

A True Global Contrast Method for IR Small Target Detection Under Complex Background

Jinhui Han¹, Saed Moradi¹, Bo Zhou¹, Wei Wang², Qian Zhao, Member, IEEE, and Zhen Luo¹

Abstract— Local contrast methods are efficient approaches for infrared (IR) small target detection. However, since its baseline is obtained locally, a local contrast method can work well only when the target is locally prominent, which is not always the case when the background is very complex. In this article, a new detection framework named true global contrast measure (TGCM) is proposed, in which its baseline is obtained by global decomposition, so that it can handle situations where the target is not locally salient. For example, when it is adjacent to some bright background. The complex edges/corners and noises that are easily lost during decomposition will be suppressed through some simple but ingenious design. A Gaussian filtering operation is adopted to suppress noises, and a local max dilation is utilized on the separated baseline to maintain as much edge/corner information as possible. The residual information during decomposition will be used as a weighting function to further suppress clutters. Experiments on both real and simulated IR images demonstrated the effectiveness of the proposed method. Compared to some existing algorithms, it has advantages in target enhancement and background suppression. In particular, it can achieve good performance in certain scenarios where the target is not locally prominent.

Index Terms— Global decomposition, infrared (IR) small target, local max dilation (LMD), target detection, true global contrast measure (TGCM).

I. INTRODUCTION

INFRARED (IR) detection systems have advantages of working all day, penetrating cloud and smoke, identifying camouflaged target, and so on. It has become an effective supplement or alternative to traditional visible light and radar detection systems. However, in some important applications such as guidance [1], air defense [2], monitoring/surveillance [3], [4], [5], and anti-drone [6], targets are very far from the detector and occupy only a few pixels in the

Received 1 February 2025; revised 25 April 2025; accepted 22 May 2025. Date of publication 12 June 2025; date of current version 20 June 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 61802455, Grant 62003381, and Grant 62472464; in part by the Natural Science Foundation of Henan Province under Grant 252300420399 and Grant 242300421718; and in part by the Foundation of the Science and Technology Department of Henan Province under Grant 192102210089, Grant 222102210077, Grant 232102320066, and Grant 252102210231. (Corresponding author: Jinhui Han.)

Jinhui Han and Qian Zhao are with the College of Physics and Telecommunication Engineering, Zhoukou Normal University, Zhoukou 466001, China (e-mail: hanjinhui@zknu.edu.cn).

Saed Moradi is with the Vision and Image Processing Laboratory, Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada.

Bo Zhou is with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China.

Wei Wang and Zhen Luo are with the School of Artificial Intelligence, Zhoukou Normal University, Zhoukou 466001, China.

Digital Object Identifier 10.1109/TGRS.2025.3579206

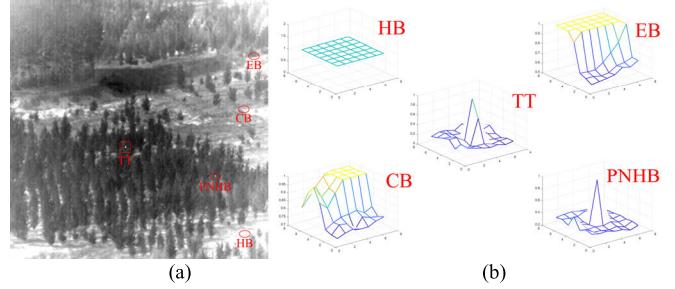


Fig. 1. Distributions of real IR small target (labeled with rectangle) and common interference (labeled with ellipses). (a) Raw IR image. (b) Distributions of different components, TT—true target, HB—high-brightness background, EB—edge of background, CB—corner of background, and PNHB—pixel-size noise with high brightness.

output image [7], commonly referred to as IR small targets. The detection of IR small targets faces great difficulties [8] (see Fig. 1). First, the size of the target is very small and it usually lacks significant shape or texture information. Second, due to the affection of distant atmosphere, the brightness of a small target is usually very weak. Third, in practical applications, various complex backgrounds such as buildings, clouds, and waves often appear in the field of view of the IR detectors, which may have high brightness and complex edges and corners, bringing many false alarms; Finally, some bad detector pixels and the random electrical noise during work may bring some pixel-size noise with high brightness (PNHB), which is very similar to real small targets and may also cause false alarms.

To address the problem of IR small target detection under complex backgrounds, many researchers have proposed a lot of methods, which can be divided into data-driven methods and model-driven methods [1]. The data-driven methods, including the two-stage methods (e.g., region convolutional neural networks (RCNN) [9], [10]), the single-stage methods (e.g., you only look once (YOLO) [11], [12], [13]), and the deep unfolding methods (e.g., robust principal component analysis net (RPCANet) [14]), typically use a trained deep convolution network to classify input data as targets or backgrounds. However, data-driven methods require a large amount of the input data and training data to have the same distribution function, but this is not always the case in real applications. In addition, deep neural networks usually have a large number of parameters, which is difficult to train and deploy on edge devices.

The model-driven methods typically utilize a handy-crafted feature model to distinguish between the true small target

and the complex background interference first and then use a threshold operation to get true targets in the feature saliency map (SM). Compared to data-driven methods, they have the advantages of being easy to understand and implement in real applications. Therefore, model-driven methods still attract a lot of attention nowadays.

In model-driven methods, the feature model can be designed between consecutive frames or within a single frame, which are referred to as sequence-based methods and frame-based methods, respectively [15]. The sequence-based methods [16], [17], [18], [19] usually have better performance since they simultaneously utilize information from multiple frames for target detection. However, due to the large amount of information that needs to be considered, such algorithms typically have larger computational costs and require more processing resources. The frame-based algorithms perform target detection within a single frame, which usually needs less computation and storage space compared to the sequence-based algorithms, making them easier to implement. Moreover, some sequence-based methods also utilize a frame-based method as the basic module. Therefore, we will focus on the frame-based methods in this article.

According to the different types of information utilized during feature extraction, existing frame-based model-driven methods can be further divided into two categories: the local type and the global type.

A. Local Type Algorithms

Local-type algorithms assume that the grayscale of a background pixel is usually close to its local neighboring pixels, while there is a significant difference between a target pixel and its neighborhood pixels. By taking the neighboring pixels of each pixel as a baseline and extracting the differences between each pixel and its baseline, the target can be easily identified. This type of algorithm mainly includes background estimation methods, morphological methods, directional derivative/gradient methods, local contrast methods, and so on.

Background estimation methods consider that the gray value of each pixel in an IR image consists of three components: the background value, the target value, and the random noise. If the background value of a pixel can be estimated using its neighboring pixels and subtracted from it, the target will be easily obtained in the residual. Current background estimation methods include not only some traditional methods such as median filtering [20] and max mean/max medium filtering [21] but also some adaptive background estimation methods such as 2-D least mean square (TDLMS) [22].

The morphological methods first design a specific shaped structural element based on the characteristics of the true small targets and then use this window to traverse the entire image. Through some morphological operations such as erosion and dilation, the target will be highlighted, while the background and noise will be suppressed. The ring-like new top hat structure window proposed by Bai and Zhou [23] is an excellent representative of morphological structural element for IR small target detection, which has been further developed by

Deng et al. [24] and Zhu et al. [25]. Zhang et al. [26] designed two dilate structures in which one is used for enhancing target and the other is used for suppressing clutters. Peng et al. [27] designed a dual structure morphological element with eight directions to distinguish background edges and so on.

Directional derivative/gradient methods treat the gray value of an IR image as a scalar field and calculate the directional derivative or gradient at each pixel. Since it contains the directional information in the directional derivative/gradient, the number of features that can be used to distinguish between the target and background will be increased. For example, Lu et al. [28] used a small kernel model to obtain directional derivatives, Bi et al. [29] introduced the concept of high-order directional derivatives, Yang et al. [30] proposed a weighted multidirectional gradient method that adopted the multidirectional difference measure as a weighting function, and Hao et al. [31] designed a gradient method, which can adaptively deal with targets of unknown size and so on.

The local contrast methods are inspired by the contrast mechanism of human eyes. Biologists found that the human eyes are more sensitive to local contrast information in the field of view rather than the brightness information, so humans can quickly and accurately capture targets that are prominent in a local area in the image [32]. By simulating the mechanism of human visual system and using local contrast instead of brightness as the basis for target extraction, better detection performance can be achieved. The core of the local contrast is the dissimilarity between a current pixel and its baseline (i.e., some neighboring pixels). According to different definitions of local contrast, existing algorithms can be further divided into three categories: difference-based local contrast methods, ratio-based local contrast methods, and joint ratio and difference-based local contrast methods.

- 1) Difference-based local contrast uses the gray value difference between several central pixels and several neighboring baseline pixels as local contrast information, which can effectively eliminate the flat background areas to zero and make the target detection easier. For example, Shao et al. [33] used the Laplacian of Gaussian (LoG) convolution template, which has positive center coefficients and negative surrounding coefficients for contrast calculation. Wang et al. [34] proposed a similar but simpler convolution template named difference of Gaussian (DoG) filter. Han et al. [35] used elliptical Gabor functions instead of circular Gaussian functions as kernel functions to distinguish background edges. Wei et al. [36] proposed a multiscale patch-based contrast measure (MPCM) to achieve multiscale detection capability for targets of unknown size. Chen et al. [37] combined LoG with MPCM and other methods together and so on.

- 2) Ratio-based local contrast takes the grayscale ratio of several central pixels to several neighboring baseline pixels as an enhancement factor to enhance the central pixels. The enhancement factor of real targets is usually greater than 1 because a target is usually slightly brighter than its surrounding neighbors. Thus, ratio-type local contrast can effectively enhance real targets. Especially,

in order to better suppress background edges, some algorithms also divided the local neighborhood into multiple directions for separating calculations. For example, Chen et al. [38] proposed a local contrast measure (LCM) algorithm based on a double-layer window with eight directions. Han et al. [39] proposed an improved LCM (ILCM) algorithm, which introduced an averaging operation to suppress PNHB; and Qin and Li [40] proposed a novel LCM (NLCM) algorithm that only averages the largest few pixels to avoid smoothing true small targets, etc.

- 3) Joint ratio and difference-based local contrast combines difference and ratio operations together to achieve the goal of enhancing true targets and eliminating complex backgrounds simultaneously. For example, Han et al. [41] proposed the definition of relative LCM (RLCM); Guan et al. [42] added multiscale Gaussian filtering as preprocessing stage; Han et al. [43], [44] designed two tri-layer windows for the ratio-difference joint local contrast calculation, which was adopted by Liu et al. [45]; and Han et al. [46] extended the local gray contrast to local feature contrast, etc.

Currently, local contrast methods have received a lot of attention due to their advantages such as theoretical simplicity, ease of implementation, and ability to enhance true targets and suppress complex backgrounds simultaneously. Some researchers also combine the local contrast with other methods to improve the detection performance further. For example, Cui et al. [47] combined with the support vector machines; Deng et al. [48], [49] and Yao et al. [50] adopted local information entropy as a weighting function; Du and Hamdulla [51] used local smoothness for weighting; Xiong et al. [52] and Zhou and Wang [53] combined local contrast with local gradient; Kou et al. [54] combined local contrast with density peak; Wang et al. [55] used the variance for weighting; and Wei et al. [56] adopted facet filtering as preprocessing and so on.

However, a local contrast method only works well when the target is locally prominent (in other words, the target must be brighter than all of its neighboring pixels), which may not always be true when the background is very complex in some practical applications. For example, a real target may appear at a position nearing some very bright backgrounds. In this case, the target will be darker than some of its neighboring pixels, so the difference between the target and its neighboring baseline pixels may be smaller than zero, while the ratio between them may be smaller than one; thus, the true target will be submerged. Some improvements have been made to deal with this situation. An important idea is to change the baseline selection during local contrast calculation. For example, Han et al. [57] used TDLMS to get the background baseline, Tang et al. [58] used a dilation operator with a ring structure to get a background baseline in a small area, and Han et al. [44] utilized the gray value closest to the center in eight surrounding directions as the baseline and so on. However, these methods can only alleviate but not solve this problem, as they still obtained baseline locally.

B. Global-Type Algorithms

The global-type algorithms assume that the global features (such as frequency bands and data spaces) of the target and background do not overlap with each other, so they can be separated from the entire image using methods such as frequency filtering, sparse representation, and sparse and low-rank decomposition.

The frequency filter methods assume that the background in the original image is usually continuously distributed, occupying the low-frequency band in the frequency domain, while the target has a significant discontinuity with its neighborhood, corresponding to the high-frequency band. The noise component is usually caused by some random factors, corresponding to the highest frequency band. Therefore, if we first transform the original image into the frequency domain, we will be able to filter out the target's frequency band. For example, Yang et al. [59] used Butterworth high-pass filters for filtering, Qi et al. [60] proposed the quaternion Fourier transform as the frequency transformation method, Gregoris et al. [61] introduced the idea of wavelet transform, and Kong et al. [62] used the diagonal detail information of Haar wavelet to detect the sea-sky line first and then proceeded with the next steps of target detection and so on. However, in complex scenes, a large number of background edges/corners also contain certain high-frequency information, which is sometimes difficult to distinguish from the target area.

The sparse representation methods use a preconstructed overcomplete dictionary to linearly represent the original data. If the correlation between the original data and some atoms in the dictionary is high, the corresponding coefficients of these atoms will be relatively large, and the coefficient vector will show obvious sparsity. Based on this, Zhao et al. [63] constructed an overcomplete dictionary in which each atom is a simulated target sample. He et al. [64] proposed a representation method with low-rank constraints based on background features. Zhang et al. [65] first used a particle filter to obtain an SM and then represented the SM. Qin et al. [66] and Liu et al. [67] constructed two dictionaries to represent the targets and backgrounds. Chen et al. [68] adopted an adaptive group sparse representation model for IR image denoising to help with small target detection and so on. However, in different scenes, the shapes of the target and the background are changing, making it extremely difficult to construct a dictionary that covers all situations, which limits the universality of sparse representation methods.

The sparse and low-rank decomposition methods consider that the background has a certain degree of nonlocal self-similarity, and thus, the rank of the background image will be relatively low. The small target size is very small, so the target image will be very sparse. If we decompose the original image into a matrix with low rank and a matrix with sparsity, the target can be easily identified in the sparse one. According to the original data used during decomposition, these methods can be divided into two categories: image patch type and tensor type.

The image patch-type sparse and low-rank decomposition methods first divide the original image into many image patches and stretch them into column vectors, then form all

these vectors as a new matrix, and decompose it. Gao et al. [69] first constructed the infrared patch image (IPI) model and used the robust principal component analysis (RPCA) algorithm to decompose it. Dai et al. [70] only focus on some small singular values during RPCA iterations to eliminate background edges. Wang et al. [71], Fang et al. [72], and Zhu et al. [73] introduced regularization factors to constrain the decomposition results. Hao et al. [74] used the proximal gradient algorithm to build a novel continuation strategy, and thus, strong edges can be suppressed. Liu et al. [75] considered the local contrast of the sparse part during decomposition and so on.

The tensor-type sparse and low-rank decomposition methods first stack some image patches into a 3-D tensor and then decompose it by iterations. Dai and Wu [76] first constructed the patch tensor model and used some local priors as a weighting function, Zhang et al. [77] simplified this model and abandoned the weighting function, and Fan et al. [78] proposed a new anisotropic background feature weight function for the IR patch tensor model and so on.

In general, global-type algorithms directly extract global features of the raw IR image, so they can achieve better detection performance even when the target is not locally prominent, which is a significant advantage compared to local-type algorithms. However, when the background is very complex, some background edges/corners and noises also have a certain degree of data sparsity, which can be mistakenly decomposed into sparse foreground, seriously interfering with the detection of true targets. Although some further operations such as data reorganization (from patch to tensor) and regularization constraints are introduced, this issue remains the main factor affecting the detection performance of such algorithms.

In recent years, some researchers have tended to combine local and global methods together to achieve better performance. For example, Chen et al. [79] combined local contrast with some frequency domain algorithms; Li et al. [80] first calculated the local contrast of the raw image as SM and then globally decomposed the SM into two parts and searched targets in the sparse one; Guan et al. [81] and Qiu et al. [82] utilized global sparsity to weight local contrast; and Lu et al. [83] adopted some gradient information as weighting function during global decomposition of the image patch tensor, while Zhang et al. [84] adopted the local entropy as weighting function and so on. However, in this research, the local algorithm and the global algorithm are completely independent modules, i.e., one is utilized as the other's weighting function, thus, and if one of them fails, the entire detection algorithm will completely fail.

C. Work in This Article

In this article, we will propose a new detection framework named true global contrast measure (TGCM). We do not call it as GCM because this name had been used in [58], but after further analysis, we found that it utilized a baseline obtained by some local but not global operations during contrast information calculation, which is totally different from this article). Instead of utilizing some neighboring pixels as a baseline for

a current pixel, we introduce the idea of global decomposition and take the separated global background estimation as the baseline for contrast information calculation so that it can deal with situations where the small target is not locally prominent, for example, when the target is darker than some of its neighboring bright background pixels. The common interference, such as complex background edges/corners and PNHB, will be suppressed by some simple but ingenious designs and weighting operations. The framework of the proposed method is shown in Fig. 2.

The main work and contributions of this article can be described as follows.

- 1) A new detection framework named TGCM is proposed, which does not use one algorithm as a parallel weighting function for another algorithm, but has a “meshing” structure. First, it uses some global methods to separate the background image from the entire raw IR image, and then, it takes the background image as baseline and calculates the contrast between the raw image and the baseline image. Since its baseline is not obtained locally, it can solve the problem of detection failure when targets are not locally prominent.
- 2) The IPI model and the sparse and low-rank decomposition are adopted as the global background estimation method, in which the low-rank part will be taken as the global background baseline image. However, it is usually very difficult to maintain edge and corner information in the low-rank image. In this article, we demonstrate the reason for errors and propose that some local operations can also be applied on the global background. Then, a simple but effective local max dilation (LMD) algorithm is proposed and applied on the low-rank background image to maintain as much edge/corner information as possible.
- 3) We propose a definition of a joint ratio and difference TGCM between the raw image and its global background baseline, in which both ratio and difference operations are used. The true target can be enhanced by the ratio operation, and the flat background area can be suppressed by the difference operation. Especially, to suppress PNHB better, the idea of matched filter is adopted and a Gaussian filtering is applied on the raw image before TGCM calculation.
- 4) To improve detection performance further, the residue image after the global decomposition (i.e., the sparse foreground matrix) is processed with a simple nonnegative constraint and then used as a weighting function on the result of TGCM so that complex clutters can be suppressed better.

Experiments on many IR images demonstrate the effectiveness and robustness of the proposed method. Compared to baseline algorithms, it has advantages in target enhancement and background suppression. In particular, it can achieve good performance in some scenes that the background is very complex or the target is not locally prominent.

The organization of this article is as follows. In Section II, some related works are introduced, such as the theory of a typical local contrast method, and the theory of a typical

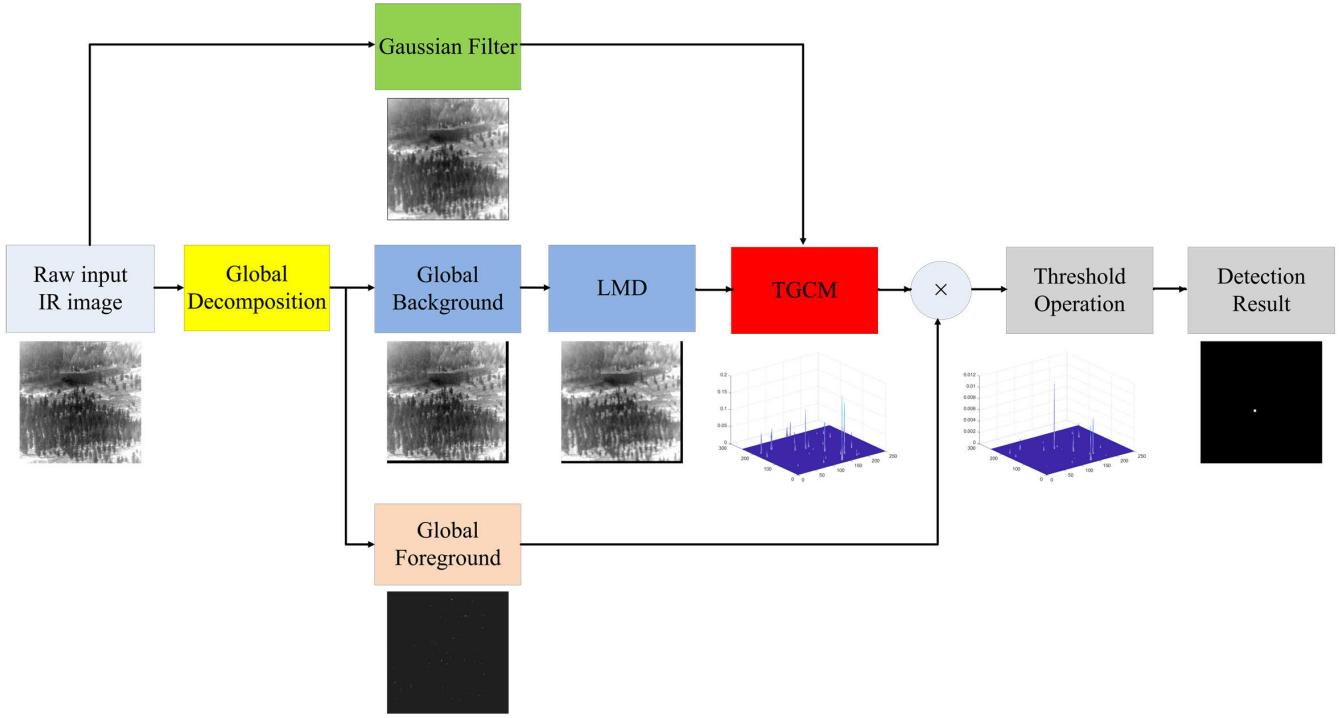


Fig. 2. Framework of the proposed method. First, a global decomposition method is used to separate the input image into the global background and the global foreground. Then, the global background, after the LMD operation, will be taken as the baseline to calculate the TGCM with the raw image after the Gaussian filtering. The global foreground part will be taken as a weighting function to the TGCM to suppress clutters further. Finally, a threshold operation will be used to obtain the target.

global decomposition method. In Section III, the calculation of the proposed TGCM algorithm 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. The experimental results are presented in Section V, and finally, Section VI concludes this article.

II. RELATED WORKS

A. Theory of a Typical Local Contrast Method

A local contrast method usually puts a nested window (see Fig. 3) on the raw IR image and slides it from left to right and top to bottom, at each position, take the gray value in one of the surrounding orientations of the window as baseline (usually the orientation with the largest gray value), and then calculate the contrast information between the central cell and the baseline. A kind of averaging operation is usually applied before contrast calculation to suppress PNHB better.

Taking the RLCM algorithm [41] as a case study, it utilizes the nested window in Fig. 3(a) and defines the contrast information as

$$\begin{aligned} \text{RLCM} &= \min\left(\frac{I_{\text{mean}_0}}{I_{\text{mean}_i}} I_{\text{mean}_0} - I_{\text{mean}_0}\right) \\ &= \min(f_i I_{\text{mean}_0} - I_{\text{mean}_0}), \quad i = 1, 2, \dots, 8 \end{aligned} \quad (1)$$

where f_i can be understood as an enhancement factor for the central cell [i.e., cell(0)] in the i th direction and I_{mean_0} or I_{mean_i} denotes the average gray of the K_1 or K_2 max pixels

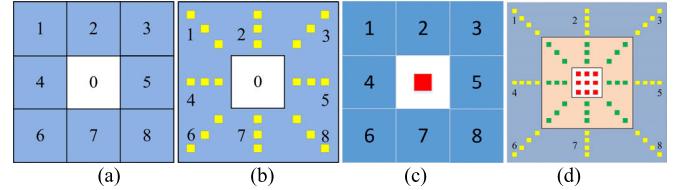


Fig. 3. Some commonly used nested windows in local contrast method. (a) and (b) Two double-layer windows. (c) and (d) Two tri-layer windows.

in cell(0) or cell(i), as shown in the following equations:

$$I_{\text{mean}_0} = \frac{1}{K_1} \sum_{j=1}^{K_1} G_0^j \quad (2)$$

$$I_{\text{mean}_i} = \frac{1}{K_2} \sum_{j=1}^{K_2} G_i^j, \quad i = 1, 2, \dots, 8 \quad (3)$$

where K_1 and K_2 are the numbers of maximal gray values considered and G_0^j or G_i^j is the j th maximal gray value of cell(0) or cell(i).

However, the effectiveness of local contrast methods for IR small target detection requires a basic assumption that the target must be the most prominent in the local area; otherwise, the detection will fail. Fig. 4 shows a detection failure example when the target is near some high-brightness background.

B. Theory of a Typical Global Decomposition Method

In a widely used imaging model, a raw IR image can be considered as the summation of three components: the

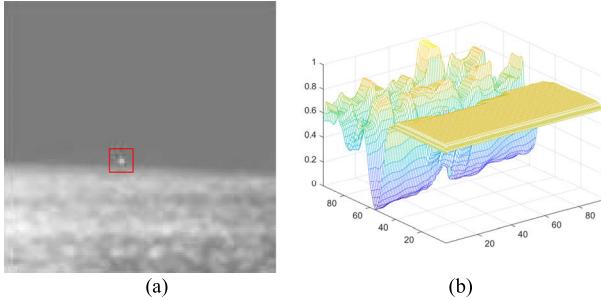


Fig. 4. Detection failure example using RLCM. (a) Raw IR image in which the target is near some high-brightness background. (b) SM after RLCM calculation.

background image, the target image, and the noise image

$$I(x, y) = I_B(x, y) + I_T(x, y) + I_N(x, y) \quad (4)$$

where (x, y) is the coordinate of each pixel in the image, I is the raw image, I_B is the background image, I_T is the target image, and I_N is the noise image.

In the field of IR small target detection, since the background has a certain degree of nonlocal self-similarity, I_B should have a low rank. The size of small target is usually very small, so I_T should be very sparse. If we use the theory of sparse and low-rank decomposition such as RPCA to divide an input IR image into a sparse matrix and a low-rank matrix, we will be able to identify the target easily in the sparse one. This problem can be described as

$$\min_{B, T} \text{rank}(B) + \lambda \|T\|_0, \quad \text{s.t. } D = B + T \quad (5)$$

where D is the data to be decomposed, i.e., the raw image I ; B is the low-rank matrix containing background, i.e., I_B ; and T is the sparse matrix containing targets, i.e., I_T .

However, solving this problem is NP-hard since the rank function and l_0 -norm are both nonconvex and discontinuous. A relaxed form is usually utilized instead, as shown in the following equation:

$$\min_{B, T} \|B\|_* + \lambda \|T\|_1, \quad \text{s.t. } D = B + T. \quad (6)$$

Here, the nuclear norm is used to approximate the rank function, and the l_1 -norm is used to approximate the l_0 -norm. They are much easier for solving.

Considering the effect of random noise, the augmented Lagrange multipliers of this problem are usually written as

$$L(B, T) = \|B\|_* + \lambda \|T\|_1 + \frac{\mu}{2} \|D - B - T\|_F^2. \quad (7)$$

Then, this problem will be transferred to an unconstrained problem and some optimization algorithms (e.g., the accelerated proximal gradient (APG) [85]) can be used to solve it. Especially, to deal with background edges better, the IPI [69] model is usually used on the raw image before separation. The flowchart of IPI algorithm is given in Fig. 5.

First, a window is put on the raw image and slid at a given step, so a series of image patches are obtained. Then, these patches will be stretched into vectors and formed as a new matrix \mathcal{D} , and decomposition will be applied on this new

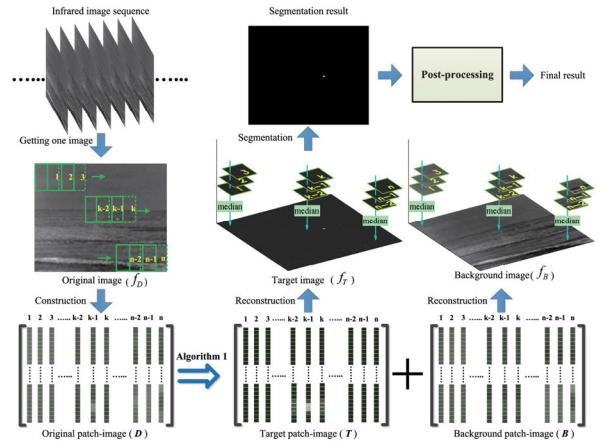


Fig. 5. Flowchart of IPI algorithm. The figure is taken from [69].

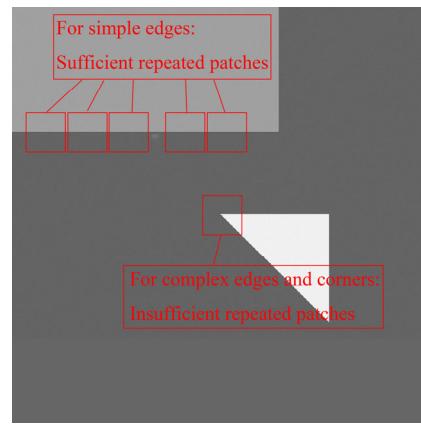


Fig. 6. Patches for simple edges and complex edges/corners using IPI model.

matrix according to the following equation to get the low-rank part \mathcal{B} and the sparse part \mathcal{T} :

$$L(\mathcal{B}, \mathcal{T}) = \|\mathcal{B}\|_* + \lambda \|\mathcal{T}\|_1 + \frac{\mu}{2} \|\mathcal{D} - \mathcal{B} - \mathcal{T}\|_F^2. \quad (8)$$

Finally, the background matrix B and target matrix T will be recovered between different patches in \mathcal{B} and \mathcal{T} , respectively. The value of each pixel is output as the median value of different overlapped patches at the same pixel position.

In the IPI model, since edges will repeat in many patches, they will have a certain low-rank characteristic in \mathcal{D} and will tend to be separated into the low-rank part \mathcal{B} , i.e., the background. Compared to separating the raw image directly, this is an advantage of IPI model. However, after further analysis, we find that the IPI model only works well for simple horizontal/vertical edges, but for complex edges and corners, it will fail since there are not sufficient repeated patches containing them (see Fig. 6). In these cases, complex edges and corners will be easily separated into sparse foreground and interfere with true target detection. Besides, noise is an interference too.

III. CALCULATION OF TGCM ALGORITHM

A. Global Decomposition

As a typical global method, the sparse and low-rank decomposition can divide the raw image into the sparse foreground

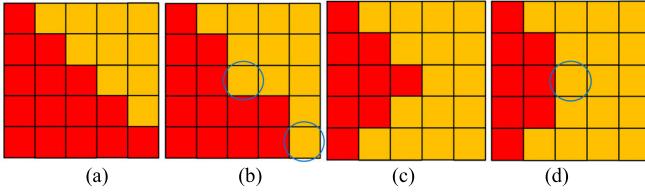


Fig. 7. Low-rank conversion at edges and corners. (a) 5×5 image with rank of 5 containing a diagonal edge. (b) Image after conversion with rank of 3. (c) 5×5 image with rank of 3 containing a corner. (d) Image after conversion with rank of 2.

and the low-rank background. Researchers usually focus on the foreground and try to directly extract the target from it. Especially, to deal with edges better, the IPI model will be used. However, some complex background edges/corners and noises also have a certain degree of data sparsity, which can easily be mistakenly decomposed into the sparse foreground, seriously interfering with the detection of real targets.

In this article, the sparse and low-rank decomposition will be adopted as the global background separation method. Instead of searching target in sparse part directly, we focus on the low-rank part, take it as the global background, and use it as a baseline in the calculation of TGCM in the next step. The main problem is how to maintain as much edge/corner information as possible in the low-rank part. We analyze the reason why we cannot obtain precise background at complex edges and corners in the low-rank part. In linear algebra, the rank of a matrix is defined as the maximum of its linearly independent columns or rows. If we want to reduce the rank of a matrix and get the low-rank part of it, we must convert some columns or rows to the linear combination of its other columns or rows. Let us consider two simple samples and ignore the processing procedures of IPI, see Fig. 7. In Fig. 7(a), the original rank of a raw image that contains a diagonal edge is 5; if we forcefully convert it to a low-rank image whose rank is 3, two rows will be changed and error data (which will be separated into the sparse part) emerges near the edge, see Fig. 7(b). In Fig. 7(c), the original rank of a raw image that contains a corner is 3; if we forcefully convert it to a low-rank image whose rank is 2, one row will be changed and error data emerges, too, see Fig. 7(d).

Since different columns or rows in a flat background area are similar to each other, it is reasonable to deduce that these data changes usually occur near complex background edges and corners only, and due to the constraint function, these data changes usually occur within a small local area. Therefore, a simple but effective LMD operation is proposed and used on the separated low-rank background image to recover as much edge/corner information as possible. The procedure of the global decomposition is shown in Algorithm 1. The IPI model [69] and APG algorithm [85] are adopted here, svd means the singular value decomposition, and \mathcal{S} means a soft-thresholding operation.

The main idea of LMD is to apply some local operations on the nonlocal background. Compared to some other nonlocal operations, such as regularization constraints and partial sum of the nuclear norm [70], [71], LMD has advantages of less calculation and is easier to understand.

Algorithm 1 Global Decomposition

Input: The input raw IR image I with resolution $M \times N$, patch size $p \times q$, the parameter λ using in APG, and the dilation radius r in LMD.
Output: The global background image B_{LMD} after local max dilation, and the residual image T .

- 1: Slide the patch window on the raw image to obtain a series of image patches, then stretch them into vectors and form them as a new matrix \mathcal{D} .
- 2: Decompose \mathcal{D} into a low rank part \mathcal{B} and a sparse part \mathcal{T} using APG algorithm:
- 3: $\mathcal{B}_0 = \mathcal{B}_{-1} = 0$; $\mathcal{T}_0 = \mathcal{T}_{-1} = 0$; $a_0 = a_{-1} = 1$; $\mu_0 > 0$; $\mu_a > 0$; $\eta < 1$.
- 4: **while** not converged **do**
- 5: $Y_k^{\mathcal{B}} = \mathcal{B}_k + \frac{a_{k-1}-1}{a_k}(\mathcal{B}_k - \mathcal{B}_{k-1})$
- 6: $Y_k^{\mathcal{T}} = \mathcal{T}_k + \frac{a_{k-1}-1}{a_k}(\mathcal{T}_k - \mathcal{T}_{k-1})$
- 7: $G_k^{\mathcal{B}} = Y_k^{\mathcal{B}} - \frac{1}{2}(Y_k^{\mathcal{B}} + Y_k^{\mathcal{T}} - \mathcal{D})$
- 8: $(U, S, V) = \text{svd}(G_k^{\mathcal{B}})$, $\mathcal{B}_{k+1} = U \mathcal{S} \frac{\mu_k}{2} [S] V^t$
- 9: $G_k^{\mathcal{T}} = Y_k^{\mathcal{T}} - \frac{1}{2}(Y_k^{\mathcal{B}} + Y_k^{\mathcal{T}} - \mathcal{D})$
- 10: $\mathcal{T}_{k+1} = \mathcal{S} \frac{\lambda \mu_k}{2} [G_k^{\mathcal{T}}]$
- 11: $a_{k+1} = \frac{1+\sqrt{4a_k^2+1}}{2}$, $\mu_{k+1} = \max(\eta\mu_k, \mu_a)$
- 12: k=k+1
- 13: **end while**
- 14: Recover low rank global background image B , and sparse residual image T from \mathcal{B} to \mathcal{T} respectively according to IPI model.
- 15: LMD operation for B :
- 16: **for** $x = 1:M$
- 17: **for** $y = 1:N$
- 18: $B_{LMD}(x, y) = \max(B(x-r:x+r, y-r:y+r))$
- 19: **end**
- 20: **end**
- 21: Output B_{LMD} and T .

The dilation radius r is a key parameter in LMD. It is obvious that the larger r , the more the edge/corner information will be recovered in the separated background. However, if the target is adjacent to a high-brightness background and r is larger than the distance between them, the target may be submerged by the dilated background. To avoid this situation as much as possible, in this article, we decide to set r to 1.

B. Matched Filtering

LMD can maintain as much edge/corner information as possible in the separated low-rank background; however, random noise (especially PNHB) is also a common interference that may easily be separated into the sparse part, and it can be hardly recovered in the separated background by the LMD operation. In this article, we turn to the raw IR image and use a filtering operation on it to suppress random noise better before the calculation of the contrast information.

The theory of matched filter tells us that the SNR of an image can be maximally improved when the filter template is similar to the signal shape [86]. Since the true IR small target usually has a shape similar to the 2-D Gaussian function,

	1	2	1
2		4	2
1	2		1

16

Fig. 8. Gaussian filtering template used in this article.

we decide to use a typical normalized Gaussian convolution template (see Fig. 8) on the raw image, and the filtering result is defined as

$$I_{GS}(x, y) = \sum_{l=-1}^1 \sum_{k=-1}^1 GS(l, k) I(x + l, y + k) \quad (9)$$

where I is the raw image, (x, y) is the coordinate of each position in the raw image, GS is the Gaussian template in Fig. 8, and I_{GS} is the filtering result.

There is another aspect to understand the matched filter. Equation (9) can be taken as a weighted gray sum in a small local area in the raw image. If there is a true target, the result will be large since a true target usually has an area larger than one pixel due to the point spread function of the optical system. Otherwise, if there is a random noise, the result will be small even if the gray of the noise is equal or slightly larger than the true target since noise is usually caused by some random reasons and usually emerges as a single pixel. Therefore, the matched filtering operation can effectively suppress random noise.

C. Joint Ratio and Difference TGCM

The TGCM will be calculated between the image after matched filtering and the global background separated according to Algorithm 1. Inspired by our former work [41], [44] in the field of local contrast, in this article, we propose a definition of joint ratio and difference TGCM, in which both ratio and difference operations are used, as shown in the following equations:

$$R(x, y) = \frac{I_{GS}(x, y)}{\max[B_{LMD}(x, y), \tau]} \quad (10)$$

$$D(x, y) = \max[I_{GS}(x, y) - B_{LMD}(x, y), 0] \quad (11)$$

$$TGCM(x, y) = R(x, y)D(x, y). \quad (12)$$

Please note that there is a small value τ (in this article, it is 5 for an 8-bit digital image) in the ratio operation to avoid division by zero. Meanwhile, there is a nonnegative constraint in the difference operation to suppress clutters.

After the TGCM calculation, the true target can be enhanced by the ratio operation, and the flat background (including high-brightness background) area can be suppressed by the difference operation.

D. Weighting Function

In Section III-A, the whole image is divided into two parts, i.e., the low-rank background part B , and the sparse

foreground part T . Although we do not directly search target in T like other researchers did, we do find that there is plenty of target information in it. Therefore, to improve detection performance further, the sparse image T will be processed with a simple nonnegative operation and then will be used as a weighting function for the result of TGCM so that complex clutters can be suppressed better. The weighting processing is given in the following equations:

$$W(x, y) = \max[T(x, y), 0] \quad (13)$$

$$TGCM(x, y) = TGCM(x, y)W(x, y). \quad (14)$$

Algorithm 2 gives the main steps for TGCM calculation.

Algorithm 2 TGCM Calculation

Input: A raw input image I with resolution $M \times N$.

Output: The TGCM matrix.

- 1: Decompose I according to Algorithm 1.
 - 2: **for** $x = 1:M$
 - 3: **for** $y = 1:N$
 - 4: Calculate $I_{GS}(x, y)$ according to (9).
 - 5: Calculate $TGCM(x, y)$ according to (10) ~ (12).
 - 6: Weight $TGCM(x, y)$ according to (13) and (14).
 - 7: **end**
 - 8: **end**
 - 9: **end**
 - 10: Normalize TGCM matrix to $0 \sim 1$ and output it.
-

IV. ANALYSIS OF DETECTION ABILITY

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

- 1) If (x, y) is a true target, since the size of a true target is usually very small, most of its information will be separated into the spare parts, and it can be easily deduced that

$$I_{GS}(x, y) > B_{LMD}(x, y) \quad (15)$$

and

$$T(x, y) > 0. \quad (16)$$

Therefore, there will be

$$R(x, y) > 1 \quad (17)$$

$$D(x, y) > 0 \quad (18)$$

and

$$TGCM(x, y) > 0. \quad (19)$$

Here, (17) means that the true target can be enhanced. Please note that when the target emerges near complex background edges and becomes not locally prominent, (15)–(19) will be still true as long as the dilation radius r in B_{LMD} is smaller than the distance between the target and the neighboring high-brightness background. It is a clear advantage of the proposed method over existing local contrast methods.

- 2) If (x, y) is pure flat background since most of its information will be separated into the low-rank matrix, there will be

$$I_{GS}(x, y) \approx B_{LMD}(x, y) \quad (20)$$

and

$$T(x, y) \approx 0. \quad (21)$$

Therefore, there will be

$$R(x, y) \approx 1 \quad (22)$$

$$D(x, y) \approx 0 \quad (23)$$

and

$$TGCM(x, y) \approx 0. \quad (24)$$

Here, (23) means that the flat background can be effectively eliminated.

Please note that (20)–(24) will be always true no matter what the background gray value is, it means that the proposed method can deal with high-brightness background properly.

- 3) If (x, y) is near a background edge or corner, although some of its information will be separated into the sparse matrix, the LMD operation on the low-rank matrix can maintain as much edge/corner information as possible, i.e.,

$$B_{LMD}(x, y) \geq I(x, y). \quad (25)$$

Considering

$$I_{GS}(x, y) \approx I(x, y). \quad (26)$$

There will be

$$B_{LMD}(x, y) \geq I_{GS}(x, y). \quad (27)$$

Therefore, we can get that

$$R(x, y) \leq 1 \quad (28)$$

$$D(x, y) \approx 0 \quad (29)$$

and

$$TGCM(x, y) \approx 0. \quad (30)$$

It means that complex background edges and corners can be effectively suppressed by the proposed algorithm.

- 4) If (x, y) is a PNHB with a gray value equal to or slightly larger than the true target, the case will be similar to the true target. However, since noise usually emerges as a single pixel, it can be suppressed to some extension by the Gaussian filtering before TGCM calculation, so it can be easily deduced that the TGCM of a PNHB pixel will be smaller than a true target. Therefore, we can conclude that the proposed method can suppress random noise.

From the discussions above, it is clear that the true target will be the most salient in the TGCM result, while other interferences are all inhibited. Therefore, in this article, we treat the TGCM result as the SM and use an adaptive threshold

operation to extract the true target. Here, the threshold Th is defined as

$$Th = \xi \cdot \max_{SM} + (1 - \xi) \text{mean}_{SM} \quad (31)$$

where \max_{SM} and mean_{SM} are the max and mean values of SM, respectively. ξ is a parameter range 0–1, in our experiments, a ξ from 0.75 to 0.95 is proper for most single-target cases. However, for multitarget cases, ξ should be set to a smaller value since different targets may have different saliency.

The last step to fulfill a successful detection would be to apply the threshold Th on SM and label the pixels larger than Th by 1; otherwise, label them by 0. In the final binary map, each connected area will be taken as a detected target (a dilation operation may be needed to eliminate clutters).

Algorithm 3 summarizes the main steps for the threshold operation.

Algorithm 3 Threshold Operation

Input: The saliency map SM with resolution $M \times N$, and the parameter ξ .

Output: The target positions.

1: Calculate the threshold Th according to (31).

```

2: for  $x = 1:M$ 
3:   for  $y = 1:N$ 
4:     if  $SM(x, y) < Th$ 
5:        $SM(x, y) = 0$ 
6:     else
7:        $SM(x, y) = 1$ 
8:     end
9:   end
10: end
```

11: Apply a dilation operation on SM to suppress clutters.

12: Output the connected areas in SM as targets.

V. EXPERIMENTAL RESULTS

In this section, 13 real IR sequences, a simulated sequence, and a single-frame dataset are used to verify the performance of the proposed algorithm. First, the details of the data used in this study are introduced. Then, we will give the processing results of the proposed algorithm step by step. To further illustrate the effectiveness of the proposed algorithm, we compare the performance and the computational complexity of the proposed algorithm with some existing algorithms. Then, we conduct some ablation experiments to verify the effectiveness of some important modules of our algorithm and select one sequence to test the robustness to noises of the proposed algorithm. Finally, a case of detection failure is given and analyzed. All the experiments are conducted on a PC with 8-GB random access memory and a 3.1-GHz Intel i5 processor, and the codes are implemented in MATLAB R2020b.

A. Data

Thirteen real IR sequences, which contain different types of targets and backgrounds, are used in our experiments. Fig. 9 illustrates some samples of them, and Table I reports the

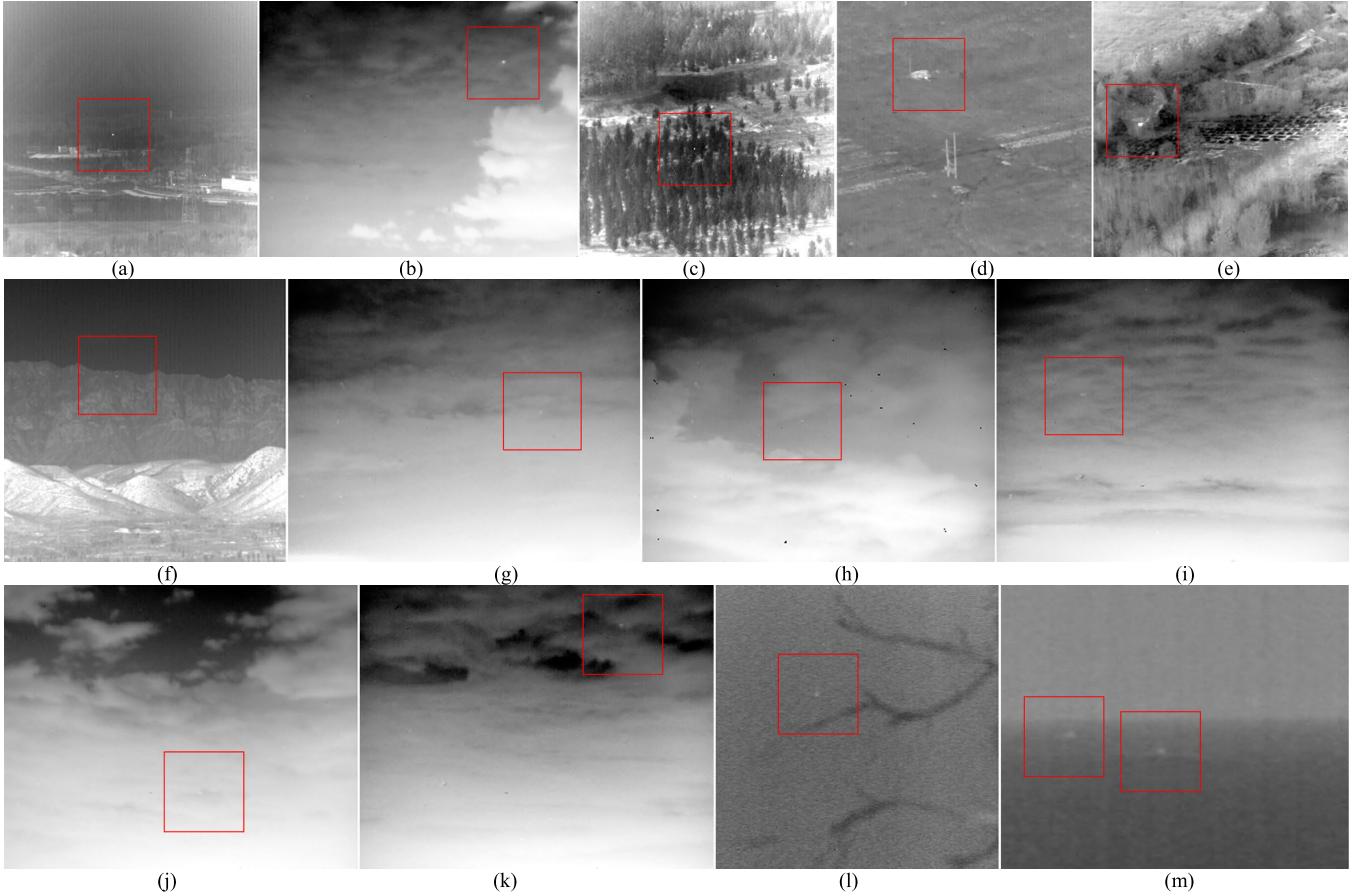


Fig. 9. Samples of the 13 real IR sequences. (a)–(m) Seq. 1–Seq. 13, respectively.

details. From Fig. 9 and Table I, it is clear that in the raw IR images, the targets are usually very small and dim, while the backgrounds are usually very complex with strong edges and corners. Also, there are heavy noises in some images, too.

To fully verify the detection performance when the target is not locally prominent, a simulated sequence that contains 400 frames is generated. Using the widely used 2-D Gaussian model, the steps of generating simulated images are listed as follows.

- 1) Generate a 500×500 matrix with a uniform value I_{NB} as the normal background. Create two high-brightness background areas in it: one is rectangle with a gray value I_{HB1} and the other is triangle with a gray value I_{HB2} . Denote this matrix by I_{BK} .
- 2) Generate a 500×500 matrix that contains a simulated small target, denoted by I_{TGT} . In this article, we use the 2-D Gaussian model to generate the small target, which can be described as

$$I_{TGT}(x, y) = I_{MAX} \exp\left(-\frac{(x - i)^2 + (y - j)^2}{2\sigma^2}\right) \quad (32)$$

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 500×500 matrix that contains random Gaussian white noises with standard deviation σ_N , and denote this matrix by I_{NOI} .
- 4) Generate a simulated image according to the following equation:

$$I = I_{BK} + I_{TGT} + I_{NOI}. \quad (33)$$

In this article, for I_{BK} , I_{NB} is randomly set to 90–100, I_{HB1} is randomly set to 150–160, and I_{HB2} is randomly set to 230–240. For I_{NOI} , σ_N is randomly set to 0–1. For I_{TGT} , I_{MAX} is randomly set to 20–60, and σ is randomly set to 2.3–3.7. To make the target not locally salient, we put the target near the rectangular high-brightness background (randomly 3–6 pixels away from the edge), with a moving speed of 1 pixel per frame from left to right.

The 152nd and the 172nd frames of the simulated sequence are shown as two samples in Fig. 10. In the 152nd frame, the target is very dim and it is weaker than its neighboring high-brightness background, while in the 172nd frame, the target has a gray value close to its neighboring high-brightness background but still weaker than the highest background in the whole image.

Besides the 14 sequences, a single-frame dataset (23 frames), which contains different types of targets and background, is used for some tests, too. Seven samples are shown in Fig. 11.

TABLE I
FEATURES OF THE 13 REAL IR SEQUENCES

	Frames	Image Resolution	Target ID	Target Size	Target Details	Background Details
Seq. 1	400	256×256	Only 1	Varies from 2×3 to 3×3	<ul style="list-style-type: none"> • Unmanned aerial vehicle target. • Located in sky. • Moving fast. • Bright enough but small. 	<ul style="list-style-type: none"> • Ground-Building background. • Heavy clutters including bright areas. • Complex edges and corners. • A few bad pixels. • Changed slowly.
Seq. 2	330	320×256	Only 1	About 4×5	<ul style="list-style-type: none"> • Plane target. • A long imaging distance. • Located in cloudy sky. • Keeping little motion. 	<ul style="list-style-type: none"> • Sky-Cloud background. • Heavy clouds including bright areas. • Some bad pixels. • Almost unchanged.
Seq. 3	110	256×256	Only 1	Varies from 2×1 to 2×3	<ul style="list-style-type: none"> • Unmanned aerial vehicle target. • Located in sky. • Moving fast. • Very small. 	<ul style="list-style-type: none"> • Ground-Tree background. • Heavy clutters including bright areas. • Complex edges and corners. • A few bad pixels. • Changed slowly.
Seq. 4	330	256×256	Only 1	Varies from 2×3 to 3×3	<ul style="list-style-type: none"> • Unmanned aerial vehicle target. • Located in sky. • Moving fast. • Near to high brightness background in some frames. 	<ul style="list-style-type: none"> • Ground background. • Heavy clutters including bright areas. • Complex edges and corners. • Changed fast.
Seq. 5	470	256×256	Only 1	Varies from 3×3 to 3×4	<ul style="list-style-type: none"> • Unmanned aerial vehicle target. • Located in sky. • Moving fast. 	<ul style="list-style-type: none"> • Ground-Tree background. • Heavy clutters including bright areas. • Complex edges and corners. • Changed fast.
Seq. 6	210	256×256	Only 1	Varies from 2×1 to 3×2	<ul style="list-style-type: none"> • Unmanned aerial vehicle target. • Located in sky. • Moving fast. • Very small and dim. 	<ul style="list-style-type: none"> • Ground-Hill background. • Heavy clutters including bright areas. • Complex textures. • Changed slowly.
Seq. 7	200	320×256	Only 1	About 3×3	<ul style="list-style-type: none"> • Plane target. • A long imaging distance. • Located in cloudy sky. • Keeping little motion. • Very small and dim. 	<ul style="list-style-type: none"> • Sky-Cloud background. • Heavy clouds including bright areas. • Complex textures. • Some bad pixels. • Almost unchanged.
Seq. 8	260	320×256	Only 1	About 3×3	<ul style="list-style-type: none"> • Plane target. • A long imaging distance. • Located in cloudy sky. • Keeping little motion. • Very small and dim. 	<ul style="list-style-type: none"> • Sky-Cloud background. • Heavy clouds including bright areas. • Complex textures. • Many bad pixels. • Change slowly.
Seq. 9	300	320×256	Only 1	About 3×5	<ul style="list-style-type: none"> • Plane target. • A long imaging distance. • Located in cloudy sky. • Keeping little motion. • Very small and dim. 	<ul style="list-style-type: none"> • Sky-Cloud background. • Heavy clouds including bright areas. • Complex textures. • Some bad pixels. • Almost unchanged.
Seq. 10	280	320×256	Only 1	About 3×4	<ul style="list-style-type: none"> • Plane target. • A long imaging distance. • Located in cloudy sky. • Keeping little motion. • Very small and dim. 	<ul style="list-style-type: none"> • Sky-Cloud background. • Heavy clouds including bright areas. • Many broken clouds. • Some bad pixels. • Almost unchanged.
Seq. 11	330	320×256	Only 1	About 3×5	<ul style="list-style-type: none"> • Plane target. • A long imaging distance. • Located in cloudy sky. • Keeping little motion. • Very small and dim. 	<ul style="list-style-type: none"> • Sky-Cloud background. • Heavy clouds including bright areas. • Complex textures. • Many bad pixels. • Almost unchanged.
Seq. 12	90	256×256	Only 1	About 3×5	<ul style="list-style-type: none"> • Truck target. • A long imaging distance. • Located in homogeneous ground. • Keeping little motion. • Very small and dim. 	<ul style="list-style-type: none"> • Ground-Tree background. • Heavy noises. • Complex textures. • Changed fast.
Seq. 13	300	280×228	Target 1	About 5×6	<ul style="list-style-type: none"> • Ship target. • A long imaging distance. • Located in homogeneous sea. 	<ul style="list-style-type: none"> • Sea-Sky background. • Heavy wave clutters. • Heavy noises.
			Target 2	About 5×5	<ul style="list-style-type: none"> • Two targets, one is moving and the other is stationary. 	<ul style="list-style-type: none"> • Almost unchanged.

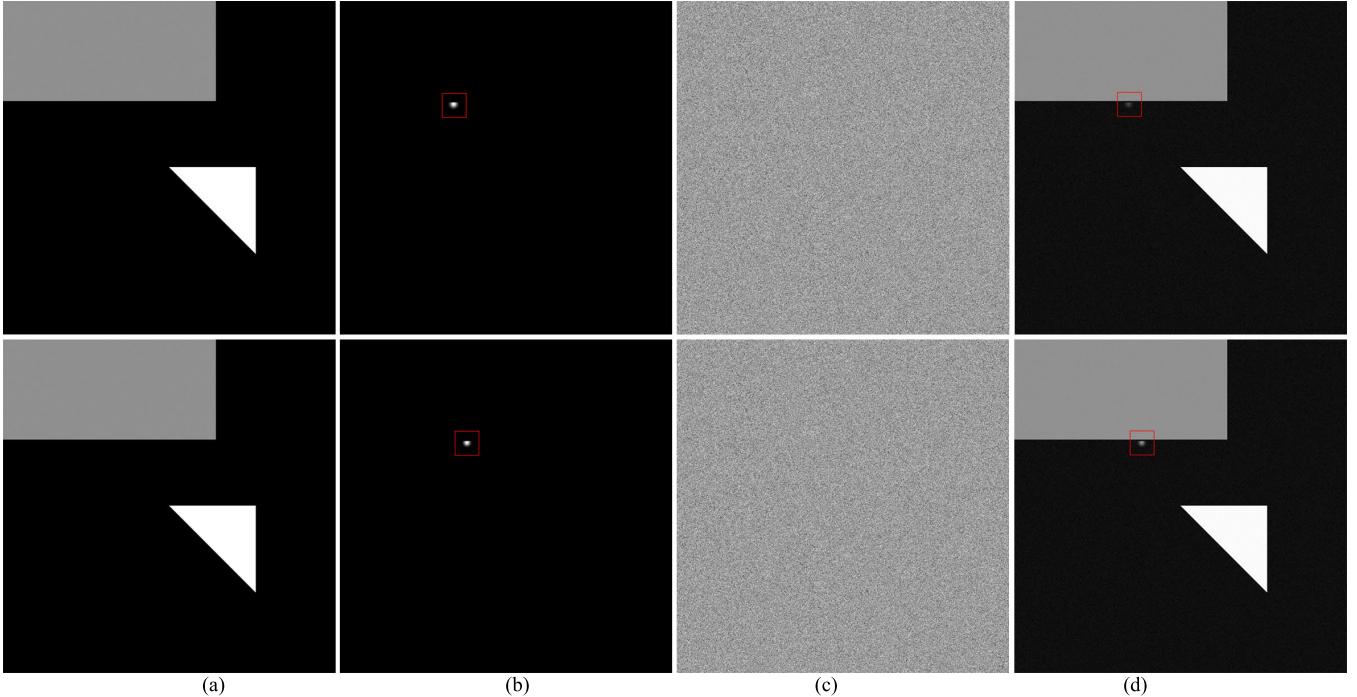


Fig. 10. Samples of two simulated images in the simulated sequence, the upper line is the 152nd frame and the bottom line is the 172nd frame. Here, I_{NB} is set to 100, I_{HB1} is set to 160, I_{HB2} is set to 240, I_{MAX} is set to 20 for the 152nd frame and 60 for the 172nd frame, σ is set to 3 for the 152nd frame and 2.7 for the 172nd frame, and σ_N is set to 1. (a) Background part I_{BK} . (b) Target part I_{TGT} . (c) Noise part I_{NOI} . (d) Simulated image I .

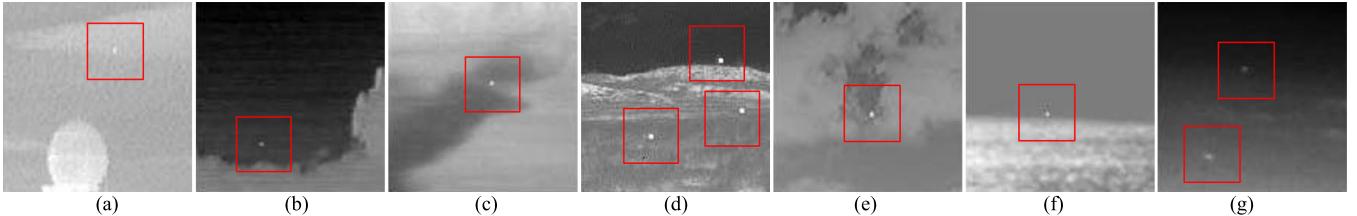


Fig. 11. Seven samples of the single-frame dataset. (a)–(g) Sample 1–Sample 7.

B. Processing Results of the Proposed Algorithm

First, we tested the detection ability of the proposed algorithm. Fig. 12 shows the processing results of the proposed algorithm step by step for the 14 sequences. The same sample images with Figs. 9 and 10 are given (note that two samples are listed for the simulated sequence).

It can be seen from Fig. 12 that the following conditions hold.

In raw IR image I , the targets are usually small and dim, while backgrounds are usually bright, with complex edges and corners. Meanwhile, heavy noises (including PNHB) exist in some sequences, too [see Fig. 12(a)].

The Gaussian filter can improve image quality to a certain extent first, especially for noise suppression [see Fig. 12(b)].

The global foregrounds T separated from the raw image contain certain amount of target information. However, some edges/corners and noise information also exist in some sequences, so it may not be wise to directly search targets in it [see Fig. 12(c)].

The global backgrounds B separated from raw image mainly contain background information, which are suitable to be considered as a baseline for contrast calculation. However,

some important information of complex edges/corners is lost since they are easily separated into T [see Fig. 12(d)].

After LMD, the global backgrounds B_{LMD} can maintain as much edges/corners information as possible, so they are more suitable to be considered as a baseline than B [see Fig. 12(e)].

In the TGCM result, true targets become salient and easy to identify. However, in some sequences that contain complex backgrounds, there are still some clutter residues [see Fig. 12(f)].

After the weighting operations by T [as shown in Fig. 12(g)], clutter residues are suppressed further, and true targets have become the most salient now [see Fig. 12(h)].

Finally, after the threshold operation, all the true targets are extracted successfully without any false targets, which proves the effectiveness of the proposed algorithm [see Fig. 12(i)].

In order to verify the effectiveness of the proposed algorithm under more different types of targets and backgrounds, Fig. 13 gives the processing results of the proposed algorithm for the single-frame dataset, and the same sample images with Fig. 11 are given here. It is clear that all the true targets are output without any false targets, too.

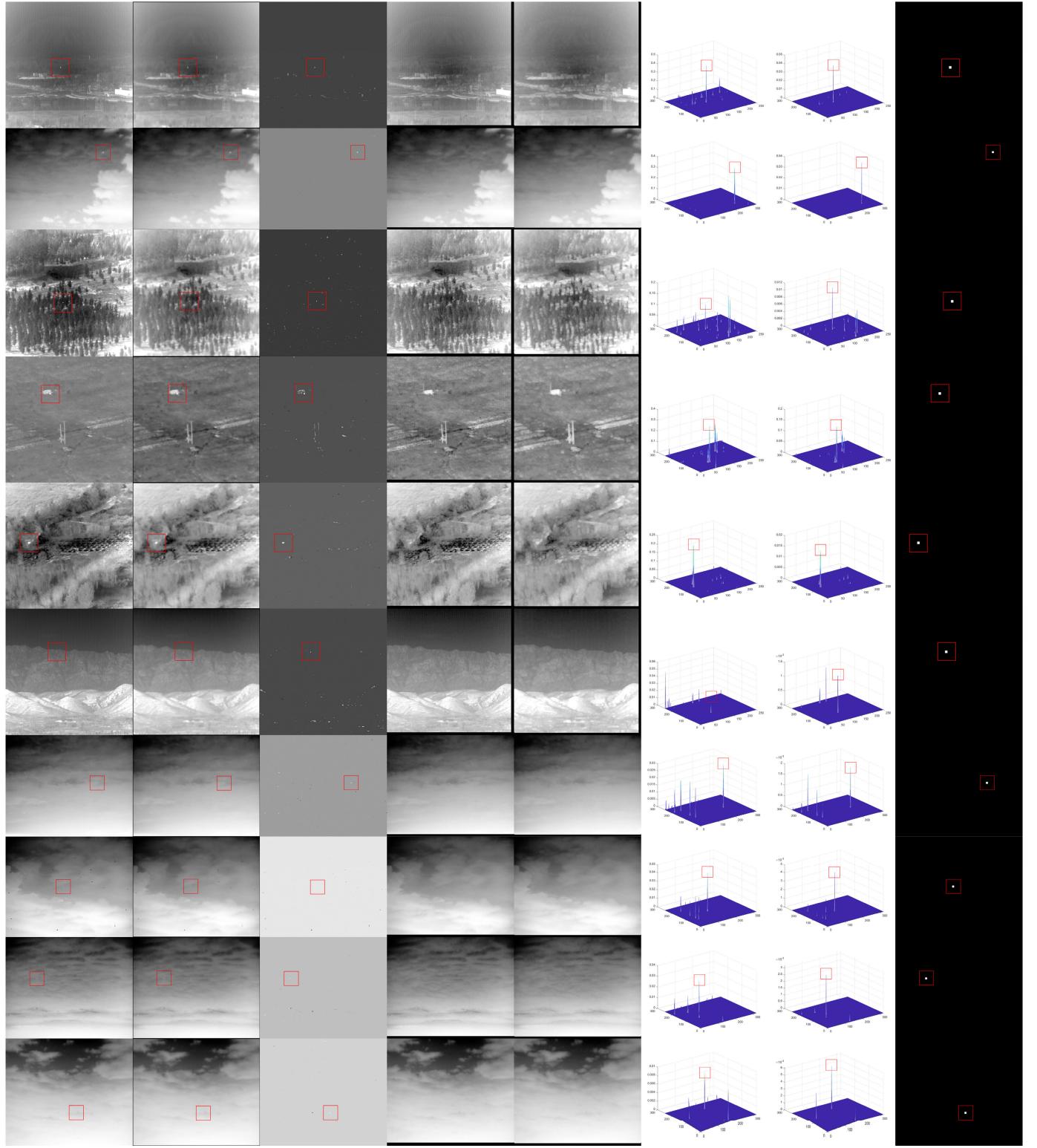


Fig. 12. Processing results of the proposed algorithm for the samples of the 14 sequences (from top to bottom: Seq. 1–Seq. 13, and two samples of the simulated sequence). (a) Raw IR images. (b) Images after Gaussian filtering. (c) Global foreground T . (d) Global background B . (e) Global background B_{LMD} after LMD operation. (f) GCM results between (b) and (e). (g) GCM results after weighting by (c). (h) Detection results after threshold operation.

C. Comparisons With Other Algorithms

To verify the advantages of the proposed algorithm, we chose some existing algorithms for comparison. Ten local contrast algorithms are chosen, including DoG [34], ILCM [39], NLCM [40], MPCM [36], RLCM [41], weighted

local difference measure (WLDM) [48], multidirectional TDLMS (MDTDLMS) [57], tri-layer LCM (TLLCM) [43], enhanced closest mean background estimation (ECMBE) [44], and GCM [58]. Besides, two global decomposition algorithms are chosen too, including IPI [69] and low-rank and sparse

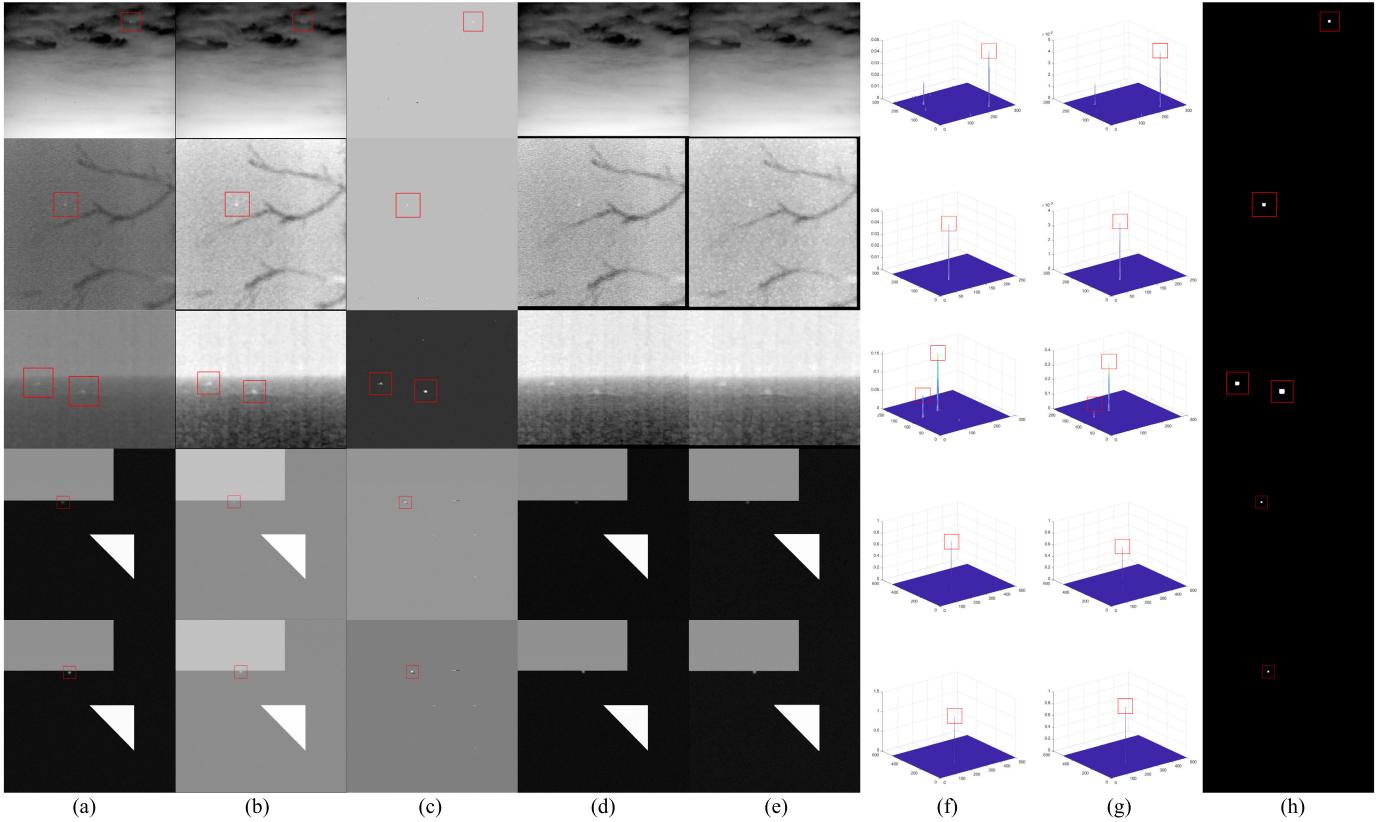


Fig. 12. (Continued.) Processing results of the proposed algorithm for the samples of the 14 sequences (from top to bottom: Seq. 1–Seq. 13, and two samples of the simulated sequence). (a) Raw IR images. (b) Images after Gaussian filtering. (c) Global foreground T . (d) Global background B . (e) Global background B_{LMD} after LMD operation. (f) GCM results between (b) and (e). (g) GCM results after weighting by (c). (h) Detection results after threshold operation.

decomposition with high local variance (HiLV) [75]. Here is a short description of each baseline method.

- 1) DoG is a traditional difference type local contrast method.
- 2) ILCM is a ratio-type local contrast method, and orientation information is utilized to suppress background edges in it.
- 3) NLCM is similar to ILCM, but it adopted only a few pixels for calculation to avoid smoothing the target.
- 4) MPCM is a multiscale method.
- 5) RLCM is a multiscale joint ratio and difference local contrast method.
- 6) WLDM is a multiscale weighted method that introduces the local entropy as a weighting function.
- 7) MDTDLMS combines the local contrast method with background estimation. However, its background is got in local.
- 8) TLLCM combines the local contrast method with background estimation, too, and its background is obtained using a multiscale tri-layer window. However, its background is still local.
- 9) ECMBE proposes a closest mean principle for background estimation, so it can effectively alleviate the problem of target submergence caused by the high-brightness background.
- 10) GCM, although declares itself as a global contrast method, still obtains the baseline locally for contrast calculation. However, its calculation area is very small,

so it can alleviate the problem of target submergence caused by the high-brightness background too.

- 11) IPI is a global decomposition algorithm, but it focuses on the sparse foreground part and directly search target in it.
- 12) HiLV is another global decomposition algorithm, and it focuses on the sparse foreground part and directly searches target in it too.

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

1) Comparisons of Detection Performance: First, Tables III and IV report the SCR gain (SCRG) and the background suppression factor (BSF) for the sample images of the 14 sequences using different algorithms. SCRG is an objective indicator that can describe the target enhancement ability of an algorithm, and BSF is another objective indicator that can describe the background suppression ability of an algorithm. They are defined as follows [87]:

$$\text{SCRG} = \frac{\text{SCR}_{\text{out}}}{\text{SCR}_{\text{in}}} \quad (34)$$

$$\text{BSF} = \frac{\sigma_{\text{in}}}{\sigma_{\text{out}}} \quad (35)$$

where SCR_{in} and SCR_{out} are the SCR [defined as (36)] of the raw image and SM, respectively, and σ_{in} and σ_{out} are the standard deviation of the raw image and SM, respectively,

$$\text{SCR} = \frac{|I_t - I_{\text{nb}}|}{\sigma} \quad (36)$$

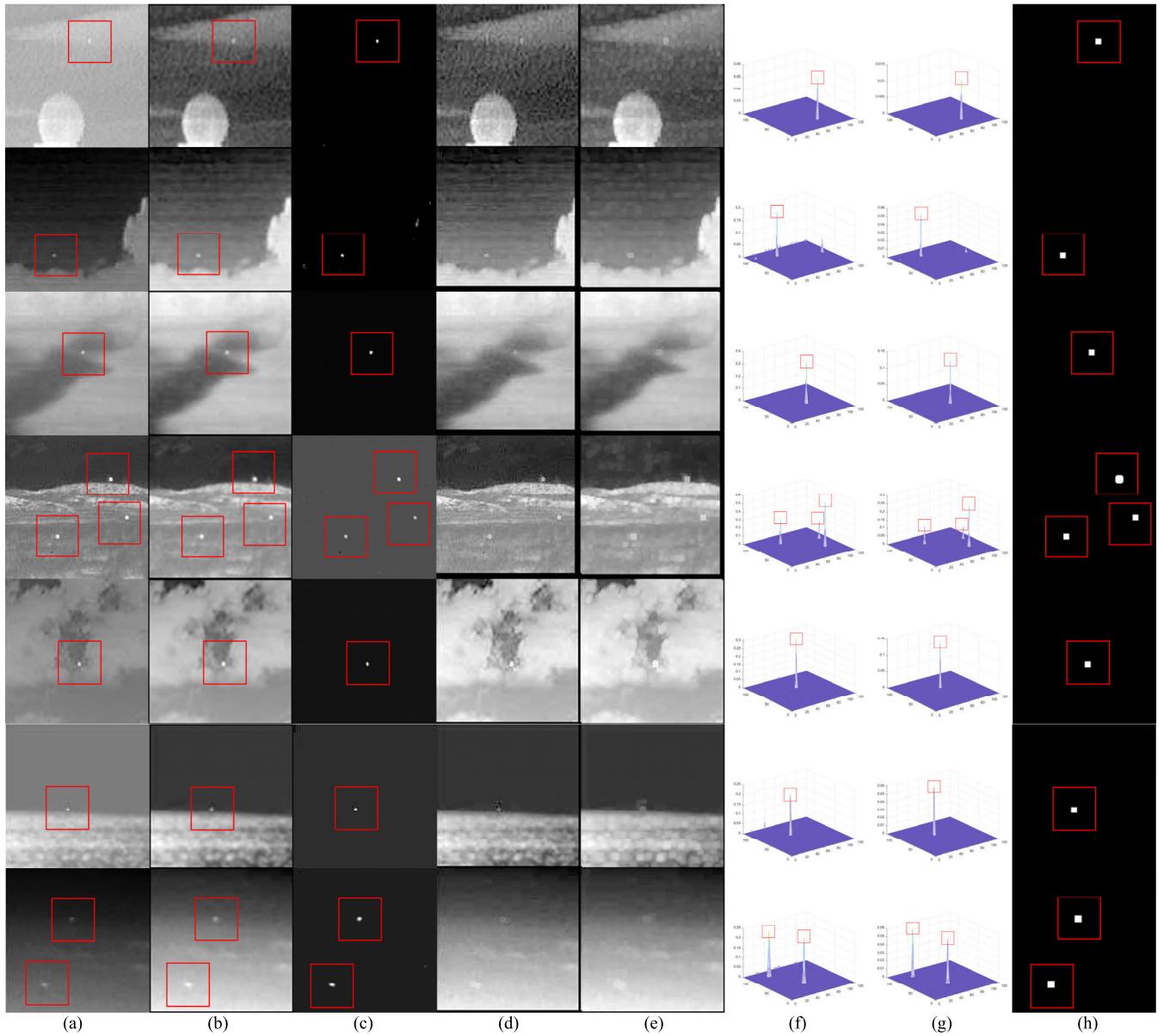


Fig. 13. Processing results of the proposed algorithm for the samples of the single-frame dataset. (a) Raw IR images. (b) Images after Gaussian filtering. (c) Global foreground T . (d) Global background B . (e) Global background B_{LMD} after LMD operation. (f) GCM results between (b) and (e). (g) GCM results weighted by (c). (h) Detection results after threshold operation.

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 stand deviation of the image.

As reported in Tables III and IV, the proposed algorithm can achieve the highest (bold) or the second highest (italic) SCRG and BSF in most cases. Especially, when the target is adjacent to high-brightness background and becomes not locally salient, the proposed algorithm can still achieve high SCRG and BSF.

Then, Figs. 14 and 15 give the salience map and the detection results for the sample images of the 14 sequences using different algorithms, and Fig. 16 gives the receiver operating characteristic (ROC) curves of different algorithms for each whole sequence; here, the false positive rate (FPR)

and the true positive rate (TPR) are defined as follows:

$$FPR = \frac{\text{number of detected false targets}}{\text{total number of pixels in the whole image}} \quad (37)$$

$$TPR = \frac{\text{number of detected true targets}}{\text{total number of real targets}}. \quad (38)$$

It can be seen from Figs. 14 to 16 that the following conditions hold.

- 1) DoG can only extract true targets in Seq. 2 and Seq. 13 in Fig. 15, and its detection performance is usually the worst in Fig. 16. This is because DoG is a traditional algorithm using only difference operation.
- 2) The performance of ILCM is slightly better than DoG; however, its performance is not satisfied in most of the sequences since it is a local method, for example, in Seq. 3, Seq. 6, Seq. 7, Seq. 8, Seq. 9, and Seq. 10.

TABLE II
PARAMETER VALUES USED IN EACH ALGORITHM

		Parameter values
DoG		A 5×5 binomial kernel as the authors 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 cell size 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, and the cell size is 3×3 , 5×5 , 7×7 and 9×9 , respectively.
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, respectively.
ECMBE		The central layer is 3×3 , the isolating layer is 7×7 , and the surrounding layer is 7×7 .
GCM		The same ring structure as the authors recommended.
IPI		The patch size is 50×50 , the step is 10, λ is $1/\sqrt{\max(m,n)}$.
HiLV		The patch size is 45×45 , the step is 20, λ is $1/\sqrt{\max(m,n)}$, C is 8, α is 1.3, β is 2.5.
Proposed		The patch size is 50×50 , the step is 10, λ is $1/\sqrt{\max(m,n)}$, r is 1, τ is 5 for an 8-bit digital image.

TABLE III
SCRG OF DIFFERENT ALGORITHMS

Seq	Target	DoG	ILCM	NLCM	MPCM	RLCM	WLDM	MDTD LMS	TLLCM	ECMBE	GCM	IPI	HiLV	Proposed	
1	1	0.43	1.80	2.65	0.35	2.99	10.94	6.15	17.85	11.58	36.52	18.72	29.22	39.98	
2	1	6.32	28.76	24.05	4.89	12.69	90.85	45.97	64.66	70.68	83.43	76.03	102.80	121.47	
3	1	0.43	0.29	0.45	1.87	1.32	8.43	1.87	8.28	6.94	29.39	20.09	20.56	42.73	
4	1	0.64	0.64	1.36	1.49	0.49	3.82	1.56	2.28	2.21	4.73	3.84	5.00	9.16	
5	1	0.96	3.10	5.79	5.31	4.39	4.67	11.28	24.85	17.27	12.38	14.61	21.05	36.62	
6	1	1.55	2.74	1.88	0.86	4.65	3.74	16.63	36.91	18.84	47.06	51.11	65.33	237.15	
7	1	10.65	10.01	39.49	2.30	9.95	98.75	74.88	203.64	121.77	159.31	147.74	272.48	341.33	
8	1	4.72	1.78	0.10	0.63	9.40	6.32	49.38	215.72	142.65	123.35	25.87	0.86	436.22	
9	1	5.68	22.87	4.66	1.08	8.08	46.39	32.41	223.46	127.33	151.63	87.51	302.10	455.64	
10	1	23.46	57.59	23.46	7.12	8.09	45.54	34.23	87.39	98.21	323.64	240.98	361.38	1.55E3	
11	1	8.94	17.29	16.47	3.69	16.56	40.90	45.08	177.40	128.83	182.58	82.05	229.42	330.43	
12	1	1.53	2.58	4.45	2.94	2.58	4.27	8.71	30.30	8.50	10.31	14.24	28.73	60.10	
13	1	10.80	14.61	9.04	11.13	6.13	34.07	35.09	44.49	48.00	47.69	60.44	59.49	33.62	
	2	6.63	11.96	15.99	9.20	7.44	23.24	31.02	60.87	47.94	51.78	66.45	83.73	98.54	
Sim	152 nd	1	3.40	1.02	0.42	0.08	1.01	2.89	61.43	0.08	116.01	175.50	51.53	8.32	425.85
Sim	172 nd	1	2.79	2.59	0.01	0.06	1.48	3.06	113.92	0.00	6.37	230.39	83.84	43.78	464.42

TABLE IV
BSF OF DIFFERENT ALGORITHMS

Seq	DoG	ILCM	NLCM	MPCM	RLCM	WLDM	MDTD LMS	TLLCM	ECMBE	GCM	IPI	HiLV	Proposed	
1	0.42	8.08	0.14	0.32	2.10	50.90	1.46E3	3.22E3	2.09E3	12.22	3.94E3	6.71E3	1.68E5	
2	3.62	27.82	0.00	0.01	7.18	11.04	3.79E5	5.31E5	5.81E5	50.80	1.04E6	1.99E6	2.62E7	
3	0.43	7.40	0.12	0.13	2.43	29.75	1.31E3	1.92E3	1.62E3	10.24	4.61E3	6.01E3	7.41E5	
4	0.18	3.49	0.08	0.22	0.85	24.65	345.70	524.66	483.09	8.72	625.09	1.33E3	6.93E3	
5	0.35	8.47	0.14	0.16	2.82	32.04	1.86E3	4.06E3	2.83E3	19.44	3.67E3	1.31E4	3.47E5	
6	0.83	20.91	0.68	0.46	4.83	116.06	2.15E3	3.87E3	2.37E3	48.12	3.63E3	4.61E4	6.68E6	
7	7.05	95.26	0.10	0.03	15.68	41.27	1.37E5	3.68E5	2.31E5	157.91	5.65E5	6.29E6	3.35E8	
8	3.43	3.98	0.00	0.02	11.14	1.90	1.49E5	5.47E5	3.62E5	100.16	4.62E5	1.50E4	2.44E8	
9	3.30	54.69	0.03	0.02	11.89	15.93	8.39E4	5.19E5	2.97E5	173.74	6.21E5	7.79E6	3.31E8	
10	4.23	89.61	0.12	0.01	12.18	22.30	1.29E5	2.78E5	2.76E5	342.46	7.49E5	1.23E8	1.45E9	
11	5.73	55.32	0.04	0.02	15.71	24.60	1.35E5	5.25E5	3.82E5	195.05	7.60E5	9.44E6	2.06E8	
12	0.73	10.47	0.83	0.57	2.84	81.89	316.08	1.08E3	306.99	14.10	2.24E3	6.84E3	5.49E5	
13	3.73	63.27	10.53	1.46	6.26	1.79E3	924.49	1.80E3	1.42E3	157.98	2.18E3	1.63E4	8.91E3	
Sim	152 nd	0.49	13.16	0.28	0.14	3.43	79.29	6.34E3	9.09E3	4.34E3	979.22	4.37E3	5.56E3	1.98E4
Sim	172 nd	0.49	13.17	0.28	0.14	3.42	79.45	4.82E3	6.74E3	495.22	930.00	4.66E3	5.26E3	1.71E4

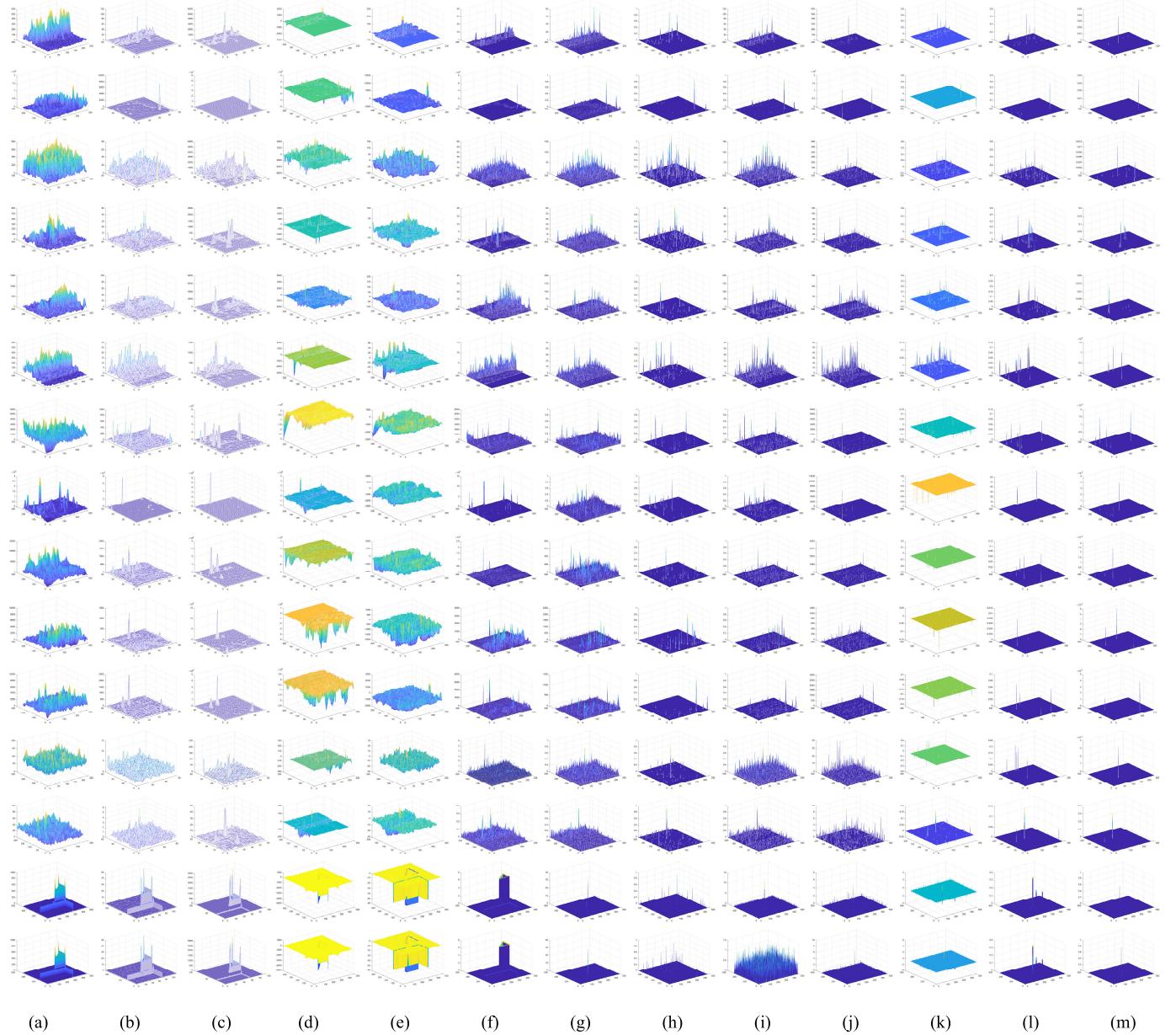


Fig. 14. SMs of the 14 sequences using different algorithms. (a) DoG. (b) ILCM. (c) NLCM. (d) MPCM. (e) RLCM. (f) WLDM. (g) MDTDLMS. (h) TLLCM. (i) ECMBE. (j) GCM. (k) IPI. (l) HiLV. (m) Proposed.

- 3) The performance of NLCM is similar to ILCM, but slightly better, for example, in Seq. 4 and Seq. 7. However, its performance is still not satisfied in many sequences, for example, in Seq. 3, Seq. 6, Seq. 8, Seq. 9, and Seq. 10.
- 4) MPCM can output true targets in some sequences. However, if the background is very complex, different scales of interference will be enhanced and output too, and its performance will decrease obviously, for example, in Seq. 1, Seq. 4, Seq. 6, and Seq. 10.
- 5) RLCM utilizes both ratio and difference operations, so it can achieve a better performance in some cases, for example, in Seq. 8, Seq. 9, and Seq. 11. However, its performance in some sequences is not satisfied, especially when the target is not locally salient, for example, in Seq. 4 and the simulated sequence.
- 6) WLDM utilizes the local entropy as the weighting function; however, when the background is very complex, and some interference will be enhanced and output, for example, in Seq. 5–Seq. 12, and the simulated sequence.
- 7) MDTDLMS can achieve a good detection performance in some cases, such as in Seq. 11 and Seq. 12. However, its background is obtained locally, so its performance is still not good in some sequences, especially when the target is not locally salient, for example, in Seq. 4 and the simulated sequence.
- 8) TLLCM can achieve good performance in some sequences, but it fails when the target is near to high-brightness background, such as in Seq. 4 and the simulated sequence. Besides, its performance is not good when the target is very small dim and the background

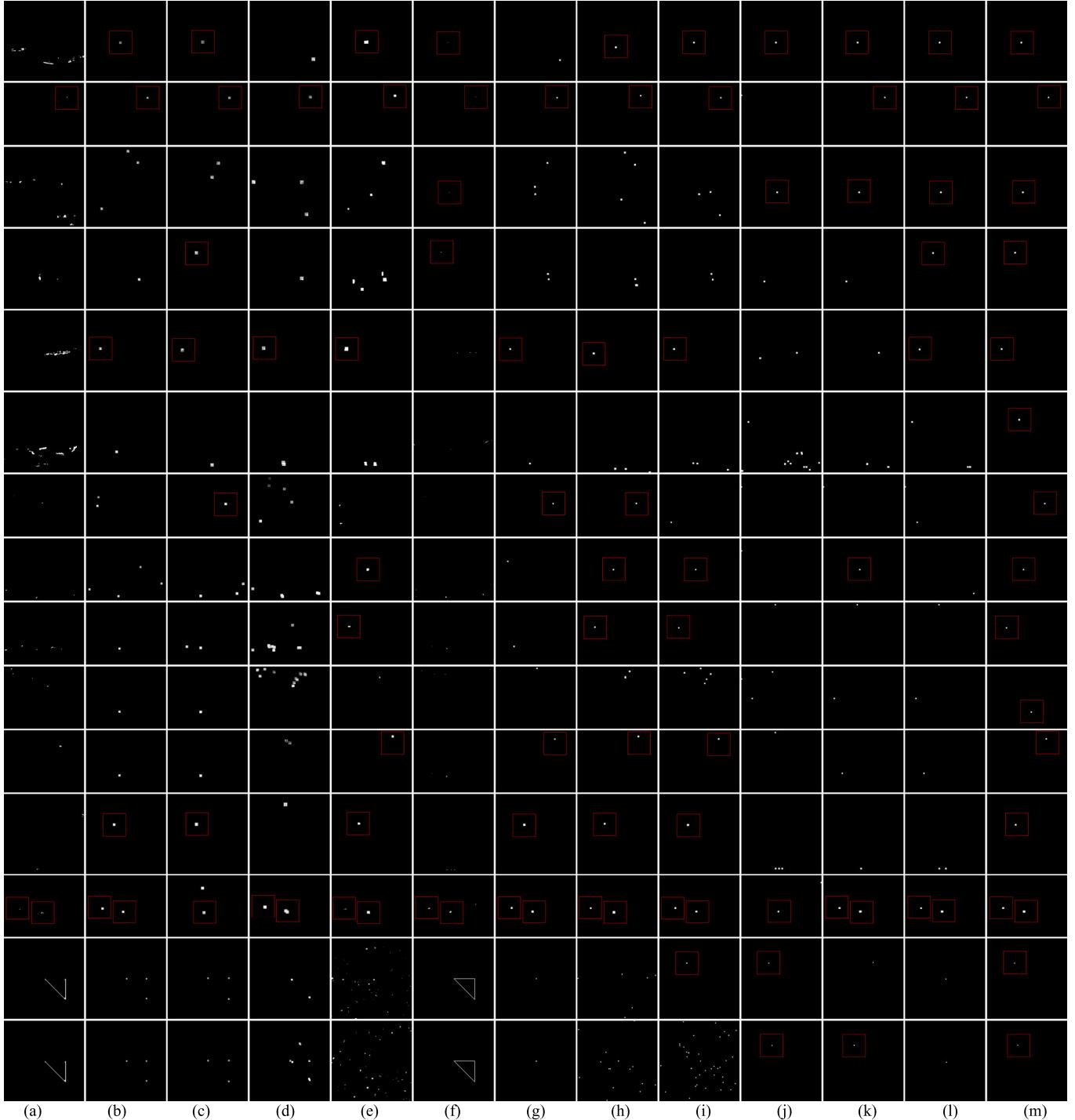


Fig. 15. Detection results of the 14 sequences using different algorithms. (a) DoG. (b) ILCM. (c) NLCM. (d) MPCM. (e) RLCM. (f) WLDM. (g) MDTDLMS. (h) TLLCM. (i) ECMBE. (j) GCM. (k) IPI. (l) HiLV. (m) Proposed.

is very complex, for example, in Seq. 3, Seq. 6, and Seq. 10.

- 9) ECMBE can achieve good detection performance in some cases, even when the target is not locally salient, for example, in the 152nd frame of the simulated sequence. However, ECMBE is still a local method, and it can only alleviate the problem but not solve it. For example, in Seq. 4 and the 172nd frame of the simulated sequence, when the target has a gray value attached

to its neighboring high-brightness background, it fails.

- 10) GCM can detect the targets near high-brightness background in the simulated sequence since it calculates the local contrast information in a very small area. However, its performance is not satisfied when the background is complex, for example, in Seq. 2 and Seq. 4–Seq. 12.
- 11) As a global algorithm, IPI can successfully detect the targets even when the targets are not locally salient, for

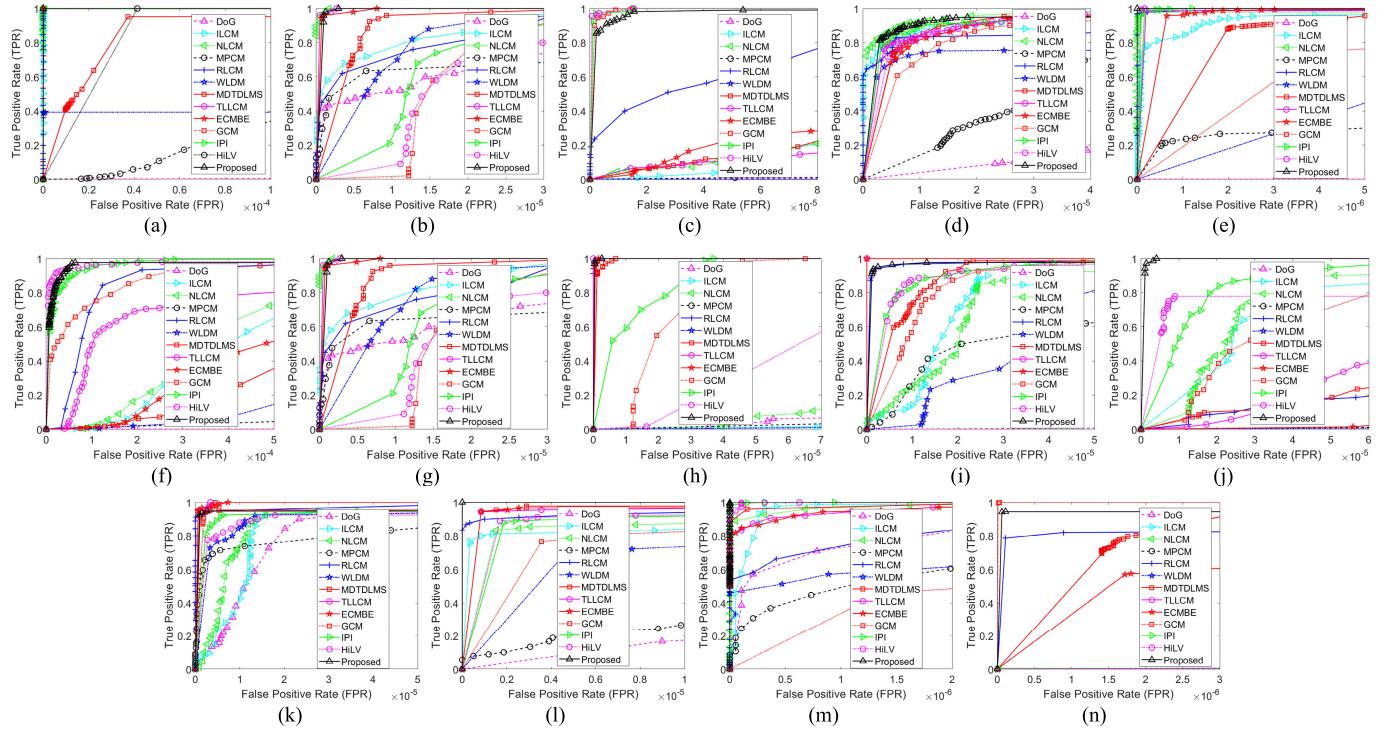


Fig. 16. ROC curves of different algorithms in different sequences. (a)–(m) Seq. 1–Seq. 13, respectively. (n) Simulated sequence.

example, in the 172nd frame of the simulated sequence. However, it is sensitive to complex edge/corner information and when the target is somewhat weak, these clutters may submerge the true target, for example, in Seq. 6, Seq. 7, and the 152nd frame of the simulated sequence.

- 12) HiLV, as another global algorithm, can detect the targets in Seq. 1–Seq. 5 and Seq. 13. However, since it introduces local contrast information during global decomposition, it cannot detect the targets near high-brightness background, for example, in the simulated sequence.
- 13) Compared to the existing algorithms, the proposed algorithm can extract all the true targets in the samples of the 14 sequences, and its performance in ROC curves is always in the forefront even when the target is weak and the clutter is complex, with no big fluctuation between different sequences. Especially, when the target is not locally salient, it can still achieve a good performance, such as in Seq. 4 and the simulated sequence.

2) *Comparisons of Computational Complexity:* Here, we will analyze the computational complexity for different algorithms. For simplicity, suppose that the resolution of the raw IR image is $X \times Y$, and the scale of the patch window or cell is $(2L + 1)^2$. For multiscale local algorithms, such as MPCM, RLCM, WLDM, and TLLCM, S means the scale number, L_i ($i = 1, 2, \dots, S$) means the L for the i th scale, and L_S is the largest L .

For DoG, there will be $(2L + 1)^2$ multiplications and $(2L + 1)^2$ additions for each pixel, so its computational complexity will be $O(L^2 XY)$.

For ILCM and NLCM, since they both use DoG as preprocessing and the latter subblock-stage processings consume less calculations, their computational complexity will be $O(L^2 XY)$.

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

For RLCM, the sort operation within a cell will cost $(2L_i + 1)^2 \log(2L_i + 1)^2$ calculations for each scale, so 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, but the entropy calculation will need a sort operation within a cell, which will cost $(2L_i + 1)^2 \log(2L_i + 1)^2$ calculations for each scale, so 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, and it is reported as $O(L^2 XY)$.

For TLLCM, the computational complexity is analyzed in the original paper, and it is reported as $O(SL_S^2 \log(L_S^2) XY)$.

For ECMBE, the computational complexity is analyzed in the original paper, and it is reported as $O(LXY)$.

For GCM, the computational complexity is analyzed in the original paper, and it is reported as $O[(X + m - 1) \times (Y + n - 1)]$. Here, m and n are the ring structuring element size.

For IPI, the computational complexity is analyzed in the original paper, and it is reported as $O[Nkmn \log(mn) + rc(p + 1)]$. Here, m is the number of pixels of the patch window, i.e., $(2L + 1)^2$. n is the number of patches, k is the number of nonzero singular value (rank) of G_k^T , N is the iteration number, and p is the overlapping pixel number during the transformation from the target/background patch image to the reconstruction image. r and c are the row and column

TABLE V
COMPUTATIONAL COMPLEXITY OF DIFFERENT ALGORITHMS

Algorithm	DoG	ILCM	NLCM	MPCM	RLCM
Computational Complexity	$O(L^2XY)$	$O(L^2XY)$	$O(L^2XY)$	$O(SL_s^2XY)$	$O(SL_s^2\log(L_s^2)XY)$
Algorithm	WLDM	MDTDLMS	TLLCM	ECMBE	GCM
Computational Complexity	$O(SL_s^2\log(L_s^2)XY)$	$O(SL_s^2\log(L_s^2)XY)$	$O(L^2XY)$	$O(LXY)$	$O((X+m-1) \times (Y+n-1))$
Algorithm	IPI	HiLV	Proposed		
Computational Complexity	$O(NkL^2n\log(L^2n) + XY(p+1))$	$O(NkL^2n\log(L^2n) + XY(p+1))$	$O(NkL^2n\log(L^2n) + XY(p+23))$		

TABLE VI
ABLATION EXPERIMENTS: THE SCRG AND BSF

Gaussian filter				√		√		√	
LMD		√				√		√	
Weighting operation		√		√				√	
Seq	Target	SCRG	BSF	SCRG	BSF	SCRG	BSF	SCRG	BSF
1	1	41.65	2.27E4	38.95	<i>1.47E5</i>	24.40	1.03E5	39.98	1.68E5
2	1	140.75	1.04E7	121.75	<i>2.36E7</i>	76.43	1.65E7	121.47	2.62E7
3	1	57.08	6.13E4	34.38	<i>5.97E5</i>	18.32	3.18E5	42.73	7.41E5
4	1	6.60	3.04E3	8.47	<i>4.78E3</i>	6.60	4.09E3	9.16	6.93E3
5	1	20.78	1.30E5	31.31	<i>2.53E5</i>	35.90	2.41E5	36.62	3.47E5
6	1	67.47	3.95E5	82.07	<i>1.94E6</i>	116.53	2.04E6	237.15	6.68E6
7	1	240.85	4.25E7	359.21	<i>2.32E8</i>	262.83	1.58E8	341.33	3.35E8
8	1	289.71	3.76E7	434.16	<i>2.10E8</i>	277.09	1.55E8	436.22	2.44E8
9	1	275.24	4.43E7	352.20	<i>2.56E8</i>	326.72	2.24E8	455.64	3.31E8
10	1	682.66	1.63E8	1.19E3	<i>6.96E8</i>	1.03E3	<i>7.13E8</i>	1.55E3	1.45E9
11	1	244.95	5.01E7	301.10	<i>1.85E8</i>	238.98	1.46E8	330.43	2.06E8
12	1	15.27	5.57E4	63.73	<i>4.18E5</i>	61.10	3.18E5	60.10	5.49E5
13	1	40.45		36.52	<i>8.31E3</i>	50.57		33.62	
	2	103.91	8.28E3	96.08		76.35	6.81E3	98.54	8.91E3
Sim(152 nd)	1	203.19	1.02E4	196.97	2.93E5	421.70	1.57E4	425.85	1.98E4
Sim(172 nd)	1	357.24	1.32E4	260.42	8.44E4	457.59	1.63E4	464.42	1.71E4

numbers of the original image, i.e., X and Y , respectively. Therefore, we can rewrite its computational complexity as $O[NkL^2n \log(L^2n) + XY(p + 1)]$ here.

For HiLV, the computational complexity is analyzed in the original paper, and it is reported as $O(mn^2)$. Here, m is the number of pixels of the patch window, i.e., $(2L + 1)^2$. n is the number of patches. However, after further analysis, we found that it is just the computational cost of SVD operation. In fact, the HiLV algorithm is based on IPI, so its true computational complexity should be $O[NkL^2n \log(L^2n) + XY(p + 1)]$ too.

The proposed algorithm consists of four main steps: Gaussian filtering operation, global decomposition, TGCM calculation, and weighting operation. For the Gaussian filtering operation, there will be nine multiplications and one addition for each pixel, totally $10XY$ operations for the whole image; for global decomposition, we have the same computational complexity as the IPI algorithm; for TGCM calculation, there will be eight comparisons for LMD operation, one division, one subtraction, and one multiplication for ratio-difference joint TGCM calculation, totally $11XY$ operations for the whole image; for the weighting operation, there

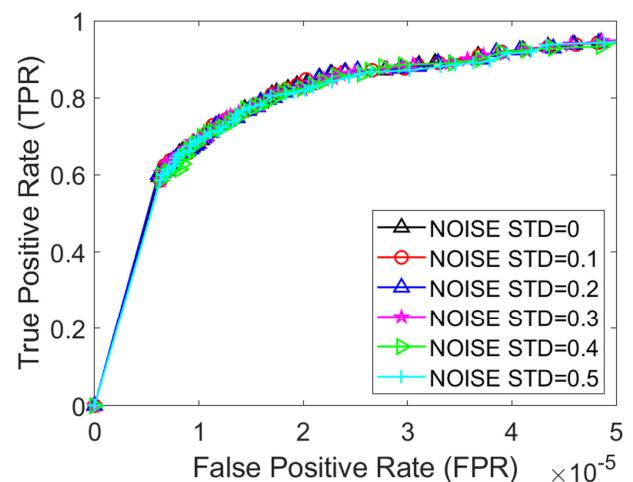


Fig. 17. Robustness to noises of the proposed algorithm. Seq. 6 is selected for test.

will be 1 multiplication for each pixel, totally XY operations for the whole image. Therefore, the computational complexity of the proposed algorithm will be $O[NkL^2n \log(L^2n) + XY(p + 23)]$.

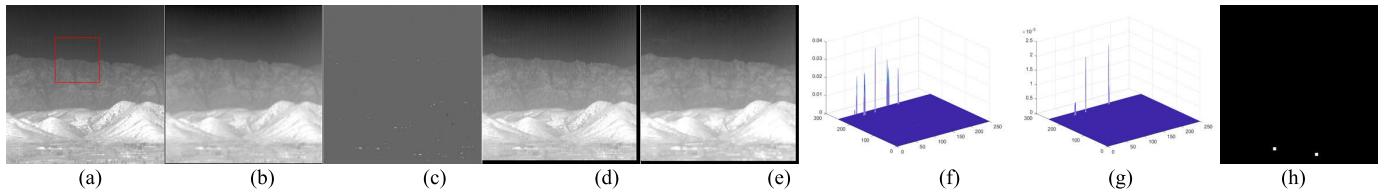


Fig. 18. Sample of detection failure. (a) Raw IR images. (b) Images after Gaussian filtering. (c) Global foreground T . (d) Global background B . (e) Global background B_{LMD} after LMD operation. (f) GCM results between (b) and (e). (g) GCM results weighted by (c). (h) Detection results after threshold operation.

Table V reflects the summary of the comparisons of computational complexity for different algorithms.

D. Ablation Experiments

There are three important modules in the proposed algorithm: the Gaussian filter, the LMD, and the weighting operation. To verify the effectiveness of each important module, some ablation experiments are conducted, and the results are shown in Table VI.

From Table VI, we can see that in most cases, the algorithm with all three modules can achieve the best (bold) or the second best (*italic*) SCRG and BSF, which means that these modules are all effective for improving detection performance.

E. Robustness to Noises

Noise is a key factor affecting IR small dim target detection. In this article, we selected one sequence (Seq. 6), added different levels of random noise to it, and then tested the robustness of the proposed algorithm. All these noises are Gaussian white noise with a mean value of 0, but with different standard deviations (ranging from 0.1 to 0.5). The ROC results are shown in Fig. 17. It can be seen that the performance of the proposed algorithm remains almost unchanged after adding noises.

F. Sample of Detection Failure

To illustrate the performance of the proposed algorithm further, an example of detection failure is provided here, as shown in Fig. 18. It can be seen that in the raw image, the target is too small and weak, even human eyes can hardly identify it. Therefore, after processing, the target is missed, but some clutters are output. In the future, we will continue to research and utilize more information (e.g., the time-domain information) to detect it.

VI. CONCLUSION

In this article, a new contrast method framework named TGCM is proposed for IR small target detection. Different from existing local contrast methods, its baseline for contrast information calculation is obtained globally, so it can deal with the situation when the target is near to some high-brightness background and is not locally salient. The sparse and low-rank global decomposition is utilized to acquire the global background baseline using the IPI model, and, to solve the problem of information loss at edge/corner, we analyzed the loss reason and proposed that some simple local operations can

be applied on the global background to recover as much edge/corner information as possible. Then, the joint ratio and difference TGCM is calculated between the raw image after Gaussian filtering and the global background baseline. The residual information of the global decomposition is adopted as a weighting function to suppress clutters further. Finally, a threshold operation is used on the SM to output true targets.

Experimental results on many real and simulated sequences show that the proposed algorithm can achieve good detection performance under different types of targets and backgrounds, and compared to some existing algorithms, the proposed algorithm can usually achieve better performance in target enhancement and background suppression (i.e., SCRG and BSF), so its detection rate and false alarm rate are usually better. Besides, ablation experiments are conducted to illustrate the effectiveness of some important modules of the proposed method, and the robustness to noises of the proposed algorithm is tested too.

The MATLAB code of this article has been made public on our Github, and readers can download it from the following website: <https://github.com/moradisaed/TGCM>.

ACKNOWLEDGMENT

Some of the testing infrared (IR) sequences used in this article are acquired from [88]. The authors would like to acknowledge the authors for their kind sharing. They would like to express their sincere appreciation to the editor and reviewers who provided valuable comments to help improve this article. Additionally, they are grateful to the researchers who provided comparative methods, such as GCM, infrared patch image (IPI), and high local variance (HiLV).

REFERENCES

- [1] R. Kou et al., "Infrared small target segmentation networks: A survey," *Pattern Recognit.*, vol. 143, Nov. 2023, Art. no. 109788, doi: [10.1016/j.patcog.2023.109788](https://doi.org/10.1016/j.patcog.2023.109788).
- [2] Y. Luo, X. Li, J. Wang, and S. Chen, "Clustering and tracking-guided infrared spatial-temporal small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5002520, doi: [10.1109/TGRS.2024.3384440](https://doi.org/10.1109/TGRS.2024.3384440).
- [3] Y. Hu et al., "Dynamic center point learning for multiple object tracking under severe occlusions," *Knowl.-Based Syst.*, vol. 300, Sep. 2024, Art. no. 112130, doi: [10.1016/j.knosys.2024.112130](https://doi.org/10.1016/j.knosys.2024.112130).
- [4] Y. Wang, J. Hsieh, P. Chen, M. Chang, H. So, and X. Li, "SMILEtrack: Similarity learning for occlusion-aware multiple object tracking," 2022, [arXiv:2211.08824](https://arxiv.org/abs/2211.08824).
- [5] Y. Zhang, H. Zhang, N. M. Nasrabadi, and T. S. Huang, "Multi-metric learning for multi-sensor fusion based classification," *Inf. Fusion*, vol. 14, no. 4, pp. 431–440, Oct. 2013, doi: [10.1016/j.inffus.2012.05.002](https://doi.org/10.1016/j.inffus.2012.05.002).
- [6] Z. Liu et al., "Robust multi-drone multi-target tracking to resolve target occlusion: A benchmark," *IEEE Trans. Multimedia*, vol. 25, pp. 1462–1476, 2023, doi: [10.1109/TMM.2023.3234822](https://doi.org/10.1109/TMM.2023.3234822).

- [7] D. Pang, T. Shan, W. Li, P. Ma, R. Tao, and Y. Ma, "Facet derivative-based multidirectional edge awareness and spatial-temporal tensor model for infrared small target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5001015, doi: [10.1109/TGRS.2021.3098969](https://doi.org/10.1109/TGRS.2021.3098969).
- [8] T. Ma, Z. Yang, J. Wang, S. Sun, X. Ren, and U. Ahmad, "Infrared small target detection network with generate label and feature mapping," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 6505405, doi: [10.1109/LGRS.2022.3140432](https://doi.org/10.1109/LGRS.2022.3140432).
- [9] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 580–587, doi: [10.1109/CVPR.2014.81](https://doi.org/10.1109/CVPR.2014.81).
- [10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: [10.1109/TPAMI.2016.2577031](https://doi.org/10.1109/TPAMI.2016.2577031).
- [11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788, doi: [10.1109/CVPR.2016.91](https://doi.org/10.1109/CVPR.2016.91).
- [12] Y. Liu et al., "Research on infrared dim target detection based on improved YOLOv8," *Remote Sens.*, vol. 16, no. 16, p. 2878, Aug. 2024, doi: [10.3390/rs16162878](https://doi.org/10.3390/rs16162878).
- [13] K. Xu, C. Song, Y. Xie, L. Pan, X. Gan, and G. Huang, "RMT-YOLOv9s: An infrared small target detection method based on UAV remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024, doi: [10.1109/LGRS.2024.3484748](https://doi.org/10.1109/LGRS.2024.3484748).
- [14] F. Wu, T. Zhang, L. Li, Y. Huang, and Z. Peng, "RPCANet: Deep unfolding RPCA based infrared small target detection," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Waikoloa, HI, USA, Jan. 2024, pp. 4797–4806, doi: [10.1109/WACV57701.2024.00474](https://doi.org/10.1109/WACV57701.2024.00474).
- [15] Y. Luo, X. Li, S. Chen, C. Xia, and L. Zhao, "IMNN-LWEC: A novel infrared small target detection based on spatial-temporal tensor model," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5004022, doi: [10.1109/TGRS.2022.3230051](https://doi.org/10.1109/TGRS.2022.3230051).
- [16] P. Zhang, L. Zhang, X. Wang, F. Shen, T. Pu, and C. Fei, "Edge and corner awareness-based spatial-temporal tensor model for infrared small-target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10708–10724, Dec. 2021, doi: [10.1109/TGRS.2020.3037938](https://doi.org/10.1109/TGRS.2020.3037938).
- [17] 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, doi: [10.1109/LGRS.2019.2954715](https://doi.org/10.1109/LGRS.2019.2954715).
- [18] C. Dang, Z. Li, C. Hao, and Q. Xiao, "Infrared small marine target detection based on spatiotemporal dynamics analysis," *Remote Sens.*, vol. 15, no. 5, p. 1258, Feb. 2023, doi: [10.3390/rs15051258](https://doi.org/10.3390/rs15051258).
- [19] D. Pang, T. Shan, P. Ma, W. Li, S. Liu, and R. Tao, "A novel spatiotemporal saliency method for low-altitude slow small infrared target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: [10.1109/LGRS.2020.3048199](https://doi.org/10.1109/LGRS.2020.3048199).
- [20] W. Yang and Z. Shen, "Small target detection and preprocessing technology in infrared image sequences," *Infr. Laser Eng.*, vol. 27, no. 1, pp. 23–28, Feb. 1998.
- [21] S. D. Deshpande, M. H. Er, R. Venkateswarlu, and P. Chan, "Max-mean and max-median filters for detection of small targets," *Proc. SPIE*, vol. 3809, pp. 74–83, Jul. 1999, doi: [10.1117/12.364049](https://doi.org/10.1117/12.364049).
- [22] 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, doi: [10.1117/1.oe.54.11.113107](https://doi.org/10.1117/1.oe.54.11.113107).
- [23] X. Bai and F. Zhou, "Analysis of new top-hat transformation and the application for infrared dim small target detection," *Pattern Recognit.*, vol. 43, no. 6, pp. 2145–2156, Jun. 2010, doi: [10.1016/j.patcog.2009.12.023](https://doi.org/10.1016/j.patcog.2009.12.023).
- [24] L. Deng, J. Zhang, G. Xu, and H. Zhu, "Infrared small target detection via adaptive M-estimator ring top-hat transformation," *Pattern Recognit.*, vol. 112, Apr. 2021, Art. no. 107729, doi: [10.1016/j.patcog.2020.107729](https://doi.org/10.1016/j.patcog.2020.107729).
- [25] H. Zhu, S. Liu, L. Deng, Y. Li, and F. Xiao, "Infrared small target detection via low-rank tensor completion with top-hat regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1004–1016, Oct. 2020, doi: [10.1109/TGRS.2019.2942384](https://doi.org/10.1109/TGRS.2019.2942384).
- [26] Y. Zhang, Z. Li, A. Siddique, A. Azeem, and W. Chen, "A real-time infrared small target detection based on double dilate contrast measure," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 17, pp. 16005–16019, 2024, doi: [10.1109/JSTARS.2024.3421646](https://doi.org/10.1109/JSTARS.2024.3421646).
- [27] L. Peng, Z. Lu, T. Lei, and P. Jiang, "Dual-structure elements morphological filtering and local Z-score normalization for infrared small target detection against heavy clouds," *Remote Sens.*, vol. 16, no. 13, p. 2343, Jun. 2024, doi: [10.3390/rs16132343](https://doi.org/10.3390/rs16132343).
- [28] R. Lu, X. Yang, W. Li, J. Fan, D. Li, and X. Jing, "Robust infrared small target detection via multidirectional derivative-based weighted contrast measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: [10.1109/LGRS.2020.3026546](https://doi.org/10.1109/LGRS.2020.3026546).
- [29] Y. Bi, J. Chen, H. Sun, and X. Bai, "Fast detection of distant, infrared targets in a single image using multiorder directional derivatives," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 56, no. 3, pp. 2422–2436, Jun. 2020, doi: [10.1109/TAES.2019.2946678](https://doi.org/10.1109/TAES.2019.2946678).
- [30] P. Yang, L. Dong, and W. Xu, "Detecting small infrared maritime targets overwhelmed in heavy waves by weighted multidirectional gradient measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: [10.1109/LGRS.2021.3080389](https://doi.org/10.1109/LGRS.2021.3080389).
- [31] C. Hao, Z. Li, Y. Zhang, W. Chen, and Y. Zou, "Infrared small target detection based on adaptive size estimation by multidirectional gradient filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5007915, doi: [10.1109/TGRS.2024.3502421](https://doi.org/10.1109/TGRS.2024.3502421).
- [32] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 11, pp. 1254–1259, Nov. 1998, doi: [10.1109/34.730558](https://doi.org/10.1109/34.730558).
- [33] 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," *Infr. Phys. Technol.*, vol. 55, no. 5, pp. 403–408, Sep. 2012, doi: [10.1016/j.infrared.2012.06.001](https://doi.org/10.1016/j.infrared.2012.06.001).
- [34] X. Wang, G. Lv, and L. Xu, "Infrared dim target detection based on visual attention," *Infr. Phys. Technol.*, vol. 55, no. 6, pp. 513–521, Nov. 2012, doi: [10.1016/j.infrared.2012.08.004](https://doi.org/10.1016/j.infrared.2012.08.004).
- [35] 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, doi: [10.1109/LGRS.2016.2519144](https://doi.org/10.1109/LGRS.2016.2519144).
- [36] 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, doi: [10.1016/j.patcog.2016.04.002](https://doi.org/10.1016/j.patcog.2016.04.002).
- [37] J. Chen, Z. Zhu, H. Hu, L. Qiu, Z. Zheng, and L. Dong, "Multi-scale local contrast fusion based on LOG in infrared small target detection," *Aerospace*, vol. 10, no. 5, p. 449, May 2023, doi: [10.3390/aerospace10050449](https://doi.org/10.3390/aerospace10050449).
- [38] 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, doi: [10.1109/TGRS.2013.2242477](https://doi.org/10.1109/TGRS.2013.2242477).
- [39] 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, doi: [10.1109/LGRS.2014.2323236](https://doi.org/10.1109/LGRS.2014.2323236).
- [40] 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, doi: [10.1109/LGRS.2016.2616416](https://doi.org/10.1109/LGRS.2016.2616416).
- [41] 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, doi: [10.1109/LGRS.2018.2790909](https://doi.org/10.1109/LGRS.2018.2790909).
- [42] X. Guan, Z. Peng, S. Huang, and Y. Chen, "Gaussian scale-space enhanced local contrast measure for small infrared target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 327–331, Feb. 2020, doi: [10.1109/LGRS.2019.2917825](https://doi.org/10.1109/LGRS.2019.2917825).
- [43] 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, doi: [10.1109/LGRS.2019.2954578](https://doi.org/10.1109/LGRS.2019.2954578).
- [44] J. Han, C. Liu, Y. Liu, Z. Luo, X. Zhang, and Q. Niu, "Infrared small target detection utilizing the enhanced closest-mean background estimation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 645–662, 2021, doi: [10.1109/JSTARS.2020.3038442](https://doi.org/10.1109/JSTARS.2020.3038442).
- [45] L. Liu, Y. Wei, Y. Wang, H. Yao, and D. Chen, "Using double-layer patch-based contrast for infrared small target detection," *Remote Sens.*, vol. 15, no. 15, p. 3839, Aug. 2023, doi: [10.3390/rs15153839](https://doi.org/10.3390/rs15153839).
- [46] J. Han et al., "A ratio-difference local feature contrast method for infrared small target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: [10.1109/LGRS.2022.3157674](https://doi.org/10.1109/LGRS.2022.3157674).

- [47] Z. Cui, J. Yang, S. Jiang, and J. Li, "An infrared small target detection algorithm based on high-speed local contrast method," *Infr. Phys. Technol.*, vol. 76, pp. 474–481, May 2016, doi: [10.1016/j.infrared.2016.03.023](https://doi.org/10.1016/j.infrared.2016.03.023).
- [48] 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, doi: [10.1109/TAES.2015.140878](https://doi.org/10.1109/TAES.2015.140878).
- [49] 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, doi: [10.1016/j.patcog.2016.07.036](https://doi.org/10.1016/j.patcog.2016.07.036).
- [50] H. Yao, L. Liu, Y. Wei, D. Chen, and M. Tong, "Infrared small-target detection using multidirectional local difference measure weighted by entropy," *Sustainability*, vol. 15, no. 3, p. 1902, Jan. 2023, doi: [10.3390/su15031902](https://doi.org/10.3390/su15031902).
- [51] 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, doi: [10.1109/LGRS.2019.2922347](https://doi.org/10.1109/LGRS.2019.2922347).
- [52] B. Xiong, X. Huang, and M. Wang, "Local gradient field feature contrast measure for infrared small target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 3, pp. 553–557, Mar. 2021, doi: [10.1109/LGRS.2020.2976208](https://doi.org/10.1109/LGRS.2020.2976208).
- [53] D. Zhou and X. Wang, "Research on high robust infrared small target detection method in complex background," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 6007705, doi: [10.1109/LGRS.2023.3297523](https://doi.org/10.1109/LGRS.2023.3297523).
- [54] R. Kou, C. Wang, Q. Fu, Y. Yu, and D. Zhang, "Infrared small target detection based on the improved density peak global search and human visual local contrast mechanism," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6144–6157, 2022, doi: [10.1109/JSTARS.2022.3193884](https://doi.org/10.1109/JSTARS.2022.3193884).
- [55] H. Wang et al., "Infrared small target detection based on weighted improved double local contrast measure," *Remote Sens.*, vol. 16, no. 21, p. 4030, Oct. 2024, doi: [10.3390/rs16214030](https://doi.org/10.3390/rs16214030).
- [56] H. Wei, P. Ma, D. Pang, W. Li, J. Qian, and X. Guo, "Weighted local ratio-difference contrast method for detecting an infrared small target against ground-sky background," *Remote Sens.*, vol. 14, no. 22, p. 5636, Nov. 2022, doi: [10.3390/rs14225636](https://doi.org/10.3390/rs14225636).
- [57] 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, doi: [10.1109/LGRS.2019.2898893](https://doi.org/10.1109/LGRS.2019.2898893).
- [58] Y. Tang, K. Xiong, and C. Wang, "Fast infrared small target detection based on global contrast measure using dilate operation," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, 2023, Art. no. 8000105, doi: [10.1109/LGRS.2023.3233958](https://doi.org/10.1109/LGRS.2023.3233958).
- [59] 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, doi: [10.1049/el:20045204](https://doi.org/10.1049/el:20045204).
- [60] S. Qi, J. Ma, H. Li, S. Zhang, and J. Tian, "Infrared small target enhancement via phase spectrum of quaternion Fourier transform," *Infr. Phys. Technol.*, vol. 62, pp. 50–58, Jan. 2014, doi: [10.1016/j.infrared.2013.10.008](https://doi.org/10.1016/j.infrared.2013.10.008).
- [61] 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, doi: [10.1117/12.170058](https://doi.org/10.1117/12.170058).
- [62] X. Kong, L. Liu, Y. Qian, and M. Cui, "Automatic detection of sea-sky horizon line and small targets in maritime infrared imagery," *Infr. Phys. Technol.*, vol. 76, pp. 185–199, May 2016, doi: [10.1016/j.infrared.2016.01.016](https://doi.org/10.1016/j.infrared.2016.01.016).
- [63] J. Zhao, Z. Tang, J. Yang, and E. Liu, "Infrared small target detection using sparse representation," *J. Syst. Eng. Electron.*, vol. 22, no. 6, pp. 897–904, Dec. 2011.
- [64] Y. He, M. Li, J. Zhang, and Q. An, "Small infrared target detection based on low-rank and sparse representation," *Infr. Phys. Technol.*, vol. 68, pp. 98–109, Jan. 2015, doi: [10.1016/j.infrared.2014.10.022](https://doi.org/10.1016/j.infrared.2014.10.022).
- [65] X. Zhang, K. Ren, M. Wan, G. Gu, and Q. Chen, "Infrared small target tracking based on sample constrained particle filtering and sparse representation," *Infr. Phys. Technol.*, vol. 87, pp. 72–82, Dec. 2017, doi: [10.1016/j.infrared.2017.10.003](https://doi.org/10.1016/j.infrared.2017.10.003).
- [66] H. Qin et al., "Infrared small moving target detection using sparse representation-based image decomposition," *Infr. Phys. Technol.*, vol. 76, pp. 148–156, May 2016, doi: [10.1016/j.infrared.2016.02.003](https://doi.org/10.1016/j.infrared.2016.02.003).
- [67] D. Liu, Z. Li, B. Liu, W. Chen, T. Liu, and L. Cao, "Infrared small target detection in heavy sky scene clutter based on sparse representation," *Infr. Phys. Technol.*, vol. 85, pp. 13–31, Sep. 2017, doi: [10.1016/j.infrared.2017.05.009](https://doi.org/10.1016/j.infrared.2017.05.009).
- [68] J. Chen, Z. Zhu, H. Hu, L. Qiu, Z. Zheng, and L. Dong, "A novel adaptive group sparse representation model based on infrared image denoising for remote sensing application," *Appl. Sci.*, vol. 13, no. 9, p. 5749, May 2023, doi: [10.3390/app13095749](https://doi.org/10.3390/app13095749).
- [69] 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, doi: [10.1109/TIP.2013.2281420](https://doi.org/10.1109/TIP.2013.2281420).
- [70] Y. Dai, Y. Wu, Y. Song, and J. Guo, "Non-negative infrared patch-image model: Robust target-background separation via partial sum minimization of singular values," *Infr. Phys. Technol.*, vol. 81, pp. 182–194, Mar. 2017, doi: [10.1016/j.infrared.2017.01.009](https://doi.org/10.1016/j.infrared.2017.01.009).
- [71] 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 Vis. Comput.*, vol. 63, pp. 1–9, Jul. 2017, doi: [10.1016/j.imavis.2017.04.002](https://doi.org/10.1016/j.imavis.2017.04.002).
- [72] H. Fang, M. Chen, X. Liu, and S. Yao, "Infrared small target detection with total variation and reweighted regularization," *Math. Problems Eng.*, vol. 2020, pp. 1–19, Jan. 2020, doi: [10.1155/2020/1529704](https://doi.org/10.1155/2020/1529704).
- [73] H. Zhu, H. Ni, S. Liu, G. Xu, and L. Deng, "TNLRS: Target-aware non-local low-rank modeling with saliency filtering regularization for infrared small target detection," *IEEE Trans. Image Process.*, vol. 29, pp. 9546–9558, 2020, doi: [10.1109/TIP.2020.3028457](https://doi.org/10.1109/TIP.2020.3028457).
- [74] X. Hao, X. Liu, Y. Liu, Y. Cui, and T. Lei, "Infrared small-target detection based on background-suppression proximal gradient and GPU acceleration," *Remote Sens.*, vol. 15, no. 22, p. 5424, Nov. 2023, doi: [10.3390/rs15225424](https://doi.org/10.3390/rs15225424).
- [75] Y. Liu, X. Liu, X. Hao, W. Tang, S. Zhang, and T. Lei, "Single-frame infrared small target detection by high local variance, low-rank and sparse decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5614317, doi: [10.1109/TGRS.2023.3291435](https://doi.org/10.1109/TGRS.2023.3291435).
- [76] Y. Dai and Y. Wu, "Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3752–3767, Aug. 2017, doi: [10.1109/JSTARS.2017.2700023](https://doi.org/10.1109/JSTARS.2017.2700023).
- [77] X. Zhang, Q. Ding, H. Luo, B. Hui, Z. Chang, and J. Zhang, "Infrared small target detection based on an image-patch tensor model," *Infr. Phys. Technol.*, vol. 99, pp. 55–63, Jun. 2019, doi: [10.1016/j.infrared.2019.03.009](https://doi.org/10.1016/j.infrared.2019.03.009).
- [78] X. Fan, A. Wu, H. Chen, Q. Huang, and Z. Xu, "Infrared dim and small target detection based on the improved tensor nuclear norm," *Appl. Sci.*, vol. 12, no. 11, p. 5570, May 2022, doi: [10.3390/app12115570](https://doi.org/10.3390/app12115570).
- [79] 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. 38996–39004, 2019, doi: [10.1109/ACCESS.2019.2906076](https://doi.org/10.1109/ACCESS.2019.2906076).
- [80] W. Li, M. Zhao, X. Deng, L. Li, L. Li, and W. Zhang, "Infrared small target detection using local and nonlocal spatial information," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 9, pp. 3677–3689, Sep. 2019, doi: [10.1109/JSTARS.2019.2931566](https://doi.org/10.1109/JSTARS.2019.2931566).
- [81] X. Guan, L. Zhang, S. Huang, and Z. Peng, "Infrared small target detection via non-convex tensor rank surrogate joint local contrast energy," *Remote Sens.*, vol. 12, no. 9, p. 1520, May 2020, doi: [10.3390/rs12091520](https://doi.org/10.3390/rs12091520).
- [82] Z. Qiu, Y. Ma, F. Fan, J. Huang, and L. Wu, "Global sparsity-weighted local contrast measure for infrared small target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, doi: [10.1109/LGRS.2022.3196433](https://doi.org/10.1109/LGRS.2022.3196433).
- [83] Z. Lu, Z. Huang, Q. Song, K. Bai, and Z. Li, "An enhanced image patch tensor decomposition for infrared small target detection," *Remote Sens.*, vol. 14, no. 23, p. 6044, Nov. 2022, doi: [10.3390/rs14236044](https://doi.org/10.3390/rs14236044).
- [84] Z. Zhang, C. Ding, Z. Gao, and C. Xie, "ANLPT: Self-adaptive and non-local patch-tensor model for infrared small target detection," *Remote Sens.*, vol. 15, no. 4, p. 1021, Feb. 2023, doi: [10.3390/rs15041021](https://doi.org/10.3390/rs15041021).
- [85] Z. Lin, A. Ganesh, J. Wright, L. Wu, M. Chen, and Y. Ma, "Fast convex optimization algorithms for exact recovery of a corrupted low rank matrix," in *Proc. CAMSAP*, Sep. 2009, pp. 1–18.
- [86] S. Moradi, P. Moallem, and M. F. Sabahi, "Scale-space point spread function based framework to boost infrared target detection algorithms," *Infr. Phys. Technol.*, vol. 77, pp. 27–34, Jul. 2016, doi: [10.1016/j.infrared.2016.05.007](https://doi.org/10.1016/j.infrared.2016.05.007).

- [87] C. Hilliard, "Selection of a clutter rejection algorithm for real-time target detection from an airborne platform," *Proc. SPIE*, vol. 4048, pp. 74–84, Jul. 2000, doi: [10.1117/12.392022](https://doi.org/10.1117/12.392022).
- [88] B. Hui et al., "A dataset for infrared detection and tracking of dim-small aircraft targets under ground / air background," *China Sci. Data*, vol. 5, no. 3, Sep. 2020, Art. no. 21, doi: [10.11922/csdata.2019.0074.zh](https://doi.org/10.11922/csdata.2019.0074.zh).



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 an Associate Professor with the College of Physics and Telecommunication Engineering, Zhoukou Normal University, Zhoukou, China. His research interests are target detection and image processing.



Saed Moradi received the bachelor's degree in electrical engineering from Urmia University, Urmia, Iran, in 2009, the master's degree in electronics engineering with a specialization in mixed-signal integrated circuits design from the Ferdowsi University of Mashhad (FUM), Mashhad, Iran, in 2012, and the Ph.D. degree in electronics engineering, focusing on computer vision and image processing, from the University of Isfahan, Isfahan, Iran, in 2018.

Following his doctoral studies, he undertook a two-year post-doctoral fellowship at Université Laval, where he worked on 3-D data processing. He is currently a Post-Doctoral Research Associate with the University of Waterloo, Waterloo, ON, Canada. His research interests lie in the areas of computer vision, machine learning, and optical metrology.



Bo Zhou received the Ph.D. degree in information and communication engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2011.

He is currently an Associate Professor with the School of Electronic Information and Communications, Huazhong University of Science and Technology. His research interests include infrared image processing and infrared detection system development.



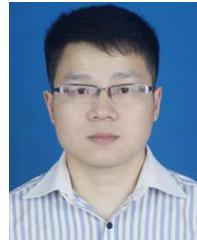
Wei Wang received the M.D. degree from Southwest Jiaotong University, Chengdu, China, in 2011, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2015.

He is currently the Director of Henan Big-data Development and Innovation Laboratory for Intelligent Multi-Modal Perception and the Associate Dean of the School of Artificial Intelligence, Zhoukou Normal University, Zhoukou, China. His research interests include image processing, 3-D computer vision, and pattern recognition.



Qian Zhao (Member, IEEE) received the Ph.D. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2017.

He is currently working as a Lecturer with the School of Physics and Telecommunication Engineering, Zhoukou Normal University, Zhoukou, China. His research interests include reinforcement learning, vision language navigation, and image processing.



Zhen Luo received the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2019.

He is currently a Lecturer with the School of Artificial Intelligence, Zhoukou Normal University, Zhoukou, China. His research interests are dim target detection and tracking.