Reweighted Infrared Patch-Tensor Model With Both Nonlocal and Local Priors for Single-Frame Small Target Detection

Yimian Dai and Yiquan Wu

Abstract—Many state-of-the-art methods have been proposed for infrared small target detection. They work well on the images with homogeneous backgrounds and high-contrast targets. However, when facing highly heterogeneous backgrounds, they would not perform very well, mainly due to: 1) the existence of strong edges and other interfering components, 2) not utilizing the priors fully. Inspired by this, we propose a novel method to exploit both local and nonlocal priors simultaneously. First, we employ a new infrared patch-tensor (IPT) model to represent the image and preserve its spatial correlations. Exploiting the target sparse prior and background nonlocal self-correlation prior, the target-background separation is modeled as a robust low-rank tensor recovery problem. Moreover, with the help of the structure tensor and reweighted idea, we design an entrywise local-structure-adaptive and sparsity enhancing weight to replace the globally constant weighting parameter. The decomposition could be achieved via the elementwise reweighted higher order robust principal component analysis with an additional convergence condition according to the practical situation of target detection. Extensive experiments demonstrate that our model outperforms the other state-of-the-arts, in particular for the images with very dim targets and heavy clutters.

Index Terms—Infrared patch-tensor model, infrared small target detection, local structure prior, nonlocal self-correlation prior, reweighted higher order robust principal component analysis.

I. INTRODUCTION

I NFRARED small target detection is a key technique for many applications, including early-warning system, precision guided weapon, missile tracking system, and maritime surveillance system [1]–[3]. Traditional sequential detection methods, such as three-dimensional (3-D) matched filter [4], improved 3-D filter [5], and multiscan adaptive matched filter [6],

are workable in the case of static background, exploiting the target spatial-temporal information. Nevertheless, with the recent fast development of high-speed aircrafts [7] like antiship missiles, the imaging backgrounds generally change quickly due to rapid relative motion between the imaging sensor and the target. The performance of the spatial-temporal detection method degrades rapidly. Therefore, the research of single-frame infrared small target detection is of great importance and has attracted a lot of attention in recent years.

Different from general object or saliency detection tasks, the main challenge of infrared small target detection is lacking enough information. Due to the long imaging distance, the target is always small without any other texture or shape features. As the target type, imaging distance, and neighboring environment differ a lot in real scenes, the target brightness could vary from extremely dim to very bright (see Fig. 5, for example). In the absence of spatial-temporal information and the target features like shape and size, the characteristics of the background [8] and the relation between the background and target are very important priors for single-frame infrared small target detection. Thus, how to design a model to incorporate and exploit these priors is vital for infrared small target detection in a single image.

A. Prior Work on Single-Frame Infrared Small Target Detection

The previously proposed single-frame infrared small target detection methods could be roughly classified into two categories. In the first type, a local background consistent prior is exploited, assuming the background is slowly transitional and nearby pixels are highly correlated. As a result, the target is viewed as the one that breaks this local correlation. Under this assumption, the classical methods, such as 2-D least mean square (TDLMS) filter [9] and Max-Median filter [10], enhance the small target by subtracting the predicted background from the original image. Unfortunately, besides the targets, they enhance the edges of the sky-sea surface or heavy cloud clutter as well, since these structures also break the background consistency as the target does. To differentiate the real target and high-frequency change, some edge analysis approaches [11], [12] have been proposed to extend these methods to estimate the edge direction in advance and preserve the edges. Bai *et al.* [13] designed a new Top-Hat transformation using

Y. Dai is with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: yimian.dai@gmail.com).

Y. Wu is with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China, and also with Key Laboratory of Spectral Imaging Technology CAS, Xi'an Institute of Optics and Precision Mechanics of CAS, Xi'an 710000, China (e-mail: nuaaimagestrong@163.com).

two different but correlated structuring elements. Another class of local prior-based methods exploits the local contrast, which is computed by comparing a pixel or a region only with its neighbors. The seminal work of Laplacian of Gaussian (LoG) filter-based method [14] has motivated a broad range of studies on the Human Visual System (HVS), and has led to a series of HVS-based methods, e.g., difference of Gaussians (DoG) [15], second-order directional derivative (SODD) filter [16], local contrast measure (LCM) [17], improved local contrast measure (ILCM) [18], multiscale patch-based contrast measure (MPCM) [19], multiscale gray difference weighted image entropy [20], improved difference of Gabors (IDoGb) [21], local saliency map (LSM) [22], weighted local difference measure (WLDM) [23], local difference measure (LDM) [24], etc.

The second type of single-frame infrared small target detection methods which has not been explored extensively, exploits the nonlocal self-correlation property of background patches, assuming that all background patches come from a single subspace or a mixture of low-rank subspace clusters. Then, target-background separation can be realized with the low-rank matrix recovery [25]. Essentially, this type of methods attempts to model the infrared small target as an outlier in the input data. To this end, Gao *et al.* [26] generalized the traditional infrared image model to a new infrared patch-image model via local patch construction. Then, the target-background separation problem is reformulated as a robust principal component analysis (RPCA) [27] problem of recovering low-rank and sparse matrices, achieving a state-of-the-art background suppressing performance. To correctly detect the small target located in a highly heterogeneous background, He *et al.* [28] proposed a low-rank and sparse representation model under the multisubspace-cluster assumption.

B. Motivation

Existing methods detect the infrared small target by either utilizing the local pixel correlation or exploiting the nonlocal patch self-correlation. From our observation, the unsatisfying performance of local prior methods [20], [23] in detecting the dim target under complicated background mainly lies in their imperfect grayscale-based center-difference measures. The saliency of a dim but true target would be easily overwhelmed by the measured saliency of some rare structures. We call this phenomenon the *rare structure effect*. In contrast, the nonlocal prior methods [26], [29], [30] suffer from the salient edge residuals. Its intrinsic reason is because the strong edge is also a sparse component as the same as the target due to lack of sufficient similar samples. Since the target might be dimmer than the edge, they would simultaneously be wiped out by simply increasing the threshold.

Our key observation is that the nonlocal prior and local prior are not equivalent, and in fact they are complementary for the problem of infrared small target detection, as illustrated in Fig. 1. The often appearing false alarm components in local (nonlocal) prior methods could be well suppressed by the nonlocal (local) prior methods. For example, the stubborn strong edges in the nonlocal prior-based methods can be easily identified by the
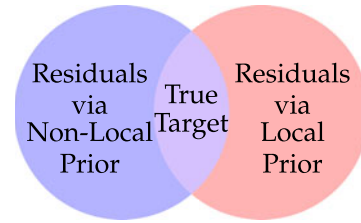

Fig. 1. Illustration of our motivation behind exploiting both nonlocal and local priors.

local edge analysis approaches. Naturally, an intuitive way to solve abovementioned dilemma is to extract the local structure information and merge it into the nonlocal prior-based detection framework. Therefore, how to simultaneously and appropriately utilize both the local and nonlocal priors has been an important issue for improving the detection performance under very complex backgrounds. To the best of our knowledge, very few single-frame infrared small target detection methods concern this problem.

To address this issue, we propose a single-frame small target detection framework with reweighted infrared patch-tensor model (RIPT). Our main contributions consist of the following three folds:

1) To dig out more information from the nonlocal self-correlationship in patch space, we generalize the patch-image model to a novel infrared patch-tensor model (IPT) and formulate the target-background separation task as an optimization problem of recovering low-rank and sparse tensors.
2) To incorporate the local structure prior into the IPT model, an elementwise weight is designed based on structure tensor, which helps us to suppress the remaining edges and preserve the dim target simultaneously.
3) To reduce the computing time, we adopt a reweighted scheme to enhance the sparsity of the target patch-tensor. Considering the particularity of infrared small target detection, an additional stopping criterion is used to avoid excessive computation.

The proposed RIPT model is validated on different real infrared image datasets. Compared with the state-of-the-art methods, our proposed model achieves a better background suppressing and target detection performance.

The remaining of this paper is organized as follows. We propose the nonlocal correlation-based IPT model in Section II. The details of the local structure weight construction are described in Section III. In Section IV, we further propose the reweighted IPT model and its detailed optimization scheme is also provided. Section V presents detailed experimental results and discussion. Finally we conclude this paper in Section VI.

II. NONLOCAL CORRELATION DRIVEN INFRARED PATCH-TENSOR MODEL

To dig out more spatial correlationships, we develop a novel target-background separation framework named infrared

TABLE I
MATHEMATICAL NOTATIONS

Notation	Explanation
$\mathcal{X}, \boldsymbol{X}, \boldsymbol{x}, x$	tensor, matrix, vector, scalar.
$\boldsymbol{X}_{(n)}$	mode-n matricization of tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_N}$, obtained by arranging the mode-n fibers as the columns of the resulting matrix of size $\mathbb{R}^{I_n \times \prod_{k \neq n} I_k}$.
$\mathrm{vec}(\mathcal{X})$	vectorization of tensor $\mathcal{X}$.
$\langle \mathcal{X}, \mathcal{Y} \rangle$	inner product of tensor $\mathcal{X}$ and $\mathcal{Y}$, which is defined as $\langle \mathcal{X}, \mathcal{Y} \rangle := \mathrm{vec}(\mathcal{X})^\top \mathrm{vec}(\mathcal{Y})$.
$\|\mathcal{X}\|_0$	ℓ_0 norm of tensor $\mathcal{X}$ which counts the number of non-zero elements.
$\|\mathcal{X}\|_1$	ℓ_1 norm of tensor $\mathcal{X}$.
$\|\mathcal{X}\|_\mathrm{F}$	Frobenius norm of tensor $\mathcal{X}$, which is defined by $\|\mathcal{X}\|_\mathrm{F} := \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}$.
$\mathrm{fold}_i(\boldsymbol{X})$	returns tensor $\mathcal{Z}$ that $\boldsymbol{Z}_{(i)} = \boldsymbol{X}$.
$\|\boldsymbol{X}\|_*$	nuclear norm of matrix $\boldsymbol{X}$, which is defined by $\|\boldsymbol{X}\|_* = \sum_i \boldsymbol{\sigma}_i$, where the SVD of $\boldsymbol{X} = \boldsymbol{U}\mathrm{diag}(\boldsymbol{\sigma})\boldsymbol{V}^\top$.
$\mathcal{S}_\mu(\boldsymbol{x})$	element-wise shrinkage operator is defined as $\mathcal{S}_\mu(\boldsymbol{x}) := \mathrm{sign}(\boldsymbol{x}) \max(|\boldsymbol{x}| - \mu, 0)$. $\mathcal{S}_\mu(\boldsymbol{x})$ is the closed-form solution of the problem: $\hat{\boldsymbol{y}} = \arg\min_{\boldsymbol{y}} \|\boldsymbol{x} - \boldsymbol{y}\|_\mathrm{F}^2 + \lambda\|\boldsymbol{y}\|_1$ [31].
$\mathcal{D}_\mu(\boldsymbol{X})$	matrix singular value thresholding operator: $\mathcal{D}_\mu(\boldsymbol{X}) := \boldsymbol{U}\mathrm{diag}(\bar{\sigma})\boldsymbol{V}^\top$, where $\boldsymbol{X} = \boldsymbol{U}\mathrm{diag}(\sigma)\boldsymbol{V}^\top$ is the SVD of $\boldsymbol{X}$ and $\bar{\sigma} := \max(\sigma - \mu, 0)$. $\mathcal{D}_\mu(\boldsymbol{X})$ is the closed-form solution of the problem: $\hat{\boldsymbol{Y}} = \arg\min_{\boldsymbol{Y}} \|\boldsymbol{X} - \boldsymbol{Y}\|_\mathrm{F}^2 + \frac{\mu}{2}\|\boldsymbol{Y}\|_*$ [32].
$\mathcal{T}_{i,\mu}(\mathcal{X})$	$\mathcal{T}_{i,\mu}(\mathcal{X}) := \mathrm{fold}_i(\mathcal{D}_\mu(\boldsymbol{X}_{(i)}))$.

patch-tensor model in this section. Before describing the details, several mathematical notations are defined first in Table I.

Given an infrared image, it could be modeled as a linear superposition of target image, background image, and noise image:

$$f_\mathrm{F} = f_\mathrm{B} + f_\mathrm{T} + f_\mathrm{N} \qquad (1)$$

where f_F, f_B, f_T, and f_N represent the input image, background image, target image, and noise image, respectively. Via a window sliding from the top left to the bottom right over an image, we stack the patches into a 3-D cube (see the construction step in Fig. 4). Then, (1) is transferred to the patch space with spatial structure preserved:

$$\mathcal{F} = \mathcal{B} + \mathcal{T} + \mathcal{N} \qquad (2)$$

where $\mathcal{F}, \mathcal{B}, \mathcal{T}, \mathcal{N} \in \mathbb{R}^{I \times J \times P}$ are the input patch-tensor, background patch-tensor, and target patch-tensor, respectively. I and J are the patch height and width, P is the patch number.

Background patch-tensor $\mathcal{B}$: Generally, the background is considered as slowly transitional, which implies high correlations among both the local and nonlocal patches in the image, as illustrated in Fig. 2(a). Although patches p_1, p_2, p_3 locate in the different region of the image, they are equivalent. Based on this nonlocal correlationship, the state-of-the-art IPI model imposed the low-rank constraint to background patch-image. As a patch-image is the mode-3 unfolding matrix of a patch-tensor, the patch-image model could be viewed as a special case of the proposed patch-tensor model essentially. Since the main difficulty of detecting the infrared small target in a single image is lacking enough information, only considering the low-rank structure in one unfolding is insufficient to deal with highly difficult scenes. Naturally, it motivates us to think whether we can utilize the other two unfolding modes. Actually, the mode-1 and mode-2 unfolding matrices of the infrared patch-tensor are also low-rank. In Fig. 2(b)–(d), the singular values of all the unfolding matrices rapidly decrease to zero, which demonstrates that every unfolding mode of the background patch-tensor is intrinsically low-rank. Therefore, we can consider the background patch-tensor $\mathcal{B}$ as a low-rank tensor, and their unfolding matrices are
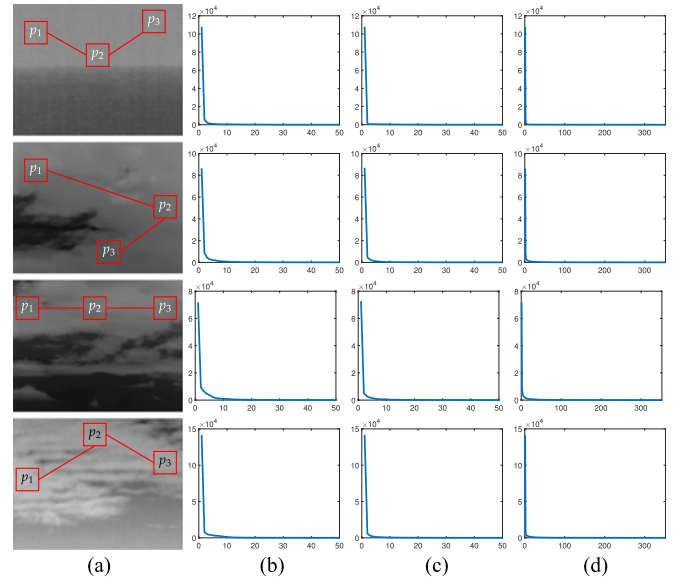

Fig. 2. Illustration of the nonlocal similarity and the low-rank property of the mode-i ($i = 1, 2, 3$) unfolding of the patch-tensor. (a) Four representative background images. (b)–(d) Singular values of the mode-1, mode-2, and mode-3 unfolding matrices of the corresponding background patch-tensors.

also all low-rank defined as:

$$\mathrm{rank}(\boldsymbol{B}_{(1)}) \leq r_1, \ \mathrm{rank}(\boldsymbol{B}_{(2)}) \leq r_2, \ \mathrm{rank}(\boldsymbol{B}_{(3)}) \leq r_3 \qquad (3)$$

where r_1, r_2, and r_3 all are constants, representing the complexity of the background image. The larger their values are, the more complex the background is.

Target patch-tensor $\mathcal{T}$: Since the small target merely occupies several pixels in the whole image, the target patch-tensor is an extremely sparse tensor in fact. That is:

$$\|\mathcal{T}\|_0 \leq k \qquad (4)$$

where k is a small integer determined by the number and size of the small target.

Noise patch-tensor $\mathcal{N}$: In this paper, we just assume that the noise is additive white Gaussian noise and $\|\mathcal{N}\|_{\mathrm{F}} \leq \delta$ for some $\delta > 0$. Thus, we have

$$\|\mathcal{F} - \mathcal{B} - \mathcal{T}\|_{\mathrm{F}} \leq \delta. \qquad (5)$$

It should be noted that although the values of parameters k, r, δ are different depending on the images, we would not directly use them in the following sections.

Ideally, we would like to obtain a low-rank and sparse decomposition and solve the following problem:

$$\min_{\mathcal{B},\mathcal{T}} \operatorname{rank}(\boldsymbol{B}) + \lambda\|\mathcal{T}\|_0, \text{ s.t. } \mathcal{B} + \mathcal{T} = \mathcal{F}. \qquad (6)$$

Unfortunately, the rank computation of a given tensor is a NP-hard problem in general [33]. In [34], Goldfarb and Qin proposed the higher order RPCA (robust tensor recovery) through replacing the rank by a convex surrogate Tucker-rank CTrank($\mathcal{B}$), and $\|\mathcal{T}\|_0$ by $\|\mathcal{T}\|_1$ to make the above problem tractable. In the singleton model, the tensor rank regularization term is defined as the sum of all the nuclear norms of the mode-i unfoldings, i.e., CTrank($\mathcal{B}$) $= \sum_i \|\boldsymbol{B}_{(i)}\|_*$, $i = 1, 2, 3$. With this relaxation, our proposed IPT model with random noise assumption can be solved by minimizing the following convex problem:

$$\min_{\mathcal{B},\mathcal{T}} \sum_{i=1}^3 \|\boldsymbol{B}_{(i)}\|_* + \lambda\|\mathcal{T}\|_1, \text{ s.t. } \|\mathcal{F} - \mathcal{B} - \mathcal{T}\|_{\mathrm{F}} \leq \delta. \qquad (7)$$

λ is a weighting parameter that controls the global tradeoff between the background patch-tensor and the target patch-tensor. Larger λ can shrink those nontarget but sparse components to zeros in the target patch-tensor. Nevertheless, it will also overshrink the dim target which should be preserved. On the contrary, a smaller λ does retain the dim target, but it keeps the strong cloud edges as well. Therefore, adopting a global constant weighting parameter λ is not a reasonable scheme for detecting the infrared small target in a complex scene. Naturally, it motivates us to design an entrywise weighting scheme.

III. INCORPORATING LOCAL STRUCTURE PRIOR

In this section, we focus our emphasis on combining the local structure prior and nonlocal correlation prior together into our model. We construct a local structure weight and interpret it as an edge salience measure. For the sake of simplicity, the local structure weight is designed on the basis of the image structure tensor. Structure tensor is widely used in many partial differential equation (PDE)-based methods [35], [36] to estimate the local structure information in the image, including edge orientation. To integrate the local information, the structure tensor is constructed based on a local regularization of a tensorial product, which is defined as

$$\boldsymbol{J}_\alpha(\nabla u_\sigma) = G_\alpha * (\nabla \boldsymbol{u}_\sigma \otimes \nabla \boldsymbol{u}_\sigma) = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} \qquad (8)$$

where $\boldsymbol{u}_\sigma$ is a Gaussian-smoothed version of a given image $\boldsymbol{u}$. $\sigma > 0$ is the standard deviation of the Gaussian kernel which denotes the noise scale, making the edge
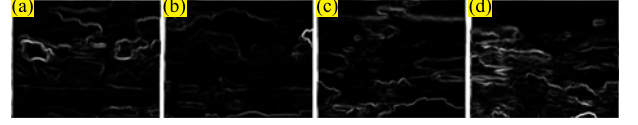

Fig. 3. Illustrations of the structure tensor-based local structure weight map. (a)–(d) Illustrations of Fig. 5(a)–(d).

detector ignorant of small details. $\boldsymbol{J}_\alpha$ is a symmetric and positive semidefinite matrix, which has two orthonormal eigenvectors denoted as $\boldsymbol{w}$ and $\boldsymbol{v} = \boldsymbol{w}^\perp$, where $\boldsymbol{w} = (2J_{12}, J_{22} - J_{11} + \sqrt{(J_{22} - J_{11})^2 + 4J_{12}^2})^\top$, $\boldsymbol{w} = \boldsymbol{w}/|\boldsymbol{w}|$. $\boldsymbol{w}$ points to the maximum contrast direction of the geometry structure while $\boldsymbol{v}$ points to the minimum direction [37]. Their corresponding eigenvalues λ_1 and λ_2 can be calculated via

$$\lambda_1, \lambda_2 = (J_{11} + J_{22}) \pm \sqrt{(J_{22} - J_{11})^2 + 4J_{12}^2}. \qquad (9)$$

These two values can be used as two feature descriptors of the local geometry structure, where at the flat region, $\lambda_1 \approx \lambda_2 \approx 0$; at the edge region, $\lambda_1 \gg \lambda_2 \approx 0$; at the corner region, $\lambda_1 \geq \lambda_2 \gg 0$. For the sake of low computational cost, we take $\lambda_1 - \lambda_2$ as the edge awareness feature since its value of the edge pixel is much larger than that belongs to the flat region and corner. By applying (8) and (9) to every pixel in the input image f_{F}, two matrices $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ can be obtained, which consists of the large and small structure tensor eigenvalues of all the pixels, respectively. Then, we can transform $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$ to their corresponding patch-tensors $\mathcal{L}_1$ and $\mathcal{L}_2$. Finally, we define the local structure weight patch-tensor as follows:

$$\boldsymbol{\mathcal{W}}_{\mathrm{LS}} = \exp\left(h \cdot \frac{(\mathcal{L}_1 - \mathcal{L}_2) - d_{\min}}{d_{\max} - d_{\min}}\right) \qquad (10)$$

where h is a weight stretching parameter, $d_{\max}$ and $d_{\min}$ are the maximum and minimum of $\mathcal{L}_1 - \mathcal{L}_2$, respectively. Fig. 3 displays the edge awareness maps of Fig. 5, which demonstrates that the structure tensor-based local structure weight has a good performance in identifying the edges. It should be noticed that for the sake of displaying effect, Fig. 3 is created via a normalized version of $-\exp\left(-h \cdot \frac{(\mathcal{L}_1 - \mathcal{L}_2) - d_{\min}}{d_{\max} - d_{\min}}\right)$. In the proposed algorithm, we still calculate the local structure weight via (10).

With the help of $\boldsymbol{\mathcal{W}}_{\mathrm{LS}}$, we can rewrite (7) into a weighted IPT model:

$$\min_{\mathcal{B},\mathcal{T}} \sum_{i=1}^3 \|\boldsymbol{B}_{(i)}\|_* + \lambda\|\boldsymbol{\mathcal{W}}_{\mathrm{LS}} \odot \mathcal{T}\|_1, \text{ s.t. } \|\mathcal{F} - \mathcal{B} - \mathcal{T}\|_{\mathrm{F}} \leq \delta \qquad (11)$$

where $\odot$ is the notation of Hadamard product. Thanks to the weighted IPT model (11), the strong edges could be well suppressed in the target image.
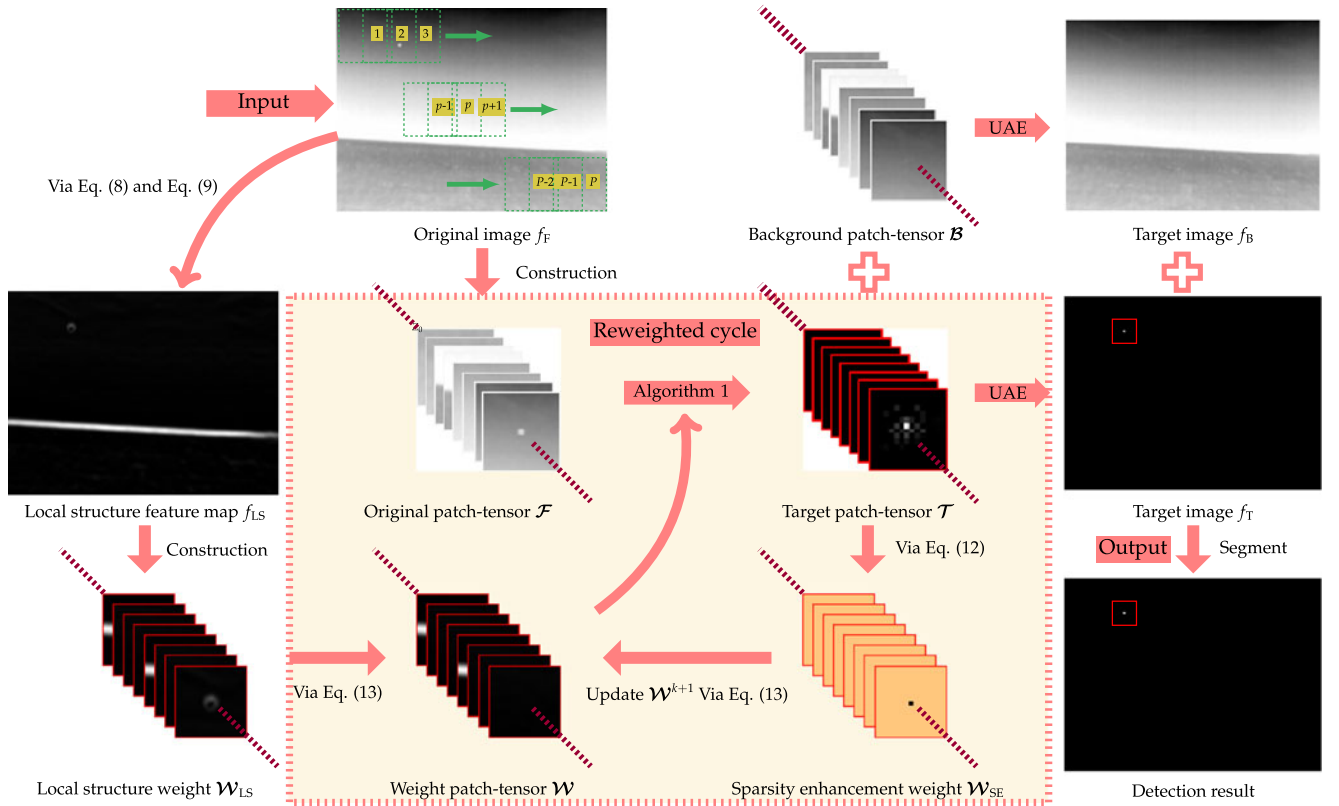
Fig. 4.   Overview of the proposed reweighted infrared patch-tensor model in this paper.
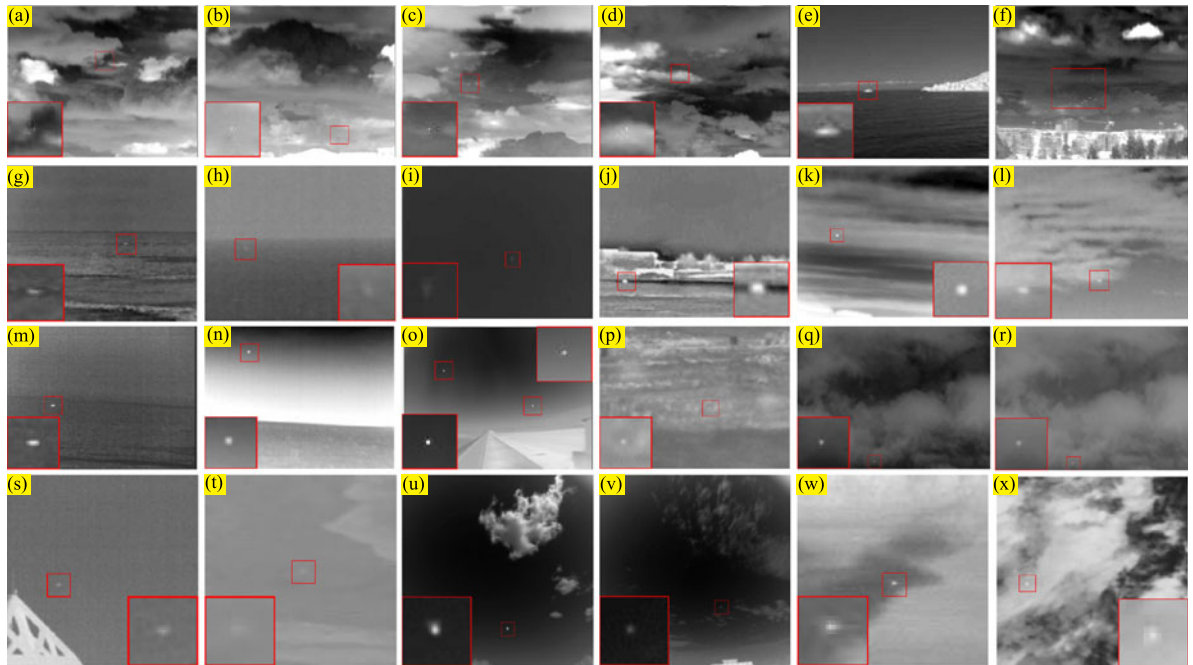

Fig. 5.   Representative infrared images for experiments. (a)–(x) Original infrared small target images against various backgrounds. For better visualization, the demarcated area is enlarged, which is better to be seen by zooming on a computer screen.

IV. REWEIGHTED INFRARED PATCH-TENSOR MODEL AND ITS SOLUTION

A. Reweighted Infrared Patch-Tensor Model

The computing time is also a major concern in infrared small target detection. Generally, the stopping criterion of a RPCA algorithm is that the reconstruction error is less than a certain small value. To meet this criterion, WIPT needs dozens of iterations, which is still time-consuming. An interesting phenomenon we find is that the nonzero entry number in the target patch-tensor has ceased to grow before the algorithm converges. In fact, in the target image, true target merely occupies the

brightest of the few. In the second half of the iteration, their values barely change. Therefore, considering the particularity of infrared small target, it is reasonable to replace the reconstruction error with the target patch-tensor sparsity in our proposed model. The algorithm stops iteration once the nonzero entry ceases to grow. Then, the sparsity of the target patch-tensor becomes critical in reducing the computing time. We hope the nonzero entries keep decreasing as the iteration goes on, leaving the final target image the sparsest. Unfortunately, the real situation is against our expectation in IPT and WIPT, where the target image deteriorates as the algorithm converges. Naturally, it motivates us to take a sparsity enhancing approach to solve this problem.

In [38], Candès proposed a reweighted ℓ_1 minimization for enhancing sparsity. By minimizing a sequence of weighted ℓ_1 norm, a significant performance improvement is obtained on sparse recovery. Inspired by it, many improved RPCA models have been proposed [39]–[42]. Motivated by these state-of-the-art models, we adopt a similar reweighted scheme for the values in the target patch-tensor. The large weights discourage nonzero entries, and the small weights preserve nonzero elements. The sparsity enhancement weight is defined as follows:

$$\mathcal{W}_{\mathrm{SE}}^{k+1}(i,j,p) = \frac{1}{|\mathcal{T}^k(i,j,p)| + \epsilon} \quad (12)$$

where $\epsilon > 0$ is a preset constant to avoid division by zero. Then, besides the relative error $\|\mathcal{F} - \mathcal{B}^{k+1} - \mathcal{T}^{k+1}\|_{\mathrm{F}}/\|\mathcal{F}\|_{\mathrm{F}} < \varepsilon$, we could add a new end condition that counts the nonzero entry element, namely $\|\mathcal{T}^{k+1}\|_0 = \|\mathcal{T}^k\|_0$. With the help of this empirical observation, the computing time could be largely decreased, as illustrated in Fig. 10 and Table IV.

Another intrinsic characteristic that both [26] and [29] neglect is the fact that the small target is always brighter than its neighborhood environment in infrared images due to the physical imaging mechanism [43]. Therefore, besides the sparsity constraint [44]–[46] of the target patch-tensor, it is reasonable to assume that all the entries in $\mathcal{T}$ are nonnegative. To this end, we incorporate this target nonnegative prior into the reweighted IPT model via rewriting (12) as follows:

$$\mathcal{W}_{\mathrm{SE}}^{k+1}(i,j,p) = \delta\left(\mathcal{T}^k(i,j,p)\right)$$
$$= \begin{cases} \frac{1}{\mathcal{T}^k(i,j,p)+\epsilon}, & \text{if } \mathcal{T}^k(i,j,p) > 0; \\ \infty, & \text{if } \mathcal{T}^k(i,j,p) \le 0 \end{cases} \quad (13)$$

where $\delta(\cdot)$ is an indicator function. We combine the local structure weight $\mathcal{W}_{\mathrm{LS}}$ and sparsity enhancing weight $\mathcal{W}_{\mathrm{SE}}^k$ to get the adaptive weight as follows:

$$\mathcal{W}^k = \mathcal{W}_{\mathrm{LS}} \odot \mathcal{W}_{\mathrm{SE}}^k. \quad (14)$$

Finally, we generalize the proposed IPT model and WIPT model to a novel reweighted infrared patch-tensor model (RIPT) as follows:

$$\min_{\mathcal{B},\mathcal{T}} \sum_{i=1}^3 \|\mathbf{B}_{(i)}\|_* + \lambda\|\mathcal{W} \odot \mathcal{T}\|_1, \text{ s.t. } \mathcal{B} + \mathcal{T} = \mathcal{F}. \quad (15)$$

B. Solution of RIPT Model

In this section, we show how to solve the proposed RIPT model as a reweighted robust tensor recovery problem via an Alternating Direction Method of Multipliers (ADMM) [47]. The augmented Lagrangian function of (15) is defined as

$$\mathcal{L} = \sum_{i=1}^N \|\mathbf{B}_{i,(i)}\|_* + \lambda\|\mathcal{W} \odot \mathcal{T}\|_1 +$$
$$\sum_{i=1}^N \frac{1}{2\mu}\|\mathcal{B}_i + \mathcal{T} - \mathcal{F}\|^2 - \langle\mathcal{Y}_i, \mathcal{B}_i + \mathcal{T} - \mathcal{F}\rangle \quad (16)$$

where $\mathcal{Y}_i \in \mathbb{R}^{I \times J \times P}$, $i = 1, 2, 3$ are the Lagrange multiplier tensors, and μ is a positive penalty scalar. ADMM decomposes the minimization of $\mathcal{L}$ into two subproblems that minimize $\mathcal{B}_i$ and $\mathcal{T}$, respectively. More specifically, the iterations of ADMM go as follows:

Updating $\mathcal{B}_i$ with the other terms fixed

$$\mathcal{B}_i^{k+1} = \arg\min_{\mathcal{B}_i} \|\mathbf{B}_{i,(i)}\|_* + \frac{1}{2\mu}\|\mathcal{B}_i - (\mathcal{F} + \mu\mathcal{Y}_i^k - \mathcal{T}^k)\|_{\mathrm{F}}^2. \quad (17)$$

Updating $\mathcal{T}$ with the other terms fixed

$$\mathcal{T}^{k+1} = \arg\min_{\mathcal{T}} \lambda\|\mathcal{W}^k \odot \mathcal{T}\|_1 +$$
$$\sum_{i=1}^N \frac{1}{2\mu}\|\mathcal{T} - (\mathcal{F} + \mu\mathcal{Y}_i^k - \mathcal{B}_i^{k+1})\|_{\mathrm{F}}^2. \quad (18)$$

Updating the multiplier $\mathcal{Y}_i^{k+1}$ with the other terms fixed

$$\mathcal{Y}_i^{k+1} = \mathcal{Y}_i^k + \frac{1}{\mu^k}\left(\mathcal{F} - \mathcal{B}_i^{k+1} - \mathcal{T}^{k+1}\right), i = 1, \ldots, N. \quad (19)$$

The subproblems (17) and (18) can be solved via the following two operators, respectively,

$$\mathcal{B}_i^{k+1} = \mathcal{T}_{i,\mu}\left(\mathcal{F} + \mu\mathcal{Y}_i^k - \mathcal{E}^k\right) \quad (20)$$
$$\mathcal{T}^{k+1} = \mathcal{S}_{\frac{\mu\lambda}{N}\mathcal{W}^k}\left[\frac{1}{N}\sum_{i=1}^N\left(\mathcal{F} + \mu\mathcal{Y}_i^k - \mathcal{B}_i^{k+1}\right)\right]. \quad (21)$$

From (21), it could be observed that the weighting parameter determines the soft-threshold, controlling the tradeoff between the target patch-tensor and background patch-tensor. Therefore, our elementwise adaptive weight tensor could simultaneously preserve the small target and suppress the strong edges. Finally, the solution of the proposed model is given in Algorithm 1.

C. Detection Procedure

In Fig. 4, we present the whole procedure of detecting the infrared small target via the model proposed in this paper. The detailed steps are as follows:
1) Given an infrared image f_{F}, its local structure feature map f_{LS} is calculated via (9).
2) The original patch-tensor $\mathcal{F}$ and local structure weight patch-tensor $\mathcal{W}_{\mathrm{LS}}$ are constructed from f_{F} and f_{LS}.

Algorithm 1: Target-Background Separation via RIPT model.

Input: $\mathcal{F}$, $\mathcal{W}_{\text{LS}}$, λ, ε, N
Initialize: $\mathcal{T}^0 := 0$; $\mathcal{B}_i^0 := \mathcal{F}$, $\mathcal{Y}_i^0 := 0$, $i := 1, \cdots, N$;
$\mathcal{W}_{\text{SE}}^0 := 1$, $\mathcal{W}^0 := \mathcal{W}_{\text{LS}} \odot \mathcal{W}_{\text{SE}}^0$; $\mu := 5 \cdot \text{std}(\text{vec}(\mathcal{F}))$,
$k := 0$

    while not converged **do**

        ▷ **Step 1:** Fix the others and update $\mathcal{B}_i$ by

        for $i = 1$ to N **do**

            $\mathcal{B}_i^{k+1} := \mathcal{T}_{i,\mu}\left(\mathcal{F} + \mu \mathcal{Y}_i^k - \mathcal{E}^k\right)$;

        end for

        ▷ **Step 2:** Fix the others and update $\mathcal{T}$ by

        $\mathcal{T}^{k+1} := \mathcal{S}_{\frac{\mu\lambda}{N}\mathcal{W}^k}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\mathcal{F} + \mu \mathcal{Y}_i^k - \mathcal{B}_i^{k+1}\right)\right]$;

        ▷ **Step 3:** Fix the others and update $\mathcal{Y}_i$ by

        for $i = 1$ to N **do**

            $\mathcal{Y}_i^{k+1} := \mathcal{Y}_i^k + \frac{1}{\mu^k}\left(\mathcal{F} - \mathcal{B}_i^{k+1} - \mathcal{T}^{k+1}\right)$;

        end for

        ▷ **Step 4:** Fix the others and update $\mathcal{W}$ by

        for $(i, j, p) \in [1, \ldots, I] \times [1, \ldots, J] \times [1, \ldots, P]$ **do**

            $\mathcal{W}_{\text{SE}}^{(k+1)}(i, j, p) := \delta\left(\mathcal{T}^k(i, j, p)\right)$;

        end for

        $\mathcal{W}^{k+1} := \mathcal{W}_{\text{LS}} \odot \mathcal{W}_{\text{SE}}^{k+1}$;

        ▷ **Step 5:** Update μ by

        $\mu^{k+1} := \mu^k / \rho$;

        ▷ **Step 6:** Check the convergence conditions

        $\frac{\|\mathcal{F} - \mathcal{B}^{k+1} - \mathcal{T}^{k+1}\|_{\text{F}}}{\|\mathcal{F}\|_{\text{F}}} < \varepsilon$ or $\left\|\mathcal{T}^{k+1}\right\|_0 = \left\|\mathcal{T}^k\right\|_0$;

        ▷ **Step 7:** Update k

        $k := k + 1$;

    end while
Output: $\frac{1}{N}\left(\sum_{i=1}^{N}\mathcal{B}_i^k\right)$, $\mathcal{T}^k$

---

3) Algorithm 1 is performed to decompose the patch-tensor $\mathcal{F}$ into the background patch-tensor $\mathcal{B}$ and target patch-tensor $\mathcal{T}$.

4) The background image f_{B} and target image f_{T} are reconstructed from the background patch-tensor $\mathcal{B}$ and target patch-tensor $\mathcal{T}$, respectively. For the sake of implementation convenience, we adopt the uniform average of estimators (UAE) reprojection scheme [48].

5) The target is segmented out as the same as [26]. The adaptive threshold is determined by

$$t_{\text{up}} = \max(v_{\min}, \bar{f}_{\text{T}} + k\sigma) \qquad (22)$$

where $\bar{f}_{\text{T}}$ and σ are the average and standard deviation of the target image f_{T}. k and $v_{\min}$ are constants determined empirically.

V. EXPERIMENTAL VALIDATION

To fully evaluate the proposed algorithm, we conduct a series of experiments using images of various scenarios and include ten state-of-the-art methods for comparison.

A. Experimental Setup

Datasets: We test the proposed model on extensive real infrared images to cover different scenarios, as illustrated in Fig. 5, varying from the flat backgrounds with salient targets to complex backgrounds with heavy clutters and extremely dim targets. All targets are labeled with red boxes. Since some targets are so dim that could hardly be observed by human eyes directly, we enlarge the demarcated area. Taking into account that the biggest difficulty of current infrared small target detection is how to detect those very dim targets with strong clutters, a good detection performance on those extremely complex images is more convincing than the satisfying result on relatively simple images. Therefore, in the following experiments, our main focus is put on the datasets with complex scenes that Fig. 5(a)–(d) and (l) belong to. The detailed characteristics of these five sequences are presented in Table II.

Baselines and Parameter Settings: The proposed algorithm is compared with ten state-of-the-art solutions, including three filtering-based methods (Max-Median [10], Top-Hat [49], TDLMS [9]), three HVS-based methods (PFT [50], MPCM [19], WLDM [23]), and four recently developed low-rank methods (IPI [26], PRPCA [51], WIPI [29], NIPPS [30]). Table III summarizes all the methods involved in the experiments and their detailed parameter settings. For all the low-rank methods, i.e., IPI, PRPCA, WIPI, NIPPS, IPT, and RPIT, they are all solved via ADMM. All the algorithms are implemented in MATLAB 2016b on a PC of 3.4 GHz and 4GB RAM. The source code of our method is publicly available at https://github.com/YimianDai/DENTIST.

Evaluation Metrics: For a comprehensive evaluation, four metrics including the local signal-to-noise ratio gain (LSNRG), background suppression factor (BSF), signal to clutter ratio gain (SCRG), and receiver operating characteristic (ROC) curve are adopted in comparison of background suppressing performance. LSNRG measures the local signal-to-noise ratio (LSNR) gain, which is defined as

$$\text{LSNRG} = \frac{\text{LSNR}_{\text{out}}}{\text{LSNR}_{\text{in}}} \qquad (23)$$

where LSNR_{in} and LSNR_{out} are the LSNR values before and after background suppression. LSNR is given as $\text{LSNR} = P_{\text{T}}/P_{\text{B}}$. P_{T} and P_{B} are the maximum grayscales of the target and neighborhood, respectively. BSF measures the background suppression quality using the standard deviation of the neighborhood region. It is defined as

$$\text{BSF} = \frac{\sigma_{\text{in}}}{\sigma_{\text{out}}} \qquad (24)$$

where σ_{in} and σ_{out} are the standard variances of background neighborhood before and after background suppression. The most widely used SCRG is defined as the ratio of signal-to-clutter ratio (SCR) before and after processing:

$$\text{SCRG} = \frac{\text{SCR}_{\text{out}}}{\text{SCR}_{\text{in}}} \qquad (25)$$

where SCR represents the difficulty of detecting the infrared small target, and it is defined by $\text{SCR} = |\mu_{\text{t}} - \mu_{\text{b}}|/\sigma_{\text{b}}$. μ_{t} is the

TABLE II
TARGET AND BACKGROUND CHARACTERISTICS OF REAL INFRARED IMAGE SEQUENCES

	# Frame	Image Resolution	Target Shape	Target characteristics	Background characteristics
Sequence 1–4	400	255×320	Gaussian	◇ Tiny, very dim with low contrast. ◇ Moves along the cloud edges or buried in the cloud.	◇ Strong undulant cloud backgrounds. ◇ Approximately noise-free.
Sequence 5	30	200×256	Rectangular	◇ Small due to long imaging distance. ◇ Brightness, contrast, and size vary a lot.	◇ Heavy cloudy-sky background clutters in uniform backgrounds with noise.

TABLE III
DETAILED PARAMETER SETTINGS FOR 12 TESTED METHODS

| No.	Methods	Acronyms	Parameter settings
1	Max-Median filter [10]	Max-Median	Support size: 5×5
2	Top-Hat method [49]	Top-Hat	Structure shape: square, structure size: 3×3
3	Phase spectrum of Fourier Transform [50]	PFT	Disk radius: 3
4	Multiscale Patch-based Contrast Measure [19]	MPCM	$N = 1, 3, \ldots, 9$
5	Weighted Local Difference Measure [23]	WLDM	$L = 4, m = 2, n = 2$
6	Two-Dimensional Least Mean Square filter [9]	TDLMS	Support size: 5×5, step size: $\mu = 5 \times 10^{-8}$
7	Infrared Patch-Image Model [26]	IPI	Patch size: 50×50, sliding step: 10, $\lambda = \frac{L}{\sqrt{\min(I,J,P)}}, L = 3, \varepsilon = 10^{-7}$
8	Patch-level RPCA method [51]	PRPCA	Patch size: 50×50, sliding step: 10, $\lambda = \frac{L}{\sqrt{\min(I,J,P)}}, L = 3, \varepsilon = 10^{-7}$
9	Weighted Infrared Patch-Tensor Model [29]	WIPI	Patch size: 51×51, sliding step: 10, smoothing parameter $h = 15, \varepsilon = 10^{-7}$
10	Non-negative IPI model via Partial Sum minimization of singular values [30]	NIPPS	Patch size: 50×50, sliding step: 10, $\lambda = \frac{L}{\sqrt{\min(I,J,P)}}, L = 2, r = 5 \times 10^{-3}$
11	Infrared Patch-Tensor Model	IPT	Patch size: 50×50, sliding step: 10, $\lambda = \frac{L}{\sqrt{\max(I,J,P)}}, L = 3$
12	Reweighted Infrared Patch-Tensor Model	RIPT	Patch size: 50×50, sliding step: 10, $\lambda = \frac{L}{\sqrt{\min(I,J,P)}}, L = 1, h = 10, \epsilon = 0.01, \varepsilon = 10^{-7}$
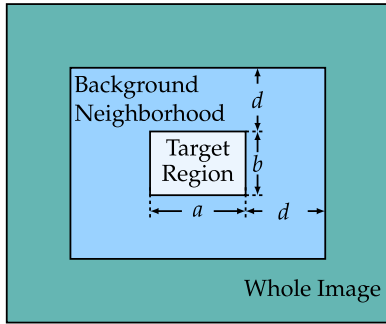
Fig. 6. Target and background neighborhood regions of a small target.

average target grayscale. μ_b and σ_b are the average grayscale and standard deviation of the neighborhood region. For all these three metrics, the higher their values are, the better background suppression performance the detection method has. All three metrics are computed in a local region, as illustrated in Fig. 6. The target size is $a \times b$, and d is the neighborhood width, we set $d = 20$ in this paper. Among all the existing metrics, the detection probability P_d and false-alarm rate F_a are the key performance indicators, which are defined as follows:

$$P_d = \frac{\text{number of true detections}}{\text{number of actual targets}} \quad (26)$$

$$F_a = \frac{\text{number of false detections}}{\text{number of images}}. \quad (27)$$

The ROC curve shows the tradeoff between the true detections and false detections.

B. Validation of the Proposed Method

In this ection, we take a closer look at the proposed method by validating their robustness against various scenes and noisy cases. At last, the roles of the patch-tensor, local structure weight, and sparsity enhancement weight are examined in depth to evaluate each prior individually.

1) Robustness to Various Scenes: In Fig. 7, we show the separated target images for the images of Fig. 5. Observing Fig. 7, it can be clearly seen that the background clutters are completely wiped out, leaving the target the only sole component in the target image. Since Fig. 5 contains a lot of different scenarios, it is fair to say that the proposed method is quite robust to various scenes.

2) Robustness to Noise: Noise is another key influence factor. In Fig. 8, we evaluate the proposed method's performance in the case of noise with different levels. When the noise standard deviation is 10, the proposed method could well enhance the targets and suppress the clutters and noise. As the noise standard deviation increases to 20, RIPT still detects the target in Fig. 8(m)–(n) and (q)–(r), but fails in Fig. 8(o)–(p). Nevertheless, this failure is acceptable, since the target is totally overwhelmed by the noise in Fig. 8(o)–(p) (see the enlarged box). Therefore, the noise influence depends not only on the intensity of the noise itself but also on the original contrast of the target. As long as the polluted target can maintain a weak contrast like Fig. 8(c)–(d), the proposed method is still able to detect it.

3) Roles of Components in the Proposed Model: To further understand the effects of the components in the proposed RIPT model, we evaluate each prior individually with exper-
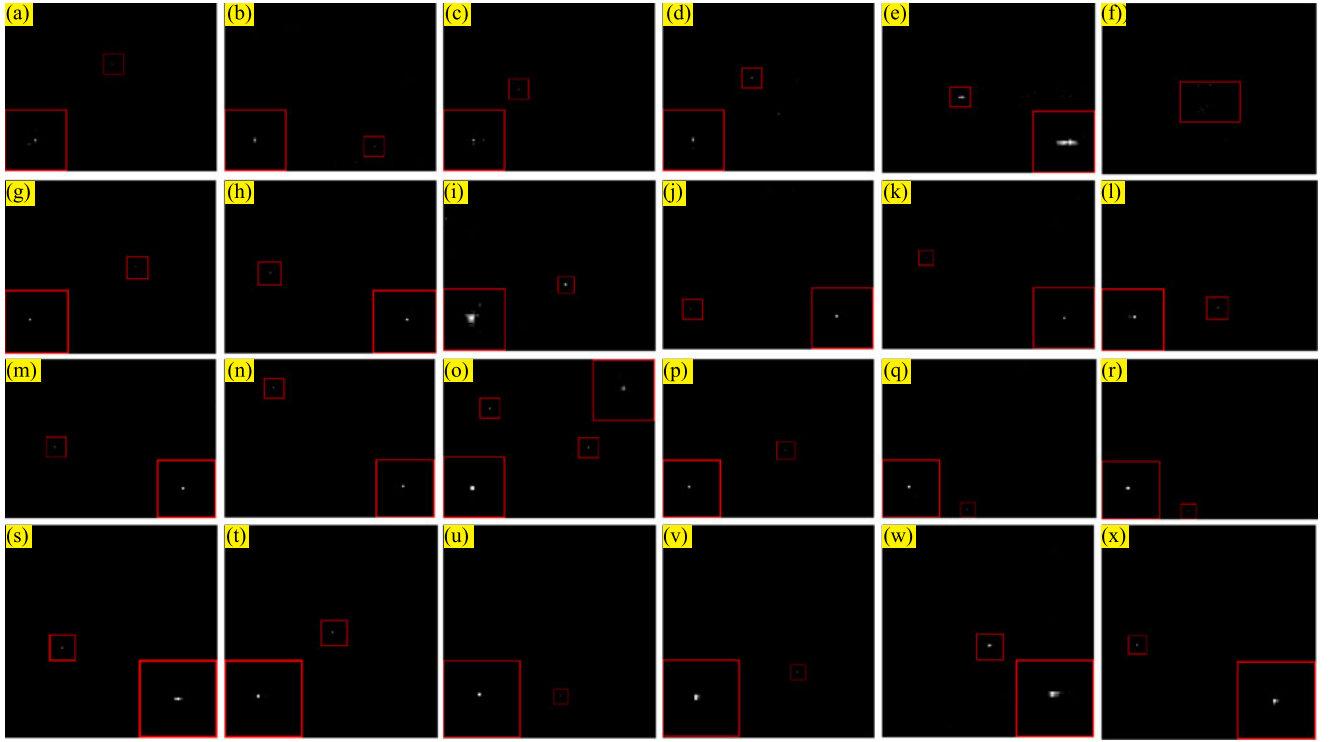
Fig. 7. Separated target images by the proposed RIPT model. (a)–(x) Separated target images of Fig. 5(a)–(x). For better visualization, the demarcated area is enlarged, which is better to be seen by zooming on a computer screen.
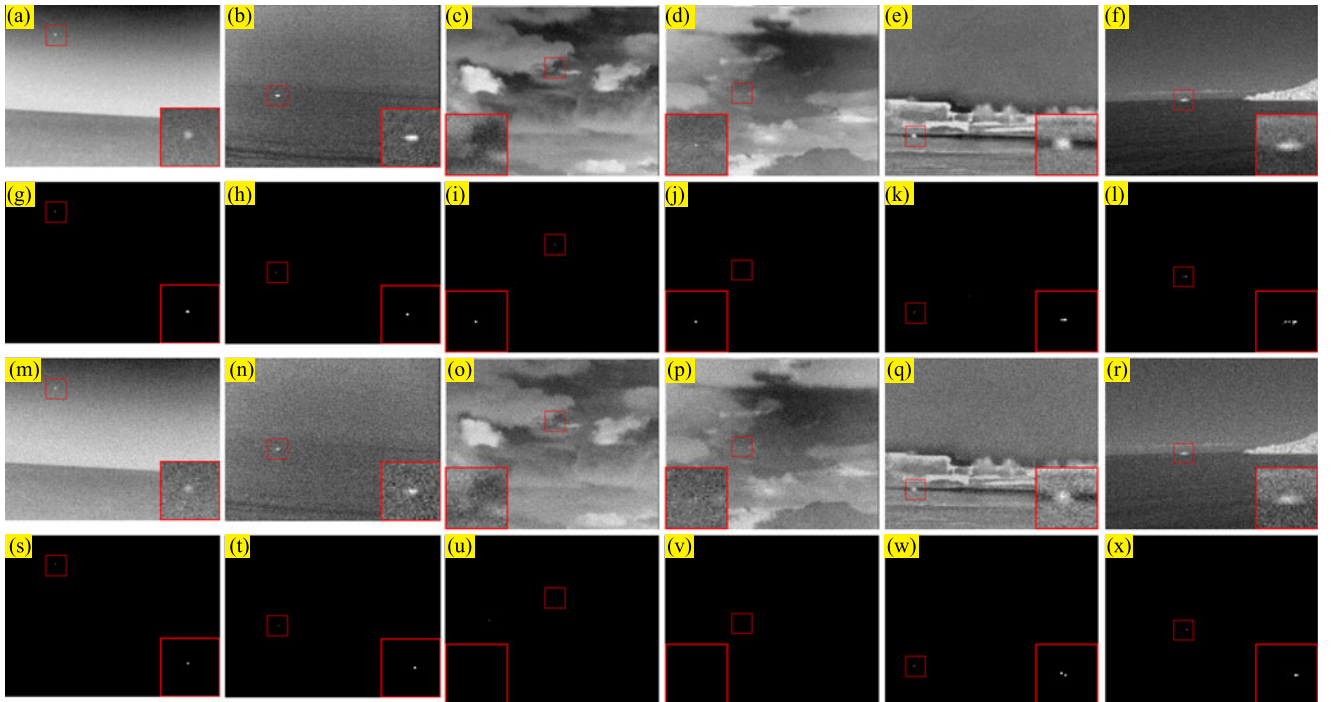

Fig. 8. First and third rows are images contaminated by additive white Gaussian noise with standard deviation of 10 and 20, and the second and fourth rows are the corresponding detection results by the proposed method.

iment to investigate how these priors influence the final detection performance. The ROC curves of IPI, IPT, IPT with sparsity enhancement weight (SIPT), WIPT, and RIPT for Sequence 1–4 are given in Fig. 9, leading to the following observations.

1) The four patch-tensor-based methods outperform the state-of-the-art IPI method, which demonstrates that the patch-tensor model, involving mode-1 and mode-2 unfolding matrices, does contribute to the final detection performance.
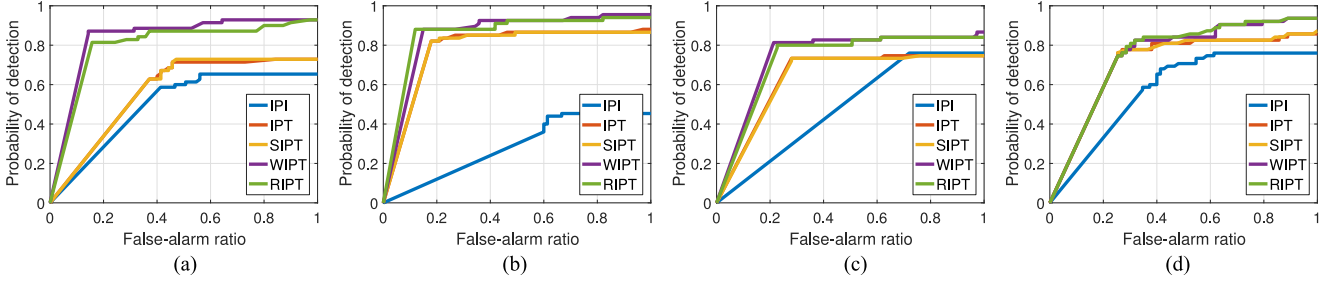
Fig. 9. Illustration for the effects of the introduced priors. (a)–(d) are the ROC curves on Sequences 1–4.
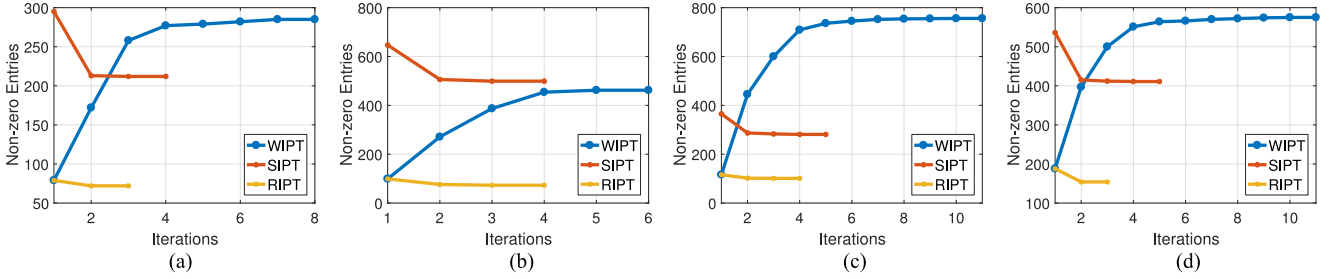

Fig. 10. Illustration for the effect of the sparsity enhancement weight: (a) Sequence 1, (b) Sequence 2, (c) Sequence 3, and (d) Sequence 4.

TABLE IV
ALGORITHM COMPLEXITY AND COMPUTATIONAL TIME COMPARISONS OF DIFFERENT METHODS

	TDLMS	PFT	MPCM	WLDM	IPI	WIPI	NIPPS	IPT	WIPT	RIPT
Complexity	$\mathcal{O}(L^2MN)$	$\mathcal{O}(MN\log MN)$	$\mathcal{O}(L^3MN)$	$\mathcal{O}(L^3MN)$	$\mathcal{O}(mn^2)$	$\mathcal{O}(mn^2)$	$\mathcal{O}(mn^2)$	$\mathcal{O}(mn^2)$	$\mathcal{O}(mn^2)$	$\mathcal{O}(mn^2)$
Time (s)	0.162	0.025	0.083	6.059	16.998	52.995	15.515	8.598	6.932	3.169

2) By comparing WIPT and RIPT with IPT and SIPT, we see that incorporating local structure prior significantly improves the detection probability.

3) Although the sparsity enhancement weight does not contribute to the final detection performance, it significantly reduces the iteration number, as illustrated in Fig. 10.

These observations indicate that the introduced priors are effective, and, when combined together, lead to excellent performance as reported in the next sections.

C. Algorithm Complexity and Computational Time

The proposed model is solved via ADMM, which has been proved a $\mathcal{O}(1/k)$ convergence [52]. Therefore, our solving algorithm is ensured to converge. The algorithm complexity and computational time for Fig. 5(a) with various methods are given in Table IV. The image size is $M \times N$, and m, n are the rows and columns of the patch-image or mode-3 unfolding. Although the computational complexities of these methods seem the same, their computing time differs a lot. For the filtering and HVS-based methods, the difference in computing time lies in whether the code could be vectorized. For the low-rank methods, the dominant factor is the iteration number. It can be seen from the data in Table IV that the low-rank methods

are generally slower than the filtering and HVS-based methods. Nevertheless, considering low-rank method could handle more difficult scenes, this tradeoff is acceptable. Among the low-rank methods, the RIPT costs the least time. The underlying reason is that both the local structure weight and sparsity enhancement weight help to reduce the iteration number. In addition, unlike the weight in WIPI, the time for constructing the weight is neglectable in RIPT.

D. Parameters Analysis

For the proposed model, the related parameters, such as the patch size, sliding step, weight stretching parameter h, weighting parameter λ, and penalty factor μ, are all important factors, which usually affects the model fitness on the real databases. Therefore, a better performance can be obtained by finely tuning these parameters. Nevertheless, the optimal values are always related to the infrared image content. In Fig. 11, we give the ROC curves for different model parameters on Sequences 1–4 to evaluate their influence. These parameters are validated to obtain a local optimal value with other parameters fixed. The stepped shape of our ROC curves might seem a bit odd. It is because we have adopted a much larger weighting parameter λ than normal RPCA-based foreground-background separation
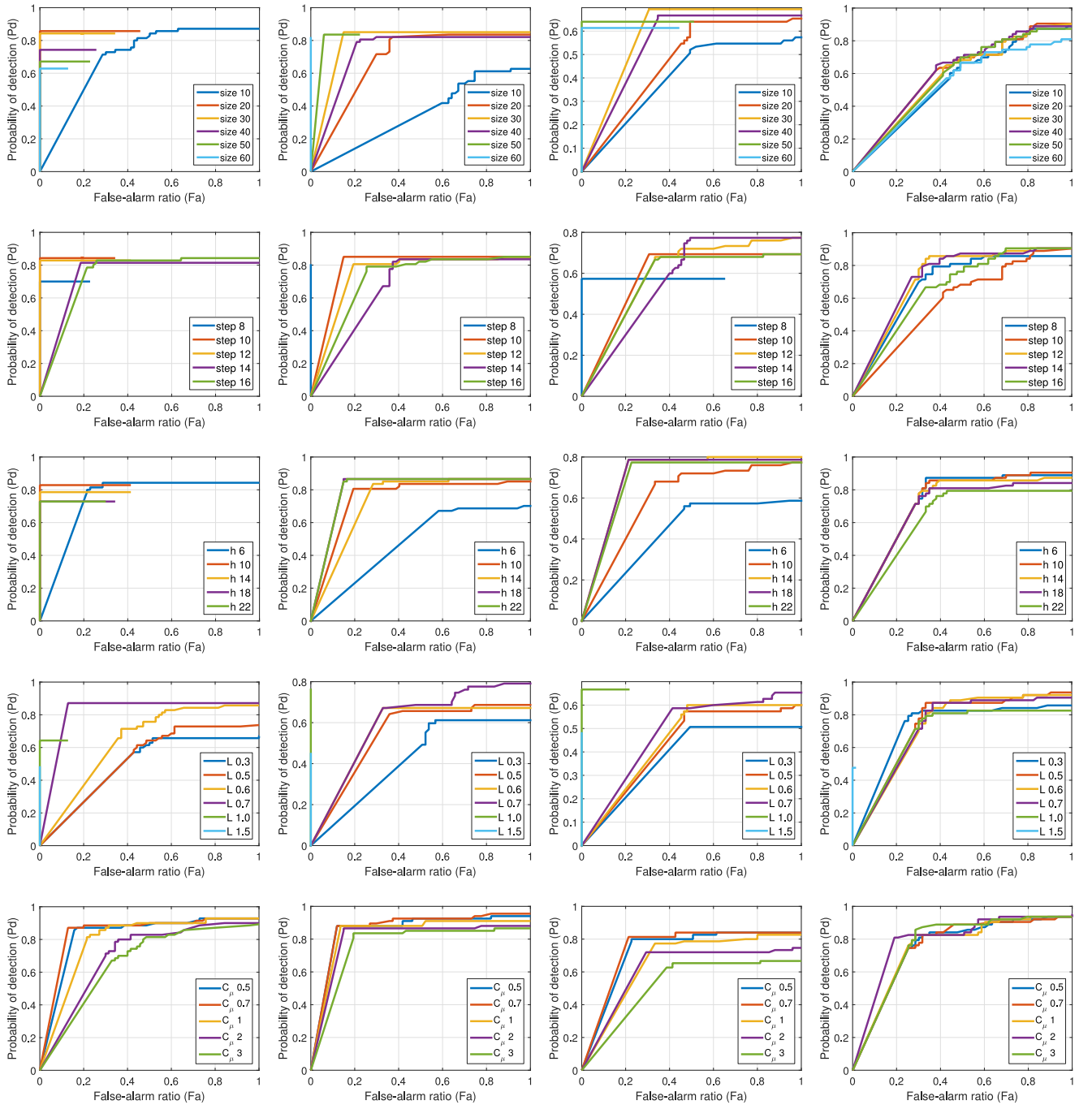
Fig. 11. ROC curves for sequences 1–4 with respect to different patch sizes, sliding steps h, weighting parameters and penalty factors. Row 1: Different patch sizes, Row 2: Different sliding steps, Row 3: Different h, Row 4: Different weighting parameters, Row 5: Different penalty factors.

tasks in order to better fit the actual situation of single-frame infrared small target detection.

1) Patch Size $I \times J$: It not only has a large impact on the separation, but also influences the computational complexity. The matrix size of mode-1 and mode-2 unfoldings of the patch-tensor is $I \times (J \cdot P)$; the matrix size of mode-3 unfolding is $J \times (I \cdot P)$. Obviously, a smaller patch size will lead to a smaller computational complexity. On one hand, we hope a larger patch size to ensure that the target is sparse enough. On the other hand, a larger patch size reduces the correlationships between

the nonlocal patches, which degrades the separation results. To seek a balance between a low computational burden, target sparsity, and background correlationship, we vary the patch size P from 10 to 60 with ten intervals and provide the ROC curves in the first row of Fig. 11. By observing the ROC curves, we can have the following conclusions. First, a smaller patch size is easier to raise false alarms while a larger patch size may lead to a relatively lower detection probability, which just demonstrates our above analysis about the patch size. Second, the proposed RIPT method is not very sensitive to the patch size. The detection

result of the patch size among 20–60 is acceptable. Third, 30 seems a good choice for Sequences 1–4 since it achieves the best performance in ROC.

2) Sliding Step: The sliding step influences the patch-tensor size as well. To reduce the computational complexity, we prefer a larger sliding step, which means smaller matrices to perform SVD. Nevertheless, a larger sliding step also reduces the redundancy of the original patch-tensor $\mathcal{T}$ and undermines the final detection results since our proposed model is based on the nonlocal repentance of correlated patches. To investigate its influence, we vary the sliding step S from 8 to 16 with two intervals. The results are displayed in the second row of Fig. 11. It could be observed that the ROC curve of small sliding step like 8 tends to have a more sharp shape, but its overall detection probability remains relatively low. The best value for sliding step is among 12 to 14, here we pick 12. In addition, by comparing the first row with the second row of Fig. 11, we can conclude that the algorithm is quite robust to the variation of step length.

3) Weight Stretching Parameter h: It controls the local structure weight's suppression degree to the clutter edges. We vary h from 6 to 22 in the experiment and illustrate the ROC curves in the third row of Fig. 11. Generally, we would like a larger h which suppresses the undesirable nontarget components thoroughly. Nevertheless, since the target-clutter distinguishing scheme is not always perfect, an overlarge h would also wipe out some targets. A typical example is the different performance of $h = 18$ or 22 among four sequences. For Sequences 2 and 3, $h = 18$ or 22 achieves the best performance, but they perform the worst for Sequences 1 and 4. It is because the target moves along the cloud edge in many frames of Sequences 1 and 4, and an overlarge h would easily mistake the target as the edge and suppress it completely, resulting in a lower detection probability. On the contrary, a smaller h might preserve the small target, but it also retains some nontarget components, making the false-alarm ratio relatively high. For Sequences 2 and 3, when the detection probability is fixed, the false-alarm ratio of $h = 6$ is the largest. In order to seek a balance, we set the optimal h as 10 in the following experiment.

4) Weighting Parameter λ: Despite the usage of local structure weight, fine tuning of λ is still of great importance. We show the effects of λ in the fourth row of Fig. 11. Since λ is set as $L/\sqrt{\min(I, J, P)}$ in our model, instead of directly varying λ, we vary L from 0.3 to 1.5. From the illustration, it can be observed that a large λ does keep the false-alarm ratio being quite low like. For example, the ROC curves of $L = 1.0$ and $L = 1.5$ for Sequence 2 are straight line segments, but their detection probabilities are also low because many dim targets are suppressed by the overlarge threshold. On the contrary, when the detection probability is fixed, the false-alarm ratio of $L = 0.3$ is always higher than the others, suggesting that a too small L is also not a good choice.

5) Penalty Factor μ: It is precisely the shrinking threshold of (21), which influences the low-rank property of the background patch-tensor. With a smaller μ, more details are preserved in the background patch-tensor. Thus, fewer nontarget components are left in the target patch-tensor. Nevertheless, the small target

might be preserved in the background patch-tensor as well, resulting no target in the target image. On the contrary, a larger μ would lead to more nontarget components lying in the target patch-tensor. Thus, it is necessary to choose an appropriate value for μ to keep the balance between detection probability and false-alarm ratio. Since we set $\mu = C_\mu \text{std}(\text{vec}(\mathcal{F}))$, instead of varying μ directly, we investigate the influence of the penalty factor on Sequences 1–4 by varying C_μ from 0.5 to 3. The results are shown in the last row of Fig. 11, from which we can observe that an overlarge or too small μ is not an optimal choice and the best value for our four sequences is about 0.7.

E. Comparison With State-of-the-Arts

In this section, we first compare the proposed model with the other state-of-the-art methods on the ability of clutter suppression. Figs. 12–15 show the separated target images by 12 tested methods for four representative frames of Sequences 1–4 in Fig. 5. It can be seen that the classical Max-Median filter does enhance the tiny targets in Figs. 12(b)–15(b). Nevertheless, many nontarget pixels are also enhanced simultaneously, especially in Figs. 13(b) and 14(b), which would raise many false alarms. In Figs. 12(a)–15(a) produced by TDLMS, the phenomenon of enhancing nontarget isolated points does not exist, but the cloud edges are highlighted, making them much brighter than the small target. Since the given target size matches the real target size exactly, the Top-Hat operator succeeds to enhance the target region. If not match, the Top-Hat operator is likely to lose the target. No matter whether the given and real target sizes match, Top-Hat cannot well suppress the background clutters. Many strong clutters still remain in resulting images, as illustrated in Figs. 12(c)–15(c). Although PFT can retain the target to a certain extent, the target is not necessarily the brightest and there are also many nontarget salient residuals, as shown in Figs. 12(d)–15(d). MPCM and WLDM failed to achieve good results because they suffered from a phenomenon we named rare structure effect which is caused by the inaccuracy of the local dissimilarity measure and often happens when the target is extremely dim. In next section, we will further discuss this phenomenon.

In fact, the common and intrinsic reason behind the unsatisfactory performance of all these six methods lies in their preset assumption about the target shape, namely a hot spot brighter than its neighborhood. Nevertheless, when the target is too dim to maintain its significant contrast over nontarget components, just like Fig. 5(a)–(d), they might not perform as well as they usually do.

The last six tested methods are all low-rank based methods. Comparing with above six methods, their results contain fewer background details. Relatively speaking, the effects of PRPCA and WIPI are not as good as the other four methods. Different from the other low-rank based methods that all build their low-rank assumptions on the data structure composed of patches, PRPCA supposes the individual patch is low-rank. Thus in PRPCA, each patch is applied to an individual RPCA process. Then, all the separated target patches are synthesized into a target image. By comparing Figs. 13(g) and 15(g) with Fig. 13(h) and 15(h), it can be seen that fewer edges were left
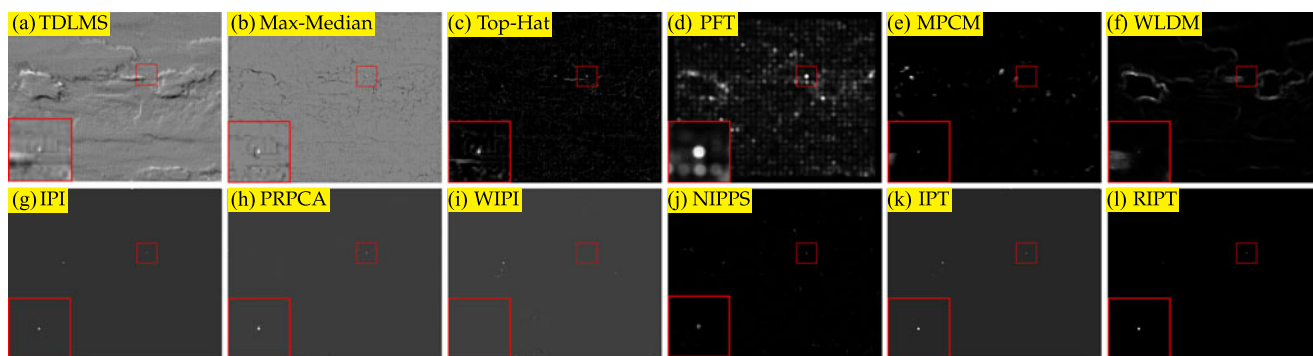
Fig. 12. Separated target images of the 65th frame in Sequence 1 by 12 tested methods. For better visualization, the demarcated area is enlarged in the left bottom corner. It is better to be seen by zooming on a computer screen.
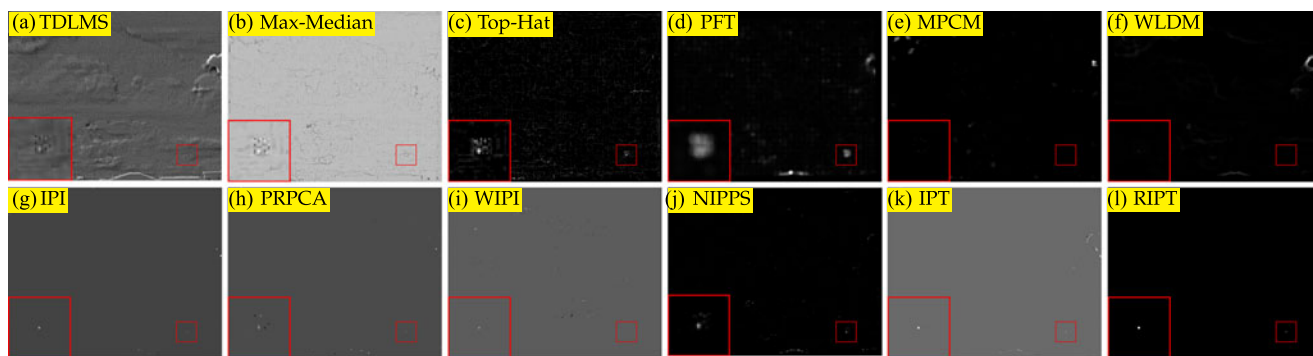

Fig. 13. Separated target images of the 52th frame in Sequence 2 by 12 tested methods. For better visualization, the demarcated area is enlarged in the left bottom corner. It is better to be seen by zooming on a computer screen.
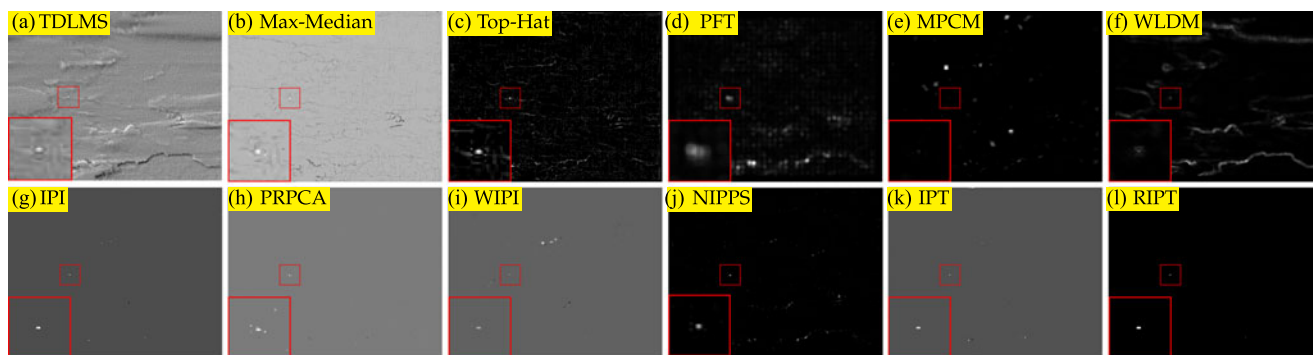

Fig. 14. Separated target images of the 53th frame in Sequence 3 by 12 tested methods. For better visualization, the demarcated area is enlarged in the left bottom corner. It is better to be seen by zooming on a computer screen.
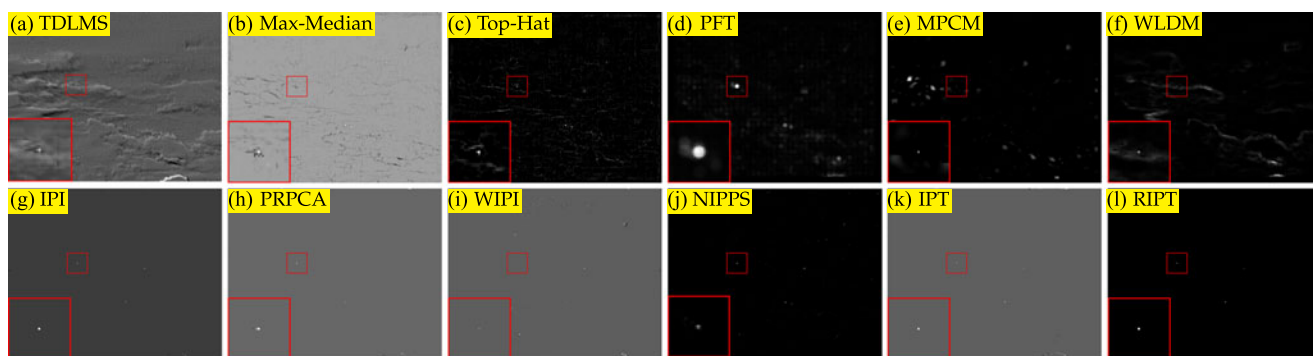

Fig. 15. Separated target images of the 56th frame in Sequence 4 by 12 tested methods. For better visualization, the demarcated area is enlarged in the left bottom corner. It is better to be seen by zooming on a computer screen.

TABLE V
QUANTITATIVE COMPARISON OF THE TESTED METHODS FOR THE REPRESENTATIVE IMAGES OF SEQUENCES 1–4

Method	65th frame of Sequence 1			52th frame of Sequence 2			53th frame of Sequence 3			56th frame of Sequence 4		
	LSNRG	SCRG	BSF	LSNRG	SCRG	BSF	LSNRG	SCRG	BSF	LSNRG	SCRG	BSF
Max-Median	5.49	10.69	12.10	1.87	5.46	7.37	2.96	6.21	11.27	7.54	9.81	16.66
Top-Hat	3.47	13.47	12.34	2.10	12.55	8.08	3.10	9.48	11.24	4.04	22.13	21.06
PFT	4.83	53.01	7.43	1.37	10.96	3.22	0.68	7.48	3.38	9.16	113.25	18.77
MPCM	1.48	7.69	3.46	1.62	11.84	15.98	0.38	1.91	2.15	1.88	14.17	4.68
WLDM	0.87	2.00	1.99	2.22	9.95	12.23	1.94	8.19	3.95	3.11	12.11	3.56
TDLMS	1.36	3.44	3.53	1.76	4.27	3.38	2.61	4.39	4.49	1.99	4.77	4.50
IPI	220.38	5215.82	19256.20	10.72	104.34	172.90	**Inf**	**Inf**	**Inf**	**Inf**	2788.19	4939.03
PRPCA	5.17	382.58	20179.12	1.30	26.88	2628.42	1.68	31.85	1982.80	2.83	267.31	20966.41
CWRPCA	2.67	36.77	602.45	4.62	40.59	58.23	7.69	98.57	201.89	52.92	441.65	2065.42
NIPPS	15.95	315.08	670.65	3.89	66.99	81.05	20.59	343.06	735.19	87.92	2280.13	3103.00
IPT	9.80	2096.70	87797.82	2.14	332.56	17488.27	3.22	**Inf**	**Inf**	2.86	**Inf**	**Inf**
RIPT	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**	**Inf**

*Different from the filtering based methods, for the low-rank based methods, **Inf**, namely infinity, is quite common, which just means that the target neighborhood completely shrinks to zero.
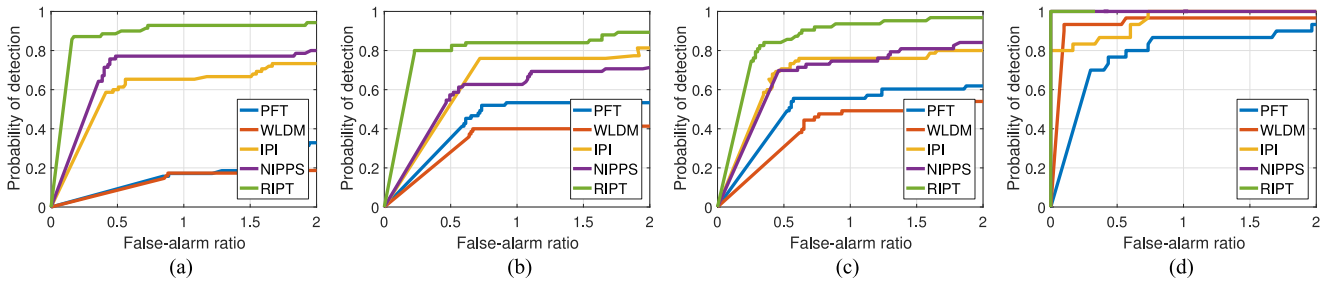

Fig. 16. ROC curves of detection results of four real infrared sequences: (a) Sequence 1, (b) Sequence 3, (c) Sequence 4, and (d) Sequence 5.

by IPI than PRPCA. It is because the rare structure in a patch is not necessarily rare in the patch-image due to the redundancy of the whole image. Therefore, the results of IPI and IPT are much better than those of PRPCA. As to WIPI, considering the targets in Sequences 1–4 is much dimmer than those in [29], it is fair to say that the steering kernel-based patch-level weight is still not robust enough to handle all of the complex infrared backgrounds. From Figs. 12(l) and 15(l), we can see that with the help of the local structure weight, the nontarget components were suppressed completely via our proposed model. For example, the cloud residuals in Fig. 12(g) by IPI is brighter than its target, while in Fig. 12(l), it is wiped out thoroughly. Based on above comparisons, it is fair to conclude that the proposed RIPT model achieves the most satisfying performance in infrared background suppression among 12 tested methods.

For infrared small target detection, the biggest difficulty is the interference of complex backgrounds. These undesirable background clutters raise the false alarm rates, and might even overwhelm the dim targets. Therefore, the ability of successfully suppressing the background clutters is a major concern in evaluating an infrared small target detection method. Quantitative evaluating indices are also widely used to assess this ability. Table V shows the experimental data of all 12 tested methods for Fig. 5(a)–(d). The gray-scale of every separated target image is rescaled to the range 0–255. It could be observed that our pro-

posed method gets the highest scores among all indices and all tested images. Different from the filtering-based methods, for the low-rank-based methods, **Inf**, namely infinity, is quite common, which just means that the target neighborhood completely shrinks to zero. In addition, it should be noted that the high scores in these three quantitative indices merely reflect the good suppression performance in a local region, and not necessarily mean a good global suppression ability.

To further reveal the advantage of the proposed model, we display the ROC curves of Sequence 1 and Sequences 3–5 for comparison in Fig. 16. The most interesting point is the performances of the state-of-the-art WLDM on Sequences 1, 3, 4 and 5 are very different. For Sequence 5, WLDM performs very well but fails in Sequences 1–4. We believe the reason lies in the rare structure effect which is a born problem for local contrast method. NIPPS's performance is slightly better than the IPI model. Finally, the proposed algorithm achieves the highest detection probability for the same false-alarm ratio, meaning that the proposed RIPT model has a better performance than the other models.

VI. CONCLUSION

To further suppress the strong edges while preserving the spatial correlation, a reweighted infrared patch-tensor model for small target detection is developed in this paper, simultane-

ously combining nonlocal redundant prior and local structure prior together. A local structure weight is designed based on the structure tensor and served as an edge indicator in the weighted model. In addition, a sparsity enhancement scheme is adopted to avoid the target image being contaminated. Then, the target-background separation task is modeled as a reweighted robust tensor recovery problem, which can be efficiently solved via ADMM. Detailed experimental results show that our proposed model is robust to various scenarios and obtains the clearest separated target images compared with the state-of-the-art target-background separation methods.

ACKNOWLEDGMENT

REFERENCES

[1] S. Kim and J. Lee, "Scale invariant small target detection by optimizing signal-to-clutter ratio in heterogeneous background for infrared search and track," *Pattern Recognit.*, vol. 45, no. 1, pp. 393–406, 2012.

[2] Z. Liu *et al.*, "Iterative infrared ship target segmentation based on multiple features," *Pattern Recognit.*, vol. 47, no. 9, pp. 2839–2852, 2014.

[3] X. Bai *et al.*, "Infrared ship target segmentation based on spatial information improved FCM," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3259–3271, Dec. 2016.

[4] I. S. Reed *et al.*, "Optical moving target detection with 3-D matched filtering," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, no. 4, pp. 327–336, Jul. 1988.

[5] M. Li *et al.*, "Moving weak point target detection and estimation with three-dimensional double directional filter in IR cluttered background," *Opt. Eng.*, vol. 44, no. 10, pp. 107 007–107 007–4, 2005.

[6] K. A. Melendez and J. W. Modestino, "Spatiotemporal multiscan adaptive matched filtering," *Proc. SPIE*, vol. 2561, pp. 51–65, 1995.

[7] J. Zheng *et al.*, "Radar high-speed target detection based on the frequency-domain deramp-keystone transform," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 1, pp. 285–294, Jan. 2016.

[8] X. Guo *et al.*, "Parallel computation of aerial target reflection of background infrared radiation: Performance comparison of openmp, openacc, and cuda implementations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 4, pp. 1653–1662, Apr. 2016.

[9] M. Hadhoud and D. Thomas, "The two-dimensional adaptive LMS (TDLMS) algorithm," *IEEE Trans. Circuits Syst.*, vol. 35, no. 5, pp. 485–494, May 1988.

[10] S. D. Deshpande *et al.*, "Max-mean and max-median filters for detection of small targets," *Proc. SPIE*, vol. 3809, pp. 74–83, 1999.

[11] T.-W. Bae *et al.*, "Edge directional 2D LMS filter for infrared small target detection," *Infrared Phys. Technol.*, vol. 55, no. 1, pp. 137–145, 2012.

[12] Y. Cao *et al.*, "Small target detection using two-dimensional least mean square (TDLMS) filter based on neighborhood analysis," *Int. J. Infrared Millim. Waves*, vol. 29, no. 2, pp. 188–200, 2008.

[13] 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, 2010.

[14] S. Kim *et al.*, "Small target detection utilizing robust methods of the human visual system for IRST," *J. Infrared, Millim. THz Waves*, vol. 30, no. 9, pp. 994–1011, 2009.

[15] X. Wang *et al.*, "Infrared dim target detection based on visual attention," *Infrared Phys. Technol.*, vol. 55, no. 6, pp. 513–521, 2012.

[16] S. Qi *et al.*, "A robust directional saliency-based method for infrared small-target detection under various complex backgrounds," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 495–499, May 2013.

[17] C. L. P. Chen *et al.*, "A local contrast method for small infrared target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 574–581, Jan. 2014.

[18] J. Han *et al.*, "A robust infrared small target detection algorithm based on human visual system," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2168–2172, Dec. 2014.

[19] Y. Wei *et al.*, "Multiscale patch-based contrast measure for small infrared target detection," *Pattern Recognit.*, vol. 58, pp. 216–226, 2016.

[20] H. Deng *et al.*, "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.

[21] J. Han *et al.*, "An infrared small target detecting algorithm based on human visual system," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 452–456, Mar. 2016.

[22] Y. Chen and Y. Xin, "An efficient infrared small target detection method based on visual contrast mechanism," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 7, pp. 962–966, Jul. 2016.

[23] H. Deng *et al.*, "Small infrared target detection based on weighted local difference measure," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4204–4214, Jul. 2016.

[24] H. Deng *et al.*, "Entropy-based window selection for detecting dim and small infrared targets," *Pattern Recognit.*, vol. 61, pp. 66–77, 2017.

[25] M. Fornasier *et al.*, "Low-rank matrix recovery via iteratively reweighted least squares minimization," *SIAM J. Optim.*, vol. 21, no. 4, pp. 1614–1640, 2011.

[26] 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.

[27] E. J. Candès *et al.*, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 11:1–11:37, Jun. 2011.

[28] Y. He *et al.*, "Small infrared target detection based on low-rank and sparse representation," *Infrared Phys. Technol.*, vol. 68, pp. 98–109, 2015.

[29] Y. Dai *et al.*, "Infrared small target and background separation via column-wise weighted robust principal component analysis," *Infrared Phys. Technol.*, vol. 77, pp. 421–430, 2016.

[30] Y. Dai *et al.*, "Non-negative infrared patch-image model: Robust target-background separation via partial sum minimization of singular values," *Infrared Phys. Technol.*, vol. 81, pp. 182–194, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1350449516303723

[31] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[32] J.-F. Cai *et al.*, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[33] J. Hstad, "Tensor rank is np-complete," *J. Algorithms*, vol. 11, no. 4, pp. 644–654, 1990.

[34] D. Goldfarb and Z. T. Qin, "Robust low-rank tensor recovery: Models and algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 35, no. 1, pp. 225–253, 2014.

[35] J. Weickert, "Coherence-enhancing diffusion filtering," *Int. J. Comput. Vis.* vol. 31, no. 2/3, pp. 111–127, 1999.

[36] Z. Wu *et al.*, "Structure tensor total variation-regularized weighted nuclear norm minimization for hyperspectral image mixed denoising," *Signal Process.*, vol. 131, pp. 202–219, 2017.

[37] W.-Z. Shao and Z.-H. Wei, "Research on filtering behavior of structure tensor based image modeling approaches," *Tien Tzu Hsueh Pao/Acta Electron. Sin.*, vol. 39, no. 7, pp. 1556–1562, 2011.

[38] E. J. Candès *et al.*, "Enhancing sparsity by reweighted l_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5, pp. 877–905, 2008.

[39] Y. Peng *et al.*, "Reweighted low-rank matrix recovery and its application in image restoration," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2418–2430, Dec. 2014.

[40] Y. Xie *et al.*, "Weighted schatten p-norm minimization for image denoising and background subtraction," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4842–4857, Oct. 2016.

[41] C. Lu *et al.*, "Nonconvex nonsmooth low rank minimization via iteratively reweighted nuclear norm," *IEEE Trans. Image Process.*, vol. 25, no. 2, pp. 829–839, Feb. 2016.

[42] S. Gu *et al.*, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vis.*, vol. 121, pp. 183–208, 2016.

[43] C. Gao *et al.*, "Infar dataset: Infrared action recognition at different times," *Neurocomputing*, vol. 212, pp. 36–47, 2016.

[44] Z. Wen *et al.*, "Joint sparse recovery with semisupervised music," *IEEE Signal Process. Lett.*, vol. 24, no. 5, pp. 629–633, May 2017.

[45] J. Chen *et al.*, "Unsupervised high-level feature extraction of SAR imagery with structured sparsity priors and incremental dictionary learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 10, pp. 1467–1471, Oct. 2016.

[46] J. Chen *et al.*, "High-level feature selection with dictionary learning for unsupervised sar imagery terrain classification," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 10, no. 1, pp. 145–160, Jan. 2017.

[47] S. Boyd *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[48] J. Salmon and Y. Strozecki, "Patch reprojections for non-local methods," *Signal Process.*, vol. 92, no. 2, pp. 477–489, 2012.

[49] J. Rivest and R. Fortin, "Detection of dim targets in digital infrared imagery by morphological image processing," *Opt. Eng.*, vol. 35, no. 7, pp. 1886–1893, 1996.

[50] C. Guo *et al.*, "Spatio-temporal saliency detection using phase spectrum of quaternion Fourier transform," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[51] C. Wang and S. Qin, "Adaptive detection method of infrared small target based on target-background separation via robust principal component analysis," *Infrared Phys. Technol.*, vol. 69, pp. 123–135, 2015.

[52] Z. Wen *et al.*, "Discriminative dictionary learning with two-level low rank and group sparse decomposition for image classification," *IEEE Trans. Cybern.*, to be published, doi: 10.1109/TCYB.2016.2581861.
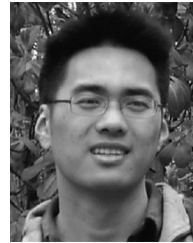
Yimian Dai received the B.S. degree in information engineering from Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, China, in 2013, where he is currently pursuing the Ph.D. degree in signal and information processing.

His current interests include target detection, image restoration, and machine learning.

Yiquan Wu received the B.S degree in instrumentation and test system from Nanjing University of Aeronautics and Astronautics (NUAA) in 1984. He received the M.S. and Ph.D. degrees in signal and information processing from Nanjing University of Aeronautics and Astronautics in 1987 and 1998, respectively.

He is at present a professor and Ph.D. supervisor in the Department of Information and Communication Engineering at the Nanjing University of Aeronautics and Astronautics, where he is involved in teaching and research in the areas of image processing and recognition, target detection and tracking, and intelligent information processing.