




Article

Infrared Small Target Detection Based on Tensor Tree Decomposition and Self-Adaptive Local Prior

Guiyu Zhang ^{1,2,3,†} , Zhenyu Ding ^{4,†}, Qunbo Lv ^{1,2,3,†}, Baoyu Zhu ^{1,2,3}, Wenjian Zhang ^{1,2,3}, Jiaao Li ^{1,2,3} and Zheng Tan ^{1,2,3,*}

- ¹ Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100094, China; zhangguiyu21@mails.ucas.ac.cn (G.Z.); lvqunbo@aoe.ac.cn (Q.L.); zhangwenjian21@mails.ucas.ac.cn (W.Z.); lijiaao21@mails.ucas.ac.cn (J.L.)
- ² Department of Key Laboratory of Computational Optical Image Technology, Chinese Academy of Sciences, Beijing 100094, China
- ³ School of Optoelectronics, University of Chinese Academy of Sciences, Beijing 100049, China
- ⁴ School of Artificial Intelligence, Xi'an Jiaotong University, Xi'an 710049, China; dzyxjtu@stu.xjtu.edu.cn
- * Correspondence: tanzheng@aircas.ac.cn
- † These authors contributed equally to this work.

Abstract: Infrared small target detection plays a crucial role in both military and civilian systems. However, current detection methods face significant challenges in complex scenes, such as inaccurate background estimation, inability to distinguish targets from similar non-target points, and poor robustness across various scenes. To address these issues, this study presents a novel spatial-temporal tensor model for infrared small target detection. In our method, we introduce the tensor tree rank to capture global structure in a more balanced strategy, which helps achieve more accurate background estimation. Meanwhile, we design a novel self-adaptive local prior weight by evaluating the level of clutter and noise content in the image. It mitigates the imbalance between target enhancement and background suppression. Then, the spatial-temporal total variation (STTV) is used as a joint regularization term to help better remove noise and obtain better detection performance. Finally, the proposed model is efficiently solved by the alternating direction multiplier method (ADMM). Extensive experiments demonstrate that our method achieves superior detection performance when compared with other state-of-the-art methods in terms of target enhancement, background suppression, and robustness across various complex scenes. Furthermore, we conduct an ablation study to validate the effectiveness of each module in the proposed model.

Keywords: infrared small target detection; tensor tree decomposition; self-adaptive local prior; spatial-temporal total variation



Citation: Zhang, G.; Ding, Z.; Lv, Q.; Zhu, B.; Zhang, W.; Li, J.; Tan, Z. Infrared Small Target Detection Based on Tensor Tree Decomposition and Self-Adaptive Local Prior. *Remote Sens.* **2024**, *16*, 1108. <https://doi.org/10.3390/rs16061108>

Academic Editors: Silvia Liberata Ullo and Li Zhang

Received: 17 February 2024

Revised: 17 March 2024

Accepted: 18 March 2024

Published: 21 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

In comparison to active radar imaging, infrared imaging offers the benefits of enhanced portability and improved concealment. Meanwhile, compared with visible light systems, it boasts a range of advantages, such as exceptional anti-interference features and the ability to operate all throughout the day [1]. Owing to the superior benefits of infrared imaging, infrared dim and small target detection plays a significant role in military and civil applications, such as aerospace technology [2], security surveillance [3], and forest fire prevention [4]. However, due to the lengthy detection distance, infrared targets usually occupy only a few pixels and lack shape information and textural features. In addition, infrared images in complex scenes often contain a variety of interferences (e.g., heavy clutter and prominent suspicious targets), resulting in a weak signal-to-clutter ratio (SCR) [5]. Therefore, infrared small target detection remains a challenging issue and has attracted widespread research interests.

1.1. Related Works

In general, infrared small target detection algorithms primarily include single-frame detection and sequential-frame detection [6]. For a long time, many single-frame detection approaches have been developed to address the challenges in infrared small target detection. Single-frame detection methods can be divided into four categories: (1) background consistency-based methods; (2) human visual system (HVS)-based methods; (3) deep learning (DL)-based methods; and (4) low-rank and sparse decomposition (LRSD)-based methods.

- Background consistency-based methods achieve target enhancement and background suppression based on the assumption of background consistency. Typical methods include the Top-hat filter [7], Max-Mean and Max-Median filters [8], and the high-pass filter [9]. Hadhoud and Thomas [10] extended the LMS algorithm [11] and proposed a two-dimensional adaptive least mean square filter (TDLMS). Cao and Sun [12] utilized the maximum inter-class variance method to improve morphological filtering. Although these methods are capable of achieving fast detection speeds, they are unsuitable for application in complex scenes.
- Contrast is the most crucial factor encoded in our visual system; HVS-based methods generally utilize visual saliency features to distinguish the target from the background. Chen et al. [13] proposed a local contrast method (LCM) to describe the difference between the target and its neighborhood. Inspired by LCM, many methods based on local contrast improvement have been proposed. Starting from the perspective of image patch difference, Wei et al. [14] presented a multiscale patch-based contrast measure (MPCM). Shi et al. [15] proposed a high-boost-based multiscale local contrast measure (HBMLCM). Han et al. [16] designed a multiscale tri-layer local contrast measure (TLLCM) to compute comprehensive contrast. Han et al. [17] improved the detection accuracy by utilizing the Laplacian filter and proposed a coarse-to-fine structure (MCFS) for infrared small moving target detection. However, when the image contains background edges and pixel-sized noises with high brightness (PNHB), such algorithms usually display high false alarms.
- With the development of artificial neural networks, DL-based methods have received extensive attention for their application in infrared target detection. Fan et al. [18] improved the convolutional neural network to extract infrared image features, aiming to improve detection accuracy and efficiency. Zhao et al. [19] designed an architecture of generative adversarial network (GAN), which models the detection problem issue as an image-to-image translation problem. In [20], a novel Dim2Clear network (Dim2Clear) was proposed to solve the problem of noise interference. Recently, Ying et al. [21] developed a label evolution framework with single point supervision. Although DL-based methods can achieve good detection performance under training scenes, their generalization to practical applications remains a challenge.
- In recent years, LRSD-based methods have achieved great success and can now effectively separate the low rank background and the sparse target of infrared image. Gao et al. [22] first proposed an infrared patch-image model (IPI) by constructing local patches. Consequently, infrared small target detection is transformed into an optimization problem. However, as the nuclear norm minimization (NNM) uses the same threshold to shrink singular values, an over-shrinkage problem may occur in complex backgrounds full of interference [23]. Furthermore, besides the target, edges and corners in the background are also considered as sparse components under the l_1 -norm [24]. To handle the above problems, Dai et al. [25] constructed a non-negative infrared patch-image model (NIPPS) by adding a non-negative constraint to the sparse target. Wang et al. [26] introduced the total variation regularization that better removes the noise and proposed a total variation regularization and principal component pursuit model (TV-PCP). Zhang et al. [27] designed a nonconvex rank approximation minimization (NRAM) by utilizing the $l_{2,1}$ -norm to constrain the remaining edges. Assuming that the background comes from multiple subspaces, the

stable multi-subspace learning model (SMSL) [28] and the self-regularized weighted sparse model (SRWS) [29] were proposed to improve detection performance. In order to better extract the image structure information and meet the practical demand for fast detection speed, Dai and Wu [30] adopted the tensor structure and proposed a reweighted infrared patch-tensor model (RIPT). Zhang and Peng [31] combined the partial sum of the tensor nuclear norm (PSTNN) and the local prior to effectively improve detection efficiency. In [32], the tensor fibered nuclear norm based on the Log operator (LogTFNN) was used to nonconvex approximate the tensor rank, which helps suppress background and noise. Zhang et al. [33] constructed a non-local block tensor and an adaptive compromising factor based on the image local entropy. Then, a self-adaptive and non-local patch-tensor model (ANLPT) was proposed for infrared small target detection.

Although the above LRSD-based single-frame detection methods have achieved good results, they ignore temporal information. Traditional sequential-frame detection methods, such as 3D matched filtering [34], dynamic programming algorithms [35], the spatiotemporal saliency approach [36], and trajectory consistency [37], face challenges in effectively separating the background from the target. In addition, these methods usually require prior knowledge, which is difficult to obtain in practical applications. In order to exploit the spatial-temporal information that is neglected in LRSD-based single-frame detection approaches, Sun et al. [38] stacked images from successive adjacent frames. Inspired by this, Zhang et al. [39] proposed a novel spatial-temporal tensor model with edge-corner awareness to further improve detection ability. Considering that the Laplace operator can approximate the tensor rank more accurately, Hu et al. [40] proposed a multi-frame spatial-temporal patch-tensor model (MFSTPT). Wang et al. [41] integrated the nonoverlapping patch spatial-temporal tensor model (NPSTT) and the tensor capped nuclear norm (TCNN) for detection results with low false alarms. Further, Liu et al. [42] designed a nonconvex tensor Tucker decomposition method, in which factor prior was used to obtain accurate background estimation and reduce computational complexity.

1.2. Motivation

Compared with background consistency-based approaches and HVS-based approaches, low-rank and sparse tensor decomposition (LRSTD)-based methods can better enhance small target features and suppress background clutter interference. Among these approaches, single-frame detection methods only consider single-frame image to construct the optimization model and struggle to achieve accurate results in various challenging environments with dynamic change or heavy clutter. Considering the significance of combining contextual information in the spatial-temporal domain, this article primarily concentrates on sequential-frame infrared target detection. While currently available methods have achieved relatively good detection performance, there are still some issues that need to be addressed.

First, due to the complex multilinear structure of the tensor, the exact approximation of the background tensor rank is always a major difficulty. To improve the accuracy of background estimation, these LRSTD-based methods [30,31,43] focus on designing more accurate tensor rank constraints, such as the sum of nuclear norm (SNN), tensor nuclear norm (TNN), and tensor train nuclear norm. Nevertheless, it has been proven that SNN fails to accurately approximate the tensor rank [44]. According to the definition of TNN, it lacks flexibility and the ability to measure low-rankness from multiple modes [45]. Although tensor train rank has a well-balanced matricization scheme, it suffers from higher storage requirements [46]. In summary, the approximation of the background tensor rank still needs to be improved. Therefore, we apply tensor tree rank to separate target and background. Compared with the above strategies, tensor tree decomposition is a more balanced method that splits the modes of a tensor in a hierarchical way.

In addition to accurate background tensor estimation, the suppression of strong edges and corner points is key to achieving good detection performance. The local structure

prior is often used to suppress interference. The RIPT only focuses on the edge structure information of the background, which may lead to false alarms. Likewise, the fixed prior weights used in PSTNN and MFSTPT cannot effectively suppress clutter in diverse scenes with different levels of interference. It is important to balance the enhancement of the target and the suppression of the interference from edges and corners in different scenes. To solve this problem, we propose a self-adaptive local prior method to adaptively suppress background clutter. Moreover, we use spatial-temporal total variation (STTV) to explore local smooth information. This strategy helps us to better remove the background noise. By combining tensor tree decomposition, self-adaptive local prior, and STTV, our method can accurately detect small targets. In the following sections, we refer to the proposed method as the TTALP-TV method. We present the results of qualitative and quantitative experiments to demonstrate that TTALP-TV surpasses other state-of-the-art methods in terms of target enhancement and background suppression in various complex scenes. Figure 1 presents the flowchart of our method. The main contributions of this article can be summarized as follows:

- (1) In order to approximate the tensor rank function more flexibly and accurately, we introduce tensor tree decomposition to exploit spatial and temporal correlation through a hierarchical structure.
- (2) The self-adaptive local prior is proposed as a target weight, which can not only better extract target information but also more effectively remove background clutter. Simultaneously, we impose STTV regularization constraint on the background to preserve image details and reduce noise interference.
- (3) We integrate the tensor tree rank, self-adaptive local prior, and STTV for infrared small target detection. An efficient optimization scheme using the alternating direction multiplier method (ADMM) is introduced to solve the proposed model.

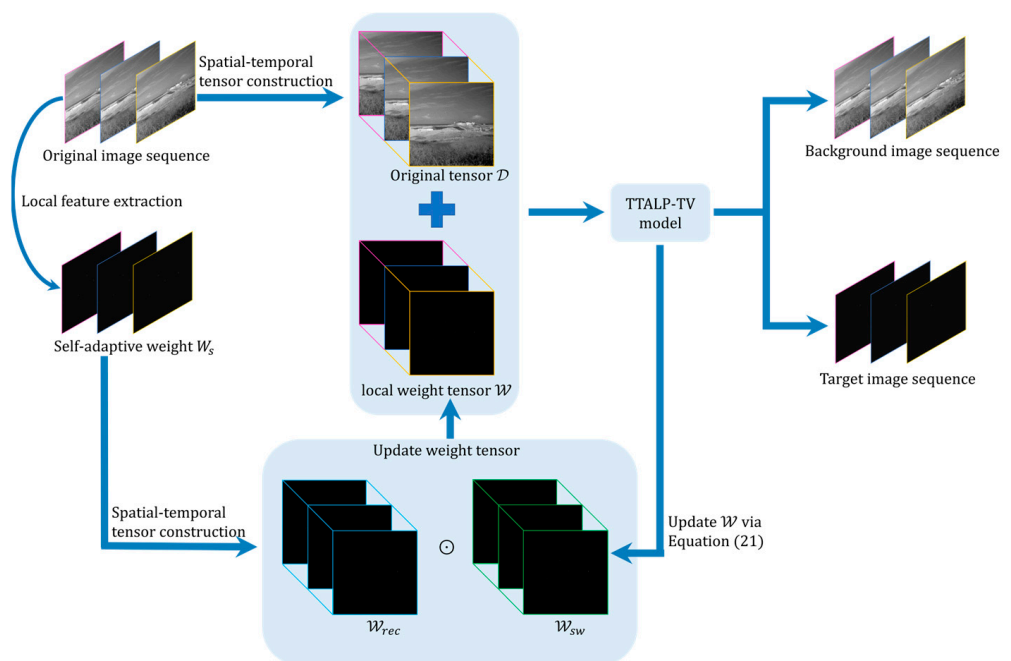


Figure 1. Flowchart of the proposed TTALP-TV model for infrared small target detection.

The remaining sections of this article are organized as follows. Section 2 summarizes the notations and preliminaries of tensor tree decomposition. Section 3 introduces the TTALP-TV model and describes its optimization procedure in detail. In Section 4, we demonstrate the effectiveness of the proposed algorithm through extensive experiments and analyses. Finally, Section 5 concludes this article and discusses the future work.

2. Notations and Preliminaries

This section introduces the essential notations and preliminaries used in this research. In this paper, we use lowercase letters (e.g., x), boldface lowercase letters (e.g., \mathbf{x}), boldface capital letters (e.g., \mathbf{X}), and Euler script (e.g., \mathcal{X}) to represent scalars, vectors, matrices, and tensors, respectively. The D th-order tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$, and $\mathbb{D} = \{1, \dots, D\}$. The node- q tensor-matrix multiplication can be denoted as $\mathcal{Y} = \mathcal{X} \times_q U$, where $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$, $U \in \mathbb{R}^{I \times I_q}$, $\mathcal{Y} \in \mathbb{R}^{I_1 \times \cdots \times I_{q-1} \times I \times I_{q+1} \times \cdots \times I_D}$, and q is the node in the tensor tree format.

The specific explanations of the symbols used are given in Table 1.

Table 1. Mathematical symbols.

| Notation | Explanation |
|--|--|
| $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ | 3th-order tensor |
| $\mathcal{X}(i, j, k) / x_{i,j,k}$ | Its (i, j, k) -th element |
| $\langle \mathcal{X}, \mathcal{Y} \rangle$ | The inner product of two tensors, $\sum_{i_1, i_2, i_3} x_{i_1, i_2, i_3} y_{i_1, i_2, i_3}$ |
| $\ \mathcal{X}\ _0$ | The l_0 -norm, the number of non-zero elements in \mathcal{X} |
| $\ \mathcal{X}\ _1$ | The l_1 -norm, the sum of non-zero elements in \mathcal{X} |
| $\ \mathcal{X}\ _F$ | The Frobenius norm, the square root of the sum of the squares of all elements in \mathcal{X} |
| $\ X\ _*$ | The matrix nuclear norm, the sum of all singular values in X |

Tensor Tree Network

Definition 1 (Tensor tree structure) [47]. For D th-order data, we define a binary tree \mathbb{T} with root \mathbb{D} as its dimension tree. Each node $\mathbb{C}_q \subset \mathbb{T}$, $q = 1, \dots, Q$ possesses the following attributes:

1. The node with only one entry is a leaf, i.e., $\mathbb{C}_p = \{d\}$. The set of all leaf nodes can be represented as follows:

$$\mathbb{F}(\mathbb{T}) = \{\mathbb{C}_p \subset \mathbb{T} \mid \mathbb{C}_p \text{ is a leaf node of } \mathbb{T}, p = 1, \dots, P\} \quad (1)$$

where P is the number of leaves.

2. The node consisting of two disjoint successors is an interior node. The set of all interior nodes is denoted by:

$$\mathbb{E}(\mathbb{T}) = \mathbb{T} \setminus \mathbb{F}(\mathbb{T}) \quad (2)$$

And $Q - P$ represents the number of interior nodes.

3. The tree distance $h(\mathbb{C}_q)$ is the distance between the node $\mathbb{C}_q \subset \mathbb{T}$ and the root, with a maximum depth of H . At depth h , P_h and Q_h denote the number of leaves and total nodes, respectively.

Definition 2 (Matricization) [47]. Given a node of dimension indices $\mathbb{C}_q \subset \mathbb{D}$ and its complement $\bar{\mathbb{C}}_q = \mathbb{D} \setminus \mathbb{C}_q$, the matricization of a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$ is defined as:

$$X_{[q]} \in \mathbb{R}^{I_{\mathbb{C}_q} \times I_{\bar{\mathbb{C}}_q}} \quad (3)$$

where $I_{\mathbb{C}_q} = \prod_{c \in \mathbb{C}_q} I_c$ and $I_{\bar{\mathbb{C}}_q} = \prod_{\bar{c} \in \bar{\mathbb{C}}_q} I_{\bar{c}}$.

Definition 3 (Tensor tree rank) [47]. Let \mathbb{T} be a dimension tree of a D th-order tensor, the tensor tree rank is the set of ranks of the matricization for each node, in the form of:

$$\text{rank}_{\text{tree}} = \left\{ k_q \mid k_q = \text{rank}(X_{[q]}), \forall \mathbb{C}_q \subset \mathbb{T} \right\} \quad (4)$$

Definition 4 (Tensor tree decomposition) [47]. Given $\mathcal{X} \in \mathbb{R}^{I_1 \times \dots \times I_D}$, for every node $\mathbb{C}_q \subset \mathbb{T}$, $X_{[q]}$ can be written as:

$$X_{[q]} = U_q V_q^T, \quad U_q \in \mathbb{R}^{I_{\mathbb{C}_q} \times k_{\mathbb{C}_q}} \quad (5)$$

where $k_{\mathbb{C}_q}$ is the standard matrix rank of $X_{[q]}$. For each $\mathbb{C}_q \subset \mathbb{E}(\mathbb{T})$ with two disjoint successors \mathbb{C}_{q_1} and \mathbb{C}_{q_2} , the column vectors $\mathbf{u}_q(:, l)$ of U_q can be expressed as:

$$\mathbf{u}_q(:, l) = \sum_{l_1=1}^{k_{q_1}} \sum_{l_2=1}^{k_{q_2}} \mathcal{G}_q(l, l_1, l_2) \mathbf{u}_{q_1}(:, l_1) \otimes \mathbf{u}_{q_2}(:, l_2) \quad (6)$$

where $\mathcal{G}_q(l, l_1, l_2)$ is the coefficient of the linear combination. Figure 2 graphically illustrates the tensor tree decomposition of a 4th-order tensor, providing an intuitive understanding of its structure.

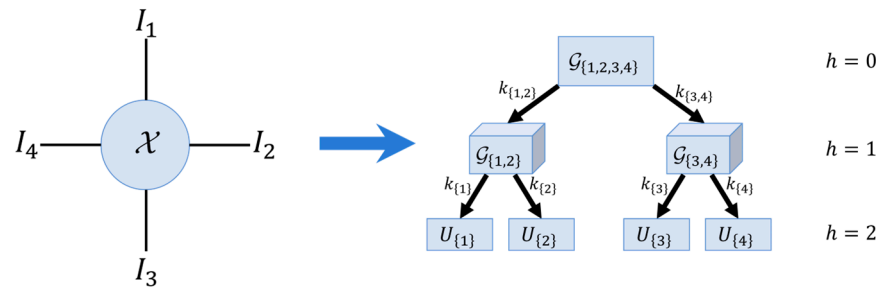


Figure 2. The diagram of tensor tree decomposition.

3. Methodology

3.1. Spatial–Temporal Infrared Tensor Model

According to the characteristic analysis in [22], the original infrared images can be linearly modeled as:

$$f_D = f_B + f_T + f_N \quad (7)$$

where f_B , f_T , f_D , and f_N denote the background image, target image, infrared image, and noise image, respectively. Equation (7) only considers spatial data and ignores the target's motion in the temporal dimension, increasing the risk of missed detections or false alarms in some complex infrared scenes. Moreover, compared with the matrix-based methods, in the tensor domain, we can explore the intrinsic relationships of the data from multiple perspectives and improve computational efficiency.

To ensure the comprehensive utilization of spatial and temporal information, we adopt the approach in [38] to construct spatial–temporal image tensor. As shown in Figure 1, the input image tensor $\mathcal{D} \in \mathbb{R}^{n_1 \times n_2 \times L}$ is constructed by stacking consecutive L frames in chronological order from the infrared sequence. Therefore, Equation (7) is written as follows:

$$\mathcal{D} = \mathcal{B} + \mathcal{T} + \mathcal{N} \quad (8)$$

where \mathcal{B} , \mathcal{T} , \mathcal{D} , and \mathcal{N} are the spatial–temporal tensor forms of f_B , f_T , f_D , and f_N , respectively. Figure 3 shows that the singular value distribution curves of the image tensor along mode 1, mode 2, and mode 3 rapidly decrease to zero. This indicates that background tensor \mathcal{B} is a low-rank tensor. Given that the infrared small targets usually occupy only a few pixels in the entire image, it can be assumed that \mathcal{T} is a sparse tensor. At the same time, it is commonly assumed that the noise in infrared images is additive Gaussian noise that satisfies $\|\mathcal{N}\|_F < \delta$. Therefore, the mathematical formula is as follows:

$$\begin{aligned} & \min_{\mathcal{B}, \mathcal{T}} \text{rank}(\mathcal{B}) + \lambda_1 \|\mathcal{T}\|_0 \\ & \text{s.t. } \|\mathcal{D} - \mathcal{B} - \mathcal{T}\|_F < \delta \end{aligned} \quad (9)$$

where λ_1 is a positive regularization parameter balancing the trade-off between the target spatial-temporal tensor and the background spatial-temporal tensor. As the optimization of l_0 -norm is NP-hard. In practice, it is usually substituted with the l_1 -norm:

$$\begin{aligned} & \min_{\mathcal{B}, \mathcal{T}} \text{rank}(\mathcal{B}) + \lambda_1 \|\mathcal{T}\|_1 \\ & \text{s.t.} \|\mathcal{D} - \mathcal{B} - \mathcal{T}\|_F < \delta \end{aligned} \quad (10)$$

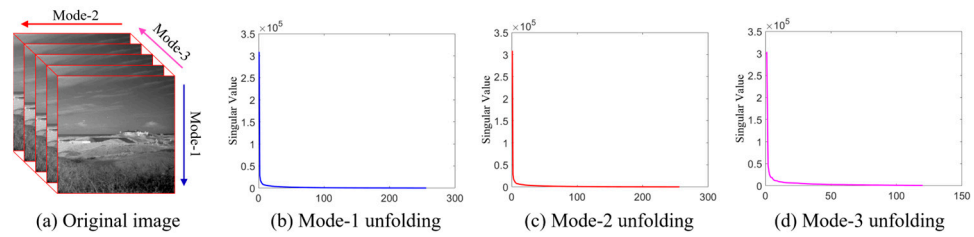


Figure 3. Singular value distribution curves of infrared spatial-temporal tensor along each mode.

3.2. Self-Adaptive Local Prior Information

In infrared images, the strong edges and corner points in the background exhibit sparsity similar to that of the target. This makes it difficult to completely distinguish them from the target when relying solely on global sparse features. Thus, it is essential to extract local prior and incorporate it into the optimization function to reduce background residuals. For this reason, the structure tensor [48] was used to depict the local geometry structure of infrared images. For an original infrared image D , the classic linear structure tensor can be calculated as follows:

$$J_\rho = K_\rho * (\nabla D_\rho \otimes \nabla D_\rho) = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix} = \begin{pmatrix} K_\rho * I_x^2 & K_\rho * I_x I_y \\ K_\rho * I_x I_y & K_\rho * I_y^2 \end{pmatrix} \quad (11)$$

$$\begin{aligned} \lambda_1 &= \frac{1}{2} \left(J_{11} + J_{22} + \sqrt{(J_{22} - J_{11})^2 + 4J_{12}^2} \right) \\ \lambda_2 &= \frac{1}{2} \left(J_{11} + J_{22} - \sqrt{(J_{22} - J_{11})^2 + 4J_{12}^2} \right) \end{aligned} \quad (12)$$

where K_ρ denotes the Gaussian kernel function with variance ρ , $*$ denotes the convolution operation, ∇ denotes the gradient, and \otimes denotes the Kronecker product. The difference between λ_1 and λ_2 reflects the image area to which the pixel belongs. When the pixel belongs to the flat region, $\lambda_1 \approx \lambda_2 \approx 0$; when the pixel belongs to the corner region, $\lambda_1 \geq \lambda_2 \gg 0$; and when the pixel belongs to the edge region, $\lambda_1 \gg \lambda_2 \approx 0$. The local prior information extracted in RIPT [30] is calculated as follows:

$$E(x, y) = \lambda_1 - \lambda_2 \quad (13)$$

where (x, y) represents the pixel position. However, as shown in row 3 of Figure 4, RIPT only captures the edge structure information of the background, which may result in background residuals and the loss of targets. Brown et al. [49] proposed the following corner-strength function to highlight target information:

$$C(x, y) = \frac{\det(ST(x, y))}{\text{tr}(ST(x, y))} = \frac{\lambda_1 \lambda_2}{\lambda_1 + \lambda_2} \quad (14)$$

where $ST(\cdot)$ means the structure tensor, and $\det(\cdot)$ and $\text{tr}(\cdot)$ represent the determinant and trace of the matrix, respectively. The PSTNN model [31] utilizes the maximum eigen-

value as the background weight function, rather than Equation (13), and combines it with Equation (14) to calculate the prior weight:

$$W_p(x, y) = C \cdot E \quad (15)$$

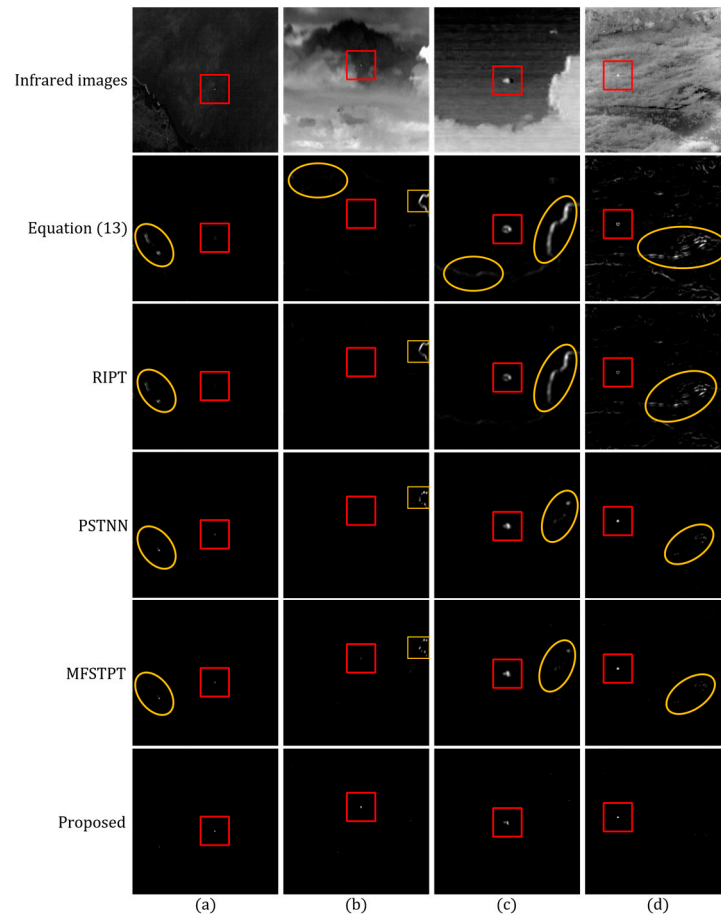


Figure 4. Comparison of different local structure priors. Row 1 shows original infrared images. Rows 2 to 6 depict different local prior maps, obtained by Equation (13), RIPT, PSTNN, MFSTPT, and the proposed method, respectively. Columns (a–d) display the prior weights extracted using different calculation methods for four infrared image sequences.

Row 4 of Figure 4 shows that the PSTNN suppresses the residual edge effect to some extent, but there is still room for improvement. In MFSTPT model [40], a weighted geometric average strategy was developed to integrate edge weighting from Equation (13) and the corner-point weighting from Equation (14), which can be expressed as follows:

$$W_c(x, y) = \sqrt[n]{C^p \cdot E^q} \quad (16)$$

However, as demonstrated in row 5 of Figure 4, strong edges are still not constrained effectively, despite the improved acquisition of target information. Based on the above analyses, we believe that the previous methods do not fully exploit the pixel information contained in λ_1 and λ_2 . Another problem is that the weight-stretching parameter artificially set in RIPT and MFSTPT cannot effectively balance the enhancement of the target and the suppression of the background. The underlying reason is the lack of consideration of the clutter information content in infrared images across different scenes. Therefore, a self-adaptive local prior method is proposed to address the above issues. Inspired by the

Frangi filter [50], we utilize the ratio and the difference of eigenvalues to highlight target information and suppress background interference:

$$R(x, y) = \frac{|\lambda_1|}{|\lambda_2|}$$

$$\beta = \text{mean}\left(\sqrt[3]{\lambda_1 - \lambda_2}\right) \quad (17)$$

where $R(x, y)$ represents the statistical measure of edges and corner points, and $|\lambda_1| > |\lambda_2|$. In edge and corner-point regions, a larger difference between the eigenvalues results in a higher R -value. In contrast, in flat regions, the similarity between the two eigenvalues leads to a lower R -value. In addition, the β -value reflects the level of background interference contained in the original image. In scenes with strong clutter, the β -value is larger. Instead, as the background clutter decreases, the β -value will also be smaller. This can be used to adaptively suppress edges and corners. Thus, the final self-adaptive prior weight is described as:

$$W_s(x, y) = \exp\left(-\frac{R^2}{2\beta}\right) \cdot \left(1 - \exp\left(-\frac{\lambda_1^2 + \lambda_2^2}{2c^2}\right)\right) \quad (18)$$

where c denotes the half of the maximum of $\sqrt{\lambda_1^2 + \lambda_2^2}$. The last row in Figure 4 shows that the proposed self-adaptive weight effectively suppresses the residual effect of strong edges and bright corner points, while also highlighting the target information. It can be seen that the adaptive factor β enhances suppression in scenes with strong clutter, resulting in a slight target shrinkage, but significantly reduces background residuals compared to other methods. Then, we construct the spatial-temporal tensor \mathcal{W}_s and normalize it as follows:

$$\mathcal{W}_s = \frac{\mathcal{W}_s - w_{\min}}{w_{\max} - w_{\min}} \quad (19)$$

where w_{\max} and w_{\min} denote the maximum and minimum values of \mathcal{W}_s , respectively. In order to accelerate the convergence speed and improve the computational efficiency, we use the reweighted scheme [51] to add a sparse weight:

$$\mathcal{W}_{sw}^{k+1} = \frac{c}{|\mathcal{T}^k| + \varepsilon} \quad (20)$$

where c denotes a non-negative constant, ε represents a small positive number preventing the denominator from being 0, and k is the number of iterations. Considering that self-adaptive prior weight in Equation (19) can suppress edges and corner points, we obtain \mathcal{W}_{rec} by taking the reciprocal of the corresponding elements in \mathcal{W}_s . Combined with the sparse weight in Equation (20), we build the final local prior tensor as follows:

$$\mathcal{W} = \mathcal{W}_{rec} \odot \mathcal{W}_{sw} \quad (21)$$

where \odot represents the Hadamard product.

3.3. Spatial-Temporal Total Variation Regularization

In real-world infrared scenes, heavy noise can be a significant interference, causing false alarms in target detection. Fortunately, the TV model effectively reduces image noise while simultaneously preserving the spatial piecewise smoothness. Introduced by Rudin [52], TV regularization can distinguish between areas with significant variations,

such as edges and textures, and smooth areas with large amounts of noise. For the matrix $X \in \mathbb{R}^{I_1 \times I_2}$, the TV norm can be mathematically expressed as follows:

$$\begin{aligned} \|X\|_{\text{TV}} = & \sum_{i=1}^{I_1-1} \sum_{j=1}^{I_2-1} \left(|X_{i+1,j} - X_{i,j}| + |X_{i,j+1} - X_{i,j}| \right) \\ & + \sum_{i=1}^{I_1-1} |X_{i+1,I_2} - X_{i,I_2}| + \sum_{j=1}^{I_2-1} |X_{I_1,j+1} - X_{I_1,j}| \end{aligned} \quad (22)$$

It can be seen from Equation (22) that the matrix-based TV framework only depicts the spatial continuity of the infrared targets and ignores the temporal continuity between successive frames. For the exploration of temporal coherence and spatio-temporal smoothing of small targets, the STTV can be obtained:

$$\|\mathcal{X}\|_{\text{STTV}} = \|D_h \mathcal{X}\|_1 + \|D_v \mathcal{X}\|_1 + \|D_z \mathcal{X}\|_1 \quad (23)$$

$$D_h \mathcal{X} = \mathcal{X}(i+1, j, k) - \mathcal{X}(i, j, k) \quad (24)$$

$$D_v \mathcal{X} = (\mathcal{X}|i, j+1, k) - \mathcal{X}(i, j, k) \quad (25)$$

$$D_z \mathcal{X} = \mathcal{X}(i, j, k+1) - \mathcal{X}(i, j, k) \quad (26)$$

where D_h , D_v , and D_z represent the horizontal, vertical, and temporal difference operators, respectively. This spatiotemporal form of TV can be seen as an effective regular item, and it exhibits a degree of resilience against noise while preserving the image details. Furthermore, it not only emphasizes the spatial smoothness of the local region in the image but also considers that the target remains temporally consistent among successive frames.

3.4. The Proposed TTALP-TV Model

In tensor robust principal component analysis (TRPCA) problems, the rank function is a nonconvex objective to solve. Therefore, the approximation of low-rank background tensor \mathcal{B} in Formula (10) is a crucial issue. A recent study [53] shows that employing the tensor tree-based TRPCA method can measure low-rankness of each mode and reduce memory requirements. In this article, we leverage the advantages of tensor tree rank and present the following optimization model:

$$\begin{aligned} \min_{\mathcal{B}, \mathcal{T}} & \mathbf{w}^T \text{rank}_{\text{tree}}(\mathcal{B}) + \lambda_1 \|\mathcal{T}\|_1 \\ \text{s.t.} & \|\mathcal{D} - \mathcal{B} - \mathcal{T}\|_F < \delta \end{aligned} \quad (27)$$

where $\text{rank}_{\text{tree}}(\mathcal{B}) = [k_1, \dots, k_Q]^T$ is the tensor tree rank and the weighting vector \mathbf{w}^T meets $\sum_{q=1}^Q w_q = 1$. The direct minimization of tensor tree ranks in Formula (27) is NP-hard. As such, we can use their matrix nuclear norms as convex surrogates:

$$\begin{aligned} \min_{\mathcal{B}, \mathcal{T}} & \sum_{q=1}^Q w_q \|B_{[q]}\|_* + \lambda_1 \|\mathcal{T}\|_1 \\ \text{s.t.} & \|\mathcal{D} - \mathcal{B} - \mathcal{T}\|_F < \delta \end{aligned} \quad (28)$$

Furthermore, we incorporate the local prior tensor and STTV regularization to obtain the prior information and suppress background noise, respectively. The proposed TTALP-TV model is as follows:

$$\begin{aligned} \min_{\mathcal{B}, \mathcal{T}} & \sum_{q=1}^Q w_q \|B_{[q]}\|_* + \lambda_1 \|\mathcal{W} \odot \mathcal{T}\|_1 + \lambda_2 \|\mathcal{B}\|_{\text{STTV}} \\ \text{s.t.} & \|\mathcal{D} - \mathcal{B} - \mathcal{T}\|_F < \delta \end{aligned} \quad (29)$$

where λ_2 is a positive regularization parameter.

3.5. Optimization Procedure

The objective function (29) can be solved effectively using the ADMM [54] method. By introducing four auxiliary variables, \mathcal{X} , Z_1 , Z_2 , and Z_3 , we obtain the following model:

$$\begin{aligned} \min_{\mathcal{B}, \mathcal{T}} \sum_{q=1}^Q w_q \|X_{[q]}\|_* + \lambda_1 \|\mathcal{W} \odot \mathcal{T}\|_1 + \lambda_2 (\|Z_1\|_1 + \|Z_2\|_1 + \|Z_3\|_1) \\ \text{s.t. } \|\mathcal{D} - \mathcal{B} - \mathcal{T}\|_F < \delta, \mathcal{X} = \mathcal{B}, Z_1 = D_h \mathcal{B}, Z_2 = D_v \mathcal{B}, Z_3 = D_z \mathcal{B} \end{aligned} \quad (30)$$

Based on the inexact augmented Lagrangian multiplier (IALM) [55], Equation (30) is written as:

$$\begin{aligned} L_A(\mathcal{B}, \mathcal{T}, \mathcal{X}, Z) = \sum_{q=1}^Q w_q \|X_{[q]}\|_* + \lambda_1 \|\mathcal{W} \odot \mathcal{T}\|_1 + \lambda_2 (\|Z_1\|_1 + \|Z_2\|_1 + \|Z_3\|_1) \\ + \langle y_1, \mathcal{D} - \mathcal{B} - \mathcal{T} \rangle + \langle y_2, \mathcal{X} - \mathcal{B} \rangle + \langle y_3, Z_1 - D_h \mathcal{B} \rangle \\ + \langle y_4, Z_2 - D_v \mathcal{B} \rangle + \langle y_5, Z_3 - D_z \mathcal{B} \rangle + \frac{\mu}{2} (\|\mathcal{D} - \mathcal{B} - \mathcal{T}\|_F^2 \\ + \|\mathcal{X} - \mathcal{B}\|_F^2 + \|Z_1 - D_h \mathcal{B}\|_F^2 + \|Z_2 - D_v \mathcal{B}\|_F^2 + \|Z_3 - D_z \mathcal{B}\|_F^2) \end{aligned} \quad (31)$$

where y_1, y_2, y_3, y_4 , and y_5 are the Lagrangian multipliers, and μ represents a positive penalty parameter. Using the ADMM framework, we can divide the Equation (31) into the following subproblems:

(a) Updating \mathcal{X} with other variables being fixed:

$$\begin{aligned} \mathcal{X} = \arg \min_{\mathcal{X}} \sum_{q=1}^Q w_q \|X_{[q]}\|_* + \langle y_2, \mathcal{X} - \mathcal{B} \rangle + \frac{\mu}{2} \|\mathcal{X} - \mathcal{B}\|_F^2 \\ = \arg \min_{\mathcal{X}} \sum_{q=1}^Q w_q \|X_{[q]}\|_* + \frac{\mu}{2} \left\| \mathcal{X} - \left(\mathcal{B} - \frac{y_2}{\mu} \right) \right\|_F^2 \end{aligned} \quad (32)$$

Let $\tau = \frac{w_q}{\mu}$ and $\mathcal{S} = \mathcal{B} - \frac{y_2}{\mu}$, Equation (32) can be rewritten as:

$$\mathcal{X} = \arg \min_{\mathcal{X}} \sum_{q=1}^Q \tau \|X_{[q]}\|_* + \frac{1}{2} \left\| X_{[q]} - \mathcal{S}_{[q]} \right\|_F^2 \quad (33)$$

For each node $\mathbb{C}_q \subset \mathbb{T}$, the solution of $X_{[q]}$ can be obtained by the singular value thresholding (SVT) [56]:

$$X_{[q]} = \text{SVT}_{\tau}(\mathcal{S}_{[q]}) = U \text{sth}_{\tau}(\Sigma) V^H \quad (34)$$

where $\text{sth}_{\tau}(s) = \text{sgn}(s) \max(|s| - \tau, 0)$ and H denotes the complex conjugate.

According to the tensor tree structure of \mathcal{X} , U_q can be used to represent the updated node value instead of directly updating $X_{[q]}$. Moreover, after updating the two successor nodes $\mathbb{C}_{q_1}, \mathbb{C}_{q_2} \subset \mathbb{C}_q$, we can update the transfer tensor \mathcal{G}_q to represent U_q for each interior node $\mathbb{C}_q \subset \mathbb{T}$, where \mathcal{G}_q is obtained by applying SVT to the new tensor $\mathcal{C} = \mathcal{S} \times_{q_1} U_{q_1} \times_{q_2} U_{q_2}$. In summary, we can utilize tensor tree decomposition to update \mathcal{X} , and the solution details are summarized in Algorithm 1.

Algorithm 1: The updating of \mathcal{X} from leaves to roots.

Input: \mathcal{S}, τ
for $p = 1, \dots, P$
 $[U_p, k_p] = \text{SVT}_{\tau}(S_{[p]})$
end for
 $\mathcal{C}_{H-1} = \mathcal{S} \times_1 U_1 \times_2 U_2 \cdots \times_P U_P$
for $h = H-1, \dots, 0$ **do**
for $q_h = 1, \dots, Q_h$ **do**
if (the q_h node is an interior one)
 $[\hat{U}_{q_h}, \hat{k}_{q_h}] = \text{SVT}_{\tau_{q_h}}(C_{[q_h]})$
 $\mathcal{G}_{q_h} = \text{reshape}(\hat{U}_{q_h}, \hat{k}_{q_h}, \hat{k}_{q_{h,1}}, \hat{k}_{q_{h,2}})$
end if
end for
 $\mathcal{C}_{h-1} = \mathcal{C}_h \times_1 \hat{U}_1 \times_2 \hat{U}_2 \cdots \times_{(Q_h-P_h)} \hat{U}_{(Q_h-P_h)}$
end for
 \mathcal{X} can be constructed from \mathcal{G}_q and U_q
Output: \mathcal{X}

(b) Updating \mathcal{B} with other variables being fixed:

$$\begin{aligned} \mathcal{B} = \arg \min_{\mathcal{B}} \frac{\mu}{2} & \left(\left\| \mathcal{D} - \mathcal{B} - \mathcal{T} + \frac{y_1}{\mu} \right\|_F^2 + \left\| \mathcal{X} - \mathcal{B} + \frac{y_2}{\mu} \right\|_F^2 \right. \\ & \left. + \left\| Z_1 - D_h \mathcal{B} + \frac{y_3}{\mu} \right\|_F^2 + \left\| Z_2 - D_v \mathcal{B} + \frac{y_4}{\mu} \right\|_F^2 + \left\| Z_3 - D_z \mathcal{B} + \frac{y_5}{\mu} \right\|_F^2 \right) \end{aligned} \quad (35)$$

The closed form solution of Equation (35) is expressed as follows:

$$\mathcal{B} = \mathcal{F}^{-1} \left(\frac{\mathcal{F}(L + \theta_1 + \theta_2 + \theta_3)}{2 + \sum_{i \in \{h,v,z\}} \mathcal{F}(D_i)^H \mathcal{F}(D_i)} \right) \quad (36)$$

where $L = \mathcal{D} - \mathcal{T} + \mathcal{X} + (y_1/\mu) + (y_2/\mu)$, $\theta_1 = D_h^T(Z_1 + (y_3/\mu))$, $\theta_2 = D_v^T(Z_2 + (y_4/\mu))$, and $\theta_3 = D_z^T(Z_3 + (y_5/\mu))$. \mathcal{F} and \mathcal{F}^{-1} represent the fast nFFT operator and the inverse nFFT operator, respectively.

(c) Updating \mathcal{T} with other variables being fixed:

$$\mathcal{T} = \arg \min_{\mathcal{T}} \lambda_1 \|\mathcal{W} \odot \mathcal{T}\|_1 + \frac{\mu}{2} \left\| \mathcal{D} - \mathcal{B} - \mathcal{T} + \frac{y_1}{\mu} \right\|_F^2 \quad (37)$$

Using the element-wise shrinkage approach [57], \mathcal{T} is updated by:

$$\mathcal{T} = \text{TH}_{\frac{\lambda_1}{\mu}} \left(\mathcal{D} - \mathcal{B} + \frac{y_1}{\mu} \right) \quad (38)$$

where $\text{TH}(\cdot)$ denotes the element-wise shrinkage operator.

(d) Updating Z_1 , Z_2 , and Z_3 with other variables being fixed:

$$\begin{cases} Z_1 = \arg \min_{Z_1} \lambda_2 \|Z_1\|_1 + \frac{\mu}{2} \left\| Z_1 - D_h \mathcal{B} + \frac{y_3}{\mu} \right\|_F^2 \\ Z_2 = \arg \min_{Z_2} \lambda_2 \|Z_2\|_1 + \frac{\mu}{2} \left\| Z_2 - D_v \mathcal{B} + \frac{y_4}{\mu} \right\|_F^2 \\ Z_3 = \arg \min_{Z_3} \lambda_2 \|Z_3\|_1 + \frac{\mu}{2} \left\| Z_3 - D_z \mathcal{B} + \frac{y_5}{\mu} \right\|_F^2 \end{cases} \quad (39)$$

The Equation (39) can be solved by the element-wise shrinkage operator:

$$\begin{cases} Z_1 = \text{TH}_{\frac{\lambda_2}{\mu}}\left(D_h \mathcal{B} - \frac{y_3}{\mu}\right) \\ Z_2 = \text{TH}_{\frac{\lambda_2}{\mu}}\left(D_v \mathcal{B} - \frac{y_4}{\mu}\right) \\ Z_3 = \text{TH}_{\frac{\lambda_2}{\mu}}\left(D_z \mathcal{B} - \frac{y_5}{\mu}\right) \end{cases} \quad (40)$$

(e) Updating Lagrangian multipliers y_1, y_2, y_3, y_4 , and y_5 with other variables being fixed:

$$\begin{cases} y_1 = y_1 + \mu(\mathcal{D} - \mathcal{B} - \mathcal{T}) \\ y_2 = y_2 + \mu(\mathcal{X} - \mathcal{B}) \\ y_3 = y_3 + \mu(Z_1 - D_h \mathcal{B}) \\ y_4 = y_4 + \mu(Z_2 - D_v \mathcal{B}) \\ y_5 = y_5 + \mu(Z_3 - D_z \mathcal{B}) \end{cases} \quad (41)$$

(f) Updating penalty parameter μ by $\mu = \min(\rho\mu, \mu_{\max})$.

The complete process of the ADMM optimization method is given in Algorithm 2.

Algorithm 2: TTALP-TV algorithm

Input: The spatial-temporal tensor $\mathcal{D} \in \mathbb{R}^{n_1 \times n_2 \times L}$, parameters $\lambda_1, \lambda_2, \mu$

Initialize: $\mathcal{B}^0 = \mathcal{D}$, $\mathcal{T}^0 = Z_i^0 = 0, i = 1, 2, 3, y_i^0 = 0, i = 1, 2, 3, 4, 5, \mu^0 = 5e - 3, \mu_{\max} = 1e6, \rho = 1.2, \zeta = 1e - 6$, maximum iteration step $K = 100$.

While not converged do

1: Update \mathcal{X}^{k+1} by **Algorithm 1**

2: Update \mathcal{B}^{k+1} via Equation (36)

3: Update \mathcal{T}^{k+1} via Equation (38)

4: Update \mathcal{W}^{k+1} via Equation (21)

5: Update $Z_1^{k+1}, Z_2^{k+1}, Z_3^{k+1}$ via Equation (40)

6: Update Lagrangian multipliers $y_i^{k+1}, i = 1, 2, 3, 4, 5$ via Equation (41)

7: Update penalty parameter μ via

$$\mu = \min(\rho\mu, \mu_{\max})$$

8: Check the convergence condition

$$\frac{\|\mathcal{D} - \mathcal{B}^{k+1} - \mathcal{T}^{k+1}\|_F^2}{\|\mathcal{D}\|_F^2} \leq \zeta$$

9: Update $k = k + 1$

End while

Output: Background component \mathcal{B} and target component \mathcal{T} .

3.6. Steps of Detection Method

Figure 1 elaborates the whole process of the proposed TTALP-TV model, which is described as follows:

1. Self-adaptive local prior extraction. Given an infrared image, the self-adaptive prior weight W_s is calculated by Equation (18).
2. Spatial-temporal tensor construction. The spatial-temporal infrared tensor $\mathcal{D} \in \mathbb{R}^{n_1 \times n_2 \times L}$ and local prior tensor $\mathcal{W} \in \mathbb{R}^{n_1 \times n_2 \times L}$ are constructed by stacking consecutive L frames in chronological order from the original image sequence and the prior weight map, respectively.
3. Background and target separation. The spatial-temporal infrared tensor \mathcal{D} is decomposed into background tensor \mathcal{B} and target tensor \mathcal{T} through Algorithm 2.

Image reconstruction. Contrary to the construction process, the target image f_T is reconstructed from \mathcal{T} .

4. Experimental Results

In this section, we first discuss the datasets used in infrared target detection experiments. Then, we introduce evaluation metrics and analyze the effects of several important parameters on the TTALP-TV model. Finally, we evaluate the detection ability and robustness of the proposed algorithm and compare it with eight state-of-the-art methods in six complex scenes.

4.1. Experiment Data

The dataset used in the experiments consists of six infrared image sequences, including complex scenes such as sky, sea, clouds, mountains, and buildings. The infrared sequences 1, 3, 4, and 6 are from [58,59]. In order to carry out an objective assessment of TTALP-TV from diverse scenes, we simulated infrared sequences 2 and 5 using the strategy in [22]. As shown in Figure 5, the images are uniformly scaled to the same size to improve target visibility. Meanwhile, each small target is marked by a red rectangle and magnified in the bottom right corner of the image. It can be seen that, in most scenes, the targets occupy a few pixels and lack shape information and texture features. Due to heavy clutter interference in complex scenes, it is difficult to distinguish the target from the background. The specific descriptions of sequences are presented in Table 2. Additionally, the entire experiment framework was implemented using MATLAB R2020a in Windows 10 based on AMD Ryzen 7 5800H 3.20 GHz CPU with 16GB memory.

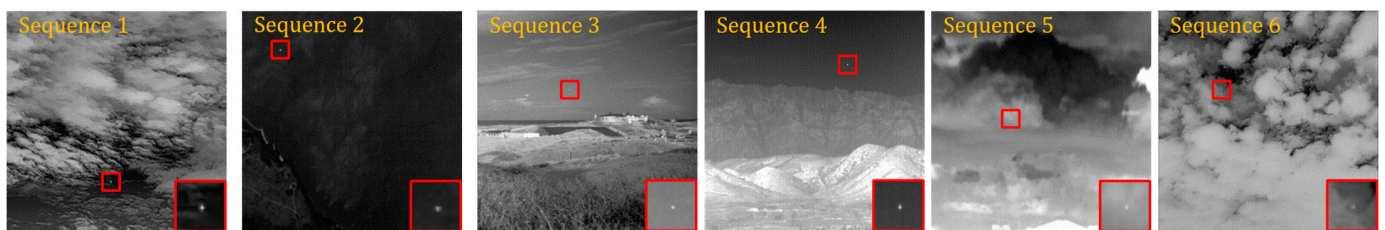


Figure 5. Representative frames corresponding to the six infrared sequences used in the experiments.

Table 2. Characteristics of the dataset.

| Sequence | Frames | Image Size | Target Descriptions | Background Descriptions |
|----------|--------|------------|--|--|
| 1 | 120 | 256 × 256 | Slow-moving and small | Ground background with fierce clouds and noise |
| 2 | 120 | 256 × 256 | Slow-moving and weak airplane | Ground background with sea and islands |
| 3 | 120 | 256 × 256 | Fast-moving, small and dim | Ground background with bright buildings |
| 4 | 120 | 256 × 256 | Fast-moving, small and regular shape | Ground background with reflective mountains |
| 5 | 120 | 256 × 205 | Fast-moving, irregularly shaped aircraft | Ground background with reflective clouds |
| 6 | 120 | 296 × 237 | Fast-moving, small and dim | Ground background with multilayer clouds |

4.2. Evaluation Metrics and Baselines

We evaluate the detection performance of the TTALP-TV method using three evaluation metrics: 3D receiver operating characteristic (3D ROC) [60], signal-to-clutter ratio gain (SCRG), and background suppression factor (BSF). The 3D ROC curve consists of three parameters, including false alarm rate F_a , detection probability P_d , and threshold τ . The P_d evaluates the target detection capability, while the F_a assesses the background suppression capability, as defined below:

$$P_d = \frac{TD}{AT} \quad (42)$$

where TD and AT denote the number of detected targets and the number of actual targets, respectively.

$$F_a = \frac{FD}{NP} \quad (43)$$

where FD and NP denote the number of false detections and the number of image pixels, respectively. Due to the intersections between ROC curves, we calculate the AUC values of three 2D ROC curves, $AUC_{(F_a, P_d)}$, $AUC_{(\tau, P_d)}$, and $AUC_{(\tau, F_a)}$, for a more accurate performance assessment. The values of $AUC_{(F_a, P_d)}$ and $AUC_{(\tau, P_d)}$ range from zero to one, where values closer to one indicate better target detection capability. Conversely, the value of $AUC_{(\tau, F_a)}$ ranges from one to zero, where the value closer to zero represents a better ability to suppress background clutter. Therefore, the above three AUC values are combined to comprehensively evaluate the overall accuracy (OA) and the signal-to-noise probability ratio (SNPR), which are defined as follows:

$$AUC_{OA} = AUC_{(F_a, P_d)} + AUC_{(\tau, P_d)} - AUC_{(\tau, F_a)} \quad (44)$$

$$AUC_{SNPR} = \frac{AUC_{(\tau, P_d)}}{AUC_{(\tau, F_a)}} \quad (45)$$

where $AUC_{OA} \in [0, 2]$ and $AUC_{SNPR} \in [0, +\infty]$. Meanwhile, higher AUC_{OA} and AUC_{SNPR} denote a stronger ability to detect targets and suppress background clutter, respectively.

In addition, the SCRG and BSF can also be used to measure an algorithm's ability to enhance the target and suppress the background, respectively. Both SCRG and BSF are calculated in the neighborhood of the target. As shown in Figure 6, if the target size is $a \times b$, then $(a + 2d) \times (b + 2d)$ denotes the size of the target neighborhood. In the experiments of this paper, we follow [32] to set $d = 65$.

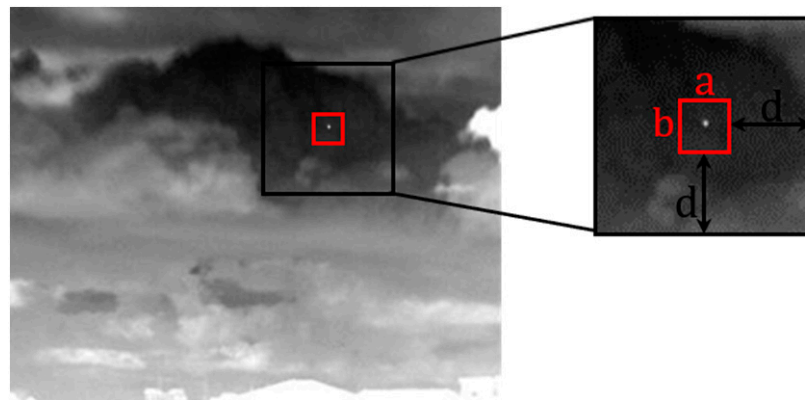


Figure 6. Diagram of the target neighborhood.

The SCRG represents the SCR of the detection result and the original image, which is expressed as:

$$SCRG = \frac{SCR_{out}}{SCR_{in}} \quad (46)$$

where SCR reflects the degree of discrimination between the target and the background clutter in the image, which can be calculated as:

$$SCR = \frac{|\bar{\mu}_0 - \bar{\mu}_1|}{\sigma_1} \quad (47)$$

where $\bar{\mu}_0$ denotes the target's average gray value, $\bar{\mu}_1$ denotes that of the target neighborhood, and σ_1 denotes the gray standard deviation of the target neighborhood.

The BSF can evaluate the background suppression ability, which is defined as follows:

$$\text{BSF} = \frac{\sigma_{\text{in}}}{\sigma_{\text{out}}} \quad (48)$$

where σ_{out} and σ_{in} represent the standard deviations of the target neighborhood in the detection result and the original image, respectively.

4.3. Parameter Analyses

The settings of different parameters in the model have a great impact on the detection performance. Therefore, this section aims to explore the appropriate parameters for the TTALP-TV method in sequences 1–6. According to [61], we set $\lambda_2 = 0.01$. Then, we detail the effects of L and H on the detection capability of our proposed method.

4.3.1. Adjacent Frames Number L

In the construction of the spatial–temporal tensor, the adjacent frame number L determines the utilization of temporal domain information. In order to investigate the influence of different L values on the detection performance of the TTALP-TV model, we adjust L from 2 to 6 with a step of 1. Figure 7 shows the analysis results of various L values using the 3D ROC. Increasing the L values can incorporate more temporal information, which ensures the low-rankness of the spatial–temporal tensor. At the same time, over-large adjacent frame numbers will lead to redundant and repetitive information, resulting in high false alarms. Figure 7 shows that $L = 3$ is the most suitable for the proposed model.

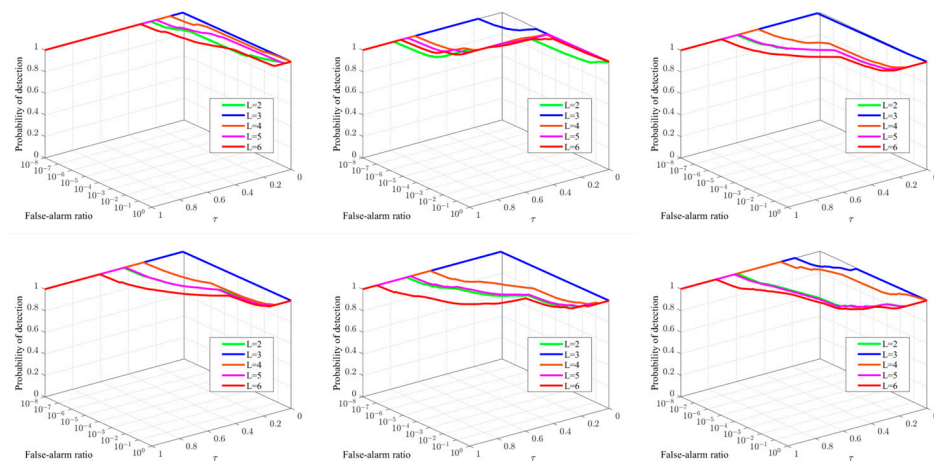


Figure 7. Three-dimensional ROC curves corresponding to different parameters of L in the six sequences.

4.3.2. Tuning Parameter H

The compromising parameter λ_1 controls the balance between the sparse target and the low-rank background in the framework. Following [62], we set $\lambda_1 = H / \sqrt{\max(n_1, n_2) * L}$, where H is a crucial tuning parameter. We change H from 4 to 12 with a step of 2. The 3D ROC analysis results of H are shown in Figure 8. It can be seen from Figure 8 that when H values increase, the false alarms decrease, which indicates that H assists in the suppression of background residuals. Meanwhile, if H is too large (e.g., $H = 12$), some necessary information may be lost, resulting in the degradation of detection performance. Based on the 3D ROC analysis results shown in Figure 8, we set $H = 10$.

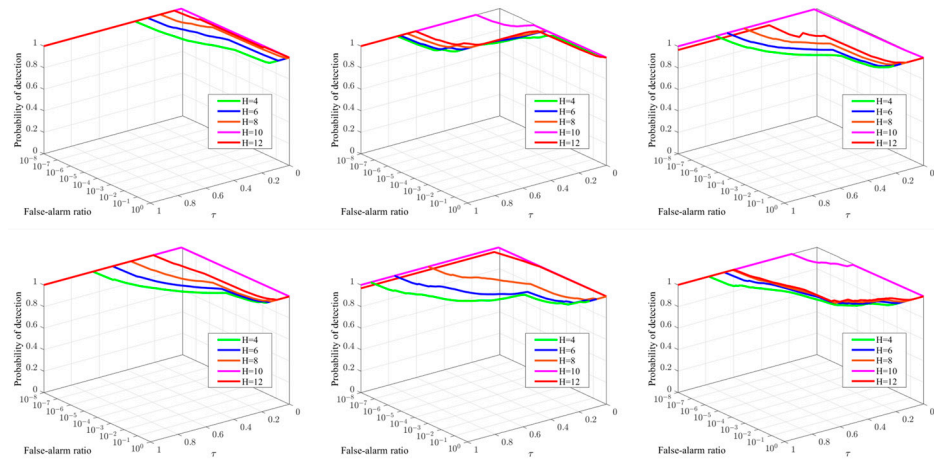


Figure 8. Three-dimensional ROC curves corresponding to different parameters of H in the six sequences.

4.4. Ablation Study

In order to validate the effectiveness of the self-adaptive local prior and STTV regularization in the proposed TTALP-TV method, we conducted an ablation study, as shown in Figure 9. The TTALP-TV framework consists of three parts: the tensor tree-based spatiotemporal tensor model, the self-adaptive local prior tensor, and STTV regularization. As illustrated in Figure 9, we compare the 3D ROC analysis results of four versions of the TTALP-TV method in sequences 1–6: (1) the tensor tree-based spatiotemporal tensor model (TTSTT), (2) incorporating self-adaptive local prior tensor into the tensor tree-based spatiotemporal tensor model (TTALP), (3) imposing STTV regularization constraint on the background component in the tensor tree-based spatiotemporal tensor model (TTSTT-TV), and (4) integrating the self-adaptive local prior tensor and STTV regularization into the tensor tree-based spatiotemporal tensor model (TTALP-TV). Figure 9 shows that leveraging the self-adaptive local prior does in fact improve target detection performance to a certain extent. Moreover, the STTV regularization constraint on the background helps better remove background clutter and noise while preserving image details. The results of the ablation experiments intuitively demonstrate the significance of any single module and provide guidance for further attempts to improve the optimization model.

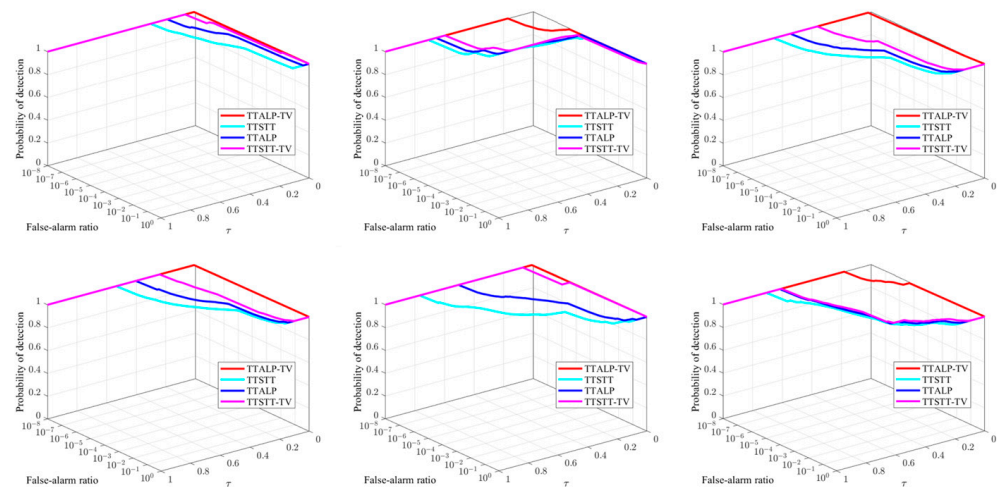


Figure 9. Ablation results of the six sequences in 3D ROC curves.

4.5. Noise Robustness Validation of the Proposed TTALP-TV Method

Due to the influence of the real-world environment on the sensor, infrared images usually contain noise. Therefore, it is essential to evaluate the robustness of the TTALP-

TV model to noise. To evaluate the noise robustness of TTALP-TV under different noise intensities, Gaussian white noise of $\sigma = 5$ and $\sigma = 15$ was added to six scenes, respectively. The second and fourth rows of Figure 9 show the visual detection results of $\sigma = 5$ and $\sigma = 15$, respectively. Figure 10 shows that TTALP-TV can accurately detect targets and suppress noise of different intensities, demonstrating its robustness to noisy scenes.

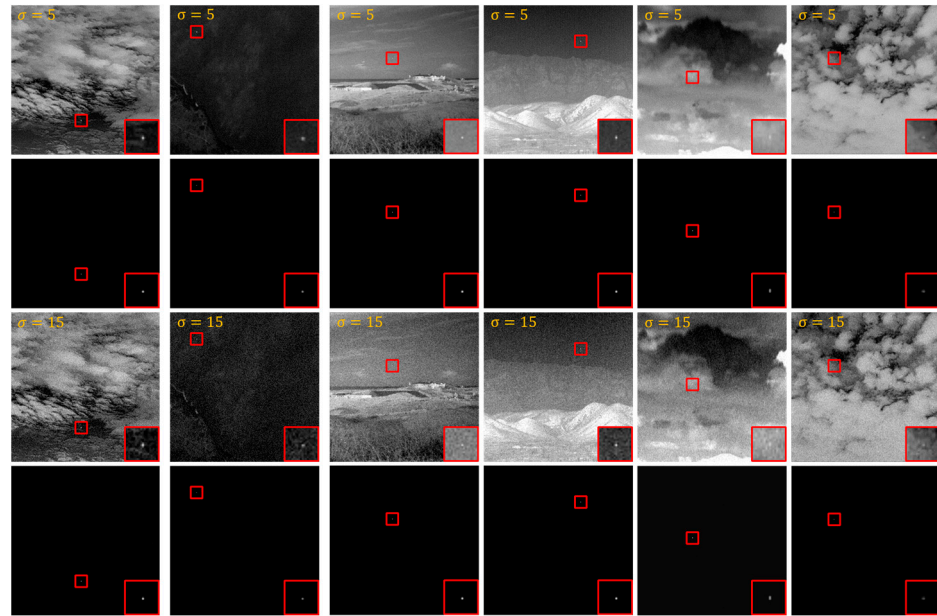


Figure 10. Detection results of TTALP-TV under different noise intensities.

4.6. Comparison with State-of-the-Art Methods

In order to assess the advantages of the TTALP-TV method, we compare it with eight representative baseline methods. These methods can be categorized into background consistency-based methods (Top-hat [7]), HVS-based methods (TLLCM [16]), LRSD-based single-frame detection methods (IPI [22], PSTNN [31], NTFRA [32], and ANLPT [33]) and LRSD-based sequential-frame detection methods (ASTTV-NTLA [63] and NFTDGSTV [42]). Table 3 lists the detailed parameter settings of these methods.

Table 3. Detailed parameters of nine methods.

| Method | Parameters |
|------------|--|
| Top-hat | Shape: disk, structure size: 5×5 . |
| TLLCM | Different filtering window: $3 \times 3, 5 \times 5, 7 \times 7$. |
| IPI | Patch size: 50×50 , step: 10, $\lambda = 1/\sqrt{\min(m,n)}$, $\varepsilon = 10^{-7}$. |
| PSTNN | Patch size: 40×40 , step: 40, $\lambda = 0.7/\sqrt{\min(n_1, n_2) * n_3}$, $\varepsilon = 10^{-7}$. |
| ASTTV-NTLA | $L = 3, H = 6, \lambda_{tv} = 0.005, \lambda_s = \frac{H}{\sqrt{\max(M,N)*L}}, \lambda_3 = 100$. |
| NTFRA | Patch size: 40×40 , step: 40, $\lambda = 1/\sqrt{\min(n_1, n_2) * n_3}$, $\beta = 0.05, \mu = 200$. |
| ANLPT | Patch size: 50×50 , step: 50, region: 10, channel: 3, $\mu = 10^{-3}$. |
| NFTDGSTV | $L = 3, H = 4, \lambda_1 = 0.01, \lambda_2 = \frac{H}{\sqrt{\max(M,N)*L}}, \lambda_s = 0.001$. |
| Proposed | $L = 3, H = 10, \lambda_1 = \frac{H}{\sqrt{\max(n_1, n_2)*L}}, \lambda_2 = 0.01$. |

4.6.1. Visual Comparison

Figures 11 and 12 show the detection results of eight compared methods and our method in six infrared sequences. From Figures 11 and 12, we can see that Top-hat has a lot of clutter and noise residuals in its detection results. The main reason for this is that the structure size of the Top-hat is fixed, meaning it cannot adapt to the dynamics of complex scenes. In contrast, TLLCM suppresses clutter to a certain extent but still has

background residuals in complex scenes. Compared with the background consistency and HVS methods, the matrix-based LRSD method IPI contains fewer background residuals, but its background is gray. As can be seen from Figures 11 and 12, the PSTNN and ANLPT methods can achieve relatively better target detection performance (e.g., sequences 1, 4, and 6), but they are basically unable to completely suppress background.

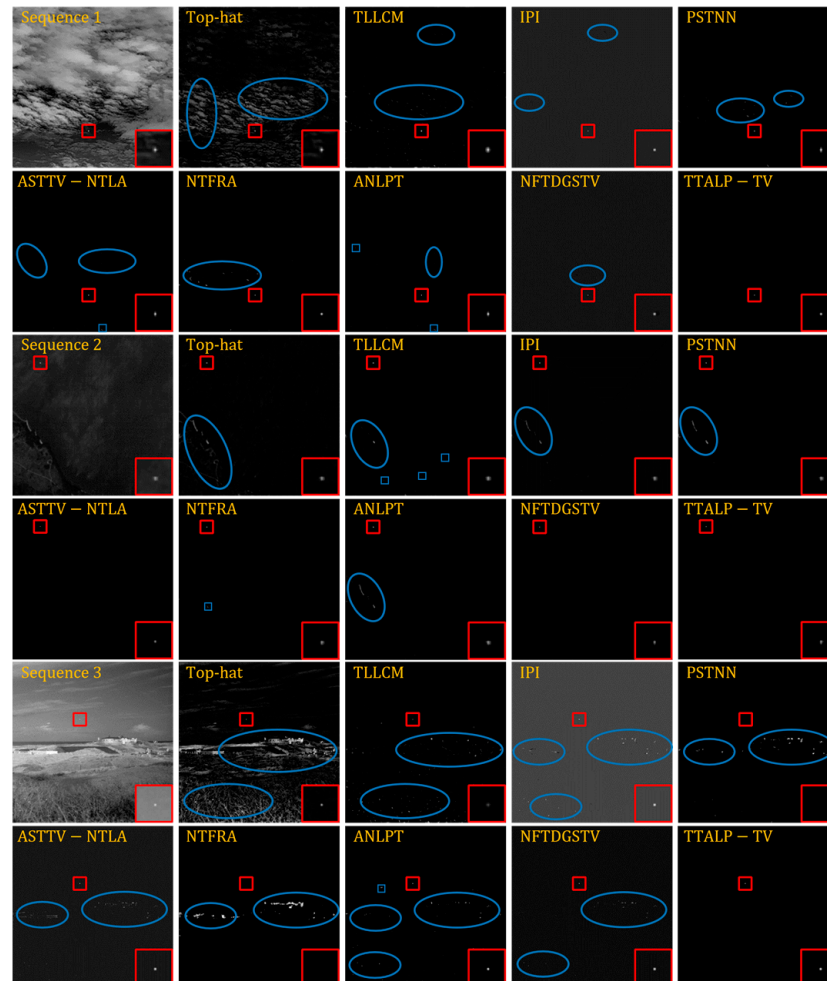


Figure 11. Detection results of nine methods in sequences 1–3. The red rectangles denote target areas, and the blue ellipses denote noise and background residuals.

At the same time, we can see that NTFRA can better preserve targets and suppress background interference but fails in complex scenes with highlighted line edges (e.g., sequences 3–4). These single-frame detection methods effectively utilize spatial information to separate the target from the background. However, using only inter-frame information results in low robustness to various complex scenes with dynamic changes and heavy clutter. Therefore, many researchers have combined spatial–temporal information to improve detection ability and remove background interference. It can be seen from Figures 11 and 12 that ASTTV-NTLA and NFDGSTV present exceptional target detection and background suppression abilities in scenes with little clutter interference (e.g., sequence 2). However, when faced with complex scenes with high-brightness clutter and heavy noise (e.g., sequences 3, 4, and 6), their detection performance will degrade significantly. In contrast, the proposed TTALP-TV method is not only able to accurately extract the target and preserve a relatively complete shape, but it can also mostly suppress strong edges and bright corner-point noise in complex scenes.

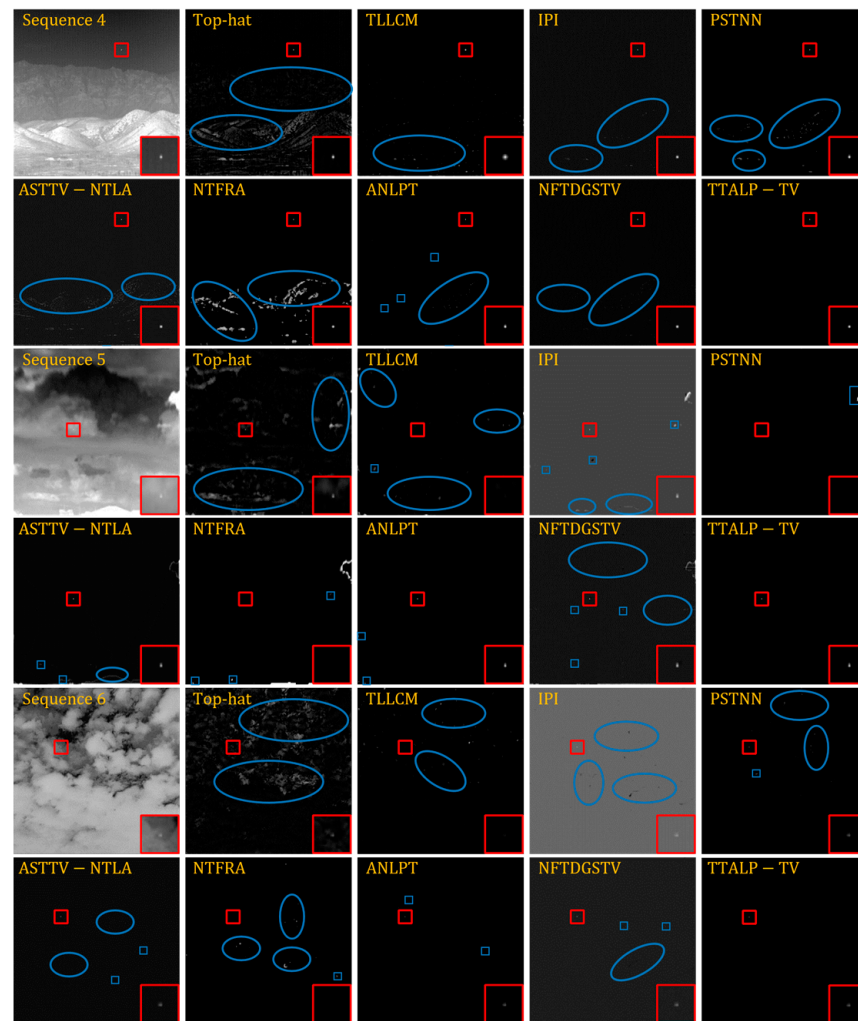


Figure 12. Detection results of nine methods in sequences 4–6. The red rectangles denote target areas, and the blue ellipses denote noise and background residuals.

4.6.2. Quantitative Analysis

In addition to the qualitative analysis in Figures 11 and 12, we adopt 3D ROC, AUC_{OA} , AUC_{SNPR} , SCRG, and BSF, a total of five evaluation metrics, to compare nine methods quantitatively. Figure 13 shows the 3D ROC curves of all comparison methods in complex and noisy scenes (e.g., sequences 1–6). In order to clearly depict the differences among the nine methods, the logarithmic scale is used for the false alarm rate axis. As shown in Figure 13, the proposed TTALP-TV method is closer to the top-right corner, indicating that it has superior detection performance. The single-frame detection method ANLPT also achieves good detection performance in sequence 5. Meanwhile, other sequential-frame detection methods, ASTTV-NTLA and NFDGSTV, exhibit performance similar to our method in sequence 2 and sequence 6, but were not good enough in the rest of the sequences. To further assess which method has the best performance, we use AUC_{OA} and AUC_{SNPR} to evaluate target detection ability and background suppression ability, respectively. In each sequence, the highest value is highlighted in red, and the second highest value is marked in green. Tables 4 and 5 show that our method achieves the highest AUC_{OA} and AUC_{SNPR} values.

In Tables 6 and 7, the SCRG and BSF of nine methods in six sequences are displayed, with the highest and second highest values of SCRG and BSF marked in red and green, respectively. The results show that the ANLPT model achieves the highest SCRG values in sequence 5. On the other hand, the SCRG and BSF values of our model surpass other methods for more complex scenes (e.g., sequences 1–4 and 6). In summary, the above quan-

titative analyses demonstrate the effectiveness of our algorithm in both target enhancement and background suppression, particularly in complex scenes.

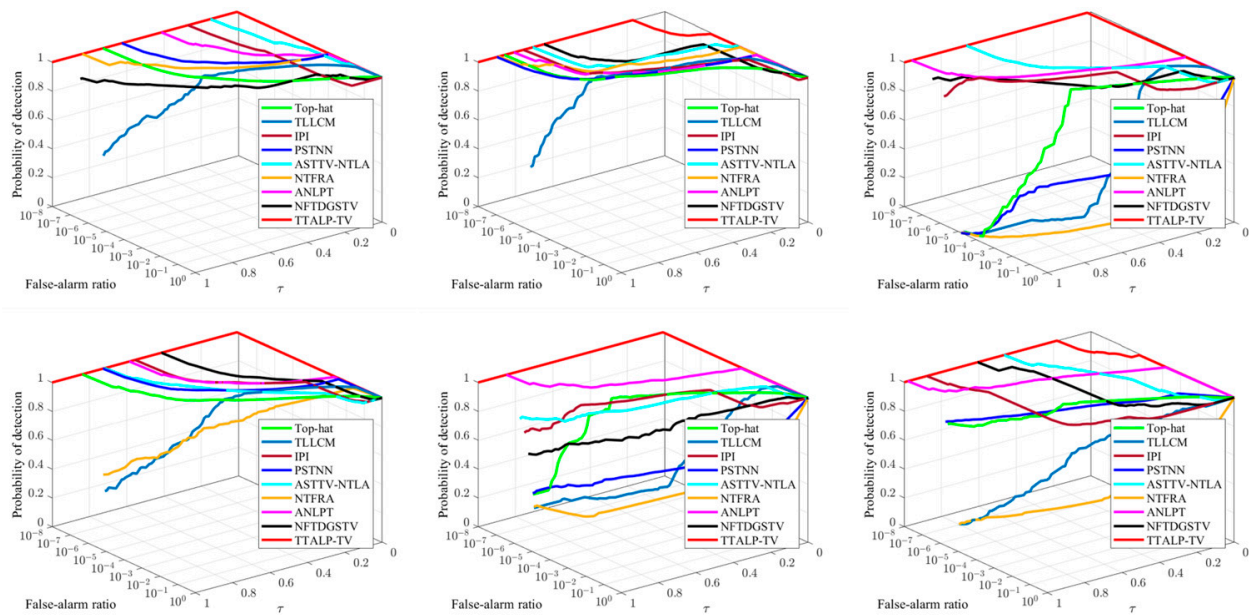


Figure 13. Three-dimensional ROC curves of nine methods in sequences 1–6.

Table 4. AUC_{OA} and AUC_{SNPR} of nine methods in sequences 1–3.

| Method | Sequence 1 | | Sequence 2 | | Sequence 3 | |
|------------|------------|--------------|------------|--------------|------------|--------------|
| | AUC_{OA} | AUC_{SNPR} | AUC_{OA} | AUC_{SNPR} | AUC_{OA} | AUC_{SNPR} |
| Top-hat | 1.9279 | 13.9053 | 1.9722 | 36.1500 | 1.6818 | 9.5483 |
| TLLCM | 1.8827 | 115.6874 | 1.9228 | 158.0823 | 1.3038 | 39.1759 |
| IPI | 1.8275 | 5.8048 | 1.9445 | 18.0709 | 1.7117 | 3.4998 |
| PSTNN | 1.9945 | 187.9084 | 1.9943 | 180.1801 | 0.8804 | 39.4834 |
| ASTTV-NTLA | 1.9943 | 182.1271 | 1.9947 | 197.3122 | 1.8817 | 8.4660 |
| NTFRA | 1.9944 | 186.0783 | 1.9947 | 195.7341 | 0.4936 | 0.6221 |
| ANLPT | 1.9946 | 193.5168 | 1.9943 | 181.1781 | 1.9942 | 178.3936 |
| NFDGSTV | 1.8983 | 10.8409 | 1.9762 | 42.2723 | 1.9065 | 12.3141 |
| Proposed | 1.9948 | 198.0273 | 1.9948 | 198.0127 | 1.9944 | 183.0405 |

Table 5. AUC_{OA} and AUC_{SNPR} of nine methods in sequences 4–6.

| Method | Sequence 4 | | Sequence 5 | | Sequence 6 | |
|------------|------------|--------------|------------|--------------|------------|--------------|
| | AUC_{OA} | AUC_{SNPR} | AUC_{OA} | AUC_{SNPR} | AUC_{OA} | AUC_{SNPR} |
| Top-hat | 1.9456 | 18.4296 | 1.8556 | 31.0480 | 1.8923 | 14.3259 |
| TLLCM | 1.8404 | 98.9758 | 1.4616 | 74.5013 | 1.4474 | 77.1831 |
| IPI | 1.8934 | 9.3979 | 1.7779 | 4.8744 | 1.5130 | 2.0540 |
| PSTNN | 1.9936 | 161.9100 | 1.1649 | 81.6425 | 1.7827 | 162.4838 |
| ASTTV-NTLA | 1.8760 | 8.0798 | 1.9490 | 22.9404 | 1.9047 | 10.5164 |
| NTFRA | 1.8710 | 47.7343 | 1.0384 | 37.7261 | 0.8706 | 45.0366 |
| ANLPT | 1.9944 | 184.8268 | 1.9947 | 196.1699 | 1.9947 | 196.9744 |
| NFDGSTV | 1.9435 | 17.7605 | 1.7477 | 5.9553 | 1.7946 | 4.8739 |
| Proposed | 1.9948 | 198.0272 | 1.9947 | 198.0339 | 1.9948 | 198.0076 |

Table 6. SCRG and BSF of nine methods in sequences 1–3.

| Method | Sequence 1 | | Sequence 2 | | Sequence 3 | |
|------------|------------|-------|------------|------|------------|------|
| | SCRG | BSF | SCRG | BSF | SCRG | BSF |
| Top-hat | 17.87 | 1.46 | 1.48 | 1.17 | 0.86 | 0.89 |
| TLLCM | 99.67 | 4.40 | 2.85 | 1.97 | 5.17 | 4.08 |
| IPI | 131.12 | 12.07 | 2.49 | 2.12 | 7.57 | 4.79 |
| PSTNN | 114.29 | 7.98 | 2.59 | 1.98 | 1.15 | 3.24 |
| ASTTV-NTLA | 219.24 | 15.33 | 2.63 | 2.52 | 9.01 | 4.63 |
| NTFRA | 81.11 | 6.20 | 2.34 | 2.27 | 0.16 | 1.56 |
| ANLPT | 178.28 | 11.19 | 2.59 | 1.95 | 9.37 | 4.23 |
| NFTDGSTV | 174.78 | 14.05 | 2.78 | 2.78 | 13.47 | 6.65 |
| Proposed | 235.51 | 17.83 | 3.56 | 3.46 | 19.51 | 8.86 |

Table 7. SCRG and BSF of nine methods in sequences 4–6.

| Method | Sequence 4 | | Sequence 5 | | Sequence 6 | |
|------------|------------|-------|------------|-------|------------|-------|
| | SCRG | BSF | SCRG | BSF | SCRG | BSF |
| Top-hat | 7.79 | 3.66 | 14.00 | 4.29 | 9.63 | 2.17 |
| TLLCM | 20.94 | 6.33 | 32.75 | 10.46 | 27.26 | 6.69 |
| IPI | 19.83 | 11.11 | 26.56 | 11.49 | 25.01 | 7.94 |
| PSTNN | 25.85 | 12.63 | 29.74 | 17.10 | 62.06 | 10.56 |
| ASTTV-NTLA | 18.18 | 10.45 | 36.50 | 15.77 | 71.84 | 14.40 |
| NTFRA | 20.50 | 10.64 | 23.89 | 17.50 | 13.06 | 9.49 |
| ANLPT | 22.97 | 10.82 | 41.95 | 16.01 | 77.48 | 14.15 |
| NFTDGSTV | 23.18 | 12.18 | 29.53 | 12.56 | 55.45 | 12.09 |
| Proposed | 27.39 | 13.88 | 40.79 | 17.83 | 96.11 | 18.30 |

4.6.3. Running Time

In addition to the above evaluation metrics, computational efficiency is also a crucial factor in infrared target detection algorithms. Table 8 presents the average running time of all comparison methods on six sequences (per frame). It should be noted that the image size of sequences 1–4 is 256×256 , and the image size of sequences 5–6 is 256×205 and 296×237 , respectively. In general, the larger image size results in the longer running time. Based on Table 8, we can find that Top-hat has the shortest time cost. This is because Top-hat adopts a simple model architecture. It is worth noting that tensor-based algorithms are significantly quicker than the matrix-based IPI algorithm. Among the tensor-based methods, the running time of sequential-frame detection methods (e.g., ASTTV-NTLA, NFTDGSTV) is longer than that of single-frame detection methods (e.g., PSTNN, NTFRA, ANLPT). This is mainly because sequential-frame detection methods require more time to process the temporal domain information. From Table 8, it can be seen that the proposed method has a longer running time than ASTTV-NTLA and NFTDGSTV. This is because computing the self-adaptive prior in TTALP-TV increases costs in terms of time. Based on the qualitative and quantitative results shown in Figures 11–13 and Tables 4–7, it can be concluded that our method has better detection performance than the compared methods. Therefore, the extra running time of our method is acceptable.

Table 8. Running time(s) of the nine methods.

| Method | Sequence 1 | Sequence 2 | Sequence 3 | Sequence 4 | Sequence 5 | Sequence 6 |
|------------|------------|------------|------------|------------|------------|------------|
| Top-hat | 0.0958 | 0.0989 | 0.1008 | 0.1013 | 0.1098 | 0.1011 |
| TLLCM | 1.1209 | 1.0979 | 1.1193 | 1.1235 | 0.8603 | 1.2962 |
| IPI | 5.1073 | 4.7903 | 5.1192 | 5.8094 | 3.9534 | 5.9438 |
| PSTNN | 0.3605 | 0.2402 | 0.3105 | 0.3168 | 0.2845 | 0.2813 |
| ASTTV-NTLA | 2.0889 | 2.1002 | 2.0552 | 2.1045 | 1.4487 | 2.2719 |
| NTFRA | 1.4493 | 1.3904 | 1.4145 | 1.4965 | 1.2829 | 1.6751 |
| ANLPT | 1.5682 | 1.5437 | 1.4939 | 1.5961 | 1.2954 | 1.5260 |
| NFTDGSTV | 1.9258 | 2.0223 | 1.9534 | 1.8297 | 1.8025 | 2.4437 |
| Proposed | 2.3115 | 2.2845 | 2.3740 | 2.3182 | 1.7406 | 2.5226 |

5. Conclusions

In this article, the TTALP-TV model is proposed for infrared small target detection in complex scenes. Based on the theorem that the tensor tree decomposition can exploit the data structure in a more balanced strategy, we introduce tensor tree rank to obtain more accurate background estimation. It reduces storage costs and retains spatial and temporal correlation through a hierarchical method. In addition, a novel local prior weight is proposed for adaptively assigning weights to targets, which helps to better distinguish targets from similar objects. Meanwhile, STTV is used as a joint regularization term to remove noise while preserving image details. Therefore, the separation of target and background is converted into an optimization problem. Finally, we provide an efficient ADMM-based framework for solving the proposed TTALP-TV model. Extensive experiments demonstrate that the proposed algorithm not only can accurately detect the target but also effectively suppresses background clutter and noise in various complex scenes. However, the real-time performance of our method still needs to be improved due to the prior weight calculation in the model. In the future, our work will focus on establishing more efficient mechanisms to further simplify the calculation and improve detection efficiency.

Author Contributions: Conceptualization, G.Z.; methodology, G.Z. and Z.D.; software, G.Z.; investigation, G.Z., Z.D., B.Z., W.Z. and J.L.; validation, G.Z.; writing—original draft preparation, G.Z.; writing—review and editing, G.Z., Q.L. and Z.T.; project administration, Q.L.; funding acquisition, Z.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Key Program Project of Science and Technology Innovation of the Chinese Academy of Sciences (no. KGFZD-135-20-03-02) and this research was funded by the Innovation Foundation of Key Laboratory of Computational Optical Imaging Technology, CAS (no. CXJJ-23S016).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the time limitation.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Rawat, S.S.; Verma, S.K.; Kumar, Y. Review on recent development in infrared small target detection algorithms. *Procedia Comput. Sci.* **2020**, *167*, 2496–2505. [\[CrossRef\]](#)
2. Xiao, S.; Peng, Z.; Li, F. Infrared Cirrus Detection Using Non-Convex Rank Surrogates for Spatial-Temporal Tensor. *Remote Sens.* **2023**, *15*, 2334. [\[CrossRef\]](#)
3. Gao, J.; Wang, L.; Yu, J.; Pan, Z. Structure Tensor-Based Infrared Small Target Detection Method for a Double Linear Array Detector. *Remote Sens.* **2022**, *14*, 4785. [\[CrossRef\]](#)
4. Pang, D.; Shan, T.; Ma, P.; Li, W.; Liu, S.; Tao, R. A Novel Spatiotemporal Saliency Method for Low-Altitude Slow Small Infrared Target Detection. *IEEE Geosci. Remote Sens. Lett.* **2022**, *19*, 7000705. [\[CrossRef\]](#)
5. Du, J.; Lu, H.; Zhang, L.; Hu, M.; Chen, S.; Deng, Y.; Shen, X.; Zhang, Y. A Spatial-Temporal Feature-Based Detection Framework for Infrared Dim Small Target. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 3000412. [\[CrossRef\]](#)
6. Eysa, R.; Hamdulla, A. Issues on Infrared Dim Small Target Detection and Tracking. In Proceedings of the 2019 International Conference on Smart Grid and Electrical Automation (ICSGEA), Xiangtan, China, 10–11 August 2019; pp. 452–456.
7. Rivest, J.-F.; Fortin, R. Detection of dim targets in digital infrared imagery by morphological image processing. *Opt. Eng.* **1996**, *35*, 1886–1893. [\[CrossRef\]](#)
8. Deshpande, S.D.; Er, M.H.; Venkateswarlu, R.; Chan, P. Max-mean and max-median filters for detection of small targets. In Proceedings of the Optics & Photonics, Denver, CO, USA, 4 October 1999.
9. Yang, L.; Yang, J.; Yang, K. Adaptive detection for infrared small target under sea-sky complex background. *Electron. Lett.* **2004**, *40*, 1. [\[CrossRef\]](#)
10. Hadhoud, M.M.; Thomas, D.W. The two-dimensional adaptive LMS (TDLMS) algorithm. *IEEE Trans. Circuits Syst.* **1988**, *35*, 485–494. [\[CrossRef\]](#)
11. Widrow, B.; Glover, J.R.; McCool, J.M.; Kaunitz, J.; Williams, C.S.; Hearn, R.H.; Zeidler, J.R.; Dong, J.E.; Goodlin, R.C. Adaptive noise cancelling: Principles and applications. *Proc. IEEE* **1975**, *63*, 1692–1716. [\[CrossRef\]](#)
12. Cao, M.; Sun, D. Infrared weak target detection based on improved morphological filtering. In Proceedings of the 2020 Chinese Control And Decision Conference (CCDC), Hefei, China, 22–24 August 2020; pp. 1808–1813.
13. Chen, C.L.P.; Li, H.; Wei, Y.; Xia, T.; Tang, Y.Y. A Local Contrast Method for Small Infrared Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 574–581. [\[CrossRef\]](#)

14. Wei, Y.; You, X.; Li, H. Multiscale patch-based contrast measure for small infrared target detection. *Pattern Recognit.* **2016**, *58*, 216–226. [\[CrossRef\]](#)
15. Shi, Y.; Wei, Y.; Yao, H.; Pan, D.; Xiao, G. High-Boost-Based Multiscale Local Contrast Measure for Infrared Small Target Detection. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 33–37. [\[CrossRef\]](#)
16. Han, J.; Moradi, S.; Faramarzi, I.; Liu, C.; Zhang, H.; Zhao, Q. A Local Contrast Method for Infrared Small-Target Detection Utilizing a Tri-Layer Window. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1822–1826. [\[CrossRef\]](#)
17. Ma, Y.; Liu, Y.; Pan, Z.; Hu, Y. Method of Infrared Small Moving Target Detection Based on Coarse-to-Fine Structure in Complex Scenes. *Remote Sens.* **2023**, *15*, 1508. [\[CrossRef\]](#)
18. Fan, Z.; Bi, D.; Xiong, L.; Ma, S.; He, L.; Ding, W. Dim infrared image enhancement based on convolutional neural network. *Neurocomputing* **2018**, *272*, 396–404. [\[CrossRef\]](#)
19. Zhao, B.; Wang, C.; Fu, Q.; Han, Z. A Novel Pattern for Infrared Small Target Detection with Generative Adversarial Network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 4481–4492. [\[CrossRef\]](#)
20. Zhang, M.; Zhang, R.; Zhang, J.; Guo, J.; Li, Y.; Gao, X. Dim2Clear Network for Infrared Small Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5001714. [\[CrossRef\]](#)
21. Ying, X.; Liu, L.; Wang, Y.; Li, R.; Chen, N.; Lin, Z.; Sheng, W.; Zhou, S. Mapping Degeneration Meets Label Evolution: Learning Infrared Small Target Detection with Single Point Supervision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 15528–15538.
22. Gao, C.; Meng, D.; Yang, Y.; Wang, Y.; Zhou, X.; Hauptmann, A.G. Infrared patch-image model for small target detection in a single image. *IEEE Trans. Image Process.* **2013**, *22*, 4996–5009. [\[CrossRef\]](#)
23. Xie, Y.; Gu, S.; Liu, Y.; Zuo, W.; Zhang, W.; Zhang, L. Weighted Schatten p -norm minimization for image denoising and background subtraction. *IEEE Trans. Image Process.* **2016**, *25*, 4842–4857. [\[CrossRef\]](#)
24. Dai, Y.; Wu, Y.; Song, Y. Infrared small target and background separation via column-wise weighted robust principal component analysis. *Infrared Phys. Technol.* **2016**, *77*, 421–430. [\[CrossRef\]](#)
25. Dai, Y.; Wu, Y.; Song, Y.; Guo, J. Non-negative infrared patch-image model: Robust target-background separation via partial sum minimization of singular values. *Infrared Phys. Technol.* **2017**, *81*, 182–194. [\[CrossRef\]](#)
26. Wang, X.; Peng, Z.; Kong, D.; Zhang, P.; He, Y. Infrared dim target detection based on total variation regularization and principal component pursuit. *Image Vis. Comput.* **2017**, *63*, 1–9. [\[CrossRef\]](#)
27. Zhang, L.; Peng, L.; Zhang, T.; Cao, S.; Peng, Z. Infrared small target detection via non-convex rank approximation minimization joint $l_{2,1}$ norm. *Remote Sens.* **2018**, *10*, 1821. [\[CrossRef\]](#)
28. Wang, X.; Peng, Z.; Kong, D.; He, Y. Infrared dim and small target detection based on stable multisubspace learning in heterogeneous scene. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5481–5493. [\[CrossRef\]](#)
29. Zhang, T.; Peng, Z.; Wu, H.; He, Y.; Li, C.; Yang, C. Infrared small target detection via self-regularized weighted sparse model. *Neurocomputing* **2021**, *420*, 124–148. [\[CrossRef\]](#)
30. Dai, Y.; Wu, Y. Reweighted infrared patch-tensor model with both nonlocal and local priors for single-frame small target detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 3752–3767. [\[CrossRef\]](#)
31. Zhang, L.; Peng, Z. Infrared small target detection based on partial sum of the tensor nuclear norm. *Remote Sens.* **2019**, *11*, 382. [\[CrossRef\]](#)
32. Kong, X.; Yang, C.; Cao, S.; Li, C.; Peng, Z. Infrared small target detection via nonconvex tensor fibered rank approximation. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5000321. [\[CrossRef\]](#)
33. Zhang, Z.; Ding, C.; Gao, Z.; Xie, C. ANLPT: Self-Adaptive and Non-Local Patch-Tensor Model for Infrared Small Target Detection. *Remote Sens.* **2023**, *15*, 1021. [\[CrossRef\]](#)
34. Reed, I.S.; Gagliardi, R.M.; Stotts, L.B. Optical moving target detection with 3-D matched filtering. *IEEE Trans. Aerosp. Electron. Syst.* **1988**, *24*, 327–336. [\[CrossRef\]](#)
35. Tonissen, S.M.; Evans, R.J. Performance of dynamic programming techniques for track-before-detect. *IEEE Trans. Aerosp. Electron. Syst.* **1996**, *32*, 1440–1451. [\[CrossRef\]](#)
36. Li, Y.; Zhang, Y.; Yu, J.-G.; Tan, Y.; Tian, J.; Ma, J. A novel spatio-temporal saliency approach for robust dim moving target detection from airborne infrared image sequences. *Inf. Sci.* **2016**, *369*, 548–563. [\[CrossRef\]](#)
37. Zhao, F.; Wang, T.; Shao, S.; Zhang, E.; Lin, G. Infrared moving small-target detection via spatiotemporal consistency of trajectory points. *IEEE Geosci. Remote Sens. Lett.* **2019**, *17*, 122–126. [\[CrossRef\]](#)
38. Sun, Y.; Yang, J.; Long, Y.; An, W. Infrared small target detection via spatial-temporal total variation regularization and weighted tensor nuclear norm. *IEEE Access* **2019**, *7*, 56667–56682. [\[CrossRef\]](#)
39. Zhang, P.; Zhang, L.; Wang, X.; Shen, F.; Pu, T.; Fei, C. Edge and Corner Awareness-Based Spatial–Temporal Tensor Model for Infrared Small-Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2021**, *59*, 10708–10724. [\[CrossRef\]](#)
40. Hu, Y.; Ma, Y.; Pan, Z.; Liu, Y. Infrared Dim and Small Target Detection from Complex Scenes via Multi-Frame Spatial–Temporal Patch-Tensor Model. *Remote Sens.* **2022**, *14*, 2234. [\[CrossRef\]](#)
41. Wang, G.; Tao, B.; Kong, X.; Peng, Z. Infrared Small Target Detection Using Nonoverlapping Patch Spatial–Temporal Tensor Factorization with Capped Nuclear Norm Regularization. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5001417. [\[CrossRef\]](#)
42. Liu, T.; Yang, J.; Li, B.; Wang, Y.; An, W. Infrared Small Target Detection via Nonconvex Tensor Tucker Decomposition with Factor Prior. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5617317. [\[CrossRef\]](#)

43. Wu, F.; Yu, H.; Liu, A.; Luo, J.; Peng, Z. Infrared Small Target Detection Using Spatiotemporal 4-D Tensor Train and Ring Unfolding. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 5002922. [\[CrossRef\]](#)
44. Romera-Paredes, B.; Pontil, M. A New Convex Relaxation for Tensor Completion. In Proceedings of the Advances in Neural Information Processing Systems, Harrahs and Harveys, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 2967–2975.
45. Yang, J.-H.; Zhao, X.-L.; Ji, T.-Y.; Ma, T.-H.; Huang, T.-Z. Low-rank tensor train for tensor robust principal component analysis. *Appl. Math. Comput.* **2020**, *367*, 124783. [\[CrossRef\]](#)
46. Zhao, Q.; Zhou, G.; Xie, S.; Zhang, L.; Cichocki, A. Tensor ring decomposition. *arXiv* **2016**, arXiv:1606.05535.
47. Hackbusch, W.; Kühn, S. A new scheme for the tensor representation. *J. Fourier Anal. Appl.* **2009**, *15*, 706–722. [\[CrossRef\]](#)
48. Bigun, J.; Granlund, G.H.; Wiklund, J. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 775–790. [\[CrossRef\]](#)
49. Brown, M.; Szeliski, R.; Winder, S. Multi-image matching using multi-scale oriented patches. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–25 June 2005; Volume 511, pp. 510–517.
50. Frangi, A.F.; Niessen, W.J.; Vincken, K.L.; Viergever, M.A. Multiscale vessel enhancement filtering. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI'98, Cambridge, MA, USA, 11–13 October 1998; Wells, W.M., Colchester, A., Delp, S., Eds.; Springer: Berlin/Heidelberg, Germany, 1998; pp. 130–137.
51. Candès, E.J.; Eldar, Y.C.; Strohmer, T.; Voroninski, V. Phase retrieval via matrix completion. *SIAM Rev.* **2015**, *57*, 225–251. [\[CrossRef\]](#)
52. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenom.* **1992**, *60*, 259–268. [\[CrossRef\]](#)
53. Liu, Y.; Long, Z.; Zhu, C. Image Completion Using Low Tensor Tree Rank and Total Variation Minimization. *IEEE Trans. Multimed.* **2019**, *21*, 338–350. [\[CrossRef\]](#)
54. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends® Mach. Learn.* **2011**, *3*, 1–122.
55. Lin, Z.; Chen, M.; Ma, Y. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv* **2010**, arXiv:1009.5055.
56. Cai, J.-F.; Candès, E.J.; Shen, Z. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* **2010**, *20*, 1956–1982. [\[CrossRef\]](#)
57. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202. [\[CrossRef\]](#)
58. Hui, B.; Song, Z.; Fan, H.; Zhong, P.; Hu, W.; Zhang, X.; Lin, J.; Su, H.; Jin, W.; Zhang, Y.; et al. *A Dataset for Infrared Image Dim-Small Aircraft Target Detection and Tracking under Ground/Air Background*; China Scientific Data: Beijing, China, 2019. [\[CrossRef\]](#)
59. Sun, X.; Guo, L.; Zhang, W.; Wang, Z.; Yu, Q. Small Aerial Target Detection for Airborne Infrared Detection Systems Using LightGBM and Trajectory Constraints. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 9959–9973. [\[CrossRef\]](#)
60. Chang, C.-I. An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis. *IEEE Trans. Geosci. Remote Sens.* **2020**, *59*, 5131–5153. [\[CrossRef\]](#)
61. Sun, L.; Zhan, T.; Wu, Z.; Jeon, B. A Novel 3D Anisotropic Total Variation Regularized Low Rank Method for Hyperspectral Image Mixed Denoising. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 412. [\[CrossRef\]](#)
62. Lu, C.; Feng, J.; Chen, Y.; Liu, W.; Lin, Z.; Yan, S. Tensor robust principal component analysis with a new tensor nuclear norm. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 925–938. [\[CrossRef\]](#)
63. Liu, T.; Yang, J.; Li, B.; Xiao, C.; Sun, Y.; Wang, Y.; An, W. Nonconvex Tensor Low-Rank Approximation for Infrared Small Target Detection. *IEEE Trans. Geosci. Remote Sens.* **2022**, *60*, 5614718. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.