Multiple Spectral-Spatial Representation Based on Tensor Decomposition for HSI Anomaly Detection

Yujian Wang, Dan Li, Hanjie Wu, Xiaojun Li, Fanqiang Kong, and Qiang Wang, *Member, IEEE*

***Abstract*—To exploit the spectral-spatial information of hyperspectral image (HSI) and achieve higher the detection accuracy, a novel multiple spectral-spatial representation based on tensor decomposition method is proposed for HSI anomaly detection (AD) in this article. First, tensor decomposition is employed to separate the original HSI dataset into anomalous tensor from background tensor to, respectively, utilize the priors of background and anomaly. Then, the segmented smoothing prior is characterized by l0-l1 hybrid total variation regularization for the spatial dimensions of the background tensor, which can not only complement the image details but also fully consider the global spectral-spatial structure. Moreover, the low-rank prior is represented with truncated nuclear norm regularization for the spectral dimension of the background tensor to make full utilization of global information in the background and reduce data redundancy. Next, l2,1-norm regularization is applied to characterize the sparse prior for the anomaly tensor, which is beneficial to inscribe the local spectral structure of each image element. Finally, all these priors are fuses to model a unified convex optimization problem and obtain the final results by alternating direction method of multipliers, which can fully explore the discriminative spectral-spatial priors and improve the AD accuracy significantly. Experiments carried out on several real data indicate that our presented method reaches excellent performance in comparison to a few advanced AD methods.**

***Index Terms*—$_1$-norm, hyperspectral image (HIS) anomaly detection, l_0-l_1 hybrid TV (l_0-l_1HTV), l_2, truncated nuclear norm (TNN).**

I. Introduction

COMPARED with traditional two-dimensional (2-D) image, hyperspectral image (HSI) contains numerous narrow and continuous spectral bands spanning various wavelengths, which provides a wealth of information on the spatial spectrum for the observation of terrestrial objects [1], [2]. As each feature has different characteristics of electromagnetic reflection at various wavelengths, target detection allows the application of their spectral information [3]. Based on whether prior knowledge of target features is employed, target detection could be classified as supervised target detection and unsupervised anomaly target detection [4]. In practical applications, the target detection algorithms are more limited due to the difficulty of obtaining accurate prior information about the target to be detected; in contrast, anomaly detection methods are more practical and have a more comprehensive range of applications [5]–[7]. Anomaly detection (AD) is commonly adopted in various applications, such as geological survey [9], intelligence, environmental surveillance, and military defense [6]–[10].

Finally, numerous HSI anomaly detectors have been proposed. The best-known AD method for HSI is Reed-Xiaoli (RX) algorithm [11], [12], which supposes the background to a multivariable Gaussian distribution. The similarity of the tested pixel to the surrounding background is determined through the calculation of the Mahalanobis distance. In general, global RX detection (GRX) [13]–[15] uses all pixels in the scene to approximate the background statistics, whereas local RX detection (LRX) is estimated considering only the neighboring pixels of the test pixel. In practice, however, the background components of HSI are often complex and heterogeneous. In addition, the calculated statistics are susceptible to contamination due to the presence of noise and anomalies. On the basis of these limitations, a couple of modified RX-based methods are presented. For instance, Kernel RX [16]–[18] characterizes non-Gaussian distributions over high dimensional eigenspace by reprojecting the primitive data onto the high-dimensional space with the kernel function. The subspace-based RX [19] method exploits the expected difference in spectral variability between background and anomalies to achieve more accurate detection performance. To enhance the robustness of background estimation to anomalies, weighted RX [20] and blocking adaptive computation efficient outlier nominator [21] are presented. However, the detection results are restricted by these methods since the background-specific distribution assumptions are usually invalid.

To break through this limitation, attention has been cast to representation-based methods devoid of any statistical assumptions [22]–[25]. Such approaches are based on the fact that backgrounds could be characterized by the constructed dictionaries, while anomalies cannot. The aim of this type of method is to construct an ultracomplete dictionary with the least amount of data information possible to briefly represent the entire image. Nevertheless, these methods consider only spectral information without spatial information, thus, it is challenging to acquire satisfying detection performance. Li and Du [26] proposed a collaborative-representation-based detector (CRD) for anomaly detection. In essence, CRD means that the background image

Manuscript received March 16, 2022; revised April 11, 2022; accepted April 19, 2022. Date of publication May 3, 2022; date of current version May 13, 2022. *(Corresponding author: Dan Li.)*

Yujian Wang, Dan Li, Hanjie Wu, and Fanqiang Kong are with the College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: yujianwang@nuaa.edu.cn; lidanhit@163.com; wuhanjie@nuaa.edu.cn; kongfq@nuaa.edu.cn).

Xiaojun Li is with National Key Laboratory Science and Technology Space Microwave, China Academic Space Technology, Xian 710018, China (e-mail: lixj@cast504.com).

Qiang Wang is with Control Science and Engineering, Harbin Institute of Technology, Harbin 150001, China (e-mail: wangqiang@hit.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2022.3170057

elements are well indicated by their surroundings, while anomalies are not. Compared with CRD, collaborative representation-based with outlier removal anomaly detector [27] method eliminates outliers that differ markedly from the background, and enables a more accurate background approximation. Du *et al.* [28] proposed local summation anomaly detection model, which fuses spatial-spectral features with numerous local windows around the tested pixel to take advantage of the spatial nearest-neighbor relationship of its local background distribution. Apart from CRD, several other detectors also have incorporated spatial information into their detection. Li *et al.* [29] proposed background joint sparse representation detector, which automatically chooses the best typical background bases in each partial domain and projects test pixels into the orthogonal subspace formed by these background bases to detect anomalies. Minh *et al.* [30] developed a framework to combine spatial, spectral, temporal, and feature domain information for detecting spectral anomalies and applied it to soil detection. Generally speaking, most of methods mentioned above detect pixels in an independent manner and do not exploit pixel-to-pixel correlation in hyperspectral images.

More recently, low-rank-based approaches have also been adopted for HSI anomaly detection. They utilize the low-rank property of background and the sparsity of anomaly [31]. The classical robust principal component analysis (RPCA) [32] detects anomaly by separating them from the low rank background. Chen *et al.* [33] proposed RPCA-RX anomaly detection algorithm, which applies the classical RX method to sparsity matrix to improve the detection accuracy. Later, low-rank representation (LRR) method is established [31], [34], whereby it supposes its data is derived from multiple subspaces. Owing to the sophisticated background characteristics of the actual data, it is more suitable for hyperspectral image detection. The LRR model is initially taken up for the problem of hyperspectral anomaly detection in low-rank and sparse representation (LRASR) [35]. Sparse components are modeled by adopting a dictionary construction strategy, and the l_2-norm of the sparse components is chosen to characterize the anomaly. LRaSMD-based Mahalanobis distance method for hyperspectral anomaly detection (LSMAD) [29], [36] essentially exploits the low-rank property of the background to compute background statistics, which enhances detection efficiency by mitigating anomalous contamination and inverse covariance matrix problems. However, most low-rank and sparse decomposition related methods only take into account the low-rank property of the background and the sparsity of the anomaly, but ignore the structural features of HSI itself and the segmental smoothness in the background. Total variation (TV) regularization-based image restoration technique is one of the more popular techniques nowadays, which can effectively enhance spatial smoothness and preserve image edge information [37]–[39]. Cheng and Wang [40] introduced a TV regularization term to improve detection rates by taking into account the spatial homogeneity of background. Besides, Cheng and Wang [41] proposed a graph and TV regularized LRR (GTVLRR) method that preserves regional geometric features and spatial information in HSI. Then, TV and sparsity regularized decomposition model is established by Cheng *et al.* [42]. They characterize background spatial smoothness and anomaly sparsity by TV and sparse regularization, respectively, to improve the detection performance. Even though the problem of segmental smoothness is solved by TV regular terms, the 3-D spatial structure features of HSI are not fully utilized due to matrix transformation. Also, global feature extraction is not considered. Li *et al.* [43] proposed a prior-based tensor approximation (PTA) model, which employs the traditional linear TV parametrization instead of the original TV parametrization to enhance the smoothness in HSI. However, it still has the limitation that it cannot extract image details. To address this issue, l_0-l_1 hybrid TV(l_0-l_1HTV) is introduced to characterize global spectral-spatial smoothness to complement image details such as sharp edges and valuable textures.

In addition to the aforementioned methods, tensor decomposition-based methods that exploit the structural information of the 3-D data are extensively applied recently [44], [45]. Low-rank tensor decomposition-based anomaly detection designed by Li *et al.* [46] utilizes Tucker decomposition to further decompose the low-rank tensor for anomaly detection. Xu *et al.* [47] employed tensor robust principal component analysis model, which comprehensively describes the spectral–spatial information in HSI. Zhang *et al.* [48] proposed a TenB algorithm, which finds the first maximum PC along each mode to remove background features in the greatest amount possible and model the surplus data as the anomaly. These methods fully consider the low rank and sparse properties of HSI, while neglect its segmental smoothness.

In this article, a novel multiple spectral-spatial representation based on tensor decomposition (MSSR) model is proposed for hyperspectral anomaly detection. First, tensor decomposition is employed to separate the original HSI dataset into anomalous components and background components, which is beneficial to utilize the individual priors of background and anomaly. Then, the background tensor is decomposed into two spatial dimensions and one spectral dimension to fully explore the discriminative 3-D spatial structure feature of HSI. Next, the low-rank prior for the spectral dimension of background tensor is modeled by utilizing truncated nuclear norm (TNN) regularization [49] with the aim of making reasonable utilization of the global background information and reduces data redundancy. Moreover, the segmented smoothing prior is characterized by l_0-l_1HTV regularization for the two spatial dimensions of the background tensor, which can not only complement the image details but also fully consider the global spectral spatial structure. l_0-l_1HTV is a globally and locally integrated TV regularization that captures local information across spatial and spectral domains to recover the overall structure of HSIs. The TV regularization compensate each other to overcome artifacts or oversmoothing drawbacks. Segmental smoothing prior is caused by the smooth variation of feature targets in HSI, whose distribution tends to remain coherent and homogeneous. Mutations are only present at a few feature edges or anomalies. Therefore, l_0-l_1HTV regularization is constructed to strengthen the segmental smoothing of the background in the spatial dimension to improve detection accuracy. In addition, an $l_{2,1}$-norm regularization on the spectral dimension is employed to represent the sparse prior for the anomaly tensor, which can inscribe the local spectral structure

of each image element. All these prior regularizations are fused and modeled into a unified convex optimization problem to fully explore more discriminative spectral-spatial prior information and improve the AD accuracy significantly. Eventually, the presented model is optimized by the alternating direction method of multipliers (ADMM) method to obtain the final results. Three HSI datasets are tested for the purpose of verifying the presented method and the results demonstrate that our presented method detects more accurately than several advanced methods.

Our proposed method mainly has the following contributions.

1) The low-rank prior in the spectral dimension of the background tensor is characterized by truncated kernel norm, which can not only adequately explore the global spatial information of the background but also reduce the data redundancy.
2) l_0-l_1HTV regularization is constructed to model the segmental smoothing prior to the spatial dimensions of the background tensor, which can fully exploit the global spectral-spatial smoothness and the local spatial-spectral structures to complement image details and improve the detection accuracy.
3) $l_{2,1}$-norm regularization is employed to represent the sparse prior for the anomaly tensor and accurately use the local spectral structure of each image element.
4) The unified convex optimization problem constrained with low-rank prior, segmental smoothing prior and sparse prior is modeled to fully explore discriminative spectral-spatial prior information and solved by ADMM, whose results are outstanding over several advanced methods.

The rest of this article is organized as follows. Notations and preliminaries of TV are introduced in Section II. The detailed introduction of our presented method is given in Section III. Section IV provides the results and interpretations of the experiments in three true datasets. Finally, Section V concludes this article.

II. Related Works

A. Preliminaries and Notations

In this article, boldface capital letters are applied to denote matrices, e.g., $\mathbf{X}$. Vectors, scalars, and tensor are denoted by boldface lowercase letters e.g., $\mathbf{x}$, lowercase letters, e.g., x, and Euler script letters, e.g., $\mathcal{X}$, respectively. For a 3-D hyperspectral image tensor $\mathcal{X} \in \mathbb{R}^{x\times y\times d}$, where x, y, d denote the height, the width and the number of spectral bands, respectively. x_{mn} is denoted as the (m, n)th item of the matrix, and use x_{mnl} to represent the (m, n, l)th item of the tensor. The lth largest singular value of the matrix $\mathbf{X}$ is defined by $\omega_k(\mathbf{X})$. The Frobenius norm is written as $\|\mathbf{X}\|_F = (Tr(\mathbf{X}^T\mathbf{X}))^{1/2}$, where $\mathbf{X}^T$ is the transpose of the matrix $\mathbf{X}$ and $Tr(\mathbf{.})$ represents the trace of the matrix.

Definition 1: (mixed $l_{1,0}$ pseudonorm [50]): Let $z \in \mathbb{R}^{\mathrm{h}}$ be a vector, and $\zeta_1, \ldots, \zeta_i, \ldots, \zeta_w (1 \le w \le h)$ be the index sets satisfying the following situations.

1) Each ζ_i is a subset of $\{1, \ldots, h\}$.
2) $\zeta_i \cap \zeta_j = \emptyset$ for any $i \neq j$.
3) $\cup_{i=1}^{w} \zeta_i = 1, \ldots, h$.

The mixed $l_{1,0}$ pseudonorm of $\mathbf{z}$ is defined by

$$\|\mathbf{z}\|_{1,0}^{\zeta} = \left\|\left(\|\mathbf{z}_{\zeta_1}\|_1, \ldots, \|\mathbf{z}_{\zeta_i}\|_1, \ldots, \|\mathbf{z}_{\zeta_w}\|_1\right)\right\|_0 \tag{1}$$

where $\mathbf{z}_{\zeta_i}$ indicates a subvector of $\mathbf{z}$ whose entries are designated by ζ_i. $\|\cdot\|_1$ is described as the sum of the absolute values of all terms of $(\cdot)$, and $\|\cdot\|_0$ is described as an amount of nonzero terms of $(\cdot)$.

Definition 2: (function $\ell(\mathbf{z})$ of the inequality constraint on the mixed $l_{1,0}$ pseudonorm [50] is defined by

$$\ell_{\left\{\|\xi\cdot\|_{1,0}^{\zeta}\le\eta\right\}}(z) = \begin{cases} 0, & \|\xi z\|_{1,0}^{\zeta} \le \eta \\ \infty, & \text{otherwise} \end{cases} \tag{2}$$

where ξ is a diagonal tensor with binary entries (0 or 1) in this article.

B. l_0 Hyperspectral TV

Xu *et al.* [51] introduced the l_0 gradient minimization model to maintain the edge information of the image and sharpen image edges effectively. For a traditional 2-D image $\mathbf{X} \in \mathbb{R}^{x\times y}$, the l_0 gradient is written by

$$\mathrm{Grad}l_0 = \sum_{m}^{x}\sum_{n}^{\mathrm{y}} \mathcal{G}(|\mathbf{X}_{m+1,n} - \mathbf{X}_{m,n}| + |\mathbf{X}_{m,n+1} - \mathbf{X}_{m,n}|) \tag{3}$$

where $\mathcal{G}(\mathbf{X})$ denotes a binary function calculating the amount of nonzero imagery gradients and $\mathcal{G}(\mathbf{X}_{m,n}) = 1$, if $\mathbf{X}_{m,n} \neq 0$; $\mathcal{G}(\mathbf{X}_{m,n}) = 0$, otherwise.

III. Proposed Method

The general illustration of our presented MSRR is shown in Fig. 1. As shown in Fig. 1, there are three major procedures in our method: tensor decomposition, modeling of the proposed method, and optimization by ADMM. More details of each step are described below.

A. Objective Function

The 3-D HSI is assumed to be a third-order tensor $\mathcal{Y}$, which allows for decomposition to a background tensor $\mathcal{X}$ and an anomaly tensor $\mathcal{S}$. It can be represented by

$$\mathcal{Y} = \mathcal{X} + \mathcal{S}. \tag{4}$$

Since adjacent pixels in HSI have similar spectral properties and their spectral features are highly correlated, the background tensor of HSI shows low-rank property as prior. Aiming at depicting the global structure of the background and characterize the low-rank prior in the spectral dimension of the background, TNN regularization is adopted for constraint. Segmental smoothing property is caused by the smooth variation of anomaly targets in hyperspectral images, whose distribution tends to remain coherent and homogeneous. Mutations only occur in a few feature edges or anomalies. Therefore, l_0-l_1HTV regularization is constructed to strengthen the segmental smoothing of the background along the spatial dimension. For the anomaly tensor, as anomalies are only a small fraction compared to the background, it shows sparse characteristics. In

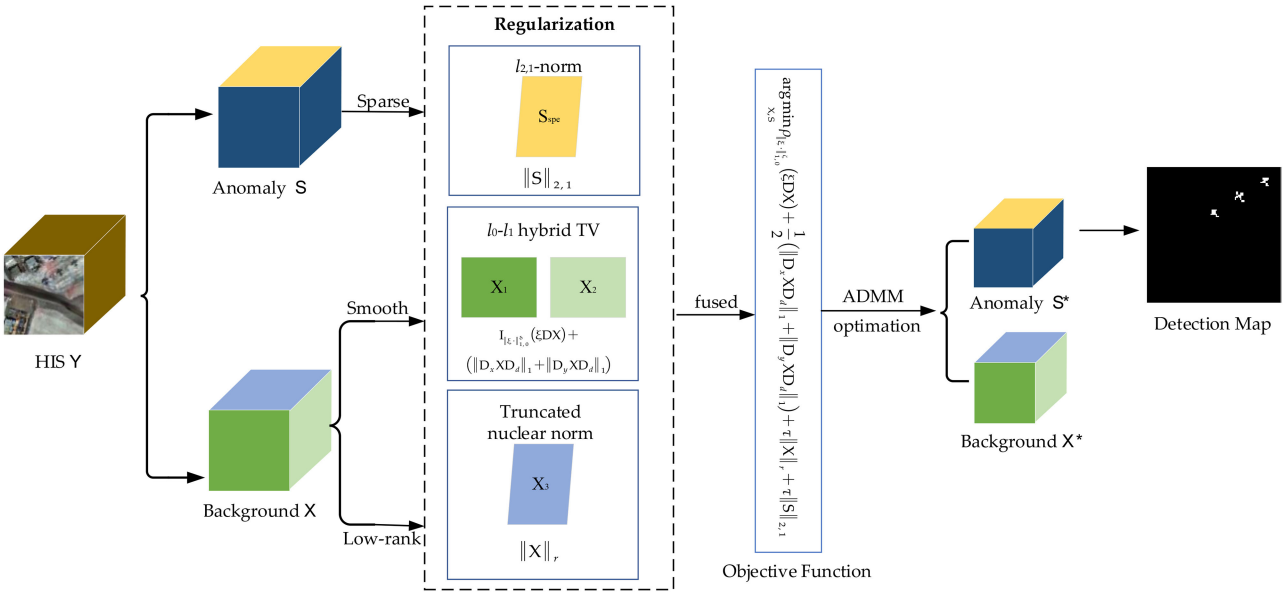


Fig. 1. Illustration of the proposed MSRR.

view of the accurate representation of each image element, the sparse constraint is placed on its representation coefficients, and the local spectral structure of each image element is inscribed. In this article, these prior constraints are fused to create a novel optimization model that exploits spatial-spectral information to improve detection accuracy, which can be expressed as

$$\arg\min_{X,S} \rho_{\|\xi\cdot\|_{1,0}^{\zeta}}(\xi \mathrm{DX}) + \frac{1}{2}(\|D_x \mathrm{XD}_d\|_1 + \|D_y \mathrm{XD}_d\|_1) + \tau\|X\|_r + \tau\|S\|_{2,1} \quad (5)$$

where D_x and D_y denote the 2-D difference operators for the length and width of the image along the spatial dimension, respectively. D_d is a 1-D finite- difference operator that employs the spectral features of the image. This regularization explores the correlation between both spatial and spectral dimensions. The first two items of the objective function in (5) denote l_0-l_1HTV regularization, and the third and the fourth items are TNN regularization and $l_{2,1}$-norm regularization, respectively. The regularization constraints employed in this model are, respectively, described in detail as follows.

1) l_0-l_1 Hybrid TV Regularization: Despite the fact that the traditional l_0 gradient leverages the spatial smoothing of HSI, it disregards the spectral correlation of HSI. The edges are strengthened by limiting the amount of nonzero gradients, and simultaneously realize smoothing globally. For HSI $\mathcal{X} \in \mathbb{R}^{x\times y\times d}$, $TV_{l_0}(\mathcal{X})$ is defined as

$$TV_{l_0}(\mathcal{X}) = \sum_{m}^{x}\sum_{n}^{y}\mathcal{G} \times \left(\sum_{l}^{d}(|\mathcal{X}_{m+1,n,l} - \mathcal{X}_{m,n,l}| + |\mathcal{X}_{m,n+1,l} - \mathcal{X}_{m,n,l}|)\right) \quad (6)$$

where $\mathcal{G}(\mathcal{X}) := 1$, if $x \neq 0$; $\mathcal{G}(\mathcal{X}) := 0$, otherwise. Boundary values of gradients are given by $|\mathcal{X}_{m+1,n,l} - \mathcal{X}_{m,n,l}| = 0$, if $m = x$, and $|\mathcal{X}_{m,n+1,l} - \mathcal{X}_{m,n,l}| = 0$, if $n = y$. In short, nonzero gradients in the spatial dimension are calculated utilizing all spectral bands, which greatly enhances the spatial edges of HSI.

For the mixed $l_{1,0}$ pseudonorm characterized by Definition 1, it calculates the quantity of subvectors where the l_1-norms are nonzero. Then, another formulation of the gradient is presented as

$$TV_{l_0}(\mathrm{X}) = \|\xi DX\|_{1,0}^{\zeta} \quad (7)$$

where $\mathcal{D}$ is a discrete difference operator with periodic boundaries that maps HSI to their vertical and horizontal discrete differences. ξ is a diagonal matrix with binary elements (0 or 1) that forces the discrete differences between opposite boundaries (due to the periodic boundary conditions) to be zero.

With continuous iterative optimization, the edges of HSI will gradually get sharper. However, if the l_0-norm is overpenalized, the eventual result will be oversmoothed. To make fuller utilization of global and local spatial information, l_0-l_1 hybrid TV (l_0-l_1HTV) is constructed as

$$l_0\text{-}l_1\mathrm{HTV}(\mathrm{X}) = \rho_{\|\xi\cdot\|_{1,0}^{\zeta}}(\xi \mathrm{DX}) + (\|D_x \mathrm{XD}|_d\|_1 + \|D_y \mathrm{XD}_d\|_1) \quad (8)$$

where $\|D_x \mathrm{XD}_d\|_1$ and $\|D_y \mathrm{XD}_d\|_1$ are designed to characterize local smoothing in the spatial dimension of the data and $\rho_{\|\xi\cdot\|_{1,0}^{\zeta}}(\xi DX)$ is designed to enhance image edges.

The l_0-l_1HTV employs an l_0 penalty term to gauge gradient sparsity, and enhances convergence by applying l_1 at most positions in the image. The novel regularization integrates l_0 and l_1 effectively and, thus, has the advantages of both. l_0-l_1HTV is a globally and regionally integrated TV regularization that is specifically designed to extract local information across spatial and spectral domains to recover the overall structure of the HSI. The global and local aspects of TV regularization compensate

each other to overcome artifacts or oversmoothing drawbacks. l_0-l_1TV can extract edge information better than conventional linear TV, thus enabling more accurate detection of anomalous targets and improving detection accuracy.

2) TNN Regularization: The traditional nuclear norm-based approach minimizes the sum of all singular values, but TNN regularization minimizes only the smallest $\min(x, y) - r$ singular values due to the fact that the rank of the matrix is only for the first r nonzero singular values. The nuclear norm cannot perfectly approximate the rank function. All nonzero singular values contribute about the same in rank(X), but contribute quite differently in the nuclear norm $\|\mathbf{X}\|_*$ of the HSI. In this manner, a more accurate and robust approximation of the rank function can be obtained. TNN can adequately describe the global structure information of background and reduce data redundancy.

The low-rank prior of $X \in \mathbb{R}^{x\times y}$ is stated in terms of truncated nuclear norm as regularization. Given two matrices $\mathbf{P} \in \mathbb{R}^{x\times y}$, $\mathbf{Q} \in \mathbb{R}^{x\times y}$ where $\mathbf{P}\mathbf{P}^T = \mathbf{Q}\mathbf{Q}^T = I_{r\times r}$. $\|\mathbf{X}\|_r$ is denoted as the sum of $\min(x, y) - r$ minimum singular values by

$$\|\mathbf{X}\|_r = \sum_{k=r+1}^{\min(x,y)} \omega_k(\mathbf{X}). \tag{9}$$

Since the rank of the matrix is not impacted with respect to the maximum nonzero singular values, they are neglected in the novel design of the truncated nuclear norm and focus only on minimizing the sum of the smallest $\min(x, y) - r$singular values. Equation (9) can be expressed as

$$\|\mathbf{X}\|_r = \|\mathbf{X}\|_* - \max_{\mathbf{P}\mathbf{P}^T=\mathbf{Q}\mathbf{Q}^T=\mathbf{I}_r} \mathrm{Tr}(\mathbf{P}\mathbf{X}\mathbf{Q}^T) \tag{10}$$

where $\|\mathrm{X}\|_* = \sum_{k=1}^{\min(x,y)} \omega_k$ is defined as the nuclear norm of $\mathbf{X}$, and ω_k is the kth largest singular value of $\mathbf{X}$. Generally speaking, the part corresponding to a large singular value will contain more information. The main part of them can be extracted by using singular value decomposition. After singular value decomposition (SVD) decomposition, the main singular values are retained to remove redundant information and dependencies between data. Therefore, $\|\mathbf{X}\|_r$ can be obtained with the following two-step method.

Step 1: First, $\mathbf{P}$and $\mathbf{Q}$ are obtained by maximizing the first r singular values via SVD.

Step 2: Then fix $\mathbf{P}$ and $\mathbf{Q}$ to update $\mathbf{X}$. The variables are optimized in the abovementioned two steps alternately and iteratively through the algorithm.

3) $l_{2,1}$-Norm Regularization: Compared to the background, only a small quantity of anomalies in hyperspectral images is relatively infrequent and, thus, it has a sparse characteristic. Applying the square of l_2-norm as the loss function is insensitive to smaller outliers but sensitive to larger ones, while employing the l_1-norm as the loss function is insensitive to larger outliers but sensitive to smaller ones. In order to address these issues, $l_{2,1}$-norm is introduced so that the model can better handle outliers. The local spectral structure of each image element is, thus, inscribed.

The anomaly tensor S is expanded in its spectral dimension and the $l_{2,1}$-norm regularization is exploited to characterize the sparse prior of the data. The $l_{2,1}$-norm regularization employed for the expanded matrix $\mathbf{S}_{\mathrm{spe}} \in \mathbb{R}^{d\times(xy)}$ appends the l_2-norm to the spectral dimension and the l_1-norm to the spatial dimension, which is defined as

$$\|\mathbf{S}_{\mathrm{spe}}\|_{2,1} = \sum_{n=1}^{xy} \sqrt{\sum_{m=1}^{d} s_{mn}^2} \tag{11}$$

where s denotes the element of the *l*th band of the *i*th row and *j*th column of the anomaly tensor. The solution to the $l_{2,1}$-norm regularization would be derived by addressing the following:

$$\arg\min_S \|\mathbf{S}_{\mathrm{spe}}\|_{2,1} + \frac{1}{2}\|\mathbf{S}_{\mathrm{spe}} - \mathbf{Y}\|_F^2. \tag{12}$$

B. Optimization Procedure

The novel model constructed in this article can be approximated by ADMM to alternately minimizes $\mathcal{X}$ and $\mathcal{S}$. The auxiliary variables $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_3$, $\mathbf{A}_4$, and $\mathbf{A}_5$ are introduced to enable the objective function to be independent and unconstrained, where $\mathbf{X}_i = \mathbf{A}_i$, $i = 1, 2, 3, 4$, and $\mathbf{A}_4 = \mathbf{A}_5$. Moreover, $\mathbf{E} = D_x\mathrm{X}\mathrm{D}_d$, $\mathbf{F} = D_y\mathrm{X}\mathrm{D}_d$, and $\mathbf{G} = DX$. The augmented Lagrangian function of (5) is rewritten by

$$\begin{aligned} &L(X, S, \mathbf{A}_i, i = 1, \ldots, 5, \mathbf{B}_1, \mathbf{B}_2, \mathbf{\Phi}, \mathbf{P}, \mathbf{Q}) \\ &= \rho_{\|\xi_\bullet\|_{1,0}^{\zeta}}(\mathbf{G}) + \frac{1}{2}(\|\mathbf{E}\|_1 + \|\mathbf{F}\|_1) + \alpha\|\mathbf{E} - D_x\mathrm{X}\mathrm{D}_d - \mathbf{A}_1\|_F^2 \\ &\quad + \alpha\|\mathbf{F} - D_y\mathrm{X}\mathrm{D}_d - A_2\|_F^2 + \alpha\|\mathbf{G} - DX - \mathbf{A}_3\|_F^2 \\ &\quad + \tau\left(\|\mathbf{A}_4\|_* - \max_{\mathbf{P}\mathbf{P}^T=\mathbf{Q}\mathbf{Q}^T=\mathbf{I}_r} \mathrm{Tr}(\mathbf{P}\mathbf{A}_5\mathbf{Q}^T)\right) + \tau\|\mathbf{S}_{\mathrm{spe}}\|_{2,1} \\ &\quad + \mu(\|\mathbf{X} - \mathbf{A}_4 + \mathbf{B}_1\|_F^2 + \|\mathbf{A}_5 - \mathbf{A}_4 + \mathbf{B}_2\|_F^2) \\ &\quad + \sigma(\mathbf{Y}_{\mathrm{spe}} - \mathbf{X}_{\mathrm{spe}} - \mathbf{S}_{\mathrm{spe}} + \mathbf{\Phi}) \end{aligned} \tag{13}$$

where $\mathbf{\Phi}$, $\mathbf{B}_1$, and $\mathbf{B}_2$ are Lagrange multipliers, a and τ are two positive constants designed to balance the contribution of each term. μ and σ are non-negative tradeoff parameters. $\mathbf{Y}_{\mathrm{spe}}$ and $\mathbf{X}_{\mathrm{spe}}$ denote the hyperspectral matrix and the background matrix expanded along the spectral dimension, respectively. The augmented Lagrangian function (13) is solvable by fixing the other variables to update only one variable alternatively. In particular, in the *k*th iteration, the optimization could be achieved in the below ten steps.

1) Update $\mathbf{E}$: The optimization problem for $\mathbf{E}$ can be written by

$$\arg\min_{\mathbf{E}} \frac{1}{2}\|\mathbf{E}\|_1 + \alpha\|\mathbf{E} - D_x\mathrm{X}\mathrm{D}_d - \mathbf{A}_1\|_F^2 \tag{14}$$

where the symbolic operator sgn(x) is the definition of

$$\mathrm{sgn}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0 \end{cases}. \tag{15}$$

The solutions are derived from (16) via the application of operator $S_\varepsilon[x] = sgn(x) * max(|x| - \varepsilon, 0)$

$$\mathbf{E} = S_{\alpha/2}[D_x \mathrm{X} \mathrm{D}_d + \mathbf{A}_1]. \tag{16}$$

2) Update $\mathbf{F}$: Similar to $\mathbf{E}$, the updated solution of $\mathbf{F}$can be obtained by the derivative calculation as

$$\mathbf{F} = S_{\alpha/2}[D_y \mathrm{X} \mathrm{D}_d + \mathbf{A}_2] \tag{17}$$

3) Update $\mathbf{G}$: The optimization problem for $\mathbf{G}$ is the same as the constrained minimization problem as

$$\begin{aligned} &\min_{\mathbf{G}} \|\mathbf{G} - \mathrm{DX} - \mathbf{A}_3\|_F^2 \\ &\text{s.t. } \|\xi \mathbf{G}\|_{1,0}^{\zeta} \leq \eta. \end{aligned} \tag{18}$$

The parameter α is neglected since it has no effect on the optimization. First, $\mathbf{G}'$ is initialized to $\mathbf{G}' = \xi(DX + \mathbf{A}_3)$, and at the kth iteration, $\mathbf{G}'^{(k)}$ is in declining order. $\mathbf{G}'_{\theta_{\gamma+1}} = 0, \ldots, \mathbf{G}'_{\theta_n}$ are shaped like zero vectors, preserving sharper image edges. $\mathbf{I}$ is defined as the unit matrix.

The optimization problem for $\mathbf{G}$ to be solved by

$$\mathbf{G} = \mathbf{G}' + (\mathbf{I} - \mathbf{B})(\mathrm{DX} + \mathbf{A}_3). \tag{19}$$

4) Update $\mathbf{A}_3$: The optimization problem for $\mathbf{A}_3$ can be solved by

$$\begin{aligned} &\mathbf{A}_3^{k+1} = \arg\min_{\mathbf{A}_3} \|\mathbf{A}_3\|_* \\ &+\mu(\left\|\mathbf{X}^k - \mathbf{A}_3^k + \mathbf{B}_1\right\|_F^2 + \left\|\mathbf{A}_4^k - \mathbf{A}_3^k + \mathbf{B}_2\right\|_F^2) \\ &= arg\min_{A_3} \alpha \left\|\mathbf{A}_3^k\right\|_* \\ &+\mu \left\|\mathbf{A}_3^k - \frac{\mathbf{X}^k + \mathbf{A}_3^k + \mathbf{B}_1^k + \mathbf{B}_2^k}{2}\right\|_F^2. \end{aligned} \tag{20}$$

5) Update $\mathbf{A}_4$: The optimization problem for $\mathbf{A}_4$ can be solved by

$$\mathbf{A}_4^{k+1} = \arg\min_{\mathbf{A}_4} -\mathrm{Tr}(\mathbf{P}^k \mathbf{A}_4^k \mathbf{Q}^{kT}) + \mu \left\|\mathbf{A}_4^k - \mathbf{A}_3^{k+1} + \mathbf{B}_2^k\right\|. \tag{21}$$

The updated solution of $\mathbf{A}_4^{k+1}$ can be obtained by the derivative calculation as

$$\mathbf{A}_4^{k+1} = \mathbf{A}_3^{k+1} - \mathbf{B}_2^k + \frac{\mathbf{P}^k \mathbf{Q}^{kT}}{\mu}. \tag{22}$$

6) Update $\mathbf{P}$ and $\mathbf{Q}$: $\mathbf{P}$ and $\mathbf{Q}$ would be obtained utilizing singular values decomposition of

$$[\mathbf{P}^{\mathrm{k}+1}, \Sigma, \mathbf{Q}^{(k+1)T}] = \mathrm{SVDs}(\mathbf{A}_4^{k+1}, r) \tag{23}$$

where Σ denotes the singular values matrix.SVDs represents the singular values decomposition function.

7) Update $\mathcal{X}$: $\mathcal{X}_{k+1}$ is updated for solutions. The subproblem associated with $\mathcal{X}$ is reformulated as its vector by

$$\begin{aligned} \mathbf{x} &= \arg\min_{\mathbf{x}} F(\mathbf{x}) \\ &= \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{x} - \mathbf{s}\|_F^2 + \alpha \|\mathbf{e} - \Gamma_x \mathbf{x} - \mathbf{a}_1\|_F^2 \\ &\quad + \alpha \|\mathbf{f} - \Gamma_y \mathbf{y} - \mathbf{a}_2\|_F^2 + \alpha \|\mathbf{g} - D\mathbf{x} - \mathbf{a}_3\|_F^2 \end{aligned} \tag{24}$$

Where $\mathbf{s}, \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ are the vector forms corresponding to their matrices, respectively. $\Gamma_x = D_x \otimes D_d, \Gamma_y = D_y \otimes D_d$, and the Kronecker product is denoted by $\otimes$. Let the partial derivative $F(x)$ be zero, and then the linear equation is obtained by

$$\begin{aligned} &(1 + \alpha D^* D + \alpha \Gamma_x^* \Gamma_x + \alpha \Gamma_y^* \Gamma_y)\mathbf{x} \\ &= \mathbf{y} - \mathbf{s} + \alpha(\mathbf{p} - \mathbf{a}_1 + \mathbf{q} - \mathbf{a}_2 + \mathbf{g} - \mathbf{a}_3 + \mathbf{a}_4 - \mathbf{b}_1). \end{aligned} \tag{25}$$

Equation (25) can be solved by the least square method [52].

8) Update $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{X}_3$

$$\begin{cases} X_1^{k+1} = L_1(X^k) \\ X_2^{k+1} = L_2(X^k) \\ X_3^{k+1} = L_3(X^k) \end{cases} \tag{26}$$

where $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{X}_3$ are the expansions of the tensor $\mathcal{X}$ along the spatial and spectral dimensions, respectively.

9) Update S_{spe} : The optimization problem for S_{spe} can be solved by

$$S_{\mathrm{spe}}^{k+1} = \arg\min_{S_{\mathrm{spe}}} \tau \left\|S_{\mathrm{spe}}^k\right\|_{2,1} + \sigma \left\|Y_{\mathrm{spe}} - X_{\mathrm{spe}} - S_{\mathrm{spe}} + \mathbf{\Phi}\right\|. \tag{27}$$

10) Update Lagrange Multiplier $\mathbf{\Phi}$, $\mathbf{A}_i$, and $\mathbf{B}_i$ by

$$\begin{aligned} \mathbf{\Phi}^{k+1} &= \mathbf{\Phi}^k + (Y_{\mathrm{spe}} - X_{\mathrm{spe}}^{k+1} - S_{\mathrm{spe}}^{k+1}) \\ \mathbf{A}_1^{k+1} &= \mathbf{A}_1^k + (D_x \mathrm{X} \mathrm{D}_d - \mathbf{E}) \\ \mathbf{A}_2^{k+1} &= \mathbf{A}_2^k + (D_y \mathrm{X} \mathrm{D}_d - \mathbf{F}) \\ \mathbf{B}_1^{k+1} &= \mathbf{B}_1^k + (\mathbf{X}_1^{k+1} - \mathbf{A}_4^{k+1}) \\ \mathbf{B}_2^{k+1} &= \mathbf{B}_2^k + (\mathbf{A}_5^{k+1} - \mathbf{A}_4^{k+1}). \end{aligned} \tag{28}$$

Eventually, the background tensor $\mathcal{X}^*$ and anomaly tensor $\mathcal{S}^*$ optimizations are accomplished when the iteration reaches convergence. The anomaly detection map is then obtained.

IV. Experiments and Analysis

In this section, our presented MSSR method is compared to a series of advanced and widely employed anomaly detection algorithms, including auto-AD, PTA, LRX, CRD, LRASR, LSMAD, and GTVLRR. All the experiments are performed with a personal computer with Intel Core i7-6700H 2.40 GHz and 16 GB RAM.

To quantitatively assess the detection results, receiver operating characteristic (ROC) is applied to represent the tradeoff between true positives and false positives. Additionally, the area under the ROC curve (AUC) value represents the region encompassed by the ROC curve. As the AUC value increases, the result of the detector gets better. All experiments are conducted ten times and the results are obtained by taking the average.

A. HSI Dataset Descriptions

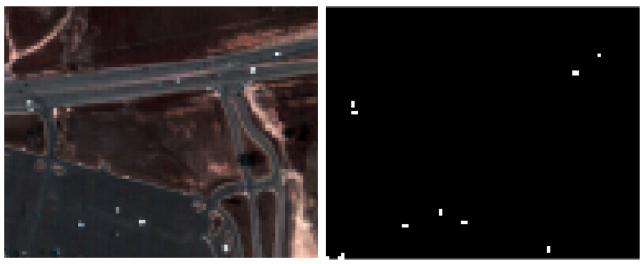
The first HSI dataset is acquired from the hyperspectral digital image acquisition experiment (HYDICE) sensor. There are 210 bands in this image with a spectral resolution of 10 nm and a spatial resolution of 1 m. After removing the absorption regions and low SNR bands, 174 bands are retained. It covers an urban area, including a vegetated area, a built-up area, and several roads. A subimage of size 80×100 is employed in this experiment. In total, 21 anomalous pixels are about vehicles of different sizes. The false color image of the whole scene and the ground truth map are shown in Fig. 2.

Algorithm 1: MSSR for HSI Anomaly Detection.

Input: HSI tensor Y, parameters: r, α, τ, μ, σ, η, λ and maxIter.

Initialization: $\mathcal{X} = 0$, $S = 0$, $\mathbf{E} = \mathbf{F} = \mathbf{G} = 0$, $\mathbf{A}_i = 0$, $\mathbf{B}_i = 0$, $X_i = \mathrm{U}_i(X)$, $\varepsilon = 10^{-6}$, $k = 1$,
while not converged, do
Update $\mathbf{E}$ with Eq. (16).
Update $\mathbf{F}$ with Eq. (17).
Update $\mathbf{G}$ with Eq. (18).
Update $\mathbf{A}_3$ with Eq. (20).
Update $\mathbf{A}_4$ with Eq. (21).
Update $\mathbf{P}$ and $\mathbf{Q}$ with Eq. (23).
Update $\mathcal{X}$ with Eq. (25).
Update $\mathcal{X}_1$, $\mathcal{X}_2$, and $\mathcal{X}_3$ with Eq. (26).
Update S_{spe} with Eq. (27).
Update $\mathbf{\Phi}$, $\mathbf{A}_i$ and $\mathbf{B}_i$ with Eq. (28).
  if $\|Y_{\mathrm{spe}} - X_{\mathrm{spe}}^{k+1} - S_{\mathrm{spe}}^{k+1}\|_F^2 \leq \varepsilon$ or $k > \mathrm{maxIter}$ **then**
  end if
end while
$S^* = \mathrm{U}_{\mathrm{spe}}(S)$, and final map $\mathcal{T}$ is derived from Eq. (29).

Output: an optimal solution $(\mathsf{X}^*, \mathsf{S}^*)$, anomaly map T.



(a) (b)

Fig. 2. HYDICE HSI: (a) false color image; (b) ground truth.

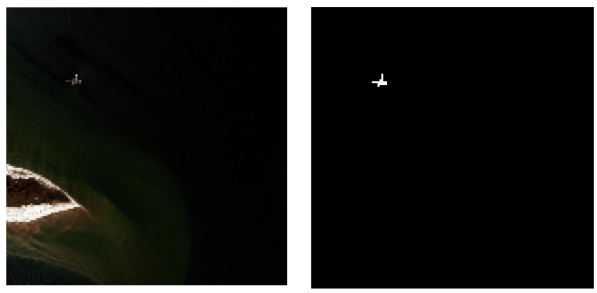
The second HSI dataset is collected from San Diego. It is captured by the AVIRIS sensor over the San Diego airport region in California, USA. Its spatial resolution of roughly 3.5 m, with 224 spectral bands spanning wavelengths from 0.37 to 2.51 μm. In total, 189 bands are maintained by stripping out the bands corresponding to the absorption and low SNR sections. A subarea of 100 $\times$ 100 pixels is experimentally chosen, and the background type consisted primarily of tarmac, roads, roofs, and shadows. In the following experiments, three aircraft are the detected anomalous targets. The false color image of the whole scene and the ground truth map are shown in Fig. 3.

The third hyperspectral dataset are Airport-Beach-Urban (ABU) scenes, consisting of four airports, four beaches, and five urban scenes. The HSI with 100 $\times$ 100 or 150 $\times$ 150 pixels is mostly collected by the AVIRIS sensor, but beach-4 images are captured by the ROSIS-03 sensor. The datasets applied in this article are Beach-1 and Urban-1 covering the wavelength range of 0.43 μm to 0.86 μm. In the following experiments, a 100 $\times$ 100 region is chosen. The false color image and the ground truth map of two datasets are presented in Figs. 4 and 5. Some parameters of these HSI are presented in Table I.

The fourth HSI dataset is MUUFL Gulfport, which is acquired from an airplane platform over the University of Southern Mississippi Gulf Park campus by the CASI-1500 sensor. The

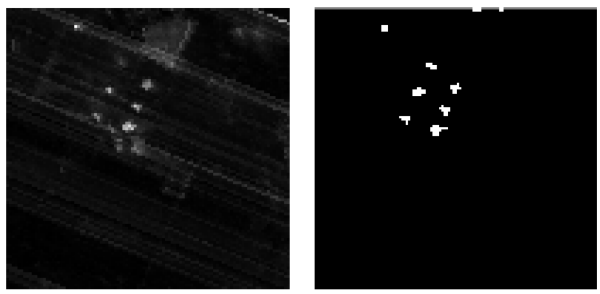
(a) (b)

Fig. 3. San Diego HSI: (a) false color image; (b) ground truth.



(a) (b)

Fig. 4. Beach-1 HSI: (a) false color image; (b) ground truth.

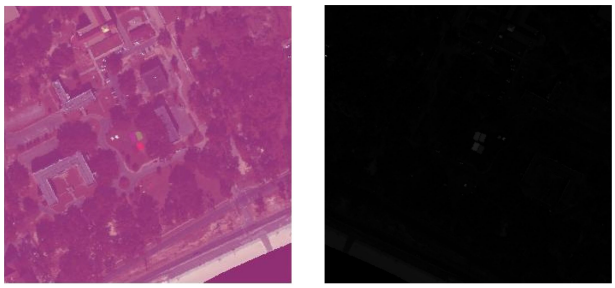


(a) (b)

Fig. 5. Urban-1 HSI: (a) false color image; (b) ground truth.

TABLE I
SELECTION OF PARAMETERS FOR EACH DATASET

Dataset	r	α	τ	μ	σ
San Diego	1	1050	0.0002	0.0002	0.50
Beach-1	2	9.8	0.0001	0.02	0.57
Urban-1	1	32.2	0.0001	0.005	0.59
HYDICE	2	160	0.0001	0.55	0.09
MUUFL	1	12	0.002	0.5	0.0002



(a) (b)

Fig. 6. MUUFL HSI: (a) false color image; (b) ground truth.

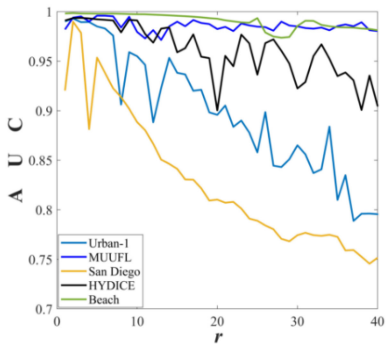


Fig. 7. Low rank of truncation r in different datasets.

wavelength range is 375 to 1050 nm, and the experimentally selected spatial size is 325 $\times$ 220 with 64 bands. There are 269 anomalous pixels available for detection at the MUUFL Gulfport. The false color image of the whole scene and the ground truth map are shown in Fig. 6.

B. Parameter Discussion

In our proposed MSSR method, there are a number of main parameters that affect the final detection results, including the low-rank of truncation r and some parameters, i.e., α , τ, μ, and σ. Aiming to assess the influence that each parameter has on the detection performance, simulations are performed by utilizing several HSI datasets and the optimal values of these parameters are discussed in our proposed method.

First, the effect of low-rank of truncation r on detection performance is investigated with other parameters fixed. The best values are sought in the range 0–40. For each of different datasets, Fig. 7 illustrates the evolution of AUC values with r. In the San Diego and Beach-1 datasets, the AUC tends to decrease as r increases. However, there is essentially little change in the HYDICE , MUUFL, and the Urban-1 HSI. It can be inferred from the figure that the optimal r is always less than 10 for most HSI. This suggests that there are not many singular values that minimize the smallest min(x, y) -r singular value, so that the first r nonzero singular values obtained match the rank of the matrices, i.e., the rank is small. In the San Diego, MUUFL, and Urban-1 datasets, the best AUC is derived when $r = 1$. In the HYDICE and the Beach-1 HSI, AUC reaches its maximum value when $r = 2$. However, in practice, it is still necessary to determine whether it is less than 10 for an unknown dataset.

All these parameters are adjusted in the range $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4\}$. Each parameter is tuned ON the basis of fixing the others, and finally the whole process of tuning is iterated until convergence. Fig. 8 shows how the AUC of the three datasets changes as each parameter varies. It is not difficult to observe that the performance tends to be smooth for α and τ, which is insensitive on the two datasets of ABU, MUUFL, and HYDICE. However, α and τ are not stable in the San Diego HSI. The opposite trend of α and τ indicates that for the background tensor, the detection performance improves with increasing the proportion of segmented smoothing constraint. Nevertheless, it gradually decreases as the weight of the low-rank constraint increases. Compared to α and τ, μ and σ are more sensitive, so the modulations for them need to vary depending on the dataset so as to achieve the best detection performance. Thus, we choose α= 9.8, τ= 0.0001 and α= 32.2, τ= 0.0001 for Beach-1 and Urban-1 datasets, respectively. For the San Diego HSI, α is chosen 1050 and τ is chosen 0.0002. As for the HYDICE, α is chosen 160 and τ is chosen 0.0001. For the MUUFL, α is chosen 12 and τ is chosen 0.002. We set μ= 0.02, σ= 0.57 and μ= 0.005, σ= 0.59 for the Beach-1 and the Urban-1 dataset, respectively. Besides, we choose μ= 0.0002, σ= 0.5 for San Diego HSI and μ= 0.55, σ= 0.09 for HYDICE. As for MUUFL, σ= 0.0002 and μ= 0.5. The specific parameters are selected, as shown in Table I.

TABLE II
AUC VALUES OF ABLATION EXPERIMENTS FOR EACH DATA SET

Dataset	Sparse and Low rank	Sparse and Smoothness	Low rank and Smoothness
San Diego	0.4284	0.3316	0.2345
Beach-1	0.4118	0.2971	0.2053
Urban-1	0.4029	0.3181	0.2098
HYDICE	0.4517	0.3628	0.2476
MUUFL	0.4285	0.3274	0.2269

C. Ablation Experiment

In this article, each of the three priors is constrained by regularization and incorporated into a model framework to better detect anomalies. To verify the impact of the three regularization terms on the model, the ablation experiments have conducted.

The experiments have conducted on different datasets with sparse, low rank, and smooth prior removed, respectively. The results of the experiments are illustrated by the following Table II. First, the sparse term and low-rank term are retained and the smoothing term is removed, thus verifying the role of l_0-l_1HTV regularization in the overall model. The experimental results show that the value of AUC without the smoothing term is the lowest. After that the smoothing term and the low-rank term are preserved as well as the sparse term is eliminated, in order to verify the role of $l_{2,1}$-norm regularization in the whole model.

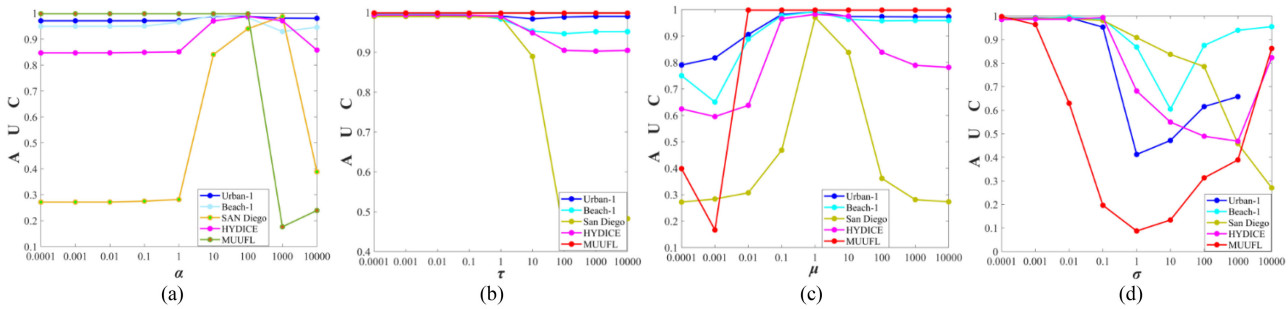


Fig. 8. Tuning parameters in different datasets. (a) α. (b) τ. (c) μ. (d) σ.

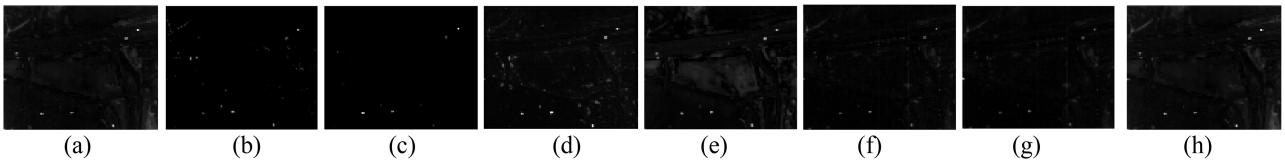


Fig. 9. Detection maps of various anomaly detectors for HYDICE dataset. (a) Auto-AD (b) PTA (c) LRX (d) CRD (e) LRASR (f) LSMAD (g) GTVLRR (h) MSRR.

Finally, with the aim of verifying the role of TNN regularization in the overall model, the smoothing term and the sparse term are maintained, and the low-rank term is deleted.

The abovementioned experimental analysis indicates that all three constraint terms are beneficial to the improvement of anomaly detection accuracy, and none of them is unavailable.

D. Detection Performance

The proposed MSSR method is compared with seven advanced detectors to evaluate the detection performance, including auto-AD, PTA, LRX, GRX, CRD, LSMAD, LRASR, and GTVLRR. The sizes of dual-windows in LRX (w_{out}, w_{in}) is set as (13,11) for San Diego HSI, (5,3) for two ABU datasets (beach-1 and urban-1). For CRD, the sizes of dual-windows are set as (7,5) while the parameter λ is set as 1 for all datasets. Besides, LSMAD have no specific parameters that significantly affect the performances. In LRASR, there are three significant parameters, two tradeoff parameters λ, β, and the number of clusters K. [λ, β]$\in$[0.01,1] in five datasets, and K is set as 6 as the same. In regards to GTVLRR algorithm, the clustering count K is 15 with 20 pixels retained in each category. The important parameters in GTVLRR are: λ= 0.1, β= 0.01, and γ= 0.5. In PTA, there are four significant parameters r, α, β, and μ. $r\in$[1], [20] and [α, β, μ]$\in$[0.0001,1] for all datasets.

As can be seen from Fig. 9, for the HYDICE dataset, the visual effect of our proposed method is significantly better than that of CRD and LRASR. These methods have a high false alarm rate compared to the other methods and do not accurately extract anomalous targets. As shown in the ROC curve in Fig. 14, MSSR has a low false alarm rate based on a high detection rate. From the comparison of the AUC methods, it can be seen that our proposed method has the largest AUC value and achieves a fairly good detection accuracy of 99.36%. In summary, our proposed method is an efficient detection method.

With regard to San Diego HSI, Fig. 10 shows the performance of the various anomaly detectors. Apparently, the detection result of our presented method significantly outperforms the compared LRX, LRASR, and GTVLRR in visual effect. None of these four methods can fully extract the anomalies from the background. Additionally, numerous false alarms appear in the LRASR results. CRD and LSMAD have better visual effect, but still inferior to the proposed method. The size of the aircraft can affect the detection performance. CRD is similar to the presented MSSR method in terms of background suppression. As far as the ROC curves in Fig. 15 are concerned, the proposed method remains outstanding. The AUC values for each method are shown in Table III and our presented method has the remarkable AUC value, which achieves a pretty good detection accuracy with 98.98%. Although the AUC value of the MSSR method is slightly inferior to the PTA method, they still have advantages compared to other methods. Fig. 15 illustrates that the ROC curves of CRD, LRASR, LSMAD, and GTVLRR are approximately similar, whereas the performance of MSSR appears to be more significant. Combining the visual observations, ROC curves and AUC values, MSSR proves to be a valid method.

For the Urban-1 HSI, MSSR also has the best visual effect than other compared methods shown in Fig. 11. The visual effects of the other methods are approximately the same, but MSSR has the highest AUC value of 99.30%. From the ROC curves in Fig. 16, the position of detection of MSSR is smaller than that of CRD at the beginning, but the presented MSSR method curve is the highest as the false alarm rate is greater than 3%, which means that it achieves the greatest detection result.

For the Beach-1 HSI, CRD, and RX have a closer visual effect to the proposed method with regard to visual effects, but MSSR is more distinct and visible shown in Fig. 12. It is noticeable that the segmentation between background and anomalies is successful from the detection results of MSSR. As can be seen by the ROC curves in Fig. 17, MSSR has the best performance compared to

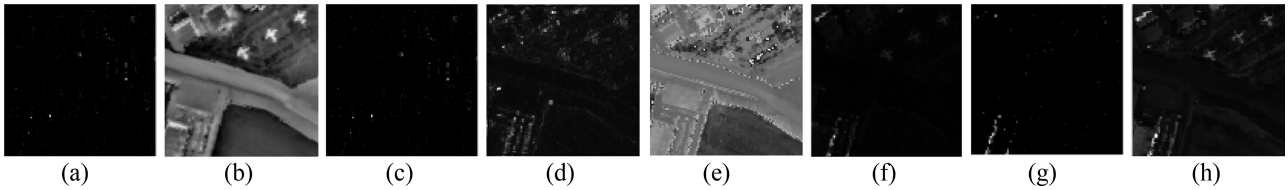


Fig. 10. Detection maps of various anomaly detectors for San Diego dataset. (a) Auto-AD. (b) PTA. (c) LRX. (d) CRD. (e) LRASR. (f) LSMAD. (g) GTVLRR. (h) MSRR.

TABLE III
AUC VALUES FOR THREE REAL DATASETS COMPARED BY DIFFERENT METHODS

Datasets	Methods							
	PTA	LRX	CRD	LRASR	LSMAD	GTVLRR	Auto-AD	MSSR
San Diego	0.99205	0.79437	0.95053	0.91233	0.97934	0.97710	0.97842	**0.98980**
Beach	0.98321	0.94657	0.9831	0.90824	0.97708	0.98460	0.99279	**0.99588**
Urban	0.98065	0.98687	0.99268	0.94961	0.98232	0.95081	0.98755	**0.99302**
HYDICE	0.98677	0.99132	0.94822	0.94317	0.98995	0.9788	0.99330	**0.99364**
MUUFL	0.98791	0.95813	0.92375	0.98124	0.62583	0.95356	0.99592	**0.99845**

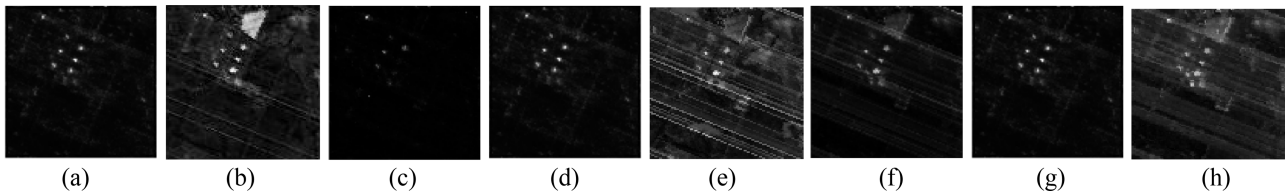


Fig. 11. Detection maps of various anomaly detectors for the Urban-1 dataset. (a) Auto-AD. (b) PTA. (c) LRX. (d) CRD. (e) LRASR. (f) LSMAD. (g) GTVLRR. (h) MSRR.

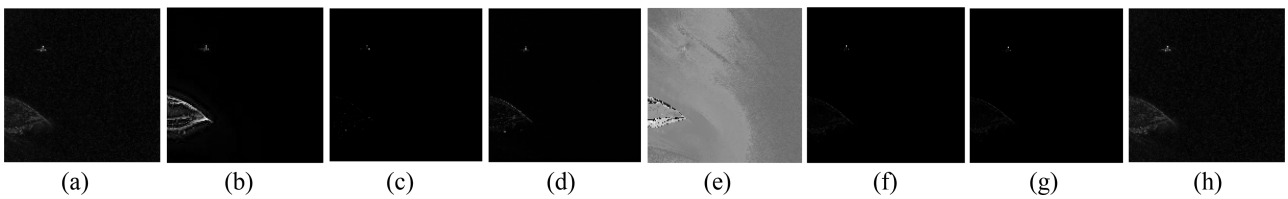


Fig. 12. Detection maps of various anomaly detectors for Beach-1 dataset. (a) Auto-AD. (b) PTA. (c) LRX. (d) CRD. (e) LRASR. (f) LSMAD. (g) GTVLRR. (h) MSRR.

the other detectors. Moreover, the AUC value of MSSR is still the highest among all methods with 99.59%.

With regard to MUUFL Gulfport HSI, Fig. 13 shows the performance of the various anomaly detectors. Obviously, the detection results of our presented method are remarkably better than the compared LRX, LRASR, and GTVLRR methods in terms of visual effects. It is observed from the ROC curves in Fig. 18 that MSSR has the greatest performance compared to the other detectors. In addition, the AUC value of MSSR remained the highest among all methods with 99.85%. In conclusion, the performance of MSSR is remarkable.

As shown in Table IV, although the execution time of the proposed method is less than LRX, GTVLRR, and auto-AD for each dataset, it is still computationally expensive, especially for the MUUFL and ABU datasets. The reason is due to the fact that the detector extracts more features, such as edge information, which makes the computation time longer.

To sum up, our presented method has the greatest performance on these datasets. The main reasons for the MSSR compared to these comparative algorithms can be categorized by the following points: 1) Our proposed method does not require any assumptions about the background, which breaks the main limitations of traditional statistical-based RX methods. 2) Abnormal pollution could trigger false alarms and leakages during detection of LRX, which affects detection performance.

Moreover, CRD, auto-AD, and PTA can effectively detect the center of anomalous targets, but tends to ignore the edges. Our proposed method exploits l_0-l_1HTV to efficiently extract edge

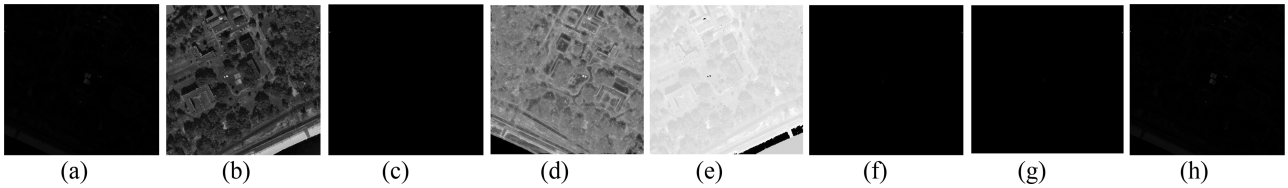


Fig. 13. Detection maps of various anomaly detectors for MUUFL dataset. (a) Auto-AD. (b) PTA. (c) LRX. (d) CRD. (e) LRASR. (f) LSMAD. (g) GTVLRR. (h) MSRR.

TABLE IV
EXECUTION TIME OF VARIOUS ANOMALY DETECTORS USING DIFFERENT DATASETS (SECONDS)

Datasets	Methods							
	PTA	LRX	CRD	LRASR	LSMAD	GTVLRR	Auto-AD	MSSR
San Diego	34.2417	49.4641	19.5122	21.3831	9.0235	98.6137	112.7081	45.7244
Beach	91.2623	89.1669	32. 7432	68.5516	20.3538	149.7873	172.0588	87.5237
Urban	36.4035	55.7084	21.3380	29.0071	10.0532	80.7802	92.5041	52.3602
HYDICE	50.3932	27.3284	15.0128	24.7859	6.5306	60.8106	67.81	27.6534
MUUFL	157.8684	65.1373	28.9275	406.47065	44.59534	1110.2036	1427.0541	59.7731

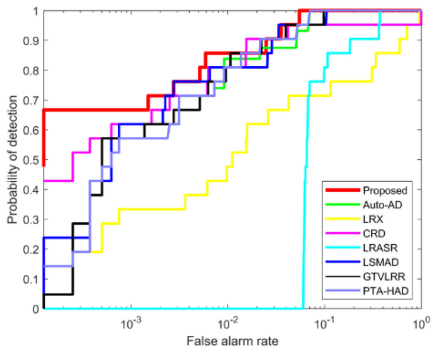


Fig. 14. ROC curves of the HYDICE data by different methods.

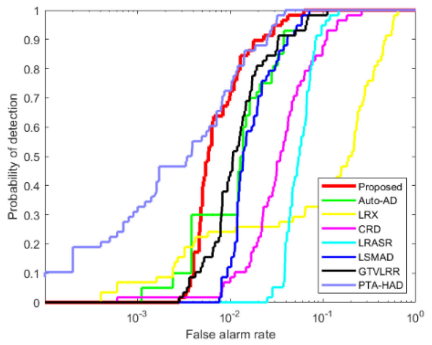


Fig. 15. ROC curves of the San Diego data by different methods.

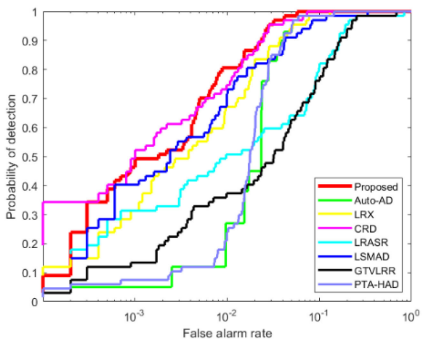


Fig. 16. ROC curves of the Urban-1 data by different methods.

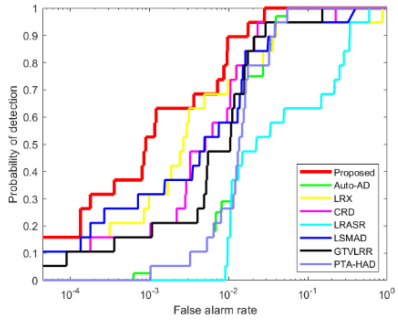


Fig. 17. ROC curves of the Beach-1 data by different methods.

information, making detection more accurate. 3) For LSMAD, it is based on LRaSMD, which can cause some false alarms due to the sparse background and the presence of decomposition errors. Our proposed method employs $l_{2,1}$-norm regularization to represent the sparse prior for the anomaly tensor and accurately use the local spectral structure of each image element to improve detection performance. 4) However, LRASR and GTVLRR based on LRR fail to adequately take into account the structural features of the data, so the detection performance is still inadequate compared to our proposed method. Compared to LRASR and GTVLR, our proposed method fully considers the structural features of the data by applying various prior, so the detection performance is superior to other detectors.

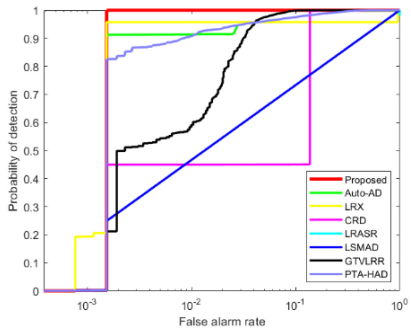


Fig. 18. ROC curves of the MUUFL data by different methods.

V. Conclusion

In this article, a novel hyperspectral anomaly detection method named MSSR is constructed. It cleverly combines the prior properties (low rank, sparsity, and segmentation smoothing) with HSI tensor decomposition. Different regularizations are adopted for different dimensions of the tensor to embed these priors in order to fully extract the spectral and spatial information of the hyperspectral data and, thus, improve the detection accuracy. In particular, the low-rank property along the background spectral dimension is represented by TNNR, the sparse prior anomaly components are expressed by $l_{2,1}$-norm regularization, and the segmented smoothing in the background spatial dimension is denoted by l_0-l_1 hybrid TV regularization to better extract the edge information. In addition, the tensor decomposition representation efficiently extracts global structural features to better separate anomalies from the background. All regularization constraints are fused into a convex optimization function and optimized iteratively utilizing the ADMM algorithm. Ultimately, when the iterations have converged, the detection map is obtained. The experimental results confirm that our proposed MSSR is capable of extracting more discriminative spectral and spatial features to further improve the detection performance. In addition, MSSR achieves the highest detection accuracy and significantly outperforms other state-of-the-art anomaly detection methods. Nevertheless, there are still some areas where further improvements can be made. In future work, we will focus on how to optimize the algorithmic framework in order to extract more discriminative spectral-spatial features in different application scenarios and further improve the detection performance.

References

[1] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021.

[2] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 29–43, Jan. 2002.

[3] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 5–27, Jul. 2010.

[4] D. Hong *et al.*, "More diverse means better: Multimodal deep learning meets remote sensing imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4340–4354, May 2021.

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

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

[7] D. Hong *et al.*, "Interpretable hyperspectral artificial intelligence: When nonconvex modeling meets hyperspectral remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 9, no. 2, pp. 52–87, Jun. 2021.

[8] A. Ertürk, M.-D. Iordache, and A. Plaza, "Sparse unmixing-based change detection for multitemporal hyperspectral images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Ses.,* vol. 9, no. 2, pp. 708–719, Feb. 2016.

[9] R. Zhao, B. Du, and L. Zhang, "A robust nonlinear hyperspectral anomaly detection approach," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1227–1234, Apr. 2014.

[10] S. Khazai, A. Safari, B. Mojaradi, and S. Homayouni, "An approach for subpixel anomaly detection in hyperspectral images," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 769–778, Apr. 2013.

[11] 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.

[12] W. M. Liu and C. I. Chang, "Multiple-Window anomaly detection for hyperspectral imagery," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 6, no. 2, pp. 644–658, Apr. 2013.

[13] S. Matteoli, M. Diani, and G. Corsini, "A kurtosis-based test to efficiently detect targets placed in close proximity by means of local covariance-based hyperspectral anomaly detectors," in *Proc. 3rd Workshop Hyperspectral Image Signal Process., Evol. Remote Sens.*, Jan. 2011, p. 4.

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

[15] J. Chen and I. Reed, "A detection algorithm for optical targets in clutter," *IEEE, Trans. Aerosp. Electron. Syst.*, vol. AES-23, no. 1, pp. 46–59, Jan. 1987.

[16] F. Li, X. W. Zhang, L. Zhang, D. Jiang, and Y. Zhang, "Exploiting structured sparsity for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 4050–4064, Jul. 2018.

[17] 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.

[18] H. Goldberg, H. Kwon, and N. M. Nasrabadi, "Kernel eigenspace separation transform for subspace anomaly detection in hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 581–585, Oct. 2007.

[19] A. Schaum, "Hyperspectral anomaly detection: Beyond rX," in *Proc. Algorithms Technol. Multispectral, Hyperspectral, Ultraspectral Imagery XIII*, Orlando, FL, USA, vol. 6565, Apr. 2007, pp. 1–13.

[20] 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. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2351–2366, Jun. 2014.

[21] 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.

[22] H. Su, Z. Wu, Q. Du, and P. Du, "Hyperspectral anomaly detection using collaborative representation with outlier removal," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 11, no. 12, pp. 5029–5038, Dec. 2018.

[23] C. Wu, B. Du, and L. Zhang, "Hyperspectral anomalous change detection based on joint sparse representation," *ISPRS J. Photogrammetry Remote Sens.*, vol. 146, pp. 137–150, Dec. 2018.

[24] S. R. Soofbaf, M. R. Sahebi, and B. Mojaradi, "A sliding window-based joint sparse representation (SWJSR) method for hyperspectral anomaly detection," *Remote Sens.*, vol. 10, no. 3, Mar. 2018, Art. no. 434.

[25] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "'Sparse representation for target detection in hyperspectral imagery," *IEEE J. Sel. Top. Signal Process.*, vol. 5, no. 3, pp. 629–640, Jun. 2011.

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

[27] M. Vafadar and H. Ghassemian, "Hyperspectral anomaly detection using outlier removal from collaborative representation," in *Proc. 3rd Int. Conf. Pattern Recognit. Image Anal.*, Shahrekord, Iran, 2017, pp. 13–19.

[28] B. Du *et al.*, "A spectral-spatial based local summation anomaly detection method for hyperspectral images," *Signal Process.*, vol. 124, pp. 115–131, Jul. 2016.

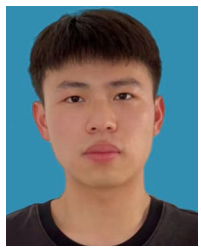
[29] J. Li, H. Zhang, L. Zhang, and L. Ma, "Hyperspectral anomaly detection by the use of background joint sparse representation," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2523–2533, Jun. 2015.
[30] D. Minh, C. Kwan, K. Koperski, and G. Marchisio, "A joint sparsity approach to tunnel activity monitoring using high resolution satellite images," in *Proc. 8th IEEE Annu. Ubiquitous Comput., Electron. Mobile Commun. Conf.*, Columbia, NY, USA, 2017, pp. 322–328.
[31] Y. Niu and B. Wang, "Hyperspectral anomaly detection based on low-rank representation and learned dictionary," *Remote Sens.*, vol. 8, no. 4, Apr. 2016, Art. no. 289.
[32] E. J. Candes *et al.*, "Robust principal component analysis?," *J. ACM*, vol. 58, no. 3, pp. 1–37, 2011.
[33] S. Y. Chen, S. M. Yang, K. Kalpakis, and C. I. Chang, "Low rank decomposition-based anomaly detection," in *Proc. Algorithms Technol. Multispectral, Hyperspectral, Ultraspectral Imagery XIX*, Baltimore, MD, USA, 2013, Art. no. 8743ON.
[34] 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.
[35] Y. Xu, Z. B. Wu, J. Li, A. Plaza, and Z. We, "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.
[36] 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.
[37] M. Wang *et al.*, "Hyperspectral image mixed noise removal based on Multidirectional low-rank modeling and spatial-spectral total variation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 1, pp. 488–507, Jan. 2021.
[38] M. Wang, Qiang Wang, and D. Hong, "l0-l1 hybrid total variation regularization and its applications on hyperspectral image mixed noise removal and compressed sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 9, pp. 7695–7710, Sep. 2021.
[39] M. Wang *et al.*, "Total variation regularized weighted tensor ring decomposition for missing data recovery in remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Apr. 2022, Art no. 6002505.
[40] T. K. Cheng and B. Wang, "Hyperspectral anomaly detection based on total variation and structured dictionary," in *Proc. IEEE Int. Symp. Geosci. Remote Sens.*, Yokohama, Japan, 2019, pp. 2252–2255.
[41] T. K. 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.
[42] T. K. Cheng and B. Wang, "Total variation and sparsity regularized decomposition model with union dictionary for hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1472–1486, Feb. 2021.
[43] L. Li, W. Li, Y. Qu, C. Zhao, R. Tao, and Q. Du, "Prior-Based tensor approximation for anomaly detection in hyperspectral imagery," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 1037–1050, Mar. 2022.
[44] M. Wang, Q. Wang, and J. Chanussot, "Tensor low-rank constraint and l0 gradient for hyperspectral image mixed noise removal," *IEEE J. Sel. Top. Signal Process.*, vol. 15, no. 3, pp. 718–733, Apr. 2021.
[45] D. T. Eason and M. Andrews, "Total variation regularization via continuation to recover compressed hyperspectral images," *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 284–293, Jan. 2015.
[46] S. J. Li, W. Wang, H. R. Qi, B. Ayhan, C. Kwan, and S. Vance, "Low-rank tensor decomposition based anomaly detection for hyperspectral imagery," in *Proc. IEEE Int. Conf. Image Process.*, Quebec City, Canada, 2015, pp. 4525–4529.
[47] Y. Xu, Z. B. 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.
[48] X. Zhang, G. Wen, and W. Dai, "A tensor decomposition-based anomaly detection algorithm for hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5801–5820, Oct. 2016.
[49] Y. Hu, D. Zhang, J. Ye, X. Li, and X. He, "Fast and accurate matrix completion via truncated nuclear norm regularization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 9, pp. 2117–2130, Sep. 2013.
[50] S. Ono, "L-0 Gradient projection," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1554–1564, Apr. 2017.
[51] L. Xu, S. Zheng, and J. Jia, "Unnatural L-0 sparse representation for natural image deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Portland, OR, USA, 2013, pp. 1107–1114.
[52] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. X. Zhu, "Joint and progressive subspace analysis (JPSA) with spatial-spectral manifold alignment for semisupervised hyperspectral dimensionality reduction," *IEEE Trans. Cybern.*, vol. 51, no. 7, pp. 3602–3615, Jul. 2021.

Yujian Wang received the B.S. degree in optoelectronic information science and engineering in 2020 from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, where he is currently working toward the M.S. degree in optical engineering with the Department of Optoelectronic Information.

His research interests include hyperspectral target detection, machine learning, and low-rank representation.

Dan Li received the B.S., M.S., and Ph.D. degrees in control science and engineering from the Harbin Institute of Technology, Harbin, China, in 2012, 2014, and 2018, respectively.

Since 2018, she has been a Lecturer with the College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing, China. Her research interests include hyperspectral image classification, signal processing, sparse sampling, and reconstruction technology.



Hanjie Wu received the B.S. degree in electronic information engineering from the Anhui University of Architecture, Anhui, China, in 2020. He is currently working toward the M.S. degree in optical engineering with the Department of Optoelectronic Information, Nanjing University of Aeronautics and Astronautics, Nanjing, China.

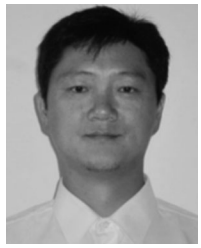
His research interests include hyperspectral image classification, machine learning, and sparse sampling.

Xiaojun Li received the Ph.D. degree in electronic science and technology from Xidian University, Xi'an, China, in 2003.

He is currently a Professor with the National Key Laboratory of Science and Technology on Space Microwave. His research interests include high-rate data transmission and satellite communication.

Fanqiang Kong received the Ph.D. degree in information and communication engineering from Xidian University, Xi'an, China, in 2008.

He is currently an Associate Professor with the College of Astronautics, Nanjing University of Aeronautics and Astronautics, Nanjing, China. His research interests include spectral image coding and image analysis, artificial intelligence, and pattern recognition.



Qiang Wang (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in control science and engineering from the Harbin Institute of Technology (HIT), Harbin, China, in 1998, 2000, and 2004, respectively.

Since 2008, he has been a Professor with the Department of Control Science and Engineering, HIT. His research interests include hyperspectral image denoising, signal/image processing, multisensor data fusion, wireless sensor networks, and intelligent detection technology.