Research Article

Dim and Small Targets Detection in Sequence Images Based on Spatiotemporal Motion Characteristics

Fan Xiangsuo,¹ Hongwei Guo ,² Xu Zhiyong,³ and Biao Li³

¹School of Electrical and Information Engineering, Guangxi University of Science and Technology, Liuzhou 545006, China

²School of Engineering, Honghe University, Mengzi 661100, China

³Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu 610209, China

Correspondence should be addressed to Hongwei Guo; ghw202@foxmail.com

Received 5 May 2020; Revised 19 August 2020; Accepted 25 September 2020; Published 17 October 2020

Academic Editor: A. M. Bastos Pereira

Copyright © 2020 Fan Xiangsuo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to effectively enhance the low detection rates of dim and small targets caused by dynamic backgrounds, this paper proposes a detection algorithm for dim and small targets in sequence images based on spatiotemporal motion characteristics. With regard to the spatial domain, this paper proposes an improved anisotropic background filtering algorithm that makes full use of the gradient differences between the target and the background pixels in the eight directions of the spatial domain and selects the mean value of the three directions with the lowest diffusion function in the eight directions as the differential filter to obtain a differential image. Then, the paper proposes a directional energy correlation enhancement algorithm in the time domain. Finally, based on the above preprocessing operations, we construct a dim and small targets detection algorithm in sequence images with local motion characteristics, which achieves target detection by determining the number of occurrences of the target, the number of displacements, and the total cumulative area in these sequential images. Experiments show that the proposed detection algorithm in this paper can effectively improve the detection of dim and small targets in dynamic scenes.

1. Introduction

Photoelectric imaging detection systems obtain information on the target object by detecting reflected or radiant energy from the object's surface. Such systems are widely used in photoelectric astronomy and remote sensing-based navigation. The detection of the target is the primary task of any photoelectric imaging detection system; whether or not the target can be discovered decides if the system can monitor the target. However, interference from strong background radiation and dynamic clutter may cause dim and small objects to completely mask by an undulant background. Therefore, research on detection algorithms for dim and small objects in such scenarios is crucial to improving the performance of photoelectric imaging detection systems.

Many researchers have worked on detection algorithms that can effectively detect dim and small objects. Their work has mainly focused on single-frame detection and multi-frame motion correlation detection. Single-frame detection

mainly uses filtering and machine learning algorithms to separate and extract target points from a single-frame image. Single-frame detection algorithms can be grouped into two broad categories: background modeling methods and machine learning-based estimation methods. Background modeling methods mainly include two-dimensional least mean square (TDLMS) filtering [1], adaptive Butterworth filtering [2], improved Top-Hat filtering [3], improved bilateral filtering [4], direction support vector filtering [5], and background modeling methods based on statistical characteristics [6]. These algorithms work by using filtering algorithms to perform background estimation on the scene image before differentiating the estimated background image from the original image, thus obtaining a differential image that contains only the target object and a small amount of noise. Such algorithms are effective at background modeling for slow-moving scenes but perform poorly when applied to dynamic scenes with large spans. Machine learning-based detection algorithms mainly

include detection algorithms based on visual saliency [7, 8], detection algorithms based on low-rank and sparse representation [9–15], and detection algorithms based on convolutional neural networks [16, 17]. These algorithms are able to perform well only in certain scenes. For instance, detection algorithms based on visual saliency require a strong contrast between the background and the target object. Detection algorithms based on low-rank and sparse representation cannot adapt to complex dynamic scenes when applied to undulant, nonstationary backgrounds, which comprise their low-rank characteristics. In order to meet the needs of complex scene detection, for example, Wang et al. proposed a TV-PCP detection method, which uses the isotropic total variation algorithm to constrain the PCP model to constrain the low-rank characteristics of the background block. However, when the background is in dynamic change, the dynamic changing area will be mistaken as “sparse,” and there are many false targets in the detection results. In order to further constrain the sparsity of the target, [14, 15] adopt different kernel norms to suppress the strong undulating background and strengthen the constraint on the sparse term to largely eliminate the interference of the undulating background on the target and achieve good detection results. In addition, sparse representation requires the construction of a super-complete dictionary of targets and backgrounds, which in turn requires a large number of training samples and increases the complexity of the model. Detection algorithms based on convolutional neural networks are often impractical, because they require a large number of training samples to build a good representational capacity, which is highly time-consuming. In addition, dim and small target objects lack well-defined shapes and textures, making it difficult to build training models. Furthermore, the vast majority of real-world applications involve dynamic scenes. The parameters of pretrained models often have trouble adapting to such scenes.

Only a portion of the information contained in the image is used in single-frame detection. In order to make full use of the interframe motion information of sequential images, researchers have proposed multiframe motion correlation detection algorithms, which mainly include detect-before-track (DBT) algorithms and track-before-detect (TBD) algorithms. For instance, Lian et al. proposed a novel “pipeline filtering method” [18], which detects targets by first using grayscale feature analysis to extract candidate targets from the image before establishing a pipeline around each candidate target point. Since this method is based on local grayscale features, effective detection and tracking can be achieved only when there is a strong contrast between the target and the local background. Moyer et al. proposed a projective transformation method [19], which uses the Hough transform to correlate multiple-frame images, extract all suspicious target trajectories, and finally accumulate energy for each suspicious target trajectory. When the energy of a target on a certain trajectory is greater than some preset threshold, it is extracted and identified as the final target object. Chen proposed a detection algorithm for dim and small target objects based on dynamic programming

[20]. This method uses dynamic programming to accumulate energy for targets along all suspicious trajectories and achieves target detection by identifying the trajectory with the largest amount of accumulated energy. Xiu proposed an improved high-order correlation method [21], where an algorithm extracts target points by calculating their high-order interframe correlations. Such algorithms can effectively detect slow-moving small targets. However, when such targets are in scenes with a low signal-to-noise ratio, the target would be lost in background clutter, making its correlation levels close to background differences, making it difficult to detect the target point.

The multiframe detection algorithms discussed above provide the present paper with an excellent detection strategy. For dynamic scenes with low signal-to-noise ratios, it is possible to achieve the detection of dim and small target objects by first using preprocessing algorithms to remove most of the interference on the target caused by noise before applying multiframe motion correlations. The present paper proposes an algorithm based on spatiotemporal motion characteristics that improve the detection of dim and small target objects in dynamic scenes. First, we deploy an improved anisotropic background filtering algorithm that makes full use of the gradient differences between the target and the background pixels along eight directions in the spatial domain. The algorithm will then take the average value of the three directions with the lowest spread function values as the differential filter. This algorithm is able to remove most background interference while effectively utilizing spatial information. Then, we deploy a directional energy correlation enhancement algorithm to boost the target signal using the temporal motion characteristics of the target. This algorithm seeks the maximal energy value by integrating the interframe motion of the target to achieve multiframe energy accumulation. Finally, based on the results of such preprocessing, we build a detection algorithm for dim and small target objects in sequential images with local motion characteristics, which achieves target detection by determining the number of times the target appears in these sequential images, the number of displacements, and the total cumulative area.

2. Improved Anisotropic Filtering

The target signal in the image is an outward diffusion process of the central pixels. The gradient relationship of the anisotropy in each pixel with different directions is similar to that of the point diffusion. Therefore, it can be used to tell the difference between the target and the background and thus effectively distinguish the target from the background [22]. In order to obtain the difference image, [23] uses the following diffusion function to realize the difference filtering. The specific formula is as follows:

$$\begin{aligned} c_1(\nabla f) &= 1 - e^{-(\nabla f/k)^2}, \\ c_2(\nabla f) &= 1 - \frac{1}{1 + (\nabla f/k)^2}, \end{aligned} \quad (1)$$

where ∇f refers to the gradient between two pixels in the image; k is a constant; c_1 and c_2 are two different diffusion functions.

Literature [23] constructs the gradient difference of the target in four different directions combining with diffusion functions to achieve the difference image. The expression of gradient difference of pixel points in four directions is as follows:

$$\begin{aligned}\Delta f_U &= f(i, j) - f(i - \text{step}, j), \\ \Delta f_D &= f(i, j) - f(i + \text{step}, j), \\ \Delta f_L &= f(i, j) - f(i, j - \text{step}), \\ \Delta f_R &= f(i, j) - f(i, j + \text{step}),\end{aligned}\quad (2)$$

where f is the input image; Δf_U , Δf_D , Δf_L , and Δf_R are the gradient differences in four directions (up, down, left, and right) with the pixel $f(i, j)$ as the center; step refers to the step size between two pixels. The differential filter function is defined based on the gradient difference in four different directions as follows:

$$\begin{aligned}f'(i, j) &= \frac{1}{4} [c(\Delta f_U) \times \Delta f_U + c(\Delta f_D) \times \Delta f_D \\ &\quad + c(\Delta f_L) \times \Delta f_L + c(\Delta f_R) \times \Delta f_R],\end{aligned}\quad (3)$$

where (i, j) refers to the coordinate position of the pixel point; $f'(i, j)$ refers to the difference image; other variables are defined as above.

Research has found that if the mean value of diffusion functions in four directions in the spatial domain is simply calculated as the difference image; when the pixel is in the edge contour, there are at least two large diffusion function values in different directions. The simple processing of the mean values will narrow the difference in diffusion function values between pixels in this area and the target area. It is difficult to preserve the edge contour in the background modeling process, resulting in much edge noise in the differential image, which is not conducive to extracting target points [22]. To effectively solve this situation, the differential filter function above is improved. The gradient differences between the target and the background in eight directions of the spatial domain are deeply analyzed. Three

smallest diffusion function values are selected to filter the image. In this case, the small diffusion coefficient will be used for the background and edge contour, which is possible to preserve these two parts of information, and the difference image can effectively reduce the interference of these two parts of information to the target signal. The gradient relationships of the target in eight different directions are shown in Figure 1.

The gradient relationships in eight directions in the spatial domain are constructed according to Figure 1. The specific formulas are as follows:

$$\begin{aligned}\Delta f_U &= f(i, j) - f(i - \text{step}, j), \\ \Delta f_D &= f(i, j) - f(i + \text{step}, j), \\ \Delta f_L &= f(i, j) - f(i, j - \text{step}), \\ \Delta f_R &= f(i, j) - f(i, j + \text{step}), \\ \Delta f_{LU} &= f(i, j) - f(i - \text{step}, j - \text{step}), \\ \Delta f_{RU} &= f(i, j) - f(i - \text{step}, j + \text{step}), \\ \Delta f_{LD} &= f(i, j) - f(i + \text{step}, j - \text{step}), \\ \Delta f_{RD} &= f(i, j) - f(i + \text{step}, j + \text{step}),\end{aligned}\quad (4)$$

where Δf_U , Δf_D , Δf_L , Δf_R , Δf_{LU} , Δf_{RU} , Δf_{LD} , and Δf_{RD} indicate the gradient differences with the pixel $f(i, j)$ as the center: up, down, left, right, upper left, upper right, lower left, and lower right.

According to the gradient difference of eight directions constructed by formula (4), they are substituted into the diffusion function c_1 . By comparing the gradient differences of eight directions in different regions of the image, it is found that the gradient in eight directions in the stationary background region is smaller and the gradient in eight directions in the target region is larger, while in the edge contour area at least three directions have smaller gradient values, and other directions have larger gradient values. Therefore, according to the above characteristics, the mean value of three directions with smaller diffusion function value can be selected for differential filtering, which will effectively enhance the target signal. The flow chart of improved differential filtering is shown in Figure 2.

According to Figure 2, the improved anisotropic difference filter expressions are as follows:

$$\begin{aligned}[\min_1, \min_2, \min_3] &= \text{Sort_ascend}\{c_1(\Delta f_U) \cdot \Delta f_U, c_1(\Delta f_D) \cdot \Delta f_D, c_1(\Delta f_L) \cdot \Delta f_L, c_1(\Delta f_R) \cdot \Delta f_R, \\ &\quad c_1(\Delta f_{LU}) \cdot \Delta f_{LU}, c_1(\Delta f_{RU}) \cdot \Delta f_{RU}, c_1(\Delta f_{LD}) \cdot \Delta f_{LD}, c_1(\Delta f_{RD}) \cdot \Delta f_{RD}\}, \\ f'(i, j) &= \frac{1}{3} [\min_1 + \min_2 + \min_3],\end{aligned}\quad (5)$$

where $\text{Sort_ascend}(\cdot)$ represents a function to sort array elements from small to large; $\min_1$, $\min_2$, and $\min_3$ are the three parameters with the smallest diffusion function values

in the eight directions. The mean values of these three parameters are used to filter the image pixel by pixel, which can effectively highlight the signal of dim and small targets.

	$f(i - \text{step}, j - \text{step})$			$f(i - \text{step}, j)$			$f(i - \text{step}, j + \text{step})$	
	$f(i, j - \text{step})$			$f(i, j)$			$f(i, j + \text{step})$	
	$f(i + \text{step}, j - \text{step})$			$f(i + \text{step}, j)$			$f(i + \text{step}, j + \text{step})$	

FIGURE 1: Gradient relationship in eight directions.

FIGURE 2: The flow chart of improved differential filtering.

3. Directional Energy Correlation Enhancement Algorithm

Conventional multiframe energy enhancement across the time domain simply takes the average of the sum of multiple image frames, which does not make full use of the spatial information of the target. To this end, [24] proposes an energy enhancement algorithm that incorporates such spatial information, which achieves energy accumulation by obtaining the maximal energy value of the target based on 12

kinds of motion templates in the spatial domain, thus achieving a better enhancement effect. Researchers have found that such methods, which achieve energy enhancement by adding a filtering template, will result in the loss of some of the target's energy, because the filtering template will only retain the energies corresponding to elements with a pixel value of 1. In order to retain more of the target signal, and considering the long-distance imaging of the system, the target usually moves between frames at a speed less than or equal to 2 pixels per second. Hence, we can build 25

directional models of the target's interframe motion (as shown in Figure 3, where the 12th directional motion follows a $1 \rightarrow 2 \rightarrow 12$ mode). We first horizontally translate the previous frame according to these 25 directional motions to obtain 25 directional motion diagrams. Thereafter, we calculate the degree of energy correlation with the next frame and finally obtain the maximal energy value for the target for these 25 directions, thus achieving multiframe energy accumulation.

The advantage of the proposed enhancement algorithm is that it makes full use of the spatiotemporal information of the pixel for energy enhancement. Firstly, the target is moved according to 25 motion modes in the spatial domain, and 25 motion mode images are obtained. Then, the energy correlation between the 25 kinds of motion mode images and the image at the next frame is calculated. In this way, the spatial motion information of the target is effectively integrated into the energy enhancement. Finally, the multiframe images in the temporal domain are accumulated to achieve the target energy enhancement. The specific formula is as follows:

$$\begin{aligned}
 T_k(x, y, t_n) &= \text{Move}(f'(x, y, t_n)), \\
 f_k(x, y, t_n) &= \frac{1}{r * r} \sum_{i=-fix(r/2)}^{i=fix(r/2)} \sum_{j=-fix(r/2)}^{j=fix(r/2)} T_k(x, y, t_n) \\
 &\quad \times f'(x + i, y + j, t_{n+1}), \\
 f_E(x, y, t_n) &= \max(f_1(x, y, t_n), f_2(x, y, t_n), \\
 &\quad \dots, f_{25}(x, y, t_n)), \\
 f'_E &= \frac{1}{N} \times (f_E(x, y, t_1) + f_E(x, y, t_2) + \dots \\
 &\quad + f_E(x, y, t_N)),
 \end{aligned} \tag{6}$$

where (x, y) represents the rank of the pixel, t_n is the n th frame, f' represents the frame difference graph for the t_n frame, and T_k is the k -th directional motion. $k = 1, 2, \dots, 25$ are the images obtained after horizontal translation in the 25 different directions, r is the radius of the neighboring area, f_E is the maximal energy value of the 25 directions, N is the cumulative number of frames, and f'_E is the image before and after energy enhancement.

4. Detection Algorithm

In this paper, the improved anisotropic algorithm is used to obtain the difference image f' based on the spatial information of the pixel. Then, the energy enhancement algorithm based on direction energy correlation is used to realize the energy enhancement in time domain, and the double window segmentation algorithm of [25] is used to segment the enhanced image f'_E and obtain the sequential binary image. Finally, combined with the idea of pipeline filtering [26], based on the interrelated characteristics of the target's motion on the sequence image, a detection algorithm of dim

17	16	15	14	13
18	5	4	3	12
19	6	1	2	11
20	7	8	9	10
21	22	23	24	25

FIGURE 3: Models of 25 movement directions.

and small targets in the sequence image with local motion characteristics is constructed. The algorithm detects the target by getting the occurrence frequency O_{sum} of the target, the number of the target displacements M_{sum} , and the total cumulative area of the target A_{sum} within the spatial domain $(2r + 1) \times (2r + 1)$ of the continuous N_2 frame images. The specific formula is as follows:

$$\begin{aligned}
 O_{\text{sum}} &= \sum_{t_n=1}^{N_1} O(F(x, y, t_n)), \\
 M_{\text{sum}} &= \sum_{t_n=1}^{N_1} M(F(x, y, t_n)), \\
 A_{\text{sum}} &= \sum_{t_n=1}^{N_1} A(F(x, y, t_n)),
 \end{aligned} \tag{7}$$

where (x, y) is the centroid position of a candidate block; $F(x, y, k)$ is a binary image at the time k ; O records the occurrence frequency that the centroid point of a candidate block appears in its neighborhood of $(2r + 1) \times (2r + 1)$; O_{sum} is the total number of times that the centroid point of a candidate block appears on continuous N_1 frames; M records the number of times that the centroid point of a candidate block moves within its neighborhood of $(2r + 1) \times (2r + 1)$; M_{sum} refers to the total number of times that the centroid point of a candidate block moves on continuous N_1 frames; A refers to the area of a candidate block; O_{sum} refers to the area of a candidate block on continuous N_1 frames.

The flow chart of the algorithm is shown in Figure 4.

According to the above flow chart, the pseudocode of the detection algorithm is given as Algorithm 1.

5. Results and Analysis

The experiment involved in this paper is done on a PC with 2 GB of memory at Pentium 2.10 GHz and completed by MATLAB 2012B software. The images in this paper are mainly dynamic images with more edge contours. The images are generally composed of three parts, background + target + edge contour area, as shown in Figure 5.

FIGURE 4: Flow chart of the proposed target detection method.

Initialization: The constant k in diffusion function is $k = 100$; The step size between two pixels is $\text{step} = 4$; Enhanced cumulative frames $N = 3$; Accumulated frames in time domain $N_1 = 5$; The radius of neighborhood $r = 3$; Target occurrence threshold $Th_1 = N_1/2$; Threshold of target displacement times $Th_2 = N_1/2$; The sum threshold of target cumulative area $Th_3 = 30$.

Input: Original images

(1) Using the improved anisotropy to filter the image with spatial information to obtain the difference image f' . Then, the direction energy correlation enhancement algorithm is used to realize the energy enhancement in the time domain and obtain the enhanced image f'_E .

(2) The segmentation algorithm of reference [25] was used to segment the enhanced image f'_E to obtain a sequence binary image.

(3) Combined with formula (7), dim and small targets detection in sequence image with local motion characteristics is used to calculate the number of target occurrence O_{sum} , number of target displacement M_{sum} and the sum of target cumulative area A_{sum} on consecutive N_1 frames images. If the conditions $O_{\text{sum}} > Th_1$ and $M_{\text{sum}} > Th_2$ and $A_{\text{sum}} \geq Th_3$ are met, it is determined as the target point, and updated the N_1 frames images until all the images are processed;

Output: Sequence detection results

ALGORITHM 1: Pseudocode of detection algorithm

It can be seen from Figure 5 that, due to the high local correlation in the background area, the gradient difference in the eight directions is small; the target imaging process satisfies the mode of the point spread function, and the energy is the process of diffusion from the central point to the outside, so the gradient difference between the central pixel and the eight directions is large. The edge contour area is in the transition area between the stationary background and the nonstationary background. This is partly caused by the uneven distribution of brightness. Therefore, at least three of the eight directions have smaller gradient values, and other directions have larger gradient values. In this paper, the corresponding algorithm is improved based on

the gradient difference of the above-mentioned different regions in order to remove most of the interference of the background and edge contour area on the target signal. The existing algorithms are insufficient for these scenes that dynamically change as the light changes. In order to verify the effectiveness of the algorithm proposed in this paper, the selected scenes are dynamically changing scenes.

5.1. Analysis of Background Modeling Results. To compare the background estimation effects of different algorithms, the structural similarity (SSIM) [27], contrast gain, and background inhibitory factors [28] are used to evaluate the background modeling effects. All indicators are as follows:

FIGURE 5: Image analysis.

$$\begin{aligned}
 \text{SSIM} &= \frac{(2\mu_R\mu_F + \varepsilon_1)(2\sigma_{RF} + \varepsilon_2)}{(\mu_R^2 + \mu_F^2 + \varepsilon_1)(\sigma_R^2 + \sigma_F^2 + \varepsilon_2)}, \\
 R &= \frac{|E_t - E_b|}{|E_t + E_b|}, \\
 I &= \frac{R_{\text{out}}}{R_{\text{in}}}, \\
 B_{\text{IF}} &= \frac{\sigma_{\text{in}}}{\sigma_{\text{out}}},
 \end{aligned} \tag{8}$$

where μ_R and σ_R are the mean value and the standard deviation of the input image; σ_{RF} is the covariance between the input image and the background image; ε_1 and ε_2 are constants; E_t and E_b are the mean values of the target area and the background area; R_{in} and R_{out} are the contrast between the original image and the differential image; I is contrast gain; σ_{in} and σ_{out} are the mean square deviations between the input image and the differential image, respectively; B_{IF} is background inhibitory factor.

The background estimation algorithm proposed in this paper is compared with the improved Top-Hat in [3], the robust principal component analysis (RPCA) method in [9], the low-rank decomposition (LRD) method in [11], and the anisotropic filtering (AF) in [12]. The same image is selected to evaluate the background estimation effects of different algorithms, as shown in Figure 6, where (a) shows the input image; (b) shows the background image and the differential image obtained by the improved Top-Hat; (c) shows the background image and the differential image obtained by the PRCA method; (d) shows the background image and the differential image obtained by the LRD; (e) shows the background image and the differential image obtained by the anisotropic filtering; (f) shows the background image and the differential image obtained by the algorithm proposed in this paper.

It can be seen from Figure 6 that the improved Top-Hat results in blurred background images even though it utilizes the advantages of combining internal and external structuring elements and makes full use of structuring elements of

different scales to effectively perform background modeling. For images with a low signal-to-noise ratio, differential images are left with more noise after image filtering. The background modeling effect of the RPCA algorithm is related to the number of frames selected for the training sample. Building decomposition matrices using more frames will result in a better differential image. However, RPCA is poor at background suppression when applied to complex scenes, because the image data distribution is composed of multiple linear subgroups. This increases the number of false targets in the differential image. The LRD algorithm is easily affected by background light. In the event of dynamic changes to the background, the differential image will contain more background clutter. Anisotropic filtering simply averages the spread functions for four directions in the neighboring area, which causes differential images to contain more peripheral contours, making it difficult to distinguish the targets. The backgrounds obtained using the algorithm proposed in the present paper preserve, to a large degree, a stationary background and peripheral contours, allowing the target signal to be effectively extracted from the differential image.

At the same time, we select three evaluation indexes to assess the differences between these algorithms from a quantitative perspective. The present paper selects images with different signal-to-noise ratios to compute the three evaluation indexes, with experimental results shown in Tables 1 to 3.

From the indicators in Tables 1 to 3, it can be seen that, for images from different scenes, the background estimation algorithm proposed in the present paper performs better at background estimation compared to conventional algorithms. The structural similarity index (SSIM), contrast gain (I), and background suppression factor (B_{IF}) are, respectively greater than 0.993, 16, and 5.56. Further analysis shows that, based on the gradient differences between the target and the background pixels along eight directions in the spatial domain, the algorithm generates differential images by building an improved anisotropic filtering algorithm, which effectively utilizes spatial information while removing most background interference, thus achieving better background modeling results.

5.2. Analysis of Energy Enhancement Results. In order to verify the enhancement effect of the algorithm proposed in this paper, we select some differential images for experimental analysis. We set the radius of the neighboring area to be $r = 3$ and the length of the accumulated frames to be $N = 3$. To analyze image enhancement effects, we select two assessment indicators: average grayscale of the target and the signal-to-noise ratio of the image. In addition, we perform comparisons with conventional multiframe energy enhancement algorithms and the enhancement algorithm in [24]. The experimental results are shown in Figure 5, where (a) shows the differential image and three-dimensional diagram; (b) shows the enhancement results diagram and three-dimensional graph obtained from a conventional multiframe energy enhancement algorithm; (c) shows the

FIGURE 6: Continued.

FIGURE 6: Background image and differential image obtained by different algorithms. (a) Input image. (b) Improved Top-Hat. (c) PRCA. (d) LRD. (e) AF. (f) Proposed method.

TABLE 1: SSIM results.

SNR	ITH [3]	PRCA [9]	LRD [11]	AF [12]	Proposed method
0.90	0.8436	0.9346	0.9423	0.8364	0.9938
1.27	0.8438	0.9349	0.9434	0.8426	0.9942
1.69	0.8464	0.9358	0.9451	0.8583	0.9949
2.00	0.8579	0.9372	0.9574	0.8646	0.9957
1.78	0.8443	0.9368	0.9478	0.8599	0.9954

TABLE 2: Contrast gain results.

SNR	ITH [3]	PRCA [9]	LRD [11]	AF [12]	Proposed method
0.90	13.9616	15.4999	15.5126	13.4105	16.0029
1.27	14.4466	15.5782	15.6832	13.3920	17.6247
1.69	14.8628	15.8294	15.9221	12.3488	18.3428
2.00	16.9810	17.8823	17.9257	15.6928	19.8290
1.78	15.5970	16.4400	16.5708	15.4074	19.5730

enhancement results diagram and three-dimensional graph obtained by the enhancement algorithm in [24]; and (d) shows the enhancement results diagram and three-dimensional graph obtained by the enhancement algorithm

proposed in this paper. Table 4 shows the average grayscales (AG) of the targets and signal-to-noise ratios (SNR) of the images for the different algorithms. From Figures 7 and 8, it can be seen that conventional multiframe energy

TABLE 3: Background inhibitory factor results.

SNR	ITH [3]	PRCA [9]	LRD [11]	AF [12]	Proposed method
0.90	2.1840	3.0113	4.0302	2.0715	5.5618
1.27	2.2298	3.7749	4.1562	2.1947	5.6990
1.69	2.2836	4.1054	4.2590	2.2661	5.7012
2.00	2.9642	4.3572	4.4993	2.7104	5.9944
1.78	2.3939	4.3241	4.3926	2.2128	5.7690

TABLE 4: Enhancement results of different algorithms.

Difference image		Conventional enhancement algorithm		Enhancement algorithm in [24]		Proposed method	
AG	SNR (dB)	AG	SNR (dB)	AG	SNR (dB)	AG	SNR (dB)
51	2.20	68	3.11	198	13.91	255	17.15
60	2.31	75	3.43	215	14.02	255	17.22

enhancement algorithms and the enhancement algorithm in [24] have managed to enhance small and dim targets. On the whole, the enhancement algorithm proposed in this paper produces greater enhancement effects, as well as far greater improvements to signal-to-noise ratio of the enhanced image, making for a far better overall performance.

It can be seen from Table 5 that the conventional multiframe energy enhancement algorithm and the enhancement algorithm in [24] are both able to enhance the target signal to a certain extent when used to process images that have undergone differential filtering. In different images, the AG and SNR of the traditional enhancement algorithm are 68 and 3.11 dB and 75 and 3.43 dB, respectively. The AG and SNR of the enhancement algorithm in [21] are 198 and 13.91 dB and 214 and 14.02 dB, respectively. The algorithm in this paper effectively combines spatiotemporal motion characteristics of the pixels to further enhance the target signal and achieves AG and SNR of, respectively, 255 and 17.15 dB and 255 and 17.22 dB.

5.3. Analysis of Detection Results

5.3.1. Detection Results. In order to evaluate the performance of the detection algorithm proposed in this paper, we selected four dynamic scenes for experimentation. Among them, the target in scene 1 is highly maneuverable. At first, the target moves upward and then moves downward; then it suddenly accelerates and moves diagonally upward and finally suddenly turns around and moves downward; the target in scene 2 moves diagonally upward in a dynamically changing scene, the target in scene 3 moves in a straight line from top to bottom in the background of changing illumination, and the target in scene 4 moves diagonally downward in the background of dynamically changing clouds. The information of each scene is shown in Table 5.

This paper compares a variety of detection algorithms with the proposed algorithm, which include the improved Top-Hat in [3], the RPCA in [9], the low-rank decomposition in [11], and the anisotropic filtering in [12]. The results

of the algorithms applied to the 4 scenes are, respectively, shown in Figures 9–12, where (a)–(f), respectively, show the input image, improved Top-Hat, RPCA algorithm, LRD algorithm, anisotropic algorithm, and the detection algorithm proposed in this paper.

It can be seen from Figures 9–12 that the improved Top-Hat uses two different scaled structuring elements for filtering, which makes full use of the scale characteristics of internal and external structuring elements to effectively improve background estimation. However, it performs poorly when applied to more undulated backgrounds and produces differential images with more noise, as shown in Figures 9(b), 10(b), 11(b), and 12(b). The RPCA algorithm needs to use data from the scene to build a low-rank decomposition model and works well only when the background is a single linear subspace. When the background is dynamic in nature, the model fails to adapt well, and there are many dynamic background elements left in the detection results, as illustrated in Figures 9(c), 10(c), 11(c), and 12(c). The LDR algorithm accounts for the fact that the image is composed of multiple linear subspaces and effectively reflects the low-rank subspaces of different components from which the background data originates. It does a better job in detection than the RPCA algorithm but remains susceptible to undulations in the background. When the background changes with the light, the low-rank properties of the decomposition model will be affected, which will adversely impact background suppression and in turn cause more false targets in the detection results, as illustrated in Figures 9(d), 10(d), 11(d), and 12(d). Because anisotropic filtering merely takes the mean value of the spread functions of four directions in the spatial domain as the background estimation result, it is difficult for the algorithm to distinguish the differences between nonstationary peripheral contour areas and the area around the target when the scene has more nonstationary edge contours. This then causes more edge contour areas to be left in the detection results, as illustrated in Figures 9(e), 10(e), 11(e), and 12(e). The detection algorithm proposed in this paper first uses improved anisotropic filtering to perform differential filtering using the

FIGURE 7: Enhancement results of different algorithms. (a) The differential image. (b) Conventional enhancement algorithm. (c) Enhancement algorithm in [24]. (d) The proposed method.

FIGURE 8: Enhancement results of different algorithms. (a) The differential image. (b) Conventional enhancement algorithm. (c) Enhancement algorithm in [24]. (d) The proposed method.

TABLE 5: Image attributes.

Scene	Average SNR (dB)	Frame length	Target size
1	2.45	113	3×3
2	2.60	100	3×3
3	1.24	100	3×3
4	1.50	388	5×5

FIGURE 9: Detection results of different algorithms for scene 1. (a) First frame. (b) Improved Top-Hat. (c) RPCA. (d) LRD. (e) AF. (f) Proposed method.

FIGURE 10: Detection results of different algorithms for scene 2. (a) First frame. (b) Improved Top-Hat. (c) RPCA. (d) LRD. (e) AF. (f) Proposed method.

FIGURE 11: Continued.

FIGURE 11: Detection results of different algorithms for scene 3. (a) First frame. (b) Improved Top-Hat. (c) RPCA. (d) LRD. (e) AF. (f) Proposed method.

FIGURE 12: Continued.

FIGURE 12: Detection results of different algorithms for scene 4. (a) First frame. (b) Improved Top-Hat. (c) RPCA. (d) LRD. (e) AF. (f) Proposed method.

spatial information of the pixels before deploying the directional energy-related enhancement algorithm to enhance the target signal in the time domain. Finally, the algorithm uses multiframe motion correlation to effectively detect the final target point by building a detection algorithm for dim and small target objects in sequential images with local motion characteristics, as illustrated in Figures 9(f), 10(f), 11(f), and 12(f).

To further explore the features of these algorithms, the ROC curves of three scenes are plotted. The horizontal axis is the detection rate (P_d) and the vertical axis is the false alarm rate (P_f), as shown in formula (9). NTDT refers to the number of the real targets detected, NFDT refers to the number of false alarm targets detected, NT refers to the total number of real targets present in the image, and NP refers to the total number of targets detected in the image. The ROC curves of the three scenes are shown in Figure 13.

$$\begin{cases} P_d = \frac{\text{NTDT}}{\text{NT}} \times 100\%, \\ P_f = \frac{\text{NFDT}}{\text{NP}} \times 100\%. \end{cases} \quad (9)$$

As can be seen from the ROC curves of the four scenes in Figure 10, compared to other algorithms, this paper first uses improved anisotropic filtering to remove most of the background interference before enhancing the target signal

using the directional energy correlation enhancement algorithm. Then, the final target point is obtained using the detection algorithm for dim and small target objects in sequential images with local motion characteristics in order to obtain better detection results. At a false alarm rate $P_f = 0.02$, the detection rate P_d of the detection algorithm proposed in this paper is greater than 90%, compared to less than 85% for other algorithms, as illustrated in Figure 13(a). In the same scene, because the LRD algorithm builds a low-rank decomposition model based on image components comprised of multisubspace data, it generates a detection result superior to the RPCA algorithm, which relies solely on image elements in a single linear subspace. Given the same false alarm rate in the same situation, the detection rate of an LRD algorithm is higher than that of an RPCA algorithm. For example, when $P_f = 0.0024$, the detection rate of the LRD algorithm P_d is higher than 85%, while the RPCA algorithm is lower than 82%, as shown in Figure 13(b). The improved Top-Hat algorithm can only be applied to static or slowly changing scenes. In dynamic scenes, the algorithm has a detection rate lower than its peers. For example, when $P_f = 0.0024$, the improved Top-Hat detection rate P_d is less than 70%, while P_d for other algorithms are higher than 80%, as shown in Figure 13(c). Because the anisotropy detection algorithm merely takes the mean value of the spread functions of four directions in the spatial domain as the background estimation result, the algorithm cannot effectively remove nonstationary peripheral contours and has the

FIGURE 13: ROC curves of different algorithms in 4 scenes. (a) Scene 1. (b) Scene 2. (c) Scene 3. (d) Scene 4.

TABLE 6: Operational efficiency of different algorithms.

Method	Scene 1	Scene 2	Scene 3	Scene 4
Top-Hat	0.05	0.07	0.65	0.20
PRCA	0.73	0.82	0.80	1.62
LRD	1.60	1.75	1.70	2.64
AF	0.17	0.20	0.15	0.50
Proposed method	0.52	0.67	0.42	0.91

lowest detection rates. When $P_f = 0.003$, the improved Top-Hat detection rate P_d is less than 70%, while P_d of other algorithms are all greater than 75%, as shown in Figure 13(d).

5.3.2. Running Time. In order to verify the efficiency of different algorithms, we calculate the efficiency of each method (s/frame), with the specific results shown in Table 6. From the table, it can be seen that the conventional Top-Hat

and AF algorithms are more efficient due to their simplicity. RPCA and LRD require sample data to build a low-rank decomposition model, which is highly time-consuming. Meanwhile the algorithm proposed in the present paper requires corresponding preprocessing algorithms to obtain differential images and enhance the target signal, which also requires a certain amount of time. Compared to conventional algorithms, our proposed algorithm is less efficient but more accurate. However, the algorithm proposed in this paper requires far less computation time than the low-rank decomposition model.

6. Conclusion

This paper proposed a detection algorithm for dim and small targets in sequential images based on spatiotemporal motion characteristics. First, the gradient differences between the target and the background pixels along eight directions in the spatial feature are analyzed in depth, and the average value of the three directions with the lowest spread function values is selected as the differential filter, which preserves the target signal while effectively removing most of the background interference. Second, we propose a directional energy correlation enhancement algorithm, which enhances the target signal by multiframe accumulation in the time domain by obtaining the maximal energy value for the target for these 25 directions. Finally, on the foundation of preprocessing operations, the final target point is detected by building a detection algorithm for dim and small target objects in sequential images with local motion characteristics. Experiments show the following:

- (1) Through in-depth analysis of the differences in the eight directions of different regions in the image, an improved anisotropic differential filter function is constructed to effectively retain the target signal. From the three evaluation indexes of background modeling, the improved anisotropic filtering achieves better filtering effect, and SSIM, I , and B_{IF} are greater than 0.993, 16, and 5.56, respectively.
- (2) The direction energy correlation enhancement algorithm firstly moves according to the target's 25 motion modes in spatial feature, which effectively takes the spatial motion information into account in the energy enhancement. Then, it calculates the energy correlation between the shifted image and the next moment image and finally achieves the multiframe energy in the temporal domain by calculating the maximum energy of the target in 25 motion directions. Compared with the traditional multiframe enhancement algorithm, the AG and SNR of the image obtained by the proposed algorithm are 255 and 17.15 dB, respectively.
- (3) Through experiments on 4 dynamic scenes, it is demonstrated that, based on the preprocessing of the image, the detection algorithm proposed in this paper deeply analyzes the difference of the local motion characteristics of the target and noise in the sequence image and constructs the detection

algorithm of multiframe motion association, which can effectively eliminate noise interference and detect actual target points.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

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

Acknowledgments

This work was partly supported by the Guangxi Science and Technology Base and Talent Project (no. AD19245130), the National Natural Science Foundation of China under Grants 62001129 and 62061015, Yunnan Local Colleges Applied Basic Research Project (no. 2018FH001-056), and the Foundation of Yunnan Educational Committee under Grant 2020J0675.

References

- [1] L. L. Wan and M. Wang, "Infrared target detection using TDLMS filter based on neighborhood information," *Journal of Huazhong University of Science and Technology*, vol. 43, no. 1, pp. 178–181, 2015.
- [2] L. Yang, *Infrared Small Target Detection and Tracking Algorithm under Complex Background*, Shanghai Jiaotong University, Shanghai, China, 2006.
- [3] D. W. Li, *Infrared Dim and Small Targets Detection Based on Complex Background*, Harbin University of Technology, Harbin, China, 2013.
- [4] Y. Q. Zeng and Q. Chen, "Dim and small target background suppression based on improved bilateral filtering for single infrared image," *Infrared Technology*, vol. 33, no. 9, pp. 245–249, 2011.
- [5] C. Yang, J. Ma, S. Qi, J. Tian, S. Zheng, and X. Tian, "Directional support value of Gaussian transformation for infrared small target detection," *Applied Optics*, vol. 54, no. 9, pp. 2255–2265, 2015.
- [6] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, no. 1, pp. 4–21, 2014.
- [7] Y. Chen, B. Song, X. Du, and M. Guizani, "Infrared small target detection through multiple feature analysis based on visual saliency," *IEEE Access*, vol. 7, pp. 1–9, 2019.
- [8] L. Dong, B. Wang, M. Zhao, and W. Xu, "Robust infrared maritime target detection based on visual attention and spatiotemporal filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 5, pp. 3037–3050, 2017.
- [9] Z. Wang, X. Yang, and X. Gu, "Small target detection in a single infrared image based on RPCA," *Acta Armamentarii*, vol. 37, no. 9, pp. 1753–1760, 2016.
- [10] Y. He, M. Li, and J. Zhang, "Clutter suppression of infrared image based on three-component lowrank matrix decomposition," *Optics and Precision Engineering*, vol. 23, no. 7, pp. 2069–2078, 2015.

- [11] O. Oreifej, X. Li, and M. Shah, "Simultaneous video stabilization and moving object detection in turbulence," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 2, pp. 450–462, 2013.
- [12] Y. Lu, S. Huang, and W. Zhao, "Sparse representation based infrared small target detection via an online-learned double sparse background dictionary," *Infrared Physics & Technology*, vol. 99, pp. 14–27, 2019.
- [13] X. Wang, Z. Peng, D. Kong, P. Zhang, and Y. He, "Infrared dim target detection based on total variation regularization and principal component pursuit," *Image and Vision Computing*, vol. 63, pp. 1–9, 2017.
- [14] L. Zhang, L. Peng, T. Zhang, S. Cao, and Z. Peng, "Infrared small target detection via non-convex rank approximation minimization joint l_2 , l_1 norm," *Remote Sensing*, vol. 10, no. 11, p. 1821, 2018.
- [15] T. Zhang, H. Wu, Y. Liu, L. Peng, C. Yang, and Z. Peng, "Infrared small target detection based on non-convex optimization with L_p -Norm constraint," *Remote Sensing*, vol. 11, no. 5, p. 559, 2019.
- [16] Z. Gao, J. Dai, and C. Xie, "Dim and small target detection based on feature mapping neural networks," *Journal of Visual Communication and Image Representation*, vol. 62, pp. 206–216, 2019.
- [17] K. Qian, H. Zhou, H. Qin, S. Rong, D. Zhao, and J. Du, "Guided filter and convolutional network based tracking for infrared dim moving target," *Infrared Physics & Technology*, vol. 85, no. 1, pp. 431–442, 2017.
- [18] K. Lian, H. Wang, and D. Li, "Pipeline filtering method based on analysis of local grayscale characteristics of infrared targets," *Journal of Projectiles, Rockets, Missiles and Guidance*, vol. 31, no. 4, pp. 200–206, 2011.
- [19] L. R. Moyer, J. Spak, and P. Lamanna, "A multi-dimensional Hough transform-based track-before-detect technique for detecting weak targets in strong clutter backgrounds," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 4, pp. 3062–3068, 2011.
- [20] M. Chen, *Dim and Small Target Tracking Algorithm of before Detection Based on Dynamic Programming*, pp. 43–47, University of Electronic Science and Technology of China, Chengdu, China, 2014.
- [21] P. Xiu, *Small Target Detection and Tracking in Deep Space*, pp. 19–26, University of Chinese Academy of Sciences, Beijing, China, 2015.
- [22] X. Fan, *Dim and Small Targets Detection and Tracking Algorithms in Sequence Image*, pp. 30–34, University of Electronic Science and Technology of China, Chengdu, China, 2019.
- [23] H. X. Zhou, Y. Zhao, H. L. Qin et al., "Infrared dim and small target detection algorithm based on multi-scale anisotropic diffusion equation," *Acta Photonica Sinica*, vol. 44, no. 9, pp. 146–150, 2015.
- [24] X. Fan, Z. Xu, J. Zhang, Y. Huang, and Z. Peng, "Dim small targets detection based on self-adaptive caliber temporal-spatial filtering," *Infrared Physics & Technology*, vol. 85, pp. 465–47, 2017.
- [25] H. J. Jiang, W. Liu, and Z. H. Liu, "Segmentation algorithm for infrared dim small targets based on double window," *Acta Photonica Sinica*, vol. 36, no. 11, pp. 2168–2171, 2007.
- [26] X. M. Zhao, S. C. Yuan, X. L. Ma et al., "Infrared small target detection technique based on moving pipeline filtering," *Infrared Technology*, vol. 31, no. 5, pp. 295–297, 2009.
- [27] Q. Zhang, J. Cai, and Q. Zhang, "Anisotropic infrared background prediction method," *High Power Laser and Particle Beams*, vol. 24, no. 2, pp. 301–306, 2012.
- [28] H. Qin, H. Zhou, S. Liu et al., "Dim and small target background suppression using bilateral filtering," *High Power Laser and Particle Beams*, vol. 21, no. 1, pp. 25–28, 2009.