

Hyperspectral Anomaly Detection Based on Low-Rank Representation With Data-Driven Projection and Dictionary Construction

Xiaoxiao Ma, Xiangrong Zhang^{1b}, Senior Member, IEEE, Xu Tang^{2b}, Member, IEEE, Huiyu Zhou, and Licheng Jiao^{1b}, Fellow, IEEE

Abstract—Hyperspectral image anomaly detection is an increasingly important research topic in remote sensing images understanding and interpretation. Recently, low-rank representation-based methods have attracted extensive attention and achieved promising performances in hyperspectral anomaly detection. These methods assume that the hyperspectral data can be decomposed into two parts: the low-rank component representing the background and the residual part indicating the anomaly. In order to improve the separability of the background and anomaly, we propose a novel hyperspectral anomaly detection based on low-rank representation with dictionary construction and data-driven projection. To construct a robust dictionary that contains all categories of the background objects whilst excluding the anomaly's influence, we adopt a superpixel-based tensor low-rank decomposition method to generate a comprehensive and pure background dictionary. Considering the spectral redundancy in the hyperspectral data, data-driven projection is introduced to the low-rank representation to project the original data to a low-dimensional feature space to better separate the anomaly and the background. Experimental results on four real hyperspectral datasets show that the proposed anomaly detection method outperforms the other anomaly detectors.

Index Terms—Data-driven projection, hyperspectral image (HSI) anomaly detection, low-rank representation (LRR), tensor decomposition.

I. INTRODUCTION

HYPERSPECTRAL image (HSI) has the capability to provide not only the spatial positions and structure infor-

Manuscript received January 21, 2020; revised March 31, 2020; accepted April 11, 2020. Date of publication May 11, 2020; date of current version May 29, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61772400, Grant 61501353, Grant 61772399, Grant 91438201, and Grant 61573267 and in part by the Key Research and Development Program in the Shaanxi Province of China under Grant 2019ZDLGY03-08. The work of Huiyu Zhou was supported in part by the U.K. EPSRC under Grant EP/N011074/1, in part by the Royal Society Newton Advanced Fellowship under Grant NA160342, and in part by European Union's Horizon 2020 research and innovation program under the Marie-Sklodowska-Curie under Grant 720325. (Corresponding author: Xiangrong Zhang.)

Xiaoxiao Ma, Xiangrong Zhang, Xu Tang, and Licheng Jiao are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an 710071, China (e-mail: xx_ma@stu.xidian.edu.cn; xrzhang@mail.xidian.edu.cn; tangxu128@gmail.com; lchjiao@mail.xidian.edu.cn).

Huiyu Zhou is with the Department of Informatics, University of Leicester, Leicester LE1 7RH, U.K. (e-mail: hz143@leicester.ac.uk).

Digital Object Identifier 10.1109/JSTARS.2020.2990457

mation but also the high-resolution spectral information for discriminating ground objects [1]. They have been successfully applied in military and civilian domains [2]. A lot of research works on HSI range from dimensionality reduction [3], spectral unmixing [4], HSI classification [5], [6] to detection task [7], [8], [9]. There are two application scenarios in the detection task: target detection and anomaly detection. Given the spectral prior, hyperspectral target detection is to find the object of interest with supervised learning. However, in many cases, the prior information is hard to obtain. In hyperspectral anomaly detection, the characteristic distribution or spectral prior is unavailable. Anomaly detection is usually to find pixels or subpixels that differ from its surroundings, e.g., ships at sea or aircraft in paddy fields, which can be considered as anomalies [10], [11]. Anomaly detection has broad application prospects. However, complex background, low spatial resolution, and high-dimensional spectral characteristics may pose challenges to anomaly detection in HSI.

Researchers have proposed various methods for hyperspectral anomaly detection. A typical algorithm is the RX detector [12], which hypothesizes the HSI data as the normal distribution and performs the detection according to the Mahalanobis distances between the test pixels and the background. However, the statistical characteristics estimation of background is often inaccurate due to the existence of noise and anomalies. A lot of RX's variant algorithms have been proposed. Guo *et al.* [13] proposed a weighted RX detector and linear filter-based RX detector that can provide more accurate background statistical characteristic estimation and decrease the influence of anomalous pixels and noise. Moreover, some computation-efficient methods also were discussed in the causal RX detector [14]. Single-feature-based anomaly detector [15] provides a framework for subpixel anomaly detection. Liu and Chang [16] proposed a multiple-window anomaly detection method to utilize the spectral information. However, the statistic-based anomaly detection methods assume that the hyperspectral data follows a Gaussian distribution, which may pose a limitation for real data.

To overcome the disadvantages of inaccurate distribution hypothesis, the representation-based methods have been proposed and indicated favorable performances in HSI analysis. Chen *et al.* proposed dictionary-based sparse representation method for HSI classification [17]. In [18], Li *et al.* utilized the

representation residual to define anomalies by constraining the representation coefficients matrix with collaborative representation detector (CRD), and extended the basic CRD model into nonlinear kernel space. Considering the spatial and spectral similarities in HSI, Xu *et al.* developed the nonlocal patch tensor-based sparse representation for HSI super-resolution [19]. Nowadays, the sparse representation with low-rank constraint models has been explored. In [20], Zhao *et al.* applied the sparse coding to explore the global redundancy in the spatial domain and the local correlation in the spectral domain, moreover, the low-rank constraint is employed to overcome the spectral distortion problem for HSI denoising. Furthermore, the coupled sparse denoising and unmixing with low-rank constraint model was presented in [21]. In [22], the $l_{1/2}$ regularized low-rank and sparse representation LRASR-based graph cuts method was proposed to precisely describe the low-rank structure of data for HSI classification. In [23], the tensor LRASR framework was proposed for HSI-compressed sensing reconstruction and anomaly detection. The representation-based algorithms can avoid the distribution estimation for the real data, sparse and low-rank constraints can inspire us to make full use of the prior information. Recently, with the superior ability in exploring the deep features, deep learning has been applied in remote sensing image analysis including supervised learning [24] and weakly supervised learning [25]. Hyperspectral anomaly detection can be considered as an unsupervised learning task, Li *et al.* [26] exploited the transferred deep convolutional neural network framework to detect the anomalous pixels in HSI. However, deep learning-based methods require huge labeled samples to participate in the model training. In general, the labeled samples are difficult or time consuming to obtain.

In hyperspectral anomaly detection, researchers have made efforts on the LRR-based approaches. It assumes that the background follows a low-rank prior and the anomalies are sparse distributed. The robust principal component analysis (RPCA) model aimed at factorizing the observed matrix into low-rank background component and sparse anomaly component [27], [28]. As the background includes complex ground object materials, the RPCA method cannot completely separate the anomaly from background. To obtain a relatively accurate background estimation, the low-rank and sparse matrix decomposition model was applied in [29]. Dictionary-based representation is of great help in depicting background characteristics. In [30], the LRR method was described. For simplicity, the whole original data were adopted to construct the background dictionary in LRR. Obviously, the anomalous pixels are included in the dictionary. To generate a comprehensive and powerful dictionary, dictionary construction strategies were explored. In [31], Xu *et al.* introduced clustering and atoms selection algorithms to construct a background dictionary for LRASR model. Furthermore, to depict the local structure information, the l_1 norm was applied on the representation coefficients matrix. In order to learn a representative dictionary, in LRRaLD (LRR and learned dictionary) model, the sparse coding gradient descent is adopted to learn and establish a dictionary [32]. The above LRR-based methods perform the matrix decomposition in the original spectral dimension

which does not consider the spectral redundancy. In addition, the constructed dictionaries are not pure for the noise interference. These two factors restrict the detectors' performances.

In order to reduce the data redundancy of HSI, dimensionality reduction [3] and band selection [33] are two applicable technologies. In [34], researchers extended the endmembers set extraction idea into bands optimal subset selection, then the basic RX detector was applied to the dimensionality-reduced data for anomaly detection. In fact, there may exist many mixed pixels in HSI. For the subpixel anomalies, the researchers in [35] applied the unmixing technologies to obtain an abundance matrix that can reflect more distinctive characteristics about background and anomalies. Meanwhile, the original data space was transferred into the abundance feature space, then the dictionary-based low-rank model was applied in the new feature space to separate anomalies. Transferring the original HSI data into a low-dimensional feature space can reduce the data redundancy and complexity efficiently, which is beneficial for the following anomalies detection. However, the dimensionality reduction and anomaly detection are two separate processes in general, which cause that the detection result relies on the obtained low-dimensional feature data. To reduce the redundancy of data and preserve the critical information, the projection-based methods were applied for dimensionality reduction of complex data [36], [37]. In [38], the RX detector was derived into the projection domain for hyperspectral anomaly detection. The features are transferred into a new space via the projection matrix. In our work, to reduce the redundancy of data and separate anomalies, the projection-based LRR model is applied for anomaly detection.

In order to realize dimensionality reduction and separate anomalies simultaneously in a model and to construct a comprehensive and pure dictionary for background modeling, in this article, a novel data-driven projection LRR method with dictionary construction is presented. Projecting the original HSI into a new low-dimensional feature space can enable us to extract the active and influential components of the data for the anomaly detection task. In dictionary construction, in order to construct a comprehensive dictionary that contains all of the background categories of the ground objects, we adopt a superpixel segmentation-based strategy to construct a dictionary candidate set. In order to make the dictionary candidates set more pure, we adopt the tensor low-rank decomposition algorithm to compress the interference of noise since the tensor decomposition-based framework can preserve the intrinsic structure information of data and reduce the noise including sparse noise efficiently. The whole scheme of the data-driven projection LRR with dictionary construction method is summarized in Fig. 1.

The contributions of our proposed data-driven projection LRR with dictionary construction method can be summarized as follows.

- 1) Different from traditional dimensionality reduction methods that are performed as a preprocessing step, we propose a data-driven projection LRR model that performs dimensionality reduction and anomaly detection simultaneously. In this model, the projection matrix is optimized under the framework of LRR that ensures the projection direction is

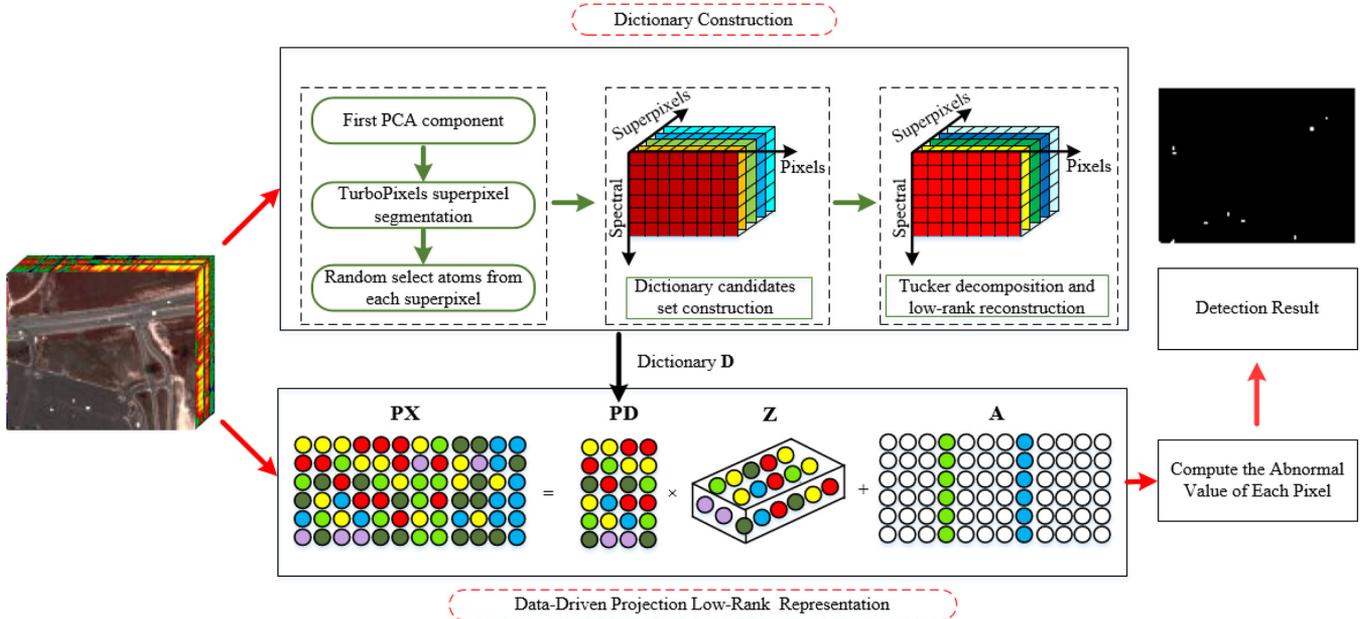


Fig. 1. Whole scheme of DPLR method for HSI anomaly detection.

consistent with the representation characteristics of the data.

- 2) A novel dictionary construction method is presented to depict the background characteristics. a) To obtain a comprehensive dictionary that contains all of the background object classes, a superpixel segmentation-based atoms selection strategy is used to generate a dictionary candidates cube. 2) To make the background dictionary pure enough, a tensor Tucker decomposition method is applied to the dictionary candidates set to compress the interference of sparse noise.

The remainder of this article is organized as follows. The basic of the LRR model is reviewed in Section II. In Section III, we detailed the proposed data-driven projection low-rank representation (DPLR) for HSI anomaly detection task including model formation and optimization. Section IV describes the experiments implemented on four real HSI datasets, and the discussions about the parameters are presented. Finally, conclusions are given in Section V.

II. LRR MODEL

Suppose $\mathbf{X} = \mathbf{L} + \mathbf{E}$, it aims at recovering the low-rank component \mathbf{L} from the acquired data \mathbf{X} with sparse noise \mathbf{E} interference [27]. The optimization problem is given by

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{E}} \text{rank}(\mathbf{L}) + \tau_0 \|\mathbf{E}\|_0 \\ \text{s.t. } \mathbf{X} = \mathbf{L} + \mathbf{E} \end{aligned} \quad (1)$$

where \mathbf{X} is the observed data and $\mathbf{X} \in R^{B \times N}$ with N samples in R^B . $\mathbf{L}, \mathbf{E} \in R^{B \times N}$ are the low-rank item and the sparse noise term, respectively. To balance the two parts, the weight parameter τ_0 ($\tau_0 > 0$) is introduced. In (1), $\text{rank}(\cdot)$ is the rank

function, $\|\cdot\|_0$ is the l_0 norm, which is the number of nonzero elements of matrix entries.

To solve the NP-hard problem in (1), the convex relax form of (1) is

$$\begin{aligned} \min_{\mathbf{L}, \mathbf{E}} \|\mathbf{L}\|_* + \tau_0 \|\mathbf{E}\|_1 \\ \text{s.t. } \mathbf{X} = \mathbf{L} + \mathbf{E} \end{aligned} \quad (2)$$

where $\|\cdot\|_*$ is the nuclear norm, which is the sum of its singular values. $\|\cdot\|_1$ is the l_1 norm, which is the sum of nonzero elements of a matrix.

Recent research works on the LRR demonstrate that a reliable dictionary is closely related to the robust representation of the data. For the LRR model in [30], researchers adopted the data itself as the dictionary to represent the low-rank component, then (2) can be rewritten as

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}} \|\mathbf{Z}\|_* + \tau_1 \|\mathbf{E}\|_{2,1} \\ \text{s.t. } \mathbf{X} = \mathbf{XZ} + \mathbf{E} \end{aligned} \quad (3)$$

where \mathbf{Z} is the representation coefficients matrix of \mathbf{X} , $\mathbf{Z} \in R^{N \times N}$, τ_1 is the weight parameter. $\|\cdot\|_{2,1}$ is called as the $l_{2,1}$ norm, and it can be seen as the l_1 norm of l_2 norm of its columns. In general, the l_1 norm induced solutions can only exploit the unstructured sparsity of data. Compared with the l_1 norm on matrix \mathbf{E} in (2), the $l_{2,1}$ norm can generate the sparsity on the columns [39]. The joint-sparsity structure is imposed on the matrix by $l_{2,1}$ norm, as the $l_{2,1}$ norm restricts most of the columns' values of the corresponding matrix to close zero. In (3), the $l_{2,1}$ norm on the sparse noise term \mathbf{E} can inspire us to search the corrupted specific samples.

III. LRR WITH DATA-DRIVEN PROJECTION AND DICTIONARY CONSTRUCTION

Hyperspectral anomaly detection can be considered as an unsupervised classification task. Compared with the anomalies, background pixels make up the major components of the image, even more than 90%. The data are imbalanced between the two classes. Furthermore, the background includes various ground objects which make the division task complex. These factors increase the difficulty of anomaly detection, and the commonly used classifiers are no longer applicable here.

Considering the characteristics of background and anomalies, the hyperspectral anomaly detection task can be formulated with LRR model, i.e., the background is modeled as a low-rank part and the anomaly is modeled as the sparse part. The dictionary-based representation methods have been successfully applied in HSI analysis [17], [31], [32]. To better characterize the background, we introduce dictionary \mathbf{D} that is constructed based on superpixel segmentation and tensor low-rank decomposition. The dictionary-based LRR model is as follows:

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}} \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1} \\ \text{s.t. } \mathbf{X} = \mathbf{DZ} + \mathbf{E} \end{aligned} \quad (4)$$

where $\mathbf{X} \in R^{B \times N}$ is the hyperspectral data with B bands and N pixels. The dictionary matrix $\mathbf{D} = [d_1, d_2, \dots, d_m]$ includes m atoms. $\mathbf{E} = [e_1, e_2, \dots, e_N]$ represents the remaining part which includes the anomaly. λ ($\lambda > 0$) is a parameter to balance the low-rank part and the sparse part. This model decomposes the HSI data into background and anomaly. The low-rank constraint on the background part can help us explore the global information of data by strengthening the correlation between the vectors of \mathbf{Z} . As each hyperspectral pixel is a column vector, we adopt the $l_{2,1}$ norm to explore the structure sparsity of matrix \mathbf{E} along its column.

A. Data-Driven Projection LRR

Hyperspectral data typically have continuous spectral bands per pixel with high spectral resolution. This makes the acquired HSI data redundant and highly correlated between the bands. There are two applicable techniques to reduce the spectral redundancy. One is based on band selection by selecting representative bands from the original bands set to replace the original data. The other is based on dimensionality reduction by transforming the original spectral data into a lower dimensional feature space. Above all, the process of reducing the data redundancy is a separated step, which can be viewed as a preprocessing step before performing a specific task (for examples, classification or target detection). The quality of dimensionality reduction will affect the following classification or detection accuracy. In order to perform the dimensionality reduction and separate anomalies from the background, we propose a data-driven projection LRR model. Our goal is to preserve the anomaly information as much as possible while dimensionality reduction is performed, and we will explore the low-rank characteristic of background and the sparsity of anomalies in a low-dimensional feature space.

Suppose a hyperspectral data cube can be denoted as $\mathcal{X} \in R^{B \times V \times H}$, where the spectral dimension is B and the spatial size is $V \times H$, we convert the data cube into a two-dimensional matrix $\mathbf{X} \in R^{B \times N}$, N is the total number of pixels, $N = V \times H$ (we denote cubic data or tensors by curlicue letters, e.g., \mathcal{X}, \mathcal{D} ; matrices are expressed with general italicized bold capital letters, e.g., \mathbf{X}, \mathbf{D}). In order to reduce the HSI spectral redundancy and detect anomalies, we introduce low-dimensional projection $\mathbf{P} \in R^{b \times B}$ into LRR model, which can transform the original data with B spectral dimensions into b ($b < B$) dimensional feature space. To better preserve the anomaly information and enhance the separability of background and anomalies, we impose an orthogonality constraint on the transformation matrix \mathbf{P} . The optimization problem is defined as

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{A}, \mathbf{P}} \|\mathbf{Z}\|_* + \lambda \|\mathbf{A}\|_{2,1} \\ \text{s.t. } \mathbf{PX} = \mathbf{PDZ} + \mathbf{A}, \mathbf{P}^T \mathbf{P} = \mathbf{I} \end{aligned} \quad (5)$$

where $\mathbf{P} \in R^{b \times B}$ ($b < B$), the dictionary $\mathbf{D} \in R^{B \times m}$ includes m atoms, $\mathbf{Z} \in R^{m \times N}$ is the representation coefficients matrix, \mathbf{A} is the residual part including anomaly, $\mathbf{A} \in R^{b \times N}$. $\lambda > 0$ is a weight parameter of the sparse term relative to the low-rank part. In (5), \mathbf{PX} represents the low-dimensional data after projection, \mathbf{PDZ} is the corresponding background component in the low-dimensional feature space, and \mathbf{A} is the corresponding remaining part. Compared with the model in (4), our proposed model can project the background and the anomaly into the corresponding low-dimensional feature space, whilst the intrinsic low-rank characteristic of background and the sparse prior to anomalies are exploited effectively.

Introducing the projection matrix with an orthogonality constraint on HSI data is expected to scatter the data as much as possible and remove the intercorrelation to better detect the anomaly. Model (5) provides an approach to detect anomalies in a low-dimensional feature space. The representation coefficients matrix \mathbf{Z} on the dictionary is low-rank as the background component has low-rank invariance in the low-dimensional feature space. The anomaly item \mathbf{A} is constrained with $l_{2,1}$ norm.

In order to solve the data-driven projection LRR model in (5), we apply the alternating direction method of multipliers (ADMM) algorithm [40]. To make the objective function separable, the auxiliary variable \mathbf{H} is introduced, the problem in (5) can be converted as the following problem:

$$\begin{aligned} \min_{\mathbf{H}, \mathbf{A}, \mathbf{P}, \mathbf{Z}} \|\mathbf{H}\|_* + \lambda \|\mathbf{A}\|_{2,1} \\ \text{s.t. } \mathbf{PX} = \mathbf{PDZ} + \mathbf{A}, \mathbf{P}^T \mathbf{P} = \mathbf{I}, \mathbf{Z} = \mathbf{H}. \end{aligned} \quad (6)$$

The unfolded augmented Lagrangian function of (6) is given as:

$$\begin{aligned} L(\mathbf{H}, \mathbf{Z}, \mathbf{A}, \mathbf{P}, \mathbf{Y}_1, \mathbf{Y}_2, \mu) \\ = \|\mathbf{H}\|_* + \lambda \|\mathbf{A}\|_{2,1} + \langle \mathbf{Y}_1, \mathbf{PX} - \mathbf{PDZ} - \mathbf{A} \rangle + \langle \mathbf{Y}_2, \mathbf{Z} - \mathbf{H} \rangle \\ + \frac{\mu}{2} (\|\mathbf{PX} - \mathbf{PDZ} - \mathbf{A}\|_F^2 + \|\mathbf{Z} - \mathbf{H}\|_F^2) \\ = \|\mathbf{H}\|_* + \lambda \|\mathbf{A}\|_{2,1} + \frac{\mu}{2} (\|\mathbf{PX} - \mathbf{PDZ} - \mathbf{A} + \mathbf{Y}_1/\mu\|_F^2 \end{aligned}$$

$$+ \|\mathbf{Z} - \mathbf{H} + \mathbf{Y}_2/\mu\|_F^2) - \frac{1}{2\mu}(\|\mathbf{Y}_1\|_F^2 + \|\mathbf{Y}_2\|_F^2)$$

s.t. $\mathbf{P}^T \mathbf{P} = \mathbf{I}$ (7)

where \mathbf{Y}_1 and \mathbf{Y}_2 are Lagrangian multipliers, μ ($\mu > 0$) is a penalty parameter. We update the variables \mathbf{H} , \mathbf{A} , \mathbf{P} , and \mathbf{Z} in turn, by minimizing the Lagrangian function with other variables fixed. Equations (8)–(11) detail the update formulas of variables with ADMM algorithm.

- 1) Fix \mathbf{Z} , \mathbf{A} , \mathbf{P} and update \mathbf{H} , the objective function is as follows:

$$\min_{\mathbf{H}} \|\mathbf{H}\|_* + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{H} + \mathbf{Y}_2/\mu\|_F^2. \quad (8)$$

- 2) Fix \mathbf{H} , \mathbf{Z} , \mathbf{P} and update \mathbf{A} , the objective function is as follows:

$$\min_{\mathbf{A}} \lambda \|\mathbf{A}\|_{2,1} + \frac{\mu}{2} \|\mathbf{P}\mathbf{X} - \mathbf{P}\mathbf{D}\mathbf{Z} - \mathbf{A} + \mathbf{Y}_1/\mu\|_F^2. \quad (9)$$

- 3) Fix \mathbf{H} , \mathbf{Z} , \mathbf{A} and update \mathbf{P} , the objective function is as follows:

$$\min_{\mathbf{P}} \frac{\mu}{2} \|\mathbf{P}\mathbf{X} - \mathbf{P}\mathbf{D}\mathbf{Z} - \mathbf{A} + \mathbf{Y}_1/\mu\|_F^2 \text{ s.t. } \mathbf{P}^T \mathbf{P} = \mathbf{I}. \quad (10)$$

- 4) Fix \mathbf{H} , \mathbf{A} , \mathbf{P} and update \mathbf{Z} , the objective function is as follows:

$$\min_{\mathbf{Z}} \frac{\mu}{2} (\|\mathbf{P}\mathbf{X} - \mathbf{P}\mathbf{D}\mathbf{Z} - \mathbf{A} + \mathbf{Y}_1/\mu\|_F^2 + \|\mathbf{Z} - \mathbf{H} + \mathbf{Y}_2/\mu\|_F^2). \quad (11)$$

The solution of \mathbf{H} is $\Theta_{(1/\mu)}(\mathbf{Z} + \mathbf{Y}_2/\mu)$ where Θ is the singular value thresholding shrinkage operator, then we update \mathbf{A} by $\Omega_{(\lambda/\mu)}(\mathbf{P}\mathbf{X} - \mathbf{P}\mathbf{D}\mathbf{Z} + \mathbf{Y}_1/\mu)$ where Ω is the $l_{2,1}$ minimization operator. As the solution of \mathbf{P} is a classic orthogonal ProCrustes problem that has been referred in [41], we first compute the singular value decomposition (SVD) of matrix $(\mathbf{A} - \mathbf{Y}_1/\mu)(\mathbf{X} - \mathbf{D}\mathbf{Z})^T$ as $(\mathbf{A} - \mathbf{Y}_1/\mu)(\mathbf{X} - \mathbf{D}\mathbf{Z})^T = \mathbf{U}\mathbf{S}\mathbf{V}^T$ and then let $\mathbf{P} = \mathbf{U}\mathbf{V}^T$. We solve the variable \mathbf{Z} in (11) by computing the derivative with respect \mathbf{Z} and setting the derivative to be zero, then we obtain $\mathbf{Z} = [(\mathbf{P}\mathbf{D})^T \mathbf{P}\mathbf{D} + \mathbf{I}_m]^{-1} [(\mathbf{P}\mathbf{D})^T (\mathbf{P}\mathbf{X} - \mathbf{A} + \mathbf{Y}_1/\mu) + (\mathbf{H} - \mathbf{Y}_2/\mu)]$. The algorithm flow of solving (7) is described in Algorithm 1.

So far, we obtain a matrix $\mathbf{A} \in R^{b \times N}$ including anomaly, where b is the dimension after projection, N is the total number of pixels. The abnormal value of each pixel can be calculated as follows:

$$R(\mathbf{x}_i) = \|\mathbf{A}_{:,i}\|_2 = \sqrt{\sum_j ([\mathbf{A}]_{j,i})^2}. \quad (12)$$

The abnormal value $R(\mathbf{x}_i)$ can reflect the abnormality level of the pixel \mathbf{x}_i . A pixel with larger value means it is will likely to be an abnormal pixel.

B. Dictionary Construction Based on Superpixel Segmentation and Tensor Low-Rank Decomposition

In our model, dictionary-based LRR can better depict the background characteristics to separate anomalies from the background. The aim of this section is to construct a comprehensive

Algorithm 1: Data-Driven Projection Low-Rank Representation.

Input: HSI dataset $\mathbf{X} \in R^{B \times N}$, Dimensionality b

Initialize: $\mathbf{H}_0 = \mathbf{A}_0 = \mathbf{Z}_0 = \mathbf{Y}_{1,0} = \mathbf{Y}_{2,0} = 0$, $\mathbf{P} = \mathbf{I}$, $\mu_{\max} = 10^6$, $\mu = 0.01$, $\rho = 1.1$, $\varepsilon = 10^{-6}$

1. Dictionary construction as depicted in Algorithm 2

While not converged do

2. Fix variables \mathbf{Z} , \mathbf{A} , \mathbf{P} and update \mathbf{H} by solving (8)

3. Fix variables \mathbf{H} , \mathbf{Z} , \mathbf{P} and update \mathbf{A} by solving (9)

4. Fix variables \mathbf{H} , \mathbf{Z} , \mathbf{A} and update \mathbf{P} by solving (10)

5. Fix variables \mathbf{H} , \mathbf{A} , \mathbf{P} and update \mathbf{Z} by

$$\mathbf{Z} := [(\mathbf{P}\mathbf{D})^T \mathbf{P}\mathbf{D} + \mathbf{I}_m]^{-1} [(\mathbf{P}\mathbf{D})^T (\mathbf{P}\mathbf{X} - \mathbf{A} + \mathbf{Y}_1/\mu) + (\mathbf{H} - \mathbf{Y}_2/\mu)]$$

6. Update \mathbf{Y}_1 and \mathbf{Y}_2 by

$$\mathbf{Y}_1 := \mathbf{Y}_1 + \mu(\mathbf{P}\mathbf{X} - \mathbf{P}\mathbf{D}\mathbf{Z} - \mathbf{A})$$

$$\mathbf{Y}_2 := \mathbf{Y}_2 + \mu(\mathbf{Z} - \mathbf{H})$$

7. Update μ by

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

8. Check the convergence conditions

$$\|\mathbf{P}\mathbf{X} - \mathbf{P}\mathbf{D}\mathbf{Z} - \mathbf{A}\|_F < \varepsilon, \|\mathbf{Z} - \mathbf{H}\|_F < \varepsilon$$

End while

Output: \mathbf{A} , \mathbf{P} , \mathbf{Z} .

Algorithm 2: Dictionary \mathbf{D} Construction Algorithm Based on Superpixel Segmentation and Tensor Low-Rank Decomposition

Input: HSI dataset $\mathcal{X} \in R^{B \times V \times H}$, superpixel number J and selected atoms number K

1. PCA algorithm is applied to the HSI and the first main component is obtained, denoted as $\mathbf{PC1}$

$$(\mathbf{PC1} \in R^{V \times H})$$

2. Segment $\mathbf{PC1}$ into J superpixels by applying TurboPixel superpixel segmentation algorithm and select K atoms from each superpixel

3. Construct a 3-D dictionary candidates set \mathcal{Y} with spectral distribution, $\mathcal{Y} \in R^{B \times J \times K}$

4. Compute the lower rank dimension (r_1, r_2, r_3) of \mathcal{Y}

5. Solve (14) using ALS algorithm

6. Solve (15) to obtain the dictionary \mathcal{D}

7. Convert \mathcal{D} into matrix \mathbf{D}

Output: dictionary \mathbf{D} .

and pure background dictionary to enhance the discrimination of background and anomalies. In detail, we propose a new superpixel segmentation-based tensor low-rank decomposition algorithm to construct dictionary \mathbf{D} . Our dictionary construction method consists of two steps. In step 1, similar to the cluster-based dictionary construction method of [31], we adopt superpixel segmentation to construct a comprehensive dictionary candidates set. Moreover, in order to compress the interference of noise (including sparse anomalies), in step 2, the tensor Tucker decomposition method is applied on dictionary candidates set. The details of our proposed superpixel-based tensor low-rank decomposition algorithm are as follows.

1) *Superpixel Segmentation-Based Dictionary Candidates Set Construction*: Superpixel segmentation allows us to divide the image into small homogeneous areas. In general, the superpixel segmentation algorithms can be classified as the graph theory-based algorithms, e.g., entropy rate superpixel [42], and the gradient descent-based algorithms, e.g., SLIC [43] and TurboPixel superpixels [44]. Considering the characteristics of HSI, the TurboPixels superpixel segmentation algorithm is applied. The superpixels obtained by TurboPixels are of roughly a similar size and shape. In detail, before superpixel segmentation is conducted, we apply principal component analysis (PCA) [45] to extract the key information of the HSI data. As the pixels within superpixel have similarities in spatial and spectral, we randomly select atoms from each superpixel to construct a dictionary candidates set. In real HSI, as the categories of ground objects are limited, we set the number of superpixel J ($J > 0$) to be medium. We denote the real obtained superpixel number as \hat{J} , we randomly select K ($K > 0$) atoms from each superpixel. The atoms number of the dictionary candidates set is $\hat{J} \times K$. With spectral distribution, the dictionary candidates set are represented as $\mathcal{Y} \in R^{B \times \hat{J} \times K}$, which is a three-order tensor, B is the number of spectral bands. We denote the second mode as superpixel direction and the third mode as pixel direction.

2) *Tensor Low-Rank Decomposition-Based Dictionary Construction*: The dictionary candidates set \mathcal{Y} consists of background components and noises. As the anomalous pixels are sparse and randomly distributed with small size, the anomalies can be considered as noise. It can be modeled as $\mathcal{Y} = \mathcal{D} + \mathcal{N}$, where \mathcal{D} is the background dictionary, \mathcal{N} is the noise component, $\mathcal{D}, \mathcal{N} \in R^{B \times \hat{J} \times K}$ [46]. To obtain a pure dictionary, it aims at minimizing the following formula:

$$\min_{\mathcal{D}} \|\mathcal{Y} - \mathcal{D}\|_F^2 \quad (13)$$

where \mathcal{D} is an approximation tensor of \mathcal{Y} . We aim at estimating the dictionary \mathcal{D} using a subspace-based method. In order to reconstruct the pure background dictionary, we apply the denoising method on the dictionary candidates set \mathcal{Y} . In (13), the tensor-based denoising model can better preserve the intrinsic structure information. We here use the subspace spanned by the eigenvectors corresponding to the covariance matrix of the observation signal to solve the problem. The subspace-based method can be extended to three-order tensor data, the purpose is to find the lower rank- (r_1, r_2, r_3) of dictionary \mathcal{D} . Using the Tucker decomposition (or called the SVD) [47], the problem in (13) can be rewritten as

$$\min_{\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3} \|\mathcal{Y} \times_1 \mathbf{S}_1 \times_2 \mathbf{S}_2 \times_3 \mathbf{S}_3\|_F^2 \quad (14)$$

where \mathbf{S}_n ($n = 1, 2, 3$) consists of the first r_n ($n = 1, 2, 3$) eigenvectors corresponding to covariance matrix $\mathbf{R}_n \mathbf{R}_n^T$, where \mathbf{R}_n is the n -mode flattening of \mathcal{Y} . “ \times_n ” is the n -mode product of tensor \mathcal{Y} with a matrix \mathbf{S}_n ($n = 1, 2, 3$). From (14), we can see that the Tucker decomposition-based three-order tensor low-rank reconstruction can better explore the low-rank characteristic in both spectral and spatial directions. Then we use the alternating least square (ALS) algorithm to solve (14).

The solution of (13) can be unfolded as

$$\mathcal{D} = \mathcal{Y} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3 \quad (15)$$

where the projector upon the n -mode of \mathcal{Y} is denoted as \mathbf{U}_n , $\mathbf{U}_n = \mathbf{S}_n \mathbf{S}_n^T$ ($n = 1, 2, 3$) with $\mathbf{S}_n = [s_1, \dots, s_{r_n}]$. Specifically, we adopt the Akaike information criterion [48] algorithm to compute the subspace dimension r_n ($n = 1, 2, 3$).

Based on tensor low-rank decomposition, we obtain the dictionary \mathcal{D} with $\hat{J} \times K$ dictionary atoms, \mathcal{D} is a low-rank approximation of dictionary candidates set \mathcal{Y} . We convert the three-order tensor \mathcal{D} into two-dimensional (2-D) matrix $\mathbf{D} \in R^{B \times m}$ ($m = \hat{J} \times K$). Algorithm 2 describes the dictionary construction key techniques in detail.

C. Algorithm Complexity Analysis

The main computation cost in Algorithm 1 appears in step 2, which requires to compute the SVD of the matrix with the complexity of $\mathcal{O}(m^2 N)$, where m is the number of dictionary atoms, N is the total number of pixels. The computation complexity of step 3 is approximate $\mathcal{O}(b^2 N)$, where b is the reduced dimensionality. The computation complexity of solving \mathbf{P} is around $\mathcal{O}(BbN)$ (B is the original spectral dimensions). The computation complexity of solving \mathbf{Z} is about $\mathcal{O}(BbN)$. Therefore, the total computation complexity of Algorithm 1 is about $\mathcal{O}[\eta N(m^2 + b^2 + Bb)]$, where η is the number of iterations.

IV. EXPERIMENTS

A. Datasets

In order to verify the effectiveness of our proposed method, a series of experiments are designed on the four real hyperspectral datasets captured by different sensors. The information of datasets is as follows.

- 1) *HYDICE urban dataset* [18]: The first dataset was captured by HYDICE airborne sensor from an aircraft platform with the spectral resolution of 10 nm and spatial resolution of 1 m approximately. Before performing the detection task, we removed the low SNR and vapor absorption bands. We obtained the HYDICE urban dataset with 175 spectral bands for experiments. We present the pseudocolor map in Fig. 2(a) with a spatial size 307×307 , the subscene in red rectangle window is picked for evaluation experiments with a spatial size of 80×100 . The subscene consists of 21 anomalous pixels. Fig. 2(b) is the pseudocolor map of the subscene, in which, the cars and roofs are defined as anomalies, the position information is described in Fig. 2(c).
- 2) *AVIRIS urban scene dataset* [49]: The second dataset was collected by an Airborne Visible/Infrared Imaging Spectrometer with a spatial resolution of 17.2 m. Fig. 3(a) shows the pseudocolor image of AVIRIS urban scene which consists of 100×100 pixels. The second dataset has 207 available bands for experiments, Fig. 3(b) describes the position information of the anomalies.
- 3) *Cri hyperspectral dataset* [29]: The third dataset was acquired by the Nuance Cri hyperspectral sensor. This

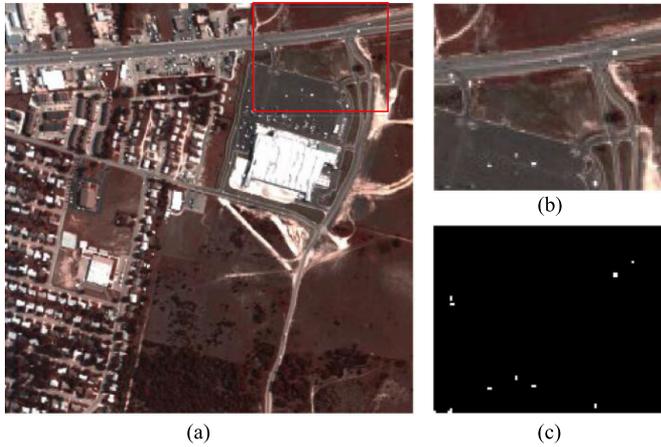


Fig. 2. HYDICE urban dataset. (a) Pseudocolor map of the whole image. (b) Pseudocolor image of the interest area. (c) Reference.

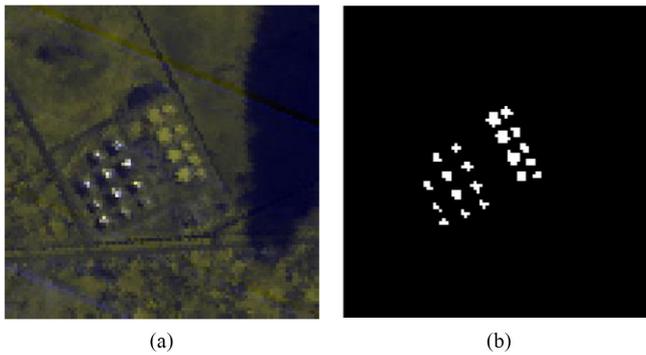


Fig. 3. AVIRIS urban scene dataset. (a) Pseudocolor image. (b) Reference.

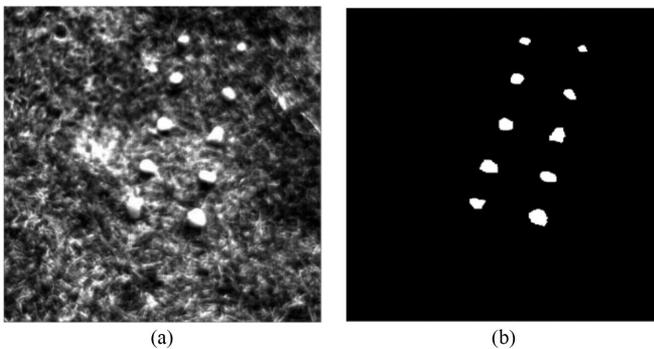


Fig. 4. Cri hyperspectral dataset. (a) Pseudocolor image. (b) Reference.

sensor can acquire imagery with a spectral resolution of 10 nm covering the wavelengths from 650 to 1100 nm. The image scene covers an area of 400×400 pixels with 46 spectral bands. The pseudocolor image is presented in Fig. 4(a). The position information of the anomalies is given in Fig. 4(b), in which the white regions are 10 rocks consisting of 2261 pixels.

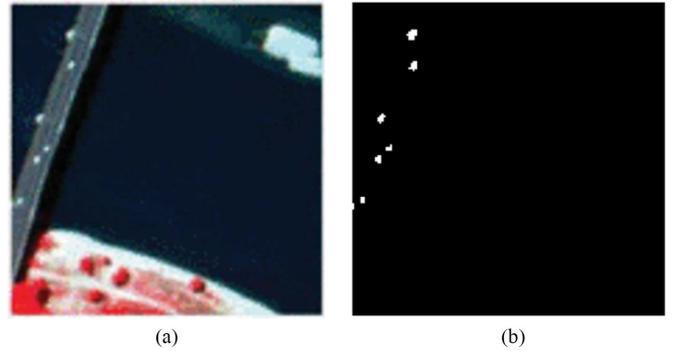


Fig. 5. Pavia beach scene dataset. (a) Pseudocolor image. (b) Reference.

- 4) *Pavia beach scene dataset* [49]: The fourth dataset was captured by the Reflective Optics System Imaging Spectrometer sensor with spatial resolution 1.3 m. The spatial size of this dataset is 150×150 , and the number of available spectral bands is 102. The pseudocolor image of this dataset is presented in Fig. 5(a), and the position information of anomalies is given in Fig. 5(b).

B. Comparison Methods and Evaluation Criteria

In the experiments, we will compare our DPLR method with eight representative methods including statistic-based methods, representation-based methods, and dimensionality reduction-based methods.

- 1) *Global RX (GRX)* [12]: It is now considered as the statistic-based benchmark algorithm, in which, the HSI data are hypothesized as a normal distribution, and then calculate the Mahalanobis distance between the test pixel and the entire image.
- 2) *Local RX (LRX)* [12]: Different from GRX, LRX calculates the Mahalanobis distance between the test pixel and its local neighbors.
- 3) *Sparse representation and linear mixture model (SR_LMM)* [50]: The background can be sparsely represented by its spatial neighbors, while anomaly cannot. In SR_LMM, the sum-to-one and threshold constraints are imposed on the coefficients vector to remove anomalies from the background dictionary.
- 4) *Collaborative representation-based detector (CRD)* [18]: The pixels are represented by their spatial neighbors and the l_2 norm is imposed on the representation weight vectors. The representation residuals are used to measure anomalies.
- 5) *RPCA* [27]: The RPCA model is as shown in (2), by which the HSI data are decomposed into a low-rank matrix and a sparse matrix. Then (12) is applied on the sparse matrix to compute the abnormal value of each pixel.
- 6) *LRR* [30]: This model has been shown in (3), and it is an extension of RPCA, in which the dictionary is introduced to represent the background. In the LRR algorithm, for simplicity, the whole data is chosen as the dictionary.

- 7) *LRASR method* [31]: LRASR is based on the LRR method with dictionary construction. In order to describe the background, LRASR adopts a cluster-based dictionary construction strategy. To depict the local characteristics, the l_1 norm is imposed on the representation coefficients matrix.
- 8) *PCA_LRR* [30], [45]: The PCA algorithm is performed to obtain a dimensionality reduced data, then the LRR algorithm is applied to separate anomaly.

The GRX and LRX are statistic-based benchmark methods. Both methods are usually adopted as comparison algorithms for HSI anomaly detection. SR_LMM, CRD, LRR, LRASR, PCA_LRR, and our DPLR method are representation-based methods. In our experiments, we select the classical statistic-based methods for comparative analysis to verify the availability of representation-based methods. SR_LMM and CRD are positive comparison algorithms based on representation. RPCA is the basic low-rank and sparse decomposition model. LRR method takes the whole data as the dictionary in the LRR part. In LRASR, the dictionary is constructed based on the clustering algorithm, moreover, the coefficients matrix is constrained with low-rank and sparse in the representation model. The proposed DPLR method is based on data-driven projection LRR with dictionary construction. To verify the effectiveness of the data-driven projection LRR model, the low-rank representation-based methods (LRR and LRASR), which process the data in the original feature space, are selected as the comparison algorithms. Moreover, the performances of different dictionary construction methods can be demonstrated by the results of LRR, LRASR, and DPLR methods. To further present the performance of detectors in the low-dimensional feature space, the PCA_LRR method is adopted for comparisons.

In order to evaluate the performance of different detectors, the receiver operating characteristic (ROC) curve is adopted as a criterion [51]. The ROC curve on a 2-D plane depicts the changes in the true positive rate (TPR) with the false positive rate (FPR). The reference values of TPR and FPR are computed as

$$FPR = \frac{FP}{FP + TN}, TPR = \frac{TP}{TP + FN}. \quad (16)$$

In the above equation, FP denotes the false positive value and TN is the true negative value; TP and FN are true positive and false negative values, respectively. Generally, the ROC curve of a detector closing to the upper-left indicates that the detector is of favorable performance. Furthermore, the AUC (area under ROC curve) value is computed, it is a positive value and usually smaller than 1.0.

C. Detection Performance

First, the AUC values obtained by different methods on the four datasets are given in Table I. The results obtained by GRX are stable whose values are more than 0.90 since it is benefited by its global background estimation. The LRX method is suitable for HYDICE Urban and Pavia beach scene datasets, in which the anomaly size is not very large. The results of SR_LMM and CRD methods demonstrate the positive background estimation

TABLE I
AUC VALUES OBTAINED BY DIFFERENT METHODS ON THE FOUR DATASETS

	HYDICE Urban	AVIRIS Urban	Cri	Pavia Beach Scene
GRX	0.9857	0.9946	0.9135	0.9538
LRX	0.9492	0.8981	0.7419	0.9391
SR_LMM	0.9699	0.8902	0.7437	0.9159
CRD	0.9836	0.9461	0.8155	0.9553
RPCA	0.9785	0.9760	0.8135	0.9215
LRR	0.9672	0.9534	0.8023	0.9322
LRASR	0.9489	0.9812	0.8298	0.8767
PCA_LRR	0.9541	0.9248	0.7983	0.9185
DPLR	0.9933	0.9970	0.9361	0.9717

by the dual windows, especially on the HYDICE urban and Pavia beach scene datasets. The AUC values of RPCA method and LRR method demonstrate the acceptable detection results due to the exploited low-rank characteristic of the background. The results obtained by the PCA_LRR method are not so pleasurable compared with the LRR method since some crucial information about anomaly is abandoned after PCA. In Table I, the numbers in boldface show that our DPLR method obtains the best score. Compared with statistic-based methods of GRX and LRX, the DPLR method demonstrates the promising performances in the four datasets, especially in the datasets of Cri and Pavia, the AUC values obtained by DPLR method are much higher with 0.2 than GRX. The RPCA method based on low-rank decomposition cannot separate the background and anomaly completely. Compared with LRR and LRASR methods based on LRR, the DPLR method can generate a comprehensive and pure background dictionary to better separate anomaly further.

Then the ROC curves of comparison algorithms and DPLR method on different datasets are presented in Fig. 6(a)–(d), and the local amplified ROC curves are also displayed. For HYDICE urban dataset in Fig. 6(a), RPCA method can get better performance when the false alarm rate is smaller than 1×10^{-3} since it can exploit the prior information of background and anomaly; when the false alarm rate is equal to 5×10^{-3} , the detection probability of DPLR method is close to 0.8, which overcomes the other methods. For AVIRIS urban scene dataset in Fig. 6(b), we can see that when the false alarm rate is more than 1×10^{-3} , the performances of the DPLR method and CRD method are better than the other methods. For the datasets of Cri and Pavia in Fig. 6(c) and (d), when the false alarm rate is greater than 0.1, our DPLR method performs better demonstrating its effectiveness.

Moreover, the visualization detection results about the four datasets are shown in Fig. 7–10, in which (b)–(j) are the detection maps obtained by different methods. In the detection maps, the deeper color part represents the background and the brighter part refers to anomalies. From the detection maps of GRX and LRX, we cannot recognize all of the anomalies since the Gaussian assumption may not be proper for the real data. In Figs. 7(e) and 10(e), the CRD method performs better than the methods based on statistics. The results obtained by the RPCA method are acceptable. In Figs. 7(g), 8(g), and 10(g) obtained by the LRR method, several anomalous pixels cannot be distinguished from the background due that the background dictionary contains

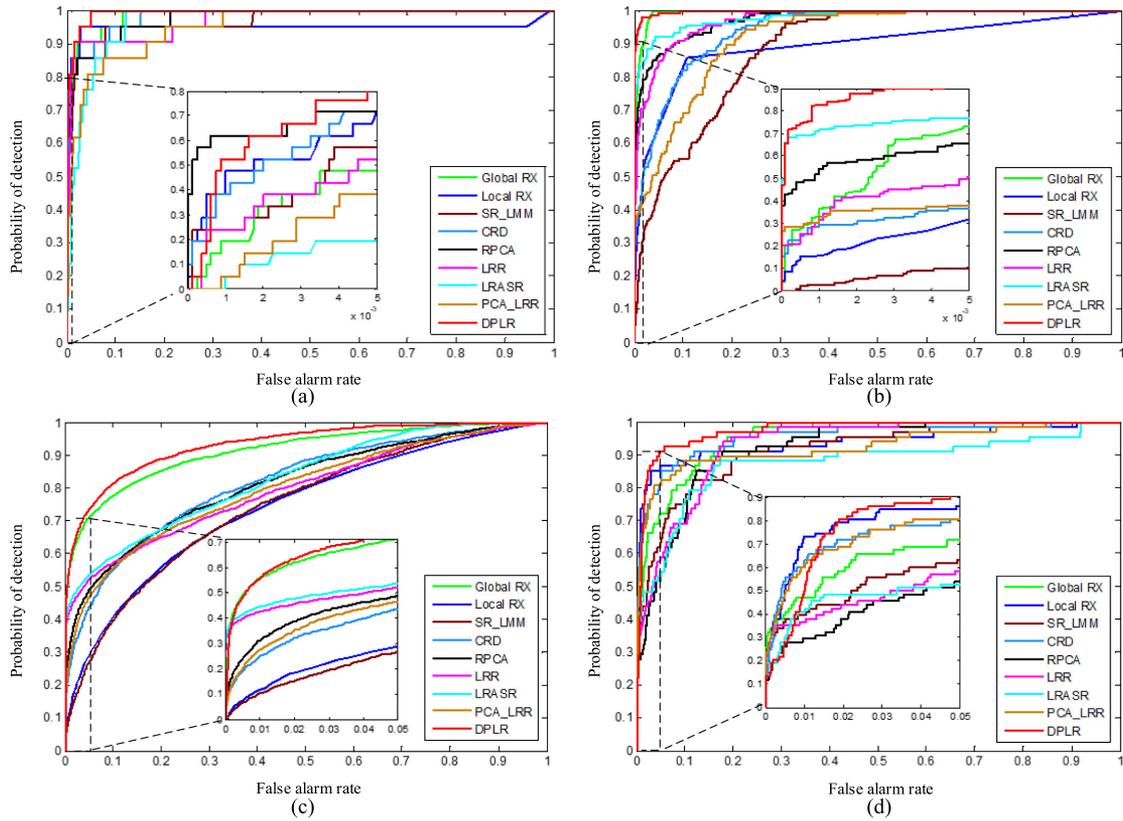


Fig. 6. ROC curves using different methods for the four datasets. (a) HYDICE urban dataset. (b) AVIRIS urban scene dataset. (c) Cri hyperspectral dataset. (d) Pavia beach scene dataset.

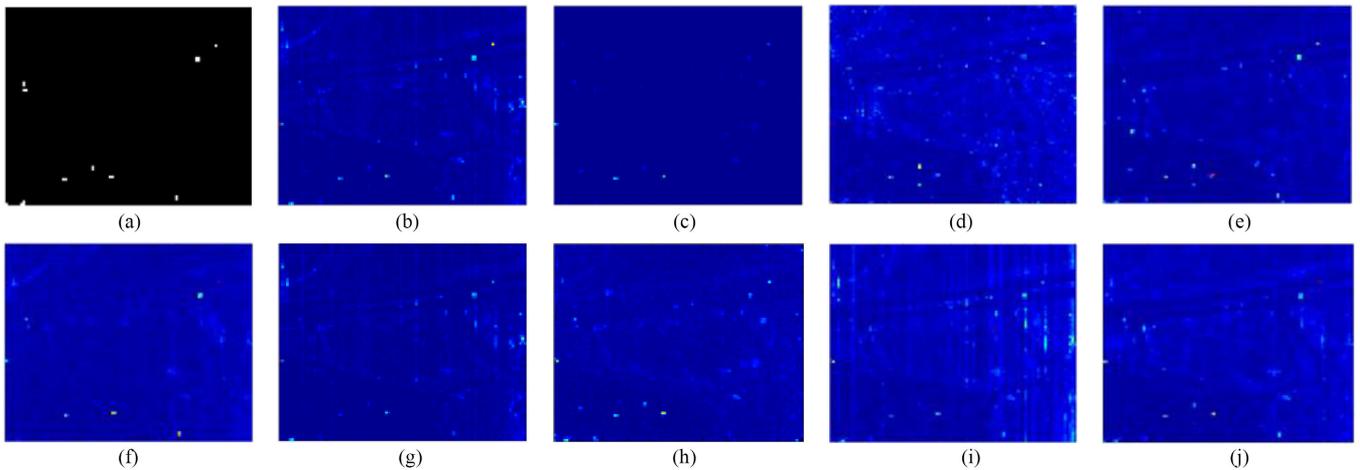


Fig. 7. Detection results using different methods for HYDICE urban dataset. (a) Reference. (b) GRX. (c) LRX. (d) SR_LMM. (e) CRD. (f) RPCA. (g) LRR. (h) LRASR. (i) PCA_LRR. (j) DPLR.

anomalous pixels. As the PCA_LRR method cannot ensure that the critical information of anomaly is preserved after PCA, the detection maps obtained by PCA_LRR are not so enjoyable, especially in Fig. 8(i). The detection maps of LRASR method and DPLR method are better than the others, and we can see that dictionary construction-based methods can efficiently depict the background to separate anomaly. Combining the results in Table I and the visualization detection maps in Figs. 7–10, we can

see that the performance of the DPLR method is better than the other methods and the detection results confirm its effectiveness for anomaly detection.

D. Parameters Settings and Computation Costs

The computer used for experiments is a 64-b quad-core Intel processor with CPU 3.20-GHz and 12 GB RAM under Windows 10. The development software we used is the MATLAB

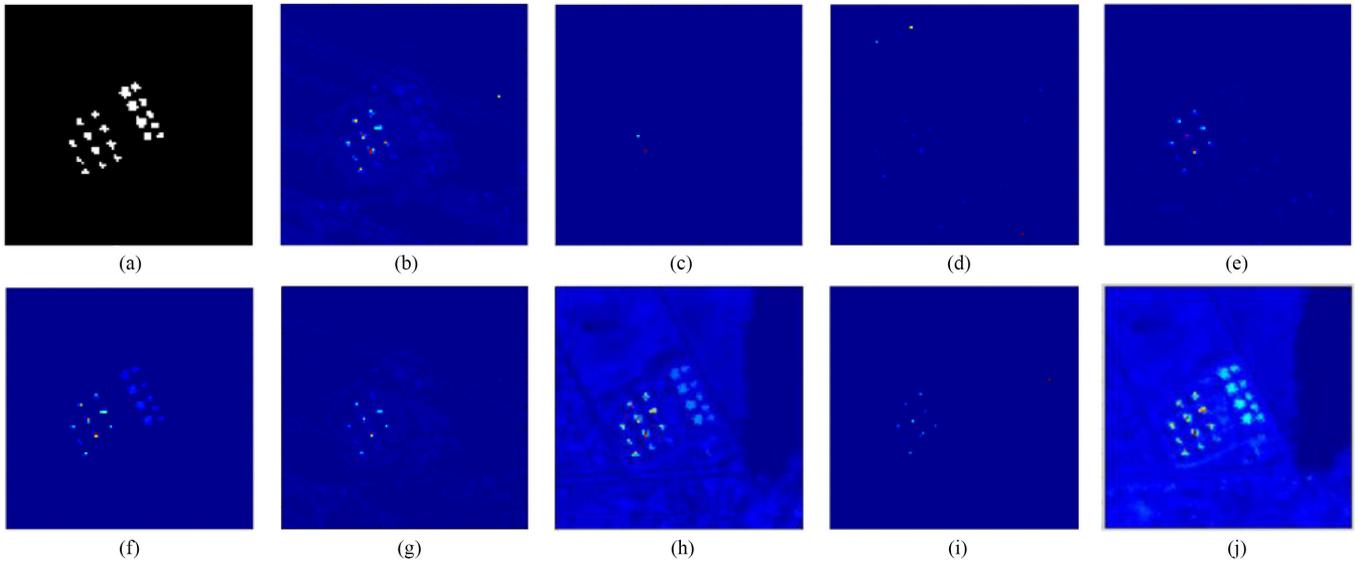


Fig. 8. Detection results using different methods for AVIRIS urban scene dataset. (a) Reference. (b) GRX. (c) LRX. (d) SR_LMM. (e) CRD. (f) RPCA. (g) LRR. (h) LRASR. (i) PCA_LRR. (j) DPLR.

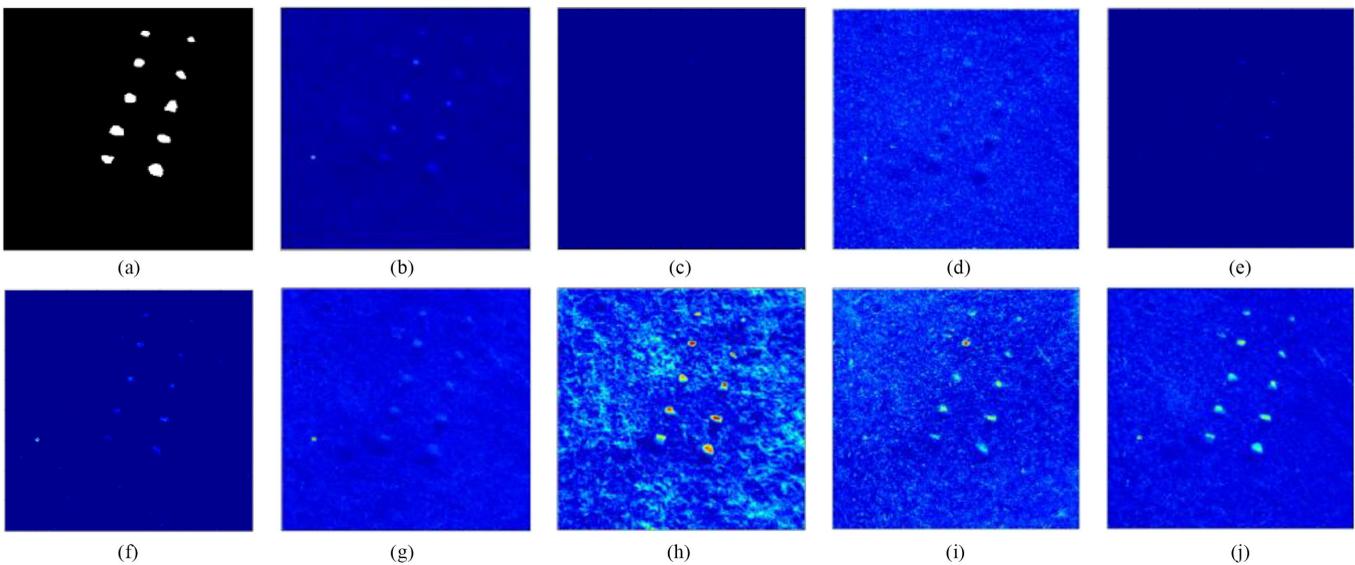


Fig. 9. Detection results using different methods for Cri hyperspectral dataset. (a) Reference. (b) GRX. (c) LRX. (d) SR_LMM. (e) CRD. (f) RPCA. (g) LRR. (h) LRASR. (i) PCA_LRR. (j) DPLR.

R2014a. Table II shows parameter settings and the running time comparisons of different methods on the four real hyperspectral datasets. As the performances of LRX, SR_LMM, and CRD methods are sensitive to the inner window size W_{in} and the outer window size W_{out} , we set W_{in} varying from 3 to 21 and W_{out} varying from 5 to 23 to experiment, the optimal window sizes of each dataset are selected. For the threshold parameter ν of SR_LMM, different values ranging in $(0, 0.25)$ are used in the experiment and the optimal value is selected. For the weight parameters of RPCA and LRR, we carry out the experiments with different values varying from 10^{-3} to 10, and the optimal parameters are recorded. For the LRASR method, just as the setting in [31], the number of clusters K and the number of the

selected atoms from each cluster P are set to be 15 and 20, respectively. The LRASR method is sensitive to the tradeoff parameter β and not sensitive to λ . In the experiment, λ is set to be 0.1 for all datasets. Varying β in $[0.01, 1]$, the optimal value for each dataset is recorded. For the PCA_LRR method, the reduced dimension after PCA is consistent with the DPLR method, and the parameter in the LRR is set as same as that in the LRR method.

We have analyzed the complexity of the DPLR algorithm in Section III. Compared with the original data space processing algorithms, our proposed algorithm is of lower computation complexity. Moreover, we compared the computation costs of different methods. In Table II, the running time of the DPLR

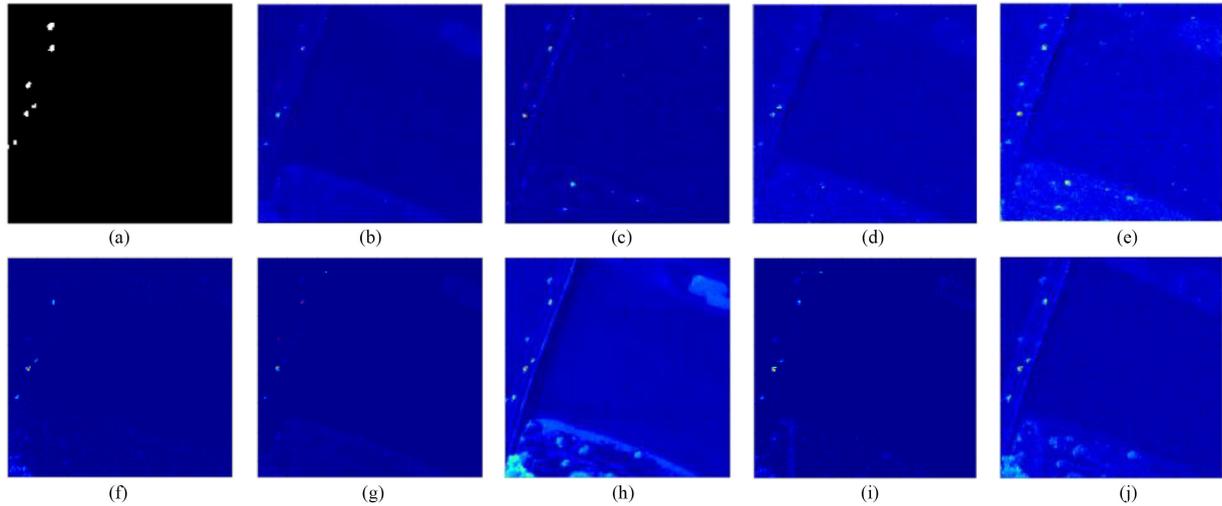


Fig. 10. Detection results using different methods for Pavia beach scene dataset. (a) Reference. (b) GRX. (c) LRX. (d) SR_LMM. (e) CRD. (f) RPCA. (g) LRR. (h) LRASR. (i) PCA_LRR. (j) DPLR.

TABLE II
PARAMETERS SETTINGS AND RUNNING TIME COMPARISONS

		HYDICE Urban	AVIRIS Urban	Cri	Pavia Beach Scene
GRX	Times	1.20	1.93	5.43	2.94
LRX	Parameters	$W_{in} = 7$ $W_{out} = 9$	$W_{in} = 7$ $W_{out} = 13$	$W_{in} = 17$ $W_{out} = 19$	$W_{in} = 7$ $W_{out} = 9$
	Times	32.41	72.36	124.43	30.50
SR_LMM	Parameters	$W_{in} = 7$ $W_{out} = 9$ $\nu = 0.01$	$W_{in} = 7$ $W_{out} = 11$ $\nu = 0.01$	$W_{in} = 13$ $W_{out} = 17$ $\nu = 0.05$	$W_{in} = 7$ $W_{out} = 9$ $\nu = 0.01$
	Times	40.29	93.78	109.57	49.71
CRD	Parameters	$W_{in} = 7$ $W_{out} = 13$	$W_{in} = 7$ $W_{out} = 9$	$W_{in} = 7$ $W_{out} = 13$	$W_{in} = 7$ $W_{out} = 9$
	Times	64.33	59.79	495.92	39.13
RPCA	Parameters	$\tau_0 = 0.01$	$\tau_0 = 0.01$	$\tau_0 = 0.01$	$\tau_0 = 0.01$
	Times	219.17	315.82	497.56	617.20
LRR	Parameters	$\tau_1 = 0.1$	$\tau_1 = 0.7$	$\tau_1 = 1.2$	$\tau_1 = 0.1$
	Times	60.51	146.19	1045.92	201.50
LRASR	Parameters	$\beta = 0.1$ $\lambda = 0.1$	$\beta = 0.5$ $\lambda = 0.1$	$\beta = 0.5$ $\lambda = 0.1$	$\beta = 0.05$ $\lambda = 0.1$
	Times	243.72	260.39	1226.84	152.92
PCA_LRR	Parameters	$b = 70$ $\tau_1 = 0.1$	$b = 100$ $\tau_1 = 0.7$	$b = 20$ $\tau_1 = 1.2$	$b = 40$ $\tau_1 = 0.1$
	Times	23.81	70.29	235.67	69.81
DPLR	Parameters	$\lambda = 1$ $b = 70$ $J = 20$ $K = 2$	$\lambda = 1$ $b = 100$ $J = 25$ $K = 2$	$\lambda = 1$ $b = 20$ $J = 35$ $K = 2$	$\lambda = 1$ $b = 40$ $J = 35$ $K = 2$
	Times	19.33	22.03	187.87	27.54

method is acceptable compared with other methods especially when processing the large data, such as the Cri dataset. In general, too high dimensions of data will bring additional computational costs. The traditional representation-based methods (such as SR_LMM, CRD, RPCA, LRR, and LRASR) process data in the original spectral dimensions. The introduced projection matrix in our model (5) can reduce the dimensions of data to compress the computation time. Thus, our proposed method has better practical application value for large datasets.

E. Effectiveness and Stability Analysis

To evaluate the effectiveness of the projection \mathbf{P} and the constructed dictionary \mathbf{D} on the detection result, three groups of

experiments for each dataset are performed. First, in our model (5), the projection \mathbf{P} is preserved, and a random dictionary [31] is used to experiment. Then we adopt the constructed dictionary \mathbf{D} without projection \mathbf{P} . Moreover, the DPLR method that includes the orthogonal projection \mathbf{P} and the novelty dictionary construction method is applied in the experiment. We repeat each group of experiments on each dataset 10 times. The AUC values are presented in Table III in which the upper and lower changes of AUC values are given. To analyze the effect of \mathbf{P} and \mathbf{D} visually, the AUC values of Table III are shown by the bar chart in Fig. 11.

From Table III, we can see that the upper and lower changes of AUC values in the second row are greater than the third and fourth rows. This is due to that the random dictionary is much more unstable than the dictionary \mathbf{D} constructed by our

TABLE III
 AUC VALUES AND THE CHANGES ON THE FOUR DATASETS

	HYDICE Urban	AVIRIS Urban	Cri	Pavia Beach Scene
P	0.9831 $\begin{smallmatrix} +0.0086 \\ -0.0328 \end{smallmatrix}$	0.9893 $\begin{smallmatrix} +0.0081 \\ -0.0192 \end{smallmatrix}$	0.8842 $\begin{smallmatrix} +0.0308 \\ -0.0370 \end{smallmatrix}$	0.9583 $\begin{smallmatrix} +0.0120 \\ -0.0213 \end{smallmatrix}$
D	0.9859 $\begin{smallmatrix} +0.0043 \\ -0.0071 \end{smallmatrix}$	0.9853 $\begin{smallmatrix} +0.0068 \\ -0.0096 \end{smallmatrix}$	0.9069 $\begin{smallmatrix} +0.0148 \\ -0.0168 \end{smallmatrix}$	0.9673 $\begin{smallmatrix} +0.0042 \\ -0.0037 \end{smallmatrix}$
P and D (DPLR)	0.9933 $\begin{smallmatrix} +0.0018 \\ -0.0017 \end{smallmatrix}$	0.9970 $\begin{smallmatrix} +0.0012 \\ -0.0020 \end{smallmatrix}$	0.9361 $\begin{smallmatrix} +0.0041 \\ -0.0061 \end{smallmatrix}$	0.9717 $\begin{smallmatrix} +0.0040 \\ -0.0065 \end{smallmatrix}$

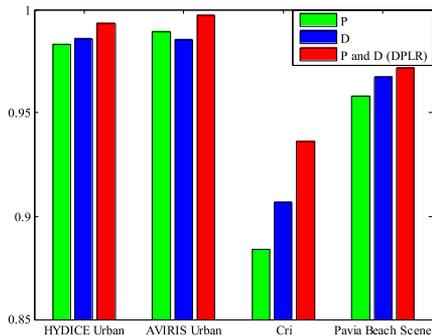


Fig. 11. AUC values on each dataset. The results of green bars are obtained by preserving projection matrix \mathbf{P} of (5) and a random dictionary. The blue bars are with the constructed dictionary \mathbf{D} without projection \mathbf{P} . The red bars are obtained by DPLR method that includes projection \mathbf{P} and the constructed dictionary \mathbf{D} .

method. For the DPLR method of the fourth row in Table III, we can observe that the changes between the maximum and minimum AUC values are within 0.011, and this confirms that our method is credible. In Fig. 11, from the green and blue bars of HYDICE urban and AVIRIS urban datasets, we can see that the factors of projection \mathbf{P} and the dictionary \mathbf{D} have a similar effect on the AUC results. On the Cri and Pavia beach scene datasets, the results demonstrate that the dictionary \mathbf{D} greatly improve the detection performance as our dictionary construction method has more advantages. In addition, the red bars in Fig. 11 demonstrate that our method with projection \mathbf{P} and the dictionary \mathbf{D} constructed by our method obtains the best result.

F. Parameters Analysis

In this part, we will discuss the impact of parameters on the detection results of the DPLR method. We divide the parameters into two groups for further discussion. First, we will discuss the impact of parameters λ and b of (5) on the four real datasets; then the parameters J and K for dictionary construction will be analyzed.

1) *Analysis of Parameters λ and b* : The impacts of parameters λ and b on the four real datasets are discussed in this section. In (5), λ is a parameter to balance the low-rank part and the sparse part, b is the reduced dimensionality. Fig. 12 illustrates the changes of AUC values when we jointly tune the parameters λ and b . In Fig. 12(a)–(d), the parameter λ is set as $\{0.01, 0.05, 0.1, 0.5, 1, 5, 10\}$. The parameter b in Fig. 12(a)–(d) is set as $\{20, 40, 60, 80, 100, 120, 140, 160\}$, $\{20, 40, 60, 80, 100, 120, 140, 160, 180, 200\}$, $\{10, 20, 30, 40\}$, and $\{20, 40, 60, 80, 100\}$,

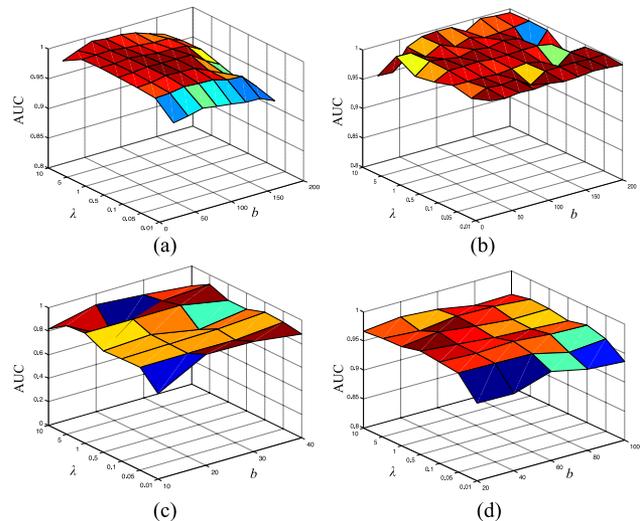


Fig. 12. AUC with the parameters tuning of λ and b . (a) HYDICE urban dataset. (b) AVIRIS urban dataset. (c) Cri hyperspectral dataset. (d) Pavia beach scene dataset.

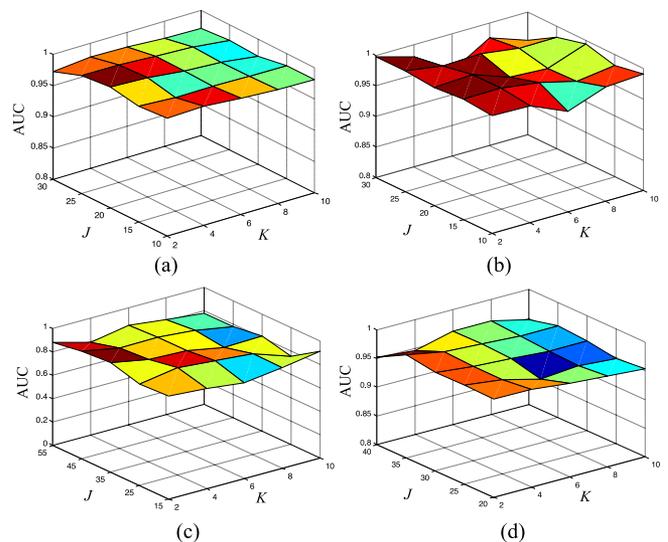


Fig. 13. AUC with the parameters tuning of J and K . (a) HYDICE urban dataset. (b) AVIRIS urban dataset. (c) Cri hyperspectral dataset. (d) Pavia beach scene dataset.

respectively. We can observe that the changes in AUC values are not significant when $\lambda \in [0.05, 5]$. For simplicity, in our experiments, we set $\lambda = 1$. When the projection dimension b is set to be the range of $1/3$ – $1/2$ of the original spectral dimension, we can achieve better detection performances.

2) *Analysis of Parameters J and K in Dictionary Construction*: Parameter J is the preset number of the superpixel. Generally, the real obtained number of superpixels is slightly more than the preset value J through TurboPixels superpixel segmentation. K is the number of randomly selected atoms from each superpixel. In Fig. 13, K is set as $\{2, 4, 6, 8, 10\}$. In Fig. 13(a)–(b), J is set as $\{10, 15, 20, 25, 30\}$. In Fig. 13(c)–(d), J is set as $\{15, 25, 35, 45, 55\}$ and $\{20, 25, 30, 35, 40\}$, respectively. Our

DPLR method is not sensitive to dictionary parameters J and K . In order to simplify the computation, in our experiment, we set $J \in [15, 35]$, $K \in [2, 6]$.

V. CONCLUSION

This article has presented a hyperspectral anomaly detection method based on a data-driven projection LRR algorithm with background dictionary construction, in which the superpixel segmentation and tensor decomposition are applied to construct a comprehensive and pure dictionary to better separate the background and anomalies. The proposed algorithm is examined on different datasets, and the experimental results indicate that the performance of our proposed algorithms is better than that of several existing algorithms.

There exist limitations as the dictionary construction and the low-rank decomposition are two separate steps, the detection performance depends on the constructed background dictionary. In addition, converting the original 3-D hyperspectral data cube into 2-D matrix will inevitably lose some structure information. Tensor representation and deep learning technologies have been successfully applied to HSI analysis. To take a closer look at the low-rank characteristics and deep features, we will use deep tensor models in our further work.

REFERENCES

- [1] 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, no. 2, pp. 6–36, Jun. 2013.
- [2] Y. Zhong *et al.*, "Mini-UAV-borne hyperspectral remote sensing: From observation and processing to applications," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 4, pp. 46–62, Dec. 2018.
- [3] J. An, X. Zhang, H. Zhou, and L. Jiao, "Tensor-based low-rank graph with multimanifold regularization for dimensionality reduction of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4731–4746, Aug. 2018.
- [4] X. Zhang *et al.*, "Hyperspectral unmixing via low-rank representation with space consistency constraint and spectral library pruning," *Remote Sens.*, vol. 10, no. 2, p. 339, Feb. 2018.
- [5] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2017.
- [6] W. Li, F. Feng, H. Li, and Q. Du, "Discriminant analysis-based dimension reduction for hyperspectral image classification: a survey of the most recent advances and an experimental comparison of different techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 6, no. 1, pp. 15–34, Mar. 2018.
- [7] H. Ren and C.-I. Chang, "Automatic spectral target recognition in hyperspectral imagery," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1232–1249, Oct. 2003.
- [8] H. Su, Z. Wu, Q. Du, and P. Du, "Hyperspectral anomaly detection using collaborative representation with outlier removal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 5029–5038, Dec. 2018.
- [9] R. Tao, X. Zhao, W. Li, H. Li, and Q. Du, "Hyperspectral anomaly detection by fractional fourier entropy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4920–4929, Dec. 2019.
- [10] D. Ma, Y. Yuan, and Q. Wang, "Hyperspectral anomaly detection via discriminative feature learning with multiple-dictionary sparse representation," *Remote Sens.*, vol. 10, no. 5, p. 745, May 2018.
- [11] 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, Jul. 2019, Art. no. 1578.
- [12] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
- [13] 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.
- [14] C.-I. Chang and S.-S. Chiang, "Anomaly detection and classification for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 6, pp. 1314–1325, Jun. 2002.
- [15] S. Khazai, A. Safari, B. Mojaradi, and S. Homayouni, "An approach for subpixel anomaly detection in hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 769–778, Apr. 2013.
- [16] W. Liu and C. I. Chang, "Multiple-window anomaly detection for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 644–658, Apr. 2013.
- [17] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
- [18] L. Wei and D. Qian, "Collaborative representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1463–1474, Mar. 2015.
- [19] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Nonlocal patch tensor sparse representation for hyperspectral image super-resolution," *IEEE Trans. Image Process.*, vol. 28, no. 6, pp. 3034–3047, Jun. 2019.
- [20] Y. Zhao and J. Yang, "Hyperspectral image denoising via sparse representation and low-rank constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 296–308, Jan. 2015.
- [21] J. Yang, Y. Zhao, J. C. W. Chan, and S. G. Kong, "Coupled sparse denoising and unmixing with low-rank constraint for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 3, pp. 1818–1833, Mar. 2016.
- [22] S. Jia, X. Zhang, and Q. Li, "Spectral-spatial hyperspectral image classification using $l_{1/2}$ regularized low-rank representation and sparse representation-based graph cuts," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2473–2484, Jun. 2015.
- [23] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Joint reconstruction and anomaly detection from compressive hyperspectral images using mahalanobis distance-regularized tensor RPCA," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2919–2930, May 2018.
- [24] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.
- [25] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Object detection in optical remote sensing images based on weakly supervised learning and high-level feature learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3325–3337, Jun. 2015.
- [26] W. Li, G. Wu, and Q. Du, "Transferred deep learning for anomaly detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 597–601, May 2017.
- [27] S. Y. Chen, S. Yang, K. Kalpakis, and C. I. Chang, "Low-rank decomposition-based anomaly detection," in *Proc. SPIE*, vol. 8743, 2013, Art. no. 87430N.
- [28] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis," *J. ACM*, vol. 58, no. 3, pp. 1–39, May 2011.
- [29] 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.
- [30] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. Int. Conf. Mach. Learn.*, 2010, pp. 663–670.
- [31] 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.
- [32] 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.
- [33] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 118–139, Jun. 2019.
- [34] L. Wang *et al.*, "Band subset selection for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 4887–4898, Sep. 2017.
- [35] Y. Qu *et al.*, "Hyperspectral anomaly detection through spectral unmixing and dictionary-based low-rank decomposition," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4391–4405, Aug. 2018.
- [36] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: Applications to image and text data," in *Proc. 7th ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2001, pp. 245–250.

- [37] J. C. Harsanyi and C.-I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.
- [38] J. E. Fowler and Q. Du, "Anomaly detection and reconstruction from random projections," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 184–195, Jan. 2012.
- [39] M. Golbabae and P. Vanderghyest, "Hyperspectral image compressed sensing via low-rank and joint-sparse matrix recovery," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2012, pp. 2741–2744.
- [40] Z. Lin, R. Liu, and H. Li, "Linearized alternating direction method with parallel splitting and adaptive penalty for separable convex programs in machine learning," *Mach. Learn.*, vol. 99, no. 2, pp. 287–325, 2015.
- [41] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graph. Statist.*, vol. 15, no. 2, pp. 262–286, 2006.
- [42] M. Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2011, pp. 2097–2104.
- [43] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [44] A. Levinstein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. Dickinson, and K. Siddiqi, "Turbopixels: Fast superpixels using geometric flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2290–2297, Dec. 2009.
- [45] I. T. Jolliffe, *Principal Component Analysis*. New York, NY, USA: Springer-Verlag, 2010.
- [46] N. Renard, S. Bourennane, and J. Blanc-Talon, "Denoising and dimensionality reduction using multilinear tools for hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 138–142, Apr. 2008.
- [47] L. de Lathauwer, B. de Moor, and J. Vanderwalle, "On the best rank-1 and rank- (r_1, r_2, \dots, r_m) approximation of higher-order tensors," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1324–1342, Mar. 2000.
- [48] N. Renard, S. Bourennane, and J. Blanc-Talon, "Multiway filtering applied on hyperspectral images," in *Proc. Int. Conf. Adv. Concepts Intell. Vis. Syst.*, Sep. 2006, vol. 4179, pp. 127–137.
- [49] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.
- [50] Q. Ling, Y. Guo, Z. Lin, and W. An, "A constrained sparse representation model for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2358–2371, Apr. 2019.
- [51] D. Manolakis and G. A. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.



Xiaoxiao Ma received the B.S. degree in communication engineering from Shijiazhuang University, Shijiazhuang, China, in 2011, and the M.S. degree in pattern recognition and intelligent system from the Southwest University of Science and Technology, Mianyang, China, in 2014. She is currently working toward the Ph.D. degree in electronic circuit and system from the School of Artificial Intelligence, Xidian University, Xian, China.

Her research interests include pattern recognition, machine learning, and hyperspectral image processing.



Xiangrong Zhang (Senior Member, IEEE) received the B.S. and M.S. degrees in computer application technology from the School of Computer Science, Xidian University, Xi'an, China, in 1999 and 2003, respectively, and the Ph.D. degree in pattern recognition and intelligent system from the School of Electronic Engineering, Xidian University, in 2006.

She is currently a Professor with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University. From January 2015 to March 2016, she was

a Visiting Scientist with the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology. Her research interests include pattern recognition, machine learning, and remote sensing image analysis and understanding.



Xu Tang (Member, IEEE) received the B.S., M.S., and Ph.D. (electronic circuit and systems) degrees from Xidian University, Xi'an, China, in 2007, 2010, 2017, respectively.

He is currently a member of the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education, Xidian University. His research interests include synthetic aperture radar (SAR) image processing and content-based retrieval.



Huiyu Zhou received the B.Eng. degree in radio technology from the Huazhong University of Science and Technology, Wuhan, China, in 1990, the M.S. degree in biomedical engineering from the University of Dundee, Dundee, U.K., in 2002, and the Ph.D. degree in computer vision from Heriot Watt University, Edinburgh, U.K., in 2006.

He is currently a Reader with the Department of Informatics, University of Leicester, Leicester, U.K. He has taken part in the consortiums of a number of research projects in medical image processing,

computer vision, intelligent systems, and data mining.



Licheng Jiao (Fellow, IEEE) received the B.S. degree from Shanghai Jiaotong University, Shanghai, China, in 1982, and the M.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1984 and 1990, respectively. He has authored or coauthored more than 150 scientific papers.

He has charged of about 40 important scientific research projects, and published more than 20 monographs and 100 papers in international journals and conferences. His research interests include image

processing, natural computation, machine learning, and intelligent information processing.