





# Dim and Small Target Detection Based on Spatio-Temporal Filtering and High-Order Energy Estimation

Fan Xiangsuo , Qin Wenlin , Li Juliu , Huang Qingnan , and Zhang Fan

**Abstract**—To improve the detection ability of infrared dim and small targets in complex cloud backgrounds, we present a novel method based on motion energy estimation. First, according to the characteristic of the gradient between the target and the background, we propose a new diffusion function and combine with the gradient difference to complete the background suppression. Then an improved high-order correlation algorithm is proposed to enhance the target signal through the motion correlation properties of the target, which build the gradient difference discrimination model between the target and four adjacent directions to establish the region of interest model, then realizes the high-order energy enhancement of the region of interest and further eliminate the clutters at the same time. Finally, the motion model of the target in the inter-frame spatial-temporal domain is constructed to estimate the motion direction of the target, then realizes the target energy accumulation along the motion direction and obtains the maximum energy of the target to detect the real target. In the experiment, we select six scenes for testing, the experimental results show that the proposed method has an excellent detection performance compared with six traditional detection algorithms in infrared dim and small target detection.

**Index Terms**—Background estimation, dim and small target, high-order energy correlation, kernel diffusion filtering, motion estimation.

## I. INTRODUCTION

THE detection of the dim and small target in complex cloud backgrounds is crucial to enhance the long capture capability of optoelectronic detection systems, and a low detection accuracy will greatly limit the detection performance of the system. Due to the long distance between the infrared imaging

equipment and target, the target usually is a bright spot with no obvious shape and relatively weak energy on the image [1]. Interference from the complex clouds or sea background clutters can easily cause the target signal to be drowned in the background clutters [2], therefore it makes the detection of weak IR targets tricky. How to better identify and extract real targets from IR images has been a hot research topic among experts and scholars for many years, and many infrared small target detection algorithms were proposed by the researchers at home and abroad in recent decades. Generally, the mainstream detection algorithms can be divided into three categories: background-modeling-based detection algorithms, visual-saliency-based detection algorithms, and deep-learning-based detection algorithms.

The background modeling-based method is to first utilize the image filtering algorithms to estimate the background of the input image, then subtracting the background estimation result from the original image to obtain the difference map, finally, the detection of weak targets can be completed by threshold segmentation of the difference map. The main filtering methods include median filtering [3], Top-hat filtering [4], two-dimensional least mean square filtering [5], bilateral filtering [6], [7], anisotropic filtering [8], [9], and so on. For example, Huang proposed a approach based on density peaks searching and maximum-gray region growing, this method achieved good results [10]. The Top-hat transform is widely used in the field of weak target detection because of its execution efficiency and filtering effect. Such as Bai [4] improved the traditional Top-hat filtering algorithm by using two different scales of structural elements to construct a ring-like structural element, compared with a single structural element, the ring-like structure element enjoys a higher adaptive capability and better background prediction effect, but it is susceptible to high-frequency interference and leaves numerous false alarms. Bae proposed an adaptive step size TDLMS algorithm for the shortcomings of the traditional TDLMS algorithm [5], the step size was adjusted according to different regions, and the background prediction effect of the algorithm was significantly improved, but the algorithm was limited in the scenes with low SNR. Bilateral filtering can highlight the target signal while preserving the background edge information of the image well, such as Qin [6] introduced the local statistical characteristics of the image to suppress the background based on traditional bilateral filtering, and Zeng [7] added a filter template to the traditional bilateral filter to estimate

Manuscript received 4 December 2022; revised 30 January 2023; accepted 3 February 2023. Date of current version 24 February 2023. This work was supported by the National Natural Science Foundation of China under Grants 62261004 and 62001129. (Corresponding author: Huang Qingnan.)

Fan Xiangsuo is with the School of Automation, Guangxi University of Science and Technology, Liuzhou 545006, China, and also with the Guangxi Collaborative Innovation Centre for Earthmoving Machinery, Guangxi University of Science and Technology, Liuzhou 545006, China (e-mail: 100002085@gxust.edu.cn).

Qin Wenlin, Li Juliu, and Huang Qingnan are with the School of Automation, Guangxi University of Science and Technology, Liuzhou 545006, China (e-mail: 221068323@gxust.edu.cn; ljiy668897@163.com; huangqingnan@gxust.edu.cn).

Zhang Fan is with the School of Electronic Engineering, Guangxi University of Science and Technology, Liuzhou 545006, China (e-mail: bitfan@gxust.edu.cn).

Digital Object Identifier 10.1109/JPHOT.2023.3242991

the background, both approaches obtained a good prediction performance, which can detect the target signal effectively, however, the bilateral filtering algorithm resides more clutter interference in the complex background. Because of the better directional gradient properties, anisotropy filtering can be used for background suppression of the image, but the traditional anisotropy algorithm has large limitations [11]. Zhang [12] improved the traditional anisotropy algorithm by analyzing the characteristics of infrared images, and used a small diffusion coefficient to suppress the background, a larger diffusion coefficient for the target region, which can effectively distinguish the target region and background region, but for the case that the gradient between target and background is close, the algorithm has a poor suppression effect; to solve this problem, Ling [13] proposed the S-type diffusion function and used the nuclear anisotropic diffusion model to suppress the background, a better background suppress effect is achieved, but the energy of the target is seriously weakened; Li [14] improved the diffusion function, but there is more edge residuals in the difference map, that is not conducive to further detection. The above algorithms can realize good background prediction in stable backgrounds, however, the background of the images in real life are complex, and the grayscale changes are drastic. So it is difficult for a single algorithm to get an ideal result, and often requires several algorithms to complete the target detection.

Visual saliency-based detection methods guide a new direction for weak target detection. The contrast of local regions in infrared images is a measure of the difference among the backgrounds and targets. Therefore, some researchers have performed small target detection based on contrast difference. For example, Guan proposed a detection method which combines the Gaussian scale-space and enhanced local contrast measure to enhance the small target and suppress background [15], the method is robust for different clutters, Chen [16] proposed the local contrast measure (LCM) algorithm, but the traditional LCM algorithm detects weak targets with the assumption that the target grayscale is higher than the background grayscale, which has a limited ability to suppress background. Wei [17] proposed the Multiscale patch-based contrast measure(MPCM) algorithm, which has a good detection performance in high-contrast scenes, but it easily smoothes out the target signal in complex backgrounds. Han [18] proposed the TLLCM (Tri-layer LCM) algorithm, which calculates the local contrast before the target enhancement, so that it can effectively discriminate the target from the background, but this algorithm can not detect the target when the target gray level is relatively low, then leading to detection failure. Mu [19] improved the TLLCM algorithm and proposed Tri-layer template local difference measure(TLLDM), which has an excellent target enhancement performance. Moradi [20] proposed absolute directional mean difference(ADMD) method, in general, the gray value of the target region is higher than background region, then the mean difference between the target region and the background region is larger, while the non-target region does not have this property. ADMD can effectively enhance the target region by constructing the mean difference between the target region and the its eight neighborhood directions, selecting the minimum value to

remove the background noise while enhancing the target region. It is different from the previous works which only took one pixel into account, the ADMD considered the local region of target, it can effectively avoid the false detection. This algorithm has a high execution efficiency, and it is suitable for practical applications because it can achieve considerable detection results in most scenes. The visual-saliency-based detection method generates different degrees of false alarms, and the application of such algorithms to practical engineering applications has certain limitations, which require further optimization and improvement of the algorithm.

In addition to the two types of the mainstream detection methods mentioned above, there are many deep-learning-based detection methods for infrared small targets developed in recent years [21], [22], [23]. For example, Cai [24] proposed an algorithm based on a two-channel feature-enhanced integrated attention network for the small target detection, Ma [25] presented a weak target detection method which based on the SSD networks, and Cai [26] proposed a detection approaches based on YOLO-FCSP network, deep-learning-based detection methods improve the detection accuracy to some extent, these methods require extensive training samples to learn the features of small targets. However, they face problems such as lack of datasets and complex background interference. When image contains the dynamic change background, the parameters of the trained model will change correspondingly and the detection accuracy will be degraded, it is hard for the deep learning model training to fully learn the characteristics of infrared small targets. Moreover, the training of samples also takes some time and has high hardware requirements.

Although most of clutters are removed by the background suppression, the energy of the target is weak and a small amount of clutters residual in the difference map. To enhance the energy of the target, the clutters can be suppressed again and the SNR of the image can be further improved. Researchers have proposed energy enhancement algorithms, for example, the higher-order accumulation algorithm can effectively enhance the target signal based on background suppression, but the traditional high-order correlation algorithm only considers the space domain information and does not fully consider the time domain information, and the enhancement effect achieved is less satisfactory. Wang [27] improved the traditional high-order correlation algorithm by using the method of interframe correlation for energy enhancement, and the algorithm can improve computing efficiency effectively. Tang [28] proposed an algorithm based on the combination of morphological filtering and high-order statistics, using the double-structured element to suppress background clutters, and combining the third-order statistics as a statistical criterion for the image segmentation, to complete the detection of infrared small targets. However, these algorithms have significantly lower energy enhancement effect when facing complex backgrounds.

In summary, the small target detection methods mentioned above achieve better detection results in the scenes with smooth background, however, the traditional background prediction methods cannot simply rely on improving a specific background suppression algorithm to achieve detection of weak targets in the

complex and variable scenes; the background prediction only retains the target and does not enhance the target. Therefore, based on the analysis of the traditional background prediction methods facing the lack of complex background suppression, this paper presented a new method and its main contributions are as follow:

- 1) Combined the gradient property between the target and the background with the diffusion function can effectively suppress the image background. This paper improves the anisotropic filtering algorithm and proposes a new diffusion function, then we utilize the proposed function combine with the gradient difference between pixels to suppress the background, and the proposed diffusion function has a stronger background suppression ability.
- 2) After the background suppression, most of background clutters is removed, but the energy of the target is also weakened. We add a gradient difference discrimination model into the traditional high-order correlation energy enhancement model. This model can further remove the clutters and enhance the candidate target region.
- 3) Based on the above image pre-processing, we construct the motion model of the target in the spatial domain to perform the energy accumulation along the motion direction, then finding a energy maximum value of the target to realize the multi-frame correlation detection. The experimental results show that our method can effectively suppress the image background, improve the signal-to-noise ratio of the image, and achieve the detection of weak infrared targets.

## II. NEW KERNEL DIFFUSION FUNCTION

In general, there is continuity between the background of the IR images, while there is a significant difference between the target and the background in local region, and this difference between the target and the background can be used to distinguish the target from the background. Traditional background suppression algorithms mainly utilize the filtering algorithms to complete background modeling, such as Difference of Gaussian filter(DoG) [29], which essentially is isotropic during the filtering process, because the directionality is not considered in the filtering process, all the pixels in the filtering window are given the same weight, while the image details (such as edges) have obvious directionality, if the same weight is applied to the background details of the image, it may cause the loss of details or blur the edges, even the position shifts. To overcome the shortcoming that the same direction weight values will blur the image edges in the isotropic algorithm, Perona and Malik proposed an anisotropic differentiation method, which controls the diffusion intensity by a diffusion function that varies with the gradient of the image. Because the gradients are different in different region of the image, the weight values of the corresponding direction will vary as the gradient changes. The essence of the anisotropy algorithm is the diffusion function combined with the gradient difference to smooth the image background, and then achieve the purpose of background suppression. The key is how to smooth the image by the diffusion function after calculating the gradient, to preserve the target and remove the background clutter at the

same time, so the selection of the diffusion equation is also important. The diffusion function given in the literature [11] has a larger coefficient in background region with a small gradient, and the coefficient is smaller in target region with a large gradient. However, the ability of background suppression of the function is insufficient in the complex infrared images. So, Zhang [12] proposed another diffusion function, the function has a excellent background suppression performance in smooth area of images, due to the difference of gradient between the target and edge contours is not clear in the scenes which have more cloud edge contours, there is more edge clutters retained in the difference map. Then Li [14] proposed a new function which has a stronger background suppression ability, but the target is easily smoothed in the scenes which have more edge contours and cause the detection failure. What the shortcoming of above diffusion function is that it is hard to distinguish the target from the background, because the gray value between the target and edge is closely in the scenes with more edge contours. Therefore, based on the above analysis, we proposed a new function which can effectively discriminate the target from the background, the expression of the function is as follow:

$$C(\nabla I) = \frac{1}{1 + 1/[\exp((\nabla I)^2/k) - 1]} \quad (1)$$

where,  $C$  is the diffusion function,  $\nabla I$  is the gradient between the pixels, and  $k$  is a constant. To calculate the gradient between pixels, the expression can be defined as follow:

$$\begin{cases} \nabla_N = I(i, j) - I(i - s, j) \\ \nabla_S = I(i, j) - I(i + s, j) \\ \nabla_W = I(i, j) - I(i, j - s) \\ \nabla_E = I(i, j) - I(i, j + s) \end{cases} \quad (2)$$

where,  $I$  is the input image,  $I(i, j)$  is one pixel value of the image,  $s$  is the step size. We select  $I(i, j)$  as the reference pixel, and take its four-neighborhood, the gradient between the reference pixel and the upper direction is denoted as  $\nabla_N$ , the gradient between the reference pixel and the lower direction is denoted as  $\nabla_S$ , the gradient between the reference pixel and the left direction is denoted as  $\nabla_W$ , the gradient between the reference pixel and the right direction is denoted as  $\nabla_E$ . Substituting gradient into the diffusion function proposed in this paper to calculate the coefficient of corresponding direction:

$$\begin{cases} C_N = 1/[1 + 1/[\exp((\nabla_N)^2/k) - 1]] \\ C_S = 1/[1 + 1/[\exp((\nabla_S)^2/k) - 1]] \\ C_W = 1/[1 + 1/[\exp((\nabla_W)^2/k) - 1]] \\ C_E = 1/[1 + 1/[\exp((\nabla_E)^2/k) - 1]] \end{cases} \quad (3)$$

where,  $C_N, C_S, C_W, C_E$  represents the diffusion coefficient of direction of upper, lower, left and right, respectively.  $k$  is a constant and  $k = 200$  is taken in this paper. We combine the diffusion coefficient with the gradient of corresponding direction for weighted summation, and obtain the result of anisotropic background suppression:

$$D_I = \frac{1}{4} \times (C_N * |\nabla_N| + C_S * |\nabla_S| + C_W * |\nabla_W| + C_E * |\nabla_E|) \quad (4)$$

In above equation,  $|\nabla_N|$ ,  $|\nabla_S|$ ,  $|\nabla_W|$ ,  $|\nabla_E|$  denotes the gradient between the center pixel and the four directions of upper, lower, left and right in the neighborhood, respectively.  $D_I$  represents the corresponding difference map of original image.

### III. IMPROVED HIGH-ORDER ENERGY CORRELATION ENHANCEMENT AND MULTI-FRAME CORRELATION DETECTION

#### A. Improved High-Order Energy Correlation Enhancement

After the process of background suppression, most of the clutters interference is suppressed, and the energy of the target is also weakened, which is still in a relatively weak state. Since in addition to the target, there is a small amount of noise interference on the differential image. If the energy accumulation method of frame accumulation is used directly, the target energy can be increased, while the clutters energy will be enhanced, too. It might cause the enhancement magnitude of the clutters exceed the enhancement magnitude of the targets, which is not conducive to the subsequent detection. In order to effectively increase the target signal in the region of interest, an improved high-order energy correlation enhancement algorithm presented in this paper, which first establishes the discriminative model of the saliency region, and the gradient difference between the target region and its local region in four directions is used to establish the saliency region, if the current image pixel is located in a region with gradient difference in three or four directions greater than the threshold, the area can be considered to contain the target, and the target would be retained as a candidate area of interest, and then utilizes high-order energy correlation to enhance the energy for the saliency region. The difference between the algorithm in this paper and the previous work is that this paper constructs the mean difference between the target region and its four neighborhood directions, and if there are three or four directions with the mean difference larger than a preset threshold, the region is considered as a candidate target region, and a higher-order correlation algorithm is used to perform the enhancement of the target region.

The high-order correlation mainly uses the correlation property of the target energy between frames for energy enhancement, and the corresponding expression is as follows [30].

$$Y(x, y, t_n) = g \left[ \sum_{i=-|v|}^{i=|v|} \sum_{j=-|v|}^{j=|v|} f(x, y, t_n) f(x+i, y+j, t_{n+1}) \right] \quad (5)$$

where  $v$  is the size of the target neighborhood,  $i, j$  represents the pixel neighborhood coordinates,  $Y(x, y, t_n)$  denotes the correlation of the target point at current moment  $n$  and the next moment  $n+1$ ,  $f(x, y, t_n)$  indicates the pixel value of the target at  $n$ ,  $f(x+i, y+j, t_{n+1})$  represents the neighborhood of the target at  $n+1$ ,  $(x, y)$  denotes the coordinates of the center pixel. The first-order correlation can be obtained by accumulating the product of the frame and the next frame. According to such an idea, the target energy can be effectively enhanced by using the correlation of multi-frame images. The specific formula can be

expressed as [30]:

$$\begin{cases} Y^{(1)}(x, y, t_n) = g \left[ \sum_{i=-|v|}^{i=|v|} \sum_{j=-|v|}^{j=|v|} Y^{(0)}(x, y, t_n) Y^{(0)}(x+i, \right. \\ \left. y+j, t_{n+1}) \right] \end{cases} \quad (6)$$

where  $Y^{(1)}(x, y, t_n)$  represents the result of first-order correlation in above equation.

The specific approach is to first build a central neighborhood block centered on the position of the target point at  $n$  moments, then select neighborhood blocks of the same scale in the four directions of the neighborhood of the central neighborhood block, calculate the mean value of these neighborhood blocks and the gradient between the mean value of the central neighborhood block and the mean value in the four directions, sort the gradient in the four directions, if the gradient in three or four directions is greater than a predetermined threshold, the image pixel is a candidate target, and the high-order correlation process is carried out to multiply the neighborhood of the current target point in the next frame with the image pixel value of the current image target point, so that the time domain information of the current frame is retained in the next frame, then the target energy is enhanced and the grayscale value is normalized at the same time; if the gradient of the image pixel does not meet the threshold judgment condition, the image pixel can be considered as noise. Specifically, see (7)–(9).

$$\begin{cases} mid = \sum_{m=-L/2}^{L/2} \sum_{n=-L/2}^{L/2} f(i+m, j+n) \\ up = \sum_{m=-(3L+1)/2}^{-(L+1)/2} \sum_{n=-L/2}^{L/2} f(i+m, j+n) \\ down = \sum_{m=(L+1)/2}^{L/2} \sum_{n=-(L/2)}^{-(L+1)/2} f(i+m, j+n) \\ left = \sum_{m=-(L/2)}^{L/2} \sum_{n=-(3L+1)/2}^{-(L+1)/2} f(i+m, j+n) \\ right = \sum_{m=-(L/2)}^{L/2} \sum_{n=(L+1)/2}^{(3L+1)/2} f(i+m, j+n) \end{cases} \quad (7)$$

In above equation,  $f$  is the difference image.  $i, j$  represents the coordinate of the pixel,  $L$  is the space neighborhood radius of the target.  $mid$  denotes the neighborhood of the target, and  $up, down, left, right$  represents the neighborhood of target neighborhood in its four adjacent directions, respectively.

$$\begin{cases} d_1 = \frac{1}{L*L} \times |mid - up| \\ d_2 = \frac{1}{L*L} \times |mid - down| \\ d_3 = \frac{1}{L*L} \times |mid - left| \\ d_4 = \frac{1}{L*L} \times |mid - right| \\ d = [d_1 \ d_2 \ d_3 \ d_4] \\ [d_{1st\_min}, d_{2nd\_min}] = sort(d, ascend) \end{cases} \quad (8)$$

In above equation,  $d_1, d_2, d_3, d_4$  is the gradient between the central neighborhood and its four adjacent directions.  $d$  records

0 0 0 0 0	0 0 1 0 0	0 0 0 0 1	1 0 0 0 0
0 0 0 0 0	0 0 1 0 0	0 0 0 1 0	0 1 0 0 0
1 1 1 1 1	0 0 1 0 0	0 0 1 0 0	0 0 1 0 0
0 0 0 0 0	0 0 1 0 0	0 1 0 0 0	0 0 0 1 0
0 0 0 0 0	0 0 1 0 0	1 0 0 0 0	0 0 0 0 1
0 0 0 0 0	0 0 0 0 1	0 0 1 0 0	0 0 1 0 0
0 0 0 1 1	0 0 1 1 0	0 0 1 0 0	0 0 1 0 0
0 0 1 0 0	0 0 1 0 0	0 0 1 0 0	0 0 1 0 0
0 1 0 0 0	0 1 0 0 0	0 1 0 0 0	0 0 0 1 0
1 0 0 0 0	0 0 0 0 0	1 0 0 0 0	0 0 0 0 1
0 1 0 0 0	0 0 0 1 0	0 1 0 0 0	0 0 1 0 0
0 0 1 0 0	0 0 1 0 0	0 0 1 0 0	0 0 0 1 0
0 0 1 0 0	0 0 1 0 0	0 0 1 1 0	0 0 0 0 1
0 0 1 0 0	0 1 0 0 0	0 0 0 0 1	0 0 0 0 1

Fig. 1. The 12 motion models used in this paper.

the value of  $d_1, d_2, d_3, d_4$ , respectively.

$$\begin{cases}
 target1 = f(i, j, t_n) \\
 target2 = f1(i + m, j + w, t_{n+1}) \\
 if \quad d_{1st\_min} > T | d_{2nd\_min} > T \\
 Result(i, j, t_n) = \sum_{m=-r}^m \sum_{w=-r}^w target1 \times target2 \\
 else \\
 Result(i, j, t_n) = 0
 \end{cases} \quad (9)$$

In above equation,  $target1$  represents the location of the target at time  $n$ ,  $target2$  represents the location of the target at time  $n + 1$ .  $r$  is the search neighborhood radius of the target.  $d_{1st\_min}$  represents the minimum value after sorting from small to large in the four directions of  $d$ ,  $d_{2nd\_min}$  represents the second smallest value after sorting from small to large in the four directions of  $d$ .  $T$  denotes the threshold of the gradient, and the selection of threshold is crucial. If the threshold is too large, the target will be filtered out, while if the threshold is too small, the noise interference can not be effectively suppressed and lead to detection failure.  $Result(i, j, t_n)$  is the result of high-order correlation.

### B. Multi-Frame Correlation Detection Based on Motion Estimation

The target image obtained after the high-order correlation energy correlation enhancement still remains a small amount of noise, in order to distinguish the target from noise, the motion continuity of the target in the time domain and the randomness of the noise appearance are mainly used to effectively distinguish the target. Considering that the target's moving speed  $v$  is less than 2 pixel/s in the long-distance imaging process, and considering the out-of-focus phenomenon caused by the influence of atmospheric aerosols in the imaging process, 12 motion models are constructed and shown in Fig. 1, which make full use of the motion characteristics of the target in the spatial and temporal domains, and finally extract the real target by finding the energy maxima of the target motion in the 12 motion models. Where

the continuity of target motion between frames is taken into account in the first four of these models, and the last eight motion models consider the spatial motion characteristics caused by the irregularity of target scale. The motion estimation algorithm proposed in the paper is based on the following assumption: the motion direction of the target is known, and the corresponding movement is done according to different motion models, and if the motion model is consistent with the motion direction of the target, the energy of the target will be further enhanced during the motion process from the current moment to the next moment. Based on the above hypothetics, the idea of the multi-frame detection algorithm based on motion energy estimation proposed in the paper is as follows: based on the high-order energy correlation enhancement in Section III-A, the result of energy enhancement is multiplied by different motion models, and the detection of the real target is achieved by finding the energy maxima. The specific formulation is as follows.

$$\begin{cases}
 f_k(x, y, t_i) = \sum_{m=-2}^{m=2} \sum_{n=-2}^{n=2} T_k(x, y) \times F_0(x + m, y + n, t_i) \\
 f_k(x, y, t_{i+1}) = \sum_{m=-2}^{m=2} \sum_{n=-2}^{n=2} T_k(x, y) \times F_0(x + m, y + n, t_{i+1}) \\
 f_k(x, y, t_{i+2}) = \sum_{m=-2}^{m=2} \sum_{n=-2}^{n=2} T_k(x, y) \times F_0(x + m, y + n, t_{i+2}) \\
 E_{k1}(x, y, t_i) = \sum_{p=-r}^{p=r} \sum_{q=-r}^{q=r} f_k(x, y, t_i) \times f_k(x + p, y + q, t_{i+1}) \\
 E_{k2}(x, y, t_i) = \sum_{p=-r}^{p=r} \sum_{q=-r}^{q=r} E_{k1}(x, y, t_i) \times f_k(x + p, y + q, t_{i+2}) \\
 E_{k3}(x, y, t_i) = E_{k2}(x, y, t_i) / (L * L) \\
 E_k = \sum_{i=1}^N E_{k3}(x, y, t_i) \\
 E'_k = \arg \max_{k=1,2,\dots,12} E_k
 \end{cases} \quad (10)$$

where,  $r$  denotes the search radius of the neighborhood.  $i = 1, 2, \dots, N$  denotes the number of images,  $k = 1, \dots, 12$  represents the number of 12 different motion models,  $T_k(x, y)$  represents 12 different motion models,  $F_0$  represents the energy-enhanced image,  $t_i$  represents the  $i$ th frame image,  $E_{k1}$  represents the result of performing high-order correlation on the first two frames,  $E_{k2}$  represents the result of performing high-order correlation on the first two frames and then high-order correlation on the third frame,  $E_{k3}$  represents the mean value of  $E_{k2}$ ,  $E_k$  represents the energy accumulation value along the estimated motion direction,  $E'_k$  represents the energy maxima of high-order correlation among the 12 motion models, and this energy maxima corresponds to the motion pattern, which can be regarded as the motion direction of the target in the adjacent frames. In the case that there is the same energy maxima in the 12 models, we take an "AND" operation for the images to solve this problem and then obtain the motion trajectory of the target.

## IV. EXPERIMENT RESULTS

To quantitatively evaluate the background prediction effect of the algorithm, SSIM (structural similarity), SNR (global signal-to-noise ratio), and BSF (background suppression factor) are

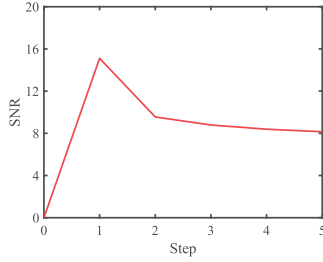


Fig. 2. The relationship between the moving step and SNR.

used as evaluation index in this paper. SSIM is used to evaluate the background prediction effect, and the larger the SSIM, the better the background prediction effect. BSF is utilized to evaluate the effectiveness of background suppression, and the higher the index of BSF, the better the background suppression effect of the proposed algorithm. The three indicators are calculated as follows [14]:

$$SSIM = \frac{(2\mu_R\mu_F + \varepsilon_1)(2\sigma_{RF} + \varepsilon_2)}{(\mu_R^2 + \mu_F^2 + \varepsilon_1)(\sigma_R^2 + \sigma_F^2 + \varepsilon_2)} \quad (11)$$

$$SNR = 10 \times \log_{10}^{(\mu_T - \mu_B)/\sigma_B} \quad (12)$$

$$BSF = \frac{\sigma_{in}}{\sigma_{out}} \quad (13)$$

In formula (11),  $\mu$  denotes the mean value,  $\mu_R$  denotes the mean value of the original image,  $\mu_F$  denotes the mean value of the predicted image.  $\sigma$  represents the standard deviation and  $\sigma_{RF}$  represents the covariance. In Formula (12),  $\mu_T$  denotes the mean value of target region,  $\mu_B$  denotes the mean value of background region, and  $\sigma_B$  represents the variance of background region. In formula (13),  $\sigma_{in}$  denotes the variance of the original image,  $\sigma_{out}$  denotes the variance the difference image, and the SNR calculated in this paper is the global SNR.

- 1) *Analysis of the selection of background suppression parameters:* Anisotropy realizes background suppression by calculating the gradient difference between the central pixel and its surrounding pixel combined with the diffusion equation. Therefore, the selection of the moving step of the image pixel is crucial. When the target is the smooth background, the grayscale difference between the target and the surrounding background is relatively stable, so the effect of the step size on the background suppression effect is negligible; while the target is in a non-smooth background (for example, there are more cloud edges in the background), the grayscale value of the background spans a large range. If the moving step is selected too large, the background image elements around the target may cross from the smooth background region with a smaller gradient to the edge region with a larger gradient, and then the edge region will be treated as the smooth background region, and more edge clutter will be left in the differential map, which leads to poor background suppression effect. For this reason, the relationship curve between global SNR and the moving step is plotted in this paper (Fig. 2), and it can be seen from the figure that when the step is 1, the

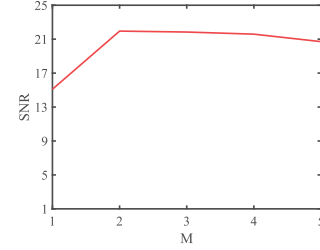


Fig. 3. The relationship between the cumulative frame number M and SNR.

SNR reaches 15.1 dB, and as the moving step increases, the SNR gradually decreases, and the lower SNR will be detrimental to the target detection, so the step=1 is selected for the experiment in this paper.

- 2) *Analysis of the selection of energy enhancement parameters:* It is found that the target energy enhancement effect is closely related to the cumulative frame number M. If the target moves slowly, the overlapping area of the target in the neighborhood between frames is larger, and a larger cumulative frame number M can achieve a better enhancement effect, while the target moves faster, the overlap area of the target in the neighborhood between frames gradually decreases. As the cumulative frame number M increases, the enhancement effect will be greatly reduced. So the relationship between the cumulative number of frames and the SNR is plotted in the paper (Fig. 3), from the figure it can be seen that when the cumulative number of frames  $M = 2$ , the SNR reaches 21.95 dB. As the cumulative frame number M increases, the target in the frame neighborhood overlap area reduces, and the SNR starts to decline, the energy enhancement effect of the algorithm decreases will be detrimental to the subsequent detection. Therefore, the cumulative frame number  $M = 2$  is chosen for the experiment in the paper.

#### A. Analysis of the Effect of Background Suppression

In order to test the effectiveness of the background suppression of the proposed algorithm, the algorithm in this paper is compared with several algorithms, the background and difference maps obtained from experiments based on different algorithms are performed to reflect the feasibility of the algorithm. These comparison algorithms include WLDM (Weighted Local Difference Measure) [31], Top-hat [4], MGDWE (Multiscale Gray Difference Weighted Entropy) [32], MPCM (Multiscale patch-based contrast measure) [17], TLLCM (Tri-layer local contrast measure) [18], NTFRA (Non-convex tensor fibered rank approximation) [33], NTLA (Non-Convex Tensor Low-Rank Approximation) [34], PSTNN (Partial sum of the tensor nuclear norm) [35], TDLMS (Two dimension least mean square) [5], ADMD (Absolute directional mean difference) [20] and TLLDM (Tri-layer template local difference measure) [19]. Six scenes were used to experiment in this paper, and the data set was obtained from the literature [36]. The relevant parameters of the six scenes are shown in Table I, and  $k = 200$ ,

TABLE I  
RELEVANT PARAMETERS OF THE SIX SCENES

	Scene1	Scene2	Scene3	Scene4	Scene5	Scene6
Size	512*640	512*640	512*640	512*640	512*640	513*641
Frames	200	150	200	200	200	100
Target size	3*3	3*3	3*3	3*3	3*3	1*1
Background	Less edge	Smooth cloud	More edge	More edge	Buildings and sky	More edge

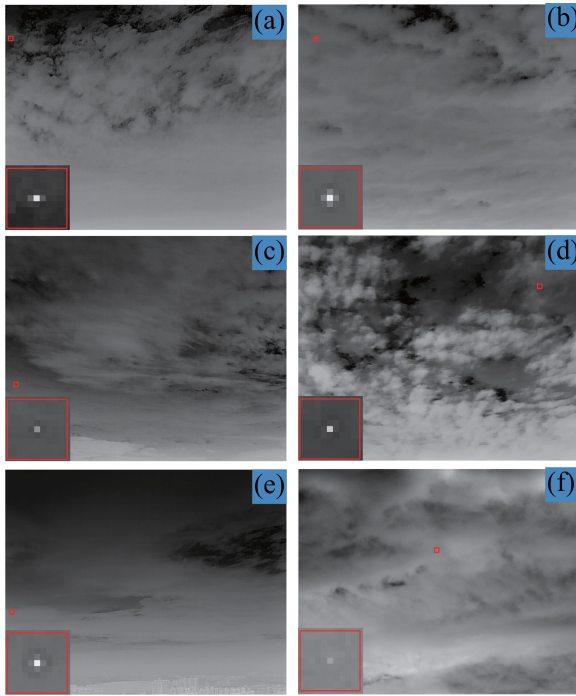


Fig. 4. Original Images.

step size  $s = 1$  are selected in the experiment. The targets in the original images have marked with a red box, see Fig. 4.

In scene 1 and scene 2, we can observe the target signal clearly in the images, and the corresponding experimental results are referred to Figs. 5 and 6. All algorithms can achieve better background suppression effect. The traditional spatial-temporal filtering methods cause energy loss due to the difference with the original image after background estimation, so the target signal can be seen to be weaker, such as Top-hat filtering and TDLMS filtering algorithms, the upper part of scene 1 has more cloud edges, resulting in more background clutter remaining in the difference map, and the intensity of clutters are even higher than the intensity of the target, while the background in scene 2 is smoother, and this phenomenon does not occur in the difference images obtained by the two algorithms; in the differential maps obtained by MPCM, MGDWE, TLLCM, and WLDM, although the targets can be detected, they also have obvious block effects at the same time, and those edge contours with higher energy will be mistaken for the targets, as can be seen in Fig. 5 that the target surrounded by clutter with higher energy, and the TLLCM algorithm has a better effect, but there is still a small amount of clutter, ADMD and TLLDM show better target enhancement, and the differential map has only

a small amount of clutter, and the target grayscale is much higher than the grayscale of the background clutter as seen in the 3D map, while in scene 2 all algorithms can effectively remove clutters; PSTNN, NTLA and NTFRA algorithms can achieve good background suppression effect in images with high signal-to-noise ratio, especially for the background with strong edge contours, scene 1 and scene 2 are both high signal-to-noise ratio images. Therefore, the three algorithms can remove most of the background clutter, and only a small number of clutters remain. In the evaluation index data table, the SNR of most of algorithms is higher than the algorithm in this paper, and the SSIM of TLLDM is higher than the algorithm in this paper. The improved background suppression algorithm in this paper can effectively remove most of the background clutter and highlight the target signal. The comparison data of evaluation indexes in Tables II and III show that the SSIM of the algorithm in this paper is lower than the TLLDM in scene 1, the SNR is higher than most of the comparison algorithms, and the BSF is highest, and the BSF and SSIM in scene 2 is highest, and the SNR is higher than most of the comparison algorithms.

The backgrounds of scene 3 and scene 4 move with the target and span a wide range. Among them, scene 3 has a weak target energy and scene 4 has more strong edge contours. The experimental results are shown in Figs. 7 and 8. Top-hat filtering and TDLMS filtering can effectively predict the image background in scene 3, and the target signal can be observed in the background suppressed image, but the suppression ability for background clutters is not sufficient, resulting in more residual clutters with higher energy, but the two algorithms can effectively suppress the background edge contour in scene 4; although the WLDM, MGDWE, MPCM, and TLLCM methods can detect the target, however, more clutters remained in the differential map due to the inconspicuous contrast between the target and the background in scene 3 and the interference of strong edge contours in scene 4; NTLA and NTFRA methods have more residual edge contours while suppressing the background because of the low energy of the target in scene 3, but in scene 4 they can effectively highlight the target signal, the NTLA method also resides a large area of background contours; PSTNN maintains excellent background suppression ability in these two scenes, with only small amount of clutters; the target enhancement effect of ADMD and TLLDM is reduced because the target scale is too small in scene 3, the target grayscale is lower than that of the background clutter as seen in the 3D map. The effect of PSTNN algorithm on background suppression is better than the ADMD and TLLDM algorithm. The target enhancement effect of ADMD algorithm in scene 4 is not obvious, and the clutter interference in the 3D map is serious, while the

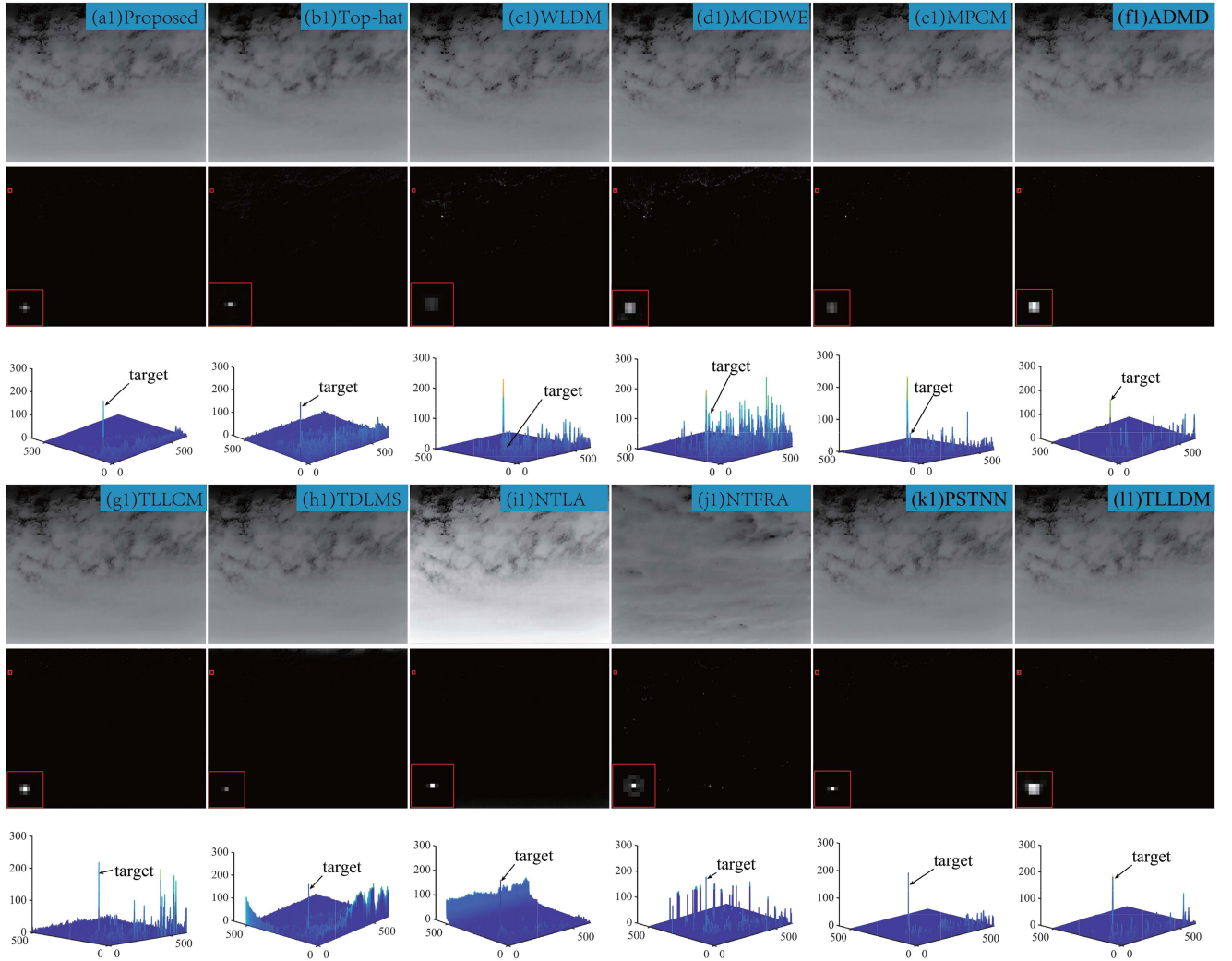


Fig. 5. Comparison of the background suppression effect of scene 1. The three images from the top to the bottom are the background image, the difference image and the 3D image of the difference image obtained by the corresponding algorithm, respectively.

TABLE II  
COMPARISON OF THE EVALUATION INDICATORS OF SCENE 1

	Proposed	WLDM [31]	Top-hat [4]	MGDWE [32]	MPCM [17]	TLLCM [18]	TDLMS [5]	NTLA [34]	NTFRA [33]	PSTNN [35]	ADMD [20]	TLLDM [19]
SSIM	0.9993	0.9971	0.9938	0.9940	0.9992	0.9989	0.9847	0.9962	0.9988	0.9992	0.9993	<b>0.9995</b>
SNR	15.10	9.99	8.32	13.17	14.60	16.04	13.91	21.11	19.11	14.35	19.81	<b>21.13</b>
BSF	<b>268.74</b>	92.02	93.42	62.98	170.43	173.88	80.38	117.32	204.60	240.45	199.41	236.01

The bold entities denote the max value of the evaluation indicators.

TABLE III  
COMPARISON OF THE EVALUATION INDICATORS OF SCENE 2

	Proposed	WLDM [31]	Top-hat [4]	MGDWE [32]	MPCM [17]	TLLCM [18]	TDLMS [5]	NTLA [34]	NTFRA [33]	PSTNN [35]	ADMD [20]	TLLDM [19]
SSIM	<b>0.9997</b>	0.9964	0.9963	0.9969	0.9988	0.9992	0.9751	0.9941	0.9992	0.9996	0.9995	0.9996
SNR	19.12	10.42	12.06	16.7	17.17	19.15	13.21	<b>26.11</b>	22.13	19.15	22.25	22.74
BSF	<b>437.06</b>	65.94	121.66	78.83	118.13	197.74	53.48	93.91	249.16	363.48	263.56	236.01

The bold entities denote the max value of the evaluation indicators.



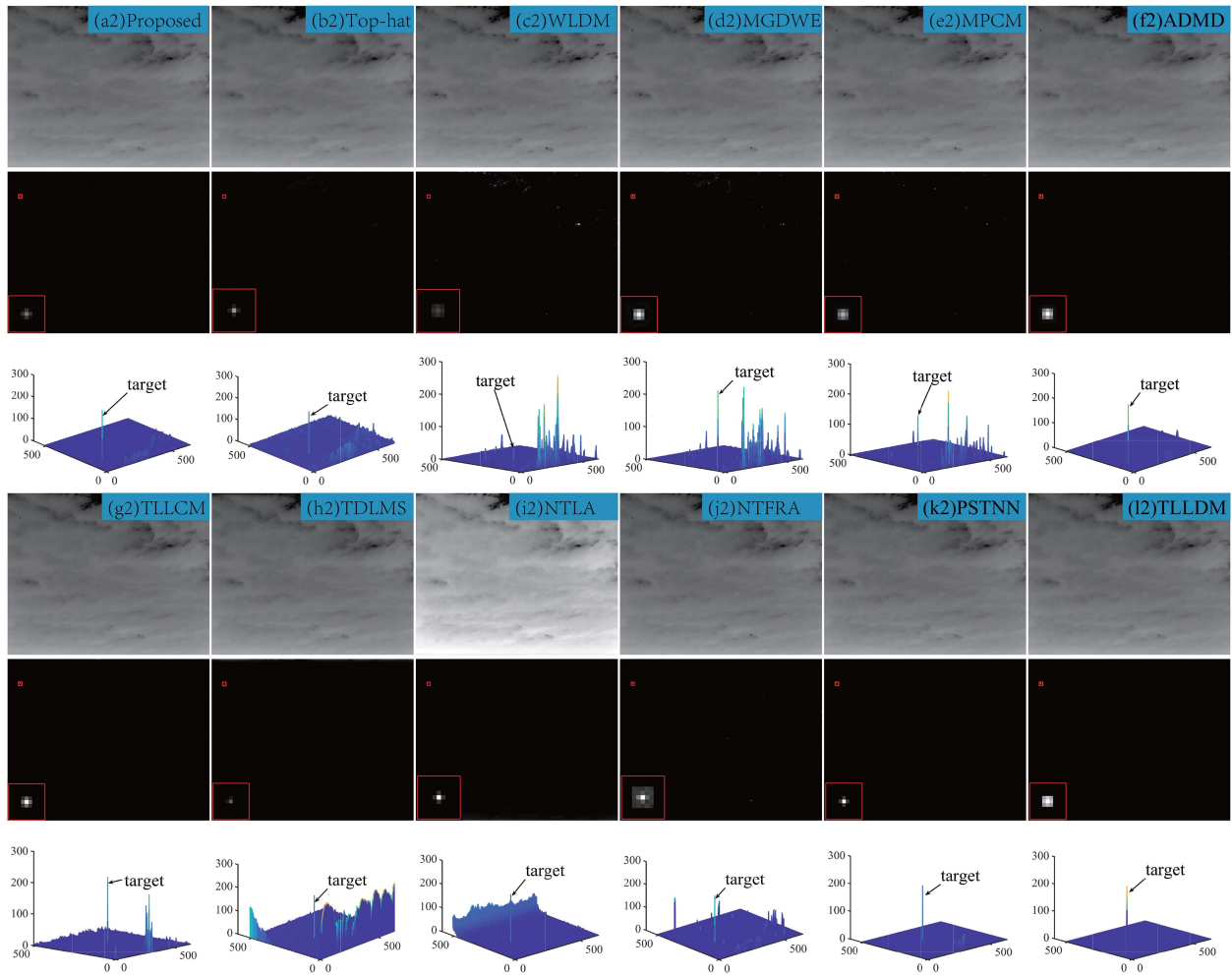


Fig. 6. Comparison of the background suppression effect of scene 2. The three images from the top to the bottom are the background image, the difference image and the 3D image of the difference image obtained by the corresponding algorithm, respectively.

TABLE IV  
COMPARISON OF THE EVALUATION INDICATORS OF SCENE 3

	Proposed	WLD	Top-hat	MGDWE	MPCM	TLLCM	TDLMS	NTLA	NTFRA	PSTNN	ADMD	TLLDM
		[31]	[4]	[32]	[17]	[18]	[5]	[34]	[33]	[35]	[20]	[19]
SSIM	<b>0.9997</b>	0.9898	0.9955	0.9929	0.9980	0.9978	0.9804	0.9933	0.9710	0.9990	0.9990	0.9968
SNR	13.16	-5.53	13.94	0.6	2.84	9.63	10.51	6.89	11.09	12.68	7.97	<b>15.33</b>
BSF	<b>379.59</b>	51.06	113.74	63.45	104.54	144.27	67.58	89.57	41.82	195.63	200.78	102.89

The bold entities denote the max value of the evaluation indicators.

background suppression effect of PSTNN and TLLDM is better, and the algorithm proposed in this paper is not affected by the target energy or edge contours, which can be seen in the two scenes also achieved the ideal background suppression effect. Tables IV and V show that our method has the highest SSIM in the two scenes; in terms of SNR, this paper’s algorithm is lower than Top-hat and TLLDM algorithms in scene 3, and only lower than the NTFRA and PSTNN method in scene 4; the BSF of this paper’s algorithm is highest in scene 3 and scene 4.

The background in scene 5 consists of some buildings and clouds, and the target energy is high. The experimental results are shown in Fig. 9. It can be seen that the Top-hat and TDLMS

algorithms can effectively detect the target points, but the suppression ability for strong edge contours of buildings is poor; while the MPCM, TLLCM, MGDWE, and WLD algorithms are influenced by the edge contours of clouds, and more strong clutter interference is distributed around the detected target signals in the differential map, which leads to the undesirable background suppression effect; NTLA and NTFRA do not suppress the edge contours of buildings sufficiently; PSTNN suppresses the background sufficiently in this scene without being affected by the edge contours of buildings; the ADMD and TLLDM algorithms remove the background clutter better. There is only a small amount of building clutter in the 3D map, and the performance of our algorithm shows a stronger background

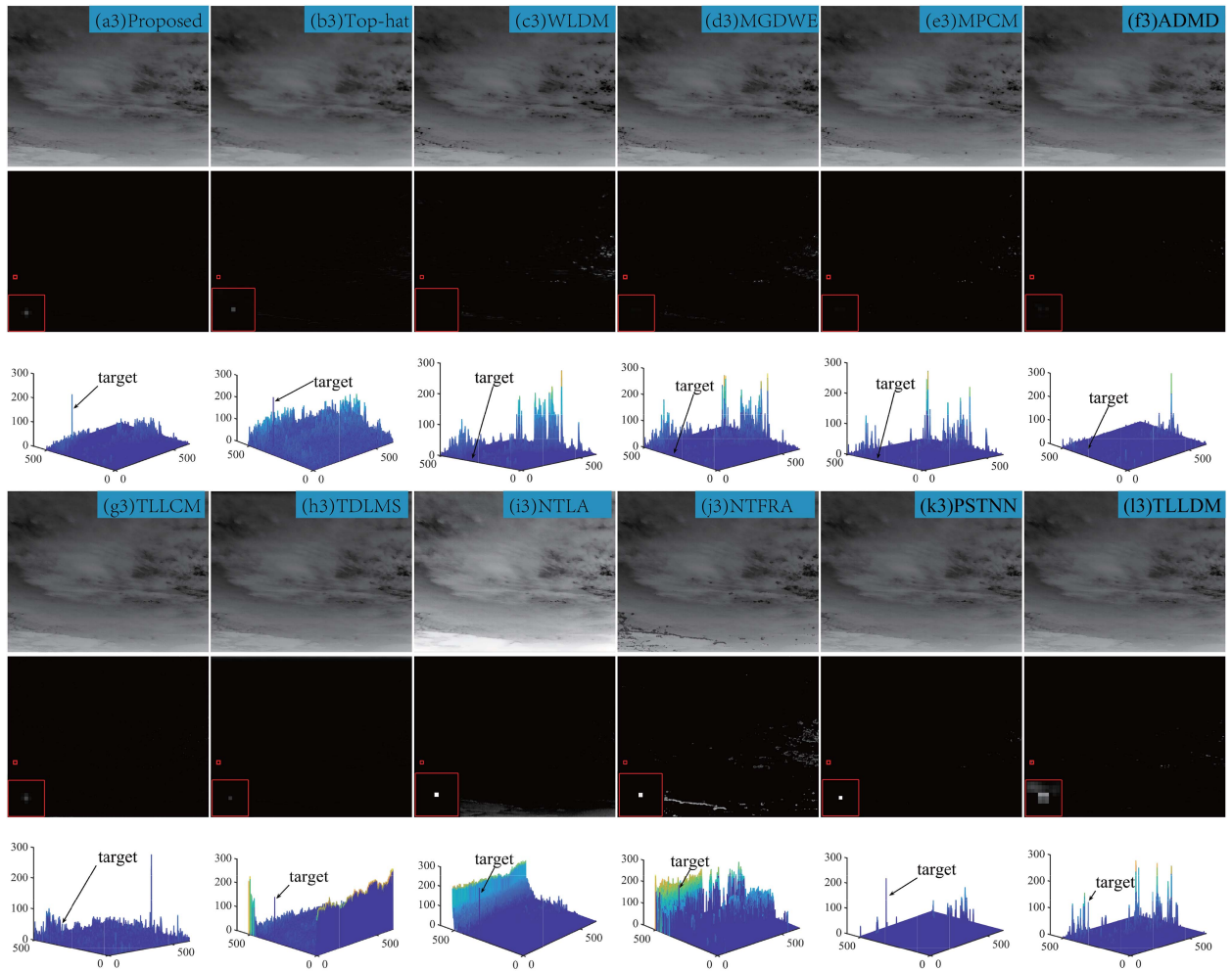


Fig. 7. Comparison of the background suppression effect of scene 3. The three images from the top to the bottom are the background image, the difference image and the 3D image of the difference image obtained by the corresponding algorithm, respectively.

TABLE V  
COMPARISON OF THE EVALUATION INDICATORS OF SCENE 4

	Proposed	WLDM [31]	Top-hat [4]	MGDWE [32]	MPCM [17]	TLLCM [18]	TDLMS [5]	NTLA [34]	NTFRA [33]	PSTNN [35]	ADMD [20]	TLLDM [19]
SSIM	<b>0.9996</b>	0.9860	0.9949	0.9941	0.9961	0.9984	0.9921	0.9823	0.9992	0.9995	0.9953	0.9992
SNR	13.57	-2.57	5.69	7.85	8.22	13.07	13.34	13.11	18.28	12.90	9.96	<b>19.58</b>
BSF	<b>341.05</b>	51.94	105.83	79.96	94.18	144.61	97.61	57.81	254.94	313.67	95.11	199.10

The bold entities denote the max value of the evaluation indicators.

suppression ability compared with other comparative algorithms besides the PSTNN algorithm.

Through the Table VI, the SSIM and BSF of our algorithm have better performance, and the SNR is only lower the TLLDM algorithm.

In scene 6, the target is affected by the severe cloud interference and a low contrast. The experimental results are shown in Fig. 10. The differential maps of the Top-hat and TDLMS algorithms weaken the intensity of the target signal, thus more strong clutters remains; similar to scene 3, MPCM, WLDM, MGDWE, TLLCM shows poor background suppression ability in scene 6, which is caused by the contrast between the target and the surrounding background is not large enough, and the

gray-value of the target is weak; the contrast of the target is low, therefore the target enhancement effect of TLLDM and ADMD algorithms is limited, however, the NTLA, NTFRA, and PSTNN algorithms can clearly observe the target signal in the differential map of scene 6, but the suppression ability

for some clouds is still insufficient; and the algorithm in this paper can greatly suppress the background clutter in this scene, but the weakening effect for the target energy is also obvious so there is some clutters with higher energy, but the target energy is relatively higher. From the Table VII, it can be seen that the BSF of our method is better than other comparison algorithms; the SSIM are also higher than other comparison algorithms, and the SNR is only lower than ADMD algorithm. Therefore, it can be

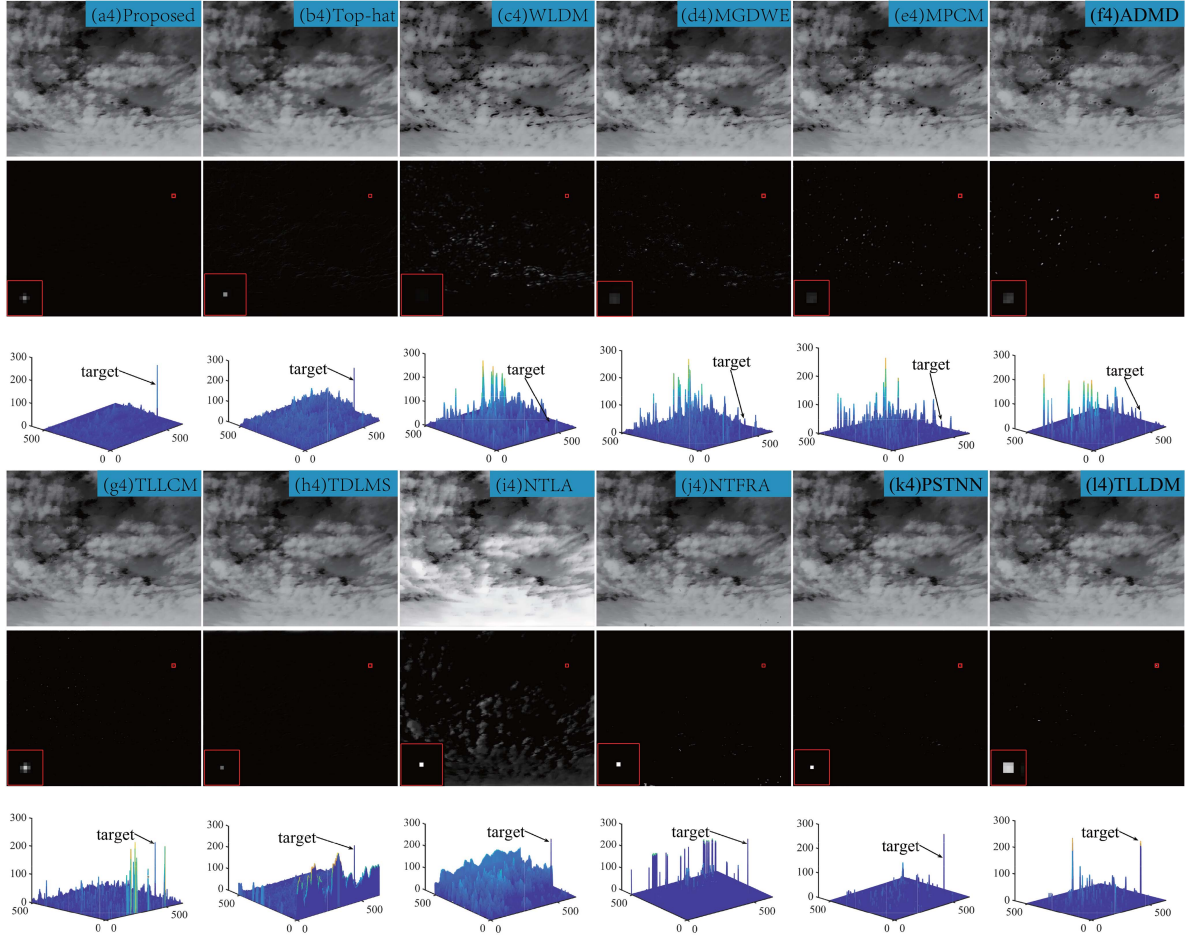


Fig. 8. Comparison of the background suppression effect of scene 4. The three images from the top to the bottom are the background image, the difference image and the 3D image of the difference image obtained by the corresponding algorithm, respectively.

TABLE VI  
COMPARISON OF THE EVALUATION INDICATORS OF SCENE 5

	Proposed	WLD	Top-hat	MGDWE	MPCM	TLLCM	TDLMS	NTLA	NTFRA	PSTNN	ADMD	TLLDM
		[31]	[4]	[32]	[17]	[18]	[5]	[34]	[33]	[35]	[20]	[19]
SSIM	<b>0.9999</b>	0.9929	0.9986	0.9961	0.9991	0.9985	0.9909	0.9950	0.9980	0.9998	0.9995	0.9995
SNR	17.15	9.66	9.58	12.3	15.19	14.30	4.44	16.11	16.89	15.61	15.56	<b>19.75</b>
BSF	<b>752.34</b>	76.34	206.19	110.97	226.1	182.81	119.06	104.77	157.77	446.54	310.82	273.26

The bold entities denote the max value of the evaluation indicators.

TABLE VII  
COMPARISON OF THE EVALUATION INDICATORS OF SCENE 6

	Proposed	WLD	Top-hat	MGDWE	MPCM	TLLCM	TDLMS	NTLA	NTFRA	PSTNN	ADMD	TLLDM
		[31]	[4]	[32]	[17]	[18]	[5]	[34]	[33]	[35]	[20]	[19]
SSIM	<b>0.9998</b>	0.9890	0.9854	0.9857	0.9990	0.9727	0.9874	0.9694	0.9651	0.9985	0.9997	0.9946
SNR	9.75	4.22	7.88	-8.06	5.37	2.66	5.2	4.22	9.31	8.33	<b>10.90</b>	-1.05
BSF	<b>485.74</b>	69.30	61.07	54.66	200.54	40.88	63.30	43.90	37.81	174.00	369.72	96.18

The bold entities denote the max value of the evaluation indicators.

seen that the algorithm proposed in this paper has an excellent background suppression performance after analyzing the above experiment results.

### B. Analysis of the Effect of Energy Enhancement

Among them, a, b, and c denote the conventional high-order correlation energy enhancement effect image, the 3D image of

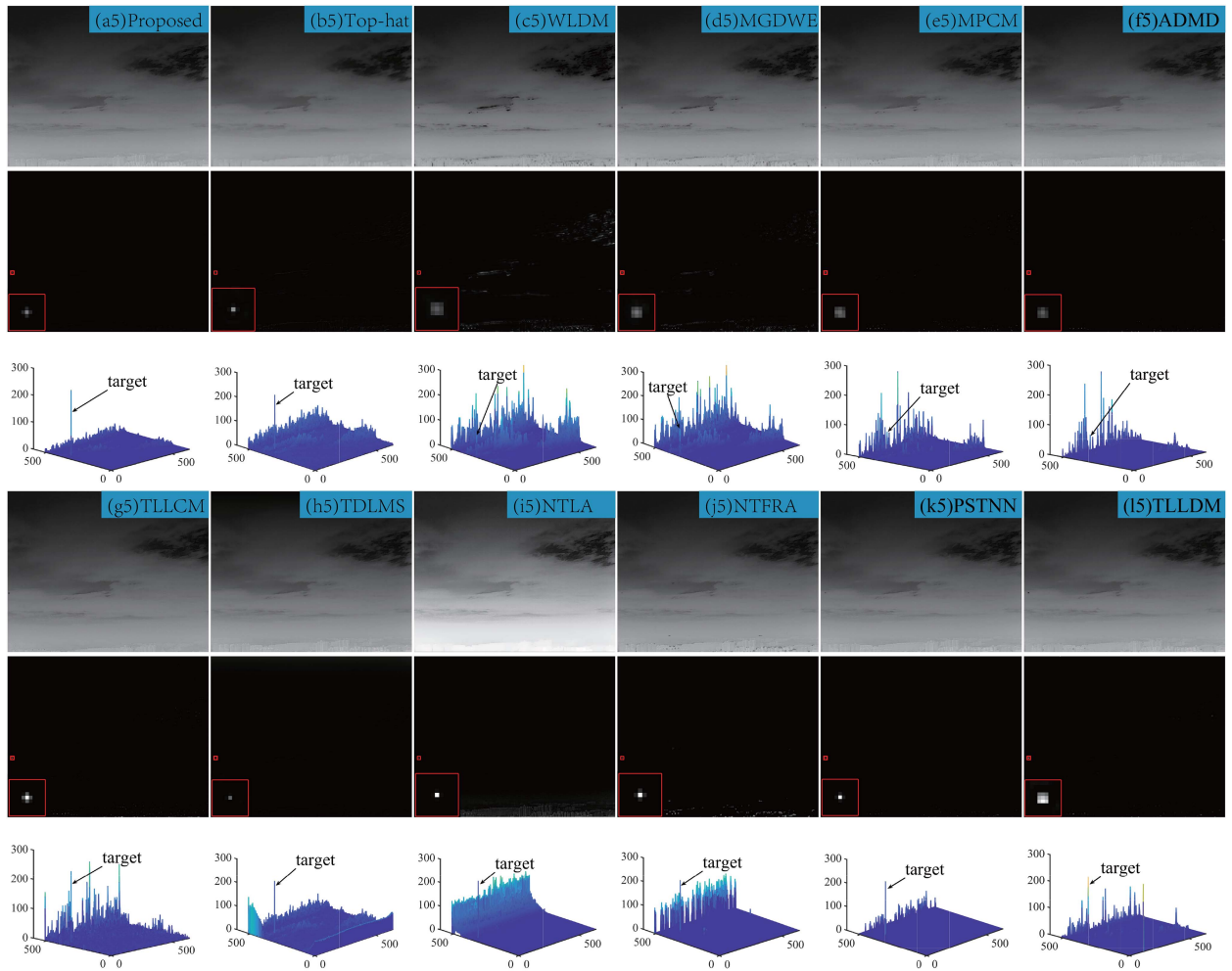


Fig. 9. Comparison of the background suppression effect of scene 5. The three images from the top to the bottom are the background image, the difference image and the 3D image of the difference image obtained by the corresponding algorithm, respectively.

anisotropic background suppression, and the 3D image of conventional energy enhancement algorithm, respectively. d, e, and f denote the improved high-order correlation energy enhancement effect image, the 3D image of anisotropic background suppression, and the 3D image of improved energy enhancement algorithm proposed in this paper, respectively. The energy of the target can be enhanced effectively by using high-order correlation energy accumulation. From the following figures, it can be seen that the target point of the difference map is not very obvious, which is not conducive to observation. After the energy enhancement by the traditional higher-order correlation algorithm, the energy of the target point is enhanced while the energy of the background clutter is also enhanced; while the target is significantly enhanced by the improved high-order correlation algorithm, not only the target energy is effectively enhanced, but also the background clutter is well removed to highlight the target signal. From Figs. 11–16, we can see that the energy enhancement directly using the traditional high-order correlation algorithm, because there is no candidate targets screening, all signals will be enhanced, which includes clutter and noise. It means that if the energy of clutter is higher than the energy of target, it will lead to more serious interference in the

detection process; and the algorithm we improved to screen out the candidate targets and eliminate the false targets before the energy enhancement, which can effectively reduce the clutters and increase the reliability of the detection algorithm. Noting that the same enhancement intensity of the traditional algorithm, the number of clutter in this paper is significantly less than the traditional algorithm, therefore the enhancement effect of the improved algorithm is

better than the traditional algorithm, regardless of the change in the number of clutter or the energy of the target after energy enhancement. Table VIII shows the global signal-to-noise ratios of the original image, anisotropic difference image and the improved high-order correlation energy enhancement image, and the signal-to-noise ratio calculation formula is shown in (12). It is calculated that the signal-to-noise ratio is improved by 6.85 dB for scene 1, 2.18 dB for scene 2, 7.22 dB for scene 3, 7.75 dB for scene 4, 3.95 dB for scene 5, and 3.05 dB for scene 6.

### C. Analysis of the Effect of Detection Result

To verify the effectiveness of the detection algorithm in this paper, six scenes are selected for experiments and

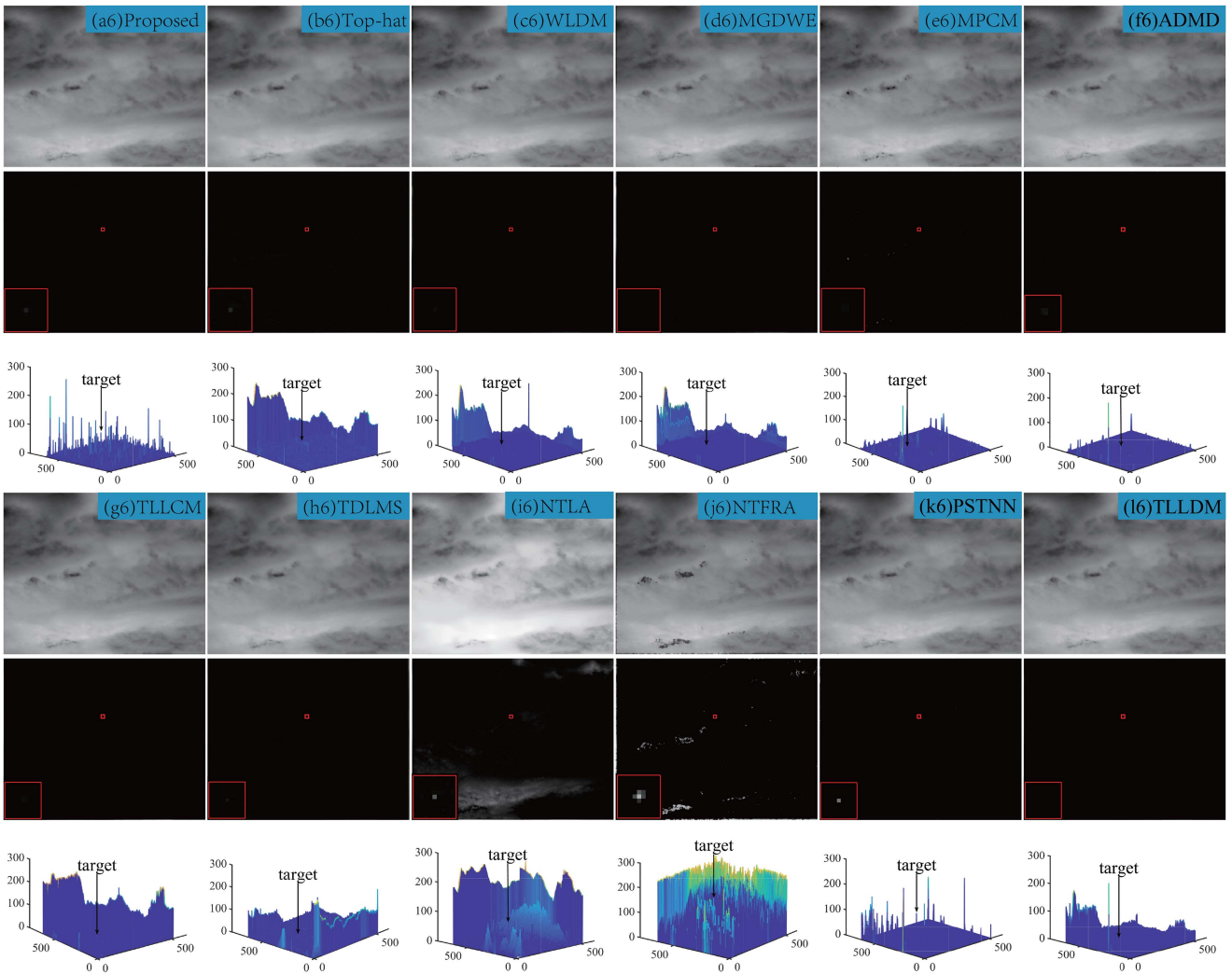


Fig. 10. Comparison of the background suppression effect of scene 6. The three images from the top to the bottom are the background image, the difference image and the 3D image of the difference image obtained by the corresponding algorithm, respectively.

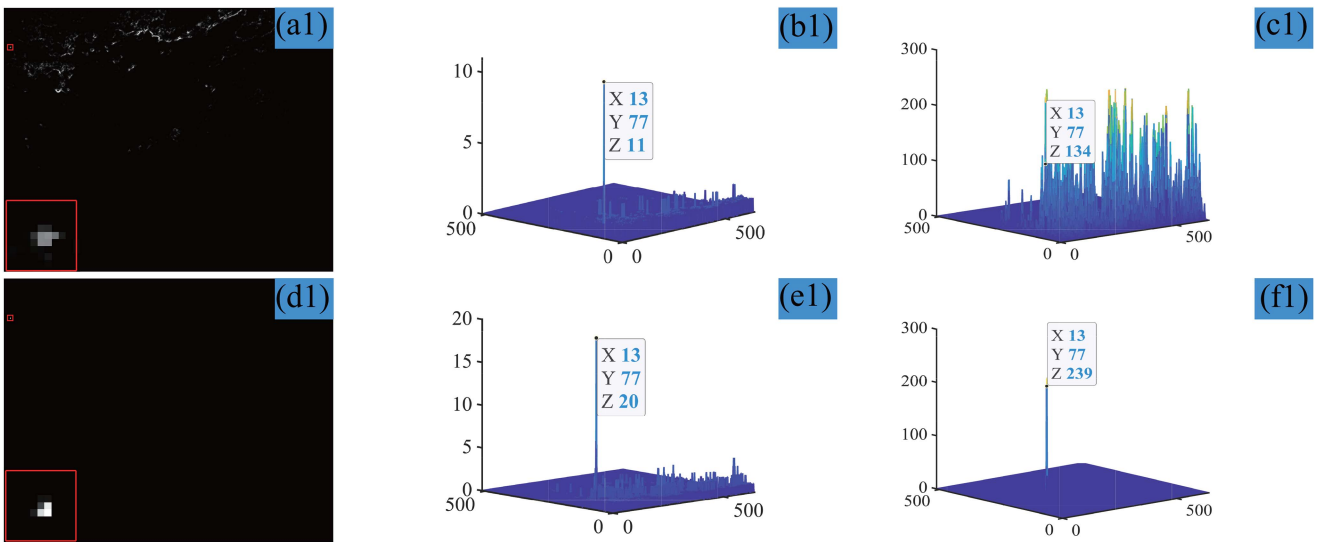


Fig. 11. Comparison of the energy enhancement effect of scene 1.

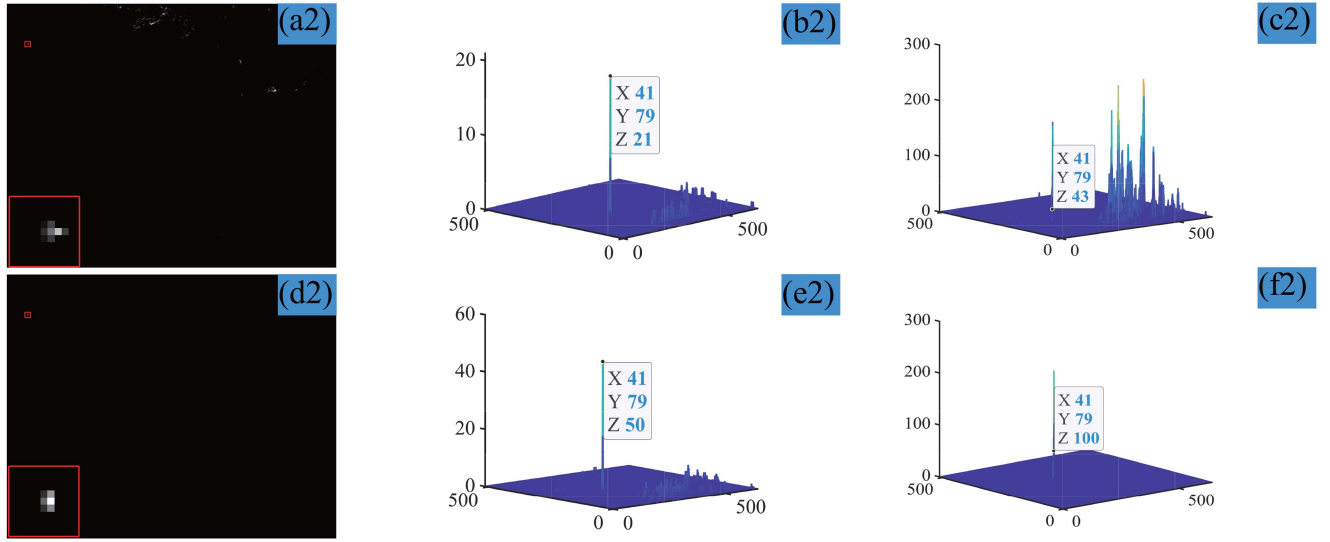


Fig. 12. Comparison of the energy enhancement effect of scene 2.

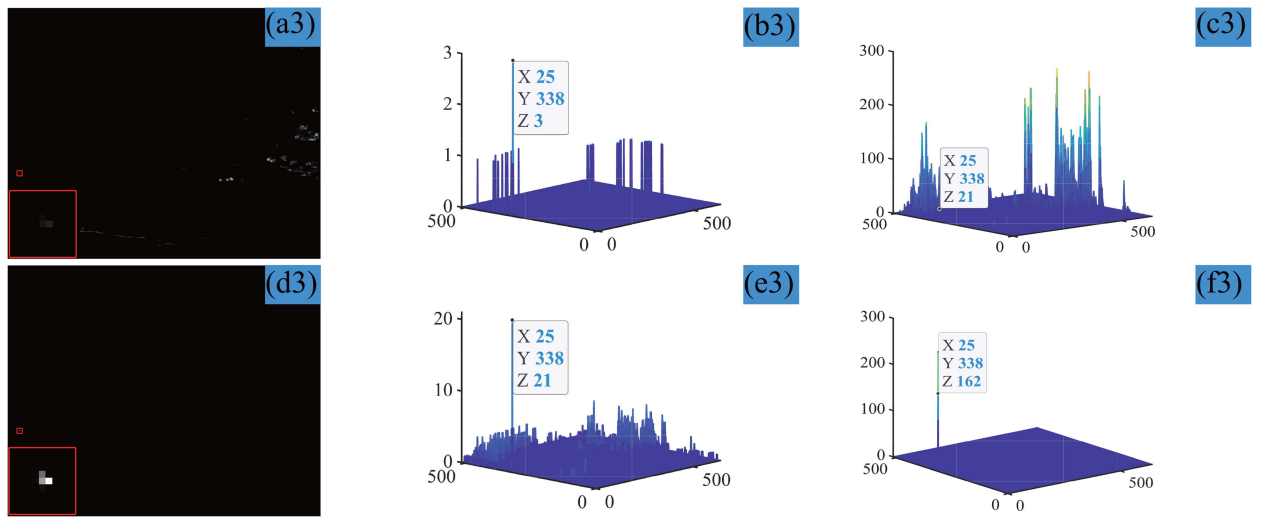


Fig. 13. Comparison of the energy enhancement effect of scene 3.

TABLE VIII  
THE SNR BEFORE AND AFTER ENERGY ENHANCEMENT (dB)

	Scene 1	Scene 2	Scene 3	Scene 4	Scene 5	Scene 6
Origin image	5.24	7.4	3.12	3.43	4.96	0.63
Background suppression image	15.10	19.12	13.16	13.57	17.15	9.75
Energy enhancement image	21.95	21.30	20.38	21.32	21.10	12.80

compared with the traditional algorithms and some algorithms proposed in recent years, which are WLDM [31], Top-hat [4], MGDWE [32], MPCM [17], TLLCM [18], NTFRA [33], NTLA [34], PSTNN [35], TDLMS [5], ADMD [20] and TLLDM [19]. The parameters used in the algorithms in this paper are shown in Table IX. Where  $M$  is the cumulative number of frames,  $L$  is the spatial neighborhood size of the target, and  $r$  is the radius of the target search neighborhood with reference to (9). The experimental results are shown in Figs. 17–22.

In scene 1, the detection effect of WLDM, MPCM, MGDWE, TLLCM and NTFRA is poor, which is due to the slow movement of the background of the subsequent images; the detection effect of Top-hat and TDLMS is better, but there is also a small amount of clutter interference; and the NTLA and PSTNN also has a good performance, the ADMD and TLLDM algorithms can detect the target and extract the target motion trajectory well. In scene 2, all algorithms obtained good result, which is the reason for the almost non-moving background and high target energy.

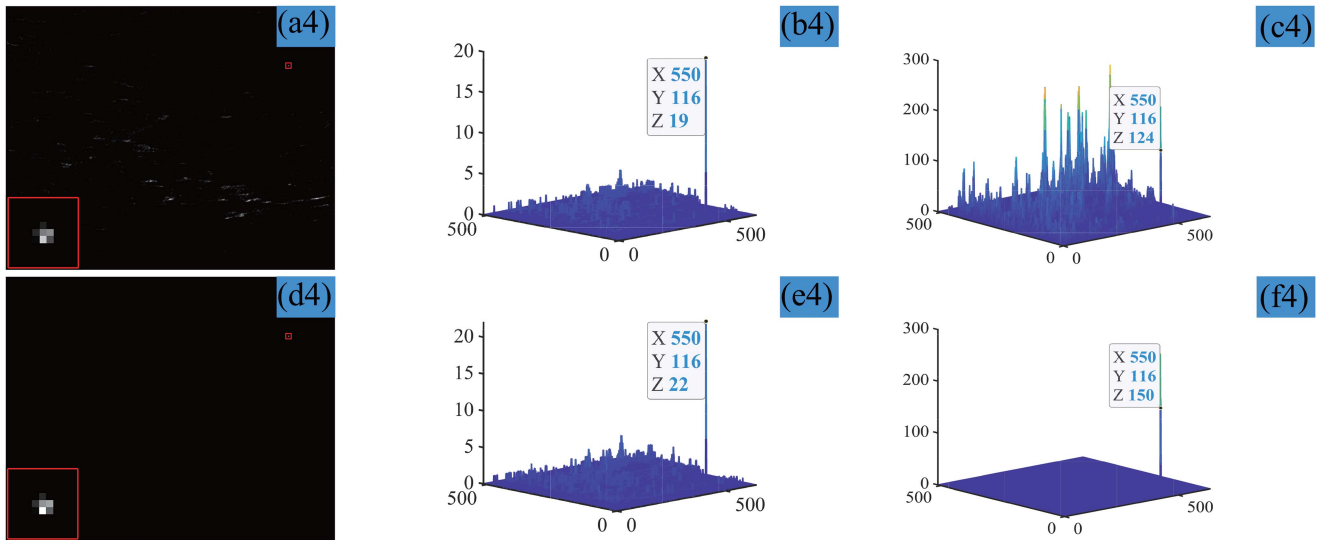


Fig. 14. Comparison of the energy enhancement effect of scene 4.

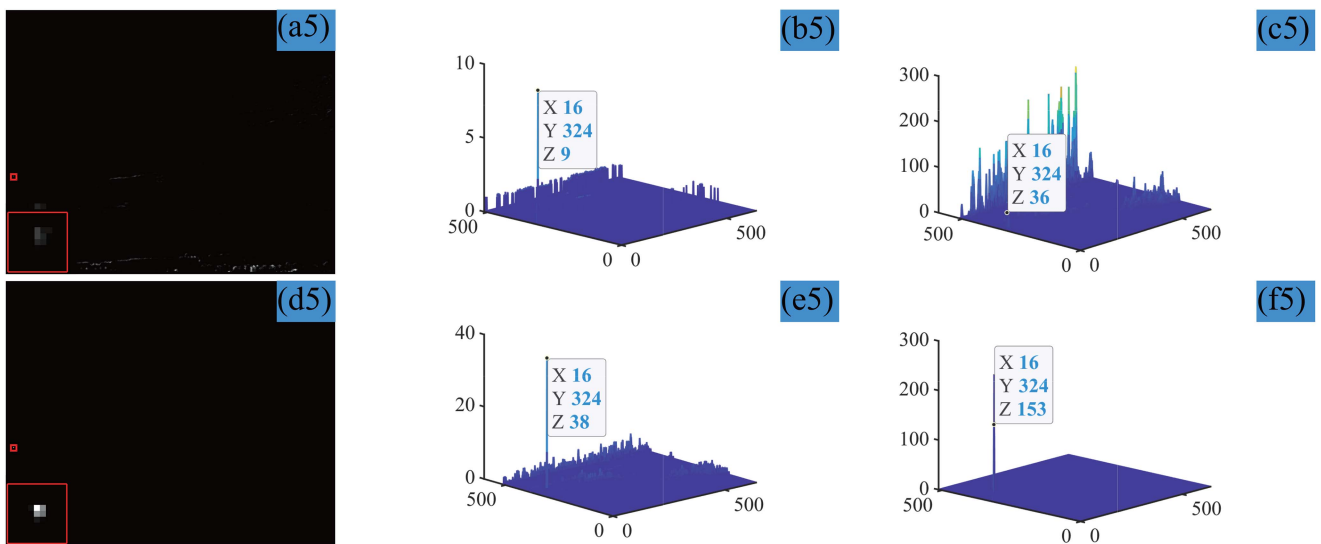


Fig. 15. Comparison of the energy enhancement effect of scene 5.

In scene 3 and scene 4, the WLDM, MPCM, MGDWE, and TLLCM can extract the target motion trajectory, but there are large continuous background clutter; the Top-hat and TDLMS algorithms have relatively good target trajectories, but there is target frame loss; the NTLA, NTFRA, and PSTNN algorithms have better detection effects than the above algorithms, but there are still different degree of clutter interference and target loss, the ADMD and TLLDM have missed detection, resulting in discontinuous target trajectory. This is due to the fact that the backgrounds of the two scenes move together with the target, and the span of the background movement is large, which increases the difficulty of detection. Scene 5 has a small number of

low-altitude buildings in the background, but the target is located at high altitude and the background changes slowly, so all algorithms can extract the target's motion trajectory better.

Scene 6 has a lower signal-to-noise ratio, but the background is relatively stable, so most algorithms can extract the target motion trajectory completely, but there are still different degrees of interference. Compared with the traditional detection algorithms, the detection algorithm proposed in this paper can suppress the background clutter more effectively in the above six scenes and extract the target motion trajectory completely. From the aforementioned analysis, it can be seen that WLDM and MGDWE algorithms mainly use local difference to detect targets in images, thus easily leading to higher false detection rate and missed detection rate, while MPCM and ADMD algorithm achieves detection through local contrast information of multi-scale blocks, but the algorithm tends to recognize darker backgrounds as targets, and although TLLCM and TLLDM algorithm uses a three-layer filtering window, the generalization ability of this

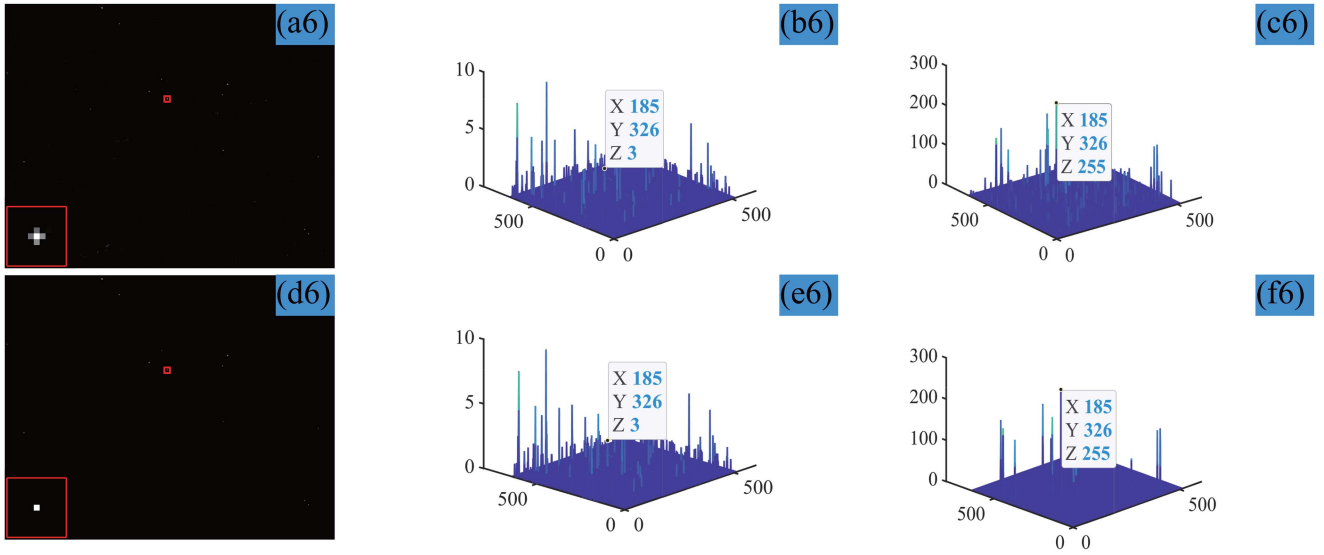


Fig. 16. Comparison of the energy enhancement effect of scene 6.

TABLE IX  
THE RELATED PARAMETERS OF MULTI-FRAME CORRELATION DETECTION

	The number of images $M$	Space neighborhood $L$	Target search neighborhood $R$	Features
Scene 1	3	5*5	9*9	Slow movement of target and background
Scene 2	3	5*5	9*9	The target moves slowly and the background barely moves
Scene 3	3	11*11	19*19	The target is moving fast and the background is changing slowly
Scene 4	3	3*3	7*7	Fast target movement and the background with a large span
Scene 5	3	7*7	17*17	Fast target movement and the background barely moves
Scene 6	3	5*5	9*9	The target moves slowly and the background barely moves

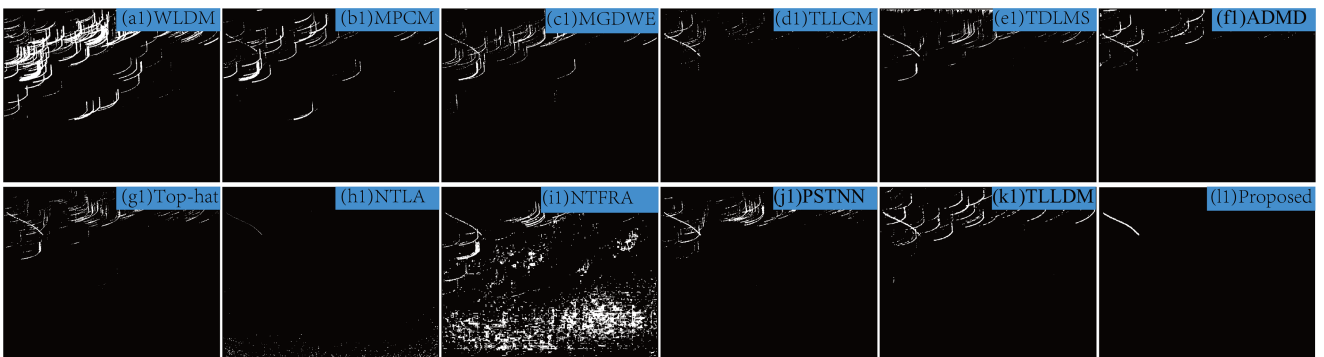


Fig. 17. The motion trajectories obtained by different algorithms of Scene 1.



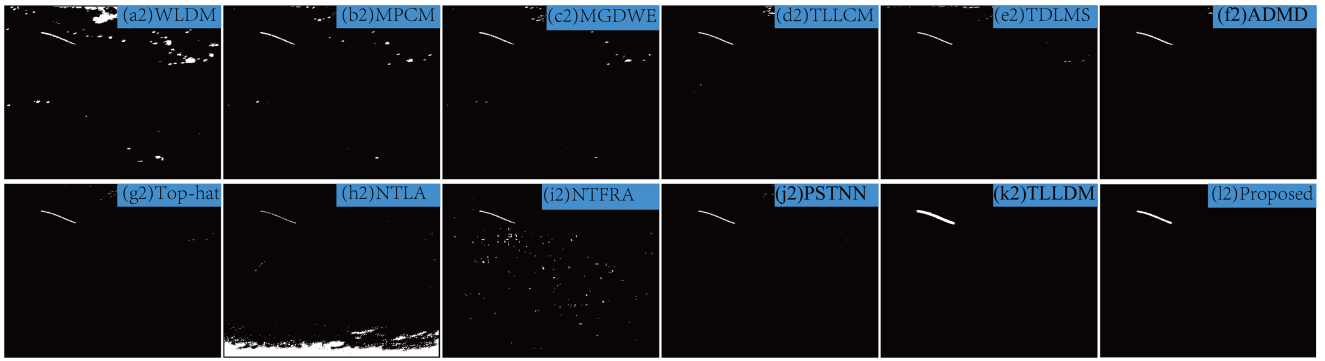


Fig. 18. The motion trajectories obtained by different algorithms of Scene 2.

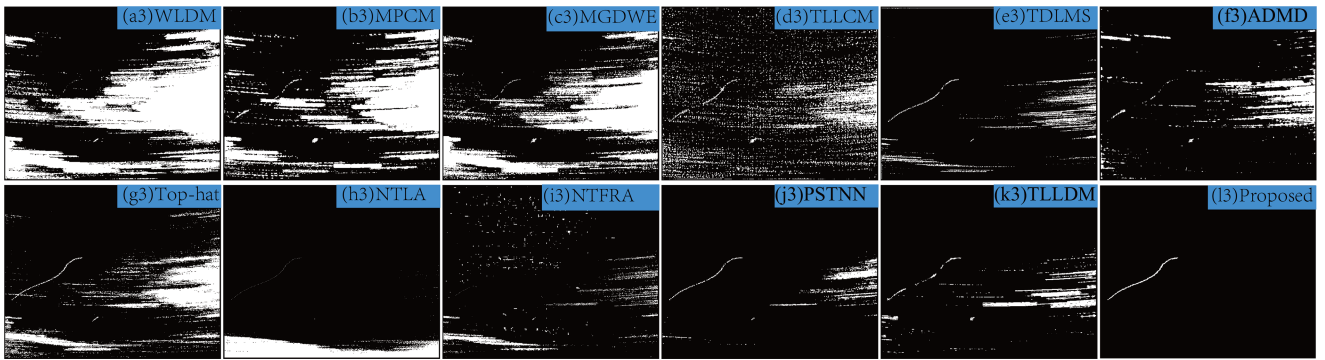


Fig. 19. The motion trajectories obtained by different algorithms of Scene 3.

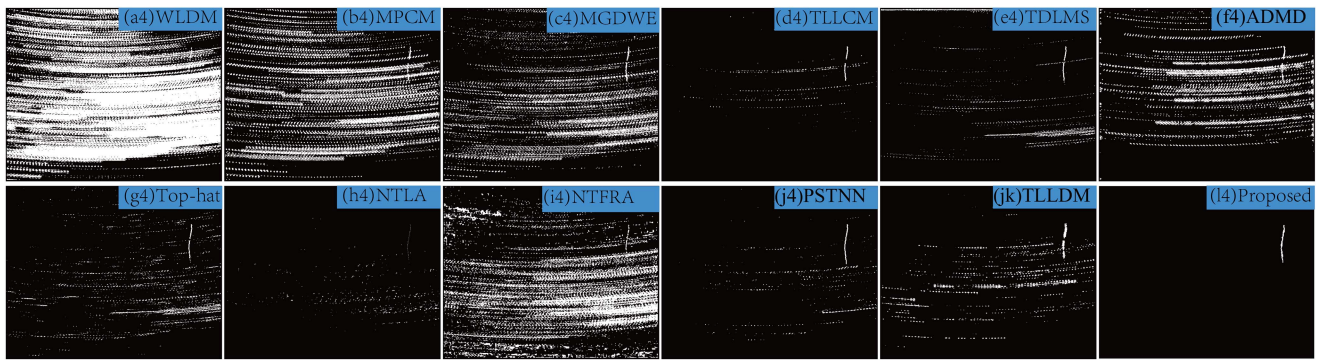


Fig. 20. The motion trajectories obtained by different algorithms of Scene 4.

algorithm is poor, so the local contrast measurement algorithm is suitable for smooth background with large contrast, but not for non-smooth and dramatically changing scenes, which is due to the fact that the local contrast measurement algorithm mainly uses the local features of the target to achieve the detection of weak targets, so it has high requirements on the contrast of the background, while the realistic background is often complex and variable. top-hat algorithm relies heavily on the selection of structural elements, TDLMS algorithm relies on the iteration step, so the traditional filtering algorithm is suitable for the target scale is small, uniform and smooth background, but the

detection effect is greatly reduced in the scene with complex background components. The low-rank algorithms represented by PSTNN, NTLA, and NTFRA algorithms can effectively suppress the strong edge contour background in scenes with high signal-to-noise ratio. If the image signal-to-noise ratio is too low and the target energy is weak, the distinction between the target and the surrounding background is not high, the sparse characteristics of the target and the low-rank characteristics of the background will be destroyed, which leads to undesirable detection results. In contrast, the algorithm in this paper is not only suitable for scenes with smooth background, but also can

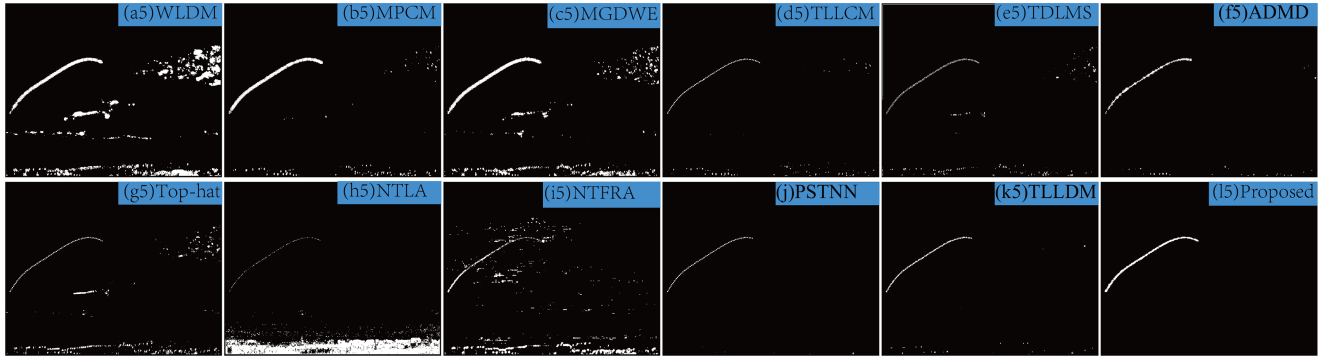


Fig. 21. The motion trajectories obtained by different algorithms of Scene 5.

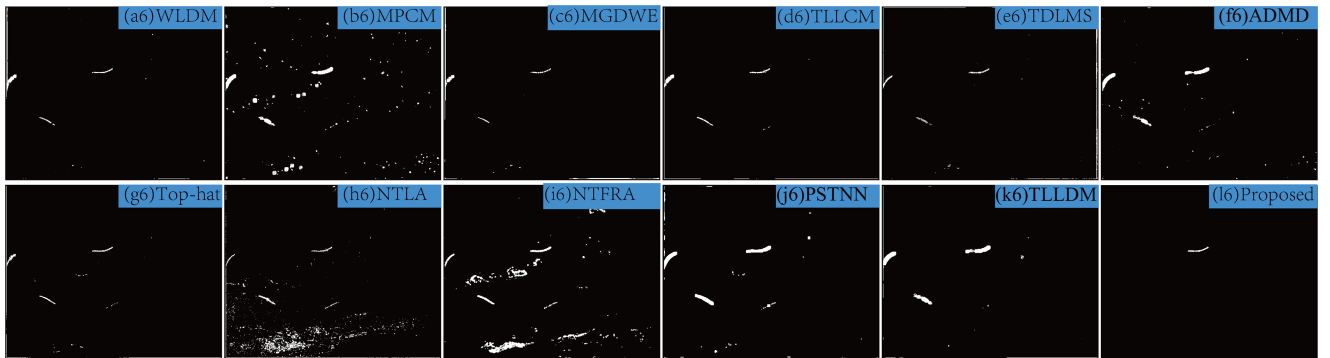


Fig. 22. The motion trajectories obtained by different algorithms of Scene 6.

obtain good detection results in scenes with large span and dynamic changes.

#### D. Analysis of the ROC Curves

In addition to the SNR, SSIM, and BSF metrics used in Section IV-B to evaluate the detection effectiveness of the algorithm, this section further compares the detection effectiveness of the algorithm by plotting the ROC curve. The horizontal coordinate of the following figure is the false alarm rate, and the vertical coordinate is the detection rate. The higher the detection rate, the better the detection effect and the more beneficial to the target recognition detection. The higher the false alarm rate, the more serious the noise interference is, which is not conducive to the detection and extraction of targets. The specific calculation formula is as follows [37].

$$P_d = \frac{NTDT}{NT} \times 100\% \quad (14)$$

$$P_f = \frac{NFDT}{NP} \times 100\% \quad (15)$$

where  $NT$  is the total number of targets in the sequence image,  $NTDT$  is the number of targets that can be detected,  $NFDT$  represents the number of false targets detected in the sequence image, and  $NP$  represents the number of all pixels in the whole image.

It can be seen from Fig. 23(a)–(f) that the our method presented in this paper achieves good results. In scene 1, the detection rate of our method reaches 100%, and the false alarm rate is 0, the result is same as the Top-hat algorithm. The detection rates of other algorithms are lower than 85%, and the false alarm rate is higher than the proposed algorithm in this paper; in scene 2, the detection rate in this paper is 98.67%, which is higher than other algorithms, and also has a lower false alarm rate; in scene 3, the detection rate of the algorithm proposed in this paper reaches 100%, and the detection rate of other comparison algorithms is lower than 85%, while the false alarm rate of the algorithm proposed in this paper is lower than the other comparison algorithms; Scene 4, the detection rate of the algorithm in this paper is 100%, and the false alarm rate is 0, and the detection rate of the other comparison algorithms is more than 80%, and the false alarm rate is higher than our algorithm; in scene 5, the detection rate of our algorithm is 100% and the false alarm rate is 0, at the same time, most of the comparison algorithms have a detection rate over 60% and the false alarm rate is higher than the algorithm presented in this paper; in scene 6, the detection rate of all the comparison algorithms over 80%, the performance of our algorithm is same as the NTLA and Top-hat, and higher than the other comparison algorithms. As a whole, the algorithm in this paper performs better in six scenes, by analyzing and comparing the ROC curves of the above algorithms, it is clear that our proposed algorithm is a better detection algorithm.

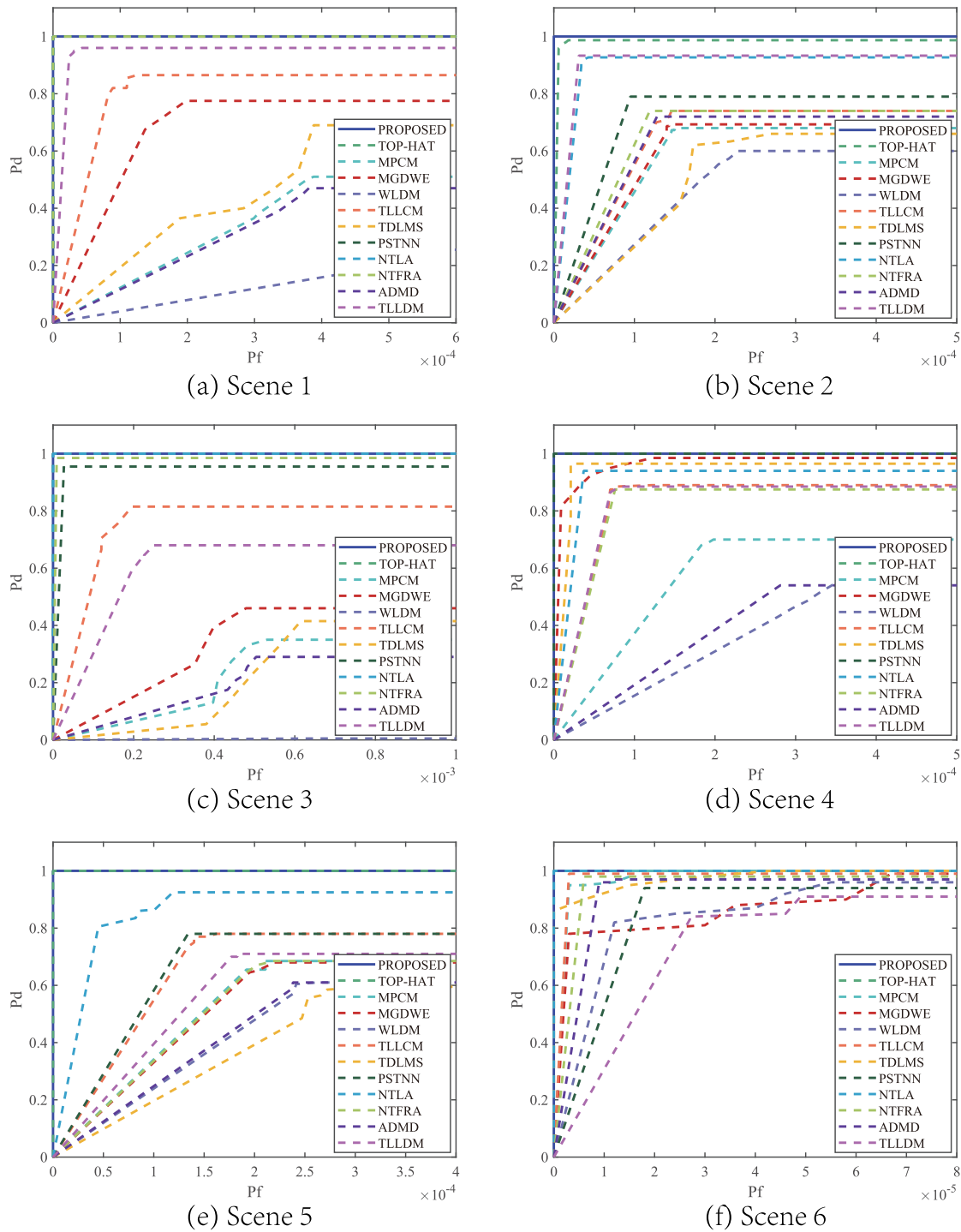


Fig. 23. The ROC curves of the six scenes

## V. CONCLUSION

To achieve effective detection of infrared weak and small targets, this paper proposes a multi-frame weak target detection algorithm based on motion energy estimation. Firstly, the traditional anisotropic diffusion function is improved and a new kernel diffusion function is proposed to complete the background suppression. Then, on the basis of background suppression, an

improved higher-order correlation energy enhancement method is used to enhance the target energy to further improve the SNR, which is beneficial to the subsequent target extraction. Finally, based on image pre-processing combined with the motion characteristics of the target in the spatio-temporal domain, a multi-frame small target detection algorithm based on motion energy estimation is proposed to realize multi-frame correlation detection. After the experimental comparison with the

traditional weak target detection methods, the following conclusions are drawn.

1. After background suppression by the kernel diffusion function proposed in this paper, the SSIM of the six scenes reach 0.9993, 0.9997, 0.9997, 0.9996, 0.9999, 0.9998, the SNR reach 15.10 dB, 19.12 dB, 13.16 dB, 13.57 dB, 17.15 dB, 9.75 dB, and the background suppression factor reach 268.74, 437.06, 379.59, 341.05, 752.34, 485.74.

2. Base on background suppression, the target energy is enhanced with improved high-order correlation, and the SNR of the four scenes reaches 21.95 dB, 21.30 dB, 20.38 dB, 21.32 dB, 21.10 dB and 12.80 dB, respectively. And the SNR is further improved.

3. On the basis of image pre-processing, the multi-frame detection algorithm proposed in the paper is used to complete multi-frame correlation detection. From the ROC curve analysis, it can be observed that the presented algorithm can achieve a high detection rate with a low false alarm rate, the detection rates of all four scenes reach more than 85%. The experimental results show that the proposed algorithm has certain practicality.

#### REFERENCES

- [1] R. Eysa and A. Hamdulla, "Issues on infrared dim small target detection and tracking," in *Proc. IEEE Int. Conf. Smart Grid Elect. Automat.*, 2019, pp. 452–456.
- [2] G. Wang, Z. Chen, and Q. Li, "A review of infrared weak and small targets detection under complicated background," *Infrared Technol.*, vol. 28, no. 5, pp. 287–292, 2006.
- [3] F. Yue, "Infrared dim and small target detection based on max-median filtering and k-means clustering," *Electro-Optic Technol. Appl.*, vol. 33, no. 5, pp. 41–43, 2018.
- [4] X. Bai and F. Zhou, "Analysis of new top-hat transformation and the application for infrared dim small target detection," *Pattern Recognit.*, vol. 43, no. 6, pp. 2145–2156, 2010.
- [5] T.-W. Bae, F. Zhang, and I.-S. Kweon, "Edge directional 2D LMS filter for infrared small target detection," *Infrared Phys. Technol.*, vol. 55, no. 1, pp. 137–145, 2012.
- [6] H. L. Qin, H. X. Z., S. Liu, and F. Li, "Dim and small target background suppression using bilateral filtering," *High Power Laser Particle Beams*, vol. 21, no. 1, pp. 25–28, 2009.
- [7] Y. Q. Zeng and Q. Chen, "Dim and small target background suppression based on improved bilateral filtering for single infrared image," *Infrared Technol.*, vol. 33, no. 9, pp. 537–540, 2011.
- [8] F. Lu, Y. Li, X. Chen, G. Chen, and P. Rao, "Weak target detection for PM model based on top-hat transform," *Syst. Eng. Electron.*, vol. 40, no. 7, pp. 1417–1422, 2018.
- [9] T. Wang, F. Chen, and X. Su, "Research of infrared background suppression method based on anisotropic bilateral filtering," *J. Hunan Univ. (Natural Sci.)*, vol. 45, no. 2, pp. 119–126, 2018.
- [10] S. Huang, Z. Peng, Z. Wang, X. Wang, and M. Li, "Infrared small target detection by density peaks searching and maximum-gray region growing," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1919–1923, Dec. 2019.
- [11] P. Pietro and M. Jitendra, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 7, pp. 629–639, Jul. 1990.
- [12] Q. Zhang, J. Cai, and Q. Zhang, "Anisotropic infrared background prediction method," *High Power Laser Part. Beams*, vol. 24, no. 02, pp. 301–306, 2012.
- [13] Q. Ling, S. Huang, X. Wu, and Y. Zhong, "Infrared small target detection based on kernel anisotropic diffusion," *High Power Laser Part. Beams*, vol. 27, no. 1, pp. 93–98, 2015.
- [14] J. Li, X. Fan, H. Chen, B. Li, L. Min, and Z. Xu, "Dim and small target detection based on improved spatio-temporal filtering," *IEEE Photon. J.*, vol. 14, no. 1, Feb. 2022, Art. no. 7801211.
- [15] X. Guan, Z. Peng, S. Huang, and Y. Chen, "Gaussian scale-space enhanced local contrast measure for small infrared target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 327–331, Feb. 2020.
- [16] C. L. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Tang, "A local contrast method for small infrared target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 574–581, Jan. 2014.
- [17] Y. Wei, X. You, and H. Li, "Multiscale patch-based contrast measure for small infrared target detection," *Pattern Recognit.*, vol. 58, pp. 216–226, 2016.
- [18] J. Han, S. Moradi, I. Faramarzi, C. Liu, H. Zhang, and Q. Zhao, "A local contrast method for infrared small-target detection utilizing a tri-layer window," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1822–1826, Oct. 2020.
- [19] J. Rao et al., "Infrared small target detection using tri-layer template local difference measure," *Opt. Precis. Eng.*, vol. 30, no. 7, pp. 869–882, 2022.
- [20] S. Moradi, P. Moallem, and M. F. Sabahi, "Fast and robust small infrared target detection using absolute directional mean difference algorithm," *Signal Process.*, vol. 177, 2020, Art. no. 107727.
- [21] J. Du, H. Lu, M. Hu, L. Zhang, and X. Shen, "CNN-based infrared dim small target detection algorithm using target-oriented shallow-deep features and effective small anchor," *IET Image Process.*, vol. 15, no. 1, pp. 1–15, 2021.
- [22] L. Ding, X. Xu, Y. Cao, G. Zhai, F. Yang, and L. Qian, "Detection and tracking of infrared small target by jointly using SSD and pipeline filter," *Digit. Signal Process.*, vol. 110, 2021, Art. no. 102949.
- [23] T. Xiaozhong, S. Bei, W. Junyu, Z. Zhen, and S. Shaojing, "EAAU-Net: Enhanced asymmetric attention u-net for infrared small target detection," *Remote Sens.*, vol. 13, no. 16, 2021, Art. no. 3200.
- [24] Y. Cai and Y. Zhang, "Infrared dim and small target detection based on dual-channel feature-enhancement integrated attention network," *Air Space Defense*, vol. 4, no. 4, pp. 14–22, 2021.
- [25] Y. Ma, C. Huang, J. Huang, and W. Wang, "Research of infrared dim and small target detection based on SSD framework," *Laser Infrared*, vol. 51, no. 10, pp. 1342–1347, 2021.
- [26] W. Cai, P. Xu, Z. Yang, X. Jiang, and B. Jiang, "Dim-small target detection of infrared images in complex background," *J. Appl. Opt.*, vol. 42, no. 4, pp. 643–650, 2021.
- [27] B. Wang and Y. Yong, "Dim target detection based on modified high order correlation method," *Infrared Laser Eng.*, vol. 35, no. S4, pp. 258–263, 2006.
- [28] Y. Tang, S. Huang, Y. Zhong, and J. Wu, "Moving dim target detection based on morphology and high order statistics in infrared image," *Modern Defense Technol.*, vol. 44, no. 2, pp. 151–156, 2016.
- [29] X. Wang, G. Lv, and L. Xu, "Infrared dim target detection based on visual attention," *Infrared Phys. Technol.*, vol. 55, no. 6, pp. 513–521, 2012.
- [30] Y. Yong, "Technologies of detection and recognizing small target," Ph.D. dissertation, Inst. of Opt. and Electron., Chinese Academy Sci., Beijing, China, 2003.
- [31] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Small infrared target detection based on weighted local difference measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4204–4214, Jul. 2016.
- [32] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Infrared small-target detection using multiscale gray difference weighted image entropy," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 1, pp. 60–72, Feb. 2016.
- [33] X. Kong, C. Yang, S. Cao, C. Li, and Z. Peng, "Infrared small target detection via nonconvex tensor fibered rank approximation," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5000321.
- [34] T. Liu et al., "Nonconvex tensor low-rank approximation for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5614718.
- [35] L. Zhang and Z. Peng, "Infrared small target detection based on partial sum of the tensor nuclear norm," *Remote Sens.*, vol. 11, no. 4, 2019, Art. no. 382.
- [36] X. Sun et al., "Infrared dim and small target detection data set in complex background," 2021. [Online]. Available: <https://www.scidb.cn/en/detail?dataSetId=808025946870251520>
- [37] L. Min, X. Fan, J. Li, Z. Xiang, and Q. Wu, "Dim and small target detection based on Gaussian Markov random field motion direction estimation," *IEEE Access*, vol. 10, pp. 48913–48926, 2022.