Hyperspectral Anomaly Detection Based on Adaptive Low-Rank Transformed Tensor

Siyu Sun, Jun Liu, *Senior Member, IEEE*, Ziwei Zhang, and Wei Li, *Senior Member, IEEE*

Abstract— Hyperspectral anomaly detection, which is aimed at distinguishing anomaly pixels from the surroundings in spatial features and spectral characteristics, has attracted considerable attention due to its various applications. In this article, we propose a novel hyperspectral anomaly detection algorithm based on adaptive low-rank transform, in which the input hyperspectral image (HSI) is divided into a background tensor, an anomaly tensor, and a noise tensor. To take full advantage of the spatial–spectral information, the background tensor is represented as the product of a transformed tensor and a low-rank matrix. The low-rank constraint is imposed on frontal slices of the transformed tensor to depict the spatial–spectral correlation of the HSI background. Besides, we initialize a matrix with predefined size and then minimize its $l_{2,1}$-norm to adaptively derive an appropriate low-rank matrix. The anomaly tensor is constrained with the $l_{2,1,1}$-norm to depict the group sparsity of anomalous pixels. We integrate all regularization terms and a fidelity term into a non-convex problem and develop a proximal alternating minimization (PAM) algorithm to solve it. Interestingly, the sequence generated by the PAM algorithm is proven to converge to a critical point. Experimental results conducted on four widely used datasets demonstrate the superiority of the proposed anomaly detector over several state-of-the-art methods.

Index Terms— Anomaly detection, group sparsity, hyperspectral images (HSIs), low rank, tensor, transform.

Nomenclature

x	Scalar.
$\mathbf{x}$	Vector.
$\mathbf{X}$	Matrix.
$\mathcal{X}$	Tensor.
$(\cdot)^T$	Transpose.
$\mathbf{X}_{ij}$	(i, j)th element.
$\mathcal{X}_{ijk}$	(i, j, k)th element.
$\mathcal{X}(i, j, :)$	(i, j)th tube.
$\|\mathbf{x}\|_2$	$\sqrt{\sum_i x_i^2}$.
$\|\mathbf{X}\|_{2,1}$	$\sum_i \|\mathbf{X}(:, i)\|_2$.
$\|\mathbf{X}\|_F$	$\sqrt{\sum_{ij} \mathbf{X}_{ij}^2}$.
$\|\mathbf{X}\|_*$	Matrix nuclear norm.
$\|\mathcal{X}\|_{2,1,1}$	$\sum_{ij} \|\mathcal{X}(i, j, :)\|_2$.
$\|\mathcal{X}\|_F$	$\sqrt{\sum_{ijk} \mathcal{X}_{ijk}^2}$.
rank$(\mathbf{X})$	Rank of $\mathbf{X}$.
tr$(\mathbf{X})$	Trace of $\mathbf{X}$.

I. Introduction

HYPERSPECTRAL images (HSIs) consisting of hundreds of contiguous and narrow bands record radiations from various materials captured by hyperspectral imaging sensors, which have attracted great attention in recent years. As 3-D data, the HSI covers a wide range of wavelengths with a high spectral resolution, which provides abundant spatial information and precise spectral characteristics, and allows for distinguishing subtle differences between objects. The superiority of the HSI makes it widely used in military and civilian fields [1], [2] and promotes various fundamental researches, such as hyperspectral target detection [3], [4] and hyperspectral anomaly detection [5], [6]. Hyperspectral target detection is usually carried out in a supervised manner, which means some prior knowledge of desired targets is necessary. On the contrary, hyperspectral anomaly detection is implemented without any prior information. On many occasions, hyperspectral anomaly detection is more practical due to the unavailable prior information of targets of interest.

Generally, the anomaly in HSIs is different from its local background in spatial features and spectral characteristics and distributed in the image with a very low probability. The main task of hyperspectral anomaly detection is to suppress background and highlight sparse anomalies. To achieve this, various methods have been proposed. As a milestone, the well-known Reed-Xiaoli (RX) detector [7] is a classical statistic-based hyperspectral anomaly detector, which can be classified into two versions, i.e., global RX (GRX) and local RX (LRX) [8]. A very important assumption of the GRX is that all the background pixels in HSI obey a multivariate Gaussian distribution. Disregard of the sparse abnormal pixels, the statistics of the Gaussian distribution, mean and covariance, are estimated via all the pixels in the HSI. Then the Mahalanobis distance between each pixel under test (PUT) and the background is considered as the anomaly value. Compared with the GRX, the LRX obtains an improved performance

due to the utilization of a dual-window strategy. Specifically, the pixels between the inner window and the outer window centered on the PUT are exploited to estimate the statistics of the local background. As a result, the local background of each PUT replaces the entire HSI in measuring the anomaly score. To better formulate the complicated background, various variants of the RX detector have been proposed, such as kernel RX (KRX) [9] and cluster KRX [10].

In addition to the statistic-based RX and its variants, detectors based on low-rank and sparse matrix decomposition (LRaSMD) have been proposed to take full advantage of structural information of different components of the HSI, such as the low-rank and piecewise smoothness properties of the background and the sparsity of the anomaly. By virtue of the robust principal analysis component (RPCA) algorithm [11], the RPCA-RX was proposed in [12], in which the input HSI was first divided into a low-rank matrix and a sparse matrix, and then the RX was applied to the obtained sparse matrix. Li et al. [13] exploited the truncated nuclear norm instead of the nuclear norm in RPCA to better formulate the low-rank background. Moreover, the total variation (TV) regularization was used to depict the piecewise smoothness of the background. To obtain a robust separation of the superposition of different components, the so-called "Go Decomposition" (GoDec) algorithm [14] was proposed with the consideration of Gaussian noise. By virtue of this GoDec algorithm, the low-rank background, sparse anomaly, and dense noise were effectively separated in [15] and [16]. In [16], six types of RX detectors were proposed with different combinations of the derived low-rank matrix and the sparse matrix. A more robust method was proposed in [17] by exploiting the mixture of Gaussian model to formulate the anomaly and complicated noise of the HSI.

Considering different types of materials in an HSI, the background is usually assumed to lie in a union of multiple subspaces. The low-rank representation (LRR) [18] has been widely used in hyperspectral anomaly detection [19], [20], [21], [22]. The main idea of the LRR-based anomaly detectors is that the background pixel can be well-approximated by a linear combination of atoms of a given background dictionary, while the anomalous pixel cannot and is treated as a samplewise outlier. Usually, a low-rank constraint is imposed on the representation coefficient matrix of the background. To better describe the local structure of the background, some other constraints are also incorporated into the LRR model, such as l_1-norm regularization [19] and TV regularization [20], [21]. Qu et al. [23] found that the differences between the abundance vectors of anomaly and background are more distinguished, and then applied the LRR to the abundant matrix of the HSI. The performances of the LRR-based anomaly detectors are closely related to the background dictionary. Various background construction methods have been proposed to include the basis of different background materials as much as possible while excluding the anomaly [24], [25], [26].

Although LRaSMD-based detectors and representation-based detectors achieve considerable results, the operation of unfolding a 3-D HSI into a matrix destroys the spatial structure of the HSI and then degrades the detection performance. Recently, the tensor has attracted more and more attentions

and has been widely used in HSI processing due to its potential to preserve the spatial–spectral information of the HSI, such as restoration [27], [28] and super-resolution [29], [30], [31]. Different tensor decomposition models can be applied in HSI processing, such as Tucker decomposition [32] and CANDECOMP/PARAFAC (CP) decomposition [33]. To make full use of the inherent information of the HSI, various improved versions based on the two decomposition methods were proposed in [34] and [35]. The two models have also been introduced for hyperspectral anomaly detection [36], [37], [38], [39]. However, the performance of these detectors is limited by the corresponding models of tensor decomposition. For the Tucker decomposition, the unfolding operation is necessary to obtain the individual correlation information of each mode, which destroys the multidimensional structure and then cannot give a comprehensive description of the spatial–spectral information of the HSI. For the CP composition, the calculation of the CP rank is non-deterministic polynomial. To obtain an approximate representation of a tensor and avoid the loss of information caused by the unfolding operation, tensor–tensor product (t-product) and tensor singular value decomposition (t-SVD) were defined in [40]. Based on the t-product, Xu et al. exploited the induced tensor tubal rank [41] and tensor nuclear norm (TNN) [42] to capture the joint correlation between different modes of the HSI background [43]. Sun et al. [44] proposed a low average rank with a TV regularization anomaly detection (LARTVAD) method, where the background is expressed as the product of a coefficient tensor and a background dictionary, and a different TNN [45] constraint is imposed on the coefficient tensor instead of the background tensor.

The main idea of the TNN-based method is to integrate the different dimensions of information into frontal slices of the transformed tensor with the aid of the discrete Fourier transform. In [46], a more general t-product and its induced TNN based on any invertible linear transforms were proposed. Moreover, a noninvertible framelet transform was exploited to fuse multidimensional information [47]. However, these transforms depend on predefined matrices, which cannot fully describe the data-specific information. As a result, several data-dependent transforms were proposed to solve this problem [48], [49], in which data-adaptive dictionaries were learned. However, satisfactory performance can be achieved when the dictionary consists of a large number of atoms, which increases the number of frontal slices of the transformed tensor and then leads to additional computational costs. In this article, by virtue of the transform, we propose a novel hyperspectral anomaly method-based adaptive low-rank transformed tensor (ALRTT). The joint correlation of the spatial and spectral dimensions of the HSI background is formulated by imposing the nuclear norm on frontal slices of the transformed tensor. To directly use the strong spectral correlation, we express the background as the product of the transformed tensor and a low-rank matrix instead of the matrix with a large number of columns. To adaptively obtain the low-rank matrix, we minimize the $l_{2,1}$-norm of an initial matrix with an appropriate number of columns. It is noted that the number of columns of the initial matrix can be much less than the number of bands of an HSI. For the anomaly tensor, we exploit the $l_{2,1,1}$-

norm to depict the group sparsity of anomalous pixels. All the regularization terms and the fidelity term are integrated into a non-convex problem, and a proximal alternating minimization (PAM) algorithm is exploited to optimize it. Finally, the anomaly detection map is derived from the anomaly tensor.

The main contributions of this work can be summarized as follows.

1) We develop a hyperspectral anomaly detection method based on the tensor transform. The spatial information and spectral information are integrated into frontal slices of a transformed tensor. The spatial–spectral correlation of the background is depicted by minimizing the nuclear norm of each frontal slice of the transformed tensor.

2) The background tensor is expressed as the product of a transformed tensor and a low-rank matrix. We develop a method that can adaptively learn the low-rank matrix. Compared with the traditional transform, the low-rank matrix allows for directly describing the spectral correlation of the background and improving the computational efficiency.

3) We develop a PAM algorithm to solve the proposed problem and give convergence analysis of the proposed algorithm. Extensive experiments are conducted on several datasets, and the result demonstrates that the proposed anomaly detector outperforms its counterparts.

The rest of this article is organized as follows. Some notations and definitions are introduced in Section II. The proposed ALRTT method including optimization procedure and convergence analysis is presented in Section III. The experimental results and discussions are provided in Section IV. Finally, we conclude this article in Section V.

II. NOTATIONS

Some notations and the corresponding interpretations are listed in Nomenclature. Besides, the $l_{2,0}$-norm of $\mathbf{A}$ denoted by $\|\mathbf{A}\|_{2,0}$ is the number of nonzero columns. The derivative of the function f is denoted by ∇f, and the derivative of f with respect to $\mathbf{X}$ is represented by $\nabla_{\mathbf{X}} f$. The computational complexity is denoted by $\mathcal{O}(n)$ which is the implementation requiring a number of ops proportional to n. The mode-i unfolding operator of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, denoted by $\text{unfold}_i(\mathcal{X}) = \mathbf{X}_{(i)} \in \mathbb{R}^{n_i \times \prod_{k=1,k\neq i}^{3} n_k}$, can be realized via $\mathbf{X}_{(i)}(l_i, j) = \mathcal{X}(l_1, l_2, l_3)$ with $j = 1 + \sum_{k=1,k\neq i}^{3} (l_k - 1) J_k$ and $J_k = \prod_{m=1,m\neq i}^{k-1} l_m$. The inverse operator of $\text{unfold}_i(\mathcal{X})$ is $\mathcal{X} = \text{fold}_i(\mathbf{X}_{(i)})$. For an N-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \dots \times n_N}$, its mode-$k$ product with $\mathbf{A} \in \mathbb{R}^{J \times n_k}$ can be calculated by $(\mathcal{X} \times_k \mathbf{A})_{i_1,\dots,i_{k-1},j,i_{k+1},i_N} = \sum_{i_k=1}^{n_k} x_{i_1,i_2,\dots,i_N} a_{j,i_k}$. The discrete Fourier transform along the third mode of $\mathcal{X}$ is denoted by $\bar{\mathcal{X}}$.

The block circulant matrix of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is defined as follows:

$$\text{bcirc}(\mathcal{X}) = \begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_{n_3} & \cdots & \mathbf{X}_2 \\ \mathbf{X}_2 & \mathbf{X}_1 & \cdots & \mathbf{X}_3 \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{X}_{n_3} & \mathbf{X}_{n_3-1} & \cdots & \mathbf{X}_1 \end{bmatrix} \quad (1)$$

where $\mathbf{X}_i$ is the ith frontal slice of $\mathcal{X}$.

Definition 1 (t-Product [40]): Given two three-order tensors, $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{Y} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$, the t-product of $\mathcal{X}$ and $\mathcal{Y}$ is a three-order tensor $\mathcal{Z} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$, which is calculated by

$$\mathcal{Z} = \mathcal{X} * \mathcal{Y} = \text{fold}_3(\text{bcirc}(\mathcal{X}) \cdot \text{unfold}_3(\mathcal{Y})). \quad (2)$$

Definition 2 (Identity Tensor [40]): The identity tensor $\mathcal{I} \in \mathbb{R}^{n_1 \times n_1 \times n_3}$ is the tensor whose first frontal slice is an $n_1 \times n_1$ identity matrix and other frontal slices are all zeros.

Definition 3 (Orthogonal Tensor [40]): The tensor $\mathcal{Q}$ is an orthogonal tensor if

$$\mathcal{Q}^T * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^T = \mathcal{I}. \quad (3)$$

Definition 4 (f-Diagonal [40]): The tensor is f-diagonal if each frontal slice of which is a diagonal matrix.

Theorem 1 (t-SVD [40]): A tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ can be decomposed as

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T \quad (4)$$

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

Definition 5 (Tensor Tubal Rank [42]): For a three-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with its t-SVD $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$, the tensor tubal rank of $\mathcal{X}$ is defined as the number of nonzero tubes of $\mathcal{S}$.

Definition 6 (TNN [42]): For tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ with its t-SVD $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^T$, the TNN of $\mathcal{X}$ is defined as

$$\|\mathcal{X}\|_{\text{tnn}} = \sum_{i=1}^{n_3} \|\bar{\mathbf{X}}_i\|_* \quad (5)$$

where $\bar{\mathbf{X}}_i$ is the ith frontal slice of $\bar{\mathcal{X}}$.

III. PROPOSED ANOMALY DETECTION ALGORITHM

A. Problem Formulation

An HSI consisting of b spectral bands with spatial size $h \times w$ is denoted by a three-order tensor $\mathcal{Y} \in \mathbb{R}^{h \times w \times b}$. Assume that the HSI is mixed with additive Gaussian noise $\mathcal{N} \in \mathbb{R}^{h \times w \times b}$. Denoting the background tensor and anomaly tensor by $\mathcal{B} \in \mathbb{R}^{h \times w \times b}$ and $\mathcal{S} \in \mathbb{R}^{h \times w \times b}$, respectively, we can express the observed data $\mathcal{Y}$ as

$$\mathcal{Y} = \mathcal{B} + \mathcal{S} + \mathcal{N}. \quad (6)$$

In this article, we want to separate $\mathcal{S}$ and $\mathcal{B}$ from the observed data $\mathcal{Y}$ under the contamination of $\mathcal{N}$. To this end, we establish the following detection model to make full use of the potential information of the background and anomaly, which can be written as

$$\min_{\mathcal{B},\mathcal{S}} \frac{1}{2} \|\mathcal{Y} - \mathcal{B} - \mathcal{S}\|_F^2 + \mathcal{R}_1(\mathcal{B}) + \mathcal{R}_2(\mathcal{S}) \quad (7)$$

where the first term $(1/2)\|\mathcal{Y} - \mathcal{B} - \mathcal{S}\|_F^2$ is the fidelity item to avoid greatly deviation of $\mathcal{B} + \mathcal{S}$ from $\mathcal{Y}$, and the second term $\mathcal{R}_1(\mathcal{B})$ and the third term $\mathcal{R}_2(\mathcal{S})$ are the regularized terms allowing for integrating the potential information of the background and anomaly, respectively.
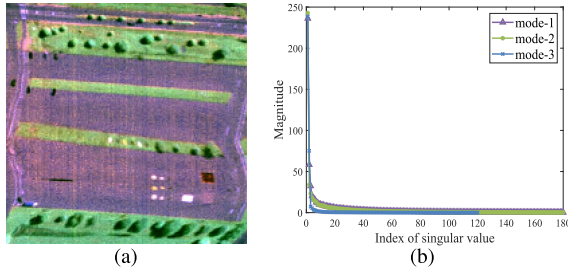
Fig. 1. (a) False-color image of the SpecTIR dataset. (b) Singular values of the unfolding matrices along different modes.

1) Spatial–Spectral Correlation: First, the HSI consisting of a large number of contiguous and narrow bands contains a lot of redundant information, which leads to a strong spectral correlation of the background. Second, the image of each band can be considered a gray image, and the background is the low-rank major component. That is to say, the correlation also exists in two spatial directions of the background. As shown in Fig. 1, the singular values of the unfolding matrices along different directions decrease very fast. As a result, the hyperspectral background is considered to possess the spatial–spectral correlation. Usually, the Tucker decomposition can be exploited to depict this property. However, the unfolding operation will destroy the multiway structure of the three-order HSI. Therefore, we use transform to preserve the structural integrity of the three-order background and formulate the spatial–spectral correlation. Specifically, the background tensor $\mathcal{B}$ can be written as

$$\mathcal{B} = \mathcal{M} \times_3 \mathbf{A} \tag{8}$$

where $\mathbf{A} \in \mathbb{R}^{b \times d}$ and $\mathcal{M} \in \mathbb{R}^{h \times w \times d}$. With (8), the spatial information and spectral information of the background are fused into each frontal slice of $\mathcal{M}$. Then the spatial–spectral correlation can be characterized by taking $\sum_{i=1}^{d} \|\mathbf{M}_i\|_*$ as a regularization, where $\mathbf{M}_i$ is the ith frontal slice of $\mathcal{M}$.

2) Adaptive Low-Rank Transform: In most transform-based applications, the matrix $\mathbf{A}$ is a predefined orthogonal matrix [46], [50] or matrix with $d \gg b$ [48], [49]. For hyperspectral anomaly detection, however, such a matrix $\mathbf{A}$ cannot directly describe the high spectral correlation of the background as discussed earlier. Besides, a large d is not conducive to the computational efficiency of anomaly detection, because it is time-consuming to compute the nuclear norm. For $\mathcal{B} = \mathcal{M} \times_3 \mathbf{A}$, unfolding both sides of the equation along the spectral direction gives the term

$$\mathbf{B}_{(3)} = \mathbf{A}\mathbf{M}_{(3)} \tag{9}$$

then we have $\text{rank}(\mathbf{B}_{(3)}) \leq \text{rank}(\mathbf{A})$. That is to say, if we constrain $\mathbf{A}$ to be a low-rank matrix, then $\mathbf{B}_{(3)}$ is also forced to be low rank. It is noted that $\text{rank}(\mathbf{A}) \leq \|\mathbf{A}\|_{2,0}$, then we can achieve the low-rank constraint by minimizing $\|\mathbf{A}\|_{2,0}$. However, the minimization of $\|\mathbf{A}\|_{2,0}$ is an NP problem. Therefore, we adopt a relaxation strategy and replace $\|\mathbf{A}\|_{2,0}$ with $\|\mathbf{A}\|_{2,1}$. To verify the low rankness of the background can be well-depicted by minimizing $\|\mathbf{A}\|_{2,1}$, we give the following proposition.

Proposition 1: For the problem

$$\min \ \|\mathbf{A}\|_{2,1}$$
$$\text{s.t. } \mathbf{B}_{(3)} = \mathbf{A}\mathbf{M}_{(3)} \tag{10}$$

where $\text{rank}(\mathbf{B}_{(3)}) = r$, we have $\text{rank}(\mathbf{A}^*) = \text{rank}(\mathbf{B}_{(3)})$, where $\mathbf{A}^*$ is an optimal solution of (10).

Proof: On one hand, we have $\text{rank}(\mathbf{A}^*) \geq \text{rank}(\mathbf{B}_{(3)}) = r$. On the other hand, assume that $\text{rank}(\mathbf{A}^*) > r$. It is noted that $\mathbf{B}_{(3)} = \mathbf{A}^*\mathbf{M}_{(3)}$, which indicates $\text{span}(\mathbf{B}_{(3)}) \subseteq \text{span}(\mathbf{A}^*)$, where $\text{span}(\cdot)$ denotes the column space. We can select r columns from $\mathbf{A}^*$ that are the basis of $\text{span}(\mathbf{B}_{(3)})$. Denoting the selected columns by $\bar{\mathbf{A}}$, we can easily obtain an $\bar{\mathbf{M}}_{(3)}$ with $\mathbf{B}_{(3)} = \bar{\mathbf{A}}\bar{\mathbf{M}}_{(3)}$. Then we have $\|\bar{\mathbf{A}}\|_{2,1} < \|\mathbf{A}^*\|_{2,1}$, which means $\text{rank}(\mathbf{A}^*) > r$ is not true. Finally, we have $\text{rank}(\mathbf{A}^*) = \text{rank}(\mathbf{B}_{(3)})$. ∎

For problem (10), we always have $(1/c)\|\mathbf{A}\|_{2,1} < \|\mathbf{A}\|_{2,1}$ with $c > 1$. With the nuclear norm constraint imposed on frontal slices of $\mathcal{M}$, the nonzero column of $\mathbf{A}$ does not tend to be arbitrarily small. Finally, the regularization on the background tensor can be written as

$$\mathcal{R}_1(\mathcal{B}) = \mathcal{R}_1(\mathbf{A}, \mathcal{M}) = \lambda\|\mathbf{A}\|_{2,1} + \beta \sum_{i=1}^{d} \|\mathbf{M}_i\|_* \tag{11}$$

where λ and β are two regularized parameters. The utilization of $l_{2,1}$-norm on $\mathbf{A}$ has two main advantages. First, it enables us to update each column of $\mathbf{A}$ in sequence. Second, the parameter d can be initialized with an appropriate value ($d \ll b$). In the iterative process, all the elements of some columns of $\mathbf{A}$ will become zero, and then a data-dependent matrix $\mathbf{A}$ can be obtained.

3) Group Sparsity: The anomaly pixels usually occur with a very low probability and are sample-specific sparse, which is the so-called group sparsity. Then the regularization on the anomaly tensor can be written as

$$\mathcal{R}_2(\mathcal{S}) = \gamma \|S\|_{2,1,1} \tag{12}$$

where γ is a regularization parameter. The norm $\|S\|_{2,1,1}$ is actually the sum of the l_2-norm of all the tubes of $\mathcal{S}$. To obtain a small value of $\|S\|_{2,1,1}$, the l_2-norm of some tubes should become zero, and then the value of all the elements in the corresponding tubes should be zero, which leads to the group sparsity of the anomaly tensor.

Finally, inserting (8), (9), and (12) into (7), we have the anomaly detection model as

$$\min_{\mathcal{M},\mathbf{A},\mathcal{S}} \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \times_3 \mathbf{A} - \mathcal{S}\|_F^2 + \lambda\|\mathbf{A}\|_{2,1} + \beta \sum_{i=1}^{d} \|\mathbf{M}_i\|_*$$
$$+ \gamma \|\mathcal{S}\|_{2,1,1}. \tag{13}$$

B. Optimization

Note that (13) is a non-convex optimization problem. In this section, we exploit the PAM algorithm to solve this problem. The original problem is divided into three subproblems, i.e., $\mathcal{M}$ subproblem, $\mathbf{A}$ subproblem, and $\mathcal{S}$ subproblem. In each iteration, we update one of the three variables with the other

two fixed. The iterative process of the proposed ALRTT is as follows.

1) $\mathcal{M}$ *Subproblem:* The objective function can be written as

$$\mathcal{M}^{t+1} = \arg\min_{\mathcal{M}} \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \times_3 \mathbf{A}^t - \mathcal{S}^t\|_F^2$$
$$+ \beta \sum_{i=1}^{d} \|\mathbf{M}_i\|_* + \frac{\rho}{2}\|\mathcal{M} - \mathcal{M}^t\|_F^2. \tag{14}$$

It is noted that

$$\|\mathcal{Y} - \mathcal{M} \times_3 \mathbf{A} - \mathcal{S}\|_F^2 = \|\mathbf{Y}_{(3)} - \mathbf{A}\mathbf{M}_{(3)} - \mathbf{S}_{(3)}\|_F^2$$
$$= \|\mathbf{Y}_{(3)} - \sum_{i=1}^{d} \mathbf{a}_i \mathbf{m}_i^T - \mathbf{S}_{(3)}\|_F^2 \tag{15}$$

where $\mathbf{m}_i = \text{vec}(\mathbf{M}_i)$ is the vectorization of $\mathbf{M}_i$, and $\mathbf{a}_i$ is the ith column of $\mathbf{A}$. We regard $\mathbf{M}_k$'s as d independent variables and update them by solving a series of subproblems. Then, we have

$$\{\mathbf{M}_1^{t+1}, \ldots, \mathbf{M}_d^{t+1}\}$$
$$= \arg\min_{\mathbf{M}_1, \ldots, \mathbf{M}_d} \frac{1}{2}\left\|\mathbf{Y}_{(3)} - \sum_{i=1}^{d} \mathbf{a}_i^t \mathbf{m}_i^T - \mathbf{S}_{(3)}^t\right\|_F^2$$
$$+ \beta \sum_{i=1}^{d} \|\mathbf{M}_i\|_* + \frac{\rho}{2} \sum_{i=1}^{d} \|\mathbf{M}_i - \mathbf{M}_i^t\|_F^2. \tag{16}$$

Define

$$\mathbf{X}_k^t = \mathbf{Y}_{(3)} - \mathbf{S}_{(3)}^t - \sum_{i=1}^{k-1} \mathbf{a}_i^t (\mathbf{m}_i^{t+1})^T - \sum_{i=k+1}^{d} \mathbf{a}_i^t (\mathbf{m}_i^t)^T \tag{17}$$

then, $\mathbf{M}_k$'s subproblems can be written as

$$\mathbf{M}_k^{t+1} = \arg\min_{\mathbf{M}_k} \frac{1}{2}\|\mathbf{X}_k^t - \mathbf{a}_k^t \mathbf{m}_k^T\|_F^2 + \beta\|\mathbf{M}_k\|_*$$
$$+ \frac{\rho}{2}\|\mathbf{M}_k - \mathbf{M}_k^t\|_F^2$$
$$= \arg\min_{\mathbf{M}_k} \beta\|\mathbf{M}_k\|_* + \frac{\rho}{2}\text{tr}\left(\mathbf{M}_k \mathbf{M}_k^T - 2\mathbf{M}_k(\mathbf{M}_k^t)^T\right)$$
$$+ \frac{1}{2}\text{tr}\left(\mathbf{a}_k^t \mathbf{m}_k^T \mathbf{m}_k (\mathbf{a}_k^t)^T - 2(\mathbf{X}_k^t)^T \mathbf{a}_k^t \mathbf{m}_k^T\right). \tag{18}$$

It is noted that $\text{tr}(\mathbf{x}\mathbf{y}^T) = \text{tr}[\text{vec}^{-1}(\mathbf{x})(\text{vec}^{-1}(\mathbf{y}))^T]$, where $\text{vec}^{-1}(\cdot)$ is the inverse operator of $\text{vec}(\cdot)$, then we have

$$\text{tr}\left(\mathbf{a}_k^t \mathbf{m}_k^T \mathbf{m}_k (\mathbf{a}_k^t)^T - 2(\mathbf{X}_k^t)^T \mathbf{a}_k^t \mathbf{m}_k^T\right)$$
$$= (\mathbf{a}_k^t)^T \mathbf{a}_k^t \text{tr}\left(\mathbf{m}_k^T \mathbf{m}_k - 2(\mathbf{X}_k^t)^T \mathbf{a}_k^t \mathbf{m}_k^T\right)$$
$$= (\mathbf{a}_k^t)^T \mathbf{a}_k^t \text{tr}(\mathbf{M}_k \mathbf{M}_k^T) - 2\text{tr}\left[\text{vec}^{-1}\left((\mathbf{X}_k^t)^T \mathbf{a}_k^t\right)\mathbf{M}_k^T\right]. \tag{19}$$

Inserting (19) into (18), we can obtain

$$\mathbf{M}_k^{t+1} = \arg\min_{\mathbf{M}_k} \beta\|\mathbf{M}_k\|_* + \frac{(\mathbf{a}_k^t)^T \mathbf{a}_k^t + \rho}{2}\text{tr}(\mathbf{M}_k \mathbf{M}_k^T)$$
$$- \text{tr}\left[\text{vec}^{-1}\left((\mathbf{X}_k^t)^T \mathbf{a}_k^t\right)\mathbf{M}_k^T + \rho\mathbf{M}_k^t \mathbf{M}_k^T\right]$$
$$= \arg\min_{\mathbf{M}_k} \beta\|\mathbf{M}_k\|_*$$
$$+ \frac{\tau}{2} \cdot \left\|\mathbf{M}_k - \frac{\text{vec}^{-1}\left((\mathbf{X}_k^t)^T \mathbf{a}_k^t\right) + \rho\mathbf{M}_k^t}{\tau}\right\|_F^2 \tag{20}$$

where

$$\tau = (\mathbf{a}_k^t)^T \mathbf{a}_k^t + \rho. \tag{21}$$

Finally, $\mathbf{M}_k^{t+1}$ can be derived by

$$\mathbf{M}_k^{t+1} = \mathcal{D}_{\frac{\beta}{\tau}}\left(\frac{\text{vec}^{-1}\left((\mathbf{X}_k^t)^T \mathbf{a}_k^t\right) + \rho\mathbf{M}_k^t}{\tau}\right) \tag{22}$$

where $\mathcal{D}_{(\beta/\tau)}(\cdot)$ is the singular value thresholding operator [51].

2) $\mathbf{A}$ *Subproblem:* The objective function can be written as

$$\mathbf{A}^{t+1} = \arg\min_{\mathbf{A}} \frac{1}{2}\|\mathcal{Y} - \mathcal{M}^{t+1} \times_3 \mathbf{A} - \mathcal{S}^t\|_F^2 + \lambda\|\mathbf{A}\|_{2,1}$$
$$+ \frac{\rho}{2}\|\mathbf{A} - \mathbf{A}^t\|_F^2. \tag{23}$$

We also update $\mathbf{A}$ by solving a series of $\mathbf{a}_k$'s subproblems. We can rewrite (23) as

$$\{\mathbf{a}_1^{t+1}, \ldots, \mathbf{a}_d^{t+1}\}$$
$$= \arg\min_{\mathbf{a}_1, \ldots, \mathbf{a}_d} \frac{1}{2}\left\|\mathbf{Y}_{(3)} - \sum_{i=1}^{d} \mathbf{a}_i (\mathbf{m}_i^{t+1})^T - \mathbf{S}_{(3)}^t\right\|_F^2$$
$$+ \lambda \sum_{i=1}^{d} \|\mathbf{a}_i\|_2 + \frac{\rho}{2} \sum_{i=1}^{d} \|\mathbf{a}_i - \mathbf{a}_i^t\|_2^2. \tag{24}$$

Define

$$\mathbf{Z}_k^t = \mathbf{Y}_{(3)} - \mathbf{S}_{(3)}^t - \sum_{i=1}^{k-1} \mathbf{a}_i^{t+1} (\mathbf{m}_i^{t+1})^T - \sum_{i=k+1}^{d} \mathbf{a}_i^t (\mathbf{m}_i^{t+1})^T \tag{25}$$

then, we have

$$\mathbf{a}_k^{t+1} = \arg\min_{\mathbf{a}_k} \frac{1}{2}\|\mathbf{Z}_k^t - \mathbf{a}_k(\mathbf{m}_k^{t+1})^T\|_F^2 + \lambda\|\mathbf{a}_k\|_2$$
$$+ \frac{\rho}{2}\|\mathbf{a}_k - \mathbf{a}_k^t\|_2^2$$
$$= \arg\min_{\mathbf{a}_k} \lambda\|\mathbf{a}_k\|_2 + \frac{\rho}{2}\text{tr}\left(\mathbf{a}_k \mathbf{a}_k^T - 2\mathbf{a}_k(\mathbf{a}_k^t)^T\right)$$
$$+ \frac{1}{2}\text{tr}\left(\mathbf{a}_k (\mathbf{m}_k^{t+1})^T \mathbf{m}_k^{t+1} \mathbf{a}_k^T - 2\mathbf{Z}_k^t \mathbf{m}_k^{t+1} \mathbf{a}_k^T\right)$$
$$= \arg\min_{\mathbf{a}_k} \lambda\|\mathbf{a}_k\|_2 + \frac{(\mathbf{m}_k^{t+1})^T \mathbf{m}_k^{t+1} + \rho}{2}$$
$$\times \left\|\mathbf{a}_k - \frac{\mathbf{Z}_k^t \mathbf{m}_k^{t+1} + \rho\mathbf{a}_k^t}{(\mathbf{m}_k^{t+1})^T \mathbf{m}_k^{t+1} + \rho}\right\|_2^2. \tag{26}$$

Finally, (26) can be solved by

$$\mathbf{a}_k^{t+1}$$
$$= \text{soft}_{2.1}\left(\frac{\mathbf{Z}_k^t \mathbf{m}_k^{t+1} + \rho \mathbf{a}_k^t}{(\mathbf{m}_k^{t+1})^T \mathbf{m}_k^{t+1} + \rho}, \frac{\lambda}{(\mathbf{m}_k^{t+1})^T \mathbf{m}_k^{t+1} + \rho}\right) \tag{27}$$

where

$$\text{soft}_{2.1}(\mathbf{x}, a) = \max\left(1 - \frac{a}{\|\mathbf{x}\|_2}, 0\right) \cdot \mathbf{x}. \tag{28}$$

3) $\mathcal{S}$ *Subproblem:* The objective function can be written as

$$\mathcal{S}^{t+1} = \arg\min_{\mathcal{S}} \frac{1}{2}\|\mathcal{Y} - \mathcal{M}^{t+1} \times_3 \mathbf{A}^{t+1} - \mathcal{S}\|_F^2$$
$$+ \gamma \|\mathcal{S}\|_{2.1.1} + \frac{\rho}{2}\|\mathcal{S} - \mathcal{S}^t\|_F^2$$
$$= \arg\min_{\mathcal{S}} \gamma \|\mathcal{S}\|_{2.1.1}$$
$$+ \frac{1+\rho}{2}\left\|\mathcal{S} - \frac{\mathcal{Y} - \mathcal{M}^{t+1} \times_3 \mathbf{A}^{t+1} + \rho \mathcal{S}^t}{1+\rho}\right\|_F^2. \tag{29}$$

Define

$$\tilde{\mathcal{S}}^t = \frac{\mathcal{Y} - \mathcal{M}^{t+1} \times_3 \mathbf{A}^{t+1} + \rho \mathcal{S}^t}{1+\rho} \tag{30}$$

we have

$$\mathcal{S}^{t+1}(i, j, :) = \text{soft}_{2.1}\left(\tilde{\mathcal{S}}^t(i, j, :), \frac{\gamma}{1+\rho}\right). \tag{31}$$

The iteration is terminated as t reaches the predefined $t_{\max}$. Then the detection map can be calculated by

$$\mathbf{R}_{ij} = \|\mathcal{S}(i, j, :)\|_2. \tag{32}$$

The pseudocode for solving (13) is summarized in Algorithm 1.

Algorithm 1: Optimization Procedure for Solving (13)

Input: tensor $\mathcal{Y}$, tradeoff parameters β, λ, γ, proximal parameter ρ, and d.
Initialization: $\mathcal{M}^0$, $\mathbf{A}^0$, $\mathcal{S}^0$, $t_{\max}$.
1: **for** $t = 1$ to $t_{\max}$ **do**
2:   Update $\mathcal{M}$ by (22);
3:   Update $\mathbf{A}$ by (27);
4:   Update $\mathcal{S}$ by (31).
5: **end for**
Output: obtain detection map with (32).

C. Complexity and Convergence Analysis

The computational complexity of updating $\mathcal{M}$, $\mathbf{A}$, and $\mathcal{S}$ in each iteration is $\mathcal{O}(whbd^2 + dwh\min(w,h))$, $\mathcal{O}(whbd^2)$, and $\mathcal{O}(dwhb)$, respectively. As a result, the computational complexity of Algorithm 1 in each iteration is $\mathcal{O}(whbd^2 + dwh\min(w,h))$.

To prove the sequence $\{\mathcal{M}^t, \mathbf{A}^t, \mathcal{S}^t\}$ generated by Algorithm 1 converges to a critical point of the objective function, we define

$$f(\mathcal{M}, \mathbf{A}, \mathcal{S}) = f_0(\mathcal{M}, \mathbf{A}, \mathcal{S}) + f_1(\mathcal{M}) + f_2(\mathbf{A}) + f_3(\mathcal{S}) \tag{33}$$

where

$$\begin{cases} f_0(\mathcal{M}, \mathbf{A}, \mathcal{S}) = \frac{1}{2}\|\mathcal{Y} - \mathcal{M} \times_3 \mathbf{A} - \mathcal{S}\|_F^2 \\ f_1(\mathcal{M}) = \beta \sum_{i=1}^d \|\mathbf{M}_i\|_* \\ f_2(\mathbf{A}) = \lambda \|\mathbf{A}\|_{2.1} \\ f_3(\mathcal{S}) = \gamma \|\mathcal{S}\|_{2.1.1}. \end{cases} \tag{34}$$

Theorem 1: The sequence $\{\mathcal{M}^t, \mathbf{A}^t, \mathcal{S}^t\}$ generated by Algorithm 1 converges to a critical point of (12) if the following conditions are satisfied.
   a) The sequence $\{\mathcal{M}^t, \mathbf{A}^t, \mathcal{S}^t\}$ is bounded.
   b) The function f is a semi-algebraic function.
   c) $f_i(\cdot), i = 1, 2, 3$ are proper and lower semi-continuous function.
   d) $\inf f > -\infty$, $\inf f_i > -\infty, i = 0, 1, 2, 3$.
   e) For variables $\mathcal{M}$, $\mathbf{A}$, and $\mathcal{S}$, $\nabla_{\mathcal{M}} f_0$, $\nabla_{\mathbf{A}} f_0$, and $\nabla_{\mathcal{S}} f_0$ are Lipschitz continuous with Lipschitz constants L_m, L_a, and L_s, respectively.
   f) f_0 has a continuous first-order derivative, and ∇f_0 is Lipschitz continuous on any bounded set.
Proof: It is noted that

$$\begin{cases} \mathcal{M}^{t+1} = \\ \quad \arg\min_{\mathcal{M}} f_0(\mathcal{M}, \mathbf{A}^t, \mathcal{S}^t) + f_1(\mathcal{M}) + \frac{\rho}{2}\|\mathcal{M} - \mathcal{M}^t\|_F^2 \\ \mathbf{A}^{t+1} = \\ \quad \arg\min_{\mathbf{A}} f_0(\mathcal{M}^{t+1}, \mathbf{A}, \mathcal{S}^t) + f_2(\mathbf{A}) + \frac{\rho}{2}\|\mathbf{A} - \mathbf{A}^t\|_F^2 \\ \mathcal{S}^{t+1} = \\ \quad \arg\min_{\mathcal{S}} f_0(\mathcal{M}^{t+1}, \mathbf{A}^{t+1}, \mathcal{S}) + f_3(\mathcal{S}) + \frac{\rho}{2}\|\mathcal{S} - \mathcal{S}^t\|_F^2 \end{cases} \tag{35}$$

then, we have

$$\begin{cases} f_0(\mathcal{M}^{t+1}, \mathbf{A}^t, \mathcal{S}^t) + f_1(\mathcal{M}^{t+1}) + \frac{\rho}{2}\|\mathcal{M}^{t+1} - \mathcal{M}^t\|_F^2 \\ \quad \leq f_0(\mathcal{M}^t, \mathbf{A}^t, \mathcal{S}^t) + f_1(\mathcal{M}^t) \\ f_0(\mathcal{M}^{t+1}, \mathbf{A}^{t+1}, \mathcal{S}^t) + f_2(\mathbf{A}^{t+1}) + \frac{\rho}{2}\|\mathbf{A}^{t+1} - \mathbf{A}^t\|_F^2 \\ \quad \leq f_0(\mathcal{M}^{t+1}, \mathbf{A}^t, \mathcal{S}^t) + f_2(\mathbf{A}^t) \\ f_0(\mathcal{M}^{t+1}, \mathbf{A}^{t+1}, \mathcal{S}^{t+1}) + f_3(\mathcal{S}^{t+1}) + \frac{\rho}{2}\|\mathcal{S}^{t+1} - \mathcal{S}^t\|_F^2 \\ \quad \leq f_0(\mathcal{M}^{t+1}, \mathbf{A}^{t+1}, \mathcal{S}^t) + f_3(\mathcal{S}^t). \end{cases} \tag{36}$$

Summing up the three inequalities in (36) gives the term

$$f(\mathcal{M}^{t+1}, \mathbf{A}^{t+1}, \mathcal{S}^{t+1}) + \frac{\rho}{2}(\|\mathcal{M}^{t+1} - \mathcal{M}^t\|_F^2$$
$$+ \|\mathbf{A}^{t+1} - \mathbf{A}^t\|_F^2 + \|\mathcal{S}^{t+1} - \mathcal{S}^t\|_F^2) \leq f(\mathcal{M}^t, \mathbf{A}^t, \mathcal{S}^t). \tag{37}$$

Following (37), it is easy to obtain

$$f(\mathcal{M}^{t+1}, \mathbf{A}^{t+1}, \mathcal{S}^{t+1}) \leq f(\mathcal{M}^0, \mathbf{A}^0, \mathcal{S}^0) < \infty. \tag{38}$$

Then we have that $f_1(\mathcal{M})$, $f_2(\mathbf{A})$, and $f_3(\mathcal{S})$ are bounded. Considering that λ, β, and γ are constants greater than zero, we have the sequence $\{\mathcal{M}^t, \mathbf{A}^t, \mathcal{S}^t\}$ is bounded. Then condition (a) holds.

Note that f_0 is a polynomial function, and hence a semi-algebraic function. It is well-known that the l_2-norm is a semi-algebraic function, and the composition of semi-algebraic functions is also a semi-algebraic function. Then, f_2 and f_3 are semi-algebraic functions. Also, f_1 is a semi-algebraic function as the nuclear norm is a semi-algebraic function [52, Proposition 4]. Finally, we know that f is a semi-algebraic function. Condition (b) holds.

It is obvious that conditions (c) and (d) are satisfied.

Taking the derivative of f_0 with respect to $\mathcal{M}$ leads to

$$\nabla_{\mathcal{M}} f_0 = -2(\mathcal{Y} - \mathcal{S}) \times_3 \mathbf{A}^T + 2\mathcal{M} \times_3 (\mathbf{A}^T\mathbf{A}). \quad (39)$$

Letting $L_m = 2\|\mathbf{A}\|_F^2$ and considering $\mathbf{A}$ is bounded, we obtain that $\nabla_{\mathcal{M}} f_0$ is L_m-Lipschitz continuous. Similarly, it is easy to check $\nabla_{\mathbf{A}} f_0$ and $\nabla_{\mathcal{S}} f_0$ are Lipschitz continuous with Lipschitz constants $L_a = 2\|\mathcal{M}\|_F^2$ and $L_s = 2$, respectively. Then condition (e) holds.

Note that f_0 is a polynomial function, and hence has a continuous first-order derivative. We know that ∇f_0 is Lipschitz continuous on any bounded set naturally holds according to condition (e). Then condition (f) is satisfied. According to [53, Th. 3.7], the sequence $\{\mathcal{M}^t, \mathbf{A}^t, \mathcal{S}^t\}$ generated by Algorithm 1 converges to a critical point of (13). The proof is completed. ∎

IV. EXPERIMENTAL RESULTS

The experimental results conducted on four real datasets are presented in this section. All the experiments are implemented in MATLAB 2019a on a personal computer with an Intel Core i7-4790k 4.00-GHz central processing unit, 16-GB memory, and 64-bit Windows 10.

A. Datasets' Description

Four real datasets are used to evaluate the performance of the proposed anomaly detector. The first dataset was captured by the Nuance Cri hyperspectral sensor [54]. It contains 400×400 pixels and 46 spectral channels. Several stones in the grass are considered anomalies. The second dataset is the Cuprite dataset, which was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor [4]. A subarea with spatial size 250×191 is exploited in our experiment, and each pixel in this image contains 188 bands. Different types of minerals in the scene are anomalies. The third dataset, the Los Angeles dataset, was collected by the AVIRIS sensor [17]. The image contains 100×100 pixels with 205 spectral channels. The fourth dataset consisting of 180×180 pixels and 120 bands was captured by the SpecTIR hyperspectral airborne Rochester experiment [17]. The false-color images and their corresponding ground-truth maps are shown in Fig. 3.

B. Evaluation Metrics

Different types of widely used metrics are used in our experiment to obtain comprehensive performance evaluations of the proposed ALRTT. The first metric is the detection map, which intuitively presents the detection results of different detectors. The varying colors in the detection map mean different detection values of pixels. The second metric is the receiver operating characteristic (ROC) curve. In this article, we use two types of ROC curves, i.e., a curve that is plotted by the false alarm rate (FAR) and probability of detection (PD) and a curve that is plotted by the detection threshold and the FAR. The ROC curve associated with the FAR and PD will be close to the upper left corner of the coordinate axis if a detector has high detection probabilities. Meanwhile, the ROC curve associated with the threshold and FAR will be close to the lower left corner of the coordinate axis if a detector is good at suppressing the background. The third metric is the area under the ROC curve (AUC). Denote the PD, FAR, and threshold by P_D, P_F, and τ, respectively, then we can obtain three different AUC values, i.e., $\text{AUC}_{(P_D, P_F)}$, $\text{AUC}_{(P_D, \tau)}$, and $\text{AUC}_{(P_F, \tau)}$ [55]. And larger $\text{AUC}_{(P_D, P_F)}$ and $\text{AUC}_{(P_D, \tau)}$ indicate better performance of a detector, while a lower $\text{AUC}_{(P_F, \tau)}$ means a better ability in suppressing the background. Besides, we combine these three AUC values to obtain another two measurements, i.e., AUC_{OD} and AUC_{SNR}, which are calculated by

$$\text{AUC}_{\text{OD}} = \text{AUC}_{(P_D, P_F)} + \text{AUC}_{(P_D, \tau)} - \text{AUC}_{(P_F, \tau)} \quad (40)$$

and

$$\text{AUC}_{\text{SNR}} = \frac{\text{AUC}_{(P_D, \tau)}}{\text{AUC}_{(P_F, \tau)}} \quad (41)$$

respectively. AUC_{OD} and AUC_{SNR} are two effective metrics for evaluating the overall performance of a detector [55]. Higher values of $\text{AUC}_{(P_D, P_F)}$ and $\text{AUC}_{(P_D, \tau)}$ or a lower value of $\text{AUC}_{(P_F, \tau)}$ lead to a higher value of AUC_{OD}. AUC_{SNR} is similar to the signal-to-noise ratio (SNR). Higher values of $\text{AUC}_{(P_D, P_F)}$ and $\text{AUC}_{(P_D, \tau)}$ indicate better performance of a detector. The fourth metric is the separability map. The detection result is first normalized to the range of [0, 1], and then the distributions of detection values of the background and anomaly are illustrated with the aid of boxes with different colors and dotted lines. Specifically, the detection values of the background or anomaly distributed in the range of [10%, 90%] are presented by the box, and the detection values distributed in the ranges of [0%, 10%] and [90%, 100%] are represented by the dotted line at the bottom and the dotted line at the top, respectively. The gap between the background box and the anomaly box implies the ability of a detector to separate the background and the anomaly. Finally, the running time of all the detectors is given in Table II.

C. Experimental Setup

1) Initialization of the ALRTT: Tensor $\mathcal{S}^0$ is initialized by a zero tensor. For a input tensor $\mathcal{Y}$, the SVD of $\mathbf{Y}_{(3)}$ is denoted by $\mathbf{Y}_{(3)} = \mathbf{U}\Sigma\mathbf{V}^T$, and then the tensor $\mathcal{M}^0$ and matrix $\mathbf{A}^0$ are initialized by $\mathcal{M}^0 = \text{fold}_3(\Sigma_{1:d,1:d}\mathbf{V}_{:,1:d}^T)$ and $\mathbf{A}^0 = \mathbf{U}_{:,1:d}$, respectively. The number $t_{\max}$ of iteration is related to the efficiency of the ALRTT. To select a proper $t_{\max}$, we plot the values of the objective function (13) and $\text{AUC}_{(P_D, P_F)}$ with an increasing number of iterations in Fig. 2. It is noted that the values of the objective function are normalized by the corresponding maximum for a visual purpose. As we

TABLE I
AUC SCORES ON THE FOUR DATASETS

| Datasets	AUC Values	ALRTT	LARTVAD	sparseHAD	RGAE	PTA	GTVLRR	TDAD	RPCA-RX	GRX
Cri	$AUC_{(P_D, P_F)}$	**0.9809**	0.9640	0.9672	0.9197	0.9350	0.5262	0.8527	0.9155	0.9134
	$AUC_{(P_D, \tau)}$	0.3827	0.3840	0.1517	0.1783	0.3727	0.2031	0.0489	0.0916	0.0683
	$AUC_{(P_F, \tau)}$	0.0456	0.0756	0.0215	0.0143	0.0845	0.1816	0.0244	0.0197	0.0235
	AUC_{OD}	**1.3180**	1.2724	1.0975	1.0818	1.2232	0.5477	0.8772	0.9874	0.9582
	AUC_{SNR}	8.3983	5.0801	7.0642	**12.4714**	4.4121	1.1183	2.0027	4.6524	2.9053
Cuprite	$AUC_{(P_D, P_F)}$	**0.9991**	0.9982	0.9912	0.8350	0.9840	0.9494	0.9974	0.9968	0.9974
	$AUC_{(P_D, \tau)}$	0.6208	0.5861	0.1547	0.0368	0.3325	0.2407	0.3340	0.1623	0.3346
	$AUC_{(P_F, \tau)}$	0.0037	0.1130	0.0200	0.0185	0.0136	0.0872	0.0757	0.0227	0.0677
	AUC_{OD}	**1.6162**	1.4712	1.1258	0.8533	1.3028	1.1028	1.2558	1.1364	1.2643
	AUC_{SNR}	168.0393	5.1844	7.7247	1.9880	24.3649	2.7588	4.4135	7.1441	4.9413
Los Angeles	$AUC_{(P_D, P_F)}$	**0.9970**	0.9947	0.9852	0.9948	0.9966	0.9419	0.9876	0.9865	0.9887
	$AUC_{(P_D, \tau)}$	0.1609	0.1480	0.0342	0.0395	0.1767	0.0848	0.1033	0.1039	0.0891
	$AUC_{(P_F, \tau)}$	0.0106	0.0148	0.0017	6.0587e-4	0.0253	0.0154	0.0219	0.0117	0.0114
	AUC_{OD}	1.1473	1.1280	1.0178	1.0337	**1.1481**	1.0113	1.0690	1.0786	1.0664
	AUC_{SNR}	15.1942	10.0274	20.5257	**65.1766**	6.9969	5.4931	4.7091	8.8514	7.8008
SpecTIR	$AUC_{(P_D, P_F)}$	**0.9998**	0.9998	0.9986	0.8777	0.9591	0.9875	0.9892	0.9974	0.9914
	$AUC_{(P_D, \tau)}$	0.4208	0.4425	0.3479	0.0889	0.2735	0.3544	0.3869	0.2644	0.3683
	$AUC_{(P_F, \tau)}$	0.0003	0.0179	0.0070	0.0176	0.0565	0.0361	0.0293	0.0180	0.0249
	AUC_{OD}	1.4202	**1.4244**	1.3395	0.9490	1.1761	1.3057	1.3468	1.2438	1.3348
	AUC_{SNR}	**1345.0816**	24.7143	49.6870	5.0393	4.8414	9.8166	13.2224	14.6518	14.8183
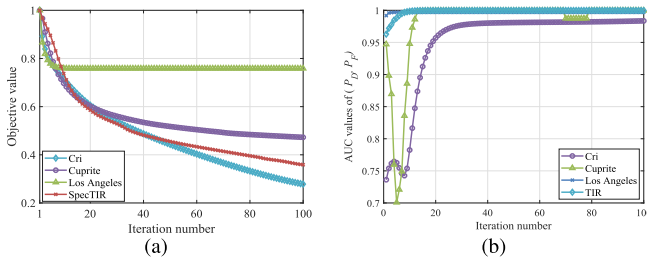
Fig. 2. (a) Values of the objective function with increasing number of iterations. (b) Values of $AUC_{(P_D, P_F)}$ with increasing number of iterations.

can see, the values of the objective function decrease during the iteration, and the values of $AUC_{(P_D, P_F)}$ increase after a certain number of iterations. A slight fluctuation appears on the Cuprite dataset, but it can eventually return to the maximum. As a result, considering the effectiveness and efficiency, t_{max} is set to 50 in our experiments.

2) Methods in Comparison: Several different types of hyperspectral anomaly detectors are used as comparisons to obtain an explicit evaluation of the proposed method, including GRX [7], RPCA-RX [12], tensor-decomposition-based anomaly detector (TDAD) [37], graph and TV regularized LRR (GTVLRR) [20], prior-based tensor approximation (PTA) [13], sparse coding-inspired generative adversarial network for hyperspectral anomaly detection (sparseHAD) [56], robust graph auto-encoder (RGAE) detector [57], and LART-VAD [44]. The parameters in these methods in our experiments are set as the recommended parameters or tuned to optimal within given sets according to the corresponding references. For the RPCA-RX, the parameter λ is set to $1/\min(w, h)$. For the GTVLRR, we set $\lambda = 0.5$, $\beta = 0.2$, and $\gamma = 0.05$ according to the suggestion [20]. Besides, two parameters K and P associated with the background dictionary are set to 15 and 20, respectively. For the PTA, we set $\alpha = 1$, $\tau = 1$, and $\beta = 0.01$. The parameters μ and truncated

rank r are tuned according to reference [13]. The parameters in LARTVAD are searched from given ranges in [44]. The sparseHAD and RGAE are two deep-learning-based detectors, and the parameters are also tuned to optimal according to [56] and [57].

D. Performance

The experimental results on the four datasets are shown in Figs. 3–6 and Table I.

1) Cri Dataset: Detection maps of the proposed ALRTT and other methods on the Cri dataset are presented in the first row of Fig. 3. As can be seen, GRX, RPCA-RX, TDAD, RGAE, and sparseHAD filter the background well, but the anomaly is also suppressed. The GTVLRR method fails to separate the background and anomaly. The PTA, LARTVAD, and ALRTT can clearly distinguish anomalies from the background. In contrast, the ALRTT performs better in detecting anomalies and suppressing the background, which can be inferred from the ROC curves of (P_D, P_F) and (P_F, τ) shown in Figs. 4(a) and 5(a), respectively. The ROC curve of (P_D, P_F) of the ALRTT is above that of other detectors, while the ROC curve of (P_F, τ) of the ALRTT is below that of the PTA and LARTVAD. Different AUC values of all the methods are presented in Table I. The value of $AUC_{(P_D, P_F)}$ of the ALRTT is 0.9809, which is much higher than those of other methods. The AUC score of AUC_{OD} of the ALRTT is also the highest among all the detectors. The RGAE achieves a highest value of AUC_{SNR} due to a low value of (P_F, τ). However, the response to the anomaly of the RGAE is very low. The separability map on the Cri dataset is shown in Fig. 6(a), where the methods for comparison fail to separate the background boxes and anomaly boxes, while the proposed ALRTT achieves a better separation.

2) Cuprite Dataset: For the Cuprite dataset, the GTVLRR and RGAE methods fail to detect the anomaly. The GRX, RPCA-RX, TDAD, PTA, and sparseHAD obtain very low
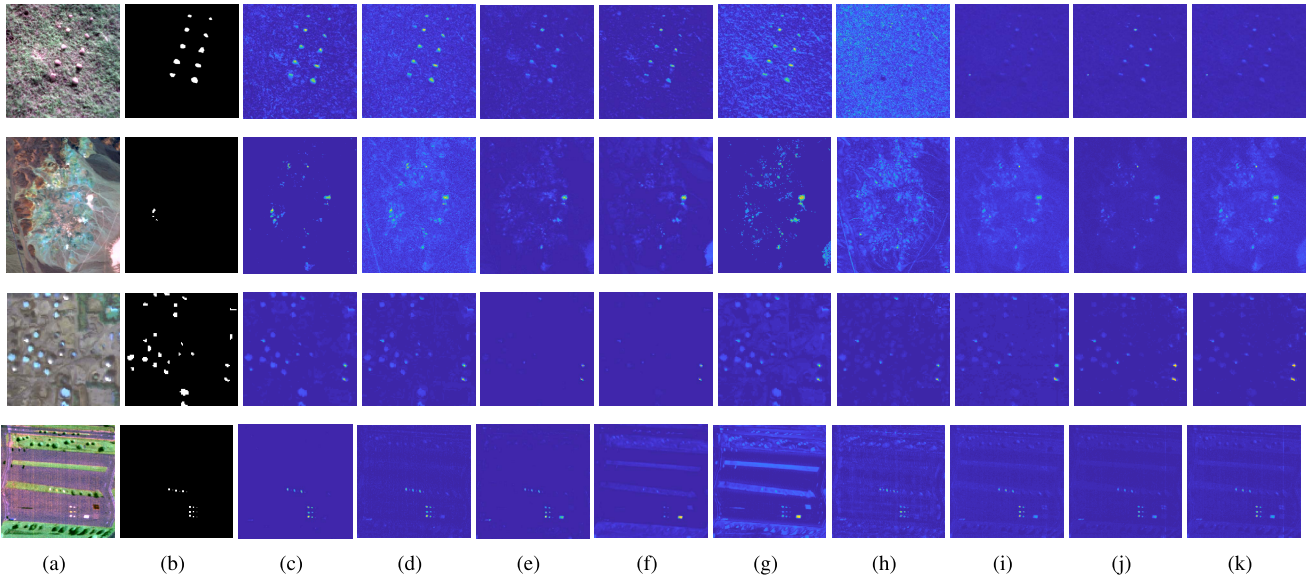
Fig. 3. Detection maps of different methods on the four test datasets. The four datasets from top to bottom are Cri, Cuprite, Los Angeles, and SpecTIR datasets. (a) False-color images. (b) Ground truth. (c) ALRTT. (d) LARTVAD. (e) sparseHAD. (f) RGAE. (g) PTA. (h) GTVLRR. (i) TDAD. (j) RPCA-RX. (k) GRX.
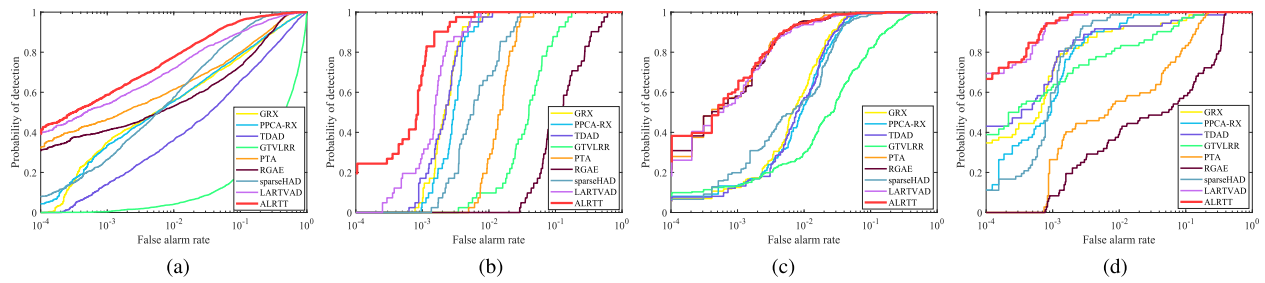

Fig. 4. ROC curves of (P_D, P_F) with different spatial information on the four test datasets. (a) Cri. (b) Cuprite. (c) Los Angeles. (d) SpecTIR.
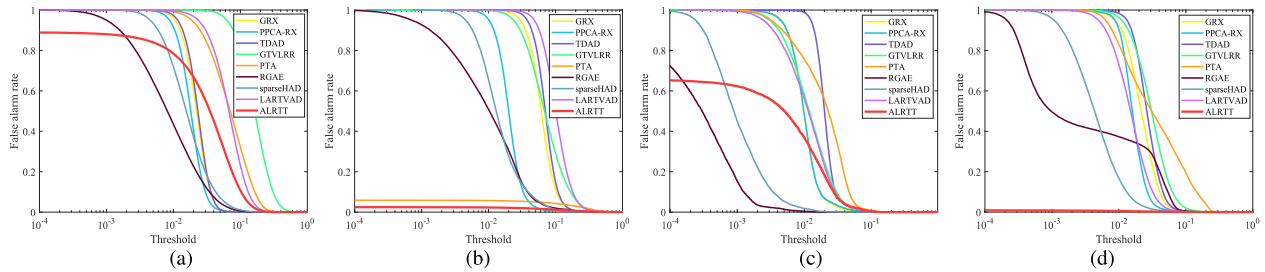

Fig. 5. ROC curves of (P_F, τ) with different spatial information on the four test datasets. (a) Cri. (b) Cuprite. (c) Los Angeles. (d) SpecTIR.

responses to the anomaly. The LARTVAD and proposed ALRTT are able to locate the anomaly. The detection results of LARTVAD are seriously contaminated by the background. For the ALRTT, although some false alarms appear in the detection map, the anomaly is well-detected and most of the background pixels are suppressed successfully. The AUC curves of (P_D, P_F) and (P_F, τ) are shown in Figs. 4(b) and 5(b), respectively. It is obvious from Fig. 4(b) that ALRTT achieves the highest PD in each FAR among all the methods, and it can detect all the anomalies in a very low FAR. Besides, the AUC curves of (P_F, τ) of ALRTT and PTA are always below than others. As can be seen in Table I, ALRTT has the best AUC

values. Its value of $\text{AUC}_{(P_D, \tau)}$ is 0.0037, which is much smaller than that of other detectors. The ALRTT achieves the highest AUC_{SNR}, which means that the overall performance of ALRTT on the Cuprite dataset is the best. The separability map on the Cuprite dataset is shown in Fig. 6(b). As can be seen, both PTA and ALRTT suppress the background in very small ranges with small values, and ALRTT obtains a larger gap between the background box and the anomaly, which indicates that it has a better ability to separate the background and anomaly.

3) Los Angeles Dataset: The detection maps on the Los Angeles dataset are displayed in the third row of Fig. 3. The RGAE and sparseHAD perform well in suppressing the
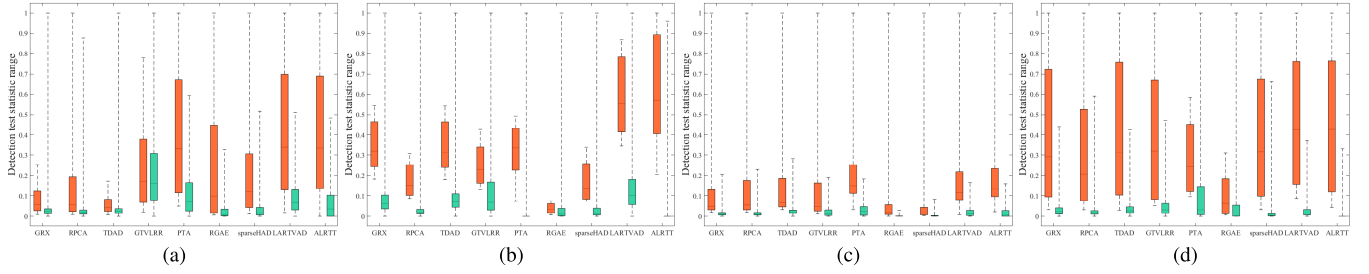
Fig. 6. Anomaly background separability maps of the four test datasets. (a) Cri. (b) Cuprite. (c) Los Angeles. (d) SpecTIR.

TABLE II

EXECUTION TIME OF DIFFERENT ALGORITHMS ON THE FOUR DATASETS (UNIT: S)

Methods	ALRTT	LARTVAD	RGAE	PTA	GTVLRR	TDAD	RPCA-RX	GRX
Cri	43.3163	904.8516	1214.5412	432.8005	5033.3527	5.9211	24.0834	0.8992
Cuprite	136.0729	448.9272	913.9440	517.9346	1350.5545	5.6178	29.0870	0.9280
Los Angeles	10.7522	55.0890	158.0701	66.3982	145.1065	1.3058	9.1942	0.1383
SpecTIR	39.2139	261.7866	438.9951	151.2777	611.1332	2.5234	9.7216	0.3069
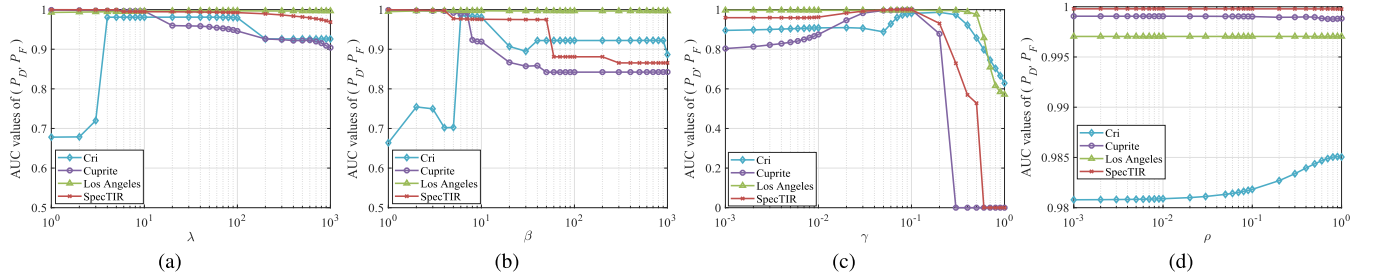
Fig. 7. Values of $\text{AUC}_{(P_D, P_F)}$ with varying parameters. (a) λ. (b) β. (c) γ. (d) ρ.

background and achieve values of $\text{AUC}_{(P_F, \tau)}$ of 6.0587×10^{-4} and 0.0017, respectively. However, most anomalous pixels are also filtered. The GRX, RPCA-RX, TDAD, and GTVLRR are able to locate all the anomalies but with low contrasts between the background and the anomaly. The responses of anomalies are similar in the detection maps of the PTA, LARTVAD, and ALRTT, but the PTA is less effective in eliminating the background. As shown in Fig. 4(c), the ROC curves of (P_D, P_F) of the PTA, LARTVAD, and ALRTT are close and much higher than that of others. The AUC values are consistent with the test result. The value of $\text{AUC}_{(P_D, P_F)}$ of the proposed ALRTT is 0.9970, which is higher than those of other methods. The RGAE achieves the highest AUC_{SNR} due to the extremely low $\text{AUC}_{(P_F, \tau)}$, but its AUC_{OD} is very low because of poor performance in locating the anomaly. The value of AUC_{OD} of the PTA is slightly higher than that of the ALRTT, but its value of AUC_{SNR} is much lower than that of the ALRTT, because the ALRTT filters the background well. The separability map on the Los Angeles dataset is presented in Fig. 6(c). The gaps between the background boxes and the anomaly boxes of PTA and ALRTT are about equal, where the normalized responses of anomalies of PTA are slightly higher, while the normalized responses of the background of ALRTT are slightly lower.

4) SpecTIR Dataset: For the SpecTIR dataset, as shown in the fourth row of Fig. 3, the GRX, RPCA-RX, TDAD, and

RGAE cannot locate anomalies of small size. The detection map of the PTA is seriously contaminated by the background. The GTVLRR, sparseHAD, LARTVAD, and ALRTT succeed to distinguish the anomaly, of which ALRTT performs best in suppressing the background. In Fig. 4(d), LARTVAD and ALRTT have similar ROC curves of (P_D, P_F) and achieve higher PDs. The ALRTT first achieves a PD of 100% with a very low FAR of 0.1980%. Besides, as shown in Fig. 5(d), the FARs of the ALRTT are always close to 0, which indicates that the ALRTT performs well in suppressing the background. For the AUC values, the LARTVAD achieves the same value of $\text{AUC}_{(P_D, P_F)}$ of 0.9998 as the proposed ALRTT, and its value of AUC_{OD} of 1.4244 is slightly higher than that of 1.4202 of ALRTT. The value of AUC_{SNR} of ALRTT is 0.0003, which is much larger than that of others. The separability map on the SpecTIR dataset is shown in Fig. 6(d). The LARTVAD and ALRTT obtain satisfying separations. The background box of ALRTT is compressed to a line in the separability map.

Table II lists the computing time of different methods on the four test HSIs. As can be seen, the GRX detector has the shortest running time. The proposed ALRTT is more time-consuming than the TDAD and RPCA-RX methods. Compared with LARTVAD, PTA, and GTVLRR, the ALRTT method is more efficient. The transform is adopted in both LARTVAD and the proposed ALRTT; however, the proposed
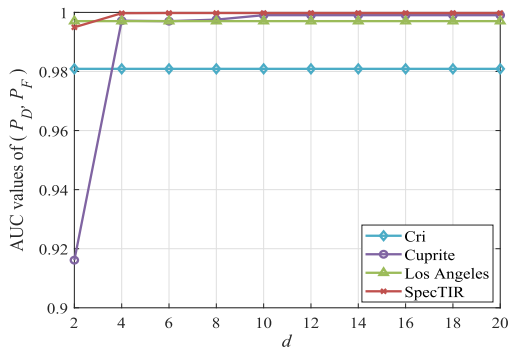
Fig. 8. AUC values of (P_D, P_F) with different d.

TABLE III
PARAMETER SETTINGS ON THE FOUR DATASETS

Datasets	Cri	Cuprite	Los Angeles	SpecTIR
λ	10	1	100	1
β	10	1	10	1
γ	0.1	0.1	0.1	0.1
ρ	0.01	0.01	0.01	0.01
d	$\lfloor 0.1b \rfloor = 4$	$\lfloor 0.1b \rfloor = 18$	$\lfloor 0.1b \rfloor = 20$	$\lfloor 0.1b \rfloor = 12$

ALRTT is more efficient due to direct utilization of the spectral correlation of the background.

E. Parameter Tuning

The proposed ALRTT includes five parameters, i.e., three regularized parameters λ, β, and γ, a proximal parameter ρ, and the initial number of columns d of matrix $\mathbf{A}$. To analyze the influence of parameters on ALRTT, the values of $\text{AUC}_{(P_D, P_F)}$ with varying parameters are illustrated in Fig. 7, where one parameter is changed with other parameters fixed in each subfigure. We change λ and β from 1 to 10 with an interval of 1, from 10 to 100 with an interval of 10, and from 100 to 1000 with an interval of 100. The parameters γ and ρ are changed from 0.001 to 0.01 with an interval of 0.001, from 0.01 to 0.1 with an interval of 0.01, and from 0.1 to 1 with an interval of 0.1.

The parameter λ is the weight of the regularization on matrix $\mathbf{A}$. As can be seen in Fig. 7(a), the values of $\text{AUC}_{(P_D, P_F)}$ of the Los Angeles and SpecTIR datasets are nonsensitive. The values of $\text{AUC}_{(P_D, P_F)}$ of the Cuprite dataset decrease when $\lambda > 1$. A stable result is obtained on the Cri dataset when λ varies from 4 to 100. For the parameter β shown in Fig. 7(b), the results of three datasets except the Los Angeles dataset are sensitive to β. In Fig. 7(c), the ALRTT first achieves better performance with the increase in γ, and then cannot work when γ is too large. Satisfactory performances of the four datasets are available when $\gamma = 0.1$. The result of the proximal parameter ρ is shown in Fig. 7(d). It is obvious that the performance of ALRTT is nonsensitive to ρ. The parameter d is searched from 2 to 20 with an interval of 2. As can be seen in Fig. 8, the values of $\text{AUC}_{(P_D, P_F)}$ of the Cuprite and Los Angeles datasets increase first with the increase in d and achieve maxima at $d = 10$ and $d = 6$, respectively. The results

of the Cri and SpecTIR datasets are stable under different values of d. It is obvious that the performance of ALRTT does not depend on the parameter d if d is large enough. It is worth noting that the computational complexity of ALRTT is related to d. A default setting of d can be selected as the largest integer not exceeding $0.1b$, which is denoted by $d = \lfloor 0.1b \rfloor$. For clarity, the parameter settings are listed in Table III.
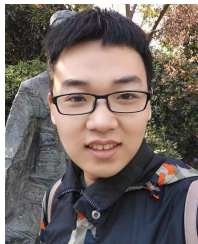
V. CONCLUSION

In this article, the ALRTT method based on tensor transform was proposed for hyperspectral anomaly detection. In the proposed ALRTT, the HSI background is represented as the product of a transformed tensor and a matrix. Each frontal slice of the transformed tensor is considered to integrate the spatial and spectral information of the background, which allows for taking full advantage of the spatial–spectral correlation by minimizing the nuclear norm of the frontal slices. Considering the strong spectral correlation of the HSI background, we exploited a low-rank matrix instead of an orthogonal matrix in the tensor transform to directly formulate the low-rank property, which also greatly improves the computational efficiency of the algorithm due to fewer frontal slices. Besides, we imposed the $l_{2,1}$-norm constraint to an initial matrix with predefined size to adaptively derive an appropriate low-rank matrix. The group sparsity of the anomaly was depicted by the $l_{2,1,1}$-norm imposing on the anomaly tensor. All the regularization terms and the fidelity term were integrated into a non-convex problem, which was solved by the PAM algorithm. The computational complexity and convergence analysis were given in this article. Finally, experimental results have demonstrated the superiority of ALRTT over its counterparts. As for future work, more spatial information can be integrated into the proposed model, such as the piecewise smoothness and nonlocal similarity.

REFERENCES

[1] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

[2] S. M. Schweizer and J. M. F. Moura, "Hyperspectral imagery: Clutter adaptation in anomaly detection," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1855–1871, Aug. 2000.

[3] Y. Dong, W. Shi, B. Du, X. Hu, and L. Zhang, "Asymmetric weighted logistic metric learning for hyperspectral target detection," *IEEE Trans. Cybern.*, vol. 52, no. 10, pp. 11093–11106, Oct. 2022.

[4] W. Xie, X. Zhang, Y. Li, K. Wang, and Q. Du, "Background learning based on target suppression constraint for hyperspectral target detection," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5887–5897, 2020.

[5] J. Liu, Z. Hou, W. Li, R. Tao, D. Orlando, and H. Li, "Multipixel anomaly detection with unknown patterns for hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 10, pp. 5557–5567, Oct. 2022.

[6] T. Jiang, W. Xie, Y. Li, J. Lei, and Q. Du, "Weakly supervised discriminative learning with spectral constrained generative adversarial network for hyperspectral anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6504–6517, Nov. 2022.

[7] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.

[8] J. M. Molero, E. M. Garzón, I. García, and A. Plaza, "Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 801–814, Apr. 2013.

[9] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.

[10] J. Zhou, C. Kwan, B. Ayhan, and M. T. Eismann, "A novel cluster kernel RX algorithm for anomaly and change detection using hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 11, pp. 6497–6504, Nov. 2016.

[11] E. J. Candes, X. L. Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.

[12] L. Wei and D. Qian, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.

[13] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, "Prior-based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1037–1050, Mar. 2022.

[14] T. Zhou and D. Tao, "GoDec: Randomized low-rank & sparse matrix decomposition in noisy case," in *Proc. Int. Conf. Mach. Learn.*, Bellevue, WA, USA, 2011, pp. 1–16.

[15] Y. Zhang, B. Du, L. Zhang, and S. Wang, "A low-rank and sparse matrix decomposition-based Mahalanobis distance method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1376–1389, Mar. 2016.

[16] C.-I. Chang, H. Cao, S. Chen, X. Shang, C. Yu, and M. Song, "Orthogonal subspace projection-based go-decomposition approach to finding low-rank and sparsity matrices for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2403–2429, Mar. 2021.

[17] L. Li, W. Li, Q. Du, and R. Tao, "Low-rank and sparse decomposition with mixture of Gaussian for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 51, no. 9, pp. 4363–4372, Sep. 2020.

[18] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[19] Y. Xu, Z. Wu, J. Li, A. Plaza, and Z. Wei, "Anomaly detection in hyperspectral images based on low-rank and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 4, pp. 1990–2000, Apr. 2015.

[20] T. Cheng and B. Wang, "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 391–406, Jan. 2020.

[21] R. Feng, H. Li, L. Wang, Y. Zhong, L. Zhang, and T. Zeng, "Local spatial constraint and total variation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5512216.

[22] L. Zhuang, L. Gao, B. Zhang, X. Fu, and J. M. Bioucas-Dias, "Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, Dec. 2021.

[23] Y. Qu et al., "Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4391–4405, Aug. 2018.

[24] W. Xie, X. Zhang, Y. Li, J. Lei, J. Li, and Q. Du, "Weakly supervised low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Cybern.*, vol. 51, no. 8, pp. 3889–3900, Aug. 2021.

[25] X. Zhang, X. Ma, N. Huyan, J. Gu, X. Tang, and L. Jiao, "Spectral-difference low-rank representation learning for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10364–10377, Dec. 2021.

[26] N. Huyan, X. Zhang, H. Zhou, and L. Jiao, "Hyperspectral anomaly detection via background and potential anomaly dictionaries construction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2263–2276, Apr. 2019.

[27] J. Xue, Y. Zhao, W. Liao, and J. C.-W. Chan, "Nonlocal low-rank regularized tensor decomposition for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 5174–5189, Jul. 2019.

[28] H. Zeng, X. Xie, H. Cui, H. Yin, and J. Ning, "Hyperspectral image restoration via global L_{1-2} spatial–spectral total variation regularized local low-rank tensor recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 4, pp. 3309–3325, Apr. 2021.

[29] S. Li, R. Dian, L. Fang, and J. M. Bioucas-Dias, "Fusing hyperspectral and multispectral images via coupled sparse tensor factorization," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4118–4130, Aug. 2018.

[30] R. Dian, S. Li, and L. Fang, "Learning a low tensor-train rank representation for hyperspectral image super-resolution," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 9, pp. 2672–2683, Sep. 2019.

[31] R. Dian, S. Li, L. Fang, T. Lu, and J. M. Bioucas-Dias, "Nonlocal sparse tensor factorization for semiblind hyperspectral and multispectral image fusion," *IEEE Trans. Cybern.*, vol. 50, no. 10, pp. 4469–4480, Oct. 2020.

[32] L. R. Tucker, "Implications of factor analysis of three-way matrices for measurement of change," in *Problems in Measuring Change*. Madison, WI, USA: Univ. of Wisconsin Press, 1963, pp. 122–137.

[33] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multimodal factor analysis," in *Proc. UCLA Work. Papers Phonetics*, 1970, pp. 1–84.

[34] J. Xue, Y. Zhao, S. Huang, W. Liao, J. C.-W. Chan, and S. G. Kong, "Multilayer sparsity-based tensor decomposition for low-rank tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6916–6930, Nov. 2022.

[35] J. Xue, Y. Zhao, Y. Bu, J. C.-W. Chan, and S. G. Kong, "When Laplacian scale mixture meets three-layer transform: A parametric tensor sparsity for tensor completion," *IEEE Trans. Cybern.*, vol. 52, no. 12, pp. 13887–13901, Dec. 2022.

[36] S. Li, W. Wang, H. Qi, B. Ayhan, C. Kwan, and S. Vance, "Low-rank tensor decomposition based anomaly detection for hyperspectral imagery," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 4525–4529.

[37] X. Zhang, G. Wen, and W. Dai, "A tensor decomposition-based anomaly detection algorithm for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5801–5820, Oct. 2016.

[38] J. Dai, C. Deng, W. Wang, and X. Liu, "Low-rank and sparse tensor recovery for hyperspectral anomaly detection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 1141–1144.

[39] J. Wang, Y. Xia, and Y. Zhang, "Anomaly detection of hyperspectral image via tensor completion," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 6, pp. 1099–1103, Jun. 2021.

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

[41] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, Feb. 2013.

[42] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-SVD," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3842–3849.

[43] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Joint reconstruction and anomaly detection from compressive hyperspectral images using Mahalanobis distance-regularized tensor RPCA," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2919–2930, May 2018.

[44] S. Sun, J. Liu, X. Chen, W. Li, and H. Li, "Hyperspectral anomaly detection with tensor average rank and piecewise smoothness constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Mar. 4, 2022. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9728731, doi: 10.1109/TNNLS.2022.3152252.

[45] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 5249–5257.

[46] E. Kernfeld, M. Kilmer, and S. Aeron, "Tensor–tensor products with invertible linear transforms," *Linear Algebra Appl.*, vol. 485, pp. 545–570, Nov. 2015.

[47] T.-X. Jiang, M. K. Ng, X.-L. Zhao, and T.-Z. Huang, "Framelet representation of tensor nuclear norm for third-order tensor completion," *IEEE Trans. Image Process.*, vol. 29, pp. 7233–7244, 2020.

[48] B.-Z. Li, X.-L. Zhao, J.-L. Wang, Y. Chen, T.-X. Jiang, and J. Liu, "Tensor completion via collaborative sparse and low-rank transforms," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 1289–1303, 2021.

[49] T.-X. Jiang, X.-L. Zhao, H. Zhang, and M. K. Ng, "Dictionary learning with low-rank coding coefficients for tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Aug. 31, 2021, doi: 10.1109/TNNLS.2021.3104837.

[50] C. Lu, "Transforms based tensor robust PCA: Corrupted low-rank tensors recovery via convex optimization," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 1145–1152.

[51] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.

[52] C. Xu, Z. Lin, and H. Zha, "A unified convex surrogate for the Schatten-p norm," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 926–932.

[53] C. Bao, H. Ji, Y. Quan, and Z. Shen, "Dictionary learning for sparse coding: Algorithms and convergence analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 7, pp. 1356–1369, Jul. 2016.

[54] S. Chang, B. Du, and L. Zhang, "A subspace selection-based discriminative forest method for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4033–4046, Jun. 2020.

[55] C.-I. Chang, "An effective evaluation tool for hyperspectral target detection: 3D receiver operating characteristic curve analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5131–5153, Jun. 2020.

[56] Y. Li, T. Jiang, W. Xie, J. Lei, and Q. Du, "Sparse coding-inspired GAN for hyperspectral anomaly detection in weakly supervised learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5512811.

[57] G. Fan, Y. Ma, X. Mei, F. Fan, J. Huang, and J. Ma, "Hyperspectral anomaly detection with robust graph autoencoders," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5511314.

Ziwei Zhang was born in 1999. She received the B.S. degree in electronic and information engineering from Xidian University, Xi'an, China, in 2021.

She is currently a Graduate Student in information and communication engineering with the University of Science and Technology of China, Hefei, China. Her research interests include radar object detection and hyperspectral image processing.

Siyu Sun was born in 1996. He received the B.S. degree in electronic and information engineering from the Hefei University of Technology, Hefei, China, in 2018, where he is currently pursuing the Ph.D. degree in information and communication engineering.

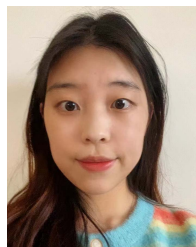His research interests include radar target detection and hyperspectral image processing.

Jun Liu (Senior Member, IEEE) received the B.S. degree in mathematics from the Wuhan University of Technology, Wuhan, China, in 2006, the M.S. degree in mathematics from the Chinese Academy of Sciences, Beijing, China, in 2009, and the Ph.D. degree in electrical engineering from Xidian University, Xi'an, China, in 2012.

From July 2012 to December 2012, he was a Post-Doctoral Research Associate with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. From January 2013 to September 2014, he was a Post-Doctoral Research Associate with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ, USA. He is currently an Associate Professor with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, China. He is the coauthor of two books *Advances in Adaptive Radar Detection and Range Estimation* (Springer) in 2022 and *Adaptive Detection of Multichannel Signals Exploiting Persymmetry* (CRC Press) in 2023. His research interests include statistical signal processing, image processing, and machine learning.

Dr. Liu is a member of the Sensor Array and Multichannel (SAM) Technical Committee and the IEEE Signal Processing Society. He was a recipient of the Best Paper Award from the IEEE WCSP 2021. He is currently an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS and a Member of the Editorial Board of the *Signal Processing* (Elsevier).

Wei Li (Senior Member, IEEE) received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2007, the M.S. degree in information science and technology from Sun Yat-sen University, Guangzhou, China, in 2009, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in 2012.

Subsequently, he spent one year as a Post-Doctoral Researcher at the University of California, Davis, CA, USA. He is currently a Professor with the School of Information and Electronics, Beijing Institute of Technology, Beijing, China. His research interests include hyperspectral image analysis, pattern recognition, and target detection.

Dr. Li received IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING Best Reviewer in 2016 and IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Best Reviewer Award in 2020 from the IEEE Geoscience and Remote Sensing Society (GRSS), and the Outstanding Paper Award at the IEEE International Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (Whispers), in 2019. He has served as an Associate Editor for the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING and IEEE SIGNAL PROCESSING LETTERS. He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.