Kernel-Based Decomposition Model With Total Variation and Sparsity Regularizations via Union Dictionary for Nonlinear Hyperspectral Anomaly Detection

Ziyu Wu and Bin Wang, Senior Member, IEEE

Abstract—Many linear approaches have been extensively proposed for the anomaly detection problem in hyperspectral images (HSIs), while nonlinear approaches have been rarely studied although most practical cases are nonlinear. Moreover, these existing nonlinear methods simply nonlinearly map each pixel into a high-dimensional space, which does not describe complex light scattering effects between endmembers. To address the above issues, this article proposes an endmember-kernel-based decomposition model with total variation (TV) and sparsity regularizations via a union dictionary for the nonlinear anomaly detection in HSIs. The proposed decomposition model utilizes endmember-kernel theory to handle nonlinear interactions between atoms in the dictionary, allowing for the effective characterization of complex light scattering effects. By using this endmember-kernel-based decomposition model, hyperspectral imagery can be decomposed into three components: anomaly, background, and noise. To separate these components effectively, the TV and sparsity regularizations are incorporated into the decomposition model to characterize the spatial properties of the background and the anomaly, respectively. Besides, we present a novel construction framework of union dictionary that combines superpixel segmentation and clustering methods sequentially to achieve more accurate dictionary representation capabilities. Finally, the anomalous level of a tested pixel is calculated by the abundances associated with the anomaly and the background, whereas the anomalies cannot. Unfortunately, there is no general rule to choose the appropriate window size so that each HSI has an appropriate window size setting, which is not feasible in practical situations. Furthermore, RX and CRD

manner. The HSIs have a high spectral resolution because the sensor acquires a spectral vector with hundreds or thousands of elements for each pixel of the HSIs [1], [2]. In fact, the different materials have completely different spectral curves, which leads to the development of several practical applications, such as unmixing and target detection [1], [2], [3], [4].

Target detection is one of the most widely used HSI processing techniques, and according to whether the target spectra are known, it is divided into unsupervised and supervised types. In the case that the target spectra are unknown, unsupervised target detection, which is also known as anomaly detection, has to be applied. It is used to identify uncommon objects with a significant difference from their surrounding background. Considering the fact that the specific spectral curves of the targets are difficult to obtain, anomaly detection is more suitable for practical applications, such as environmental monitoring and military detection [5], [6].

Many approaches have been proposed to solve the problem of anomaly detection for HSIs. In view of the convenience of calculation and physical interpretation, most of them are researched on the basis of the hyperspectral linear mixing model (LMM). The most typical anomaly detection method is the Reed–Xiaoli (RX) detector [7], which computes the anomalous level of the tested pixel by the Mahalanobis distance between the tested pixel and the mean of the background under the assumption of multivariate Gaussian distribution. Several modified anomaly detectors based on RX have emerged, including local RX (LRX) [8], weighted RX [9], segmented RX [10], regularized RX [11], and so on. However, in real HSIs, it is difficult to satisfy the assumption of a multivariate Gaussian distribution in RX. Besides, the estimates of the mean of the background are susceptible to noises and anomalies. Compared with the above RX-based detectors, the representation-based methods do not make any specific assumptions about the HSIs. The collaborative-representation-based detector (CRD) proposed in [12] is based on the concept that background pixels can be represented by their spatial neighbors, whereas the anomalies cannot. Unfortunately, there is no general rule to choose the appropriate window size so that each HSI has an appropriate window size setting, which is not feasible in practical situations.
calculate the anomalous level for each tested pixel separately, thus ignoring the global statistical properties of the entire hyperspectral imagery.

In recent years, robust principal component analysis (RPCA) [13] was utilized to decompose the HSIs into sparse and low-rank components [14]. However, in HSIs, the spectra are generally contaminated by noise, which results in the high false alarm of RPCA. In terms of this issue, the low-rank and sparse matrix decomposition (LRaSMD) [15] method models the anomalies and noise separately, but it still cannot separate the noise and weak anomalies effectively. In [16], [17], and [18], to represent the other anomalies in HSIs, the anomaly dictionaries are constructed by using some anomalous pixels that are significantly different from the surrounding background. In this way, the anomaly component can be well separated from the noise since the noise cannot be described by the latent anomaly dictionary. Considering the spatial similarity of background pixels in the neighboring regions, the total variation (TV) and sparsity regularized decomposition model (TVSADM) [18] incorporates the TV regularization to make the adjacent pixels' abundances close. Besides, sparsity regularization is incorporated to restrain the spatial sparsity of anomalies distributed in the whole HSIs [18]. Moreover, the anomaly detector with local spatial constraint and TV (LSC-TV) [19] divides the whole hyperspectral imagery into several superpixels and makes the background abundances close in each superpixel. However, the above detectors can only be used to solve the anomaly detection problem with LMM. In reality, the model errors caused by the inherent nonlinear characteristics [20] in real HSIs reduce the performance of LMM-based anomaly detectors significantly. Therefore, the nonlinear models must be considered.

Recently, a number of nonlinear mixture models (NMMs) have been presented to characterize the nonlinear interactions in HSIs. The well-known Hapke model [21] assumes that the multiple scattering is isotropic and utilizes a set of physical parameters to accurately calculate the single scattering, and thus, it can be used to solve the nonlinear problem. Due to the difficulty of obtaining accurate physical parameters in advance, this model is rarely applied in practical situations. Consequently, in order to avoid presetting physical parameters, the recently proposed NMMs extend the LMM by adding a nonlinear term to characterize the nonlinearity in spectral mixing. The Fan model (FM) [22] takes the second-scattering between two different endmembers into consideration, and then, the generalized bilinear model (GBM) [23] introduces nonlinear parameters for each interaction component to control the weight of the nonlinear mixing component. Furthermore, the polynomial postnonlinear model (PPNM) [24] considers that the second-order scattering phenomenon also exists between two same endmembers. Since the nonlinear interactions in real HSIs are mainly generated by second-order scattering, these bilinear mixture models (BMMs) are adequate in some practical applications. However, the high-order scattering effect cannot be ignored by others. In [25], a multilinear mixing (MLM) model is derived by introducing a probability parameter about multiple interactions, which extends the PPNM to an infinite number of degrees.

The kernel theory can be utilized to extend the LMM-based anomaly detectors to their nonlinear versions by mapping the hyperspectral data into a high-dimensional space. Kernel RX (KRX) [26] first executes the nonlinear mapping through kernel transform, which enhances the separability of the data, and then, the RX algorithm is applied in this high-dimensional space. Similar to the KRX, the nonlinear version of CRD, namely, kernel CRD (KCRD) [12], achieves the collaborative representation by the kernel function implicitly. Considering the fact that anomalies are far from most of the backgrounds in the high-dimensional space, a novel nonlinear anomaly detector with kernel isolation forest (KIFD) [27] is proposed. Moreover, the support vector data description (SVDD) [28] constructs a minimum hypersphere in the hyperspectral data space, and then, the pixels falling outside of this hypersphere are regarded as anomalies. The above nonlinear methods map the hyperspectral data into a high-dimensional space to extract nonlinear features. Although such methods can improve the separability of HSIs, the inherent nonlinear interactions in HSIs are ignored. Besides, there is no general rule to select the kernel function, which results in very unstable anomaly detection results. In terms of this issue, the KType [29] proposes the endmember-kernel theory to realize the nonlinear interactions between endmembers. On this basis, the KNUD [30] utilizes the endmember-kernel theory with a dual window to implement anomaly detection. However, the detection result of KNUD is affected to some extent due to the absence of global spatial properties, and the deficiency of the dual window has still not been solved.

The specific spectral information of the anomaly and background is unknown in the case of anomaly detection. Therefore, in order to distinguish the background and anomalies effectively, the existing anomaly detection algorithms generally construct a background dictionary or a union dictionary that consists of an anomaly dictionary and a background dictionary [16], [17], [18]. In order to express other anomalies in the hyperspectral imagery, the anomaly dictionary is composed of anomalies with significantly different spectra from their surroundings [17], [18], [31]. Besides, the global clustering methods consider the spatial homogeneity [32] of background pixels and group all the pixels into several clusters, such as mean shift [33], k-means [34], and k-means++ [35]. Then, the background dictionary can be constructed by selecting the representative pixels in each cluster. However, it is unrealistic to make the assumption that the cluster number is known. Moreover, the global clustering methods have a common drawback that cannot correctly distinguish the sparse background and anomalies because of their similar proportions in the whole hyperspectral imagery.

To sum up, there are several problems with the existing hyperspectral anomaly detectors.

1) The traditional kernel theories map the hyperspectral data into a high-dimensional space, which is inconsistent with the inherent nonlinear interactions of HSIs based on the NMMs.

2) The spatial information of HSIs is not well exploited in nonlinear anomaly detection algorithms, which is very important for effective model decomposition.
3) The existing dictionary construction methods have a common drawback that cannot correctly distinguish the sparse background and anomalies, reducing the accuracy of anomaly detection.

In this article, in order to solve the above problems, we propose a novel kernel-based decomposition model with TV and sparsity regularizations via a union dictionary for nonlinear anomaly detection in HSIs, named KDM-TVS. The proposed decomposition model utilizes endmember-kernel theory to handle nonlinear interactions between atoms in the dictionary so that the complex light scattering effect can be effectively described. Compared with KNUD, which also utilizes endmember-kernel theory to implement anomaly detection in the dual window, the proposed KDM-TVS decomposes the whole hyperspectral imagery into three components: anomaly, background, and noise. In order to separate these components effectively, the spatial sparsity of the anomalies is enforced by means of sparsity regularization. Besides, TV regularization is utilized to make the abundance of adjacent pixels in the background similar so that the spatial smoothness of the background is realized. Moreover, in order to enhance the accuracy of model decomposition, we present a novel construction framework of union dictionary, in which the hyperspectral imagery is first divided into several blocks using a superpixel-based segmentation method, and then apply a local clustering method within each block. Unlike other global clustering methods that cannot distinguish the sparse backgrounds and anomalies, the proposed construction framework of union dictionary extracts background dictionary atoms in each block, and thus, a complete background dictionary can be obtained. Meanwhile, the anomaly dictionary can be constituted easily by picking out some strong anomalous pixels. By means of the optimal abundances associated with the latent anomaly dictionary, precise detection results can be obtained.

The main contributions of this article can be briefly summarized as follows.

1) A novel endmember-kernel-based decomposition model for nonlinear hyperspectral anomaly detection is proposed to handle the nonlinear coactions between dictionary atoms. By fully considering the inherent nonlinear interactions of hyperspectral imagery based on the NMMs, the decomposition model is more conform with the physical process of light scattering effect and more physically interpretable compared with the traditional kernel methods that map the hyperspectral data into a high-dimensional space. Besides, the sparsity and the TV regularizations are incorporated into the decomposition model to characterize the spatial properties of anomalies and background, respectively, to further achieve the effective separation of the anomaly, background, and noise.

2) In view of the difficulty of distinguishing sparse backgrounds and anomalies in the existing methods, we present a novel construction framework of the union dictionary, which combines the superpixel segmentation and clustering methods sequentially. By means of local clustering, the interference of anomalies can be easily eliminated, and the sparse backgrounds can be effectively singled out to constitute a background dictionary.

The rest of this article is arranged as follows. Section II briefly introduces the related works. In Section III, the proposed nonlinear anomaly detection method will be described in detail. The experiments on both synthetic and real hyperspectral datasets are conducted, which demonstrate the superiority of the proposed method in Section IV. Finally, Section V draws the conclusions of this article.

II. RELATED WORKS

A. Hyperspectral Mixing Models

Each pixel in HSIs may consist of several endmembers on account of the low spatial resolution, namely, the spectra of the pixel are the interactions of several types of the endmember spectra. Denote $E = [e_1, e_2, \ldots, e_P] \in R^{L \times P}$ an endmember matrix with $L$ spectral bands and $P$ endmembers. $y \in R^{L \times 1}$ is a tested pixel; the form of LMM is shown as follows:

$$y = \sum_{i=1}^{P} \alpha_i e_i + n$$

s.t. $\alpha_i > 0$, $\sum_{i=1}^{P} \alpha_i = 1$ (1)

where $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_P]^T$ denotes the abundance vector of $y$ and $n$ represents the noise. Besides, the abundance vector $\alpha$ needs to satisfy two constraints, i.e., abundances’ sum-to-one constraint (ASC) and abundances’ nonnegativity constraint (ANC) [36].

However, the LMM is insufficient because of the multiple light scattering effects. In order to characterize the nonlinear interactions in HSIs, the FM shown in (2) considers the second-order scattering effects between two different endmembers

$$y = E\alpha + \sum_{i=1}^{P-1} \sum_{j=i+1}^{P} \alpha_i \alpha_j e_i \odot e_j + n$$ (2)

where $\odot$ represents the Hadamard product and $e_i \odot e_j = [e_{i,1} e_{j,1}, e_{i,2} e_{j,2}, \ldots, e_{i,L} e_{j,L}]^T$ models the second-order interaction between two endmembers. In order to control the weight of the nonlinear mixing component, the GBM shown in (3) introduces the nonlinear parameter $\gamma$

$$y = E\alpha + \sum_{i=1}^{P-1} \sum_{j=i+1}^{P} \gamma_{ij} \alpha_i \alpha_j e_i \odot e_j + n.$$ (3)

The FM and GBM only consider the nonlinear interactions between different endmembers. Recently, the PPNM shown in (4) considers that the second-order scattering phenomenon also exists between two same endmembers [25]

$$y = E\alpha + \bar{\varepsilon} \sum_{i=1}^{P} \sum_{j=1}^{P} \alpha_i \alpha_j e_i \odot e_j + n.$$ (4)
To sum up, the aforementioned nonlinear mixing models limit the light scattering phenomenon in two times, and a general form can be summarized as follows:

\[ y = E\alpha + \sum_{i=1}^{P} \sum_{j=1}^{P} \tau_{ij} \alpha_i \alpha_j e_i \otimes e_j + n \]

where \( \tau_{ij}, \theta_1, \) and \( \theta_2 \) denote the parameters of each nonlinear mixing models, and their specific definition is shown in Table I.

The nonlinear interactions are indeed dominated by the second-order scattering for some practical applications. However, the high-order scattering should be considered in complex scenes with significant altitude differences. In order to consider the high-order scattering effects present in HSIs, the MLM model [26] extends PPNM by introducing a probability parameter \( q \):

\[ y = (1 - q)E\alpha + (1 - q)q^{P} \sum_{i=1}^{P} \sum_{j=1}^{P} \alpha_i \alpha_j e_i \otimes e_j + \sum_{i=1}^{P} \sum_{j=1}^{P} \sum_{k=1}^{P} \alpha_i \alpha_j \alpha_k e_i \otimes e_j \otimes e_k + \ldots. \]  

Both the PPNM and MLM demonstrate that the nonlinear interactions among the endmembers can be represented by the endmember-kernel theory. By means of the alternating direction method of multipliers (ADMM) [39], the abundances associated with the union dictionary can be obtained. Then, the anomalous level of a tested pixel can be calculated by the abundance of the latent anomaly dictionary. In order to obtain a union dictionary precisely, a novel construction framework of union dictionary, which consists of the local clustering method and the global anomaly information, is presented. The graphical description of the proposed KDM-TVS is shown in Fig. 1.

### III. PROPOSED METHOD

A detailed description of the proposed KDM-TVS is given in this section. We first present a novel kernel-based decomposition model to separate the hyperspectral imagery into three components: anomaly, background, and noise. To separate these components effectively, the TV and sparsity regularizations are incorporated to characterize the spatial properties of the background and the anomaly, respectively. The nonlinear interactions among the components can be represented by the endmember-kernel theory. By means of the alternating direction method of multipliers (ADMM) [39], the abundances associated with the union dictionary can be obtained. Then, the anomalous level of a tested pixel can be calculated by the abundance of the latent anomaly dictionary. In order to obtain a union dictionary precisely, a novel construction framework of union dictionary, which consists of the local clustering method and the global anomaly information, is presented. The graphical description of the proposed KDM-TVS is shown in Fig. 1.

### A. KDM-TVS for Anomaly Detection

Let \( A = [a_1, a_2, \ldots, a_R] \in R^{L \times R} \) and \( B = [b_1, b_2, \ldots, b_P] \in R^{L \times P} \) denote the anomaly dictionary and the background dictionary, respectively. Then, the union dictionary can be expressed as \( E = [A, B] = [a_1, a_2, \ldots, a_R, b_1, b_2, \ldots, b_P] \in R^{L \times (P + R)} \), where \( P + R \) is the number of atoms in the union dictionary. The nonlinear mixing model can be formulated by the union dictionary and abundance matrix based on (7)

\[ Y = BX + AZ + \psi(E) + N \]  

where \( Z \in R^{R \times M} \) and \( X \in R^{P \times M} \) are the abundance coefficient matrix associated with the anomaly dictionary \( A \) and the background dictionary \( B \), respectively. Thus, the union abundance coefficient matrix can be expressed as \( \alpha = [Z, X] \in R^{(P + R) \times M} \).

Usually, the natural background is spatially smooth generally, that is, the abundances of the adjacent pixels are similar. Thus, the TV regularization [40], [41], which makes the adjacent pixels’ abundances close [42], [43], [44], is incorporated to characterize the background component in the anomaly detection of HSIs.

### Table I

| Model | \( \tau_{ij} \) | \( \theta_1 \) | \( \theta_2 \) |
|-------|-----------------|----------------|----------------|
| FM    | 1               | \( P - 1 \)     | \( i + 1 \)     |
| GBM   | \( \gamma_{ij} \)| \( P - 1 \)     | \( i + 1 \)     |
| PPNM  | \( \xi \)       | \( P \)         | 1              |

where \( \langle \cdot, \cdot \rangle_{\mathcal{H}} \) denotes the inner product in the RKHS \( \mathcal{H} \) and \( k_n(\cdot, \cdot) \) is the used reproducing kernel function. The nonlinear function \( \psi \) can be expressed by the mapping results of other spectral bands. Detailed proof of this can be found in [29].

The objective function of K-Hype is formulated as follows:

\[
\begin{aligned}
\arg\min_{\alpha, \psi} & \frac{1}{2} \left( \| \alpha \|^2 + \| \psi \|^2_{\mathcal{H}} + \frac{1}{\mu} \sum_{l} \text{err}_l^2 \right) \\
\text{s.t.} & \quad \text{err}_l = y_l - e_l^T \alpha - \psi(e_{l2}) \\
& \quad \alpha \geq 0, \quad 1^T \alpha = 1
\end{aligned}
\]  

which can be solved by standard duality theory [38].
Besides, the anomalies in HSIs show spatial sparsity, that is, the abundance matrix $Z$ should be column sparse. Therefore, the $l_{2,1}$-norm [45], which is defined as the sum of $l_2$-norm of each column in the abundance matrix, can be used to characterize the anomaly component.

In the meantime, we expand (10) to the whole HSIs. Besides, in order to separate the objective function, we introduced two auxiliary variables, $Z_1$ and $X_1$. In this way, the problem of the proposed KDM-TVS can be described as

$$
\text{arg min}_{\alpha, \psi} \sum_{i} \left( \frac{1}{2} \left( \| \alpha_i \|^2 + \| \psi_i \|^2 + \frac{1}{\mu} \sum_{l=1}^{L} \text{err}_{il}^2 \right) + \eta_1 \| X_1 \|_{HTV} + \eta_2 \| Z_1 \|_{2,1} \right) 
$$

s.t. $err = Y - BX - AZ - \psi((A, B))$

(11)

Besides, the anomalies in HSIs show spatial sparsity, that is, the anomalies in HSIs show spatial sparsity, that is, $\frac{1}{\mu} \sum_{l=1}^{L} \text{err}_{il}^2$.

The problem (12) can be transformed into the corresponding Lagrange function as follows:

$$
\mathcal{L}(Z, Z_1, X_1, D_1, D_2, \eta_1, \eta_2) = \sum_{i} \left( \frac{1}{2} \left( \| \alpha_i \|^2 + \| \psi_i \|^2 + \frac{1}{\mu} \sum_{l=1}^{L} \text{err}_{il}^2 \right) + \eta_1 \| X_1 \|_{HTV} + \eta_2 \| Z_1 \|_{2,1} \right) + \frac{1}{2} \left( \| Z - Z_1 - D_1 \|_F^2 + \| X_1 - X_2 \|_F^2 \right) - \sum_{i=1}^{M} \sum_{l=1}^{L} \beta_{il} \left( \text{err}_{il} - y_{il} + \psi_i^T \alpha_i + \psi_i (e_{zi}) \right) - \sum_{p=1}^{P+R} \gamma_{ip} \alpha_{ip} + \lambda_i \left( 1_{P+R}^T \alpha_i - 1 \right) 
$$

(13)

where $\beta_{il}, \gamma_{ip},$ and $\lambda_i$ are the Lagrange multipliers, and $1_{P+R}$ is a $(P+R)$-dimensional column vector with an element of 1. $D_1$ is the auxiliary variable that denotes the differences between the original variable $X$ and the corresponding auxiliary variable $X_1$. Similarly, $D_2$ represents the differences between $Z$ and $Z_1$. Each variable in (13) can be updated iteratively by means of ADMM [39]. The subproblems of the Lagrange function are presented as follows.

1) The abundance coefficient vector $\alpha_i$ associated with the tested pixel $y_i$ can be formulated as

$$
\alpha_i^{(t+1)} = \text{arg min}_{\alpha_i} \frac{1}{2} \left( \| \alpha_i \|^2 + \| \psi_i \|^2 + \frac{1}{\mu} \sum_{l=1}^{L} \text{err}_{il}^2 \right) + \frac{1}{2} \left( \| Z_i - Z_1 - D_1 \|_F^2 + \| X_i - X_2 \|_F^2 \right) - \sum_{l=1}^{L} \beta_{il} \left( \text{err}_{il} - y_{il} + \psi_i^T \alpha_i + \psi_i (e_{zi}) \right) - \sum_{p=1}^{P+R} \gamma_{ip} \alpha_{ip} + \lambda_i \left( 1_{P+R}^T \alpha_i - 1 \right) 
$$

(14)

By taking the derivative of the original variables in (14), the optimality conditions are [29]

$$
\begin{align*}
Z_i^* &= \frac{1}{2} \left( \sum_{l=1}^{L} \beta_{il}^* A_{zi} + y_{zi}^* - \lambda_i^* 1_R + Z_1^{(t)} + D_1^{(t)} \right) \\
X_i^* &= \frac{1}{2} \left( \sum_{l=1}^{L} \beta_{il}^* B_{zi} + y_{bi}^* - \lambda_i^* 1_R + X_1^{(t)} + D_2^{(t)} \right) \\
\psi_i^* &= \sum_{l=1}^{L} \beta_{il}^* k (\cdot, e_{zi}) \\
\text{err}_{il}^* &= \mu \beta_{il}^* 
\end{align*}
$$

(15)
The dual problem (16), as shown at the bottom of the page, can be obtained by substituting (15) into (14). In (16), $K$ denotes the Gram matrix defined as $K_{ij} = k(e_{zi}, e_{zj})$.

In order to handle the nonlinear interactions between endmembers, we introduce the second-order polynomial kernel and the Gaussian kernel, which correspond with the description of nonlinear interactions in PPNM and MLM, respectively.

$$k_{\text{polynomial}}(e_{zi}, e_{zj}) = \left(a \cdot e_{zi}^T e_{zj} + b\right)^2$$

$$k_{\text{gaussian}}(e_{zi}, e_{zj}) = \exp\left(-\frac{\|e_{zi} - e_{zj}\|^2}{2\sigma^2}\right)$$

where $a = a^T a$, $b = a^T b$, and $\sigma$ are the parameters of the polynomial kernel, and $a$ and $b$ are the parameters of the Gaussian kernel, respectively.

The computational complexity of Algorithm 1 mainly depends on the update of the abundance $\alpha$. In each iteration, (15) and (16) are operated for each pixel, and their computational complexities are $O((P + R)L)$ and $O((L + P + R^2)^2)$, respectively. Thus, the computational complexity of $\alpha$ is $O((L + P + R^2)^2M)$. About the auxiliary variables $X_1$ and $Z_1$, their time complexities are $O(PM)$ and $O(RM)$, respectively. Here, $M$ and $L$ are the number of pixels and spectral bands, respectively, and $P$ and $R$ denote the number of atoms in the background dictionary and the anomaly dictionary,

$$D_2^{t+1} = D_2^t - \left(X^{t+1} - X_1^{t+1}\right)$$

$$\rho^{t+1} = \min(\rho^{t+1}, \rho_{\max})$$

$$\eta_1^{t+1} = \frac{\eta_1^{t+1}}{\rho^{t+1}}$$

$$\eta_2^{t+1} = \frac{\eta_2^{t+1}}{\rho^{t+1}}.$$
respectively. Therefore, Algorithm 1 has the computational complexity of $O((L + P + R)^2M)$.

B. Construction of Union Dictionary

Let the sketch map with $8 \times 8$ pixels denote hyperspectral imagery shown in Fig. 2(a). Due to the similar proportion of sparse backgrounds and anomalies, the global clustering-based methods cannot distinguish the sparse backgrounds and anomalies effectively. Considering the fact that the spatial distributions of the sparse backgrounds and anomalies are very different, we present a novel construction framework of union dictionary. First, we utilize a superpixel segmentation method to divide the hyperspectral imagery into several superpixel blocks, as shown in Fig. 2(b), and the proportion of sparse backgrounds in a superpixel block increases significantly compared with the proportion of sparse backgrounds in the original hyperspectral imagery because of the compact spatial distribution, whereas the proportion of anomaly does not change. Then, the separation of the sparse backgrounds and anomalies can be achieved by a local clustering method in each superpixel block. Simple linear iterative clustering (SLIC) [48] is a superpixel segmentation method proposed in recent years, which integrates spatial distance and spectral distance, so as to obtain an excellent segmentation result focusing on boundary maintenance. Therefore, it is adopted to execute the superpixel segmentation for the union dictionary construction in this article. Besides, in order to avoid presetting the number of clusters in clustering methods, the density peak-based clustering (DPBC) method [49], which is only based on the similarity between data points, is utilized to execute the local clustering operation.

In order to construct a representative latent anomaly dictionary $A$, which can be used to represent other anomalies in HSIs [17], [18], [31]. Traditional global anomaly detectors are used to pick out strong anomalies that are significantly different from their surroundings. Considering the complexity and efficiency, this article utilizes the first $R$ pixels of the RX [7] detection results to construct an anomaly dictionary. Finally, superior detection results will be obtained by combining the union dictionary with the proposed method.

The proposed construction framework of union dictionary is summarized in Algorithm 2.

**Algorithm 2** Procedure for the Construction of Union Dictionary

**Input:** Data $Y \in \mathbb{R}^{L \times M}$, the number of superpixel blocks $N_s$, the number of atoms in the latent anomaly dictionary $R$.

**Background dictionary $B$:**

1) Divide $Y$ into $N_s$ superpixel blocks by the superpixel segmentation method.
2) By means of the local clustering method, the background dictionary $B$ can be constituted by using representative pixels in each superpixel block.

**Anomaly dictionary $A$:** Apply the global anomaly detector RX.

**Output:** Anomaly dictionary $A$, and background dictionary $B$.

![Fig. 2. Difference between the sparse backgrounds and anomalies. (a) Distribution map with four ground objects for an $8 \times 8$ pixel image. (b) Presegmentation map by the superpixel method.](image-url)

IV. EXPERIMENTAL RESULTS

In this section, a synthetic hyperspectral dataset is utilized to analyze the specific procedures and the performance of KDM-TVS comprehensively. Then, three real datasets are used to verify the effectiveness of our proposed KDM-TVS in the actual applications.

Since the receiver operating characteristic (ROC) curve [50] and the area under the ROC curve (AUC) [51] can intuitively evaluate the detection results of an anomaly detector. Therefore, the ROC curve and the AUC are utilized to compare the KDM-TVS with other several anomaly detection methods quantificationally. All the experiments are implemented on a computer with an Intel Core i9-11900K CPU at 3.5 GHz and 32-GB RAM, MATLAB R2018b.

A. Experiments on Synthetic Dataset

The synthetic hyperspectral dataset is produced based on the San Diego dataset, which was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS). After removing the water vapor absorption and low signal-to-noise ratio (SNR) bands (i.e., 1–6, 33–35, 94–97, 107–113, 153–166, and 221–224) from the original 224 spectral bands, 186 bands are retained in the following experiments. The original whole scene with $400 \times 400$ pixels is displayed in Fig. 3(a), and a subimage with $100 \times 100$ pixels in the red square is cropped to generate the synthetic dataset. The anomaly pixel $t$ is selected from an airplane in the original scene, whose signature is shown in Fig. 3(b). Here, the anomalous pixels are generated by the target implantation method [52]. Specifically, a synthetic anomalous pixel $z$ is generated by fractionally embedding the selected anomaly pixel $t$ into a given background pixel based on PPNM

$$z = f \cdot t + (1 - f) \cdot b + \zeta f (1 - f) \cdot (t \odot b)$$

(25)

where $f$ denotes the abundance, which belongs to [0.05, 0.1, 0.2, 0.4, and 0.5]. Each of them generates three anomaly blocks with different sizes of $1 \times 1$, $1 \times 2$, and $2 \times 2$, and the positions of these blocks are completely random, considering the true distribution of anomalies in real hyperspectral imagery. $\zeta$ denotes the tradeoff between linear and nonlinear components in PPNM, which is randomly generated in the...
Fig. 3. Synthetic dataset. (a) Pseudocolor image of the whole scene. (b) Spectral signature of the selected anomaly \( t \) from an airplane in the whole scene. (c) Pseudocolor image of the detection area with embedded anomalies. (d) Ground-truth map of the embedded anomalies.

Fig. 4. Segmentation maps of the synthetic dataset with different clustering methods. (a) Segmentation map obtained by the proposed local clustering method. (b) Segmentation map obtained by the global clustering method (cluster number = 5). (c) Segmentation map obtained by the global clustering method (cluster number = 10).

Correspondingly, the segmentation map of the global clustering method with a small cluster number is shown in Fig. 4(b). It can be seen that the sparse background at the left bottom is classified as a road. With the increase in the cluster number, the global clustering method is able to separate the sparse background, but many anomalies are also classified separately because of the similar proportion of sparse backgrounds and anomalies, which is shown in Fig. 4(c).

To sum up, the proposed construction framework of union dictionary with the local clustering method can pick up the sparse backgrounds in the background dictionary without the interference of anomalies, which overcomes the imperfection of global clustering methods.

As for the anomaly dictionary, the predetector is just utilized to pick out several high-purity anomalies. With this latent anomaly dictionary of strong anomalies, the proposed KDM-TVS is able to represent other weak anomalies in HSIs. Here, we compare the performances of RX, LRX, CRD, and KRX in the process of anomaly dictionary construction. The detection maps based on different union dictionaries are shown in Fig. 5, and the ROC curves are shown in Fig. 6. The AUC values of these detectors and the AUC values of KDM-TVS(G) based on these detectors are displayed in Table II. It can be seen that, although the AUC values of these detectors are not good due to the limitation of themselves, the detection results of the proposed KDM-TVS(G) based on these detectors are satisfactory, which indicated that many efficient anomaly detectors can be applied to pick out strong anomalies. Furthermore, the KRX has a kernel parameter \( \sigma \) that needs to be adjusted, the sizes of dual windows in LRX and CRD also need to be preset, and it can be observed that it takes less time to build an RX-based union dictionary than other detectors. Therefore, to improve the stability and efficiency,
Fig. 5. Detection maps obtained by the proposed KDM-TVS(G) with different union dictionaries.

Fig. 6. ROC curves of the proposed KDM-TVS(G) on the synthetic dataset with different union dictionaries.

of the predetection, we use the simple and efficient RX to generate the latent anomaly dictionary.

Fig. 7 shows the variation curve of AUC values with the changes of $R$. The strongest anomalies can be picked out by traditional anomaly detectors. Then, these strongest anomalies can represent other anomalies in the whole HSIs, and thus, the proposed KDM-TVS obtains a satisfactory detection performance with a few of anomaly atoms. In order to guarantee the representation capability of the anomaly dictionary by the proposed Algorithm 1, the value of $R$ should not be too small to construct a complete anomaly dictionary. However, the latent anomaly detection dictionary probably contains some sparse backgrounds when $R$ increases, resulting in the deterioration of the anomaly detection results. Therefore, the value of $R$ should take a relatively appropriate value. It can be observed from Fig. 7 that the proposed KDM-TVS can obtain a satisfactory detection result in a relatively wide range of $R$. To avoid the effect of sparse background and cover the strong anomalous pixels as many as possible in the meantime, we set $R = 7$ in the follow-up experiments.

2) Parameter Setting: We evaluate the influence of the parameters of the proposed KDM-TVS (Gaussian kernel) on the detection result. $\eta_1$ and $\eta_2$ are tradeoff parameters used to balance the regularization terms, which do not need to be adjusted because they get smaller as the iteration proceeds.

Next, we design a parametric experiment on the tradeoff parameter $\mu$, which is used to balance the error term. The adjustment range of $\mu$ belongs to [0.001, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 1], and the corresponding AUC values are listed in Table III. It reveals that the change of AUC value is not severe when the tradeoff parameter $\mu$ is set in a relatively wide range. Therefore, we empirically fix $\mu = 0.2$ in the follow-up experiments for the sake of simplicity.

3) Ablation on Regularization Terms: We investigate the contributions of the incorporated regularizations, i.e., the sparsity ($l_{2,1}$-norm) term and the TV term. The proposed

| $\mu$ | 0.001 | 0.01 | 0.05 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 1    |
|-------|-------|------|------|-----|-----|-----|-----|-----|------|
| AUC   | 0.9913 | 0.9954 | 0.9959 | 0.9972 | 0.9963 | 0.9938 | 0.9905 | 0.9724 |
KDM-TVS without only the sparsity regularization terms ($\eta_1 = 5$ and $\eta_2 = 0$) and without any regularization term ($\eta_1 = 0$ and $\eta_2 = 0$) is implemented on the synthetic dataset, respectively. The detection maps are displayed in Fig. 8(a) and (b), respectively. It can be seen that the TV term makes the detection result distinct for its superior performance at keeping the background smooth. Another observation is that there are a number of false alarms in the detection result. The reason is that the $l_2,1$-norm can restrain the number of numerically large points in the anomalous abundance map. The intact anomaly detection result with two incorporated regularizations is shown in Fig. 5.

4) Comparison of Detection Results: Here, we compare the detection performance of the proposed KDM-TVS with other anomaly detectors by utilizing the synthetic dataset. The competitors include the linear detectors (LRX [8], CRD [12], TVSDM [18], and LSC-TV [19]) and the nonlinear detectors (KRX [26], KCRD [12], KNUD [30], and KIFD [27]). LRX and KRX are two improved versions of RX, which are recognized as benchmarks among the statistical approaches. CRD and its kernel version KCRD are based on the representation in the neighborhood, which accounts for the relationship between adjacent pixels. TVSDM incorporates the TV and sparsity regularizations to the decomposition model based on LMM. LSC-TV also makes the background abundances close in each superpixel. KIFD applies an isolation forest algorithm in the high-dimensional space to detect anomalies. KNUD utilizes the endmember-kernel theory with a dual window to implement anomaly detection, which takes the nonlinear mixing between endmembers into account.

We set the involved parameters according to the values recommended in the literature. Specifically, the dual window sizes ($\omega_{in}$, $\omega_{out}$) in LRX [8], CRD [12], KCRD [12], and KNUD [30] are all set to (3, 9). The tradeoff parameter $\lambda$ in CRD and KCRD is set to $10^{-6}$ as recommended in [12]. For KNUD, the number of anomaly atoms and the proportion of the selected pixels in the window are set to 15 and 95% according to [30], respectively. Similarly, for TVSDM [18], the number of anomaly atoms and background atoms in each class are set to 20 and 20, respectively. Besides, we set $\lambda = 10^{-3}$, $\beta = 5 \times 10^{-3}$, $P = 20$, and $S = 15$ for LSC-TV as recommended in [19]. The number of trees and the subsampling size of the tree in KIFD are set to 1000 and 3% according to [27], which are fixed in the follow-up experiments.

The detection maps of the different detectors on the synthetic dataset are displayed in Fig. 9. It can be observed that the CRD and LRX cannot identify the anomalies, that is, the weak anomaly detection case is severe for these two linear detectors. Besides, a large number of false alarms are included in the detection results of KRX, KCRD, and KIFD. It should be due to the inappropriate nonlinear mapping, which ignores the inherent nonlinear interactions among endmembers. Both the KNUD and the proposed KDM-TVS consider the nonlinear interaction between dictionary atoms, while the detection results of KNUD are slightly dim. Correspondingly, the detection results of the proposed KDM-TVS are distinct, which gives the credit to the consideration of spatial relations by utilizing the TV and sparsity regularizations.

The ROC curves of each anomaly detector are displayed in Fig. 10 to evaluate the detection performance quantitatively. It is obvious that the detection rate of the proposed KDM-TVS is higher than that of other detectors at nearly all false alarms apart from the first 2%. Moreover, the proposed KDM-TVS can achieve a 100% detection rate when the false alarm rate is lower than 2%, which shows the significant improvement of the proposed KDM-TVS with other detectors. Besides, the AUC values of different detectors are shown in Table IV, which indicates the superiority of the proposed KDM-TVS quantitatively. It is worth noting that the detection result of KDM-TVS(G) is slightly worse than that of KDM-TVS(P), just because the synthetic dataset is generated on the basis of the PPNM, which only considers the second-order interactions between atoms.

The running times of different detectors are also provided in Table IV. Compared with several nonlinear detectors with the large computational cost of kernel matrix, the linear detectors besides LSC-TV are time-saving. Here, we analyze the time-consuming items in each detector. The LSC-TV updates the guided matrix $\hat{Z}$ in each superpixel block. The KIFD generates 1000 binary trees to detect anomalies. The KCRD computes the kernel matrix in each local region, while the KRX does it based on the whole HSIs, which results in that the running time of KCRD is much shorter than that of KRX. Besides, the KNUD utilizes the linear detector to obtain the background dictionary in each of the local regions, resulting in the time cost. The proposed method generates and solves (16) in each iteration, which is time-consuming but still acceptable compared with other detectors.
TABLE V  
AUC VALUES (MEAN ± STD) OF ALL THE DETECTORS ON THE SYNTHETIC DATASET WITH DIFFERENT LEVELS OF ZERO-MEAN WHITE GAUSSIAN NOISE (20 REPLICATES FOR EACH LEVEL OF NOISE)

| Algorithm | LRX       | CRD       | TVSDM     | LSC-TV    | KRX       | KCRD      | KIFD      | KNUD(P)   | KNUD(G)   | KDM-TVS(P) | KDM-TVS(G) |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|------------|------------|
| 40dB      | 0.9480±   | 0.9608±   | 0.9821±   | 0.9855±   | 0.9756±   | 0.9832±   | 0.9745±   | 0.9906±   | 0.9908±   | 0.9944±    | 0.9963±    |
| 35dB      | 0.9382±   | 0.9557±   | 0.9805±   | 0.9823±   | 0.9647±   | 0.9812±   | 0.9563±   | 0.9869±   | 0.9893±   | 0.9921±    | 0.9942±    |
| 30dB      | 0.9132±   | 0.9189±   | 0.9739±   | 0.9783±   | 0.9594±   | 0.9658±   | 0.9213±   | 0.9765±   | 0.9876±   | 0.9870±    | 0.9890±    |

where \( E[\cdot] \) computes the expectation of what it contains, and \( y \) and \( n \) represent the original pixel and the noise, respectively. The ROC curves are displayed in Fig. 11. It can be seen that the proposed KDM-TVS still outperforms other detectors at nearly all false alarms. Finally, we perform 20 replicates for each level of noise and then list the mean value and the standard deviation (std) of AUC values in Table V. It can be observed that KDM-TVS(P) and KDM-TVS(G) are more robust to the noise. Especially, KDM-TVS(G) is superior to KDM-TVS(P) in all noisy cases, which could be explained by the additive noise affecting the original nonlinear mixing properties, while the Gaussian kernel is able to explain the high-order nonlinear mixings, which leads to that the KDM-TVS(G) is more robust to noise than the KDM-TVS(P).

In summary, the proposed KDM-TVS outperforms other linear and nonlinear anomaly detection methods according to the experimental results. Moreover, the proposed KDM-TVS shows superior performance in noisy situations, and its running time is in an acceptable range.

B. Experiments on Real Hyperspectral Dataset

In this section, the anomaly detection performance of the proposed KDM-TVS is evaluated by three widely used real hyperspectral datasets, which were collected by the different sensors. The areas selected are the urban and airport scenarios, and the altitude difference of landcovers widely exists in these real HSIs due to the various man-made structures. These real HSIs are influenced by the light scattering effect, and thus, the nonlinear interaction has to be emphatically taken into consideration during anomaly detection.

The first real hyperspectral dataset covers an urban residential area and was captured by the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor. The original hyperspectral imagery displayed in Fig. 12(a) contains 400 × 400 pixels with 174 spectral bands, and the spatial resolution is 1 m/pixel. A subimage (with 80 × 100 pixels) in the red square is adopted in our experiments, and the corresponding pseudocolor image is displayed in Fig. 12(b).
The anomalies are roofs and vehicles in this dataset, and a portion of them are subanomalies for their small size.

The second dataset used in our experiments is a subimage (with $100 \times 100$ pixels) captured by AVIRIS, whose pseudocolor image is shown in Fig. 12(c). The anomalous targets to be detected are three airplanes with 72 pixels. It is worth pointing out that there are a number of subpixel anomalies in this dataset due to the mixing between the edge of planes and the background.

The third one was collected over Pavia by the reflective optics system imaging spectrometer (ROSIS) sensor. The original hyperspectral imagery has a size of $1096 \times 1096$ pixels with 102 spectral bands, and the spatial resolution is 1.3 m/pixel. A subimage with $100 \times 100$ pixels is adopted in the experiments, whose pseudocolor image is displayed in Fig. 12(d). The anomalies are the vehicles on the bridge, and the high-order scattering effects exist in this hyperspectral imagery for the prominent altitude difference between the river, bridge, and vehicles.
The dual window sizes \(w_{\text{in}}, w_{\text{out}}\) in LRX [8], CRD [12], KCRD [12], and KNUD [30] are set to (11, 13), (11, 13), (11, 13), and (9, 15), respectively. The detection maps of different detectors on the A VIRIS dataset are displayed in Fig. 16. Similarly, the differences between anomalies and backgrounds in the detection results of the nonlinear detectors are more obvious compared with the linear detectors. The high-order scattering effects exist in this dataset due to the evident altitude differences between rivers, bridges, and vehicles, leading to that the detection performance of KDM-TVS(G) is better than that of KDM-TVS(P). The ROC curves displayed in Fig. 17(c) show that

| Algorithm | LRX  | CRD  | TVSDM | LSC-TV | KRX  | KCRD | KIFD | KNUD(P) | KNUD(G) | KDM-TVS(P) | KDM-TVS(G) |
|-----------|------|------|-------|--------|------|------|------|---------|---------|------------|------------|
| HYDICE    | 0.9877 | 0.9892 | 0.9982 | 0.9969 | 0.8692 | 0.9944 | 0.9952 | 0.9915 | 0.9925 | 0.9977 | 0.9984 |
| San Diego | 0.9668 | 0.9785 | 0.9975 | 0.9967 | 0.9897 | 0.9952 | 0.9861 | 0.9965 | 0.9961 | 0.9994 | 0.9991 |
| Pavia     | 0.9842 | 0.9899 | 0.9959 | 0.9954 | 0.9930 | 0.9941 | 0.9851 | 0.9951 | 0.9967 | 0.9951 | 0.9976 |

TABLE VI
AUC VALUES OF DIFFERENT DETECTORS ON THE THREE REAL DATASETS
the detection rates of the KDM-TVS(G) are highest on almost all of the false alarm rates apart from the first 4%. The AUC values listed in Table VI also indicate that the proposed KDM-TVS are superior to other detectors.

C. Summary

The superiority of the proposed KDM-TVS has been fully validated by the experiment results on both the synthetic and real hyperspectral datasets. Three main advantages of the proposed method can be briefly summarized as follows.

1) Effectiveness: By means of the proposed decomposition model with endmember-kernel theory, each tested pixel can be represented by the nonlinear characteristics between union dictionary atoms appropriately based on the NMMs (i.e., PPNM and MLM). This strategy is more conform to the physical process of light scattering effect in HSIs. In order to achieve better separation performance, the sparsity and TV regularization terms are incorporated into the decomposition model to characterize the spatial properties of anomaly and background, respectively. Besides, a novel construction framework of union dictionary, which combines the superpixel segmentation and clustering methods sequentially, is designed to distinguish sparse backgrounds and anomalies effectively. The experiment results demonstrate that the anomalies can be identified by the proposed KDM-TVS from the background (even the sparse background) effectively and can achieve a 100% detection rate with a very low false alarm rate.

2) Robustness to the Noise: In order to evaluate the robustness of the proposed KDM-TVS in noisy cases, the synthetic dataset is corrupted by three levels of Gaussian noise. The incorporated TV regularization term is especially effective for suppressing noise pollution, and the experimental results on the synthetic dataset show that the proposed KDM-TVS remains satisfactory even when the SNR is very low.

3) Convenience in Parameters Setting: The experimental results on the change in parameters demonstrate that the detection performance of the proposed KDM-TVS is insensitive to the choices of algorithm parameter $\mu$ and dictionary parameter $R$ in a relatively large range. Moreover, $\mu$ and $R$ are fixed throughout the experiments on all of the datasets and still yield satisfactory detection results.

V. CONCLUSION

In this article, we have proposed a novel endmember-kernel-based nonlinear anomaly detection method (named KDM-TVS). Compared with the traditional nonlinear methods
that map the hyperspectral data into a high-dimensional space, the proposed KDM-TVS considers the inherent nonlinear characteristics between dictionary atoms, which is a more physically interpretable strategy and achieved by the endmember-kernel theory based on the NMMs implicitly. By means of the proposed decomposition model, hyperspectral imagery can be decomposed into three components: anomaly, background, and noise. Moreover, in order to obtain the optimal separation result, the sparsity and the TV regularizations are incorporated into the decomposition model to characterize the spatial property of background and anomaly, respectively. Besides, we have presented a novel construction framework of union dictionary, which combines the superpixel segmentation method and clustering method sequentially to represent the background and anomaly, which ensures that the proposed framework can include the sparse background into the background dictionary and eliminate the influence of anomalies when generating the background dictionary. Compared with other state-of-the-art detectors, the effectiveness and robustness of the proposed KDM-TVS have been verified adequately by the experimental results on both synthetic and real hyperspectral datasets.

It can be concluded from the experiments that the choices of kernel result in the difference in detection results on account of the various mixing modes in various datasets. Therefore, in order to accommodate the demand of the anomaly detection problem in various datasets, multiple kernel learning is a relevant strategy worth researching. Besides, the accuracy of the constructed dictionary would directly affect the accuracy of the subsequent solution of abundance. To make dictionaries more adaptive, the study of iterative strategies combining dictionary learning with coefficient learning should be our future work.

REFERENCES

[1] J. M. Bioncas-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] P. Ghanisian et al., “Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art,” IEEE Geosci. Remote Sens. Mag., vol. 5, no. 4, pp. 37–78, Dec. 2017.

[3] S. Matteoli, M. Diani, and G. Corsini, “A tutorial overview of anomaly detection in hyperspectral images,” IEEE Aeron. Electron. Syst. Mag., vol. 25, no. 7, pp. 5–28, Jul. 2010.

[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. 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.

[6] S. M. Schweizer and J. M. F. Moura, “Hyperspectral imaging: Clutter adaptation in anomaly detection,” IEEE Trans. Inf. Theory, vol. 46, no. 5, pp. 1855–1871, Aug. 2000.

[7] 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.

[8] D. Borghys, I. Käsen, 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.

[9] Q. Guo, B. Zhang, Q. Ran, L. Gao, J. Li, and A. Plaza, “Weighted-RX 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.

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

[11] N. M. Nasrabadi, “Regularization for spectral matched filter and RX anomaly detector,” Proc. SPIE, vol. 6966, Apr. 2008, Art. no. 696604.

[12] 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.

[13] T. Candès, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” J. ACM, vol. 58, no. 3, p. 11, May 2011.

[14] W. Wang, S. Li, H. Qi, B. Ayhan, C. Kwan, and S. Vance, “Identify anomaly component by sparsity and low rank,” in Proc. IEEE WHISPERS, Tokyo, Japan, Jun. 2015, pp. 1–4.

[15] 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.

[16] 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.

[17] M. Dao, C. Kwan, B. Ayhan, and T. D. Tran, “Burn scar detection using cloudy MODIS images via low-rank and sparsity-based models,” in Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP), Washington, DC, USA, Dec. 2016, pp. 177–181.

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

[19] R. Feng, H. Li, L. Wang, Y. Zhong, L. Zhang, and T. Zeng, “Local spatial constraint and total variation for hyperspectral anomaly detection,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–16, 2022.

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

[21] B. Hapke, “Bidirectional reflectance spectroscopy: I. Theory,” J. Geophys. Res., vol. 86, no. B4, pp. 3039–3054, Apr. 1981.

[22] W. Fan, B. Hu, J. Miller, and M. Li, “Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data,” Int. J. Remote Sens., vol. 30, no. 11, pp. 2951–2962, Jun. 2009.

[23] A. Halimi, Y. Altmann, N. Dobigeon, and J.-Y. Tourneret, “Nonlinear unmixing of hyperspectral images using a generalized bilinear model,” IEEE Trans. Geosci. Remote Sens., vol. 49, no. 11, pp. 4153–4162, Nov. 2011.

[24] Y. Altmann, A. Halimi, N. Dobigeon, and J.-Y. Tourneret, “Supervised nonlinear spectral unmixing using a postnonlinear mixing model for hyperspectral imagery,” IEEE Trans. Image Process., vol. 21, no. 6, pp. 3017–3025, Jun. 2012.

[25] R. Heylen and P. Scheunders, “A multilinear mixing model for nonlinear spectral unmixing,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 1, pp. 240–251, Jan. 2016.

[26] 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.

[27] 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.

[28] 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.

[29] J. Chen, C. Richard, and P. Honeine, “Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model,” IEEE Trans. Signal Process., vol. 61, no. 2, pp. 480–492, Jan. 2013.

[30] Y. Gao, J. Gu, T. Cheng, and B. Wang, “Kernel-based nonlinear anomaly detection via union dictionary for hyperspectral images,” IEEE Trans. Geosci. Remote Sens., vol. 60, pp. 1–15, 2022, doi: 10.1109/TGRS.2021.3093591.

[31] Z. Huang and S. Li, “From difference to similarity: A manifold ranking based hyperspectral anomaly detection framework,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 10, pp. 8118–8130, Oct. 2019.

[32] Y. Niu and B. Wang, “Hyperspectral anomaly detection based on low-rank representation and learned dictionary,” Remote Sens., vol. 8, no. 4, p. 289, 2016.
[33] 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.

[34] 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.

[35] D. Ma, Y. Yuan, and Q. Wang, “Hyperspectral anomaly detection via discriminative feature learning with multiple-dictionary sparse representation,” Remote Sens., vol. 10, no. 5, p. 745, May 2018.

[36] D. C. Heinz and C.-I. Chang, “Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery,” IEEE Trans. Geosci. Remote Sens., vol. 39, no. 3, pp. 529–545, Mar. 2001.

[37] N. Aronszajn, “Theory of reproducing kernels,” Trans. Amer. Math. Soc., vol. 68, no. 3, pp. 337–404, 1950.

[38] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[39] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” Found. Trends Mach. Learn., vol. 3, no. 1, pp. 1–122, Nov. 2011.

[40] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin, “An iterative regularization method for total variation-based image restoration,” Multiscale Model. Simul., vol. 4, no. 2, pp. 460–489, Jun. 2005.

[41] 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.

[42] 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.

[43] M.-D. Iordache, J. Bioucas-Dias, and A. Plaza, “Total variation spatial regularization for sparse hyperspectral unmixing,” IEEE Trans. Geosci. Remote Sens., vol. 50, no. 11, pp. 4484–4502, Nov. 2012.

[44] S. Yang and Z. Shi, “Hyperspectral image target detection improvement based on total variation,” IEEE Trans. Image Process., vol. 25, no. 5, pp. 2249–2258, May 2016.

[45] 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.

[46] D. G. Luenberger and Y. Ye, Linear and Nonlinear Programming, New York, NY, USA: Springer-Verlag, 2008.

[47] A. Beck and M. Teboulle, “Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems,” IEEE Trans. Image Process., vol. 18, no. 11, pp. 2419–2434, Nov. 2009.

[48] Y. Liang, P. P. Markopoulos, and E. S. Saber, “Subpixel target detection in hyperspectral images with local matched filtering in SLIC superpixels,” in Proc. 8th Workshop Hyperspectral Image Signal Processing: Evol. Remote Sens. (WHISPERS), Aug. 2016, pp. 1–5.

[49] A. Rodriguez and A. Laio, “Clustering by fast search and find of density peaks,” Science, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.

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

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

[52] S. M. Schweitzer and J. M. F. Moura. “Efficient detection in hyperspectral imagery,” IEEE Trans. Image Process., vol. 10, no. 4, pp. 584–597, Apr. 2001.

Ziyu Wu received the B.S. degree in electronic information science and technology from Fudan University, Shanghai, China, in 2019, where he is currently pursuing the Ph.D. degree with the School of Information Science and Technology. His research interests include hyperspectral target detection, machine learning, and pattern recognition.

Bin Wang (Senior Member, IEEE) received the B.S. degree in electronic engineering and the M.S. degree in communication and electronic systems from Xidian University, Xi’an, China, in 1985 and 1988, respectively, and the Ph.D. degree in system science from Kobe University, Kobe, Japan, in 1999. After his graduation in 1988, he was a Teacher with Xidian University. From 1999 to 2000, he was a Research Fellow with the Communications Research Laboratory, Ministry of Posts and Telecommunications, Kobe, Japan, working on magnetoencephalography signal processing and its application for brain science. He was a Senior Supervisor with the Department of Etching, Tokyo Electron AT Ltd., Tokyo, Japan, from 2000 to 2002, dealing with the development of advanced process control systems for etching semiconductor equipment. Since September 2002, he has been with the Department of Electronic Engineering, Fudan University, Shanghai, China, where he is currently a Full Professor and the Leader of the Image and Intelligence Laboratory. He has published more than 150 scientific papers in important domestic and international periodicals. He is the holder of several patents. His main research interests include multispectral/hyperspectral image analysis, automatic target/object detection and recognition, pattern recognition, signal detection and estimation, and machine learning.

Dr. Wang is an Associate Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (IEEE JSTARS).