文章编号：1001-9014(XXXX)XX-0001-12

DOI:10.11972/j.issn.1001-9014.XXXX.XX.001

Infrared small target detection algorithm via partial sum of the tensor nuclear norm and direction residual weighting

SUN Bin[1*], XIA Xing-Ling[1], FU Rong-Guo[2], SHI Liang[3]

(1. College of Artificial Intelligence and Automation, Nanjing University Posts and Telecommunications, Nanjing 210023, China;
2. College of Electronic Engineering and Optoelectronic Technology, Nanjing University of Science and Technology, Nanjing 210094, China;
3. National Key Laboratory of Space Environment and Matter, Lanzhou Institute of Physics, Lanzhou 730030, China)

Abstract: Aiming at the problem that infrared small target detection faces low contrast between the background and the target and insufficient noise suppression ability under the complex cloud background, an infrared small target detection method based on the tensor nuclear norm and direction residual weighting was proposed. Based on converting the infrared image into an infrared patch tensor model, from the perspective of the low-rank nature of the background tensor, and taking advantage of the difference in contrast between the background and the target in different directions, we designed a double-neighborhood local contrast based on direction residual weighting method (DNLCDRW) combined with the partial sum of tensor nuclear norm (PSTNN) to achieve effective background suppression and recovery of infrared small targets. Experiments show that the algorithm is effective in suppressing the background and improving the detection ability of the target.

Key words: infrared small target detection, infrared patch tensor model, partial sum of the tensor nuclear norm, direction residual weighting

基于张量奇异值部分和与方向残差加权的红外小目标检测算法

孙 斌[1*], 夏星玲[1], 富容国[2], 史 亮[3]

(1. 南京邮电大学 自动化学院、人工智能学院,江苏 南京 210023;
2. 南京理工大学 电子工程与光电技术学院,江苏 南京 210094;
3. 兰州空间技术物理研究所 空间环境与物质作用全国重点实验室,甘肃 兰州 730030)

摘要:针对红外弱小目标检测在复杂云背景下面临的背景与目标对比度低、噪声抑制能力不足等问题,提出了一种基于张量核规范和方向残差加权的红外小目标检测方法。在将红外图像转换为红外块张量模型的基础上,从背景张量的低秩特性出发,利用背景与目标在不同方向上的对比度差异,设计了一种基于方向残差加权的双邻域局部对比度方法(double-neighborhood local contrast based on direction residual weighting method, DNLCDRW),并结合张量核范数部分和方法(partial sum of tensor nuclear norm, PSTNN),实现了对红外小目标的有效背景抑制和恢复。实验表,明该算法能有效抑制背景,提高对目标的检测能力。

关 键 词:红外弱小目标检测;红外块张量模型;张量核范数的部分和;方向残差加权

中图分类号:TP751 **文献标识码**:A

Introduction

Infrared small target detection technology plays a key role in practical applications, such as early warning systems, precision-guided weapons, missile tracking systems[1][2][3], and other military applications due to its advantages of good concealment and strong anti-interference capability. Generally speaking, small targets are

Received date: 2024-06-25, **revised date**: 2024-07-25 **收稿日期**:2024-06-25,**修回日期**:2024-07-25

Foundation items: Supported by Key Laboratory Fund for Equipment Pre-Research (6142207210202)

Biography: SUN Bin(1986-), male, Nanjing, associate professor. Research area involves optoelectronic detection and signal processing. E-mail: ffgz366@163.com

* **Corresponding author**: E-mail: ffgz366@163.com

submerged in the complex cloud or ground background with point-like information without obvious structure, and their intensity will be weakened by the interference of external clutter, coupled with the fact that the pixels of small targets account for a small percentage of the whole infrared image[4], which makes the task of infrared small target detection difficult. Therefore, scholars at home and abroad have made a lot of efforts to suppress the complex background, enhance the target strength, and improve the detection rate.

Infrared small target detection algorithms with good detection performance have been proposed in the past few years, which can be categorized into three groups: deep learning-based, local contrast-based, and low-rank sparse decomposition-based[5].

With the increased interest in the field of deep learning, more scholars proposed different deep learning frameworks to detect infrared small targets by training models with numerous samples. Wang et al. [6] used YOLO-based feature extraction backbone network for IR small target detection. Shi et al. [7] treated noise as a small target and used a denoising self-encoder with end-to-end modeling for denoising tasks. Wu et al. [8] proposed a novel multi-modal feature fusion network (MFFN), by fusing morphological features, infrared radiation and motion characteristics. However, the deep learning method is too dependent on a large amount of sample data learning, resulting in a long time, which is not suitable for application in daily production.

Inspired by the human visual system(HVS), the local contrast-based method extracts a local prior by mining the feature differences between the infrared small targets and the background to process the infrared images[9]. Chen et al. [10] proposed a detection method based on local contrast measure (LCM), which traverses the whole image using a nine-grid window and calculates the maximum and mean values of the grayscale of the local region through the non-similarity between the target pixel and its neighborhood, but the oversimplification of the processing leads to sensitivity to highlight noise, as well as time-consuming. Wei et al. [11] proposed a new multi-scale patch-based contrast measure (MPCM) metric for detecting small targets, which utilizes the product of the difference in the mean gray values of the center and edge sub-blocks in two opposite directions to suppress the background edges, but it cannot be correctly processed for low-contrast images, which is prone to cause false alarms. Zhang et al. [12] designed the local intensity and gradient (LIG) map by utilizing the intensity value of the target pixel greater than the pixel value of the local adjacent region and the regular distribution of the target gradient, which eliminates the effect of background edge clutter while extracting the target. Han et al. [13] proposed a multi-scale tri-layer local contrast measure (TLLCM), which was designed with a core layer, a retention layer, and a surrounding layer. By applying Gaussian filtering to the core layer, the target is effectively enhanced by combining the target shape and enhancing the pixel differences between the core layer and the surrounding local background. In addition, a double-layer local contrast measurement (DLCM) algorithm with a three-layer window structure consisting of a center window, an inner window, and an outer window was first proposed by Pan et al[14], which simultaneously achieves background suppression and reduces the elapsed time by calculating the difference in pixel contrast between the center window and the other two layers of the window, but does not achieve the enhancement of target intensity.

In recent years, from the perspective of low-rank sparse decomposition, the separation of target and background in infrared images has been successfully achieved by fully considering the sparsity of the target and the low-rank of the background and combining the optimization theoretical methods such as matrix complementation and tensor decomposition. Gao et al. [15] for the first time built the target detection on the infrared patch image (IPI) model, by constructing the infrared image into an infrared block image[16], but due to the lack of information in the infrared block image, resulted in unstable detection results. Dai et al. first extended the IPI model to the tensor domain and proposed the infrared patch tensor (IPT) [17] model, and innovatively introduced the local prior information, and further proposed the reweighted IPT (RIPT) model[17], which can effectively make up for the insufficiency of IPI for feature information extraction. Subsequently, Zhang et al. [18] proposed the NRAM model, which utilizes γ norms to eliminate the background residuals and increase the robustness. Zhang et al. [19] also proposed the self-regularized weighted spares (SRWS) model, which separates the background based on a dictionary-based algorithm, and separates the target by using the overlapping edge information (OEI) and l_1 norms of the background structural information, but this method is time-consuming, and it is not able to recognize some highlight backgrounds.

To improve the accuracy of the algorithm and suppress the highlighted background, this paper mainly makes the following contributions.

Firstly, the infrared image problem is converted into an infrared block tensor problem avoiding the problem of lost pixels converted to column vectors. Moreover, to better constrain the background, the joint regularization constraints and the partial sum of the tensor nuclear norm (PSTNN) to more appropriately characterize the low-rank property of background tensor, are used to achieve effective background suppression and recovery of infrared weak targets.

Secondly, to avoid the problem of insufficient structural information extraction, we focus on the local contrast difference between the target and the background, proposing a new double-neighborhood local contrast based on direction residual weighting method (DNLCDRW), which improves the algorithm's ability to suppress bright and complex backgrounds by constraining the sparse terms through DNLCDRW, and better improves the ability of IR weak target detection.

Thirdly, an effective solver is designed for the proposed model using the alternating direction multiplier

method (ADMM) algorithm and tensor singular value decomposition (t-SVD). Meanwhile, the convergence strategy of the algorithm is improved to realize the win-win situation of the algorithm in terms of complexity and computation time.

1 Obtaining the image tensor model

Generally, an infrared image with small target can be represented as follows[20]:

$$f_D = f_B + f_T + f_N \quad , \quad (1)$$

where f_D, f_B, f_T and f_N represent original image, background image, target image, and noise. On this basis, Gao et al. [16] proposed that the idea of traditional infrared weak target detection can be transformed into the problem of robust principal component analysis (RPCA) of matrices to construct the IPI model of the image as:

$$D = B + T + N \quad , \quad (2)$$

where D, B, T, N donate the infrared original image, background image, target image, and random noise. To obtain more information about the infrared image, as indicated in Fig. 1 we choose to construct it into an infrared patch-tensor model by traversing the infrared image from left to right through the application sliding window of size $k \times k$ according to the sliding step of size p, obtaining n patch images, and constructing the IPT model $D \in R^{k \times k \times n}$:

$$\mathcal{D} = \mathcal{B} + \mathcal{T} + \mathcal{N} \quad , \quad (3)$$

where $\mathcal{D}, \mathcal{B}, \mathcal{T}$, and $\mathcal{N}$ correspond to the infrared patch tensor of D, B, T, N. It is usually assumed that the background patch-tensor is low-rank, to verify this conjecture, the unfolding diagrams of the three modes $\mathcal{B}_1, \mathcal{B}_2$ and $\mathcal{B}_3$ are obtained by unfolding the infrared background patch-tensor, and according to the performance of each modal curve steeply decreasing and stabilizing near 0 in Fig. 2, it is proved that the infrared background patch-tensor along each mode unfolding is of low rank, and so the conjecture is valid. As well, the target patch-tensor is a sparse matrix for the reason that it only accounts for a tiny part of the entire image. Satisfying $\|T\|_0 \leqslant s$, in which s is an integer dependent on the number and size of small targets. In addition, we are aware that we can make full use of the feature information of different mode unfolding matrices to improve the performance of the detection model effectively.

Considering noise is additive white Gaussian noise, that $(N)_F \delta$, the small target detection problem is transformed into a convex optimization function solution problem for tensor models:

$$\min_{B,T} rank(B) + \lambda \|T\|_0 \text{ s.t } D = B + T \quad , \quad (4)$$

where λ represents a compromising parameter and $\|\cdot\|_0$ denotes the l_0 norm.

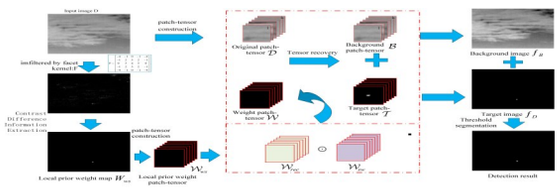


Fig. 1 Original image patch tensor construction
图1 红外块张量构建

2 Algorithm Design

2.1 Double-neighborhood local contrast based on direction residual weighting method(DNLCDRW)

The algorithm based on numerous human visual systems (HVS) has obvious edge suppression advantages in local prior algorithms and can more fully extract the prior information, so it can be utilized to obtain the target saliency map, which can be used as the local structure weights to weight the sparse target tensor in the IPT model. The performance of the algorithm can be improved to a certain extent by skillfully introducing the local a prior method into the infrared weak target detection algorithm based on the non-local a prior and effectively combining the advantages of the two.

In this chapter, a double-neighborhood local contrast detection model is designed for calculating the local weights to effectively suppress the large smooth background. Before that, face kernel filtering is utilized to enhance the difference between the target and the background, and the preprocessing operation Q is shown below:

$$Q = I(x,y) \times F^2 \quad , \quad (5)$$

where $I(x, y)$ represents the original infrared image and the facet kernel is defined as F:

$$F = \begin{bmatrix} -4 & -1 & 0 & -1 & -4 \\ -1 & 2 & 3 & 2 & -1 \\ 0 & 3 & 4 & 3 & 0 \\ -1 & 2 & 3 & 2 & -1 \\ -4 & -1 & 0 & -1 & -4 \end{bmatrix} \quad . \quad (6)$$

After the preprocessing operation, the algorithm is

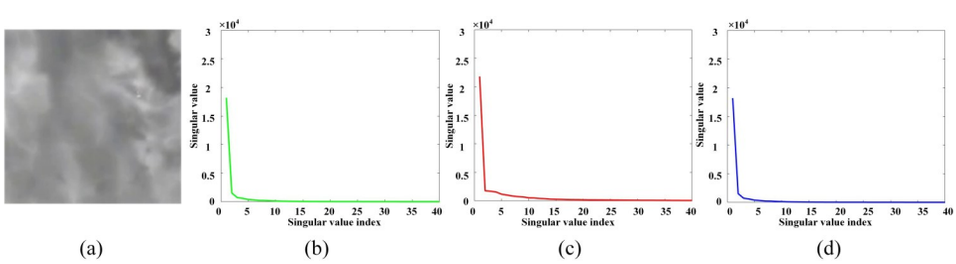
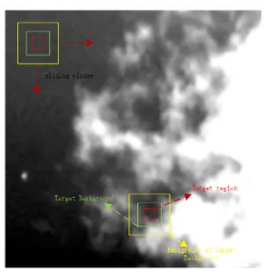


Fig. 2 Illustration of the low-rank nature of the infrared background patch-tensors: (a)one typical scene; (b)-(d): singular values of mode-1, mode-2, and mode-3 unfolding matrices of the corresponding patch-tensors
图2 红外背景块张量的低秩性质说明:(a)一个典型场景;(b)~(d) 相应光斑张量的模式-1、模式-2 和模式-3 展开矩阵的奇异值

designed using a sliding window as shown in Fig. 3 below, the double-neighborhoods sliding window contains a total of 5×5 sub-windows such that each sub-window contains $n \times n$ pixels. The central region is denoted as T and is called the target region, wrapped by two layers of proxies. The inner proxies are denoted as $IB_1 - IB_8$, and the outer proxies are denoted as $OB_1 - OB_{16}$.



OB_1	OB_2	OB_3	OB_4	OB_5
OB_{16}	IB_1	IB_2	IB_3	OB_6
OB_{15}	IB_8	T	IB_4	OB_7
OB_{14}	IB_7	IB_6	IB_5	OB_8
OB_{13}	OB_{12}	OB_{11}	OB_{10}	OB_9

Fig. 3 Structure of a sliding window with double-neighbourhoods
图3 双层滑动窗口的结构

To identify the target efficiently and suppress the highlighted background, the product of the difference in contrast between the target region and the two neighborhood domains, and between the inner neighborhood and the outer neighborhood, is used. CR defined as follows:

$$CR = DI \times DO \times \sum_{i=1}^{8} IO_i \quad , \quad (7)$$

$$DI = \begin{cases} A_k - \max\left(m_{IL_i}\right), A_k \max\left(m_{IL_i}\right) \\ 0, \quad \text{else} \end{cases} \quad , \quad (8)$$

$$DO = \begin{cases} A_k - \max\left(m_{OL_j}\right), A_k \max\left(m_{OL_j}\right) \\ 0, \quad \text{else} \end{cases} \quad , \quad (9)$$

$$IO_i = m_{IB_i} - m_{OB_{2i-1}}, \; i = 1,2,...,8 \quad , \quad (10)$$

where A_k is the average of the previous largest pixel grey values of the central block, m_{IL_i} and m_{OL_j} are the average of the pixel grey values of the ith inner background block and the average of the pixel grey values of the jth outer background block, respectively.

To further strengthen the weights and suppress the noise residue in the target, a new definition of t-he strengthened weight function E is proposed as follows:

$$E = \frac{\max\left(e_T - e_{IB_i}\right)}{\max\left(m_{IB_i}\right)\max\left(m_{OB_i}\right)} \quad . \quad (11)$$

$$e_T = \max\left(m_T\right) - \text{mean}\left(m_T\right)$$

$$e_{IB_i} = \max\left(m_{IB_i}\right) - \text{mean}\left(m_{IB_i}\right)$$

Therefore, the double-neighborhood contrast weighting algorithm based on directional residuals is defined as follows:

$$\text{DNLCDRW} = CR \times E \quad . \quad (12)$$

The local contrast a prior weights $DNLCDRW$ are then normalized and the tensor form $\mathcal{W}_p$ is then obtained by constructing a block tensor, with the W_p of the normalized prior weight map defined as:

$$W_p = \frac{\text{DNLCDRW} - \min(\text{DNLCDRW})}{\max(\text{DNLCDRW}) - \min(\text{DNLCDRW})} \quad . \quad (13)$$

Candès et al[21] proposed l_1 replace l_0 reweighting strategy to minimize the number of a-parameters that successfully enhance the sparsity of the target component. This optimization method mitigates the problem of decreasing the bounding ability of the number of l_1 when there are strong disturbances in the sparse tensor by mitigating the problem of the imbalance in the penalties between the coefficients and improving the computational time, the specific definitions are as follows:

$$\mathcal{W}_{sw}^{l+1} = \frac{c}{\left|\mathcal{T}^l\right| + \xi} \quad , \quad (14)$$

$$\mathcal{W} = \mathcal{W}_{rec} \odot \mathcal{W}_{sw}$$

where ξ is a tiny constant that prevents the denominator from being 0, c is mostly set to 1, $\mathcal{W}_{sw}$ is the value of the weight matrix at the current position, l represents the number of iterations, $\mathcal{W}_{rec}$ is the tensor corresponding to elementwise reciprocals of the corresponding elements $\mathcal{W}_p$, $\odot$ denotes Hadamard accumulation.

2.2 Partial sum of the tensor nuclear norm

In the IPT model, given that the IR image background is low-rank, its low-rankness can be reflected by the tensor rank. How decomposing the tensor leads to the definition of tensor rank is not unique, and there are two mainstream definitions nowadays, which are CP rank[22] and Tucker rank][23]. Compared with CP decomposition which cannot be calculated directly, scholars prefer Tucker rank, so minimizing Tucker rank becomes one of the keys to achieving background separation.

In the RIPT model, dai[17] constrains the background term with the sum of nuclear norms (SNN), which can be defined as $\sum_i \left(\mathcal{X}_{(i)}\right)_*$. This calculation is only applicable to simple matrix nuclear norms, for complex matrix nuclear norms multidimensional structural data is lost. The proposal of TNN[24] is defined as $\sum_{i=1}^{n_3} \left(\bar{X}^{(i)}\right)_*$ for $X \in R^{n_1 \times n_2 \times n_3}$ compromises the shortage of SNN. However, minimizing the TNN may lead to some unavoidable biases resulting in insufficient algorithmic detection performance. For this reason, Oh et al[25]. used the partial sum of minimized singular values (PSSV) instead of minimizing the nuclear norm. Jiang et al. [26] generalized the PSSV to the tensor version by using the partial sum of the nuclear norm of the tensor as $\mathcal{X} \in R^{n_1 \times n_2 \times n_3}$ nonconvex approximation of the tensor defined as :

$$\left\|\mathcal{X}\right\|_{\text{PSTNN}} = \left\|\bar{\mathcal{X}}^{(i)}\right\|_{p=N} \quad , \quad (15)$$

where $\left\|\cdot\right\|_{p=N}$ is PSSV, $\left\|\bar{X}^{(i)}\right\|_{p=N} = \sum_{k=p+1}^{\min(n_1,n_2)} \sigma_k\left(\bar{X}^{(i)}\right)$, $\sigma_k^{(i)}\left(\bar{\mathcal{X}}^{(i)}\right)$ is the ith largest singular value of $\bar{\mathcal{X}}^{(i)}$. From Eq. 15, it is found that for better highlighting the details of the image, PSTNN uses unequal distribution law. Smaller weights are assigned to larger singular values and larger weights are assigned to smaller singular values. Considering the complexity of the infrared background image and the fact that small targets make up a small per-

centage of the whole infrared image, a fixed energy ratio is used to determine the parameter N.

Thus the objective function is rewritten as:

$$\min(\mathcal{B})_{\text{PSTNN}} + \lambda\left(\mathcal{W}\odot\mathcal{T}\right)_1 \quad \text{s.t}\ \mathcal{D}=\mathcal{B}+\mathcal{T} \quad \mathcal{W}=\mathcal{W}_{\text{rec}}\odot\mathcal{W}_{\text{sw}} \quad . \quad (16)$$

2.3 The solution of the proposed model

There are two main existing optimization solution algorithms for solving the RPCA-based model, namely Accelerated Proximal Gradient (APG)[27] and Alternating Direction Method of Multipliers (ADMM)[28], which are chosen to solve the proposed model in this paper given the fast convergence and high solution accuracy performance of ADMM, and the augmented Lagrange function of Eq. 17 is defined as:

$$\mathcal{L}(\mathcal{B},\mathcal{T},\mathcal{W},\mathcal{Y}) = \|\mathcal{B}\|_{\text{PSTNN}} + \lambda\|\mathcal{W}\odot\mathcal{T}\|_1 + \langle\mathcal{Y},\mathcal{D}-\mathcal{B}-\mathcal{T}\rangle + \frac{\mu}{2}\|\mathcal{D}-\mathcal{B}-\mathcal{T}\|_F^2 \quad , \quad (17)$$

where $\mathcal{Y}$ is the Lagrange multiplier, μ is the penalty factor, $\langle\cdot\rangle$ denotes the inner product of the two tensors, and $\|\cdot\|_F$ is the Frobenius norm. ADMM decomposes the solution of the $\mathcal{L}(\mathcal{B},\mathcal{T},\mathcal{W},\mathcal{Y})$ into several subproblems and updates $\mathcal{B},\mathcal{T}$ by iterating, and at step $k+1$, $\mathcal{B},\mathcal{T}$ is updated as:

$$\mathcal{B}^{k+1} = \text{argmin}\|\mathcal{B}\|_{\text{PSTNN}} + \frac{\mu^k}{2}\left\|\mathcal{B}+\mathcal{T}^{k+1}-\mathcal{D}+\frac{\mathcal{Y}^k}{\mu^k}\right\|_F^2 , \quad (18)$$

$$\mathcal{T}^{k+1} = \text{argmin}\lambda\|\mathcal{W}\odot\mathcal{T}\|_1 + \frac{\mu^k}{2}\left\|\mathcal{B}^k+\mathcal{T}-\mathcal{D}+\frac{\mathcal{Y}^k}{\mu^k}\right\|_F^2 . \quad (19)$$

Theorem 1 (soft thresholding operator)[28] Let $\tau>0$, Given matrix $X, Y\in R^{n_1\times n_2}$, Define the l_1 norm minimization problem as:

$$\text{argmin}\tau\|X\|_1 + \frac{1}{2}\|X-Y\|_F^2 \quad . \quad (20)$$

The optimal solution $X=S_\tau(Y)$ can be found by using Eq. 21:

$$S_\tau(Y) = \text{sign}(Y)\times\max(|Y|-\tau,0) \quad . \quad (21)$$

Therefore, according to the theoretical $\mathcal{T}^{k+1}$ can be solved as:

$$\mathcal{T}^{k+1} = \mathcal{S}_{\frac{\lambda\mathcal{W}}{\mu^k}}\left(\mathcal{D}-\mathcal{B}^k-\frac{\mathcal{Y}^k}{\mu^k}\right) \quad . \quad (22)$$

Theorem 2 (partial singular value thresholding operator, PSVT)[29] Let $\tau>0$, $l=\min(n_1,n_2)$. Given matrix $X, Y\in R^{n_1\times n_2}$ can be decomposed by SVD, $Y=Y_1+Y_2=U_{Y_1}D_{Y_1}V_{Y_1}^{\text{H}}+U_{Y_2}D_{Y_2}V_{Y_2}^{\text{H}}$. U_{Y_1} and V_{Y_1} are the singular vector matrices corresponding to the N largest singular values, and U_{Y_2} and V_{Y_2} correspond to the ones from the (N+1)th singular value to the last singular value, thus the complex minimization problem of PSSV is defined as:

$$\text{argmin}\,\tau\|X\|_{p=N} + \frac{\beta}{2}(X-Y)_F^2 \quad . \quad (23)$$

The optimal solution of Eq. 23 is represented by the PSVT operator:

$$P_{N,\tau}(Y) = U_Y\left(D_{Y_1}+S_\tau\left[D_{Y_2}\right]\right)V_Y^{\text{H}} = Y_1 + U_{Y_2}S_\tau\left[D_{Y_2}\right]V_{Y_2}^{\text{H}}, \quad (24)$$

where $\tau=\lambda/\beta$, $D_{Y_1}=\text{diag}(\sigma_1^Y,\ldots,\sigma_N^Y,0,\ldots0)$, $D_{Y_2}=\text{diag}(0,\ldots0,\sigma_{N+1}^Y,\ldots,\sigma_l^Y)$, , and $S_\tau[x]$ is the soft-thresholding operator.

Hence, we can solve Eq. 15 by pseudocode given in Algorithm 1

Algorithm 1: Solve ($\mathcal{B}^{k+1}$) by PSVT

Input: $\mathcal{Z}^k=\mathcal{D}-\mathcal{T}^{k+1}-\frac{\mathcal{Y}}{\mu^k}\in\mathbb{R}^{n1\times n2\times n3}, \lambda, \mu^k$

Initialize: $\mathcal{B}=\text{zeros}(n_1,n_2,n_3)$

1. compute $\overline{\mathcal{Z}}^k=\text{fft}(\mathcal{Z}^k,[],3)$;
2. compute $\overline{\mathcal{B}}^{k+1}$ for $i=1,\ldots,n_3$

$\left(\overline{\mathcal{B}}^{k+1}\right)=\mathcal{P}_{N,\lambda/\mu^k}\left(\left(\overline{\mathcal{Z}}^k\right)^{(i)}\right)$ (operator $\mathcal{P}(\cdot)$

end for

3. compute $\mathcal{B}^{k+1}=\text{ifft}\left(\overline{\mathcal{B}}^{k+1},[],3\right)$;

At the same time, $\mathcal{Y}$ and μ update as:

$$\mathcal{Y}^{k+1} = \mathcal{Y}^k + \mu^k\left(\mathcal{D}-\mathcal{B}^{k+1}-\mathcal{T}^{k+1}\right) \quad , \quad (25)$$

$$\mu^{k+1} = \mu^k \quad . \quad (26)$$

From Algorithm 2, we can straightforwardly see the entire optimization process of the proposed model.

Algorithm 2 ADMM for solving the proposed model

Input: $\mathcal{D}, \mathcal{W}_p, \mu, \lambda, \alpha, N$

Output: $\mathcal{B}, \mathcal{T}$

Initialize: $\mathcal{B}_0=\mathcal{T}_0=\mathcal{Y}_0=0$, $\mathcal{W}_{\text{sw}}=1$, $\mathcal{W}=\mathcal{W}_{\text{rec}}\odot\mathcal{W}_{\text{sw}}$, $\mu=3\times10^{-3}$, $\rho=1.1$, $c=1$, $k=0$

while not converged

1. Fix the others and update $\mathcal{T}^{k+1}$ by Equation 22;
2. Fix the others and update $\mathcal{B}^{k+1}$ by Algorithm 1;
3. Fix the others and update $\mathcal{Y}^{k+1}$ by Equation (25);
4. Fix the others and update $\mathcal{W}^{k+1}$ by Equation (14);
5. Fix the others and update μ by Equation (26);
6. Check the convergence conditions:

$\left(\mathcal{T}^{k-1}\right)_0=\left(\mathcal{T}^k\right)_0$ or $\frac{\left(\|\mathcal{B}^{k-1}+\mathcal{T}^{k-1}-\mathcal{D}\|\right)_F^2}{\left(\|\mathcal{D}\|\right)_F^2}\varepsilon$;

7. Update k: $k=k+1$;

end while

8. Output: $\mathcal{B}=\mathcal{B}^k$, $\mathcal{T}=\mathcal{T}^k$

2.4 The whole procedure of the proposed model

The process of the proposed model is shown in Fig. 4 and described as follows:

(1) Local a prior weight map generation: first preprocessing through the surface kernel F, after processing through the double-neighborhood local contrast weighting algorithm to extract the prior information, to obtain the prior weight map W_p.

(2) Infrared patch-tensor construction: Utilizing the sliding window of size $k\times k$ from left to right, from top to bottom, n times, as shown in Fig. 3. The original infrared image D and the previous weight map W_p are transformed into the infrared patch tensor $\mathcal{D}\in R^{k\times k\times n}$ and weighted patch tensor $\mathcal{W}_p\in R^{k\times k\times n}$, respectively,

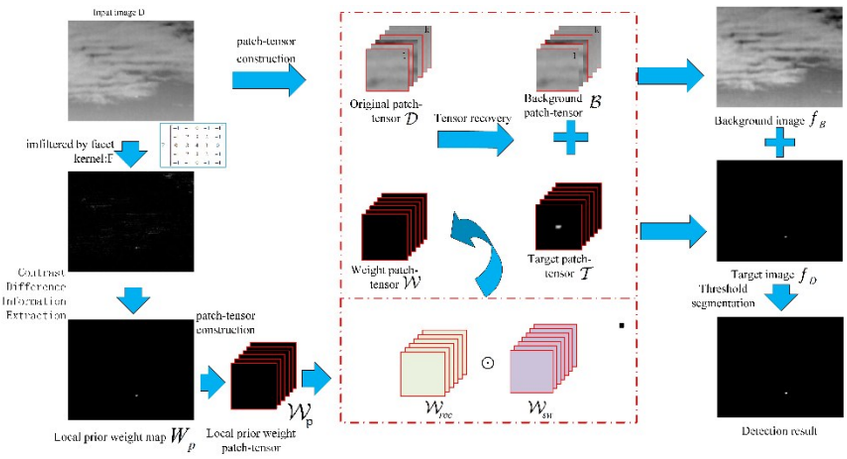


Fig. 4 The whole procedure of the proposed model
图4 所提模型整个流程

(3) Separation of background and target: Recover the low-rank patch-tensor $\mathcal{B}$ and the sparse patch-tensor $\mathcal{T}$ from $\mathcal{D}$ according to the algorithm 1.

(4) Image reconstruction: In contrast to the previously constructed approach, and for the overlapping parts with one-dimensional mean filtering, the infrared patch-tensor $\mathcal{B}$ and $\mathcal{T}$ reconstructed into background image f_B and target image f_T, with threshold segmentation to achieve the final weak target detection.

3 Experimental results

3.1 Experimental environment and parameter settings

We compare this paper's algorithm with the algorithms of Top-hat[31], MPCM[11], TLLCM[13], LIG[12], IPI[15], RIPT[17], NRAM[18] and SRWS[19]. The comparison experiments are conducted with 10 real infrared single-frame images as shown in Fig. 5, which have complex cloud cover, strong ground clutter interference, and wooded background, respectively, to reflect the comprehensive performance of the algorithm. All the experimental programs were run on a laptop computer configured with Intel Core i7-12700HQ processor and 16 G RAM, and the programming software used was MATLAB R2018b. In this chapter, experiments will be conducted on several existing public datasets, including SIRST dataset provided by Dai et al[33], and a dataset for infrared detection and tracking of dim-small aircraft targets underground / air background provided by Hui et al[34].

In addition to ten single-frame infrared images, three sequence infrared images are also selected in this paper to demonstrate the superiority of the algorithm in this paper more intuitively. Figures 6 (k)-(m) show the representative frames of the three infrared sequence images, where the targets and the zoomed-in image are described in the same red border. Table 1 is the description of the three infrared sequence images.

Several important parameter settings of the algorithm can have an impact on the performance of the algorithm, like sub-window size, patch size, sliding step, penalty factor μ, and compromising parameter λ, which have an impact on the robustness of different scenarios. To accommodate variations in target size, we set the sub-window size of CR to 3×3, and the size of each sub-window E to 5×5. we set the sliding window size to 40 × 40 and the sliding step size to 40. Setting this size balances the detection rate and real-time performance at the same time. The penalty factor μ controls the balance between the various components of the model background and the target. When the penalty factor is large, more of the target images can be preserved, but sparse clutter is intro-

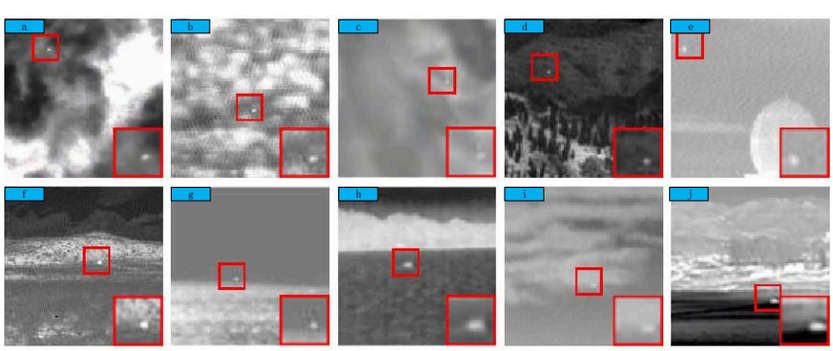


Fig. 5 Ten single-frame infrared images from different scenes(a)-(j)
图5 十幅不同场景的单幅红外图像(a)-(j)

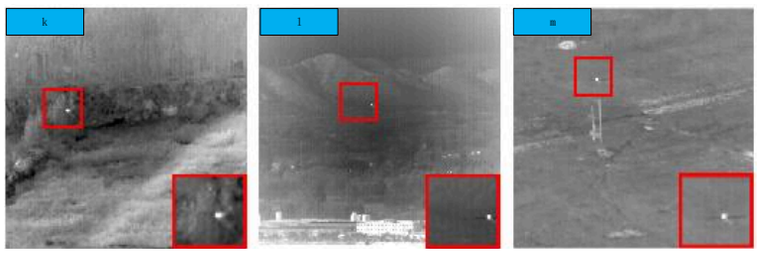


Fig. 6 Three infrared sequences of infrared image representative frames (k) - (m)
图6 红外图像代表帧(k)～(m)的三个红外序列

Table 1 Detailed descriptions of three real sequences
表1 三个真实序列的详细说明

	Frame Number	Size	Background Description	Target Description
Sequence 1	50	256x256	Vegetation cover and bright background	Tiny and regular shape
Sequence 2	50	256x256	Strong interference from artificial buildings	Single and tiny
Sequence 3	50	256x256	Strong ground clutter	Single and size change

duced. When the penalty factor is small, the introduction of non-target sparse clutter can be avoided, but the target may be over-sized. Hence, the best choice for the penalty factor μ is 3×10^{-3}. The compromising parameter λ also controls the balance between the low-rank background component and the sparse target component. When the appropriate value is chosen, the whole infrared image will retain the integrity of the target while suppressing the background clutter, so $\lambda = \frac{0.6}{\sqrt{\max(n1 \times n2) \times n3}}$ is proper.

3.2 Evaluation metrics

In this subsection, we select signal-to-clutter ratio gain (SCRG), and the background suppression factor (BSF) to evaluate the ability of background suppression and target enhancement, and in addition, the detection performance of the algorithm is verified by combining with received operating characteristic (ROC) curves. The dynamic relationship between the Probability of detection (Pd) and the False alarm rate (Fa) is quantitatively described. The three metrics are defined as follows:

$$\mathrm{SCRG} = \frac{\mathrm{SCR}_{\mathrm{out}}}{\mathrm{SCR}_{\mathrm{in}}}, \quad (27)$$

$$\mathrm{SCR} = \frac{|\mu_t - \mu_b|}{\sigma_b}, \quad (28)$$

$$\mathrm{BSF} = \frac{\sigma_{\mathrm{in}}}{\sigma_{\mathrm{out}}}, \quad (29)$$

where μ_t and μ_b denote the mean grayscale values of the target and its neighboring background in the local region, σ_b denotes the standard deviation of the grayscale of the target's neighboring background in the local region, σ_{in} and σ_{out} are the standard deviation of the grayscale of the background region of the image before and after processing by the algorithm, respectively. The higher the values of SCRG and BSF, the better the performance of background suppression and target enhancement.

According to the definition of the ROC curve, the algorithm with good performance requires that Pd is larger but Fa is smaller. Fig. 7 shows the local region, where we assume the size of the region target as $\alpha \times b$, the whole region size as $(a + 2d) \times (b + 2d)$, and $d = 20$.

$$\mathrm{Pd} = \frac{\text{Number of targets actually detected}}{\text{True number of targets in the sequence}}, \quad (30)$$

$$\mathrm{Fa} = \frac{\text{Number of incorrectly detected pixels}}{\text{Total number of pixels in the sequence}}. \quad (31)$$

3.3 Analysis of experimental results

Figures 8 and 9 show the detection results of this paper's algorithm for ten different single infrared scenes. It

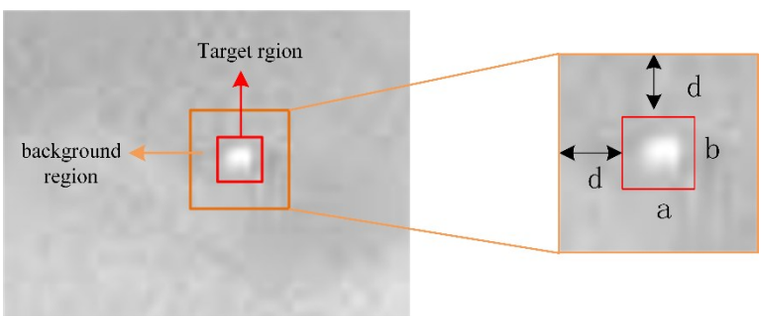


Fig. 7 Schematic representation of the target in the local area of the infrared image and the background area of its neighborhood
图7 红外目标在红外图像局部区域及其邻近背景区域的示意图

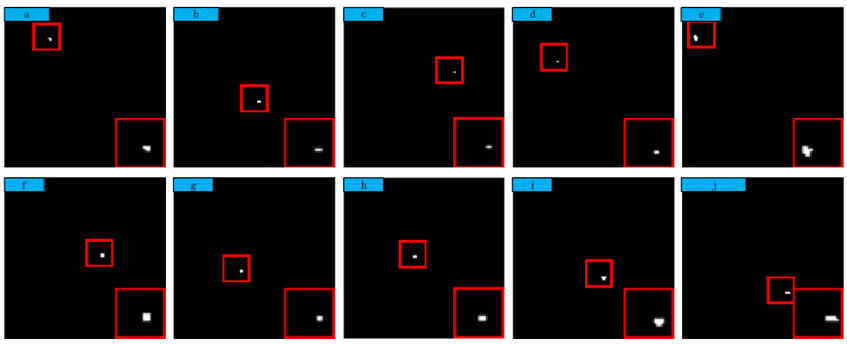


Fig. 8 Detection results of this paper's algorithm in ten different scenes
图8 本文算法在十种不同场景中的检测结果

can be seen that the small targets in the recovered target images are well-preserved, and various background components such as noise and clutter are also sufficiently suppressed, which illustrates that this paper's algorithm has a certain degree of infrared small target detection capability.

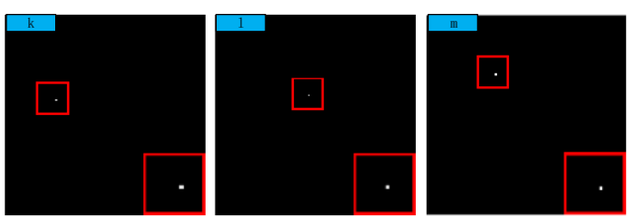


Fig. 9 Detection results of this paper's algorithm in three sequence scenes
图9 本文算法在三个序列场景中的检测结果

3.4 Robustness to various scenes

The variability of scenes causes the difficulties of target detection, so qualified detection algorithms must have the ability to adapt to different scenes, that is, scene robustness. From Fig. 8, it can be seen that the algorithm in this paper can realize the infrared detection task in ten different scenes. To make a better comparison, the five pictures Fig. (a)-(e) are picked out from them to be analyzed in detail. Figures (a)-(d) are all complex backgrounds, especially (a) and (b) surrounded by high brightness backgrounds and chaotic backgrounds; Figures(c) and (d) are in the dim backgrounds and have low target salience, and the surroundings in (e) are surrounded by the interference from man-made structures.

From Fig. 10 to Fig. 11, it can be seen that Top-hat handles the target very roughly, maintaining many background components while the target is detected. MPCM, TLLCM, and LIG are all HVS-based algorithms. MPCM utilizes the local patch difference to suppress the background clutter, but it is more sensitive to the background under strong noise interference. For example, when processing with the interference of the edge of the banded cloud in (a) and (b) are prone to cause false alarms. The ability of TLLCM for background edge and highlight background processing is still poor, for the scene Fig. 12 with uniform and single background change and low contrast. Due to the limitation of contrast variation, it is difficult for contrast-based methods to perform the target detection task better. IPI-based algorithms can detect most targets, but cannot ensure target integrity, which can be seen in Fig. 11. IPI also cannot suppress the background well when facing a bright background (the background is gray). NRAM has improved IPI to achieve background clutter suppression in different backgrounds and achieve good detection results. However, when facing the bright background in Fig. 10, RIPT, as a method that extends the detection problem to tensor direction, can better maintain the integrity of the target. However, when facing strong corners, background clutter will still exist. SRWS meets the detection needs of most scenarios and achieves good background suppression while detecting the target, but it cannot achieve the enhancement of the target intensity and will shrink the target too much. Therefore, it can be seen that our proposed algorithms perform better than the above algorithms and can extract the target well while suppressing the background.

3.5 Robustness to different noise

In practical applications, infrared target detection tasks are not only challenging due to the complex background but also invariably subject to the interference from diverse types of noise[32]. Therefore, whether proposed methods can suit changeable noises becomes another indispensable index of detection missions. In this paper, we experimentally evaluate the anti-noise performance of the model.

We choose to add Gaussian noise with mean 0 and standard deviations 4, 8, 12, 16, and 20 to the two infrared scenes. Comparison samples for each infrared image were obtained. Using the algorithm proposed in this paper, we generated the corresponding recovered images. Finally, the restoration effect of the infrared image was compared and referenced through the three-dimensional gray scale distribution map. Figures 15 and 16 show the detection results of this paper's algorithm on infrared small target images after adding Gaussian noise of different sizes. From the figure, it can be seen that the contrast of the target in the five infrared images after add-

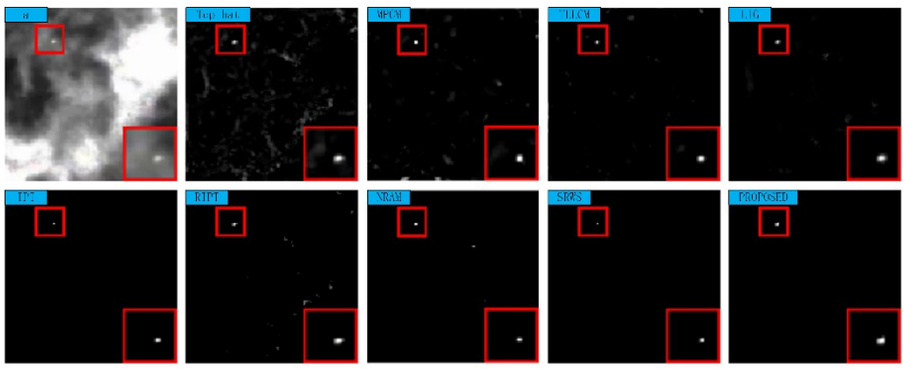


Fig. 10 Detection results of different algorithms for image (a)
图10 不同算法对图像(a)的检测结果

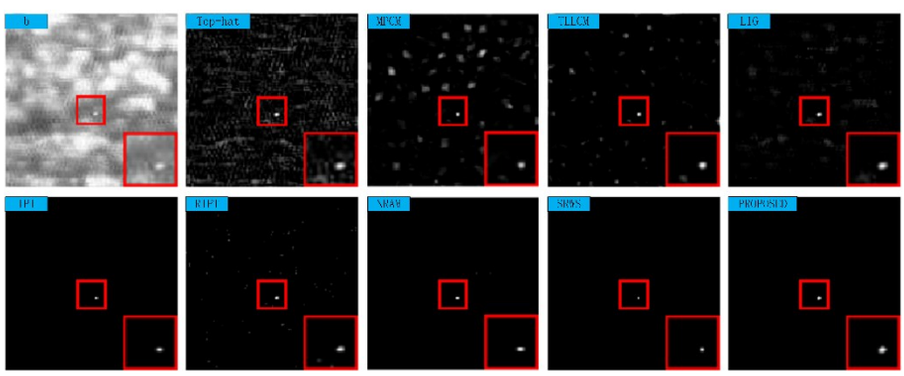


Fig. 11 Detection results of different algorithms for image (b)
图11 不同算法对图像(b)的检测结果

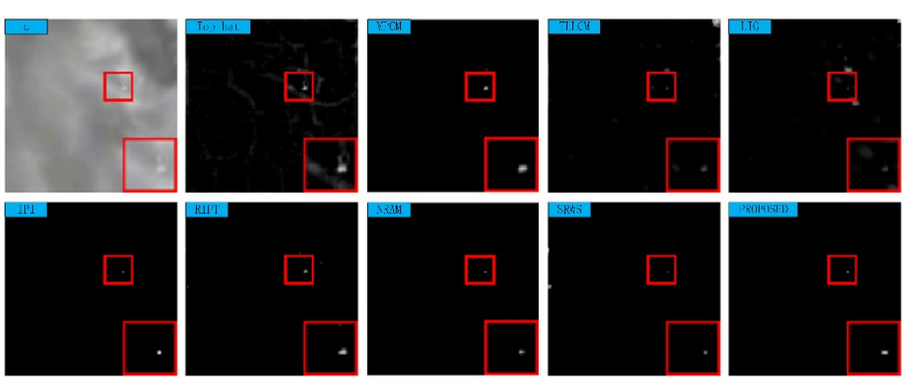


Fig. 12 Detection results of different algorithms for image (c)
图12 不同算法对图像(c)的检测结果

ing noise decreases, and target detection is more difficult. In all the added noise infrared images, the algorithm in this paper can correctly detect the target and effectively suppress the noise. However, when the noise is added to the infrared images 12 and 20, the integrity of the target is damaged to a certain extent due to the serious interference of strong noise. According to the above experiments, it is obvious that the algorithm in this paper has a certain degree of robustness when dealing with the noisy infrared images with variance within a certain range.

3.6 Quantitative analysis

In this subsection, to further evaluate the superiority of our algorithm and make a comprehensive comparison between nine algorithms, we utilize the quantitative evaluation metrics, including signal-to-clutter ratio gain (SCRG), background suppression factor (BSF), and ROC curves on the three infrared image sequences. From Table 2, it can be seen that although our algorithm is slightly inferior to NRAM in SCRG on seq. 3, it is much better than NRAM in background suppression. The proposed algorithm achieves the highest values of SCRG and BSF on seq. 1 and seq. 2. Thus, it can be

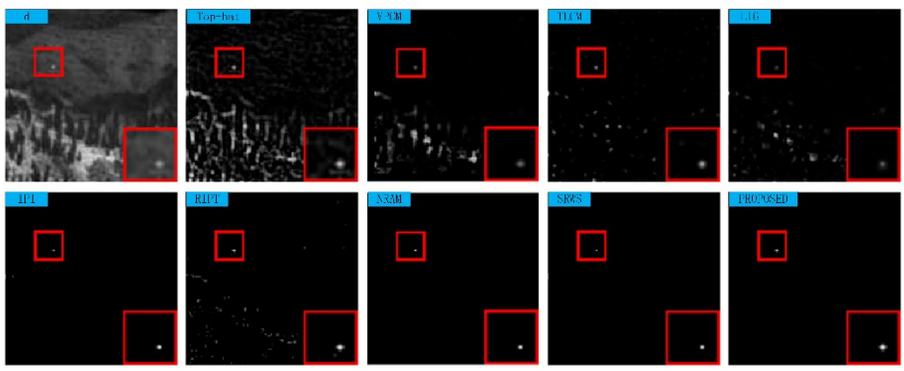


Fig. 13 Detection results of different algorithms for image (d)
图13 不同算法对图像(d)的检测结果

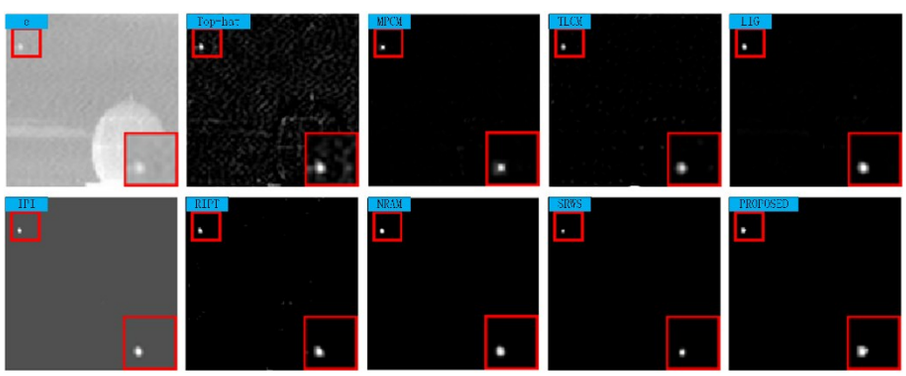


Fig. 14 Detection results of different algorithms for image (e)
图14 不同算法对图像(e)的检测结果

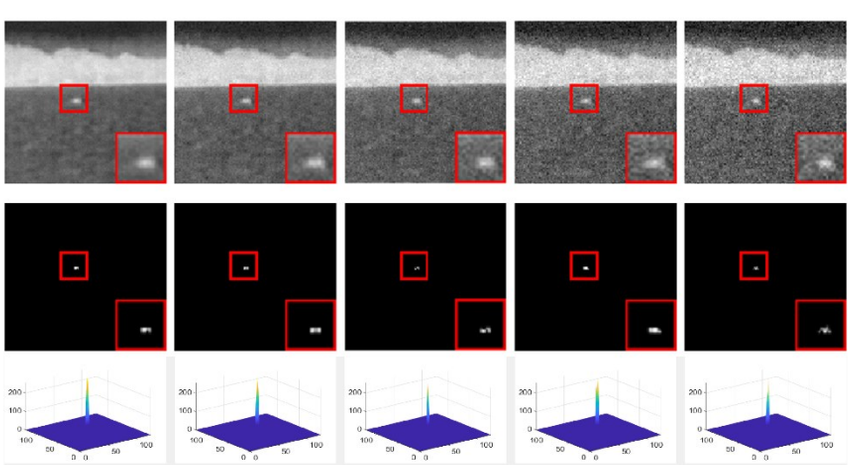


Fig. 15 Detection results of the proposed algorithm under Gaussian noise added to (h) with 4, 8, 12, 16, 20
图15 所提算法在图像(h)上添加4、8、12、16,20高斯噪声下的检测结果

seen that the algorithm proposed in this paper shows a great advantage in Table 2.

The detection performance of the external weak target detection algorithm mainly includes two aspects: detection capability and detection efficiency. The detection capability can be analyzed by the ROC curve, and the detection efficiency can be analyzed by the running time of the algorithm. In this experiment, we set thresholds of different sizes to obtain the detection rate and false alarm rate of the segmented image respectively, reflecting another detection ability of our method. As can be seen from Fig. 17 of the ROC curve, the LIG algorithm utilizes the pixel and gradient properties and achieves a high detection rate in most cases. The eight previous algorithms all suffer from the problem of large fluctuation of detection rate and false alarm rate on different sequences to some extent, and none of them is as stable as the algorithm proposed in this paper. The speed of the algorithms is also an important index in infrared small target detection. We tested and calculated the average

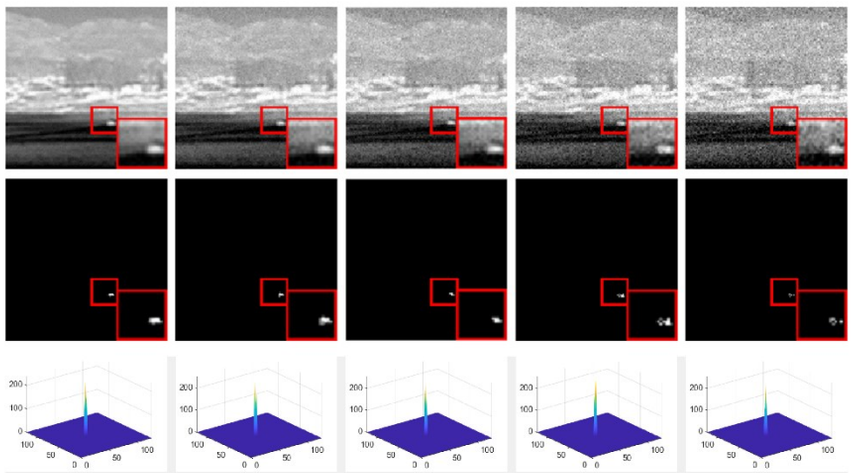


Fig. 16 Detection results of the proposed algorithm under Gaussian noise added to (i) with 4, 8, 12, 16, 20
图16 所提算法在图像(i)上添加 4、8、12、16,20高斯噪声下的检测结果

Table 2 SCRG and BSF values of the nine methods
表2 九种算法的SCRG和BSF值

Algorithm	Sequence 1 (k)		Sequence 2 (l)		Sequence 3 (m)	
	SCRG	BSF	SCRG	BSF	SCRG	BSF
TOPHAT	2. 3839	4. 10	7. 00	3. 718	3. 60	1. 04
LIG	19. 44	46. 60	201. 93	16. 49	38. 47	3. 98
TLCM	14. 29	18. 87	44. 71	27. 71	13. 34	4. 39
MPCM	10. 41	25. 52	77. 95	24. 80	13. 93	7. 78
IPI	2. 24	22. 45	2. 468	19. 76	9. 38	4. 83
RIPT	4. 78	15. 59	201. 34	12. 56	15. 34	2. 84
NRAM	29. 34	26. 01	200. 54	25. 05	185. 71	6. 65
SRWS	23. 82	34. 07	98. 93	24. 26	146. 25	10. 38
PROPOSED	180. 09	199. 46	205. 34	423. 12	157. 57	107. 85

Table 3 Average time of a single detected image frame for the nine algorithms (unit: s)
表3 九种算法检测单帧图像的平均时间(单位:s)

Algorithm	Sequence 1 (k)	Sequence 2 (l)	Sequence 3 (m)
TOPHAT	0. 0663	0. 1639	0. 0623
MPCM	0. 1619	0. 0968	0. 0979
TLLCM	0. 6792	0. 7470	0. 6939
LIG	0. 5079	0. 5051	0. 5048
IPI	3. 3639	3. 6799	2. 8455
RIPT	1. 0468	0. 8504	1. 7024
NRAM	0. 9831	1. 3488	1. 3830
SRWS	1. 5197	1. 4283	1. 6335
PROPOSED	0. 3568	0. 3836	0. 4448

single-frame running time of each algorithm on three sequences, and the traditional filtering algorithm Top-hat and the HVS-based algorithm MPCM have faster running time. The rest of the matrix optimization-based algorithms are slower. The algorithms in this paper are matrix optimization algorithms and still run slower compared to the traditional filtering-based methods, but this shortcoming is acceptable considering the performance advantages.

Based on all experimental results, our designed algorithm is compared to the above-mentioned algorithms, which confirms that partial sum of the tensor nuclear norm can better realize the constraints on the background image and recover the low-rank background. The target image is constrained using the direction residual weighting l_1 norm, which ensures the integrity of the target and improves the brightness of the target at the same time. The joint use of the two can achieve a better level

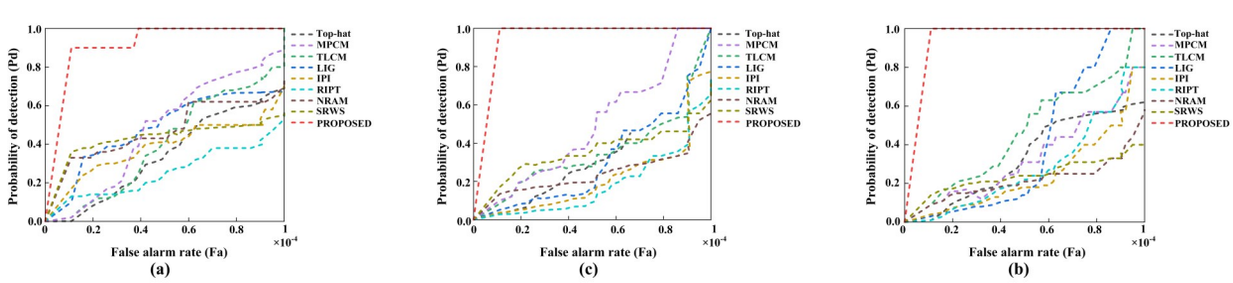


Fig. 17 ROC curves of detection results of three real sequences: (a) sequence 1; (b) sequence 2; (c) sequence 3
图17 三个真实序列检测结果的 ROC 曲线:(a) 序列 1;(b) 序列 2;(c) 序列 3

of performance and reduce algorithmic time consumption.

4 Conclusions

To simultaneously meet the need to improve the detection performance, speed up the processing speed, and further improve the robustness to noise and various scenes, this paper proposes an infrared small target detection algorithm based on partial sum of the tensor nuclear norm and direction residual weighting. Combining the local prior information extraction with the matrix optimization algorithm, the performance of the algorithm is evaluated through qualitative and quantitative analysis by selecting ten single-frame infrared images of typical scenes and three sequences of infrared images, and additionally selecting eight detection algorithms as baseline algorithms for comparison. Numerous experiments show that the algorithm in this paper has stronger scene robustness and noise robustness, more excellent detection performance, and both higher detection rate and lower false alarm rate in different scenes while completing the detection task in a reasonable time. In summary, the improvement of the algorithm in this paper has certain effectiveness and superiority.

References

[1] Kim S, Lee J. Scale invariant small target detection by optimizing signal-to-clutter ratio in heterogeneous background for infrared search and track [J]. Pattern Recognition, 2012, 45(1): 393-406.

[2] Liu Z, Zhou F, Chen X, et al. Iterative infrared ship target segmentation based on multiple features [J]. Pattern Recognition, 2014, 47(9): 2839-2852.

[3] Bai X, Chen Z, Zhang Y, et al. Infrared ship target segmentation based on spatial information improved FCM [J]. IEEE transactions on cybernetics, 2015, 46(12): 3259-3271.

[4] Mao X, Diao W. Criterion to evaluate the quality of infrared small target images [J]. Journal of Infrared, Millimeter, and Terahertz Waves, 2009, 30: 56-64.

[5] Zhang K, Ni S, Yan D, et al. Review of dim small target detection algorithms in single-frame infrared images [C]. 2021 IEEE 4th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). IEEE, 2021, 4: 2115-2120.

[6] Wang K, Li S, Niu S, et al. Detection of infrared small targets using feature fusion convolutional network [J]. IEEE Access, 2019, 7: 146081-146092.

[7] Shi M, Wang H. Infrared dim and small target detection based on denoising autoencoder network [J]. Mobile Networks and Applications, 2020, 25(4): 1469-1483.

[8] Wu D, Cao L, Zhou P, et al. Infrared small-target detection based on radiation characteristics with a multimodal feature fusion network [J]. Remote Sensing, 2022, 14(15): 3570.

[9] Kim S, Yang Y, Lee J, et al. Small target detection utilizing robust methods of the human visual system for IRST [J]. Journal of Infrared, Millimeter, and Terahertz Waves, 2009, 30: 994-1011.

[10] Chen C L P, Li H, Wei Y, et al. A local contrast method for small infrared target detection [J]. IEEE Transactions on Geoscience and Remote Sensing, 2013, 52(1): 574-581.

[11] Wei Y, You X, Li H. Multiscale patch-based contrast measure for small infrared target detection [J]. Pattern Recognition, 2016, 58: 216-226.

[12] Zhang H, Zhang L, Yuan D, et al. Infrared small target detection based on local intensity and gradient properties [J]. Infrared Physics & Technology, 2018, 89: 88-96.

[13] Han J, Moradi S, Faramarzi I, et al. A local contrast method for infrared small-target detection utilizing a tri-layer window [J]. IEEE Geoscience and Remote Sensing Letters, 2019, 17(10): 1822-1826.

[14] Han J Han J, Moradi S, Faramarzi I, et al. Infrared small target detection based on the weighted strengthened local contrast measure [J]. IEEE Geoscience and Remote Sensing Letters, 2020, 18(9): 1670-1674.

[15] Gao C, Meng D, Yang Y, et al. Infrared patch-image model for small target detection in a single image [J]. IEEE Transactions on Image Processing, 2013, 22(12): 4996-5009.

[16] Dai Y, Wu Y, Song Y. Infrared small target and background separation via column-wise weighted robust principal component analysis [J]. Infrared Physics & Technology, 2016, 77: 421-430.

[17] Dai Y, Wu Y. Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection [J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2017, 10(8): 3752-3767.

[18] Zhang L, Peng L, Zhang T, et al. Infrared small target detection via non-convex rank approximation minimization joint l 2, 1 norm [J]. Remote Sensing, 2018, 10(11): 1821.

[19] Zhang T, Peng Z, Wu H, et al. Infrared small target detection via self-regularized weighted sparse model [J]. Neurocomputing, 2021, 420: 124-148.

[20] Gu Y, Wang C, Liu B, et al. A kernel-based nonparametric regression method for clutter removal in infrared small-target detection applications [J]. IEEE Geoscience and Remote Sensing Letters, 2010, 7(3): 469-473.

[21] Candes E J, Wakin M B, Boyd S P. Enhancing sparsity by reweighted ℓ 1 minimization [J]. Journal of Fourier Analysis and Applications, 2008, 14: 877-905.

[22] Carroll J D, Chang J J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition [J]. Psychometrika, 1970, 35(3): 283-319.

[23] Tucker L R. Some mathematical notes on three-mode factor analysis [J]. Psychometrika, 1966, 31(3): 279-311.

[24] Semerci O, Hao N, Kilmer M E, et al. Tensor-based formulation and nuclear norm regularization for multienergy computed tomography [J]. IEEE Transactions on Image Processing, 2014, 23(4): 1678-1693.

[25] Oh T H, Kim H, Tai Y W, et al. Partial sum minimization of singular values in RPCA for low-level vision [C]. Proceedings of the IEEE International Conference on Computer Vision. 2013: 145-152.

[26] Jiang T X, Huang T Z, Zhao X L, et al. Multi-dimensional imaging data recovery via minimizing the partial sum of tubal nuclear norm [J]. Journal of Computational and Applied Mathematics, 2020, 372: 112680.

[27] Lin Z, Ganesh A, Wright J, et al. Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix [J]. Coordinated Science Laboratory Report no. UILU-ENG-09-2214, DC-246, 2009.

[28] Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers [J]. Foundations and Trends® in Machine learning, 2011, 3(1): 1-122.

[29] Hale E T, Yin W, Zhang Y. Fixed-point continuation for ℓ_1-*minimization*: Methodology and convergence [J]. SIAM Journal on Optimization, 2008, 19(3): 1107-1130.

[30] Zhang Z, Ely G, Aeron S, et al. Novel methods for multilinear data completion and de-noising based on tensor-SVD [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 3842-3849.

[31] Tom V T, Peli T, Leung M, et al. Morphology-based algorithm for point target detection in infrared backgrounds [C]. Signal and Data Processing of Small Targets 1993. SPIE, 1993, 1954: 2-11.

[32] Guan X, Zhang L, Huang S, et al. Infrared small target detection via non-convex tensor rank surrogate joint local contrast energy [J]. Remote Sensing, 2020, 12(9): 1520.

[33] Dai Y, Wu Y, Zhou F, et al. Asymmetric contextual modulation for infrared small target detection [C]. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 950-959.

[34] Hui B, Song Z, Fan H, et al. A dataset for infrared detection and tracking of dim-small aircraft targets under ground/air background [J]. China Sci. Data, 2020, 5(3): 291-302.