Detection of Moving Infrared Small Target Based on Fusion of Multi-gradient Filter and Vibe Algorithm

Guofeng Zhang[1,2] and Askar Hamdulla[1(✉)]

[1] Institute of Information Science and Engineering, Xinjiang University, Urumqi 830046, China
askar@xju.edu.cn
[2] Xinjiang Changji Vocational and Technical College, Changji 831100, China

Abstract. Aiming at the problem that dim small targets are submerged to complicated background in infrared images, it is difficult to complete extraction from background and noise clutter. An improved vibe algorithm is proposed for small target detection and tracking. First, target areas are extracted and stored by using vibe algorithm in every frame of video, meanwhile local multi-gradient filter is used to detect and store prominent edge information in each same frame of video. Then, fusion image is obtained through vibe algorithm and multi-gradient filter. Finally, a threshold separation technique is used to further eliminate background clutter and extract small targets. The experimental results show that proposed algorithm can quickly eliminate ghosts and is effective for detecting moving small targets. Compare to other background difference method, gaussian mixture model, experimental evaluation results show that our method outperforms vibe, background difference method and gaussian mixture model methods in terms of both tracking accuracy and computation speed for detection infrared small targets.

Keywords: Vibe algorithm · Ghost · Gradient filter · Image fusion

1 Introduction

High efficiency and high precision target detection and tracking technology have been widely applied in military and civil fields. For long-range infrared imaging with a distance of 13800 km, imaging target occupies only a few pixels and no shape, texture or other characteristics. Average brightness is low, motion law is indeterminate, the size of target is less than 9×9 pixels.

At the same time, due to jitter of infrared imaging equipment and effect of noise from imaging equipment, weak and small targets are submerged in the background. These problems make it more difficult to search and detect small targets in infrared sequence images. Traditional algorithms such as background difference method (BDM), is not friendly to adapt environmental change. It takes more time to build background model. Gaussian mixture model (GMM) is often used to detect moving small targets in complex background, However, it needs to extract a lot of frame of sequence images to

H. Song and D. Jiang (Eds.): SIMUtools 2020, LNICST 370, pp. 600–612, 2021.
https://doi.org/10.1007/978-3-030-72795-6_48

complete background model initialization, and real-time performance of the algorithm is not guaranteed. Vibe algorithm is mainly used for tracking of large targets, such as vehicle detection and tracking, face recognition, pedestrian detection, significance detection, etc. which is very effective [1], but there is no existing reference for infrared small target detection. There are two disadvantages in the application of vibe algorithm to detect small targets: one is that moving targets are regarded as background modeling in the first frame, which leads to generation of ghost in infrared images. The other is that ghost phenomenon caused by sudden movement targets which are stationary for a long time in previous several frames. Based on two problems, this paper proposes vibe algorithm of multi-gradient filter (GFVibe) to eliminate ghost and realize real-time target detection. The designing flow chart is shown in Fig. 1. Firstly, foreground feature map of moving target is obtained by using vibe algorithm, but detection result may generate ghost and background clutter. Therefore edge significant information of small targets is obtained by multi-gradient filter operator in the same time, this is target feature map. Then, fusion of foreground feature map and target feature map generates more accurate small targets. Finally, threshold separation technique is applied to eliminate residual background clutter and get real target area. Compared with traditional vibe algorithm, GFVibe algorithm proposed can compensate the disadvantage that moving targets generated ghost region, and also overcome flicker pixel point problem which is difficult to completely eliminate various noises when using vibe algorithm to detect small targets.

2 Feature Analysis for Point Target Image

The image of point target under complex conditions can be described as three parts [2, 3]:

$$I(x, y, t) = B(x, y, t) + T(x, y, t) + N(x, y, t) \tag{1}$$

where $I(x, y, t)$ represents infrared image contained point target, $B(x, y, t)$ represents background areas, $T(x, y, t)$ represents dim point targets, $N(x, y, t)$ represents random noise clutter, (x, y) represents current pixel location, t denotes different intervals of frames. Fig. 1 represents the detection flow chart of proposed algorithm.

3 The Vibe Algorithm [3, 4]

Vibe algorithm is a background modeling algorithm based on pixel level, which is mainly used for target detection. It has some advantages that other algorithms do not have, such as fast operation speed and small memory usage. The algorithm adopts random neighborhood pixel value to establish background modeling. Firstly, it establishes a background model sample set for each pixel. The elements of sample set come from random neighborhood pixel. Then, new pixel value at the next frame is compared with a set of values taken in the past at the same location or in the neighborhood. Euclidean distance is used to determine whether new pixel value belongs to the background, vibe algorithm adapts the model by choosing randomly which values to substitute from the background model. That is classification process of background pixel. The algorithm mainly consists of three core parts: background model initialization, extraction of foreground feature map and background model updating mechanism.

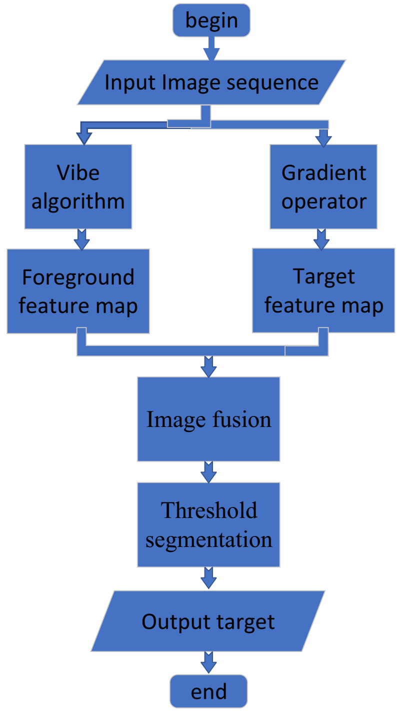


Fig. 1. The detection flow chart of proposed algorithm.

3.1 Background Model Initialization

In vibe algorithm, the essence of background model initialization is the process of establishing a sample set for each pixel. Since pixels in the image cannot contain spatio-temporal information, vibe algorithm uses image sequence and neighborhood pixel information to construct spatio-temporal distribution characteristics. Specific implementation method is as follows: each pixel in the image randomly selects gray value of its neighboring pixel as its sample values. In order to ensure that background model conforms to statistical laws, the range of neighborhood should be large enough. Figure 2 is a sample set of a center pixel point N(x, y) filled with the values of eight neighborhood pixels from background model.

$N(2)$	$N(2)$	$N(3)$
$N(4)$	$N(x, y)$	$N(5)$
$N(6)$	$N(7)$	$N(8)$

Fig. 2. Eight neighbor pixels of N(x, y).

3.2 Extraction of Foreground Feature Map

No matter whether there are targets in the image, the first frame is considered as background by default for vibe algorithm, and foreground feature map is detected from the second frame image. Moving target detection consists of following two steps:

The first step is target candidate, and the second step is target confirmation. In the first step, difference value $I_t(x, y)$ of grayscale is calculated and stored between the current pixel value $N_t(x, y)$ and all sample values N(k) that have been stored in the past in this neighborhood. Compared with the pre-setting Euclidean distance R. If the difference value $I_t(x, y)$ is greater than the Euclidean distance R, it is considered candidate target. The second step is to count the number of NUM which is more than Euclidean distance R in the first-step storage. If NUM is less than pre-set #min, it is confirmed as final target. The flow chart is shown in Fig. 3.

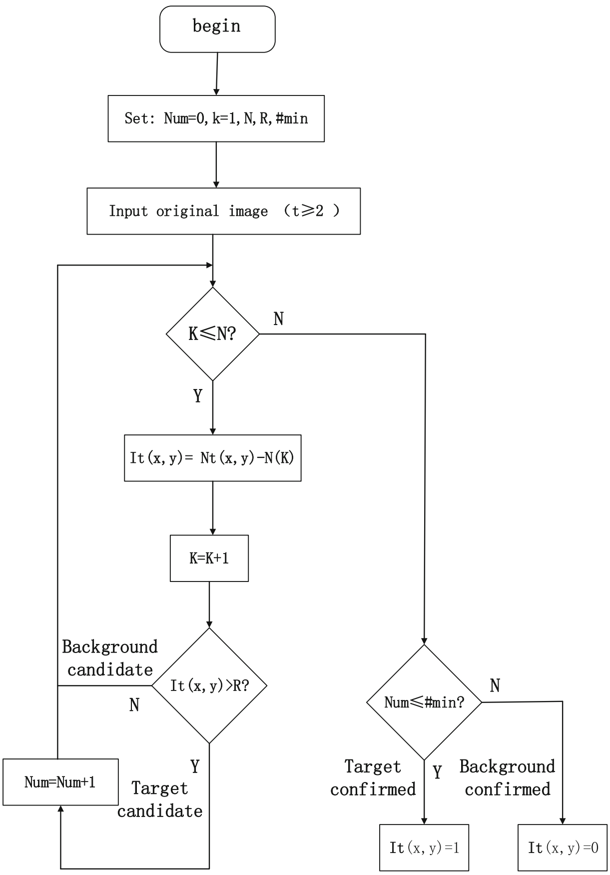


Fig. 3. Small target detection flow chart based on vibe algorithm.

3.3 Improvement of Model Update

In order to adapt to background environment of possible mutations, vibe establishes the time-sampling updating strategy [5] and the spatial neighboring updating mechanism [6]. Update strategy of time-sampling depends on time factor Ø, what can do is not need to update processing when process data of each frame, but according to a specific probability to update the background model. When a pixel is classified as background, it has a probability of 1/Ø to update background model samples or background samples of a neighboring pixel. Spatial neighboring update strategy: When background samples need to be updated, it is updated by randomly selecting a pixel value in the neighboring domain and replacing selected background model with newly randomly selected pixel. It is worth noting that when updating N sample values in a pixel sample's set, it is random to choose one from N samples to update. Therefore, the probability of the occurrence of sample value in the model decays exponentially, the (N − 1)/N probability of a sample present in the model is not updated at time t. Assuming time continuity and the absence of memory in the selection procedure, we can derive a similar probability, denoted as P(t, t + dt), for any further time t + dt. This probability is equal to (2).

$$p(t, t + dt) = \left(\frac{N-1}{N}\right)^{(t+dt)-t} = e^{-\ln(\frac{N}{N-1})dt} \tag{2}$$

Narayanan first proposed that whether update of a sample value in the model has nothing to do with time t, and random update strategy is reasonable. In other words, the past has no effect on the future. This property, called the memory less property, is known to be applicable to an exponential density. Therefore, the update of vibe algorithm is random in both time and space. Because there is no memory update mechanism [6], the dynamic adaptability of the algorithm is improved, and the sample set ensures the smooth attenuation of the life cycle of each pixel value. To solve mutation of image content caused by special reasons such as weather change and light change, vibe algorithm sets a specific threshold. When certain conditions are met, it determines whether the background model needs to be re-initialized. In general, since the difference in background changes is not obvious, the number of NUM_{update} for each background model update is similar. Therefore, depending on the number of background model updates each time, it is possible to determine whether required initialization of the model. Assume that the number of updates in the first frame is SUM_{update}, and then compare the number of updates in each frame from NUM_{update} to SUM_{update}.

$$\left|NUM_{update} - SUM_{update}\right| > 0.3 * SUM_{update} \tag{3}$$

When the difference value satisfies the above formula, the model can be re-initialized. After the model initialization is completed, target extraction can continue from the second frame. More and more attention has been paid to this method because of its fast and effective characteristics.

4 Multi-gradient Filter

Four gradient operators H1, H2, H3 and H4 (45°, 135°, 225°, 315°) are defined in this paper. Templates are respectively used to carry out convolution operation for each frame on the video image to obtain gradient convolution filter graph in four directions. Formula (4) below represents the convolution operation.

$$H_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad H_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad H_3 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad H_4 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$GF(x, y, t) = f(x, y, t) * H \tag{4}$$

Where f(x, y, t) represents sequence images in video at time t, GF(x, y, t) represents gradient convolution filter graph for corresponding sequence image in video at time t. H represents gradient operators. The size of image block is defined as W × W, and each gradient convolution filter graph $G_k (1 \le k \le 4)$ is traversed from left to right and from top to bottom. Each image block is divided into four sub-image blocks $\psi_k (1 \le k \le 4)$ on average, the gradient mean square value is respectively calculated in the corresponding direction of each sub-image block under the gradient convolution filter graph, it is denoted as g_{ψ_k} $(1 \le k \le 4)$, and calculate the maximum(Gmax) and minimum(Gmin) of the gradient mean square value as follows:

$$G_k = \frac{1}{N_j} \sum_{j=1}^{N_j} \left\| g^j_{\Psi_k} \right\| \tag{5}$$

$$G_{\max} = \max_{1 \le k \le 4} G_k \tag{6}$$

$$G_{\min} = \min_{1 \le k \le 4} G_k \tag{7}$$

Where N_j represents the number of pixels in the sub-image block. After the gradient operation is realized, the threshold is set as η. According to Eq. (8), gradient filter graph (GF) can be obtained.

$$G = \begin{cases} \sum_{i=1}^{4} G_i & \frac{G_{\min}}{G_{\max}} > \eta \\ 0 & 其它 \end{cases} \tag{8}$$

5 Image Fusion

The image obtained by vibe algorithm and GF image are fused for each frame in the video to obtain a new fusion image. This effectively suppresses presence of ghost and problem of flicker pixels point. The blue rectangle represents the sliding window and the red rectangle represents small target. The calculation process of fusion is shown in Eq. (9) (Fig. 4).

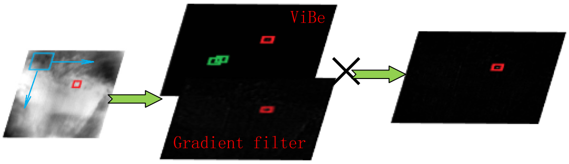


Fig. 4. The process of image fusion. (Color figure online)

$$GFVibe(x, y, t) = GF(x, y, t) \times Vibe(x, y, t) \tag{9}$$

Vibe algorithm can quickly and effectively extract moving small targets, and the gradient vector of small targets roughly points to the target center, but background clutter does not have such characteristics, so gradient operator filtering has a good performance of detecting edges, while fusion technology can effectively remove ghost or suppress isolated noise points to enhance the target areas.

6 Threshold Segmentation Method

After fusion image is obtained, in order to extract the small target, the segmentation method is used in fusion image, which is defined as Eq. 10.

$$T = \mu + k \times \sigma \tag{10}$$

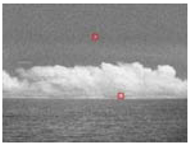
Where μ and σ is mean and standard deviation of the final salient GFVibe map, k is a adjustment parameter. Our experiments show that the optimal range of k is from 3 to 10. If the pixel value in the GFVibe map is greater than T, we affirm that it is a target, otherwise, it should be considered as background.

7 Experimental Evaluation

7.1 Selection of Parameters

Setting of some parameters in this experiment consults the reference [6, 7]. The size of background sample (N equals 20) and the minimum cardinality (#min equals 2) are both selected as the default. The initial Euclidean distance is improved to $R = 0.5 \times \sigma_m$ and is set in the integer range [20,40], in order to better adapt to the changes for background clutter. The random update factor is set to 4, the updating probability of background model sample is 25%. As the detected target tends to be a point target, the image to be tested in the experiment is two sets of sequence images with a size of 240 × 320 video under different backgrounds of 200 frames. Each frame of video image 1 contains two small targets and video image 2 contains one small target. Figure 5 and Fig. 6 show tracking results of GMM, background difference method, vibe and proposed algorithm (GFVibe) at two different backgrounds. Extraction results of 20, 60, 120 and 180 frames are shown in Fig. 5 and Fig. 6.

In this paper, neighborhood pixels is set to 24, the gradient operator is improved on the basis of Robert operator [7–9]: the center coefficient is negative and the sum of all



(a) Original infrared image

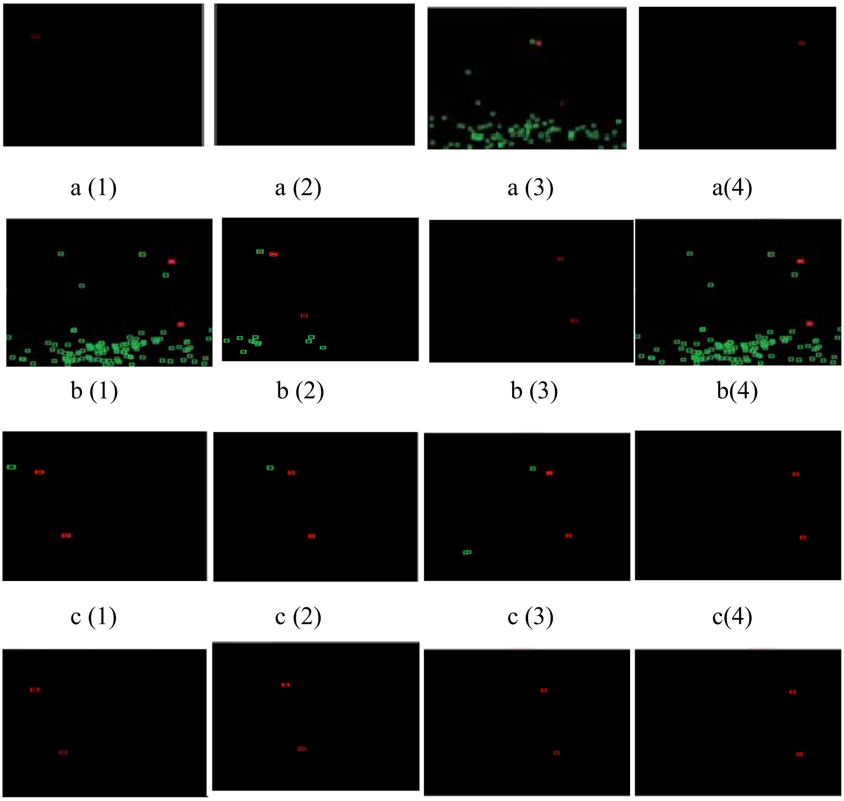
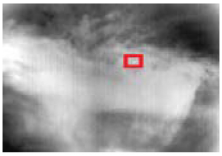


Fig. 5. Tracking results of different algorithm in video image 1.

coefficients is 0. The center of four gradient operators is set as the coefficient −1, and four diagonal edges (45°, 135°, 225°, 315°) are set as 1. The gradient is estimated according to the difference in any mutually perpendicular direction. Compared with horizontal and vertical edge detection, diagonal edge detection is more suitable for those edges with strong undulations whose boundary is not obvious. Gradient operator template matches gradient vector of small target region and is a template with a center, so the size of template must be odd. In order to achieve higher efficiency, the template in this paper adopts 3×3 for convolution filtering. According to literature [10–12], when the size of



(b) Original infrared image

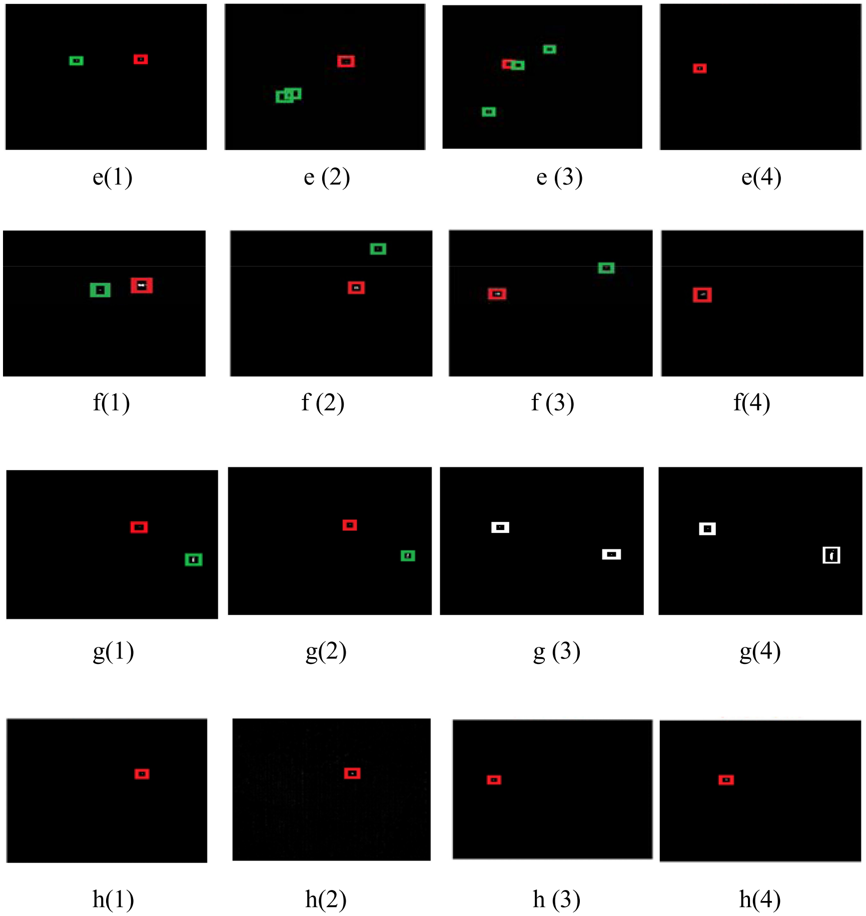


Fig. 6. Tracking results of different algorithm in video image 2.

image block increases beyond small target size, the new gradient data does not change significantly, so the size of local gradient sliding window only needs to be larger or closer to the target size. In order to evenly distribute 4 sub-image blocks, W is taken as an even number, that is 10×10, so that the sizes of 4 sub-image blocks are all set to 5×5, which can also satisfy the asymptotic change of gradient. After experimental data analysis, if it is target region, the maximum and minimum values of the gradient in the four directions have little difference. The edge area is quite different. Setting enough big is easy to miss targets, too small setting results in high false alarm rate. Experimental results show that η equal to 0.2 is the best effect. In the experimental image, the green rectangular represents the false targets and red rectangular represents real targets. The trajectory of small target is a curve, video image 1 and video image 2 are the tracking results of GMM (a1−a4, e1−e4), Background difference method (b1−b4, f1−f4),Vibe(c1−c4, g1−g4), GFVibe (d1−d4, h1−h4) in Fig. 5 and Fig. 6, respectively.

Vibe also achieved good tracking effect, but in video image 1, there is still a false alarm point in the 20th, 60th and frame 120th frame, until the false alarm point disappears in the 180th frame. It can be seen that successful background modeling of vibe requires a certain amount of frame accumulation to eliminate ghost and flicker pixels. The false alarm rate of GMM is very high in the complex sea background, and false negative rate is very serious, small targets are lost in 20th, 60th and 180 frames, respectively. Background difference method can correctly track small target in video image 1, but the false alarm rate is also very high in the 60th and 180th frame.The proposed algorithm GFVibe can correctly track small moving targets in the complex sea background. In video image 2, background difference method, GMM and vibe demonstrate good tracking results, The false alarm rate of GMM is slightly tall, ghost problems of vibe doesn't disappear for the same place in video image 2, the background difference method is slightly better than the vibe, to some extent, the background difference method can eliminate the ghosts, but it cannot eliminate flicker pixels, GFVibe not only can better eliminates ghost phenomenon existed, but also can restrains flicker pixel point, and achieve a good tracking effect.This good inhibition ability of GFVibe is due to good edge detection ability of gradient operator. Thus it can be seen that the proposed algorithm can correctly track moving small target under two different backgrounds. When there is a small target in the first frame. Ghost and flicker pixel removal for clutter noise is effective, robust, and excellent.

7.2 Quantitative Evaluation

Experimental results verify the comparison of runtime for GMM, background difference method, Vibe, GFVibe algorithm in two groups of video images, as shown in Table 1 and Table 2. T1, T2, T3 and T4 are the results of runtime of 200 frame sequence images under two different algorithms when the SNR (Signal-to-Noise Ratio) is 4, 3, 2, 1 dB. T(means) represents the average runtime of the same algorithm for 200 frame image sequences.

Table 1. Comparison of the mean runtime of different methods under video image 1 (unit: s).

Time	T1	T2	T3	T4	T(means)
GMM	76.368	267.739	268.142	269.153	220.351
BDM	77.587	269.766	270.586	269.616	221.888
ViBe	59.338	60.390	58.3390	59.300	59.342
GFVibe	**57.235**	**56.817**	**58.022**	**56.579**	**57.163**

Table 2. Comparison of the mean runtime of different methods under video image 2 (unit: s).

Time	T1	T2	T3	T4	T(means)
GMM	73.386	117.43	156.41	192.025	127.266
BDM	75.410	119.421	105.186	131.16	107.794
ViBe	53.783	50.190	51.431	54.012	52.354
GFVibe	**56.708**	**50.017**	**51.022**	**46.675**	**51.106**

Compared with GMM, BDM, Vibe algorithm consumes less time in video image 1 and in video image 2. It takes time about 0.297 s and 0.262 s to process a frame in video 1 and in video 2, and real-time is better. The runtime of GFViBe algorithm is basically stable under in two different background. It is shown in bold for Table 1 and Table 2. In Table 1 that average processing time of one frame is about 0.286 s and 0.256 s in video 1 and in video 2, which is about 4% and 2.3% shorter in video 1 and in video 2. Compared with the other three algorithms, the shorter runtime of the proposed method justifies its use in real time applications.

In addition, we also choose two common evaluation indicators for comparison, which are signal-to-clutter ratio gain (SCRG) and background suppression factor (BSF). The indicators are defined as follows:

$$SCRG = \frac{SCR_{\text{out}}}{SCR_{in}}, \quad BSF = \frac{C_{\text{in}}}{C_{out}}$$

where SCR_{in} and SCR_{out} represent the SCR values of the raw image and GFViBe map, respectively, and σ_{in} and σ_{out} represent the standard deviation of the raw image and GFViBe map, respectively. To a certain extent, SCRG can measure clutter suppression and target preservation, and BSF is metric which reflects the performance for clutter removal without the target information. Generally, the higher the SCRG is, the stronger ability of clutter suppression and target enhancement the method has.

Table 3. The average results of different detection methods for each sequence.

Seq.	Indicators	GMM	BDM	ViBe	GFVibe
1	$\overline{SCRG}$	14.57	24.06	49.42	**55.92**
	$\overline{BSF}$	6.632	8.54	23.65	**28. 87**
2	$\overline{SCRG}$	20.50	23.48	38.49	**42.31**
	$\overline{BSF}$	17.62	19.15	27.36	**38.47**

Table 3 presents the average results of different methods for each sequence. SCRG and BSF represent the average results. It can be seen that vibe achieve a better performance than the other baseline methods, because they have the steps of backgrounds prediction or preservation, and the predicted backgrounds are subtracted from the original image. Thus, they can achieve large SCRG and BSF. However, **GMM** and **BDM** do not specialize in the background suppression, thus their performance is not superior to others. The proposed approach achieves the best performance to enhance the target areas. On the two metrics. SCRG and BSF of our method are large extremely, because GFVibe fuses gradient operator on the basis of vibe, which has extremely capability of edge detection to remove ghost or flicker pixel points.

8 Conclusion

This paper focuses on the key technology of small target tracking, which involves the detection and tracking of moving small targets. From experimental evaluation, it can be seen that the proposed algorithm in this paper has certain practical significance in eliminating ghost and suppressing flicker pixel points. GFVibe has strong ability of background modeling and performance of good edge detection duo to introduction of gradient operator. Fusion image has obvious advantages in eliminating ghost and isolated noise. The proposed algorithm can be relatively stable and correctly track moving small target. This algorithm is the least time-consuming. In this paper, we only extract the simple small targets in the static background. Whether small targets can be accurately extracted in complex background will be explored in future work. Compared with GMM, BDM, Vibe tracking detection algorithms, GFVibe has better stability and shorter runtime.

Acknowledgments. This work has been supported by the National Natural Science Foundation of China (under grant number of 61563049).

References

1. Li, J., Li, S.-J., Zhao, Y.-J., Ma, J.-N., Huang, H.: Background suppression for infrared dim and small target detection using local gradient weighted filtering. In: International Conference on Electrical Engineering and Automation (2016)
2. Marvasti, F.S., Mosavi, M.R., Nasiri, M.: Flying small target detection in IR images based on adaptive toggle operator. IET Comput. Vis. **12**(4), 527–534 (2018)
3. Zhang, H., Zhang, L., Yuan, D., et al.: Infrared small target detection based on local intensity and gradient properties. Infrared Phys. Technol. **89**, 88–96 (2017)
4. Barnich, O., Van Droogenbroeck, M.: ViBe: a universal background subtraction algorithm for video sequences. IEEE Trans. Image Process. **20**(6), 1709–1724 (2011)
5. Henriques, J.F., Caseiro, R., Martins, P., Batista, J.: High-speed tracking with kernelized correlation filters. IEEE Trans. Pattern Anal. Mach. Intell. **37**(3), 583–596 (2015)
6. Bai,K., Wang, Y., Song, Q.: Patch similarity based edge-preserving background estimation for single frame infrared small target detection. In: IEEE International Conference on Image Processing, pp. 181–185 (2016)
7. Zhang, L.Y., Du, Y.X., Li, B.: Research on threshold segmentation algorithm and its application on infrared small target detection algorithm. In: International Conference on Signal Processing, pp. 678–682 (2015)
8. Wei, H., Tan, Y., Lin, J.: Robust infrared small target detection via temporal low-rank and sparse representation. In: International Conference on Information Science, pp. 583–587 (2015)

9. Yao, Y., Hao, Y.: Small infrared target detection based on spatio-temporal fusion saliency. In: 2017 17th IEEE International Conference on Communication Technology
10. Zhang, H., Niu, Y., Zhang, H.: Small target detection based on difference accumulation and Gaussian curvature under complex conditions. Infrared Phys. Technol. **87**, 55–64 (2017)
11. Xie, K., Fu, K., Zhou, T., Yang, J., Wu, Q., He, X.: Small target detection using an optimization-based filter. In: International Conference on Acoustic, pp. 1583–1587 (2015)
12. Deng, L., Zhu, H., Zhou, Q., Li, Y.: Adaptive top-hat filter based on quantum genetic algorithm for infrared small target detection. Multimed. Tools. Appl. **77**, 10539–10551 (2018)