Infrared Small Target Detection Utilizing the Enhanced Closest-Mean Background Estimation

Jinhui Han, Chengyin Liu, Yuchun Liu, Zhen Luo, Xiaojian Zhang, and Qifeng Niu

***Abstract*—Background estimation is an efficient infrared (IR) small target detection method. However, to deal with unknown targets, the estimation window in existing algorithms should be adjusted to perform multiscale detection and requires a lot of calculations. Besides, the stages during and after estimation have received wide attention in existing algorithms, but the research on the stages before estimation is insufficient. Moreover, existing algorithms typically regard the maximum value of different orientations as the estimation value. However, when a dim target is adjacent to high-brightness background, it is easily submerged. This article proposes a three-layer estimation window to detect targets of different sizes with only a single-scale calculation. The enhanced closest-mean background estimation method is then proposed and carefully designed before, during, and after the estimation. Before estimation, the matched filter is adopted to improve the image signal-to-noise ratio. During estimation, the principle of closest-mean is proposed to suppress high-brightness background. After estimation, a ratio-difference operation is performed to enhance the true target and suppress the background simultaneously. A simple checking mechanism is proposed to further improve the detection performance. Experiments on some IR images demonstrate the effectiveness and robustness of the proposed method. Compared with existing algorithms, the proposed method has better target enhancement, background suppression, and computational efficiency.**

***Index Terms*—Background estimation, closest-mean, infrared (IR) small target, matched filter, three-layer window.**

I. Introduction

IN INFRARED (IR) precise guidance [1], early warning [2], space tracking [3], maritime target searching [4], and other fields, the target is usually very far from the detector. Due to the limitation of the optical system and focal plane array performance, the target usually occupies only a few pixels and has a weak gray value, which is difficult to detect and leads to a low detection rate [5]. In addition, due to the presence of trees, houses, clouds, sea waves, and other clutter, there are usually various types of complex backgrounds in the raw IR image, such as high-brightness background [6], background edge [7], and pixel-sized noise with high brightness (PNHB) [8], and they may cause severe interferences and bring low signal-to-clutter ratio (SCR) [9], which may lead to a high false alarm rate.

Several algorithms have been proposed, including sequence-based and frame-based algorithms [10] to achieve precise IR small target detection with high detection rate and low false alarm rate. Sequence-based algorithms use the difference information between consecutive frames to detect a moving target in a stationary background, and achieve good detection performance even if the target is occluded in some frames. However, sequence-based algorithms are usually time-consuming and have limitations in some applications (for example, when the target is stationary and/or the background is changing fast). By contrast, the frame-based algorithms only use the information within a single frame and usually have a faster detection speed. Moreover, frame-based algorithms can be the basic module in some sequence-based algorithms; hence, we focus on frame-based detection algorithms in this study.

Of the several frame-based detection algorithms, those based on background estimation first divide the raw image into background and foreground, and then detect small targets in the foreground. In theory, the background estimation algorithms are consistent with the imaging model. Moreover, they are simple in structure and easy to implement, attracting extensive attention.

Existing background estimation algorithms can be divided into nonlocal and local background estimations [11]. The nonlocal background estimation algorithms decompose the entire image into the background and foreground directly. For example, the wavelet transform [12] and Butterworth high-pass filter [13] decompose the image through some frequency-domain methods; the sparse representation [14] and the robust principal component analysis [15] decompose the whole image under constraints of sparse and low rank. However, these algorithms require huge computations since the whole image information is needed to estimate the background of each pixel.

By contrast, the local background estimation algorithms first use local neighborhood pixel information (but not the whole image) of each pixel to estimate the background and, then, get the foreground by calculating the difference information between the raw image and the background. Some of them are also called local contrast methods. Compared with nonlocal algorithms,

Manuscript received August 18, 2020; revised October 10, 2020 and October 31, 2020; accepted November 7, 2020. Date of publication November 17, 2020; date of current version January 6, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61802455 and in part by the Foundation of the Science and Technology Department of Henan Province under Grant 192102210089. *(Corresponding author: Jinhui Han.)*

Jinhui Han, Xiaojian Zhang, and Qifeng Niu are with the College of Physics and Telecommunication Engineering, Zhoukou Normal University, Zhoukou 466001, China (e-mail: hanjinhui@zknu.edu.cn; xiaojiandr@yeah.net; niuqifeng2011@163.com).

Chengyin Liu is with Wuhan Electronic Information Institute, Wuhan 430019, China (e-mail: kworms@foxmail.com).

Yuchun Liu is with the School of Mechanical and Electrical Engineering, Zhoukou Normal University, Zhoukou 466001, China (e-mail: lycdgp@163.com).

Zhen Luo is with the School of Network Engineering, Zhoukou Normal University, Zhoukou 466001, China (e-mail: zluo@zknu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2020.3038442

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

local background estimation algorithms are easier in both theory and application and are more popular.

Numerous local background estimation algorithms have been proposed to date. For example, the Laplacian of Gaussian filter [16], [17] and the difference of Gaussian (DoG) filter [18]–[21] use the weighted sum of neighborhood pixels as the background; however, they are sensitive to edges. To suppress the edges better, the improved difference of Gabor (IDoGb) filter [22] and accumulated center-surround difference measure [23] divide the local area into eight orientations, calculate the weighted sum of neighborhood pixels in different orientations and, then, take their maximum value as the background.

Several algorithms have adopted the idea of orientation segmentation. For example, the max-mean and max-median filters [24] first calculate the mean and median values in different lines, respectively, then take the maximum value as the background. The local contrast measure (LCM) [25], improved LCM (ILCM) [8], and multiscale patch-based contrast measure (MPCM) [26] adopt an image patch with a 3×3 cell estimation window. They first calculate the average value of the eight surrounding cells and, then, take the maximum value as the background. The relative LCM (RLCM) [27] and novel LCM (NLCM) [28] use some largest pixels in different cells to calculate the average value. The 2-D least mean square (TDLMS) [29]–[31] and the multidirectional 2-D least mean square (MDTDLMS) [32] use iterations to estimate the background according to the principle of least mean square error. However, since the target size is usually unknown in real applications, it is necessary to adjust the estimation window size to perform multiscale calculations, making the calculation very complex.

Recent research considers the stages during and after estimation to achieve better performance and uses some weighting functions to further suppress clutters. For example, the weighted local difference measure (WLDM) [33]–[35] introduced the local entropy as a weighting function; the local self-similar [36] used self-similarity as a weighting function; the homogeneity weighted LCM (HWLCM) [37] used local homogeneity as a weighting function; the regional intensity level (RIL) [38] used the difference of RIL as a weighting function. However, the weighting function calculation is usually complex since the window size needs to be adjusted to deal with unknown size targets.

In general, existing local background estimation algorithms for IR small target detection still have some drawbacks. First, to deal with unknown size targets, the estimation window should be adjusted to perform multiscale detection and calculate the weighting function, which requires a lot of calculations. Second, there has been much research on the stages during and after estimation, but the significant effect of the preprocessing before background estimation is usually overlooked. Third, during estimation, most algorithms typically take the maximum value of different orientations as the background. However, if a dim target is adjacent to high-brightness background, it will be easily submerged by the background.

In this study, an enhanced closest-mean background estimation (ECMBE) method is proposed and carefully designed before, during, and after estimation. First, a three-layer estimation window consisting of a central layer, an isolating layer, and a surrounding layer is proposed to deal with unknown size targets by single-scale calculation. Before estimation, the idea of matched filter [39] is adopted and a Gaussian filter is applied in the central layer to improve the image signal-to-noise ratio (SNR), so as to detect the target more easily. During estimation, the closest-mean principle is proposed for the eight orientations of the surrounding layer to suppress high-brightness background. After estimation, a ratio-difference operation is used between the central and surrounding layers to enhance true target and suppress background simultaneously, and a simple weighting function utilizing the isolating layer is proposed as a checking mechanism to suppress clutters further.

The contributions of this article can be summarized as follows.

1) A new estimation window consisting of three layers is proposed. The isolating layer setting ensures that it can deal with small targets of different sizes using only single-scale calculation.
2) Not only the stages during and after the background estimation, but also the stage before estimation are fully considered, and the image SNR is purposefully improved according to the shape of the true target.
3) Aiming at the problem of target submergence caused by the high-brightness background, the closest-mean principle for different orientations of the surrounding layer is proposed, effectively alleviating the problem.
4) A weighting function utilizing the isolating layer is proposed that does not require estimation window size adjustment according to the target size, making the calculations simpler.

Experiments on some real and simulated IR images demonstrate the effectiveness of the proposed method in successfully detecting small targets under complex background. Compared to existing state-of-the-art algorithms, the proposed method has advantages in both target enhancement and background suppression and achieves good detection performance in terms of detection rate and false alarm rate. Besides, the computation is reduced significantly.

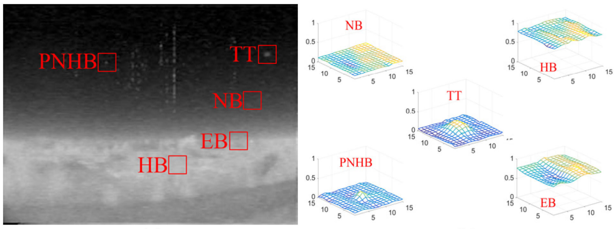
The organization of rest of this article is as follows. In Section II, some related research works are introduced, such as the imaging model and theoretical basis of the background estimation methods, features of the different components in the IR image, and some most common estimation windows. In Section III, the calculation of the ECMBE is described in detail. In Section IV, the detection ability of the proposed algorithm is analyzed, and the threshold operation for small target detection is described. Simulations and experimental results are presented in Section V, and the conclusion is presented in Section VI.

II. Related Research Works

A. Theoretical Basis of Background Estimation Methods

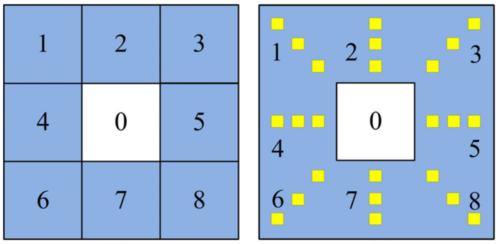
According to the widely used imaging model, a raw IR image consists of three parts, the background image, the target image, and the noise image [40]

$$I(x,y) = I_B(x,y) + I_T(x,y) + I_N(x,y) \tag{1}$$



(a) (b)

Fig. 1. (a) Typical real IR image sample. (b) 3-D distributions of different types of components in a local 15 × 15 area. Here, TT represents true small target, NB represents normal background, HB represents high-brightness background, EB represents edge of background, and PNHB represents pixel-sized noise with high brightness.



(a) (b)

Fig. 2. (a) Common double-layer estimation window for background estimation, in which the surrounding layer is divided into eight orientations. (b) Another common double-layer estimation window, in which only some active pixels (the yellow squares) in the surrounding layer are used.

where (x, y) is the coordinate of each pixel in the image; I is the raw image; I_B is the image with backgrounds, such as clouds, buildings, and sea waves and is usually self-correlated; I_T is the image containing the true small target and is usually sparse and independent of I_B; I_N is a random noise image.

Traditional background estimation algorithms first estimate I_B using some methods, then subtract I_B from the raw image to eliminate background (some algorithms choose to use the ratio of the raw image and I_B to enhance target), and the target can be found in the result. In theory, the background estimation algorithms are consistent with the imaging model, and they are simple in structure and easy to implement. However, random noise can be a serious interference factor, needing special attention.

B. Features of Different Types of Components

To intuitively explain the differences between true small target and various interferences, a typical sample of real IR image [41] is given in Fig. 1(a). The normalized 3-D local distribution of the components in Fig. 1(a) is given in Fig. 1(b), including true target (denoted by TT), normal background (denoted by NB), high-brightness background (denoted by HB), edge of background (denoted by EB), and PNHB.

From Fig. 1, we derive the following conclusions.

1) TT is usually brighter than its immediate neighbors, since the target is usually hotter in most practical applications. In addition, due to the optics point spread function (PSF) of the detector [42], TT usually has a small area between 3 × 3 and 9 × 9 (a target larger than 80 pixels is no longer a small target, which is beyond the scope of this study). It attenuates from the center without anisotropy, similar to the 2-D Gaussian function [33].
2) NB is usually dark and flat, so the difference between the central position and its neighborhood pixels is not obvious.
3) HB is also flat, although it may have a large gray value (probably much larger than TT). Therefore, the difference between the central position and its neighborhood pixels is also not obvious.
4) EB has different gray values on each side. However, in a small local area, EB is usually distributed in a specific direction, significantly different from TT in the directional feature.
5) The pattern of PNHB is similar to TT. However, PNHB is usually caused by random factors, and only emerges as a single pixel, different from TT.

C. Most Common Estimation Windows

Fig. 2 gives the two most common estimation windows used in existing algorithms. They both have two layers, in which the central layer is used to mask the whole target, and the surrounding layer is used to capture the background near the target. Especially, in Fig. 2(b), only some active pixels in different orientations are used, which is more advantageous in processing stripe noise in the IR image (more common in images obtained by scanning devices [43]).

There are two main defects in the two windows in Fig. 2. First, when the target size is unknown, it will be necessary to adjust the window size to carry out multiscale detection, which requires many calculations; additionally, the calculation of the weighting function will be very complex. Second, to suppress background edges better, the surrounding layer is divided into eight orientations, and the maximum value in the eight orientations will be taken as the final background estimation of the central position of the window. However, if a dim target is adjacent to high-brightness background, it will be easily submerged by the background.

III. Calculation of ECMBE

In this section, based on the analysis in Section II, a new estimation window, which consists of three layers, is proposed first. Subsequently, the Gaussian filtering of the central layer (before estimation) and closest-mean principle of the surrounding layer (during estimation) are described in detail. Finally, the weighting function utilizing the information of the isolating layer is introduced briefly, and the definition of the ECMBE is finally given (after estimation).

A. Proposed Estimation Window

This article proposes a new estimation window consisting of three layers, as shown in Fig. 3. The central layer is used

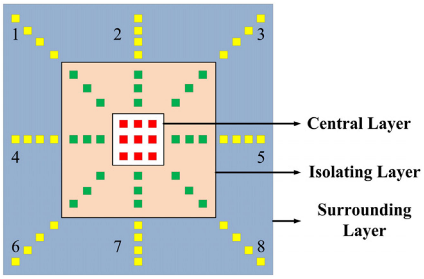


Fig. 3. Proposed three-layer estimation window.

to capture the target, which can be the whole target (for target smaller than the central layer) or the main energy around the target center (for target larger than the central layer). The surrounding layer is used to capture the local neighborhood background pixels and is divided into eight orientations for edge suppression. The isolating layer separates the target from its neighborhood background when the target is larger than the central layer.

The main improvement of the new estimation window is the setting of the isolating layer. As described in Section II-C, existing algorithms usually use a double-layer window, in which the central layer is used to capture the whole target. However, in practical applications, the target size is usually unknown; hence, existing algorithms must adjust the window size to carry out multiscale detection, making the calculation very complex. In this study, by setting an isolating layer between the central and surrounding layer, the estimation window does not need to be adjusted as long as the small target does not exceed the isolating layer, and small targets of unknown sizes can be processed by single-scale calculation. It simplifies not only the estimation process, but also the calculation of the weighting function.

B. Gaussian Filtering of the Central Layer

As described in Section II-A, random noise is a serious interference factor in background estimation methods. Hence, it is better to suppress noise before estimation, but existing algorithms usually do not fully consider the stage before estimation. In this study, a filtering operation is applied in the central layer to suppress random noise and improve the image SNR before estimation.

According to the matched filter theory, SNR can be improved best when the filter template is similar to the signal shape [39]. From the analysis in Section II-B, it can be seen that the true target usually has a 2-D Gaussian shape near its center, so a typical normalized Gaussian filtering template (as seen in Fig. 4) is first applied to the central layer. The filtering result of the central layer is defined as

$$I_{\mathrm{GS}}(x,y)=\sum_{l=-1}^{1}\sum_{k=-1}^{1}\mathrm{GS}(l,k)I(x+l,y+k) \tag{2}$$

where (x, y) is the coordinate of the central pixel in the central layer, I is the raw image, GS is the Gaussian template in Fig. 4, and I_{GS} is the filtering result.

$$\frac{1}{16}\begin{array}{|c|c|c|}\hline 1 & 2 & 1 \\ \hline 2 & 4 & 2 \\ \hline 1 & 2 & 1 \\ \hline\end{array}$$

Fig. 4. Normalized 2-D Gaussian filter template.

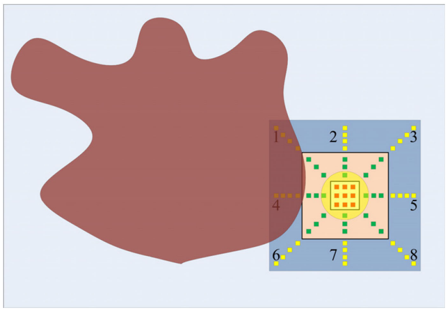


Fig. 5. Special case where the dim target (the yellow circle) is adjacent to the bright background (the brown area). According to the max-mean principle, orientation 1 or 4 will be output as the estimated background. But according to the closest-mean principle, other orientations will be output.

C. Closest-Mean Background Estimation (CMBE) of the Surrounding Layer

The surrounding layer of the estimation window is used to capture the local neighborhood background pixels. For each orientation, to suppress random noise, the background estimation (BE) value is defined as the mean gray of the active pixels, i.e., the yellow pixels in Fig. 3

$$\mathrm{BE}_i(x,y)=\frac{1}{p}\sum_{j=1}^{p}\mathrm{APIX}_j^i, i=1,2,\ldots,8 \tag{3}$$

where (x, y) is the coordinate of the central pixel in the central layer, i denotes the ith orientation, APIX_j^i is the gray value of the jth active pixel in the ith orientation, and p is the number of active pixels in each orientation.

After calculating each orientation, existing algorithms usually take the maximum value of different orientations as the final background estimation value, so that the background edges can be suppressed. However, when the target gray is very dim and background gray is very high and they are adjacent to each other, the background estimation value will be very high, and the target will be easily submerged, as seen in Fig. 5.

In this study, the closest-mean principle for different orientations of the surrounding layer is proposed to alleviate this problem. The CMBE of a current position is defined as

$$\mathrm{CMBE}\,(x,y)=\underset{\mathrm{BE}_i(x,y)}{\arg\min}\left|\mathrm{BE}_i(x,y)-I_{\mathrm{GS}}(x,y)\right|. \tag{4}$$

In short, the closest-mean principle can be described as: From the mean values of different orientations, select the value closest to the central value as the final estimated background, so that high-brightness background will not participate in the calculation of the next stages, as seen in Fig. 5.

D. Weighting Function Utilizing the Isolating Layer

From the analysis in Section II-B, we find that a true target usually attenuates from the center. Therefore, in this study, a simple weighting function utilizing the information of the isolating layer is proposed to suppress clutters further

$$\mathrm{IL}_i(x,y) = \frac{1}{q}\sum_{j=1}^{q}\mathrm{IPIX}_j^i, i = 1, 2, \ldots, 8 \tag{5}$$

$$W(x,y) = \begin{cases} 1, & \text{if } I_{\mathrm{GS}}(x,y) \geq \max[\mathrm{IL}_i(x,y)] \\ 0, & \text{otherwise} \end{cases} \tag{6}$$

where (*x*, *y*) is the coordinate of the central pixel in the central layer, *i* denotes the *i*th orientation, IPIX_j^i is the gray value of the *j*th inactive pixel (i.e., the green pixels in Fig. 3) in the *i*th orientation, totally there are *q* inactive pixels in each orientation in the isolating layer. *W* is the weighting parameter, set to 0 when the central value is smaller than the isolating layer to suppress clutters further.

E. Calculation of ECMBE

The basic idea of the background estimation method is to obtain the foreground by calculating the ratio or difference between the raw image and estimated background, in which the ratio operation can enhance the true target, and the difference operation can eliminate background [27]. In addition, a weighting function is used to achieve better performance sometimes. In this study, both ratio and difference operations are used, along with a weighting function, so the ECMBE of a pixel is defined as

$$\mathrm{ECMBE}(x,y) = R(x,y)D(x,y)W(x,y) \tag{7}$$

where

$$R(x,y) = \frac{I_{\mathrm{GS}}(x,y)}{\max[\mathrm{CMBE}(x,y), \xi]} \tag{8}$$

$$D(x,y) = \max[I_{\mathrm{GS}}(x,y) - \mathrm{CMBE}(x,y), 0] \tag{9}$$

and *W* is the weighting function in (6). There is a small value ξ (in this article, it is 5 for an 8-b digital image) in the ratio operation to avoid a denominator of 0, and a nonnegative constraint in the difference operation to suppress clutters.

Apply the estimation window on the raw IR image from left to right and top to bottom pixel by pixel, the ECMBE is calculated for each pixel, and a new matrix named saliency map (SM) is formed. Algorithm 1 gives the main steps for the ECMBE calculation.

IV. Discussions and Threshold Operation

It is necessary to discuss the following different cases when (*x*, *y*) is different types of pixels.

1) If (*x*, *y*) is a TT: Because TT is usually brighter than its immediate neighbors and attenuates from the center, it can be easily deduced that

$$I_{\mathrm{GS}}(x,y) > \mathrm{CMBE}(x,y) \tag{10}$$

and

$$I_{\mathrm{GS}}(x,y) \geq \max[\mathrm{IL}_i(x,y)] \tag{11}$$

as long as it does not exceed the isolating layer. Hence

$$R(x,y) > 1 \tag{12}$$

$$D(x,y) > 0 \tag{13}$$

$$W(x,y) = 1 \tag{14}$$

and

$$\mathrm{ECMBE}(x,y) > 0. \tag{15}$$

Here, (12) means that the true target can be enhanced.

2) If (*x*, *y*) is an NB: Since background is usually flat, there will be

$$I_{\mathrm{GS}}(x,y) \approx \mathrm{CMBE}(x,y). \tag{16}$$

Hence

$$D(x,y) \approx 0 \tag{17}$$

and

$$\mathrm{ECMBE}(x,y) \approx 0. \tag{18}$$

Here, (17) means that the flat background can be eliminated.

3) If (*x*, *y*) is an HB: Because HB is similar to NB

$$\mathrm{ECMBE}(x,y) \approx 0. \tag{19}$$

4) If (*x*, *y*) is an EB: If (*x*, *y*) is on the higher side, according to the closest-mean principle, the estimated background will be high and

$$I_{\mathrm{GS}}(x,y) \approx \mathrm{CMBE}(x,y). \tag{20}$$

Hence

$$D(x,y) \approx 0 \tag{21}$$

and

$$\mathrm{ECMBE}(x,y) \approx 0. \tag{22}$$

If (*x*, *y*) is on the darker side, the estimated background will be low and the result will be similar to (20)–(22).

1) If (*x*, *y*) is a PNHB: If (*x*, *y*) is located in a flat background, it will be similar to TT, i.e.

$$\mathrm{ECMBE}(x,y) > 0. \tag{23}$$

However, it can be easily deduced that

$$\mathrm{ECMBE}_{\mathrm{PNHB}} < \mathrm{ECMBE}_{\mathrm{TT}} \tag{24}$$

when the PNHB has a gray value equal to or slightly larger than TT, because PNHB usually emerges as a single pixel and can be suppressed to some extent by Gaussian filtering.

Fig. 5 shows a special case where the target is adjacent to the high-brightness background, which has a gray value far larger than the true target. Subsequently, according to the closest-mean principle, the high-brightness background in orientation 1 and

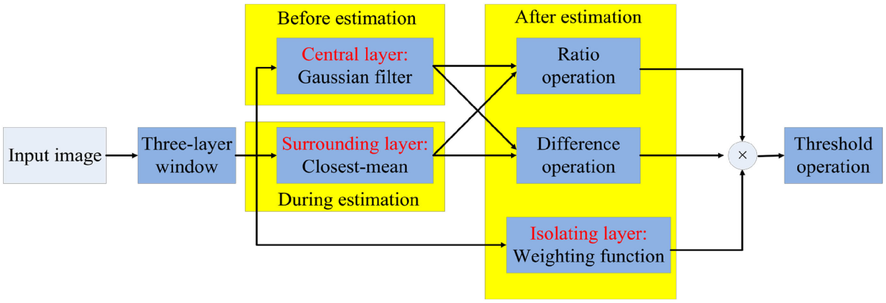


Fig. 6. Flowchart of the proposed algorithm.

```
Algorithm 1: ECMBE Calculation.
Input: A raw input image $I$ with resolution $M \times N$, and the new estimation window.
Output: The saliency map SM.
for $x = 1{:}M$
    for $y = 1{:}N$
        Place the new estimation window on the image, and take $(x, y)$ as its center pixel.
        Calculate $I_{\mathrm{GS}}(x, y)$ for the central layer, according to (2).
        Calculate $\mathrm{CMBE}(x, y)$ for the surrounding layer, according to (3) and (4).
        Calculate $W(x, y)$ for the isolating layer, according to (5) and (6).
        Calculate $\mathrm{ECMBE}(x, y)$, according to (7)–(9).
    end
end
Form the results as a new matrix SM, and normalize it to (0, 1)
```

```
Algorithm 2: Threshold Operation.
Input: The saliency map SM with resolution $M \times N$, and the parameter $\lambda$.
Output: The target positions.
Calculate the threshold Th according to (25).
for $x = 1{:}M$
    for $y = 1{:}N$
        if $\mathrm{SM}(x, y) < \mathrm{Th}$
            $\mathrm{SM}(x, y) = 0$
        else
            $\mathrm{SM}(x, y) = 1$
        end
    end
end
Apply a dilation operation on SM to suppress clutters.
Output the connected areas in SM as target positions.
```

4 will be ignored, and the darker background in other orientations will be taken as CMBE. Therefore, the target will not be submerged, and (10)–(15) are still true.

From the discussions earlier, it can be seen that the true target will be the most salient in SM. Therefore, an adaptive threshold operation will be used to extract the true target. In this study, the threshold Th is defined as

$$\mathrm{Th} = \lambda \times \mathrm{max}_{\mathrm{SM}} + (1 - \lambda)\mathrm{mean}_{\mathrm{SM}} \tag{25}$$

where $\mathrm{max}_{\mathrm{SM}}$ and $\mathrm{mean}_{\mathrm{SM}}$ are the max and mean value of SM, respectively. λ is a parameter in the range 0–1. Our experiments show that a λ between 0.7 and 0.9 will be appropriate for most single-target cases; however, for multitarget cases, λ should be set to a smaller value since different targets may have different saliency.

In SM, the pixels with larger values than Th will be labeled as 1, or else as 0. In the final binary result, each connected area will be taken as a detected target (to eliminate clutters, a dilation operation may be needed).

Algorithm 2 summarizes the main steps for the threshold operation, and Fig. 6 shows the flowchart of the proposed method.

V. Experimental Results

In this section, 11 real IR sequences, a simulated sequence, and a single-frame dataset are used to verify the performance of the proposed algorithm. First, the processing steps of the proposed algorithm are given. To further illustrate the effectiveness of the proposed algorithm, comparisons between the proposed and state-of-the-art algorithms are given. Finally, the noise immunity of the proposed algorithm is tested. All the experiments are conducted on a computer with 8-GB random access memory and 2.6-GHz Intel i5 processor, and the codes are implemented in MATLAB R2016b.

A. Testing Data Used in This Article

Eleven real IR sequences with different targets and backgrounds are used for the experiments. The samples of these sequences are shown in Fig. 7, and the details are listed in Table I. From Fig. 7 and Table I, it can be seen that in the raw IR images, the targets are usually very small and dim, but they still have a small area (equal to or larger than 3 × 3, but smaller than 7 × 7 in most cases), and the backgrounds are usually very complex. Additionally, there are heavy noises in some images.

To verify the detection performance when the target is adjacent to high-brightness background, a simulated sequence that

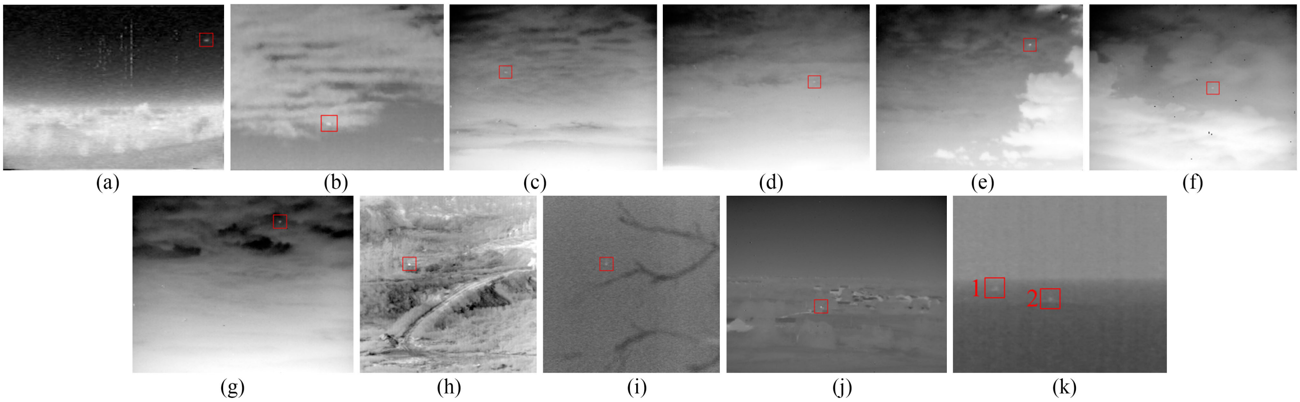


Fig. 7. Samples of the 11 real IR sequences. (a)–(k): Seq. 1–Seq. 11.

TABLE I
FEATURES OF THE 11 REAL IR SEQUENCES

	Frames	Image Resolution	Target ID	Target Size	Target Details	Background Details
Seq. 1	200	320 × 240	Only 1	Varies from 7 × 5 to 4 × 3	• Plane target. • Long imaging distance. • Located in homogeneous sky. • Keeping little motion.	• Sky-Cloud background. • Heavy clutters. • Heavy random noises. • Many bad pixels. • Almost unchanged.
Seq. 2	70	256 × 200	Only 1	Varies from 4 × 5 to 4 × 10	• Plane target. • Long imaging distance. • Target size changes. • Moving from homogeneous sky to cloud.	• Sky-Cloud background. • Heavy clutter. • Change from cloudy sky to homogeneous sky.
Seq. 3	200	320 × 256	Only 1	3 × 5	• Plane target. • Long imaging distance. • Located in cloudy sky. • Keeping little motion.	• Sky-Cloud background. • Heavy clutters. • Some bad pixels. • Almost unchanged.
Seq. 4	200	320 × 256	Only 1	3 × 3	• Plane target. • Long imaging distance. • Located in cloudy sky. • Keeping little motion.	• Sky-Cloud background. • Heavy clutters. • Some bad pixels. • Almost unchanged.
Seq. 5	200	320 × 256	Only 1	4 × 5	• Plane target. • Long imaging distance. • Located in cloudy sky. • Keeping little motion.	• Sky-Cloud background. • Heavy clutters. • Some bad pixels. • Almost unchanged.
Seq. 6	200	320 × 256	Only 1	3 × 3	• Plane target. • Long imaging distance. • Located in cloudy sky. • Keeping little motion.	• Sky-Cloud background. • Heavy clutters. • Many bad pixels. • Change slowly.
Seq. 7	200	320 × 256	Only 1	3 × 5	• Plane target. • Long imaging distance. • Located in cloudy sky. • Keeping little motion.	• Sky-Cloud background. • Heavy clutters. • Many bad pixels. • Almost unchanged.
Seq. 8	300	256 × 256	Only 1	Varies from 3 × 3 to 3 × 4	• Unmanned aerial vehicle target. • Located in sky. • Moving fast.	• Ground-Tree background. • Heavy clutters. • Changed fast.
Seq. 9	90	256 × 256	Only 1	3 × 5	• Truck target. • Long imaging distance. • Located in homogeneous ground. • Keeping little motion.	• Ground-Tree background. • Heavy clutters. • Changed slowly.
Seq. 10	300	320 × 256	Only 1	3 × 3	• Car target. • Located in ground. • Keeping little motion.	• Building-Tree background. • Heavy clutters. • Almost unchanged.
Seq. 11	200	280 × 228	Target 1	5 × 6	• Ship target. • Long imaging distance. • Located in homogeneous sea. • Two targets, one is moving and the other is stationary.	• Sea-Sky background. • Heavy clutter. • Almost unchanged.
			Target 2	5 × 5		

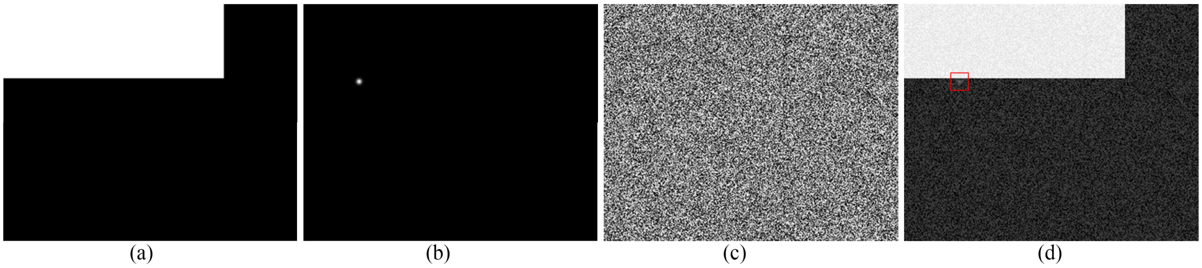


Fig. 8. Generation of the simulated images (the 42nd frame is presented), here I_{NB} is set to 100, I_{HB} is set to 200, I_{MAX} is set to 20, σ is set to 2, and σ_N is set to 20. (a) Background part I_{BK}. (b) Target part I_{TGT}, a 5 × 5 target is shown here. (c) Noise part I_{NOI}. (d) Simulated image *I*.

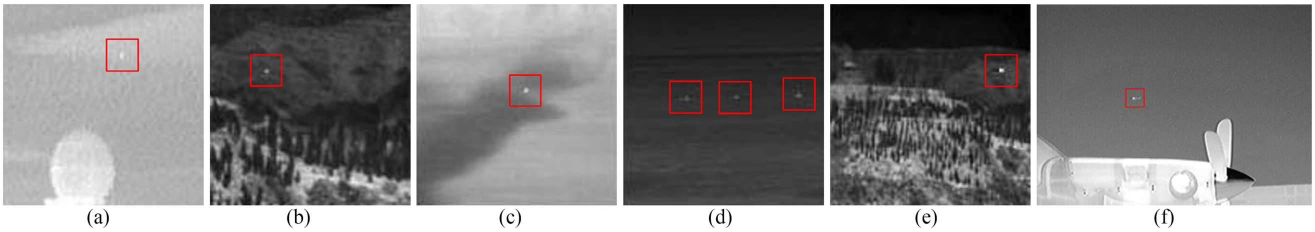


Fig. 9. Samples of the single-frame dataset.

contains 200 frames is generated. Using the widely used 2-D Gaussian model, the steps of generating simulated images are listed as follows.

1) Generate a 256 × 320 matrix with a uniform value I_{NB} as the normal background. Subsequently, generate a high-brightness background area (upper left corner with area 80 × 240) with a gray value I_{HB} in it. Denote this matrix by I_{BK}.
2) Generate a 256 × 320 matrix, which contains a simulated small target, denoted by I_{TGT}. In this study, we use the 2-D Gaussian model to generate a small target, which can be described as

$$I_{\mathrm{TGT}}(x,y) = I_{\mathrm{MAX}} \exp\left(-\frac{(x-i)^2+(y-j)^2}{2\sigma^2}\right) \tag{26}$$

where I_{MAX} is the maximum gray value of the target, (*i*, *j*) is the target center, and σ determines the target size.
3) Generate a 256 × 320 matrix, which contains random Gaussian white noises with standard deviation σ_N; denote this matrix as I_{NOI}.
4) Generate a simulated image, according to

$$I = I_{\mathrm{BK}} + I_{\mathrm{TGT}} + I_{\mathrm{NOI}}. \tag{27}$$

In this study, I_{NB} is set to 100, I_{HB} is set to 200, I_{MAX} is set to 20, σ is set to 2 (the target size is about 5 × 5), and σ_N is set to 20. The target is placed at location (84, 20) in the first frame and moved from left to right one pixel per frame.

A sample of the simulated images is shown in Fig. 8.

Besides the 12 sequences, a single-frame dataset (23 frames), which contains different types of targets and backgrounds, is used in some tests, too. Some samples are shown in Fig. 9.

B. Processing Results of the Proposed Algorithm

First, we tested the detection ability of the proposed algorithm. Three parameters are important and need to be discussed, i.e., the sizes of the central, isolating, and surrounding layers.

The central layer is used to capture the main energy of a target. A true target usually has an area larger than 3 × 3 due to the optics PSF of the detector [42]; hence, we set the central layer to 3 × 3.

The isolating layer is used to separate the target from its neighborhood background when the target is larger than the central layer; hence, it should not be smaller than the maximum target size in practical applications. However, if it is too large, the computational area will be very large and unnecessary clutters may be introduced, reducing the detection ability. Hence, the isolating layer is set to be equal to the general maximum size of the small target; 7 × 7 or 9 × 9 are both suitable values. In this study, considering the target size is usually smaller than 7 × 7 (see Table I), we set the isolating layer to 7 × 7, i.e., *q* in (5) is set to 2.

The surrounding layer is used to capture the neighboring background near a target. The larger the surrounding layer, the more active pixels will be considered in the mean operation in (3) to achieve better suppression of the random white noise. However, if the surrounding layer is too large, the computational area will be very large, and some unnecessary clutters may also be introduced, reducing the detection ability. Several experiments have been conducted to achieve a balance between noise suppression and clutter introduction. The results show that a surrounding layer between 11 × 11 and 17 × 17 will be appropriate. In this study, we set the surrounding layer to 13 × 13, i.e., *p* in (3) is set to 3.

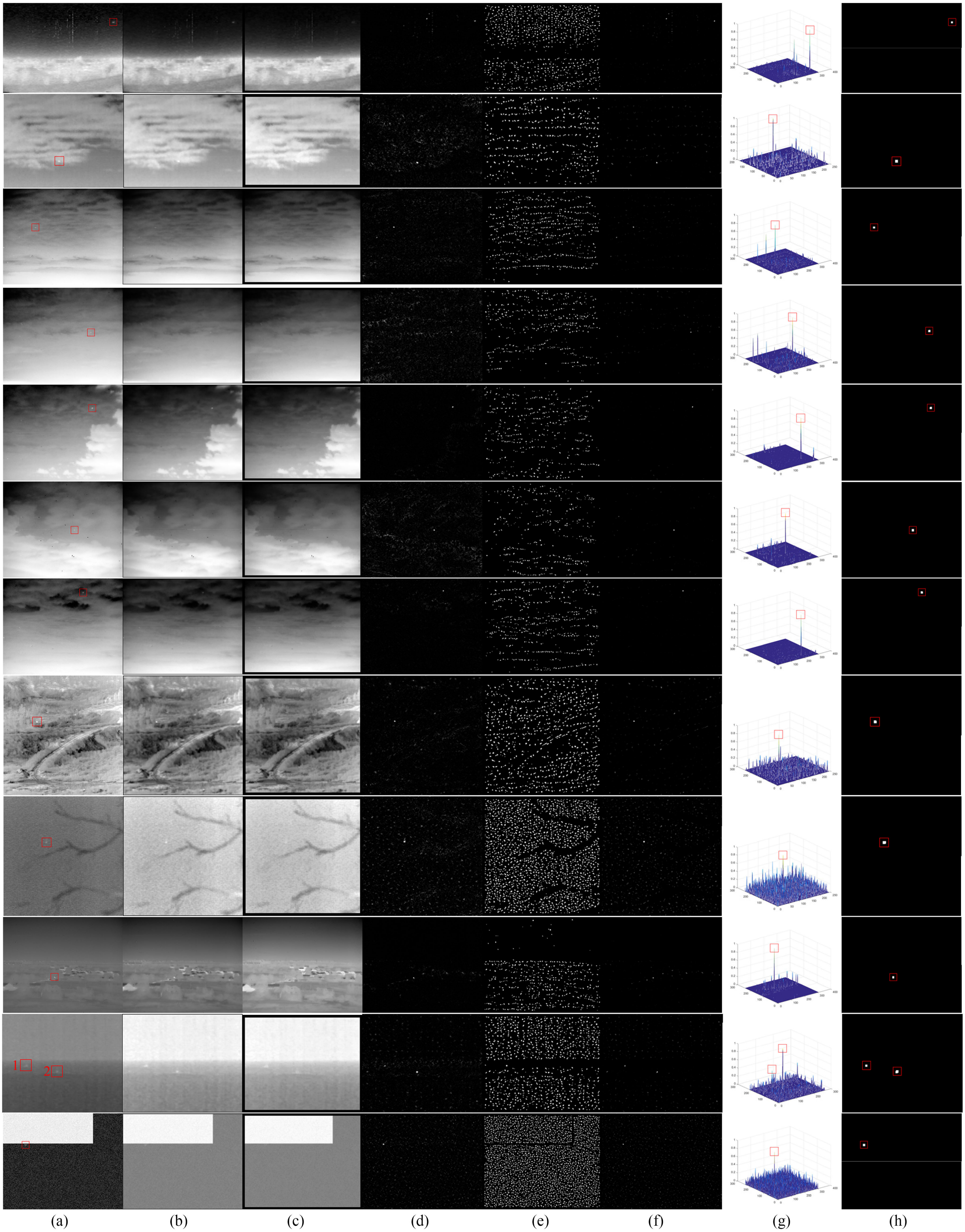


Fig. 10. Processing results of the proposed algorithm for the samples of the 12 sequences (from top to bottom: Seq. 1-Seq. 11, and the simulated sequence). (a) Raw IR images. (b) Images after Gaussian filtering. (c) Estimated backgrounds. (d) Results of ratio-difference operations between (b) and (c). (e) Weighting functions. (f) Final saliency maps of ECMBE. (g) 3-D display of (f). (h) Detection results after threshold operation.

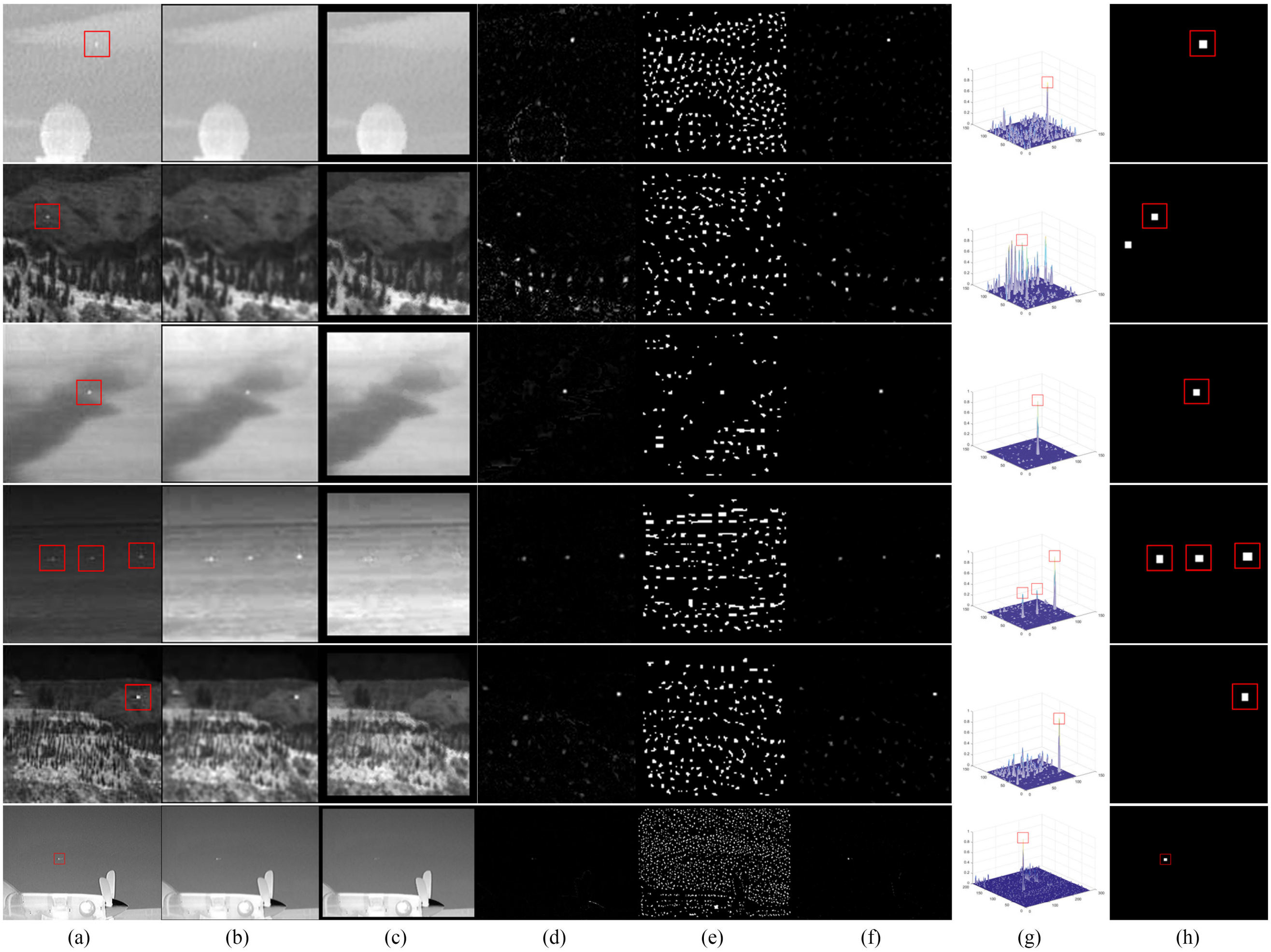


Fig. 11. Processing results of the proposed algorithm for the samples of the single-frame dataset. (a) Raw IR images. (b) Images after Gaussian filtering. (c) Estimated backgrounds. (d) Results of ratio-difference operations between (b) and (c). (e) Weighting functions. (f) Final saliency maps of ECMBE. (g) 3-D display of (f). (h) Detection results after threshold operation.

Fig. 10 shows the step-by-step processing results of the proposed algorithm for the 12 sequences. The same sample images used in Figs. 7 and 8 are given.

Following can be seen from Fig. 10.

1) The Gaussian filter can improve the image quality to a certain extent first, as seen in Fig. 10(b).
2) In the estimated background, the target position is different from the original data, as seen in Fig. 10(c).
3) The ratio-difference calculation between the filtered image and estimated background shows that the target is very salient but with some residues of complex background, as seen in Fig. 10(d).
4) The weighting function, as seen in Fig. 10(e), can suppress the complex background further, making the target more salient after weighting, as seen in Fig. 10(f) and (g).
5) After the threshold operation, all the real targets are found successfully without any false targets, which prove the effectiveness of the proposed algorithm, as seen in Fig. 10(h).

To verify the effectiveness of the proposed algorithm in more different types of targets and backgrounds, Fig. 11 gives the processing results of the proposed algorithm for the single-frame dataset, and the same sample images with Fig. 9 are given here. It can be seen that all the real targets are output, and only one false target is output when the background is very complex.

C. Comparisons With Other Algorithms

To verify the advantage of the proposed algorithm, eight state-of-the-art algorithms are chosen for comparison, namely DoG [18], ILCM [8], NLCM [28], MPCM [26], RLCM [27], WLDM [33], MDTDLMS [32], and the recent, multiscale trilayer LCM (TLLCM) [44].

1) DoG is considered as a traditional background estimation method without orientation segmentation.
2) ILCM and NLCM are background estimation methods with orientation segmentation; however, they use the maximum value in different orientations as the estimated background.
3) MPCM is a multiscale method.
4) RLCM is a multiscale method in which both ratio and difference operations are used.
5) WLDM is a multiscale weighted method that introduces the local entropy as a weighting function.

TABLE II
PARAMETER VALUES USED IN EXISTING ALGORITHMS

	Parameter values
DoG	A 5×5 binomial kernel as the author recommended. C_{Th}=0.15 for single-target situation and 0.01 for multi-target situation.
ILCM	The same 5×5 binomial kernel with DoG, and the cell size is 8 × 8.
NLCM	The same 5 × 5 binomial kernel with DoG. The cell size is 10 × 10, and K is 4.
MPCM	Three scales with a cell size of 3 × 3, 5 × 5 and 7 × 7 are used.
RLCM	The cell size is 9 × 9. Three scales are used, and (K_1, K_2) is set to (2, 4), (5, 9), and (9, 16), respectively.
WLDM	Entropy window size is 5 × 5. Four scales are used for LDM calculation, with a cell size of 3 × 3, 5 × 5, 7 × 7 and 9 × 9.
MDTDLMS	The inner window is 7 × 7, the outer window is 11 × 11, and μ is set to 10^{-7} for 8-bit images.
TLLCM	Three scales are used, in which s is set to 5, 7, and 9.

TABLE III
SCRG OF DIFFERENT ALGORITHMS

Seq	Target	DoG	ILCM	NLCM	MPCM	RLCM	WLDM	MDTDLMS	TLLCM	Proposed
1	1	6.2885	7.2875	10.1899	7.6446	15.7705	18.5764	51.2914	**102.5329**	*100.6701*
2	1	1.9132	6.0963	5.6193	2.9264	2.5565	7.8443	12.6830	**28.1468**	*23.0421*
3	1	8.3011	13.0714	18.8765	2.8125	11.8627	46.2286	48.9359	**171.0698**	*130.7640*
4	1	15.4798	23.8531	51.7499	2.4423	11.0850	115.9090	81.4976	**276.3283**	*181.2071*
5	1	8.1177	7.4718	10.3798	6.2368	13.2762	75.9957	59.4319	*87.3026*	**92.5705**
6	1	6.1230	1.8996	1.6719	2.5088	11.8251	3.4102	87.0347	**280.0638**	*232.7211*
7	1	13.9985	46.9499	44.1210	8.3706	20.1072	96.6349	134.0081	**231.6447**	*224.2902*
8	1	1.1871	3.9251	7.8695	1.5108	3.3806	8.1257	11.9812	**22.0774**	*17.2770*
9	1	1.5289	2.5832	4.4538	2.9411	2.5795	4.2662	*8.7117*	**30.2972**	8.4955
10	1	0.5052	1.9404	1.6612	2.0665	1.5181	**19.0926**	3.6358	*10.1938*	8.8434
11	1	14.9848	13.6793	14.2903	10.8889	5.4331	**46.9801**	37.9182	42.6979	*44.0365*
	2	9.1796	8.4822	12.0151	9.0312	7.2304	24.8754	34.3315	**59.1686**	*49.6438*
Sim	1	14.4232	0.6540	1.8608	11.6526	**23.1682**	5.1178	0.5058	0	*21.3447*

The bold means the largest value, the italics-underlined means the second largest value.

6) MDTDLMS uses iterations to estimate the background according to the TDLMS principle, then uses both ratio and difference operations between the original image and estimated background. However, iterations cost times.
7) TLLCM uses another trilayer window to estimate the background. However, it is still a multiscale algorithm and costs a lot of computations. Besides, TLLCM uses the maximum value in different orientations as the final estimated background. Moreover, the information of the isolating layer is wasted in TLLCM.

The key parameters of each algorithm are listed in Table II.

1) Comparing Detection Performance: SCR gain (SCRG) and background suppression factor (BSF) are objective indicators that can describe the target enhancement and background suppression ability of an algorithm and defined, respectively, in

$$\mathrm{SCRG} = \frac{\mathrm{SCR}_{\mathrm{out}}}{\mathrm{SCR}_{\mathrm{in}}} \tag{28}$$

$$\mathrm{BSF} = \frac{\sigma_{\mathrm{in}}}{\sigma_{\mathrm{out}}} \tag{29}$$

where $\mathrm{SCR}_{\mathrm{in}}$ and $\mathrm{SCR}_{\mathrm{out}}$ are the SCR [defined as (30)] of the raw image and SM respectively, σ_{in} and σ_{out} are the standard deviation of the raw image and SM, respectively

$$\mathrm{SCR} = \frac{|I_t - I_{nb}|}{\sigma} \tag{30}$$

where I_t is the maximal gray in the 9 × 9 area around the target center, I_{nb} is the average gray of the neighboring background between 15 × 15 and 9 × 9 area around the target center, and σ is the standard deviation of the image.

As seen in Tables III and IV, the proposed algorithm can achieve the highest or second-highest SCRG and BSF in most cases. TLLCM may achieve better performance than the proposed algorithm sometimes. However, as it uses the maximum of different orientations as the estimated background, it fails to detect targets adjacent to high-brightness background. For example, for the simulated sequence (denoted by Sim in the tables), the SCRG of TLLCM is 0, which means that the target is totally submerged.

Figs. 12 and 13 show the saliency map and detection results for the sample images of the 12 sequences using different algorithms. Fig. 14 gives the receiver operating characteristic (ROC) curves [45] of different algorithms for each whole sequence. Here, the false positive rate (FPR) and the true positive rate (TPR) are defined, respectively, as

$$\mathrm{FPR} = \frac{\text{number of detected false targets}}{\text{total number of pixels in the whole image}} \tag{31}$$

$$\mathrm{TPR} = \frac{\text{number of detected true targets}}{\text{total number of real targets}}. \tag{32}$$

Following can be seen from Figs. 12–14.

1) The detection performance of DoG is usually the worst because it is a traditional algorithm without orientation segmentation.

TABLE IV
BSF OF DIFFERENT ALGORITHMS

Seq	DoG	ILCM	NLCM	MPCM	RLCM	WLDM	MDTDLMS	TLLCM	Proposed
1	2.1241	33.9468	1.4121	0.5767	6.1397	681.6723	3.4236E3	*4.3918E3*	**4.4826E3**
2	0.7403	17.4796	1.0426	0.3002	2.7740	112.1502	660.7865	**1.4532E3**	*1.1913E3*
3	3.2310	49.5869	0.0361	0.0202	11.1200	15.1272	1.3470E5	**4.6379E5**	*3.5544E5*
4	7.0589	102.7397	0.1017	0.0294	15.7150	41.2982	1.2874E5	**4.3056E5**	*2.8272E5*
5	3.6008	25.8584	0.0052	0.0107	7.7217	12.5982	3.6880E5	*5.4004E5*	**5.7287E5**
6	2.9415	8.8598	0.0019	0.0191	11.2321	1.9188	1.7400E5	**5.5720E5**	*4.6306E5*
7	5.7067	46.3694	0.0220	0.0192	15.1353	22.8578	3.8940E5	**6.7007E5**	*6.4900E5*
8	0.4343	7.9812	0.0943	0.1349	2.4911	37.9672	1.6373E3	**3.0041E3**	*2.3546E3*
9	0.7329	10.4656	0.8334	0.5660	2.8416	81.8911	*316.0840*	**1.0835E3**	306.9916
10	0.4814	9.1742	0.0344	0.1368	2.1982	16.5951	4.2422E3	**1.1835E4**	*1.0283E4*
11	3.9246	59.0583	10.9018	1.4717	6.1755	*1.6762E3*	992.0782	**1.6974E3**	1.4245E3
Sim	0.6529	16.8767	0.5002	0.4047	4.1183	45.6642	689.5115	**7.2121E3**	*1.4888E3*

The bold means the largest value, the italics-underlined means the second largest value.

2) The performance of ILCM is usually better than DoG, because of the utilization of the directional information to suppress the complex background edges better. However, its performance is still not satisfied in some sequences, for example, Seq. 1, Seq. 3, and Seq. 6.

3) NLCM is similar to ILCM and has similar performances in most sequences.

4) MPCM adopts multiscale detection to achieve a good performance. However, if the background is very complex, some interferences will be enhanced and output, as in Seq. 8 and Seq. 9.

5) RLCM can usually achieve a better detection performance since it utilizes both ratio and difference operations. However, its performance in some sequences is not satisfied, especially when the target is adjacent to high-brightness background, for example, the simulated sequence. This is because RLCM takes the maximum value in different orientations as the final background estimation.

6) WLDM is a multiscale weighted method using the local entropy as the weighting function; however, when the background is very complex, some interferences will be enhanced and output. Sometimes, its performance is even worse than RLCM, which does not have a weighting function.

7) MDTDLMS can achieve a good detection performance in most cases. However, some false alarms emerge in Seq. 1 and the simulated sequence.

8) TLLCM, as a newly published method, can achieve good performance in 11 real sequences. However, in the simulated sequence, it fails because it takes the maximum value in different orientations as the final estimated background.

9) The proposed algorithm can achieve good detection performance. Compared to existing state-of-the-art algorithms, its performance is always at the forefront, and there is no big fluctuation. Particularly, when the target is adjacent to high-brightness background (i.e., the simulated sequence), it is the only algorithm that can still output the true target successfully.

2) Comparison of Computational Complexity and Time Consumption: The computational complexity of the different algorithms is analyzed in this section. For simplicity, suppose the resolution of the IR image is $X \times Y$, and scale of the filter window or cell is $(2L{+}1)^2$. For multiscale algorithms, such as MPCM, WLDM, RLCM, and TLLCM, S represents the scale number, L_i ($i = 1, 2, \ldots, S$) represents the L for the ith scale, and L_S is the largest L.

For DoG, with $(2L{+}1)^2$ multiplications and $(2L{+}1)^2$ additions for each pixel, the computational complexity is $O(L^2XY)$.

For ILCM and NLCM, since they both use DoG in preprocessing, and the latter subblock-stage processing has fewer calculations, their computational complexity will be $O(L^2XY)$.

For MPCM, the average operation will cost $(2L_i{+}1)^2$ additions for each pixel at each scale; hence, its total computational complexity will be $O(SL_S{}^2XY)$.

For RLCM, the sort operation within a cell will cost $(2L_i{+}1)^2\log(2L_i{+}1)^2$ calculations for each scale; hence, its computational complexity will be $O[SL_S{}^2\log(L_S{}^2)XY]$.

For WLDM, the mean operation will cost $(2L_i{+}1)^2$ additions for each pixel at each scale. In addition, the entropy calculation requires a sort operation within a cell, costing $(2L_i{+}1)^2\log(2L_i{+}1)^2$ calculations for each scale; hence, its computational complexity will be $O[SL_S{}^2\log(L_S{}^2)XY]$.

For MDTDLMS, the computational complexity is analyzed in the original paper [32] and is $O(L^2XY)$.

For TLLCM, the computational complexity is analyzed in the original paper [44] and is $O(SL_S{}^2\log(L_S{}^2)XY)$.

The proposed algorithm consists of four steps: Gaussian filtering operation for the central layer, background estimation for the surrounding layer, ratio-difference operation between them, and weighting function using the isolating layer. There are $3 \times 3 = 9$ pixels in the central layer, $7 \times 7 - 9 = 40$ pixels in the isolating layer, and $L - 3 - 1 = L$ - 4 pixels in each orientation of the surrounding layer. Hence, for the Gaussian filtering operation, there will be 9 multiplications and 1 addition, totaling 10 operations; for the background estimation, there will be L - 4 additions and 1 division for the average of each orientation, and 8 comparisons between eight orientations, totaling 8(L - 3) + 8 operations; for the ratio-difference operation, there will be 1 division and 1 subtraction, totaling 2 operations; for the weighting operation, there will be 40 additions and 1 division for the average of the isolating layer, 1 comparison and 1 multiplication, totaling 43 operations. If the resolution of the IR image is $X \times Y$, there will

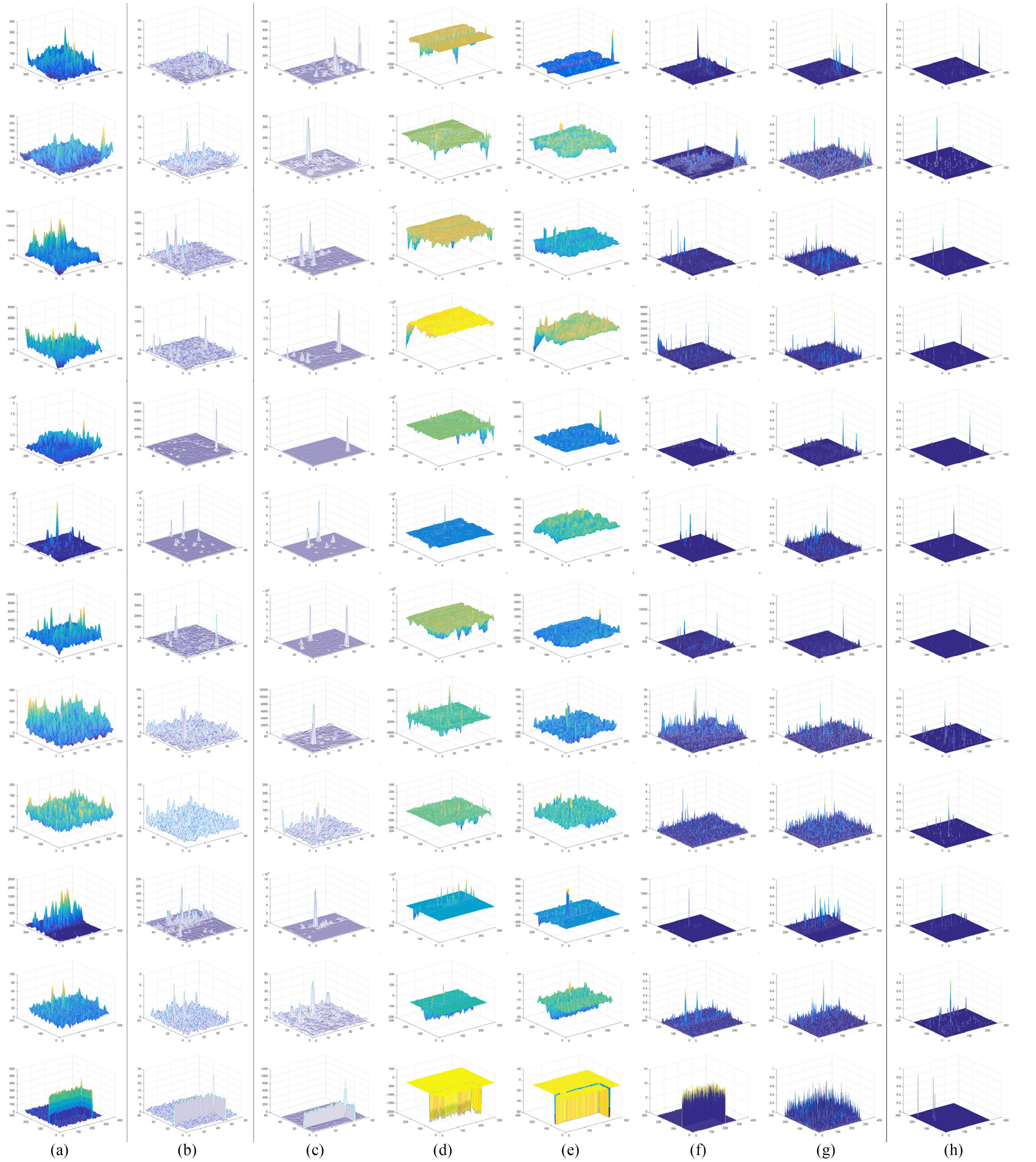


Fig. 12. Saliency maps of the 12 sequences using different algorithms. (a) DoG. (b) ILCM. (c) NLCM. (d) MPCM. (e) RLCM. (f) WLDM. (g) MDTDLMS. (h) TLLCM.

be a total of $[10 + 8(L - 3) + 8 + 2 + 43]XY$ operations for the proposed algorithm. Thus, the computational complexity of the proposed algorithm will be $O(LXY)$. The computational complexity of different algorithms is given in Table V. The proposed algorithm has the smallest computational complexity, which helps to reduce the detection time and improve detection speed.

Table VI gives the average detection time of different algorithms for one frame.

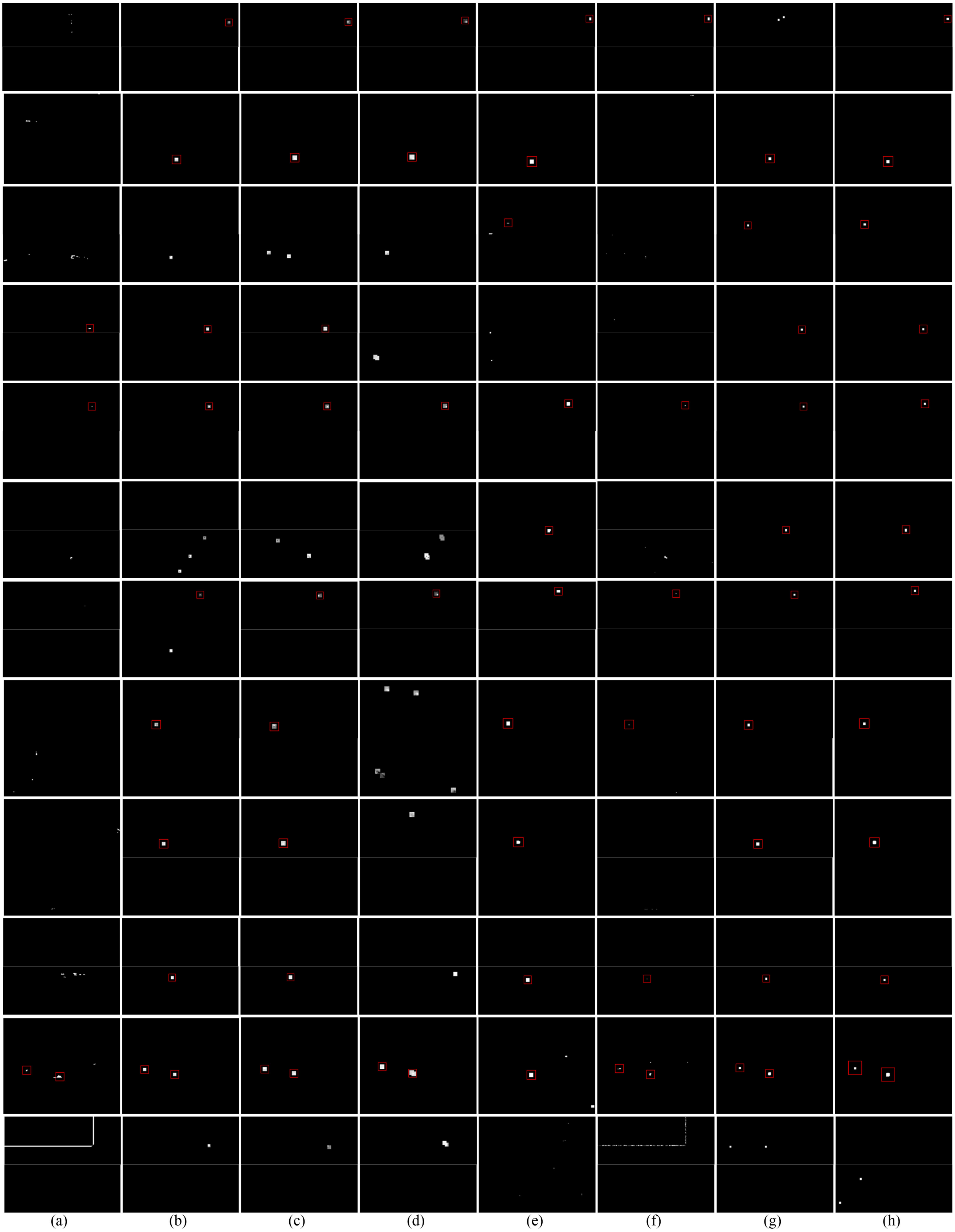


Fig. 13. Detection results of the 12 sequences using different algorithms. (a) DoG. (b) ILCM. (c) NLCM. (d) MPCM. (e) RLCM. (f) WLDM. (g) MDTDLMS. (h) TLLCM.

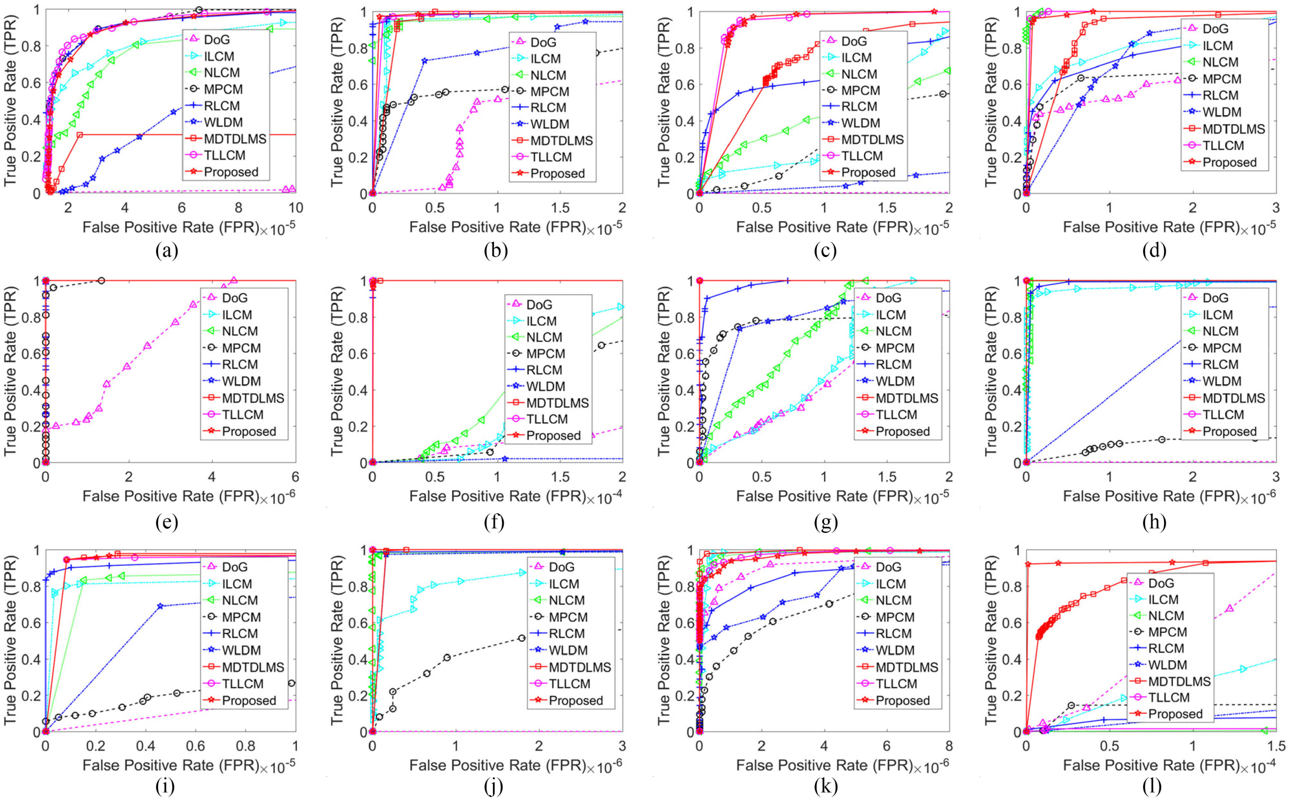


Fig. 14. ROC curves of different algorithms in different sequences. (a)–(k) Seq. 1–Seq. 11. (l) Simulated sequence.

TABLE V
COMPUTATIONAL COMPLEXITY OF DIFFERENT ALGORITHMS

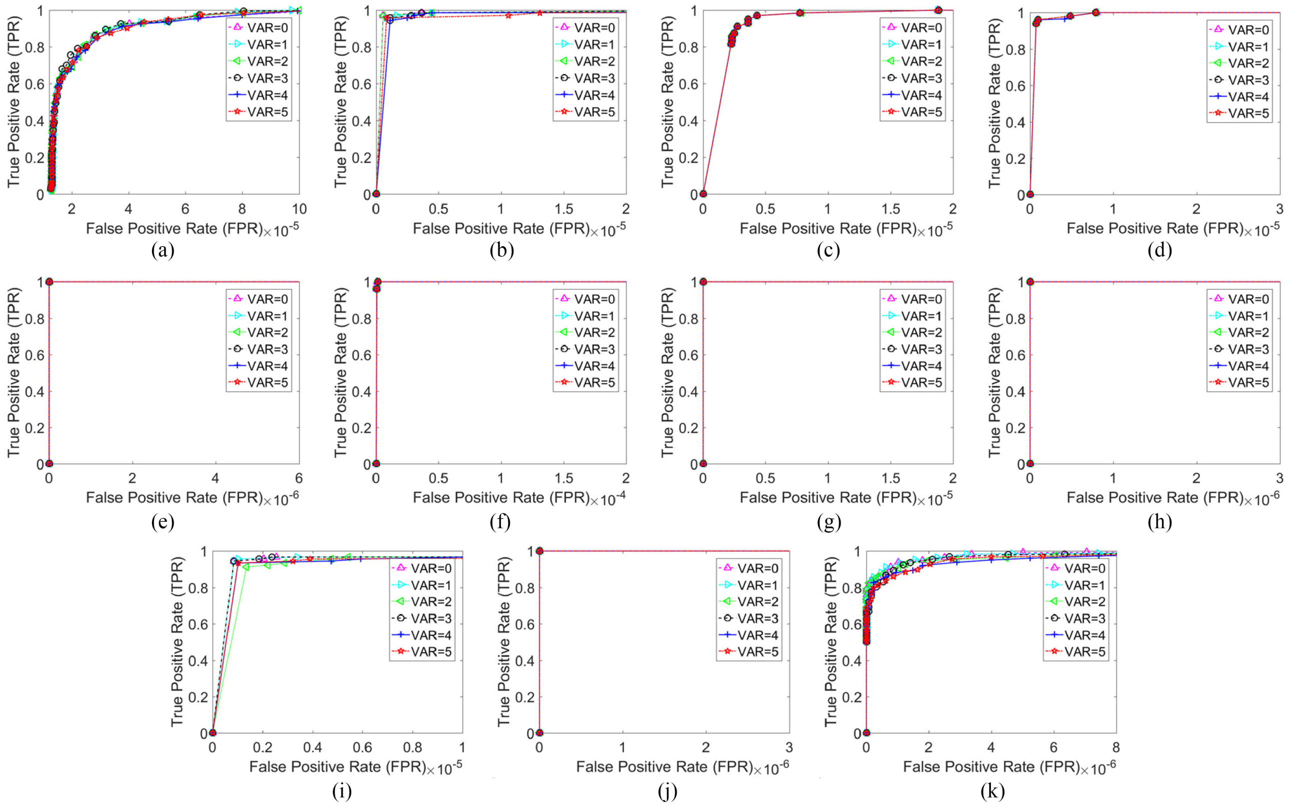
DoG	ILCM	NLCM	MPCM	RLCM	WLDM	MDTDLMS	TLLCM	Proposed
$O(L^2XY)$	$O(L^2XY)$	$O(L^2XY)$	$O(SL_S^2XY)$	$O[SL_S^2\log(L_S^2)XY]$	$O[SL_S^2\log(L_S^2)XY]$	$O(SL_S^2\log(L_S^2)XY)$	$O(L^2XY)$	$O(LXY)$

TABLE VI
AVERAGE DETECTION TIME OF DIFFERENT ALGORITHMS FOR ONE FRAME (IN SECONDS)

Seq	DoG	ILCM	NLCM	MPCM	RLCM	WLDM	MDTDLMS	TLLCM	Proposed
1	0.173	**0.043**	*0.033*	1.463	1.252	2.850	11.831	26.682	0.448
2	0.117	**0.029**	*0.023*	0.899	0.771	1.794	10.988	25.618	0.317
3	0.162	*0.032*	**0.034**	1.382	1.317	2.274	11.252	26.661	0.488
4	0.150	*0.035*	**0.045**	0.983	1.236	2.295	12.006	24.473	0.492
5	0.126	**0.046**	*0.031*	1.117	1.021	2.422	11.371	27.185	0.490
6	0.132	**0.039**	*0.036*	1.353	1.276	2.215	11.308	26.841	0.487
7	0.147	**0.069**	*0.050*	1.142	1.054	2.501	10.874	24.386	0.477
8	0.120	**0.048**	*0.047*	0.988	0.973	1.977	10.292	26.004	0.411
9	0.094	**0.053**	*0.049*	0.996	0.943	1.982	9.883	25.097	0.413
10	0.152	**0.041**	*0.035*	1.200	1.181	2.335	11.578	27.386	0.482
11	0.155	**0.059**	**0.059**	1.106	1.072	1.806	11.041	25.465	*0.380*
12	0.218	**0.077**	*0.052*	1.124	1.281	2.553	10.856	28.377	0.478

The bold means the largest value, the italics-underlined means the second largest value.

The comparison results show that some existing algorithms (such as DoG, ILCM, NLCM, etc.) can achieve a fast detection speed; however, their performances in target enhancement and background suppression are bad, so the detection rate and false alarm rate are not satisfied. By contrast, some existing algorithms (such as RLCM, MDTDLMS, TLLCM, etc.) can achieve good detection performances but at an increased cost. The proposed algorithm can achieve a good detection performance in all of the 12 sequences, and its average time consumption is only less than 0.5 s per frame. Particularly, its performance is much better



(a) (b) (c) (d) (e) (f) (g) (h) (i) (j) (k)

Fig. 15. ROC curves in different sequences using the proposed algorithm after the addition of random noise. (a)–(k) Seq. 1–Seq. 11.

than those of other algorithms in the simulated sequence when the target is adjacent to high-brightness background.

D. Noise Immunity of the Proposed Algorithm

Fig. 15 gives the performances of the proposed algorithm under different levels of random noises, a key influence factor. Eleven real IR sequences are tested, and zero-mean Gaussian white noises with variances 1, 2, 3, 4, and 5 are added.

The ROC curves in Fig. 15 show that the performance of the proposed algorithm is usually stable or only decreases slightly with noise, which confirms the robustness of the proposed algorithm against noises.

VI. Conclusion

In this article, a new background estimation method for IR small target detection is proposed. A three-layer estimation window, consisting of a central layer, an isolating layer, and a surrounding layer, deals with targets of different sizes using a single-scale calculation. A Gaussian filtering operation is adopted for the central layer to improve the image SNR, a closest-mean principle is proposed for the surrounding layer to suppress high-brightness background, and a simple weighting function utilizing the information of the isolating layer is proposed to suppress clutters further. Experiments on real and simulated images show the effectiveness of the proposed algorithm under different types of targets and backgrounds and robustness to random noise. Compared with eight existing state-of-the-art algorithms, the proposed algorithm can achieve better target enhancement and background suppression (i.e., SCRG and BSF) and improved detection and false alarm rates. In addition, the computational complexity is reduced, and its average time consumption is only less than 0.5s per frame, much less than those of some multiscale algorithms.

References

[1] X. Wang, Z. Peng, D. Kong, and Y. He, “Infrared dim and small target detection based on stable multisubspace learning in heterogeneous scene,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5481–5493, Oct. 2017.

[2] S. Qi, J. Ma, C. Tao, C. Yang, and J. Tian, “A robust directional saliency-based method for infrared small-target detection under various complex backgrounds,” *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 495–499, May 2013.

[3] J. Gao, Z. Lin, and W. An, “Infrared small target detection using a temporal variance and spatial patch contrast filter,” *IEEE Access*, vol. 7, pp. 32217–32226, Mar. 2019.

[4] L. Dong, B. Wang, M. Zhao, and W. Xu, “Robust infrared maritime target detection based on visual attention and spatiotemporal filtering,” *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 3037–3050, May 2017.

[5] X. Bai and Y. Bi, “Derivative entropy-based contrast measure for infrared small target detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2452–2466, Apr. 2018.

[6] C. Gao, T. Zhang, and Q. Li, “Small infrared target detection using sparse ring representation,” *IEEE Aerosp. Electron. Syst. Mag.*, vol. 27, no. 3, pp. 21–30, Mar. 2012.

[7] Y. Dai and Y. Wu, “Reweighted infrared patch-tensor model with both nonlocal can local priors for single-frame small target detection,” *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3752–3767, Aug. 2017.

[8] J. Han, Y. Ma, B. Zhou, F. Fan, K. Liang, and Y. Fang, “A robust infrared small target detection algorithm based on human visual system,” *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2168–2172, Dec. 2014.

[9] X. Cao, C. Rong, and X. Bai, “Infrared small target detection based on derivative dissimilarity measure,” *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 8, pp. 3101–3116, Sep. 2019.

[10] P. Du and A. Hamdulla, “Infrared moving small-target detection using spatial-temporal local difference measure,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1817–1821, Oct. 2020.

[11] W. Li, M. Zhao, X. Deng, L. Li, and W. Zhang, “Infrared small target detection using local and nonlocal spatial information,” *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3677–3689, Sep. 2019.

[12] D. J. Gregoris, S. K. Yu, and S. Tritchew, “Wavelet transform-based filtering for the enhancement of dim targets in FLIR images,” *Proc. SPIE*, vol. 2242, pp. 573–583, Mar. 1994.

[13] L. Yang, J. Yang, and K. Yang, “Adaptive detection for infrared small target under sea-sky complex background,” *Electron. Lett.*, vol. 40, no. 17, pp. 1083–1085, Aug. 2004.

[14] D. Pang, T. Shan, W. Li, P. Ma, S. Liu, and R. Tao, “Infrared dim and small target detection based on greedy bilateral factorization in image sequences,” *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 3394–3408, Jun. 2020.

[15] C. Wang and S. Qin, “Adaptive detection method of infrared small target based on target-background separation via robust principal component analysis,” *Infrared Phys. Technol.*, vol. 69, pp. 123–135, Mar. 2015.

[16] S. Kim, Y. Yang, J. Lee, and Y. Park, “Small target detection utilizing robust methods of the human visual system for IRST,” *J. Infrared, Millimeter, Terahertz Waves*, vol. 30, no. 9, pp. 994–1011, Sep. 2009.

[17] X. Shao, H. Fan, G. Lu, and J. Xu, “An improved infrared dim and small target detection algorithm based on the contrast mechanism of human visual system,” *Infrared Phys. Technol.*, vol. 55, no. 5, pp. 403–408, Sep. 2012.

[18] X. Wang, G. Lv, and L. Xu, “Infrared dim target detection based on visual attention,” *Infrared Phys. Technol.*, vol. 55, no. 6, pp. 513–521, Nov. 2012.

[19] 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, Jul. 2014.

[20] X. Dong, X. Huang, Y. Zheng, L. Shen, and S. Bai, “Infrared dim and small target detecting and tracking method inspired by human visual system,” *Infrared Phys. Technol.*, vol. 62, pp. 100–109, Jan. 2014.

[21] Y. Deng and M. Wang, “Infrared dim and small target detection based on the human visual attention mechanism,” in *Proc. 2nd Int. Conf. Mechatronics Autom. Control*, 2015, pp. 1153–1161.

[22] J. Han, Y. Ma, J. Huang, X. Mei, and J. Ma, “An infrared small target detecting algorithm based on human visual system,” *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 452–456, Mar. 2016.

[23] K. Xie, K. Fu, T. Zhou, J. Zhang, J. Yang, and Q. Wu, “Small target detection based on accumulated center-surround difference measure,” *Infrared Phys. Technol.*, vol. 67, pp. 229–236, Nov. 2014.

[24] S. Deshpande, H. Meng, and R. Venkateswarlu, “Max-mean and max-median filters for detection of small targets,” *Proc. SPIE*, vol. 3809, pp. 74–83, Oct. 1999.

[25] C. L. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, “A local contrast method for small infrared target detection,” *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 574–581, Jan. 2014.

[26] Y. Wei, X. You, and H. Li, “Multiscale patch-based contrast measure for small infrared target detection,” *Pattern Recognit.*, vol. 58, pp. 216–226, Oct. 2016.

[27] J. Han, K. Liang, B. Zhou, X. Zhu, J. Zhao, and L. Zhao, “Infrared small target detection utilizing the multiscale relative local contrast measure,” *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 612–616, Apr. 2018.

[28] Y. Qin and B. Li, “Effective infrared small target detection utilizing a novel local contrast method,” *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1890–1894, Dec. 2016.

[29] Y. Cao, R. Liu, and J. Yang, “Small target detection using two-dimensional least mean square (TDLMS) filter based on neighborhood analysis,” *J. Infrared, Millimeter, Terahertz Waves*, vol. 29, no. 2, pp. 188–200, Feb. 2008.

[30] H. Ding and H. Zhao, “Adaptive method for the detection of infrared small target,” *Opt. Eng.*, vol. 54, no. 11, Nov. 2015, Art. no. 113107.

[31] P. Lv, S. Sun, C. Lin, and G. Liu, “Space moving target detection and tracking method in complex background,” *Infrared Phys. Technol.*, vol. 91, pp. 107–118, Mar. 2018.

[32] 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.

[33] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, “Small infrared target detection based on weighted local difference measure,” *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4204–4214, Jul. 2016.

[34] 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.

[35] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, “Entropy-based window selection for detecting dim and small infrared targets,” *Pattern Recognit.*, vol. 61, pp. 66–77, Jan. 2017.

[36] Y. Chen and Y. Xin, “An efficient infrared small target detection method based on visual contrast mechanism,” *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 7, pp. 962–966, Jul. 2016.

[37] P. Du and A. Hamdulla, “Infrared small target detection using homogeneity-weighted local contrast measure,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 3, pp. 514–518, Mar. 2020.

[38] P. Lv, S. Sun, C. Lin, and G. Liu, “A method for weak target detection based on human visual contrast mechanism,” *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 2, pp. 261–265, Feb. 2019.

[39] S. Moradi, P. Moallem, and M. F. Sabahi, “Scale-space point spread function based framework to boost infrared target detection algorithms,” *Infrared Phys. Technol.*, vol. 77, pp. 27–34, Jul. 2016.

[40] 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.

[41] *IEEE OTCBVS WS Series Bench; Roland Miezianko, Terravic Research Infrared Database*. Accessed: Jul. 17, 2016. [Online]. Available: http://vcipl-okstate.org/pbvs/bench/index.html

[42] Z. Cui, J. Yang, S. Jiang, and J. Li, “An infrared small target detection algorithm based on high-speed local contrast method,” *Infrared Phys. Technol.*, vol. 76, pp. 474–481, May 2016.

[43] Q. Zeng, H. Qin, X. Yan, and H. Zhou, “Fourier spectrum guidance for stripe noise removal in thermal infrared imagery,” *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1072–1076, Jun. 2020.

[44] 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.

[45] J. Davis and M. Goadrich, “The relationship between precision-recall and ROC curves,” in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 233–240.

Jinhui Han received the Ph.D. degree in electrical circuit and system from the Huazhong University of Science and Technology, Wuhan, China, in 2016.

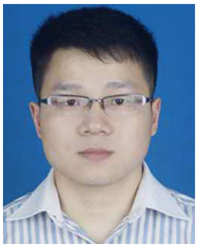
He is currently a Lecturer with the College of Physics and Telecommunication Engineering, Zhoukou Normal University, Zhoukou, China. His current research interests include target detection and image processing.

Chengyin Liu received the B.S. degree in electronic and information engineering and the Ph.D. degree in circuits and systems from the Huazhong University of Science and Technology, Wuhan, China, in 2012 and 2018, respectively.

He is currently a Lecturer with the Wuhan Electronic Information Institute, Wuhan, China. His research interests include hyperspectral image analysis, infrared image processing, and radar signal processing.

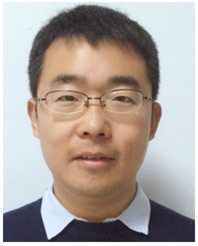
Yuchun Liu received the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2013.

He is currently an Associated Professor with the School of Mechanical and Electrical Engineering, Zhoukou Normal University, Zhoukou, China. His current research interests include bistatic and passive radar imaging.



Zhen Luo received the Ph.D. degree in signal and information processing from the National Key Laboratory of Radar Signal Processing, Xidian University, Xi'an, China, in 2019.

He is currently a Lecturer with the College of Network Engineering, Zhoukou Normal University, Zhoukou, China. His research interests include track-before-detect, multisensor information fusion, and space–time adaptive processing.



Xiaojian Zhang received the Ph.D. degree in armament science and technology from Northwestern Polytechnical University, Xi'an, China, in 2019.

He is currently a Lecturer with the College of Physics and Telecommunication Engineering, Zhoukou Normal University, Zhoukou, China. His current research interests include information processing and guidance.

Qifeng Niu received the Ph.D. degree in information and communication engineering from the Changchun University of Technology, Changchun, China, in 2020.

She is currently a Lecturer with the College of Physics and Telecommunication Engineering, Zhoukou Normal University, Zhoukou, China. Her current research interests include laser-phased array radar technology, information processing, etc.