Dim and Small Target Detection Based on Improved Bilateral Filtering and Gaussian Motion Probability Estimation

Fan Xiangsuo, Qin Wenlin, Feng Gaoshan, Huang Qingnan, and Min Lei

***Abstract*—Dim and small target detection plays an important role in infrared target recognition systems. In this paper, we present a dim and small target detection algorithm based on improved bilateral filtering and Gaussian motion probability estimation, aiming to improve the detection efficiency of the detection system. First, a bilateral filtering algorithm based on image patch analysis is proposed to complete the background modeling, compare with single pixel, image patch contains more neighborhood information. Then, we use the Gaussian process combining the target position of consecutive n frames to predict the target position of the $(n+1)$th frame, and the target energy is accumulated along the trajectory direction at the same time. Finally, we construct the grayscale probability model to realize the multi-frame correlation detection, which combining the grayscale features and the motion characteristics of the target. Six scenes and eleven comparison algorithms are selected for experiments, experimental results show the effectiveness and robustness of the proposed algorithm.**

***Index Terms*—Bilateral filtering, dim and small target, gaussian process, motion estimation.**

I. Introduction

INFRARED automatic search and tracking system weak target detection technology, as an important part of the weak signal detection field, is widely used in military and civilian fields, such as space debris detection, early warning, missile guidance, and so on [1]. However, due to factors such as long-range imaging, clutter interference, and noise generated by imaging devices, the detection of weak targets has become increasingly difficult, and therefore has attracted the attention of research scholars in recent years. Many scholars have been successful in the field of weak target detection, and the current detection algorithms are mainly divided into two categories: single-frame detection algorithms and multi-frame detection algorithms [2].

The grayscale value of the central pixel can be predicted by the grayscale information of the neighborhood background, which is the idea of background prediction. Through background modeling, the target and noise can be separated from the background. Hu [3] proposed an improved non-local mean background modeling method, which assgins different weights based on the similarity of pixels in the filter window between two adjacent frames. Song [4] improved the traditional propagation filter based on image patch analysis, which predicts the grayscale values of pixels by calculating the similarity between patchs, and achieves a better background prediction effect. Han [5] constructed a new background prediction window consisting of an intermediate layer, an isolation layer, and a neighboring background layer, which can detect targets of different scales using only one window, the role of the isolation layer is to separate the target from the background completely. Han [6] proposed a multi-direction TDLMS algorithm, which can generate more accurate reference results compared with the traditional TDLMS algorithm, effectively reducing prediction errors and obtaining better background modeling effect. In addition to space domain filtering, frequency domain filtering can also effectively remove background clutter [7]. Usually, the target grayscale is higher than the background grayscale, so the noise generated by the imaging device and the target belong to the high-frequency component of the image, the background belongs to the low-frequency component of the image, we can design high-pass filter or low-pass filter to separate the target from the background in the frequency domain. For example, Yang [8] proposed an entropy-based adaptive high-pass filtering algorithm for background suppression of infrared images, and Wang [9] used wavelet transform to suppress background clutter, then combined with higher-order statistical analysis to detect dim and small target.

Human visual saliency-based detection methods [10], [11]achieve weak target detection based on the contrast between the target and the background. Ren [12] proposed an improved double-layer local contrast measure (IDLCM) method to suppress the image background, followed by further noise interference removal by multidirectional gradient (MG), and extracts the target using singular value decomposition. Lu [13] defined a fusion of differential scaling (RDLCM) and differential limiting (CDLCM) to enhance the target signal and suppress

Manuscript received 1 July 2024; accepted 9 August 2024. Date of publication 14 August 2024; date of current version 28 August 2024. This work was supported by the National Natural Science Foundation of China under Grant 62261004 and Grant 62001129. *(Corresponding author: Qin Wenlin.)*

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 and Huang Qingnan are with the School of Automation, Guangxi University of Science and Technology, Liuzhou 545006, China (e-mail: qinwenlin_123@163.com; huangqingnan@gxust.edu.cn).

Feng Gaoshan is with the Dongfeng Liuzhou Motor Company, Ltd., Liuzhou 545005, China (e-mail: fenggs@dflzm.com).

Min Lei is with the Institute of Optics and Electronics Chinese Academy of Sciences, Chengdu 610209, China (e-mail: minlei1986@163.com).

Digital Object Identifier 10.1109/JPHOT.2024.3443239

background clutter, and then complete the target detection by an adaptive threshold segmentation algorithm. Kou [14] used an improved density peak global search method (IDPGSM) to extract the candidate target region, to enhance the candidate region's contrast through a double weighted enhancement (DWELCM) method.

Detection methods based on low-rank and sparse theory [15], [16] realize the detection of weak target based on the sparse target and the low-rank background, and these methods have a strong ability to suppress strong background clutters of the image. Yang [17] introduced a comprehensive target saliency measure, which first extracts the candidate target regions in the image by the cross-window standard deviation (CSD), and then remove the background clutter of the image by the low-rank representation (LRR) method to enhance the target and suppress the background, and finally determine the real target by the iterative thresholding method. He [18] presented a low-rank and sparse representation (LRSR) method, which constructs the super-complete dictionary by two-dimensional Gaussian model, then obtains the target, background and noise components of the image by the LRSR model, and obtains the target image by combining thresholding method. Gao [19] proposed the IPI model, which constructs the image local block, then recovering the image background through the IPI model and acquiring the target image to achieve weak target detection.

Deep learning-based detection methods [20], [21] utilize convolutional neural networks to train data samples, construct detection models, and then detect weak targets. These methods have the advantages of fast speed and high detection accuracy compared with traditional image feature extraction methods. For example, Zhang [22] presented a data-driven approach called Attention-Guided Pyramid Context Network (AGPCNet) to against the complex background. Hou [23] proposed a robust infrared weak target detection network (RISTDnet) based on deep learning that constructs a feature extraction network combining manual feature extraction methods as well as combining threshold segmentation methods to extract targets. Yu [24] introduced a multiscale local contrast learning network (MLCL-net), which incorporates a bilinear feature pyramid network to overcome the problem that the target scale is too small and the slight pixel offset leads to a serious decrease in accuracy. Shi [25] proposed a coordinated attention and feature fusion combination network (CAFF-Net), to capture both low-level texture structure features and high-level semantic information of dim and small target with strong anti-interference ability, and can effectively avoid false and missed detection in complex background.

The single-frame detection methods can complete the detection by using only a single frame, but there are some limitations. Such as the filter-based method is easily disturbed by the complex background clutters. The local contrast measurement method has certain requirements on the contrast of the target and the scale of the target, otherwise the highlighted non-target areas are easily mistaken for the target causing detection failure. The strong edge contour background in low signal-to-noise ratio can destroy the sparse characteristic of the target and the low-rank characteristic of the background leading to the false detection. In addition, the deep learning method requires sufficient training samples, and the training process takes some time, so the timeliness of the algorithm is difficult to be satisfied. In summary, the information we can utilize in a single frame is limited, to detect target in low signal-to-noise ratio (SNR), it is necessary to combine the time-domain information of the target.

The multi-frame detection method makes full use of the space-domain information and time-domain information, which include the grayscale features and motion features of the target. The classical track-before-detect (TBD) algorithms mainly include dynamic programming, higher-order correlation and multi-level hypothesis testing, and so on. Yang [26] classified the structure of infrared images into five categories, firstly constructed a multi-directional filtering window with three layers of different gray levels, then combined morphological filtering with median filtering to reduce the algorithm running time, while proposed a target extraction method to determine the candidate targets, and finally implemented the target detection by a parameter optimization method based on fuzzy control theory. The method achieves better tracking effect. Huang [27] used the maximum background prediction algorithm to complete the background modeling, then extracting the suspected target, and proposed an improved Kalman filtering algorithm to predict the position of the target in the next image frame and track the target, which showed a high detection performance in different complex scenes. Shaik [28] proposed a tracking method based on frequency domain information correlation and Bayesian estimation to determine the real target by calculating the maximum probability of the target trajectory from the motion trajectories of multiple candidate targets. The method has high real-time performance and achieves good tracking results for both stationary and moving targets. Chen [29] combined the advantages of two-dimensional empirical modal decomposition and time-domain differential filtering methods, and the two-dimensional empirical modal decomposition has a strong adaptive capability, the algorithm can retain the target information in the high-frequency component, then the effective enhancement of the target intensity through the time-domain information between adjacent frames to achieve weak target detection. The method has better anti-clutter interference capability and better detection effect in low signal-to-noise ratio and complex scenes. Ren [30] proposed a detection method based on 3D collaborative filtering and spatial inversion, which first removes the background clutter by 3D collaborative filtering, followed by an energy accumulation algorithm to enhance the target signal, and finally extracts the real target by the spatial inversion method. This method effectively solves the problem of false alarm or false detection phenomenon caused by extremely similar target grayscale and background noise grayscale under low contrast, but the complexity of the algorithm is high. A representative algorithm of detect-before-track (DBT) is pipeline filtering. Li [31] improved adaptive pipeline filtering, which adaptively adjusts the pipe diameter according to the moving speed of the target, and the algorithm has stronger robustness in the case of severe background noise interference. Dong [32] first used Difference of Gaussian (DoG) algorithm to extract the interest points, then tracking these points in adjacent frames by HVS, and finally the tracking results were used to determine the real target by clustering algorithm (R-means). Lei [33] combining background estimation and frame difference method, which first

estimates the background of an image by the median of the pixels in the neighborhood of the pixel, then calculates the difference between the current frame and its corresponding background image, and fuses the results to achieve target detection. This method has good real-time performance. However, due to the limitation of median filtering, it is easily disturbed by complex background and the detection performance is degraded.

Both DBT and TBD algorithms require background modeling to obtain the difference map. The difference is that TBD does not know the a priori information of the target and needs to track all suspicious target to determine the real target through the posteriori probability, therefore it can achieve detection and tracking of weak targets in low SNR, but the storage and computation of TBD are large. DBT, on the other hand, first determines the target information of a single frame, and then combines the target motion information and grayscale information between frames to determine the targets on subsequent images to achieve tracking of weak targets. The structure of DBT algorithms is simple and it is easy to implement in hardware, but the anti-interference performance is insufficient. By comparing the above methods, a novel detection method is proposed in this paper. The main contributions are as follow:

1) In order to retain the target information while minimizing the background information, the traditional bilateral filtering is improved in this paper. The traditional bilateral filtering only uses the information of a single pixel to calculate the filter coefficients, compared with a single pixel, the mean value of image patch is more representative for the background information around the target, so this paper constructs the mean information of image patches at different scales, calculates the mean value of these patches and we use these mean values intead of single pixel to complete the background modeling.

2) The background modeling may weaken the target energy, the target intensity should be enhanced. In this paper, based on the historical position information of the target, the new target position is predicted by Gaussian process to accumulate the energy of the target along the motion direction, and the energy enhancement can furtherimprove the contrast of the image.

3) Finally, after energy enhancement, a multi-frame detection model based on grayscale weighted probability model is constructed by combining the motion characteristics and grayscale features of the target between consecutive frames, and calculate the probability of candidate targets, the real target is determined by the maximum grayscale probability.

II. Propose Methods

A. Improved Bilateral Filtering Background Modeling

Gaussian filtering only considers the distance between the neighborhood pixel and the central pixel, which tends to blur the edge details of the image. Therefore, Tomasi and Manduchi proposed bilateral filtering [34], which introduces a grayscale similarity function to denoise while preserving the edges of the image well. The dim and small target, as a "singularity" in the image, is distinct from the surrounding background, therefore the background information of the image is maximally preserved by bilateral filtering to highlight the target and improve the saliency.

Yaqiong Zeng [35] proposed an improved bilateral filtering algorithm for background estimation of infrared images, the authors construct a filtering template and participate in the filtering process. The algorithm has a better background prediction effect and better anti-interference ability than the traditional bilateral filtering algorithm. However, the filtering template constructed by this algorithm has a single structure, and applying it to scenes where the target scale changes will reduce its effectiveness in background estimation. In addition, the algorithm only consider the information of single pixel in the neighborhood, which achieves better background modeling when the scene is in a small background fluctuation, when the background fluctuation is large, the influence of the neighborhood information of single image element on the central pixel is not sufficiently considered, which leads to undesirable prediction results. In order to more fully exploit the neighborhood information of the target and cope with different scale variations of the target between adjecent frames at the same time, an improved bilateral filtering background modeling method is proposed in this paper.

As shown in Fig. 1, A is a local window consisting of 3*3pixels, now it is expanded into a local window B consisting of 3*3 region blocks, each of which corresponds to A, and the size of the region block is 3*3. The mean value of each region block in B is calculated and replaced by the grayscale value of the corresponding pixel in A, and then bilateral filtering is performed. In order to maximize the preservation of the background structure of the image, especially in the edge contour region, this paper expands the single pixel into a patch, and the patch analysis contains more information than a single pixel, so that better background prediction results can be obtained. The two filtering templates we constructed in the paper effectively solve the problem of inaccurate prediction due to the possible shift of pixels when the target scale is changed by the single-scale filtering template. The specific expressions of the background modeling algorithm in this paper are as follows:

$$\begin{cases} M_1 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}, M_2 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \end{cases} \tag{1}$$

In the above equation, M_1 and M_2 are the filtering templates constructed in this paper, respectively.

$$\begin{cases} P_1(i,j) = \sum\limits_{i,j \in N_{x_1,y_1}} e^{\frac{-[(i-x_1)^2+(j-y_1)^2]}{2\sigma_{d_1}^2}} e^{\frac{-\left|I_1'(i,j)-I_1'(x_1,y_1)\right|^2}{2\sigma_{r_1}^2}} \\ \cdot I_1''(i,j) \cdot M_1 \cdot \frac{1}{C_1} \\ C_1 = \sum\limits_{i,j \in N_{x_1,y_1}} e^{\frac{-[(i-x_1)^2+(j-y_1)^2]}{2\sigma_{d_1}^2}} e^{\frac{-\left|I_1'(i,j)-I_1'(x_1,y_1)\right|^2}{2\sigma_{r_1}^2}} \\ \cdot M_1 \\ I_1'(i,j) = \frac{1}{r_1 * r_1} \times \sum\limits_{m_1=-(r_1/2)}^{m_1=r_1/2} \sum\limits_{n_1=-(r_1/2)}^{n_1=r_1/2} I(i+m_1, j+n_1) \\ I_1''(i,j) = \frac{1}{r_1 * r_1} \times \sum\limits_{m_1=-(r_1/2)}^{m_1=r_1/2} \sum\limits_{n_1=-(r_1/2)}^{n_1=r_1/2} I_1'(i+m_1, j+n_1) \end{cases} \tag{2}$$

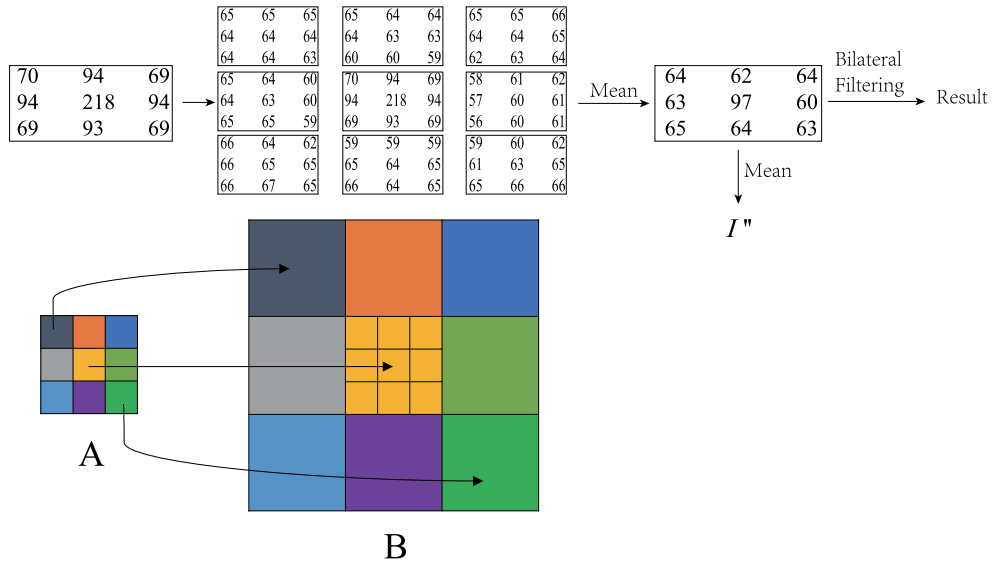

Fig. 1. Schematic diagram of background modeling.

I denotes the original image, (i, j) denotes the central pixel, I_1' denotes the mean value of the area, and I_1'' denotes the mean of the 3*3 blocks in the area. P_1 denotes the result of bilateral filtering, N_{x_1,y_1} denotes the range of the filter window, (x_1, y_1) denotes the neighborhood size, and r_1 denotes the neighborhood radius. σ_{d_1} is the standard deviation of the spatial neighborhood, and σ_{r_1} is the standard deviation of the pixel value. C_1 is a constant, which is the weighted sum of the product of spatial weights and grayscale weights.

$$\begin{cases} P_2(i,j) = \sum_{i,j \in N_{x_2,y_2}} e^{\frac{-[(i-x_2)^2+(j-y_2)^2]}{2\sigma_{d_2}^2}} e^{\frac{-\left|I_2'(i,j)-I_2'(x_2,y_2)\right|^2}{2\sigma_{r_2}^2}} \\ \cdot I_2''(i,j) \cdot M_2 \cdot \frac{1}{C_2} \\ C_2 = \sum_{i,j \in N_{x_2,y_2}} e^{\frac{-[(i-x_2)^2+(j-y_2)^2]}{2\sigma_{d_2}^2}} e^{\frac{-\left|I_2'(i,j)-I_2'(x_2,y_2)\right|^2}{2\sigma_{r_2}^2}} \\ \cdot M_2 \\ I_2'(i,j) = \frac{1}{r_2 * r_2} \times \sum_{m_2=-(r_2/2)}^{m_2=r_2/2} \sum_{n_2=-(r_2/2)}^{n_2=r_2/2} I(i+m_2, j+n_2) \\ I_2''(i,j) = \frac{1}{r_2 * r_2} \times \sum_{m_2=-(r_2/2)}^{m_2=r_2/2} \sum_{n_2=-(r_2/2)}^{n_2=r_2/2} I_2'(i+m_2, j+n_2) \end{cases} \tag{3}$$

I denotes the original image, (i, j) denotes the central pixel, I_2' denotes the grayscale mean of the area where the pixel is located, and I_2'' denotes the mean of the 5*5 blocks in the area. P_2 denotes the result of bilateral filtering, N_{x_2,y_2} denotes the range of the filter window, (x_2, y_2) denotes the neighborhood size, and r_2 denotes the neighborhood radius. σ_{d_2} is the standard deviation of the spatial neighborhood, and σ_{r_2} is the standard deviation of the pixel value. C_2 is a constant, which is the weighted sum of the product of spatial weights and grayscale weights.

$$P = P_1 * P_2 \tag{4}$$

where P denotes the result of P_1 multiplying with P_2, the bilateral filtered background prediction image.

B. Gaussian Motion Estimation

Most of the clutter interference can be removed and the contrast of the image is improved after background suppression. However, because the limitations of the spatio-temporal filtering algorithm, the suppression of some strong edge contours and noise interference is insufficient, resulting in more clutter retaining in the difference map, and the contrast between the target and the surrounding background is not obvious, which is not conducive to detect the target. To improve the detection efficiency of the algorithm, the contrast between the target and the background clutter can be improved by enhancing the intensity of the target. The energy enhancement algorithms are mainly time-domain energy accumulation or space-domain energy accumulation algorithms [36], [37], which enhance the target signal to a certain extent but do not consider the time-space domain information of the target sufficiently; therefore, it is crucial to fully incorporate the spatio-temporal information in the target energy enhancement process.

Shaik [38] used the initial position of the target and its historical position information to predict the target position in the next frame based on the Bayesian conditional probability principle, which enables the detection and tracking of weak target. Gaussian process is widely used in the field of target tracking, and it is mainly used for the prediction of target position [39]. The prediction of Gaussian process is an interpolation method, which assumes that the function is smooth and does not change significantly between observation points. Therefore, when the points we need to predict are far away from the observed points, the reliability of the prediction result will be reduced. Since the Gaussian process requires to calculate the inverse matrix of the covariance matrix, this method is feasible when the number of

observations is small. Too many observations will increase the execution time of the algorithm.

Inspired by Gaussian process, in long-range imaging, the change of the position of small target is small between frames and the moving direction has certain regularity, Gaussian process can be applied to predict the target position. Assuming that the target position of the N frames is known, the target position of the N+1th frame can be predicted by Gaussian process, and then the candidate target energy of the N frames is accumulated along the trajectory, which can greatly enhance the intensity of the target. The Gaussian process is briefly described as follows.

Suppose that the values of the function at different points are random variables that obey a Gaussian distribution with the following prior distribution [39]:

$$f(x) \sim GP\left(m(x), k\left(x, x'\right)\right) \tag{5}$$

Where $m(x)$ is the prior mean function of $f(x)$ and $k(x, x')$ is the covariance function of $f(x)$. Typically, we choose the radial basis function based covariance function as follows:

$$k\left(x, x'\right) = \sigma_f^2 \exp\left(-\frac{1}{2}\left(\frac{\|x - x'\|}{\ell}\right)^2\right) \tag{6}$$

Where σ_f^2 and ℓ are hyperparameters that denote the variance and length scale of the function, respectively.

Given the previous n coordinates $X = [x_1, x_2, \ldots, x_n]$ and the corresponding function values $Y = [y_1, y_2, \ldots, y_n]$, we can calculate the covariance matrix $K(X, X)$ at these points and the vectors $k(x_{n+1}, X)$ associated with them as follows:

$$K(X, X) = \begin{bmatrix} k\left(x_1, x_1\right) & k\left(x_1, x_2\right) & \cdots & k\left(x_1, x_n\right) \\ k\left(x_2, x_1\right) & k\left(x_2, x_2\right) & \cdots & k\left(x_2, x_n\right) \\ \vdots & \vdots & \ddots & \vdots \\ k\left(x_n, x_1\right) & k\left(x_n, x_2\right) & \cdots & k\left(x_n, x_n\right) \end{bmatrix} \tag{7}$$

$$k(x_{n+1}, X) = [k(x_{n+1}, x_1)k(x_{n+1}, x_2) \cdots k(x_{n+1}, x_n)] \tag{8}$$

We can use these values to calculate the predicted distribution of the $(n+1)th$ coordinates given the $(n)th$ coordinates:

$$f(x_{n+1})|(X, Y, x_{n+1}) = N(\mu(x_{n+1}), \sigma^2(x_{n+1})) \tag{9}$$

Where $\mu(x_{n+1})$ and $\sigma^2(x_{n+1})$ are:

$$\begin{cases} \mu(x_{n+1}) = k(\mathrm{x}_{n+1}, X)[K(X, X) + \sigma_n^2 I]^{-1} Y \\ \sigma^2(x_{n+1}) = k(x_{n+1}, x_{n+1}) - k(x_{n+1}, X) \\ [K(X, X) + \sigma_n^2 I]^{-1} k(X, x_{n+1}) \end{cases} \tag{10}$$

Where I is the unit matrix and σ_n^2 is the variance of the noise. This allows the Gaussian process to be used to predict the value of the function of the $(n+1)$th coordinate.

To reduce the program running time, the differential map is segmented using the double-window segmentation algorithm before energy enhancement. The selection of segmentation threshold follows the following principles: retain the target region while minimizing the interference of background noise, too much clutter will increase the running time. After the threshold segmentation, only a small number of interest points are left on the image, and the target signal intensity is enhanced by energy accumulation, and the target signal enhancement will also eliminate part of the clutter interference. Relevant expressions are as follow [40]:

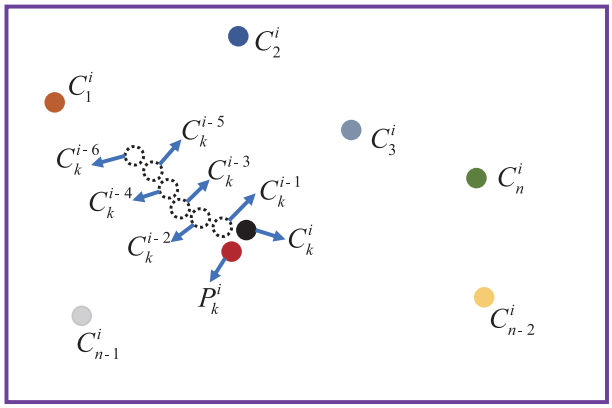

Fig. 2. Gaussian prediction schematic.

$$\begin{cases} r_1 = 1, r_2 = 5 \\ q_1 = \frac{1}{r_1^2} \times \sum_{m=-(r_1/2)}^{m=(r_1/2)} \sum_{n=-(r_1/2)}^{n=(r_1/2)} P(i+m, j+n) \\ q_2 = \frac{1}{r_2^2} \times \sum_{m=-(r_1/2)}^{m=(r_1/2)} \sum_{n=-(r_1/2)}^{n=(r_1/2)} P(i+m, j+n) \\ q_3 = \frac{q_1 - q_2}{r_2^2 - r_1^2} \\ if q_3 > T \\ D(i, j) = 1 \\ \text{else} \\ D(i, j) = 0 \end{cases} \tag{11}$$

In the above equation, P represents the difference image. r_1 and r_2 represent the radius size of the inner and outer windows, respectively. q_1 and q_2 represent the average gray value of the inner and outer windows, and q_3 represents the difference of mean value between q_1 and q_2. T represents the segmentation threshold, which is selected empirically. The basic principle is to ensure that the target region is preserved while the amount of noise or clutter is least. D represents the segmented image.

As shown in Fig. 2, n represents the number of candidate target regions of the current, i denotes the current frame, C_k^i is the kth candidate target of the current frame, assuming that the true position $C_k^{i-6}, \ldots, C_k^{i-1}$ of its previous frames is known, and P_k^i is obtained by Gaussian process prediction, by comparing the distance between P_k^i and the candidate targets, the closest distance is the true position of the kth candidate target of the current frame, then the target energy of the previous frames is accumulated to the kth candidate target, and the target signal enhancement is completed. The reason why the target position can be predicted by Gaussian process is that the position of weak targets does not change significantly between consecutive frames, however, in some scenes there are cases where the background moves, so the energy of these pseudo-targets will

also be enhanced. The specific expressions are as follows:

$$\begin{cases} L'_{n+1}(x_{n+1}, y_{n+1}) = GP_{Pre}(L_1(x_1, y_1), \\ L_2(x_2, y_2), \ldots, L_n(x_n, y_n)) \\ d^{ii}_{n+1} = \sqrt{(x_{ii} - x_{n+1})^2 + (y_{ii} - y_{n+1})^2} \\ L_{n+1}(x_{n+1}, y_{n+1}) = \min d^{ii}_{n+1}(x_{ii}, y_{ii}) \\ G_k = D_k * I_k \\ G'_{n+1} = \sum_{k=1}^{n+1} G_k(x_k, y_k) \end{cases} \tag{12}$$

GP_{Pre} denotes the Gaussian process prediction function, $L_n(x_n, y_n)$ denotes the centroid of the candidate target in the nth frame, $L'_{n+1}(x_{n+1}, y_{n+1})$ denotes the predicted centroid of the $(n+1)$th frame, ii denotes the number of candidate targets on the $(n+1)$th frame, d denotes the distance between the predicted centroid and the centroid of all candidate target regions in the image, and $L_{n+1}(x_{n+1}, y_{n+1})$ is the real centroid of the $(n+1)$th frame. G_k denotes the gray value of the candidate target region, and G'_{n+1} denotes the accumulation of the gray value of the previous n-frame target region corresponding to the current candidate target region to the candidate target region corresponding to the $(n+1)$th frame, which is the energy enhancement result.

C. Effective Displacement Energy Probability

After energy enhancement of the candidate target area, a small number of points of interest will be retained. These interest points may be background, random noise or target, and the grayscale of the target may be lower than the grayscale of the strong edge contour background or random noise points. For this reason, the target needs to be separated from the noise based on the continuity of the target motion and the randomness of the noise. In traditional multi-frame detection methods (such as pipeline filtering [31]), the number of target appears and target moves between consecutive frames are usually counted, and if the number exceeds a threshold, the current point is the real target. The algorithm can achieve better detection results in scenes where the background moves slowly or even does not move, once the background moves rapidly with the target, the method is prone to missed or false detection, reducing the reliability of the detection algorithm.

Based on the continuity of the target movement between frames, this paper proposes a detection approach which is the effective displacement energy probability model. Assuming that the target does continuous motion in 5-9 consecutive frames, the target appears in the current frame then will also appear in the small neighborhood of the corresponding position in the next frame, and the number of target appears between consecutive frames is recorded. Also, if the number of target moves between consecutive frames reaches half of the frames, the target is considered to be effectively move. When the number of occurrences of the target between consecutive frames reaches the set threshold and it is a effective move, the sum of the energy of the point between consecutive frames is calculated, and the sum of the energy of all candidate target points in the image is also calculated. Considering that the noise is randomly distributed, the probability of appearing in consecutive frames is small, and almost no effective movement occurs; while the background clutter usually does not move, but there is a possibility of effective movement; however, the target is continuously moving between frames, so the probability of effective movement of the target is larger, and the corresponding energy probability is also larger, so the candidate target point with the largest energy probability can be considered the real target.

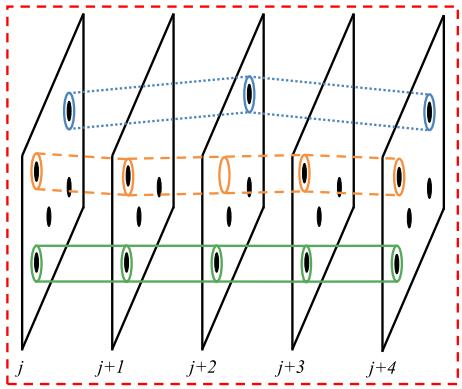

Fig. 3. Effective probability diagram.

Fig. 3 shows the detection model constructed in this paper, and the number of image frames L that the model can accommodate is 5. Now to determine the real target from the n suspicious targets in the first frame, assume that the ith candidate target in the first frame is in the small neighborhood of the same position in the second frame, and the number of target appears a_{num} plus 1, and if the position is moved, the number of target moves m_{num} plus 1. Then we continue to search the third frame with the target position of the second frame as the center, until we traverse the whole pipeline and output the results of a_{num} and m_{num}. When both a_{num} and m_{num} are more than half of the pipe length, we define the target point as the effective displacement. The effective displacement energy of the target E_i is calculated and the point with the highest probability of effective displacement energy T_i is the real target point. For the case that the target is lost in the current frame, the target position of the previous frame is used as the center to search the next frame. The specific expressions are as follows:

$$\begin{cases} \text{if } a_{\text{mum}} > (L/2) \quad \&\& \quad m_{\text{mum}} > (L/2) \\ E^i = \sum_{k=1}^{l} E^i_k \quad (l < L) \\ T_i = \frac{E^i}{\sum E^i} \\ F = \max\{T_1, T_2, \ldots, T_n\} \end{cases} \tag{13}$$

The above equation, E denotes the image after energy enhancement, k is the frame number where the effective displacement occurs, n denotes the number of candidate targets in the image, and E_i denotes the energy accumulation value of the effective movement of the ith candidate target. T_i denotes the energy probability of the candidate target. F denotes the real target, which candidate target with the highest energy probability.

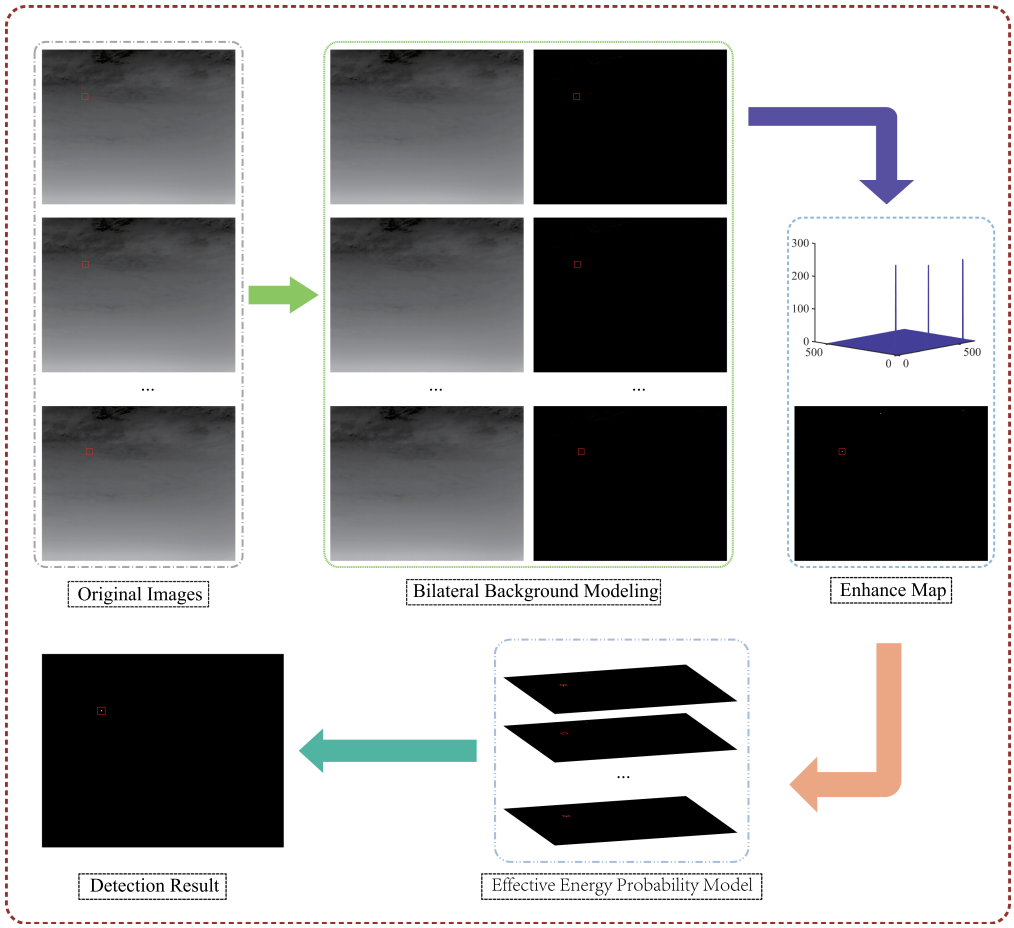

Fig. 4. Flowchart of the algorithm in this paper.

D. Algorithm Summary

This section summarizes the algorithm of this paper, which follows the ideas of background modeling, energy enhancement and multi-frame correlation detection. First, considering that the traditional bilateral filtering using a single pixel for background prediction does not fully utilize the background information of the local neighborhood, this paper extends the single pixel of the traditional filtering into a block for background prediction and incorporates the filtering results of two different scale filtering templates, and the improved background modeling algorithm fully exploits the neighborhood information of the target. Then, the candidate target region is obtained by the double-window segmentation method, and the Gaussian process is used to predict new position, and realize the accumulation of the candidate target energy along the direction of the motion trajectory. Finally, according to the continuity of target motion and the randomness of noise, combining the motion characteristics and gray-scale characteristics of the target, an effective displacement energy probability model of the target is proposed, and the candidata target with the largest energy probability is the real target point. The related experimental results are in Part III. Fig. 4 shows the flow chart of the algorithm in this paper, and Table I shows the pseudo-code related to the algorithm in this paper.

III. Experiment

A. Experimental Setup

1) Dataset: To verify the feasibility of our algorithm, six scenes are used to experimental validation. These scenes contain clouds with different degrees of complexity and motion speed of target. The specific information is referred to Table II. The dataset is obtained from the literature [41]. The targets are marked in the original image, and the target area is enlarged and placed in the lower left corner of the image (as shown in Fig. 5).

2) Evaluation Indicators: In addition to the intuitive background modeling, energy enhancement and multi-frame correlation detection to reflect the feasibility of the algorithm, the evaluation indexes such as background structure similarity(SSIM), background suppression factor(BSF) and signal-to-noise ratio(SNR) are also selected to evaluate the effectiveness of the algorithm. Among them, the background modeling effect is analyzed by the background suppression factor and background structure similarity, and the larger the value, the better the background suppression effect. The energy enhancement effect is analyzed by the SNR and the average gray level of the target region. Finally, the ROC curve can be used to evaluate the detection performance of the algorithm, which can better reflect

TABLE I
ALGORITHM PSEUDO-CODE

Algorithm 1: Spatio-temporal filtering and energy probability model
Input: Original images f
Output: Resulting images f_T
1: Set parameters σ_d=2, L=5
2: for i=1 to m
3: for j=1 to n
4: Background modeling use Equation (1), (2), (3) and (4)
5: P=Bilateral(f)
6: end
7: end
8: G=f-P
9: Image segmentation by Equation (11)
10: E=segment(G)
11: K=E*f
12: Target enhance by Equation (12)
13: D=enhance(K)
14: Target exaction by Equation (13)
15: f_T=probability(D)

TABLE II
DATASET DETAILS

	Image frames	Image size	Target size	Background
Scene 1	300	512*640	3*3	Slow movement
Scene 2	300	512*640	3*3	Slow movement
Scene 3	300	512*640	3*3	Fast movement
Scene 4	500	512*640	3*3	Slow movement
Scene 5	500	512*640	3*3	Slow movement
Scene 6	500	512*640	3*3	Fast movement

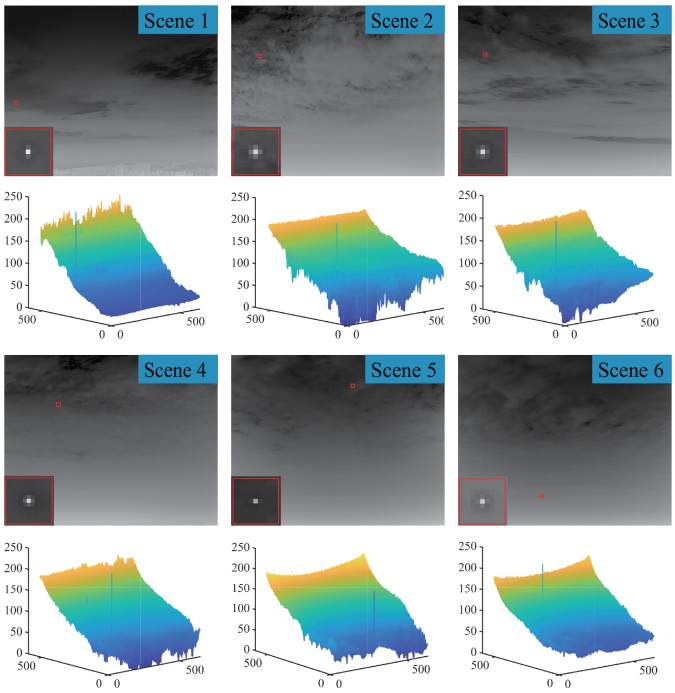

Fig. 5. The original image and its 3D map.

TABLE III
COMPARISON ALGORITHM PARAMETER SETTINGS

Algorithm	parameter
Bilateral [35]	Filter size: 7×7, $\sigma_d = 2$, $\sigma_s = 0.1$
TDLMS [43]	Support size: 5×5, step size: 1×10^{-8}
MPCM [10]	$N = [3, 5, 7, 9]$
TLLCM [11]	$C = 3$, $k = 9$, $R \in [5, 7, 9]$
PSTNN [45]	Patch size: 40×40, sliding step: 40, $\lambda = 0.7$
MGDWE [44]	Mean filter size: 7×7
Anisotropy [46]	$K = 120, step = 4, M = 40$
NTFRA [47]	Patch size: 40×40, sliding step: 40, $\beta = 0.01$, $\lambda = 0.6$
TLLDM [48]	Filter size: 15×15, $K = 3$
ELUM [50]	Filter size: 3×3, $K_1 = 5$, $K_2 = 3$
LGDC [49]	$K = 0.65$, $N = 3$, $\lambda = 0.5$
Proposed	Filter size: 3×3 and 5×5, $\sigma_d = 2$

the relationship between detection rate and false alarm rate. The relevant calculation equation is as follows [42]:

$$SSIM = \frac{(2\mu_R\mu_F + \varepsilon_1)(2\sigma_{RF} + \varepsilon_2)}{(\mu_R^2 + \mu_F^2 + \varepsilon_1)(\sigma_R^2 + \sigma_F^2 + \varepsilon_2)} \tag{14}$$

$$BSF = \frac{\sigma_{in}}{\sigma_{out}} \tag{15}$$

Equation (14), μ is the mean value, μ_R is the original image mean, μ_F is the predicted image mean. σ is the standard deviation, σ_{RF} is the covariance. ε_1 and ε_2 are constants. Equation (15), σ_{in} is the variance of the original image, σ_{out} is the variance of the difference image.

$$SNR = 10 \times \log_{10}^{(\mu_T - \mu_B)/\sigma_B} \tag{16}$$

Equation (16), μ_T is the target region mean, μ_B is the background region mean, and σ_B is the background region variance. The SNR calculated in this paper is the global SNR.

$$P_d = \frac{NTDT}{NT} \times 100\% \tag{17}$$

$$P_f = \frac{NFDT}{NP} \times 100\% \tag{18}$$

Equation (17), P_d is the detection rate, NT is the total number of targets in the sequence image, and $NTDT$ is the number of targets that can be detected. Equation (18), P_f is the false alarm rate, $NFDT$ represents the sum of false pixels, and NP is the sum of the total pixels of the sequence image.

3) Comparison Algorithms: Our algorithm is compared with 11 state-of-the-art algorithms, which are conventional bilateral filtering [35], TDLMS [43], MPCM [10], MGDWE [44], TLLCM [11], PSTNN [45], anisotropy [46], NTFRA [47], TLLDM [48], LGDC [49] and ELUM [50]. The relevant parameters of these algorithms are shown in Table III.

4) Analysis of Filter Window Scale Design: In the experiments, we constructed two filtering templates of 3*3 and 5*5, respectively, because the size of weak targets is generally about 3*3–5*5. However, it is found that the background modeling effect achieved by a single 3*3 or 5*5 filter template is insufficient, and a better background modeling effect will be achieved

Fig. 6. Comparison of SNR for different filter template scales.

if the background prediction results of two filter templates are fused. The following Fig. 6 shows the SNR comparison of the six sequence images at different scales of filtering templates, and the results show that the SNR obtained by the fusion of two filtering templates is better than that of a single filtering template, therefore the fusion results of the two filtering templates are chosen for the experiments in this paper.

B. Analysis of Background Modeling Effect

1) Qualitative Evaluation: To evaluate the background suppression effect of our algorithms, it is compared with 11 algorithms, which include classical background modeling algorithms such as bilateral filtering algorithms, two-dimensional least mean square filtering algorithms, and anisotropic background suppression algorithms; including local contrast measurement algorithms such as MPCM, MGDWE, TLLCM, LGDC, ELUM and TLLDM; and low rank and sparse theory algorithms such as PSTNN and NTFRA algorithms. The first row shows the background prediction images obtained by the different algorithms, the second row shows the difference maps corresponding to the original images, and the third row shows the 3D images corresponding to the difference maps.

Fig. 7 shows the comparison of the background modeling effect of scene 1. The traditional bilateral filtering achieves a better background modeling effect, but a small amount of clutter remains; TDLMS highlights the target area better, and the clutter suppression effect is not enough for the image edges, and also retains a large amount of background with lower intensity; MPCM and MGDWE algorithm have similar background suppression effect, although the target can be observed, most of the background energy is enhanced, and the background grayscale is higher than the target grayscale, which is caused by the uneven distribution of grayscale and the small scale of the target, the TLLCM algorithm has less clutter, the background suppression effect is better than MPCM and MGDWE algorithms; the PSTNN, LGDC and NTFRA algorithms have stronger suppression ability of background clutter, at the same time can retain the target information better, and the difference map contains less strong clutter; the anisotropy algorithm has better retention effect for the target, but the background suppression ability is not enough and the background interference is serious; the TLLDM and ELUM algorithm not only remove most of the background clutter, but also enhance the target area significantly.

Fig. 8 shows the comparison of the background modeling effect of scene 2. The traditional bilateral filtering, LGDC, PSTNN, NTFRA and TLLDM algorithms all show excellent background suppression ability, although the differential map still contains some clutter, but these clutter intensity is low and can be removed by simple threshold segmentation; the TDLMS obtained differential image has obvious targets, and the number of clutter is relatively less, but still cannot remove the interference for the image edges; while the MPCM, MGDWE, TLLCM and ELUM algorithms are close to the background suppression effect in this scene, and can clearly observe the target signal, and have a better enhancement effect on the target signal, but the number of clutter is large; the anisotropic algorithm has a weak background suppression ability in this scene, although it can retain the target information better, the background clutter interference is serious.

Fig. 9 shows the comparison of the background modeling effect of scene 3. Most of the algorithms achieve better background suppression effect, such as traditional bilateral filtering, PSTNN, ELUM and TLLDM algorithms obtain difference maps with almost no clutter, MPCM, TLLCM, LGDC and NTFRA algorithms also contain less background clutter, compared with the above two scenes, TDLMS, MGDWE and anisotropic algorithms also achieve better background modeling effect or enhancement effect, but the background interference is still serious compared with other algorithms.

Fig. 10 shows the comparison of the background modeling effect of scene 4, which has a relatively smooth background, high target contrast and fewer cloud edges, so most algorithms also obtain ideal background suppression effect in this scene, the background suppression effect of TLLDM and LGDC algorithm is better than other comparison algorithms, the traditional bilateral filtering and PSTNN algorithms are second, MPCM, TLLCM, ELUM and NTFRA algorithms also highlight the target clearly. The background suppression effect of TDLMS, MGDWE and anisotropic algorithms is insufficient, and all have different degrees of noise interference.

Fig. 11 shows the comparison of the background modeling effect of scene 5. In this scene, the background is smooth, but there are a few edge contours. Although TDLMS, MPCM, MGDWE and TLLCM retain the target information better, due to the limitations of the algorithms, the difference maps obtained by these algorithms also have more background interference, which will have a certain impact on the subsequent detection process. Other comparison algorithms achieve ideal background modeling effect, not only retaining the target information, but also effectively suppressing the background clutter.

Fig. 12 shows the comparison of the background modeling effect of scene 6. Similar to scene 5, the background is smooth, but the target is not obvious enough, while the background moves quickly with the target. Except for TDLMS and TLLCM with serious background interference, other algorithms achieve effective background suppression even though a small amount of clutter remains, and the target energy is lost after the

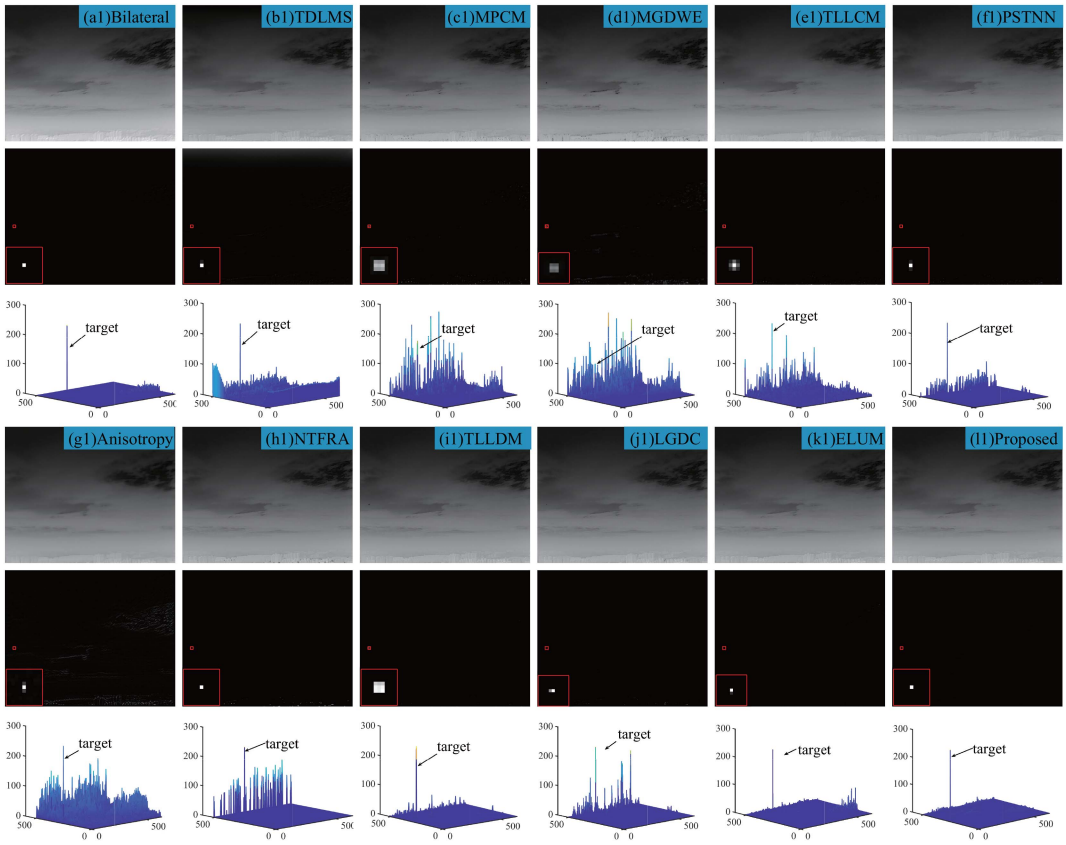

Fig. 7. Background modeling effect of scene 1.

TABLE IV
SCENE 1 EVALUATION INDICATORS

	Proposed	Bilateral [35]	TDLMS [43]	MPCM [10]	TLLCM [11]	PSTNN [45]	Anisotropy [46]	NTFRA [47]	TLLDM [48]	MGDWE [44]	LGDC [49]	ELUM [50]
SSIM	**0.9999**	0.942	0.9909	0.9986	0.9992	0.9997	0.9993	0.9987	**0.9999**	0.9965	0.9996	**0.9999**
BSF	**783.88**	94.448	118.9	180.34	252.77	385.7	291.94	192.38	381.13	116.47	359.07	605.50

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

TABLE V
SCENE 2 EVALUATION INDICATORS

	Proposed	Bilateral [35]	TDLMS [43]	MPCM [10]	TLLCM [11]	PSTNN [45]	Anisotropy [46]	NTFRA [47]	TLLDM [48]	MGDWE [44]	LGDC [49]	ELUM [50]
SSIM	**0.9996**	0.9208	0.9865	0.9985	0.9993	0.9995	0.9977	0.9994	0.9995	0.9960	0.9995	0.9989
BSF	**315.11**	296.9	79.07	137.52	230.43	294.3	149.51	281.69	245.72	90.0408	300.50	134.9602

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

background suppression of MPCM, MGDWE, and ELUM algorithm. Compared with other algorithm, the proposed algorithms can achieve better background modeling effect in all six scenes and has a strong background suppression ability, the number of background clutter in the difference map is less.

2) Quantitative Evaluation: Two evaluation indexes, BSF and SSIM, are selected to evaluate the background modeling effect of the algorithms. Table IV shows the comparison of algorithm evaluation metrics for scene 1. It can be seen that the SSIM of the TLLDM and ELUM algorithm is higher than the other comparison algorithms, while the BSF of the ELUM algorithm is higher than the other comparison algorithms. The SSIM of the proposed algorithm is the same as the TLLDM and ELUM algorithm, but the BSF are all higher than those of the other comparison algorithms, which is better compared with the traditional bilateral filtering algorithm.

Table V shows the comparison of algorithm evaluation indexes for scene 2. In this scene, the SSIM of the TLLDM algorithm is still higher than the other comparison algorithms, while the BSF of the traditional bilateral filtering algorithm

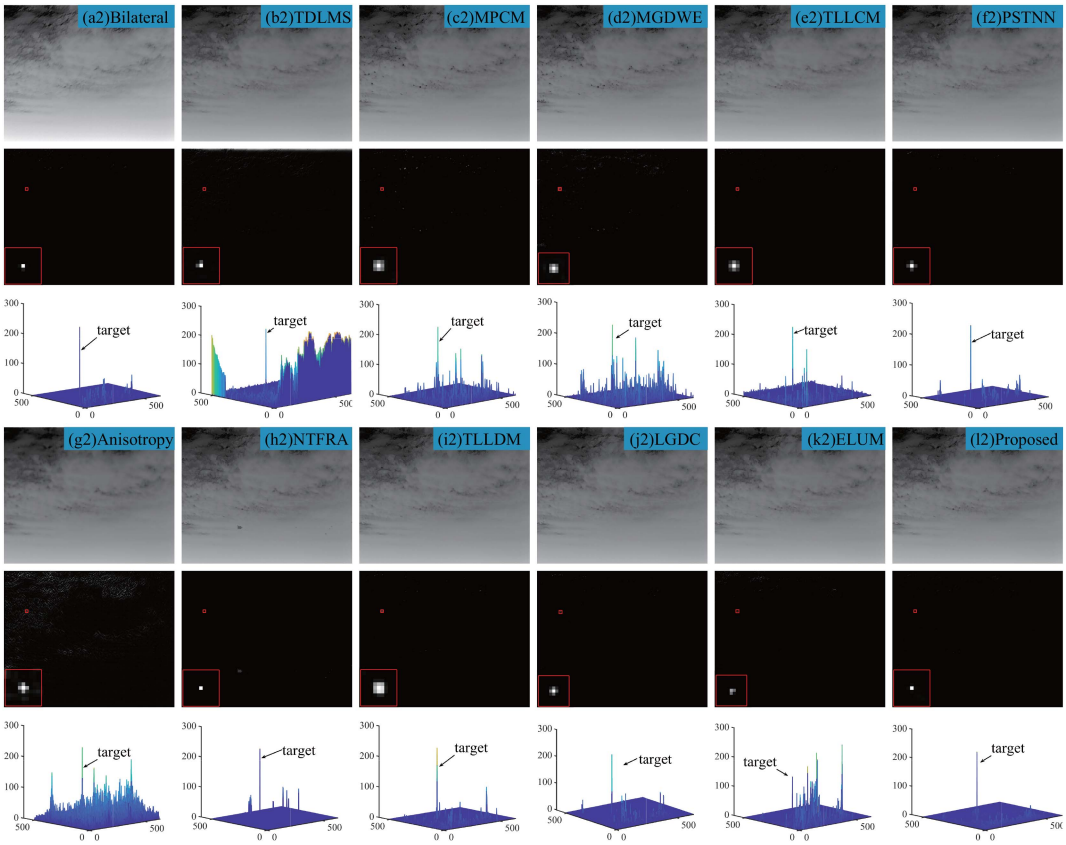

Fig. 8. Background modeling effect of scene 2.

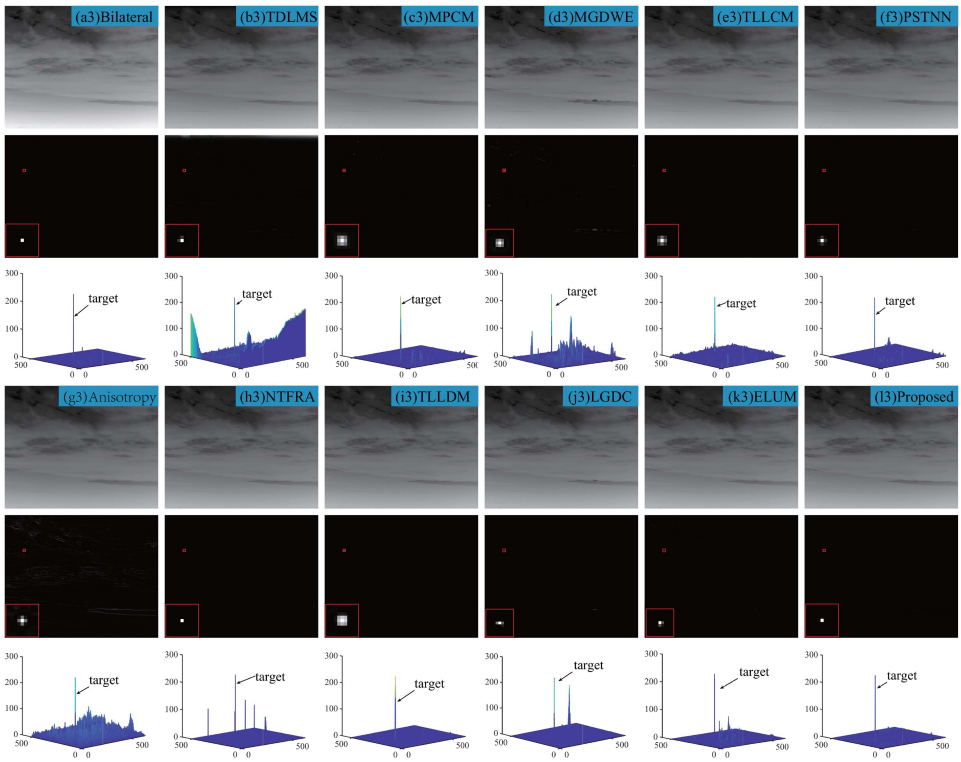

Fig. 9. Background modeling effect of scene 3.

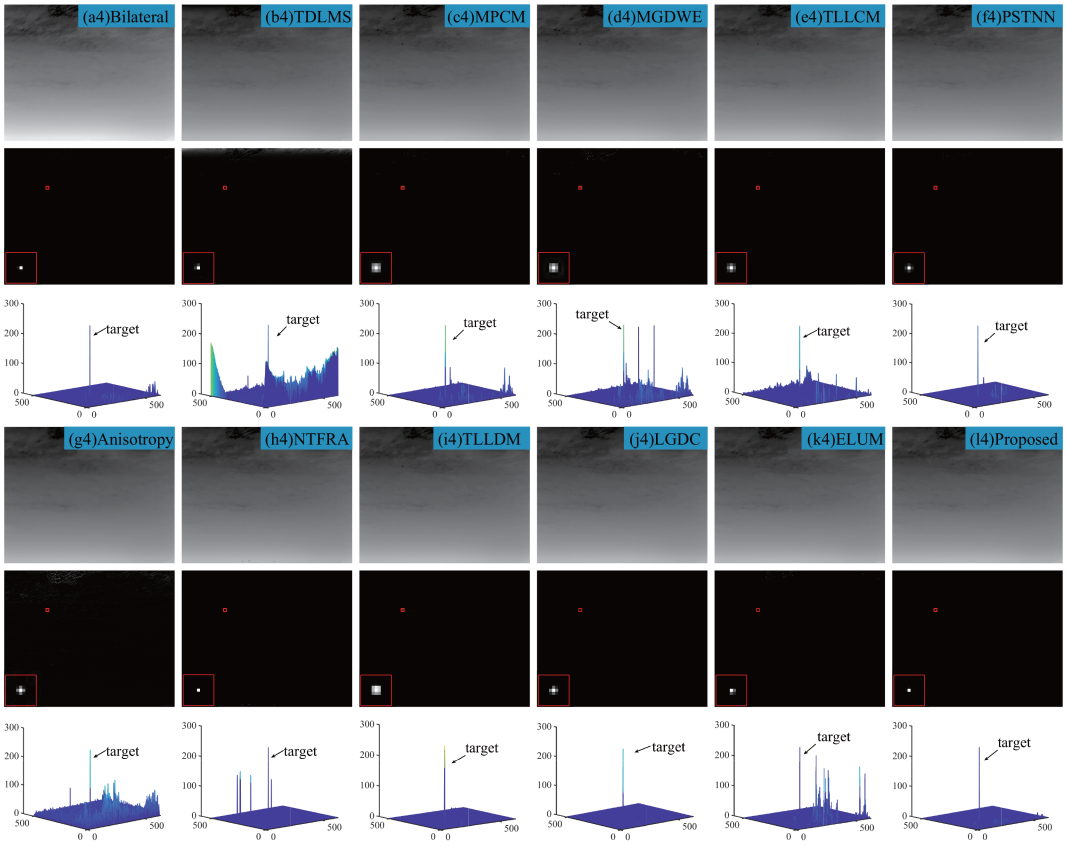

Fig. 10. Background modeling effect of scene 4.

TABLE VI
SCENE 3 EVALUATION INDICATORS

	Proposed	Bilateral [35]	TDLMS [43]	MPCM [10]	TLLCM [11]	PSTNN [45]	Anisotropy [46]	NTFRA [47]	TLLDM [48]	MGDWE [44]	LGDC [49]	ELUM [50]
SSIM	**0.9999**	0.9248	0.9883	0.9998	0.9998	0.9998	0.9992	**0.9999**	**0.9999**	0.9982	0.9997	0.9998
BSF	582.91	**694.16**	88.83	357.36	487.71	534.34	268.7	575.21	372.49	153.00	390.19	404.66

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

TABLE VII
SCENE 4 EVALUATION INDICATORS

	Proposed	Bilateral [35]	TDLMS [43]	MPCM [10]	TLLCM [11]	PSTNN [45]	Anisotropy [46]	NTFRA [47]	TLLDM [48]	MGDWE [44]	LGDC [49]	ELUM [50]
SSIM	**0.9999**	0.9478	0.988	0.9997	0.9997	0.9998	0.9994	0.9998	**0.9999**	0.9992	**0.9999**	0.9996
BSF	**670.97**	447	91.51	309.99	379.05	507.34	287.84	463.52	391.36	179.349	661.69	209.13

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

is higher than the other comparison algorithms. Although the difference of the BSF between our algorithm and the ELUM algorithm is small, the SSIM is much higher than the ELUM algorithm.

Table VI shows the comparison of algorithm evaluation indexes for scene 3. Most of the comparison algorithms achieve better background modeling effect in this scene, such as the SSIM of NTFRA and TLLDM reaches 0.9999. Although the SSIM of traditional bilateral filtering algorithm is lower, its BSF is higher than the other comparison algorithms including our algorithm. The BSF of our algorithm is lower than the traditional bilateral filtering algorithm, it is still higher than the other algorithms, and the SSIM is also higher than the traditional bilateral filtering.

Table VII shows the comparison of algorithm evaluation indexes for scene 4. In this scene, all the comparison algorithms, except the traditional bilateral filtering algorithm, achieve high SSIM, these algorithms achieve the desired background modeling effect. For the BSF, all the comparison algorithms obtain high BSF except for the TDLMS algorithm, which has a low BSF. Both the SSIM and BSF of the proposed algorithm are higher than the other comparison algorithms.

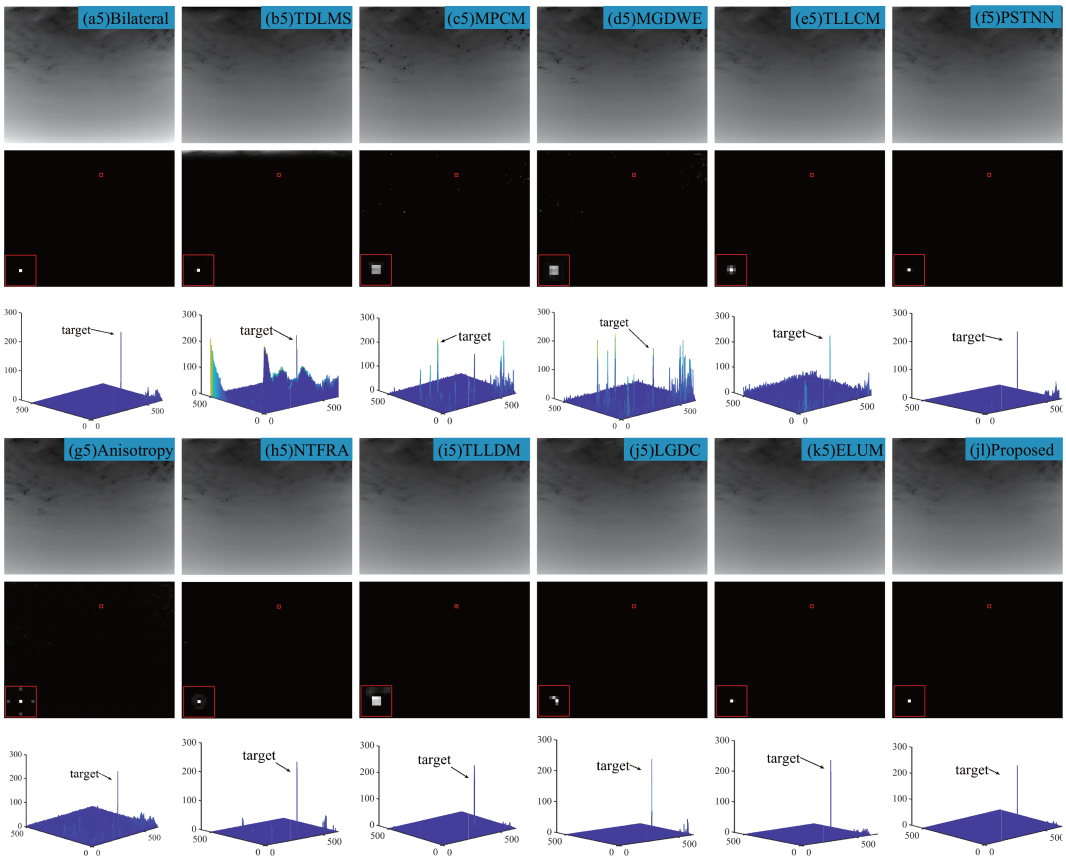

Fig. 11. Background modeling effect of scene 5.

TABLE VIII
SCENE 5 EVALUATION INDICATORS

	Proposed	Bilateral [35]	TDLMS [43]	MPCM [10]	TLLCM [11]	PSTNN [45]	Anisotropy [46]	NTFRA [47]	TLLDM [48]	MGDWE [44]	LGDC [49]	ELUM [50]
SSIM	**0.9999**	0.9460	0.9884	0.9991	0.9993	**0.9999**	0.9978	0.9998	**0.9999**	0.9980	**0.9999**	**0.9999**
BSF	737.19	450.29	31.24	155.93	244.49	606.13	165.43	472.21	331.49	110.85	620.55	**741.47**

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

TABLE IX
SCENE 6 EVALUATION INDICATORS

	Proposed	Bilateral [35]	TDLMS [43]	MPCM [10]	TLLCM [11]	PSTNN [45]	Anisotropy [46]	NTFRA [47]	TLLDM [48]	MGDWE [44]	LGDC [49]	ELUM [50]
SSIM	**0.9999**	0.9465	0.9880	**0.9999**	0.9981	0.9998	0.9972	0.9998	0.9998	0.9997	**0.9999**	0.9998
BSF	**504.71**	280.53	24.84	217.72	156.16	483.36	143.40	438.93	334.62	180.60	380.48	295.16

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

Table VIII shows the comparison of algorithm evaluation indexes of scene 5. In this scene most of the algorithms have better evaluation indexes, but some data is relatively low, such as TDLMS algorithm. In this scene, the SSIM of PSTNN, TTLDM, LGDC and ELUM algorithms reaches 0.9999, which is equal to the proposed algorithm. Meanwhile, the BSF of the algorithm in this paper is slightly lower than that of ELUM algorithm, but higher than that of other compared algorithms.

Table IX shows the comparison of algorithm evaluation indexes of scene 6. In this scene, the SSIM of most algorithms is more than 0.99, MPCM, and LGDC even reaches 0.9999, which is the same as the SSIM of our algorithm, but in terms of BSF, the proposed algorithm is higher than the other comparison algorithms. On the whole, the algorithm in this paper has better background modeling effect and strong background suppression ability.

C. Analysis of Energy Enhancement Effect

In the process of acquiring the difference image, some of the target energy is weakened, so it is necessary to improve the contrast of the target. To avoid increasing the execution time,

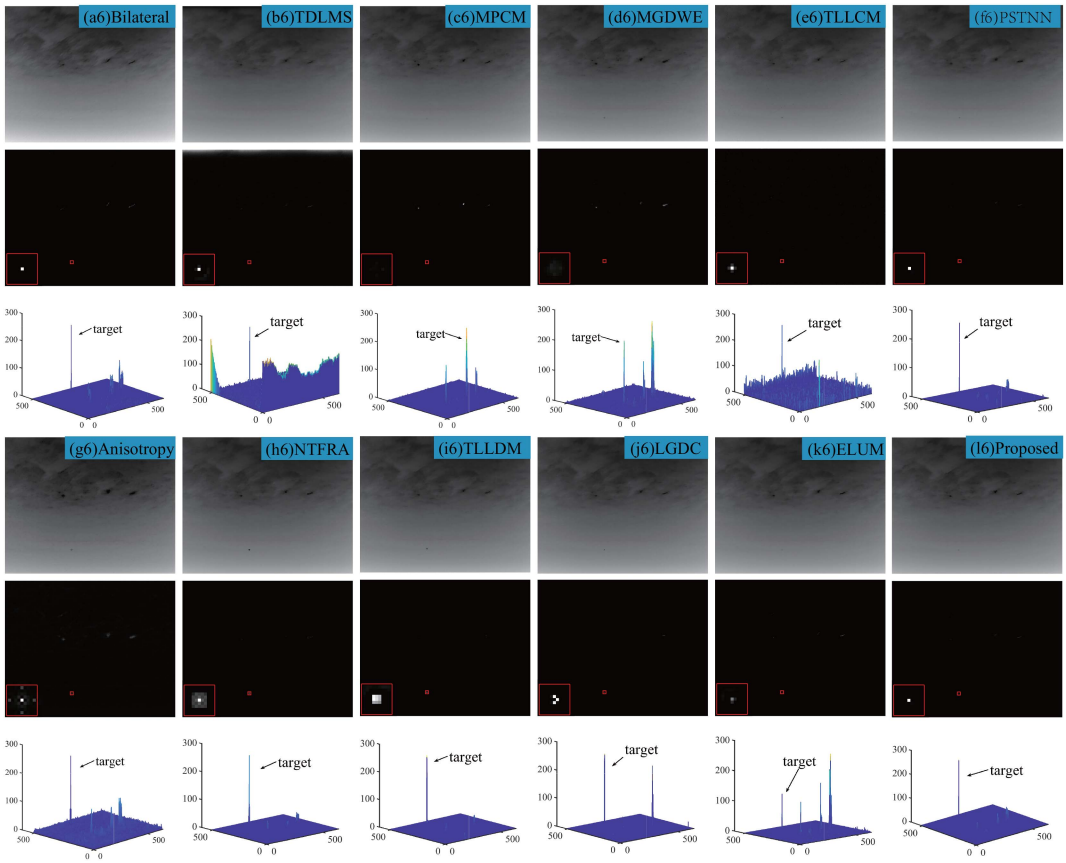

Fig. 12. Background modeling effect of scene 6.

a double window segmentation algorithm is used to extract the candidate targets. Because the target and the surrounding background have certain differences in grayscale, most of the background clutter interference can be removed by these features, thus retaining some points of interest that are closer to the grayscale features of the target. The coordinate of the new target is predicted by its historical motion coordinate, and then the historical energy of corresponding candidate target is accumulated to the new predicted target. The method fully combines the spatio-temporal domain information of the target to improve the target energy, and it can be seen from Fig. 13 that the grayscale of the target is low before enhancement, and the grayscale value is significantly increased after the enhancement. In scene 1, the target has the highest grayscale value before enhancement, and the target grayscale is lower than the background clutter grayscale after enhancement, because this paper takes the grayscale information of consecutive multi-frame images for enhancement, in which the background region is larger and the target region is smaller, and the magnitude of background energy enhancement is higher than that of target energy enhancement in the process of energy accumulation. In contrast, scene 2 and scene 3 do not appear similar cases. In scene 4, although some candidate target energies are also enhanced, the target grayscale is still the highest. The energy enhancement effect of scene 5 and scene 6 energy enhancement effects are similar, there is a small amount of noise in addition to the target before the energy enhancement, but the energy enhancement effect of the noise is less pronounced relative to the target, whereas the energy enhancement effect of the target is obvious. In order to objectively evaluate the enhancement effect of the proposed algorithm. Target area average grayscale and SNR are used to compare the effect of the target before and after enhancement, see Table X. The target area average grayscale of the six scenes before target enhancement are 0.5556, 4.8889, 5.2222, 6.3333, 1.2222 and 5.5556,respectively, after enhancement they reach 55.7778, 81.2222, 75.0000, 70.3333, 82.4444 and 107.8889, respectively. The SNR are −6.70, 2.93, 3.22, 4.06, 17.86 and 17.83 for the six scenes before target enhancement, and 13.36, 15.05, 14.72, 14.39, 21.12 and 22.16 after enhancement, which indicate that the energy enhancement algorithm proposed in this paper has better target energy enhancement effect.

D. Detection Results

After completing the clutter suppression and energy enhancement, the multi-frame correlation detection is implemented by the energy probability model proposed in this paper to determine the real target and extract its corresponding motion trajectory. To verify the effectiveness of the detection algorithm in this paper, we compare the detection results with 11 detection algorithms, including bilateral filtering algorithm, TDLMS algorithm, MPCM algorithm, MGDWE algorithm,

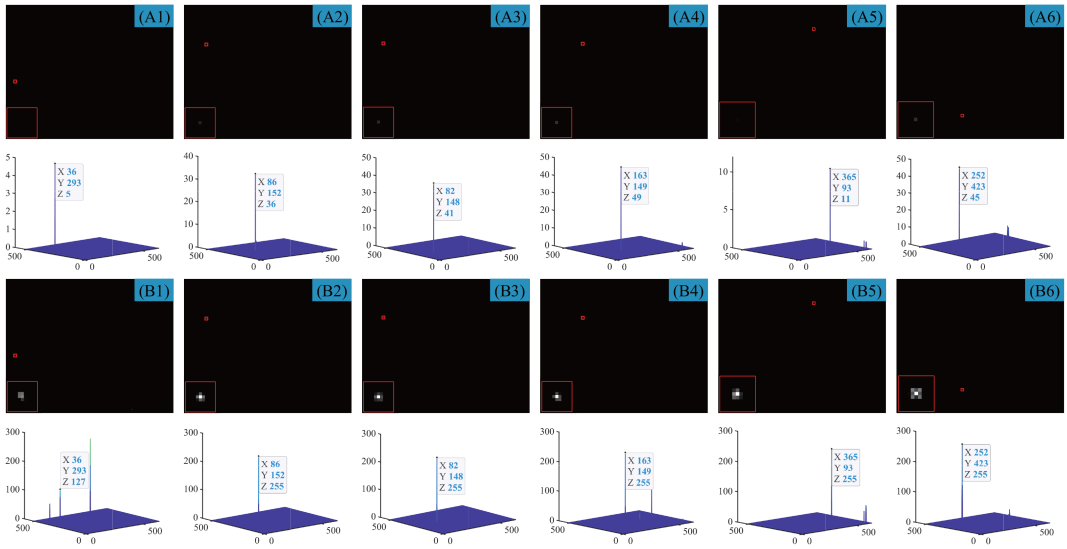

Fig. 13. Target enhancement effect.

TABLE X
TARGET ENHANCEMENT EVALUATION INDICATORS

		Scene 1	Scene 2	Scene 3	Scene 4	Scene 5	Scene 6
Average grayscale	Before	0.5556	4.8889	5.2222	6.3333	1.2222	5.5556
	After	55.7778	81.2222	75.0000	70.1111	82.4444	107.8889
SNR(dB)	Before	-6.7	2.93	3.22	4.06	17.86	17.83
	After	13.36	15.05	14.72	14.39	21.12	22.16

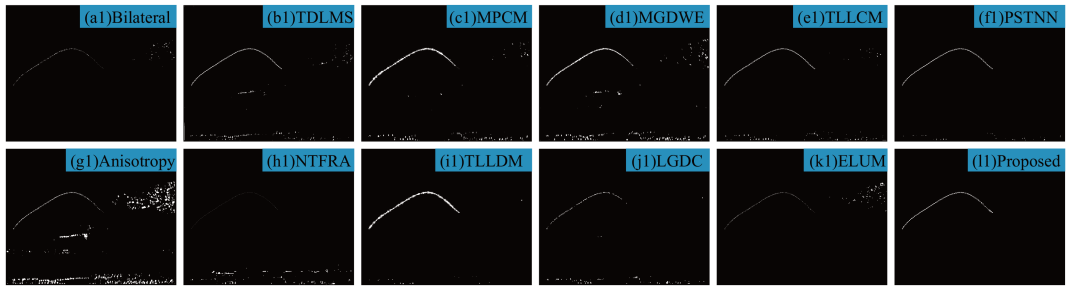

Fig. 14. Detection result of scene 1.

TLLCM algorithm, PSTNN algorithm, Anisotropy algorithm, NTFRA algorithm, LGDC algorithm, ELUM algorithm and TLLDM algorithm.

As can be seen from Fig. 14, most of the algorithms can locate the target in scene 1, but there are different degrees of miss detection and more clutter interference. TLLDM algorithm and PSTNN algorithm have better detection performance among all compared algorithms with less clutter, followed by bilateral filtering algorithm, TDLMS, MPCM, MGDWE, TLLCM and ELUM algorithms. However, these algorithms have different degrees of miss detection and more strong clutter interference, anisotropy can not completely remove the background clutter, LGDC and NTFRA can accurately locate the target, but also has a miss detection phenomenon. In contrast, the algorithm in this paper can effectively eliminate the noise interference and extract the target accurately.

From the Fig. 15, we can see that the TDLMS, TLLCM, PSTNN and TLLDM algorithms achieve better detection results, and the extracted target motion trajectories are relatively complete, despite less interference. In contrast, the bilateral filtering algorithm, MPCM, MGDWE,LGDC and NTFRA algorithms miss detection is more seriously, the extracted target motion trajectories are incomplete. The anisotropic and ELUM algorithm also reside more background clutter, and it can be seen that part of the target trajectory is covered by the background clutter. Our algorithm achieves better detection effect in this scene.

Fig. 16 shows the comparison of the detection results of scene 3. Since the background of this scene changes rapidly, the obvious background motion trajectory is visible in the comparison algorithm, but most algorithms can still detect the target and extract the target motion trajectory completely. The difference

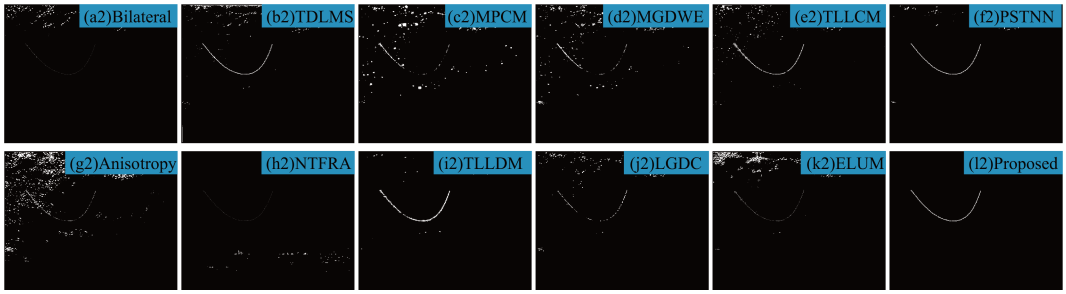

Fig. 15. Detection result of scene 2.

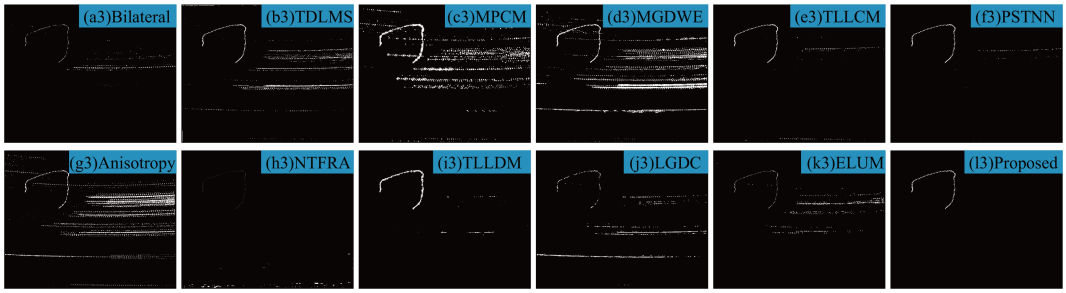

Fig. 16. Detection result of scene 3.

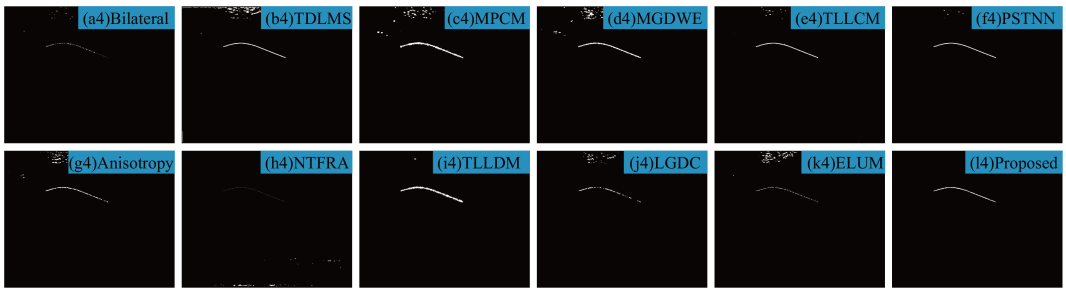

Fig. 17. Detection result of scene 4.

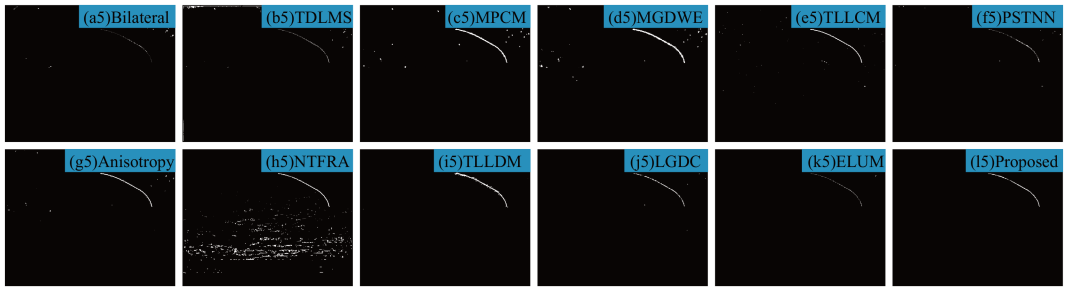

Fig. 18. Detection result of scene 5.

is that the energy probability model of the algorithm in this paper can eliminate the false target, the motion trajectory of the background can be removed.

Fig. 17 shows the comparison of the detection results of scene 4, where the background is relatively smooth and moves slowly, so it can be seen that all algorithms can obtain ideal detection results in this scene.

Fig. 18 shows the comparison of the detection results of scene 5. All algorithms in this scene detect the target signal, but the false alarm rate of NTFRA is high, and the other

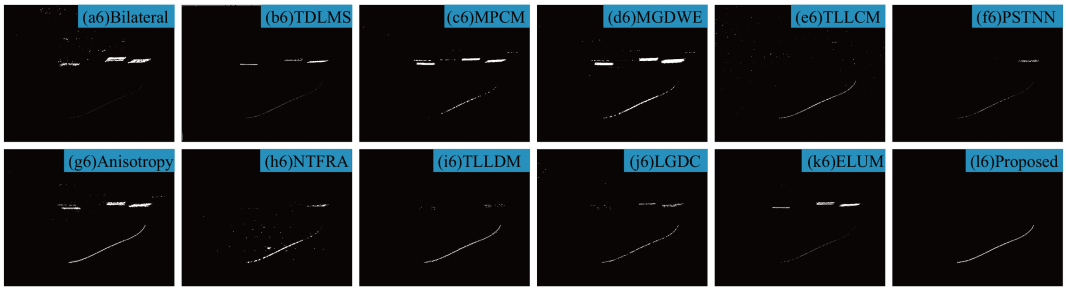

Fig. 19. Detection result of scene 6.

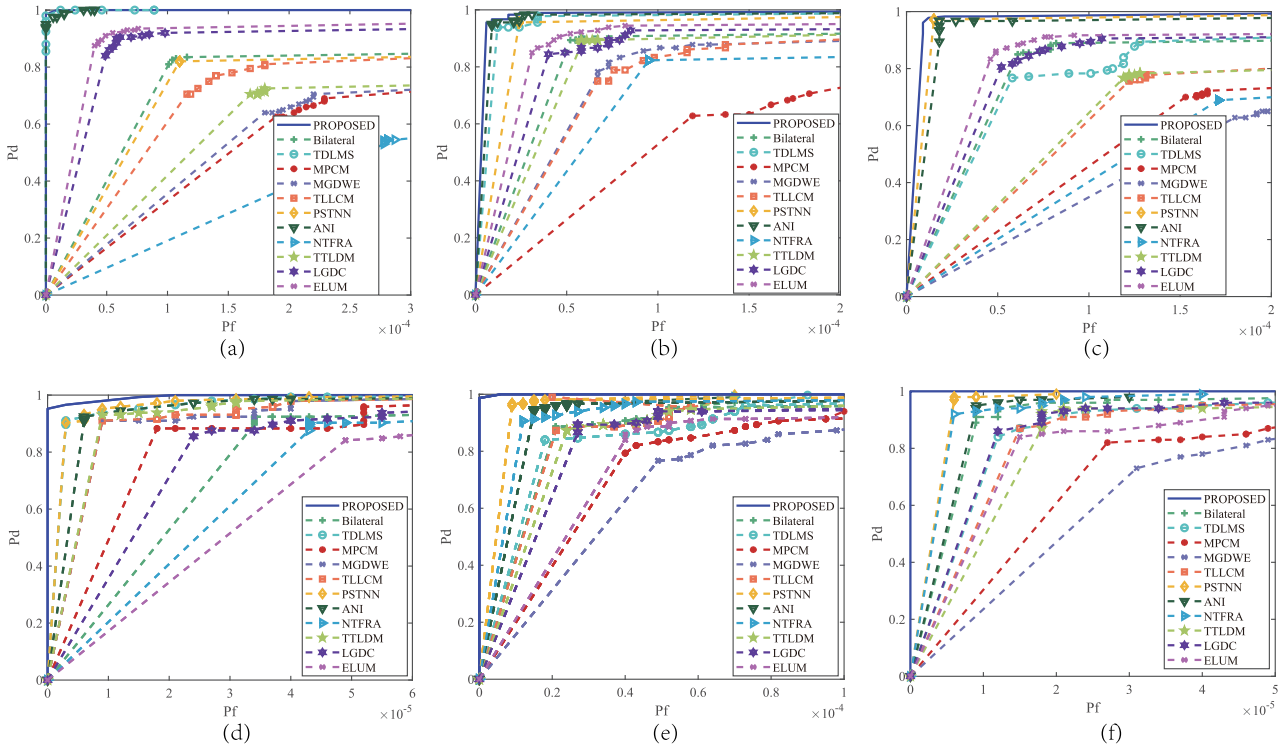

Fig. 20. Comparison of ROC curves.

comparison algorithms also have a small amount of false alarms.

Fig. 19 shows the comparison of the detection effect of scene 6. In this scene, the detection effect is less satisfactory because the background moves faster and there are a few small patches similar to the target in the image interfering with the detection process, although the target is successfully detected, there is still a high false alarm rate and missed detection rate. Analysis of the above experimental results shows that our algorithm utilizes the motion characteristics of the target motion and the grayscale features, incorporates the spatio-temporal domain information of the target, and is able to effectively remove background clutter and successfully detect the target even in scenes where the target and the background are moving rapidly.

E. ROC Curve Analysis

In the previous background modeling and energy enhancement sections, three evaluation metrics, SNR, SSIM and BSF, were used to evaluate the background clutter suppression ability of the algorithm. However, dim and small target detection is more concerned with the detection rate and correctness of the algorithm, in the detection section, detection rate and false alarm rate are used to evaluate the detection effectiveness, and the six ROC plots in Fig. 20 correspond to the six scenes selected in the paper. From Fig. 20(a), we can see that the detection rate of most of the comparison algorithms more than 70%, while the detection rate of the proposed algorithm is higher than other comparison algorithms and the false alarm rate is lower than other algorithms; from Fig. 20(b), we can see that the detection rate of all comparison algorithms exceeds 70%,

and most of them more than 85%, the detection rate of our algorithm is the highest; from Fig. 20(c) the detection rate of all algorithms exceeds 60%, and the detection rate and false alarm rate of this paper's algorithm are better than other comparison algorithms; in Fig. 20(d) it can be seen that the detection rate of all algorithms exceeds 85%, and the false alarm rate of the proposed algorithm is the lowest, despite the small difference in the false alarm rate of each algorithm; in Fig. 20(e) and (f), most of algorithms exceeds 80%, the proposed methods in this paper still exhibit an excellent performance in detection rate and false alarms rate. It demonstrates the detection efficiency of the algorithm proposed in this paper is high and has some feasibility.

IV. Discussion

Analysis of the above experimental results shows that the proposed algorithm achieves better background modeling and detection results with other comparison algorithms, but there are still some limitations of these algorithms, such as multi-edge contour scenes with low SNR, and the grayscale difference between the target and the edge contour is small, which can easily lead to detection failure. Therefore, this section focuses on the applicability and limitations of these algorithms.

The traditional bilateral filtering algorithm may lead to the loss of some detailed information when smoothing the image, which may affect the features of the target, thus reducing the detection rate of target detection. The TDLMS filter has a small error between the expected and predicted values in smooth background, and the algorithm converges faster. In the complex background, the error between the expected and predicted values at the edge contours is larger, and the algorithm has difficulty converging, and more clutter remains in the difference map, which leads to a higher false alarm rate. If the difference between the target and the background in the IR image background is small, the anisotropic diffusion filter may not be able to distinguish significantly between the target and the background, thus limiting the effectiveness of detection of weak targets, the selection of the step size is also crucial. In multi-edge contour scenes, too large or too small a step size can leave more background clutter. The MPCM algorithm is unable to highlight the interference of the background in scenes with dim and small target, and enhancing the target also enhances the background, resulting in a higher false alarm rate, despite it uses multi-scale windows to accommodate different target scales. In infrared images, the texture and grayscale distributions of the target and the background may be very similar, when the difference of the local entropy between the target and the background is small, which may lead to difficulties for the MGDWE algorithm to distinguishing the target from the background. In particular, the detection rate of the algorithm may be affected by the small target scales. Both the TLLCM algorithm and the TLLDM algorithm use a three-layer local window, the TLLCM detects the target by local contrast, the TLLDM identifies the target based on the ratio difference of local contrast. Therefore, its effect is better than the TLLCM algorithm, but both algorithms are prone to false detection and enhance the intensity of background clutter in a bright background. LGDC can enhance the contrast between the target and the background, in complex backgrounds with numerous regions exhibiting similar gradient and curvature characteristics, background textures may interfere with target detection. Consequently, the LGDC method may be unable to effectively distinguish the target from the background. Although the ELUM utilizes local uncertainty measurements to suppress complex backgrounds, the algorithm's robustness and stability may be insufficient when dealing with highly variable background features or high levels of noise, leading to false alarms or missed detections. The PSTNN algorithm is weak in suppressing strong edge contour backgrounds during tensor decomposition and low-rank approximation in some scenes, which leads to low robustness of the algorithm. NTFRA algorithms usually require high computational complexity, because the solution of nonconvex optimization problems often requires the use of iterative algorithms and may require significant computational resources and time to reach convergence.

The bilateral filtering based on the neighborhood block level proposed in this algorithm can smooth the image better because it considers both the similarity and spatial relationship of pixels within the neighborhood patch. Compared with between single pixel, the neighborhood patch can provide greater contextual information, thus producing better background modeling results. The Gaussian motion estimation energy enhancement algorithm can effectively compensate for the loss of target energy caused by background modeling, and the effective displacement probability effectively excludes noise interference and extracts the real target. However, the neighborhood patch filtering may encounter challenges when dealing with large targets. Due to the local nature of bilateral filtering, it may not be able to accurately handle large targets spanning multiple neighborhood blocks within a single neighborhood block. In addition, bilateral filtering at the neighborhood block level involves more pixels and more complex computations than traditional bilateral filtering at the individual pixel level. This may lead to increase computational complexity, which is an area for improvement in our next work.

V. Conclusion

A dim and small target detection method for sequence images is proposed in the paper, which consists of three parts. Firstly, the difference image is acquired through a spatio-temporal filtering method, then the target signal intensity should be improved, in order to reduce the running time during the energy enhancement, a threshold segmentation algorithm is used to further remove the background clutter. The new target position is predicted by Gaussian process based on the historical position data of the target, and the target energy is accumulated along the trajectory direction. Finally, to realize the correlation detection of targets between consecutive frames, a grayscale probability model is constructed in this paper, which can effectively reject noise and extract the real target. After a comparative analysis with other detection algorithms, the following conclusions are drawn:

1) Firstly, this paper fully considers the neighborhood information of the target and constructs two different scale filter windows to achieve a better background modeling

effect. The SSIM of six scenes are 0.9999, 0.9996, 0.9999, 0.9999, 0.9999 and 0.9999, the BSF are 783.88, 315.11, 582.91, 670.97, 737.19 and 504.71, respectively.

2) Then, the target energy accumulation was achieved by using the historical motion information and grayscale information of the target. And the SNR and the average grayscale of the target area were selected to evaluate the energy enhancement effect. The average grayscale of the six scenes before the energy enhancement was 0.5556, 4.8889, 5.2222, 6.3333, 1.2222 and 5.5556, respectively. After the enhancement was 55.7778, 81.2222, 75.0000, 70.1111, 82.4444 and 107.8889, respectively. The SNR are −6.70, 2.93, 3.22, 4.06, 17.86 and 17.83 for the six scenes before target enhancement, and 13.36, 15.05, 14.72, 14.39, 21.12 and 22.16 after enhancement, respectively.
3) Finally, the detection of weak targets is achieved by the target energy probability model, and the detection rate of the algorithm in this paper reaches more than 90% in all six scenes.

In summary, the algorithm proposed in this paper is an effective and robust weak target detection method. However, the proposed method also has some limitations, such as the processing time of background modeling method based on image patches is long and the diminished effectiveness when handling large targets spanning multiple neighboring blocks. Based on Gaussian estimation, the energy enhancement method exhibits better adaptability. But the inherent limitations of Gaussian process prediction, significant prediction errors occur after a certain number of frames, leading to inaccurate prediction. And the detection effect is depending on the pre-processeffect. Therefore, it is necessary to optimize the execution efficiency and refine the structure of our algorithm in future work.

REFERENCES

[1] Y. Yang, C. Xu, Y. Ma, and C. Huang, "A review of infrared dim small target detection algorithms with low SNR," *Laser Infrared*, vol. 49, no. 6, pp. 643–649, 2019.

[2] J. Han et al., "Infrared dim and small target detection: A review," *Infrared Laser Eng.*, vol. 51, no. 4, 2022, Art. no. 20210393.

[3] J. Hu, Y. Yu, and F. Liu, "Small and dim target detection by background estimation," *Infrared Phys. Technol.*, vol. 73, pp. 141–148, 2015.

[4] Q. Song, Y. Wang, K. Dai, and K. Bai, "Single frame infrared image small target detection via patch similarity propagation based background estimation," *Infrared Phys. Technol.*, vol. 106, 2020, Art. no. 103197.

[5] J. Han, C. Liu, Y. Liu, Z. Luo, X. Zhang, and Q. Niu, "Infrared small target detection utilizing the enhanced closest-mean background estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 645–662, 2021.

[6] J. Han, S. Liu, G. Qin, Q. Zhao, H. Zhang, and N. Li, "A local contrast method combined with adaptive background estimation for infrared small target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 9, pp. 1442–1446, Sep. 2019.

[7] G. Wang, B. Tao, X. Kong, and Z. Peng, "Infrared small target detection using nonoverlapping patch spatial–temporal tensor factorization with capped nuclear norm regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2021, Art. no. 5001417.

[8] L. Yang, J. Yang, and K. Yang, "Adaptive detection for infrared small target under sea-sky complex background," *Electron. Lett.*, vol. 40, pp. 1083–1085, 2004.

[9] X. Wang and Z. Tang, "Combining wavelet packets with higher-order statistics for infrared small targets detection," *Infrared Laser Eng.*, vol. 38, no. 5, pp. 915–920, 2009.

[10] 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.

[11] 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.

[12] L. Ren, Z. Pan, and Y. Ni, "Double layer local contrast measure and multi-directional gradient comparison for small infrared target detection," *Optik*, vol. 258, 2022, Art. no. 168891.

[13] Z. Lu, Z. Huang, Q. Song, H. Ni, and K. Bai, "Infrared small target detection based on joint local contrast measures," *Optik*, vol. 273, 2023, Art. no. 170437.

[14] R. Kou, C. Wang, Q. Fu, Y. Yu, and D. Zhang, "Infrared small target detection based on the improved density peak global search and human visual local contrast mechanism," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6144–6157, 2022.

[15] S. Xiao, Z. Peng, and F. Li, "Infrared cirrus detection using non-convex rank surrogates for spatial-temporal tensor," *Remote Sens.*, vol. 15, no. 9, 2023, Art. no. 2334.

[16] S. Cao, J. Deng, J. Luo, Z. Li, J. Hu, and Z. Peng, "Local convergence index-based infrared small target detection against complex scenes," *Remote Sens.*, vol. 15, no. 5, 2023, Art. no. 1464.

[17] P. Yang, L. Dong, and W. Xu, "Infrared small maritime target detection based on integrated target saliency measure," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 2369–2386, 2021.

[18] Y. J. He, M. Li, J. Zhang, and Q. An, "Small infrared target detection based on low-rank and sparse representation," *Infrared Phys. Technol.*, vol. 68, pp. 98–109, 2015.

[19] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, and A. G. Hauptmann, "Infrared patch-image model for small target detection in a single image," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4996–5009, Dec. 2013.

[20] Y. Dai, Y. Wu, F. Zhou, and K. Barnard, "Attentional local contrast networks for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9813–9824, Nov. 2021.

[21] C. Yu et al., "Pay attention to local contrast learning networks for infrared small target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 3512705.

[22] T. Zhang, L. Li, S. Cao, T. Pu, and Z. Peng, "Attention-guided pyramid context networks for detecting infrared small target under complex background," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 4, pp. 4250–4261, Aug. 2023.

[23] Q. Hou, Z. Wang, F. Tan, Y. Zhao, H. Zheng, and W. Zhang, "RISTDNet: Robust infrared small target detection network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2021, Art. no. 7000805.

[24] C. Yu et al., "Infrared small target detection based on multiscale local contrast learning networks," *Infrared Phys. Technol.*, vol. 123, 2022, Art. no. 104107.

[25] Q. Shi, C. Zhang, Z. Chen, F. Lu, L. Ge, and S. Wei, "An infrared small target detection method using coordinate attention and feature fusion," *Infrared Phys. Technol.*, vol. 131, 2023, Art. no. 104614.

[26] X. Yang, Y. Zhou, D. Zhou, R. Yang, and Y. Hu, "A new infrared small and dim target detection algorithm based on multi-directional composite window," *Infrared Phys. Technol.*, vol. 71, pp. 402–407, 2015.

[27] K. Huang and X. Mao, "Detectability of infrared small targets," *Infrared Phys. Technol.*, vol. 53, no. 3, pp. 208–217, 2010.

[28] J. Shaik and K. M. Iftekharuddin, "Detection and tracking of targets in infrared images using Bayesian techniques," *Opt. Laser Technol.*, vol. 41, no. 6, pp. 832–842, 2009.

[29] Z. Chen, T. Deng, L. Gao, H. Zhou, and S. Luo, "A novel spatial–temporal detection method of dim infrared moving small target," *Infrared Phys. Technol.*, vol. 66, pp. 84–96, 2014.

[30] X. Ren, J. Wang, T. Ma, K. Bai, M. Ge, and Y. Wang, "Infrared dim and small target detection based on three-dimensional collaborative filtering and spatial inversion modeling," *Infrared Phys. Technol.*, vol. 101, pp. 13–24, 2019.

[31] B. Li, X. Zhiyong, J. Zhang, X. Wang, and X. Fan, "Dim-small target detection based on adaptive pipeline filtering," *Math. Problems Eng.*, vol. 2020, 2020, Art. no. 8234349.

[32] X. Dong, X. Huang, Y. Zheng, S. Bai, and W. Xu, "A. novel infrared small moving target detection method based on tracking interest points under complicated background," *Infrared Phys. Technol.*, vol. 65, pp. 36–42, 2014.

[33] L. Liu and Z. Huang, "Infrared dim target detection technology based on background estimate," *Infrared Phys. Technol.*, vol. 62, pp. 59–64, 2014.

[34] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. IEEE 6th Int. Conf. Comput. Vis.*, 1998, pp. 839–846.
[35] 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.
[36] H. Jiang, "Infrared dim target detection in complex scene," Master's thesis, Southeast Univ., Nanjing, Jiangsu, China, 2018.
[37] J. Chen, "Infrared small target detection," Master's thesis, Southeast Univ., Nanjing, Jiangsu, China, 2016.
[38] J. S. Shaik and K. M. Iftekharuddin, "Probabilistic detection and tracking of IR targets," *Proc. SPIE*, vol. 5556, pp. 90–101, 2004.
[39] W. Aftab and L. Mihaylova, "A Gaussian process regression approach for point target tracking," in *2019 IEEE 22th Int. Conf. Inf. Fusion*, 2019, pp. 1–8.
[40] H. Jiang, W. Liu, and Z. Liu, "Segmentation algorithm for infrared dim small targets based on double window," *Acta Photonica Sinica*, vol. 36, no. 11, pp. 2168–2171, 2007.
[41] 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
[42] 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.
[43] 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.
[44] 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.
[45] 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.
[46] L. Juliu, F. Xiangsuo, C. Huajin, L. Bing, M. Lei, and X. Zhiyong, "Dim and small target detection based on improved spatio-temporal filtering," *IEEE Photon. J.*, vol. 14, no. 1, Feb. 2022, Art. no. 7801211.
[47] 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, 2022, Art. no. 5000321.
[48] J. Mu, W. Li, J. Rao, F. Li, and H. Wei, "Infrared small target detection using Tri-layer template local difference measure," *Opt. Precis. Eng.*, vol. 30, no. 7, pp. 869–882, 2022.
[49] M. Wan, Y. Xu, Q. Huang, W. Qian, G. Gu, and Q. Chen, "Single frame infrared small target detection based on local gradient and directional curvature," *Proc. SPIE*, vol. 11897, pp. 99–107, 2021.
[50] E. Zhao, W. Zheng, M. Li, H. Sun, and J. Wang, "Infrared small target detection using local component uncertainty measure with consistency assessment," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6518205.