Spatial Invariant Tensor Self-Representation Model for Hyperspectral Anomaly Detection

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

Abstract—With the development of hyperspectral imaging technology, the hyperspectral anomaly has attracted considerable attention due to its significant role in many applications. Hyperspectral images (HSIs) with two spatial dimensions and one spectral dimension are intrinsically three-order tensors. However, most of the existing anomaly detectors were designed after converting the 3-D HSI data into a matrix, which destroys the multidimension structure. To solve this problem, in this article, we propose a spatial invariant tensor self-representation (SITSR) hyperspectral anomaly detection algorithm, which is derived based on the tensor–tensor product (t-product) to preserve the multidimension structure and achieve a comprehensive description of the global correlation of HSIs. Specifically, we exploit the t-product to integrate spectral information and spatial information, and the background image of each band is represented as the sum of the t-product of all bands and their corresponding coefficients. Considering the directionality of the t-product, we utilize two tensor self-representation methods with different spatial modes to obtain a more balanced and informative model. To depict the global correlation of the background, we merge the unfolding matrices of two representative coefficients and constrain them to lie in a low-dimensional subspace. Moreover, the group sparsity of anomaly is characterized by $l_{2,1,1}$ norm regularization to promote the separation of background and anomaly. Extensive experiments conducted on several real HSI datasets demonstrate the superiority of SITSR compared with state-of-the-art anomaly detectors.

Index Terms—Anomaly detection, group sparsity, hyperspectral images (HSIs), tensor, tensor self-representation model.

I. Introduction

WITH THE development of hyperspectral imaging technology, hyperspectral images (HSIs) consist of more and more contiguous and narrow bands, which provide abundant spectral information of materials. Compared with other optical images, such as panchromatic and multispectral images, HSIs convey more precise spectral characteristics of land covers due to their wider spectral range and higher-spectral resolution, which bring great convenience to distinguish ground objects. This superiority makes the study of HSI become a topic, such as hyperspectral classification [1], [2], [3] and hyperspectral target detection [4], [5], [6], [7]. Target detection aims to identify pixels or subpixels with specific spectral signatures from input images. Usually, target detection is performed in a supervised manner, which means target and background information can be learned in advance from some available training samples. However, when there is no prior knowledge about target spectra, the target detection needs to be performed in an unsupervised manner, that is, anomaly detection. In recent years, hyperspectral anomaly detection has attracted considerable interest due to its significance in many applications, such as mineral exploration, military surveillance, and environmental surveillance [8], [9].

Anomalies in an HSI are referred to as pixels or subpixels whose spectral signatures and spatial features are different from their surroundings. In the past decades, a variety of methods have been established to achieve anomaly detection. One of the most famous detectors is the Reed–Xiaoli (RX) detector [10], which is considered as the milestone of anomaly detection. Based on the assumption that the background in an HSI obeys a multivariate Gaussian distribution, two versions of the RX detector were proposed, i.e., global RX (GRX) and local RX (LRX). For the GRX detector, all the pixels in an input image are used to estimate the mean and covariance of background distribution, and then the anomaly score is obtained by calculating the Mahalanobis distance between each test pixel and the background. Different from the GRX detector, a sliding dual window strategy is exploited in LRX to estimate the statistics of the background around the test pixel. In practice, the complicated background of a real scene cannot be described only by a multivariate Gaussian distribution. As a result, a lot of modified RX detectors have been developed, such as weighted-RX [11] and kernel-RX [12].

Recently, low-rank and sparsity-matrix decomposition (LRaSMD) has been successfully brought into hyperspectral due to its capability to formulate and separate latent data components. Li and Du [13] proposed an anomaly detection algorithm based on robust principal component analysis (RPCA) [14], where the background and anomaly are modeled by a low-rank matrix and a sparse matrix, respectively. Then, the sparse matrix was used to detect anomalies based

on the RX detector. Li et al. [15] proposed the prior-based tensor approximation (PTA) for hyperspectral anomaly detection, where the total variation norm (TV) is exploited to preserve the piecewise smoothness of the background. With consideration of the dense noise in HSIs, a more adaptive LRaSMD algorithm named go decomposition (GoDec) was proposed in [16] and introduced in hyperspectral anomaly detection [17], [18], [19]. To better describe the complicated anomaly and noise, a mixture noise model is incorporated into LRaSMD [20]. Considering that the HSI consists of types of background materials, the low-rank representation (LRR) model has been widely exploited in hyperspectral anomaly detection to model its multiple subspace structure [21], [22], [23], [24], [25], [26], [27]. The representation-based detectors are derived under the assumption that the background can be linearly represented among a candidate, while anomaly cannot. Xu et al. [21] combined low-rank and sparse representation (LRASR) for hyperspectral anomaly detection. In [22], an anomaly detection method based on graph and TV regularized LRR (GTVLRR) was proposed. In [23] and [24], the low-rank representative matrices were input into plug-and-play denoisers to fully exploit their spatial information.

The hyperspectral data with a spectral dimension and two spatial dimensions are intrinsically a three-order tensor. However, the above methods are derived after converting the cubic into a matrix, which destroys the information integrity and leads to a decay of the detection performance. To tackle this problem, different tensor decomposition methods have been proposed, such as tensor train decomposition [28], Tucker decomposition [29], and CANDECOMP/PARAFAC (CP) decomposition [30], [31], which have been successfully applied in super-resolution [32], denoising [33], complete [34], and so on. For the CP decomposition, the calculation of CP rank is NP-hard. The Tucker decomposition and CP decomposition have also been used in anomaly detection [35], [36], [37], [38], [39]. In [35] and [36], the major component in the sense of Tucker decomposition was eliminated, and the minor component was utilized to detect the anomaly. In [37], the background was suppressed by minimizing the Tucker-based nuclear norm. In [38], the CP decomposition is implemented on a set of subtensors divided from the original HSI, and the anomaly was extracted from the residual between the original tensor and the reconstructed tensor. Recently, the tensor singular value decomposition (t-SVD) model [40] based on the tensor–tensor product (t-product) has become popular. The t-SVD can be regarded as a generalization of matrix singular value decomposition and allows for taking full advantage of global information of HSIs. Based on the t-SVD, Lu et al. defined a new tensor nuclear norm (tnn) induced by t-product and extended RPCA to tensor RPCA (TRPCA) to recover a low-tubal-rank tensor and a sparse tensor from their sum. By virtue of the TRPCA, Xu et al. proposed a new framework, where hyperspectral compressive sensing and anomaly were implemented simultaneously. Besides, the tensor-based representation has been investigated in several works [41], [42], [43], [44]. In [41], [42], and [43], the tensor self-representation model was exploited to obtain a robust depiction of the global correlation of multidimensional data. Compared to the TPCA, the tensor-based representation model is more robust and can be utilized to recover the data drawn from a union of multiple tensor subspaces [44].

In this article, to take full use of the spatial information and spectral information of HSI, we propose a spatial invariant tensor self-representation (SITSR) model for hyperspectral anomaly detection. The HSI is treated as a three-order tensor with a spectral dimension and two spatial dimensions, and each frontal slice represents an image of a band. We first twist each frontal slice into a lateral slice, and then the background image of each band is represented as the sum of the t-product of all bands and their corresponding representative coefficient integrating spectral information and spatial information of a spatial mode. However, the HSI contains two spatial modes, and different twists may lead to different results due to the different spatial information included into the representative coefficient. As a result, two different representative tensors obtained under different twists of the HSI are exploited to achieve a more balanced and informative description of the spatial correlation, which indicates the self-representation model includes different spatial modes and then is spatial invariant. Each lateral slice of the representative tensor is considered as another representation of the corresponding background image with the integration of spectral information and spatial information. Then, the obtained two different representative tensors are connected under the assumption that their merged unfolding matrix lies in a low-dimensional subspace. To facilitate the separation of anomaly and background, we exploit the $l_{2,1,1}$ norm regularization to characterize the group sparsity of the anomaly.

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

1) Based on the t-product, the proposed method is implemented directly on the raw tensorial data, which preserves the multiway structure of the HSI. With the aid of the tensor self-representation model, SITSR effectively integrates spectral information and spatial information and takes full advantage of the global correlation.

2) We simultaneously utilize two self-representation methods with different spatial modes to build a more balanced and informative model. The background is suppressed well with the comprehensive utilization of spatial information. Extensive experiments conducted on several datasets indicate the superiority of SITSR compared with several state-of-the-art detectors.

The remainder of this article is organized as follows. The related work and some notations are introduced in Section II. The proposed SITSR model, including optimization procedure, is presented in Section III. The experimental results and discussions are provided in Section IV. Finally, we conclude this article in Section V.

II. RELATED WORK AND PRELIMINARIES

A. Tensor Self-Representation Model

Considering a data matrix $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$, each column of which is sampled from a union of multiple signal subspaces. The LRR model aims to find the lowest-rank representation

of $\mathbf{X}$ based on a given dictionary and segment each sample into its corresponding subspace [45]. Specifically, the matrix $\mathbf{X}$ can be written as $\mathbf{X} = \mathbf{DZ} + \mathbf{E}$, where $\mathbf{D}$ is a dictionary, $\mathbf{Z}$ is a coefficient matrix, and $\mathbf{E}$ is an error term designed for noise or outliers. And then, the LRR was established by solving the minimization problem as follows:

$$\min \ \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_l$$
$$\text{s.t.} \ \ \mathbf{X} = \mathbf{DZ} + \mathbf{E} \tag{1}$$

where λ is the tradeoff to balance the low-rank term and the error term, and $\|\cdot\|_l$ is certain matrix norm, such as l_1 norm to characterize sparse noise or $l_{2,1}$ norm to characterize sample-specific corruptions. To obtain the lowest-rank representation of the sample matrix, the dictionary matrix $\mathbf{D}$ should contain the bases of all signal subspaces. A simple way is to select the data matrix $\mathbf{X}$ itself as a dictionary, which is also named self-representation model. Then, the objective function can be written as

$$\min \ \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_l$$
$$\text{s.t.} \ \ \mathbf{X} = \mathbf{XZ} + \mathbf{E}. \tag{2}$$

However, this matrix-based representation is only suited for two-way data. For multiway data, such as HSI, one has to unfold the tensorial data into the matrix, which destroys the multiway structure. In order to tackle this problem, the tensor-based representation model has been used in several works [41], [42], [43], [44]. For the observed multidimensional data denoted by a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, it can be expressed as $\mathcal{X} = \mathcal{X} * \mathcal{Z} + \mathcal{E}$, where $\mathcal{Z}$ is a coefficient tensor, and $\mathcal{E}$ is an error term. For different types of contamination, $\mathcal{Z}$ and $\mathcal{E}$ can be derived by solving different problems. When $\mathcal{E}$ is considered as a sparse term, the problem can be written as [44]

$$\min \ \|\mathcal{Z}\|_{\text{tnn}} + \lambda \|\mathcal{E}\|_1$$
$$\text{s.t.} \ \ \mathcal{X} = \mathcal{X} * \mathcal{Z} + \mathcal{E} \tag{3}$$

where $\|\cdot\|_{\text{tnn}}$ denotes the tnn used to characterize the low-average rank of the representative tensor $\mathcal{Z}$ [46]. For dense noise, $\mathcal{Z}$ can be obtained by optimizing [41], [42], [43]

$$\min \|\mathcal{X} - \mathcal{X} * \mathcal{Z}\|_F^2 + \mathcal{R}(\mathcal{Z}) \tag{4}$$

where $\mathcal{R}(\mathcal{Z})$ is used to constraint $\mathcal{Z}$.

In this article, we exploit the tensor self-representation model for hyperspectral anomaly detection, but with a different optimization problem. We consider both sparse and dense terms in the proposed model, and the sparse term is regarded as the anomaly.

B. Tensor Notations

Some notations used in this article are listed in Table I. The mode-i unfolding matrix of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ is denoted as $\mathbf{X}_{(i)} \in \mathbb{R}^{n_i \times \prod_{k=1,k\neq i}^3 n_k}$ with $\mathbf{X}_{(i)}(l_i, j) = \mathcal{X}(l_1, l_2, l_3)$, where $j = 1 + \sum_{k=1,k\neq i}^3 (l_k - 1)J_k$ with $J_k = \prod_{m=1,m\neq i}^{k-1} l_m$. The mode-$i$ fold tensor of $\mathbf{X}_{(i)} \in \mathbb{R}^{n_i \times \prod_{k=1,k\neq i}^3 n_k}$ is $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$. For an N-order tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \ldots \times n_N}$, the k-mode product of $\mathcal{X}$ with a matrix $\mathbf{A} \in \mathbb{R}^{J \times n_k}$ can be calculated by

TABLE I
SOME NOTATIONS

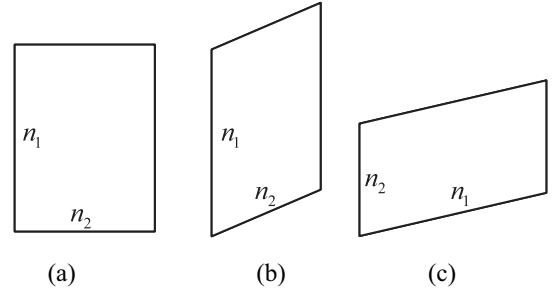| Notations	Descriptions	Notations	Descriptions
x	scalar	$\mathbf{x}$	vector
$\mathbf{X}$	matrix	$\mathcal{X}$	tensor
$\mathbf{I}_n$	$n \times n$ identity matrix	$(\cdot)^T$	transpose
$(\cdot)^\dagger$	conjugate transpose	x_{ij}	(i,j)-element
$\mathbf{X}(:,i)$	i-th column	x_{ijk}	(i,j,k)-th element
$\mathcal{X}(i,j,:)$	(i,j)-th tube	$\mathcal{X}^k$	k-th frontal slice
$\mathcal{X}(:,k,:)$	k-th lateral slice	$|x|$	absolute value
$\|\mathbf{x}\|_2$	$\sqrt{\sum_i x_i^2}$	$\|\mathbf{X}\|_{2.1}$	$\sum_i \|\mathbf{X}(:,i)\|_2$
$\|\mathbf{X}\|_*$	matrix nuclear norm	$\|\mathbf{X}\|_F$	$\sqrt{\sum_{ij} x_{ij}^2}$
$\|\mathcal{X}\|_1$	$\sum_{ijk} |x_{ijk}|$	$\|\mathcal{X}\|_{2.1.1}$	$\sum_{ij} \|\mathcal{X}(i,j,:)\|_2$
$\|\mathcal{X}\|_F$	$\sqrt{\sum_{ijk} x_{ijk}^2}$	$\mathcal{X} * \mathcal{Y}$	t-product

Fig. 1. One frontal slice of $\mathcal{X}$ and its two different twists. (a) Frontal slice of $\mathcal{X}$. (b) Lateral slice of $\mathcal{P}_1(\mathcal{X})$. (c) Lateral slice of $\mathcal{P}_2(\mathcal{X})$.

$(\mathcal{X} \times_k \mathbf{A})_{i_1,\ldots,i_{k-1},j,i_{k+1},i_N} = \sum_{i_k=1}^{n_k} x_{i_1,i_2,\ldots,i_N} a_{j,i_k}$. We use $\bar{\mathcal{X}}$ to denote the tensor obtained by taking the fast Fourier transform along the third dimension of $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$, i.e., $\bar{\mathcal{X}} = \text{fft}(\mathcal{X}, [], 3)$, and the inverse operation is $\mathcal{X} = \text{ifft}(\bar{\mathcal{X}}, [], 3)$, where fft and ifft are MATLAB commands. Besides, we define two twist manipulations, i.e., $\mathcal{P}_1$ and $\mathcal{P}_2$, which twist a tensor $\mathcal{X} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ into $\mathcal{P}_1(\mathcal{X}) \in \mathbb{R}^{n_1 \times n_3 \times n_2}$ and $\mathcal{P}_2(\mathcal{X}) \in \mathbb{R}^{n_2 \times n_3 \times n_1}$, respectively. As we can see in Fig. 1, Fig. 1(a) is the one frontal slice of $\mathcal{X}$. Fig. 1(b) and (c) are the corresponding lateral slices of $\mathcal{P}_1(\mathcal{X})$ and $\mathcal{P}_2(\mathcal{X})$, respectively.

The block circulant matrix $\text{bcirc}(\mathcal{X})$ is defined as

$$\text{bcirc}(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^1 & \mathcal{X}^{n_3} & \cdots & \mathcal{X}^2 \\ \mathcal{X}^2 & \mathcal{X}^1 & \cdots & \mathcal{X}^3 \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{X}^{n_3} & \mathcal{X}^{n_3-1} & \cdots & \mathcal{X}^1. \end{bmatrix}. \tag{5}$$

The block vectorizing operator $\text{unfold}(\mathcal{X})$ is defined as

$$\text{unfold}(\mathcal{X}) = \begin{bmatrix} \mathcal{X}^1 \\ \mathcal{X}^2 \\ \vdots \\ \mathcal{X}^{n_3} \end{bmatrix} \tag{6}$$

and its inverse operator is $\text{fold}(\text{unfold}(\mathcal{X})) = \mathcal{X}$.

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

$$\mathcal{Z} = \mathcal{X} * \mathcal{Y} = \text{fold}(\text{bcirc}(\mathcal{X}) \cdot \text{unfold}(\mathcal{Y})). \tag{7}$$

III. Proposed Anomaly Detection Algorithm

A. Problem Formulation

We denote an HSI with spatial size $w \times h$ and spectral size b by a tensor $\mathcal{Y} \in \mathbb{R}^{w \times h \times b}$. The observed tensor $\mathcal{Y}$ can be written as

$$\mathcal{Y} = \mathcal{X} + \mathcal{A} + \mathcal{N} \tag{8}$$

where $\mathcal{X} \in \mathbb{R}^{w \times h \times b}$ is the background tensor, $\mathcal{A} \in \mathbb{R}^{w \times h \times b}$ is the anomaly tensor, and $\mathcal{N} \in \mathbb{R}^{w \times h \times b}$ is the Gaussian noise. In this article, we aim to distinguish the anomaly component $\mathcal{A}$ and background component $\mathcal{X}$ from the given HSI data $\mathcal{Y}$ with the contamination of the Gaussian noise $\mathcal{N}$. This is challenging due to a lack of prior information on background pixels and anomaly pixels.

Fortunately, there are many different potential characteristics between anomaly and background, which provide convenience for anomaly detection. First, abnormal pixels occupy only a small part of all pixels and have different spectral curves and spatial features with background pixels. Second, for the background pixels in an HSI, a large number of continuous bands lead to a strong spectral correlation. Besides, just like most natural images, each band can be regarded as a gray image and then possesses a strong spatial correlation. For the sake of fully characterizing the inherent global correlation of the HSI data, we exploit the tensor-based representation model to formulate the background tensor with the aid of the t-product, i.e.,

$$\mathcal{X}_1 = \mathcal{Y}_1 * \mathcal{Z}_1 \tag{9}$$

where $\mathcal{X}_1 = \mathcal{P}_1(\mathcal{X}) \in \mathbb{R}^{w \times b \times h}$, $\mathcal{Y}_1 = \mathcal{P}_1(\mathcal{Y}) \in \mathbb{R}^{w \times b \times h}$, and $\mathcal{Z}_1 \in \mathbb{R}^{b \times b \times h}$ is the representative tensor. According to (9), we have

$$\mathcal{X}_1(:, j, :) = \mathcal{Y}_1 * \mathcal{Z}(:, j, :) = \sum_{i=1}^{b} \mathcal{Y}_1(:, i, :) * \mathcal{Z}(i, j, :) \tag{10}$$

where $\mathcal{Z}(i, j, :) \in \mathbb{R}^{1 \times 1 \times h}$ is the (i, j)th tube of the tensor $\mathcal{Z}$. As we can see, each lateral slice $\mathcal{X}_1(:, j, :) \in \mathbb{R}^{w \times 1 \times h}$, which represents the background image of band j, is expressed as the sum of the t-product of all lateral slices of $\mathcal{Y}_1$ and the corresponding representative tubes. This representation integrates the spatial information and spectral information into the representative coefficient. Define $\mathcal{A}_1 = \mathcal{P}_1(\mathcal{A}) \in \mathbb{R}^{w \times b \times h}$ and $\mathcal{N}_1 = \mathcal{P}_1(\mathcal{N}) \in \mathbb{R}^{w \times b \times h}$, we obtain the tensor self-representation model as follows:

$$\mathcal{Y}_1 = \mathcal{Y}_1 * \mathcal{Z}_1 + \mathcal{A}_1 + \mathcal{N}_1. \tag{11}$$

Compare to matrix-based representation, the tensor-based representation gets rid of vectoring each band of three-order HSI, and then better describes the correlation existing in different modes. Therefore, the representative tensor $\mathcal{Z}_1$ is more informative. Similar to the background dictionary $\mathbf{X}$ in (2), the original tensor data $\mathcal{Y}_1$ can be regarded as a tensorial dictionary [44]. Although $\mathcal{Y}_1$ contains anomalous pixels, its proportion is very small. The impact of anomalous pixels can be ignored in the process of reconstructing the background tensor with appropriate regularization imposed on $\mathcal{Z}_1$. With
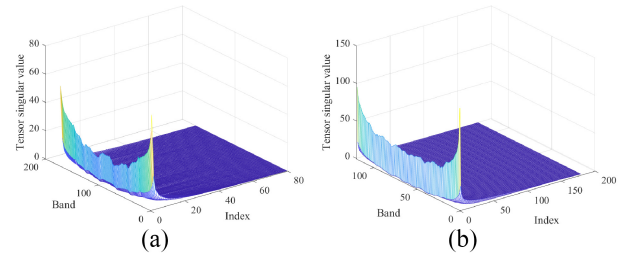

Fig. 2. Tensor singular values of the reconstructed background of the HYDICE and TIR datasets. (a) HYDICE. (b) TIR.

the tensor-self representation, each lateral slice $\mathcal{Z}_1(:, j, :)$ can be regarded as another representation of $\mathcal{X}_1(:, j, :)$, which is actually the twisted j-band background image. Therefore, the lateral slices of $\mathcal{Z}_1$ have a strong correlation. To depict this correlation, we assume that the mode-2 unfolding matrix $\mathbf{Z}_{1(2)}$ of $\mathcal{Z}_1$ lies in a low-dimensional subspace, i.e., $\mathbf{Z}_{1(2)} = \mathbf{FC}_1$, where $\mathbf{F} \in \mathbb{R}^{b \times r}$ is an r-dimensional subspace and $\mathbf{C}_1 \in \mathbb{R}^{r \times bh}$ is a representative matrix. To obtain the representative tensor $\mathcal{Z}_1$ and anomaly tensor $\mathcal{A}$, we can optimize the following problem:

$$\min \quad \frac{1}{2} \|\mathcal{Y}_1 - \mathcal{Y}_1 * \mathcal{Z}_1 - \mathcal{A}_1\|_F^2$$
$$+ \frac{\lambda}{2} \|\mathbf{Z}_{1(2)} - \mathbf{FC}_1\|_F^2 + \beta \|\mathcal{A}\|_{2.1.1}$$
$$\text{s.t.} \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}_r \tag{12}$$

where λ and β are two regularization parameters, and $\|\mathcal{A}\|_{2.1.1}$ is designed to account for the group sparsity of anomalous pixels.

The constraint on $\mathcal{Z}_1$ is different from the most existing tensor-based representation models, which exploits the tnn to characterize the low-average rank of $\mathcal{Z}_1$ [43], [44]. Our approach is mainly based on three considerations.

1) Benefiting from the calculation of the t-product, each lateral slice of the representative tensor $\mathcal{Z}_1$ has integrated spatial information and spectral information. As a result, even though we only exploit the low-rank property of mode-2 unfolding of the representative tensor $\mathcal{Z}_1$, we have characterized the global correlation of the background tensor. We plot the tensor singular values of the reconstructed background of the HYDICE and TIR datasets in Fig. 2. As we can see, most tensor singular values are very close to zero, which indicates that the spatial–spectral correlation of the HSI background is well formulated.

2) The average rank of $\mathcal{Z}_1$ is smaller than the rank of $\mathbf{Z}_{1(2)}$. That is to say, the low-rank constraint on $\mathbf{Z}_{1(2)}$ is a stronger condition, and the representative tensor satisfies this more stringent restriction.

3) This representation is convenient for us to illustrate the connection between different representative tensors corresponding to different spatial modes, which will be introduced in the next section.

B. Spatial Invariant Representation

In the tensor-based representation (9), we first turn $\mathcal{Y} \in \mathbb{R}^{w \times h \times b}$ into $\mathcal{Y}_1 \in \mathbb{R}^{w \times b \times h}$, and then utilize t-product to

integrate the spatial information along the third dimension of $\mathcal{Y}_1$ and spectral information to represent the background tensor. However, for an HSI $\mathcal{Y}$, it has two spatial dimensions. Correspondingly, there are two different ways to twist $\mathcal{Y}$, i.e., $\mathcal{Y}_1 = \mathcal{P}_1(\mathcal{Y}) \in \mathbb{R}^{w \times b \times h}$ and $\mathcal{Y}_2 = \mathcal{P}_2(\mathcal{Y}) \in \mathbb{R}^{h \times b \times w}$. Based on $\mathcal{Y}_1$ and $\mathcal{Y}_2$, we have two different representations, which can be written as

$$\begin{cases} \mathcal{Y}_1 = \mathcal{Y}_1 * \mathcal{Z}_1 + \mathcal{A}_1 + \mathcal{N}_1 \\ \mathcal{Y}_2 = \mathcal{Y}_2 * \mathcal{Z}_2 + \mathcal{A}_2 + \mathcal{N}_2 \end{cases} \quad (13)$$

where $\mathcal{A}_2 = \mathcal{P}_2(\mathcal{A})$ and $\mathcal{N}_2 = \mathcal{P}_2(\mathcal{N})$. When we only use one of these two representations for anomaly detection, such as problem (12), there are some drawbacks leading to a degeneration of detection performance. First, a single representation cannot make full use of the spatial information of HSI. The third dimensions of $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are actually the second and first dimensions of $\mathcal{Y}$, respectively. That is to say, a single representation only exploits the spatial information of one direction. Second, the representation is spatial imbalance. The result based on these two different representations will be different due to the directional calculation of the t-product, which is not intuitive.

To take full advantage of the spatial information of the background, we develop a new model which exploits both representations in (13). Then, we can obtain two different representative tensors, i.e., $\mathcal{Z}_1 \in \mathbb{R}^{b \times b \times h}$ and $\mathcal{Z}_2 \in \mathbb{R}^{b \times b \times w}$. As we discussed earlier, the lateral slices of $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are different representations of the corresponding frontal slice of $\mathcal{X}$. To describe the relation of the two representative tensors and exploit the correlation, we merge the mode-2 unfolding matrices of $\mathcal{Z}_1$ and $\mathcal{Z}_2$ into a matrix $\mathbf{Z}$ and constrain it to lie in a low-dimensional subspace. Specifically, for $\mathbf{Z} = [\mathbf{Z}_{1(2)}, \mathbf{Z}_{2(2)}]$, where $\mathbf{Z}_{2(2)}$ is the mode-2 unfolding matrix of $\mathcal{Z}_2$, we have $\mathbf{Z} = \mathbf{FC} = \mathbf{F}[\mathbf{C}_1, \mathbf{C}_2]$ with $\mathbf{C} = [\mathbf{C}_1, \mathbf{C}_2] \in \mathbb{R}^{r \times (bh+bw)}$, where $\mathbf{C}_2 \in \mathbb{R}^{r \times bw}$ is a coefficient matrix of $\mathbf{Z}_{2(2)}$. As a result, we propose a more intuitive and informative model as follows:

$$\min \sum_{i=1}^{2} \frac{1}{2} \|\mathcal{Y}_i - \mathcal{Y}_i * \mathcal{Z}_i - \mathcal{A}_i\|_F^2$$
$$+ \frac{\lambda}{2} \|\mathbf{Z} - \mathbf{FC}\|_F^2 + \beta \|\mathcal{A}\|_{2,1,1}$$
$$\text{s.t.} \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}_r \quad (14)$$

where the second term is exploited to depict the spectral correlation and spatial correlation of the background, and the third term is utilized to formulate the sparse anomaly pixels in the HSI. Compared to the conventional matrix-based models, the proposed model (14) is implemented directly on the raw tensorial data, which preserves the structural integrity of the HSI. Besides, we consider two spatial modes in the tensor self-representation, which can take full advantage of the spatial correlation. It is noted that $\mathcal{A}$, $\mathcal{A}_1$, and $\mathcal{A}_2$ are different twists of the same variable, but $\mathcal{Z}_1$ and $\mathcal{Z}_2$ are two different variables. Next, we introduce an alternating minimization method to solve the problem (14).

C. Optimization

We exploit an alternating minimization algorithm to optimize problem (14). In each step, we update one variable with the others fixed. The iteration procedure is as follows.

1) $\mathcal{Z}_i$ $(i = 1, 2)$ Subproblems: The objective function can be written as

$$\min_{\mathcal{Z}_i} \frac{1}{2} \|\mathcal{Y}_i - \mathcal{Y}_i * \mathcal{Z}_i - \mathcal{A}_i\|_F^2 + \frac{\lambda}{2} \|\mathbf{Z}_{i(2)} - \mathbf{FC}_i\|_F^2. \quad (15)$$

To solve the above problem, we first exploit a mode-2 fold operation to transfer the matrix $\mathbf{C}_i$ into a tensor $\mathcal{C}_i$ ($\mathcal{C}_1 \in \mathbb{R}^{b \times r \times h}$ and $\mathcal{C}_2 \in \mathbb{R}^{b \times r \times w}$). Define $\mathcal{B}_i = \mathcal{C}_i \times_2 \mathbf{F}$, $\|\mathbf{Z}_{i(2)} - \mathbf{FC}_i\|_F^2$ can be rewritten as $\|\mathcal{Z}_i - \mathcal{B}_i\|_F^2$. Then, we transform (15) into the Fourier domain, for each frontal slice, we have

$$\min_{\bar{\mathcal{Z}}_i^v} \frac{1}{2} \|\bar{\mathcal{Y}}_i^v - \bar{\mathcal{Y}}_i^v \bar{\mathcal{Z}}_i^v - \bar{\mathcal{A}}_i^v\|_F^2 + \frac{\lambda}{2} \|\bar{\mathcal{Z}}_i^v - \bar{\mathcal{B}}_i^v\|_F^2. \quad (16)$$

Taking the derivative of (16) with respect to $\bar{\mathcal{Z}}_i^v$ and setting it to zero, we have

$$-(\bar{\mathcal{Y}}_i^v)^\dagger \bar{\mathcal{Y}}_i^v + (\bar{\mathcal{Y}}_i^v)^\dagger \bar{\mathcal{Y}}_i^v \bar{\mathcal{Z}}_i^v + (\bar{\mathcal{Y}}_i^v)^\dagger \bar{\mathcal{A}}_i^v + \lambda(\bar{\mathcal{Z}}_i^v - \bar{\mathcal{B}}_i^v) = \mathbf{0}. \quad (17)$$

Then, we have

$$\bar{\mathcal{Z}}_i^v = \left[\lambda \mathbf{I} + (\bar{\mathcal{Y}}_i^v)^\dagger \bar{\mathcal{Y}}_i^v \right]^{-1}$$
$$\times \left[\lambda \bar{\mathcal{B}}_i^v + (\bar{\mathcal{Y}}_i^v)^\dagger \bar{\mathcal{Y}}_i^v - (\bar{\mathcal{Y}}_i^v)^\dagger \bar{\mathcal{A}}_i^v \right]. \quad (18)$$

After getting all frontal slices of $\bar{\mathcal{Z}}_i$, $\mathcal{Z}_i$ can be calculated by $\mathcal{Z}_i = \text{ifft}(\bar{\mathcal{Z}}_i, [], 3)$.

2) $\mathbf{C}_i$ $(i = 1, 2)$ and $\mathbf{F}$ Subproblems: The objective function can be written as

$$\min \|\mathbf{Z} - \mathbf{FC}\|_F^2$$
$$\text{s.t.} \quad \mathbf{F}^T \mathbf{F} = \mathbf{I}_r. \quad (19)$$

Problem (19) can be efficiently solved by iteration as in [47] and [48]. In this work, we exploit a different method and directly obtain the optimal solution of (19) by the SVD.

Proposition 1: Define

$$\begin{cases} \mathcal{J}_1 = \min_{\mathbf{F}^T\mathbf{F}=\mathbf{I}_r,\mathbf{C}} \|\mathbf{Z} - \mathbf{FC}\|_F^2 \\ \mathcal{J}_2 = \min_{\text{rank}(\mathbf{E})=r} \|\mathbf{Z} - \mathbf{E}\|_F^2 \end{cases} \quad (20)$$

then, we have $\mathcal{J}_1 = \mathcal{J}_2$.

Proof: On the one hand, for the optimal $(\mathbf{F}_*, \mathbf{C}_*)$, we can easily find an $\mathbf{E} = \mathbf{F}_*\mathbf{C}_*$ with $\text{rank}(\mathbf{E}) = r$, which indicates $\mathcal{J}_2 \leq \mathcal{J}_1$. On the other hand, the singular value decomposition of $\mathbf{Z}$ can be written as $\mathbf{Z} = \mathbf{USV}^T$. According to the Eckart–Young Theorem [49], we have the optimal $\mathbf{E}_* = \mathbf{U}(:, 1{:}r)\mathbf{S}(1:r, 1:r)\mathbf{V}(:, 1:r)^T$. Then, we can find an $\mathbf{F} = \mathbf{U}(:, 1:r)$ and a $\mathbf{C} = \mathbf{S}(1:r, 1:r)\mathbf{V}(:, 1:r)^T$, and, thus, we have $\mathcal{J}_1 \leq \mathcal{J}_2$. Overall, we have $\mathcal{J}_1 = \mathcal{J}_2$. Then, we know that $(\mathbf{F} = \mathbf{U}(:, 1{:}r), \mathbf{C} = \mathbf{S}(1:r, 1:r)\mathbf{V}(:, 1:r)^T)$ is one of the optimal solutions of (19). ∎

Algorithm 1: Optimization Procedure for SITSR

Input: data tensor $\mathcal{Y}$, tradeoff parameters β, λ, rank r.
Initialize: set the variables $\mathcal{A}$, $\mathcal{Z}_1$, and $\mathcal{Z}_2$ as zero
tensors, $k = 1$, $k_{\max} = 100$, $\varepsilon_0 = 10^{-6}$.
While ($\varepsilon < \varepsilon_0$ or $k \leq k_{\max}$) **do**
  1) Update $\mathcal{Z}_1$ and $\mathcal{Z}_2$ by Eq. (15);
  2) Update $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathbf{F}$ by Eq. (21);
  3) Update $\mathcal{A}$ by Eq. (24);
  4) Calculate ϵ by Eq. (26);
  5) $k = k + 1$.
End while
Output: obtain detection map with (27).

Finally, we have

$$\begin{cases} \mathbf{F} = \mathbf{U}(:, 1:r) \\ \mathbf{C} = \mathbf{S}(1:r, 1:r)\mathbf{V}(:, 1:r)^T \\ \mathbf{C}_1 = \mathbf{C}(:, 1:bh) \\ \mathbf{C}_2 = \mathbf{C}(:, bh+1:bh+bw). \end{cases} \quad (21)$$

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

$$\min \sum_{i=1}^{2} \frac{1}{2}\Big(\|\mathcal{Y}_i - \mathcal{Y}_i * \mathcal{Z}_i - \mathcal{A}_i\|_F^2\Big) + \beta\|\mathcal{A}\|_{2.1.1}. \quad (22)$$

Define

$$\tilde{\mathcal{A}} = \frac{1}{2}\mathcal{P}_1^{-1}(\mathcal{Y}_1 - \mathcal{Y}_1 * \mathcal{Z}_1) + \frac{1}{2}\mathcal{P}_2^{-1}(\mathcal{Y}_2 - \mathcal{Y}_2 * \mathcal{Z}_2) \quad (23)$$

where $\mathcal{P}_1^{-1}$ and $\mathcal{P}_2^{-1}$ are the inverse operators of $\mathcal{P}_1$ and $\mathcal{P}_2$, respectively, then, we have

$$\mathcal{A}(i, j, :) = \text{soft}_{2.1}\left(\tilde{\mathcal{A}}(i, j, :), \frac{\beta}{2}\right) \quad (24)$$

where

$$\text{soft}_{2.1}(\mathbf{x}, a) = \max\left(1 - \frac{a}{\|\mathbf{x}\|_2}, 0\right) \cdot \mathbf{x}. \quad (25)$$

The iteration terminates when the number of iterations k reaches a preselected $k_{\max}$ or the residual ϵ of $\mathcal{Z}_i$ is smaller than a preassigned value ϵ_0. For clarity, we abuse the meaning of superscript here and write $\mathcal{Z}_i$ in k-iteration as $\mathcal{Z}_i^k$. And then, ϵ can be calculated by

$$\epsilon = \sqrt{\left\|\mathcal{Z}_1^{k+1} - \mathcal{Z}_1^k\right\|_F} + \sqrt{\left\|\mathcal{Z}_2^{k+1} - \mathcal{Z}_2^k\right\|_F}. \quad (26)$$

Finally, the anomaly detection map $\mathbf{R}$ can be obtained by

$$\mathbf{R}(i, j) = \|\mathcal{A}(i, j, :)\|_2. \quad (27)$$

And a pixel is considered as an anomaly pixel if its detection value $\mathbf{R}(i, j) > \xi$, where ξ is a detection threshold. The pseudocode for the proposed method is summarized in Algorithm 1.

D. Computational Complexity Analysis

The computational complexity of Algorithm 1 includes the following.

1) The complexity of computing the DFT and inverse DFT of tensors with size $b \times b \times h$ in updating $\mathcal{Z}_1$ is $\mathcal{O}(b^2 h \log h)$, and the complexity of computing $\mathcal{Z}_1^v$, ($v = 1, \ldots, h$) is $\mathcal{O}(h(b^3 + b^2 w))$. Then, the cost of updating $\mathcal{Z}_1$ is $\mathcal{O}(b^2 h \log h + b^2 wh + b^3 h)$. Similarly, the cost of updating $\mathcal{Z}_2$ is $\mathcal{O}(b^2 w \log w + b^2 wh + b^3 w)$.

2) The computational cost of updating $\mathbf{C}_i$ and $\mathbf{F}$ mainly comes from the singular value decomposition of $\mathbf{Z}$, which is $\mathcal{O}(b^3(h + w))$.

3) Updating $\mathcal{A}$ needs computing the t-product of tensors, and its complexity is $\mathcal{O}(b^2 wh + bwh \log h + b^2 h \log h + bwh \log w + b^2 w \log w)$.

4) Overall, the computational complexity of Algorithm 1 is $\mathcal{O}(k(b^3 h + b^3 w + b^2 wh + b^2 h \log h + b^2 w \log w + bwh \log h + bwh \log w))$.

IV. Experimental Results

In this section, we exploit the four real hyperspectral data to evaluate the performance of the proposed SITSR. Before we input the data into the optimization procedure, we normalize the data band by band. Details of these real data are given below.

A. Dataset Description

The first dataset captures an urban area by the hyperspectral digital imagery collection experiment (HYDICE) airborne sensor [26], whose spatial resolution is about 1.56 m and spectral wavelengths range from 400 to 2500 nm. The original data has 307×307 pixels and 210 bands. A subarea of 80×100 is cropped from the original data for testing. After discarding the water absorption and low-signal-to-noise bands, 175 bands are used in our experiment. The false-color image and the corresponding ground-truth map are shown in the first row of Fig. 3.

The second dataset, the Gainesville urban dataset, was collected by the airborne visible/infrared imaging spectrometer (AVIRIS) sensor [50]. The exploited image consists of 100×100 pixels. 191 bands are preserved after the removal of several low-signal-to-noise and water absorption bands. The false-color image and the corresponding ground-truth map are displayed in the second row of Fig. 3.

The third dataset was captured by the SpecTIR hyperspectral airborne Rochester experiment (SHARE) [20]. The spectral and spatial resolutions of this data are 5 nm and 1 m, respectively. The experimental data contains 180×180 pixels and 120 bands. The false-color image and ground-truth map are shown in the third row of Fig. 3.

The fourth data was collected by the Nuance Cri hyperspectral sensor [18]. The spectral resolution of this data is 10 nm. The scene captures an area of 400×400 pixels and 46 spectral channels with wavelengths ranging from 650 to 1100 nm. The false-color image and its corresponding ground-truth map are displayed in the fourth row of Fig. 3.
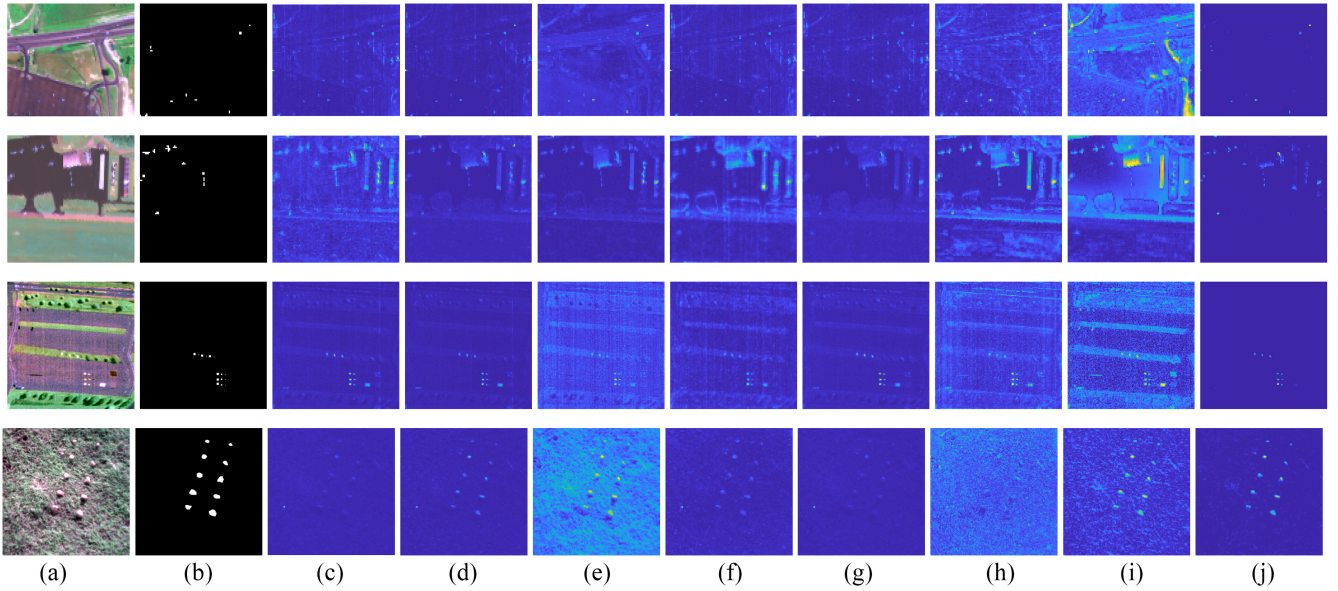
Fig. 3. Detection maps of different methods on the four test datasets. (a) False color. (b) Ground truth. (c) GRX. (d) RPCA-RX. (e) LRASR. (f) TPCA. (g) TDAD. (h) GTVLRR. (i) PTA. (j) SITSR.

B. Methods in Comparison and Evaluation Metrics

1) Methods: We compare the proposed algorithm with several types of anomaly detection methods. As a benchmark detector, the statistical-based GRX [10] is chosen as the baseline. The RPCA-RX [13] and PTA [15] algorithms are implemented based on the LRaSMD architecture, in which the latter simultaneously exploits the TV regularization. One parameter in the RPCA-RX method is set as $(1/wh)$. For the PTA detector, two critical parameters need to be tuned. The truncated low-rank r is searched from 0 to 20 and the penalty parameter μ is selected from $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3, 10^4\}$. The regularization parameters are set to the same values as [15], i.e., $\alpha = 1$, $\tau = 1$, and $\beta = 0.01$. The LRASR [18] and GTVLRR [22] are two LRR-based anomaly detection methods, and the piecewise smoothness property is also considered in the GTVLRR. The same dictionary construct method based on k-means is exploited in both methods. The number of clusters is set to 15 and 20 pixels are selected as atoms from each cluster. Two regularization parameters β and λ in LRASR are searched in the sets $\{0.001, 0.01, 0.1, 0.5, 1, 2, 3\}$ and $\{0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$, respectively. Three parameters in GTVLRR are also tuned to optimal, i.e., β and λ are selected from $\{0.005, 0.05, 0.1, 0.3, 0.5, 0.7, 1\}$, and γ is picked from $\{0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5\}$. Besides, two tensor-based algorithms, TPCA [51] and TDAD [52], are also exploited for comparisons.

2) Evaluation Metrics: To obtain a comprehensive performance evaluation, we exploit several different metrics to evaluate the anomaly detectors. First, the detection performance is intuitively presented with the aid of the detection map. Different colors on the detection map indicate the abnormal scores of pixels, and bright yellow represents a high-abnormal score while dark blue means the opposite. The target detection ability and background suppression

ability of the anomaly detectors can be visually reflected by the detection map. The second metric is the receiver operating characteristic (ROC) curve. The ROC curve shows the change of the detection probability P_D with the false alarm probability P_F, which can be calculated by

$$P_D = \frac{N_D}{N_A}, \ P_F = \frac{N_F}{N_B} \tag{28}$$

where N_D denotes the number of pixels correctly determined as abnormal, N_A is the total number of real abnormal pixels, N_F is the number of background pixels determined as abnormal pixels, and N_B represents the total number of true background pixels. Besides, for the detection threshold τ, we also give the ROC curve of (P_F, τ), which reflects the ability of a detector to suppress background [53]. The third is the area under the ROC curve (AUC). The AUC scores of the above two ROC curves can quantitatively reflect the overall detection performance of a detector.

C. Performance

The detection maps of different anomaly detectors on the four test datasets are displayed in Fig. 3. The detection maps of the HYDICE dataset are shown in the first row of Fig. 3. As we can see, the GRX, RPCA-RX, TPCA, and TDAD methods obtain low responses to abnormal pixels and the detection results are slightly contaminated by the background. Two LRR-based methods, LRASR and GTVLRR, have high responses to abnormal pixels and the former has a stronger ability to suppress the background. The PTA detector fails to detect the anomaly and suppress the background. Compared to the competitors, the proposed SITSR can detect anomalies well and has a strong ability to suppress the background. For the Gainesville dataset, as we can see in the second row of Fig. 3, the results of GRX, TPCA, GTVLRR, and PTA
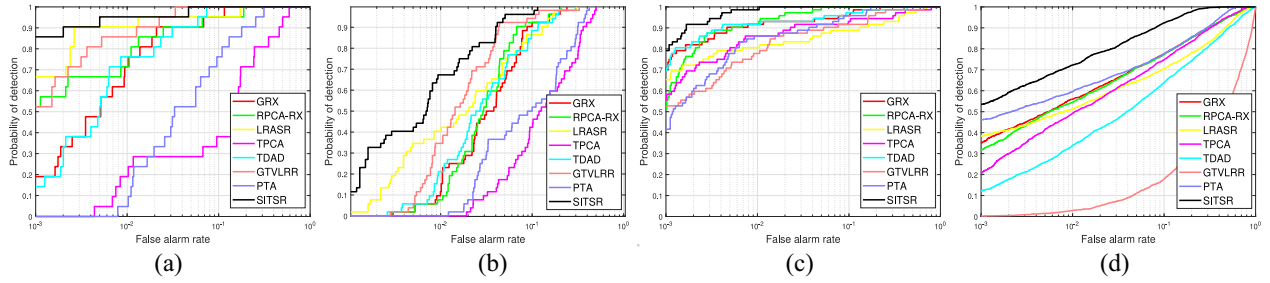
Fig. 4. ROC curves of (P_D, P_F) of different methods on the four test datasets. (a) HYDICE. (b) Gainesville. (c) TIR. (d) Cri.
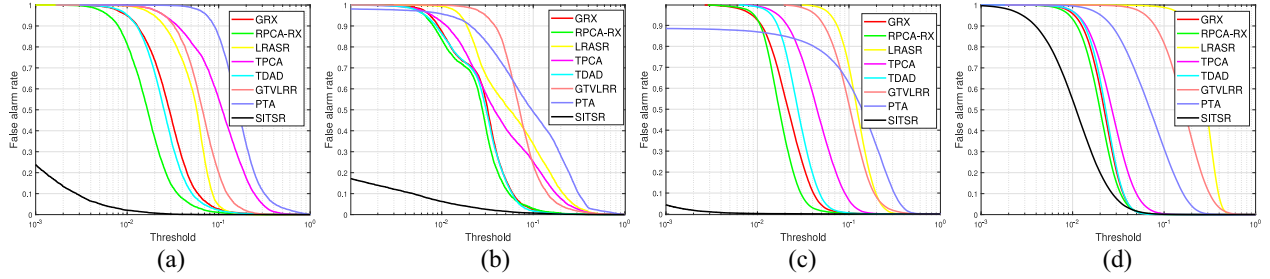

Fig. 5. ROC curves of (P_F, τ) of different methods on the four test datasets. (a) HYDICE. (b) Gainesville. (c) TIR. (d) Cri.

methods are seriously polluted by the background. The RPCA-RX, LRASR, and TDAD algorithms are also influenced by the background and have poor visual effects. By comparison, our method performs well in detecting anomalies and suppressing background. The detection maps of the TIR dataset are presented in the third row of Fig. 3. In addition to the SITSR, the compared algorithms are more or less disturbed by background and noise. A large amount of dense noise appears on the detection map of the PTA method. Vertical stripes decay the detection performances of the LRASR, TPCA, and GTVLRR detectors. Besides, all detectors except the proposed SITSR fail to suppress the band-shaped background. For the Cri dataset shown in the fourth row of Fig. 3, the LRASR and PTA methods successfully locate the anomaly but with severe background interference. The GRX, RPCA-RX, TPCA, and TDAD algorithms have a good ability to suppress background, however, the response of the anomaly is not visually obvious. The anomaly response of the GTVLRR detector is completely submerged in the background and noise. By contrast, the proposed SITSR can detect the anomaly while performing well in eliminating the impact of the background.

The detection maps present the superiority of the proposed SITSR in target detection or background suppression. As shown in Figs. 4 and 5, we exploit two types of ROC curves to further illustrate the detection performance of different methods. For the ROC curve of (P_D, P_F), the closer the curve is to the upper-left corner, the better the overall detection performance. For the HYDICE dataset shown in Fig. 4(a), SITSR achieves higher values of P_D than other methods at small values of P_F. For the Gainesville, TIR, and Cri datasets, as we can see in Fig. 4(b)–(d), the ROC curve of SITSR is always above others, which means SITSR reaches higher probabilities of detection than its competitors at all values of P_F. The ROC curve of (P_F, τ) reflects the background suppression
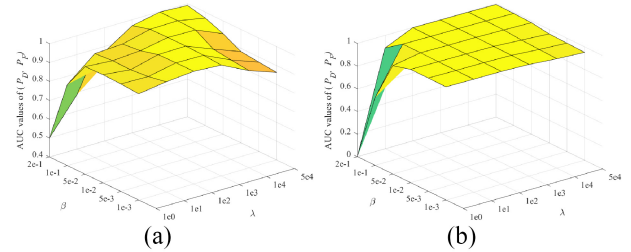

Fig. 6. AUC values of (P_D, P_F) under different combinations of λ and β. (a) HYDICE. (b) TIR.

ability of a detector and is considered to perform well if it is close to the lower-left corner. The ROC curves shown in Fig. 5 are consistent with the detection map, which verifies the superiority of SITSR in suppressing background due to the lowest values of P_F at different detection thresholds, especially, for the HYDICE, Gainesville, and TIR datasets. For a quantitative evaluation, the AUC values of (P_D, P_F) and (P_F, τ) are calculated to directly compare the detection performance of all detectors. As shown in Table II, SITSR obtains the highest-AUC scores of (P_D, P_F) and lowest-AUC scores of (P_F, τ) on all the test datasets.

The running time of all methods is presented in Table III. As we can see, running the GRX detector costs the least time. The GTVLRR is the most time-consuming. The proposed SITSR is also time-consuming due to the high-computational complexity of the tensor operation.

D. Parameter Tuning

Three parameters are related to the detection performance of the proposed SITSR. Two regularization parameters λ and β are used to balance the group sparse term

TABLE II
AUC SCORES ON THE FOUR DATASETS

Dataset	AUC values	GRX	RPCA-RX	LRASR	TPCA	TDAD	GTVLRR	PTA	TSR-h	TSR-v	SITSR-l_1	SITSR
HYDICE	$AUC(P_D, P_F)$	0.9856	0.9842	0.9905	0.8241	0.9874	0.9958	0.9265	0.9930	0.9956	0.9967	**0.9971**
	$AUC(P_F, \tau)$	0.0351	0.0212	0.0574	0.1296	0.0314	0.0749	0.1935	0.0059	0.0048	0.0038	**0.0014**
Gainesville	$AUC(P_D, P_F)$	0.9512	0.9577	0.9566	0.8225	0.9514	0.9712	0.8612	0.9732	0.9721	0.9807	**0.9855**
	$AUC(P_F, \tau)$	0.0351	0.0327	0.0922	0.0729	0.0338	0.0866	0.1437	0.0104	0.0093	0.0104	**0.0042**
TIR	$AUC(P_D, P_F)$	0.9914	0.9970	0.9564	0.9694	0.9887	0.9731	0.9866	0.9966	0.9992	0.9991	**0.9993**
	$AUC(P_F, \tau)$	0.0249	0.0197	0.1302	0.0514	0.0308	0.1156	0.1531	0.0003	0.0004	0.0020	**0.0002**
Cri	$AUC(P_D, P_F)$	0.9134	0.9125	0.8781	0.9078	0.8478	0.7906	0.9271	0.8545	0.8476	0.9747	**0.9877**
	$AUC(P_F, \tau)$	0.0235	0.0217	0.2901	0.0302	0.0242	0.1848	0.0841	0.0182	0.0222	0.0153	**0.0100**

TABLE III
EXECUTION TIME OF DIFFERENT ALGORITHMS ON THE FOUR DATASETS (UNIT: SECONDS)

Methods	GRX	RPCA-RX	LRASR	TPCA	TDAD	GTVLRR	PTA	SITSR
HYDICE	0.0619	3.2145	24.4226	14.9591	0.8274	136.0146	48.3014	109.8857
Gainesville	0.0721	5.4603	24.7638	22.2857	1.0576	176.7196	57.9303	162.3671
TIR	0.1524	9.0368	101.7068	34.3787	2.1700	640.3531	165.9385	166.1050
Cri	0.2181	25.0648	849.4161	54.1528	7.4444	5588.3063	295.7440	149.9834
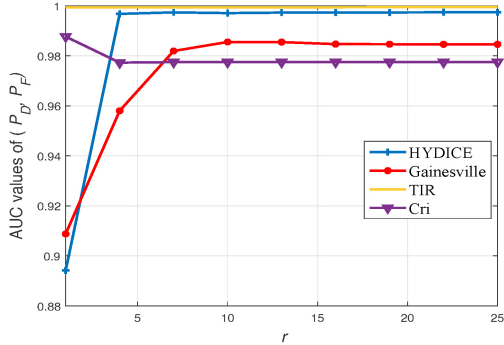
Fig. 7. AUC values of (P_D, P_F) with different rank r.

TABLE IV
PARAMETER SETTINGS ON THE FOUR DATASETS

Dataset	HYDICE	Gainesville	TIR	Cri
β	0.2	0.2	0.1	0.005
λ	10000	10000	10	50000
r	10	10	10	1

Gainesville datasets improves first with the increase of r and then tends to be stable. The detection performance of the TIR dataset is always good with different r. The background pattern of the Cri dataset is relatively simple, and the AUC value reaches the highest when $r = 1$, then declines after $r > 1$. In practice, the rank r can be set to a slightly larger value. We list the specific parameter settings of the test datasets in Table IV.

E. Ablation Study

1) Spatial Invariant Representation: Compared with the tensor self-representation model with a single spatial mode, the proposed SITSR is more informative with the consideration of double spatial information. To verify the utility of this alteration for hyperspectral anomaly detection, we design additional experiments in which only one direction of spatial information is exploited. Specifically, for the representation $\mathcal{Y}_1 = \mathcal{Y}_1 * \mathcal{Z}_1 + \mathcal{A}_1 + \mathcal{N}_1$, we conduct anomaly detection by optimizing (12), and we name this model TSR-h. The other representation $\mathcal{Y}_2 = \mathcal{Y}_2 * \mathcal{Z}_2 + \mathcal{A}_2 + \mathcal{N}_2$ is implemented in the same way and is named TSR-v. In order to achieve a fair comparison, the parameters in TSR-h and TSR-v are reselected from the candidates. We present the ROC curves of (P_D, P_F) and (P_F, τ) in Figs. 8 and 9, respectively. As we can see in Fig. 8, under a single utilization of different spatial information, the detection performance of TSR-h and TSR-v with different false alarm probabilities has its own advantages. However, when we combine the spatial information
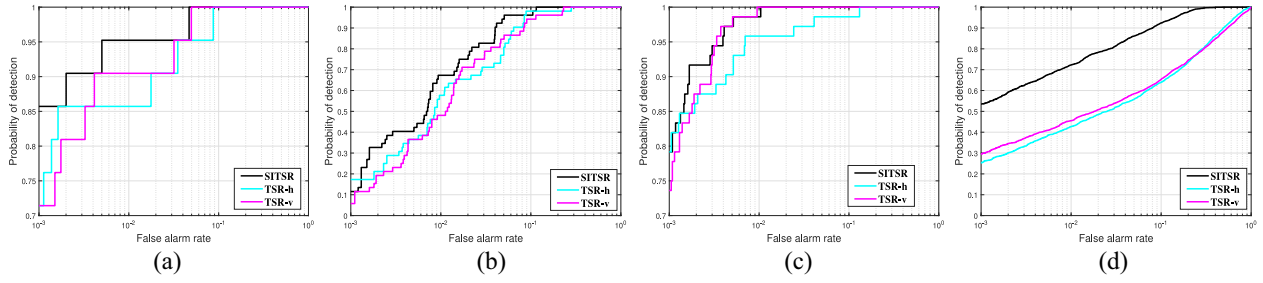
and low-rank term, respectively. A large parameter leads to a small value of its corresponding term. In order to better suppress the background and reach satisfactory anomaly detection ability, parameters λ and β are selected from the sets {1, 10, 100, 1000, 10000, 50000} and {0.001, 0.005, 0.01, 0.05, 0.1, 0.2}, respectively. In Fig. 6, we present the AUC values of (P_D, P_F) of the HYDICE and TIR datasets under different combinations of λ and β. As we can see, the detection performance of SITSR degenerates when λ and β decrease simultaneously within the predefined parameter ranges. But good performances of the HYDICE and TIR datasets can be retained with large ranges of λ and β. To obtain a satisfactory detection performance, λ and β need to be adjusted individually for each dataset.

The other parameter is the rank r of the matrix **Z**, of which each row is an another representation of the corresponding band of the background. As a result, the value of r is relevant to the complexity of the hyperspectral background. We test the influence of r on the detection performance of SITSR by varying it from 1 to 25. The AUC values of (P_D, P_F) with different r on the test datasets are shown in Fig. 7. As we can see, the detection performance of the HYDICE and

Fig. 8. ROC curves of (P_D, P_F) with different spatial information on the four test datasets. (a) HYDICE. (b) Gainesville. (c) TIR. (d) Cri.
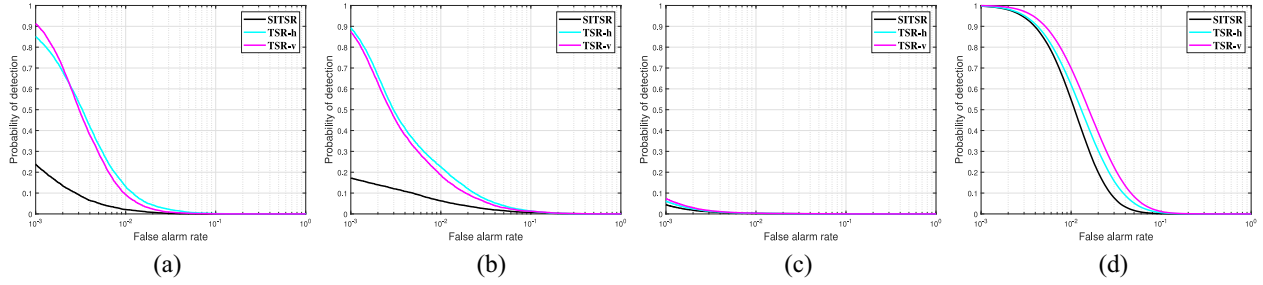

Fig. 9. ROC curves of (P_F, τ) with different spatial information on the four test datasets. (a) HYDICE. (b) Gainesville. (c) TIR. (d) Cri.
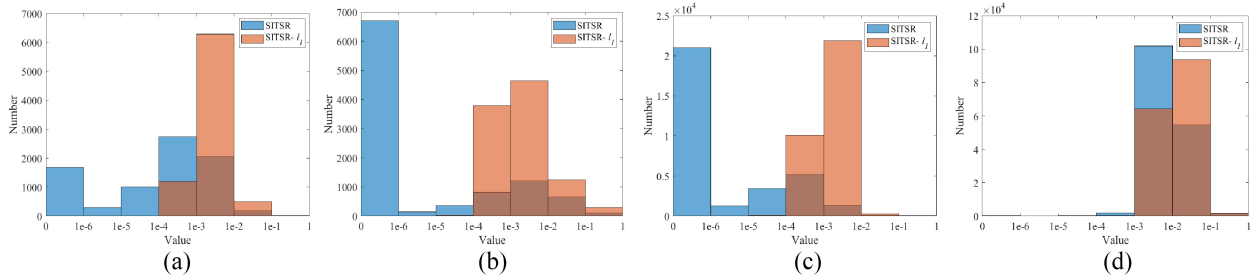

Fig. 10. Histograms of the normalized anomaly obtained by the SITSR and SITSR-l_1 on the four test datasets. (a) HYDICE. (b) Gainesville. (c) TIR. (d) Cri.

of two directions, actually SITSR, the detection performance has been improved to varying degrees, especially, for the Cri dataset. The same conclusion can be obtained from Fig. 9, which presents a better ability of SITSR to suppress background compared with TSR-h and TSR-v. The AUC values are also listed in Table II. As we can see, SITSR obtains the highest-AUC values of (P_D, P_F) and lowest-AUC values of (P_F, τ).

2) Group Sparsity: To verify the effectiveness of the $l_{2.1.1}$ norm in modeling the pixel-wise sparse anomaly, we replace it in (14) with the l_1 norm. The SITSR based on the l_1 norm is denoted as SITSR$-l_1$. Histograms of the normalized anomaly obtained by the SITSR and SITSR-l_1 are presented in Fig. 10. As we can see, for the Gainesville and TIR datasets, most values of the anomaly map of the SITSR are close to zero. For the HYDICE and Cri datasets, most values are small, and only a few values are between 0.1 and 1. Even if only a few values in the anomaly map of the Cri dataset are close to 0, the anomaly can be detected well, because the response of the anomaly is far greater than that of the background. Besides, the value of the anomaly part obtained by the SITSR is much smaller than that of the SITSR-l_1, which indicates the $l_{2.1.1}$ norm performs better than the l_1 norm in modeling the anomaly. We also list

the values of AUC(P_D, P_F) and AUC(P_F, τ) of the SITSR-l_1 in Table II, which shows that the SITSR has better overall performance.
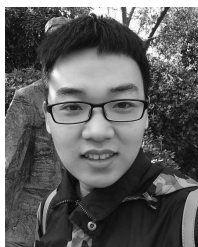
V. CONCLUSION

In this article, we presented a SITSR method for hyperspectral anomaly detection based on the tensor self-representation model. Compared with conventional matrix-based anomaly detectors derived by using the unfolding matrix of the 3-D HSI, our tensor-based method is implemented directly on the raw tensorial data, which preserves structural integrity and makes it possible to use the global correlation. The spatial information and spectral information were incorporated by virtue of the t-product. A more balanced and informative model was proposed by combining two self-representation ways with two different spatial modes. Then, the obtained two different representative coefficients were integrated into a low-rank matrix, and the low-rank property was characterized by pursuing a low-dimensional subspace. For the anomaly component, we utilized the $l_{2.1.1}$ norm to formulate its group sparsity. Extensive experiments conducted on several datasets have validated the superiority of the proposed

SITSR compared with the state-of-the-art detectors. And the utility of the double spatial self-representation model was also verified by the designed ablation study. As for future work, a tensor subspace combining coefficient tensor instead of the HSI data itself can be exploited to depict the global correlation of the background. And more regularization can be imposed on the coefficient tensor to utilize more potential information of the HSI.

REFERENCES

[1] M. Zhang, W. Li, Y. Zhang, R. Tao, and Q. Du, "Hyperspectral and LiDAR data classification based on structural optimization transmission," *IEEE Trans. Cybern.*, early access, May 13, 2022, doi: 10.1109/TCYB.2022.3169773.

[2] W. Li and Q. Du, "Gabor-filtering-based nearest regularized subspace for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1012–1022, Apr. 2014.

[3] Q. Zhu et al., "A spectral-spatial-dependent global learning framework for insufficient and imbalanced hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 52, no. 11, pp. 11709–11723, Nov. 2022.

[4] C. Jiao et al., "l_1 sparsity-regularized attention multiple-instance network for hyperspectral target detection," *IEEE Trans. Cybern.*, vol. 53, no. 1, pp. 124–137, Jan. 2023, doi: 10.1109/TCYB.2021.3087662.

[5] 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, doi: 10.1109/TCYB.2021.3070909.

[6] 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, doi: 10.1109/TNNLS.2021.3071026.

[7] 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. Syst.*, vol. 33, no. 11, pp. 6504–6517, Nov. 2022, doi: 10.1109/TNNLS.2021.3082158.

[8] 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.

[9] 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.

[10] 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. 19, pp. 1760–1770, Oct. 1990.

[11] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, "Weighted-RXD and linear filter-based RXD: Improving background statistics estimation for anomaly detection in hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, Jun. 2014.

[12] 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.

[13] W. Li and Q. Du, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.

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

[15] 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, doi: 10.1109/TNNLS.2020.3038659.

[16] T. Zhou and D. Tao, "GoDec: Randomized lowrank & sparse matrix decomposition in noisy case," in *Proc. Int. Conf. Mach. Learn.*, Bellevue, WA, USA, 2011, pp. 33–40.

[17] W. Sun, C. Liu, J. Li, Y. M. Lai, and W. Li, "Low-rank and sparse matrix decomposition-based anomaly detection for hyperspectral imagery," *J. Appl. Remote Sens.*, vol. 8, no. 1, 2014, Art. no. 83641.

[18] 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.

[19] 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.

[20] 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. 2021, doi: 10.1109/TCYB.2020.2968750.

[21] 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. 2016.

[22] 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.

[23] 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, no. 60, Dec. 2021, Art. no. 5500117, doi: 10.1109/TGRS.2020.3040221.

[24] X. Fu, S. Jia, L. Zhuang, M. Xu, J. Zhou, and Q. Li, "Hyperspectral anomaly detection via deep plug-and-play denoising CNN regularization," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9553–9568, Nov. 2021.

[25] X. Zhang, X. Ma, H. Ning, 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, doi: 10.1109/TGRS.2020.3046727.

[26] H. Ning, 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] T. Cheng and B. Wang, "Total variation and sparsity regularized decomposition model with union dictionary for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1472–1486, Feb. 2021.

[28] I. V. Oseledets, "Tensor-train decomposition," *SIAM J. Sci. Comput.*, vol. 33, no. 5, pp. 2295–2317, 2011.

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

[30] R. A. Harshman, "Foundations of the parafac procedure: Models and conditions for an 'explanatory' multi-model factor analysis," UCLA, Los Angeles, CA, USA, Working Papers in Phonetics, 1970, pp. 1–84.

[31] J. D. Carroll and J.-J. Chang, "Analysis of individual differences in multidimensional scaling via an n-way generalization of 'Eckart-Young' decomposition," *Psychometrika*, vol. 35, pp. 283–319, Sep. 1970.

[32] 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.

[33] 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.

[34] J. Xue, Y. Zhao, W. Liao, J. C.-W. Chan, and S. G. Kong, "Enhanced sparsity prior model for low-rank tensor completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 11, pp. 4567–4581, Nov. 2020.

[35] 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)*, 2015, pp. 4525–4529.

[36] 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.

[37] 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)*, 2017, pp. 1141–1144.

[38] X. Zhang, G. Wen, and W. Dai, "Anomaly detecting in hyperspectral imageries based on tensor decomposition with spectral and spatial partitioning," in *Proc. 8th Int. Congr. Image Signal Process. (CISP)*, 2015, pp. 737–741.

[39] J. Wang, Y. Xia, and Y. Zhang, "Anomaly detection of hyperspectral image via tensor completion," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, pp. 1099–1103, 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, Aug. 2011.

[41] E. Kernfeld, S. Aeron, and M. Kilmer, "Clustering multi-way data: A novel algebraic approach," 2014, *arXiv:1412.7056*.

[42] M. Cheng, L. Jing, and M. K. Ng, "Tensor-based low-dimensional representation learning for multi-view clustering," *IEEE Trans. Image Process.*, vol. 28, pp. 2399–2414, 2019.

[43] X. Xiao, Y. Chen, Y.-J. Gong, and Y. Zhou, "Low-rank preserving t-linear projection for robust image feature extraction," *IEEE Trans. Image Process.*, vol. 30, pp. 108–120, 2021.

[44] P. Zhou, C. Lu, J. Feng, Z. Lin, and S. Yan, "Tensor low-rank representation for data recovery and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 5, pp. 1718–1732, May 2021.

[45] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yong, 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.

[46] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis with a new tensor nuclear norm," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 4, pp. 925–938, Apr. 2020.

[47] 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.

[48] 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.

[49] C. Eckart and G. Young, "The approximation of one matrix by another of lower rank," *Psychometrika*, vol. 1, no. 3, pp. 211–218, 1936.

[50] W. Xie, J. Lei, J. Yang, Y. Li, Q. Du, and Z. Li, "Deep latent spectral representation learning-based hyperspectral band selection for target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 2015–2026, Mar. 2020.

[51] B. Yang, Z. Chen, and B. Wang, "A preprocessing method for hyperspectral target detection based on tensor principal component analysis," *Remote Sens.*, vol. 10, no. 7, p. 1033, 2018.

[52] X. Zhang and G. Wen, "A fast and adaptive method for determining k_1, k_2, and k_3 in the tensor decomposition-based anomaly detection algorithm," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, pp. 3–7, 2018.

[53] 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. 2021.

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 Postdoctoral Research Associate with the Department of Electrical and Computer Engineering, Duke University, Durham, NC, USA. From January 2013 to September 2014, he was a Postdoctoral 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. His research interests include statistical signal processing, image processing, and machine learning.

Dr. Liu was the recipient of the Best Paper Award from the IEEE WCSP 2021. 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. He is a member of the Sensor Array and Multichannel Technical Committee, IEEE Signal Processing Society. He is currently an Associate Editor for 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 Postdoctoral Researcher with the University of California at Davis, 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 the 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 IEEE GEOSCIENCE AND REMOTE SENSING SOCIETY, and the Outstanding Paper Award at IEEE INTERNATIONAL WORKSHOP ON HYPERSPECTRAL IMAGE AND SIGNAL PROCESSING: EVOLUTION IN REMOTE SENSING (Whispers) in 2019. He is currently serving as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. 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.

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. He is currently pursuing the Ph.D. degree in information and communication engineering with the University of Science and Technology of China, Hefei.

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