

# Local Spatial Constraint and Total Variation for Hyperspectral Anomaly Detection

Ruyi Feng<sup>ID</sup>, Member, IEEE, Hao Li<sup>ID</sup>, Lizhe Wang, Fellow, IEEE, Yanfei Zhong<sup>ID</sup>, Senior Member, IEEE,  
Liangpei Zhang<sup>ID</sup>, Fellow, IEEE, and Tieyong Zeng<sup>ID</sup>

**Abstract**—Hyperspectral anomaly detection, which is aimed at locating anomaly, has received widespread attention. In this article, a new anomaly detector, named local spatial constraint and total variation (LSC-TV), is proposed for hyperspectral imagery. In anomaly detection methods based on low-rank representation, background pixels are usually considered to have a global low-dimensional structure. However, the complex background distribution in hyperspectral images (HSIs) means that this global low-dimensional structure rarely occurs. In LSC-TV, the effective local spatial information is extracted by superpixel segmentation, and the regularization based on the F-norm is used to force the background within the same superpixel to show uniform spectral features. Moreover, each pixel is given a penalty based on the degree of anomaly determined during model iteration, while the anomaly is not considered by the background constraint. In addition, the background pixels in the neighborhood often show a high correlation, whereas the anomaly does not possess this feature. Nonisotropic TV is introduced into the proposed LSC model using the correlation of first-order neighborhoods to make it easier for anomalies to be separated. The proposed LSC-TV method and current state-of-the-art methods are tested on a set of simulated data and four sets of real data. The experimental results demonstrate that the proposed method is superior to the comparative method in terms of both color map detection and quantitative evaluation.

**Index Terms**—Anomaly detection, hyperspectral imagery, local idea, matrix decomposition, spatial information, superpixel segmentation, total variation.

## I. INTRODUCTION

WITH the great development of remote sensing imaging technology, hyperspectral remote sensors now have the capability of capturing the reflective electromagnetic spectrum in hundreds of contiguous and narrowbands [1]–[3]. The

Manuscript received April 20, 2021; revised June 11, 2021; accepted July 12, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 41925007, Grant 41701429, and Grant U1711266; in part by the Hong Kong Scholars Program under Grant XJ2020025; and in part by the Hubei Key Laboratory of Regional Development and Environmental Response (Hubei University) under Grant 2020(B)003. (Corresponding authors: Hao Li; Lizhe Wang.)

Ruyi Feng, Hao Li, and Lizhe Wang are with the School of Computer Science, China University of Geosciences, Wuhan 430074, China, and also with Hubei Key Laboratory of Intelligent GeoInformation Processing, China University of Geosciences, Wuhan 430074, China (e-mail: fengry@cug.edu.cn; lihao@cug.edu.cn; lizhe.wang@gmail.com).

Yanfei Zhong and Liangpei Zhang are with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University, Wuhan 430079, China.

Tieyong Zeng is with the Department of Mathematics, The Chinese University of Hong Kong, Hong Kong, China.

Digital Object Identifier 10.1109/TGRS.2021.3098814

abundant spectral information and available spatial distribution information of hyperspectral images (HSIs) make it possible to determine the spatial location of targets of interest based on distinct spectral signatures [4]. Due to these unique advantages, HSIs have been widely used in many practical applications, such as environmental monitoring, agricultural practices, and military defense, among others [5], [6].

Target detection, as an important research area of HSIs, has attracted substantial attention. Depending on whether the prior spectral information of the targets is available, target detection can be divided into supervised detection and unsupervised detection (i.e., anomaly detection). Since the spectra of the targets are often difficult to grasp, it is more convenient to put anomaly detection into practical applications, such as the distinction of man-made material in military defense, the discovery of rare crops in agriculture, and the detection of minerals [7], [8]. In short, anomaly detection has received increasing attention in HSI analysis. Since HSIs are often accompanied by noise and complex distributions, the extraction of anomalies can be more difficult. Fortunately, due to the distinguished characteristics of the anomaly, it can be separated from complex HSIs. Anomalies have a low probability of appearing in HSIs. Meanwhile, compared with the background spectra, anomalies have distinct spectral signatures that significantly deviate from their neighborhood [9]. Considering this information, anomaly detection can be modeled as an unsupervised binary classification problem.

Due to efficient computing and the exclusion of parameters, the global Reed–Xiaoli (RX) [10] method is the most popular statistical-based anomaly detection method, and it has been accepted as a benchmark algorithm for anomaly detection. RX assumes that the probability density function of background is consistent with a multivariate Gaussian distribution and that the final detection result will depend on the Mahalanobis distance between the under-testing pixel and the background. However, the background obeys a single normal distribution, which rarely occurs in complex HSI scenes. Meanwhile, the statistics concerning noise will cause the center of the background to deviate from the real sample center. Therefore, some improved detectors based on the RX method have been proposed, such as the local RX (LRX) [11], [12], subspace RX [13], regularized RX [14], weighted RX [15], random selection-based anomaly detector (RSAD) [16], blocked adaptive computationally efficient outlier nominators (BACONs) [17], kernel RX [18], and support

vector data description (SVDD) [19]. Some of these methods achieve the robustness of the detection model by improving the accuracy of background estimation. Others use kernel theory to deal with data in high-dimensional feature space. Besides RX and its improved methods, there are also some widely used anomaly detection methods. Wu *et al.* [20] proposed scalable low-dimensional manifold modeling (SLDMM) method by using the strategy of multiple random sampling reconstruction. Li *et al.* [21] introduced the isolation forest idea into hyperspectral anomaly detection and proposed the kernel isolation forest detection (KIFD) method.

In addition to the methods mentioned above, some representation-based detectors have attracted attention. The **background joint sparse representation detection** (BJSRD) detector [22] assumes that the background pixel can be reconstructed by a small number of atoms linear combination in an overcomplete background dictionary. Anomalies are identified by measuring the residuals of the reconstructed spectral and the tested pixel spectral. In [23], a joint sparse representation and multitask learning anomaly detection method has been proposed. However, these methods only focus on spectral information while ignoring the spatial correlation between pixels. Based on the fact that background pixels can be obtained by a linear combination of surrounding pixels, but anomaly cannot, an anomaly detection method, called the collaborative representation detector (CRD) [24], was proposed. By defining the double windows, the pixels between the inner and outer windows are used to form the background dictionary and to minimize the  $l_2$ -norm of the coefficients to complete the reconstruction. **Consistent with BJSRD, the determination of anomaly is achieved by calculating the residuals of the reconstructed spectral and the measured pixel spectral.** However, the size selection of the internal and external windows greatly affects the detection results.

Recently, inspired by robust principal component analysis (RPCA) [25], matrix (or tensor) decomposition has been proven to be an effective tool in anomaly detection. The **RPCA method assumes that the background data should be located in a low-dimensional subspace and that outliers can be separated with sparse constraints.** Based on this idea, HSIs can be decomposed into a low-rank background matrix and a sparse anomaly matrix. Li *et al.* [26] proposed a tensor decomposition method based on RPCA, which uses a low-rank tensor to represent the background subspace, in which anomaly is effectively separated. In order to avoid instability among the extracted sparse components, the **low-rank and sparse matrix decomposition-based Mahalanobis distance** (LSMAD) [27] method has been proposed to calculate the Mahalanobis distance in stable background subspace. However, RPCA considers the background data to be located in a single subspace, which rarely holds true for complex HSIs. Furthermore, some noise points isolated in different bands will also be separated into anomaly components, increasing the false alarm rate of the detector. Low-rank representation (LRR) [28] as a new technique can effectively solve this problem. The input data are used as a background dictionary, and the mixed  $l_{2,1}$ -norm constraint is imposed on the anomaly component. LRR can effectively resolve the predicament in

which hyperspectral data are distributed into multiple subspaces. To increase the robustness of the model, the low-rank and sparse representation (LRASR) [29] method introduces a background dictionary and imposes sparse constraints on the representation coefficient. To improve detection performance in highly mixed HSIs, Qu *et al.* [30] proposed a new detector named abundance- and dictionary-based low-rank (ADLR) decomposition. In ADLR, an LRR is used on the abundance matrix instead of the original HSIs. By constructing a potential background dictionary, the potential information about anomaly in HSIs can be maximized in [31]. In order to maintain the background subspace correlation, Cheng and Wang [32] applied both manifold regularization and total variation regularization to anomaly detection. However, these LRR-based detectors ignore the local structure of background data. As a result, background component extraction is not accurate.

To overcome the shortcomings of the above methods, this article proposes a new anomaly detection method, namely the local spatial constraint and total variation (LSC-TV) algorithm, which is based on superpixel segmentation and the total variation (TV) method. The idea of superpixel segmentation has been widely combined with sparse representation, LRR, and TV to explore the local structure of hyperspectral data and has achieved good performance in HSI analysis, such as HSIs denoising [33], HSIs super-resolution [34], and HSIs classification [35]. To extract local structure information from the HSIs, a superpixel segmentation method for HSIs is introduced to obtain a uniformly distributed area of the features. Within the same superpixel, the background pixels have similar spectral characteristics, but the anomalies differ greatly. By maintaining the local structure of background pixels, the isolated anomalies can be better separated from the HSI. Moreover, TV is introduced into the model to promote the smoothness of the background. While using the correlation of adjacent pixels, some isolated anomalies can be effectively removed from the background.

Therefore, the main contributions of this article can be summarized as follows.

- 1) Compared with traditional LRR-based detectors, the local structure of the background is better considered in the proposed LSC-TV method. Superpixel segmentation is used to obtain the spatial information of HSIs. The local spatial structure of the background is maintained by an effective penalty term. Meanwhile, TV is introduced to promote the smoothness of the coefficients and to exclude some anomalies from the background.
- 2) To better exclude anomaly from the background, the LSC term considers the anomaly weighting of the pixels. Some uniformly distributed pixels will be given greater weight in the LSC, helping them to effectively remain in the background. These anomalies will not be considered by the LSC term, helping them to be separated into the sparse component. The alternating direction method of multiplier (ADMM) algorithm is used to improve the accuracy and efficiency of the solution.

The remainder of this article is structured as follows. The LSC-TV model and optimization algorithm are introduced in Section II. Section III provides the experiments and analyses with synthetic and real images. Section IV draws some conclusions.

## II. LSC-TV FOR ANOMALY DETECTION

### A. LRR for Anomaly Detection

Consider an HSI data  $\mathcal{X} \in \mathbb{R}^{r \times c \times L}$ , where  $r$  and  $c$  are the spatial sizes and  $L$  is the number of bands of the HSI data cube. For convenience, this HSI data cube can be expanded into a matrix  $\mathbf{X} \in \mathbb{R}^{L \times N}$  by pixels  $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ , where  $N = r \times c$  represents the total number of pixels in the HSI and  $\mathbf{x}_i$  represents the spectral of the  $i$ th pixel. This LRR treats the detection task as a matrix decomposition process

$$\mathbf{X} = \mathbf{DZ} + \mathbf{E} \quad (1)$$

where  $\mathbf{DZ}$  represents the background component and  $\mathbf{D} = \{\mathbf{d}_i\}_{i=1}^m$ ,  $\mathbf{d}_i \in \mathbb{R}^{L \times 1}$  represents the pretrained dictionary.  $\mathbf{Z} = \{\mathbf{z}_i\}_{i=1}^N$ ,  $\mathbf{z}_i \in \mathbb{R}^{m \times 1}$  represents the representation coefficient of the dictionary. When  $\mathbf{D} = \mathbf{I}$ , the model is degraded to RPCA. In a typical LRR method,  $\mathbf{D} = \mathbf{X}$ . However, when the amount of input data is large, the computation cost becomes a problem worth considering.  $\mathbf{E}$  represents the rare pixels (i.e., anomaly target) and the noisy or modeling errors. For  $\mathbf{DZ}$ , a lot of prior spatial information is available here because of the uniform distribution of the background. For example, the background satisfies the local similarity of homogeneous regions [36], nonlocal similarity [37], piecewise smoothness [38], and low-rank property. Based on the fact that  $\mathbf{DZ}$  exists in a low-dimensional subspace, LRR applies low-rank constraints on the background components that ignore the spatial structure. In addition to the low probability of anomalies in the entire image, it is a fact that the anomalies have distinctive spectra in all bands compared to the surroundings, meaning that  $\mathbf{E}$  should be, columnwise, sparse. Based on the above analysis, the standard LRR model is given as follows:

$$\min_{\mathbf{Z}, \mathbf{E}} \text{rank}(\mathbf{Z}) + \beta \|\mathbf{E}\|_{2,1} \quad \text{s.t. } \mathbf{X} = \mathbf{DZ} + \mathbf{E} \quad (2)$$

while  $\beta \geq 0$  is the tradeoff coefficient needed to balance the background and anomaly components,  $\text{rank}(\cdot)$  is the rank function, and  $\|\cdot\|_{2,1}$  represents the mixed  $l_{2,1}$ -norm, which promotes columnwise sparsity among the columns of  $\mathbf{E}$  [39]. The definition of this mixed norm is given as follows:

$$\|\mathbf{E}\|_{2,1} = \sum_{i=1}^N \sqrt{\sum_{j=1}^L e_{j,i}^2} \quad (3)$$

where  $e_{j,i}$  is the entry of the  $j$ th row and column  $i$ th in  $\mathbf{E}$ . Since the uniform pixels (i.e., background) are constrained into multiple subspaces by minimizing the rank of a matrix,  $\|\cdot\|_{2,1}$  encourages these background pixels to be equal to zero in  $\mathbf{E}$ . Consequently, the anomaly component  $\mathbf{E}$  naturally collects spectral vectors of small- and low-probability ground objects (i.e., anomaly target). However, it is obviously impractical to directly solve problem (2). The rank function of a matrix is a discontinuous and nonconvex function. The process of

minimizing rank is a typical NP-hard problem. Fortunately, the nuclear norm [40], which is the convex function closest to the rank function, can be used to convexly relax the problem (2). Problem (2) is transformed into the following:

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

where  $\|\cdot\|_*$  represents the nuclear norm and is the sum of the values of the singular values of the matrix. Through the linearized alternating direction method with adaptive penalty (LADMAP) [41], problem (4) can be solved as a standard convex optimization problem. Once the optimal  $\mathbf{E}^*$  is found, the final detection result will be obtained immediately. The determination of anomaly is realized by computing the  $l_2$ -norm of each column in  $\mathbf{E}^*$  [29], [30]. If it is greater than the threshold set in advance, this column (i.e., pixel) will be judged as comprising anomaly. A more specific definition is given as follows:

$$\|\mathbf{E}^*(\cdot, i)\|_2 \geq \eta, \quad 1 \leq i \leq N \quad (5)$$

where  $\eta$  is a preset threshold.

### B. LSC-TV Anomaly Detection

The LRR can handle the problem of data being in multiple subspaces, and the global structure of the background data can be maintained by minimizing the nuclear norm. However, the local spatial structure and the correlation between the pixels of HSIs have not been well considered, which leads to imperfect background extraction. Some anomaly may be classified into the background to reduce the detection rate—otherwise, these pixels, which should belong to the background, can be misclassified into the anomaly component, which can, in turn, increase the false alarm rate. An anomaly detection method that can consider local spatial relationships is therefore proposed in this article.

For HSIs, pixels are mostly uniformly distributed in homogeneous regions rather than globally. This proves that these pixels will share a similar spectral, which means that in dictionary-based representation methods, the coefficients of the dictionary should be extremely similar in these homogeneous regions [42], [43]. The spectral of the anomaly in these regions should be large-contrast and easy to distinguish, and the representation coefficients of these pixels should seriously deviate from the coefficients of other uniform pixels. Taking advantage of this fact, anomaly can be better separated from the background.

Superpixel is a group of adjacent pixels with similar color, texture, and other characteristics [44]. Homogeneous regions are often obtained by superpixel segmentation considering spatial information and spectral information. The simple linear iterative clustering (SLIC) [45] method, which is a faster, much more memory-efficient superpixel segmentation algorithm, exhibits state-of-the-art boundary adherence and is a common method for generating superpixels. In this method, location information and color (or spectral) information can both be better considered. In addition, very few preset parameters, including only one parameter with a presegmented number of superpixels, are involved in this algorithm. For the above

reasons, the SLIC-based superpixel segmentation method is extended from RGB images to HSIs to acquire homogeneous regions in the proposed LSC-TV method. The SLIC algorithm can be viewed as a localized version of the  $k$ -means clustering method. Different from the  $k$ -means method, SLIC fixes the search window to a local area during the process of merging similar pixels, considering the spatial relationship. In traditional RGB images, the distance measurement of the SLIC method is based on comprehensive consideration of pixel brightness information and position information. In HSIs, the distance measurement can be changed to a compromise between spectral similarity and spatial proximity. The specific definitions are given as follows:

$$\begin{aligned} d_{\text{spatial}}(i, j) &= \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \\ d_{\text{spectral}}(i, j) &= \|y_i - y_j\|_2^2 \\ D' &= \sqrt{\left(\frac{d_{\text{spatial}}}{S}\right)^2 + \left(\frac{d_{\text{spectral}}}{t}\right)^2} \end{aligned} \quad (6)$$

where  $\mathbf{x}_i$  and  $\mathbf{x}_j$  represent the spectral vectors of  $i$ th and  $j$ th pixels, respectively,  $(x_i, y_i)$  and  $(x_j, y_j)$  represent the spatial coordinates of the  $i$ th and  $j$ th pixels, respectively,  $S$  is the preset size of the superpixel, and  $t$  is the tradeoff parameter used to balance the spatial distance and the spectral distance. It is worth noting that during the execution of the SLIC algorithm, the anomaly is far from the cluster center of each superpixel due to its distinctive spectral characteristics. At the end of each iteration, the anomaly will be merged into the nearby superpixels.

Through the above analysis, the anomaly detection model considering only local space constraint can be defined as

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}} \frac{1}{2} \|(\mathbf{Z} - \hat{\mathbf{Z}})\mathbf{W}\|_F^2 + \beta \|\mathbf{E}\|_{2,1} \\ \text{s.t. } \mathbf{X} = \mathbf{DZ} + \mathbf{E} \end{aligned} \quad (7)$$

where  $\|\mathbf{X}\|_F^2 = \text{trace}(\mathbf{XX}^T)$  is the Frobenius norm of  $\mathbf{X}$ .  $\hat{\mathbf{Z}} \in \mathbb{R}^{m \times N}$  is a guided matrix containing local spatial information, and  $\mathbf{W} = \text{diag}(\mathbf{w})$  is a weighted diagonal matrix based on anomaly probability. The solution process of  $\hat{\mathbf{Z}}$  and  $\mathbf{W}$  will be introduced in detail in the following.

The HSI-based SLIC method, as applied on HSI  $\mathbf{X}$ , can be obtained  $\mathbf{X} = \{\mathbf{X}^1, \mathbf{X}^2, \dots, \mathbf{X}^k\}$ , where  $k$  is the total number of superpixels,  $\mathbf{X}^i \in \mathbb{R}^{L \times n}$  is the  $i$ th superpixel, and  $n$  is the number of pixels contained in  $\mathbf{X}^i$ . By sharing the position information of the pixels contained in each superpixel into the coefficient matrix  $\mathbf{Z}$ , the same superpixel structure can be obtained:  $\mathbf{Z} = \{\mathbf{Z}^1, \mathbf{Z}^2, \dots, \mathbf{Z}^k\}$ ,  $\mathbf{Z}^i \in \mathbb{R}^{m \times n}$ . Through the previous analysis, the coefficient matrix in each superpixel (i.e.,  $\mathbf{Z}^i$ ) in the background pixel should share similar coefficient values. For the coefficient  $\hat{\mathbf{z}}^{(i)}$  of the  $i$ th superpixel, an estimated overall coefficient value can be derived as follows:

$$\hat{\mathbf{z}}^{(i)} = \frac{\sum_{j=1}^n \mathbf{w}_j^i \mathbf{z}_j^i}{\|\mathbf{w}^i\|_1} \quad (8)$$

where  $(\cdot)^i$  represents the set of pixel entries contained in the  $i$ th superpixel,  $(\cdot)_j^i$  represents the  $j$ th pixel entry in the  $i$ th superpixel, and  $\|\cdot\|_1$  is the  $l_1$  norm, which is used to unify

the scale of weights.  $\mathbf{w} = \exp(-(\|\mathbf{E}(:, i)\|_2/\sigma)) \in \mathbb{R}^{1 \times N}$ ,  $i = 1, 2, \dots, N$ , is a weighted vector based on anomaly. The entry of the background pixel in  $\mathbf{E}$  is close to zero, and the corresponding  $\mathbf{w}_j^i$  is close to one. However, the anomaly has a larger entry in  $\mathbf{E}$ , which makes  $\mathbf{w}_j^i$  yield a smaller value and  $\sigma$  represents the Gaussian kernel used to control the range of  $\mathbf{w}$ . It can be seen from (8) that the overall coefficient value of the current superpixel is obtained by the weighted average of all background pixels in this superpixel. Anomaly is penalized by a smaller weight, which reduces interference with the background. This guided matrix  $\hat{\mathbf{Z}}$  can be expanded to the same size as the original coefficient matrix  $\mathbf{Z}$ , which is solved as follows:

$$\hat{\mathbf{Z}}^i(:, 1:n) = \hat{\mathbf{z}}^i, \quad i = 1, 2, \dots, k \quad (9)$$

where  $\hat{\mathbf{Z}}$ , the coefficient values of pixels in the same superpixel, are given the same value, achieving the injection of local spatial detail information. However, in  $(\mathbf{Z} - \hat{\mathbf{Z}})$ , the anomaly should not be forced close to the estimated background coefficient value.  $\mathbf{W} = \text{diag}(\mathbf{w})$  weighs the background constraints and separates anomalies from the background by assigning a small weight to these pixels. By adding a guided matrix, the local geometry of the original HSI is maintained in the background. A graphical description of the LSC is shown in Fig. 1. However, the correlation between pixels is not well-considered, and the background structure using the correlation between pixels will be more uniform. TV regularization [46], [47], as prior information, has been proven to effectively promote piecewise constant (or smooth) transitions. The nonisotropic TV we adopt is defined as

$$\text{TV}(\mathbf{Z}) = \sum_{i,j \in \varepsilon} \|\mathbf{z}_i - \mathbf{z}_j\|_1 \quad (10)$$

where  $\mathbf{z}_i$  and  $\mathbf{z}_j$  represent the coefficient vector of the  $i$ th and  $j$ th pixels in the background coefficient matrix  $\mathbf{Z}$ , respectively, while  $\varepsilon$  is the first-order neighborhood of  $i$ th pixels. By minimizing (10),  $\mathbf{Z}$  will show a smoother result in the spatial domain. Consequently, the spectra of the pixels of the  $\mathbf{DZ}$  will appear more uniform. Furthermore, TV considers the correlation of pixels to have a certain suppression effect on some randomly distributed noise. By introducing TV, the final detection model is as follows:

$$\begin{aligned} \min_{\mathbf{Z}, \mathbf{E}} \frac{1}{2} \|(\mathbf{Z} - \hat{\mathbf{Z}})\mathbf{W}\|_F^2 + \lambda_{\text{TV}} \text{TV}(\mathbf{Z}) + \beta \|\mathbf{E}\|_{2,1} \\ \text{s.t. } \mathbf{X} = \mathbf{DZ} + \mathbf{E} \end{aligned} \quad (11)$$

where  $\lambda_{\text{TV}}$  and  $\beta$  are the tradeoff parameters that balance the contribution of each regular term. It is worth mentioning again that the spatial structure of the background is mainly maintained by the first data-fitting term. Hierarchical features are better considered by superpixel segmentation at different scales. The TV term is used as an additional regular term to take advantage of the correlation between pixels to obtain the background with detailed texture while combating some random noise. Furthermore, the anomaly has a weak correlation with the pixels in the neighborhood, and the TV model can effectively isolate the anomaly by using this feature. Under the premise of an accurate evaluation of the background,

the separation of anomaly is achieved by the mixed  $l_{2,1}$  norm, and the  $l_{2,1}$  norm simultaneously controls the size of the separated anomaly target. The parameter settings will be introduced in detail in Section III.

To make the model more explicit, a linear transformation operator,  $\mathbf{H}$ , is introduced.  $\mathbf{H} = [\mathbf{H}_h; \mathbf{H}_v]^T$ , where  $\mathbf{H}_h: \mathbb{R}^{m \times N} \rightarrow \mathbb{R}^{m \times N}$  represents an operator of the horizontal differences between neighboring pixels, i.e.,  $\mathbf{H}_h \mathbf{Z} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_N]$ , where  $\mathbf{d}_i = \mathbf{z}_i - \mathbf{z}_{i_h}$  and  $i_h$ th pixel is the horizontal neighbor of the  $i$ th pixel. Similarly,  $\mathbf{H}_v: \mathbb{R}^{m \times N} \rightarrow \mathbb{R}^{m \times N}$  represents an operator of the vertical differences between neighboring pixels, i.e.,  $\mathbf{H}_v \mathbf{Z} = [\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N]$ , where  $\mathbf{d}_i = \mathbf{z}_i - \mathbf{z}_{i_v}$  and  $i_v$ th pixel is the vertical neighbor of the  $i$ th pixel. With the above definition, the TV term can be expressed as  $\mathbf{HZ} = [\mathbf{H}_h \mathbf{Z}; \mathbf{H}_v \mathbf{Z}]^T$ . A meaningful detection model can be written as

$$\begin{aligned} & \min_{\mathbf{Z}, \mathbf{E}} \frac{1}{2} \|(\mathbf{Z} - \hat{\mathbf{Z}})\mathbf{W}\|_F^2 + \lambda_{\text{TV}} \|\mathbf{HZ}\|_{1,1} + \beta \|\mathbf{E}\|_{2,1} \\ & \text{s.t. } \mathbf{X} = \mathbf{DZ} + \mathbf{E}. \end{aligned} \quad (12)$$

### C. Optimization via ADMM

In this section, the ADMM-based optimization algorithm [48] is used to solve the objective function (12). The objective function is split by introducing auxiliary matrices  $\mathbf{J}$ ,  $\mathbf{P}$ , and  $\mathbf{R}$ . The separated objective function is given as follows:

$$\begin{aligned} & \min_{\mathbf{Z}, \mathbf{J}, \mathbf{P}, \mathbf{R}, \mathbf{E}} \frac{1}{2} \|(\mathbf{J} - \hat{\mathbf{Z}})\mathbf{W}\|_F^2 + \lambda_{\text{TV}} \|\mathbf{R}\|_{1,1} + \beta \|\mathbf{E}\|_{2,1} \\ & \text{s.t. } \mathbf{X} = \mathbf{DZ} + \mathbf{E} \\ & \mathbf{J} = \mathbf{Z} \\ & \mathbf{P} = \mathbf{Z} \\ & \mathbf{R} = \mathbf{HP}. \end{aligned} \quad (13)$$

By introducing the Lagrangian multipliers, a constrained objective function can be transformed into an unconstrained minimization problem

$$\begin{aligned} & L(\mathbf{Z}, \mathbf{J}, \mathbf{P}, \mathbf{R}, \mathbf{E}, \mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3, \mathbf{D}_4) \\ & = \frac{1}{2} \|(\mathbf{J} - \hat{\mathbf{Z}})\mathbf{W}\|_F^2 + \lambda_{\text{TV}} \|\mathbf{R}\|_{1,1} + \beta \|\mathbf{E}\|_{2,1} \\ & + \frac{\mu}{2} \|\mathbf{DZ} + \mathbf{E} - \mathbf{X} - \mathbf{D}_1\|_F^2 + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{J} - \mathbf{D}_2\|_F^2 \\ & + \frac{\mu}{2} \|\mathbf{Z} - \mathbf{P} - \mathbf{D}_3\|_F^2 + \frac{\mu}{2} \|\mathbf{HP} - \mathbf{R} - \mathbf{D}_4\|_F^2. \end{aligned} \quad (14)$$

In the augmented Lagrangian function (14),  $\mu > 0$  remains unchanged during the execution of the algorithm, and  $\mathbf{D}_i/\mu$  ( $i = 1, 2, 3$ , and  $4$ ) are the Lagrangian multipliers, which guarantees that the constraints of each variable in (14) can continue to hold. The ADMM algorithm is used to iteratively update each variable. When the algorithm reaches a preset condition, it stops and outputs the anomaly component  $\mathbf{E}$ . The process of ADMM-based LSC-TV is shown in Algorithm 1. The update of the variables in Algorithm 1 involves  $\mathbf{Z}$  and  $\mathbf{E}$  and intermediate auxiliary variables  $\mathbf{J}$ ,  $\mathbf{P}$ , and  $\mathbf{R}$ .

In order to update  $\mathbf{Z}$ , the following method needs to be minimized:

$$\begin{aligned} \mathbf{Z}^{k+1} & \leftarrow \arg \min_{\mathbf{Z}} \frac{\mu}{2} \left\| \mathbf{DZ} + \mathbf{E}^{(k)} - \mathbf{X} - \mathbf{D}_1^{(k)} \right\|_F^2 \\ & + \frac{\mu}{2} \left\| \mathbf{Z} - \mathbf{J}^{(k)} - \mathbf{D}_2^{(k)} \right\|_F^2 + \frac{\mu}{2} \left\| \mathbf{Z} - \mathbf{P}^{(k)} - \mathbf{D}_3^{(k)} \right\|_F^2. \end{aligned} \quad (15)$$

The solution of (17) is given as follows:

$$\begin{aligned} \mathbf{Z}^{(k+1)} & \leftarrow (\mathbf{D}^T \mathbf{D} + 2\mathbf{I})^{-1} \left( \mathbf{D}^T \left( \mathbf{X} - \mathbf{E}^{(k)} - \mathbf{D}_1^{(k)} \right) \right. \\ & \left. + \left( \mathbf{J} + \mathbf{D}_2^{(k)} \right) + \left( \mathbf{P} + \mathbf{D}_3^{(k)} \right) \right) \end{aligned} \quad (16)$$

where  $\mathbf{I}$  is the identity matrix,  $(\cdot)^{(-1)}$  represents the inverse of the matrix, and  $\mathbf{D}^T$  is the transpose of  $\mathbf{D}$ .

To update  $\mathbf{J}$ , the following equations will be optimized:

$$\mathbf{J}^{(k+1)} \leftarrow \frac{1}{2} \left\| (\mathbf{J} - \hat{\mathbf{Z}}) \mathbf{W} \right\|_F^2 + \frac{\mu}{2} \left\| \mathbf{Z}^{(k)} - \mathbf{J} - \mathbf{D}_2^{(k)} \right\|_F^2. \quad (17)$$

The minimization of  $\mathbf{J}$  is given as follows:

$$\mathbf{J}^{(k+1)} \leftarrow \left( \hat{\mathbf{Z}} \mathbf{W} \mathbf{W}^T + \mu \left( \mathbf{Z}^{(k)} - \mathbf{D}_2^{(k)} \right) \right) (\mathbf{W} \mathbf{W}^T + \mu \mathbf{I})^{-1}. \quad (18)$$

To compute  $\mathbf{P}$ , the following minimization of problem needs to be solved:

$$\begin{aligned} \mathbf{P}^{(k+1)} & \leftarrow \frac{\mu}{2} \left\| \mathbf{Z}^{(k)} - \mathbf{P} - \mathbf{D}_3^{(k)} \right\|_F^2 \\ & + \frac{\mu}{2} \left\| \mathbf{HP} - \mathbf{R}^{(k)} - \mathbf{D}_4^{(k)} \right\|_F^2 \end{aligned} \quad (19)$$

whose solution is

$$\mathbf{P}^{(k+1)} \leftarrow (\mathbf{H}^T \mathbf{H} + \mathbf{I})^{-1} \left( \mathbf{Z}^{(k)} - \mathbf{D}_3^{(k)} + \mathbf{H}^T \left( \mathbf{R}^{(k)} + \mathbf{D}_4^{(k)} \right) \right). \quad (20)$$

Since  $\mathbf{H}$  is a convolution, it can be computed efficiently using the discrete Fourier transform diagonalization.

To compute  $\mathbf{R}$ , the following minimization of problem needs to be solved:

$$\mathbf{R}^{(k+1)} \leftarrow \lambda_{\text{TV}} \|\mathbf{R}\|_{1,1} + \frac{\mu}{2} \left\| \mathbf{HP}^{(k)} - \mathbf{R} - \mathbf{D}_4^{(k)} \right\|_F^2 \quad (21)$$

whose solution is the well-known soft threshold [49]

$$\mathbf{R}^{(k+1)} \leftarrow \text{soft} \left( \mathbf{HP}^{(k)} - \mathbf{D}_4^{(k)}, \frac{\lambda_{\text{TV}}}{\mu} \right) \quad (22)$$

where  $\text{soft}(\cdot, \tau)$  denotes the componentwise application of the soft-threshold function  $\text{soft}(u, a) \equiv \text{sign}(u) \max\{|u| - a, 0\}$ .

Finally, the equation for  $\mathbf{E}$  is given as follows:

$$\mathbf{E}^{(k+1)} \leftarrow \beta \|\mathbf{E}\|_{2,1} + \frac{\mu}{2} \left\| \mathbf{DZ}^{(k)} + \mathbf{E} - \mathbf{X} - \mathbf{D}_1^{(k)} \right\|_F^2 \quad (23)$$

whose solution is the well-known soft-vector threshold [50]

$$\mathbf{E}^{(k+1)} \leftarrow \text{soft-vector} \left( \mathbf{X} - \mathbf{DZ}^{(k)} - \mathbf{D}_1^{(k)}, \frac{\beta}{\mu} \right) \quad (24)$$

where  $\text{soft-vector}(\cdot, \tau)$  denotes the soft-vector-threshold function

$$\text{soft-vector}(\mathbf{U}(:, i), \tau) = \frac{\max(\|\mathbf{U}(:, i)\|_2 - \tau, 0)}{\max(\|\mathbf{U}(:, i)\|_2 - \tau, 0) + \tau} \mathbf{U}(:, i), \quad i = 1, 2, \dots, N. \quad (25)$$

In Algorithm 1, the guided matrix  $\hat{\mathbf{Z}}$  and the weighted matrix  $\mathbf{w}$  are updated every ten generations, which can help

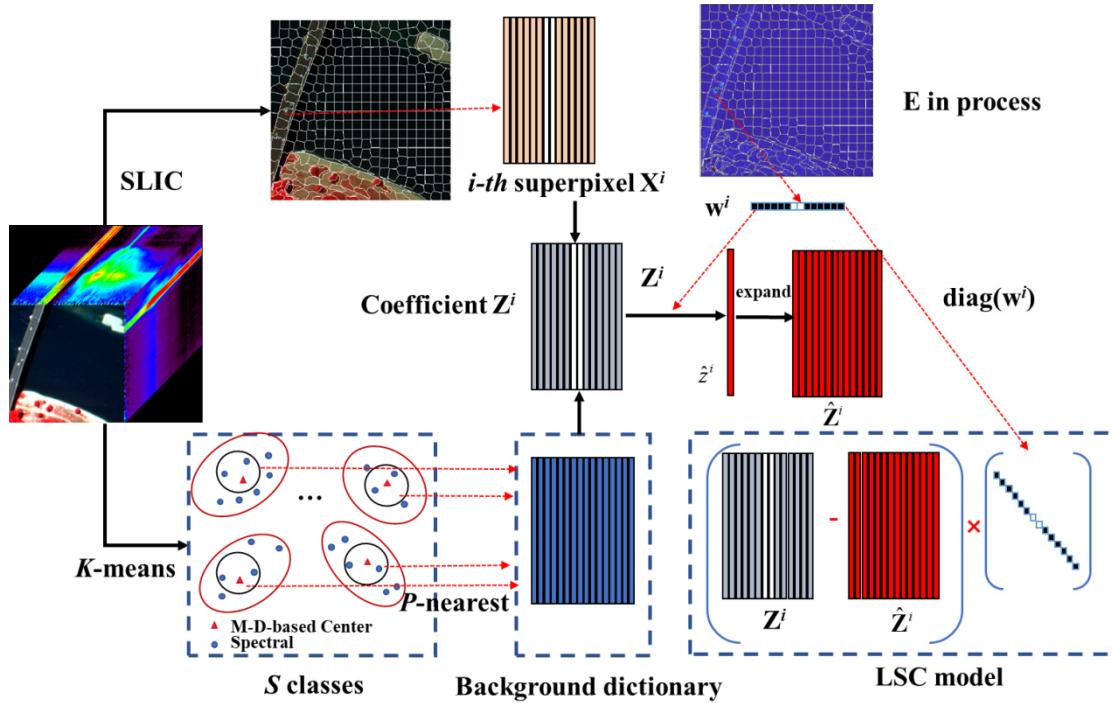


Fig. 1. Illustration of the LSC model.

obtain stable  $\mathbf{w}$  and  $\hat{\mathbf{Z}}$  while reducing the burden of updating. Regarding the stopping conditions of the algorithm, once the number of iterations reaches the maximum value of the number of iterations or the residual ( $\sigma$  is a preset threshold), the algorithm stops.

From Algorithm 1, its computational complexity depends on the update of the variables  $\mathbf{Z}$  and  $\mathbf{E}$ . Because  $(\mathbf{D}^T \mathbf{D} + 2\mathbf{I})^{-1}$  is fixed, it can be calculated in advance during the algorithm iteration process, and its time complexity is  $\mathcal{O}(m^2N)$ . About auxiliary variable  $\mathbf{J}$ , since  $\mathbf{W}$  is a diagonal matrix, in the solution process, it is only necessary to pay attention to the calculation of the entries on the diagonal, and its time complexity is  $\mathcal{O}(mN)$ . As for  $\mathbf{P}$ , its time complexity is  $\mathcal{O}(mN \log N)$ . Regarding  $\mathbf{R}$ , its time complexity is  $\mathcal{O}(N)$ ,  $\mathbf{E}$  has the time complexity of  $\mathcal{O}(NL)$ . Here,  $m$  is the number of atoms in dictionary  $\mathbf{D}$ , and  $N$  and  $L$  are the numbers of pixels and spectral bands contained in the HSI matrix  $\mathbf{X}$ , respectively. Thus, the time complexity of this algorithm is  $\mathcal{O}(m^2N)$ .

### III. EXPERIMENT RESULTS AND DISCUSSION

In this section, the proposed LSC-TV algorithm will be evaluated regarding the performance of anomaly detection on several sets of HSIs with different characteristics, including one set of simulated data and four sets of real data.

At the same time, the proposed LSC-TV algorithm will also be compared with three typical methods and a state-of-the-art method. The first is the traditional detector RX [10] and LRX [11], [12] based on statistical methods. The remaining two are representation-based detection methods, including the CRD [24] method based on collaborative representation and the LSMAD [27] algorithm based on the RPCA and Mahalanobis distance. The last is the LRASR [29] method,

---

**Algorithm 1** ADMM for LSC-TV

---

**Input:** the hyperspectral matrix  $\mathbf{X}$ , the background dictionary  $\mathbf{D}$ . Superpixel segmentation results  $\mathbf{X} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(k)}\}$ .

**Initialization:**

set  $k = 0$ ,  $\text{Max\_Iter} = 100$ , choose  $\mu = 0.01$ , and  $\mathbf{Z}^{(0)} = \mathbf{0}$ ,  $\mathbf{J}^{(0)} = \mathbf{Z}^{(0)}$ ,  $\mathbf{P}^{(0)} = \mathbf{Z}^{(0)}$ ,  $\mathbf{R}^{(0)} = \mathbf{H}\mathbf{P}^{(0)}$ ,  $\mathbf{E}^{(0)} = \mathbf{0}$ ,  $\mathbf{D}_1^{(0)} = \mathbf{E}^{(0)}$ ,  $\mathbf{D}_2^{(0)} = \mathbf{J}^{(0)}$ ,  $\mathbf{D}_3^{(0)} = \mathbf{P}^{(0)}$ ,  $\mathbf{D}_4^{(0)} = \mathbf{R}^{(0)}$ .

**Repeat:**

**Update:**

the guided matrix  $\hat{\mathbf{Z}}$ , the weighted matrix  $\mathbf{w}$ ,  $\mathbf{W} = \text{diag}(\mathbf{w})$  as (8).

**Repeat:**

**Update  $\mathbf{Z}$ :**

$$\mathbf{Z}^{(k+1)} \leftarrow \arg \min_{\mathbf{Z}} L(\mathbf{Z}, \mathbf{J}^{(k)}, \mathbf{P}^{(k)}, \mathbf{R}^{(k)}, \mathbf{E}^{(k)}, \mathbf{D}_i^{(k)})$$

**Update auxiliary matrices  $\mathbf{J}$ ,  $\mathbf{P}$ ,  $\mathbf{R}$ :**

$$\mathbf{J}^{(k+1)} \leftarrow \arg \min_{\mathbf{J}} L(\mathbf{Z}^{(k)}, \mathbf{J}, \mathbf{P}^{(k)}, \mathbf{R}^{(k)}, \mathbf{E}^{(k)}, \mathbf{D}_i^{(k)})$$

$$\mathbf{P}^{(k+1)} \leftarrow \arg \min_{\mathbf{P}} L(\mathbf{Z}^{(k)}, \mathbf{J}^{(k)}, \mathbf{P}, \mathbf{R}^{(k)}, \mathbf{E}^{(k)}, \mathbf{D}_i^{(k)})$$

$$\mathbf{R}^{(k+1)} \leftarrow \arg \min_{\mathbf{R}} L(\mathbf{Z}^{(k)}, \mathbf{J}^{(k)}, \mathbf{P}^{(k)}, \mathbf{R}, \mathbf{E}^{(k)}, \mathbf{D}_i^{(k)})$$

**Update  $\mathbf{E}$ :**

$$\mathbf{E}^{(k+1)} \leftarrow \arg \min_{\mathbf{E}} L(\mathbf{Z}^{(k)}, \mathbf{J}^{(k)}, \mathbf{P}^{(k)}, \mathbf{R}^{(k)}, \mathbf{E}, \mathbf{D}_i^{(k)})$$

**Update Lagrange multipliers:**

$$\mathbf{D}_1^{(k+1)} = \mathbf{D}_1^{(k+1)} - (\mathbf{X} - \mathbf{D}\mathbf{Z}^{(k)} - \mathbf{E}^{(k)})$$

$$\mathbf{D}_2^{(k+1)} = \mathbf{D}_2^{(k)} - (\mathbf{Z}^{(k)} - \mathbf{J}^{(k)})$$

$$\mathbf{D}_3^{(k+1)} = \mathbf{D}_3^{(k)} - (\mathbf{Z}^{(k)} - \mathbf{P}^{(k)})$$

$$\mathbf{D}_4^{(k+1)} = \mathbf{D}_4^{(k)} - (\mathbf{H}\mathbf{P}^{(k)} - \mathbf{R}^{(k)})$$

**Update iteration:**  $k \leftarrow k + 1$

**Until** the stopping condition is satisfied, and output the  $\mathbf{E}$

---

which combines LRR and sparse representation. The last one is considered a state-of-the-art algorithm, KIFD method [21].

In addition, on the simulated data, the contribution of each component of LSC-TV method will also be verified.

First, to verify the idea that considering the local structure of the HSI can effectively separate anomaly, the proposed methods without the TV term (i.e., LSC) will be compared with the standard LRR method, which considers only the global structure of the background. Subsequently, the proposed LSC-TV method will be compared with LSC to illustrate the contribution of the TV term to the effective separation of anomaly.

In order to judge the actual performance of each detector, the anomaly components extracted by each method will be displayed with a color detection map. In addition, the numerical evaluation results will be determined by the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC) [51]. The ROC curve is drawn by calculating the probability of detection ( $P_D$ ) versus probability of false alarm ( $P_{FA}$ ). Once the threshold  $\eta$  is determined, the detection result of the detector will be converted into a binary map, these detection values higher than  $\eta$  will be judged as anomaly, and those lower than  $\eta$  will be judged as background. Therefore,  $P_D$  and  $P_{FA}$  are defined as follows:

$$P_D = \frac{N_{APC}}{N_A}, \quad P_{FA} = \frac{N_{BA}}{N_B} \quad (26)$$

where  $N_{APC}$  is the number of correctly labeled anomaly,  $N_A$  is the number of all anomaly,  $N_{BA}$  is the number of background pixels labeled as anomaly, and  $N_B$  is the number of all background pixels. ROC can be obtained by plotting ( $P_{FA}$ ,  $P_D$ ), corresponding to different  $\eta$  values, into a 2-D coordinate curve. By calculating the area of the ROC and the  $P_{FA}$  axes, the AUC value can also be obtained. It should be noted that the higher  $P_D$  corresponding to the lower  $P_{FA}$  is expected by the detector—that is, the larger the AUC value—the better the performance of the detector.

#### A. Experiments on Simulated Data

The simulated data used in the experiment were generated based on a real dataset, which was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over Salinas Valley, CA, USA.<sup>1</sup> The spatial resolution of the Salinas spectrometer is 3.7 m/pixel. The size of this scene is 217 rows and 512 columns, with 224 bands. After removing some low signal-to-noise ratio and water absorption bands (bands 108–112, 154–167, and 224), the remaining 204 bands are used on the simulated data. The 120 × 120 pixel subscene located in the bottom-left corner is used here. Fig. 2 shows the HSI image of Salinas, the subscene used, and the location of the anomaly.

The target implantation method [52] was used to generate anomaly, and the generation of these simulation data followed a linear mixed model as follows:

$$\mathbf{z} = f \cdot \mathbf{t} + (1 - f) \cdot \mathbf{b} \quad (27)$$

where  $\mathbf{t}$  is the specified anomaly signature,  $\mathbf{b}$  is the background spectral, and  $f$  is the abundance of  $\mathbf{t}$  (ranging from 0.04 to 1). These simulated data have been proven to be effective in evaluating various methods of anomaly detection in [53].

<sup>1</sup>[http://www.ahu.edu/ccwintco/index.php?title=Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ahu.edu/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes)

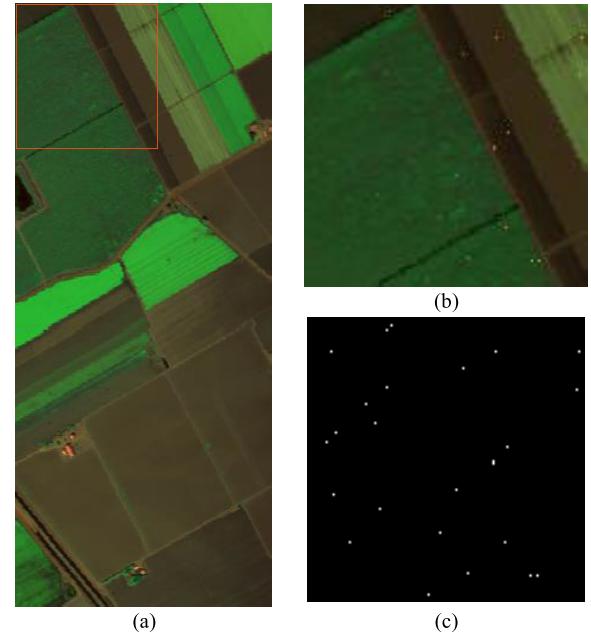


Fig. 2. Salinas dataset. (a) False-color image of the whole scene. (b) False-color image of the subscene. (c) Location map of anomaly.

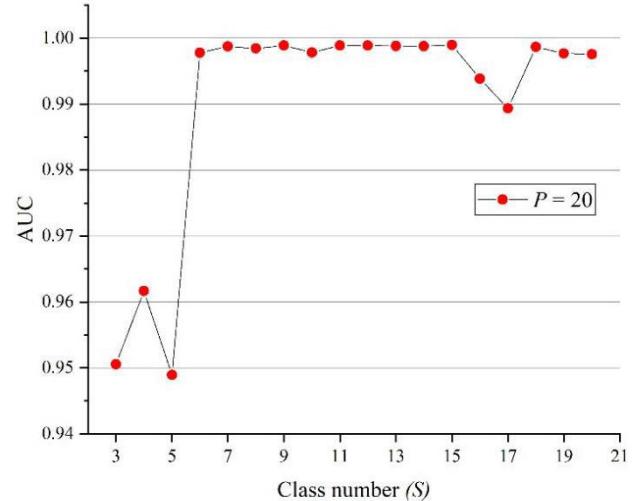


Fig. 3. Influence of the selection of the category  $S$  ( $P = 20$ ) value in the  $k$ -means method on the final detection result.

1) *Dictionary Construction*: The background dictionary was constructed using the method proposed in [29], and it is also recommended in [32]. First, the  $k$ -means clustering method was used to classify HSIs into  $S$  categories. To avoid the selection of anomaly, the Mahalanobis distance between each pixel and the current category was calculated, and the  $P$  pixels closest to the center of each category were selected as the background. In all experiments, the selection of  $P$  was 20, following the recommended values in [29] and [32]. Fig. 3 shows the AUC value of the proposed LSC-TV method on the Salinas data under different  $S$  values.

From Fig. 3, it can be seen that when the selected category  $S$  is very small, the value of AUC is relatively low, which is because the atoms of the background dictionary are not overcomplete. Many classes of atoms are not considered in the background dictionary. As the categories increase, the

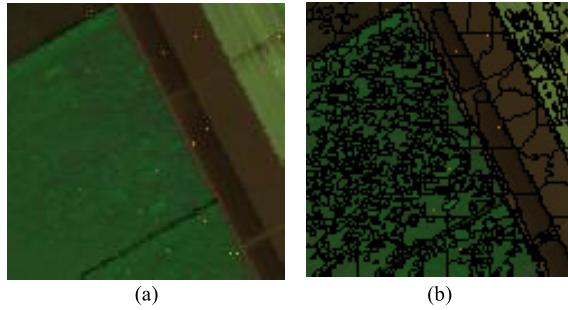


Fig. 4. Superpixel segmentation result on Salinas data. (a) False-color image of the subscene. (b) Segmentation result ( $size = 10 \times 10$  pixels).

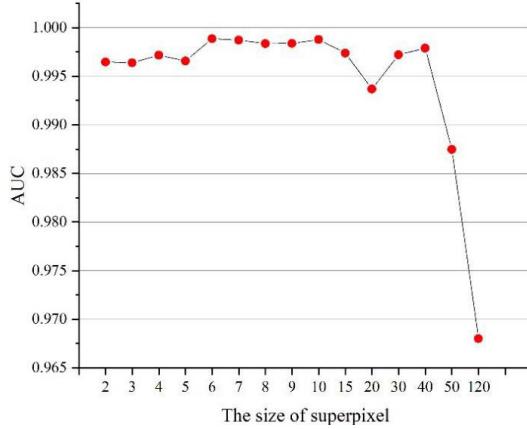


Fig. 5. Impact of different segmentation scales on final detection results.

proposed method shows a relatively stable AUC value. Based on all the experiments, the  $S$  value chosen is 15.

2) *Superpixel Segmentation*: The proposed LSC-TV method extends the SLIC method to HSIs. Fig. 4 shows the segmentation results of the simulated image under the preset superpixel size of  $10 \times 10$  pixels. In the segmentation result, it can be observed that the boundaries of the region can be effectively highlighted, and the uniformity of the distribution in the superpixel region is well maintained. This proves the effectiveness of superpixels in extracting the semantic information (i.e., joint spectral and spatial information) of HSIs. Moreover, it can be seen that the background distribution in anomaly detection is also consistent with the result of superpixel segmentation—that is, it is more inclined to present a uniform distribution locally rather than maintaining a global uniform structure like the LRR model. In order to analyze the impact of different segmentation scales on the final detection results, Fig. 5 shows the AUC values of the proposed method under simulated data at different segmentation scales.

It can be seen in Fig. 5 that selecting a small-sized superpixel tends to obtain a larger AUC value. This is because small-sized superpixels often contain more similar pixel spectra and more adjacent spatial locations. However, when the segmentation scale is set to  $120 \times 120$  pixels (i.e., the entire scene is regarded as a homogeneous region), the performance of the proposed detector will decrease sharply. This again shows that HSIs tend to show local similarity features.

3) *Contribution of Each Regular Term*: In order to verify the validity of each regular term of the proposed method,

TABLE I  
AUC VALUE OF DIFFERENT METHODS UNDER SALINAS DATA

algorithm	LRR	LSC	LSC-TV
AUC	0.9206	0.9824	<b>0.9986</b>

the following method was used for comparative experiments using the Salinas data:

- 1) LRR method (with the same background dictionary as the proposed method) [50];
- 2) LSC method (without TV regularization);
- 3) the proposed LSC-TV method.

To verify that the proposed LSC maintains the background detailed structure better than the standard global low-rank constraint used by the LRR method, a version without TV regularization is compared with the standard LRR algorithm as (4) (except using the same background dictionary with the proposed method). Furthermore, by comparing the experimental results of the LSC method and LSC-TV, the necessity of introducing TV regularization can be determined. The visual detection results of these three methods are shown in Fig. 6. The values of AUC are listed in Table I. We carefully searched the  $\beta$  parameters of LRR and LSC in  $(0, 1)$  to ensure that both methods can demonstrate the best performance. Finally, in the proposed LSC-TV method,  $\lambda_{TV} = 1e-3$  and  $\beta = 5e-3$ ; in the LSC,  $\beta = 1e-4$ ; and in the LRR,  $\beta = 5e-3$ , such that these methods will achieve the maximum AUC value under these parameters.

It can be seen in Fig. 6(a) and (b) that the proposed LSC yields better results for the separation of background and anomalies and that the estimated values of the background are limited to a lower value range. In the LRR method, because only the global structure of the background is considered, some boundaries are estimated to be extraordinary values and are as such determined to be anomalies. Moreover, it is difficult to distinguish the anomaly from the detection map. Although LSC can effectively constrain the background, some difficult-to-detect targets have not been well distinguished (the abundance value of the anomaly target is very low in the left part). By taking advantage of the high correlation of background pixels, these targets can be distinguished well in the LSC-TV. In the ROC, the proposed LSC-TV can detect all anomalies with a very low false alarm rate. Compared with the LSC method and the LRR, the LSC first distinguishes all targets, but both methods suffer high false alarm rates. All these illustrate the contribution and effectiveness of each regular term in the proposed method for the separation of anomaly.

Consistent with the previous analysis, it is noticed in Table I that the proposed LSC-TV method naturally obtains the maximum AUC value as it has the highest detection rate. LSC has a higher value than the LRR method due to effective background constraints.

4) *Detection Performance*: Some previous basic comparative experiments have verified the effectiveness of the proposed method. However, some state-of-the-art methods are also used to further compare with the proposed method. The detection results of the proposed method and the previously mentioned comparison method are shown in Fig. 7.

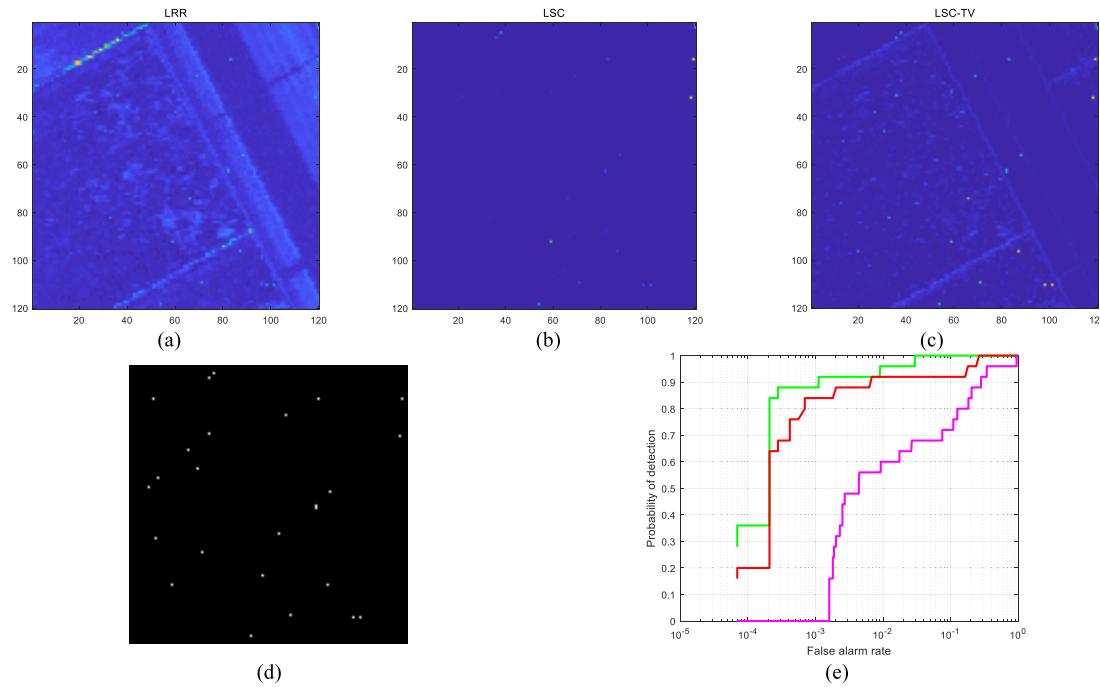


Fig. 6. Detection results on the Salinas data by different methods. (a)–(c) Result maps of LRR, LSC, and LSC-TV, respectively. (d) Ground-truth map. (e) ROC curve.

It can be seen from Fig. 7 that RX, as a benchmark method, performs the worst on the Salinas data, and only a few anomalies are detected. Because LSMAD applies the RX method to the background extracted by RPCA, these detection results can be considered a significant improvement, compared to the RX detector. At the same time, because the background extracted by RPCA contains less noise, the LSMAD method can suppress the background at a lower value range. However, because this method still uses the assumption of RX, some anomalies that have not been separated in RX remain unresolved. The inner and outer windows of LRX and CRD are set to (7, 3) and (5, 3) after careful searching, respectively. Although the background has a smaller estimated value than the previous two methods, the detection results of some anomaly are completely ignored in CRD and LRX. LRASR introduces a sparse representation method based on the standard LRR method, which is an improvement on the LRR method. However, the global similarity structure of the background is still used here, which leads to a large detection value for some backgrounds. Those anomalies that are not easy to detect are drowned in the background. The KIFD algorithm has a good preference on the low abundance of anomaly, and then, its robustness to noise is poor; there are more false alarm targets. As for the proposed LSC-TV method, due to the local similarity of the background and the correlation between the pixels effectively considered, all the anomalies are well separated from the background. Meanwhile, anomaly with higher detection values is easily distinguished. It can also be seen from the ROC diagram that compared with all other methods. Although CRD and LRX have high detection rates at very low false alarm rates, indicating that they are more sensitive to a small number of anomalies, they can detect all anomalies only at very high false alarm rates. For LSC-TV

and KIFD methods, they identify all anomalies at lower false alarm rates. Also, it can be seen that LSC-TV has a greater detection rate in most cases. Compared with other methods, the RX method has a very low detection rate in most cases. This further validates that all methods are effective compared to the benchmark method.

### B. Experiments on Real Data

1) *Data Description:* The first dataset was collected through the hyperspectral digital imagery collection experiment (HYDICE) sensor with an aircraft platform.<sup>2</sup> This scene contains  $307 \times 307$  pixels and has a spatial resolution of 2 m. Its spectral resolution is 10 nm, collected in the wavelength range of 400–2500 nm. After removing some low signal-to-noise ratio and water vapor absorption bands, the remaining 168 bands are available. An  $80 \times 100$  pixel subscene false-color map of this image is selected for experiments. The image and the reference anomaly map are shown in Fig. 8(a).

The second dataset was obtained by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight over the city of Pavia. This information is provided by Zhang *et al.* [27]. The HSI of Pavia Centre has  $1096 \times 1096$  pixels, with a spatial resolution of 1.3 m, while retaining spectral information for 102 bands. A  $150 \times 150$  pixel subscene is used for the task of anomaly detection, in which the anomaly is mainly composed of cars and piers on a bridge. The HSI of this scene, with reference to the location information of the anomaly, is shown in Fig. 8(b).

The third set of data was acquired by the Nuance Cri hyperspectral sensor, and its spectral resolution is 10 nm. This scene contains  $400 \times 400$  pixels, 46 bands in wavelengths

<sup>2</sup>[http://www.escience.cn/people/feiyunZHU/Dataset\\_GT.html#](http://www.escience.cn/people/feiyunZHU/Dataset_GT.html#)

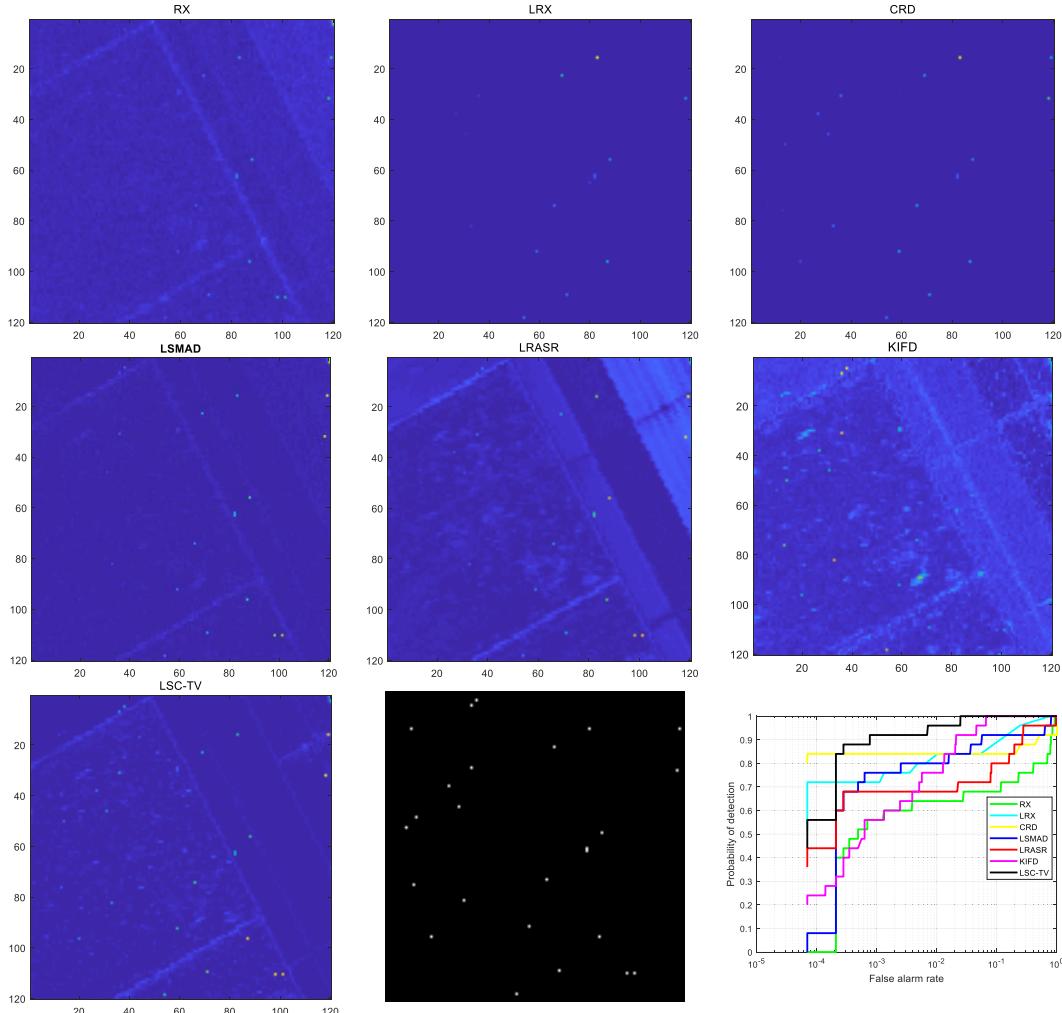


Fig. 7. Detection results on the Salinas data by different methods and the ROC curve.

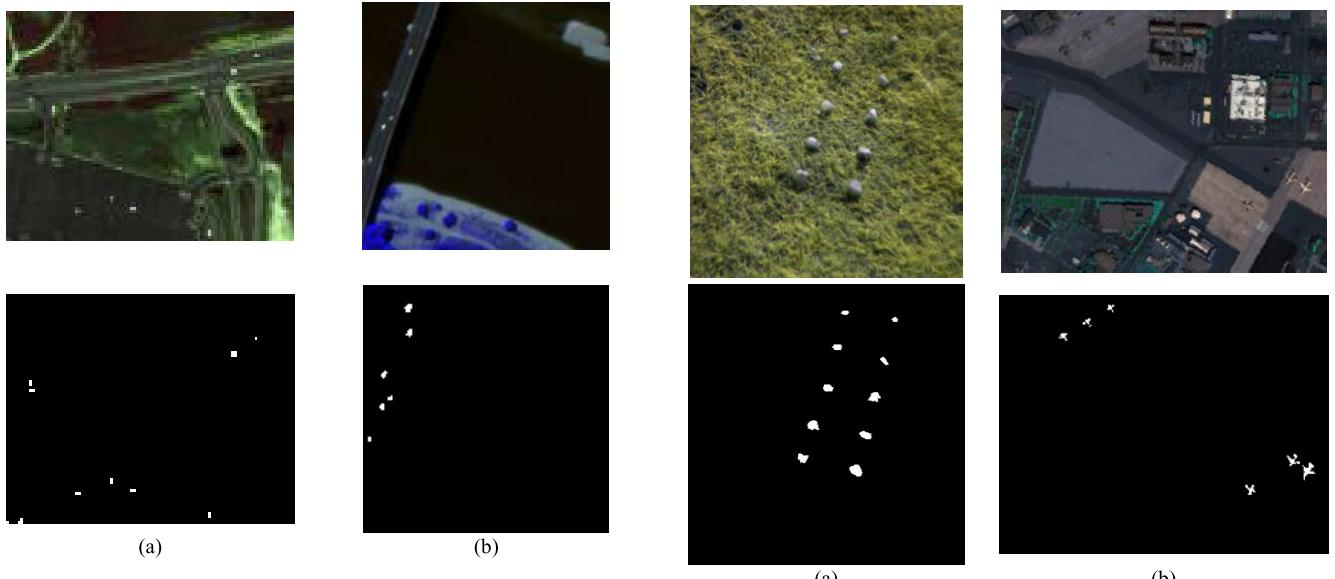


Fig. 8. False-color image and ground-truth maps. (a) HYDICE. (b) Pavia.

ranging from 650 to 1100 nm. There are ten rocks located in the grassy scene, which consist of 2216 pixels and account for 1.385% of the image. The HSI of this scene, with reference to the location information of the anomaly, is shown in Fig. 9(a).

Fig. 9. False-color image and ground-truth maps. (a) Cri. (b) Sandiego.

The fourth dataset was selected from the San Diego airport data. These data were also collected by the AVIRIS sensor. The size of the original image contains  $400 \times 400$  pixels, and

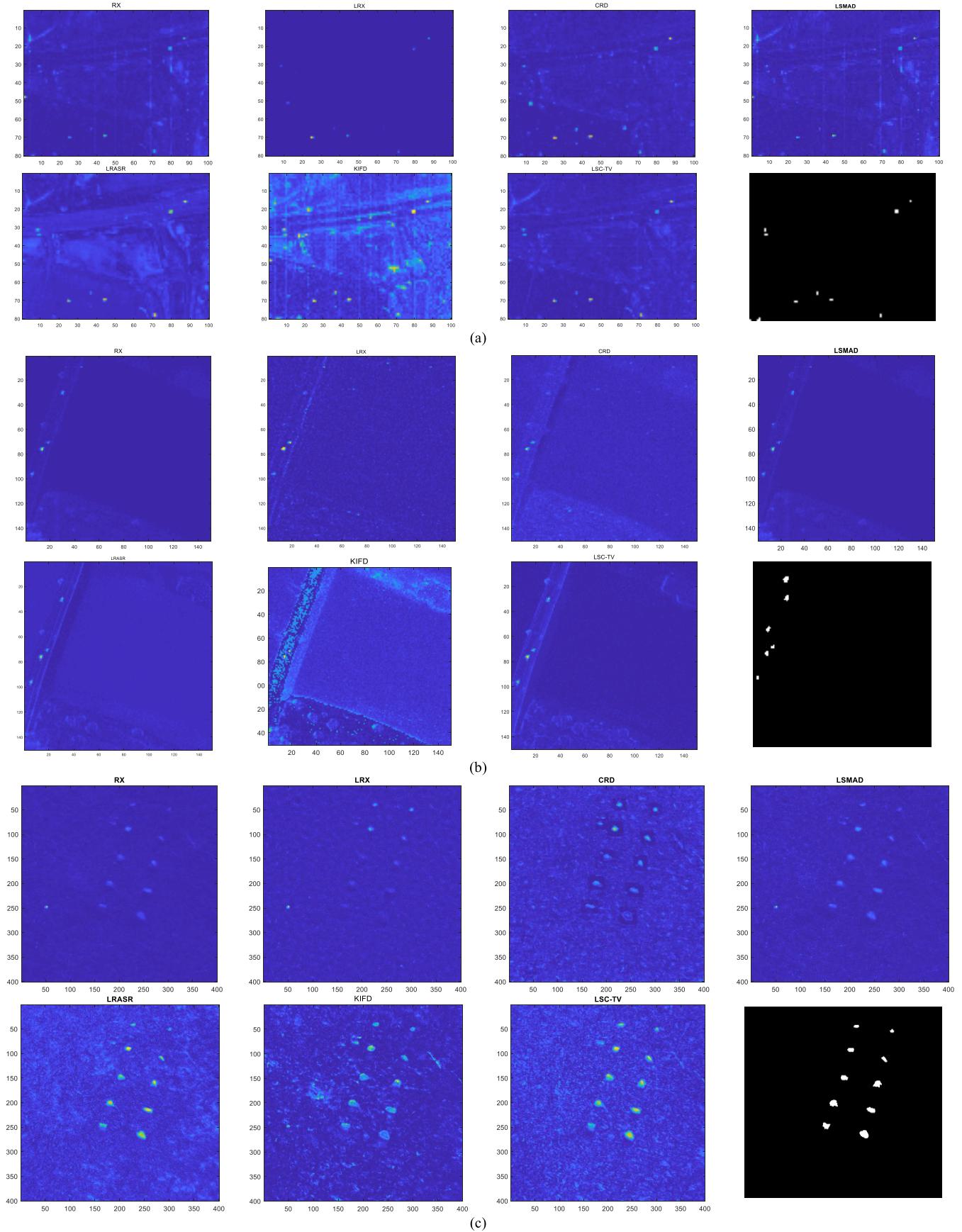


Fig. 10. Color detection maps of compared algorithms for real dataset. (a) HYDICE data. (b) Pavia data. (c) Cri data. (d) Sandiego. In each set of data (from Left to Right and from Top to Bottom), RX, LRX, CRD, LSMAD, LRASR, KIFD, the proposed LSC-TV, and the reference map.

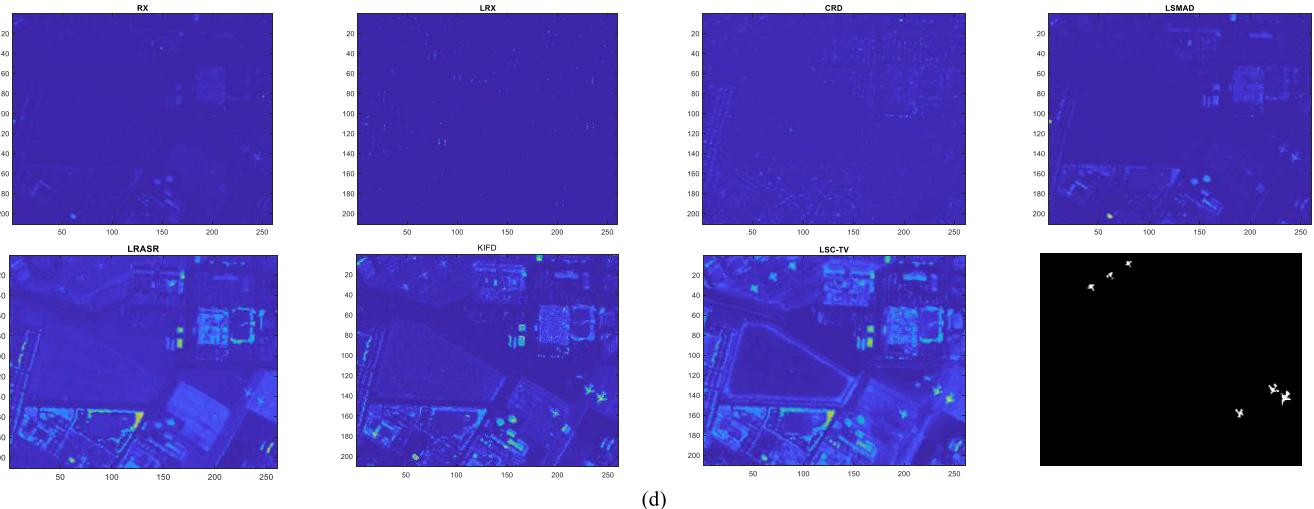


Fig. 10. (Continued.) Color detection maps of compared algorithms for real dataset. (a) HYDICE data. (b) Pavia data. (c) Cri data. (d) Sandiego. In each set of data (from Left to Right and from Top to Bottom), RX, LRX, CRD, LSMAD, LRASR, KIFD, the proposed LSC-TV, and the reference map.

the spatial resolution is 3.5 m. The original number of bands is 224 bands, distributed in wavelengths ranging from 370 to 2510 nm. After removing some useless bands, 189 bands are left for use. The anomaly of these data is several aircraft on the airport. One subscene ( $210 \times 260$  pixels) in the image contains aircraft and is used as a comparison experiment. The specific details of these data are shown in Fig. 9(b).

2) *Detection Performance*: In this section, the comparison of the results on real data is discussed in detail. Regarding the parameter-setting problem, after an extensive search of the internal and external windows of the LXR and CRD algorithms, the parameters of window size ( $w_{out}, w_{in}$ ) are set to (5, 3) for LXR and CRD on the HYDICE and Pavia, set to (21, 13) for LXR and CRD on the Cri, and set to (7, 5) for LXR and CRD on the Sandiego. It can be seen that the setting of the inner window mainly depends on the size of the target. Based on the analysis of Salinas, the construction of the background dictionary in the proposed LSC-TV method and LRASR is  $S = 15$  and  $P = 20$ . The size of superpixel segmentation is set to  $8 \times 8$  pixels.

The color maps of different detectors are shown in Fig. 10. When compared with other methods, anomaly is more easily distinguished in the detection map of LSC-TV. In the LSC-TV detection map, the homogeneous regions of the background are suppressed to a lower range. Furthermore, for the pixel-level anomaly (i.e., HYDICE and Pavia), the background area in the color map of LSC-TV contains less noise, and the anomaly can be correctly distinguished. With other methods (i.e., RX, LXR, CRD, and LSMAD), however, some anomalies are completely ignored and buried in the background. For LRASR, the problem of high false alarm rate also exists at the boundary. For KIFD, a large number of background targets were misclassified due to insufficient consideration of spatial correlation. For anomaly existing in multiple pixels (i.e., Cri and Sandiego), the LSC-TV method can still accurately distinguish them from the background. However, other methods have a poor ability to detect such targets. Especially for Sandiego data, RX, LXR, CRD, and LSMAD

can hardly detect the location of the anomaly, while LRASR can distinguish the anomaly, but not very clearly. For KIFD and LSC-TV, the anomaly is completely visible. It can be seen from all other subgraphs that the detection capability of the LSC-TV method is better, comparatively, than the RX, LXR, CRD, LSMAD, LRASR, and KIFD methods.

In addition, to compare the quantitative results of the proposed method and the comparative methods, the ROC curves of each method are plotted in Fig. 11. According to the previous analysis, the AUC values of each method on different datasets are shown in Table II. It can be seen from Fig. 11 that the LSC-TV always has the highest detection rate in the Pavia and Sandiego data. For HYDICE and Cri data, LRASR has a higher detection rate at a very lower false alarm rate, but LSC-TV has a higher detection rate at a false alarm rate of 0.1% (which is acceptable). Furthermore, LSC-TV can identify all anomalies in all data at a lower false alarm rate. This is consistent with the results of the color detection maps in Fig. 10. Naturally, in Table II, the proposed method has the largest AUC value on all real data compared with other methods. These results show that, under visual results and numerical evaluation, the proposed LSC-TV method has the best detection result on anomaly compared with state-of-the-art methods.

In order to verify the computational efficiency of the proposed method, Table III also shows the running times of different methods on simulated data and real data. The specification of the test computer is Intel Core i3-8100 CPU at 3.60 GHz with RAM 8 GB. All methods were tested on the MATLAB R2017a platform in the same environment.

From Table III, the RX algorithm has a significant advantage over other algorithms because it only needs to calculate the background variance and the mean value compared with each pixel. The running time of the CRD method is greatly affected by the difference between the inner and outer windows. When the difference between the inner and outer windows is too large (i.e., Cri), the background dictionary contains more atoms, and the running time will become larger. LSMAD uses two

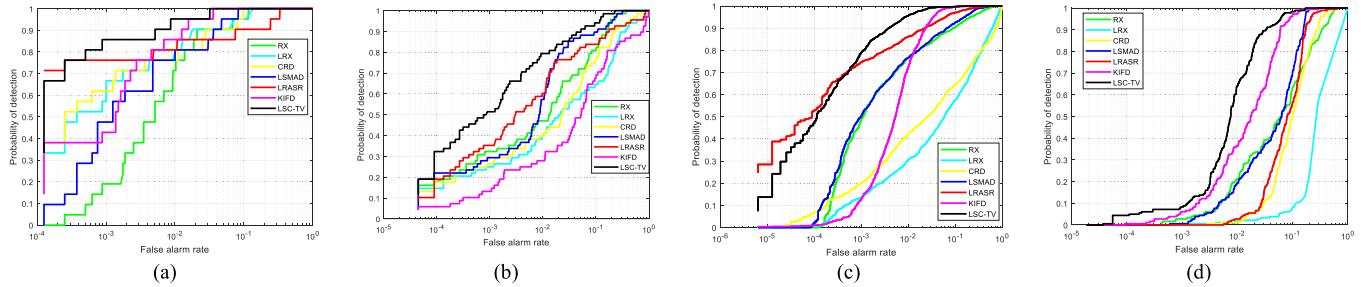


Fig. 11. Detection performance using different algorithms on Real data. (a) HYDICE data. (b) Pavia data. (c) Cri data. (d) Sandiego data.

TABLE II  
AUC FOR REAL DATA OBTAINED FROM DIFFERENT ALGORITHMS

Data	RX	LRX	CRD	LSMAD	LRASR	KIFD	LSC-TV
Salinas	0.8074	0.9613	0.8920	0.9383	0.9182	0.9919	<b>0.9986</b>
HYDICE	0.9855	0.9885	0.9879	0.9886	0.9694	0.9958	<b>0.9978</b>
Pavia	0.9538	0.8712	0.9159	0.9692	0.9314	0.8388	<b>0.9843</b>
Cri	0.9675	0.8152	0.8155	0.9758	0.9906	0.9877	<b>0.9984</b>
Sandiego	0.8807	0.6120	0.8713	0.9294	0.8953	0.9718	<b>0.9857</b>

TABLE III  
COMPUTATIONS TIMES OF LSC-TV AND OTHER METHODS IN SECONDS

Data	RX	LRX	CRD	LSMAD	LRASR	KIFD	LSC-TV
Salinas	<b>0.1604</b>	43.35	50.94	21.20	88.28	117.7	52.30
HYDICE	<b>0.1109</b>	15.29	18.34	9.110	41.20	50.86	58.11
Pavia	<b>0.1686</b>	17.69	22.04	17.78	177.7	100.6	121.7
Cri	<b>0.5531</b>	74.76	1021	52.53	1693	326.6	757.5
Sandiego	<b>0.5213</b>	128.8	144.8	70.21	574.0	162.3	273.5

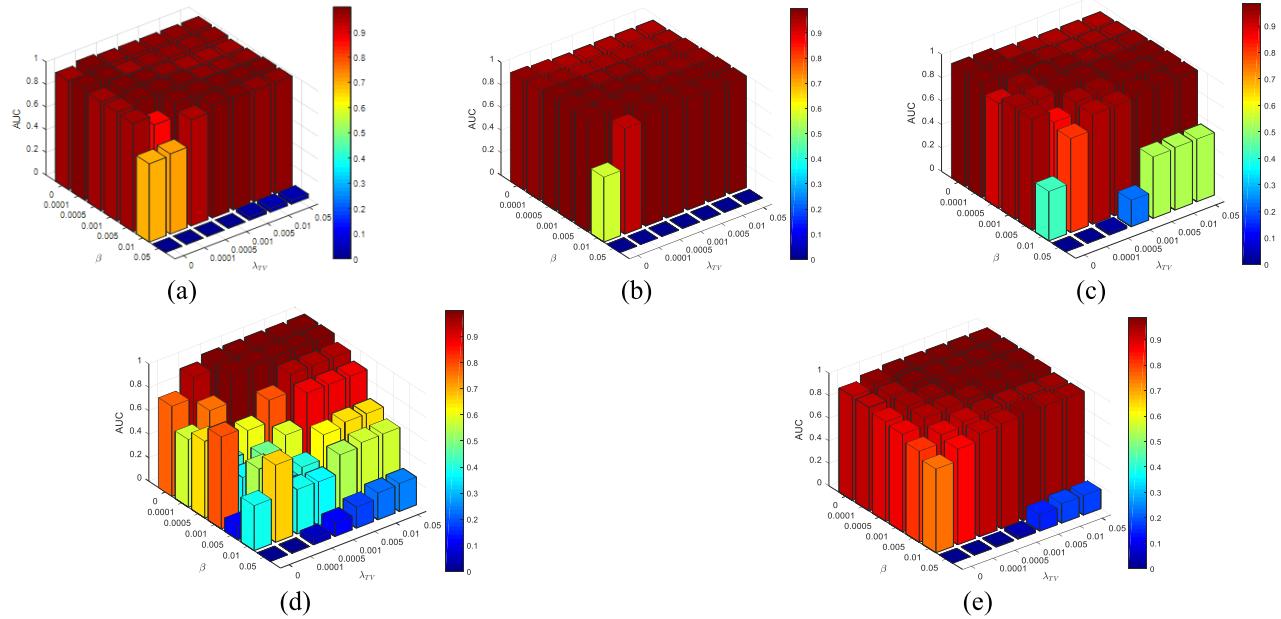


Fig. 12. AUC values for different parameter combinations in the proposed method. (a) Salinas data. (b) HYDICE data. (c) Pavia data. (d) Cri data. (e) Sandiego.

high-efficiency methods, GoDec and RX, to apply RX to the background extracted by GoDec. The running time has certain advantages compared with the other five detectors (except RX). For the data of two sets of large-scale scenes (i.e., Cri and Sandiego), the proposed method crops them properly before executing the image and then reslices the resulting image. Although the proposed LSC-TV method reduces the process of singular value decomposition (SVD) for the background

estimation, the additional TV model adds the running time of the LSC-TV method. Therefore, the proposed method has a calculation time equivalent to that of the LRASR method.

3) *Parameter Sensitivity*: The proposed method involves the selection of two parameters:  $\lambda_{TV}$  and  $\beta$ . Among them, parameter  $\lambda_{TV}$  indirectly affects the result of anomaly separation by affecting the smoothness between background pixels, and parameter  $\beta$  directly controls the effect of separation.

An excessively large value means that the anomaly target cannot be separated. When the value is too small, the background is incorrectly separated into anomaly components. Therefore, the sensitivity of these two parameters is worth discussing. By calculating the AUC under different parameter combinations between (0, 0.05), the rules of parameter selection can be explored, and the results of this exploration are shown in Fig. 12.

As expected, since parameter  $\beta$  directly affects the detection effect, the proposed method is more sensitive to  $\beta$ . Consistent with the LRR method, when  $\beta$  is set to a large value, the detector will not detect any anomaly. The requirements for columnwise sparsity of anomaly components are too strict such that all columns are equal to zero. However, within a certain range of  $\beta$ , good detection results can be achieved. For the normalized five groups of data, satisfactory performance can be obtained when  $(\lambda_{TV}, \beta)$  is set near (0.001, 0.0001). In addition, an interesting phenomenon occurs:  $\beta$  becomes equal to zero, which means that the anomaly at this time is collected through the residual of  $\mathbf{X}-\mathbf{DZ}$ . This detection result is still satisfactory, which further illustrates the powerful capabilities of the LSC-TV method regarding background extraction.

#### IV. CONCLUSION

In this article, a new method of hyperspectral anomaly detection, named LSC-TV, is proposed. Compared with the assumption that the background pixels have a global low-dimensional structure in the LRR method, the proposed method follows the fact that the background part shows a uniformly distributed feature locally. By using superpixel segmentation to extract homogeneous regions, the background estimation of homogeneous regions is obtained based on the anomaly index of each pixel. A weighted diagonal matrix, which moves the background pixels closer to the background estimation and, conversely, moves the anomaly farther away from the background estimation, is used to help separate anomaly from the background to the anomaly components. Compared with the traditional LRR method, the effectiveness of the proposed LSC model for separating anomaly can be demonstrated. Furthermore, the background pixels have a high correlation with the surrounding neighborhood, while the anomalies are weak. Taking advantage of this fact, the nonisotropic TV model has been considered to complement the original LSC method. To verify the effectiveness of the proposed algorithm, LSC-TV is compared with classical algorithm and state-of-the-art algorithm methods. In the experimental results of simulated dataset and real datasets, LSC-TV has performed better than the other six algorithms in terms of visual color detection maps and quantitative evaluation values. The proposed algorithm can detect more anomalies at a lower false alarm rate. All these demonstrate the powerful detection ability of the LSC-TV algorithm.

#### ACKNOWLEDGMENT

The authors would like to thank Prof W. Li for sharing the CRD code, Prof K. Tan for sharing the simulated Salinas

datasets, and Prof. Y. Zhang for sharing the Cri datasets. They would also like to thank editors, associated editors, and anonymous reviewers for their insightful suggestions and comments, which significantly improve this article.

#### 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] M. E. Paoletti *et al.*, "Capsule networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2145–2160, Apr. 2019.
- [3] Z. Liu, L. Ma, and Q. Du, "Class-wise distribution adaptation for unsupervised classification of hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 508–521, Jan. 2021.
- [4] N. M. Nasrabadi, "Hyperspectral target detection: An overview of current and future challenges," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 34–44, Jan. 2014.
- [5] D. Manolakis and G. S. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.
- [6] W. Xie, Y. Li, J. Lei, J. Yang, C.-I. Chang, and Z. Li, "Hyperspectral band selection for spectral-spatial anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3426–3436, May 2020.
- [7] S. Matteoli, M. Diani, and J. Theiler, "An overview of background modeling for detection of targets and anomalies in hyperspectral remotely sensed imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2317–2336, Jun. 2014.
- [8] M. Vafadar and H. Ghassemian, "Hyperspectral anomaly detection using combined similarity criteria," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4076–4085, Nov. 2018.
- [9] A. Taghipour and H. Ghassemian, "Hyperspectral anomaly detection using attribute profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1136–1140, Jul. 2017.
- [10] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.
- [11] J. M. Molero, E. M. Garzón, I. García, and A. Plaza, "Analysis and optimizations of global and local versions of the RX algorithm for anomaly detection in hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 801–814, Apr. 2013.
- [12] D. Borghys, I. Kåsen, V. Achard, and C. Perneel, "Comparative evaluation of hyperspectral anomaly detectors in different types of background," *Proc. SPIE*, vol. 8390, pp. 8390-1–8390-12, May 2012.
- [13] A. Schaum, "Joint subspace detection of hyperspectral targets," in *Proc. IEEE Aerosp. Conf.*, Mar. 2004, pp. 1818–1824.
- [14] N. M. Nasrabadi, "Regularization for spectral matched filter and RX anomaly detector," *Proc. SPIE*, vol. 6966, Apr. 2008, Art. no. 696604.
- [15] 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.
- [16] B. Du and L. Zhang, "Random-selection-based anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1578–1589, May 2011.
- [17] 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, Sep. 2000.
- [18] H. Kwon and N. M. Nasrabadi, "Kernel RX-algorithm: A nonlinear anomaly detector for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 2, pp. 388–397, Feb. 2005.
- [19] A. Banerjee, P. Burlina, and C. Diehl, "A support vector method for anomaly detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2282–2291, Aug. 2006.
- [20] Z. Wu, W. Zhu, J. Chanussot, Y. Xu, and S. Osher, "Hyperspectral anomaly detection via global and local joint modeling of background," *IEEE Trans. Signal Process.*, vol. 67, no. 14, pp. 3858–3869, Jul. 2019.
- [21] S. Li, K. Zhang, P. Duan, and X. Kang, "Hyperspectral anomaly detection with kernel isolation forest," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 319–329, Jan. 2020.
- [22] J. Li, H. Zhang, L. Zhang, and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2523–2533, Jun. 2015.

- [23] Y. Zhang, B. Du, L. Zhang, and T. Liu, "Joint sparse representation and multitask learning for hyperspectral target detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 894–906, Feb. 2017.
- [24] 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.
- [25] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, pp. 1–37, May 2011.
- [26] S. Li, W. Wang, H. Qi, B. Ayhan, C. Kwan, and S. Vance, "Low-rank tensor decomposition based anomaly detection for hyperspectral imagery," in *Proc. IEEE ICIP*, Sep. 2015, pp. 4525–4529.
- [27] 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.
- [28] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 663–670.
- [29] 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.
- [30] 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.
- [31] N. Huyan, X. Zhang, H. Zhou, and L. Jiao, "Hyperspectral anomaly detection via background and potential anomaly dictionaries construction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 4, pp. 2263–2276, Apr. 2019.
- [32] T. Cheng and B. Wang, "Graph and total variation regularized low-rank representation for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 1, pp. 391–406, Jan. 2020.
- [33] F. Fan, Y. Ma, C. Li, X. Mei, J. Huang, and J. Ma, "Hyperspectral image denoising with superpixel segmentation and low-rank representation," *Inf. Sci.*, vols. 397–398, pp. 48–68, Aug. 2017.
- [34] L. Fang, H. Zhuo, and S. Li, "Super-resolution of hyperspectral image via superpixel-based sparse representation," *Neurocomputing*, vol. 273, pp. 171–177, Jan. 2018.
- [35] T. Zhan, L. Sun, Y. Xu, G. Yang, Y. Zhang, and Z. Wu, "Hyperspectral classification via superpixel kernel learning-based low rank representation," *Remote Sens.*, vol. 10, no. 10, pp. 1639–1655, Oct. 2018.
- [36] Q. Wang, X. He, and X. Li, "Locality and structure regularized low rank representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 911–923, Feb. 2019.
- [37] L. Zhuang, L. Gao, B. Zhang, and J. M. Bioucas-Dias, "Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations," *Proc. SPIE*, vol. 10427, Oct. 2017, Art. no. 104270M.
- [38] B. Rasti, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral feature extraction using total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 6976–6985, Dec. 2016.
- [39] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.
- [40] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. NIPS*, 2009, pp. 2080–2088.
- [41] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Adv. Neural Inf. Process. Syst.*, 2011, pp. 612–620.
- [42] A. Ertürk, "Enhanced unmixing-based hyperspectral image denoising using spatial preprocessing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2720–2727, Jun. 2015.
- [43] T. Priya, S. Prasad, and H. Wu, "Superpixels for spatially reinforced Bayesian classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 1071–1075, May 2015.
- [44] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. ICCV*, vol. 1, Oct. 2003, pp. 10–17.
- [45] R. Achanta, A. Shajii, 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.
- [46] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, 1992.
- [47] M. Nikolova, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1, pp. 89–97, Jan. 2004.
- [48] J. Eckstein and D. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 3, pp. 293–318, Jun. 1992.
- [49] P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Model. Simul.*, vol. 4, no. 4, pp. 1168–1200, Jan. 2005.
- [50] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
- [51] J. Kerekes, "Receiver operating characteristic curve confidence intervals and regions," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 251–255, Apr. 2008.
- [52] S. M. Schweizer and J. M. F. Moura, "Efficient detection in hyperspectral imagery," *IEEE Trans. Image Process.*, vol. 10, no. 4, pp. 584–597, Apr. 2001.
- [53] 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, p. 1578, Jul. 2019.



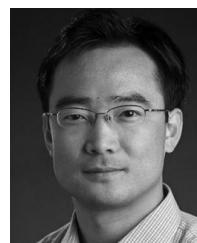
**Ruyi Feng** (Member, IEEE) received the B.S. degree in geographic information system from Hunan Normal University, Changsha, China, in 2011, and the M.S. degree in surveying and mapping engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2013 and 2016, respectively.

She has been with the School of Computer Science, China University of Geosciences, Wuhan, since 2016, where she is currently an Associate Professor. Her research interests include sparse representation, deep learning, hyperspectral image analysis, high-resolution remote sensing understanding, and intelligent interpretation of remote sensing imagery.



**Hao Li** received the B.S. degree in computer science and technology from China University of Geosciences, Wuhan, China, in 2018, where he is currently pursuing the M.S. degree in computer science and technology.

His research interests include hyperspectral image analysis, sparse representation, low-rank representation, spectral unmixing, and anomaly detection.



**Lizhe Wang** (Fellow, IEEE) received the B.S. and M.S. degrees from Tsinghua University, Beijing, China, in 1998 and 2001, respectively, and the D.E. degree (*magna cum laude*) from the University of Karlsruhe, Karlsruhe, Germany, in 2007.

He is currently the Dean and the "ChuTian" Chair Professor with the School of Computer Science, China University of Geosciences, Wuhan, China. His research interests include remote sensing data processing, digital Earth, and big data computing.

Dr. Wang is a fellow of the Institution of Engineering and Technology (IET) and the British Computer Society (BCS) and an Associate Editor of *Remote Sensing*, *International Journal of Difference Equations*, and so on. He was a recipient of the Distinguished Young Scholars of NSFC, the National Leading Talents of Science and Technology Innovation, and the 100-Talents Program of the Chinese Academy of Sciences. He is also an Associate Editor of *Remote Sensing*, the *International Journal of Digital Earth*, *ACM Computing Surveys*, the *IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS*, and the *IEEE TRANSACTIONS ON SUSTAINABLE COMPUTING*.



**Yanfei Zhong** (Senior Member, IEEE) received the B.S. degree in information engineering and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2002 and 2007, respectively.

Since 2010, he has been a Full Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing (LIESMARS), Wuhan University. He organized the Intelligent Data Extraction, Analysis and Applications of Remote Sensing (RSIDEA) research group.

He has published more than 100 research articles in international journals, such as *Remote Sensing of Environment*, *ISPRS Journal of Photogrammetry and Remote Sensing*, and *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*. His research interests include hyperspectral remote sensing information processing, high-resolution remote sensing image understanding, and geoscience interpretation for multisource remote sensing data and applications.

Dr. Zhong is a fellow of the Institution of Engineering and Technology (IET). He was a recipient of the 2010 Best Paper Theoretical Innovation Award from the International Society for Optics and Photonics (SPIE). He received the Second Place Prize in the 2013 IEEE GRSS Data Fusion Contest and the Single-view Semantic 3-D Challenge of the 2019 IEEE GRSS Data Fusion Contest. He is also serving as an Associate Editor for the *IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING* and the *International Journal of Remote Sensing*.

Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University. He was a Principal Scientist for the China State Key Basic Research Project from 2011 to 2016 appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has published more than 700 research articles and five books. He is the Institute for Scientific Information (ISI) Highly Cited Author. He holds 15 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a fellow of the Institution of Engineering and Technology (IET). He was a recipient of the 2010 Best Paper Boeing Award, the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing (ASPRS), and the 2016 Best Paper Theoretical Innovation Award from the International Society for Optics and Photonics (SPIE). His research teams won the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS) student paper contest in recent years. He is the Founding Chair of the IEEE Geoscience and Remote Sensing Society (GRSS) Wuhan Chapter. He also serves as an associate editor or editor for more than ten international journals. He is serving as an Associate Editor for the *IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING*.



**Liangpei Zhang** (Fellow, IEEE) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently a “Chang-Jiang Scholar” Chair Professor appointed by the Ministry of Education of China at the State Key Laboratory of Information



**Tieyong Zeng** received the B.S. degree from Peking University, Beijing, China, in 2000, the M.S. degree from the Ecole Polytechnique, Palaiseau, France, in 2004, and the Ph.D. degree from the University of Paris XIII, Paris, France, in 2007.

He is currently an Associate Professor with the Department of Mathematics, The Chinese University of Hong Kong, Hong Kong. His research interests include image processing, optimization, artificial intelligence, scientific computing, computer vision, machine learning, and inverse problems.