Entropy-Driven Morphological Top-hat Transformation for Infrared Small Target Detection

Lizhen Deng, *Member, IEEE,* Guoxia Xu, *Member, IEEE,* Jieke Zhang, and Hu Zhu, *Member, IEEE,*

Abstract—**Infrared small target detection is a key technique in an infrared system. In the past decade, many methods have concentrated on traditional top-hat transformation, which relies on the hand-crafted shape and value of structural elements. However, these methods are inevitably challenged by two aspects: 1) The structural elements cannot suppress heavy clutter because the construction of structural elements is always according to the prior information of the target and unable to consider the feature of clutter. 2) Adaptively extracting sufficient local feature information for background suppression is hard for the structural element. In this paper, we propose an entropy-driven top-hat transformation with guided filter kernel (EGFK top-hat transformation) for considering the features of both the clutters and background. First, we propose an entropy-driven top-hat transformation method with our proposed local mean entropy, which can be used to suppress clutter according to the local complex degree of clutter. Then, an adaptive structural element based on a guided filter kernel is further exploited to capture the local feature information of image for background suppression. Finally, an adaptive threshold is combined with our algorithm to achieve target detection in image sequences. The experimental results show that the proposed algorithm is not only robust for suppressing different kinds of backgrounds but can also obtain a higher value of the signal to clutter ratio gain (SCRG) and detection accuracy compared with some popular traditional baseline methods and related top-hat methods.**

Index Terms—**Entropy-Driven Morphological Method, Top-Hat Transformation, Local Mean Entropy, Guided Filter, Infrared Small Target Detection.**

I. INTRODUCTION

T he detection and tracking of small infrared targets is a crucial technique of infrared image systems and is widely used in surveillance [1], reconnaissance [2], early warning [3]–[5]. However, due to optics point spread function of thermal imaging systems and the long-distance imagery, the targets always appear as small and dim shapes with few texture features. In addition, the heavy clutter enhances the complexity of the background and targets always suffer from a low signal-to-clutter ratio (SCRG) [6]. As a consequence, small target detection under low quality images remains an open problem.

In the recent decades, various research works have been developed to detect infrared small targets [7]–[10] effectively.

Lizhen Deng is with National Engineering Research Center of Communication and Network Technology, Nanjing University of Posts and Telecommunications, Nanjing, 210003, China. Guoxia Xu is with Department of Computer Science, Norwegian University of Science and Technology, 2815 Gjovik, Norway. Hu Zhu and Jieke Zhang are with Jiangsu Province Key Lab on Image Processing and Image Communication, Nanjing University of Posts and Telecommunications, Nanjing 210003, China.

Since the targets are usually buried in a complex background with clutter, it is hard to track or detect targets directly [11]–[13]. Therefore, one popular method of detection is to enhance the targets by predicting the background and subtracting it from the original image. Some methods based on learning prototypes can effectively predict the background by introducing neural networks [14], SVM [15], and manifold learning [16]. Nevertheless, the mass training set is not usually easy to construct, and the training procedure requires too much time. In recent years, the trend of using deep learning techniques combined with traditional image algorithms has become more and more obvious. Deep learning technology has great advantages in feature extraction and learning. Researchers have proposed various target detection models based on deep learning technology, such as the symbiosis filter and the symbiosis filter neural network proposed by Avidan S [17]. By making full use of the spatial information in the image, it can better utilize the spatial and spectral characteristics of infrared image sequences. Guo [18] proposed a multi-scale feature learning model to deal with the target detection problem. He [19] uses domain adaptive technology to accelerate the detection efficiency; Wang [20] proposed to use fast neural architecture search to find the most suitable network model. However, deep learning-based methods have achieved good performance in many fields, but these methods rely on massive amounts of data to a large extent. Under certain conditions, marking infrared data is difficult to obtain and requires a lot of labor costs. Thus, some classic methods without training are more practical, such as top-hat transformation in morphology.

The top-hat transformation is widely applied to infrared target detection [21] and other fields of image processing [22]–[24]. Intuitively, the closing operation in the top-hat transformation is taken to reserve the high-frequency component for background prediction. However, small targets and noise are also left in the residual plot. In order to enhance the performance of top-hat transformation against noise and clutter, some methods based on modified top-hat transformations have also been proposed. In [25], a novel multiscale center-surround top-hat transformation was proposed, that used two structural elements with different shapes to perceive difference information between targets and surrounding regions. Similarly, Bai [26] utilized this difference information through hit-or-miss transformation and suppresses the false alarms effectively. However, clutter will still decreases the efficiency of hit-or-miss transformation. To overcome this problem, a toggle contrast operator combined with top-hat transformation [27] was further proposed to extract the feature of the image for target detection. The above methods can effectively improve

the quality of small target enhancement. Unfortunately, Guo [28] proved that it is significant to select structural elements in the discriminability of the detection model. The value of structural elements should be set initially to accomplish the whole small target detection.

The structural element plays an important role in top-hat transformation. In general, the selection of structural elements depends on the target and the construction of structural elements that adaptively reflect the different components of the background [28]. However, prior information on targets cannot be obtained easily when targets are buried in complex backgrounds [29]. Meanwhile, the complexity and uncertainty of the quality of image signals are always extremely unbalanced and immeasurable. Therefore, the selective structural elements should be adaptive for input infrared images. There are also some works [30], [31] that focused on the construction of adaptive structural elements in mathematical morphology. These structural elements can capture the different feature information of an image, such as local orientation information [31], local contrast information and spatial-temporal information [30]. However, they cannot be used in small target detection due to the following reasons: 1) The construction of the structural elements cannot suppress the heavy clutter because structural elements only rely on the prior information of the target and the features of clutter cannot be considered. 2) Their structural elements cannot capture the local feature information under complex background. The above two problems are the hot and difficult points in the field of small target detection. The balance ring top hat transformation method for infrared small target detection in complex background introduces the shape of the balance ring [32], [33], and USES the contrast information between the target and the surrounding background to enhance the target.

In this paper, we reconstruct an entropy-driven top hat transformation algorithm, which can overcome the above two problems.

(1) In general, the features of clutter extracted from the edge areas (e.g., some undecided areas between sky and cloud, and the edge areas of clouds in the sky) show a similar representative pattern to the target. However, in classic top-hat transformation, the erosion operation may smooth the target and some clutter together for background prediction. It is likely that the traditional methods will always cause a large false alarm in prediction. It is noticeable that small targets are usually displayed as small bright areas in infrared images [25]. And the adjacent pixels of a small target are usually dependent on each other. However, the pixels of the clutter are always random. So, the clutters will cause a different complex degree in the local region compared with targets [34]. In order to solve the classic top-hat transformation problem, we exploit the information entropy [35] to balance the contribution of those components with high gray values to the image entropy. In [35] the local entropy was presented to emphasize the importance of the complexity of gray value distribution on an image. Motivated by this, we propose an entropy-driven top-hat transformation. The novelty of this entropy-driven top-hat transformation utilizes a proposed local mean entropy to weight the erosion operation. Furthermore, the enhancement

of the proposed method can make structural element fully consider the features of clutter to overcome the effects of the clutter in the detection model.

(2) To date, guided filters have been successfully used in edge-preserving [36], image enhancement [37] and other related fields in computer vision [38], [39]. We make use of a guided filter kernel (GFK) to construct the structural elements since the guided filter kernel can change the value in the region with large or small variance adaptively [40]. A small target always disconnect with periphery and the appearance of target will cause a large difference in the local area of an image, while the background area is usually uniform and the difference is very small [5]. Therefore, the use of guided filter kernel can capture local features of both the target and background. Specifically, we will obtain the structural elements with different features at different positions. In this paper, we explore the application of GFK in infrared target detection and further illustrate that GFK-based structural elements have special characteristics for enhancing the top-hat transformation. By this means, our top-hat transformation is robust under different backgrounds.

Our paper has the following contributions:

- An entropy-driven top-hat transformation is proposed for suppressing clutters. We construct the model with our proposed local mean entropy to change the degree of erosion operation by utilizing the complex degree information in the local region.
- A guided filter kernel based structural element is proposed to capture the local feature information for background suppression. This structural element is adaptive according to the local feature of images and suppresses the background of targets.
- A novel entropy-driven morphological top-hat transformation with guided filter kernel (EGFK top-hat) is proposed for small target detection. This top-hat transformation can both consider the features of clutters and local feature information, which effectively suppresses clutter and background. Moreover, extensive experimental results demonstrate the obvious advantages over recent methods in terms of several different metrics.

The rest of this paper is organized as follows: In section 2, we review the background of the mathematical morphology operation, guided filter, and entropy. Section 3 describes the proposed method and its detail. Section 4 provides two analyses of the proposed method. Section 5 shows the experimental results and performance analysis based on extensive data. The conclusion is given in Section 6.

II. BACKGROUND

The mathematical morphology and guided filter are both bases of our proposed method. In this section, we introduce the knowledge of morphology involving dilation, erosion, opening, closing and top-hat transformation. Then the guided filter is introduced. Finally, a weighted local entropy is introduced.

A. Mathematical Morphology and Top-hat Transformation

Mathematical morphology has been widely applied in infrared image processing and other fields of computer vision.
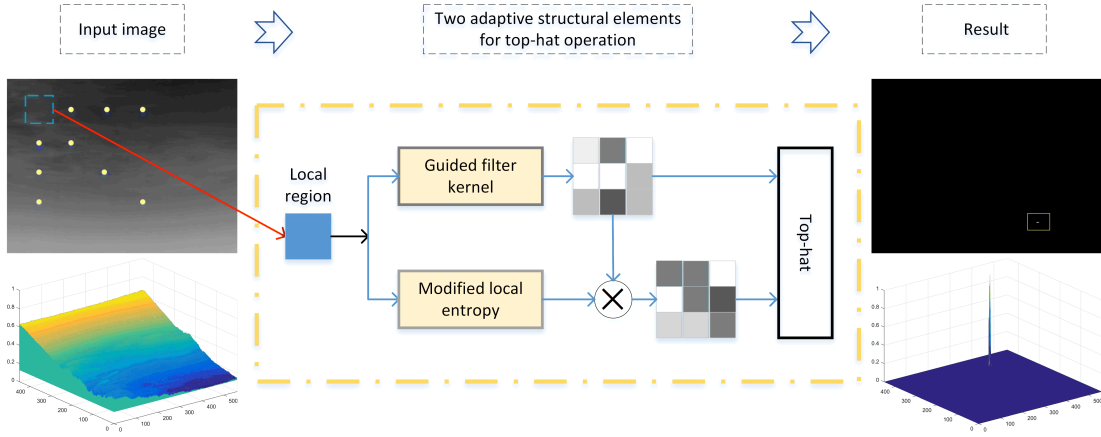
Fig. 1. Our proposed algorithm. The main procedure consists of two parts: the entropy-driven top-hat transformation and adaptive structural elements with a guided filter kernel.

Two basic operations, dilation and erosion, are defined as the gray image $f(x, y)$ and structural elements $b(m, n)$. They are denoted by $(f \oplus b)$ and $(f \ominus b)$ as follows:

$$(f \oplus b)(x, y) = \max_{m,n} (f(x - m, y - n) + b(m, n)) \quad (1)$$

$$(f \ominus b)(x, y) = \min_{m,n} (f(x + m, y + n) - b(m, n)) \quad (2)$$

Where x and y can determine the coordinates of a pixel in the image, m and n are the offsets of the coordinates of a pixel in a structural element with respect to x and y. Due to the maximum operation, the dilation operation makes the gray values of the image greater than that of the input image, while the erosion operation makes the gray values of the image smaller than that of the input image due to the minimum operation. Consequently, dilation increases the size of the bright region and reduces the size of the dark region. The results of erosion are the opposite.

Through these two basic operations, the opening operations $(f \circ b)$ and closing operations $(f \bullet b)$ can be represented by:

$$(f \circ b)(x, y) = (f \ominus b) \oplus b \quad (3)$$

$$(f \bullet b)(x, y) = (f \oplus b) \ominus b \quad (4)$$

The small bright region is removed after the opening operation. Likewise, the closing operation can eliminate the small dark hole.

In general, a small target is always presented as a small bright area. The background is easily obtained after foreground extraction through an opening operation. Subsequently, we can subtract the background from the original image to highlight the target . The above process is also described as top-hat transformation, which is defined according to the opening and closing operations:

$$OTH_{f,b}(x, y) = f(x, y) - (f \circ b)(x, y) \quad (5)$$

$$CTH_{f,b}(x, y) = (f \bullet b)(x, y) - f(x, y) \quad (6)$$

where, $OTH_{f,b}(x, y)$ and $CTH_{f,b}(x, y)$ are called the opening top-hat operation and closing top-hat operation respectively. They are widely used in bright and dark target detection.

B. Guided Filter

The guided filter is a smooth linear translation variable filter for edge -preserving smoothing. For guidance image I, input image I_{in}, and output image I_{out}, the filter procession can be expressed by introducing the filter kernel as follows:

$$\{I_{out}\}_i = \sum_j W_{ij}(I) \{I_{in}\}_j \quad (7)$$

where i and j represent the pixel indexes. If I_{in} and I are the same image, the filter process can smooth the input image and preserve the edge at the same time.

To obtain the expression of kernel W_{ij}, the key assumption the guided filter is the local linear model between output image I_{out} and guidance image I:

$$I_{out} = x_k I_i + y_k, \forall i \in \omega_k \quad (8)$$

where I_{out} is a linear transformation of I in the window ω_k that is centered at pixel k and (x_k, y_k) is the linear coefficient in ω_k. This linear model is also useful for image matting [41] and image super-resolution [42]. The solution of x_k and y_k should minimize the difference between I_{out} and I_{in} and maintain the linear model in Eq. (7), which is defined as following cost function:

$$E(x_k, y_k) = \sum_{i \in \omega_k} \left((x_k I_i + y_k - \{I_{in}\}_i) + \varepsilon x_k^2 \right) \quad (9)$$

ε is an important parameter that can adjust the effect of the filter. Add Eq. (8) to Eq. (9) and let $W_{ij} = \partial \{I_{out}\}_i / \partial \{I_{out}\}_j$, parameter y is removed:

$$W_{ij} = \frac{1}{|\omega|} \sum_{k \in \omega_i} \left(\frac{\partial x_k}{\partial \{I_{in}\}_j}(I_i - \mu_k) + \frac{\partial \{\bar{I}_{out}\}_i}{\partial \{I_{in}\}_j} \right) \quad (10)$$

The process of derivation can be seen in [40]. The filter kernel can be obtained through acquiring the solution of x and y:

$$W_{ij} = \frac{1}{|\omega|^2} \sum_{k \in \omega_i, k \in \omega_j} \left(1 + \frac{(I_i - \mu_k)(I_j - \mu_k)}{\sigma_k^2 + \varepsilon} \right) \quad (11)$$

where μ_k and σ_k denote as the mean and variance respectively of the pixel in the window ω_k. $|\omega|$ is the number of pixels

in ω_k. It can be proven that $\sum_j W_{ij}(I) = 1$, because it dose not normalize the weight. This kernel shows the different properties of the pixel in different regions of the image, and we further explore it and use it to construct the structural elements.

C. Weight Local Entropy

Generally speaking, a small target often appears in the form of a high gray value in the infrared images with the complex background. For the purpose of emphasizing the contribution of high gray value components of the information entropy in an image, a weighted information entropy is defined in [35]:

$$H(x,y) = -\sum_{i=1}^{m} (f_i - f(x,y))^2 \{I_{in}\}_i \log \{I_{in}\}_i \quad (12)$$

The gray value of the original infrared image is $f(x,y)$, and the corresponding neighboring pixel is M, which includes m kinds of gray values $f_1, f_2, ..., f_m$ and is a simple and efficient space domain preprocessing method for infrared small target images. In our paper, we propose an entropy-driven top-hat transformation that uses a novel entropy based on this transformation.

III. OUR ENTROPY-DRIVEN GUIDED MORPHOLOGICAL TOP-HAT TRANSFORMATION FOR TARGET DETECTION

In this section, firstly, an entropy driven top-hat transformation is proposed. Then two adaptive elements based on a guided filter kernel and an adjusted local entropy are used in our proposed top-hat transformation to suppress the background and enhance the target. Finally an entropy driven top-hat transformation with guided filter kernel is given.

A. The Proposed Entropy-Driven Top-hat Transformation with Local Mean Entropy

A small target appears as a small bright area in the image, which causes the discontinuity with surrounding area(as shown in Fig. 2) [43]. For the proposed top-hat transformation, the key operation is closing operation. The closing operation can smooth the bright area of the image, such as the target. If it can remove a target completely from original image, the background prediction will be precise after opening the operation. However, for the case of a complex background, some clutter regions contain clutters that shows a similar representative pattern to target. Thus, these regions will also be removed after the closing operation. In this situation, the clutter is also reserved after subtracting the prediction background from the original image.

Under the ideal case, we hope that our adaptive structural element can remove the targets completely and smooth the clutter as much as possible in the erosion operation. Thus, less clutters will be included in the resulting image. Furthermore, entropy is widely used to represent complex degree. To address this challenging problem for the opening operation, we propose a local mean entropy to weight erosion operation.

The size of the image is $N \times M$, $f(x,y)$ is the gray value at the center pixel point (x,y), and its neighborhood pixels are formulated as follows:
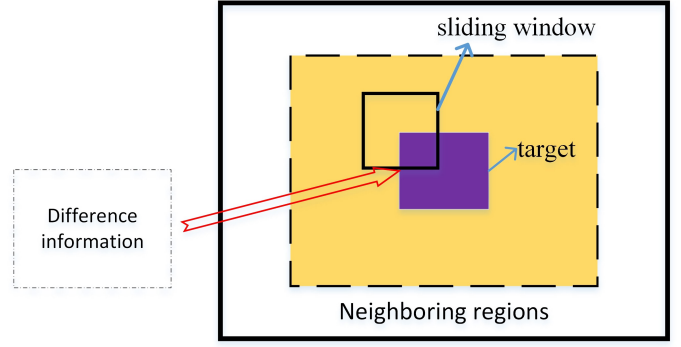

Fig. 2. The discontinuity between target and surrounding area.

$$L = \{(x,y), (I_{in}, I_{out}) \,|\, |x - I_{in}| \le n, |y - I_{out}| \le m\} \quad (13)$$

Here, L is the edge of the target region, (I_{in}, I_{out}) is the surrounding pixel, n and m represent the range of the neighborhood. For a small window with a size of $R \times R$, $f(x,y)$ is also defined as the gray value of the central pixel point (x,y). If the block contains s various of gray values, f_i, $i = 1, 2, ..., s$, then our local mean entropy is defined as follows:

$$S(x,y) = 1 - \sum_{i=1, i \in L}^{s} \left(f_i - \frac{\sum_{i=1}^{R \times R} f_i}{R \times R} \right)^2 \{I_{in}\}_i \log \{I_{in}\}_i \quad (14)$$

where μ_k is mean value of the local region. $\{I_{in}\}_i = \frac{s_i}{R \times R}$ represents the probability density function of the ith gray level. Since a target appears with high gray value and causes a large degree of complexity in local region, the pixel value of the target has a difference from the mean value of the local pixel. If f_i is a pixel of target, $f_i - \frac{1}{R \times R}\sum_{i=1}^{R \times R} f_i$ will large. However, for local region, the pixels are usually random and spare. Then the complex degree is relative smaller if compared with target region. Thus, $f_i - \frac{1}{R \times R}\sum_{i=1}^{R \times R} f_i$ will be relative small. And we will analyze of our proposed local mean entropy later.

Then, we use our proposed local mean entropy to weight the erosion operation in the opening operation. Correspondingly, it is also used for weight dilation operations in the closing operation. The formulation and our entropy-driven top-hat transformation are defined as follows:

$$E - OTH_{f,S}(x,y) = f - (f \ominus b \cdot S) \oplus b \quad (15)$$

$$E - CTH_{f,S}(x,y) = (f \oplus b) \ominus b \cdot S - f \quad (16)$$

where E is a function.

B. Adaptive Structural Elements Based on Guided Filter Kernel

A small target is normally has difference with the surrounding background areas. Consequently, the opening operation mainly utilizes this property to remove the target. This is the main reason why the top-hat transformation is well suitable for small target detection models. However, in classical top-hat

transformation, the value of the structural elements should be initialized before filtering operation. In other words, the value of a structural element is a constant that is short of properly handling the local features efficiently and effectively. Here, we propose an adaptive structural element inspired by [40]

Here, we adopt the guided filter kernel for designing the adaptive structural elements. The key term of the guided filter kernel in Eq. (11) is presented as follows:

$$G = 1 + \frac{(I_i - \mu_k)(I_j - \mu_k)}{\sigma_k^2 + \varepsilon} \qquad (17)$$

For a sliding window, I_i and I_j represent two different pixels. The value of the G varies according to the location of the window in the image.

Case 1: The window centers around the edge in the image. The terms $I_i - \mu_k$ and $I_j - \mu_k$ have the same sign $(+/-)$ if the two pixels I_i and I_j are on the same side of edge. If two pixels are on the different, then they have opposite signs. In this way, G will become smaller (even close to 0) if two pixels are on the different sides of an edge.

Case 2: If the window is located in a flat area with slight change of images, $\sigma_k^2 \ll \varepsilon$, $I_i - \mu_k \approx 0$, and $I_j - \mu_k \approx 0$. Then $G \approx 1$.

These above two cases show that the value of G will be close to the value of different types of areas in the image.

In a local region of the image, the background can be considered a small flat region with similar pixels. In this way, the pixel corresponds to the situation in case 2. The pixels of a small target in case 1 always show the discontinuity with the surrounding area.

Aiming to remove these two aspects, here, we introduce G in Eq. (17) to construct the adaptive structural elements. For a certain local window ω_k, we define $I_{x,y}$ as a pixel. Then our structural function can be represented as follows:

$$b_{R,\varepsilon}(x,y) = \frac{1}{R^2}\left(1 + \frac{(I_{x,y} - \mu_k)(I_o - \mu_k)}{\sigma_k^2 + \varepsilon}\right)_{I_{x,y}, I_o \in \omega_k} \qquad (18)$$

where R represents the size of the structural element and local window, and R^2 is used to normalize the value in a structural element. I_o is the center pixel of the local window, μ_k and σ_k^2 denote the mean and variance respectively. The output $b_{R,\varepsilon}(x,y)$ is the value of a pixel in the structural element with the corresponding index.

In this way, an adaptive structural element is obtained at each position along with a sliding window. Since the top-hat transformation consists of dilation and erosion, the structural function of these two operations is used twice in a top-hat transformation for target detection.

C. Target Detection Based on Overall Model

Our top-hat transformation is based on guided filter kernel in Eq. (18) and local mean entropy in Eq. (14). The formulation is defined as follows:

$$EGFK-OTH_{f,b_{R,\varepsilon},S}(x,y) = f - (f \ominus b_{R,\varepsilon} \cdot S) \oplus b_{R,\varepsilon} \quad (19)$$

$$EGFK-CTH_{f,b_{R,\varepsilon},S}(x,y) = (f \oplus b_{R,\varepsilon} \cdot S) \ominus b_{R,\varepsilon} - f \quad (20)$$

where the $b_{R,\varepsilon}$ is our adaptive structural element. The value of this structural element is adaptive on the position of the pixel in the image and can capture the local feature information. The S represents our proposed local mean entropy, which is used for clutter suppression. We provide further detailed information and analysis in next section.

In summary, a small infrared target and its surrounding regions is discontinuous, and it causes the large entropy in local areas. Our EGFK top-hat transformation mainly utilizes these two features to enhance the target while suppressing the clutters. The following threshold is defined to subdivide the result:

$$T = \overline{I}_o + K\sigma_{I_i} \qquad (21)$$

For output image I_o, $\overline{I}_o$ and σ_{I_i} are the mean and standard deviation, respectively, of I_o. K is a parameter that ranges from 5 to 10. Correspondingly, the EGFK top-hat based infrared small target detection method is described in Algorithm 1.

Algorithm 1   EGFK top-hat based infrared small target detection method

---

1: **Input:** Frame image f.
2: **Output:** Result image f_T and target position.
3: **Initialization:** initial parameter R and ε to calculate structural elements and local mean entropy.
4:     **for** $x = 1$ to M
5:         **for** $y = 1$ to N
6:             (1). Construct the adaptive structural elements $b_{R,\varepsilon}$ by Eq.(18).
7:             (2). Get the local mean entropy $S(x,y)$ by Eq.(14).
8:             (3). Implement the erosion operation $f_E(x,y) = (f \ominus b_{R,\varepsilon} \cdot S) \oplus b_{R,\varepsilon}$
9:         **end**
10:    **end**
11:    **for** $x = 1$ to M
12:        **for** $y = 1$ to N
13:            (1). Construct the adaptive structural elements $b_{R,\varepsilon}$ for image f_E by Eq.(18).
14:            (2). Implement the dilation operation $f_D(x,y) = f_E \oplus b_{R,\varepsilon}$
15:        **end**
16:    **end**
17: Obtain the result image f_T through top-hat transformation: $f_T = f - f_D$
18: Calculate the threshold: $T = \overline{I}_i + K\sigma_{I_i}$
19: Segment target from the result image f_T according to T:the pixel at the location (x,y) is the target if $f_T \geq T$, otherwise it is a pixel of background.
20: **Result:** Obtain the target detection.

IV. ANALYSIS OF THE STRUCTURAL ELEMENT AND ENTROPY

A. Analysis of the Adaptive Structural Elements

From the definition in Eq.(18), we can find that each value of a structural element is depend on the mean and pixel

value in the local region. Here, we take the opening top-hat transformation for example to analyze that our adaptive structural elements have different shapes at the target and background regions. We explain how this property will be helpful for background prediction in top-hat transformation.

The opening top-hat transformation takes the erosion operation first. In a small area of background, the pixels can be considered uniformly. Then we will have the following:

$$I_{x,y} \approx \mu_k, I_o \approx \mu_k \Rightarrow \frac{(I_{x,y} - \mu_k)(I_o - \mu_k)}{\sigma_k^2 + \varepsilon} \approx 0 \quad (22)$$

Thus, we obtain the each pixel of a structural element after normalization:

$$b_{R,\varepsilon}(x,y) \approx \frac{1}{R^2} \quad (23)$$

which means that the structural element is approximately flat:

$$b_{R,\varepsilon} \approx \begin{bmatrix} \frac{1}{R^2} & \cdots & \frac{1}{R^2} \\ \vdots & \ddots & \vdots \\ \frac{1}{R^2} & \cdots & \frac{1}{R^2} \end{bmatrix}_{R \times R} \quad (24)$$

Then, we use an erosion operation through this flat structural element, and the pixel value in each local region of the background can be defined as: $\mu_k - \frac{1}{R^2}$.

For the target region, the pixel value shows the discontinuity at the boundary of a target. Here, the center pixel I_o belongs to a target and a surrounding pixel $I_{x,y}$ belongs to background. Then we have the following:

$$\mu_k > I_{x,y}, I_o > \mu_k \Rightarrow \frac{(I_{x,y} - \mu_k)(I_o - \mu_k)}{\sigma_k^2 + \varepsilon} < 0 \quad (25)$$

In this way, if ε is small enough, it may result in

$$(1 + \frac{(I_{x,y} - \mu_k)(I_o - \mu_k)}{\sigma_k^2 + \varepsilon}) \approx 0 \quad (26)$$

Thus:

$$b_{R,\varepsilon}(x,y) \approx 0 \quad (27)$$

In contrast, if I_o and $I_{x,y}$ are both belonging to the target, we have:

$$\frac{(I_{x,y} - \mu_k)(I_o - \mu_k)}{\sigma_k^2 + \varepsilon} > 0 \quad (28)$$

In this way, for the target region, we obtain a sharp structural element as follows:

$$b_{R,\varepsilon} \approx \begin{bmatrix} 0 & & \cdots & & 0 \\ \vdots & t_{x,y} & t_{x,y} & & \vdots \\ & & t_{x,y} & & \\ 0 & & \cdots & & 0 \end{bmatrix} \quad (29)$$

where

$$t_{x,y} = \frac{1}{R^2}\left(1 + \frac{(I_{x,y} - \mu_k)(I_o - \mu_k)}{\sigma_k^2 + \varepsilon}\right) \quad (30)$$

The location of $t_{x,y}$ is the corresponding location of target pixel. Then, the pixel value of target is formed after the erosion operation, which is defined as follows:

$$\min(I_o - t_{x,y}, I_{x,y} - b_{R,\varepsilon}(x,y)) \quad (31)$$

Thus, a target can be removed and a flat background is obtained. Consequently, after the erosion operation, the input image is relatively flat for the local region. We use the structural function Eq. (18) again for the dilation operation and obtain the flat structural element for each local region. Finally, the pixel value in each local region could is formed as $\mu_k - \frac{1}{R^2}$ to recover the background.

B. Analysis of Our Proposed Local Mean Entropy

The local entropy refers to the variance in gray values of a local region. The homogeneous areas always show a small entropy and the heterogeneous areas are opposite. A small target usually possesses gray value information in a local area, where the local entropy of the target region is larger than that of the background areas. Let (x_t, y_t) and (x_B, y_B) denote as the pixels of target and background respectively. For the target region and background region with clutters, we will have the following equation:

$$\begin{aligned} &\left(-\sum_{i=1}^{s} I_{in\,x_t}(i)\,y_t(i)\log\{I_{in}\}_{x_t}(i)y_t(i)\right) \\ &> \left(-\sum_{i=1}^{s}\{I_{in}\}_{x_B}(i)y_B(i)\log\{I_{in}\}_{x_B}(i)y_B(i)\right) \end{aligned} \quad (32)$$

In addition, the discontinuity between the target and surrounding area is apparently shown. Therefore, we also obtain the following inequality:

$$\left(f_{x_t(i)y_t(i)} - \mu_{k_{x_t,y_t}}\right) > \left(f_{x_B(i)y_B(i)} - \mu_{k_{x_B,y_B}}\right) \quad (33)$$

In this way, we have the following equation:

$$S(x_t, y_t) > S(x_B, y_B) \quad (34)$$

After using our proposed local mean entropy to weight the erosion operation, we obtain:

$$(f \ominus b_{R,\varepsilon} \cdot S)(x_t, y_t) < (f \ominus b_{R,\varepsilon} \cdot S)(x_B, y_B) \quad (35)$$

which means that the target area is more corroded than the background area with clutter. In this way, more clutter is reserved if the target is removed from the background. Consequently, after subtracting the background from the original image, the target is enhanced.

For the flat background region, we define (x_n, y_n) as pixel. Since the gray value is uniform in the flat region, we have the following definition:

$$f_{i_{x_n,y_n}} - \mu_{k_{x_n,y_n}} \approx 0 \quad (36)$$

Thus,

$$S \approx 1, (f \ominus b_{R,\varepsilon} \cdot S)(x_n, y_n) \approx (f \ominus b_{R,\varepsilon})(x_n, y_n) \quad (37)$$

Here, the degree of the erosion operation is hardly changed by introducing our proposed improved local entropy.

In other words, our local mean entropy can control the degree of the erosion operation of top-hat transformation, which causes the erosion operation to smooth the target while retaining as much clutters as possible, which increases the ability of our adaptive structural elements.
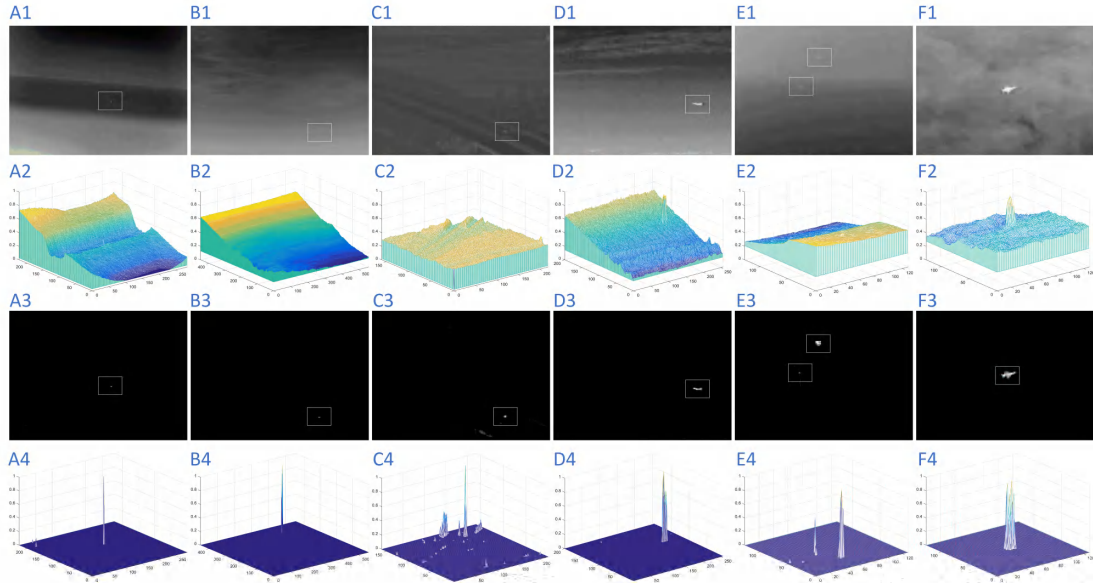
Fig. 3. (A1)-(F1) and (A2)-(F2) denote the six real image sequences and three-dimensional gray distributions, (A3)-(F3) and (A4)-(F4) denote the output image and corresponding three-dimensional distribution
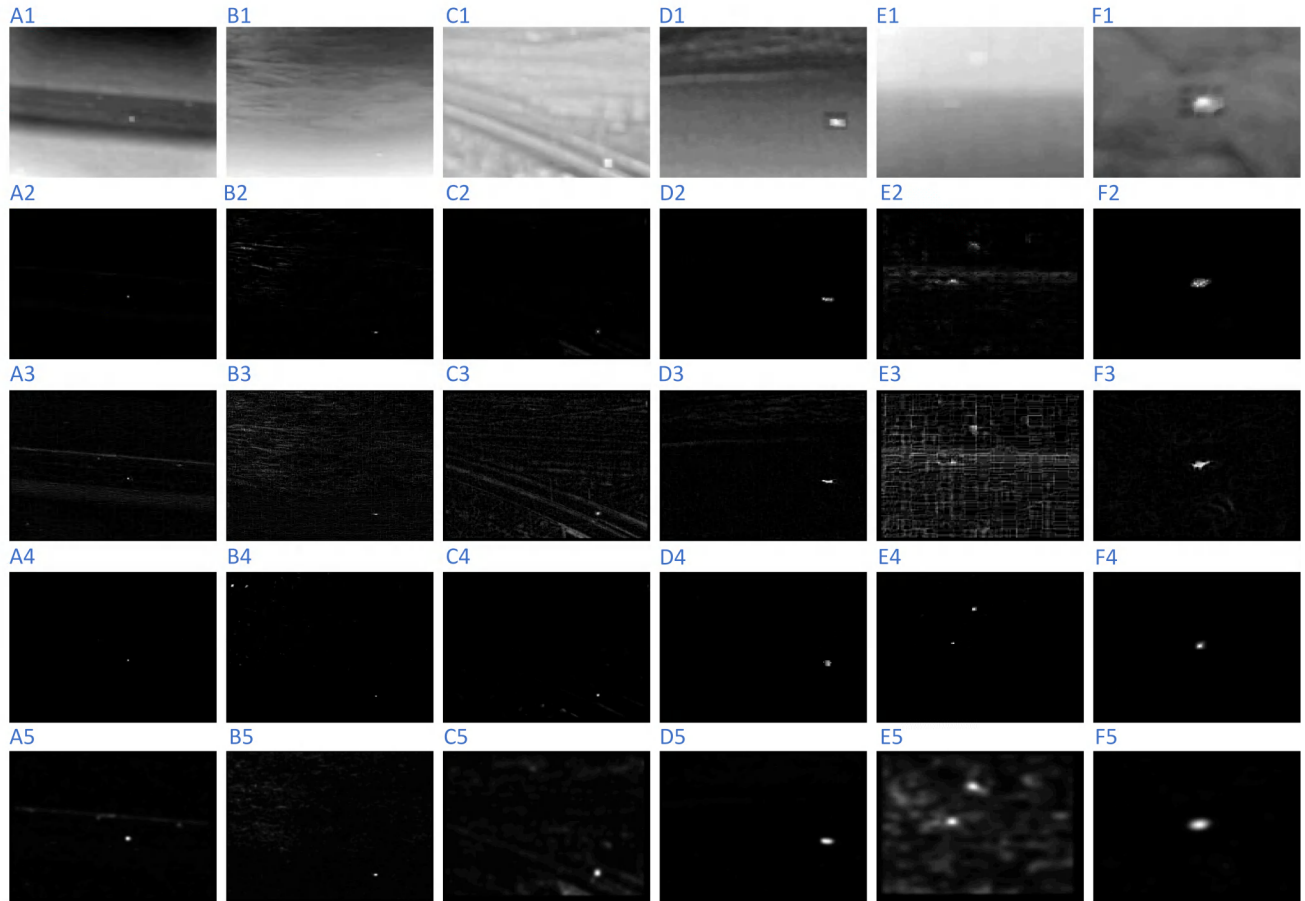

Fig. 4. Experimental results of different methods. (A1)-(F1),(A2)-(F2), (A3)-(F3), (A4)-(F4), and (A5)-F5) are the enhanced results obtained through the LCM , WLDM, NWTH, MPCM, and LACRFR respectively.

TABLE I
THE DETAILS OF REAL IMAGE SEQUENCES

	Image size(pixels) ($width \times height$)	Target size(pixels) ($width \times height$)	Target type	Target details	Background details
sequence1	286×225	$2 \times 2 \sim 3 \times 3$	A small ship	Small and motionless	Sea-sky background with little change
sequence2	561×452	$3 \times 2 \sim 4 \times 3$	Airplane	Dim and constant movement	Sky background with clouds clutter
sequence3	128×128	$4 \times 6 \sim 5 \times 6$	A small ship	A short imaging distance Keeping motion	Blurred sea-sky background with heavy noise.
sequence4	256×200	$7 \times 4 \sim 9 \times 5$	Airplane	A change size and constant movement	Sea-sky background with clouds.
sequence5	220×160	$3 \times 3 \sim 4 \times 5$	Car	A small size and motionless	Highway background
sequence6	128×128	$9 \times 5 \sim 10 \times 6$	Aircraft	A large size and motionless	Sky background with heavy clouds constant change

V. EXPERIMENT AND DISCUSSION

A. Experimental Setting

First of all, we introduce the evaluation indicators and compare several recent methods. Next, we make use of three image sequences with heavy clutter in complex background to demonstrate the contribution of our proposed adaptive structural elements and local mean entropy in top-hat transformation. The most part of dataset is collected from the [44]. We collect real infrared images and generate synthetic ones to validate the proposed model. Furthermore, we use real small target image sequences in different backgrounds to show the superiority of our method through comparison. The experimental data used in the article contains six sets of real infrared sequences, each set of 500 frames, the sequence contains a different number of real targets. Our experimental data sets are actually collected infrared sequence images. In some comparative experiments, because of insufficient data, we use simulated data for experiments. The experiments are performed on a computer with 8-GB of random access memory and an Intel(R) Core(TM) i5-7440HQ CPU with 2.80GHz processor, and the implementation is presented in Matlab 2016a.

B. Evaluation Metrics and Baseline Methods for Comparison

The main tasks for detecting targets in complex background are suppressing the background clutters and enhancing the target. Consequently, if there is less background clutter after using our method, the target will be more obvious.

Based on this fact, we use signal-to-clutter ratio gain (SCRG) and background suppression factor (BSF) to evaluate the performance of target enhancement after using different methods, since these evaluation indicators can measure the significance of the target and the degree of clutter in the background. If the value of SCRG and BSF are higher, a target is more obvious. The definitions of the SCRG and BSF are formulated as follows:

$$SCRG = \frac{(S/C)_{out}}{(S/C)_{in}} \tag{38}$$

$$BSF = \frac{C_{in}}{C_{out}} \tag{39}$$

The S and C denote average value of the pixels in image and the clutter standard deviation respectively, and C_{in}, C_{out}

denote the standard deviation of input image and output image respectively.

Additionally, the detection probability P_D and the rate of false alarm P_F are also taken to evaluate the detection performance. Their definitions are formulated as:

$$P_D = \frac{N_t}{N_C} \times 100\% \tag{40}$$

$$P_F = \frac{N_f}{N} \times 100\% \tag{41}$$

where N_t, N_C, N_f and N denote the number of correctly detected targets, actual targets, false detections, and total detection number respectively.

Our method exploits the mathematical morphology and entropy to detect target, which also uses the local mutation information caused by the emergence of small target in an image. We choose some recent methods to perform the comparative test: LCM [45], NWTH [25], MPCM [46], LACRFR [47] and WLDM [48].

C. Target Enhancement and Performance Analysis

In this section, six image sequences with different kinds of targets and background (more than 500 images) are used in our experiment. They are denoted as sequences 1-6. In Fig. 3, there are six images representing six sequences and detailed information about the target is listed in Table I. Moreover, some recent baseline methods are introduced for comparison, including:

- Local contrast method(**LCM**) [45]. This method utilizes the contrast of brightness as a standard to decide whether the method should pay attention to the areas or not.
- New white top-hat(**NWTH**) [49]. NWTH transformation utilizes structuring elements considering target region and background region. This method is superior than the white top-hat method transformation in viewpoint of target enhancement.
- Multiscale patch-based contrast measure(**MPCM**) [46]. MPCM can increase the contrast between target and background, which makes it easy to segment small target by simple adaptive thresholding method.
- Local adaptive contrast operation based on regularized feature reconstruction(**LACRFR**) [47]. Based on the closed-form solution derived from regularized feature
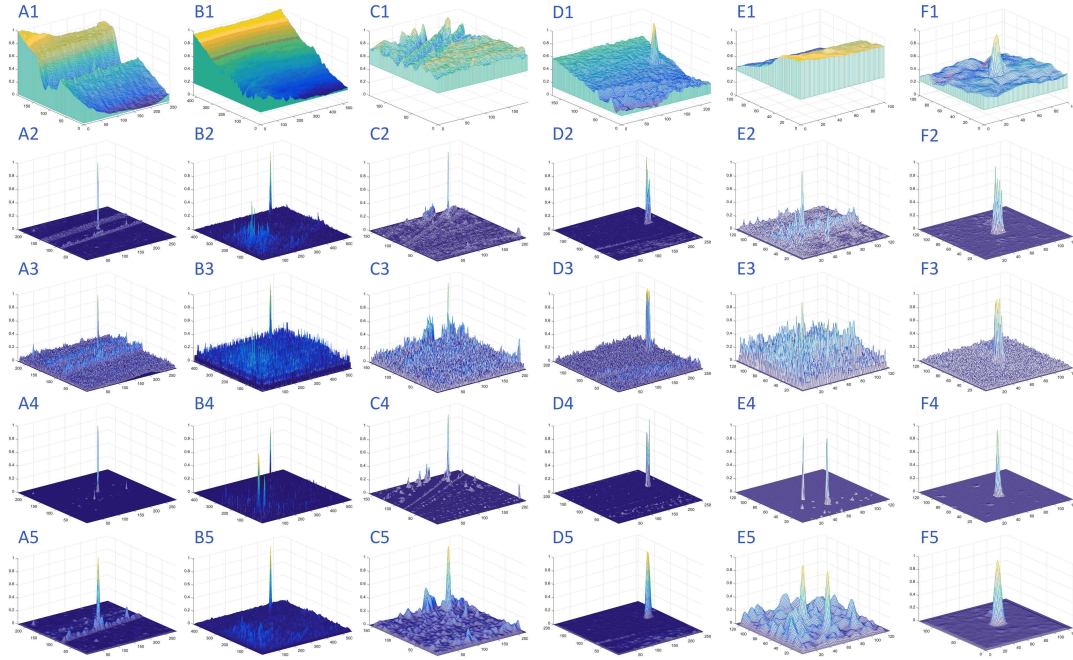
Fig. 5. The 3D distribution from other methods. All results are normalized in $[0, 1]$. (A1)-(F1),(A2)-(F2), (A3)-(F3), (A4)-(F4), and (A5)-F5) are the enhanced results obtained through the LCM , WLDM, NWTH, MPCM, and LACRFR respectively.
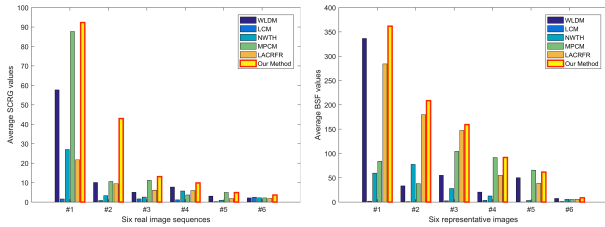

Fig. 6. The average value of SCRG and BSF between different methods .

reconstruction, this method can perform local adaptive contrast operation. Finally, the spatial saliency map and the temporal saliency map can be calculated on the spatial domain and the temporal domain.

- Weighted local difference measure(**WLDM**) [48]. This method weights a multiscale local difference contrast to separate the true target from jamming objects. Subsequently, this method utilizes a simple threshold to detect the target in the filtered result. And this measure can simultaneously enhance the target and suppress background clutters.

Here, extensive images with different targets and background are used to compare the performance of these methods and our method. Hence, the pixels of all results are normalized in $[0, 1]$. (Fig. 3, A4-F4) shows the enhanced 3D gray distributions obtained by our methods. We can see that in original image (Fig. 3, A1-F1), most targets are not obvious, and the background is complex. The corresponding 3D gray input images are displayed in (Fig. 3, A2-F2). After applying our method, the background clutter is almost suppressed. In this way, the target in our results can be detected easily. The result of other methods are displayed in Fig. 4 and Fig. 5. Our method has less clutters and the targets are more obvious

compared with other methods,.

Furthermore, the values of SCRG and BSF are used to prove the performance of these methods. The high values of SCRG and BSF indicate the better performance. We compute the average value of SCRG and BSF to objectively evaluate these six image sequences. The results of each sequence are displayed in Fig. 6. Two bar images present the average values of SCRG and BSF for six image sequences from different methods. Our method obtains higher scores, which also suggests that our method can enhance the target better than other methods. The above experimental results demonstrate that our method can achieve higher performance in terms of target enhancement and background suppression .

According to the protocol in [5], if the distance between centers of the ground-truth does not exceed 5 pixels, then we assume that the detected result is correct. The targets in sequences 2 and 4 keep moving in each frame, while the targets in other frames have little movement (Table I). For sequences 2 and 4, Fig. 7 displays the ground-truth target movement trajectories and the detected trajectories from our method. We can find that there are only a few errors and that horizontal and vertical errors are both less than 10 pixels. A conclusion can be drawn from Fig. 7 that proposed method can achieve a high detection ratio and low false alarm rates for moving targets.

The receiver operating characteristic (ROC) curve can also reflect the relationship between detection versus false alarm rates. We provide ROC curves obtained using different methods in Fig. 8, which suggests the better performance of our method in other ways.
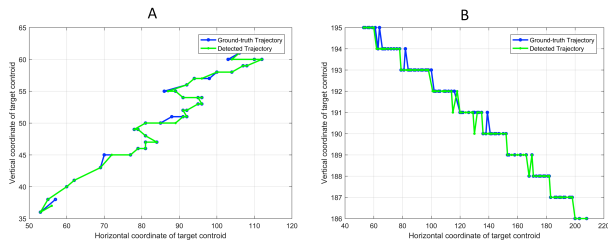
Fig. 7. Ground-truth and detected trajectories obtained by using our method: A and b Ground-truth and detected trajectories of Sequence 2 and 4.

TABLE II
THE CONTRIBUTION OF THE DIFFERENT PARTS IN OUR ALGORITHM

	Sequence a SCRG/BSF	Sequence b SCRG/BSF	Sequence c SCRG/BSF	Average Runing Time
Top-hat	1.69/2.98	1.14/2.25	1.74/2.59	1.31s
GFK top-hat	5.86/9.30	2.99/4.43	5.47/5.72	2.38s
EGFK top-hat	15.84/13.02	4.96/6.14	9.35/13.76	5.64s

D. The Contribution Analysis of the Adaptive Structural Elements and the Proposed Local Mean Entropy

Our adaptive structural elements allow our top-hat transformation to handle complex background because the proposed local mean entropy makes our algorithm more robust under clutters. Here we compare the results between top-hat transformation, top-hat transformation based on guided filter kernel (GFK top-hat transformation) and entropy-driven top-hat transformation based on a guided filter (EGFK top-hat transformation).

We adopt three representative image sequences (named sequences a, b and c) with complex backgrounds and clutters. The size of the target is ranges from 10 to 25 pixels. As shown in Fig. 9, A1-A6 are the representative three input images and 3D gray distribution images. It is obvious that there is heavy clutter in the background area. The top-hat transformation can not suppress the background thoroughly due to the simple structural elements. After using our adaptive structural elements, the background is suppressed better (second line, B1-B6). However, there is still some clutter in the results that can be seen in the 3D gray distribution result (C2,C4,C6). Finally, after using the proposed local mean entropy, the background is clear and the target is enhanced sufficiently. Moreover, the testing average values of the SCRG and BSF are listed in Table II, which shows the incremental contribution of GFK top hat transformation and EGFK top-hat transformation.

E. Sensitivity Analysis of the Crucial Parameters

In this section, we analyze the sensitivity of the crucial parameters of our algorithm. Two parameters are tested reasonably in our proposed top-hat transformation.

The size of the sliding window controls the size of the structural elements and the area size of the local entropy. If the size is too small, the window cannot cover the target. Thus, it may not sufficiently utilize the local difference and background information. This problem potentially yields a high miss rate for target detection. However, if the larger window is considered, the window covers the background clutter and may cause the false detections. In order to analyze the effects of the window size, we set the size of the window

as $3\times3, 5\times5, 7\times7, 9\times9, 11\times11$ respectively. Then we test the SCRG value and BSF value for the above three representative images. The result is shown in Fig. 10, and we find that the choice of $R = 5$ achieves good performance in the value of BSF and SCRG.

Then, we fix R and test the value of $\varepsilon = \{0.001, 0.003, 0.006, 0.01, 0.03, 0.06, 0.1, 0.2, 0.4, 0.6, 0.9\}$ from the setting $\varepsilon \in (0,1)$ [40]. We can find that $\varepsilon = 0.01$ and $\varepsilon = 0.03$ achieve the highest SCRG value and BSF value. In the above experiment, we set the $R = 5$ and $\varepsilon = 0.02$ for all test images.

F. Computational Complexity

In this part, we discuss the complexity of our method briefly. The whole framework can be shown in Fig. 1. We find that the calculation of our algorithm mainly includes two parts: the structural element operation, and the entropy operation. Assuming the size of the image as $m \times n$. The computational complexity is $O(m \times n \times R^2)$ according to the local entropy operation. For structural element construction, the complexity is also $O(m \times n \times R^2)$. Then for the dilation and erosion operation, the total computational complexity is $O(m^2 \times n^2 \times R^2)$. Therefore, on the basis of the above analysis, the total computational complexity is $O(m \times n \times R^2 + 2 \times m^2 \times n^2 \times R^2)$. The table III shows the algorithm complexity and computational time of different methods.

VI. CONCLUSION

In our paper, we present an entropy-driven top-hat transformation with a guided filter kernel for small infrared target detection. The entropy-driven top-hat transformation is used to suppress the clutters of an image capturing the features of clutters. Then structural elements based on a guided filter kernel allow our top-hat transformation to capture the local feature information for background suppression. In this way, our top-hat transformation can not only enhance the target in the image with a low SCRG value, but also effectively suppress the background with clutter. According to the threshold adaptation, our algorithm can obtain a higher detection probability and lower false alarm rate.The experimental result also shows that our algorithm can acquire a higher value of SCRG under different backgrounds compared with several other approaches.

REFERENCES

[1] C. E. Caefer, J. Silverman, and J. M. Mooney, "Optimization of point target tracking filters," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 1, pp. 15–25, Jan 2000.

[2] S. Leonov, "Nonparametric methods for clutter removal," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 3, pp. 832–848, Jul. 2001.

[3] T. Zhang, M. Li, Z. Zuo, W. Yang, and X. Sun, "Moving dim point target detection with three-dimensional wide-to-exact search directional filtering," *Pattern Recognition Letters*, vol. 28, no. 2, pp. 246 – 253, 2007.

[4] X. Bai and Y. Bi, "Derivative entropy-based contrast measure for infrared small-target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 4, pp. 2452–2466, Apr. 2018.

[5] H. Deng, X. Sun, and X. Zhou, "A multiscale fuzzy metric for detecting small infrared targets against chaotic cloudy/sea-sky backgrounds," *IEEE Transactions on Cybernetics*, vol. 49, no. 5, pp. 1694–1707, May 2018.
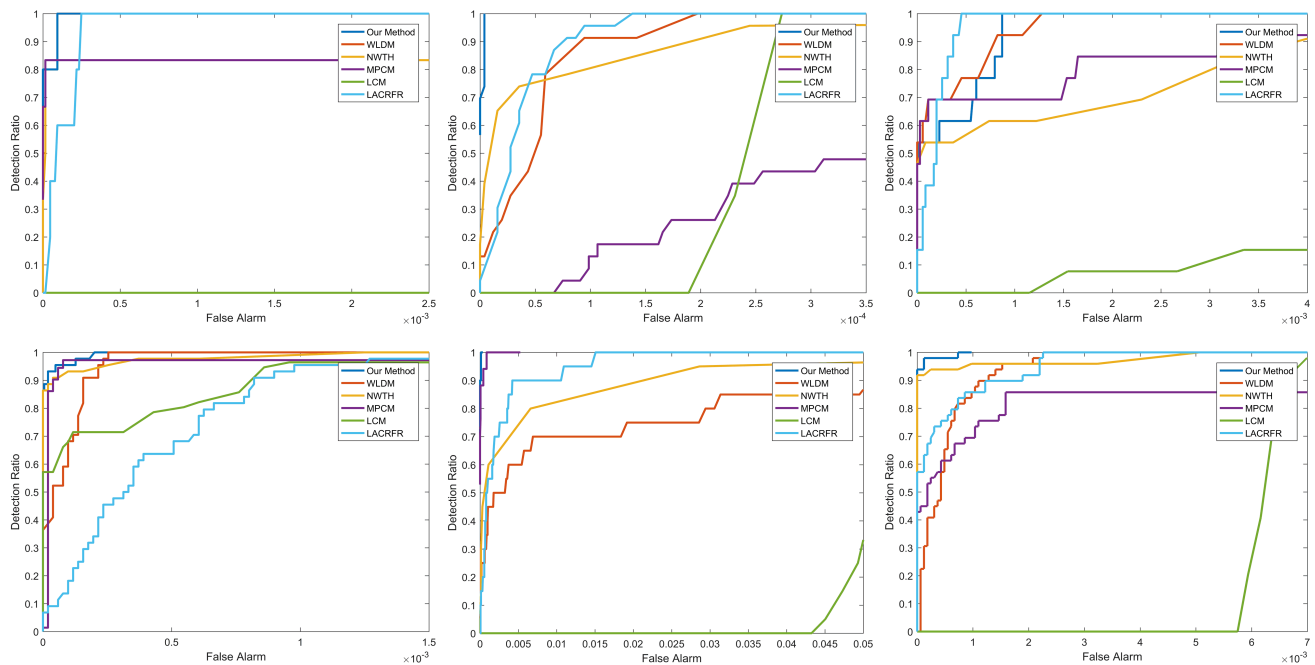
Fig. 8. The ROC curves of all sequences are obtained by using the other methods and our method.

TABLE III
THE COMPUTATIONAL COMPLEXITY AND TIME CONSUMED THROUGH DIFFERENT METHODS

	WLDM	LCM	NWTH	MPCM	LACRFR	Our Method
Complexity	$O\left(R^3 m n^2\right)$	$O(m \times n \times R^2)$	$O(m^2 \times n^2)$	$O(l \times m \times n \times R^2)$	$O\left(m^2 n^2\right)$	$O(m^2 \times n^2 \times R^2)$
Time/s	6.08	0.57	0.72	0.50	4.54	6.74

[6] 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 Transactions on Image Processing*, vol. 22, no. 12, pp. 4996–5009, Dec. 2013.

[7] X. Cao, C. Rong, and X. Bai, "Infrared small target detection based on derivative dissimilarity measure," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 8, pp. 3101–3116, Sep. 2019.

[8] Y. Qin, L. Bruzzone, C. Gao, and B. Li, "Infrared small target detection based on facet kernel and random walker," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 7104–7118, Sep. 2019.

[9] Y. Shi, Y. Wei, H. Yao, D. Pan, and G. Xiao, "High-boost-based multiscale local contrast measure for infrared small target detection," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 1, pp. 33–37, Jan. 2018.

[10] L. Lin, S. Wang, and Z. Tang, "Using deep learning to detect small targets in infrared oversampling images," *Journal of Systems Engineering and Electronics*, vol. 29, no. 5, pp. 947–952, Oct. 2018.

[11] A. Mehmood and N. M. Nasrabadi, "Kernel wavelet-reed–xiaoli: an anomaly detection for forward-looking infrared imagery," *Applied optics*, vol. 50, no. 17, pp. 2744–2751, 2011.

[12] H. Kwon and N. M. Nasrabadi, "Kernel rx-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE transactions on Geoscience and Remote Sensing*, vol. 43, no. 2, pp. 388–397, 2005.

[13] A. Mehmood and N. M. Nasrabadi, "Wavelet-rx anomaly detection for dual-band forward-looking infrared imagery," *Applied optics*, vol. 49, no. 24, pp. 4621–4632, 2010.

[14] M. V. Shirvaikar and M. M. Trivedi, "A neural network filter to detect small targets in high clutter backgrounds," *IEEE Transactions on Neural Networks*, vol. 6, no. 1, pp. 252–257, Jan. 2002.

[15] P. Wang, J. W. Tian, and C. Q. Gao, "Infrared small target detection using directional highpass filters based on LS-SVM," *Electronics Letters*, vol. 45, no. 3, pp. 156–158, Jan. 2009.

[16] H. Li, Y. Wei, L. Li, and Y. Y. Tang, "Infrared moving target detection and tracking based on tensor locality preserving projection," *Infrared Physics & Technology*, vol. 53, no. 2, pp. 77 – 83, 2010.

[17] I. Shevlev and S. Avidan, "Co-occurrence neural network," in *Proceed-*

ings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 4797–4804.

[18] C. Guo, B. Fan, Q. Zhang, S. Xiang, and C. Pan, "Augfpn: Improving multi-scale feature learning for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 595–12 604.

[19] Z. He and L. Zhang, "Domain adaptive object detection via asymmetric tri-way faster-rcnn," *arXiv preprint arXiv:2007.01571*, 2020.

[20] N. Wang, Y. Gao, H. Chen, P. Wang, Z. Tian, C. Shen, and Y. Zhang, "Nas-fcos: Fast neural architecture search for object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 11 943–11 951.

[21] X. Bai and F. Zhou, "Infrared small target enhancement and detection based on modified top-hat transformations," *Computers & Electrical Engineering*, vol. 36, no. 6, pp. 1193 – 1201, 2010.

[22] M. Zeng, J. Li, and Z. Peng, "The design of top-hat morphological filter and application to infrared target detection," *Infrared Physics & Technology*, vol. 48, no. 1, pp. 67 – 76, 2006.

[23] P. T. Jackway and M. Deriche, "Scale-space properties of the multiscale morphological dilation-erosion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 1, pp. 38–51, Jan. 1996.

[24] Y. Chen and Y. Xin, "An efficient infrared small target detection method based on visual contrast mechanism," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 7, pp. 962–966, July 2016.

[25] X. Bai and F. Zhou, "Analysis of new top-hat transformation and the application for infrared dim small target detection," *Pattern Recognition*, vol. 43, no. 6, pp. 2145 – 2156, 2010.

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

[27] X. Bai, F. Zhou, and B. Xue, "Infrared dim small target enhancement using toggle contrast operator," *Infrared Physics & Technology*, vol. 55, no. 2, pp. 177 – 182, 2012.

[28] J. Guo and G. Chen, "Analysis of selection of structural element in mathematical morphology with application to infrared point target detection," in *Infrared Materials, Devices, and Applications*, vol. 6835, 2008, p. 68350P.

[29] X. Wang, Z. Peng, D. Kong, and Y. He, "Infrared dim and small target detection based on stable multisubspace learning in heterogeneous
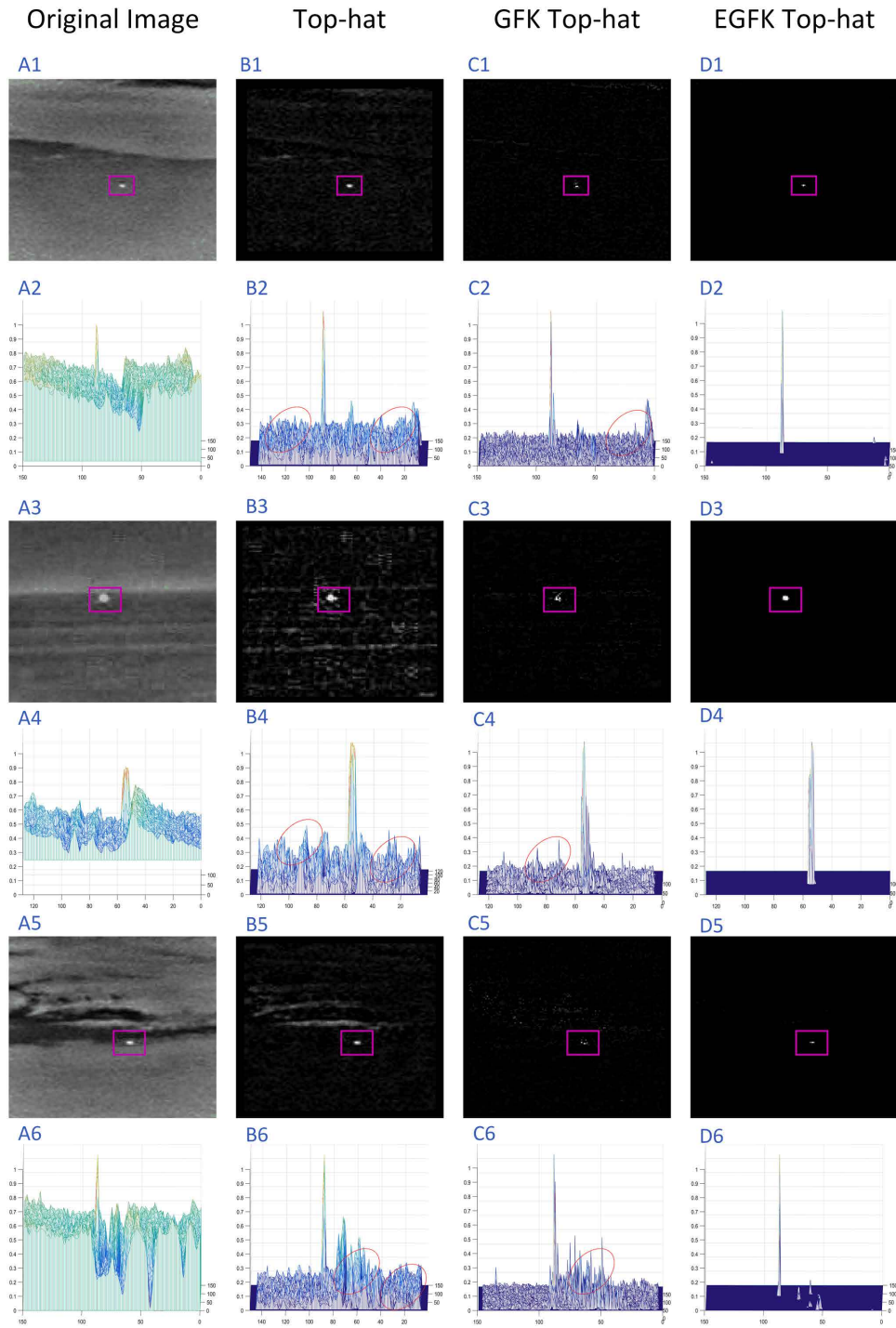
Fig. 9. The figure shows the component with/without comparison amongst the traditional top-hat, GFK top-hat, and our proposed EGFK top-hat method.

scene," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 10, pp. 5481–5493, Oct. 2017.

[30] P. Maragos and C. Vachier, "A pde formulation for viscous morphological operators with extensions to intensity-adaptive operators," in *Proc. 15th IEEE Int. Conf. Image Processing*, Dec. 2008, pp. 2200–2203.

[31] R. Verd-Monedero, J. Angulo, and J. Serra, "Anisotropic morphological filters with spatially-variant structuring elements based on image-dependent gradient fields," *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 200–212, Jan. 2011.

[32] H. Zhu, S. Liu, L. Deng, Y. Li, and F. Xiao, "Infrared small target detection via low-rank tensor completion with top-hat regularization," *IEEE Transactions on Geoence and Remote Sensing*, vol. PP, no. 99, pp. 1–13, 2019.

[33] H. Zhu, J. Zhang, G. Xu, and L. Deng, "Balanced ring top-hat transformation for infrared small target detection with guided filter kernel," *IEEE Transactions on Aerospace and Electronic Systems*, vol. PP, no. 99, pp. 1–1, 2020.

[34] H. Liu, J. You, X. You, and Q. Su, "Fuzzy petri nets using intuitionistic fuzzy sets and ordered weighted averaging operators," *IEEE Transactions on Cybernetics*, vol. 46, no. 8, pp. 1839–1850, Aug. 2016.

[35] X. Qu, H. Chen, and G. Peng, "Novel detection method for infrared small targets using weighted information entropy," *Journal of Systems Engineering and Electronics*, vol. 23, no. 6, pp. 838–842, Dec. 2012.

[36] F. Kou, W. Chen, C. Wen, and Z. Li, "Gradient domain guided image filtering," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4528–4539, Aug. 2015.
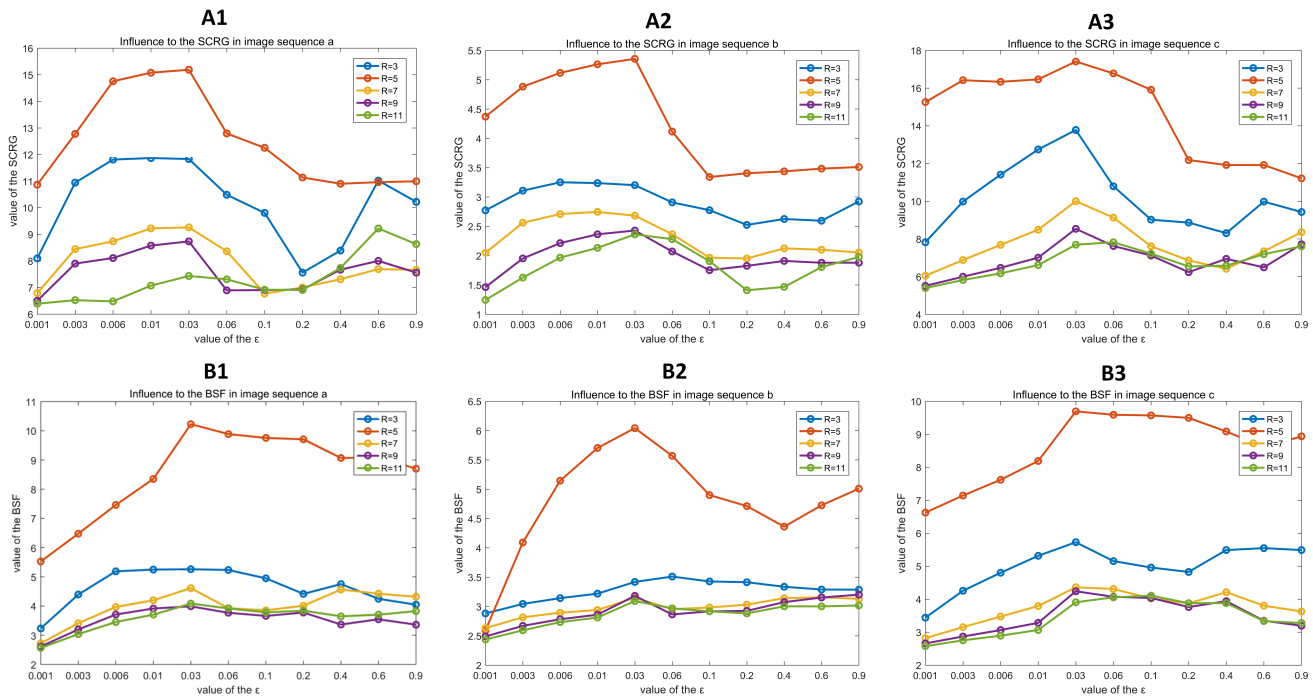
Fig. 10. The analysis of the parameters R and ε, we can find that the superior robustness of our proposed method.

[37] T. Hui and K. N. Ngan, "Depth enhancement using rgb-d guided filtering," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Jan. 2015, pp. 3832–3836.

[38] L. Caraffa, J. Tarel, and P. Charbonnier, "The guided bilateral filter: When the joint/cross bilateral filter becomes robust," *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, vol. 24, no. 4, pp. 1199–1208, April 2015.

[39] D. Du, H. Lu, L. Zhang, and F. Li, "Visual tracking via guided filter," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Sep. 2015, pp. 1781–1785.

[40] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, June 2013.

[41] A. Levin, "A closed form solution to natural image matting," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR'06)*, vol. 1, Jun. 2006, pp. 61–68.

[42] A. Zomet and S. Peleg, "Multi-sensor super-resolution," in *Proc. Sixth IEEE Workshop Applications of Computer Vision (WACV 2002)*, Dec. 2002, pp. 27–31.

[43] X. Zhao, P. Shi, and X. Zheng, "Fuzzy adaptive control design and discretization for a class of nonlinear uncertain systems," *IEEE Transactions on Cybernetics*, vol. 46, no. 6, pp. 1476–1483, Jun. 2015.

[44] H. Zhu, H. Ni, S. Liu, G. Xu, and L. Deng, "Tnlrs: Target-aware non-local low-rank modeling with saliency filtering regularization for infrared small target detection," *IEEE Transactions on Image Processing*, vol. 29, pp. 9546–9558, 2020.

[45] C. L. P. Chen, H. Li, Y. Wei, T. Xia, and Y. Y. Tang, "A local contrast method for small infrared target detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 1, pp. 574–581, Jan. 2013.

[46] Y. Wei, X. You, and H. Li, "Multiscale patch-based contrast measure for small infrared target detection," *Pattern Recognition*, vol. 58, pp. 216–226, 2016.

[47] Y. Li, Y. Zhang, J.-G. Yu, Y. Tan, J. Tian, and J. Ma, "A novel spatio-temporal saliency approach for robust dim moving target detection from airborne infrared image sequences," *information sciences*, vol. 369, pp. 548–563, 2016.

[48] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Small infrared target detection based on weighted local difference measure," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 7, pp. 4204–4214, July 2016.

[49] X. Bai and F. Zhou, "Analysis of new top-hat transformation and the application for infrared dim small target detection," *Pattern Recognition*, vol. 43, no. 6, pp. 2145–2156, 2010.

Lizhen Deng (M'17) received the B.S. degree in electronic information science and technology from Huaibei Coal Industry Teachers College, Huaibei, China, in 2007, and the M.S. degree in communication and information systems from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2010. She received her Ph.D. degree in electrical engineering from Huazhong University of Science and Technology, China, in 2014. In 2014, she joined the Nanjing University of Posts and Telecommunications, Nanjing, China. Her current research interests include image processing, computer vision, pattern recognition, and spectral data processing. E-mail: alicedenglzh@gmail.com

Guoxia Xu (M'19) received the B.S. degree in information and computer science from Yancheng Teachers University, Jiangsu Yancheng, China in 2015, and the M.S. degree in computer science and technology from Hohai University, Nanjing, China in 2018. He was a research assistant in City University of Hong Kong and Chinese University of Hong Kong. Now, he is pursing his Ph.D. degree in Department of Computer Science, Norwegian University of Science and Technology, Gjovik Norway. His research interest includes pattern recognition, image processing, and computer vision. E-mail: gxxu.re@gmail.com

Jieke Zhang received the B.S. degree in communication engineering from Qufu Normal University in 2017.He currently is pursuing his master degree in electronics and communications engineering at NanJing University of Post and Telecommunication. Now his research interest is image processing. Email: 1217012341@njupt.edu.cn

Hu Zhu (M'17) received the B.S. degree in mathematics and applied mathematics from Huaibei Coal Industry Teachers College, Huaibei, China, in 2007, and the M.S. and Ph.D. degrees in computational mathematics and pattern recognition and intelligent systems from Huazhong University of Science and Technology, Wuhan, China, in 2009 and 2013, respectively. In 2013, he joined the Nanjing University of Posts and Telecommunications, Nanjing, China. His research interests include pattern recognition, image processing, and computer vision. E-mail: peter.hu.zhu@gmail.com