Learning Tensor Low-Rank Representation for Hyperspectral Anomaly Detection

Minghua Wang, *Member, IEEE,* Qiang Wang, *Member, IEEE,* Danfeng Hong, *Senior Member, IEEE,* Swalpa Kumar Roy, *Student Member, IEEE,* and Jocelyn Chanussot, *Fellow, IEEE*

Abstract—Recently, low-rank representation (LRR) methods have been widely applied for hyperspectral anomaly detection, owing to their potentials in separating the backgrounds and anomalies. However, existing LRR models generally convert 3-D hyperspectral images (HSIs) into 2-D matrices, inevitably leading to the destruction of intrinsic 3-D structure properties in HSIs. To this end, we propose a novel tensor low-rank and sparse representation (TLRSR) method for hyperspectral anomaly detection. A 3-D TLR model is expanded to separate the LR background part represented by a tensorial background dictionary and corresponding coefficients. This representation characterizes the multiple subspace property of the complex LR background. Based on the weighted tensor nuclear norm and the $L_{F,1}$ sparse norm, a dictionary is designed to make its atoms more relevant to the background. Moreover, a principal component analysis (PCA) method can be assigned as one pre-processing step to exact a subset of HSI bands, retaining enough the HSI object information and reducing computational time of the post-processing tensorial operations. The proposed model is efficiently solved by the well-designed alternating direction method of multipliers (ADMM). A comparison with the existing algorithms via experiments establishes the competitiveness of the proposed method with the state-of-the-art competitors in the hyperspectral anomaly detection task.

keywords—Tensor low-rank, sparse representation, dictionary construction, hyperspectral anomaly detection, alternating direction method of multipliers.

I. INTRODUCTION

Unlike traditional remote sensing imaging techniques, hyperspectral imagery collected from satellite or airborne imaging sensors provides hundreds of spectral bands for the same scene, with many crucial applications in classification [1], [2], unmixing [3], [4], and target detection [5], [6]. Among these applications, hyperspectral anomaly detection has attached increasing attention in recent years, since it is constructive for civilian and military developments [7], [8]. The goal of hyperspectral anomaly detection is to separate interesting objects from backgrounds. For example, planes in a village background and vehicles in a bridge background are usually referred to as anomalies or outliers. Anomaly detection can be considered as an unsupervised two-class classification problem where anomalies occupy small areas with their surrounding background. The key to settling this matter is to explore the differences between anomalies and their background. Anomalies commonly appear with low probabilities and their spectral signatures are quite different from neighbors.

Based on investigating the statistical characteristics, the classical Reed-Xiaoli (RX) algorithm and its variants have been proposed in the last three decades. RX is a constant false-alarm rate algorithm developed from the generalized likelihood ratio, generally assuming that the background conforms to a multivariate Gaussian distribution [9]. The RX detector calculates the mean vector and covariance matrix of the samples from the HSI scene to obtain the Mahalanobis distance between testing pixels and their surrounding background. Global RX considers the statistics of background pixels from the entire image, whereas local RX uses local statistics [10], [11]. Nevertheless, two drawbacks deteriorate the detection performances of these RX algorithms: the assumption of the background described by a single normal distribution is rarely suitable for real hyperspectral applications; the mean and covariance matrix of the background are susceptible to noisy pixels and anomalies. Many modifications to the RX methods have been proposed to cope with these limitations. Schaum [12] used the principal component analysis (PCA) of the background covariance matrix to obtain subspace RX (SSRX). Matteoli *et al.* [13] proposed the segment RX model by investigating the critical factors which affect local background covariance matrix estimation. Guo *et al.* [14] developed weighted-RX and linear filter-based RX to provide better background estimation. Kwon and Nasrabadi [15] proposed the kernel RX (KRX) algorithm with the kernelization of the basic RX model in the feature space. However, KRX brings a high computational burden. Zhou *et al.* [16] presented a cluster KRX approach to aggregate background pixels into clusters via a fast eigendecomposition algorithm. Zhao *et al.* [17] used a Gaussian background purification method and an inverse-of matrix-free model to improve KRX.

Many researchers have actively attempted to study other HSI inherent properties like the low-rank representation (LRR) of the background part and the sparsity of anomalies. Due to the correlations of HSI bands, the matrix whose columns

are consisted of the vectorized bands of HSI background is LR [18]. Anomalies usually account for a small portion of an image with low probabilities, which naturally leads to an assumption that anomalies have a sparse property. Chen *et al.* [19] employed the robust principal component analysis (RPCA) [20] framework to decompose the HSI data into the LR and sparse parts. Consequently, several LR and sparse matrix decomposition (LRaSMD) approaches have been proposed [21], [22], but they excessively focus on the sparse component while neglecting the background component. To alleviate this limitation, Zhang *et al.* [23] developed an LRaSMD-based Mahalanobis distance (LSMAD) method to explore the LR prior knowledge of the background part and design a new Mahalanobis-distance-based detector. Xu *et al.* [24] characterized the background information via a new strategy for dictionary construction and used a sparsity-inducing regularized LRR to obtain more accurate sparse anomalies. Niu and Wang [25] exploited a learning dictionary to enhance the robustness of LRR to the regularization parameters. Tan *et al.* [26] added a spatial constraint in the LRR coefficients with single or multiple local window modes. Cheng and Wang [27] combined the graph [28] nd TV regularizer with the LR model to use the spatial information of neighboring pixels. Xu *et al.* [29] fused collaborative representation and the Euclidean distance into the solving procedure of the LRaSMD framework. Li *et al.* [30] investigated LRaSMD under the assumption of the mixture of Gaussian (MoG) and designed a global detector by calculating the Manhattan distance. Inspired by an enhanced LR property of the background signals in local homogeneous regions, Wang *et al.* [31] proposed a locally enhanced LR prior for hyperspectral anomaly detection. However, the detection performances of the LRaSMD-based methods are limited by vectorizing the HSI bands and discarding the useful multi-order information. Meanwhile, to exploit the priors from off-the-shelf denoisers (such as BM3D [32], weighted nuclear norm minimization (WNNM) [33], and convolutional neural network (CNN)-based denoisers [34]), Fu *et al.* [35] proposed a plug-and-play denoising CNN regularized anomaly detection (DeCNN-AD) method.

Containing two spatial dimensions and one spectral dimension, HSIs are intrinsically regarded as three-order tensors. Most existing algorithms like the above RX-based and LRR-based models ignore the 3-D tensor property of HSIs, but this property has been considered in HSI denoising [36], [37], super-resolution [38], [39] and fusion [40], [41]. The tensor-based approaches have been gaining attention for HSI anomaly detection in recent years. Li *et al.* [42] employed Tucker decomposition to obtain the core tensor of the LR part, where an unmixing approach is used to extract the spectral signatures of anomalies. Except for the same decomposition, Zhang *et al.* [43] utilized a reconstruction-error-based method to eliminate the background pixels and remain the anomaly information. Chen *et al.* [44] presented a Tensor PCA (TRPCA)-based pre-processing method to separate a principal component part and a residual part. Based on Tucker decomposition, Qin *et al.* [45] proposed a LR and sparse tensor decomposition. Song *et al.* [46] used a Tucker decomposition-based endmember extraction model to yield an abundance matrix and decomposed

anomalies from this abundance matrix. Nonetheless, Tucker decomposition has an obvious limitation about its ranks.

Until now, the above tensor-based anomaly detection methods focus on the Tucker decomposition and hardly involve other tensor representation models. Nevertheless, an HSI tensor is decomposed into one small core tensor and three matrices via Tucker decomposition [47]. Three ranks should be predefined before carrying out the computing procedure, which is bound to bring difficulties in practice. Therefore, it is imperative for the current development of the tensor-based methods to consider an efficient decomposition model. [48] proposed a tensor LRR method for data reconstruction and clustering with theoretical guarantees.

In this paper, a novel Tensor LR and Sparse representation method is proposed for hyperspectral anomaly detection. To the best of our knowledge, we develop the concept of Tensor LRR for the first time in the field of HSI processing. This representation model is expanded from a 2-D LR matrix form, which effectively preserves the HSI 3-D inherent structure. As opposed to Tucker decomposition, the solution of the LR background part relies on the tensor nuclear norm without any presetting problem of ranks. Unlike the sparse noise randomly distributed in three HSI directions, anomalies appear in the same spatial positions of all spectral bands. To some degree, they are slightly denser in the spectral direction. Thus, we constrain the joint spectral-spatial sparsity. This sparsity is also utilized in dictionary construction. Moreover, the tensorial operations tend to bring a heavier computational burden than the matrix's. The PCA pre-processing step can strike a balance between computational accuracy and time. Compared with existing matrix-based and tensor-based techniques, the main contributions of this study are summarized as follows:

- To the best of our knowledge, this is the first time to expand the concept of Tensor LR representation in hyperspectral anomaly detection, which exploits the 3-D inherent structure of HSIs. Assisted by the multi-subspace learning of the tensor domain and the sparsity constraint along the joint spectral-spatial dimensions, we separate the LR background and anomalies in a more accurate manner.
- Based on the weighted tensor nuclear norm and the $L_{F,1}$ sparsity-promoting norm, we provide a novel dictionary construction strategy. The dictionary is composed of the LR background tensor, which is preserved by the adaptive shrinkage of weighted tensor nuclear norm effectively.
- To strike the balance between computational accuracy and complexity of the tensor-based models, we serve the PCA as a pre-processing step and propose a PCA-based Tensor LR and Sparse Representation (PCA-TLRSR) model. We design the alternating direction method of multipliers (ADMM) to solve the resulting model more effectively.

The remainder of this paper is organized as follows. Some notations and preliminaries of tensors are introduced in Sec.II. Sec.III presents the details of the proposed PCA-TLR detector. The experimental results and analysis are reported in Sec.IV. Finally, the conclusions are drawn in Sec.V

II. Notations and Preliminaries

In this section, we summarize some notations and preliminaries used in this paper. The notations are introduced in Table I. An m-way tensor is defined by $\mathbb{R}^{n_1 \times n_2 \times \ldots \times n_m}$ and its order m denotes the number of its dimensions or modes. A fiber of a tensor is a 1-D section defined by fixing every index but one. A matrix has a column (mode-1 fiber) and row (mode-2 fiber). Except for column and row, a 3-D tensor has mode-3 fiber (tube). For a three-way HSI tensor $\mathcal{X} \in \mathbb{R}^{h \times v \times z}$, its conjugate transpose is obtained by conjugate transposing each front slice of $\mathcal{X}$ and then reversing the order of transposed frontal slices 2 through z. $\hat{\mathcal{X}}$ is the result of discrete Fourier transformation (DFT) of $\mathcal{X}$ along the third dimension, i.e., $\hat{\mathcal{X}} = \text{fft}(\mathcal{X}, [], 3)$. Naturally, $\mathcal{X}$ can be obtained via the inverse DFT $\mathcal{X} = \text{ifft}(\hat{\mathcal{X}}, [], 3)$. The vectorization of a matrix $\mathbf{X} \in \mathbb{R}^{h \times v}$, denoted as $\mathbf{x} = \text{vec}(\mathbf{X}) \in \mathbb{R}^{hv \times 1}$, is defined as $\mathbf{x} = [\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, ..., \mathbf{X}^{(hv)}]^{\text{T}}$. The block circulant matrix $\text{bcirc}(\mathcal{A}) \in \mathbb{R}^{hz \times vz}$ of $\mathcal{A} \in \mathbb{R}^{h \times v \times z}$, with $h \times v$ frontal slices denoted $A^{(i)}$, is defined as

$$\text{bcirc}(\mathcal{A}) = \begin{bmatrix} A^{(1)} & A^{(z)} & A^{(z-1)} & \ldots & A^{(2)} \\ A^{(2)} & A^{(1)} & A^{(z)} & \cdots & A^{(3)} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ A^{(z)} & A^{(z-1)} & \ddots & A^{(2)} & A^{(1)} \end{bmatrix}.$$

The block vectorization operation $\text{bvec}(\cdot)$ and its inverse operation $\text{bvfold}(\cdot)$ are anchored as

$$\text{bvec}(\mathcal{A}) = \begin{bmatrix} A^{(1)} \\ A^{(2)} \\ \vdots \\ A^{(z)} \end{bmatrix}, \quad \text{bvfold}(\text{bvec}(\mathcal{A})) = \mathcal{A}.$$

TABLE I: The notations used in the paper

Notation	Description
$\mathcal{X}$	a tensor
$\mathbf{X}$	a matrix
$\mathbf{x}$	a vector
x	a scalar
$\mathcal{X}^*$	the conjugate transpose of $\mathcal{X}$
$\mathbf{X}^{\text{T}}$	the transpose of $\mathbf{X}$
$\mathbf{x}^{\text{T}}$	the transpose of $\mathbf{x}$
$\mathcal{X}^{(k)}$	the k-th frontal slice of $\mathcal{X}$
$\mathcal{I}$	an identify tensor
$vec(\mathbf{X})$	$vec(\mathbf{X})$ stacks the columns of $\mathbf{X}$
$\mathcal{X} \in \mathbb{R}^{h \times v \times z}$	tensors with 3-modes
$\mathcal{X}_{h_i, v_i, z_i}$	the (h_i, v_i, z_i)-element of $\mathcal{X}$
$\mathcal{X}(i,:,:), \mathcal{X}(:,i,:)$ and $\mathcal{X}(:,:,i)$	the i-th horizontal, lateral and frontal slices
$\lVert \mathcal{X} \rVert_1 = \sum_{h_i, v_i, z_i} \lvert \mathcal{X}_{h_i, v_i, z_i} \rvert$	the l_1 norm of $\mathcal{X}$
$\lVert \mathcal{X} \rVert_F = \sqrt{\sum_{h_i, v_i, z_i} \lvert \mathcal{X}_{h_i, v_i, z_i} \rvert^2}$	the Frobenius norm of $\mathcal{X}$
$\sigma_i(\mathbf{X})$	the singular values of $\mathbf{X}$
$\lVert \mathbf{X} \rVert_* = \sum_i \sigma_i(\mathbf{X})$	the nuclear norm of $\mathbf{X}$
$\lVert \mathbf{x} \rVert_2 = \sqrt{\sum_i \lvert \mathbf{x}_i \rvert^2}$	the l_2 norm of $\mathbf{X}$
$\hat{\mathcal{X}} = \text{fft}(\mathcal{X}, [], 3)$	the Fourier transformation of $\mathcal{X}$ along mode-3

Definition 1 (Tensor product [49], [50]): The product $\mathcal{C} \in \mathbb{R}^{n_1 \times n_4 \times n_3}$ of two 3-way tensors $\mathcal{A} \in \mathbb{R}^{n_1 \times n_2 \times n_3}$ and $\mathcal{B} \in \mathbb{R}^{n_2 \times n_4 \times n_3}$ is:

$$\mathcal{C} = \text{bvfold}(\text{bcirc}(\mathcal{A}) * \text{bvec}(\mathcal{B})) \tag{1}$$

The product of two tensors is analogous to the matrix multiplication unless the multiplication operation between the elements is replaced by a circular convolution between the tubes, i.e.,

$$\mathcal{C} = \mathcal{A} * \mathcal{B} \Leftrightarrow \mathcal{C}(i,k,:) = \sum_{t=1}^{n_2} \mathcal{A}(i,j,:) \star \mathcal{B}(j,k,:) \tag{2}$$

where $\star$ represents the circular convolution between two tubes.

Definition 2 (Special tensors [50], [51])

F-diagonal tensor: A tensor $\mathcal{S}$ is f-diagonal if each slice of this tensor is a diagonal matrix.

Identity tensor: The identity tensor $\mathcal{I} \in \mathbb{R}^{h \times v \times z}$ is the tensor whose first frontal slice is an identity matrix and all other frontal slices are zeros.

Orthogonal tensor: A tensor $\mathcal{Q}$ is orthogonal when it satisfies $\mathcal{Q}^* * \mathcal{Q} = \mathcal{Q} * \mathcal{Q}^* = \mathcal{I}$.

Definition 3 (t-SVD [50], [51]) A tensor $\mathcal{X} \in \mathbb{R}^{h \times v \times z}$ has the factorization

$$\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^* \tag{3}$$

where $\mathcal{U} \in \mathbb{R}^{h \times h \times z}$, $\mathcal{V} \in \mathbb{R}^{v \times v \times z}$ are orthogonal tensors and $\mathcal{S} \in \mathbb{R}^{h \times v \times z}$ is a f-diagonal tensor. The procedure of t-SVD is described in the Algorithm 1.

Algorithm 1 t-SVD

Input: $\mathcal{X} \in \mathbb{R}^{h \times v \times z}$
1: $\hat{\mathcal{X}} = \text{fft}(\mathcal{X}, [], 3)$;
2: **for** $i = 0, 1, \ldots, [\frac{z+1}{2}]$
3:   $[\hat{\mathcal{U}}(:,:,i), \hat{\mathcal{S}}(:,:,i), \hat{\mathcal{V}}(:,:,i)] = \text{SVD}(\hat{\mathcal{X}}(:,:,i))$;
4: **end for**
5: **for** $i = [\frac{z+1}{2} + 1], \ldots, z$
6:   $\hat{\mathcal{U}}(:,:,i) = \text{conj}(\hat{\mathcal{U}}(:,:,z-i+2))$;
7:   $\hat{\mathcal{S}}(:,:,i) = (\hat{\mathcal{S}}(:,:,z-i+2))$
8:   $\hat{\mathcal{V}}(:,:,i) = \text{conj}(\hat{\mathcal{V}}(:,:,z-i+2))$;
9: **end for**
10: $\mathcal{U} = \text{ifft}(\hat{\mathcal{U}}, [], 3)$, $\mathcal{S} = \text{Ifft}(\hat{\mathcal{S}}, [], 3)$, $\mathcal{V} = \text{fft}(\hat{\mathcal{V}}, [], 3)$
Output: $\mathcal{U}, \mathcal{S}, \mathcal{V}$

Definition 4 (Tensor tubal rank [50], [51]) For a tensor $\mathcal{X} \in \mathbb{R}^{h \times v \times z}$, the tubal rank $rank_t(\mathcal{X})$ of $\mathcal{X}$ is the number of non-zeros singular tubes $\mathcal{S}$, where $\mathcal{X} = \mathcal{U} * \mathcal{S} * \mathcal{V}^*$ denotes the t-SVD of $\mathcal{X}$:

$$rank_t(\mathcal{X}) = \#\{k : \mathcal{S}(k,k,:) \neq 0\} \tag{4}$$

Definition 5 (Tensor Nuclear Norm (TNN) [50], [52]) The tensor nuclear norm of a tensor $\mathcal{X} \in \mathbb{R}^{h \times v \times z}$ is the sum of singular values of all front slices of $\hat{\mathcal{X}}$, that is,

$$\lVert \mathcal{X} \rVert_* := \sum_{k=1}^{z} \lVert \hat{\mathcal{X}}^{(k)} \rVert_*. \tag{5}$$

III. Proposed Model

The schematic diagram of the proposed method is shown in Fig. 1. The spectral dimension of the observed HSI image is reduced by performing PCA in order to lighten the computational burden. A dictionary construction strategy is conducted via the weighted tensor nuclear norm and the $L_{F,1}$ sparse-promoting norm. Finally, we employ TLRSR to exactly separate the LR background component and the

sparse anomalies without any destruction of the inherent 3-D structure.

A. TLRSR for Hyperspectral Anomaly Detection

Due to the normal existence of mixed pixels in HSIs, the HSI data tends to be drawn from multiple subspaces. LRR is more appropriate than RPCA in this context since RPCA assumes that the underlying data lie in a single subspace. An HSI $\mathcal{X} \in \mathbb{R}^{h \times v \times z}$ is reformatted as a matrix $\mathbf{X} \in \mathbb{R}^{hv \times z}$ whose columns are composed of vectorized HSI bands. LRR aims at exploring the LR linear representation of $\mathbf{X}$ by a dictionary $\mathbf{A}$:

$$\mathbf{X} = \mathbf{A}\mathbf{Z} + \mathbf{E} \tag{6}$$

where $\mathbf{A}\mathbf{Z}$ is the LR background part, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_m]$ denotes the background dictionary with m atoms, $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_z]$ is the corresponding representation coefficients, and $\mathbf{E} = [\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_z]$ represents the sparse part related to anomalies. The LRR model captures the low rankness along the spectral dimension of HSIs, yet destroys the intrinsic HSI tensor structure. To alleviate this limitation, we consider tensor LR linear representation [48], which performs more naturally and auspiciously than vector linear representation for HSI data analysis. An HSI tensor $\mathcal{X} \in \mathbb{R}^{h \times v \times z}$ is decomposed by tensor LRR:

$$\mathcal{X} = \mathcal{A} * \mathcal{Z} + \mathcal{E} \tag{7}$$

We consider the representation problem of obtaining the corresponding coefficient tensor $\mathcal{Z}$ and the remaining tensor $\mathcal{E}$ related to anomalies. Given an observed HSI $\mathcal{X}$ and a dictionary $\mathcal{A}$, there are still many feasible solutions of two unknown tensors $\mathcal{Z}$ and $\mathcal{E}$. To tackle this issue, we demand more criteria for constraining $\mathcal{Z}$ and $\mathcal{E}$. On the one hand, the spectral signature of each pixel corresponds to one type of material or mixed materials. Thus, the spectrum of every pixel lies in at least one subspace. Overall HSI spectra tend to be drawn from a union of subspaces. Tensor linear representation contributes to describe and analyze the LR subspaces, avoiding vectorizing the tensor data. The tubal rank minimization is added to constrain the tensor low rankness of the coefficient tensor $\mathcal{Z}$. On the other hand, anomalies fixedly occur in a small fraction of pixels in most bands and achieve a dense performance in the spectral direction. Naturally, we should further restrict the joint spectral-spatial sparsity. To take account of the above factors, the optimization problem of Tensor LRR is written as:

$$\min_{\mathcal{Z},\mathcal{E}} rank_t(\mathcal{Z}) + \lambda ||\mathcal{E}||_{F,1}$$
$$\text{s.t. } \mathcal{X} = \mathcal{A} * \mathcal{Z} + \mathcal{E} \tag{8}$$

where $rank_t(\cdot)$ denotes the tensor tubal rank function, λ is a regularization parameter, and $||\cdot||_{F,1}$ denotes the $L_{F,1}$ norm defined as:

$$||\mathcal{E}||_{F,1} = \sum_{i=1}^{h} \sum_{j=1}^{v} \sqrt{\sum_{k=1}^{z} \mathcal{E}(i,j,k)^2} \tag{9}$$

However, the problem (8) is non-convex and NP-hard, but it can be transformed into the following convex optimization by replacing $rank_t(\cdot)$ by TNN:

$$\min_{\mathcal{Z},\mathcal{E}} ||\mathcal{Z}||_* + \lambda ||\mathcal{E}||_{F,1}$$
$$\text{s.t. } \mathcal{X} = \mathcal{A} * \mathcal{Z} + \mathcal{E} \tag{10}$$

where $||\cdot||_*$ is TNN of $\mathcal{Z}$ solved by the tensor singular value thresholding algorithm. This algorithm considers the largest or the smallest singular value equally, hardly leading to an appropriate approximation of $rank_t(\mathcal{Z})$. Inspired by [33], we introduce the Weighted Nuclear Norm (WNN) into problem (10):

$$\min_{\mathcal{Z},\mathcal{E}} ||\mathcal{Z}||_{w,*} + \lambda ||\mathcal{E}||_{F,1}$$
$$\text{s.t. } \mathcal{X} = \mathcal{A} * \mathcal{Z} + \mathcal{E} \tag{11}$$

where the weighted TNN $||\mathcal{Z}||_{\omega,*}$ is the sum of the WNNM of all the frontal slices of $\hat{\mathcal{Z}}$:

$$||\mathcal{Z}||_{\omega,*} = \sum_{k=1}^{z} ||\hat{\mathcal{Z}}(:,:,k)||_{\omega,*} = \sum_{k=1}^{z} \sum_{i=1}^{min(h,v)} \omega(i,i,k)\sigma(i,i,k) \tag{12}$$

where $\sigma(i,i,k)$ is the singular value of each frontal slice $\hat{\mathcal{Z}}(:,:,k)$ and ω stands for the weighted parameter.

ADMM is a straightforward and efficient approach to solve problem (11) [53], [54]. By introducing an auxiliary variable $\mathcal{W}$, we decouple the variable from the objective $||\mathcal{Z}||_{w,*}$ and the constraint $\mathcal{X} = \mathcal{A} * \mathcal{Z} + \mathcal{E}$. Eq.(11) is reformulated as:

$$\min_{\mathcal{Z},\mathcal{E}} ||\mathcal{Z}||_{w,*} + \lambda ||\mathcal{E}||_{F,1}$$
$$\text{s.t. } \mathcal{Z} = \mathcal{W}, \mathcal{X} = \mathcal{A} * \mathcal{W} + \mathcal{E} \tag{13}$$

Algorithm 2 Weighted tensor singular value thresholding

Input: $\mathcal{W}, \mathcal{Q}_1, \mu$
1: $\mathcal{P} = \mathcal{W} - \frac{\mathcal{Q}_1}{\mu}$
2: $\hat{\mathcal{P}} = \text{fft}(\mathcal{P}, [], 3)$;
3: **for** $i = 0, 1, \ldots, [\frac{z+1}{2}]$
4: $\quad [\hat{\mathcal{U}}(:,:,i), \hat{\mathcal{S}}(:,:,i), \hat{\mathcal{V}}(:,:,i)] = \text{SVD}(\hat{\mathcal{P}}(:,:,i))$;
5: $\quad \omega = \frac{1}{diag(\hat{\mathcal{S}}(:,:,i)) + \epsilon}$;
6: $\quad \hat{\mathcal{S}}_{new}(:,:,i) = max(\hat{\mathcal{S}}(:,:,i) - \frac{\omega}{\mu}, 0)$;
7: **end for**
8: **for** $i = [\frac{z+1}{2} + 1], \ldots, z$
9: $\quad \hat{\mathcal{U}}(:,:,i) = \text{conj}(\hat{\mathcal{U}}(:,:,z-i+2))$;
10: $\quad \hat{\mathcal{S}}_{new}(:,:,i) = \text{conj}(\hat{\mathcal{S}}_{new}(:,:,i))$;
11: $\quad \hat{\mathcal{V}}(:,:,i) = \text{conj}(\hat{\mathcal{V}}(:,:,z-i+2))$;
12: **end for**
13: $\mathcal{U} = \text{ifft}(\hat{\mathcal{U}}, [], 3), \mathcal{S}_{new} = \text{ifft}(\hat{\mathcal{S}}_{new}, [], 3), \mathcal{V} = \text{ifft}(\hat{\mathcal{V}}, [], 3)$,
14: $\mathcal{Z} = \mathcal{U} * \mathcal{S}_{new} * \mathcal{V}^*$;

The unconstrained augmented Lagrangian function of Eq.(13) is converted to the following problem:

$$\min_{\mathcal{Z},\mathcal{W},\mathcal{E}} ||\mathcal{Z}||_{w,*} + \lambda ||\mathcal{E}||_{F,1} + <\mathcal{Q}_1, \mathcal{Z} - \mathcal{W}> + \frac{\mu}{2}||\mathcal{Z} - \mathcal{W}||_F^2$$
$$<\mathcal{Q}_2, \mathcal{X} - \mathcal{A} * \mathcal{W} - \mathcal{E}> + \frac{\mu}{2}||\mathcal{X} - \mathcal{A} * \mathcal{W} - \mathcal{E}||_F^2$$
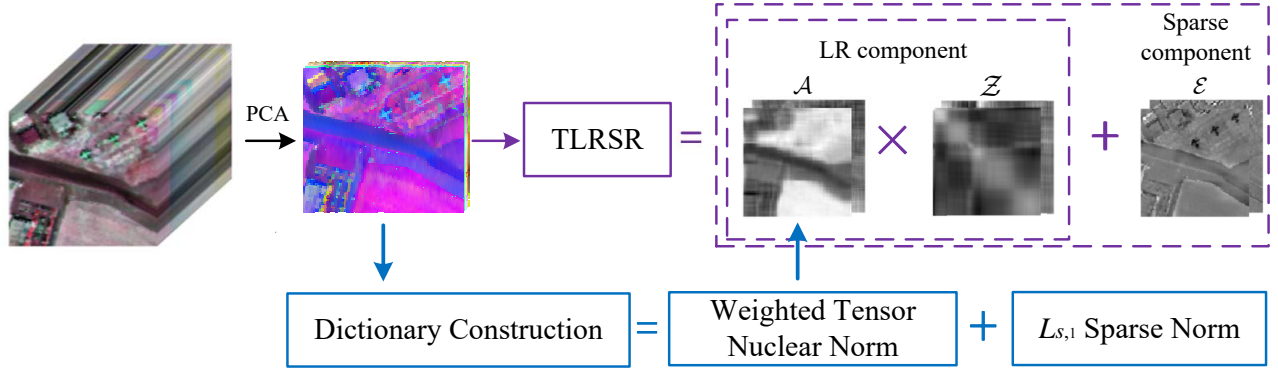
Fig. 1: The schematic diagram of the proposed PCA-TLRSR model for HSI anomaly detection.

(14)

where $\mathbf{Q}_1$ and $\mathcal{Q}_2$ are the Lagrange multipliers and μ is an auto-adjusted penalty parameter. We solve problem (14) by updating one variable with others fixed.

1) Update $\mathcal{Z}$: Exact all $\mathcal{Z}$-related items from (14):

$$\min_{\mathcal{Z}} ||\mathcal{Z}||_{w,*} + \frac{\mu}{2}||\mathcal{Z} - (\mathcal{W} - \frac{\mathcal{Q}_1}{\mu})||_F^2 \qquad (15)$$

This sub-problem is solved by the following weighted tensor singular value thresholding method, which is shown in Algorithm 2.

2) Update $\mathcal{E}$: The optimization function with respect to $\mathcal{E}$ is

$$\min_{\mathcal{E}} \lambda||\mathcal{E}||_{F,1} + \frac{\mu}{2}||\mathcal{E} - (\mathcal{X} - \mathcal{A} * \mathcal{W} + \frac{\mathcal{Q}_2}{\mu})||_F^2 \qquad (16)$$

This sub-problem has a close-form solution [55] and the k-th band of $\mathcal{E}$ is:

$$\mathcal{E}(:,:,k) = \begin{cases} \dfrac{||\mathcal{T}(:,:,k)||_F - \lambda}{||\mathcal{T}(:,:,k)||_F}\mathcal{T}(:,:,k), & \lambda < ||\mathcal{T}(:,:,k)||_F \\ 0, & otherwise \end{cases} \qquad (17)$$

where $\mathcal{T} = \mathcal{X} - \mathcal{A} * \mathcal{W} + \frac{\mathcal{Q}_2}{\mu}$.

3) Update $\mathcal{W}$ by minimizing:

$$\min_{\mathcal{W}} \frac{\mu}{2}||\mathcal{Z} - \mathcal{W} + \frac{\mathcal{Q}_1}{\mu}||_F^2 + \frac{\mu}{2}||\mathcal{X} - \mathcal{A} * \mathcal{W} - \mathcal{E} + \frac{\mathcal{Q}_2}{\mu}||_F^2 \qquad (18)$$

By setting the derivative of (18) with respect to $\mathcal{W}$ to zero and calculating a simple reduction, we obtain the following solution:

$$\mathcal{W} = (\mathcal{A}^* * \mathcal{A} + \mathcal{I})^{-1}(\mathcal{Z} + \frac{\mathcal{Q}_1}{\mu} + \mathcal{A}^* * (\mathcal{X} - \mathcal{E} + \frac{\mathcal{Q}_2}{\mu})) \qquad (19)$$

The whole procedure of the proposed method is summarized in Algorithm 3. In the beginning, the dimension reduction of HSI is carried out via classical PCA processing [56], which is proved to reveal that only the first few PCs contain significant information for HSI classification [57]. As the first step of PCA, the mean values $\mathbf{m}_i$ of each HS band vector $\mathbf{x}'_i =$

$\text{vec}(\mathcal{X}'(:,:,i)) \in \mathbb{R}^{hv}$ are subtracted to provide an appropriate normalization. The covariance of every two standardized bands $\overline{\mathbf{x}'}_i = \mathbf{x}'_i - \mathbf{m}_i$ is calculated to measure their correlation. The HS data is transformed into an uncorrelated orthogonal basis set, which is obtained by the eigenvectors of the covariance matrix $\mathbf{C}_x$. The eigenvalue decomposition of $\mathbf{C}_{x'}$ contains into two parts: the diagonal matrix $\mathbf{B}$ of eigenvalues, and the orthonormal matrix of eigenvectors $\mathbf{A}$, i.e., $\mathbf{C}_{x'} = \mathbf{ABA}^{\mathrm{T}}$. The dimension reduced image is converted by $\mathbf{X} = \mathbf{X}'\mathbf{A}'$, where $\mathbf{X}'$ is composed of $\mathbf{x}'_i$ and $\mathbf{A}'$ denotes the first K eigenvectors corresponding to the first K large eigenvalues. Represented by fewer bands with less redundancy, the PCA image $\mathcal{X} = fold(\mathbf{X}, 3)$ considers a majority of inherent information.

Due to the weighted TNN regularization, the solving process of problem (14) is non-convex. In [33], the authors just give a weak convergence result of WNNM for matrix completion. The theoretical convergence of the used ADMM algorithm is difficult to prove with more than two block variables [58]. Therefore, we define the convergence condition as follows

$$\max \left\{ \begin{array}{l} ||\mathcal{W}_k - \mathcal{W}_{k-1}||_\infty, \\ ||\mathcal{Z}_k - \mathcal{Z}_{k-1}||_\infty, \\ ||\mathcal{E}_k - \mathcal{E}_{k-1}||_\infty, \\ ||\mathcal{W}_k - \mathcal{Z}_k||_\infty, \\ ||\mathcal{X} - \mathcal{A} * \mathcal{W}_k - \mathcal{E}_k||_\infty \end{array} \right\} \le \epsilon \qquad (20)$$

and opt to analyze the convergence of our model empirically, which is shown in Section IV-B. Furthermore, μ is determined by $\min(\gamma\mu, \mu_{max})$ with $\gamma = 1.1$ in each iteration. This strategy of updating the variable μ has been supported by the recent ADMM studies [59], [60] to facilitate the ADMM convergence of non-convex optimization.

B. Dictionary construction for TLSR

A qualified dictionary $\mathcal{A}$ plays a critical role in the proposed algorithm. The simple way to design a dictionary is to directly use the HSI tensor $\mathcal{X}$ of its own, which is similar to the strategy of LRR [55]. However, the HSI tensor $\mathcal{X}$ contains anomalies, and the dictionary $\mathcal{A}$ should be composed of the LR background part. The above strategy hardly considers the real case of HSI anomaly detection, easily leading to worse

---

Algorithm 3 PCA-TLRSR for HSI anomaly detection

Input: Input HSI data $\mathcal{X}'$, dictionary $\mathcal{A}$, λ
  1: **Initializtion**: $\mathcal{W} = \mathcal{Z} = \mathcal{Q}_1 = 0$, $\mathcal{E} = \mathcal{Q}_2 = 0$, $i = 0$, $\mu = 1e-5$, $\mu_{max} = 1e+8$, $\gamma = 1.1$, $\epsilon = 10^{-6}$, maxiter=100;
  2: **while** $0 \leq i \leq$ maxiter or meet the convergence conditions
  3:    $i = i + 1$;
  4:    start a PCA pre-processing step: $\mathcal{X}$=PCA($\mathcal{X}'$);
  5:    update $\mathcal{Z}$ by solving Eq. (15) via weighted singular value thresholding method;
  6:    update $\mathcal{E}$ by calculating Eq. (17);
  7:    update $\mathcal{W}$ by calculating Eq. (19);
  8:    update the multipliers
        $\mathcal{Q}_1 = \mathcal{Q}_1 + (\mathcal{Z} - \mathcal{W})$
        $\mathcal{Q}_2 = \mathcal{Q}_2 + (\mathcal{X} - \mathcal{A} * \mathcal{Z} - \mathcal{E})$
  9:    $\mu = \min(\gamma\mu, \mu_{max})$;
 10:    check the convergence conditions (20)
 11: **end while**
 12: obtain the optimal solution $\mathcal{E}^*$;
 13: compute the anomaly detection map **M**:
     $\mathbf{M}(i,j) = \sqrt{\sum_{k=1}^{z} |\mathcal{E}^*(i,j,k)|^2}$;
Output: anomaly detection map **M**.

---

performances. The dictionary $\mathcal{A}$ is expected to be related to a pure background with anomalies. To address this issue, we adopt the $L_{F,1}$-regularized RPCA to retain the background information:

$$\min_{\mathcal{L},\mathcal{S}} ||\mathcal{L}||_{w,*} + \lambda'||\mathcal{S}||_{F,1}$$
$$\text{s.t. } \mathcal{X} = \mathcal{L} + \mathcal{S} \tag{21}$$

We solve this problem by ADMM with high computational efficiency. The augmented Lagrangian function of problem (21) is:

$$F = ||\mathcal{L}||_{w,*} + \lambda'||\mathcal{S}||_{F,1} + <\mathcal{P}, \mathcal{X} - \mathcal{L} - \mathcal{S}>$$
$$+ \frac{\beta}{2}||\mathcal{X} - \mathcal{L} - \mathcal{S}||_F^2 \tag{22}$$

where $\mathcal{P}$ is a Lagrangian multiplier. $\mathcal{L}$ and $\mathcal{S}$ are updated by minimizing F alternatively. Two sub-problems have closed-form solutions, whose procedures are similar to steps 2) and 3) of Algorithm 3. The LR background part $\mathcal{L}$ is obtained for the dictionary $\mathcal{A}$ of the follow-up TLRSR, i.e., $\mathcal{A}=\mathcal{L}$. Based on the weighted TNN and the $L_{F,1}$ sparse norm, the process for dictionary construction also avoids remodeling a 3-D HSI data into a 2-D matrix and preserves the inherent spatial-spectral structure.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we carry out real-data experiments to evaluate the performances of the proposed PCA-TLRSR for HSI anomaly detection. Three real-world HSI data sets are employed, including the Urban data set, the San Diego data set, and the airport data set.

The first data was collected from the hyperspectral digital imagery collection experiment (HYDICE) with a 10-nm spectral solution and 1-m spatial solution. An image area of 80×100 with 175 bands is selected for evaluation as shown in Fig. 2(a). In this image scene, anomalies with 21 pixels consist of vehicles and roofs in different backgrounds, whose specific positions are presented in Fig. 2(b).

The second HSI data named the San Diego data set was captured by the airborne visible/ infrared imaging spectrometer (AVIRIS), containing $100 \times 100 \times 189$ pixels. In this scene, normal classes of grass, roofs, and shadows are background mostly; in contrast to the above background, anomalies are composed of three man-made airplanes. The real-scene image and corresponding ground-truth map are displayed in Fig. 6 (a) and (b), respectively.

The last hyperspectral data was the Airport data set, whose sample images of size 100×100 were manually obtained by the AVIRIS over Los Angeles and Gulfport, USA. The environment for visualizing images (ENVI) is used to label the anomaly maps of the Airport images. Things on the Earth's surface like roads and buildings are regarded as background. Oppositely, anomalies are flights of different heights. Fig. 4 (a) and (b) give the sample images and reference detection maps.

A. Detection performance

This experiment aims at testing the detection performances of the proposed PCA-TLRSR in comparison with other baselines, including two RX methods (i.e., classic global RX [10] and discrete wavelet transform (DWT) RX [61]); four LRaSRMD methods (i.e., RPCA [19], LSMAD [23], LR and sparse representation (LRSAR) [24], graph and TV regularized LRR (GTVLRR) [27]); a tensor representation method TRPCA [44]; and a deep learning method (i.e., DeCNN-AD) [35]). Except for the visual observation of resulted anomaly maps, we adopt the receiver operating characteristic (ROC) curve [62] and the area under the ROC curve (AUC) [63] to quantitatively assess the detection accuracy of a series of compared methods. The ROC curve plots the varying relationship of the probability of detection (PD) and false alarm rate (FAR) for extensive possible thresholds. The area under this curve is calculated as AUC, whose ideal result is 1.

1) HYDICE-Urban: The detection map of different methods is visually observed in Fig. 2. Although the bridge contours of the map center are retained well by TRPCA, most important anomalies are unfortunately ignored. Global RX, DWT RX, RPCA, and LSMAD obtain similar results, but hardly detect the anomalies in the sea. LRSAR, GTVLRR, and DeCNN-AD have similar anomaly maps, which are produced better than the former algorithms. The proposed PCA-TLRSR achieves the clearest detection map among all compared algorithms. Fig. 3 presents the ROC curves of different anomaly detectors for the HYDICE-Urban data set. Consistent with the visual observations, TRPCA has the lowest value among all the approaches. RPCA and LSMAD generate better performances than global RX and DWT RX. Although the PD of PCA-TLRSR is lower than LRSAR, GTVLRR and DeCNN-AD at the beginning, the proposed PCA-TLRSR provides the best
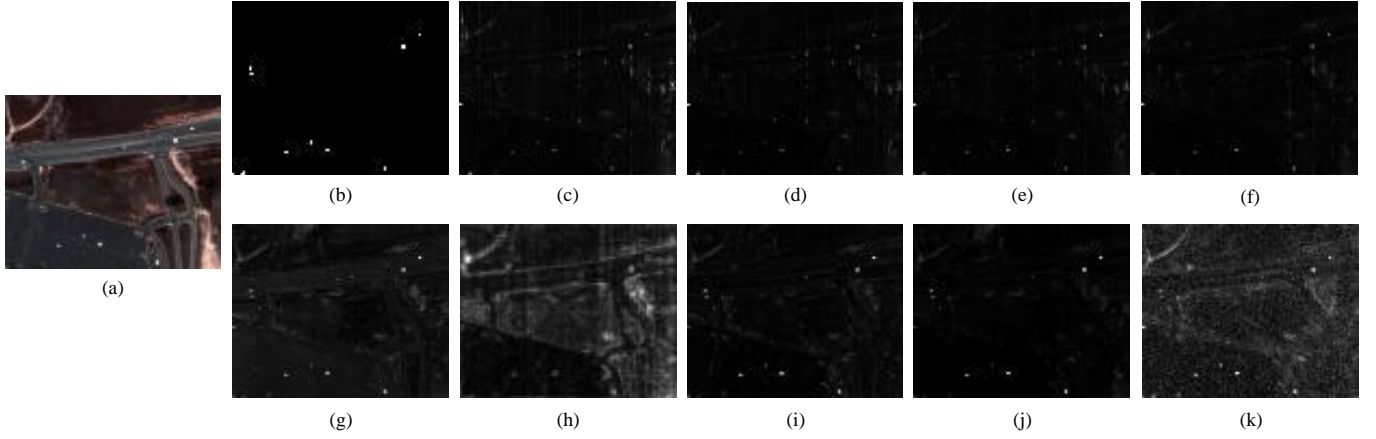
Fig. 2: Detection maps of all compared methods in HYDICE-Urban data. (a) Color composite of HYDICE-Urban scene, (b) reference map, (c) global RX, (d) DWT RX, (e) RPCA, (f) LSMAD, (g) LRSAR, (h) TPCA, (i) GTVLRR, (j) DeCNN-AD, (k) PCA-TLRSR.

TABLE II: AUC values of all compared algorithms for three data sets.

| HSI data sets	global RX	DWT RX	RPCA	LSMAD	LRSAR	TPCA	GTVLRR	DeCNNAD	PCA-TLRSR
HYDICE-Urban	0.9855	0.9899	0.9842	0.9909	0.9911	0.8242	0.9894	<u>0.9935</u>	**0.9941**
San Diego	0.8884	0.9499	0.9165	0.9772	0.9878	0.8849	0.9682	<u>0.9901</u>	**0.9957**
Airport-1	0.8219	0.8676	0.8089	0.8311	0.7529	0.8023	0.9003	0.8413	**0.9478**
Airport-2	0.8403	0.9101	0.8431	0.9197	0.8672	0.8891	0.8755	<u>0.9512</u>	**0.9697**
Airport-3	0.9287	0.9331	0.9275	0.9356	0.9153	0.9297	0.9287	<u>0.9394</u>	**0.9574**
Airport-4	0.9525	0.9765	0.9627	0.9859	0.9781	0.9432	0.9525	<u>0.9897</u>	**0.9943**
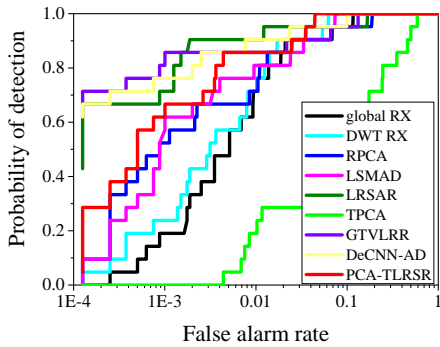
Fig. 3: ROC curves obtained by different methods for the HYDICE-Urban dataset.

performance when FAR is equal and higher than 0.03. The first row of Table II gives direct evidence that PCA-TLRSR achieves the highest AUC score in bold. The second one gained by DeCNN-AD is underline.

2) San Diego: Fig. 6 shows the detection results of all compared methods. Three flights are quite prominent in the top right corner of Fig. 6 (b). Global RX and RPCA fail to find the anomaly contours. DWT RX and TRPCA discover the locations of three anomalies, but they hardly suppress the background information. Compared to DWT RX, TRPCA, and GTVLRR, the deep learning-based approach: DeCNN-AD have better detection performances. However, the detection

map generated by the proposed PCA-TLRSR is closest to the original map among all maps. The edges of three flights are effectively conducted by the proposed algorithm. The ROC curves of PCA-TLRSR and the other compared methods are shown in Fig. 7 for a specific illustration. DeCNN-AD and three LRaSAM-based methods, i.e., LSMAD, LRSAR, and GTVLRR have higher PD values than RX-based ones. The PD of PCA-TLRSR reaches 1.0 even when FAR is 0.01, whereas most of the others are lower than 0.4. Furthermore, the AUCs of eight methods are also computed to quantitatively evaluate the detection performances, which are shown in the second row of Table II. The AUC values of global RX and TRPCA are similar and less than 0.9. DeCNN-AD and LRSAR gain the second and third largest AUCs, respectively, but they are lower than the proposed PCA-TLRSR.

3) Airport: Four airport scenes are available in this data set, where several airplanes are found in different positions. Four ground-truth detection maps are displayed in Fig. 4 (b). Here, we take airport-1 as an example to demonstrate the detection performance of PCA-TLRSR and to compare it with the other algorithms. The first four methods rarely detect the original anomalies in the scene center. Although some flights are located in the detection maps of LSMAD, LRSAR, and GTVLRR, the backgrounds surrounding these anomalies inevitably remain. DeCNN-AD suppresses the backgrounds better than the former mentioned approaches. But PCA-TLRSR obtains much clearer shapes of anomaly objects than the other algorithms, even when the object size is small. The ROC curves and corresponding AUC values are given in Fig.
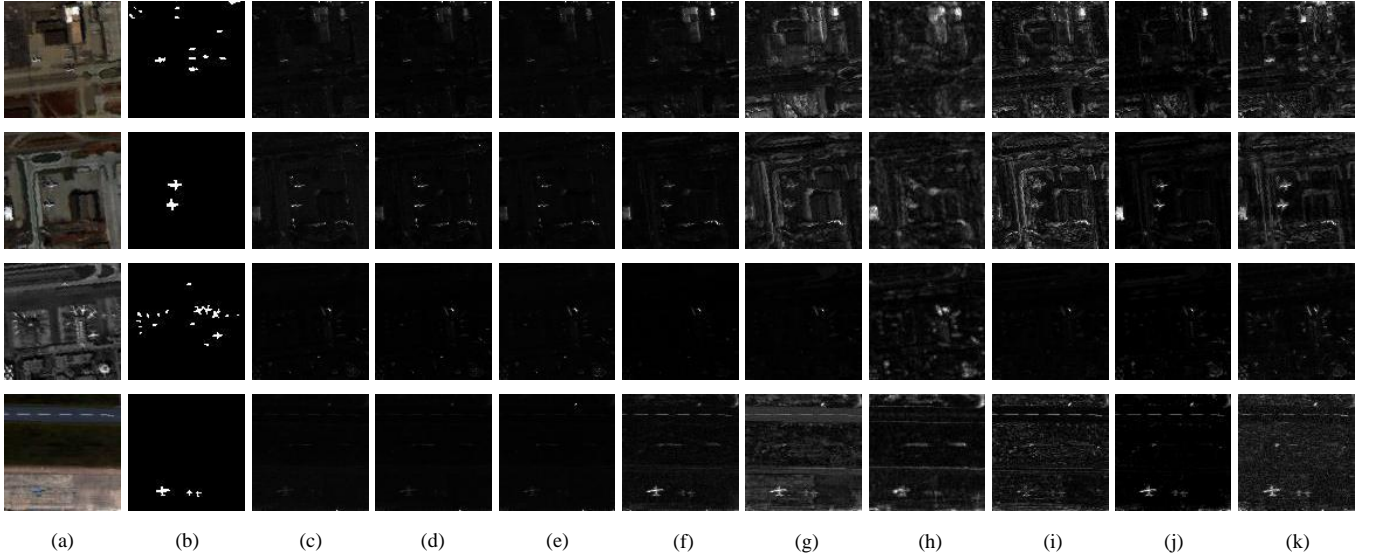
Fig. 4: Detection maps of all compared methods in Airport-1 data. (a) Color composite of airport scene, (b) reference map, (c) global RX, (d) DWT RX, (e) RPCA, (f) LSMAD, (g) LRSAR, (h) TPCA, (i) GTVLRR, (j) DeCNN-AD
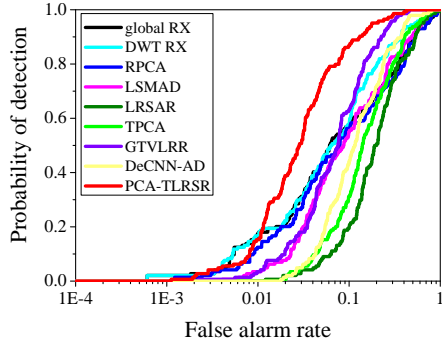, (k) PCA-TLRSR.

Fig. 5: ROC curves obtained by different methods for the airport-1 dataset.

5 and Table II. Although global RX and DWT RX exhibit slightly higher PDs than the others when FAR is less than 0.01, they have worse performances with the increase of FAR. The proposed PCA-TLRSR yields higher PDs in terms of the overall curve comparison. DeCNN-AD achieves the second-best AUCs except for Airport-1 data set. PCA-TLRSR derives the highest AUC scores for four images. Especially, the AUCs of PCA-TLRSR have a great improvement for Airport-1 and Airport-2, compared the other LRaSRMD methods.

B. Discussion

1) Effects of PCA and dictionary construction:

This section analyzes the contributions of the PCA pre-processing step and the dictionary training step, over AUC scores on each data set. Fig. 8 presents the computing time and the AUC values of the proposed method with and without the PCA pre-processing step. Due to the dimension reduction step, the running times of PCA-TLRSR are dramatically reduced

on different data sets, compared with TLRSR. Meanwhile, the detection performances of PCA-TLRSR are kept well and even improved for Urban and Airport-1, and 3 data sets. These data sets contain more complex backgrounds and more anomaly objects than the others. The backgrounds include manual bridges, buildings, and airport hardstands mixed with the anomalies, which may have similar spatial structure or spectral signatures with some of the background man-made materials. With the aid of PCA, the potential features of mixture pixels are explored and distinguished from backgrounds by learning tensor LR representation. The PCA step makes up for the spectral redundancy of original HSIs and benefits the improvement of detection performances in most scenarios. Meanwhile, the computational burden of tensorial operations is reduced efficiently via the PCA pre-processing step.

To illustrate the effectiveness of the dictionary construction strategy, we test the computing time and AUCs of PCA-TLRSR with and without this training step. The result comparison is given in Fig. 9. The computing times slightly rise with dictionary construction, while the AUC scores of PCA-TLRSR increase significantly with the dictionary training step, especially for Airport-1, 2, and 3. Obviously, dictionary construction plays a critical role in improving the detection performance.

2) Effects of weighted nuclear norm used in the PCA-TLRSR framework: We provide evidence for the effectiveness of accounting for the weighted nuclear norm used in a PCA-TLRSR framework via illustrative examples. Fig. 10 shows the computing time and the AUC values of PCA-TLRSR with TNN and with weighted TNN. The proposed methods with the standard nuclear norm and weighted nuclear norm have similar computing times. Weighted TNN used in the PCA-TLRSR framework benefits the improvement of detection performances on each data set. Compared with the
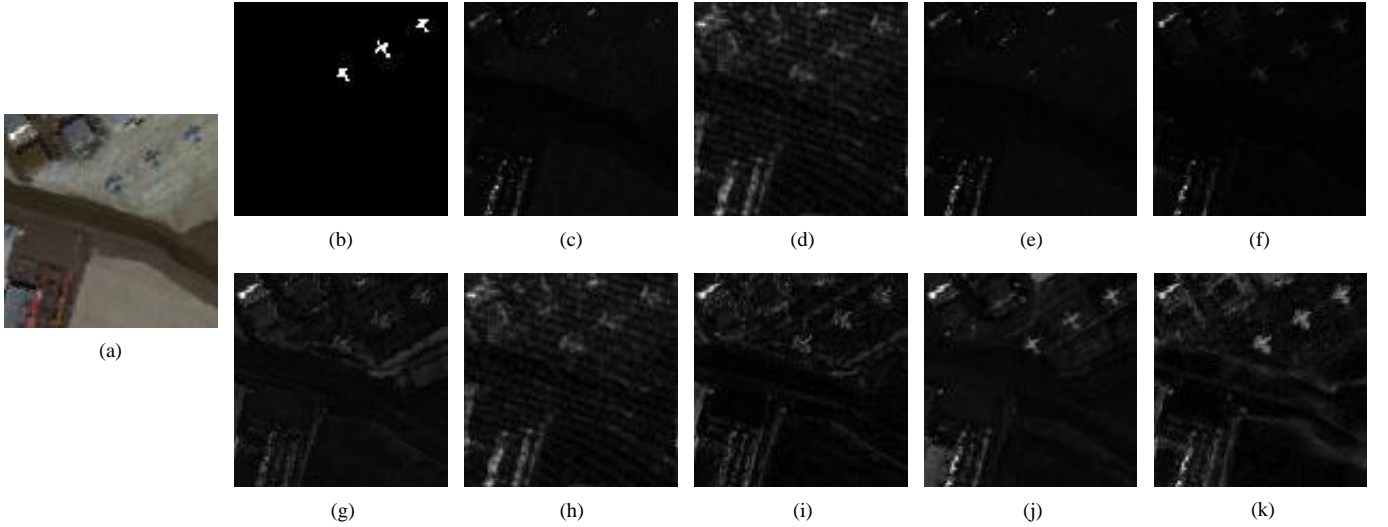
Fig. 6: Detection maps of all compared methods in San Diego data. (a) Color composite of San Diego scene, (b) reference map, (c) global RX, (d) DWT RX, (e) RPCA, (f) LSMAD, (g) LRSAR, (h) TPCA, (i) GTVLRR, (j) DeCNN-AD, (k) PCA-TLRSR.
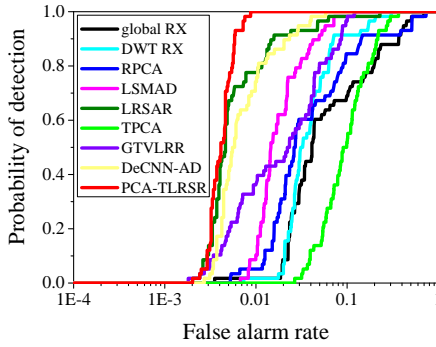

Fig. 7: ROC curves obtained by different methods for the San Diego dataset.

Fig. 8: Computing time and AUC values of TLRSR and PCA-TLRSR.

standard TNN, the weighted TNN adaptively assigns weights on different singular values.

3) Parameter analysis: We discuss the robustness of two parameters λ and λ' involved in the PCA-TLRSR on three data sets. Both parameters λ and λ' are chosen from the set {0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5} and one of them is adjusted with the other being fixed. Fig. 11 shows the contribution and sensitivity of two parameters to AUC values. It is apparent that the PCA-TLRSR solver is entirely robust to parameters λ and λ'. In Fig. 11 (a), three AUC value curves with respect to λ rise, when λ lies in the range of [0.001,0.01]. After peaking at $\lambda = 0.01$, these curves show downward trends. Similarly, the best performances of the proposed method are obtained when $\lambda' = 0.05$ in Fig. 11 (b). Therefore, we set $\lambda = 0.01$ and $\lambda' = 0.05$ in experiments.

4) Computational complexity and running time: The computational complexity of the proposed algorithm is discussed in this section: 1) Solving the sub-problem (15) depends on t-SVD operator, which takes about $O(v^3z)$ iterations. 2) For Eq. (17), tensor product spends the most time and
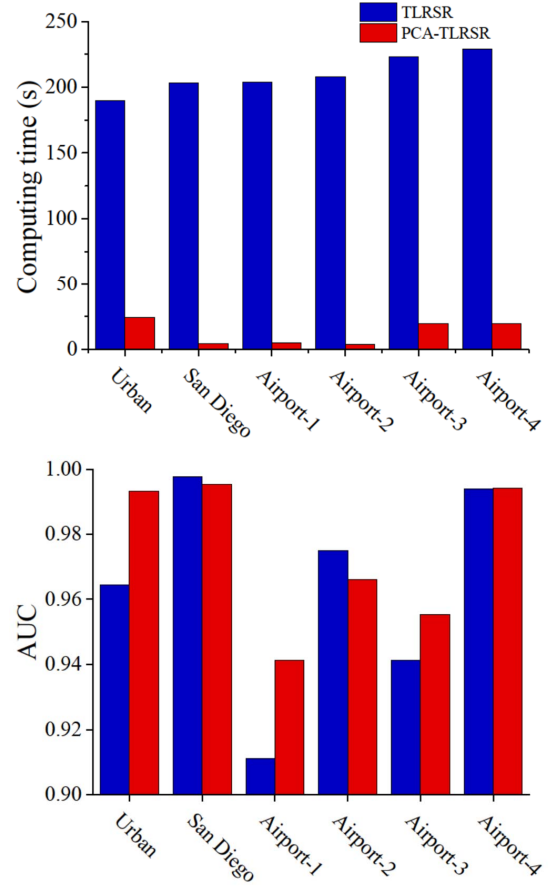
requires $O(hv^2z + v(h + v)z\log(z))$ time. 3) Updating $\mathbf{W}$ needs computing its closed-form solution, which requires $O(hv^2z + v(h + v)z\log(z))$. Therefore, the cost of the Algorithm 3 is in total $O(v^3z + 2hv^2z + 2v(h + v)z\log(z))$.

TABLE III: Running time comparison (s) of different algorithms.

HSI data sets	global RX	DWT RX	RPCA	LSMAD	LRSAR	TPCA	GTVLRR	DeCNN-AD	PCA-TLRSR
HYDICE-Urban	0.148	7.636	4.512	9.072	44.117	37.899	317.443	25.562	24.277
San Diego	0.171	10.200	4.965	14.070	39.058	52.490	339.548	27.527	4.502
Airport-1	0.188	10.476	11.083	15.297	18.283	59.798	307.961	32.554	5.115
Airport-2	0.175	10.506	10.926	14.137	18.135	61.942	341.584	30.698	4.076
Airport-3	0.180	10.810	10.114	15.663	18.734	58.519	295.933	32.817	19.647
Airport-4	0.177	10.261	5.448	12.682	15.471	56.437	340.799	28.476	19.490
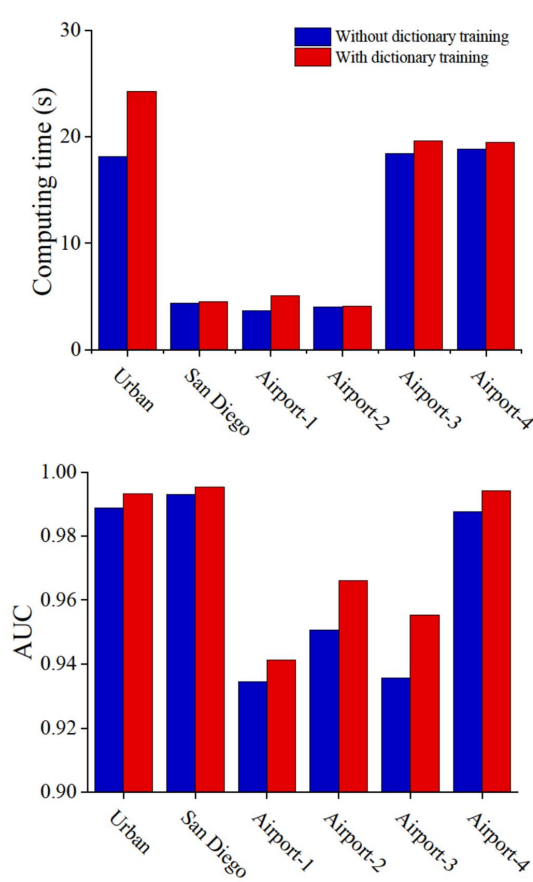
Fig. 9: Computing time and AUC values of PCA-TLRSR with and without a dictionary construction strategy.
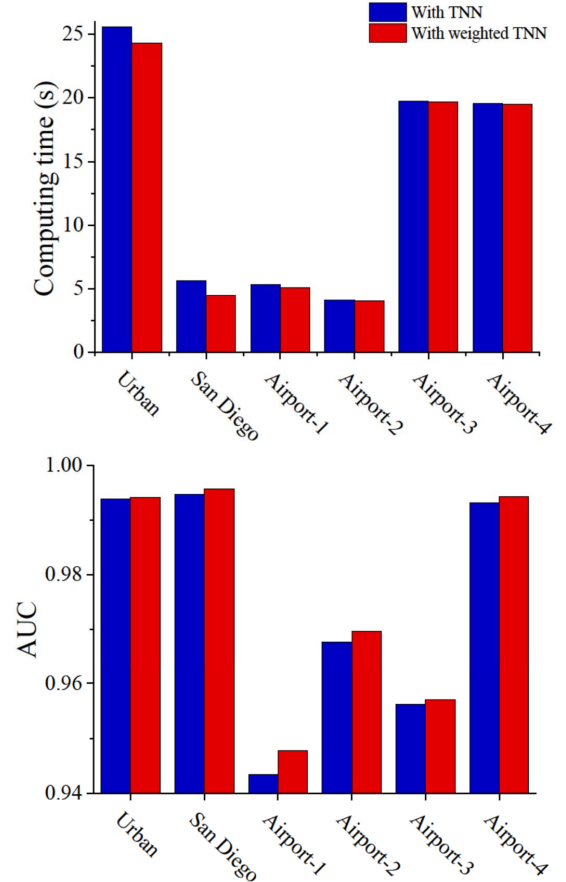
Fig. 10: Computing time and AUC values of PCA-TLRSR with TNN or with weighted TNN.

The running times of different methods on HSI anomaly detection experiments are reported in Table III. Global RX is the fastest one among all the comparing methods, but it hardly obtains satisfactory detection performances. PCA-TLRSR has similar computing times with DWT RX, RPCA, and LSMAD, and even is the second-best one for San Diego, Airport-1, and 2. As a deep learning-based method, DeCNN-AD is faster than TPCA and GTVLRR on each data set. Due to the consideration of low rankness and spatial-spectral smoothness, GTVLRR is more time-consuming than the other algorithms. With the assistance of the PCA step, the proposed algorithm does not derive superior performances but also spends proper running time.

5) Convergence analysis: The numerical convergence is

investigated on three data sets. Fig. 12 displays the relative errors given in Eq. (20) versus iterations. It can be spotted that as the number of iterations increases, the relative errors keep stable at the beginning and then monotonically decrease, finally approaching zero.

V. CONCLUSION

In this paper, we have proposed a novel PCA-TLRSR model for HSI anomaly detection. A TLR framework is used to preserve the inherent HSI 3-D structure. A PCA pre-processing step is served for acceleration of TLRSR, which is caused by various tensorial operations. A tensor dictionary strategy is designed to extract the LR background part, which is regarded as the dictionary of TLRSR. The weighted tensor nuclear
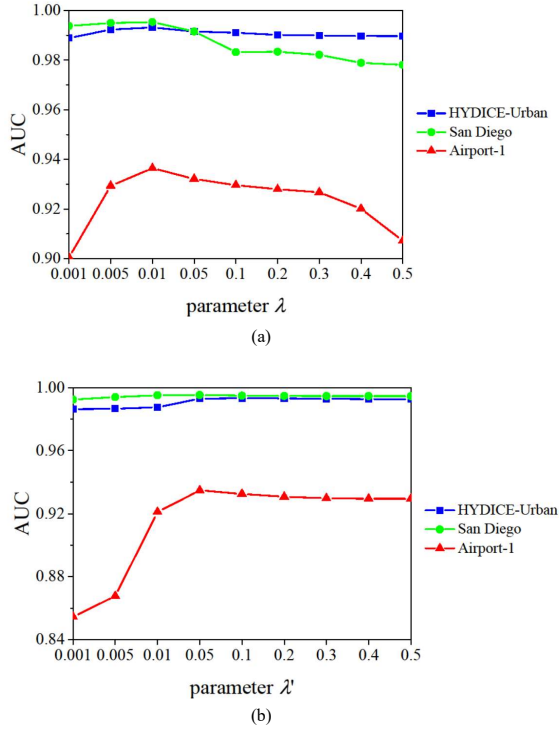
(a)

(b)

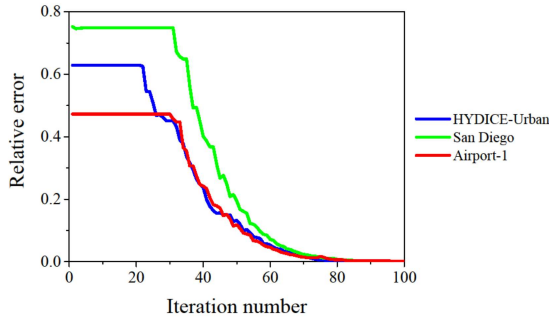**Fig. 11:** AUC value curves with respect to (a) λ, and (b) λ'.

Fig. 12: Relative errors versus iterations on three data sets.

norm and the $L_{F,1}$ norm are efficient to separate the TLR and the sparse tensor components. We develop ADMM for the optimization solving procedure. Experimental results on different HSI data sets demonstrate the superior performances of our proposed methods.

For further work, one novel convolution nuclear norm technique can be explored to exploit the high-order structure of HSIs for anomaly detection [64]. As a natural generalization of DFT, this convolution nuclear norm are more suitable for solving tensor decomposition problem than tensor nuclear norms, which undoubtedly has great significance in future research.

REFERENCES

[1] D. Hong, L. Gao, N. Yokoya, J. Yao, J. Chanussot, Q. Du, and B. Zhang, "More diverse means better: Multimodal deep learning meets remote-sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, pp. 4340–4354, May 2021.

[2] F. Liu and Q. Wang, "A sparse tensor-based classification method of hyperspectral image," *Signal Process.*, vol. 168, p. 107361, 2020.

[3] J. M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 5, pp. 354–379, Apr. 2012.

[4] D. Hong, N. Yokoya, J. Chanussot, and X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Trans. Image Process.*, vol. 28, no. 4, pp. 1923–1938, 2019.

[5] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Joint reconstruction and anomaly detection from compressive hyperspectral images using maha-lanobis distance-regularized tensor RPCA," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, pp. 2919–2930, May 2018.

[6] X. Wu, D. Hong, J. Tian, J. Chanussot, W. Li, and R. Tao, "Orsim detector: A novel object detection framework in optical remote sensing imagery using spatial-frequency channel features," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, 2019.

[7] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, pp. 6–36, Jun. 2013.

[8] D. Hong, W. He, N. Yokoya, J. Yao, L. Gao, L. Zhang, J. Chanussot, and X. Zhu, "Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, pp. 52–87, Jun. 2021.

[9] I. 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, 1990.

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

[11] Y. Taitano, B. Geier, and K. Bauer, "A locally adaptable iterative rx detector," *EURASIP J. Adv. Signal Process.*, vol. 341908, no. 11, 2010.

[12] A. Schaum, "Joint subspace detection of hyperspectral targets," in *Proc. IEEE Aerospace Conf.*, vol. 3, pp. 1818–1824, 2004.

[13] S. Matteoli, M. Diani, and G. Corsini, "Improved estimation of local background covariance matrix for anomaly detection in hyperspectral images," *Opt. Eng.*, vol. 49, no. 4, pp. 1–16, 2010.

[14] 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. Select. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, 2014.

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

[16] 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, 2016.

[17] C. Zhao, X.-F. Yao, and Y. Yan, "Modified kernel rx algorithm based on background purification and inverse-of-matrix-free calculation," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 4, pp. 544–548, 2017.

[18] T. E. Smetek and K. W. Bauer, "Finding hyperspectral anomalies using multivariate outlier detection," in *Proc. IEEE Aerosp. Conf.*, pp. 1–24, 2007.

[19] S.-Y. Chen, S. Yang, K. Kalpakis, and C.-I. Chang, "Low-rank decomposition-based anomaly detection," in *Proc. SPIE*, vol. 8743, pp. 1–24, 2013.

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

[21] M. Farrell and R. Mersereau, "On the impact of covariance contamination for adaptive detection in hyperspectral imaging," *IEEE Signal Process. Lett.*, vol. 12, no. 9, pp. 649–652, 2005.

[22] N. Billor, A. S. Hadi, and P. F. Velleman, "Bacon: Blocked adaptive computationally efficient outlier nominators," *Comput. Statist. Data Anal.*, vol. 34, no. 3, pp. 279–298, 2000.

[23] 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, 2016.

[24] 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, 2016.

[25] Y. Niu and B. Wang, "Hyperspectral anomaly detection based on low-rank representation and learned dictionary," *Remote Sens.*, vol. 8, no. 4, p. 289, 2016.

[26] K. Tan, Z. Hou, D. Ma, Y. Chen, and Q. Du, "Anomaly detection in hyperspectral imagery based on low-rank representation incorporating a spatial constraint," *Remote Sens.*, vol. 11, no. 13, 2019.

[27] 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, 2020.

[28] D. Hong, L. Gao, J. Yao, B. Zhang, P. Antonio, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, pp. 5966–5978, Jul. 2021.

[29] Y. Xu, B. Du, L. Zhang, and S. Chang, "A low-rank and sparse matrix decomposition- based dictionary reconstruction and anomaly extraction framework for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 7, pp. 1248–1252, 2020.

[30] 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.*, pp. 1–10, 2020.

[31] S. Wang, X. Wang, Y. Zhong, and L. Zhang, "Hyperspectral anomaly detection via locally enhanced low-rank prior," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 10, pp. 6995–7009, 2020.

[32] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-d transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, 2007.

[33] S. Gu, Q. Xie, D. Meng, W. Zuo, X. Feng, and L. Zhang, "Weighted nuclear norm minimization and its applications to low level vision," *Int. J. Comput. Vis.*, vol. 121, no. 2, pp. 183–208, 2017.

[34] K. Zhang, W. Zuo, and L. Zhang, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, 2018.

[35] 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, 2021.

[36] M. Wang, Q. Wang, and J. Chanussot, "Tensor low-rank constraint and l_0 total variation for hyperspectral image mixed noise removal," *IEEE J. Select. Topics Signal Process.*, vol. 15, no. 3, pp. 718–733, 2021.

[37] Y. Chen, W. He, N. Yokoya, and T.-Z. Huang, "Hyperspectral image restoration using weighted group sparsity-regularized low-rank tensor decomposition," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3556–3570, 2020.

[38] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Hyperspectral images super-resolution via learning high-order coupled tensor ring representation," *IEEE Trans. Neural Netw. Learn Syst.*, vol. 31, no. 11, pp. 4747–4760, 2020.

[39] 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, 2019.

[40] X. Li, Y. Yuan, and Q. Wang, "Hyperspectral and multispectral image fusion via nonlocal low-rank tensor approximation and sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 550–562, 2021.

[41] D. Hong, Z. Han, J. Yao, L. Gao, B. Zhang, A. Plaza, and J. Chanussot, "Spectralformer: Rethinking hyperspectral image classification with transformers," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, Nov. 2022.

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

[43] 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, 2016.

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

[45] F. Qin, Z. Wu, Y. Xu, H. Liu, Y. Zhang, and Z. Wei, "Hyperspectral anomaly detection based on low rank and sparse tensor decomposition," in *IEEE Int. Geosci. Remote Sens. Symp.*, pp. 2240–2243, 2019.

[46] S. Song, H. Zhou, L. Gu, Y. Yang, and Y. Yang, "Hyperspectral anomaly detection via tensor- based endmember extraction and low-rank decomposition," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1772–1776, 2020.

[47] R. Zhao, Q. Wang, J. Fu, and L. Ren, "Exploiting block-sparsity for hyperspectral kronecker compressive sensing: A tensor-based bayesian method," *IEEE Trans. Image Process.*, vol. 29, pp. 1654–1668, 2020.

[48] 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, 2021.

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

[50] 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, 2013.

[51] 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*, pp. 3842–3849, 2014.

[52] 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," pp. 5249–5257, 2016.

[53] M. Wang, Q. Wang, J. Chanussot, and D. Li, "Hyperspectral image mixed noise removal based on multidirectional low-rank modeling and spatialspectral total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 488–507, 2021.

[54] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends® Mach. Learn.*, vol. 3, pp. 1–122, Jan. 2011.

[55] G. Liu, Z. Lin, Y. Yu, *et al.*, "Robust subspace segmentation by low-rank representation.," in *Proc. Int. Conf. Mach. Learn.*, pp. 663–670, 2010.

[56] C. Rodarmel and J. Shan, "Principal component analysis for hyperspectral image classification," *Geo. Spat. Inf. Sci.*, vol. 62, no. 2, pp. 115–122, 2002.

[57] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philos. Trans. Royal Soc. A*, vol. 374, no. 2065, p. 20150202, 2016.

[58] C. Chen, B.-S. He, Y. Ye, and X. Yuan, "The direct extension of admm for multi-block convex minimization problems is not necessarily convergent," *Mathc Programmc*, vol. 155, p. 5779, Jan. 2016.

[59] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM J. Optim.*, vol. 26, no. 1, pp. 337–364, 2016.

[60] G. Li and T. K. Pong, "Global convergence of splitting methods for nonconvex composite optimization," *SIAM J. Optim.*, vol. 25, no. 4, pp. 2434–2460, 2015.

[61] Y. Y. Tang, Y. Lu, and H. Yuan, "Hyperspectral image classification based on three-dimensional scattering wavelet transform," *IEEE Trans. Geosci. Remote sens.*, vol. 53, no. 5, pp. 2467–2480, 2014.

[62] J. Kerekes, "Receiver operating characteristic curve confidence intervals and regions," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 251–255, 2008.

[63] S. Khazai, S. Homayouni, A. Safari, and B. Mojaradi, "Anomaly detection in hyperspectral images based on an adaptive support vector method," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 646–650, 2011.

[64] G. Liu and W. Zhang, "Recovery of future data via convolution nuclear norm minimization," *arXiv:1909.03889*, pp. 1–16, 2019.

Minghua Wang (Member, IEEE) received the B.S. degree in the department of automation and the Ph.D. degree in the department of control science and engineering, Harbin Institute of Technology (HIT), Harbin, China, in 2016 and 2021, respectively. She is currently working with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences (CAS). She was also a visiting Ph.D student in the Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, Grenoble, France (2019-2020). Her research interests include Hyperspectral image denoising, anomaly detection, tensor completion, and low-rank representations.

Qiang Wang (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in control science and engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 1998, 2000, and 2004, respectively. Since 2008, he has been a Professor with the Department of Control Science and Engineering, HIT. His research interests include Hyperspectral image denoising, signal/image processing, multisensor data fusion, wireless sensor networks, and intelligent detection technology.

Danfeng Hong (S'16–M'19–SM'21) received the Dr.-Ing degree (summa cum laude) in Signal Processing in Earth Observation (SiPEO) from the Faculty of Civil, Geo and Environmental Engineering, Technical University of Munich (TUM), Munich, Germany, in 2019.

He is currently a Professor with the Key Laboratory of Computational Optical Imaging Technology, Aerospace Information Research Institute, Chinese Academy of Sciences (CAS). Before joining CAS, he has been a Research Scientist and led a Spectral Vision Working Group at the Remote Sensing Technology Institute (IMF), German Aerospace Center (DLR), Oberpfaffenhofen, Germany. His research interests include signal / image processing, hyperspectral remote sensing, machine / deep learning, artificial intelligence, and their applications in Earth Vision.

Dr. Hong is a Topical Associate Editor of the IEEE Transactions on Geoscience and Remote Sensing (TGRS), an Editorial Board Member of Remote Sensing, and an Editorial Advisory Board Member of ISPRS Journal of Photogrammetry and Remote Sensing. He was a recipient of the Best Reviewer Award of the IEEE TGRS in 2021 and 2022, and the Best Reviewer Award of the IEEE JSTARS in 2022, the Jose Bioucas Dias Award for recognizing the outstanding paper at WHISPERS in 2021, the Remote Sensing Young Investigator Award in 2022, and the IEEE GRSS Early Career Award in 2022.

Swalpa Kumar Roy (S'15) received both the bachelors and the masters degree in Computer Science and Engineering from West Bengal University of Technology, Kolkata, India, in 2012, and Indian Institute of Engineering Science and Technology, Shibpur, Howrah, India, (IIEST Shibpur) in 2015. He is currently pursuing the Ph.D. degree jointly with Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, and Department of Computer Science and Engineering, University of Calcutta, Kolkata. He was a Project Linked Person with the Optical Character Recognition (OCR) Laboratory, Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, from July 2015 to March 2016. He is currently working as an Assistant Professor with the Department of Computer Science and Engineering, Jalpaiguri Government Engineering College, Jalpaiguri, West Bengal, India. He was nominated for Indian National Academy of Engineering (INAE) engineering teachers mentoring fellowship by INAE Fellows in 2021 and also a recipient of the Outstanding Paper Award in second workshop: Hyperspectral Sensing Meets Machine Learning and Pattern Analysis (HyperMLPA) at the Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS) in 2021. He has served as a reviewer for the IEEE Transactions on Geoscience and Remote Sensing and IEEE Geoscience and Remote Sensing Letters. His research interests include computer vision, deep learning and remote sensing.

Jocelyn Chanussot (M'04–SM'04–F'12) received the M.Sc. degree in electrical engineering from the Grenoble Institute of Technology (Grenoble INP), Grenoble, France, in 1995, and the Ph.D. degree from the Universit de Savoie, Annecy, France, in 1998. Since 1999, he has been with Grenoble INP, where he is currently a Professor of signal and image processing. His research interests include image analysis, hyperspectral remote sensing, data fusion, machine learning and artificial intelligence. He has been a visiting scholar at Stanford University (USA), KTH (Sweden) and NUS (Singapore). Since 2013, he is an Adjunct Professor of the University of Iceland. In 2015-2017, he was a visiting professor at the University of California, Los Angeles (UCLA). He holds the AXA chair in remote sensing and is an Adjunct professor at the Chinese Academy of Sciences, Aerospace Information research Institute, Beijing.

Dr. Chanussot is the founding President of IEEE Geoscience and Remote Sensing French chapter (2007-2010) which received the 2010 IEEE GRS-S Chapter Excellence Award. He has received multiple outstanding paper awards. He was the Vice-President of the IEEE Geoscience and Remote Sensing Society, in charge of meetings and symposia (2017-2019). He was the General Chair of the first IEEE GRSS Workshop on Hyperspectral Image and Signal Processing, Evolution in Remote sensing (WHISPERS). He was the Chair (2009-2011) and Cochair of the GRS Data Fusion Technical Committee (2005-2008). He was a member of the Machine Learning for Signal Processing Technical Committee of the IEEE Signal Processing Society (2006-2008) and the Program Chair of the IEEE International Workshop on Machine Learning for Signal Processing (2009). He is an Associate Editor for the IEEE Transactions on Geoscience and Remote Sensing, the IEEE Transactions on Image Processing and the Proceedings of the IEEE. He was the Editor-in-Chief of the IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (2011-2015). In 2014 he served as a Guest Editor for the IEEE Signal Processing Magazine. He is a Fellow of the IEEE, a member of the Institut Universitaire de France (2012-2017) and a Highly Cited Researcher (Clarivate Analytics/Thomson Reuters).