

# Infrared Small Target Detection Algorithm Using an Augmented Intensity and Density-Based Clustering

In Ho Lee<sup>1</sup> and Chan Gook Park<sup>1</sup>, *Member, IEEE*

**Abstract**—In infrared search and tracking (IRST) systems, small target detection is challenging because IR imaging lacks feature information and has a low signal-to-noise ratio. The recently studied small IR target detection methods have achieved high detection performance without considering execution time. We propose a fast and robust single-frame IR small target detection algorithm while maintaining excellent detection performance. The augmented infrared intensity map based on the standard deviation speeds up small target detection and improves detection accuracy. Density-based clustering helps to detect the shape of objects and makes it easy to identify centroid points. By incorporating these two approaches, the proposed method has a novel approach to the small target detection algorithm. We have self-built 300 images with various scenes and experimented with comparing other methods. Experimental results demonstrate that the proposed method is suitable for real-time detection and effective even when the target size is as small as 2 pixels.

**Index Terms**—Density-based spatial clustering, image gradient, infrared (IR) image, small target detection.

## I. INTRODUCTION

INFRARED (IR) imagery target detection technology is widely used in many applications, such as early warning systems, military surveillance, IR search and tracking (IRST), and medical imaging [1]. IR small target detection has attracted considerable attention as a key technology, and many researchers have proposed IR target detection methods over the past 20 years. IR sensors for surveillance usually include external influences, such as atmospheric scattering, refraction, lens contamination, distortion, and various noises, because the target is more than tens of km away [2], [3], [4]. The IR sensing system receives a very blurry target intensity and background clutter. In the IR image, the target generally has a characteristic that the signal is weak and small. It is also nonsmooth and nonuniform, including various background environments (sky, sea, mountains, and man-made structures) and natural weather conditions (weather, temperature, and solar radiation). Therefore, it is very difficult to detect IR small targets with a high detection rate, low false alarm rate, and high-speed computation due to general facts.

- 1) Due to the large distance between the target and the detection sensor, the target usually has a faint gray level IR image [2].

Manuscript received 27 March 2023; accepted 2 May 2023. Date of publication 12 May 2023; date of current version 9 June 2023. This work was supported by the Theater Defense Research Center funded by the Defense Acquisition Program Administration under Grant UD200043CD. (Corresponding author: Chan Gook Park.)

The authors are with the Department of Aerospace Engineering and the Automation and Systems Research Institute, Seoul National University, Seoul 08826, South Korea (e-mail: skylih@snu.ac.kr; chanpark@snu.ac.kr).

Digital Object Identifier 10.1109/TGRS.2023.3274757

- 2) Various types of interference exist in IR images, such as high-brightness backgrounds, complex backgrounds, and pixel-sized noise [3].
- 3) In actual application, since the size of the target cannot be known, multisize detection capability is required [4].

Algorithms that deal with small target detection can be generally classified into two groups: sequential detection methods and single-frame detection methods [5]. The existing sequential detection method shows excellent performance in the assumption of a static background or a consistent target in an adjacent frame, as well as small target detection with prior knowledge of the target. However, it is generally difficult to obtain preset assumptions or prior knowledge in actual IRST systems. Therefore, small target detection using a single frame is the most practical method. This method usually highlights the target by preprocessing the image and then uses a threshold to segment the target within the image. Its low computational burden and ease of implementation make it suitable for real-time applications and widely used in practice.

A general single-frame-based detection method can be categorized into three groups [6].

- 1) The background consistency-based method first estimates the background of the original image with a specific filter. The target is then extracted based on background subtraction. Therefore, the selection of the filter, such as max-median, max-mean [7], morphology opening [8], and principal component pursuit (PCP) [9], is an important factor because it directly affects the detection accuracy [10], [11]. The high-boost-based multiscale local contrast measure (HB-MLCM) [12] and the multiscale weighted local contrast measure (MWLCM) [13] can significantly enhance the contrast between the target and background, and employing it may easily distinguish the target from the background. While this approach is straightforward and has low computational complexity, it suffers from a high false alarm rate where the true background is not accurately approximated.
- 2) The patch image-based method approach directly enhances the target region and suppresses the background [6]. It is important to define the contrast between the target and the background properly. Chen et al. [14] proposed a local contrast measure (LCM) algorithm by observing that a small target has a discontinuity with a neighboring area and concentrates on a homogeneous small area. Han et al. [3] use subblocks of the IR image by changing the human visual system size. Considering the Gaussian model for the spatial distribution

of the small targets, normalized Laplacian of Gaussian (LoG) is adopted in a scale-space manner to resolve target size variation during successive IR frames [15]. This approach is sensitive to background noise due to utilizing second-order derivatives. The trilinear LCM (TLLCM) [16] performs Gaussian filtering on the center layer of the template, which effectively eliminates the influence of noise. This method overcomes situations where the target is easily overwhelmed owing to very bright backgrounds. Recently, the density peaks' searching [17] method was proposed for the detection of small targets. It uses a density feature map that combines the first- and second-order local tetra patterns to obtain candidate targets.

- 3) The deep learning-based methods can automatically learn the hierarchical features of images by neural networks having the ability to approximate arbitrary functions using data. Using the advantages of neural networks, research is being conducted to apply various deep learning techniques, such as convolution neural network (CNN), generative adversarial networks (GANs) [18], and you only look once (YOLO) [19] in IR target detection. CNN-based methods can learn features of IR small targets in a data-driven manner. The first CNN-based IR small target detection [20] method designed a multilayer perception network. Then, Krišto et al. [19] fine-tuned several existing generic object detection networks for IR small target detection. DNA-Net [21] designed a tridirection dense nested interactive module with an attention model to achieve feature enhancement.

In general, the background consistency-based method can be applied in real-time to a simple scene with a soft background. On the other hand, the patch image-based method is suitable for complex scenes with background noise by applying additional techniques to improve overall performance. However, real-time performance is poor. Finally, deep learning-based methods are highly data-dependent. GAN is used for data augmentation of the original images to improve small targets, but these methods are dependent on the quality and quantity of the training dataset. If the quality of the dataset is poor, the detection performance deteriorates. Thus, it needs to have good quality of many IR images, including vivid small targets. YOLO detectors based on CNNs are very bad for small target detection, not to mention smaller targets up to a few pixels in IR images.

In this article, we present a fast and robust single-frame IR image multismall target detection method with a novel approach. The proposed method detects small targets through three steps, and effective techniques are applied to each step. The main contributions of this article are given as follows.

- 1) A new IR map is created to improve the detection speed and accuracy of the target by using the standard deviation of IR intensity to increase the contrast of the entire image without using patch-based contrast properties using the local features of the image.
- 2) Our newly proposed detection algorithm uses density-based clustering, which can accurately recognize geometric object forms or very small objects of

$2 \times 1$  size and classify them as one object. To date, there has been no research on detecting targets using clustering, but multitarget detection can be achieved through clustering.

- 3) Using the characteristics of the size and IR intensity of the small target, we construct a layered window rather than a sliding window. It is possible to quickly extract a small target from among several candidating objects without a complex equation.
- 4) We develop 300 single-frame IR images by photographing the drone with a thermal imaging camera. Our dataset is composed of numerous target shapes, various target sizes, and diverse backgrounds.
- 5) We compare the performance of the proposed method with existing algorithms using a self-generated IR imaging dataset and publicly available datasets. Compared to existing methods, our method is more robust to the variations of complex background, target size, target number, and target shape. In addition, it shows a high detection rate and a low false positive rate while having real-time performance.

The structure of this article is given as follows. The motivation for small target detection in IR images is introduced in Section II-A. The relevant methodologies for this research are described in Sections II-B and II-C. Section III presents the details of the proposed method. We introduce our self-generated dataset in Section IV and use the public datasets in Section IV-A, comparing the parameter settings of the other baseline methods in Section IV-B, and the evaluation metric is represented in Section IV-C. We describe experimental comparison results and analysis in Sections V-A and V-B. Finally, conclusion and future work are drawn in Section VI.

## II. MOTIVATION AND METHODOLOGY

### A. Motivation

The IR sensor converts IR wavelengths into gray-scale intensity. It is sensitive to the temperature distribution. If the temperature of an object and around the object does not differ significantly, the contrast of the image is lowered. IR imagery has less information than a colored image because of a single channel. Thus, it needs to perform histogram equalization for the recognition of an object from the background. Several studies show that sliding window-based contrast improvement methods handle low contrast problems. However, the sliding window method takes a long time to calculate because it has to sweep all the pixels of an image. Since the target has relatively larger intensity pixels than the intensity of neighboring pixels, we have adopted a method that complements the contrast through the amount of variation in average intensity for each image. This effect also helps detect pixels where the target exists. The intensity of a pixel is mapped into a deviation dimension by means of intensity. In IR images, pixels belonging to large objects form a large proportion and cause large deviations at the boundary. Contrary to this, the target pixels only form a tiny proportion and have large deviations at a few positions. These distribution characteristics of small targets motivate us to apply density-based clustering to locate the

small target. We apply to traditional DBSCAN algorithm [22] and make it suitable for the needs of small target detection.

### B. Standard Deviation of Intensity Variation

In image processing, mean and standard deviation are used to describe the intensity distribution of an image. The mean is simply the average intensity of all the pixels in an image, while the standard deviation is a measure of how much the intensity of the pixels varies from the mean. The average of intensity  $m$  is corresponding to the measuring points at  $I(i, j)$  as

$$m = \frac{1}{w \times h} \sum_{i=1}^w \sum_{j=1}^h I(i, j) \quad (1)$$

where  $w$  and  $h$  are the horizontal and vertical sizes of the images. The IR intensity average is obtained by dividing the sum of the intensity values of all pixels by the number of all image pixels. Variance is calculated using the average value, which is the sum of the squared differences between the pixel intensity value and the average value. Then, dividing by the number of pixels is the variance, and the standard deviation is calculated as the square root of the variance

$$\sigma = \sqrt{\frac{1}{w \times h} \sum_{i=1}^w \sum_{j=1}^h [I(i, j) - m]^2}. \quad (2)$$

The mean and standard deviation can be used to characterize the overall contrast of an image. The mean is used to describe the brightness or darkness of an image. The standard deviation is an indirect method of calculating the dispersion of gray level intensity in black and white photographs. An image with a high mean and low standard deviation will have a small range of intensity values and, therefore, appear very flat. Conversely, an image with a low mean and high standard deviation will have a large range of intensity values and appear very in contrast. For image processing, these properties are used for making filters referred to as spatial filters and noise reduction filters. In addition, it can detect the edge according to the value using the average and standard deviation of the image.

The standard deviation can indicate whether a significant change in image gradient is present in IR images. Fig. 1(a) shows the application of the standard deviation effect, and Fig. 1(b) shows the normal IR intensity value. In the case of small targets, a large change in IR intensity value occurs in a small area so that it can be easily detected through the standard deviation effect. However, if the IR average value is similar to the IR value of a small target, the effect of the standard deviation disappears. We applied augmented IR values to solve this problem. A detailed description thereof is described in Section III-B.

### C. DBSCAN Method

K-means [23] and hierarchical [24] achieve clustering based on the distance between the two points. Although the calculation is simple, the number of clusters must be specified, and there are limitations in classifying clusters of different sizes or various geometric forms. On the other hand, density-based clustering generates clusters using location information

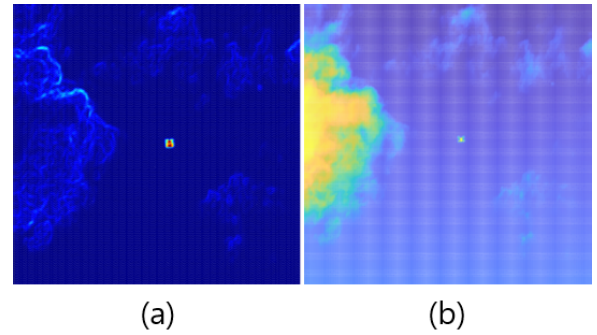


Fig. 1. Effect of standard deviation. (a) Standard deviation. (b) Normal.

between data and classifies places where points are concentrated into one cluster [22]. DBSCAN aims to identify clusters in terms of high-density regions separated by low-density regions. There are two parameters: a distance threshold  $\epsilon$  and the least minimum number of points  $\text{MinPts}$

$$C_\epsilon(p) = \{q \in D | \text{dist}(p, q) \leq \epsilon\} \quad (3)$$

$$\text{dist}(p, q) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2} \quad (4)$$

where  $p$  and  $q$  are any points, and a distance metric is the Euclidean distance

$$|C_\epsilon(p)| \geq \text{MinPts}(N). \quad (5)$$

If there are more than  $\text{MinPts}$  in the distance threshold  $\epsilon$  based on a core point, it is recognized as a cluster. Points that do not enter any cluster are recognized as noise and excluded. Density-based clustering does not need to specify the number of clusters in advance and can effectively exclude noise. Therefore, it is possible to robustly process general characteristics of IR images with poor image quality, including noise. In addition, it has the advantage of being able to find clusters with complex shapes due to unspecified distributions. Obtaining a center point for each cluster with the average position of pixels classified into the same cluster can be easy.

In the example, as shown in Fig. 2, the DBSCAN algorithm divides three clusters  $C_1 = \{P_1, P_2, \dots\}$  (red circles),  $C_2 = \{P_3, \dots\}$  (blue circles), and  $C_3 = \{P_4, \dots\}$  (yellow circles) with two parameters  $\epsilon$  and  $\text{MinPts}$ . The algorithm starts by choosing an arbitrary point, such as  $P_1$ , and checks the number of points within the distance  $\epsilon$ . If that number is greater than or equal to the  $\text{MinPts}$  within the distance  $\epsilon$ , it considers all those points as one cluster. Next, it expands the cluster by checking each point in that cluster, such as from  $P_1$  to  $P_2$ . In contrast,  $P_5$  is an outer point in the cluster because it does not satisfy the  $\text{MinPts}$ . That point is marked as noise. This process is repeated recursively until it runs out of points in that region. If no point is longer satisfying the condition, it is classified as one cluster  $C_1$ . The same process is performed at any point in another region repeatedly. In this example, a total of three clusters were found. In addition, since each cluster knows the location of all the points, it is possible to easily calculate the center point of the cluster, as shown diamond mark in Fig. 2.

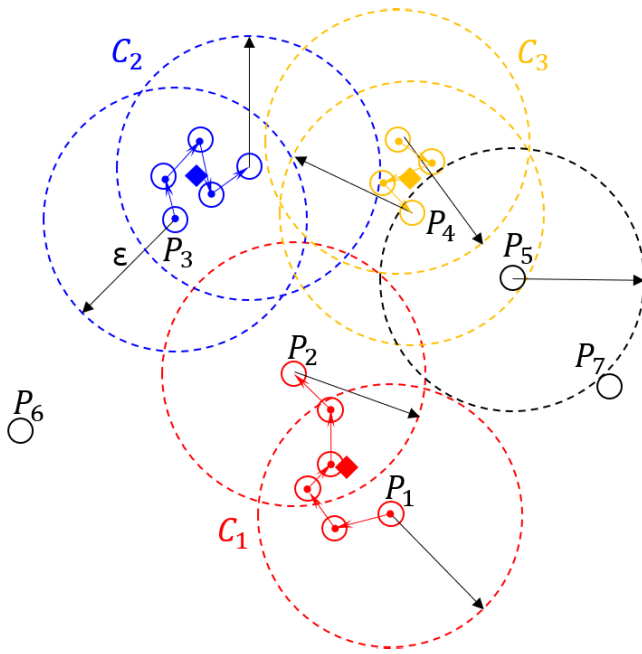


Fig. 2. Example of DBSCAN with parameter  $\epsilon$  and MinPts = 4.

### III. IMPLEMENTATION DETAILS

This section presents the small target detection algorithm in IR images. The overall algorithm sequence can be found.

#### A. Overall Detection Method

As shown in Fig. 3, the proposed approach consists of three modules. First, with IR image preprocessing, small targets usually have a greater intensity value than surrounding pixels. Other objects also have greater intensity values than background pixels. Using this property, each image's relative IR value deviation is standardized to find pixels that may have a target. The detected pixels may classify a candidating region through the first density-based clustering [see Fig. 3(a)]. Second, since there is a target in the candidating area, a boundary box of an appropriate size is set to recognize objects through the second density-based clustering of pixels with large gray values in the boundary box. When several candidating regions are concentrated, a boundary box may be overlapped. In this case, one boundary box is combined in consideration of the IOU. As shown in Fig. 3(b), it can be seen that objects are classified, and very close pixels with high gray intensity values are recognized as one object. Third, we hope to detect small targets, so we exclude large objects by measuring the size of the objects. In order to detect a final target from among the selected small candidating targets, we have designed a new window by adding a layer of core cell and peripheral neighborhood cell. The entire window is divided into two regions: the core cell is the target cell, and the neighborhood cell is the background cell. Using the property that a small target has a greater intensity value than the surrounding pixels when the target locates in the center of a window, the intensity difference will be reflected in the two regions regardless of the size of the small target, as shown in Fig. 3(c). We can detect small targets with this algorithm.

#### B. Preprocessing: Candidating Area

In the IR image, IR intensity values are distributed for each pixel according to the temperature of the object. It is necessary to detect a candidating area in which a target may exist by using a change in intensity value. A considerable number of pixels may be detected when detecting simply using an intensity value above the threshold using (6), as shown in Fig. 4

$$I(x, y) > \text{Max}\{I(0 < i < w, 0 < j < h)\} \times \alpha \quad (6)$$

where  $w$  and  $h$  are the width and height of the image, respectively.  $\alpha$  is the threshold rate between 0 and 1. The same color means the same cluster. The area to be reviewed may be widened, and the number of pixels detected is too high, resulting in poor real-time performance.

To prevent this, the calculation burden is lowered by excluding unnecessary parts through boundary detection, where the amount of change in IR intensity becomes more than a threshold value. However, when the contrast among the IR images is low, the amount of change in the intensity value may be small. In this case, since very few pixels may be detected, the problem is solved by detecting pixels with relatively large deviations using the standard deviation of IR intensity variation for each image

$$\text{Std}(x, y) > \text{Max}\{\text{Std}(0 < i < w, 0 < j < h)\} \times \alpha \quad (7)$$

where Std indicates the standard deviation filter application.

Since the standard deviation represents the amount of change based on the average, it can also be used for edge detection. Based on (6) and (7), we propose an augmented IR intensity that can detect targets well, taking into account IR intensity and standard deviation. The combination of IR intensity and standard deviation represents a high IR value and a high edge value and, thus, may be used to detect pixels of a target. This is called augmented intensity (AI)

$$\text{AI}(x, y) = I(x, y) * \text{Std}(x, y) \quad (8)$$

where  $*$  is the operator of multiplication between matrix elements.

Fig. 5 compares three different maps: the AI map, the standard deviation map, and the original intensity map. The original intensity map contains high IR values, but the distribution may be wide. Since the standard deviation map shows an area with a big change in IR values, it is detected mainly on the boundary line. The proposed AI map simultaneously shows high IR values and pixels with high variation in IR values. Therefore, the AI map can select the characteristics of the target in a very short time in the entire image.

Utilizing the second-order differential operator [25], which is widely used for boundary detection, has the advantage of enabling clear boundary detection but has the disadvantage of being vulnerable to noise. Accordingly, a large number of pixels with noise may be sporadically detected with respect to a general IR image having poor image quality. When LoG [26] is used, the Gaussian filter is utilized to smooth the IR intensity to alleviate the noise component. This has the effect of lowering the contrast of the image, and there is a



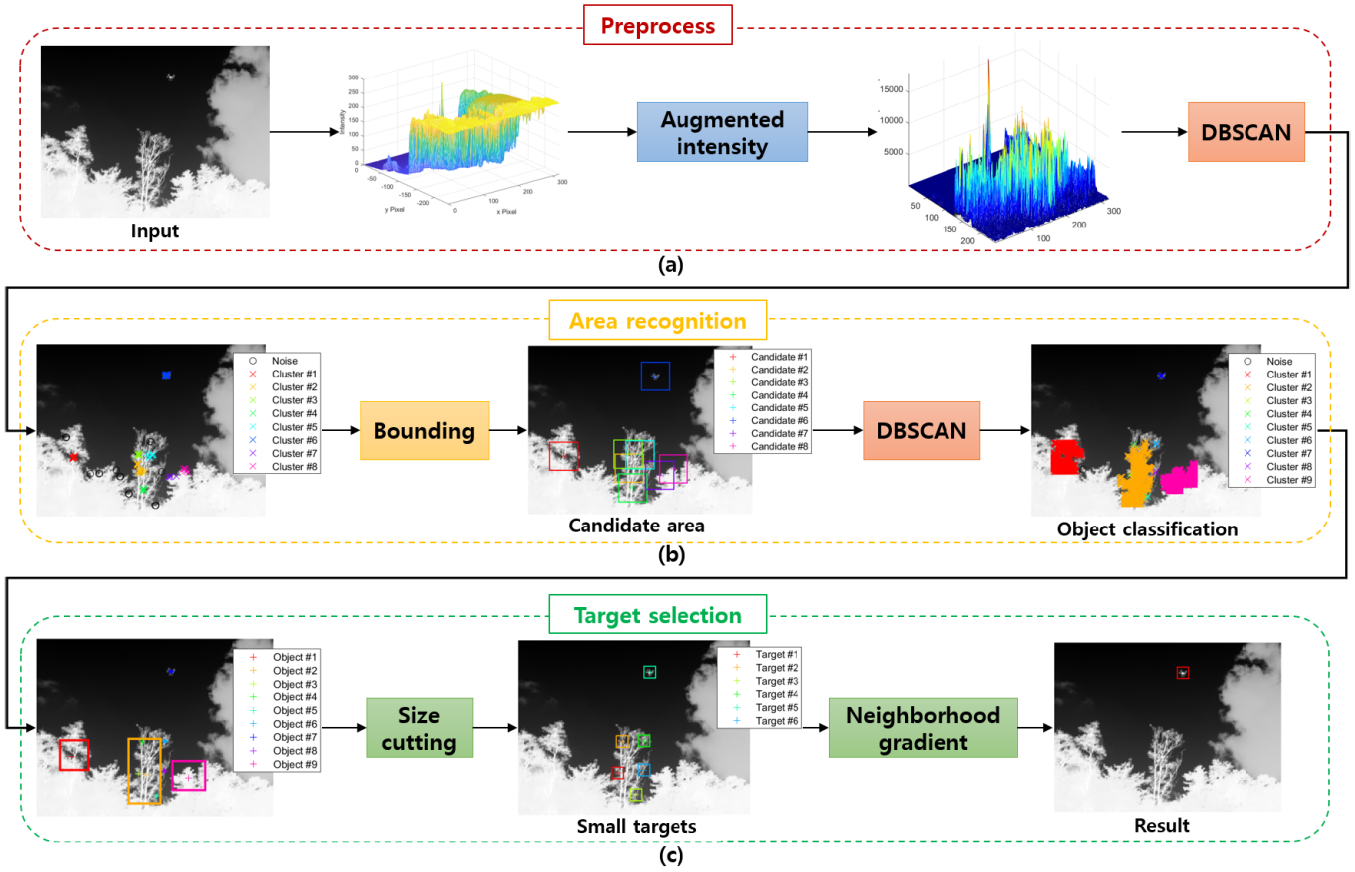


Fig. 3. Overview of the proposed method. (a) Preprocessing stage. (b) Area recognition stage. (c) Target selection stage.

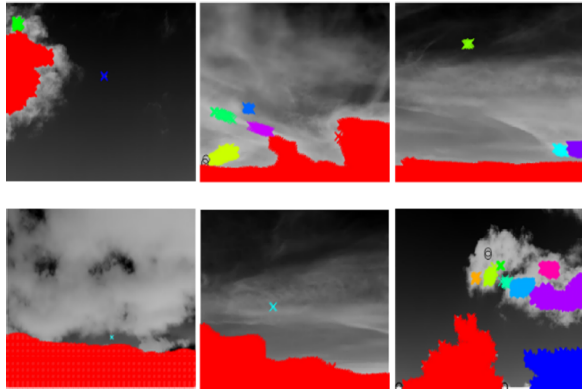


Fig. 4. Candidating area detection using the threshold of IR intensity.

concern that the IR intensity value of the small target may be lost. Therefore, by selecting a candidating area using the amount of change in the standard deviation, the area where the target may exist is detected.

Since objects have similar temperatures depending on their shape, the clustering of high-intensity pixels makes them classifiable. The distance between pixels in the shape of an object is very close, so distance-based clustering may be used, but the shape is irregular because the boundary is detected through the amount of change in the standard deviation. Therefore, density-based clustering is more suitable than distance-based clustering. If there are more than two points within the

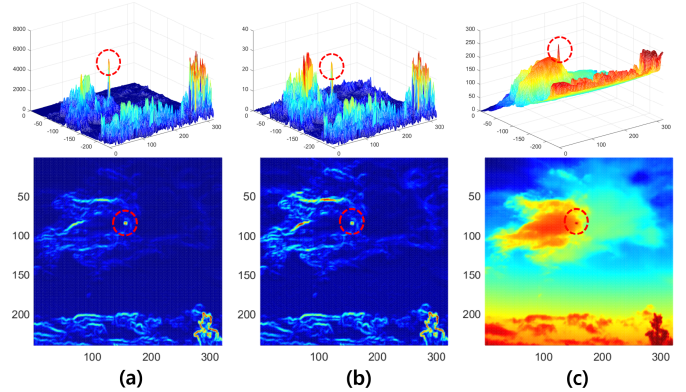


Fig. 5. Comparison maps. (a) AI map. (b) Standard deviation map. (c) Original intensity map.

distance  $\epsilon$  between the pixels, it is set as one cluster. Having only one pixel or data point severely limits the ability to differentiate between noise and a target signal, as it lacks the necessary information for proper analysis. Therefore, it is set to detect a target whose pixel size is at least  $2 \times 1$  or more

$$C_\epsilon(p) = \{q \in D | \text{dist}(p, q) \leq \epsilon\}, \text{ where } p, q = \text{AI}(x, y). \quad (9)$$

Fig. 6 shows the results of detecting pixels through the amount of the AI and performing density-based clustering of the detected several pixels. The rectangular red box shows the part with a large amount of augmented IR intensity.

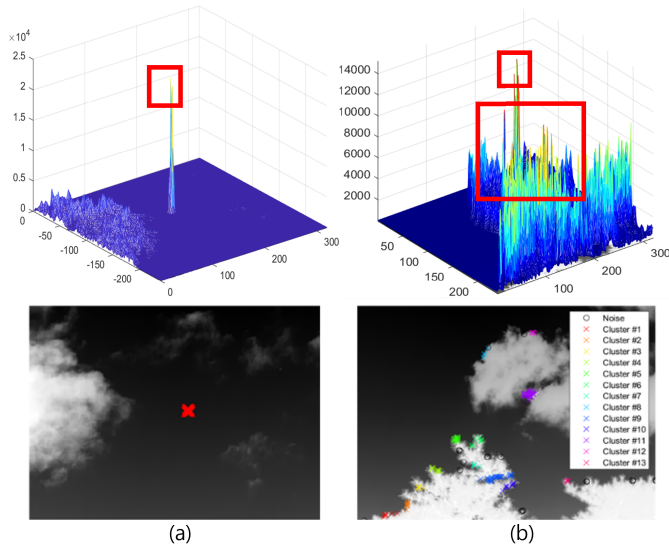


Fig. 6. Candidating area detection using AI and density-based clustering. (a) One target with a simple background. (b) One target with complex background.

Fig. 6(a) has a target and clouds, but the boundary of the cloud is not clear, so the cloud is not detected. In Fig. 6(b), one target, clouds, and trees have a clear boundary. Trees are detected in the AI map due to high-intensity values and clear boundaries. The same color means the same cluster. In addition, Fig. 6(a) and (b) can be seen that pixels detected in a narrow area form one clustering through density-based clustering. That is, a smaller amount of pixels is detected in the candidating area of Fig. 6 compared to Fig. 4. We can predict that the target is likely to exist among the detected candidating regions. Candidating areas without targets become false alarms.

### C. Area Recognition: Object Classification

Based on the pixel's location, the candidating area's center point can be obtained because it is classified as an object through density-based clustering. A bounding box is needed around the candidating area to find the exact center point according to the object's shape. Since the candidating areas were represented through the detection of the AI, it is not the exact location of the objects. In Fig. 7, the true target is located in candidate #11 (purple rectangular). In the upper left corner of Fig. 7, as the target is enlarged, it can be seen that the central point is caught at the boundary of the object.

Considering that the bounding box is for a small target, we chose 15% of the total size of the image that was selected. In order to find a pixel indicating an object in a plurality of candidating areas, a pixel having an original IR intensity value higher than a threshold value is rediscovered in the bounding box

$$I(x', y') > \text{Max}\{I(0 < i < w_B^i, 0 < j < h_B^i)\} \times \alpha \quad (10)$$

where  $w_B^i$  and  $h_B^i$  are width and height of the  $i$ th bounding box, respectively.

There may be places where candidating areas overlap, such as from candidate #1 to candidate #7 and from candidate #9

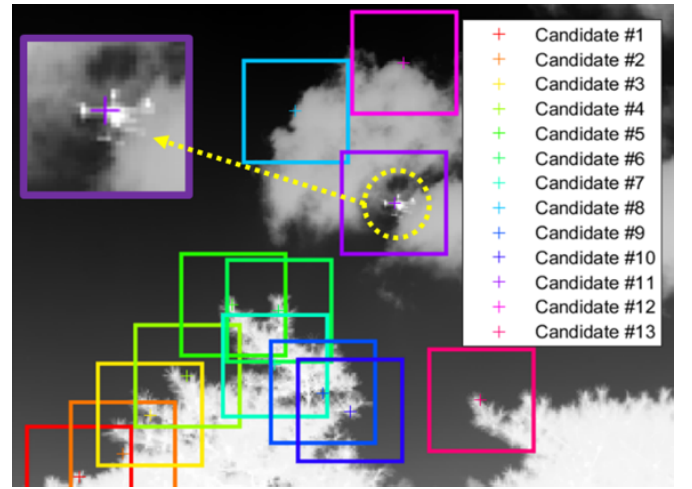


Fig. 7. Creating a bounding box to find the centroid of an object.

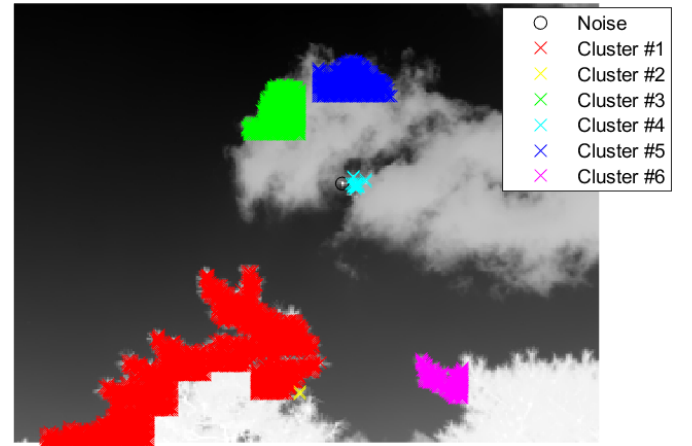


Fig. 8. Pixel classification for searching an object.

to candidate #10 in Fig. 7. The intersection over union (IOU) between boundary boxes is calculated. It is integrated into one candidating area and designated if it is a high value

$$\text{IOU} = \frac{\text{Area of intersection}}{\text{Area of union}}. \quad (11)$$

Since multiple objects may exist within one candidating area, density-based clustering is performed again for each candidating area to classify objects from the background

$$O_\varepsilon(p) = \{q \in D | \text{dist}(p, q) \leq \varepsilon\}, \text{ where } p, q = I(x', y'). \quad (12)$$

Fig. 8 shows the clustering of pixels with high IR intensity in the candidating area. Since a pixel with a high IR intensity value is detected near the boundary, it may be confirmed that no unnecessary pixel is detected compared to Fig. 4. In Fig. 8, Cluster #4 is the true target. Cluster #1 is wide because it is detected in the integrated candidating area from Fig. 7. Other clusters are not bigger than the bounding box size.

### D. Target Selection: Eliminating False Alarm

Since all clustered pixels are known, a center point may be calculated for each cluster. The same clustering points' mean

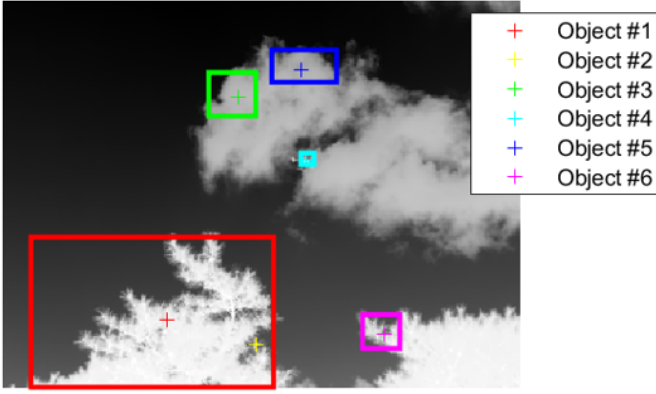


Fig. 9. Object detection with identified size and center point.

position can obtain the object's center position. In addition, the size of each cluster can be expressed, as shown in Fig. 9. Due to the IOU, it can be confirmed that the integrated object has a larger size than the boundary box, and the other object has a smaller size. This enables the recognition and classification of meaningful objects in the IR image.

Among the detected objects, we are interested in a small target. Although the criteria for small size are different, extracting features of objects less than  $9 \times 9$  pixels in the image [27] is generally difficult. Therefore, in this article, objects larger than 12 pixels are excluded by heuristically giving a margin of 3 pixels to the object size of 9 pixels

$$O_\varepsilon(p) = \begin{cases} \text{width}(O_\varepsilon(p)) \leq 12 \\ \text{height}(O_\varepsilon(p)) \leq 12. \end{cases} \quad (13)$$

In Fig. 9, object #1, object #3, and object #5 are large in size, so they are excluded from the target, and only three objects, object #2, object #4, and object #6, are selected. A target and portions of objects with a large IR intensity may be selected among the three objects. Therefore, we try to filter out the small target by generating a layered window of the size defined in (13). Small targets are distinguished from objects because they have a stronger IR intensity than the background. An object with a size larger than a layered window may have a high intensity similar to that of its surroundings.

To distinguish between targets and objects, we designed a new layered window where the central cell captures the entire target, and the peripheral cell captures the background (see Fig. 10). The layered window does not slide in the image. The maximum size of the target cell ( $T$ ) must be set to a minimum of  $12 \times 12$  pixels to capture the entire target. One cell comprises  $3 \times 3$  pixels, and the target cell is divided into 16 cells in Fig. 10. Background cell ( $B_i$ ) has the same configuration as the target cell.  $B_i$  around  $T$  is the neighborhood background with a total of eight cells. We can measure IR contrast by making the most of the differences between the two regions. The minimum gradient between the target cell and the background cell can be used as a definition of contrast

$$G_{TB} = \begin{cases} m_T - \max(m_{B_i}), & m_T > \max(m_{B_i}) \\ 0, & \text{else} \end{cases} \quad (14)$$

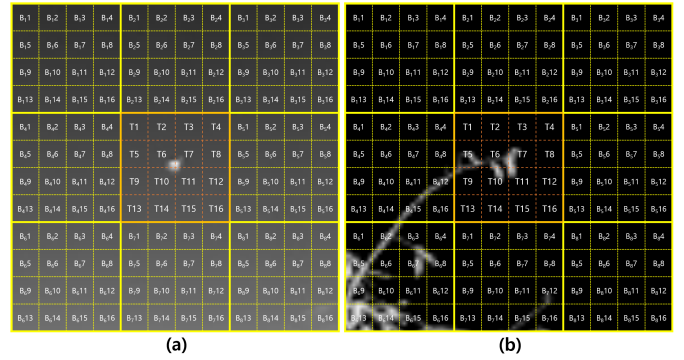


Fig. 10. Illustration of the window. (a) Window detects the small target. (b) Window detects an object.

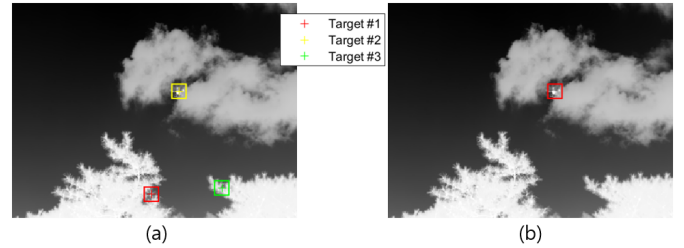


Fig. 11. Result of target detection. (a) Eliminating big size. (b) Eliminating big size and zero gradients.

where  $m_T$  and  $m_{B_i}$  are mean IR value of the target cell and the  $i$ th background cell, respectively. Use the mean IR difference to calculate the gradient, which can effectively distinguish between the target and the object.

Fig. 10(a) shows that the target cell has a target, so the value of  $G_{TB}$  has a positive value. Fig. 10(b) shows an object in the target cell, but, since a part of the object is captured in the background cell, the value of  $G_{TB}$  becomes zero.

Ultimately, it can be seen that candidating targets excluding large objects are remained, as shown in Fig. 11(a), and Fig. 11(b) shows that the final target is detected as a result of excluding targets with zero value of  $G_{TB}$  in the layered window.

#### IV. DATASET AND EVALUATION ENVIRONMENT

In this section, we introduce our single-frame dataset and other public datasets in Section IV-A. Then, we show baseline methods to compare our proposed method and the evaluation metrics in Sections IV-B and IV-C, respectively.

##### A. Dataset

We test the proposed algorithm on typical IR images under different complex scenarios: five sequence IR datasets and a single-frame IR dataset. We use the self-made small targets' IR dataset for experiments. The thermal imaging camera used to obtain IR images is an E75 model from FLIR Systems manufacturer. This camera can measure temperatures from  $-20^\circ\text{C}$  to  $1000^\circ\text{C}$ , and long-wavelength IR rays with a spectrum range of  $7.5\text{--}14.0\ \mu\text{m}$  are used. The drones used as targets were DJI's model MAVIC and Matrice200. As shown in Fig. 12, drones were sent far away to reduce the size of



---

**Algorithm 1** Small Target Detection of the Proposed Method
 

---

```

input      : Infrared image  $I$ ,
               The number of images  $N$ 
output     : Target position  $T(x, y)$ 
begin
1   for  $i = 1 : N$  do
2     1) Augmented intensity map:
3     Calculate  $Z_i = Std(I_i)$ 
4     Augmented intensity  $AI = I * Z_i$ 
5     Select pixel:
6      $P_i = \text{find}(AI_i(x, y) \geq \text{threshold} \times \max(AI_i))$ 
7     2) DBSCAN:
8      $C(p) = \{P_i | \text{distance}(p, q) \leq \epsilon\}$ 
9     for  $j = 1 : M$  do
10    | Calculated the mean point of  $C_j(p)$ 
11    | Make bounding box
12    end
13    Calculate IOU
14    If  $IOU \geq 0.3$ 
15    | Merge candidating area
16    End if
17    3) Do again DBSCAN to find a center
18    position of objects
19    for  $j=1:M$  do
20    | Select pixel:
21    |  $P'_j = \text{find}(I_i(x', y') \geq \text{threshold} \times \max(I_j))$ 
22    | Clustering:
23    |  $O_j(p) = \{P'_j | \text{distance}(p, q) \leq \epsilon\}$ 
24    end
25    4) Eliminating false alarm
26    for  $k=1:C$  do
27    | Size of  $O_k \leq 12$ 
28    | Apply a layered window
29    | Calculate  $G_{TB}^k$ 
30    |  $T(x, y) = \{T_k(x, y) | G_{TB}^k > 0\}$ 
31    end
32  end
33 end

```

---

the target in the image. The image output from the camera has a size of  $320 \times 240$  pixels, while the target sizes are all under  $12 \times 12$  pixels, and 300 IR images are obtained. We photographed drones under several environmental conditions (e.g., sky, clouds, and ground backgrounds). Dataset 1 is composed of 100 images with a size of  $320 \times 240$  with a light cloudy-sky background. Dataset 2 consists of 100 images with a size of  $320 \times 240$  with a heavy cloudy-sky background. Dataset 3 is made of 599 images with a resolution of  $256 \times 256$  pixels having clear sky background. Dataset 4 is composed of 399 images with a size of  $256 \times 256$  pixels with a ground background. Dataset 5 is made of

100 images with a resolution of  $256 \times 256$  pixels and contains complex clusters of continuous ground background. Table I presents detailed information on these five datasets and our dataset.

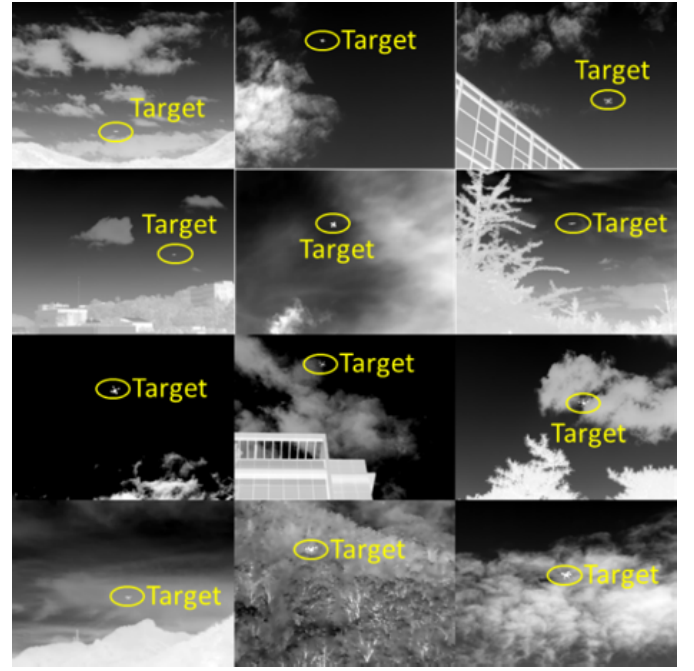


Fig. 12. Samples of dataset 6 (our dataset).

100 images with a resolution of  $256 \times 256$  pixels and contains complex clusters of continuous ground background. Table I presents detailed information on these five datasets and our dataset.

### B. Methods in Comparison and Experiment Setting

To evaluate the superiority of the proposed method, we have compared it with ten IR small target detection methods: local contrast method (LCM) [14], double-neighborhood gradient method (DNGM) [28], dense nested attention network (DNA-Net) [21], maximum mean filter method (MaxMean) [7], maximum median filter method (MaxMedian) [7], accelerated multiscale weighted LCM (AMWLCM) [13], local intensity and gradient (LIG) properties method [29], absolute average gray difference (AAGD) method [30], TLLCM method [16], and local energy factor (LEF) method [31]. The detailed parameter settings of these methods are represented in Table II. All experiments were conducted on MATLAB 2022a with 64 GB of main memory and 3.61 GHz of Intel Core i7-12700K CPU in Windows 11.

*Remark 1:* Among the detection methods used as a comparative group, DNA-Net is a technique for detecting IR small targets using an artificial intelligence network. Since the proposed algorithm is model-driven, comparing it with a data-driven method is somewhat unreasonable. However, since object detection using deep learning has recently shown good performance in the field of imaging, this article used DNA-Net, state-of-the-art technology with the best performance, to compare model- and data-driven.

### C. Evaluation Metrics

Many evaluation indicators have been studied to evaluate the detection performance of IR small target detection. It can



TABLE I  
INFORMATION OF SIX IR SMALL TARGET DATASETS

No.	Frame	Image Resolution	Target number	Target Size	Target Description	Background Description
dataset 1	sequence 100	320×240	1	2×1	an airplane	sky scene with light cloudy.
dataset 2	sequence 100	320×240	1	5×3	an airplane	sky scene with heavy cloudy.
dataset 3	sequence 599	256×256	2	2×2 / 3×3	two airplanes	sky scene.
dataset 4	sequence 399	256×256	1	5×5	an airplane	ground scene.
dataset 5	sequence 100	256×256	1	3×3	an airplane	ground scene.
dataset 6	single 300	320×240	1 or 2	difference	two drones or an airplane	various scene.

TABLE II  
PARAMETER SETTINGS FOR DIFFERENT DETECTORS

No.	Method	Acronym	Parameter setting
1	Local contrast method [14]	LCM	Local window size: 3 × 3, 5 × 5, and 7 × 7
2	Double-neighborhood gradient method [28]	DNGM	Window size: 15 × 15, cell size: 3 × 3
3	Accelerated multi-scale weighted local contrast [13]	AMWLCM	Different scales: 3 × 3, 4 × 4, and 5 × 5, $\beta = 0.8$
4	Absolute average gray difference method [30]	AAGD	Scale: 3, 5, 7, and 9
5	Dense nested attention network [21]	DNA-Net	Learning rate: 0.05, batch size: 16, and epoch size: 1,500
6	Maximum mean filter method [7]	MaxMean	Local window size: 5 × 5
7	Maximum median filter method [7]	MaxMedian	Local window size: 5 × 5
8	Local intensity and gradient properties method [29]	LIG	Local window size: 11 × 11, threshold = 0.2
9	Tri-layer local contrast measure method [16]	TLLCM	Scale: 3, 5, 7, and 9
10	Local energy factor method [31]	LEF	$\alpha = 0.5$ , $h = 0.2$ , scale: 1, 3, 5, 7, and 9
11	Our method	Proposed	$\alpha = 0.65$ , $\beta = 0.9$ , Window size: 12 × 12

be largely classified into a qualitative evaluation method and a quantitative evaluation method. Qualitative evaluation methods based on human vision play an essential role in evaluating the performance of various detection methods. Depending on a number of criteria, such as whether the target was detected correctly and the amount of background clutter, qualitative evaluation methods can be consistently compared mainly by visual detection. Unlike the qualitative evaluation method, the quantitative evaluation approach is not easily influenced by the subjective judgment of the observer.

The most important evaluation metrics for target detection, the detection probability  $P_d$  and the false alarm rate  $F_a$ , are utilized. They are defined as follows:

$$P_d = \frac{\text{number of pixels in the detected targets}}{\text{number of pixels in the actual targets}} \quad (15)$$

$$F_a = \frac{\text{number of pixels in the false targets}}{\text{total number of detected pixels}}. \quad (16)$$

The detected target is considered true if the pixel distance between the center of ground truth and the result is less than 2 pixels (Euclidean distance). Receiver operating characteristic (ROC) curves are widely used as indicators to quantitatively explain the dynamic relationship between false alarm rates  $F_a$  and detection probabilities  $P_d$  [32]. The closer the ROC curve is to the upper left corner, the better the performance of the corresponding detection method. Here,  $P_d$  and  $F_a$  are represented by the ordinate and abscissa of the ROC curve, respectively. Another important indicator is the area under the ROC curve (AUC). In general, the higher the AUC, the better the method.

The other metric is the signal-to-clutter ratio (SCR). It can describe the degree of difficulty of IR small target detection [33]. The higher the SCR, the easier the target is to detect. SCR calculates the neighboring region around the target, which is mathematically defined as follows:

$$\text{SCR} = \frac{|\mu_t - \mu_b|}{\sigma_b} \quad (17)$$

where  $\mu_t$  denotes the average pixel value of the target and  $\sigma_b$ , and  $\mu_b$  represents the standard deviation and the average pixel value, respectively, in the background area. The higher the SCR gain (SCRG), the more information about the target is extracted from the original image, indicating better detection performance. The SCRG is defined as follows:

$$\text{SCRG} = \frac{\text{SCR}_{\text{out}}}{\text{SCR}_{\text{in}}} \quad (18)$$

where  $\text{SCR}_{\text{in}}$  and  $\text{SCR}_{\text{out}}$  are the SCR values of the original IR image and the processed IR image by the detection algorithm, respectively.

IOU is an index for evaluating the accuracy of object location estimation. When pixels of actual objects are called the truth area and pixels of the predicted object are called the predicated area, it is a method of evaluating the size of the area in which the two pixels overlap. The wide area in which the two pixels overlap means that the model well estimates the location of the object. IOU has a value between 0 and 1. In general, if IOU is 0.5 or more, it can be considered to have

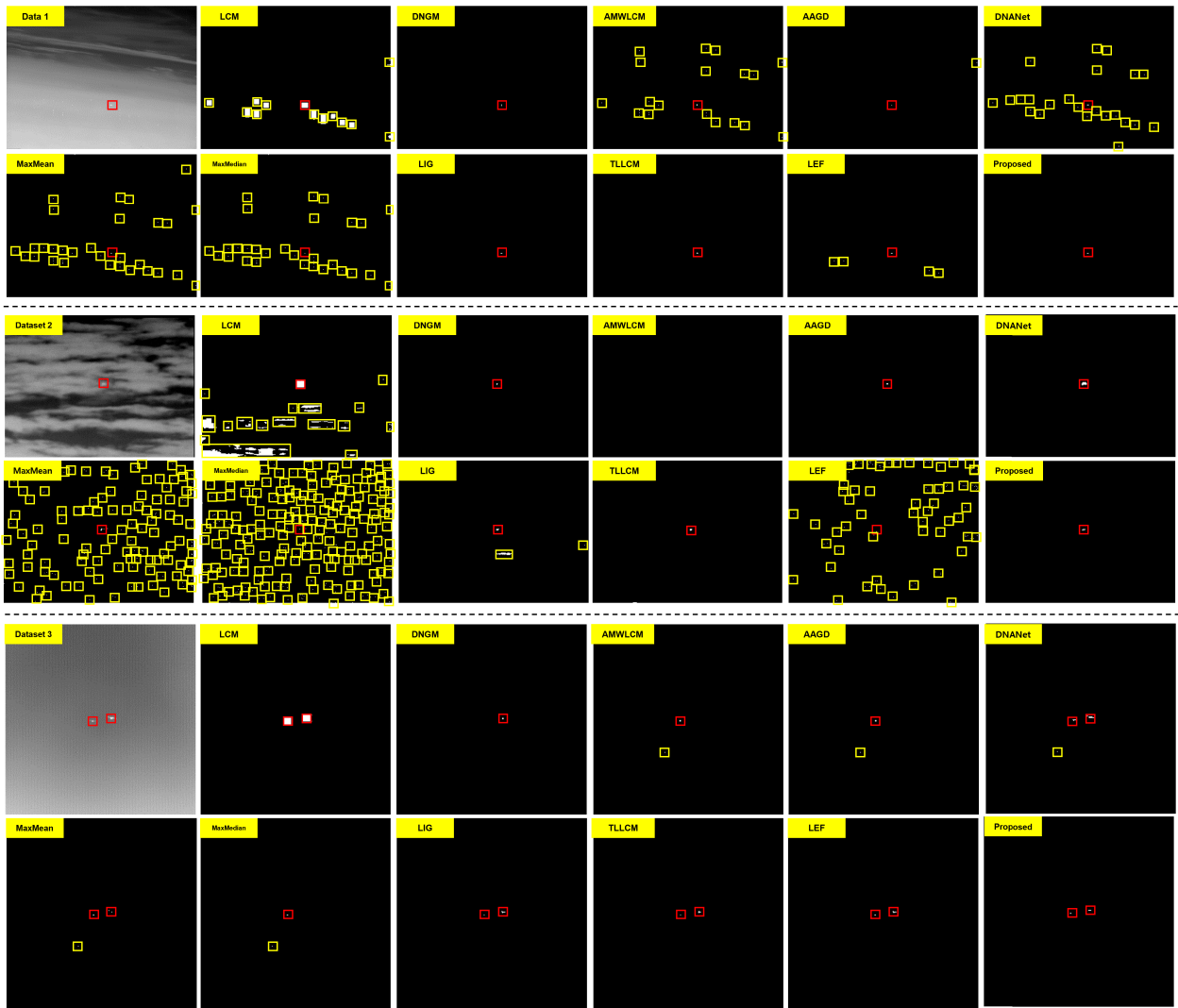


Fig. 13. Target detection results achieved by different methods based on datasets 1–3. Dataset 1 has a target in a lightly cloudy sky scene. Dataset 2 has a target in a cloudy sky scene. Dataset 3 has two targets in a clear sky. Ground-truth targets are shown in red rectangles, and false alarms are shown in yellow rectangles.

been correctly predicted

$$F1 \text{ score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (19)$$

Precision and recall are useful in measuring the performance of the model. The F1 score is an indicator that can explain how effective the model is. The F1 score is the harmonic mean of precision and recall. The formula is given as (19). The harmonic mean is used to ensure that the model's performance is not good by reflecting both indicators in a balanced way when either precision or recall is low.

The last metric to be compared is the computational speed, which compares the time required to detect the target in real time. The time consumed per dataset is employed to describe the computational load.

## V. ANALYSIS OF EXPERIMENTAL RESULTS

In this section, we verify the effectiveness of the proposed method via extensive experiments. Quantitative and qualitative analyses are presented in Sections V-A and V-B, respectively.

### A. Qualitative Analysis

Small target detection results achieved by 11 different methods based on six datasets are shown in Figs. 13–15, which can visually provide a more intuitive comparison. It can be seen that the proposed method provides impressive results in different scenes compared to the baseline methods. The results of the proposed method filter out small clutter and residual noise from various backgrounds, and the target is noticeably detected. This is attributed to density-based clustering and neighborhood gradient in the augmented IR map.

Fig. 13 is based on representative images in three sequence datasets. Six existing methods for datasets 1–3, such as LCM, AMWLCM, DNA-Net, LEF, MaxMean, and MaxMedian methods, are inferior to other methods due to a large number of false alarm detection. In contrast, targets are detected by the proposed method in dense foggy skies or clouded backgrounds without false alarms. It can be seen that the sparse clusters remain in the AMWLCM, LEF, AAGD, and LIG methods through datasets 1 and 2. DNA-Net, MaxMean, MaxMedian, AMWLCM, and LCM methods can achieve good

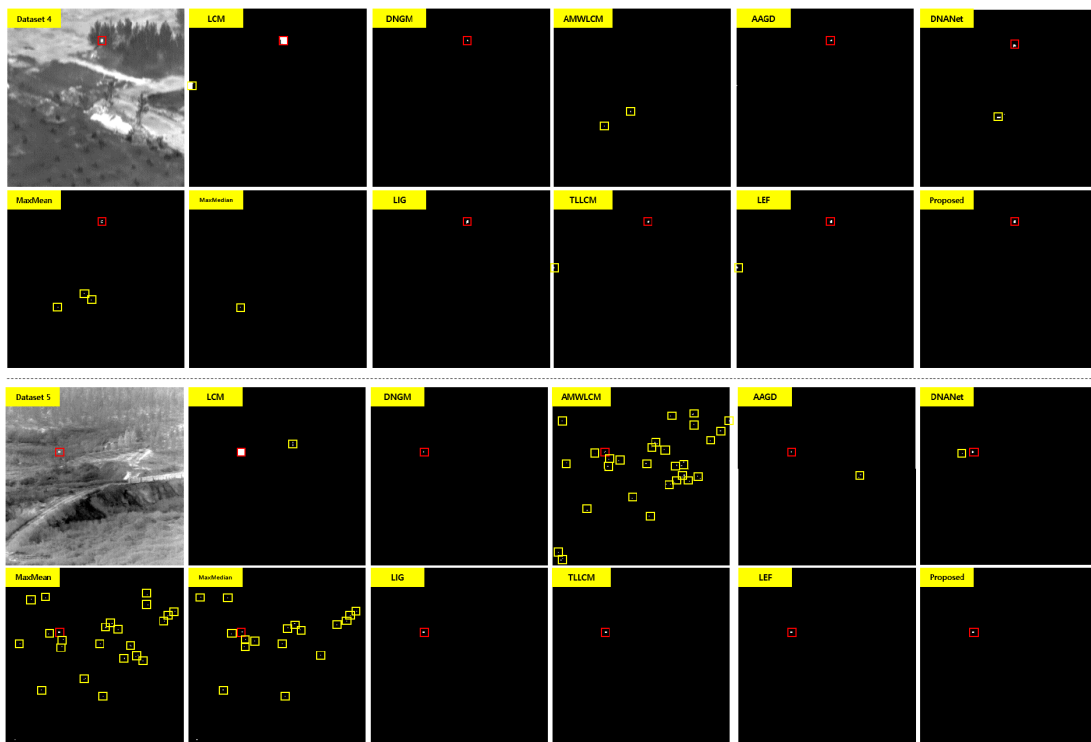


Fig. 14. Target detection results achieved by different methods based on datasets 4 and 5. Target detection results were achieved by different methods based on datasets 4 and 5. Datasets 4 and 5 have one target in the terrain. Ground-truth targets are shown in red rectangles, and false alarms are shown in yellow rectangles.

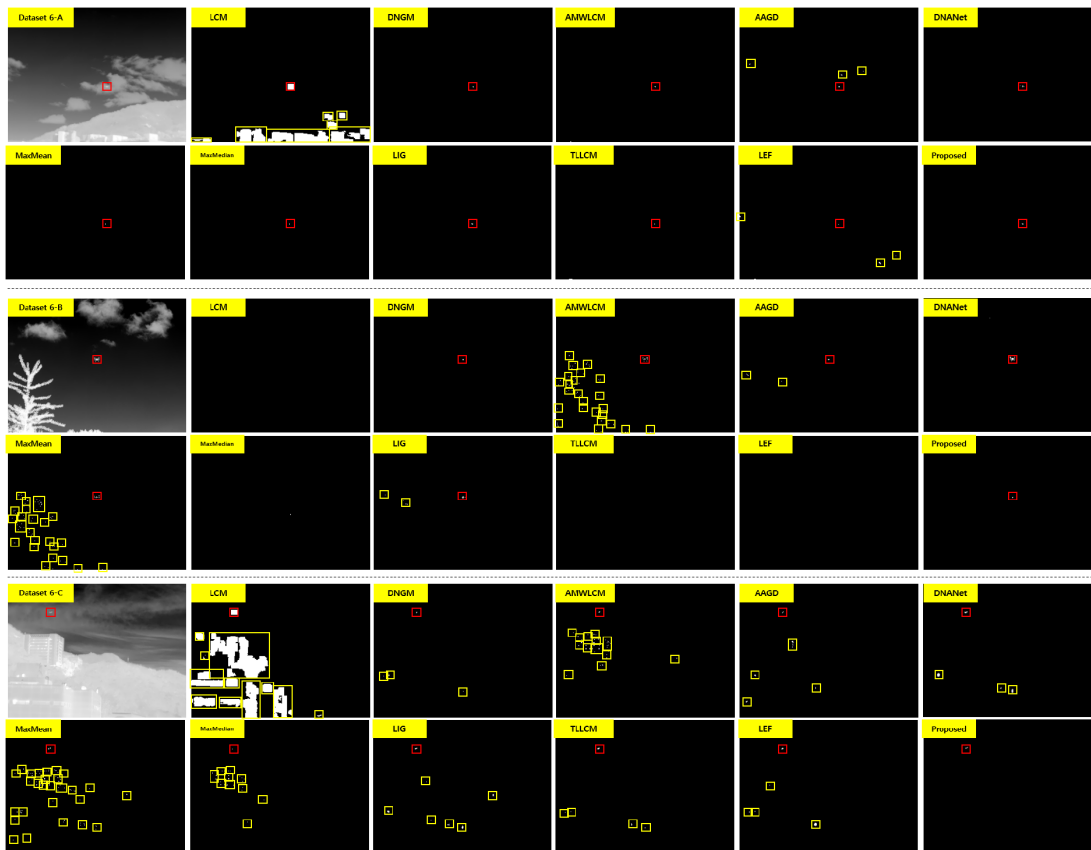


Fig. 15. Target detection results achieved by different methods based on datasets 6-A, B, and C. 6-A has a target in a complex background with sky and mountains, and 6-B has an obstruction at close range. 6-C has a target in the building background. Ground-truth targets are shown in red rectangles, and false alarms are shown in yellow rectangles.

TABLE III  
AVERAGE AOU, SCRG, IOU, F1 SCORE, AND TIME OBTAINED THROUGH DIFFERENT METHODS IN DATASETS 1–6

dataset	Metric	Proposed	LCM	DGRAD	AMWLCM	AAGD	DNA-Net	MaxMean	MaxMedian	LIG	TLLCM	LEF
dataset 1	AUC	0.9773	0	0.9646	0.6695	0.9639	0	0.5620	0.6449	0.9157	0.9898	<b>1</b>
	SCRG	<b>10.609</b>	0.6280	2.4598	0.4729	2.8180	0.5103	0.4772	0.5011	1.4918	<u>3.2128</u>	2.4331
	IOU	0.3688	0.0054	<b>0.6148</b>	0.0818	0.3326	0.0176	0.0511	0.0587	0.3157	<u>0.5870</u>	0.1923
	F1 score	0.4906	0.0104	<b>0.7226</b>	0.1469	0.4674	blue:0.0345	0.0938	0.1088	0.4344	<u>0.7058</u>	0.2978
	Frame rate (Hz)	122.8	41.061	20.053	0.2812	28.932	18.311	<u>152.6</u>	<b>175.16</b>	1.3876	0.6503	0.3037
dataset 2	AUC	<b>1</b>	0	<b>1</b>	0	0.8412	0.5788	0	0.4658	<u>0.9923</u>	<b>1</b>	0
	SCRG	<b>10.787</b>	1.1067	0.5403	2.9424	1.0768	2.0855	0.5448	0.7794	<u>4.5404</u>	0.7395	0.6794
	IOU	0.3486	0.0082	0.3156	0.0002	0.1175	0.1515	0.0049	0.0339	0.0759	<b>0.4115</b>	0.0056
	F1 score	<u>0.4711</u>	0.0162	0.4705	0.0003	0.1812	0.2492	0.0096	0.0642	0.1300	<b>0.5767</b>	0.0109
	Frame rate (Hz)	62.363	26.417	<b>86.229</b>	0.6838	30.889	18.348	<b>86.229</b>	<u>80.366</u>	1.5460	0.9325	0.2668
dataset 3	AUC	<b>0.7577</b>	0	0.4720	0.7199	0.5888	0.5222	0.6134	0.4362	0.4085	0.4852	0.4680
	SCRG	<b>14.033</b>	0.4318	1.7251	1.0684	<u>12.905</u>	0.3128	0.7237	0.5315	1.2200	0.8516	0.5886
	IOU	<b>0.7443</b>	0.0157	0.3507	0.1692	<u>0.5640</u>	0.1547	0.2339	0.3404	0.5318	0.4691	0.4483
	F1 score	<b>0.8267</b>	0.0309	0.5027	0.2822	<u>0.6736</u>	0.2632	0.3481	0.4945	0.6710	0.6199	0.6002
	Frame rate (Hz)	<b>174.63</b>	43.285	34.289	0.9451	59.923	33.578	146.81	<u>148.00</u>	2.3436	1.4382	0.3621
dataset 4	AUC	<b>0.9987</b>	0	<u>0.9737</u>	0.8279	0.9338	0.3818	0.9302	0.7680	0.9087	0.9464	0.5567
	SCRG	<u>6.3622</u>	0.8607	<b>10.175</b>	2.7678	6.3510	0.9378	0.8875	0.7652	5.3899	3.3225	2.3101
	IOU	<b>0.5390</b>	0.0163	0.2551	0.0435	0.3090	0.3932	0.1056	0.3466	<u>0.4944</u>	0.4102	0.4102
	F1 score	<b>0.6308</b>	0.0316	0.3755	0.0746	0.4435	0.5375	0.1395	0.4740	<u>0.6184</u>	0.5347	0.5347
	Frame rate (Hz)	<b>261.96</b>	17.028	17.132	0.9072	22.769	40.553	185.87	<u>191.19</u>	2.2723	0.9391	0.3588
dataset 5	AUC	<b>1</b>	0	0.8638	0	0.3247	0.6327	0.8457	0.4471	<b>1</b>	0.6677	<u>0.9204</u>
	SCRG	0.7691	0.5199	<b>1.9964</b>	0.9048	<u>1.8687</u>	0.4507	0.6909	0.6682	0.6500	0.8299	0.4722
	IOU	<u>0.7072</u>	0.0183	0.5123	0.0439	0.3083	0.532	0.0745	0.2356	0.7043	<b>0.7298</b>	0.6645
	F1 score	<u>0.8748</u>	0.0357	0.8638	0.0782	0.4555	0.6728	0.1220	0.3692	0.8122	<b>0.8232</b>	0.7838
	Frame rate (Hz)	<b>98.106</b>	30.251	21.872	0.8177	41.451	27.855	<u>95.265</u>	89.936	2.0131	1.0855	0.3268
dataset 6	AUC	<b>0.9974</b>	0.7134	<u>0.9902</u>	0.9657	0.9607	0.796	0.8953	0.9221	0.9909	0.9862	0.9948
	SCRG	<b>3.9500</b>	0.6181	<u>2.5684</u>	0.6909	2.6268	0.2554	2.4994	0.4788	1.6063	1.8384	2.2450
	IOU	<u>0.3767</u>	0.1437	0.1329	0.1901	0.1290	0.4298	0.1001	0.2186	<b>0.4367</b>	0.2031	0.3333
	F1 score	<u>0.4960</u>	0.2322	0.2185	0.3053	0.2180	0.5876	0.1690	0.3426	<b>0.5731</b>	0.3017	0.4507
	Frame rate (Hz)	<u>143.84</u>	35.824	29.524	0.7775	39.834	18.666	142.95	<b>150.97</b>	0.9817	1.6512	0.2925

Notes. Bold texts indicate the best results. Underlined texts indicate second-best results.

target detection results in clean sky backgrounds without clutter, as shown in dataset 3 of Fig. 13, but fail to eliminate false alarms because clutter is present in the ground background, as shown in Fig. 14 (datasets 4 and 5). DNGM, LEF, TLLCM, and LIG methods are recently developed detection methods that use sling windows to enhance the IR intensity of targets and suppress the background. They show significantly higher detection performance, similar to the proposed method of detecting targets. DNA-Net is a data-driven small target detection method using dense nested U-net-based deep learning [34]. It has good small target detection performance but cannot eliminate noise. It is shown that false alarms are frequently detected due to the lack of training datasets. Compared with the baseline methods, the proposed method performs better on target detection and filters false alarms for all five datasets. Specifically, the images processed by the proposed method have fewer clusters, and all clusters are removed. As shown in Figs. 13 and 14, our model is suitable for small target detection in complex backgrounds and can also eliminate background clutter.

As described in Section IV-A, Fig. 15 shows three representative images of dataset 6, which have targets against the background of sky, trees, and buildings (dataset 6-A, dataset 6-B, and dataset 6-C, respectively). Almost all baseline methods detect targets well in the sky background [see Fig. 15 (dataset 6-A)]. In Fig. 15 (dataset 6-B), DNGM, DNA-Net, and LIG methods that suppress the background in a complex background with trees have shown the results of eliminating

false alarms. However, in the background of the building in dataset 6-C of Fig. 15, only the proposed method accurately detects the target and eliminates the false alarm. Reflective materials, such as glass windows in structures, can easily cause sunlight to reflect, resulting in heat sources. This is a major cause of false alarms in small target detection. The baseline method suppresses the background, so it is detected by recognizing the false alarm reflected on the structures as a target. Still, the proposed method does not suppress the background and removes the effect of reflection due to the augmented IR map. Therefore, it can be seen that the proposed method is very robust to the surrounding conditions and selects only the characteristics of the target well.

### B. Quantitative Analysis

We quantitatively evaluate the detection performance of the proposed method and other baseline methods using AUC, SCRG, IOU, F1 score, frame rate, and ROC curves on six datasets. We exclude background suppression factor (BSF) performance indicators because the proposed method does not suppress IR values in the background. The average AUC, SCRG, IOU, F1 score, and frame rate results obtained by our method, and the baseline methods are summarized in Table III. The results of the quantitative experiments show that methods have their own advantages and disadvantages in different aspects. In addition, the adaptability of the methods to various scenes may also differ.



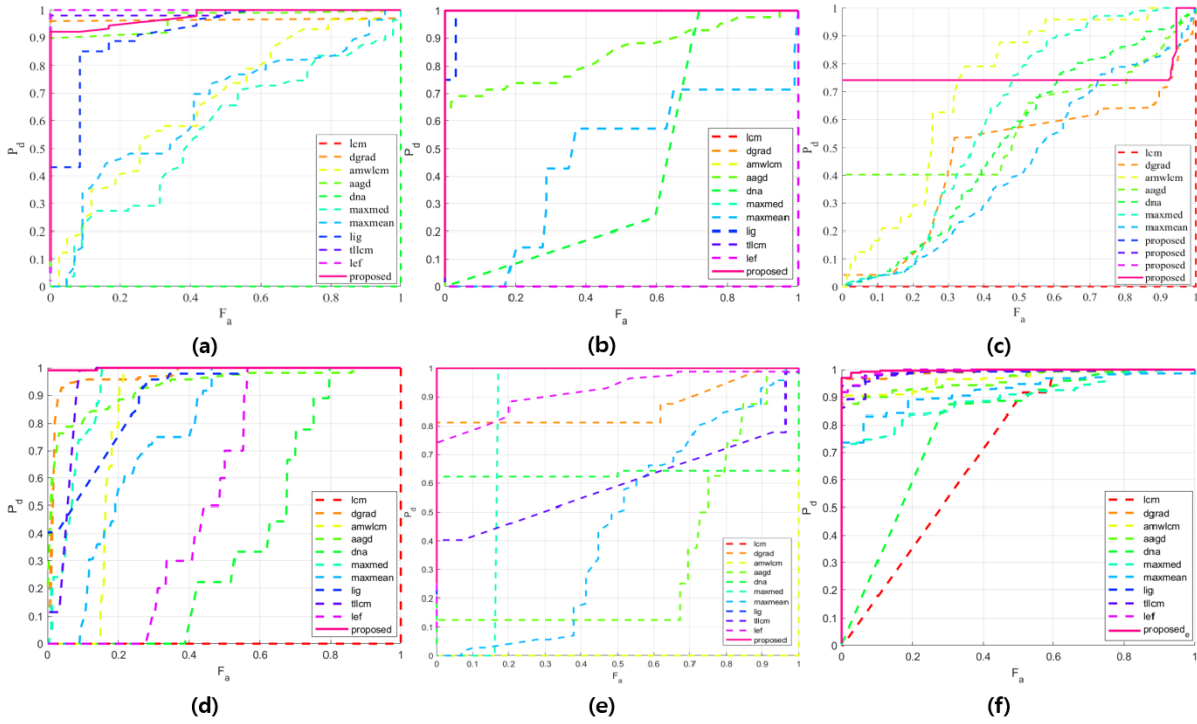


Fig. 16. ROC curve obtained by different methods on (a) dataset 1, (b) dataset 2, (c) dataset 3, (d) dataset 4, (e) dataset 5, and (f) dataset 6.

The proposed method yielded the highest score in terms of SCRG for datasets 1, 2, 3, and 6. In addition, it achieved the second-highest score in dataset 4. With high SCRG, the proposed method is demonstrated to have superior performance in removing false alarms and detecting targets than other baseline methods. Regarding IOU and F1 scores, the proposed method also yielded the highest score in datasets 3 and 4 and the second highest in datasets 2, 5, and 6. The high IOU and F1 score mean that the pixel position of the target detected by the proposed method is very close to the ground truth.

The ROC curve is used to evaluate the performance of the proposed method. ROC is a function of the false alarm ratio  $F_a$  and represents the detection probability  $P_d$ . Fig. 16 shows the result of the ROC curve for six datasets. It can be seen that the ROC curve calculated by the proposed method is close to the upper left corner. That is, the proposed method is superior to other baseline methods in terms of  $F_a$  and  $P_d$ . The index calculating the under area of the ROC curve is AOU in Table III. The proposed method yielded the highest AUC values in datasets 2–6 and the second highest in dataset 1, but close to 1. Thus, the proposed method can obtain the best performance on six datasets, meaning that the proposed method detects targets more robustly against various clutter and noisy backgrounds.

We have run on six datasets with all detection methods by the same computer. The execution time (i.e., frame rate) of each dataset for each method is shown in Table III. The execution time of MaxMean and MaxMedian methods comes out the fastest in some datasets, but it can be seen that the detection performance is very low. In the method of detecting a target based on a sliding window, the calculation time increases proportionally as the resolution of the image

increases. The proposed method does not use a sliding window and is not related to image resolution, so it is possible to find target pixels at a very high speed. Therefore, efficiency is much higher than other baseline methods, and real-time performance is secured.

*Remark 2:* All detection algorithms, except for DNA-Net, were computed using a model-driven method on the CPU. However, because DNA-Net uses a data-driven method based on multiple IR images, it required improvements using PyTorch on a server equipped with a GeForce RTX 3090 Ti of 24 GB memory. Although the different computational processing processes make it difficult to compare frame rates fairly, it should be noted that the trained neural network is much larger than the model-driven source code and requires high-end hardware. Therefore, from an economic perspective, the model-driven method is more cost-effective compared to detection performance.

## VI. CONCLUSION

This article has developed an augmented IR map by combining IR intensity and standard deviation variables and proposed a new, efficient small target detection method through density-based clustering. Unlike previous studies, the proposed method has not used a sliding window, nor does it suppresses background IR intensity values. As a result, the proposed method detects target pixels in real time with less computational burden and demonstrates robust performance even in complex backgrounds. Experimental results on multiple published sequence datasets and self-generated single-frame datasets show that the proposed method can efficiently detect targets on various real-world IR images. Future research will

focus on real-time tracking of targets detected in IR images and information fusion that calculates the location of targets based on IR images measured from various angles for the same target.

## REFERENCES

- [1] 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.
- [2] C. Yang, J. Ma, S. Qi, J. Tian, S. Zheng, and X. Tian, "Directional support value of Gaussian transformation for infrared small target detection," *Appl. Opt.*, vol. 54, no. 9, pp. 2255–2265, Mar. 2015.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] R. Venkateswarlu, "Max-mean and max-median filters for detection of small targets," *Proc. SPIE*, vol. 3809, pp. 74–83, Jul. 1999.
- [8] M. Zeng, J. Li, and Z. Peng, "The design of top-hat morphological filter and application to infrared target detection," *Infr. Phys. Technol.*, vol. 48, no. 1, pp. 67–76, Apr. 2006.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] Y. Shi, Y. Wei, H. Yao, D. Pan, and G. Xiao, "High-boost-based multiscale local contrast measure for infrared small target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 1, pp. 33–37, Jan. 2018.
- [13] J. Liu, Z. He, Z. Chen, and L. Shao, "Tiny and dim infrared target detection based on weighted local contrast," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 11, pp. 1780–1784, Nov. 2018.
- [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. 2013.
- [15] S. Kim and J. Lee, "Scale invariant small target detection by optimizing signal-to-clutter ratio in heterogeneous background for infrared search and track," *Pattern Recognit.*, vol. 45, no. 1, pp. 393–406, Jan. 2012.
- [16] 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.
- [17] Q. Zhu, S. Zhu, G. Liu, and Z. Peng, "Infrared small target detection using local feature-based density peaks searching," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.
- [18] B. Zhao, C. Wang, Q. Fu, and Z. Han, "A novel pattern for infrared small target detection with generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4481–4492, May 2021.
- [19] M. Kristo, M. Ivacic-Kos, and M. Pobar, "Thermal object detection in difficult weather conditions using YOLO," *IEEE Access*, vol. 8, pp. 125459–125476, 2020.
- [20] M. Liu, H.-Y. Du, Y.-J. Zhao, L.-Q. Dong, and M. Hui, "Image small target detection based on deep learning with SNR controlled sample generation," in *Current Trends in Computer Science and Mechanical Automation*, vol. 1. 2018, pp. 211–220.
- [21] B. Li et al., "Dense nested attention network for infrared small target detection," *IEEE Trans. Image Process.*, vol. 32, pp. 1745–1758, 2023.
- [22] M. Ester et al., "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. KDD*, 1996, vol. 96, no. 34, pp. 226–231.
- [23] A. Ahmad and L. Dey, "A K-mean clustering algorithm for mixed numeric and categorical data," *Data Knowl. Eng.*, vol. 63, no. 2, pp. 503–527, Nov. 2007.
- [24] F. Murtagh and P. Contreras, "Algorithms for hierarchical clustering: An overview," *WIREs Data Mining Knowl. Discovery*, vol. 2, no. 1, pp. 86–97, Jan. 2012.
- [25] M. Aubry, S. Paris, S. W. Hasinoff, J. Kautz, and F. Durand, "Fast local Laplacian filters: Theory and applications," *ACM Trans. Graph.*, vol. 33, no. 5, pp. 1–14, Sep. 2014.
- [26] H. Kong, H. C. Akakin, and S. E. Sarma, "A generalized Laplacian of Gaussian filter for blob detection and its applications," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1719–1733, Dec. 2013.
- [27] J. Han, K. Liang, B. Zhou, X. Zhu, J. Zhao, and L. Zhao, "Infrared small target detection utilizing the multiscale relative local contrast measure," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 4, pp. 612–616, Apr. 2018.
- [28] 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.*, vol. 18, no. 8, pp. 1476–1480, Aug. 2021.
- [29] 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.
- [30] S. Aghaziyarati, S. Moradi, and H. Talebi, "Small infrared target detection using absolute average difference weighted by cumulative directional derivatives," *Infr. Phys. Technol.*, vol. 101, pp. 78–87, Sep. 2019.
- [31] C. Xia, X. Li, L. Zhao, and R. Shu, "Infrared small target detection based on multiscale local contrast measure using local energy factor," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 1, pp. 157–161, Jan. 2020.
- [32] J. Gao, Y. Guo, Z. Lin, W. An, and J. Li, "Robust infrared small target detection using multiscale gray and variance difference measures," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 5039–5052, Dec. 2018.
- [33] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Small infrared target detection based on weighted local difference measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4204–4214, Jul. 2016.
- [34] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," in *Proc. Deep Learn. Med. Image Anal. (DLMA) Workshop*. Granada, Spain: Springer, Jul. 2018, pp. 3–11.



**In Ho Lee** received the B.S. and M.S. degrees in aerospace engineering from Korea Aerospace University, Goyang, South Korea, in 2010 and 2015, respectively. He is currently pursuing the Ph.D. degree with the Department of Aerospace Engineering, Seoul National University, Seoul, South Korea.

From 2015 to 2021, he worked at the Hanwha Aerospace Research Center, Seongnam-si, South Korea. His research interests include target tracking, filtering techniques, and vision-aided navigation.



**Chan Gook Park** (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in control and instrumentation engineering from Seoul National University, Seoul, South Korea, in 1985, 1987, and 1993, respectively.

In 1998, he was a Post-Doctoral Fellow with Prof. J. L. Speyer about peak seeking control for a formation flight with the University of California at Los Angeles, Los Angeles, CA, USA. From 1994 to 2003, he was an Associate Professor with Kwangwoon University, Seoul. In 2003,

he joined the faculty of the School of Mechanical and Aerospace Engineering, Seoul National University, where he is currently a Professor. In 2009, he was a Visiting Scholar with the Department of Aerospace Engineering, Georgia Institute of Technology, Atlanta, GA, USA. His research interests include advanced filtering techniques, high-precision inertial navigation systems (INSSs), visual-inertial odometry, INS/GNSS/IBN integration, and smartphone-based/foot-mounted pedestrian dead reckoning systems.

Dr. Park was the Chair of the IEEE AES Korea Chapter until 2009.