Infrared Physics & Technology 85 (2017) 465–477

Contents lists available at ScienceDirect

Infrared Physics & Technology

journal homepage: www.elsevier.com/locate/infrared

Regular article

Dim small targets detection based on self-adaptive caliber temporal-spatial filtering

Xiangsuo Fan [a,b,c,*], Zhiyong Xu [a], Jianlin Zhang [a], Yongmei Huang [a], Zhenming Peng [b]

[a] Key Laboratory of Beam Control Chinese Academy of Sciences, the Institute of Optics and Electronics the Chinese Academy of Sciences, Chengdu 610209, China
[b] School of Optoelectronic Information, University of Electronic Science and Technology of China, Chengdu 610054, China
[c] University of Chinese Academy of Sciences, Beijing 100039, China

HIGHLIGHTS

- An improved anisotropy for background prediction is proposed.
- An improved high-order cumulates is proposed, which increased the SNR of images.
- Compared with other methods, the proposed method achieved a noticeable improvement.
- The detection algorithm is proposed, which can detect the lowest SNR is 0.37.

ARTICLE INFO

Article history:
Received 14 February 2017
Revised 12 August 2017
Accepted 13 August 2017
Available online 16 August 2017

Keywords:
Dim small target detection
Background suppression
Improved anisotropy for background prediction
Improved high-order accumulation
Self-adaptive caliber
Time-space domain

ABSTRACT

To boost the detect ability of dim small targets, this paper began by using improved anisotropy for background prediction (IABP), followed by target enhancement by improved high-order cumulates (HQS). Finally, on the basis of image pre-processing, to address the problem of missed and wrong detection caused by fixed caliber of traditional pipeline filtering, this paper used targets' multi-frame movement correlation in the time-space domain, combined with the scale-space theory, to propose a temporal-spatial filtering algorithm which allows the caliber to make self-adaptive changes according to the changes of the targets' scale, effectively solving the detection-related issues brought by unchanged caliber and decreased/increased size of the targets. Experiments showed that the improved anisotropic background predication could be loyal to the true background of the original image to the maximum extent, presenting a superior overall performance to other background prediction methods; the improved HQS significantly increased the signal-noise ratio of images; when the signal-noise ratio was lower than 2.6 dB, this detection algorithm could effectively eliminate noise and detect targets. For the algorithm, the lowest signal-to-noise ratio of the detectable target is 0.37.

1. Introduction

At present, detection on small moving targets in the strong-clutter environment is a core technology of the photo-electronic imaging system and other systems, and also the research hotspot and difficulty of this field. The main existing detection algorithms include single-frame detection and multi-frame detection. Single-frame detection mainly includes median filter [1], top-hat transform algorithm [2], the adaptive filter technology, the detection algorithm of directional support value of Gaussian transformation [3], the detection algorithm of multi-scale facet fitting model [4], etc. Median filter has a stable filter performance as it can effectively hold target signals while eliminating random noise. However, limited by its structure, median filter can only eliminate random noise whose pulse width is smaller than the width of the filter window, and its structure background cannot be changed. The top-hat transform algorithm is a practical nonlinear background prediction technology which has a better effect than median filter. However, the top-hat transform algorithm has a poor adaptively because it requires knowing prior knowledge of the images in advance. To improve the adaptively of this algorithm, some scholars introduce the adaptive filtering technology into the research on image preprocessing, such as the two-dimensional least-mean square (TDLMS) filter [5], the least squares support vector machine (LS-SVM) [6], the adaptive

* Corresponding author at: Key Laboratory of Beam Control Chinese Academy of Sciences, the Institute of Optics and Electronics the Chinese Academy of Sciences, Chengdu 610209, China.
E-mail addresses: ffanxs@163.com (X. Fan), xzy158@163.com (Z. Xu), huangym@ioe.ac.cn (Y. Huang).

http://dx.doi.org/10.1016/j.infrared.2017.08.010
1350-4495/

lattice filter [7], the kalman filter [8], the wiener filter [9], etc. The top-hat transform algorithm predicts according to relevance of the background in the neighborhood, adaptively adjusts parameters of the filter according to the errors between the predicted background and the actual background, and then realizes background suppression by comparing the predicted background and the original image. This adaptive algorithm does not require understanding prior knowledge of images and has a simple structure. However, its shortage is that it requires the statistic characteristics of the background keep unchanged or change slowly. For the large-span backgrounds and the quick changing scenes, some scholars try to realize detection by separating the "gray singularity" of the target area formed due to "gray disturbance" of images under the effect of the target. For example, Wang et al. [10] separated the gradient characteristic differences of the target and the background through anisotropic diffusion filtering, so as to improve the signal-to-noise ratio of images. However, it is a pity that this algorithm adopts unilateral diffusion can only negatively hold but cannot enhance the target signal. On the other hand, this algorithm does not require making assumptions on the statistic characteristics and gray characteristics of the background, and it has a simple structure and has a large advantage in the field of engineering application. In recent years, when the gray singularity of the target and the background texture structure in the complicated background are exactly similar, it is difficult for the traditional single-frame detection algorithms (such as top-hat, max-mean filter, max-median filter, min-local-Laplacian of Gaussian (LoG) filter, as well as LS-SVM) to effectively suppress the complicated background and highlight the target. Thus, to effectively detect the small targets in the complicated background, Yang et al. proposed many detection algorithms [3,4], and verified that these two algorithms can obtain better detection effects than the traditional method in case of the complicated background.

The single-frame detection algorithm can effectively detect the target to a certain extent, but it does not take full advantage of movement information of the target among different frames. To eliminate noise interference by using the track continuity of the small moving targets sufficiently and to further improve the detection rate, the multi-frame detection algorithm is used extensively at present. The pipeline filter algorithm is a classical multi-frame detection algorithm. The pipe diameter of the traditional pipeline filter algorithm is fixed, but the dimensions of the actual small moving targets keep changing. Thus, the pipeline filter algorithm can only detect the targets whose dimensions are smaller than the pipe diameter and cannot adapt to changes of the target dimensions. As a result, the pipeline filter algorithm is heavily limited in the practical application. The targets cannot be detected effectively unless the pipe center moves along with the movement of the target. Liu et al. [11] proposed the moving weighted pipeline filter algorithm which can effectively suppress the interference of edge noise to target detection. Dong et al. [12] proposed the motion estimation pipeline filter algorithm which can substitute the actually detected position of the small target with the estimated position of the small target when encountering obstructers. However, generally the pipe diameter of these algorithms keep unchanged, which obtains a good target detection effect when the target dimensions are small than the pipe diameter, but leads to decrease in the detection performance when the target dimensions are bigger than the pipe diameter because the target would overflow from the pipe. Therefore, Dong et al. [13] proposed to simplify the target into a point, so as to ensure that the pipe can constantly hold the target. However, losing the ability of comparing the target and noise, this simplification method may lead to difficulty in distinguishing the target point and the noise point and increase the detection risk. In fact, whether the target becomes bigger or smaller, the detection method with the fixed pipe diameter will cause problems. In case that the pipe diameter keeps unchanged, the proportion of noise of the pipe will increase correspondingly when the target becomes smaller, and the target may overflow from the pipe when the target becomes bigger.

The sizes of the small moving target keep changing actually, which is caused by two reasons. The first one is the relative motion between the target and the observer, and the second is the influence of the detection distance of the imaging system on the target size. For this realistic condition, this paper presents a detection algorithm with the pipe diameter adapting to the dimension change of the target. To improve the detection ability of the algorithm for the follow-up targets, the improved anisotropy is used to detect the background, then the improved high-order cumulates is adopted to enhance the target, and finally the scale space algorithm is introduced to adaptively adjust the pipe diameter, thereby effectively improving the detection performance.

2. Improved anisotropy for background prediction

The characteristics of the anisotropy lie in that it can smoothen and stabilize the background area and maintain the edge details and sudden change area of the background. The diffusion equation of anisotropy is:

$$\frac{\partial u(x,y,t)}{\partial t} = div[c(\nabla u)]\nabla u \tag{1}$$

where u is the grayscale image, ∇u is the gradient, $c(\nabla u)$ is the edge stop function, and div is the divergence operator. The edge stop function $c(\nabla u)$ is used to calculate the smoothing factor according to the gradient relations in different directions. Literature [14] presents the anisotropic edge stopping function as Eq. (2):

$$c(\nabla u) = \exp[-(|\nabla u|/k)^2] \tag{2}$$

where k is a constant; for the flat area with a small gradient, the value of $c(\nabla u)$ is large, and more smoothening treatment is adopted; while for the sudden change area with a big gradient, the value of $c(\nabla u)$ is small, little or no smoothening treatment is adopted, so that these areas are maintained.

After analyzing the image of the small targets, it can be seen that differentiation treatment of different characteristic areas can be realized by virtue of the characteristic difference between the target area and other areas in the directions. The gradient operator of the local area where the small targets are located is shown in formula (3):

$$\begin{cases} Up\_Grad = f(i,j) - f(i-step,j) \\ Down\_Grad = f(i,j) - f(i+step,j) \\ Left\_Grad = f(i,j) - f(i,j-step) \\ Right\_Grad = f(i,j) - f(i,j+step) \end{cases} \tag{3}$$

where *Up_Grad*, *Down_Grad*, *Left_Grad* and *Right_Grad* are the upper gradient, the lower gradient, the left gradient and the right gradient with $f(i,j)$ as the center.

If the mean value of two minimum parameters $\min_1$ and $\min_2$ of the original anisotropic edge-stopping function in the four directions is chosen for pixel-by-pixel filter of the images, it is found that the parameter values of the stable background and the non-stable background are relatively large, while the parameter value of the singular area (target signal) is relatively small, which however can only negatively hold but cannot enhance the target signal. The filter formula is as follows:

$$\begin{cases} \min_1 = \min\{c(Up\_Grad), c(Down\_Grad), \\ \quad c(Left\_Grad), c(f(Right\_Grad)\} \\ \min_2 = \min\{\{c(Up\_Grad), c(Down\_Grad), \\ \quad c(Left\_Grad), c(f(Right\_Grad)|not\ \min_1\} \end{cases} \tag{4}$$

Research finds that as the gradient value of the stable area in the four directions is small, the value of the edge-stopping function given in the literature [14] is large; while as the gradient value of the singular area in the four directions is large, the value of the edge-stopping function is small. If the edge-stopping function in the literature [14] is used for filtering, the function value of the area will be small. Thus, to highlight the singular area, weaken the stable area and realize the purpose of enhancing the target signal, the edge-stopping function is improved as follows:

$$c(\nabla u) = 1 - \exp[-(|\nabla u|/k)^2] \tag{5}$$

Finally, formula (5) is put into formula (3), and then formula (4) is used to calculate the mean value of the two minimum parameters $\min_1$ and $\min_2$ in the four directions for pixel-by-pixel filter of the images, thereby smoothly highlighting the small target.

3. Improved HQS

High-order cumulates can effectively accumulate energy in the time and space domains, suppress Gauss noise and enhance transient signals, thereby realizing the purpose of enhancing energy of the small target. Literature [15] gives the definition for the M frame high-order cumulates:

$$C_{Mf} = E\{F_0(x,y,t_1) + F_0(x,y,t_2) + \cdots + F_0(x,y,t_M)\} \tag{6}$$

where $F_0(x,y,t_M)$ represents the image with the background removed, t_M represents the $t_M (M = 1,2,\ldots,n)$ frame image, and M represents the number of image accumulated frames.

The original high-order cumulates only consider the time-domain characteristics, so that it certainly influences the enhancement effect. To better accumulate its energy in the space domain, the space-domain characteristics (motion information of the target) should be considered. The motion of the target in the adjacent frames can be described as the 12 patterns shown in Fig. 1. The first five forms are horizontal movement. The middle five forms are vertical movement and the last two forms are diagonal movement, as shown in Fig. 1:

Whatever directions, the target always moves in the continuous neighborhoods of adjacent frames. Energy accumulation of the moving target can be realized by accumulating the maximum energy value of M continuous frames of images in the motion neighborhood. The motion energy accumulation of the target can be described as:

$$\begin{aligned} f_p(x,y,t_M) &= \sum_{m=-r,n=-r}^{m=r,n=r} T_p(x,y) \times F_0(x+m,y+n,t_M) \\ P_0(x,y,t_M) &= \arg \max_p f_p(x,y,t_M) \end{aligned} \tag{7}$$

where T_p represents the target movement model (see Fig. 1), r represents the radius of the accumulation window, $F_0(x,y,t_M)$ represents the sequence image with background removed, $P_0(x,y,t_M)$ represents the maximum value of the t_M $(M = 1,2,\ldots,n)$ frame image among the 12 movement model. The improved M frame high-order cumulates can be defined as:

$$C_{Mf} = E\{P_0(x,y,t_1) + P_0(x,y,t_2) + \cdots + P_0(x,y,t_M)\} \tag{8}$$

4. Detection algorithm

4.1. Scale-space theory

(1) From the scale space point of view, the image can be decomposed by convolving the image$f(x,y)$ with the Gaussian kernel of some scale σ. That is, in scale space the scaled image $L(x,y,\sigma)$ is:

$$L(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\pi\sigma^2} * f(x,y) \tag{9}$$

(2) Constructing differential scale space with difference of Gaussian (DoG), i.e.:

$$D(x,y,\sigma) = L(x,y,n\sigma) - L(x,y,\sigma) \tag{10}$$

wherein σ is the scale factor, $f(x,y)$ represents the image, $L(x,y,n\sigma)$ means the scale space whose scale factor is $n\sigma$. For a small target of 2×2, the minimum value of σ is 0.5. To show continuity of the scale, multiple adjacent scale factors of each scale are made be n , it is to say that we will have$\sigma_n = n\sigma$. The value of n is typically the n-th power of 1.2.

(3) Using the DoG algorithm to calculate the extreme spatial points. The extreme values of all adjacent points of its space domain and scale domain can be found by comparing every pixel of the medium layer with 8 adjacent pixels at the same layer and 9 pixel points of the two immediate upper and lower layers. Obtaining the extreme value point, it can be selected as a candidate target point and its location coordinates (x,y) and corresponding scale σ are recorded. So the diameter d of the target could be calculated as:

$$d = 2\sqrt{2}\sigma \tag{11}$$

wherein σ is the corresponding scale factor, and d represents the diameter of the candidate target.

(4) Using local signal-to-clutter ratio threshold value detection method for an accurate location of the candidate target points. The representation is as below:

$$\begin{cases} LSCR(x,y,\sigma) = \frac{\mu_t - \mu_b}{\sigma_b} \\ LSCR(x,y,\sigma) > \theta \end{cases} \tag{12}$$

wherein θ usually falls within the range from 1 to 3. $LSCR(x,y,\sigma)$ is the target point's local signal-to-clutter ratio.(x,y) is the coordinate of the candidate point. σ is the corresponding scale, μ_t is the candidate point's local mean value, μ_b is the mean value of the local background, and σ_b is the local background's standard deviation.

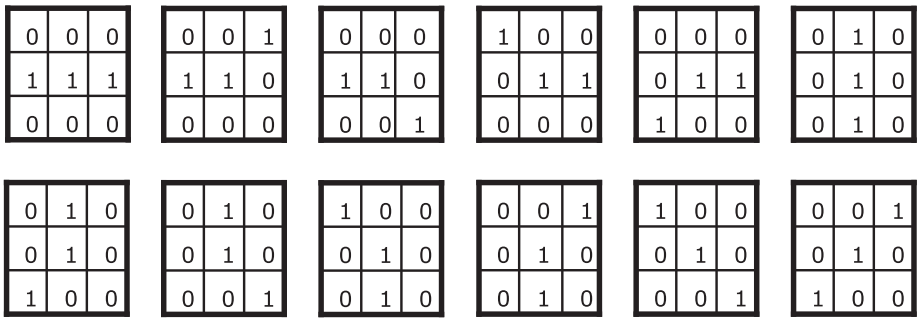


Fig. 1. Target movement model.

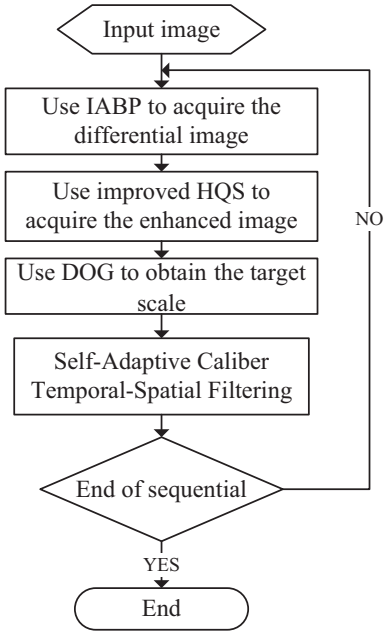


Fig. 2. The frame flow chart.

The size of the target region is $2\sqrt{2}\sigma$. The size of local background region is taken as $6\sqrt{2}\sigma$.

4.2. Procedure of the algorithm

To improve the detection ability against the small targets, the sequence image should be pretreated, including removing background interference and enhancing energy. This paper acquires the differential image by virtue of improved anisotropy, and then enhances the differential image by using the improved high-order cumulates, and finally realizes scale change of the enhanced differential image during the detection process via the target scale obtained through the DOG algorithm described in Section 4.1. The frame flow chart is shown in Fig. 2:

By combining the frame flow chart of Fig. 2, the specific algorithm description is as follows

Algorithm.

(1) **Initialize the parameters:** R_1-target search neighborhood region, R_2-target movement neighborhood region, Num-number frames of images, The lower limit and upper limit of the restricted candidate target area are $Data_1$ and $Data_2$, respectively. R_1=15; R_2=6; $Data_1$=2; $Data_1$=25;
(2) Use IABP to acquire the difference graph. Then use improved HQS to enhancement the difference graph. Finally, use literature [16] presents the local maximum value partitioning (LMVP) to obtain the binary image of the sequence
(3) **Input:** Input N (usually the range of N is 5 to 9) frames of binary image. Use the pipeline filtering algorithm to acquire the true target(x,y). x_1=x; y_1=y;
(4) **Output:** the target in every frame of image's center (x,y)
1: FOR T=1:Num
2: Record the T's width as m and high as n
3: FOR row=1:m
4: FOR col=1:n
5: IF row<x- R_1 |row>x+ R_1 | col<y- R_1 | col>y+ R_1
6: Record all candidate targets x_i (i=1, 2, 3, ...)
7: END IF
8: END FOR
9: END FOR
10: Record a candidate target point x_i (i=1, 2, 3, ...) as (X,Y)
11: Record a candidate target point x_i (i=1, 2, 3, ...) area as Ar
12: IF X<x- R_2 | X>x+ R_2 | Y<y- R_2 | Y>y+ R_2
13: continue
14: END IF
15: IF Ar>=$Data_1$ & Ar<$Data_2$
16: Record all points meeting the two conditions above like as:
17: data_ncenter=[data_ncenter,X]
data_mcenter=[data_mcenter,Y]
data=length(data_mcenter(:))
18: IF data==1
x=data_ncenter
y=data_mcenter
x_2= x; y_2= y;
19: ELSE
20: Use the Kalman algorithm to predict the target, [cc,cr] = kalman_pre(x,y)
21: Use the closest proximity method to acquire the target point in the predicted neighborhood region and record the target (x,y)
22: [cc_new, cr_new]=min_distance(data_mcenter,data_ncenter,cr,cc)
23: x=cc_new
24: y=cr_new
25: x_2= x; y_2= y;
26: END IF
27: END IF
28: Update the next frame of R_1, use DOG algorithm in Section 4.1 to acquire the diameter d, $R_1 = c_1 \times d$
29: Update R_2, $R_2 = c_2 \times \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$,(x_2, y_2)-the current frame's target point, (x_1, y_1)-the previous frame's target point
30: END FOR
(5) Record the points (x,y) of the target in every frame of image and export corresponding detection result

5. Experiment results and analysis

5.1. Performance improvement evaluation

For evaluation of background prediction result, in this study, we use three indexes, Mean Squared Error (MSE) [17], Structural Similarity (SSIM) [18] and local Signal-to-noise Ratio Gain (GSNR) [19], to evaluate the effect of image background prediction. The enhancement effect of high-order accumulation is evaluated by using the target's average grayscale value (AGV) and the image's local signal-to-noise ratio (LSNR).

(1) MSE is used to calculate the average error between each pixel value of the predicted background image and the real background image. The equation is as below:

$$\mathrm{MSE} = \frac{1}{MN}\sum_{x=1}^{N}\sum_{y=1}^{M}[R(x,y) - F(x,y)]^2 \tag{13}$$

where F is the predict background image; R is the real background image (because of dim and small infrared target is very weak, so use

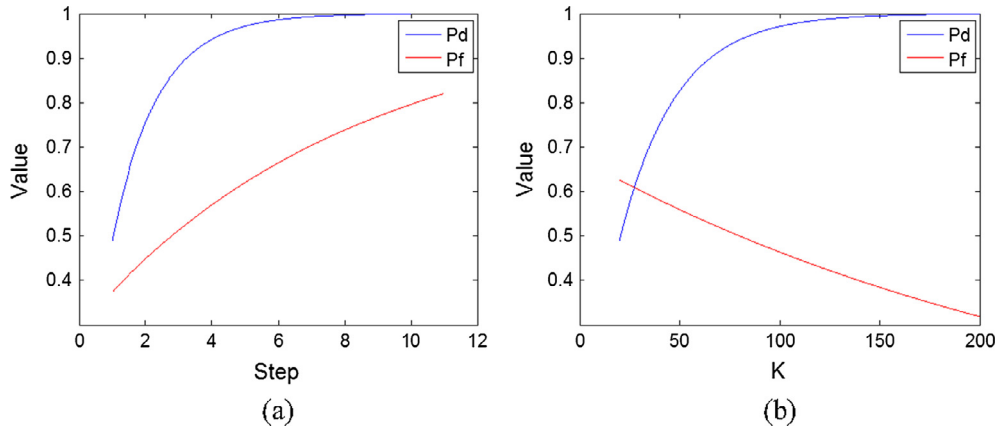


Fig. 3. Relations between *step*, *k*, detection rate, and false alarm rate.

infrared image as real background image); M and N are image width and height respectively.

(2) SSIM is used to evaluate the degree of similarity of geometric structure information of the predicted and the real background, the parameters are very effective for the evaluation of the performance of the image background prediction. The equation is as below:

$$\begin{cases} \mu_R = \frac{1}{NM}\sum_{x=1}^{N}\sum_{y=1}^{M} R(x,y) \\ \sigma_R = \sqrt{\frac{1}{NM-1}\sum_{x=1}^{N}\sum_{y=1}^{M}[R(x,y)-\mu_R]^2} \\ \sigma_{RF} = \frac{1}{NM-1}\sum_{x=1}^{N}\sum_{y=1}^{M}[R(x,y)-\mu_R][F(x,y)-\mu_F] \\ SSIM = \frac{(2\mu_R\mu_F+\varepsilon_1)(2\sigma_{RF}+\varepsilon_2)}{(\mu_R^2+\mu_F^2+\varepsilon_1)(\sigma_R^2+\sigma_F^2+\varepsilon_2)} \end{cases} \tag{14}$$

where *F, R, M, N* are as defined above; μ_R represents the real background pixels mean; σ_R represents the real background of standard deviation, σ_{RF} is the background covariance; ε_1 and ε_2 is a small constant to ensure denominator is not 0.

(3) GSNR is the mean value of signal to noise ratio in the sequence frames. The equation is as below:

$$\begin{cases} SNR = 10 \times \log_{10}((g_t - g_b)/\sigma) \\ GSNR = \frac{1}{N}\sum_{i=1}^{N}(SNR_{out}^{i}/SNR_{in}^{i}) \end{cases} \tag{15}$$

where g_t is the maximum value of the target area; μ_b is the mean value of the local region of the target; σ is the standard deviation of the local region of the target.

(4) The formula for AGV and LSNR are as follows:

$$\begin{cases} AGV = \frac{\sum_{(i,j)\in N(o_x,o_y)} I_{i,j}}{num} \\ LSNR = 10 \times \log_{10}\left(\frac{\mu_t - \mu_k}{\sigma_k}\right) \end{cases} \tag{16}$$

where (o_x, o_y) is the position of target center. $N(o_x, o_y)$ denotes the neighbor area of target, and its size is*num*. $(i,j) \in N(o_x, o_y)$ refer all the pixels belonging to $N(o_x, o_y)$. $I(i,j)$ is the image grayscale at pixel (i,j). μ_t is the local mean value of target. μ_k is the mean value of the local background around target, and σ_k is standard deviation of the local background. The size of the local background area is generally 3 times of the target area.

5.2. Background prediction results and analysis

5.2.1. Parameter selection analysis

The main parameters that impact the effects of improved anisotropy for background prediction are *step* and *k* value of the edge stop function. The relations between *step*, detection rate (Pd), and false alarm rate (Pf) are shown in Fig. 3(a). The relations between the edge stop function *k* value, detection rate (Pd), and false alarm rate (Pf) are shown in Fig. 3(b). From Fig. 3(a) it can be seen that when the *step* gradually increases, the detection rate also gradually increases. When the *step* reaches 4, the detection rate becomes flat. When the *step* gradually increases, the false alarm rate gradually increases. From Fig. 3(b) it can be seen that with the increase of *k* value, the detection rate gradually increases and the false alarm rate gradually decreases. When the *k* value reaches 120, the detection rate and false alarm rate both become flat. Based on the idea of compromise, the experiment sets *step* = 4 and *k* = 120.

5.2.2. Results and analysis

The image sequence obtained in the actual scene was adopted, and 6 frames with different signal-to-noise ratios were selected as the experiment images. The background was predicted by aid of the improved anisotropic background prediction method, with the parameter setting being$k = 120$, $step = 4$. Besides, background prediction was subject to comparative analysis with Top-hat [2], TDLMS [20], multi-scale morphology [21], anisotropy [8], improved top-hat [22] and gradient inverse weighted filter [23]. The background prediction effect of the images was evaluated with three indexes (MSE, SSIM and GSNR) were used to evaluate the image's background prediction effect. Wherein, Top-Hat adopted the 5×5 "square" structure, and multi-scale morphology used "square" structures from 1×1 to 7×7. Other methods were set up based on the bibliography[2,8,20–23]. Each experiment result is shown in Tables 1–4.

The smaller the MSE value was, the smaller the error was, indicating better background prediction effect; The closer the SSIM value approached to 1, the more similar the predicted background

Table 1
Signal-noise ratio of 6 frames of images.

NO.	1	2	3	4	5	6
SNR	0.37	0.49	0.82	2.21	1.91	2.14

Table 2
MSE comparison between various background prediction methods.

MSE	1	2	3	4	5	6
Top-Hat	149.88	121.64	118.76	85.43	92.52	87.89
TDLMS	126.94	116.85	112.18	63.84	69.74	65.73
MSM	75.67	73.88	70.14	54.66	61.32	56.63
ABP	16.34	15.42	13.04	11.16	12.23	11. 46
ITH	15.24	14.57	12.54	10.76	11.42	10. 88
GIWF	17.65	15.93	13.71	11.24	12.45	11.73
IABP	**9.28**	**9.71**	**8.43**	**6.59**	**7.62**	**6.83**

Table 3
SSIM comparison between various background prediction methods.

SSIM	1	2	3	4	5	6
Top-Hat	0.573	0.594	0.655	0.782	0.742	0.767
TDLMS	0.672	0.686	0.717	0.872	0.789	0.857
MSM	0.719	0.752	0.765	0.876	0.823	0.853
ABP	0.853	0.865	0.893	0.948	0.925	0.934
ITH	0.873	0.887	0.903	0.952	0.932	0.945
GIWF	0.827	0.843	0.873	0.946	0.919	0.923
IABP	**0.969**	**0.975**	**0.983**	**0.998**	**0.995**	**0.997**

Table 4
GSNR Comparison between Various Background Prediction Methods.

	Top-Hat	TDLMS	MSM	ABP	ITH	GIWF	**IABP**
GSNR	5.24	6.43	7.37	9.24	9.56	8.32	**10.85**

was to the true background; the bigger the GSNR value was, the better the background prediction's resultant difference image's target enhancement effect. By comparing and analyzing these three performance indexes, MSE, SSIM and GSNR, it could be seen that the improved anisotropy background prediction method was significantly better than other background prediction algorithms.

Meanwhile, a frame of image whose signal-noise ratio was 0.82 was selected, and the methods mentioned in this paper were used to predict its background. Fig. 4 shows the experimental results. Fig. 4(a) shows the original image, and the white rectangle was used to mark out the target position. Fig. 4(b)–(h) show background prediction result by Top-Hat, TDLMS, multi-scale morphology, anisotropy, improved anisotropy, improved Top-Hat, and gradient inverse weight Filtering, respectively. In Fig. 4(b)–(h), the Left column is background prediction results; the Middle column is the difference graph results; the Right column is the three-dimensional graph results.

The background obtained through the traditional background prediction methods (Hop-Hat, TDLMS and multi-scale morphology) is blurred and has the obvious blocking effect. The improved Top-Hat realizes filtering by fundamentally changing the structural elements of mathematical morphology, but it leads to more noise for scenes with a small signal-noise ratio after filtering. Gradient inverse weighted filter has a good detail preserving and strong clutter resisting ability, but the key parameters cannot be adjusted timely according to local clutters, the adaptability is poor and there are too many noisy points. In contrast, the improved anisotropic background prediction method can effectively eliminate most of the backgrounds in the images. Besides, this method not only retains the edge contour of the stable backgrounds and the non-stable backgrounds, but also is free from the problems of the blocking effect and target drift. After differentiation with the original image, the improved anisotropic background prediction method can effectively extract candidate targets and reduce the false alarm rate of targets.

5.3. Enhancement result and analysis

5.3.1. Parameter selection analysis

The main parameters of high-order cumulates are radius of the accumulation window r and the accumulated frame length M. To realize the effective accumulation of a moving target's energy, the cumulative window radius, the frame length, and the target moving velocity v must satisfy the following relation.

$$r \geqslant (M-1) \times v \tag{17}$$

In order to reduce the accumulation of noise energy in the image, it is desirable to select a window radius r that meets the minimum value of the Eq. (17). With formula (8), the signal-to-noise ratio of the accumulated image is calculated once a frame of image is added, and then the relationship between the cumulative frame length M and the signal-to-noise ratio is drawn according to the signal-to-noise ratio results (as shown in Fig. 5). Research finds that when the cumulative frame length $M = 4$, the signal-to-noise ratio SNR is bigger than 4, and the requirement of performance detection can be met; while with continuous increase of the cumulative frame length, the noise energy will also be accumulated, which increases the false alarm rate of images and is not favorable for improvement of the performance detected. Thus, after a comprehensive consideration, the value of M is set at 4. In the experiment, the length M of the accumulated frames is 4, considering that the movement of an infrared small target in remote imaging is slow (usually $v \leqslant 2pix/s$), the length r of the accumulated window radius is 6.

5.3.2. Results and analysis

To verify the enhancement effect of the high-order cumulates, a frame of image with the signal-noise ratio of 1.91 was selected for conducting a simulation experiment. Values of major parameters of the algorithm: accumulation window radius: $r = 6$; cumulates length: $M = 4$ frames. Fig. 6(a) is the difference graph obtained by the improved anisotropy method and the corresponding 3D view; (b) is the image obtained after applying original high-order cumulates to (a) and the corresponding 3D view; (c) is the image obtained after applying the improved high-order cumulates (IHQS) to (a) and the corresponding 3D view, with the target position being marked out. Table 5 is the comparison between AGV and LSNR of the image applying the original high-order cumulates and the improved high-order cumulates, respectively. Fig. 6 and Table 5 show that both the original high-order cumulates and the improved high-order cumulates managed to enhance the dim targets. In general, the improved high-order cumulates showed better enhancement performance.

5.4. DOG detection result and analysis

The DoG scale space algorithm mainly tackles 2 problems: (1) locate the target in the image in the case that the target size is unknown; (2) locate the targets of appointed scale and rule out other targets and similar disturbances in the case that the target size is known. The two experiments below were designed:

Experiment I:

$\sigma = 0.5$, k = 1.2,and the number of layers in the DoG space was 5. The image contains three small targets whose sizes were 2×2, 2×2 and 3×3 respectively. Fig. 7(a) shows that the DoG algorithm successfully separated the three targets, indicating that the DoG algorithm has the ability to detect targets of different scales. In other words, if not specifying the target size, the DoG algorithm would locate all small spot targets of different sizes.

Experiment II:

$\sigma = 0.5$, k = 1.2, and the number of layers in the DoG space was 10. The image contained a 5×5 spot target and several similar disturbances of different scales. Fig. 7(b) shows that, when specifying the target size, the DoG algorithm could complete the job of locating targets of specified sizes.

5.5. Detection result and analysis

Experiments are implemented in the MATLAB2012B environment on a PC with a 2.1 GHz, Pentium CPU and 2 GB memory. To

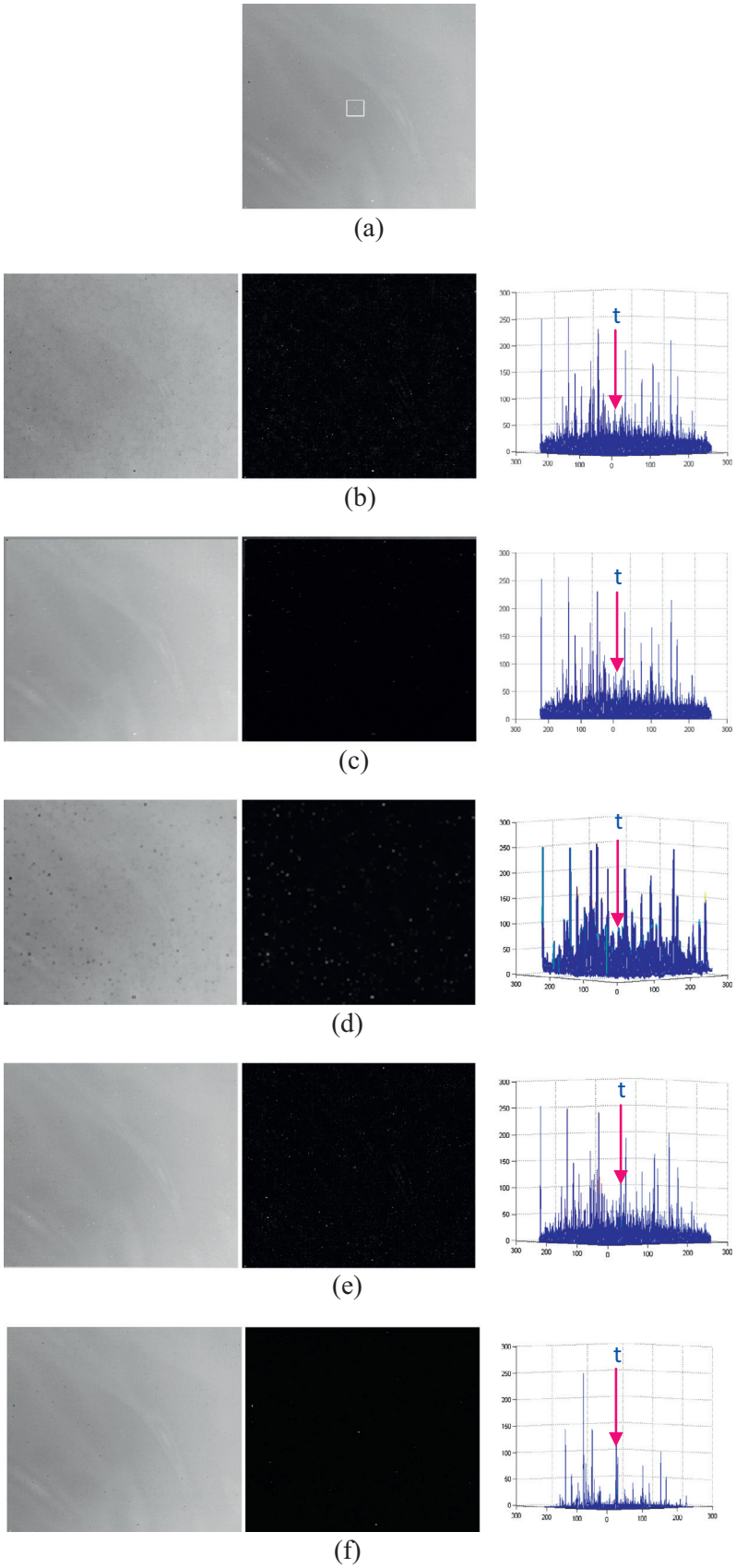
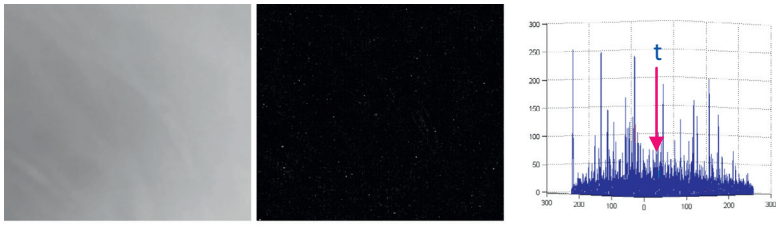


Fig. 4. Background obtained by different background prediction methods and the resultant difference graph. (a) Original image. (b)–(h) Show background prediction result by Top-Hat, TDLMS, multi-scale morphology, anisotropy, improved anisotropy, improved Top-Hat, and gradient inverse weight Filtering, respectively. In (b)–(h), the Left column is background prediction results; the Middle column is the difference graph results; the Right column is the three-dimensional graph results.



(g)

(h)

Fig. 4 (*continued*)

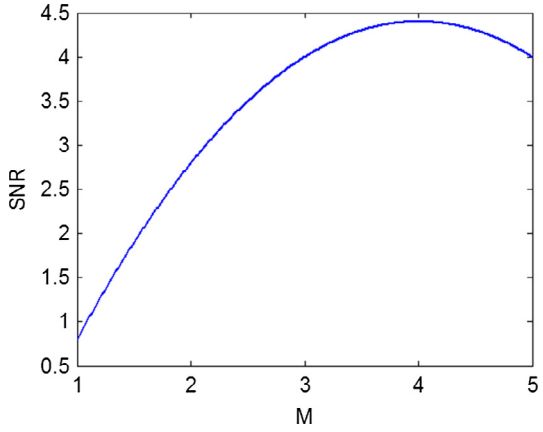


Fig. 5. Relationship between SNR gain and accumulated frame length *M*.

verify the effectiveness of the algorithm proposed in this paper, the sequence images of three groups of actual scenes A, B and C were used to detect the small targets. The targets in scene A had a strong maneuver, and they moved upwards and then downwards, followed by sudden accelerated motion in the obliquely upward direction, and finally the targets turned suddenly to move downwards. The targets in scene B did uniformly accelerated rectilinear motion in the oblique direction. The targets in scene C moved randomly in the lateral direction surrounding a certain point. The signal-to-noise ratio of the images in each scene and the image information are shown in Table 6.

Figs. 8–10 show the detection results of the three scenes A, B and C. Figs. 8(a), 9(a) and 10(a) show the original image, and the red rectangle was used to mark out the target position. Fig. 8(b)–(g), 9(b)–(g) and 10(b)–(g) show detection results by Top-Hat, TDLMS, multi-scale morphology, anisotropy, pipeline filtering and proposed method, respectively. From Figs. 8–10, it can be seen that the three methods of Top-Hat, TDLMS and multi-scale morphology lead to a high false alarm rate, and there are too many false target points in the detection results. The false alarm rate of Anisotropy is smaller than that of the above three methods. The pipeline filter algorithm can effectively reduce the false alarm rate of images, but since the fixed pipe diameter is adopted in the detection process, this algorithm may lead to overflow of the targets from the pipe and failure of detection. For example, in Fig. 8(f), the displacement of the target which turned suddenly exceeds the fixed value of the pipe diameter in the pipeline filter, leading to failure of subsequent detection; in Fig. 9(f), as the fixed pipe diameter is adopted, when the noise intensity is bigger than the target, the noise will be wrongly considered as the target; and in Fig. 10(f), when the pipe diameter keeps unchanged but the target becomes smaller, the proportion of noise in the pipe will be increased correspondingly, and partial noise will be wrongly considered as the target. The algorithm presented in this paper, on the basis of pretreatment in the earlier stage, introduces the scale space theory to adaptively adjust the pipe diameter, thereby effectively eliminating interference from image noise and accurately detecting the targets.

To further explore the performance of these algorithms, the ROC (Receiver Operating Characteristic) curve is used to evaluate the performance of these algorithms in the three scenes. These algorithms include Top-Hat [2] and TDLMS [20], multi scale morphology (MSM) [21], anisotropic [8], Pipeline filtering [11] and Proposed method. Pd represents the probability of detection rate (Eq. (18)); and Pf represents the probability of false alarm rate (Eq. (19)). NTDT is the number of true detected targets; NFDT is the number of false detected targets; NT is the quantity of true targets existing in images; NP is the quantity of targets detected in images. The ROC curves of the three scenes are shown in the following Fig. 11(a)–(c).

$$P_d = \frac{NTDT}{NT} \times 100\% \tag{18}$$

$$P_f = \frac{NFDT}{NP} \times 100\% \tag{19}$$

In Fig. 11(a), the Pd of the proposed method obtains more than 90% at the Pf = 0.015, while the other methods only obtain less than 75%. In Fig. 11(b), the Pd of the proposed method exceeds 92% at the Pf = 0.025, in contrast, the Pd of other methods only obtain less than 80%. In Fig. 11(c), the four compared methods fall behind the new method obviously. In Fig. 11(c) shows that when Pf = 0.015%, the Pd of the proposed method is 74%, in contrast, other methods have much lower Pd. This experiment shows the proposed method has better false alarms suppression than other four methods to low SNR images while keeping a stable detection rate.

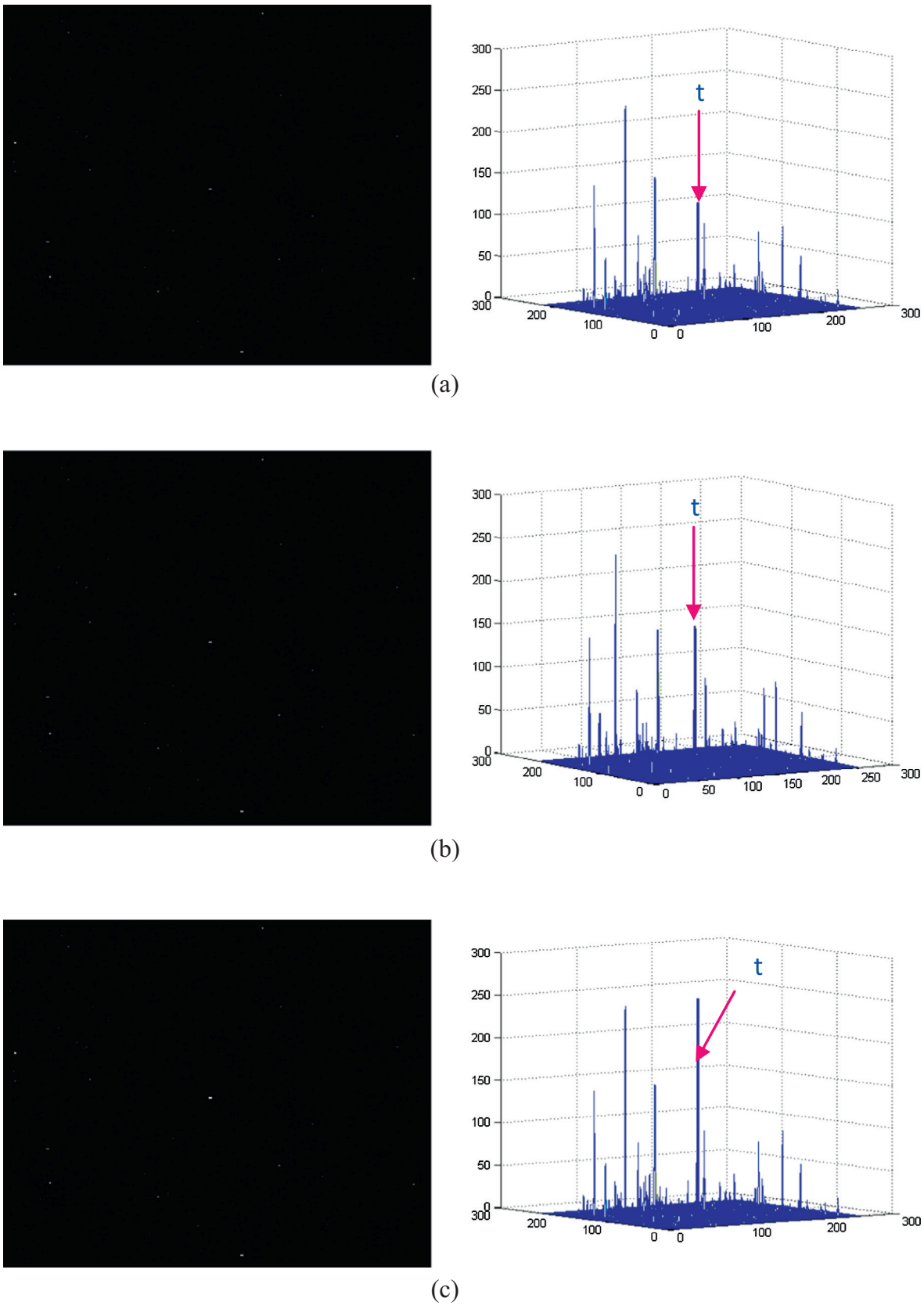


Fig. 6. Results before and after enhancement and corresponding 3D views.

It can be seen from Table 7 that the algorithms presented in this paper need pretreatment in the earlier stage, including eliminating the image backgrounds through improved anisotropy, then enhancing the images with backgrounds eliminated through improved high-order cumulates, and finally obtaining the target scale of the enhanced images through the DOG scale space theory,

Table 5
Comparison between the enhancement effects of the original high-order cumulates and the improved high-order cumulates.

Difference Graph		HQS		IHQS	
AGV	LSNR	AGV	LSNR	AGV	LSNR
95	4.57	124	8.26	**242**	**14.23**

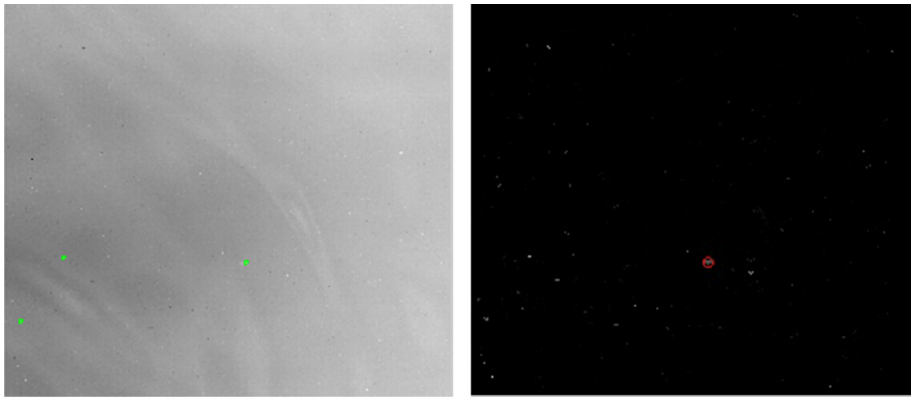


Fig. 7. DoG detection result.

Table 6
Attribute of the experiment sequence.

Scene	Average SNR	Frame length	Image size	Target size
A	2.6 dB	114	278×246	3×3
B	1.5 dB	389	320×240	5×5
C	1.2 dB	85	258×226	2×2

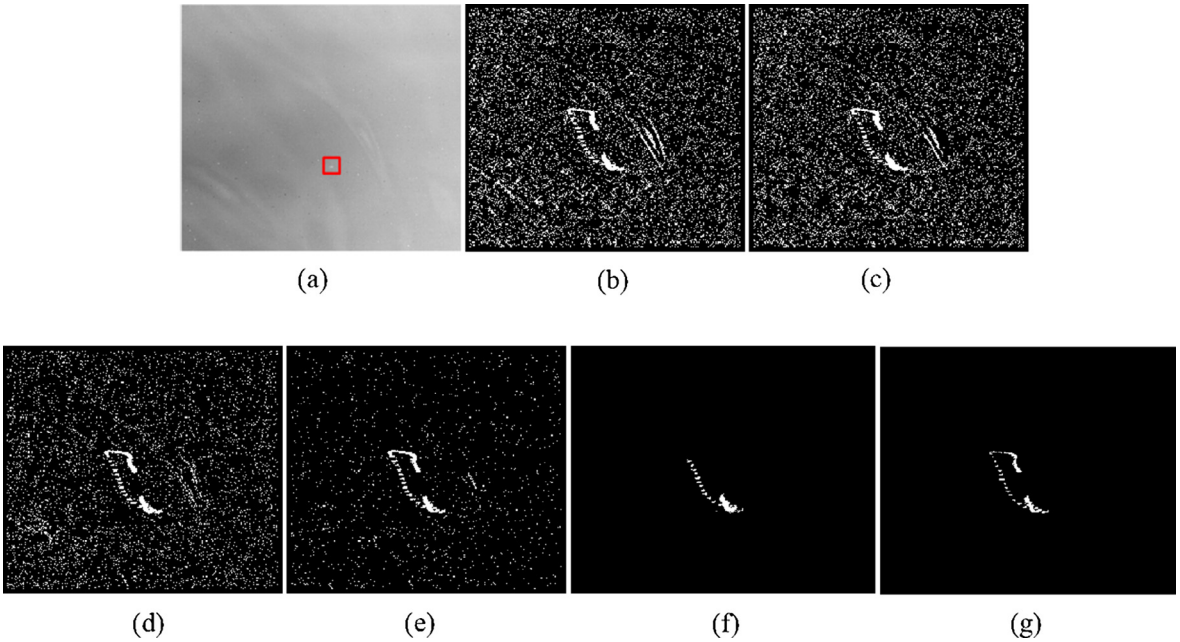


Fig. 8. Detection results of scene A. (a) Original image. (b)–(g) Show detection results by Top-Hat, TDLMS, multi-scale morphology, anisotropy, pipeline filtering and proposed method, respectively.

so as to correct the target scale in real time. All of these can increase the operation time and complexity of the algorithms. The frame frequency of the algorithms presented in this paper for the scenes A, B and C are 4 frames/s, 3 frames/s and 5 frames/s, respectively, indicating that these algorithms consume more time than other algorithms.

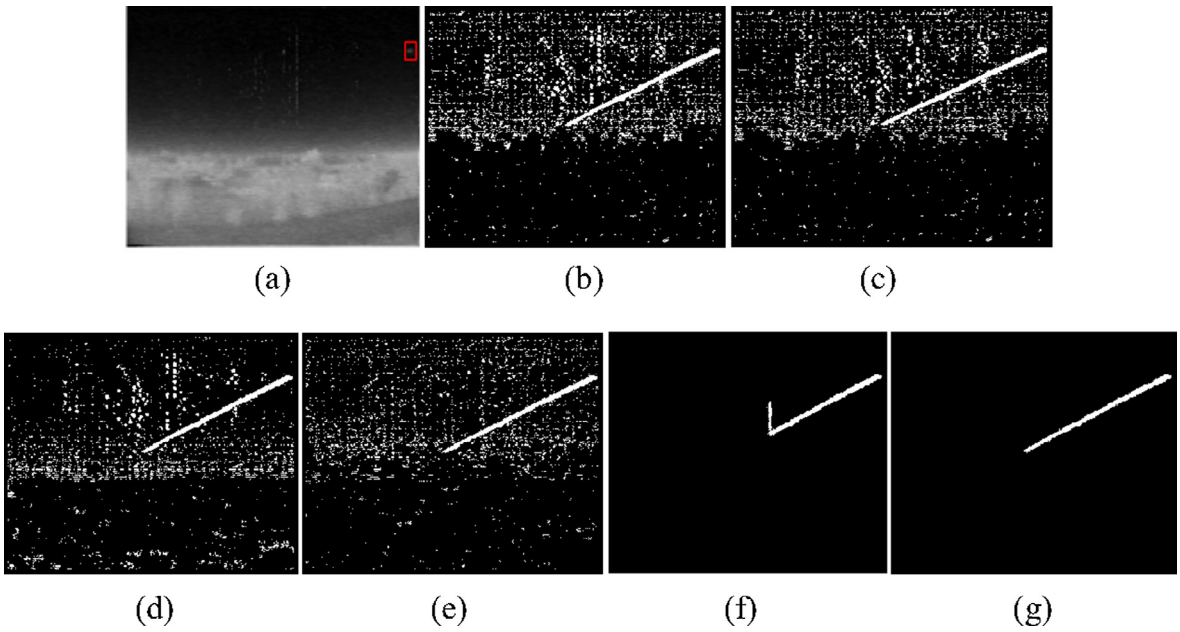


Fig. 9. Detection results of scene B. (a) Original image. (b)–(g) Show detection results by Top-Hat, TDLMS, multi-scale morphology, anisotropy, pipeline filtering and proposed method, respectively.

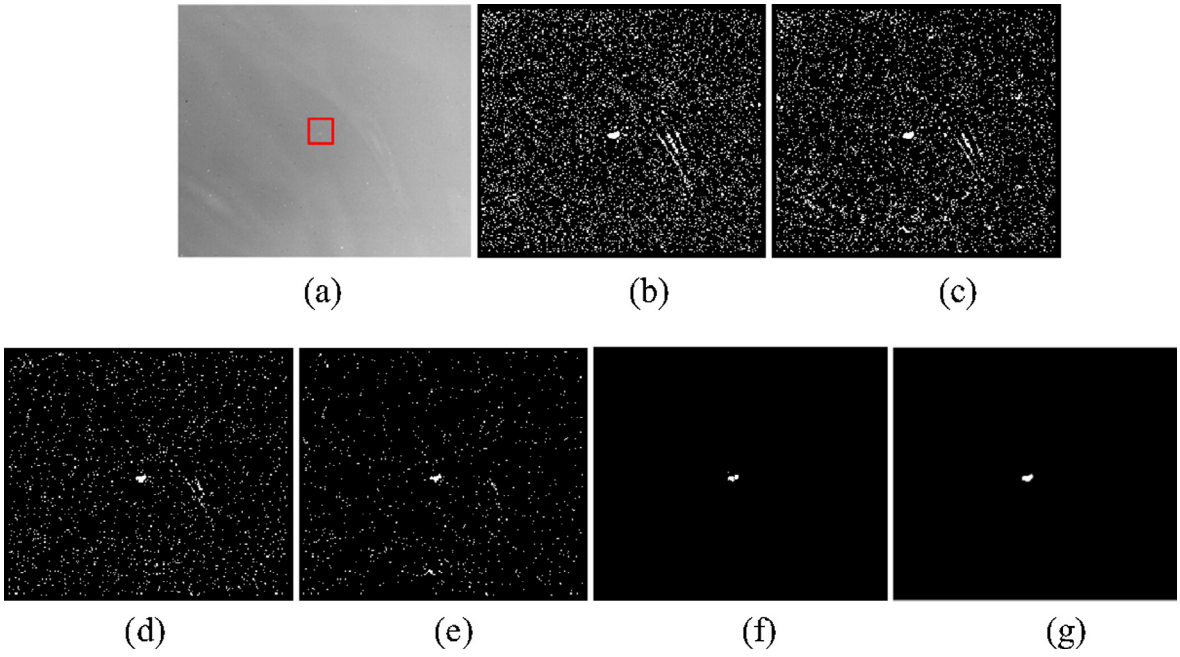


Fig. 10. Detection results of scene C. (a) Original image. (b)–(g) Show detection results by Top-Hat, TDLMS, multi-scale morphology, anisotropy, pipeline filtering and proposed method, respectively.

6. Conclusions

In order to improve the detection and recognition ability of small targets in images, this paper first uses the improved anisotropy to predict the background, then adopts the improved high-order cumulates to enhance the target, and finally, on the basis of image background suppression and target enhancement, puts forward a spatial-temporal filtering algorithm where the pipe diameter can be adaptively adjusted with the change of the target size. The simulation experiment shows that:

(1) Overall performance of the improved anisotropy is better than other background prediction methods. For different SNR images, MSE of the improved anisotropy are all less than 10, and MSE is lower when SNR is higher. SSIM of the improved anisotropy are all greater than 0.96 for different SNR images. For low SNR images, such as SNR = 0.37, SSIM also achieves good results, up to 0.969. GSNR achieves good results among the improved anisotropy, reaching 10.85.
(2) For the image without the background, the target energy is still very weak and there is noise interference, which

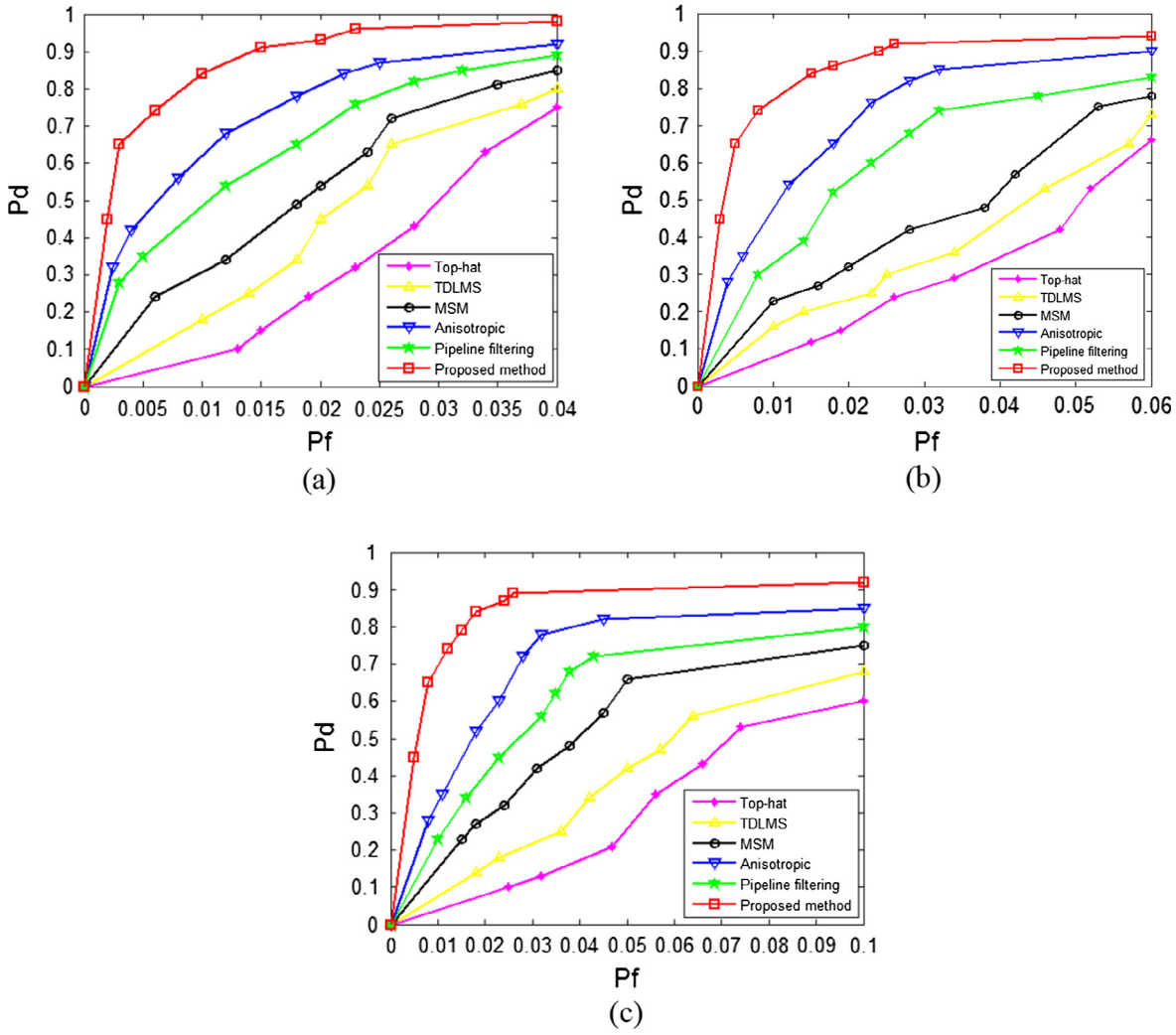


Fig. 11. The ROC curves under different Scene. The Scene of (a), (b), (c) is A, B, C respectively.

Table 7
Different algorithms of Frame frequency (FPS, frames/s).

Scene	Top-Hat	TDLMS	MSM	Anisotropy	Pipeline Filtering	Proposed method
A	18	9	6	8	5	4
B	14	10	4	6	4	3
C	20	12	9	10	7	5

is not easy to be accurately segmented and extracted. By reinforcing the high-order cumulate method and fully considering the energy accumulation in the neighborhood of the target's movement, the effect is better than the original high-order cumulates. The average grayscale and the local SNR of the target are respectively 124 and 8.26 before the improvement, while those are respectively 242 and 14.23 after the improvement, which significantly improve image grayscale of small targets and the local SNR.

(3) DOG scale space is introduced in this paper and can realize the self-adaptive modification of the pipe diameter, which effectively eliminates the noise and detects the target as well as has superiority in low SNR conditions. The method can detect a target with the lowest SNR of 0.37.

Acknowledgment

This work was partly supported by the Key Laboratory of Beam Control Chinese Academy of Sciences, by the National Natural Science Foundation of China (61571096), by Foundation of Key Laboratory of beam control Chinese Academy of Sciences (2014LBC002).

Conflict of interest

The authors declared that they have no conflicts of interest to this work.

References

[1] W.P. Yang, Z.K. Shen, Preprocessing technology for small target detection in infrared image sequences, Infrared Laser Eng. 27 (1) (1998) 23–28.
[2] R.C. Gonzalez, R.E. Woods, Ditical Image Processing, second ed., Prentice Hall, USA, 2003.
[3] C.C. Yang, J.Y. Ma, S.X. Qi, J.W. Tian, et al., Directional support value of gaussian transformation for infrared small target detection, Appl. Opt. 54 (9) (2015) 2255–2265.
[4] C.C. Yang, J.Y. Ma, S.X. Qi, J.W. Tian, et al., Multiscale facet model for infrared small target detection, Infrared Phys. Technol. 67 (2014) 202–209.
[5] H. Fan, C. Wen, Two-dimensional adaptive filtering based on projection algorithm, IEEE Trans. Signal Process. 52 (3) (2004) 832–838.

[6] P. Wang, J.W. Tian, C.Q. Gao, Infrared small target detection using directional high pass filters based on LS-SVM, Electron. Lett. 45 (3) (2009) 156–158.
[7] H. Youlal, M. Janati, M. Najim, Two-dimensional joint process lattices for adaptive restoration of images, IEEE Trans. Image Process. 1 (3) (1992) 366–378.
[8] H.S. Sang, X.B. Shen, C.Y. Chen, Architecture of a configurable 2-D adaptive filter used of small object detection and digital image processing, Opt. Eng. 43 (8) (2003) 2182–2198.
[9] H.S. Ne, Z.K. Shen, An infrared background suppression method based on Winer filtering, J. Natl. Univ. Defense Technol. 25 (3) (2003) 51–54.
[10] Y.H. Wang, W.N. Liu, Dim target enhancement algorithm for low-contrast image based on anisotropic diffusion, Opto-Electron. Eng. 35 (6) (2008) 15–19.
[11] X. Liu, Q.G. Liang, Infrared dim small target detection based on mobile weighted pipeline filter, Comput. Eng. Appl. 47 (30) (2011) 198–201.
[12] W.K. Dong, J.Q. Zhang, D.L. Liu, Q.X. Wang, Pipeline filtering algorithm based on motion direction estimation, Photon J. 42 (4) (2014) 471–474.
[13] X.B. Dong, X.S. Huang, A novel infrared small moving target detection method based on tracking interest points under complicated background, Infrared Phys. Technol. 65 (11) (2014) 36–42.
[14] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Trans. Pattern Anal. Mach. Intel. 12 (7) (1990) 629–639.
[15] B. Wu, H.B. Ji, P. Li, New method for moving dim target detection based on third order cumulate in infrared image, J. Infrared Millimeter Wave 25 (5) (2006) 364–367 (in Chinese).
[16] Q. Zhang, J.J. Cai, Small dim infrared targets segmentation method based on local maximum value partitioning, Infrared Technol. 33 (1) (2011) 124–130.
[17] Y.B. Tong, Q.S. Zhang, Y.P. Qi, Image quality assessing by combining PSNR with SSIM, J. Image Graphics 11 (12) (2006) 1758–1763 (in Chinese).
[18] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (2004) 600–612.
[19] Q. Cao, D.Y. Bi, Characteristic selecting filtering in infrared small target detection, Acta Optica Sinica 29 (9) (2009) 2048–2412 (in Chinese).
[20] Y. Cao, J. Yang, R.M. Liu, Infrared small target detection based on neighborhood analysis TDLMS filter, Acta Optica Sinica 28 (3) (2009) 210–235 (in Chinese).
[21] W.W. Ma, Y.Q. Zhao, G.H. Zhang, et al., Infrared dim target detection based on multi-structural element morphological filter combined with adaptive threshold segmentation, Acta Photonica Sinica 40 (4) (2011) 124–129.
[22] X.Z. Bai, F.G. Zhou, T. Jin, Enhancement of dim small target through modified top-hat transformation under the condition of heavy clutter, Signal Process. 90 (2010) 1643–1654.
[23] Z.Z. Li, N.L. Dong, G. Jin, Detection of dim and small targets based on adaptive filtering in strong fluctuation background, J. Instrum. 36 (7) (2004) 124–130 (in Chinese).