IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 18, 2021, accepted June 2, 2021, date of publication June 14, 2021, date of current version July 12, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3089376

Fast and Robust Infrared Image Small Target Detection Based on the Convolution of Layered Gradient Kernel

TUNG-HAN HSIEH[1], CHAO-LUNG CHOU[2], YU-PIN LAN[3], PIN-HSUAN TING[4], AND CHUN-TING LIN[1]

[1]Institute of Photonic System, National Yang Ming Chiao Tung University, Guiren District, Tainan 711, Taiwan
[2]Department of Computer Science and Information Engineering, Chung Cheng Institute of Technology, National Defense University, Taoyuan 33551, Taiwan
[3]Institute of Lighting and Energy Photonics, National Yang Ming Chiao Tung University, Guiren District, Tainan 711, Taiwan
[4]Institute of Imaging and Biomedical Photonics, National Yang Ming Chiao Tung University, Tainan 711, Guiren District, Taiwan

Corresponding authors: Chao-Lung Chou (chaolung.chou@gmail.com) and Chun-Ting Lin (jinting@mail.nctu.edu.tw)

This work was supported by the Ministry of Science and Technology (MOST), R.O.C., under Grant 109-2221-E-009-154-MY3 and Grant 110-2623-E-A49-001.

ABSTRACT Infrared (IR) small target detection is challenging because the IR imaging lacks detailed features, weak shape features, and a low signal-to-noise ratio (SNR). The existing small IR target detection methods usually focus on improving their high detective performance without considering the execution time. However, high-speed detection is vital for various applications, such as early warning systems, military surveillance, infrared search and track (IRST), etc. This paper proposes a fast and robust single-frame IR small target detection algorithm with a low computational cost while maintaining excellent detection performance. We propose a layered gradient kernel (LGK) based on the contrast properties of the human visual system (HVS) and model it through a three-layer patch image model. The layered gradient kernel is used to convolute with the input IR frame to obtain its gradient map. The target detection is further performed on the acquired gradient map with an adaptive threshold method. This method is compared with eight representative small target detection algorithms to evaluate the performance. Experimental results demonstrate that the algorithm is fast and suitable for real-time applications, and it is very effective even when the small target size is as small as 2×2.

INDEX TERMS Infrared (IR) small target detection, signal-to-noise ratio (SNR), infrared search and track (IRST), human visual system (HVS), layered gradient kernel (LGK), real-time.

I. INTRODUCTION

Target detection techniques in infrared (IR) images have been widely used in many applications, such as early warning systems, military surveillance, infrared search and track (IRST), medical images, and so on [1], [2]. IR images usually lack certain detailed features, such as color, texture, and shape information. By definition, the small target's size is less than 0.15% of the whole images and pixels' ranges is from 2×2 to 9×9 [3]. IR target is hard to be detected due to long distance from IR sensor and easily interference from the sharpness and texture of features. The probed image is easily affected by the IR sensor's noise or complex background, resulting in it is hard to effectively separate the target and the background under a low signal-to-noise ratio (SNR).

A lot of small IR target detection methods have been designed over the last two decades. These methods focus mostly on how to enhance the target and suppress background regions as much as possible. The techniques could be roughly divided into two categories: single frame-based methods and sequential frames-based methods [4]–[6]. Single frame-based methods are also called detection-before-track (DBT) methods. Single frames-based methods usually highlight the target through pre-processing and then use a threshold to segment the target within the image. They are low computation cost, easy for hardware implementation, suitable

The associate editor coordinating the review of this manuscript and approving it for publication was Ravibabu Mulaveesala.

for real-time applications, and are widely used in practice. Sequential frames-based approaches are also called track-before-detection (TBD) methods. Sequential frames-based methods deal with the spatial-temporal domain and separate the small target from a series of images through the prior information such as the small target's shape, the grayscale change, and the motion path. Sequential frames-based methods usually perform better than single frame-based methods in images with low SNR. However, they are more complex and need more preliminary information, make them unsuitable for real-time applications. Moreover, the sequential frame-based methods can be decomposed into single frame-based manner within a certain period of time. Therefore, the research on single frame-based methods has become the leading research direction of the infrared dim and small target detection [7].

The common single frame-based detection approaches can be classified into three main categories: background consistency-based methods, the HVS-based methods, and patch image-based methods.

A. BACKGROUND CONSISTENCY-BASED METHODS

Background consistency-based methods firstly estimate the original image background by a particular filter; then, the target is enhanced and extracted from the background. The selection of filters is crucial because it directly affects the accuracy of the detection. The common methods are morphology-based methods, such as Top-Hat [8] and Hit-or-miss transformation [9]. Statistics-based methods, such as max-mean/max-median filter [10] and median subtraction filter [11]. Several improved methods based on two-dimensional least mean square (TDLMS), such as bilateral TDLMS (BTDLMS) filter [12] and edge directional TDLMS (EDTDLMS) filter [13].

B. HVS-BASED METHODS

HVS-based methods are mainly based on the fact that the response of the human eye's ganglion cells to contrast patterns is linear. Therefore, properly define the contrast between the target and the background is the most crucial part of these methods. Chen *et al.* proposed the local contrast measure (LCM) algorithm [14] by observing that small targets have discontinuity with their neighboring regions and concentrates in a small region, which can be seen as a homogeneous compact region. Han *et al.* proposed the improved LCM (ILCM) algorithm according to the HVS size-adaptation and attention shift mechanism to reduce the computation cost by using half window size as moving steps and image sub-blocks to estimate contrast [15]. Wu *et al.* proposed the DNGM algorithm, which uses a tri-layer sliding window to calculate the double-neighborhood gradient under a fixed window scale to avoid the ''expansion effect'' of the traditional multiscale HVS-based methods [5]. Qin *et al.* proposed a novel small target detection algorithm inspired from facet kernel and random walker (FKRW) algorithm, which is derived by the heterogeneity and compactness of the targets and the directional consistency of the backgrounds. [16]. Nasiri and Chehresa proposed the VARD method based on a three-layer model and calculate each variance of different layers of neighboring patch-images [17]. The VARD method can simultaneously enhance the target and suppress the background clutter with an acceptable speed.

C. PATCH IMAGE-BASED METHODS

Patch image-based methods vectorize the original IR image into a patch image and separate the target from the background using optimization algorithms. Then, the target image was obtained by reconstruction of the target patch-image. The most representative method called the IPI algorithm [18]. IPI utilizes the non-local self-correlation to depict the background and enhance and indicate the target by transforming it into an optimization problem of recovering low-rank and sparse matrices. IPI method has remarkable results but leads to a high computation complexity. To obtain a more precise reconstruction matrix, an efficient optimization algorithm based on alternating direction method of multipliers (ADMM) plus the difference of convex (DC) programming was designed, called Non-convex Rank Approximation Minimization (NRAM) [19]. This approach not only utilizes the weighted ℓ_1 norm to improve detection performance but also exploits the $\ell_{2,1}$ norm to overcome the elimination of strong edge. Comparing with IPI, NRAM has lower computation complexity.

In general, background consistency-based methods are most suitable for real-time applications in a simple scene with a smooth background. On the contrary, the HVS-based methods and the patch image-based methods are suitable for complex scenes with background noises by adopting additional techniques to improve the overall performance. But, their computation cost is higher and is not suitable for real-time applications.

This article presents a fast and robust single-frame infrared image small target detection method that supports real-time applications. We propose the layered gradient kernel (LGK) based on the contrast properties of the HVS and model it through a three-layer patch image model [17]. The proposed layered gradient kernel convolute with the input IR frame to obtain its gradient map and conduct the small target detection by an adaptive threshold algorithm.

The main contributions of this paper are as follows:

1) The proposed layered gradient kernel is apposite to image convolution and has a lower computational cost and high detection performance, which is very suitable for real-time applications.

2) The proposed method can detect extremely dim and small targets with excellent detection performance, and the detection capability is as small as 2×2.

The remainder of this paper is organized as follows: Section 2 reviews the related works. Section 3 presents the detail of our proposed method. Experimental results and analysis are demonstrated in section 4. Section 5 gives the conclusions.

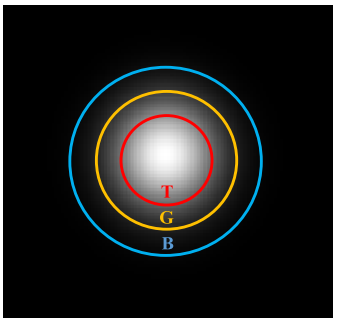


FIGURE 1. The three-layer patch-image model of a gaussian target in a simulated IR image. T denotes the target layer, G denote the guard layer and B denotes the background layer.

II. BACKGROUNDS

A. THREE-LAYER PATCH-IMAGE MODEL

According to the local intensity property, the brightness value of the target is greater than the value of its surrounding pixels in the infrared image [20]. Thus, the IR small target's intensity is usually more significant than the surrounding background and can be modeled as a Gaussian function and concentrate at the center. Common HVS-based methods considering the local intensity property usually divide the detected image into a target area and background. The VARD method subdivides the target area into two layers: the target layer and the protective (guard) layer, and proposed the three-layer patch-image model shows in Fig. 1 [17]. The target layer (T) consisting of the maximum brightness intensity values of the target in its center. The background layer (B) including the clutter of the neighboring target. The guard layer (G) includes a range of intensity values between the target and background layers.

The contrast between the target or the background in these three layers can capture more target features, so it is easier to separate the target from the background, thereby increasing the detection rate and reducing the false alarm rate. The VARD method uses a sliding window to scan the whole image to extract patch images. By utilizing the three-layer model for each patch-image to considerably enhance the targets and suppress the backgrounds based on the difference between different layers. Finally, by thresholding the different variance between the layers of the patch-images, targets are discriminated from the background.

B. GRADIENT FILTER

The gradient of an image $\nabla f\,(x, y)$ can be calculated by its directional derivatives given in (1).

$$\nabla f\,(x, y) = \left[\frac{\partial f\,(x, y)}{\partial x}, \frac{\partial f\,(x, y)}{\partial y}\right] \tag{1}$$

where $\frac{\partial f(x,y)}{\partial x}$ and $\frac{\partial f(x,y)}{\partial y}$ denote the derivatives respect to the x-axis and y-axis. The second-order derivative of an image denotes in (2).

$$\nabla^2 f\,(x, y) = \frac{\partial^2 f\,(x, y)}{\partial x^2} + \frac{\partial^2 f\,(x, y)}{\partial y^2} \tag{2}$$

The location of an edge is the same as the location of the extrema of $\frac{\partial f(x,y)}{\partial x}$ and $\frac{\partial f(x,y)}{\partial y}$, and it is equivalent to the locations of the zero-crossing of $\nabla^2 f\,(x, y)$. Zero-crossings whose strength is greater than a threshold and finding zero-crossings are much easier than finding extrema because of its local property. The approximation of a second-order derivative can be accomplished by doing a convolution between a kernel and an image to reduce computation cost.

The general expression of a convolution is given in (3).

$$g\,(x, y) = \sum_{u=-a}^{a} \sum_{v=-b}^{b} f\,(x - u, y - v) K\,(u, v) \tag{3}$$

where $g\,(x, y)$ is the filtered image, $f\,(x, y)$ is the original image and $K(u, v)$ is the filter kernel in which $-a \leq u \leq a$ and $-b \leq v \leq b$.

A 3×3 Laplacian filter is a common derivatives filter to obtain the gradient of an image given in (4) to (6) [21].

$$Lf_1 = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \tag{4}$$

$$Lf_2 = \begin{bmatrix} -1 & -0 & -1 \\ 0 & 4 & 0 \\ -1 & 0 & -1 \end{bmatrix} \tag{5}$$

$$Lf_3 = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \tag{6}$$

where Lf_1 filter can calculate the sum of four direction gradients, Lf_2 filter can calculate the sum of four diagonal direction gradients, and Lf_3 filter can calculate the sum of eight direction gradients.

Note that the values of Lf_1, Lf_2, and Lf_3 filters all add up to 0. It implies that for smooth backgrounds, the average value calculated by the kernel is around 0. If exists a target, the kernel center will enhance the target with a higher weight while decreasing the surrounding with a negative weight.

III. THE PROPOSED METHOD

A. LAYERED GRADIENT KERNEL

The local gradient property represents almost all gradients point to its center for a two-dimensional Gaussian function [20]. The distribution of gradients is different between target and background with strong edges, as shown in Fig. 2. We proposed the layered gradient kernel based on the Laplacian filter that combines the local intensity property and the local gradient property. The Laplacian filter is modified using the layered kernel to calculate the sum of all direction gradients and extract more characteristics from the kernel filter. The layered gradient kernel can be partitioned into three regions: the target layer, the guard layer, and the background layer, as shown in Fig. 3.

The target layer is used to capture the main energy of the target. The guard layer is used to separate the target from its neighbors and its size is determined by the captured image

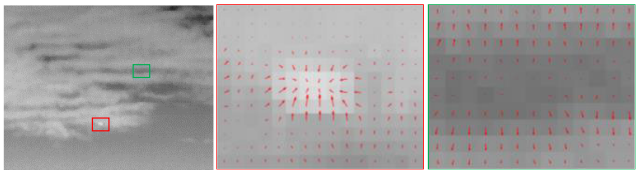


FIGURE 2. The local gradient property of an infrared image. (a) An original infrared image. (b) The gradient vectors of small target labeled as red rectangle. (c) The gradient vectors of local background labeled as green rectangle.

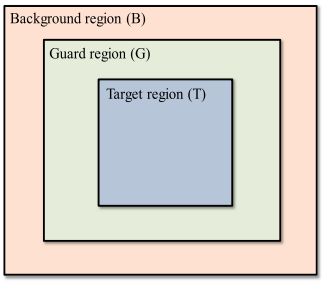


FIGURE 3. Sketch map of proposed layered gradient kernel. The internal is the target layer.

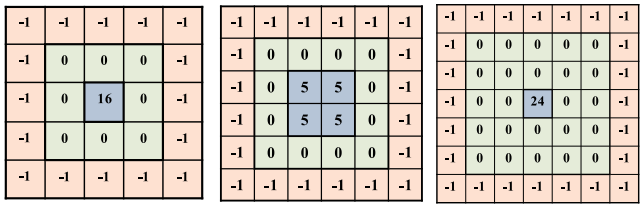


FIGURE 4. Examples of the proposed layered gradient kernel with different sizes. (a) and (c) have a 1 × 1 target layer and are expected to detect 1 × 1 infrared small targets; (b) has a 2 × 2 target layer and is expected to detect 2 × 2 infrared small targets.

resolution and its diffraction effect. The background layer is used to capture the surrounding background of the target.

The proposed layered gradient kernel size is $N \times N$, which N is determined by the expected target size and is generally 2 to 9. The size of the background layer is fixed as one pixel, and the coefficients are all set to -1. The coefficients of the guard layer are all set to 0, which represents no need to calculate gradients between the target layer and the guard layer. The interval between the target and background layers varies to keep the sum of all coefficients in the kernel equals zero.

Figure 4 shows examples of proposed layered gradient kernels with different sizes. Figure 4(a) and Figure 4(c) are the kernels both expected to detect 1×1 small infrared targets. Figure 4(b) is the kernel expected to detect 2×2 small infrared targets. The kernel for the small target of 1×1 may have the so-called pixel-size noises with high brightness (PNHB) effect and interfere with noises [16]. Therefore, the kernel size is recommended to set at least 6×6 or more for detecting small targets with a size of 2×2. Note that the larger the kernel size, the higher the computation cost required.

B. THE PROPOSED INFRARED SMALL TARGET DETECTION SYSTEM

The method consists of three stages, as shown in Fig. 5. Firstly, the original IR image convolved with the proposed layered gradient kernel and then obtains its gradient map. The pixel with a larger value represents a higher gradient. The small target has a higher gradient than the background and is located at the gradient map histogram's right side. Secondly, we utilize the adaptive threshold to transform the gradient map into a binary image and filter potential targets [14]. The potential small targets may be decomposed when the layered gradient kernel's size is larger than its size. In the last step, decomposed small targets will be combined into the final small target. The details of the proposed method are introduced below.

1) CONVOLUTION OF ORIGINAL IR IMAGE AND THE PROPOSED LAYERED GRADIENT KERNEL

In general, the gradient map's higher value represents the larger sum of local gradients and implies a higher possibility of the target's location. We use the proposed layered gradient kernel K_G to convolute with the input IR image I to obtain its gradient map (I_{GM}), which is given in (7).

$$I_{GM}(x, y) = \sum_{u=-a}^{a} \sum_{v=-b}^{b} I(x-u, y-v) K_G(u, v) \tag{7}$$

where (x, y) and (u, v) pairs are the sizes of the original IR image I and the layered gradient kernel K_G. Every element of the kernel is considered by $-a \le u \le a$ and $-b \le v \le b$.

The range of surrounding pixels and weighting are determined by the size of the layered gradient kernel and its coefficients. After the convolution, it is obvious that the smoothing background pixels concentrated on the low value in the gradient map histogram, as shown in Fig. 5. The small target has a higher value than the background and is located at the gradient map histogram's right side.

2) TARGET DETECTION THROUGH ADAPTIVE THRESHOLD

Background subtraction is a common and widely used technique whose aim is to detect changes in image sequences for generating a binary image containing the pixels belonging to foreground objects. We utilize the same idea to efficiently separate the background and the target by using the adaptive threshold to transform the gradient map into a binary image and discriminate potential targets. The adaptive threshold Th is described as (8) [14]:

$$Th = \mu_{I_{GM}} + k \times \sigma_{I_{GM}} \tag{8}$$

where k is a parameter decided by the test image, $\mu_{I_{GM}}$ and $\sigma_{I_{GM}}$ denote the mean and standard deviation from the gradient map, which are defined as (9) and (10), respectively.

$$\mu_{I_{GM}} = \frac{1}{M \times N} \sum_{i=1}^{M \times N} g_i \tag{9}$$

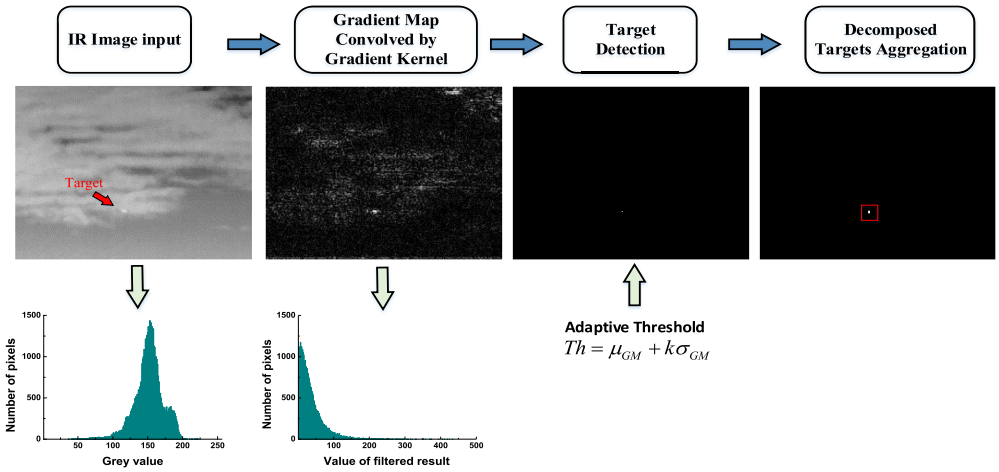


FIGURE 5. The proposed infrared small target detection system.

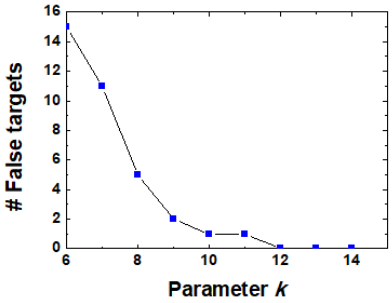


FIGURE 6. The correlation of the number of false targets and the parameter *k*.

$$\sigma_{I_{GM}} = \sqrt{\frac{1}{M \times N} \sum_{i=1}^{M \times N} (g_i - \mu_{I_{GM}})^2} \tag{10}$$

where g_i denotes the value of the its gradient map I_{GM}.

Note that the number of false targets can be effectively reduced by appropriately selecting the parameter k in (8). Fig. 6 shows an example of the correlation between the number of false targets and the parameter k for a 256×200 image, the value of k from 12 to 14 can effectively reduce the false targets. The target detection using the adaptive threshold is shown in Algorithm 1. The locations of non-zero pixel values represent possible small target locations.

3) DECOMPOSED TARGETS AGGREGATION

In generally, the feature of the infrared small target approximately seems as a spot-like target with Gaussian distribution. The target can be accurately detected in this situation. However, when the target has concrete shape, it may be separated by convolving with the proposed layered gradient kernel due to extracting the edge feature. To further suit the circumstances of relatively large size than a spot-like target and maintain the fast detection ability, we discard the concept of multiscale for layered gradient kernel. On the contrary, in the post-processing, the potential decomposed small targets are aggregated into a complete one that is shown in Algorithm 2. The decomposed targets aggregation is analog to the 4-connectivity operation in binary image morphology.

Fig. 7 shows the decompose targets aggregation procedure. Fig. 7 (b) demonstrates the Gradient map obtained by the original IR Image, which convolved with layered gradient kernel, and Fig. 7 (c) depicts the result of potential small targets aggregation.

Algorithm 1 Target Detection

Input: Original IR image $I(x, y)$.
Output: Binary image $I_B(x, y)$.

1: Convolute with the original IR image $I(x, y)$ and the layered gradient kernel K_G to obtain $I_{GM}(x, y)$ by using (7).
2: Compute the threshold Th according to (8).
3: Transform $I_{GM}(x, y)$ to a binary image $I_B(x, y)$ according to follows:

$$I_B(x, y) = \begin{cases} 1, & \text{if } I_{GM}(x, y) \geq Th, \\ 0, & otherwise, \end{cases}$$

Algorithm 2 Decomposed Targets Aggregation

Input: Binary image $I_B(x, y)$.
Output: Final Image $I_F(x, y)$.

```
for x = 1: rows do
    for y = 1: columns do
        if I_B(x, y) == 1 Then
            I_F(x, y) = 1
            I_F(x, y + 1) = 1
            I_F(x + 1, y) = 1
            I_F(x + 1, y + 1) = 1
        end if
    end for
end for
```

FIGURE 7. Illustration of decomposed targets aggregation. (a) Original IR Image (b) Gradient map obtained by original IR Image which convolved with layered gradient kernel. (c) Result of decomposed targets aggregation.

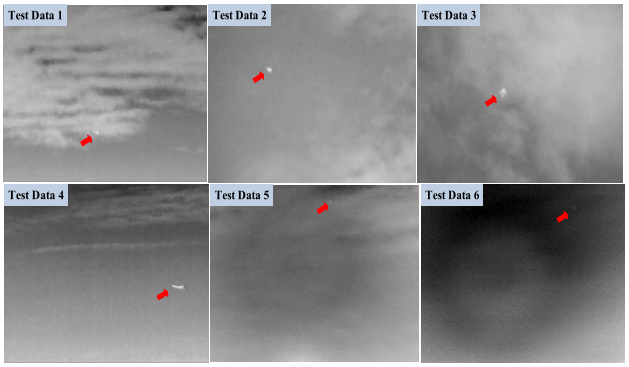


FIGURE 8. The six test data, red arrow point to the small target.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. EXPERIMENTAL SETUP

All the experiments are implemented by MATLAB software on a personal computer with 16-GB memory and 3.7-GHz Intel i3-4170 processor. To evaluate the proposed method's effectiveness and robustness, we used a total of six test data consisting of both the actual IR image sequences and the simulated images, as shown in Fig. 8. The details of the test data are listed in Table 1. Each test data exists only one target at each frame. Test Data 1–4 are actual IR image sequences with various target sizes, and Test Data 5–6 are simulated images both with only a single frame and 2×2 target sizes. Test Data 1 belongs to a complex background among six test data, and the rest of the test data belong to smooth backgrounds.

The signal-to-clutter ratio (SCR) is widely used to measure the difference between targets and backgrounds and can be frequently used to measure the detection difficulty. The SCR is defined as (11):

$$SCR = \frac{|\mu_t - \mu_b|}{\sigma_b} \tag{11}$$

where μ_t is the average pixel value of the target region, μ_b is the average pixel value of the neighboring region around the target, and σ_b is the standard deviation of neighboring region.

Generally, a lower SCR means that it is more difficult to detect the target. This implies that Test Data 2 (SCR = 8.89) is the easiest for small target detection, and Test Data 6 (SCR = 0.12) is the most difficult one.

TABLE 1. Details of the test data.

Test Data #	Image Resolution	Target Size (average)	Frame Number	SCR
1	256 × 200	4 × 6	30	1.74
2	320 × 240	8 × 9	100	8.89
3	320 × 240	11 × 11	66	2.64
4	256 × 200	4 × 15	40	5.48
5	260 × 210	2 × 2	1	1.21
6	200 × 150	2 × 2	1	0.92

B. RESULTS AND COMPARSIONS

Our method is compared with eight representative single frame-based detection methods to evaluate the performance in the experiments that is Top-Hat [8], LCM [14], ILCM [15], DNGM [5], FKRW [16], VARD [17], IPI [18], and NRAM [19]. The detailed parameter settings in the experiments are described in Table 2. Fig. 9 to Fig. 14 illustrates the detection results and the corresponding 3D display of different methods. The target location within the data set is marked with a red rectangle, and the peak in the 3D display shows the detected target.

The experimental results in Fig. 9 to Fig. 14 showed that our proposed method could efficiently detect all test data targets. From Fig. 9, the target within Test Data 1 is just at the edge of the cloud. The ILCM method reduces the contrast after sub-block averaging, resulting in missing the detection. Fig. 9 also shows that the IPI method will produce some black spots (on the top right corner), and it may require further post-processing to prevent possible false detections.

From Fig. 10 to Fig. 12, most methods have good detection performance. We observed that the LCM, ILCM, DNGM, FKRW, VARD, and NRAM methods cause the target to shrink or break down into smaller targets in the Test Data 2–4. This may phenomenon increase the detection error rate while the target is becoming smaller. Also, the FKRW method misses the detection in Test Data 4; its peak shows in an incorrect location, as shown in Fig. 12. Since the target size of Test Data 5 and Test Data 6 is small as 2×2, only our proposed method, DNGM method, IPI method, and FKRW method can detect the target, as shown in Fig. 13 and Fig. 14. Note that the DNGM method detected one false target in Test Data 5, and resulted in two peaks in Fig. 13.

To meet the practical application of infrared small target detection, such as tracking of long-distance targets in the early warning system, high-reliability detection must be achieved in the shortest time. To demonstrate the proposed method's performance against the other seven algorithms, the receiver operation characteristics (ROC) curves and time consumption are used for detection performance comparison. ROC curves represent the dynamic relationship between detection probability and false alarm probability. The detection probability P_d and false alarm probability P_f are defined as (12) and (13).

$$P_d = \frac{Number\ of\ true\ targets\ detected\ pixels}{Number\ of\ real\ targets\ pixels} \times 100\% \tag{12}$$

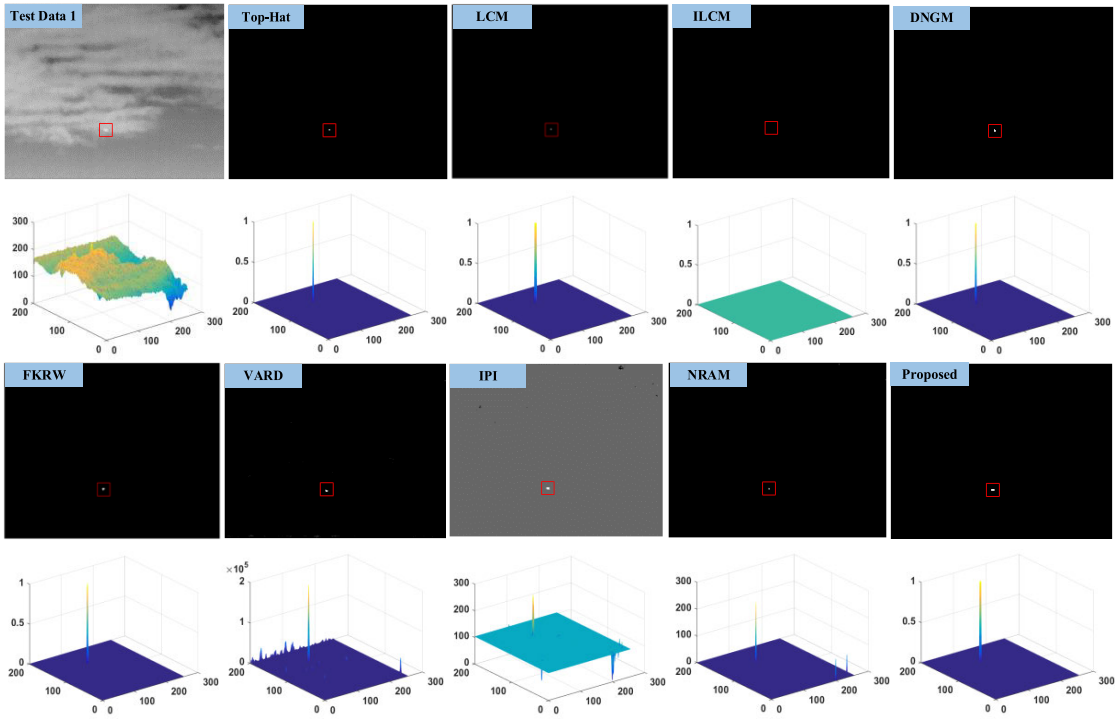


FIGURE 9. Detection results and 3D gray distribution map obtained by different methods on Test Data 1. The ILCM method missed the detection.

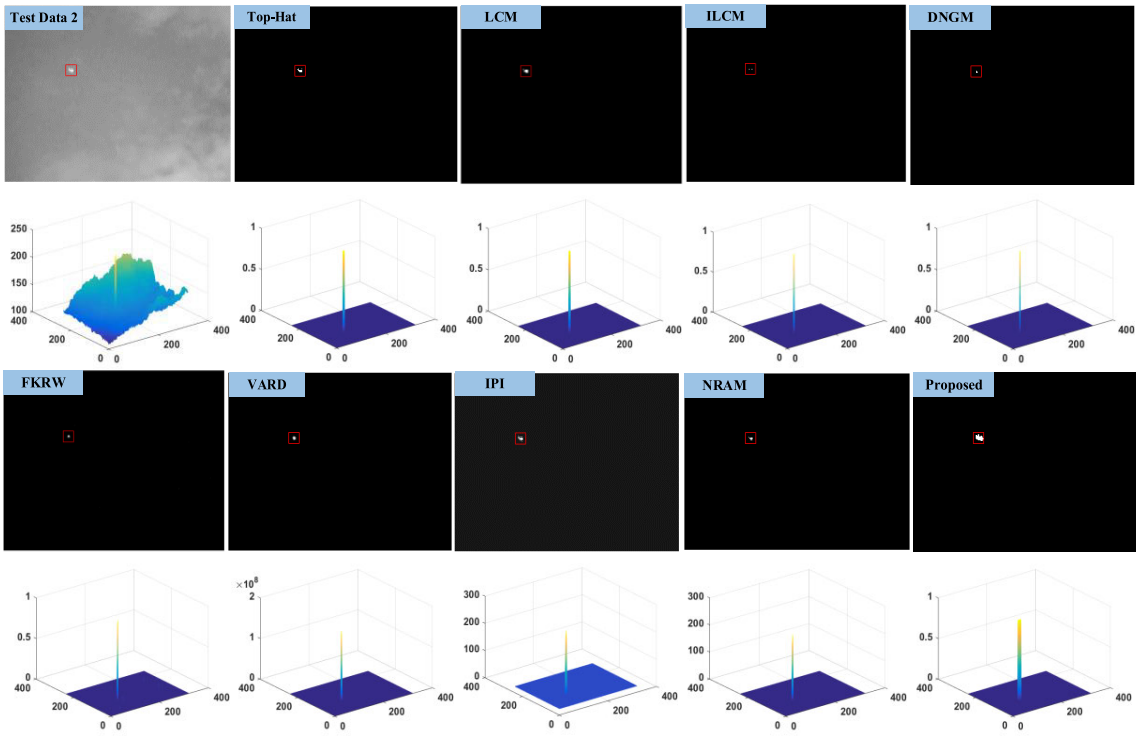


FIGURE 10. Detection results and 3D gray distribution map obtained by different methods on Test Data 2. All methods correctly detect the target.

$$P_f = \frac{Number\ of\ false\ alarm\ detected\ pixels}{Number\ of\ all\ pixels} \times 100\% \quad (13)$$

It should be noted that Test Data 5 and Test Data 6 consist of a single frame. Therefore, the experimental results of the ROC curves are merely for Test Data 1–4 which are composed of consecutive frames. Figure 15 shows the results of ROC curves of Test Data 1–4 of the proposed method and the other seven compared methods. Figure 15(a) shows that IPI and

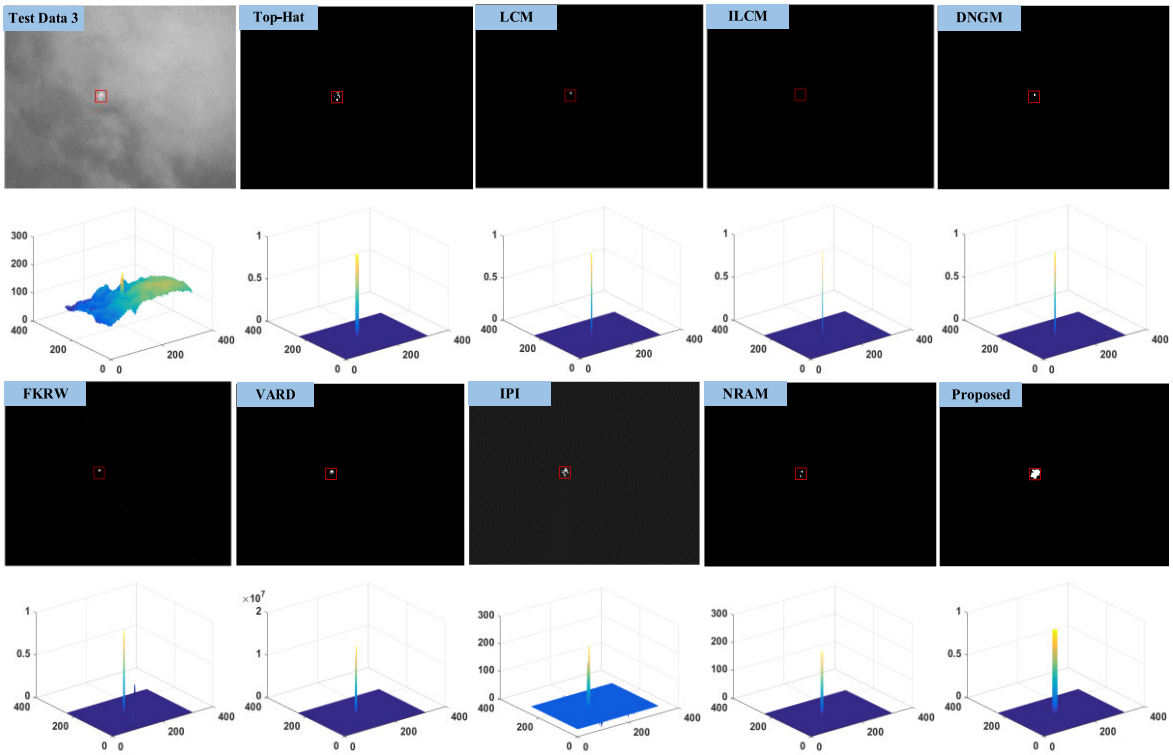


FIGURE 11. Detection results and 3D gray distribution map obtained by different methods on Test Data 3. All methods correctly detect the target.

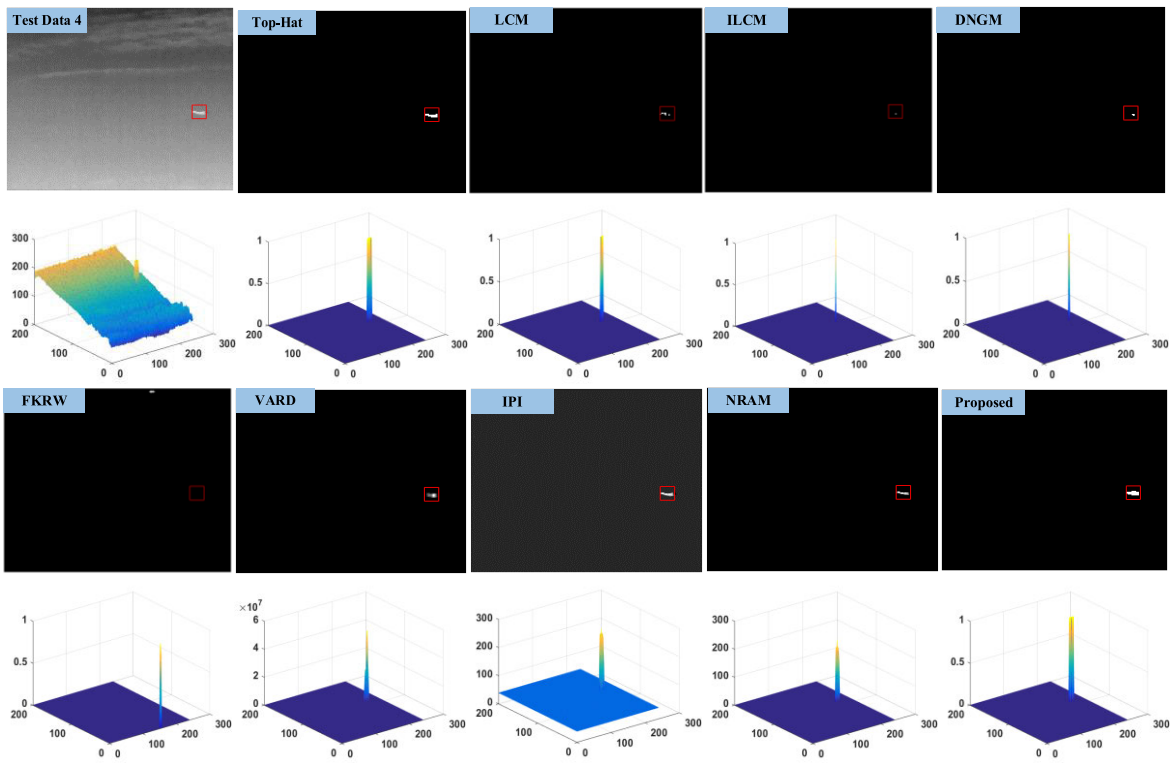


FIGURE 12. Detection results and 3D gray distribution map obtained by different methods on Test Data 4. The FKRW method missed the detection.

the proposed methods have better performance while ILCM, DNGM and VARD are much lower detection performance for Test Data 1. Figure 15(b) shows that IPI, VARD, and the proposed methods have better performance while Top-hat,

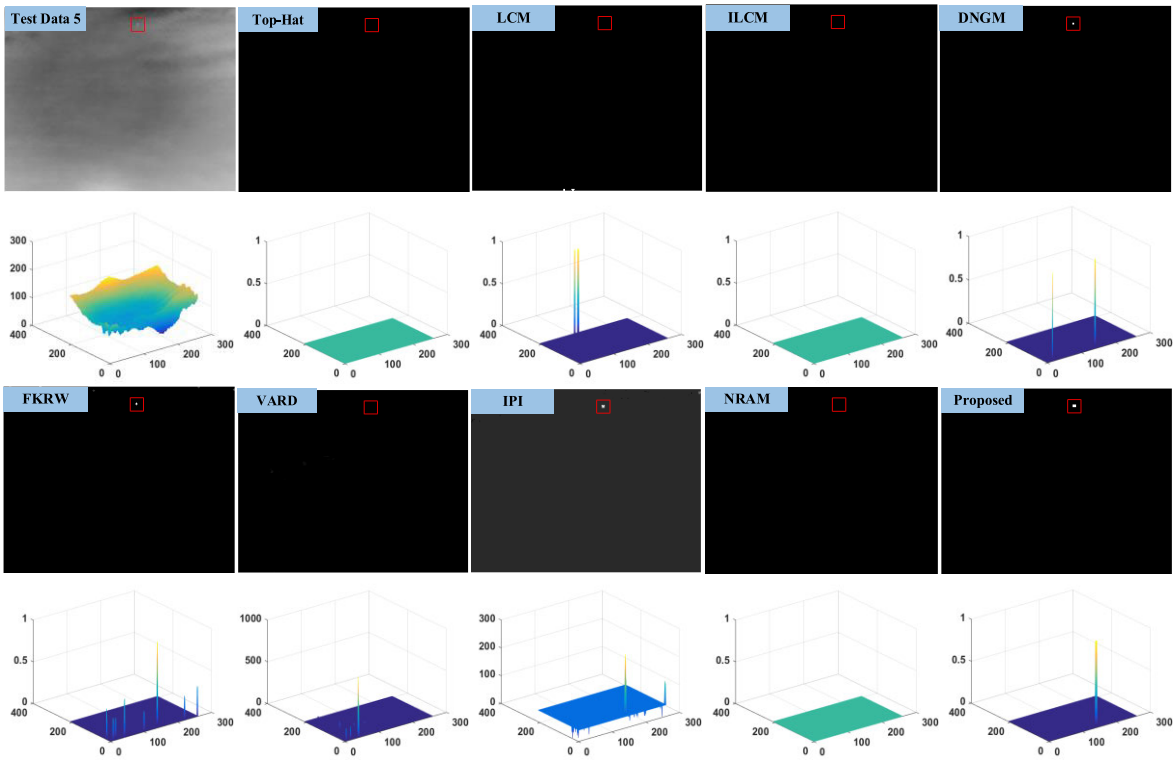


FIGURE 13. Detection results and 3D gray distribution map obtained by different methods on Test Data 5. The DNGM, FKRW, IPI and proposed methods correctly detect the target, while the Top-Hat, LCM, ILCM, VARD and NARM methods missed the detection.

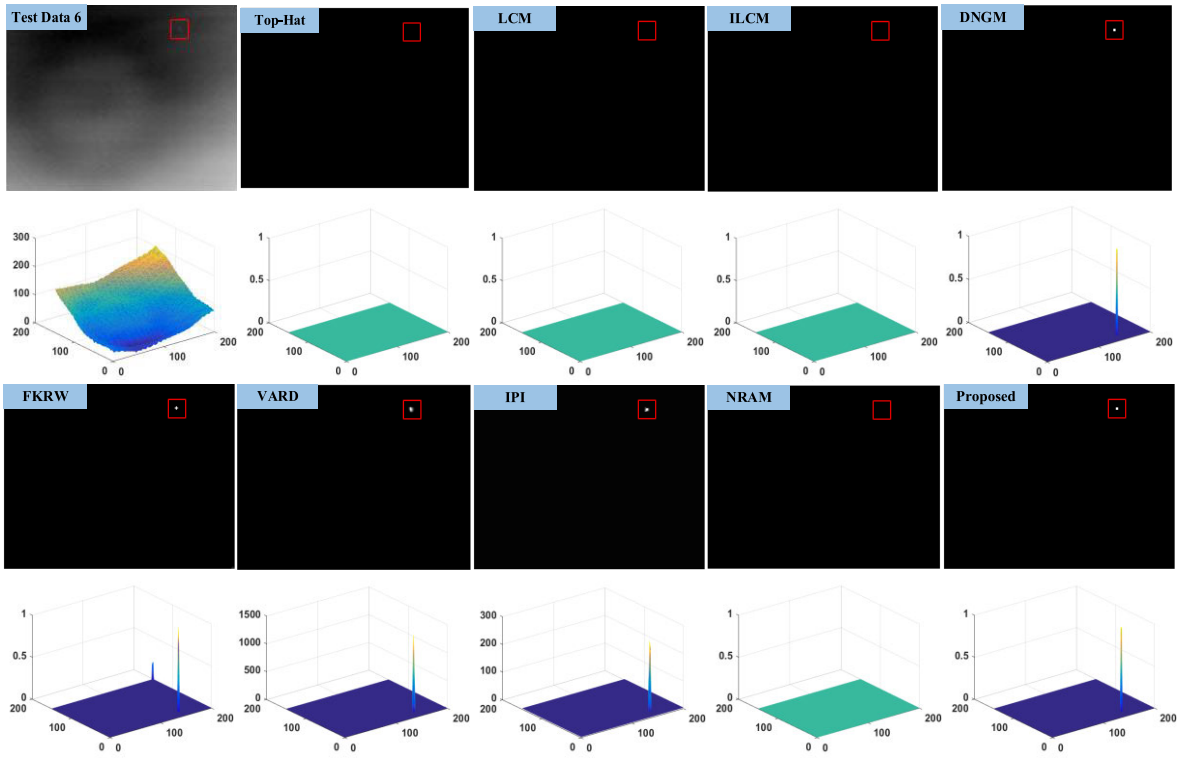


FIGURE 14. Detection results and 3D gray distribution map obtained by different methods on Test Data 6. The DNGM, FKRW, VARD, IPI and proposed methods correctly detect the target, while the Top-Hat, LCM, ILCM, VARD and NARM methods missed the detection.

LCM, ILCM, DNGM and NRAM are at lower detection performance for Test Data 2. Figure 15(c) shows that IPI, VARD, FKRW, and the proposed methods have better performance while Top-hat, ILCM, DNGM and NRAM are at

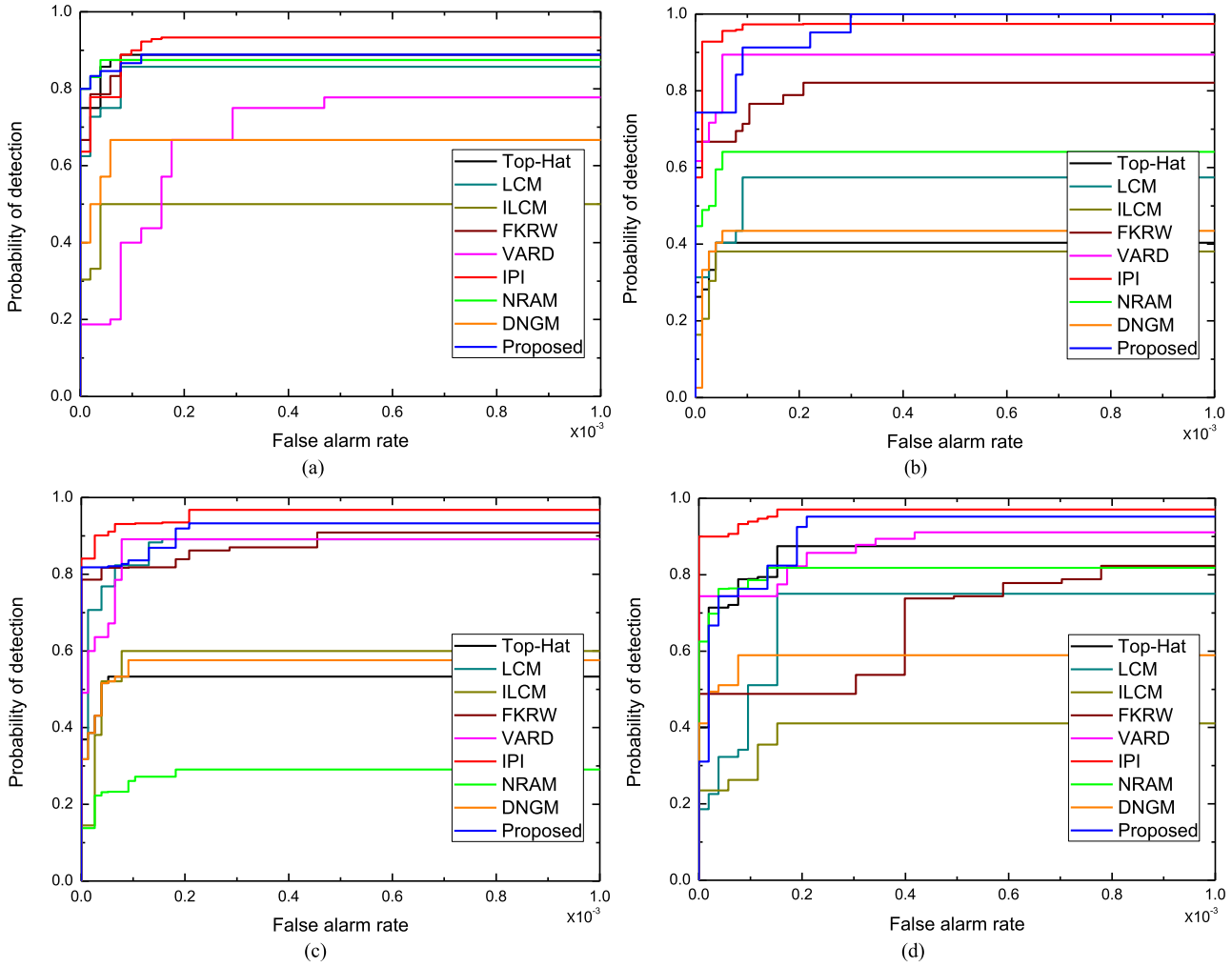


FIGURE 15. ROC curves on Test Data 1-4. (a) ROC curve of Test Data 1. (b) ROC curve of Test Data 2. (c) ROC curve of Test Data 3. (d) ROC curve of Test Data 4.

TABLE 2. Detailed parameter settings of the compared methods.

No.	Methods	Parameter settings
1	Top-hat [8]	Structure shape: disk, structure size: 2×2; Threshold: 25
2	LCM [14]	Local window size: 6×6, k=2.5~4
3	ILCM [15]	Sub-block size: 8×8, moving step: 4, k=85~115
4	DNGM [5]	window cell size: 3×3, k=40
5	FKRW [16]	k=4, p=6, patch size:11×11
6	VARD [17]	Target window: 7×7, guard window: 11×11, background window: 15×15
7	IPI [18]	Patch size: 50×50, sliding step: 10, $\lambda=1/\sqrt{min(m,n)}$, $\varepsilon=10^{-6}$
8	NRAM [19]	Patch size: 50×50, sliding step:10, $\lambda=1/\sqrt{min(m,n)}$, $\mu^0=3\sqrt{min(m,n)}$, γ=0.002, $C=\sqrt{min(m,n)}/2.5$, $\varepsilon=10^{-7}$,
9	Proposed method	Kernel size: 6×6 (Target window: 2×2, guard window: 4×4, background window: 6×6), k=12~14

lower detection performance for Test Data 3. Figure 15(d) shows that IPI and the proposed methods have better performance while LCM, ILCM, DNGM and FKRW are at lower detection performance for Test Data 4. Overall, we can see that the proposed method and IPI both take better detection probability than other methods.

The average computational cost of the different methods is compared and shown in Table 3. From Table 3, the proposed method and the VARD method achieves the fastest detection speed, which an average of 0.02 seconds. The Top-hat, ILCM, DNGM and FKRW can also achieve fast detection speed in less than one second, while

TABLE 3. Comparison of the time consumption (in seconds).

	Top-hat	LCM	ILCM	DNGM	FKRW	VARD	IPI	NRAM	Proposed
Test Data 1	0.1125	0.8912	0.1233	0.1341	0.4576	0.0245	6.2861	1.6226	**0.0217**
Test Data 2	0.1134	1.2689	0.1566	0.1375	0.4601	0.0251	14.0731	4.2481	**0.0221**
Test Data 3	0.1176	1.2875	0.1537	0.1339	0.4815	0.0271	15.5122	3.8411	**0.0225**
Test Data 4	0.1094	0.8353	0.0981	0.1295	0.5019	0.0264	5.2517	1.5967	**0.0211**
Test Data 5	0.1086	0.9313	0.1035	0.1236	0.5722	0.0249	8.8156	0.9354	**0.0236**
Test Data 6	0.1065	0.5549	0.0782	0.1164	0.3570	0.0215	3.0312	0.2298	**0.0202**

LCM, IPI and NRAM need longer time to complete the detection.

The comparison results show that background consistency-based methods (Top-hat) and HVS-based methods (LCM, ILCM, DNGM, FKRW and VARD) can achieve a fast detection speed; however, their detection performance is relatively lower for smaller targets detection. On the contrary, patch image-based methods (IPI and NRAM) can achieve better detection performance but need more computation cost. In summary, the proposed method can achieve not only the fastest detection speed but also with excellent performance in detection probability in all the eight detection methods.

V. CONCLUSION

This paper proposes a fast and robust single-frame infrared small target detection algorithm based on the human visual system and the convolution of the layered gradient kernel (LGK). Experimental results show the excellent detection performance. The experimental results also demonstrate the proposed method is especially good at detecting small infrared targets as small as 2×2. The detection speed of the proposed method is also superior to other detection methods under similar detection performance, which an average of 0.02 seconds for detecting a single frame. This proposed algorithm is quite suitable for real-time applications in single-frame small infrared target detection or further used as a basis in sequential target tracking.

ACKNOWLEDGMENT

The authors are thankful to Dr. Yuhwai Tseng and Wei-Yuan Weng for their precious comments and suggestions which are very helpful to improve the quality of the paper.

REFERENCES

[1] P. Zhang, X. Wang, X. Wang, C. Fei, and Z. Guo, "Infrared small target detection based on spatial-temporal enhancement using quaternion discrete cosine transform," *IEEE Access*, vol. 7, pp. 54712–54723, 2019.

[2] 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.

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

[4] S. S. Rawat, S. K. Verma, and Y. Kumar, "Review on recent development in infrared small target detection algorithms," *Procedia Comput. Sci.*, vol. 167, pp. 2496–2505, Jan. 2020.

[5] L. Wu, Y. Ma, F. Fan, M. Wu, and J. Huang, "A double-neighborhood gradient method for infrared small target detection," *IEEE Geosci. Remote Sens. Lett.*, early access, Jun. 29, 2020, doi: 10.1109/LGRS.2020.3003267.

[6] 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.

[7] S. Moradi, P. Moallem, and M. F. Sabahi, "Fast and robust small infrared target detection using absolute directional mean difference algorithm," *Signal Process.*, vol. 177, Dec. 2020, Art. no. 107727.

[8] V. T. Tom, T. Peli, M. Leung, and J. E. Bondaryk, "Morphology-based algorithm for point target detection in infrared backgrounds," *Proc. SPIE*, vol. 1954, pp. 2–11, Oct. 1993.

[9] X. Bai and F. Zhou, "Hit-or-miss transform based infrared dim small target enhancement," *Opt. Laser Technol.*, vol. 43, no. 7, pp. 1084–1090, Oct. 2011.

[10] S. D. Deshpande, M. H. Er, V. Ronda, and P. Chan, "Max-mean and max-median filters for detection of small-targets," *Proc. SPIE*, vol. 3809, pp. 74–83, Oct. 1999.

[11] J. Barnett, "Statistical analysis of median subtraction filtering with application to point target detection in infrared backgrounds," *Proc. SPIE*, vol. 1050, pp. 10–15, Jun. 1989.

[12] Y. Zhao, H. Pan, C. Du, Y. Peng, and Y. Zheng, "Bilateral two-dimensional least mean square filter for infrared small target detection," *Infr. Phys. Technol.*, vol. 65, pp. 17–23, Jul. 2014.

[13] T.-W. Bae, F. Zhang, and I.-S. Kweon, "Edge directional 2D LMS filter for infrared small target detection," *Infr. Phys. Technol.*, vol. 55, no. 1, pp. 137–145, Jan. 2012.

[14] 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.

[15] 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.

[16] Y. Qin, L. Bruzzone, C. Gao, and B. Li, "Infrared small target detection based on facet kernel and random Walker," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7104–7118, Sep. 2019.

[17] M. Nasiri and S. Chehresa, "Infrared small target enhancement based on variance difference," *Infr. Phys. Technol.*, vol. 82, pp. 107–119, May 2017.

[18] 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.

[19] L. Zhang, L. Peng, T. Zhang, S. Cao, and Z. Peng, "Infrared small target detection via non-convex rank approximation minimization joint $\ell_{2,1}$ norm," *Remote Sens.*, vol. 10, no. 11, p. 1821, Nov. 2018.

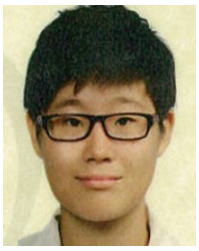
[20] H. Zhang, L. Zhang, D. Yuan, and H. Chen, "Infrared small target detection based on local intensity and gradient properties," *Infr. Phys. Technol.*, vol. 89, pp. 88–96, Mar. 2018.

[21] A. Distante and C. Distante, *Handbook of Image Processing and Computer Vision*, vol. 3. Basel, Switzerland: Springer, 2020.

TUNG-HAN HSIEH was born in Taichung, Taiwan, in 1992. He received the B.S. degree in electrophysics from National Chiayi University (NCYU), Chiayi, Taiwan, in 2014. He is currently pursuing the M.S. degree with the Institute of Photonic System, National Yang Ming Chiao Tung University (NYCU), Tainan, Taiwan. His current research interests include image processing and infrared target detection.

CHAO-LUNG CHOU received the Ph.D. degree in electrical and electronics engineering from the Chung Cheng Institute of Technology, National Defense University, Taiwan, in 2012. Since 2015, he has been an Assistant Professor with the Computer Science and Information Engineering Department, Chung Cheng Institute of Technology, National Defense University. His research interests include information security, image processing, machine learning, and biometrics.

YU-PIN LAN received the M.S. and Ph.D. degrees from the Institute of Electro-Optical Engineering, National Chiao Tung University, Hsinchu, Taiwan. She has an extensive professional career both in research and industry. Since 2016, she has been an Assistant Professor with the College of Photonics, National Yang Ming Chiao Tung University (NYCU), Taiwan. Her research interests include the laser engineering and physics.



PIN-HSUAN TING received the B.S. degree in electrical engineering from National Tsing Huang University, Hsinchu, Taiwan. She is currently pursuing the M.S. degree with the Institute of Imaging and Biomedical Photonics, National Yang Ming Chiao Tung University (NYCU), Tainan, Taiwan. Her research interests include image processing, machine learning, and communication systems.

CHUN-TING LIN received the B.S. and M.S. degrees in material science and engineering from National Tsing Huang University, Hsinchu, Taiwan, in 1997 and 2001, respectively, and the Ph.D. degree in electro-optical engineering from National Chiao-Tung University, Hsinchu, in 2007. Since 2015, he has been a Professor with the College of Photonics, National Yang Ming Chiao Tung University (NYCU), Taiwan. His research interests include radio-over-fiber systems, optical data formats, optoelectronic packages, and image processing.