Infrared Patch-Tensor Model With Weighted Tensor Nuclear Norm for Small Target Detection in a Single Frame

YANG SUN[iD]**, JUNGANG YANG, (Member, IEEE), YUNLI LONG, ZHERAN SHANG, AND WEI AN**
College of Electronic Science, National University of Defense Technology, Changsha 410073, China
Corresponding author: Jungang Yang (yangjungang@nudt.edu.cn)

ABSTRACT The robust and efficient detection of infrared small target is a key technique for infrared search and track systems. Several robust principal component analysis (RPCA)-based methods have been developed recently, which have achieved the state-of-the-art performance. However, there are still two drawbacks: 1) the false alarm ratio would raise under the heavy background clutters and noises and 2) these methods are usually time-consuming and not suitable for real-time processing. To solve this problem, an infrared patch-tensor (IPT) model based on weighted tensor nuclear norm is proposed in this paper. First, the infrared image is transformed into the IPT. Considering the sum of nuclear norms adopted in the IPT model is not the convex envelope of the tensor rank, and the solution is substantially suboptimal. The tensor nuclear norm is adopted to recover the underlying low-rank background tensor and sparse target tensor, and the computation complexity can be reduced dramatically with the help of the tensor Singular Value Decomposition. Moreover, to further suppress the background clutters, a weight tensor is incorporated with tensor nuclear norm to preserve the background edges better. Then, the separation between target and background is formulated as a convex weighted tensor RPCA model. Finally, the proposed model can be solved by the alternating direction method of multipliers. Extensive experiments demonstrate that the proposed model outperforms the other state-of-the-arts in terms of performance and efficiency.

INDEX TERMS Infrared patch-tensor model, small target detection, weighted tensor robust principal component analysis.

I. INTRODUCTION

Infrared small target detection is one of the key techniques for various applications, such as infrared search and track (IRST) systems and space-based surveillance systems (SBSS), where accuracy, robustness and efficiency are indispensable. Nevertheless, the infrared targets are always very small and dim without known shape in consequence of long imaging distance. Moreover, small targets are usually buried in a complex background clutters with low signal-to-clutter ratio (SCR), including heavy cloud clutter, sea clutter and pixel-sized noises. Therefore, the infrared small target detection is a challenging problem and has attracted much research attention [1]–[3].

A. RELATED WORK

Many state-of-the-art small infrared target detection approaches have been developed in recent years, and these methods can be classified into two categories, namely, the single-frame detection methods and sequential detection methods.

The sequential detection methods usually use the spatial and temporal information of a number of frames simultaneously to detect target. One of representative classical methods is three-dimensional (3D) matched filters [4], which can detect targets with the prior knowledge of target shape and velocity. And a new double directional filter was proposed in [5] to improve the performance of weak target detection. Then an improved 3D directional filters [6] further improved the performance. Recently, a joint spatial-temporal sparse recovery method was proposed in [7], where both the motion and morphological feature of the targets are characterized simultaneously. And a spatial-temporal local contrast filter [8] analyzed the fluctuation of target signal in temporal domain and spatial domain. To reduce the computational complexity, an idea of only using temporal features was proposed in [9]. In addition, some sequential detection

methods were developed to suppress the background clutters and enhance the small targets in singe frame, then multiframe accumulation is adopted to delete non-target residuals, such as an method based on support vector machines (SVM) [10].

This kind of methods can achieve satisfactory performance with static or slowly changing backgrounds and consistent target trajectory in the infrared sequence. Nevertheless, these preset assumptions fail in many applications, such as anti-ship missiles and airplanes targets. The background would change rapidly and the target motion might be discontinuous between frames in consequence of the relative high velocity of targets to imaging systems [11]. Compared with the sequential detection methods, the single-frame target detection methods have some superiorities. It requires both less prior knowledge and computational complexity. Therefore, the single-frame target detection methods have attracted much attention recently. The existing single-frame target detection methods can be classified into four types: the filtering methods using spatial information, the human visual system (HVS) based methods, binary classification-based methods and robust principal component analysis (RPCA) based methods.

The first type of methods attempt to predict the background component and extract targets based on spatial consistency assumption. The representative methods include Top-Hat filter [12], 2D least mean square (TDLMS) filter [13], Max-Median filter [14] and other methods [15], [16]. However, the strong edges and clutter could not be wiped out by these methods since they also break the spatial consistency like targets.

The human visual system (HVS) based methods are developed recently, where the small target is regarded as the salient point in contrast to background. Some representative method [17]–[19] measured the saliency of targets by different filters. In [20], the local contrast measure (LCM) is adopted to quantify the saliency, and some works [21], [22] have been proposed to improve the performance. However, the performance of these methods would degrade for complex background cases that the clutters are similar to the target in saliency map.

The binary classification-based methods transformed the original infrared image into mixture of features and then the targets can be discriminated by trained classifiers. Some representative works are listed as follows: artificial neural networks [23], SVM based method [24], and so on [25]. Nevertheless, the performance of these methods depends heavily on the features selecting and sufficient training samples, which restrict its potential use in practical applications.

The RPCA based methods assume that the background and target are low-rank component and sparse component, respectively. And these methods model the sparse and low-rank components recovery as a RPCA problem [26]–[30]. In the seminar work of infrared image-patch (IPI) model [31], the original infrared image is transformed into patch-image by sliding window. Then the separation problem between target and background is modeled as a RPCA problem.

IPI model has achieved satisfying performance in some scenarios. However, this method still would remain some background clutters in target image due to quantify target sparsity by l_1 norm. To address this problem, a weighted IPI model was proposed in [32] by adding the column-wise weight to l_1 norm. But it is time-consuming to compute the column-wise weights. To further improve the performance, a non-negative IPI model by minimizing the partial sum of singular values (NIPPS) was proposed in [33]. However, the performance of NIPPS is sensitive to the energy constrain ratio, and the prior knowledge of rank is hard to estimate firstly in real application. An reweighted WIPI model (ReWIPI) was proposed in [34] by incorporating the weighted nuclear norm to suppress the background better. Nevertheless, the computation complexity is much more than other low-rank based model. To utilize more spatial correlationships, a reweighted infrared patch-tensor (RIPT) model using both nonlocal and local priors was proposed in [11]. The original images are transformed into a tensor by sliding window. And the small target detection is achieved by solving the low-rank tensor recovery problem. However, the existing low-rank based methods are time-consuming, which cannot meet the real-time processing requirements.

B. MOTIVATION

The infrared patch-tensor (IPT) model was proposed in [11] with both nonlocal and local priors, and this model has achieved a state-of-the-art performance for single-frame small target detection. In [11], to minimize the 3-way background image tensor $\mathcal{B}$, a convex surrogate Trucker-rank CTrank $(\mathcal{B})$ is defined as the Sum of Nuclear Norms (SNN) [35], i.e., CTrank $(\mathcal{B})$ = $\sum_i \|\boldsymbol{B}_{(i)}\|_*$, i = 1, 2, 3, where $\boldsymbol{B}_{(i)}$ is the mode-i unfold matricization of tensor $\mathcal{B}$. Nevertheless, the SNN used in the IPT model is not the convex envelope of $\sum_i \text{rank} \left(\boldsymbol{B}_{(i)} \right)$ [36]. Therefore, the above model is still substantially suboptimal. And a great deal of matrix SVD decomposition operations increase computation complexity.

Another motivation is that the performance of vanilla nuclear norm minimization based methods, i.e., IPI model and RIPT model, could degrade when the amount of background patches containing strong edges might be deficient in practical infrared images. In this case, the deficient edge samples are also considered as sparse component globally, like true targets, due to the rare and sparse background structures' effects [11], which would raise the false alarm rate. Although the RIPT model exploited the local structure prior by incorporating the weight tensor to l_1 norm to measure the sparsity, and it does overcome this dilemma to a certain extent, it is only applicable for some specific types of strong edges. Guo *et al.* [34] analyzed that the l_1 norm based sparsity measure is not the direct reason for this problem, which demonstrated that the performance of weighted l_1 norm based method [11] is suboptimal. And the intrinsic reason is that the vanilla nuclear norm is only suitable for the cases of sufficient edge samples. To solve this problem, the weighted nuclear

TABLE 1. Metamathematical notations.

Notations	Explanation
$\mathcal{A}, \boldsymbol{A}, a, a$	tensor,matrix,vector,scalar
I_n	$n \times n$ sized identity matrix
$\mathbb{R}, \mathbb{C}$	real numbers and complex numbers
$\mathcal{I}$	The identity tensor $\mathcal{I} \in \mathbb{R}^{n \times n \times n_3}$ is the tensor with its first frontal slice being the $n \times n$ identity matrix, and other frontal slices being all zeros.
$\mathcal{A}_{ijk}$ or a_{ijk}	the (i, j, k)th entry of tensor $\mathcal{A}$
$\mathcal{A}(i, :, :)$ or $\boldsymbol{A}^{(i)}, \mathcal{A}(:, i, :), \mathcal{A}(:, :, i)$	ith horizontal, lateral and frontal slice of tensor $\mathcal{A}$
$\mathcal{A}(i, j, :)$	the tube of $\mathcal{A}$
$\mathcal{A}^*$	The conjugate transpose of $\mathcal{A} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ is the tensor $\mathcal{A}^* \in \mathbb{C}^{n_1 \times n_2 \times n_3}$, which is obtained by conjugate transposing each of the frontal slices and reversing the order of transposed frontal slices 2 through n_3.
$\text{fold}(\mathcal{A}), \text{unfold}(\mathcal{A})$	unfold operator maps $\mathcal{A}$ to matrix of size $n_1 n_3 \times n_2$, and fold is the inverse operator
$\langle \mathcal{A}, \mathcal{B} \rangle$	inner product between tensor $\mathcal{A}$ and $\mathcal{B}$, which is defined as $\sum_{i=1}^{n_3} \left\langle \boldsymbol{A}^{(i)}, \boldsymbol{B}^{(i)} \right\rangle$
$\lfloor t \rfloor$ and $\lceil t \rceil$	the nearest integer less than or equal to t and the one greater than or equal to t
$\|\mathcal{A}\|_0$	ℓ_0 norm of $\mathcal{A}$, which is the number of non-zero elements
$\|\mathcal{A}\|_1$	ℓ_1 norm of $\mathcal{A}$, which is the sum of all $|a_{ijk}|$
$\|\mathcal{A}\|_\infty$	the infinity norm of $\mathcal{A}$, which is defined as the maximum of all $|a_{ijk}|$
$\|\mathcal{A}\|_F$	the Frobenius norm of $\mathcal{A}$, which is defined as $\sqrt{\sum_{ijk} |a_{ijk}|^2}$

norm minimization (WNNM) [37] method was adopted in ReWIPI [34] to suppress the deficient simple edges in target image by penalizing the larger singular values with smaller weights. The WNNM has excellent performance in image denoising [37], and it was used in ReWIPI model to preserve the background edge in the low-rank component, namely, the background patch image. The ReWIPI model has achieved better performance than vanilla nuclear norm based methods [31] for infrared small target detection. Motivated by the satisfying background suppression performance of WNNM method, a weighted tensor nuclear norm is also adopted to further suppress the strong edges in our model, which is a natural generalization of the weighted matrix nuclear norm for tensor case.

Therefore, a weighted tensor nuclear norm with reweighted infrared patch-tensor (WNRIPT) model is proposed in this paper for better infrared small target detection performance in a single frame. Our main contributions of this work are summarized as follows:

1) Motivated by the tensor nuclear norm [38]–[40], which is to recover a tensor of low tubal rank, the solution to the convex TRPCA model can perfectly recover the low-rank background component and sparse target component.

2) Equipped with the tensor-SVD (t-SVD) factorization, the computation of t-SVD can be more efficient than matrix SVD decomposition used in [11], so that the proposed WNRIPT method can be more efficient than RIPT model.

3) Noting that the performance of vanilla nuclear norm minimization would degrade when the strong edge samples are insufficient. The weighted tensor nuclear norm minimization [34], [37] is extended for tensor case, which can further suppress the strong edges.

The experimental results demonstrate the superiority of WNRIPT method for background suppression and target detection task over the existing state-of-art methods. And the proposed model is more efficient and suitable for real-time application.

The remainder of this paper is structured as follows. Section II gives some notations and briefly introduce the IPT model. The proposed model is described in Section III. Section IV provides extensive experiments on synthetic and real infrared images to validate the effectiveness of proposed method. We conclude this work in Section V.

II. NOTATIONS AND IPT MODEL
In this section, we introduce several mathematical notations and IPT model firstly.

A. NOTATIONS
The mathematical notations used in this paper are defined in table 1.

B. IPT MODEL
Generally, a single infrared small target image can be formulated as:

$$f_D = f_B + f_T + f_N \qquad (1)$$

where f_D, f_B, f_T and f_N represent the original image, background image, target image and noise image, respectively. And the IPT model propsed in [11] used a window sliding from top left to the bottom right over original image, and the image patches are staked into a 3D tensor. Model (1) is transformed into tensor form:

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

where $\mathcal{D}, \mathcal{B}, \mathcal{T}, \mathcal{N} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ are the original patch-tensor, background patch-tensor, target patch-tensor and noise patch-tensor, respectively. n_1 and n_2 are the patch height and width, n_3 is the patch number.

The background image is usually considered to change slowly, so there are high correlation property among its local and nonlocal patches. And the low-rank property of all mode-i, $i = 1, 2, 3$ unfolding matrices of the 3-way background patch-tensor was analyzed in [11], i.e., the singular values of

all the unfolding matrices decrease to zero rapidly. The low-rank property of these unfolding matrices are depicted as:

$$\text{rank}\left(\boldsymbol{B}_{(i)}\right) \leq r_i \tag{3}$$

where r_i, $i = 1, 2, 3$ is a constant to describe the complexity of the background image.

The small target usually occupies only several pixels, so it is relatively small to the whole image. The target patch-tensor is an extremely sparse tensor, which can be depicted as:

$$\|\boldsymbol{\mathcal{T}}\|_0 \leq k \tag{4}$$

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

The noise is assumed to be additive white Gaussian noise, and the Frobenius norm of the noise patch tensor satisfies $\|\boldsymbol{\mathcal{N}}\|_F \leq \delta$. Based on the IPT model, the infrared small target detection task can be formulated to the following problem:

$$\min_{\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{T}}} \|\boldsymbol{\mathcal{B}}\|_* + \lambda \|\boldsymbol{\mathcal{T}}\|_0, \quad s.t. \ \boldsymbol{\mathcal{B}} + \boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{D}} \tag{5}$$

where λ is a weighting parameter.

The above problem is a NP-hard problem. And the SNN [35] was adopted in [11], i.e., $\text{CTrank}(\boldsymbol{\mathcal{B}}) = \sum_i \|\boldsymbol{B}_{(i)}\|_*$, $i = 1, 2, 3$. And the $\|\boldsymbol{\mathcal{T}}\|_0$ is replaced by $\|\boldsymbol{\mathcal{T}}\|_1$ to make the above problem tractable. Besides, ℓ_1 norm of the target patch-tensor was incorporated with weight tensor $\boldsymbol{\mathcal{W}}$, including the local structure weight $\boldsymbol{\mathcal{W}}_{\text{LS}}$ and sparsity enhancing weight $\boldsymbol{\mathcal{W}}_{\text{SE}}^k$. The reweighted IPT model (RIPT) can be formulated as below [11]:

$$\min_{\boldsymbol{\mathcal{B}}, \boldsymbol{\mathcal{T}}} \sum_{i=1}^{3} \|\boldsymbol{B}_{(i)}\|_* + \lambda \|\boldsymbol{\mathcal{W}}_T \odot \boldsymbol{\mathcal{T}}\|_1, \quad s.t. \ \boldsymbol{\mathcal{B}} + \boldsymbol{\mathcal{T}} = \boldsymbol{\mathcal{D}} \tag{6}$$

where $\boldsymbol{\mathcal{W}}_T = \boldsymbol{\mathcal{W}}_{\text{LS}} \odot \boldsymbol{\mathcal{W}}_{\text{SE}}^k$.

III. THE PROPOSED MODEL

Motivation part has pointed out that SNN is not the convex envelop of $\sum_i \text{rank}\left(\boldsymbol{B}_{(i)}\right)$, the solution of model (6) is substantially suboptimal to model (5). In this section, we adopt the new tensor nuclear norm [38]–[40], which can perfectly recover the low-rank background patch-tensor and sparse target patch-tensor.

A. TENSOR NUCLEAR NORM

To improve the target detection performance, we incorporate the novel tensor nuclear norm [38]–[40] to solve model (5).

Some definitions of tensor are introduced in this subsection firstly. For tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, we denote $\bar{\boldsymbol{\mathcal{A}}} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ as the result of Discrete Fourier Transformation (DFT) on $\boldsymbol{\mathcal{A}}$ along the third dimension: $\bar{\boldsymbol{\mathcal{A}}} = \text{fft}(\boldsymbol{\mathcal{A}}, [\,], 3)$. Conversely, we can compute $\boldsymbol{\mathcal{A}}$ from $\bar{\boldsymbol{\mathcal{A}}}$ by inverse FFT: $\boldsymbol{\mathcal{A}} = \text{ifft}(\bar{\boldsymbol{\mathcal{A}}}, [\,], 3)$. And we denote $\bar{\boldsymbol{A}} \in \mathbb{R}^{n_1 n_3 \times n_2 n_3}$ as a block diagonal matrix with its ith block on the diagonal as the

ith frontal slice $\bar{\boldsymbol{A}}^{(i)}$ of $\bar{\boldsymbol{\mathcal{A}}}$:

$$\bar{\boldsymbol{A}} = \text{bdiag}\left(\bar{\boldsymbol{\mathcal{A}}}\right) = \begin{bmatrix} \bar{\boldsymbol{A}}^{(1)} & & & \\ & \bar{\boldsymbol{A}}^{(2)} & & \\ & & \ddots & \\ & & & \bar{\boldsymbol{A}}^{(n_3)} \end{bmatrix}$$

where bdiag( ) is an operator that transforms the tensor to the block diagonal matrix. And we have a following important property for $\bar{\boldsymbol{A}}^{(i)}$ [41]:

$$\begin{cases} \bar{\boldsymbol{A}}^{(1)} \in \mathbb{R}^{n_1 \times n_2} \\ \text{conj}\left(\bar{\boldsymbol{A}}^{(i)}\right) = \bar{\boldsymbol{A}}^{(n_3 - i + 2)}, \quad i = 2, \cdots, \left\lfloor \frac{n_3 + 1}{2} \right\rfloor \end{cases} \tag{7}$$

Definition 1 (T-Product [42]): Let tensor $\boldsymbol{\mathcal{A}} \in \mathbb{C}^{n_1 \times n_2 \times n_3}$ and $\boldsymbol{\mathcal{B}} \in \mathbb{C}^{n_1 \times l \times n_3}$. Then the t-product $\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{B}}$ is defined to be a tensor of size $n_1 \times l \times n_3$:

$$\boldsymbol{\mathcal{A}} * \boldsymbol{\mathcal{B}} = \text{fold}\left(\text{bcirc}\left(\boldsymbol{\mathcal{A}}\right) \cdot \text{unfold}\left(\boldsymbol{\mathcal{A}}\right)\right) \tag{8}$$

where bcirc $(\boldsymbol{\mathcal{A}})$ is the block circulant matrix of size $n_1 n_3 \times n_2 n_3$, which is defined as:

$$\text{bcirc}\left(\boldsymbol{\mathcal{A}}\right) = \begin{bmatrix} \boldsymbol{A}^{(1)} & \boldsymbol{A}^{(n_3)} & \cdots & \boldsymbol{A}^{(2)} \\ \boldsymbol{A}^{(2)} & \boldsymbol{A}^{(1)} & \cdots & \boldsymbol{A}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}^{(n_3)} & \boldsymbol{A}^{(n_3 - 1)} & \cdots & \boldsymbol{A}^{(1)} \end{bmatrix}$$

Definition 2 (F-Diagonal Tensor [42]): $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ tensor is called f-diagonal if each of its frontal slices is a diagonal matrix.

Definition 3 (T-SVD [38]): Let $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. Then it can be factorized as:

$$\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^* \tag{9}$$

where $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$, where $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{n_2 \times n_2 \times n_3}$ are orthogonal, and $\boldsymbol{\mathcal{S}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is an f-diagonal tensor.

Definition 4 (Tensor Tubal Rank [38]): Let $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^*$, the tensor tubal rank of $\boldsymbol{\mathcal{A}}$ is defined as the number of nonzero singular tubes of $\boldsymbol{\mathcal{S}}$, which is defined as follow:

$$\text{rank}_t\left(\boldsymbol{\mathcal{A}}\right) = \#\{i, \boldsymbol{\mathcal{S}}\left(i, i, :\right) \neq 0\}$$

Definition 5 (Tensor Nuclear Norm [38]): Let $\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{S}} * \boldsymbol{\mathcal{V}}^*$ be the t-SVD of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. The tensor nuclear norm is defined as:

$$\|\boldsymbol{\mathcal{A}}\|_* = \langle \boldsymbol{\mathcal{S}}, \boldsymbol{\mathcal{I}} \rangle = \sum_{i=1}^{r} \boldsymbol{\mathcal{S}}\left(i, i, 1\right)$$

$$= \frac{1}{n_3} \sum_{i=1}^{r} \sum_{j=1}^{n_3} \bar{\boldsymbol{\mathcal{S}}}\left(i, i, j\right) \tag{10}$$

where $r = \text{rank}_t\left(\boldsymbol{\mathcal{A}}\right)$, and $\boldsymbol{\mathcal{S}}\left(i, i, 1\right)$ is the entries on the diagonal of the first slice of $\boldsymbol{\mathcal{S}}$, which have the decreasing order property, i.e.,

$$\boldsymbol{\mathcal{S}}\left(1, 1, 1\right) \geq \boldsymbol{\mathcal{S}}\left(2, 2, 1\right) \geq \cdots \geq \boldsymbol{\mathcal{S}}\left(n_{(2)}, n_{(2)}, 1\right) \geq 0$$

where $n_{(2)} = \min(n_1, n_2)$. And the entries on the diagonal of $\bar{\mathcal{S}}(:, :, j)$ are the singular values of $\bar{\mathcal{A}}(:, :, j)$.

B. WEIGHTED BACKGROUND IMAGE TENSOR NUCLEAR NORM

Noticing that the performance of nuclear norm minimization would degenerate when the edge samples in background image patch is insufficient. To solve this deficiency, we use the weighted tensor nuclear norm to further constrain background patch-tensor, which is defined as:

$$\|\mathcal{B}\|_{\mathcal{W}_B,*} = \frac{1}{n_3} \sum_{i=1}^{r} \sum_{j=1}^{n_3} \mathcal{W}_B(i, i, j)\, \bar{\mathcal{S}}(i, i, j) \quad (11)$$

$$\mathcal{W}_B^{k+1}(i, i, j) = \frac{1}{\bar{\mathcal{S}}^k(i, i, j) + \varepsilon_{\mathcal{B}}} \quad (12)$$

where k is the iteration reweighted counter, $\bar{\mathcal{S}}(:, :, j)$ are the singular values of $\bar{\mathcal{B}}(:, :, j)$, which can be obtained by t-SVD. $\varepsilon_{\mathcal{B}}$ is a positive constant. And $\mathcal{W}_B^{k+1} \in \mathbb{R}^{r \times r \times n_3}$ is a weight tensor for tensor nuclear norm in $(k+1)$th iteration.

C. THE PROPOSED MODEL

Based on the above analysis, we can generalize a novel reweighted infrared patch-tensor model based on weighted tensor nuclear norm (WNRIPT), which can be formulated as:

$$\min_{\mathcal{B}, \mathcal{T}} \|\mathcal{B}\|_{\mathcal{W}_B,*} + \lambda \|\mathcal{W}_T \odot \mathcal{T}\|_1, \quad s.t.\ \mathcal{B} + \mathcal{T} = \mathcal{D} \quad (13)$$

D. SOLUTION OF WNRIPT MODEL

In this section, we incorporate the Alternating Direction Method of Multipliers (ADMM) [43] and the inexact augmented Lagrangian multiplier (IALM) [44] method to solve the proposed model. Model (13) can be reformulated as follow:

$$\mathcal{L}(\mathcal{B}, \mathcal{T}, \mathcal{Y}, \mu) = \|\mathcal{B}\|_{\mathcal{W}_B,*} + \lambda \|\mathcal{W}_T \odot \mathcal{T}\|_1 \\ + \frac{\mu}{2} \|\mathcal{B} + \mathcal{T} - \mathcal{D}\|_F + \langle \mathcal{Y}, \mathcal{B} + \mathcal{T} - \mathcal{D} \rangle \quad (14)$$

where $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ denote the Lagrangian multiplier tensors, and μ is a positive penalty scalar.

The above problem (14) can be minimized by solving two subproblems by ADMM, the $(\mathcal{B}, \mathcal{T}, \mathcal{Y})$ are solved alternatively as follows:

$$\mathcal{B}^{k+1} = \arg\min_{\mathcal{B}} \|\mathcal{B}\|_{\mathcal{W}_B^k,*} \\ + \frac{\mu_k}{2} \left\| \mathcal{B} - \left(\mathcal{D} - \mathcal{T}^k - \left(\mu^k\right)^{-1} \mathcal{Y}^k \right) \right\|_F^2 \quad (15)$$

$$\mathcal{T}_{k+1} = \arg\min_{\mathcal{B}} \lambda \|\mathcal{W}_k \odot \mathcal{T}\|_* \\ + \frac{\mu_k}{2} \left\| \mathcal{B}_k - \left(\mathcal{D} - \mathcal{B}_k - (\mu_k)^{-1} \mathcal{Y}_k \right) \right\|_F^2 \quad (16)$$

$$\mathcal{Y}^{k+1} = \mathcal{Y}^k + \mu^k \left(\mathcal{B}^{k+1} + \mathcal{T}^{k+1} - \mathcal{D} \right) \quad (17)$$

The problem (15) can be solved by tensor singular value thresholding (t-SVT) [38], [40], [45], which is described

in Algorithm 1, and the calculation times of matrix SVD operations can be reduced by almost half with property (7).

$$\mathcal{B}^{k+1} = \mathcal{D}_{(\mu^k)^{-1} \mathcal{W}_B^k} \left(\mathcal{D} - \mathcal{T}^k - \left(\mu^k\right)^{-1} \mathcal{Y}^k \right) \quad (18)$$

Algorithm 1 Solve (18) by t-SVT

Input: $\mathcal{Y} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$
Output: $\mathcal{D}_\mu(\mathcal{Y})$
1. Compute $\bar{\mathcal{Y}} = \text{fft}(\mathcal{Y}, [\ ], 3)$
2. Perform matrix SVT operator on each frontal slice of $\bar{\mathcal{Y}}$ by
    for $i = 1, \cdots, \left\lceil \frac{n_3+1}{2} \right\rceil$ **do**
      $[U, S, V] = \text{SVD}\left(\bar{Y}^{(i)} \right)$
      $\bar{W}^{(i)} = U \cdot (S - \mu)_+ \cdot V^*$
    end for
    for $i = \left\lceil \frac{n_3+1}{2} \right\rceil + 1, \cdots, n_3$ **do**
      $\bar{W}^{(i)} = \text{conj}\left(\bar{W}^{(n_3-i+2)} \right)$
    end for
3. Compute $\mathcal{D}_\mu(\mathcal{Y}) = \text{ifft}\left(\mathcal{W}, \bar{[\ ]}, 3 \right)$

And problem (16) can be solved by element-wise shrinkage operator [46]:

$$\mathcal{T}^{k+1} = \mathcal{F}_{(\mu_k)^{-1} \lambda \mathcal{W}_T^k} \left(\mathcal{D} - \mathcal{B}_k - (\mu_k)^{-1} \mathcal{Y}_k \right) \quad (19)$$

Finally, the detailed solving process is described in Algorithm 2.

E. DETECTION PROCEDURE

The detailed detection procedure is described as follows:
1) The original infrared image f_D is transformed into a patch-tensor $\mathcal{D} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ by using a sliding window of size $n_1 \times n_2$.
2) The local structure weighted target patch-tensor $\mathcal{W}_{LS}$ are constructed from corresponding local structure feature map [11].
3) The patch-tensor f_D is decomposed into background patch-tenor $\mathcal{B}$ and target patch-tenor $\mathcal{T}$ by Algorithm 1.
4) The background image f_B and target image f_T are reconstructed from background patch-tenor $\mathcal{B}$ and target patch-tenor $\mathcal{T}$ by the uniform average of estimators (UAE) reprojection scheme [47].
5) Finally, the target can be segmented by the adaptive threshold [31]:

$$t_{up} = \max(\upsilon_{\min}, \mu + k\sigma) \quad (20)$$

where μ and σ are the mean and standard deviation of the target image f_T, respectively. k and $\upsilon_{\min}$ are constant determined experientially. A pixel at (x, y) can be segmented as target if $f_T(x, y) > t_{up}$.

IV. EXPERIMENTS

In this section, we conduct experiments using infrared images from different scenarios. And four state-of-the-art methods are compared with the proposed method.

Algorithm 2 Solve (13) by ADMM

Input: image tensor data $\mathcal{D}$, parameter λ
Initialize: $\mathcal{B}^0 = \mathcal{T}^0 = \mathcal{Y}^0 = 0$, $\mathcal{W}_B^0 = \mathcal{I}$, $\mathcal{W}_{SE}^0 = \mathcal{I}$, $\mathcal{W}_T^0 = \mathcal{W}_{LS} \odot \mathcal{W}_{SE}^0$, $\mu_0 = 1e - 3, \mu_{\max} = 1e7, k = 0$, $\varepsilon = 1e - 8$, $\rho = 1.1$
While: not converged do
Step 1:Update $\mathcal{B}^{k+1}$ by

$$\mathcal{B}^{k+1} = \mathcal{D}_{(\mu^k)^{-1}\mathcal{W}_B^k} \left(\mathcal{D} - \mathcal{T}^k - \left(\mu^k\right)^{-1} \mathcal{Y}^k \right)$$

Step 2: Update $\mathcal{T}^{k+1}$ by

$$\mathcal{T}^{k+1} = \mathcal{F}_{(\mu_k)^{-1}\lambda\mathcal{W}_T^k} \left(\mathcal{D} - \mathcal{B}_k - (\mu_k)^{-1}\mathcal{Y}_k \right)$$

Step 3: Update $\mathcal{Y}^{k+1}$ by

$$\mathcal{Y}^{k+1} = \mathcal{Y}^k + \mu^k \left(\mathcal{B}^{k+1} + \mathcal{T}^{k+1} - \mathcal{D} \right)$$

Step 4: Update $\mathcal{W}_B^{k+1}$ by

$$\mathcal{W}_B^{k+1}(i, i, j) = \frac{1}{\bar{\mathcal{S}}^k(i, i, j) + \varepsilon_\mathcal{B}}$$

Step 5: Update $\mathcal{W}_T^{k+1}$ by
for $(i, j, p) \in [1, \ldots, n_1] \times [1, \ldots, n_2] \times [1, \ldots, n_3]$ **do**

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

end $\mathcal{W}^{k+1} = \mathcal{W}_{LS} \odot \mathcal{W}_{SE}^{k+1}$
Step 6:Update μ^{k+1} by

$$\mu^{k+1} = \min \left(\rho\mu^k, \mu_{\max} \right)$$

Step 7: Check the convergence conditions

$$\left\| \mathcal{B}^{k+1} - \mathcal{B}^k \right\|_\infty \leq \varepsilon, \quad \left\| \mathcal{T}^{k+1} - \mathcal{T}^k \right\|_\infty \leq \varepsilon,$$

$$\left\| \mathcal{B}^{k+1} + \mathcal{T}^{k+1} - \mathcal{D} \right\|_\infty \leq \varepsilon$$
or $\operatorname{rank}_t \left(\mathcal{T}^{k+1} \right) = \operatorname{rank}_t \left(\mathcal{T}^k \right)$
Step 8: Update $k = k + 1$
end While
Output: $\mathcal{B}^k, \mathcal{T}^k$

A. EXPERIMENTAL SETTINGS

1) DATASETS
Firstly, the proposed model is tested on some real infrared images to validate its background suppression ability. And the scenarios are varying from flat background with salient targets to complex background with heavy clutters, and the

targets of the latter scenario are extremely dim. In the following experiments, we focus on the real datasets with complex background and dim targets, and the detailed characteristics of five real infrared sequences are described in Table 2.

2) METRICS
The most important metrics of evaluating the detection performance are the detection probability P_d and false-alarm rate F_a, which are defined as following [12]:

$$P_d = \frac{\text{number of true detections}}{\text{number of actual targets}} \tag{21}$$

$$F_a = \frac{\text{number of false detections}}{\text{number of images}} \tag{22}$$

And the receiver operating characteristic (ROC) curves shows the tradeoff between P_d and F_a. In addition, if a target simultaneously meet two requirements: (i) the result and a ground truth have overlap pixels; (ii) the pixel distance between centers of the ground truth and the result is within a threshold [31], it is considered to be true detection.

Then, the local signal-to-noise ratio gain (LSNRG), background suppression factor (BSF), signal to clutter ratio gain (SCRG) [11] metrics are adopted to evaluate the background suppression ability of methods in this section. Since LSNRG, BSF and SCRG need to be computed in a local neighborhood region. We adopt a variable neighboring area with respect to the target [31]. Let the size of the target be $a \times b$, the size of its neighborhood region is $(a + 2d) \times (b + 2d)$, where d is the width of neighboring area and we set $d = 20$, as described in Fig. 1. Then the three metrics are defined as follows:

$$\text{LSNRG} = \frac{\text{LSNR}_{\text{out}}}{\text{LSNR}_{\text{in}}} \tag{23}$$

where LSNR_{out} and LSNR_{in} are the LSNR values before and after processing, and $\text{LSNR} = P_T / P_B$. P_T and P_B are the maximum pixel value of the target and neighborhood, respectively. BSF is defined as below:

$$\text{BSF} = \frac{\sigma_{\text{in}}}{\sigma_{\text{out}}} \tag{24}$$

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

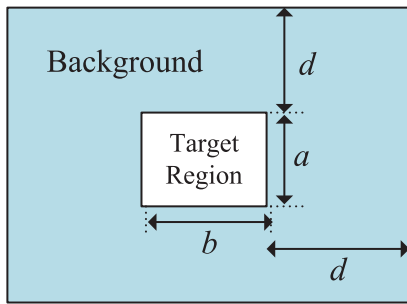$$\text{SCRG} = \frac{\text{SCR}_{\text{out}}}{\text{SCR}_{\text{in}}} \tag{25}$$

TABLE 2. Characteristics of five real infrared sequences.

Sequence	Frame number	Image resolution	Target characteristics	Background characteristics
Sequence 1	100	200×150	Tiny with low contrast	Heavy clutters with small amount of noise
Sequence 2	100	250×200	Very dim, buried in the cloud	Heavy cloud-sky clutters with small amount of noise
Sequence 3	100	250×200	Very dim, buried in the cloud	Heavy cloud-sky clutters with severe noise
Sequence 4	100	281×240	Tiny and varying size	Heavy cloud-sky clutters with small amount of noise
Sequence 5	100	281×240	Very dim, buried in the cloud	Heavy cloud-sky clutters with severe noise

FIGURE 1. The neighboring background rectangle of a small target.

the local SCR is defined as follows [48]:

$$\text{SCR} = \frac{|\mu_t - \mu_b|}{\sigma_b} \qquad (26)$$

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

Note that the computation of LSNRG, BSF and SCRG includes standard deviation which may be close to zero when the suppressed background is very clean. In this case, these metrics may approach to infinity. To address this problem, another metric, namely contrast gain (CG) is also adopted to evaluate the ability to enlarge gray value difference between background and small targets, which is helpful to extract true targets and defined as follows [49]:

$$\text{CG} = \frac{\text{CON}_{\text{out}}}{\text{CON}_{\text{in}}} \qquad (27)$$

where CON_{in} and CON_{out} are the contrast (CON) of the original and result images, respectively, and CON is

defined as:

$$\text{CON} = |\mu_t - \mu_b| \qquad (28)$$

where μ_t and μ_b are the same as those in Eq. (26).

In general, higher values means better background suppression ability for the above three metrics. And it should be noted that these three metrics only evaluate the suppression ability in local neighboring area, but not globally.

3) BASELINES AND PARAMETER SETTINGS
To further evaluate the propose method, we compare the performance of WNRIPT with four state-of-the-art methods. Considering that the RPIT outperforms many methods in [11], including TDLMS [13], three HVS-based methods (PFT [50], MPCM [22], WLDM [51]), and three low-rank methods (PRPCA [26], WIPI [32], NIPPS [33]). For briefly, our focus is put on the comparison with RIPT method. So we choose two traditional filtering-based methods (Max-Median [14], Top-Hat [12]) and the two low-rank method IPI [31], and RPIT [11]. Table 3 gives the detailed parameter settings. To make this a fair comparison, the IPI [31] is solved by ADMM. All the tested methods are implemented in MATLAB 2014a on a laptop of 2.6 GHz and 4GB RAM.

B. VALIDATION OF WNRIPT
In this section, the proposed method are validated for its robustness to different scenes, including single target scene, multiple targets scene and noisy cases.

1) PERFORMANCE OF SINGLE TARGET SCENE
In Fig.2, five real single target infrared images are given in the first row, and their corresponding separated target images

TABLE 3. Detailed parameter setting for tested methods.

| Methods	Acronyms	Parameter settings
Max-Median filter	Max-Median	Support size:5×5
Top-Hat method	Top-Hat	Structure shape: square, structure size:3×3
Infrared Patch-Image Mode	IPI	Patch size:50×50, sliding step:10, $\lambda = \frac{L}{\sqrt{\min(n_1, n_2, n_3)}}, L = 1, \varepsilon = 1e-7$
Reweighted Infrared Patch-Tensor Model	RIPT	Patch size:50×50, sliding step:10, $\lambda = \frac{L}{\sqrt{\min(n_1, n_2, n_3)}}, L = 1, h = 10, \varepsilon = 1e-7$
Reweighted Infrared Patch-Tensor Model Based on Weighted Tensor Nuclear norm	WNRIPT	Patch size:60×60, sliding step:30, $\lambda = \frac{L}{\sqrt{\max(n_1, n_2)n_3}}, L = 2, h = 10, \varepsilon = 1e-7$
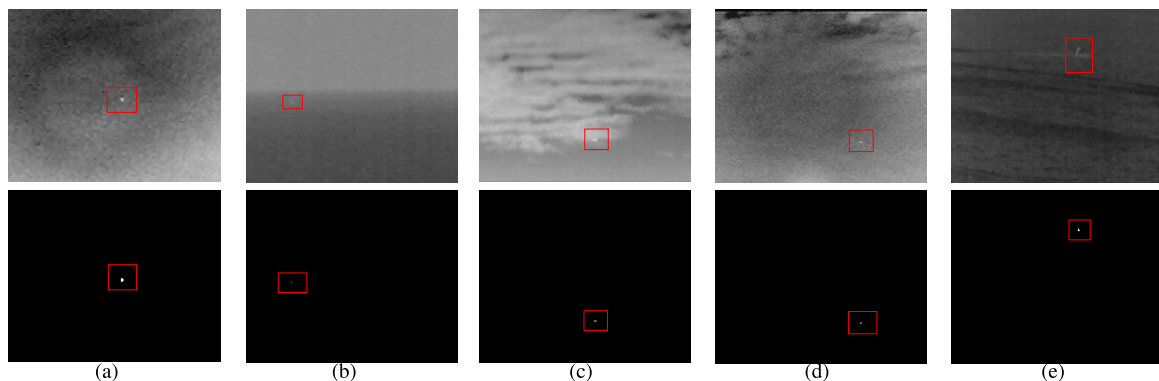
FIGURE 2. Single target scenes. The first row of (a)-(e) are five original infrared images for experiments. The second row of (a)-(e) are corresponding separated target images by the WNRIPT model.
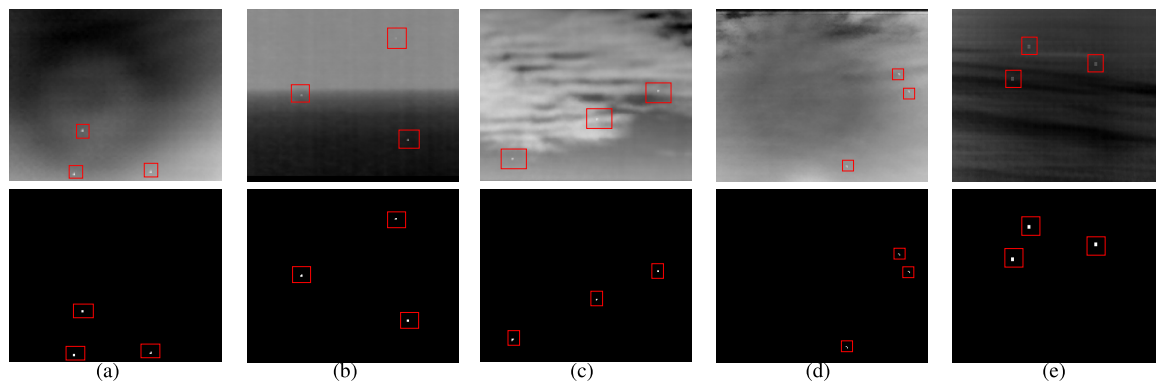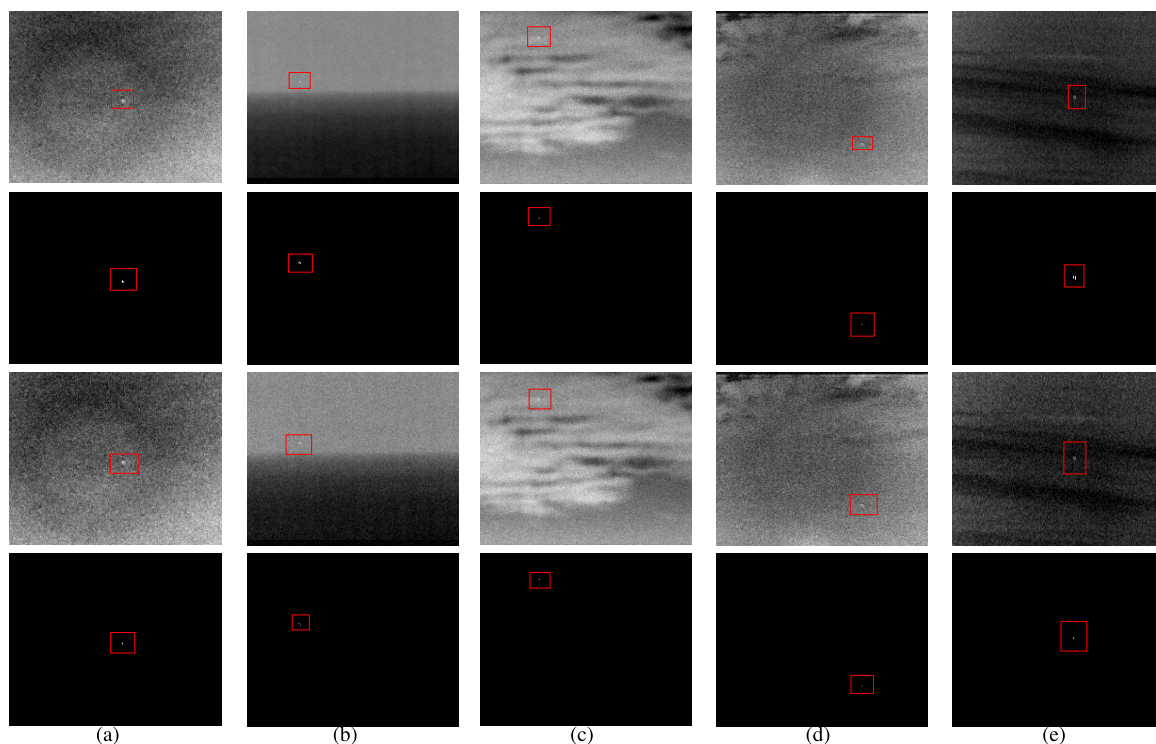
FIGURE 3. Multiple target scenes. The first row of (a)-(e) are five original infrared images for experiments. The second row of (a)-(e) are corresponding separated target images by the WNRIPT model.

FIGURE 4. The first and third rows of (a)-(e) are images contaminated by additive white Gaussian noise with standard deviation of 10 and 20 for experiments. The second and forth rows of (a)-(e) are corresponding separated target images by the WNRIPT model.

are presented in the second row. For observing conveniently, the targets are labeled with red boxes. It can be shown from Fig.2 that the background clutters are suppressed perfectly and each target is detected successfully.

2) PERFORMANCE OF MULTIPLE TARGETS SCENE

The first row of Fig.3 give five multiple targets scenes. A target is embedded into the background images as follows:

$$f_D(x, y) = \begin{cases} \max(pT_r(x - x_0, y - y_0), f_B(x, y)) \\ \qquad x \in (1 + x_0, am + x_0) \\ \qquad y \in (1 + y_0, bn + y_0) \\ f_B(x, y) \qquad \text{otherwise} \end{cases}$$

(29)

where T_r denotes resizing the targets with size of $am \times bn$ by using the bicubic interpolation.(x_0, y_0) is the position of the left upper corner of the target image in the background image produced randomly, p is weight of the target image also produced randomly within the range $[h, 255]$, h is the maximum value of the whole image [31]. And it can be seen from the second row of Fig.3 that the background clutters are suppressed clearly, leaving the true targets the only sole component.

3) PERFORMANCE OF NOISY CASES

The robustness to noise of a small target detection is another key metric for evaluating the method. We test the performance of our method in the noisy cases with different noise level, the first and third rows of Fig.4 are infrared images
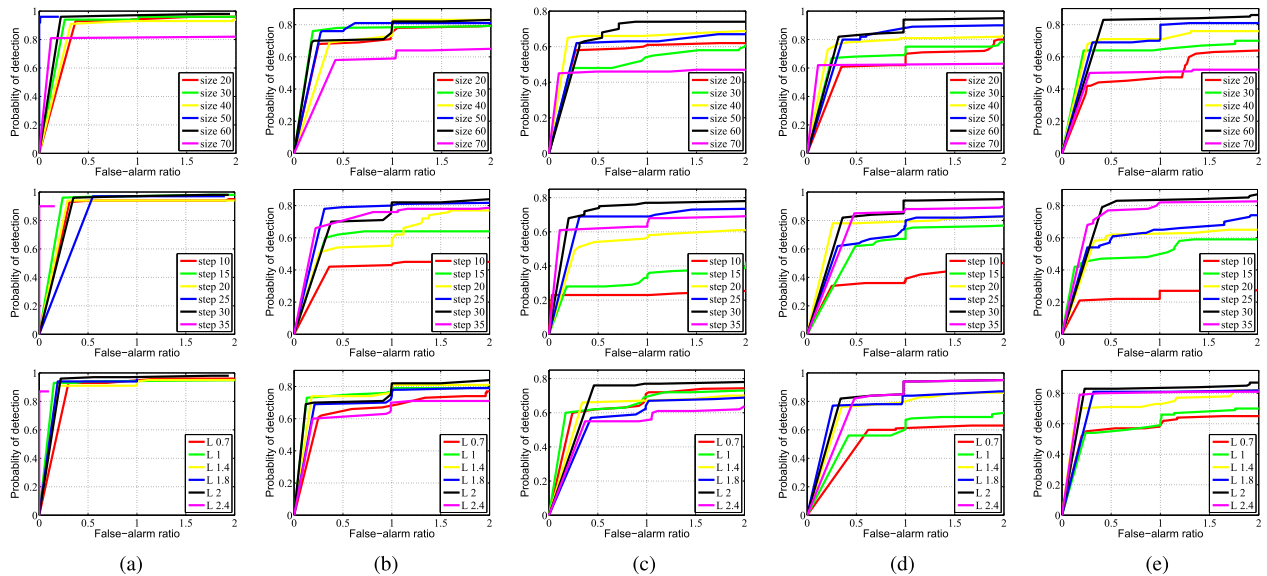
FIGURE 5. ROC curves for Sequences 1-5 with respect to different patch sizes, sliding steps and weighting parameters. Row 1: Different patch sizes. Row 2: Different sliding steps. Row 3: Different weighting parameters. (a) Sequence 1. (b) Sequence 2. (c) Sequence 3. (d) Sequence 4. (e) Sequence 5.

contaminated by additive white Gaussian noise with standard deviation of 10 and 20. And it can be seen from the corresponding second and fourth rows in Fig.4, the proposed model can still detect the dim target in severe noise cases.

C. PARAMETER ANALYSIS

To further analyze the influence of the parameters on performance of the proposed method. In this subsection, we evaluate the related parameters, including the patch size, sliding step and weighting parameter λ, which can affect the performance. And the values of parameters can be finely tunned for the optimal performance on different real infrared images. We evaluate these parameters on Sequence 1-5 by the ROC curves in Fig.5. It should be noted that the performances obtained by tunning only one parameter with others fixed can be suboptimal.

1) PATCH SIZE

Patch size is a decisive factor to the detection performance and computation complexity. A larger patch size usually can enhance the sparsity of the infrared small target. Nevertheless, it would increase the computation complexity and reduce the correlationship between the nonlocal background component. In order to find a tradeoff between target sparsity, background correlationship and computation complexity, we vary the patch size from 20 to 60 with 10 intervals and give the ROC curves in the first row of Fig.5. We can obtain the following conclusions from the results. First, the proposed method is not very sensitive to varying patch size. The performance of patch size among 30-60 is similar. Second, we can observe the results of patch size 20 and 70, it can be obtain that it's not optimal to set patch size too small or too large, because a too small patch size might cause high false-alarm rate while a too large patch size keeps a low detection probability.

2) STEP SIZE

The step size also has a great influence on the infrared image patch tensor size. Obviously, a larger step size can reduce the time of matrix SVD operation, namely, the computation complexity. In this experiment, we fix the patch size as 60×60 and var the step size from 10 to 35 with 5 intervals. The results are shown in the second row of Fig.5. It can be concluded that the performance of larger step size is better than that of smaller one. Because a smaller step size is easier to raise false alarms, especially for complex and noisy backgrounds, namely, Sequence 2-5. However, we can also observe that the performance step 35 is worse than that of step 30. Because too large step size might reduce the correlationship between background image patch-tensor and reduce detection probability.

3) WEIGHTING PARAMETER

We set $\lambda = L / \sqrt{\min (n_1, n_2) \, n_3}$ in our method. But it should be noted that it is possible to further improve the performance by tunning λ more carefully. We vary L from 0.7 to 2.4, and the results are shown in the third row of Fig.5. It can be concluded that a larger λ can obtain low false-alarm ration with acceptable performance. And the ROC curves of $L = 0.7$ and $L = 1$ suggest that it might not be suitable to set L too small, because their false-alarm ration are higher than that of other values of L.

D. COMPARISON WITH COMPETITIVE METHODS

In this subsection, the performance of the proposed model is compared with four other state-of-the-art methods to validate its superiority. The separated target images by 5 tested methods are shown in Fig.6 for five representative frames of real images, namely, Sequence 1-5. It can be observed that the classical Max-Median filter can enhance the targets
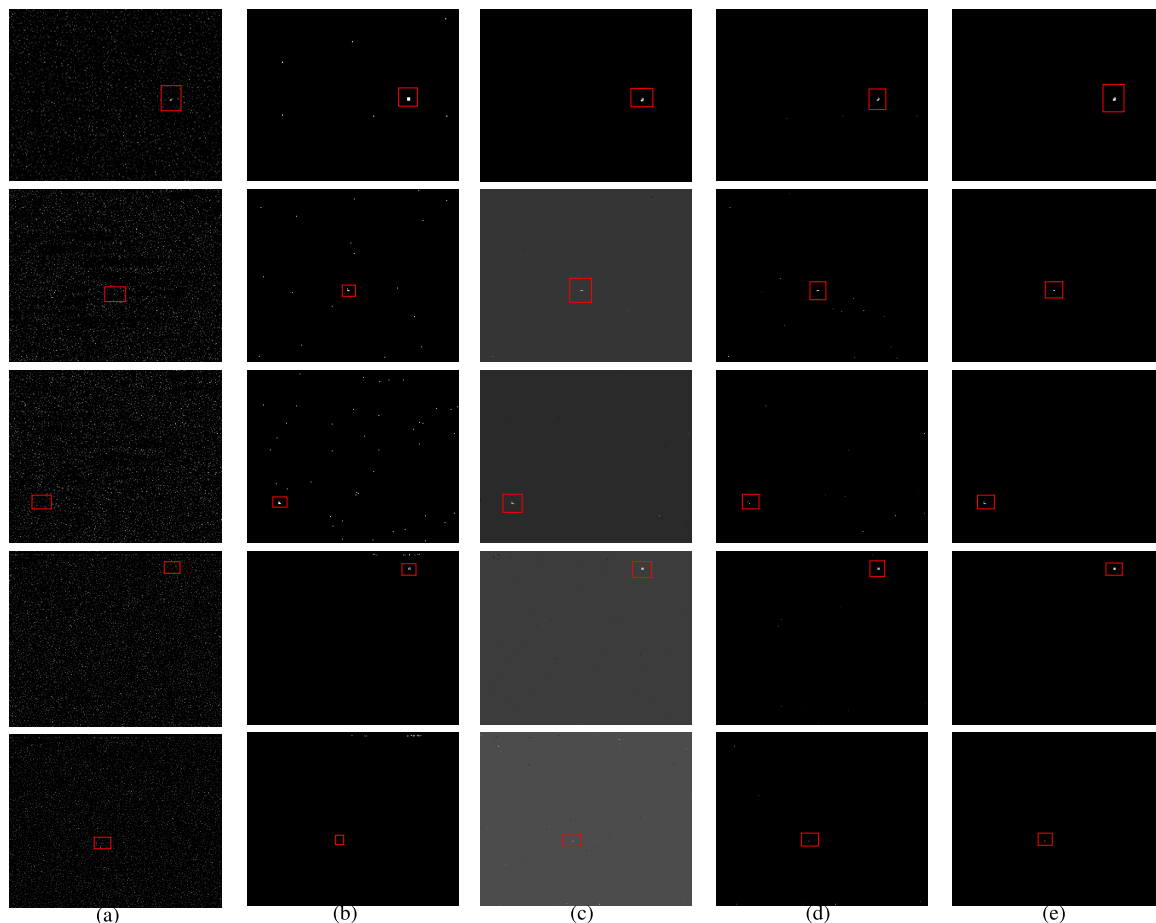
FIGURE 6. Separated target images of the representative frames in Sequence 1-5 by 5 tested methods. Row 1: 58th frame of Sequence 1. Row 2: 38th frame of Sequence 2.Row 3: 18th frame of Sequence 3.Row 4: 59th frame of Sequence 4.Row 5: 98th frame of Sequence 5. And (a)-(e) columns are the separated target images of (a) Max-median, (b) Top-hat, (c) IPI, (d) RIPT and (e) WNRIPT methods.

of Sequence 1-2 and Sequence 5 from the first column of Fig.6, however, many non-target pixels are enhanced simultaneously, which can cause high false-alarm ratios. And Max-Median filter fails in Sequence 3-4 that the targets are buried in clutters and noise. Top-Hat filter also does enhance the targets in Sequence 1-4, but many non-targets and clutters still remain in the target images. And it fails in Sequence 5. In addition, Top-Hat filter is very sensitive to the real target size. The unsatisfactory performance of the above two classical filter method is related to their strict preset target size, which is usually unknown in real datasets.

For the last three low-rank assumption based methods, the suppression of background clutters are relatively better than that of filter methods. It can be observed from the third column of Fig.6, the IPI model can detect all the targets in Sequence 1-5. Nevertheless, the separated target images still remain some background residuals in consequence of the deficiency effects. And the performance of RIPT is better than that of IPI, less background clutters remained in the target images. However, we can observe from the results on Sequence 2-5, a small amount non-target pixels are not wiped out clearly. Based on the weighted tensor nuclear norm,

we can see that the non-target pixels are suppressed throughly by the proposed model from the last column of Fig.6. For example, the target separated by RIPT is dimmer than the background residuals in the 18th frame of Sequence 3, which could raise false-alarm ratio. And we can observe the corresponding result by the proposed model, these remaining residuals are wiped out clearly. So the proposed model achieves the best performance in background clutters suppression among 5 tested methods.

In addition, the quantitative evaluation indices are used to compare the background suppression ability of the tested 5 methods. The results of 5 tested methods for the above representative frames in Sequence 1-5 are shown in Table 4 and 5. It can be concluded that the proposed method gets the highest score for all tested images and indices. And it should be noted that **Inf** are quite common for low-rank based methods, which denotes that the target neighborhood regions are suppressed to zero.

To further demonstrate the superiority of the proposed method over other methods, the ROC curves of Sequence 1-5 by 5 methods are provided for comparison in Fig.7. We can conclude that the performance of low-rank based method are

TABLE 4. Quantitative evaluation of the tested methods for the representative images of sequence 1-3.

Method	58th frame of Sequence 1				38th frame of Sequence 2				18th frame of Sequence 3			
	LSNRG	BSF	SCRG	CG	LSNRG	BSF	SCRG	CG	LSNRG	BSF	SCRG	CG
Max-Meidan	1.2252	8.4061	1.1965	1.5582	0.1579	14.4362	0.0049	1.2017	0.0856	8.5034	0.0390	1.0226
Top-Hat	2.4832	13.5647	3.2956	2.1328	1.2079	4.7691	2.4285	2.0685	1.1917	2.7428	2.0421	1.6806
IPI	**Inf**	**Inf**	**Inf**	3.3637	12.1579	12.6146	9.1429	2.7218	5.9032	5.3981	5.6811	2.0483
RIPT	**Inf**	**Inf**	**Inf**	3.3879	**Inf**	**Inf**	**Inf**	4.2005	5.9148	1713.8	31.2483	1.9463
WNRIPT	**Inf**	**Inf**	**Inf**	**3.7092**	**Inf**	**Inf**	**Inf**	**4.5659**	**Inf**	**Inf**	**Inf**	**3.2812**

TABLE 5. Quantitative evaluation of the tested methods for the representative images of sequence 4-5.

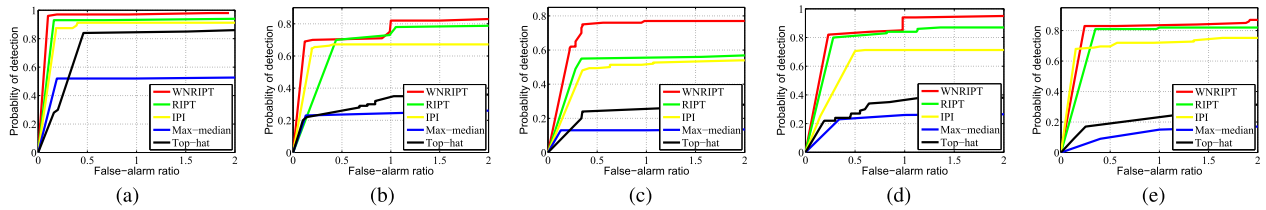| Method	59th frame of Sequence 4				98th frame of Sequence 5			
	LSNRG	BSF	SCRG	CG	LSNRG	BSF	SCRG	CG
Max-Meidan	0.4491	9.1182	0.3555	1.2264	0.8966	5.0552	0.9646	1.6940
Top-Hat	1.1633	2.4976	2.0439	2.8593	1.0374	1.7592	1.0964	2.3273
IPI	4.1893	3.2813	6.2871	2.9156	3.9082	2.5954	1.7171	2.6592
RIPT	3.3969	257.9780	46.2239	3.3992	**Inf**	**Inf**	**Inf**	4.6813
WNRIPT	**Inf**	**Inf**	**Inf**	**3.8113**	**Inf**	**Inf**	**Inf**	**4.7711**

FIGURE 7. ROC curves of detection results of five real infrared Sequences. (a) Sequence 1. (b) Sequence 2. (c) Sequence 3. (d) Sequence 4. (e) Sequence 5.

TABLE 6. Algorithm complexity and computational time comparisons of different methods.

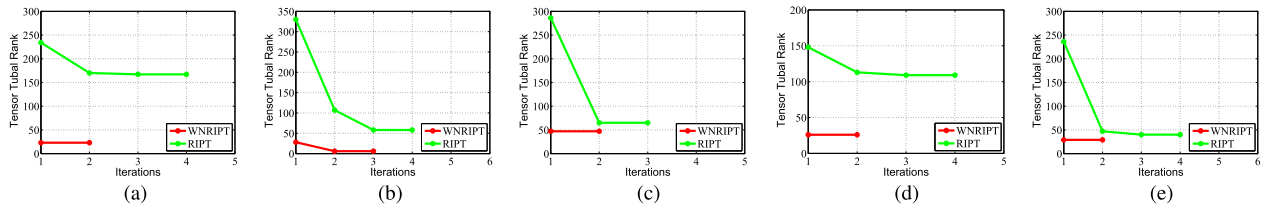| Method	Complexity	Times(s)
Max-Median	$\mathcal{O}(MNL^2)$	16.7189
Top-Hat	$\mathcal{O}(MNL^2 logL)$	0.1647
IPI	$\mathcal{O}(mn^2)$	32.9876
RIPT	$\mathcal{O}(kn_1n_2n_3 (n_1n_2 + n_2n_3 + n_1n_3))$	6.3353
WNRIPT	$\mathcal{O}(kn_1n_2n_3(logn_3 + (n_2 \lceil (n_3 + 1)/2 \rceil)/n_3))$	0.9335

FIGURE 8. Comparison of iteration rate and tubal rank of separated target images by RIPT and WNRIPT. (a) Sequence 1. (b) Sequence 2. (c) Sequence 3. (d) Sequence 4. (e) Sequence 5.

better than filter-based methods. The performance of RIPT method is better than IPI method. And the proposed method can achieve the highest detection probability than other methods for the same false-alarm ratio, which means it achieves the best performance.

E. ALGORITHM COMPLEXITY AND COMPUTATION TIME

The proposed model is equipped with the tensor nuclear norm and solved by ADMM, which can be guaranteed to converge [38]. The computation complexity and computation time for the above five representative frames in Sequence 1-5 with different methods are shown in Table 6. For Max-Median and Top-Hat filter methods, L denotes the size of the structure element, and the image size is $M \times N$. And for low-rank based methods, the computation complexity are mainly derived from the matrix SVD. For a patch-image with size of $m \times n$, the computation cost of matrix SVD is $\mathcal{O}(mn^2)$. For the image-patch tensor methods, let the input image tensor $\mathcal{D} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. The RIPT method needs to compute SVD of three matrix, namely, mode-1, mode-2 and mode-3 unfolding of input image tensor, and the corresponding matrix size are $n_1 \times (n_2 \cdot n_3)$, $n_2 \times (n_1 \cdot n_3)$ and $n_3 \times (n_1 \cdot n_2)$. The computation complexity of SVD are $n_1n_2n_3 (n_1n_2 + n_2n_3 + n_1n_3)$. Considering the iteration number k, the cost of RIPT method is $\mathcal{O}(kn_1n_2n_3 (n_1n_2 + n_2n_3 + n_1n_3))$. For the

proposed method, with the help of the property (7), it only needs to computing FFT and $\left\lceil \frac{n_3+1}{2} \right\rceil$ SVDs of $n_1 \times n_2$ matrices in each iteration, the whole complexity is $\mathcal{O}(kn_1n_2n_3(\log n_3 + (n_2 \lceil (n_3+1)/2 \rceil)/n_3))$, which can reduce the computational complexity dramatically.

Another important factor is the iteration number k. We compare iteration numbers and the tensor tubal rank of target image-patch tensor separated by RIPT and WNRIPT on the above five representative frames in Sequence 1-5. From Fig.8, it can be observed that the WNRIPT equipped with weighted tensor nuclear norm can converge with less iteration numbers and achieve smaller tensor tubal rank. The intrinsic reason is that the solution obtained by SNN is suboptimal. So the proposed model is more efficient than IPI and RIPT methods, which means WNRIPT is more suitable for real-time application.

V. CONCLUSION

To further improve the performance of infrared small target detection and background clutters suppression, a WNRIPT model is developed in this paper. By incorporating a weighted tensor nuclear norm, the low-rank background and sparse target separation problem is modeled as a tensor robust principal component model, which can be solved very efficiently by ADMM and t-SVT. Experimental results validate the superiority of our proposed method over other competitive methods in background suppression and target detection, and the proposed method is much more efficient than other low-rank based methods, which shows the potential for real-time application.

ACKNOWLEDGMENT

REFERENCES

[1] 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.

[2] X. Bai, Z. Chen, Y. Zhang, Z. Liu, and Y. Lu, "Infrared ship target segmentation based on spatial information improved FCM," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3259–3271, Dec. 2016.

[3] 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.

[4] 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.

[5] 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, p. 107007, 2005.

[6] X. Liu and Z. Zuo, "A dim small infrared moving target detection algorithm based on improved three-dimensional directional filtering," in *Proc. Chin. Conf. Image Graph. Technol.*, vol. 363, 2013, pp. 102–108.

[7] Z. Li et al., "Infrared small moving target detection algorithm based on joint spatio-temporal sparse recovery," *Infr. Phys. Technol.*, vol. 69, no. 9, pp. 44–52, 2015.

[8] L. Deng, H. Zhu, C. Tao, and Y. Wei, "Infrared moving point target detection based on spatial–temporal local contrast filter," *Infr. Phys. Technol.*, vol. 76, pp. 168–173, May 2016.

[9] D. Liu, J. Zhang, and W. Dong, "Temporal profile based small moving target detection algorithm in infrared image sequences," *Int. J. Infr. Millim. Waves*, vol. 28, no. 5, pp. 373–381, 2007.

[10] 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, pp. 076401-1–076401-9, Jul. 2006.

[11] Y. Dai and Y. Wu, "Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 8, pp. 3752–3767, Aug. 2017.

[12] J.-F. 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, Jul. 1996.

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

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

[15] 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.

[16] Y. Zhao, H. Pan, C. Du, Y. Peng, and Y. Zheng, "Bilateral two-dimensional least mean square filter for infrared small target detection," *Infr. Phys. Technol.*, vol. 65, no. 5, pp. 17–23, 2014.

[17] S. Kim, Y. Yang, J. Lee, and Y. Park, "Small target detection utilizing robust methods of the human visual system for IRST," *J. Infr., Millim., THz Waves*, vol. 30, no. 9, pp. 994–1011, Sep. 2009.

[18] X. Wang, G. Lv, and L. Xu, "Infrared dim target detection based on visual attention," *Infr. Phys. Technol.*, vol. 55, no. 6, pp. 513–521, Nov. 2012.

[19] S. Qi, J. Ma, C. Tao, C. Yang, and J. Tian, "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.

[20] 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.

[21] J. Han, Y. Ma, B. Zhou, F. Fan, K. Liang, and Y. Fang, "A robust infrared small target detection algorithm based on human visual system," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2168–2172, Dec. 2014.

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

[23] 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.

[24] Y. Bi, X. Bai, T. Jin, and S. Guo, "Multiple feature analysis for infrared small target detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 8, pp. 1333–1337, Jul. 2017.

[25] S. Kim, "Analysis of small infrared target features and learning-based false detection removal for infrared search and track," *Pattern Anal. Appl.*, vol. 17, no. 4, pp. 883–900, Nov. 2014.

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

[27] Z. Zha et al., "Analyzing the group sparsity based on the rank minimization methods," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2017, pp. 883–888.

[28] Z. Zha et al., "Non-convex weighted ℓ_p nuclear norm based admm framework for image restoration," *Neurocomputing*, vol. 311, pp. 209–224, Oct. 2018.

[29] Z. Zha, X. Zhang, Q. Wang, Y. Bai, and L. Tang. (2017). "Analyzing the weighted nuclear norm minimization and nuclear norm minimization based on group sparse representation." [Online]. Available: https://arxiv.org/abs/1702.04463

[30] Z. Zha, X. Yuan, T. Yue, and J. Zhou. (2018). "From rank estimation to rank approximation: Rank residual constraint for image denoising." [Online]. Available: https://arxiv.org/abs/1807.02504

[31] C. Q. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, and A. G. Hauptmann, "Infrared patch-image model for small target detection in a single image," *IEEE Trans. Image Process.*, vol. 22, no. 12, pp. 4996–5009, Dec. 2013.

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

[33] Y. Dai, Y. Wu, Y. Song, and J. Guo, "Non-negative infrared patch-image model: Robust target-background separation with partial sum minimization of singular values," *Infr. Phys. Technol.*, vol. 81, pp. 182–194, Mar. 2017.

[34] J. Guo, Y. Wu, and Y. Dai, "Small target detection based on reweighted infrared patch-image model," *IET Image Process.*, vol. 12, no. 1, pp. 70–79, 2018.

[35] J. Liu, P. Musialski, P. Wonka, and J. Ye, "Tensor completion for estimating missing values in visual data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 208–220, Jan. 2013.

[36] B. Romera-Paredes and M. Pontil, "A new convex relaxation for tensor completion," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2967–2975.

[37] S. Gu, L. Zhang, W. Zuo, and X. Feng, "Weighted nuclear norm minimization with application to image denoising," in *Proc. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2862–2869.

[38] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan. (2018). "Tensor robust principal component analysis with a new tensor nuclear norm." [Online]. Available: http://arxiv.org/abs/1804.03728

[39] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 5249–5257.

[40] C. Lu, C. Zhu, C. Xu, S. Yan, and Z. Lin, "Generalized singular value thresholding," *Comput. Sci.*, vol. 3, pp. 1805–1811, Jun. 2014.

[41] O. Rojo and H. Rojo, "Some results on symmetric circulant matrices and on symmetric centrosymmetric matrices," *Linear Algebra Appl.*, vol. 392, pp. 211–233, Nov. 2004.

[42] M. E. Kilmer and C. D. Martin, "Factorization strategies for third-order tensors," *Linear Algebra Appl.*, vol. 435, no. 3, pp. 641–658, 2011.

[43] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[44] Z. Lin, M. Chen, and Y. Ma. (2010). "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices." [Online]. Available: https://arxiv.org/abs/1009.5055

[45] C. Lu. (2018). *Tensor-Tensor Product Toolbox.* [Online]. Available: https://github.com/canyilu/tproduct

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

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

[48] C. Gao, T. Zhang, and Q. Li, "Small infrared target detection using sparse ring representation," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 27, no. 3, pp. 21–30, Mar. 2012.

[49] 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.

[50] C. Guo, Q. Ma, and L. Zhang, "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] 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.

JUNGANG YANG received the B.E. degree in electronic engineering, the M.E. degree, and the Ph.D. degree in information and communication engineering from the National University of Defense Technology (NUDT), Changsha, China, in 2007, 2008, and 2013, respectively. In 2011, he was a Visiting Ph.D. Student with The University of Edinburgh, Edinburgh, U.K. He is currently a Lecturer with the College of Electronic Science, NUDT. His research interests include computational imaging, image processing, compressive sensing, and sparse representation.

YUNLI LONG received the M.S. degree in electronic science and technology and the Ph.D. degree in information and communication engineering from the National University of Defense Technology, Changsha, China, in 2005 and 2012, respectively. He is currently a Post-Doctoral Fellow with the School of Electronic Science and Engineering, National University of Defense Technology. His current research interests include object detection and recognition, object tracking, and video surveillance.

ZHERAN SHANG received the B.S. and M.S. degrees from Wuhan Radar Academy, Wuhan, in 2014 and 2016, respectively. He is currently pursuing the Ph.D. degree with the National University of Defense Technology, Changsha, China. His current research interests include statistical signal processing and target detection.

YANG SUN received the B.S. and M.S. degrees in electronic engineering from Information Engineering University in 2014 and 2017, respectively. He is currently pursuing the Ph.D. degree in information and communication engineering with the National University of Defense Technology. His research interests are signal processing target detection and image processing.
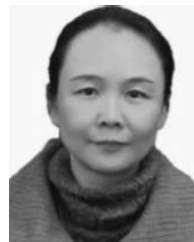
WEI AN received the Ph.D. degree from the National University of Defense Technology (NUDT), Changsha, China, in 1999. She was a Senior Visiting Scholar with the University of Southampton, Southampton, U.K., in 2016. She is currently a Professor with the College of Electronic Science, NUDT. Her current research interests include signal processing and image processing.

• • •