac{\sum\left|V(x_i)-V_j'(x_i)\right|}{N}, j=1\ldots N \quad (4)$$

Where $V_k'(x_i)$ is the 5×5 neighborhood pixel of pixel x_i, $S(s_i)$ is the dynamic index of x_i pixel, j is the j-th neighborhood pixel satisfying equation (3), k is the k-th neighborhood pixel, and $V(x_i)$ is the pixel of the window center pixel Value, $V_j'(x_i)$ is a neighborhood pixel satisfying equation (3), ε is a constant.

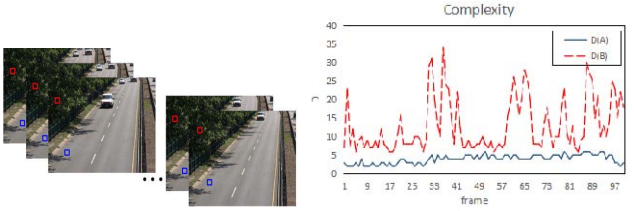
C. *Model update strategy*

Factors such as sensor error, external illumination changes and twig shaking in the video acquisition device cause the background pixel values to dynamically change and produce a large number of erroneous foreground targets. It is necessary to establish a standard to measure the degree of background dynamic change, and to compare the sample values of the sample and the background model. And, the obtained Euclidean distance and the mean value as the background complexity $C(x_i)$ satisfies the formula (5)

$$C(x_t)=\frac{V(x_t)-P(x_{t-1})}{N} \quad (5)$$

Where $V(x_t)$ is the input image pixel value at time t.

As shown in Figure 2(a), the low complexity area A (blue box mark, center coordinates are 191, 58) and the high complexity area B (red box mark, center coordinates are 59, 30), select video consecutive 100 The changes in the pixel values D(A) and D(B) of the center point of the frame image area A and the area B are as shown in Fig. 2(b).



(a) (b)

Fig.2 The pixel value changes dynamically. (a) 100-frame image sequence. (b) Region A and Region B adaptive dynamic coefficients.

Using the feedback information of the pixel-level background, the dynamic coefficients $S(x_i)$ are adaptively acquired for regions of different complexity to adapt to the background of different complexity. For high complexity background areas, the dynamic coefficient $S(x_i)$ should be large to avoid false front spots, low complexity background areas, and the dynamic coefficient $S(x_i)$ should be small to reduce misidentification of background points. The dynamic coefficient $S(x_i)$ satisfies the equation (6), where δ is a fixed coefficient.

$$S(x_i)=\begin{cases} S(x_i)\text{-}C(x_i) & S(x_i)\le\delta\cdot D(x_i) \\ S(x_i)+C(x_i) & S(x_i)>\delta\cdot D(x_i) \end{cases} \quad (6)$$

The update strategy of the background model directly affects the quality of the foreground moving target extraction. The slow update will generate the ghost phenomenon and the legacy objects cannot be correctly identified. The update will produce the moving object holes and a large number of false foregrounds. The algorithm uses the $F(x_i)$ sample frequency and $T(x_i)$ The last update time effectively improved the ghost phenomenon and reduced the hole in the foreground moving target.

The model update strategy is as follows:

Step 1: The x_i pixel points are not classified.

Step 2: x_i satisfies $abs(p_i-V(x_i))<S(x_i)$, f_i is incremented, and ti is cleared. The f_i of the other samples in $M(x_i)$ is decremented by 1, and t_i is incremented by 1.

Step 3: If $abs(p_i-V(x_i))<S(x_i)$ is not found in $M(x_i)$, insert xi into the sample set.

Step 4: When a xi pixel is classified as a front spot, directly insert xi into the sample set, replacing the element in the sample set that satisfies equation (7).

$$p_i=\{\min(F(x_j))\bigcap\max(T(t_j)\mid j=1\ldots N\} \quad (7)$$

Step 5: Update $M(x_i)$.

III. EXPERIMENT AND RESULT ANALYSIS

A. *Comparative test and analysis*

In order to verify the validity and practicability of the proposed algorithm, the algorithm is implemented under the C++ programming tool. The experimental hardware environment is Intel Core i3 3.4GHz processor with 4.0GB memory. In order to reflect the advantages of the proposed algorithm under complex texture background conditions, four background modeling algorithms, CodeBook[13], MOG[12], PBAS[15] and ViBe[14], were selected for comparison experiments.

In the experiment, the ForegroundAperture and LightSwitch in the Microsoft Wallflower paper dataset were selected; the canoe, fountain02, pedestrians and PETS2006 in the CDNet2014 dataset were divided into three types of scene testing algorithms: indoor, outdoor and complex background. The ForegroundAperture and LightSwitch resolutions is 160*120.

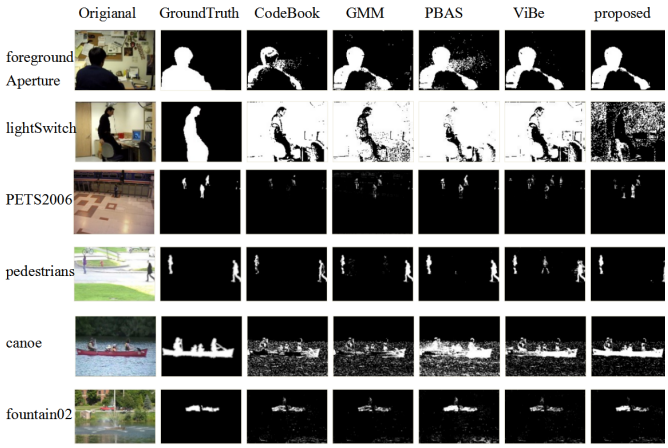


Fig.3 Comparative Experiment

The resolution of canoe is 320*240; the resolution of pedestrians is 360*240; the resolution of fountain02 is 432*288; the resolution of PETS2006 is 720*576. The

CodeBook and GMM modeling frames are 50 frames, and the result is shown in Figure 3.

B. Quantitative analysis

In order to evaluate the algorithm more accurately, the algorithm based on the real-time processing speed and accuracy of the algorithm and the classical parameter-based MOG, CodeBook, ViBe and PBAS algorithms are compared with the proposed algorithm.

The real-time processing speed is compared from M_{MT} (modeling time) to the background model initialization time and H_{HT} (handling time). The final data is shown in Table 1. It can be seen that the algorithm has the shortest modeling time and the single frame processing time is faster.

Tab.1 Processing speed comparison

Method	lightSwitch		canoe		pedestrians		PETS2006	
	M_{MT}	H_{HT}	M_{MT}	H_{HT}	M_{MT}	H_{HT}	M_{MT}	H_{HT}
Codebook	0.68	0.13	1.38	0.33	1.27	0.16	3.93	0.6
GMM	6.55	0.22	33.99	0.85	25.73	0.95	124.28	4.6
PBAS	0.08	0.68	0.23	1.39	0.42	0.63	1.181	2.35
ViBe	0.11	0.06	0.35	0.10	0.04	0.07	1.86	0.31
propose	0.04	0.13	0.19	0.70	0.21	0.56	1.02	2.25

Quantitative analysis uses five information retrieval criteria to reflect the search results in the neighborhood to evaluate five background models, including: P (Precision) indicates the correct rate of detection of the foreground target, R (Recall) indicates that all prospects have How many are detected, P_{PWC} (Precentage of Wrong Classifications) indicates the error rate of foreground and background points.

$$P = \frac{t_{tp}}{t_{tp} + f_{fp}} \tag{8}$$

$$R = \frac{t_{tp}}{t_{tp} + f_{fn}} \tag{9}$$

$$P_{pwc} = 100 \times \frac{f_{fp} + f_{fn}}{t_{tp} + t_{tn} + f_{fp} + f_{fn}} \tag{10}$$

Among them, t_{tp} (true positives) detects the correct number of pre-attractions, t_{tn} (true negatives) detects the number of spots before the error, f_{fp} (false positives) detects the correct number of background points, and f_{fn} (false negatives) detects the number of incorrect background points.

In the general scene experiment, the results of four sets of image sequences of ForegroundAperture, LightSwitch, pedestrians, and PETS2006 are shown in Table 2; the results of two sets of image sequences of complex scenes canoe and fountain02 are shown in Table 3. It can be seen that each performance index of the algorithm is consistent with Figure 3. The P_{PR}, R_{RE}, and P_{PWC} algorithms in this paper are significantly better than other algorithms, and have strong robustness in complex scenarios.

Tab.2 Ordinary background

Method	ForegroundAperture			LightSwitch			pedestrians			PETS2006		
	R	P	P_{pwc}	R	P	P_{pwc}	R	P	P_{pwc}	R	P	P_{pwc}
Codebook	0.30	0.69	21.4	0.54	0.099	84.2	0.66	0.98	0.68	0.15	0.92	0.98
GMM	0.58	0.83	13.8	0.48	0.087	86.9	0.96	0.91	0.26	0.35	0.45	1.2
PBAS	0.55	0.68	18.3	0.71	0.12	85.8	0.99	0.91	0.20	0.56	0.89	0.58
ViBe	0.48	0.83	16.0	0.54	0.11	82.8	0.93	0.65	1.13	0.30	0.49	1.2
propose	0.51	0.82	15.4	0.204	0.12	36.1	0.94	0.97	0.14	0.514	0.74	0.7

Tab.3 Complex background

Method	canoe			fountain02		
	R	P	P_{pwc}	R	P	P_{pwc}
Codebook	0.41	0.31	15.2	0.64	0.50	1.81
GMM	0.41	0.4	12.5	0.49	0.62	1.47
PBAS	0.81	0.35	17.3	0.92	0.45	2.14
ViBe	0.71	0.57	8.46	0.51	0.56	1.61
propose	0.86	0.78	3.87	0.77	0.71	0.96

IV. CONCLUSION

In this paper, a background modeling method based on multi-feature fusion is proposed. The initial background model is quickly established by using pixel neighborhood correlation in single-frame images, and various features of pixel value, frequency, update time and adaptive dynamic coefficient of video image sequence are utilized. Effectively improve the ghost phenomenon, reduce the hole of the moving target and the false foreground caused by pixel drift. By testing multiple sets of data, the experimental results show that the proposed algorithm improves the adaptability and robustness to dynamic background and complex background. The results of the algorithm in this paper are not subjected to any post-processing of the image (such as morphological processing), which leads to the partial incompleteness of the extraction foreground target, and the improvement of the algorithm such as filling holes in the later stage is needed to extract a more complete moving target.

REFERENCES

[1] Ueng S K, Chen G Z. Vision based multi-user human computer interaction[J]. Multimedia Tools and Applications, 2016(16): 1-18.

[2] HOCINE L, CAO Wei, DING Yong, et al. Adaptive learning rate GMM for moving object detection in outdoor surveillance for sudden illumination changes[J]. Journal of Beijing Institute of Technology, 2016, 25(1): 145–151. doi: 10.15918/j.jbit1004-0579.201625.0121.

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

[4] Lipton A J,Fujiyoshi H, Patil R S.Moving target classification and tracking fromreal-timevideo[C]. //Applications of Computer Vision, 1998. WACV'98.Proceedings, Fourth IEEE Workshop on. IEEE, 1998:8-14.

[5] Barron J,Fleet D, Beauchemin S. Performance of optical flow techniques[J]. InternationalJournal of Computer Vision, 1994, 12: 43-77.

[6] VIGNON D, LOVELL B C, ANDREWS R J. General purpose real-time object tracking using Hausdorff transforms[C] //9th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Annency, France, July 1-6, 2002: 1-6.

[7] NUMMIAFO K, KOLLER-MEIER E, GOOL L V. Color features for tracking non-rigid objects[J]. Acta Automatics Sinica, 2003, 29(3): 345-355.

[8] Xue G J, Song L, Sun J. Foreground estimation based on linear regression model with fused sparsity on outliers. IEEE Transactions on Circuits and Systems for Video Technology, 2013, 23(8): 1346-1357.

[9] Bi G L, Xu Z J, Chen T, et al. Complex background model and foreground detection based on random aggregation[J]. Acta Physica Sinica, 2015, 64(15): 150701.

[10] Sui Z S, Li J S, Zhang J, et al. Video foreground detection of tensor low-rank representation and spatial-temporal sparsity decomposition[J]. Optics and Precision Engineering, 2017, 25(2): 529-536.

[11] Wang Yong-Zhong, Liang Yan, Pan Quan, Cheng Yong-Mei, Zhao Chun-Hui. Spatiotemporal background modelingbased on adaptive mixture of Gaussians. Acta AutomaticaSinica, 2009, 35(4): 371-378.

[12] FAN Wen-chao, Li Xiao-yu, WEI Kai, CHEN Xing-lin.Moving Target Detection Based on Improved Gaussian Mixture Model[J]. Computer Science, 2015, 42(5): 286-288.

[13] Huo Dong-Hai, Yang Dan, Zhang Xiao-Hong, Hong Ming-Jian. Principal component analysis based codebook back-ground modeling algorithm. Acta Automatica Sinica, 2012, 38(4): 591-600.

[14] Barnich O,VAN DroogenbroeckViBe: A Universal Background Subtraction Algorithm for Video Sequences[J]. IEEE Transactions on Image Processing. 2011, 20(6): 1709-1724.

[15] Zhang Z B, Yuan X B. An improved PBAS algorithm for dynamic background[J]. Electronic Design Engineering, 2017.

[16] Wang Y, Yipierre J, Fatih P, et al. CDnet 2014: An expanded change detection benchmark dataset[C]. 2nd IEEE Change Detection Workshop, in conjunction with CVPR, 2014, 6: 393-400.