A Multiscale Fuzzy Metric for Detecting Small Infrared Targets Against Chaotic Cloudy/Sea-Sky Backgrounds

He Deng, Xianping Sun, and Xin Zhou

Abstract—In a low signal-to-clutter ratio (SCR) small-infrared-target image with chaotic cloudy-/sea-sky background, the target has very similar thermal intensities to the background (e.g., edges of clouds). In such case, how to accurately detect small targets is crucial in infrared search and tracking applications. Conventional methods based on the local difference/mutation potentially result in high miss and/or false alarm rates. Here, we propose an effective method for detecting small infrared targets embedded in complex backgrounds through a multiscale fuzzy metric that measures the certainty of targets in images. Accordingly, the detection task is formulated as a fuzzy measure issue. The presented metric is able to eliminate substantial background clutters and noise. Especially, it significantly improves SCR values of the image. Subsequently, a simple and adaptive threshold is used to segment target. Extensive clipped and real data experiments demonstrate that the proposed algorithm not only works more robustly for different target sizes, SCR values, target and/or background types, but also has better performance regarding detection accuracy, when compared with traditional baseline methods. Moreover, the mathematical proofs are provided for understanding the proposed detection method.

Index Terms—Infrared image, multiscale fuzzy metric (MFM), small target detection.

I. INTRODUCTION

SMALL infrared target detection is crucial in infrared search and tracking (IRST) systems, because the performance of whole IRST depends upon the accuracy of detection results [1], [2]. Different to the general object or saliency detection tasks (such as visible light images), the major challenge/difficulty of small target detection is short of enough prior knowledge about the target (such as, size, shape, texture, and velocity) due to the long imaging distance. A small target usually possesses several pixels (less than 0.15% of the image size) [3] and its brightness possibly changes from dim to bright [1]. On the other hand, a small target is generally submerged in heavy noise and strong background clutters (e.g., cloudy and/or sea clutters) with low signal-to-clutter ratio (SCR). Although many efforts have been concentrated on this issue over the last two decades [4]–[11], it is still a difficulty problem to tackle.

Due to the long distance between the imaging sensor and the target, the spatial–temporal information, background features or the relation between the background and target is significant for the target detection. Given this, existing techniques are roughly classified into the track-before-detect (TBD) and detect-before-track (DBT) methods [4], [5]. TBD algorithms usually process several frames to estimate targets, while need prior information. For example, 3-D matched (directional) filtering [6] could detect moving small targets with a constant velocity, which requires some knowledge about the target shape and velocity. 3-D double directional filtering [7] is effective to detect weak targets, but demands the information about the maximum target velocity. In addition, the methods based on the modified partial differential equation [8] or support vector machines [9] are used to suppress background clutters in a single image, and then eliminate false alarms using multiframe accumulation or autocorrelation.

When compared to TBD methods, DBT algorithms are more powerful because of lower computation complexity, and fewer requirements of assumptions and prior knowledge. In general, there are two necessary assumptions in most of DBT methods: one is that the background has correlation in spatial domain as well as stability in time domain. It generally occupies the low frequency portion of an image in frequency domain. The other assumption is that a small target is unrelated to the background in spatial domain, and it dominates the high frequency portion of the image. Some techniques based on these assumptions are utilized to eliminate background clutters. The finite or infinite impulse response filtering and space-time maximum likelihood algorithms [11] can suppress backgrounds with the assumption of short-time stationarity. Top hat [12] and median filtering are widely used to reduce clutters. Howbeit, these methods would cause high false alarms when SCR is low [13], [14]. Moreover, classification-based methods are able to remove clutter points [15], for example the manifold learning [16], nearest neighbor classifier [17] or convolutional neural network [18], [19]. There are still many

other techniques, including methods based on the tri-feature detector [20], or biological vision [21].

Recently, there arises a novel trend toward the fact that the emergence of small target leads to significant changes of local rather than global texture for promoting detection performance [22]–[25]. Local difference- or mutation-based algorithms have displayed great potential in diverse small target detection tasks. None the less, such algorithm requires to be ameliorated. For example, the sparse representation scheme has been exploited to characterize the local difference between the small target and neighboring backgrounds [22], but is unable to characterize the background very well [13]. Wang *et al.* [23] put forward an average gray absolute difference maximum map (AGADM) for target detection, but it roughly characterizes the background and possibly produces high false alarms [14]. Chen *et al.* [3] proposed a local contrast map (LCM) to "pop out" small targets and "neglect" background areas as much as possible. Howbeit, this algorithm assumes that the target is bright, which is not suitable for detecting dark targets. To solve this problem, we explored a local difference measure [24] to measure local mutation and then detect dark and bright targets. However, this method is time-consuming to determine the size of regional window. The above research shows that measuring local difference/mutation between the small target and neighboring background is vital in small infrared target detection.

Noticed that the ambiguity and uncertainty are unavoidably arisen during image acquisition and image transmission, a valid way representing such image should utilize human knowledge expressed heuristically. Besides, the description of object in an image has fuzziness. Owing to the ability for tackling imperfect data derived from the ambiguity and vagueness [26]–[28], fuzzy sets are widely applied in various computer vision and machine learning fields, including image segmentation [29], [30], image enhancement [31], and object tracking [26]. Accordingly, fuzzy sets can guide the design of a new metric, which could measure the certainty due to the emergence of a small target.

With these considerations in mind, a novel local mutation or difference metric (called as multiscale fuzzy metric, MFM) is designed from the perspective of fuzzy sets and local difference/mutation models in this paper, which focuses on how to pop out small targets and "suppress" backgrounds. Experimental results verify that the proposed method is effective regarding detection accuracy. Especially, it displays a distinct superiority over several widely used baseline techniques in terms of SCR and background suppression factor (BSF).

The contributions of this paper can be summarized as follows.

1) A novel local difference/mutation measure named MFM is proposed, which is able to measure the certainty of small targets in infrared images under complicate background clutters/noise. As a result, the target detection task is transformed into a fuzzy measure issue. The MFM is capable to well pop out targets, while neglect backgrounds at the same time. In particular, it significantly improves SCR and BSF values of the image.

2) A new MFM-based detection algorithm is designed. By applying such method on extensive clipped and real small target images against diverse complex backgrounds (the number of clipped images is 105, and that of real images is 1345), it demonstrates that the designed method is simple and effective regarding the detection accuracy. Moreover, the theoretical analyses of MFM-based scheme are supplemented. This gives useful foundation for comprehending the designed small-target detection method and instructing future developments. The technique presented recently is able to guarantee to robustly and efficiently tackle low SCR images with different targets types and backgrounds.

The organization of the remainder of this paper is as follows. In Section II, we explain the MFM-based small target detection method in detail. In Section III, we present theoretical analyses of MFM. In Section IV, we give extensive experimental results and discussions, and several factors of the proposed method are discussed in Section V. The conclusions and perspectives are given in Section VI.

II. Small Target Detection Based on Fuzzy Metric

In this section, we introduce a novel detection scheme. This scheme exploits a fuzzy metric to suppress background clutters and enhance targets, and then uses a simple threshold to detect targets. It is good at improving SCR values of the image.

A. Fuzzy Metric

Suppose that Θ is a finite set, a fuzzy set $A = \{\mu_\theta | \theta \in \Theta\}$ is represented by a membership function/degree μ_θ, where θ is a generic element of Θ. The μ_θ associates a real number within an interval [0,1] with each point in Θ. The function is a measure of belongingness or membership degree of an element in Θ [32].

An image can be considered as a cluster of fuzzy singletons indicating the membership value of each pixel point regarding a predefined image property (such as brightness or homogeneity) for specific applications. In a small target image, the target has signature of discontinuity with its neighboring background area [3]. For representing that discontinuity, the membership degree/function of an image with L gray levels is defined as

$$\mu_\theta = \mu(x, y) = 1 - \left[\exp\left(1 - \left|\tilde{x}^2 - \tilde{y}^2\right|\right) - 1\right]/(\exp(1) - 1)$$
$$\text{where } \tilde{x} = x/(L-1), \tilde{y} = y/(L-1) \qquad (1)$$

where x and y denote gray values at the current and neighboring points. Then, an image is converted into a fuzzy set through the function $\mu_\theta : [0, L-1] \times [0, L-1] \to [0, 1]$.

A function is a metric/distance function that defines a metric/ distance between each pair of elements of a set, if it satisfies the following conditions [33]: 1) $D(x, y) \geq 0[D(x, y) = 0$ iff $x = y]$ and 2) $D(x, y) = D(y, x)$; 3) $D(y, z) \leq D(x, y) + D(x, z)$, where x, y, and z are arbitrary elements of the set. Hence, the membership function defined using (1) satisfies the conditions for the distance function, as shown in the following proposition.

Proposition 1: Suppose that an image is with L gray levels, and x, y, and z are gray values at different points (viz., $x, y, z \in [0, L-1]$), then (1) satisfies: 1) $0 \leq \mu(x, y) \leq 1$; 2) $\mu(x, y) = 0$ iff $x = y$; 3) $\mu(x, y) = \mu(y, x)$; and 4) $\mu(y, z) \leq \mu(x, y) + \mu(x, z)$.

Proof:

1) Because $x, y \in [0, L-1]$, $\tilde{x} = x/(L-1) \in [0, 1]$, $\tilde{y} = y/(L-1) \in [0, 1]$, and $0 \leq 1 - |\tilde{x}^2 - \tilde{y}^2| \leq 1$. Thus, $0 \leq [\exp(1 - |\tilde{x}^2 - \tilde{y}^2|) - 1]/(\exp(1) - 1) \leq 1$. Accordingly, $0 \leq \mu(x, y) \leq 1$ for any x and y.

2) When $x = y$, $\mu(x, y) = 1 - (\exp(1) - 1)/(\exp(1) - 1) = 0$.

3) $\mu(x, y) = \mu(y, x)$.

4) Let $t_1 = |\tilde{x}^2 - \tilde{y}^2|$, $t_2 = |\tilde{x}^2 - \tilde{z}^2|$, $t_3 = |\tilde{y}^2 - \tilde{z}^2|$, where $\tilde{z} = z/(L-1)$, thus $0 \leq t_1 \leq 1$, $0 \leq t_2 \leq 1$ and $0 \leq t_3 \leq 1$. Now, we consider a function $g(t_1, t_2)$ on $[0, 1] \times [0, 1]$, which is defined as follows:

$$g(t_1, t_2) = 1 + \frac{\exp(1 - (t_1 + t_2)) - \exp(1 - t_1) - \exp(1 - t_2) + 1}{\exp(1) - 1}.$$

Then

$$\frac{\partial g(t_1, t_2)}{\partial t_1} = \frac{\exp(1 - t_1) - \exp(1 - (t_1 + t_2))}{\exp(1) - 1}$$
$$= \frac{\exp(1 - t_1)}{\exp(1) - 1}\left(1 - \frac{1}{\exp(t_2)}\right) \geq 0$$
$$\frac{\partial g(t_1, t_2)}{\partial t_2} = \frac{\exp(1 - t_2) - \exp(1 - (t_1 + t_2))}{\exp(1) - 1}$$
$$= \frac{\exp(1 - t_2)}{\exp(1) - 1}\left(1 - \frac{1}{\exp(t_1)}\right) \geq 0.$$

Accordingly, the function $g(t_1, t_2)$ is monotonically increasing along both t_1 and t_2 directions. As a result, when $(t_1, t_2) = (0, 0)$, $g(t_1, t_2)$ achieves the minimum value $g(0, 0) = 0$. This suggests that $g(t_1, t_2)$ is no less than 0 for any t_1 and t_2 in an interval [0,1].

Because

$$\left|\tilde{y}^2 - \tilde{z}^2\right| = \left|\left(\tilde{x}^2 - \tilde{y}^2\right) + \left(\tilde{z}^2 - \tilde{x}^2\right)\right| \leq \left|\tilde{x}^2 - \tilde{y}^2\right| + \left|\tilde{z}^2 - \tilde{x}^2\right|$$
$$t_3 \leq t_1 + t_2.$$

Thus, $1 + [(\exp(1 - t_3) - \exp(1 - t_1) - \exp(1 - t_2) + 1)/(\exp(1) - 1)] \geq g(t_1, t_2) \geq 0$, viz.,

$$\left[1 - \frac{\exp(1 - t_1) - 1}{\exp(1) - 1}\right] + \left[1 - \frac{\exp(1 - t_2) - 1}{\exp(1) - 1}\right]$$
$$\geq \left[1 - \frac{\exp(1 - (t_1 + t_2)) - 1}{\exp(1) - 1}\right].$$

Consequently, $\mu(y, z) \leq \mu(x, y) + \mu(x, z)$. ∎

Proposition 1 suggests that a function using (1) is a fuzzy metric function. However, a neighborhood around current pixel x consists of many neighboring pixels, which arises the variety of membership values according to the selection of neighboring pixels. An alternative strategy may substitute some proper rank statistics of neighborhood for the variable y in (1), such as the min, max, mean, median, or others. In this case, (1) is a membership degree that measures the difference of a pixel to an image block, or the metric/distance (or certainty) between the pixel and its neighborhood.

B. Multiscale Fuzzy Metric

The fuzzy metric (1) can measure the distance/discontinuity between a current point and its neighboring one. However, this metric is generally sensitive to noise. An effective artifice is to obtain more information from the neighborhood of the current pixel. A small target often occupies several pixels in an infrared image [23], and the target region is viewed as a homogeneous and compact area [2], [3], [14]. Thus, we substitute the average gray value of current region for the variable x in (1), and replace the variable y in (1) with the average gray value of surrounding region. This is also derived from the fact that the discontinuity between the target and neighboring background is essentially involved determining the property of average gray difference based on the neighboring pixels [23]. Accordingly, (1) turns into a function $\mu : Đ*_1 \times Đ*_2 = \{(x, y)|x \in Đ*_1, y \in Đ*_2\}$, and the exponential term is transformed from the $1 - |x^2 - y^2|$ into the $1 - |\text{mean}^2(Đ*_1)\text{-mean}^2(Đ*_2)|$, where $Đ*_1$ and $Đ*_2$ are the current and surrounding regions.

Nevertheless, it arouses another important issue (viz., how to determine the size of target region or window/patch) since the target size constantly changes along with the change of imaging distance, environments, target types, and so on. In this case, the window or patch size should change as the target size changes. However, the prior knowledge about the target size or imaging distance is scarcely acquired in applications [1]–[3]. To tackle such case, a multiscale model is adopted to smartly manipulate multiple target scales [3], [14]. Thus, an MFM is proposed to represent the distance/metric between the target region and surrounding background clutters.

Given a pixel point (p, q) in an image, its kth neighborhood is defined as follows:

$$\Omega_k = \{(a, b)||p - a| \leq k \text{ and } |q - b| \leq k\}, k = 1, 2, \ldots, K \tag{2}$$

where K is a positive integer that defines the maximum size of window. Then, the average gray value of the kth neighborhood Ω_k is defined as

$$A_k(p, q) = \frac{1}{\#\Omega_k} \sum_{(s, t) \in \Omega_k} f(s, t) \tag{3}$$

where $\#\Omega_k$ is the number of pixels contained in Ω_k, and $f(s, t)$ is the gray value at the point (s, t).

In this way, a fuzzy metric between two neighboring regions is defined as the following formula:

$$\mu(A_i, A_j) = 1 - \left[\exp\left(1 - \left|\tilde{A}_i^2 - \tilde{A}_j^2\right|\right) - 1\right]/(\exp(1) - 1)$$

where $\tilde{A}_i = A_i/(L - 1)$, $\tilde{A}_j = A_j/(L - 1)$, $i, j \in \{1, 2, \ldots, K\}$. \tag{4}

Similar to (1), (4) satisfies the conditions for distance functions (see Proposition 2).

Proposition 2: Assume that an image is with L gray levels, and A_i, A_j, and A_k are average gray values of neighborhoods centered at the point (p, q), then a fuzzy metric using (4) satisfies: 1) $0 \leq \mu(A_i, A_j) \leq 1$; 2) $\mu(A_i, A_j) = 0$ iff $A_i = $

A_j; 3) $\mu(A_i, A_j) = \mu(A_j, A_i)$; and 4) $\mu(A_i, A_k) \leq \mu(A_i, A_j) + \mu(A_j, A_k)$.

Proof: The proof of Proposition 2 is analogy to that of Proposition 1. ∎

Different to (1) and (4) measures a distance between two neighboring regions centered at the same point. When the index k changes from 1 to K, we can achieve a fuzzy distance matrix

$$\Lambda = \begin{Bmatrix} 0 & \mu(A_1, A_2) & \cdots & \mu(A_1, A_K) \\ \mu(A_2, A_1) & 0 & \cdots & \mu(A_2, A_K) \\ \cdots & \cdots & \cdots & \cdots \\ \mu(A_K, A_1) & \mu(A_K, A_2) & \cdots & 0 \end{Bmatrix}. \quad (5)$$

Matrix Λ is a symmetric matrix, where every element denotes the distance between the current region and neighboring one. A fuzzy metric with diverse scales is to meet potential changes of target size owing to the possible change of imaging distance.

Moreover, a small target region in an image has conspicuous discontinuity with its surrounding background regions [2], [3], [23]. After that, the most dissimilar point according to the fuzzy metric can be considered as a target point. Then, MFM centered at the point (p, q) is defined as

$$v(p, q) = \max\{\Lambda, 0\}. \quad (6)$$

For each point in a given scale [viz., K in (2)], its MFM is acquired according to (2)–(6). When the sliding window moves in an image from left to right and from top to down, the MFM map (MFMM) will be achieved. Each element represents the maximum discontinuity between the current region and the surrounding neighborhoods.

C. MFMM-Based Small Target Detection

As mentioned before, a small target is discontinuous with its neighboring background area [3], [23]. According to properties of the metric function, the background region in an image may result in low MFM values, while the target region may produce high one (see Proposition 3). In this case, the MFMM is able to enlarge the discontinuity between the target and its neighboring areas (see Proposition 4), and effectively enhance small target as well as suppress complex background clutters and noise (see Propositions 5 and 6). Therefore, we believe that, if an MFMM is achieved, the local region whose MFM is greater than a given threshold in some scale may be a position where target appears.

We know that fewer residues of clutters and noise determine lower false alarm rates under the same probability of detection [1]. In order to eliminate clutters and noise residual as much as possible, the final filtered result is achieved through an iterative filtering procedure (see Fig. 1). Since the average local entropy (ALE) is utilized to quantify the textural information reduction along with iterations [34], we adopt this concept to determine the number of MFMM iterations. For an image with size of $M \times N$, the ALE is defined as

$$\text{ALE} = \frac{1}{M \times N} \sum_{x=1}^{M} \sum_{y=1}^{N} \text{LE}(x, y) \quad (7)$$
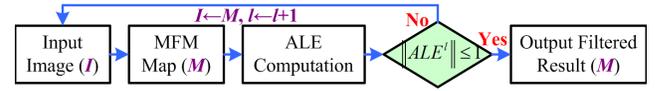
Fig. 1. Flowchart of the iterative filtering procedure.
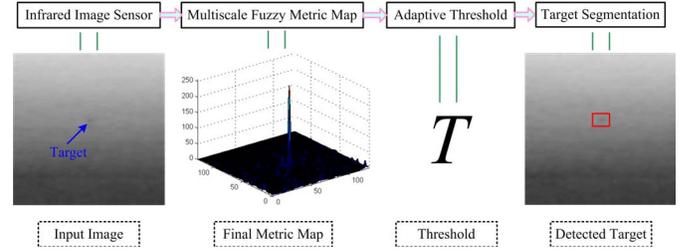
Fig. 2. Flowchart of the proposed small target detection scheme.

where $\text{LE}(x, y)$ denotes the entropy of a 9×9 window centered at the point (x, y). After that, we set a stopping criterion as

$$\left\| \text{ALE}^l \right\| \leq 1 \quad (8)$$

where ALE^l denotes the ALE after l iterations, and $\|\cdot\|$ denotes the L_2 norm operator.

After achieving a filtered result, a simple and adaptive global threshold is utilized to segment targets. The adaptive threshold is defined as

$$T = \alpha \cdot E_{\max} + \beta \cdot m_t, \quad \text{s.t. } \alpha + \beta = 1 \quad (9)$$

where $E_{\max}$ and m_t denotes the maximum and average values of the filtered result, α and β are positive constants. The constraint of coefficients (the sum is set as 1) is to ensure the rationality of threshold (viz., the threshold T is within the interval $[m_t, E_{\max}]$), avoiding producing an invalid threshold. As previously stated, the most dissimilar point can possibly be target point, as guided by MFM. Then we segment a pixel at the point (i, j) as the target pixel if the point with value $\geq T$, otherwise it is a background pixel. Since a small target region has higher MFM value than its neighboring background areas, irrespective of the bright or dark target, (9) will give an appropriate threshold to detect the target once the α and β are selected. In our experiments, the alpha is chosen in an interval [0.4,0.6], and the beta is 1-alpha.

To intuitively display the proposed MFM-based small target detection algorithm, the whole detection scheme is given in Fig. 2. In general, the user can manually and experimentally pick the maximum window size [viz., $(2K + 1) \times (2K + 1)$] to calculate the MFM in (6), α and β to compute the threshold in (9), as well as stopping criterion for practical design requirements (see Section V for details).

III. ALGORITHM ANALYSIS

Some theoretical analyses of the proposed detection method are discussed in this section. This can supply useful foundation for understanding the method and guiding future developments.

A. Algorithm Properties

From the definition, it can be seen that the fuzzy metric using (1) measures the fuzzy distance between the current point and its neighboring ones. To explain the metric as reasonable, we establish the following proposition.

Proposition 3: Assume that $\mu(x, y)$ is a fuzzy metric using (1), then $\mu(x, y)$ satisfies (where x, y, and z are the gray values at different points, and the image has L gray levels):

1) $\mu(x, y) = 1$, iff $x = L - 1, y = 0$, or $x = 0, y = L - 1$;
2) for all $x, y, z \in [0, L - 1]$, if $x \leq y \leq z$, then $\mu(x, y) \leq \mu(x, z)$, and $\mu(y, z) \leq \mu(x, z)$;
3) If $x_1 \leq y_1 \leq y_2 \leq x_2 \in [0, L - 1]$, then $\mu(y_1, y_2) \leq \mu(x_1, x_2)$.

Proof:

1) When $x = L - 1, y = 0$, or $x = 0, y = L - 1, \tilde{x} = 1, \tilde{y} = 0$ or $\tilde{x} = 0, \tilde{y} = 1$, then $\mu(x, y) = 1 - (\exp(0) - 1)/(\exp(1) - 1) = 1$. On the contrary, if $\mu(x, y) = 1$, then $|\tilde{x}^2 - \tilde{y}^2| = 1$. Because $\tilde{x} \in [0, 1], \tilde{y} \in [0, 1]$, then $\tilde{x} = 1, \tilde{y} = 0$ or $\tilde{x} = 0, \tilde{y} = 1$. This suggests that $x = L - 1, y = 0$, or $x = 0, y = L - 1$.
2) Let $t = |\tilde{x}^2 - \tilde{y}^2| \in [0, 1]$, thus (1) becomes $\mu(t) = 1 - (\exp(1 - t) - 1)/(\exp(1) - 1)$. Then $(\partial \mu(t)/\partial t) = (\exp(1 - t)/[\exp(1) - 1]) > 0$. Consequently, the function $\mu(t)$ is monotonic increasing. If $x \leq y \leq z$, then $|\tilde{x}^2 - \tilde{y}^2| \leq |\tilde{x}^2 - \tilde{z}^2|, |\tilde{y}^2 - \tilde{z}^2| \leq |\tilde{x}^2 - \tilde{z}^2|$. This indicates that $\mu(x, y) \leq \mu(x, z)$ and $\mu(y, z)$ and $\mu(x, z)$.
3) Suppose that $t_1 = |\tilde{x}_1^2 - \tilde{x}_2^2|, t_2 = |\tilde{y}_1^2 - \tilde{y}_2^2|$ (where $\tilde{x}_1 = x_1/(L - 1), \tilde{x}_2 = x_2/(L - 1), \tilde{y}_1 = y_1/(L - 1), \tilde{y}_2 = y_2/(L - 1)$), then $t_2 \leq t_1 \in [0, 1]$. Because $\mu(t)$ is an increasing function proved in the above, thus $\mu(t_2) \leq \mu(t_1)$. Accordingly, $\mu(y_1, y_2) \leq \mu(x_1, x_2)$ is valid. ∎

According to Propositions 1 and 3, the fuzzy metric is a local function. Proposition 3 indicates that the more gray difference between two points in an image owns the larger distance. This is helpful for detecting small targets because target pixels are significantly nonanalogous to its surrounding background ones that have similar imaging properties.

It is easy to prove that MFM has the similar conclusions of Proposition 3. In this case, the MFM in a target region is higher than that in a background area. Accordingly, this metric can be used to measure the certainty of small target in infrared images.

B. Detection Ability Analysis

In a small target image with low contrast, the distribution of target highly resembles that of background, which may lead to minute discrimination (such as intensity difference) between the target and neighboring background regions. However, after fuzzy metric (4), the difference is expanded (see Proposition 4).

Proposition 4: The fuzzy metric (4) can broaden the original intensity discontinuity between any two neighboring regions.

Proof: Suppose that Ω_i and Ω_j denote any two neighboring regions centered at the same point (the respective average gray values are C_i and C_j, and the image is normalized to [0,1] prior), then the original intensity difference is $|C_i - C_j|$, denoted as t. Through the metric (4), the difference becomes $1 - [\exp(1 - t) - 1]/(\exp(1) - 1)$. Now, we consider a function $f(t) = 1 - [\exp(1 - t) - 1]/(\exp(1) - 1) - t$ on [0,1]. Then

$$\frac{\partial f(t)}{\partial t} = \frac{\exp(1 - t)}{\exp(1) - 1} - 1, \frac{\partial f^2(t)}{\partial t^2} = \frac{-\exp(1 - t)}{\exp(1) - 1} \leq 0. \quad (10)$$

Hence, the function $f(t)$ has a minimum at $1 - \log(\exp(1) - 1)$, and the minimum is $(\exp(1) - 2)/(\exp(1) - 1) + \log(\exp(1) - 1) \approx 0.1233$. This indicates that the function $f(t)$ is always larger than 0. Thus, the discontinuity using (4) is greater than the original. ∎

Assume that C_b is the average intensity of a neighboring area (including m pixels) around a target region (containing n pixels, $n < m$, and the average intensity is C_t), the original discontinuity between the target and neighboring regions is $m/(m - n) \cdot |C_b - C_t|$. Because the original intensity difference (viz., $|C_b - C_t|$) could be enlarged through the fuzzy metric (4) (see Proposition 4), the difference between the target area and neighboring background ones is expanded when adopted the metric (4). Accordingly, the MFM is able to broaden the difference between the target and neighboring background regions. After multiple iterations, such difference is gradually escalated. This significantly upraises the separability of target from background clutters and noise.

Furthermore, Propositions 5 and 6 indicate that the MFM can produce less clutters and noise residual. Since false alarm rates under the same probability of detection are direct proportional to the complexity of clutters and noise residual, the MFMM can result in low false alarm rates.

Proposition 5: Suppose that a small target image owns L gray levels and (p_0, q_0) is a current point in the background, then its MFM $v^b(p_0, q_0) \approx 0$.

Proof: Because the background is related to its neighboring areas in spatial domain, the difference between the current region and neighboring ones is trivial. Then the formula is warrantable

$$A_k^b(p_0, q_0) \cong A_{k+1}^b(p_0, q_0), \quad \text{where } k = 1, 2, \ldots, K - 1 \quad (11)$$

where A_k^b is the average gray value of the kth neighborhood.

According to Proposition 2, we can acquire

$$\mu^b\left(A_i^b, A_j^b\right) \cong 0, i, j \in \{1, 2, \ldots, K\} \quad (12)$$

where $\mu^b(A_i^b, A_j^b)$ is the fuzzy metric (4) in the background areas. Then, the fuzzy distance matrix is approximately considered as a null matrix. After that, the MFM centered at the point (p_0, q_0) is near to zero, viz., $v^b(p_0, q_0) \approx 0$. ∎

Proposition 6: Assume that (i_0, j_0) is the center of a target in an infrared image, whose MFM is $v^t(i_0, j_0)$, then $0 < v^t(i_0, j_0) \leq 1$.

Proof: A small target usually concentrates in a small, compact, and homogeneous region [3], [14]. Then for a bright target, it can be found that

$$A_k^t(i_0, j_0) \geq A_{k+1}^t(i_0, j_0), \quad \text{where } k = 1, 2, \ldots, K - 1 \quad (13)$$

where A_k^t is the average gray value of the kth neighborhood.

According to Proposition 3, we can gain

$$\forall i, j, k \in \{1, 2, \ldots, K\}, \text{ and } i \leq j \leq k$$
$$\Rightarrow \mu^t\left(A_i^t, A_j^t\right) \leq \mu^t\left(A_i^t, A_k^t\right), \mu^t\left(A_j^t, A_k^t\right) \leq \mu^t\left(A_i^t, A_k^t\right) \quad (14)$$

where $\mu^t(A_i^t, A_j^t)$, $\mu^t(A_i^t, A_k^t)$ and $\mu^t(A_j^t, A_k^t)$ are fuzzy metrics using (4). In this way, the MFM centered at the point (i_0, j_0) is

$$\nu^t(i_0, j_0) = \mu^t\left(A_1^t, A_K^t\right) \in (0, 1]. \quad (15)$$

Accordingly, MFM represents the maximum difference among neighboring areas around the small target.

For a dark small target, if the current region only involves the target, (13)–(15) are tenable according to Proposition 3. On the other hand, if the current region includes both the target and neighboring background, we can find that

$$\tilde{A}_m^t(i_0, j_0) \leq \tilde{A}_{m+1}^t(i_0, j_0), m = M + 1, \ldots, K - 1$$
$$\text{and } \tilde{A}_i^t(i_0, j_0) \leq \tilde{A}_K^t(i_0, j_0), i \in \{1, 2, \ldots, K\} \quad (16)$$

where M denotes the maximum scale of the target. In this case, $\tilde{A}_1^t(x_0, y_0) \leq \tilde{A}_p^t(x_0, y_0) \leq \tilde{A}_K^t(x_0, y_0)$, $1 \leq p \leq K$. Then, the MFM is

$$\tilde{\nu}^t(i_0, j_0) = \tilde{\mu}^t\left(\tilde{A}_1^t, \tilde{A}_K^t\right) \in (0, 1]. \quad (17)$$

■

When compared to the original intensity difference in a small infrared target image, Propositions 4–6 suggest that the MFM is more significant in target regions. If the maximum window size in (2) is set appropriately, the MFM value will be close to 1 in target regions. Moreover, Propositions 5 and 6 indicate that the MFM in a target region is higher than that in a background one. This reveals twofold roles of MFM: the target of interest is enhanced, while complicated background clutters and noise are adaptively suppressed. Accordingly, the proposed MFM-based small target detection algorithm is able to work well for images against diverse backgrounds.

C. Computational Complexity

The complexity of our method is briefly discussed. As shown in Fig. 2, it is not hard to find that the computational time of the proposed algorithm mainly consists of three parts: 1) the MFMM operation; 2) the ALE computation; and 3) the target segmentation.

The computation complexity of MFMM operation primarily depends on the fuzzy metric computation (4) across all pixels in an image. For each pixel point, the MFM computation (5) needs $O(K^2 2^n)$ cost (viz., the worst-case time complexity of algorithm, which is defined as the maximum amount of time taken on the input size n), where K is the maximum size of sliding window (2). Accordingly, the entire computation complexity of this step is around $O(NrcK^2 2^n)$, where r and c are the row and column numbers of the original image, and N is the iteration number of the MFMM algorithm, respectively.

For ALE computation, we necessitate implementing the local entropy operator across all pixels in the image as well. For each pixel, the local entropy operator can be carried out in $O(h \log(n))$ time, where h is the neighborhood size.

Consequently, the total computation complexity of this step is around $O(Nrch \log(n))$.

For target segmentation, a simple threshold is used. For each pixel, only a simple comparison operator is performed. Thus, the entire cost of this step is around $O(rc)$.

According to the above analysis, the entire computational complexity of the proposed MFM-based algorithm is around $O(NrcK^2 2^n + Nrch \log(n) + rc)$. The real runtime of the proposed method is provided in the supplementary materials.

IV. EXPERIMENTAL RESULTS

In this section, we first introduce some evaluation indexes and baseline algorithms for comparison. After that, we perform experiments on small target images (against different complex backgrounds) clipped from lots of published papers to evaluate the performance of the proposed algorithm. Finally, we use real data to further test the proposed method.

A. Indexes and Baseline Methods

The probability of detection (Pd) and the false alarm rate (Fa) are the most popular indexes for evaluating the target detection performance [1]. Pd denotes the probability of detected targets in multiframe images where targets truly exist, while Fa is the rate of detected targets in multiframe images where targets do not exist. Pd and Fa are defined as

$$Pd = N_d/N_t, Fa = N_f/N \quad (18)$$

where N_d, N_t, N_f, and N denote the number of true detections, the number of actual targets, the number of false detections, and the number of images, respectively. A method with good detection performance means it owns both high Pd and low Fa values.

SCR is widely used to portray the difficulty level of target detection [1]. If an image has several targets, the average SCR (aSCR) is utilized in this section. The aSCR is

$$\text{aSCR} = 1/M \cdot \sum_{j=1}^{M} \text{SCR}_j, \text{SCR}_j = \left|\theta_t^j - \theta_b^j\right|/\sigma_b^j \quad (19)$$

where M is the number of targets, SCR_j is the SCR value of the jth target, θ_t^j is the average gray value of the jth target, θ_b^j and σ_b^j are the average and standard deviation of gray values in the neighboring area around the jth target, respectively.

Besides SCR/aSCR, the BSF is usually used to represent the residual degree of background clutters and noise [24]. The BSF is defined as

$$\text{BSF} = \sigma_I/\sigma_O \quad (20)$$

where σ_I and σ_O denote the standard deviations of gray values in the original and filtered images. Generally speaking, the higher SCR/aSCR and BSF values of an image are, the easier targets can be detected.

The fuzzy metric is similar to some local difference/mutation descriptors that represent local mutation due to the appearance of small targets. Consequently, we select the LCM [3], LMWIE [25], and AGADM [23] as three

baseline methods. Moreover, max-mean (MME), and max-median (MDM) filtering methods [35] are also adopted as baseline methods since they are well studied and widely used for assessing new methods [1], [24].

From the perspective of imaging, a small target defined by Society of Photo-Optical Instrumentation Engineers has a total spatial extent of less than 80 pixels [3]. Hence, for the proposed method, we have $K = 4$, while the number of iterations is decided dynamically according to the iterative strategy (see Fig. 1). The effects of parameters (including the window size and stopping criterion) are discussed in detail in Section V. All experiments are carried out using MATLAB on an Intel Xeon CPU E5-2407, 2.40 GHz machine with 8 GB RAM.

B. Clipped Images

Background suppression is a critical issue in detecting small targets [13]. Less clutters and noise residual is indispensable to maintain lower Fa values under the invariable Pd [1]. Hence, if the proposed algorithm can eliminate more background clutters, the target will be detected more easily.

Massive test small target images are clipped from published papers through a screenshot tool (The clipped images are color png files), then the images are transformed into the gray format (the gray range is [0, 255]). In the test clipped images, the target size possibly changes from 2×2 to more than 9×9 pixels and its brightness could vary from dim to bright due to different target types, environments, and imaging distances (the representative clipped images are shown in Figs. 3 and 4). The targets usually submerge in different complicate backgrounds, for example the images with noisy bright spots or strong sea/cloudy clutters (see Figs. 3 and 4). In addition, the deficiencies of dots per inch, screen/scanning resolution, along with other unknowns potentially produce serious disturbances on those images. This further increases difficulty to tackle such situation.

The number of clipped images is 105. In view of background types, the backgrounds of test clipped images could be simply divided into the cloudy-sky, sea-clutter, terrain-sky, and sea-sky backgrounds. According to the number of small targets, the test images are roughly classified into the single- and multitarget images. On account of the target types, the targets embedded in clipped images are categorized into the cloudy-sky targets (e.g., missile or plane), sea targets (e.g., ship or torpedo), and terrain targets (e.g., panzer or vehicle).

Four representative clipped single-target images are shown in Fig. 3(a1)–(d1), where the blue squares indicate the location of targets. It can be found that the images have low SCR values, low spatial resolution, noisy bright spots, heavy noise, and/or strong cloudy or sea clutters, which lead to great difficulties or challenges in the detection of those targets. Nevertheless, after the proposed method, such heavy disturbances are substantially eliminated, and the targets are well enhanced [as shown in Fig. 3(a2)–(d2)]. In this way, those targets are readily detected.

Four representative clipped multitarget images against sea-sky, cloudy-sky, and terrain-sky backgrounds are displayed in Fig. 4(a1)–(d1), where the white squares indicate the location

of targets and the number of targets is 2. It can be found that the images have very low contrast and the targets have very similar distribution to the background. In this case, the target detection is a difficult issue to tackle. Since the proposed method focuses on how to pop out (or enhance) small targets and neglect (or suppress) backgrounds as much as possible, the respective filtered results of Fig. 4(a1)–(d1) are shown in Fig. 4(a2)–(d2). Similar to Fig. 3, the substantial interference factors (such as clutters and/or noise) are effectively eliminated, which ensures the stability and robustness of the proposed algorithm regarding the detection performance. The conclusion derived from Figs. 3 and 4 is that the proposed method can work stably for different backgrounds, target types, target sizes, and target numbers. The filtered results of representative single- and multitarget images via baseline methods are not given because of space limitation. Please see the supplementary materials for comparisons.

Two indexes (viz., aSCR and BSF) are suitable to represent the filter performance of various methods and are used here for comparison. For each index, a higher score indicates the better performance. As for the representative clipped images, Table I lists aSCR values of the original and filtered images obtained using the baseline and proposed methods. It can be seen that the proposed algorithm can significantly improve aSCR values of image against different complex backgrounds. Table I indicates that the proposed method is superior to the baseline methods (such as the LCM, LMWIE, AGADM, MME, and MED) with regard to the improvement of aSCR values. The BSF values of filtered results through different methods are shown in Table II. We can find that the proposed detection method also shows the superiority over baseline methods in term of BSF comparisons.

Against complex infrared background, a small target usually submerges in clutters/noise. However, the target is dissimilar to surrounding clutters/noise in spatial domain. This dissimilarity is the basis of our method that aims to measure the certainty of small targets in images. It can be found from Proposition 4 that the designed fuzzy metric can broaden the difference between the target and neighboring background regions. Accordingly, the separation of targets from complex clutters and noise is easy to be resolved. Propositions 5 and 6 demonstrate clearly that the metric in background area is less than that in target region. This indicates that the proposed method can availably enhance target by effectively suppressing clutters and noise. Propositions 4–6 are able to explain theoretically the robustness of the proposed detection method. The evidences from Figs. 3 and 4, and Tables I and II authenticate that the proposed MFMM-based technique enable the problem of target detection in the cases of heavy disturbances (for example, noisy bright spots, heavy noise, low contrast, and/or strong cloud/sea clutters) to be stably handled. In addition, more experiments of test clipped images with great interferences of human-made artifacts are supplemented in the supplementary materials.

In accordance with background types, the test clipped images are separated into four groups and every group owns analogous background types (viz., the cloudy-sky, sea-clutter, terrain-sky, and sea-sky background), denoted as Group 1 to 4.
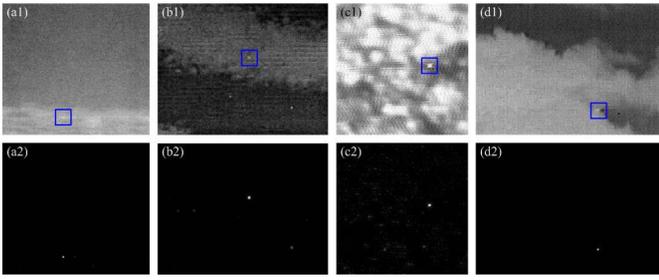
Fig. 3. Representative clipped single-target images against diverse backgrounds and filtered results obtained using the proposed method. Top row, from left to right, are the (a1) original dim-target image with low SCR and heavy noisy sea clutters, (b1) dim-bright-target image with noisy bright spots and strong cloudy clutters, (c1) bright-target image with low spatial resolution and strong noisy sea clutters, and (d1) dim-bright-target image with low resolution, and heavy noisy cloudy clutters. (a2)–(d2) Filtered results of (a1)–(d1) obtained using our method.
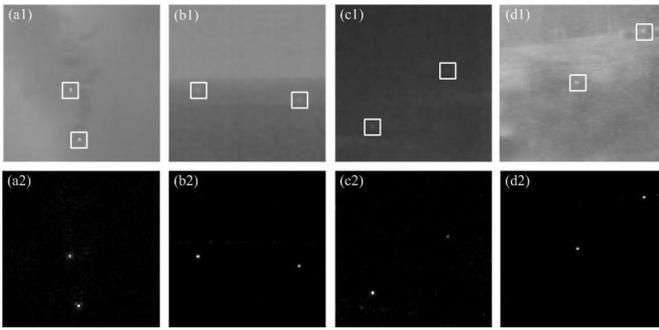
Fig. 4. Representative clipped multitarget images against various backgrounds and filtered results obtained using the proposed method. Top row, from left to right, are the (a1) original bright-targets image against cloudy-sky background, (b1) dim-surface-targets image against sea-sky background, (c1) dim-targets image against terrain-sky background, and (d1) vehicle-targets image against terrain-sky background. (a2)–(d2) Filtered results of (a1)–(d1).

TABLE I
AVERAGE SCR VALUES OF ORIGINAL AND FILTERED IMAGES OBTAINED
USING DIFFERENT METHODS FOR THE REPRESENTATIVE
CLIPPED IMAGES

Original	LCM	LMWIE	AGADM	MME	MDM	MFMM
Fig. 3						
1.80	3.07	0.05	3.86	2.46	2.30	51.00
1.56	3.35	1.31	2.49	2.30	1.98	58.62
2.19	3.51	2.44	2.10	2.06	2.21	23.57
1.11	1.44	0.21	1.28	1.26	1.32	42.89
Fig. 4						
8.66	5.52	30.66	8.28	8.49	9.09	31.87
1.60	2.39	2.51	1.46	1.47	1.56	38.92
4.31	4.91	10.97	4.56	4.56	4.46	22.80
2.68	3.43	11.79	2.46	2.56	2.74	40.45

The number of images in each group is 57, 8, 23, and 17, respectively. This division strategy is able to objectively evaluate the performance of our method in term of different complex backgrounds. Some evaluation results are listed in Tables III–V.

Table III lists the ranges of SCR and ensemble-average SCR (denoted as SCR) of each group of original and filtered images obtained by using the LCM, LMWIE, AGADM, MME, MED, and MFMM methods. It can be seen that the LMWIE produces higher SCR values in comparison with other

TABLE II
BSF VALUES OBTAINED USING DIFFERENT METHODS FOR THE
REPRESENTATIVE SINGLE-AND MULTITARGET IMAGES

LCM	LMWIE	AGADM	MME	MDM	MFMM
Fig. 3					
0.72	1.49	0.57	0.63	0.66	15.56
0.63	3.40	0.53	0.57	0.61	8.13
0.93	2.62	0.71	0.74	0.74	8.24
0.77	3.29	0.58	0.64	0.64	24.09
Fig. 4					
0.42	1.83	0.26	0.26	0.25	2.70
0.22	1.48	0.22	0.22	0.22	4.77
0.17	0.73	0.12	0.13	0.12	1.24
0.37	3.65	0.32	0.32	0.32	4.44

TABLE III
RANGES OF SCR AND ENSEMBLE-AVERAGE SCR VALUES OF FOUR
GROUPS OF ORIGINAL AND FILTERED IMAGES OBTAINED USING
BASELINE AND PROPOSED METHODS

Methods/Indexes		Group1	Group2	Group3	Group4
Original	SCR	0.21~5.35	1.16~3.98	0.81~7.87	0.12~7.98
	SCR	2.35	2.23	2.97	1.54
LCM	SCR	0.12~5.95	1.13~4.99	1.56~6.33	0.15~5.62
	SCR	3.62	2.69	3.21	1.50
LMWIE	SCR	0.12~37.94	1.09~3.91	1.05~14.35	0.18~12.26
	SCR	8.31	2.91	4.42	2.12
AGADM	SCR	0.25~7.85	0.97~4.59	0.71~7.89	0.11~8.56
	SCR	3.62	2.35	3.25	1.59
MME	SCR	0.46~7.31	0.97~4.29	0.98~7.30	0.10~7.30
	SCR	3.10	2.22	2.97	1.50
MED	SCR	0.72~5.14	1.15~4.67	0.96~8.42	0.14~8.13
	SCR	2.79	2.38	3.19	1.55
MFMM	SCR	2.12~58.12	11.14~25.14	3.38~63.07	1.56~54.46
	SCR	35.50	17.28	37.75	12.01

where SCR denotes the ensemble-average SCR value of every group.

baseline methods. However, the MFMM has the best enhancement performance ranging from the cloudy-sky to terrain-sky backgrounds. The same conclusions are also derived from Table IV that lists the BSF ranges and ensemble-average BSF (denoted as BSF) via diverse methods. Table IV suggests that the proposed algorithm outperforms the baseline methods with respect to background suppression for various types of complex background clutters.

Moreover, the probabilities of detection and false alarm rates obtained using the baseline and proposed methods are listed in Table V. We can find that the proposed algorithm can produce high Pd and low Fa values. By contrast, our algorithm can work stably for different complex backgrounds with strong clutters and noise. We resort to Propositions 4–6 to provide theoretical explanations of those superior results achieved via our method. Propositions 4–6 indicate that the difference/mutation between a small target and neighboring background regions is gradually enlarged through the iterative filtering procedure, and residual clutters and noise are availably eliminated. It gives easiness to detect targets accurately.

C. Real Images

In this section, 15 real small-target image sequences with low SCR values (denotes as sequence 1 to 15) are utilized

TABLE IV
RANGES OF BSF AND ENSEMBLE-AVERAGE BSF VALUES OF FOUR
GROUPS OBTAINED USING THE LCM, LMWIE, AGADM,
MME, MED, AND MFMM METHODS

Methods/Indexes		Group1	Group2	Group3	Group4
LCM	BSF	0.27~1.17	0.38~0.59	0.22~0.79	0.23~1.02
	$\overline{BSF}$	0.62	0.42	0.47	0.46
LMWIE	BSF	0.92~18.44	1.09~1.68	0.86~3.48	0.67~5.37
	$\overline{BSF}$	3.60	1.25	1.96	2.17
AGADM	BSF	0.23~0.99	0.31~0.36	0.22~0.76	0.23~0.86
	$\overline{BSF}$	0.52	0.33	0.38	0.36
MME	BSF	0.24~0.99	0.31~0.38	0.22~0.72	0.23~0.86
	$\overline{BSF}$	0.54	0.34	0.39	0.37
MED	BSF	0.24~0.99	0.30~0.37	0.22~0.78	0.23~0.87
	$\overline{BSF}$	0.55	0.33	0.40	0.37
MFMM	BSF	1.93~25.38	2.07~6.13	1.74~10.86	1.79~10.69
	$\overline{BSF}$	7.81	3.50	5.23	3.43

where $\underline{BSF}$ is the ensemble-average BSF value of each group.

TABLE V
PROBABILITIES OF DETECTION AND FALSE ALARM RATES
OBTAINED USING DIFFERENT ALGORITHMS

Methods/Indexes		Group1	Group2	Group3	Group4
LCM	Pd	0.73	0.88	0.72	0.63
	Fa	2.1/image	1.5/image	2.1/image	2.5/image
LMWIE	Pd	0.69	0.63	0.56	0.50
	Fa	2.2/image	2.6/image	2.8/image	3.2/image
AGADM	Pd	0.53	0.50	0.33	0.33
	Fa	3.0/image	3.1/image	4.0/image	3.9/image
MME	Pd	0.49	0.38	0.28	0.30
	Fa	3.2/image	3.6/image	4.2/image	4.1/image
MED	Pd	0.51	0.38	0.28	0.37
	Fa	3.0/image	3.6/image	4.1/image	3.6/image
MFMM	Pd	0.95	1.00	0.94	0.97
	Fa	0.4/image	0.1/image	1.2/image	0.3/image

Fig. 5. Representative real images randomly derived from the 15 real small-infrared-target-image sequences.

Fig. 6. For the representative real images, (a)–(o) denote the 3-D gray distributions of filtered results obtained using the proposed algorithm.

to test the performance of the proposed target detection algorithm. The total number of real small-infrared-target images is 1345. The details about targets and backgrounds are supplemented in the supplementary materials. Fig. 5 displays representative images randomly derived from the 15 sequences. It can be seen that most of test sequences have very low contrast or SCR, and the targets have similar distribution to the backgrounds. In such case, the proposed method is expected to pop out targets and neglect complex backgrounds well.

The backgrounds of the real sequences are roughly classified into the sea-sky backgrounds [e.g., Fig. 5(a)–(f), (n), and (o)] and cloudy-sky backgrounds [e.g., Fig. 5(g)–(m)]. We can find that the images have strong sea-/cloudy-sky clutters and heavy noise. After the proposed algorithm, the clutters and noise are effectively suppressed, and the small targets are well enhanced, as shown in Fig. 6 [viz., the respective 3-D gray distributions of filtered results of Fig. 5(a)–(o)]. Hence, the targets are easily detected. When compared with the filtered results via the LCM, LMWIE, AGADM, MME, and MED methods that are provided in the supplementary materials, our method performs better than the baseline methods regarding both target enhancement and background suppression. The SCR and BSF comparisons can also validate this conclusion, as shown in Fig. 7 and Tables VI and VII.

For the representative real images, Fig. 7(a) shows the aSCR values of the original and filtered images

via the LCM, LMWIE, AGADM, MME, MED, and MFMM methods. We can find that our algorithm significantly improves the aSCR of images with different background clutters, and target sizes and/or types. All the P-values (including $P_{\text{original versus MFMM}}$, $P_{\text{LCM versus. MFMM}}$, $P_{\text{LMWIE versus MFMM}}$, $P_{\text{AGADM versus MFMM}}$, $P_{\text{MME vresus MFMM}}$, and $P_{\text{MED versus MFMM}}$) are less than 0.0001. This indicates that the proposed MFMM algorithm is statistically significantly different from the baseline methods. As for the BSF, the performance of background suppression via different methods is shown in Fig. 7(b). It can be found that the proposed MFMM method is superior to the baseline algorithms for suppressing different intricate backgrounds, except for the comparisons between the LMWIE and MFMM algorithms for Fig. 5(n) and (o). Nevertheless, the targets in Fig. 5(n) and (o) are concealed in the filtered results obtained using the LMWIE (see supplementary materials). This causes great troubles
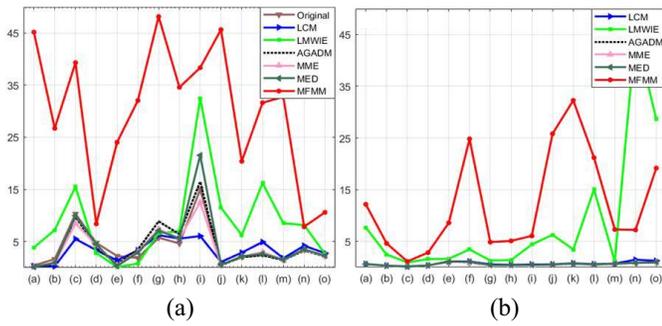
Fig. 7. (a) SCR comparisons of original and filtered images obtained using different algorithm for representative real images. (b) BSF values obtained using the LCM, LMWIE, AGADM, MME, MED, and MFMM methods for the representative real images.

for detecting the targets, that is, high miss rates. For the statistical analysis of BSF term, the $P_{\text{original versus MFMM}}$, $P_{\text{LCM versus MFMM}}$, $P_{\text{AGADM versus MFMM}}$, $P_{\text{MME versus MFMM}}$, and $P_{\text{MED versus MFMM}}$ are less than 0.0001, while $P_{\text{LMWIE versus MFMM}} = 0.335$. There is no significant statistical between the LMWIE and MFMM for suppressing backgrounds, but the LMWIE potentially masks small targets and arouse high miss and/or false detection rate.

For the real sequences, the ensemble-average SCR and BSF comparisons are displayed in Tables VI and VII. Similar to Fig. 7, the conclusions derived from Tables VI and VII demonstrate that the proposed detection algorithm has less clutters and noise residual for different background types, target types, and target sizes, when compared with the baseline methods. This indicates that the proposed detection method maintains lower false alarm rates under the same probability of detection. Since the targets submerge in the filtered results of sequences 14 and 15 obtained using the LMWIE (those filtered images are homogeneous), the ensemble-average BSF values are greater than that achieved by using the MFMM (see Table VII). Howbeit, the target detection in those cases encounters great challenges for the LMWIE (see supplementary materials).

Since small targets in sequences 3 and 13 keep motion in each frame (the number of frames are 65 and 70, respectively), Fig. 8 displays the respective ground-truth and detected trajectories of target motion obtained using the proposed method. We can find that the ground-truth trajectories are curvilinear. For sequence 3, the most horizontal detection errors are no greater than one pixel (the respective frequencies of zero-, one-, and two-pixel errors are 33, 30, and 2), and the most vertical detection errors are zero (the frequency of zero-pixel errors is 44 and that of one-pixel errors is 21). For the sequence 13, the major horizontal detection errors are zeros (the number of zero-pixel errors is 50 and that of one-pixel errors is 20), and the major vertical errors are no more than two pixels (the numbers of zero-, one-, two-, and three-pixel errors are 21, 25, 17, and 7, respectively). It can draw a conclusion from Fig. 8 that the tracked traces do almost match the trajectories of target motion.

The receiver operating characteristic (ROC) curve is adopted to test the detection performance of our method. ROC curve is a plot of the probabilities of detection (a fraction of

TABLE VI
ENSEMBLE-AVERAGE SCR VALUES OF 15 REAL IMAGE SEQUENCES THROUGH DIFFERENT ALGORITHMS

	Original	LCM	LMWIE	AGADM	MME	MED	MFMM
# 1	0.90	0.50	6.50	0.57	0.68	0.66	26.07
# 2	2.10	1.26	10.08	1.29	1.44	1.42	20.42
# 3	7.83	4.16	9.86	7.36	7.98	7.98	22.89
# 4	4.84	3.22	2.70	4.03	4.39	4.57	8.40
# 5	2.13	2.84	1.16	2.61	2.41	2.41	21.07
# 6	1.77	3.14	0.72	3.67	3.02	3.02	38.78
# 7	5.76	5.97	7.30	7.65	6.81	7.01	42.78
# 8	3.05	4.31	2.60	4.46	3.99	3.81	25.69
# 9	11.22	5.74	26.33	11.04	9.77	12.66	35.74
#10	2.23	3.23	11.31	2.24	2.42	2.29	45.00
#11	4.60	6.31	6.56	5.76	5.89	5.89	23.34
#12	5.32	4.81	22.21	5.69	5.58	5.85	32.51
#13	0.72	0.91	4.31	0.67	0.69	0.69	11.68
#14	3.44	4.18	7.34	3.27	3.41	3.53	7.61
#15	2.29	2.79	2.65	2.15	2.19	2.28	10.59

TABLE VII
ENSEMBLE-AVERAGE BSF VALUES OF 15 REAL IMAGE SEQUENCES OBTAINED USING DIFFERENT ALGORITHMS

	LCM	LMWIE	AGADM	MME	MED	MFMM
# 1	0.62	3.83	0.60	0.61	0.61	10.43
# 2	0.30	3.34	0.29	0.31	0.32	4.24
# 3	0.12	0.91	0.07	0.07	0.08	1.10
# 4	0.33	1.97	0.32	0.34	0.35	2.72
# 5	1.13	1.67	1.08	1.07	1.07	7.62
# 6	1.07	4.14	0.96	0.98	1.04	26.21
# 7	0.58	1.34	0.42	0.42	0.42	4.87
# 8	0.38	1.24	0.37	0.39	0.39	3.41
# 9	0.48	4.68	0.36	0.36	0.38	5.91
#10	0.54	6.50	0.48	0.48	0.48	23.91
#11	0.96	5.09	0.72	0.75	0.76	28.63
#12	0.62	14.27	0.53	0.53	0.54	20.87
#13	0.18	0.86	0.17	0.17	0.18	2.31
#14	1.43	45.40	0.83	0.84	0.84	7.28
#15	1.23	29.35	0.91	0.97	0.97	19.08
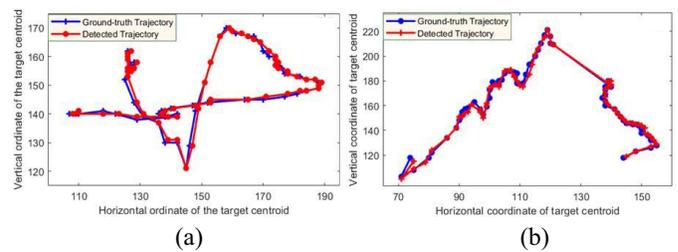
Fig. 8. (a) and (b) Ground-truth trajectories and detected trajectories obtained using the proposed detection method for sequences 3 and 13.

true positives over positives) versus the false alarm rates (a fraction of false positives over negatives). The detected results are correct if the pixel distance between centers of the ground-truth and detected results is less than a given threshold (e.g., 5 pixels [12], 4 pixels [3], and 3 pixels [2]). The smaller threshold chosen means the lower detection errors and less workload in IRST applications. Hence, we choose the threshold as 3 pixels.

We provide ROC curves obtained using the LCM, LMWIE, AGADM, MME, MED, and MFMM algorithms for the 15 real sequences in Fig. 9. The probability of detection and false alarm rate are based on (18). The ROC curves indicate that our method has better detection performance than the baseline methods. Especially for the real sequence 1, 2, 4–10, 12,
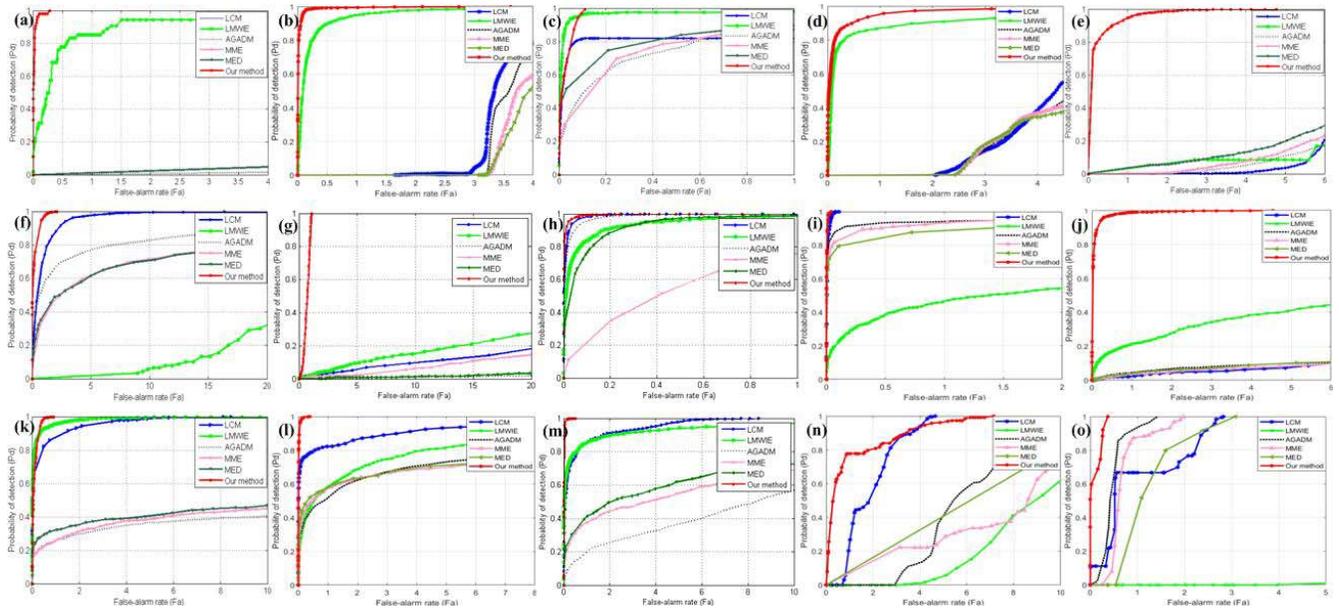
Fig. 9. (a)–(o) ROC curves of the 15 real sequences obtained using different methods.

13, and 15, the proposed method owns the highest Pd but lowest Fa values in comparisons. For the sequence 3 or 11, the LMWIE is slightly superior to the proposed method when $Fa \leq 0.06$ or $Fa \leq 0.05$. However, the MFMM method can reach 1 faster than that method when $Fa > 0.06$ or $Fa > 0.05$. For sequences 1–4, 7, 10, and 11, the LMWIE has better performance than other baseline methods. Fig. 9 also indicates that the LCM possesses higher Pd but lower Fa values, when compared with other baseline methods for sequence 6, 8, 9, and 11–13. The AGADM usually produces low Pd but high Fa values for the sequences except sequences 6 and 8–10. The comparisons derived from the ROC curves demonstrate that the proposed method is appropriate and robust to detect small targets against complex backgrounds.

D. Limitations

Qualitative and quantitative comparison results demonstrate that the proposed method is able to work stably and effectively for different complex backgrounds and target types. However, there are several limitations to this paper: first, the flexibility of the proposed method should be further explored in background cases with heavy disturbances, such as noisy bright spots and/or strong sea/cloudy clutters, although the proposed algorithm can tackle such conditions to some extent. This illustrates prospects in future study. Besides, theoretical justifications regarding the detection performance on highly difficult challenges/situations should be offered. Second, much attention should be paid to the determination of optimum window size because large window may give rise to the possibility of spatial nonuniformity in the background that causes some false detections. In this case, how to determine the optimum window/patch size that could handle multiple target sizes becomes a key issue in applications of the proposed algorithm. This points out a key for our future work. Late, the proposed method uses the multiple iterations scheme

to eliminate complex background clutters and noise step by step. Accordingly, a faster version of the current algorithm should be investigated. If the above are effectively handled, the potential of the proposed algorithm will be further strengthened.

V. DISCUSSION

In this section, we discuss the effects of the window size and stopping criterion involved in the proposed method. In addition, we discuss the noise sensitivity of the proposed algorithm.

A. Effects of Parameters

Some parameters should be pondered reasonably in the use of our method, e.g., the maximum size of window to compute MFM (6) and the stopping criterion to define number of iterations (8). The parameters can be experimentally /manually chosen. However, this is a time-consuming approach and difficult to reach the best detection result owing to criterion lack for quantitative evaluation. Alternatively, based on some reasonable assumptions, the proposed detection method can be simplified in applications.

The performance of the proposed algorithm depends on the choice of window size [viz., K in (4)]. If the window size is too small, the target is possibly masked by background clutters. This potentially yields high miss rates. On the other hand, if too large size is considered, the possibility of spatial nonuniformity in the background image may engender (that is, the assumption with regard to the spatial correlation in the background may not hold), which may cause many false detections with (6). To analyze the effects of the maximum window size in (6), we set the window size as 5×5, 7×7, 9×9, 11×11, 13×13, 15×15, and 17×17, respectively, and then, test the proposed algorithm. The results are shown in Fig. 10 and
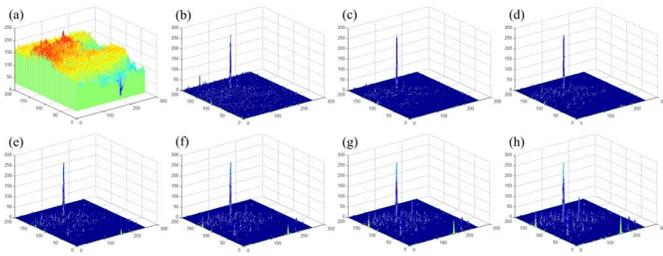
Fig. 10. 3-D gray distributions of noise-added image and filtered results obtained using the proposed method with different maximum window sizes. (a) Noise-added image of Fig. 5(m). (b)–(h) Filtered results through the proposed method with the maximum window size of 5×5, 7×7, 9×9, 11×11, 13×13, 15×15, and 17×17, respectively.
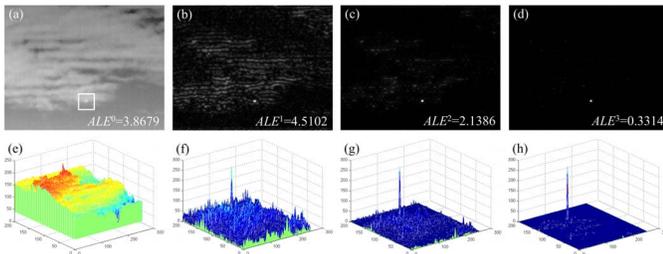
Fig. 11. Representative small target image and filtered results after different numbers of MFMM iterations, where $ALE^l (l = 0, 1, 2, 3)$ denotes the ALE after l iterations. (a) Original image. (b)–(d) Filtered results after one, two, or three MFMM iterations. (e)–(h) 3-D gray distributions of (a)–(d). The gray range of original and filtered images is normalized to [0,255].

Table VIII. Fig. 10 shows 3-D gray distributions of noise-added image [Fig. 5(m) is added a zero-mean Gaussian white noise, and the variance of noise is 0.001] and filtered results obtained using the proposed method with different maximum window sizes. Table VIII displays the number of iterations, ALE values, elapsed time, as well as SCR values of filtered results. The SCR of Fig. 10(a) is 2.03. It can be found from Fig. 10 and Table VIII that the choice of $K = 4$ can achieve a good comprise in SCR and elapsed time. In addition, a small target has a total spatial extent of less than 80 pixels [3], which is defined by Photo-Optical Instrumentation Engineers from the perspective of imaging. This suggests that the proper window size in (6) is no more than 4. Hence, the maximum window size in (6) is chosen as 9×9 (that is, K is 4) in the experiments. Extensive experimental results (see Section IV) suggest that the selection of parameter $K = 4$ can result in robust detection performance for different target types and sizes, and background types. Furthermore, the supplementary materials supplements some experiments of images with large target size.

Given prior information from a small target image (such as, a suitable fuzzy distance metric), target enhancement by multiple copies is the task of suppressing mazy background clutters and noise as much as possible. However, the large number of MFM iterations increases computational cost. Therefore, we perform some experiments to discuss the effects of stopping criterion, as shown in Fig. 11. A small target image randomly selected from sequence 13 is shown in Fig. 11(a), where the white rectangle denotes the location of target. The ALE value is 3.8679. After different

TABLE VIII
PERFORMANCE OF THE PROPOSED METHOD WITH DIFFERENT
MAXIMUM WINDOW/PATCH SIZES

	5×5	7×7	9×9	11×11	13×13	15×15	17×17
Num.	3	3	3	3	3	3	3
ALE^*	0.8810	0.4262	0.4984	0.5942	0.6966	0.7881	0.8595
Time	8.09	11.97	15.99	19.85	23.00	27.90	31.79
SCR	25.46	39.83	40.83	38.14	34.86	31.98	29.48

Note: Num. denotes the iteration stopping point based on the stopping criterion (8). ALE^* denotes the ALE value after Num. MFMM iterations.

TABLE IX
EVALUATION RESULTS OF NOISE-ADDED AND NOISE-REMOVED IMAGES

Group		# 1	# 2	# 3	# 4
Noise-added images					
SCR	Original	2.13	1.64	1.38	1.21
	MFMM	42.76	42.82	40.39	38.37
BSF		17.35	18.54	18.65	19.64
Noise-removed images					
SCR	Original	3.52	3.06	2.45	2.07
	MFMM	43.41	42.29	39.21	36.67
BSF		21.52	21.28	20.52	19.95

numbers of MFMM iterations, the filtered results are shown in Fig. 11(b)–(d), respectively. The corresponding ALE values are 4.5102, 2.1386, and 0.3314. Fig. 11(e)–(h) show the 3-D gray distributions of Fig. 11(a)–(d). It can be found that the complex background clutters and noise are distinctly attenuated after each MFMM operation (theoretical explanations resort to Propositions 4–6). Based on stopping criterion (8), the number of iterations is 3. In this case, the target is well enhanced and the clutters and noise are well eliminated [see Figs. 11(d) and (h)]. Fig. 11 indicates that the proposed stopping criterion (8) is a good compromise in practice.

B. Noise Sensitivity

To analyze the noise behavior of the proposed algorithm, we conduct a Monte-Carlo study with lots of Gaussian white noise realizations for each noise level. An image is randomly selected from the sequence 11, then Gaussian white noise with different variances is added to it (the noise mean is 0, and the variance is 0.002, 0.004, 0.006, and 0.008, respectively). In this case, four noise-added images are synthesized. The above work has been done 100 times, which synthesizes four groups of noise-added images with different noise levels (denoted as Group 1 to 4). Subsequently, we use block-matching and 3-D filtering (BM3D) [37] to denoise those noise-added images because BM3D can achieve good denoising performance. Then, four corresponding groups of noise-removed images are generated.

The SCR of original non-noise-added image is 3.65, and that becomes 44.19 for the corresponding filtered result through the proposed algorithm. Table IX displays the comparisons of the proposed algorithm on those noise-added and noise-removed images. It can be seen that the ensemble-average SCR values of four groups of noise-added images are 2.13, 1.64, 1.38, and 1.21, while that are 3.52, 3.06, 2.45, and 2.07 for the noise-removed images. After our method, the respective values are obviously improved, as displayed in Table IX. This means that the proposed method can significantly enhance the

targets as well as suppress the background clutters and noise. From the perspective of SCR values, there is little difference among the filtered results of non-noise-added images, noise-added images, and noise-removed images.

The BSF of non-noise-added image is 16.96. Table IX lists the ensemble-average BSF values of each group of noise-added and noise-removed images. Both the SCR and BSF evaluation results verify the noise insensitiveness of the proposed method. Moreover, the probabilities of detection of each group of noise-added and noise-removed sequences are 1, and the false alarm rates are 0. Therefore, the proposed method has good immunity to noise. More experimental results regarding noise sensitivity are provided in the supplementary materials.

VI. Conclusion

This paper presents an effective fuzzy distance/metric model called MFM for detecting small infrared target, according to the correlation properties of the background in spatial domain and noncorrelation properties between the target and background regions. Therefore, the target detection task is transformed into a measure issue of portraying such correlation/noncorrelation properties, which can be availably solved using the MFM that measures the certainty of targets in images. In this case, the fuzzy metric is close to 0 inside the background area, while within the area between the target and neighboring background, the fuzzy metric is greater than 0. Accordingly, the presented algorithm can eliminate substantial jamming factors, such as complex background clutters and noise. In particular, it significantly improves the SCR and BSF values of image. Thus, the presented algorithm ensures access to high probabilities of detection and low false alarm rates.

Experiments implemented on extensive clipped and real-data small infrared target images against diverse complicate cloudy- or sea-sky backgrounds (the number of clipped images is 105, and that of real images is 1345) demonstrate that the presented detection algorithm significantly outperforms several classical baseline methods, such as the LCM, LMWIE, AGADM, MME, and MED methods. The qualitative and quantitative results also demonstrate that the presented algorithm works more stably for different background types, SCR values, target types, and target sizes. In the future, we will investigate a faster version of the current method. We will also further improve the flexibility of our method in highly difficult background cases.

References

[1] C. Gao *et al.*, "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.

[2] H. Deng, X. Sun, M. Liu, C. Ye, and X. Zhou, "Infrared small-target detection using multiscale gray difference weighted image entropy," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 1, pp. 60–72, Feb. 2016.

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

[4] C. Gao, L. Wang, Y. Xiao, Q. Zhao, and D. Meng, "Infrared small-dim target detection based on Markov random field guided noise modeling," *Pattern Recognit.*, vol. 76, pp. 463–475, Apr. 2018.

[5] B. Porat and B. Friedlander, "A frequency domain algorithm for multiframe detection and estimation of dim targets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 4, pp. 398–401, Apr. 1990.

[6] I. S. Reed, R. M. Gagliardi, and L. B. Stotts, "Optical moving target detection with 3-D matched filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-24, no. 4, pp. 327–336, Jul. 1988.

[7] M. Li, T. Zhang, W. Yang, and X. Sun, "Moving weak point target detection and estimation with three-dimensional double directional filter in IR cluttered background," *Opt. Eng.*, vol. 44, no. 10, Oct. 2005, Art. no. 107007.

[8] Z. Wang, J. Tian, J. Liu, and S. Zheng, "Small infrared target fusion detection based on support vector machines in the wavelet domain," *Opt. Eng.*, vol. 45, no. 7, Jul. 2006, Art. no. 076401.

[9] B. Zhang, T. Zhang, Z. Cao, and K. Zhang, "Fast new small-target detection algorithm based on a modified partial differential equation in infrared clutter," *Opt. Eng.*, vol. 46, no. 10, Oct. 2007, Art. no. 106401.

[10] X. Bai and Y. Bi, "Derivative entropy-based contrast measure for infrared small-target detection," *IEEE Trans. Geosci. Remote Sens.*, to be published.

[11] S. C. Pohlig, "Spatial–temporal detection of electro-optic moving targets," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 31, no. 2, pp. 608–616, Apr. 1995.

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

[13] Y. Wei, X. You, and H. Li, "Multiscale patch-based contrast measure for small infrared target detection," *Pattern Recognit.*, vol. 58, no. 10, pp. 216–226, Oct. 2016.

[14] 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.

[15] S. Kim and J. Lee, "Small Infrared target detection by region-adaptive clutter rejection for sea-based infrared search and track," Sensors (Basel), vol. 14, no. 7, pp. 13210–13242, Jul. 2014.

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

[17] M. V. Shirvaikar and M. M. Trivedi, "A neural network filter to detect small targets in high clutter backgrounds," *IEEE Trans. Neural Netw.*, vol. 6, no. 1, pp. 252–257, Jan. 1995.

[18] C. Chen, M.-Y. Liu, O. Tuzel, and J. Xiao, "R-CNN for small object detection," in *Proc. ACCV*, Taipei, Taiwan, 2016, pp. 214–230.

[19] A. Takeki *et al.*, "Combining deep features for object detection at various scales: Finding small birds in landscape images," *IPSJ Trans. Comput. Vis. Appl.*, vol. 8, no. 1, pp. 1–7, Dec. 2016.

[20] P.-L. Shui, D.-C. Li, and S.-W. Xu, "Tri-feature-based detection of floating small targets in sea clutter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 2, pp. 1416–1430, Apr. 2014.

[21] D. Song and D. Tao, "Biologically inspired feature manifold for scene classification," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 174–184, Jan. 2010.

[22] W. Li, Q. Du, and B. Zhang, "Combined sparse and collaborative representation for hyperspectral target detection," *Pattern Recognit.*, vol. 48, no. 12, pp. 3904–3916, Dec. 2015.

[23] G. Wang, T. Zhang, L. Wei, and N. Sang, "Efficient method for multiscale small target detection from a natural scene," *Opt. Eng.*, vol. 35, no. 3, pp. 761–768, Mar. 1996.

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

[25] X. Qu, H. Chen, and G. Peng, "Novel detection method for infrared small targets using weighted information entropy," *J. Syst. Eng. Electron.*, vol. 23, no. 6, pp. 838–842, Dec. 2012.

[26] P. Chiranjeevi and S. Sengupta, "Detection of moving objects using multi-channel kernel fuzzy correlogram based background subtraction," *IEEE Trans. Cybern.*, vol. 44, no. 6, pp. 870–881, Jun. 2014.

[27] X. Zhao, P. Shi, and X. Zheng, "Fuzzy adaptive control design and discretization for a class of nonlinear uncertain systems," *IEEE Trans. Cybern.*, vol. 46, no. 6, pp. 1476–1483, Jun. 2016.

[28] H.-C. Liu, J.-X. You, X.-Y. You, and Q. Su, "Fuzzy Petri nets using intuitionistic fuzzy sets and ordered weighted averaging operators," *IEEE Trans. Cybern.*, vol. 46, no. 8, pp. 1839–1850, Aug. 2016.

[29] S. Balla-Arabé, X. Gao, and B. Wang, "A fast and robust level set method for image segmentation using fuzzy clustering and lattice Boltzmann method," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 910–920, Jun. 2013.

[30] G. Liu, Y. Zhang, and A. Wang, "Incorporating adaptive local information into fuzzy clustering for image segmentation," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3990–4000, Nov. 2015.

[31] Y. S. Choi and R. Krishnapuram, "A robust approach to image enhancement based on fuzzy logic," *IEEE Trans. Image Process.*, vol. 6, no. 6, pp. 808–825, Jun. 1997.

[32] T. Chaira, "Intuitionistic fuzzy segmentation of medical images," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 6, pp. 1430–1436, Jun. 2010.

[33] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2008, ch. 2, pp. 35–103.

[34] Z. Liang, W. Liu, and R. Yao, "Contrast enhancement by nonlinear diffusion filtering," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 673–686, Feb. 2016.

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

[36] W. Zhang, M. Cong, and L. Wang, "Algorithms for optical weak small targets detection and tracking: Review," in *Proc. IEEE Int. Conf. Neural Netw. Signal Process.*, vol. 12. Nanjing, China, Dec. 2003, pp. 643–647.

[37] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

Xianping Sun received the B.S. degree from Peking University, Beijing, China, in 1980.

He is currently a Professor with the Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences, Beijing. His current research interests include cybernetics, and signal acquisition and processing.

He Deng received the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2011.

He is currently an Associate Professor with the Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences, Beijing, China. His current research interests include target detection and image processing.

Xin Zhou received the Ph.D. degree from the Chinese Academy of Sciences, Beijing, China, in 2004.

He is currently a Professor with the Wuhan Institute of Physics and Mathematics, Chinese Academy of Sciences. His current research interests include scientific developments, imaging acquisition, and data processing.