Robust Dual Graph Self-Representation for Unsupervised Hyperspectral Band Selection

Yongshan Zhang, Member, IEEE, Xinxin Wang, Xinwei Jiang, Member, IEEE, and Yicong Zhou, Senior Member, IEEE

Abstract—Unsupervised band selection aims to select informative spectral bands to preprocess hyperspectral images (HSIs) without using labels. Traditional band selection methods only work well on Euclidean data, but ignore structural information of pixels and spectral bands. Moreover, they treat each HSI as a whole to exploit latent spatial information while ignoring the difference in spatial distribution between diverse homogeneous regions. In this article, we propose a robust dual graph self-representation (RDGSR) method for unsupervised band selection. RDGSR uses a superpixel segmentation technique to generate homogeneous regions of each HSI to extract spatial information. Based on the segmentation result, the superpixel-based similarity graph and band-based similarity graph are constructed from HSIs to record spatial and structural information. With this knowledge, the dual graph convolution is developed and the $\ell_{2,1}$-norm is introduced in the loss function and regularization term to eliminate the noise in rows for robust and effective band selection. The novelty of RDGSR is the joint utilization of the geometric structure of pixels with spatial consistency and the geometric structure of spectral bands to enhance the performance of band selection in a robust $\ell_{2,1}$-norm manner. An iterative optimization algorithm is designed to solve the proposed formulation. Substantial experiments on HSI datasets are conducted to verify the superiority of the proposed RDGSR over the state-of-the-art methods. The source code is available at https://github.com/ZhangYongshan/RDGSR.

Index Terms—Band selection, graph convolution, hyperspectral imagery, self-representation, unsupervised learning.

I. INTRODUCTION

WITH advanced spectral imaging technology, a great number of hyperspectral images (HSIs) are collected to archive the scenes from the observation areas [1], [2], [3]. In general, each HSI is represented by a data cube that contains hundreds of spectral bands to record the reflectance of the scene with a wide range of electromagnetic waves. Compared to red-green-blue (RGB) images, HSIs provide more abundant information to identify the distinction between diverse landcover materials. It has been proven that the recognition of HSIs can be achieved by using advanced technologies to solve many Earth observation tasks like military defense, forest protection, and anomaly detection. However, the high-dimensional HSI data inevitably contain redundant and noisy bands that may induce degenerated performance and high computational complexity in the subsequent learning [4], [5], [6]. To overcome these problems, it is necessary to reduce negative information among spectral bands.

For dimensionality reduction of HSIs, feature extraction and band selection are crucial techniques. The former one with designed learning criteria aims to find an optimal transformation matrix to project the original HSI data into a new feature space [7], [8], [9]. Representative methods are principal component analysis (PCA) [8], linear discriminant analysis (LDA) [10] and neighborhood preserving embedding (NPE) [11]. The physical meaning of the new projected representation is not clear and differs from that of the original data. The latter one endeavors to discriminate representative spectral bands from the original ones without transformation [12], [13], [14]. Compared to feature extraction, band selection can preserve the inherent properties and physical meaning of the original spectral bands. Considering these advantages, we focus on band selection in this work.

According to the presence or absence of labels, band selection approaches can be divided into supervised and unsupervised ones. Supervised approaches train learning models with completely labeled samples to select the optimal bands [15], [16], [17]. In [15], Yang et al. proposed an efficient method to select informative bands with known class labels. To conduct effective band selection, Feng et al. [16] presented a pointwise-ranking-based criterion with a nonhomogeneous hidden Markov chain model and known labels. Compared to traditional learning methods [15], [16], deep learning is also a promising technique for band selection. In [18], Feng et al. developed a dual-graph convolutional network (GCN) with band attention and sparse constraint for band selection. To avoid repeatedly training, Feng et al. [19] further devised a deep reinforcement learning model to achieve band selection. Other deep learning-based methods were also proposed for effective band selection [20], [21]. For supervised...
methods, it is difficult to obtain sufficient and accurate labels by human annotation. Thus, unsupervised approaches are more flexible when labels are not available in most cases. Compared to supervised methods learning with labeled HSI data, unsupervised methods only need unlabeled data to develop learning models based on certain criteria without using explicit labels [22], [23], [24]. In [23], Jia et al. proposed a fast density-peak-based clustering method for unsupervised band selection. To achieve robust band selection, Sui et al. [25] incorporated multiple graphs into a self-contained regression model. Most aforementioned methods only use the spectral band for learning while ignoring potential spatial information. Thus, the performance may be compromised and can be further improved.

Advanced methods utilize spectral signatures as well as auxiliary information to conduct band selection that can gain better results than solely using spectral bands [26], [27], [28]. Apart from spectral signatures, there are abundant spatial information and potential structural information hidden in HSIs. Spatial information showing the shapes, textures, and edges exhibits in the two dimensions of pixels. In [29], Roy et al. developed a dual-attention reconstruction network to capture contextual information in both spectral and spatial directions and select informative bands. With both spectral and spatial information, Feng et al. [20] devised an end-to-end network that can simultaneously achieve band selection, feature extraction and classification. These methods utilize both spectral and spatial information, but they neglect inconspicuous structural information of HSIs. Structural information reflecting the relationships between two entities exists in both pixels and spectral bands. In [28], Cai et al. presented an efficient graph convolutional self-representation for band selection with structural graph constructed from spectral bands. Further, Zhang et al. [8] proposed a marginalized graph self-representation with introduced noises and region-wise similarity graph for band selection. The utilization of auxiliary information provides additional and distinctive knowledge to better understand HSIs from other perspectives. It has been proven that reasonable using spatial and structural information is beneficial to significantly promote the performance of band selection.

Although most advanced methods have achieved promising performance with auxiliary information, they still exist the following limitations. First, simultaneous feature selection and extraction method [26] and multimodal learning feature selection method [27] regard each HSI as a whole to extract spatial information. In fact, spatially neighboring pixels with similar spectral signatures share the same land cover with high probability. To solve this issue, Feng et al. [19], [20] adopted a square window to extract potential spatial information. As shown in the grayscale image of the HSI in Fig. 1, it is clear that a square window can not accurately extract spatial information. A more reasonable solution is to extract spatial information from homogenous regions with adaptive shapes and sizes. Second, marginalized graph self-representation [8] considers only the geometric structure of pixels, while graph convolutional self-representation [28] utilizes only the geometric structure of spectral bands. For HSIs, inconspicuous structural information exists in both pixels and bands to reflect their pair-wise relationships. The effectiveness of the interactions between the geometric structures of pixels and bands is not well investigated in previous work. It is desired to jointly explore the dual geometric structures of pixels and bands in a unified model. Third, latent representation learning-based
autoencoder [22] and multiojective optimization-based self-representation [30] train learning models on the original HSI data without considering the potential outliers. In reality, outliers or noises may be introduced in HSIs due to the uncertainty occurred during data collection. When training a model with outliers or noises, the performance is inevitably degenerated. To alleviate the negative effect of outliers and noises, learning models should be designed in a more robust manner. Based on the above observations, how to accurately extract spatial information and explore dual geometric structures from both pixels and bands in a robust manner is a problem needed to be studied.

In this article, we propose a robust dual graph self-representation (RDGSR) method to achieve unsupervised band selection. The flowchart of RDGSR is shown in Fig. 1. Specifically, RDGSR utilizes superpixel segmentation to generate a segmentation map with multiple homogenous regions from the first principal component of the original HSI data. The superpixel-based and band-based similarity graphs are then constructed to record spatial and structural information among multiple segmentations and spectral bands. Using this knowledge, we develop a dual graph convolution to characterize the loss function and introduce the \(\ell_{2,1}\)-norm in both the loss function and regularization term to achieve robust and effective band selection. An efficient optimization algorithm is designed to solve the proposed formulation. Experimental results on HSI datasets demonstrate the effectiveness of the proposed RDGSR compared to the state-of-the-art methods.

In summary, the contributions of this article are as follows.

1) We propose an RDGSR method to achieve unsupervised band selection. The \(\ell_{2,1}\)-norm is introduced into the loss function of self-representation because it is less sensitive to outliers and noises than the Frobenius norm. The proposed RDGSR specializes in selecting informative bands from original ones, especially in a noisy environment.

2) The dual graph convolution with superpixel-based and band-based similarity graphs is seamlessly integrated into the proposed RDGSR. It provides a learning paradigm to directly explore the interactions between the geometric structures of pixels and bands with matrix multiplication. By doing so, structural information is introduced into the learning model.

3) The superpixel-based similarity graph is constructed from multiple homogenous regions obtained by superpixel segmentation. The construction of a superpixel-based similarity graph in our method effectively reduces searching space and computational time. Besides, local spatial consistency is recorded in this graph.

4) We devise an efficient optimization strategy as a solution to solve the proposed RDGSR. The optimization algorithm shows desired convergence and finds the optimal solution for RDGSR.

The rest of this article is organized as follows. Section II briefly surveys the related work about GCN and self-representation. Section III presents the proposed RDGSR method for unsupervised hyperspectral band selection and introduces an optimization algorithm as a solution. Extensive experiments on HSIs are conducted in Section IV. The discussion is given in Section V. Finally, Section VI concludes this article.

II. RELATED WORK

A. Graph Convolutional Network

Graph neural network (GNN) is an improved variant of a neural network that efficiently handles graph-structured data in the non-Euclidean domain. The concept of GNN was first proposed by Gori et al. [31]. Subsequently, Scarselli et al. [32] trained a GNN model with a supervised learning algorithm that is inapplicable to large-scale data. To efficiently handle graph-structured data, Bruna et al. [33] utilized the spectral property to develop the operation of graph convolution. After that, various extensions of GNN have been proposed and gained promising results. Representative models are structure2vec [34], separable recurrent multi-graph convolutional neural network (sRGCNN) [35] and graph sample and aggregate (GraphSAGE) [36]. Based on the previous work, Kipf and Welling [37] proposed a GCN with a fast approximation of graph spectral convolution that can directly encode both graph structure and node features in a unified framework. This is the most popular and commonly used model to solve the problems in graph domain.

Assume that there is a graph \(G = (\mathcal{V}, \mathcal{E})\) with a set of vertexes/nodes \(v_i \in \mathcal{V} ((|\mathcal{V}| = N)\) and a set of edges \(e_{ij} \in \mathcal{E}\) connecting vertexes \(v_i\) and \(v_j\), the matrix \(X \in \mathbb{R}^{N \times D}\) contains all \(N\) vertexes described by a \(D\)-dimensional feature vector and the graph \(A \in \mathbb{R}^{N \times N}\) records the information of edges that represents the adjacency relationships between vertexes. According to \(A\), we define the degree matrix \(D \in \mathbb{R}^{N \times N}\) with diagonal elements \(D_{ii} = \sum_j A_{ij}\). Using the above information, a one-layer GCN calculates the new \(L\)-dimensional graph embedding matrix \(H \in \mathbb{R}^{N \times L}\) as follows:

\[
H^{(1)} = \sigma (\hat{A}XW^{(0)})
\]

where \(\hat{A} = \hat{D}^{-1/2}\hat{A}\hat{D}^{-1/2}\) is the normalized symmetric adjacency matrix. Specifically, \(A = A + 1_N\) is the similarity graph with self-loops and \(\hat{D}\) is a degree matrix with diagonal elements \(\hat{D}_{ii} = \sum_j \hat{A}_{ij}\). \(W^{(0)} \in \mathbb{R}^{R \times L}\) is a trainable weight matrix and \(\sigma ()\) is an activation function. To learn a high-order graph embedding, multiple GCN models can be stacked in a layer-wise fashion as

\[
H^{(j+1)} = \sigma (\hat{A}H^{(j)}W^{(j)})
\]

where \(j\) indicates the layer number and \(H^{(0)} = X\). GCN is essentially designed for the problems in non-Euclidean domain. In recent research, it has also shown great potential in many learning problems like text classification [38], emotion recognition [39] and HSI analysis [40].

B. Self-Representation

Self-representation is an unsupervised learning framework to reconstruct the input data itself by utilizing the self-expressive property [41]. The purpose of
self-representation is to learn the smallest reconstruction residual of the self-expressive relationships of samples or features. Suppose that the data matrix is denoted as \( X = \{ x_1, x_2, \ldots, x_N \} \in \mathbb{R}^{N \times B} \) with \( N \) samples and \( B \) features, the self-representation model to learn the relationships among features is formulated as

\[
\min \mathcal{L}(X - XW) + \lambda \mathcal{R}(W)
\]

where \( W \in \mathbb{R}^{B \times B} \) is the coefficient matrix. \( \mathcal{L}(X - XW) \) is the loss function to measure the discrepancies between \( X \) and \( XW \) and \( \mathcal{R}(W) \) is the regularization term imposed on \( W \) balanced with parameter \( \lambda \). According to explicit learning purposes, different loss functions and regularization terms can be used to formulate different learning models [42], [43], [44].

Recently, self-representation has been proven to be effective for HSI analysis, such as band selection, clustering, and unmixing. Band selection is a hot research topic to preprocess high-dimensional HSI data. In [45], Sun et al. incorporated a dissimilarity-weighted regularization term into the sparse self-representation to select informative bands. To achieve fast and robust band selection, Sun et al. [46] further integrated the structured random projection into the robust self-representation. Afterward, Wei et al. [47] processed HSI data in a streaming manner to dynamically select discriminative bands to avoid the heavy storage burden. To exploit the structural relationships among spectral bands, Cai et al. [28] constructed the similarity graph recording the affinity between adjacent bands to promote band selection. Although the above methods gain superior results, they fail to exploit spatial information that can further improve performance.

Different from existing methods, our method exploits structural information with spatial consistency among multiple superpixels. Besides, our method also explores structural information from spectral bands. Instead of treating structural graph information as a regularization term, our method jointly utilizes spectral, spatial and geometric information to steer band selection in a natural and robust manner.

III. PROPOSED METHOD

In this section, we will expound on the proposed RDGSR method with a designed optimization algorithm to distinguish informative bands.

A. Structural Graph Construction

A graph is a structure consisting of vertexes to represent entities and weighted edges to denote their pair-wise relationships. The affinity between any two vertexes is measured by the weight related to the edge that connects them. It is known that graphs can effectively record the relationships among various entities that provide structural information. For HSIs, structural information exists in both pixels and bands. It is beneficial to exploit dual structural graphs for further learning.

1) Construction of Region-Wise Pixel-Based Graph: Each HSI record a scene from the observed area with a number of spectral signatures. It is clear that different land covers in diverse regions exhibit different characteristics. Spatial information is related to the distribution of land-cover materials and differs in diverse homogeneous regions [48], [49]. In light of this, we utilize superpixel segmentation technique to localize the texture of HSI data and generate homogeneous regions with spatial consistency. Due to the flexibility and efficiency, the entropy rate superpixel (ERS) segmentation method [50] is adopted to localize the homogeneity of HSI data. Initially, the 3-D HSI data cube \( X' \in \mathbb{R}^{W \times H \times B} \) with \( W \times H \) pixels and \( B \) spectral bands is reshaped to a 2-D spectral matrix \( X_1 \in \mathbb{R}^{N \times B} \) of each HSI. Then, we can generate \( S \) homogeneous regions by conducting ERS on \( I_f \) as follows:

\[
I_f = \bigcup_{s=1}^{S} \mathcal{H}_s, \quad \text{s.t.} \quad \mathcal{H}_s \cap \mathcal{H}_g = \emptyset, \quad (s \neq g)
\]

where \( \mathcal{H}_s \) is the \( s \)th segmentation region. The ERS segmentation method is originally developed for generating superpixels from RGB images. In general, ERS conducts segmentation on grayscale images converted from RGB images. It means that ERS conducts segmentation on one component instead of more components. To segment HSIs with many spectral bands (i.e., components), it is necessary to get only one component with major information on HSIs. Therefore, we learn the first principal component for HSI segmentation. Mapping the segmentation result to the original HSI, we can get the segmented HSI to construct the region-wise pixel-based graph [51].

To extract structural information with spatial consistency among superpixels in a graph \( G_F = (V_F, \mathcal{E}_F) \), pixels are treated as vertexes \( v_i \in \mathcal{V}_F \), and the relationships between pair-wise pixels are regarded as the weighted edges \( e_{ij} \in \mathcal{E}_F \) that connect two adjacent vertexes. Given any two pixels \( x_i \) and \( x_j \) from the segmented HSI, the region-wise pixel-based graph \( A_F \in \mathbb{R}^{N \times N} \) can be calculated as

\[
A_F(i,j) = \begin{cases} 
\exp \left( -\frac{||x_i - x_j||^2}{2\sigma^2} \right), & x_i, x_j \in \mathcal{H}_s \\
0, & \text{otherwise}
\end{cases}
\]

where \( \sigma \) is the kernel width of the Gaussian kernel function. According to (5), the similarity of neighboring pixels within the same region is determined by the Gaussian kernel function and zero otherwise. By doing so, the geometric structure of pixels with local spatial consistency can be preserved in \( A_F \). In most graph-based methods, the pixel-based graph is constructed by traversing the entire HSI to record the similarities of the first \( k \) nearest neighbors. It takes \( O(N^2 B + N^2 \log N) \) to construct traditional \( k \)-nearest neighbor (KNN) graph at pixel level for HSIs. It is time-consuming when the number of pixels \( N \) is very large. Different from the KNN graph, the construction of \( A_F \) is mainly developed by discovering the relationships between pixels belonging to the same segmentation. When the HSI is segmented into \( S \) nonoverlapping superpixels \( \{ \mathcal{H}_s \}_{s=1}^{S} \), each superpixel \( \mathcal{H}_s \) consists of \( n_s \) pixels and \( \sum_{s=1}^{S} n_s = N \). It takes \( O(\sum_{s=1}^{S} n_s^2 B) \) to construct \( A_F \). The affinity calculation of \( A_F \) is restricted in the same segmentation region instead of traversing the entire HSI. Therefore, it is much computationally efficient when \( S > 1 \) because \( \sum_{s=1}^{S} n_s^2 \ll N^2 = (\sum_{s=1}^{S} n_s)^2 \).
2) Construction of Band-Based Graph: Apart from the geometric structure of pixels, structural information also exists in spectral bands. To extract structural information among spectral bands, we treat spectral bands as vertices $v_i \in \mathcal{V}_B$ and quantify the relationships between neighboring spectral bands as the weighted edges $e_{ij} \in \mathcal{E}_B$ that connect two vertices. In such manner, the spectral band-based graph $A_B \in \mathbb{R}^{B \times B}$ can be obtained. Suppose that the $i$th and $j$th spectral bands are denoted as $b_i$ and $b_j$, $A_B$ can be calculated as

$$A_{B(ij)} = \begin{cases} \exp\left(\frac{\|b_i - b_j\|^2}{2\sigma^2}\right), & b_i \in \mathcal{N}_k(b_j) \text{ or } b_j \in \mathcal{N}_k(b_i) \\ 0, & \text{otherwise} \end{cases}$$

(6)

where $\mathcal{N}_k(b_j)$ represents the set of $k$ nearest neighbors of $b_j$. In detail, the similarity of neighboring spectral bands belonging to the set of $k$ nearest neighbors is quantified by the Gaussian kernel function. Otherwise, the similarity of the two bands is set as zero. The construction of $A_B$ is also computationally efficient because $B \ll N$. Using the spectral matrix $X$ and structural graphs $A_P$ and $A_B$ to represent HSI data, the band selection problem in our article is extended into the non-Euclidean domain.

B. Robust Dual Graph Self-Representation

Self-representation utilizes the self-expressive property to reconstruct the input data itself. For hyperspectral band selection, self-representation aims to learn the smallest residual and model the self-expressive relationship of spectral bands [30], [46]. According to recent publications, we found that most existing methods adopt the Frobenius norm to measure the residual. In reality, the Frobenius norm is sensitive to outliers. Considering the outliers in HSIIs, the $\ell_{2,1}$-norm is more suitable to characterize the residual of self-representation due to the capability of making row sparse [41]. Thus, a robust variant of self-representation is represented as

$$\min_W \|X - XW\|_{2,1} + \lambda \mathcal{R}(W)$$

(7)

where $\mathcal{R}(\cdot)$ is a regularization term and $\lambda$ is a regularization parameter. Equation (7) formulates a model to solve the problem with Euclidean data. When the data are represented in the non-Euclidean domain, (7) is incapable.

GCN is an improved variant of neural networks that operate directly on graph-structured data [37]. Compared to graph Laplacian regularization, graph convolution provides an effective solution to graph-based learning problems without relying on the assumption of sharing the same label among connected nodes. GCN simultaneously adopts the adjacency matrix and feature matrix in a learning model with matrix multiplication. As described in Section III-A, the region-wise pixel-based graph $A_P$ is obtained from all segmentations at the pixel level to record the relationships between pair-wise pixels with spatial consistency and the band-based graph $A_B$ is calculated at band level to extract the relationships between pair-wise spectral bands. It is beneficial to explore and exploit the interaction and effectiveness of the dual graphs from both pixel and band levels for HSI analysis. Motivated by GCN, we simultaneously integrate the dual graphs into the self-representation model to achieve robust band selection for HSIIs. Therefore, the proposed RDGSR is formulated as follows:

$$\min_W \|X - (\hat{A}_P X + \hat{X}\hat{A}_B)W\|_{2,1} + \lambda \mathcal{R}(W)$$

(8)

where $\hat{A}_P = \tilde{D}_P^{-1/2}\hat{A}_P \tilde{D}_P^{-1/2}$ is the normalized region-wise pixel-based similarity graph with the size of $N \times N$ and $\hat{A}_B = \tilde{D}_B^{-1/2}\hat{A}_B \tilde{D}_B^{-1/2}$ is the normalized band-based similarity graph with the size of $B \times B$. More specifically, $\hat{A}_P = A_P + I_N$ is the region-wise pixel-based similarity graph with self-loops and $\tilde{D}_P = \text{diag}(\hat{D}_P)$. The definition of $A_B$ and $\tilde{D}_B$ is similar to the above descriptions. Since the size of $X$ is $N \times B$, $\hat{A}_P X$ and $\hat{X}\hat{A}_B$ are both with the size of $N \times B$. Therefore, the reconstruction error in (8) can be achieved with matrix manipulation. With the size of $B \times B$, the coefficient matrix $W$ is used to model the relationship between $X$ and $(\hat{A}_P X + \hat{X}\hat{A}_B)$ that reflects the band-wise weights. Instead of simply using the spectral matrix, the proposed RDGSR of (8) is an efficient and effective model with dual graph convolution to jointly integrate spectral, spatial, and geometric information in a unified framework.

To realize band selection, we impose an $\ell_{2,1}$-norm regularization [43] on the weight matrix $W$. This will lead to the sparsity of $W$ in a row making it particularly suitable for band selection. Thus, the objective function of the proposed RDGSR is finally formulated as

$$\min_W \|X - (\hat{A}_P X + \hat{X}\hat{A}_B)W\|_{2,1} + \lambda \|W\|_{2,1}.$$ 

(9)

As shown in (9), the spectral matrix and dual structural graphs are harmoniously integrated into a unified model. It is beneficial to jointly exploit spectral and spatial information in non-Euclidean domain for band selection. With the dual similarity graphs from region-wise pixels and spectral bands, the performance of the proposed method can be further improved. After obtaining the optimal $W$, we rank the significance of spectral bands according to the descending order of $\|w_i\|_2 (i = 1, \ldots, B)$ to achieve band selection. Before evaluation, the proposed RDGSR does not introduce any explicit label information for band selection. Thus, it still belongs to the unsupervised band selection method.

C. Optimization for RDGSR

Solving the problem (9) for the proposed RDGSR is difficult due to the $\ell_{2,1}$-norm regularization of the loss function and penalty term. To clearly illustrate the solution to problem (9), we define an auxiliary matrix $Z$ to represent $(\hat{A}_P X + \hat{X}\hat{A}_B)$. Then, the formulation of RDGSR can be rewritten as

$$\min_W \|X - ZW\|_{2,1} + \|W\|_{2,1}.$$ 

(10)

To solve problem (10), we relax $\|X - ZW\|_{2,1}$ and $\|W\|_{2,1}$ by $\text{Tr}((X - ZW)^T C_L(X - ZW))$ and $\text{Tr}(W^T C_B W)$, respectively.
The matrix $C_L \in \mathbb{R}^{N \times N}$ and $C_R \in \mathbb{R}^{B \times B}$ are diagonal matrices with diagonal elements $C_{L(i,i)} = (1/2 \| x_i - z_{iW} \|_2^2 + \epsilon)$ and $C_{R(i,i)} = (1/(2 \| w_i \|_2^2 + \epsilon)$, respectively. In practice, $\epsilon$ is a very small constant introduced to prevent the overflow error. Thus, problem (10) can be reformulated as
$$J(W) = \text{Tr}((X - ZW)^T C_L (X - ZW)) + \lambda \text{Tr}(W^T C_R W).$$

To solve the problem (11), we calculate the derivative w.r.t $W$ and set it to zero as follows:
$$\frac{\partial J}{\partial W} = -2Z^T C_L (X - ZW) + 2\lambda C_R W = 0. \quad (12)$$
According to (12), the update rule of $W$ is given by
$$W = (Z^T C_L Z + \lambda C_R)^{-1} Z^T C_L X. \quad (13)$$

In the proposed formulation, the matrix $W$ is the only variable to be learned for band selection. As shown in the previous descriptions, (13) provides an efficient and effective solution to learn the optimal $W$. Initially, the matrices $C_L$ and $C_R$ are initialized with the identity matrices. The matrices $Z$ and $X$ are fixed and unchanged during learning. When conducting repeated experiments with the same parameter setting, the optimal $W$ is still the same after iterative learning. Therefore, the band selection results are unchanged and have nothing to do with the initialization of $W$. Besides, we know that the calculation of $C_L$ and $C_R$ also relies on $W$. Thus, we alternately update $W$, $C_L$, and $C_R$ in each iteration. During the iterative learning process, the objective function value of (9) monotonically decreases to reach the optimal solution to RDGSR. The pseudocode of RDGSR is exhibited in Algorithm 1.

D. Complexity Analysis

1) Time Complexity Analysis: Since there are several learning steps in the proposed RDGSR, we analyze the time complexity of RDGSR step by step. It takes $O(B^2 N + B^3)$ to perform PCA on the original HSI. The time cost of applying ERS to get multiple segmentations is $O(N \log N)$. Given $S$ nonoverlapping superpixels $\{s_1, \ldots, s_S\}$ with $n_i$ pixels, it takes $O(S^2 n_i^2 B)$ to construct the region-wise pixel-based graph $A_P$. The time cost of constructing the band-based graph $A_B$ requires $O(B^2 N + B^2 \log B)$. To calculate the auxiliary matrix $Z$, the time cost is $O(N^2 B + B^2 N)$. The overall computational cost of RDGSR is the optimization of the weight matrix $W$. To solve the optimization of $W$, it takes $O(B^3$) to compute matrix inverse and $O(B^2 N)$ to determine matrix multiplication. Suppose that there is $T$ learning iterations, the overall time cost to learn $W$ is $O(B^2 N + B^3 T)$. Since $S \ll N$, the overall computational cost of RDGSR can be simplified as $O(N^2 + B^2 + N B)$.

IV. EXPERIMENTS

A. Hyperspectral Datasets

In the experiments, five public HSI datasets are adopted. They are Indian Pines, Pavia University (PaviaU), Kennedy Space Center (KSC), Botswana, and Washington DC Mall (DC). Table I summarizes the major information.

1) Indian Pines: This dataset was gathered by the airborne visible and infrared imaging spectrometer (AVIRIS) sensor in 1992 and recorded a scene of the physiognomy from the Indian Pines test site in Northwest Indiana. There are $145 \times 145$ pixels and 224 spectral signatures within the wavelength range between 400 and 2500 nm. After removing the water-absorption and noise bands, 200 valid spectral bands remained to describe 16 types of different land covers with regular geometry in this scene.

2) Pavia University (PaviaU): This dataset was obtained by the reflective optics system imaging spectrometer (ROSIS) sensor in 2002 and recorded a scene over the University of Pavia. There are $610 \times 340$ pixels by discarding some pixels from the original data with the size of $610 \times 610$. After removing some bands with a low signal-to-noise ratio (SNR), 103 spectral bands are retained to describe nine classes of different land covers in this extracted dataset.

3) Kennedy Space Center (KSC): This dataset was captured by the National Aeronautics and Space Administration (NASA) AVIRIS instrument in 1996 and recorded a scene over KSC, Florida, USA. After the removal of water-absorption and low SNR bands, this dataset is recorded by 176 useful spectral bands to represent a scene with $512 \times 614$ pixels from a spatial...
resolution of 18 m. There are 13 types of different land covers recorded in this scene.

4) Botswana: This dataset was collected by the NASA Earth Observing-1 (EO-1) satellite in 2001 and recorded a scene from the Okavango Delta, Botswana. Requiring at 30 m pixel resolution, this scene consists of 256 × 1476 pixels. After the removal of water-absorption and low SNR bands, there are 145 available spectral bands from the portion of the spectrum within 400–2500 nm to describe 14 classes of different land covers.

5) Washington DC Mall (DC): This dataset was obtained by the hyperspectral digital imagery collection experiment (HYDICE) sensor in 1995 and recorded a scene of an urban area over the Washington DC Mall. There are 191 spectral bands left in this dataset after removing some useless bands. A subscene with 307 × 280 is extracted from the original scene and used in the experiment. There are six types of different land covers recorded in this subscene.

B. Experimental Settings

1) Compared Methods: To validate the effectiveness of the proposed RDGSR, we utilize the following unsupervised methods for comparison. AllBands is taken as a baseline using all spectral bands in the experiment. LapScore [52] selects informative spectral bands by virtue of the Laplacian matrix. Maximum-variance PCA (MVPCA) [53] selects valuable spectral bands that use maximum-variance principal analysis as the selection criterion. Enhanced fast density peak-based clustering (E_FDPC) [23] is a ranking-based method to estimate the normalized local density and intracluster distance for band selection. Regularized self-representation (RSR) [41] is a robust method with an ℓ_{2,1}-norm to discriminate informative spectral bands. Scalable one-pass self-representation learning (SOP-SRL) [47] treats HSI data as a stream to dynamically select useful spectral bands. Efficient graph convolutional self-representation (EGCSR) [28] constructs a band-based graph to steer band selection with the help of graph convolution. Optimal neighboring reconstruction (ONR) [54] considers band selection as a combinatorial optimization problem by assessing the ability of selected bands to reconstruct other ones. Fast neighborhood grouping band selection (FNGBS) [55] utilizes local density and information entropy as metrics to select the most relevant and informative bands from the partitioned band subsets. Dual-graph sparse non-negative matrix factorization (DSNMF) [56] selects representative bands by combining the dual-graph model from both pixels and bands with nonnegative matrix factorization. RDGSR is the proposed method to achieve band selection with both spectral information and structural information with spatial consistency among pixels and bands. Consistent with the unsupervised learning scheme, we adopt k-means clustering algorithm to measure the quality of selected bands instead of using classifiers like support vector machine (SVM) and KNN. Different from classifiers learning with completely labeled samples, k-means clustering algorithm only needs unlabeled samples for learning. It is more flexible and applicable in most practical cases due to the difficulty of obtaining sufficient labels. Besides, when using the k-means clustering algorithm to measure the quality of selected bands for unsupervised band selection, there is no need to divide the dataset into the training and testing sets. The random or manual partition of the training and testing sets may introduce uncertainty into the final results. Considering all these reasons, we report clustering performance obtained by k-means with the maximum iterations of 100 in our article rather than classification performance obtained by classifiers for comparison.

2) Evaluation Metrics: Since we use the k-means clustering algorithm to evaluate the band selection performance of all compared methods, clustering evaluation metrics are adopted in our setting. With the selected spectral bands, we evaluate the performance in terms of four quantitative metrics, including clustering accuracy (ACC), normalized mutual information (NMI), purity and Kappa coefficient. For all these metrics, the better quality of the selected bands contributes to the larger values.

3) Parameter Settings: There are two hyperparameters in the proposed RDGSR that are tuned with a grid-search strategy. The number of segmentations S varies within [3, 7, 15, 25, 50, 100, 200, 500, 1000] and the regularization parameter λ is chosen from [10⁻³, 10⁻², ⋯, 10³]. For hyperspectral band selection, the desired number of selected bands is unknown and difficult to determine in practice for different datasets. In our setting, the number of selected bands varies from 5 to 50 with step 5 to evaluate the capability of band selection for all methods. The hyperparameters in other compared methods are also tuned to get the best results. The reported results are averaged from ten repeated trials to alleviate the random initialization in k-means.

C. Performance Evaluation

We empirically evaluate the performance of these unsupervised band selection methods by selecting the top 20 ranked spectral bands for clustering. The detailed results on the HSI datasets are reported in Table II. From Table II, we note that the proposed RDGSR obtains the best results on the five HSI datasets compared to all band selection methods. This indicates the superiority of the proposed RDGSR for band selection. We also show the clustering maps of 20 selected bands for a part of the Indian Pines scene in Fig. 2.

From the experimental results, we can get the observations as follows. First, compared to RSR, our RDGSR obtains at least 3.28%, 1.51%, 7.35%, 4.18%, and 1.21% improvements on the five HSI datasets, respectively. This demonstrates that it is effective to consider both structural information among multiple superpixels and spectral bands in the band selection model. Second, compared to EGCSR, our RDGSR also achieves better performance in terms of the four evaluation metrics on the HSI datasets. This signifies that it is imperative to introduce structural information with spatial consistency among heterogeneous regions and remove noisy information by the ℓ_{2,1}-norm learning scheme. Third, by comparing E_FDPC, SOP-SRL, ONR, and FNGBS, our RDGSR obtains at least 11.73%, 10.76%, 10.00%, and 5.98% improvements on the Indian Pines datasets. The superiority of EGCSR and DSNMF can be also found in the comparison with E_FDPC and SOP-SRL. This indicates that it is crucial to explore structural information in non-Euclidean domain to steer band selection. Fourth, LapScore and MVPCA are inferior to other compared methods for comparison.

Authorized licensed use limited to: Universidade de Macau. Downloaded on September 30,2022 at 23:58:41 UTC from IEEE Xplore. Restrictions apply.
TABLE II
CLUSTERING PERFORMANCE COMPARISON OF DIFFERENT METHODS WITH 20 SELECTED BANDS ON HSI DATASETS

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Measure</th>
<th>AllBands</th>
<th>LapScore</th>
<th>MVPCA</th>
<th>E_FDPC</th>
<th>RSR</th>
<th>SOP SRL</th>
<th>EGCSR</th>
<th>ONR</th>
<th>FNGBS</th>
<th>DSNMF*</th>
<th>RDGSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indian Pines</td>
<td>ACC</td>
<td>0.3367</td>
<td>0.3320</td>
<td>0.3173</td>
<td>0.3443</td>
<td>0.3216</td>
<td>0.3364</td>
<td>0.3324</td>
<td>0.3380</td>
<td>0.3535</td>
<td>0.3582</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NMI</td>
<td>0.4246</td>
<td>0.4012</td>
<td>0.3955</td>
<td>0.3917</td>
<td>0.4182</td>
<td>0.3931</td>
<td>0.4243</td>
<td>0.4062</td>
<td>0.4158</td>
<td>0.4238</td>
<td>0.4323</td>
</tr>
<tr>
<td></td>
<td>Purity</td>
<td>0.5109</td>
<td>0.4995</td>
<td>0.4899</td>
<td>0.4924</td>
<td>0.5148</td>
<td>0.4910</td>
<td>0.5190</td>
<td>0.4960</td>
<td>0.5035</td>
<td>0.5214</td>
<td>0.5295</td>
</tr>
<tr>
<td></td>
<td>Kappa</td>
<td>0.2971</td>
<td>0.2555</td>
<td>0.2565</td>
<td>0.2891</td>
<td>0.2616</td>
<td>0.2942</td>
<td>0.2581</td>
<td>0.2737</td>
<td>0.2951</td>
<td>0.2978</td>
<td></td>
</tr>
<tr>
<td>PaviaU</td>
<td>ACC</td>
<td>0.3233</td>
<td>0.3626</td>
<td>0.5302</td>
<td>0.5111</td>
<td>0.5368</td>
<td>0.5543</td>
<td>0.5339</td>
<td>0.5249</td>
<td>0.5120</td>
<td>0.5627</td>
<td>0.5673</td>
</tr>
<tr>
<td></td>
<td>NMI</td>
<td>0.5416</td>
<td>0.5539</td>
<td>0.5513</td>
<td>0.5393</td>
<td>0.5524</td>
<td>0.5509</td>
<td>0.5527</td>
<td>0.5451</td>
<td>0.5391</td>
<td>0.5543</td>
<td>0.5618</td>
</tr>
<tr>
<td></td>
<td>Purity</td>
<td>0.7002</td>
<td>0.7047</td>
<td>0.7050</td>
<td>0.6981</td>
<td>0.7053</td>
<td>0.7014</td>
<td>0.7052</td>
<td>0.7024</td>
<td>0.6999</td>
<td>0.7037</td>
<td>0.7089</td>
</tr>
<tr>
<td></td>
<td>Kappa</td>
<td>0.4152</td>
<td>0.4623</td>
<td>0.437</td>
<td>0.4136</td>
<td>0.4589</td>
<td>0.4584</td>
<td>0.4575</td>
<td>0.4290</td>
<td>0.4160</td>
<td>0.4617</td>
<td>0.4697</td>
</tr>
<tr>
<td>KSC</td>
<td>ACC</td>
<td>0.5134</td>
<td>0.4666</td>
<td>0.3697</td>
<td>0.4964</td>
<td>0.5248</td>
<td>0.4785</td>
<td>0.5072</td>
<td>0.5254</td>
<td>0.4826</td>
<td>0.5154</td>
<td>0.5562</td>
</tr>
<tr>
<td></td>
<td>NMI</td>
<td>0.5536</td>
<td>0.5167</td>
<td>0.4098</td>
<td>0.5699</td>
<td>0.5695</td>
<td>0.5587</td>
<td>0.4900</td>
<td>0.5970</td>
<td>0.5621</td>
<td>0.5966</td>
<td>0.5984</td>
</tr>
<tr>
<td></td>
<td>Purity</td>
<td>0.5623</td>
<td>0.5598</td>
<td>0.4804</td>
<td>0.6287</td>
<td>0.5810</td>
<td>0.6204</td>
<td>0.4821</td>
<td>0.6467</td>
<td>0.6223</td>
<td>0.6267</td>
<td>0.6459</td>
</tr>
<tr>
<td></td>
<td>Kappa</td>
<td>0.4577</td>
<td>0.4086</td>
<td>0.2992</td>
<td>0.4471</td>
<td>0.4771</td>
<td>0.4257</td>
<td>0.2996</td>
<td>0.4793</td>
<td>0.4432</td>
<td>0.4709</td>
<td>0.5036</td>
</tr>
<tr>
<td>Botswana</td>
<td>ACC</td>
<td>0.5558</td>
<td>0.4542</td>
<td>0.5355</td>
<td>0.5329</td>
<td>0.5711</td>
<td>0.5820</td>
<td>0.5811</td>
<td>0.5972</td>
<td>0.5721</td>
<td>0.6000</td>
<td>0.6000</td>
</tr>
<tr>
<td></td>
<td>NMI</td>
<td>0.6569</td>
<td>0.5202</td>
<td>0.6104</td>
<td>0.6293</td>
<td>0.6633</td>
<td>0.6580</td>
<td>0.6671</td>
<td>0.6551</td>
<td>0.6598</td>
<td>0.6855</td>
<td>0.6855</td>
</tr>
<tr>
<td></td>
<td>Purity</td>
<td>0.6271</td>
<td>0.4987</td>
<td>0.5913</td>
<td>0.5890</td>
<td>0.6367</td>
<td>0.6387</td>
<td>0.6361</td>
<td>0.6417</td>
<td>0.6219</td>
<td>0.6261</td>
<td>0.6545</td>
</tr>
<tr>
<td></td>
<td>Kappa</td>
<td>0.5198</td>
<td>0.4099</td>
<td>0.4976</td>
<td>0.4950</td>
<td>0.5362</td>
<td>0.5480</td>
<td>0.5469</td>
<td>0.5647</td>
<td>0.5338</td>
<td>0.5370</td>
<td>0.5680</td>
</tr>
<tr>
<td>DC</td>
<td>ACC</td>
<td>0.5123</td>
<td>0.5626</td>
<td>0.5392</td>
<td>0.5111</td>
<td>0.5658</td>
<td>0.5543</td>
<td>0.5359</td>
<td>0.5299</td>
<td>0.5130</td>
<td>0.5627</td>
<td>0.5673</td>
</tr>
<tr>
<td></td>
<td>NMI</td>
<td>0.5416</td>
<td>0.5539</td>
<td>0.5513</td>
<td>0.5393</td>
<td>0.5524</td>
<td>0.5509</td>
<td>0.5527</td>
<td>0.5451</td>
<td>0.5391</td>
<td>0.5543</td>
<td>0.5618</td>
</tr>
<tr>
<td></td>
<td>Purity</td>
<td>0.7002</td>
<td>0.7047</td>
<td>0.7050</td>
<td>0.6981</td>
<td>0.7053</td>
<td>0.7014</td>
<td>0.7052</td>
<td>0.7024</td>
<td>0.6999</td>
<td>0.7037</td>
<td>0.7089</td>
</tr>
<tr>
<td></td>
<td>Kappa</td>
<td>0.4152</td>
<td>0.4623</td>
<td>0.4370</td>
<td>0.4136</td>
<td>0.4589</td>
<td>0.4584</td>
<td>0.4575</td>
<td>0.4290</td>
<td>0.4160</td>
<td>0.4617</td>
<td>0.4697</td>
</tr>
</tbody>
</table>

Fig. 2. Clustering maps of a part of the Indian Pines dataset obtained by different methods. (a) Ground truth. (b) LapScore. (c) MVPCA. (d) E_FDPC. (e) RSR. (f) SOP-SRL. (g) EGCSR. (h) ONR. (i) FNGBS. (j) DSNMF. (k) RDGSR.

Fig. 3. Clustering performance comparison of different methods w.r.t. different numbers of selected bands. (a) Indian Pines. (b) PaviaU. (c) KSC. (d) Botswana. (e) DC.

TABLE III
OPTIMAL CLUSTERING ACCURACY COMPARISON OF DIFFERENT METHODS ON HSI DATASETS

<table>
<thead>
<tr>
<th>Dataset</th>
<th>AllBands</th>
<th>LapScore</th>
<th>MVPCA</th>
<th>E_FDPC</th>
<th>RSR</th>
<th>SOP SRL</th>
<th>EGCSR</th>
<th>ONR</th>
<th>FNGBS</th>
<th>DSNMF</th>
<th>RDGSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Indian Pines</td>
<td>0.3367</td>
<td>0.3320</td>
<td>0.3173</td>
<td>0.3443</td>
<td>0.3216</td>
<td>0.3364</td>
<td>0.3324</td>
<td>0.3380</td>
<td>0.3535</td>
<td>0.3582</td>
<td></td>
</tr>
<tr>
<td>PaviaU</td>
<td>0.5134</td>
<td>0.4809</td>
<td>0.5096</td>
<td>0.5024</td>
<td>0.5248</td>
<td>0.5206</td>
<td>0.5202</td>
<td>0.5473</td>
<td>0.5026</td>
<td>0.5167</td>
<td>0.5562</td>
</tr>
<tr>
<td>KSC</td>
<td>0.5358</td>
<td>0.5563</td>
<td>0.5655</td>
<td>0.5905</td>
<td>0.5755</td>
<td>0.5820</td>
<td>0.5923</td>
<td>0.6024</td>
<td>0.6002</td>
<td>0.5972</td>
<td>0.6224</td>
</tr>
<tr>
<td>Botswana</td>
<td>0.6389</td>
<td>0.6528</td>
<td>0.6521</td>
<td>0.5933</td>
<td>0.6521</td>
<td>0.6481</td>
<td>0.6310</td>
<td>0.6410</td>
<td>0.6764</td>
<td>0.6492</td>
<td>0.6691</td>
</tr>
</tbody>
</table>

To quantify the redundancy of the selected bands, we report the redundancy rate measured by \( R_e = (1/r(r-1)) \sum_{i,j \in U, i < j} \rho_{ij} \) where \( U \) is the set of the selected \( r(r = 20) \) bands and \( \rho_{ij} \) is the correlation between the \( i \)th and \( j \)th bands. The higher value of \( R_e \) means that there is much more redundancy existing in the selected bands. The redundancy rates of different methods on the KSC and DC datasets are presented in Table IV. The results shown in Table IV demonstrate that the band subset selected by RDGSR contains much less redundancy than other methods.

Authorized licensed use limited to: Universidade de Macau. Downloaded on September 30, 2022 at 23:58:41 UTC from IEEE Xplore. Restrictions apply.
TABLE IV
REDUNDANCY RATES OF DIFFERENT METHODS ON THE KSC AND DC DATASETS. (THE LOWER THE BETTER)

<table>
<thead>
<tr>
<th>Dataset</th>
<th>LapScore</th>
<th>MVPCA</th>
<th>E_PDFC</th>
<th>RSR</th>
<th>SOP-SRL</th>
<th>EGCSR</th>
<th>ONR</th>
<th>FNGBS</th>
<th>DSNMF</th>
<th>RDGSR</th>
</tr>
</thead>
<tbody>
<tr>
<td>KSC</td>
<td>0.2924</td>
<td>0.0510</td>
<td>0.3126</td>
<td>0.0386</td>
<td>0.4370</td>
<td>0.0436</td>
<td>0.2111</td>
<td>0.3137</td>
<td>0.2794</td>
<td>0.0383</td>
</tr>
<tr>
<td>DC</td>
<td>0.3717</td>
<td>0.3979</td>
<td>0.3794</td>
<td>0.3376</td>
<td>0.3783</td>
<td>0.3513</td>
<td>0.3704</td>
<td>0.3694</td>
<td>0.3392</td>
<td>0.3372</td>
</tr>
</tbody>
</table>

Fig. 4. Distribution of the top 20 spectral bands selected by different methods. (a) KSC. (b) DC.

Corresponding to Table IV and Fig. 3 with 20 selected bands, we also visualize the distribution of the bands selected by different methods from the KSC and DC datasets in Fig. 4. For RDGSR, the latter part of spectral bands is more effective for the KSC dataset, while the previous part of spectral bands is more useful for the DC dataset. From Fig. 4, we can easily find more details for different methods.

Additionally, we also evaluate the performance of all methods by selecting different numbers of spectral bands \( r \in \{5, 10, \ldots, 50\} \) and then present the best clustering accuracy results of all methods. The corresponding results are depicted in Fig. 3 and Table III. Since we use a grid-search strategy to tune the parameters \( S \) and \( \lambda \) in the proposed RDGSR, we found that the optimal values of them to achieve the best clustering accuracy may be distinctive for different numbers of selected bands on different datasets. This is because different HSI datasets show different characteristics in spectral bands, land covers and spatial distribution. We report the values of the parameters \( S \) and \( \lambda \) for all situations in Table V. From Table III, it is easily found that the proposed RDGSR is superior to all compared methods except for FNGBS on the DC dataset. As shown in Fig. 3, the proposed RDGSR consistently outperforms other compared methods with different numbers of selected bands in most cases. When \( r \) increases, RDGSR first exhibits the increasing performance and then displays stable or even degraded results. For the Indian Pines dataset shown in Fig. 3(a), RDGSR achieves superior performance with a small \( r \) and the optimal result is obtained with \( r = 15 \).

When \( r = 20 \), the performance of RDGSR reduces compared to the situation of \( r = 15 \). This is because the most informative bands are selected in the first 15 spectral bands and the less informative or noisy bands are introduced in the latter five spectral bands. When selecting different numbers of bands, the combinations of different band subsets may result in different results. It is possible that the combination of some spectral bands may cause a negative effect, while the combination of some other spectral bands may not. Similar problems also exist in the PaviaU dataset shown in Fig. 3(b), but the situation described above is not obvious in the KSC and Botswana dataset shown in Fig. 3(c) and (d). The variance of different situations may be attributed to the characteristics of different datasets. For the KSC dataset, compared to the second best method RSR, RDGSR gains at least 7.23% enhancement by averaging the results from different numbers of selected bands. The best result of RDGSR on the KSC dataset is achieved with \( r = 20 \). For the Botswana dataset, RDGSR needs a large \( r \) to get significant performance compared to other methods. When \( r = 45 \), the performance of RDGSR is the best. The reason why RDGSR outperforms other compared methods is that the geometric structure of pixels with spatial consistency and the geometric structure of spectral bands are both considered in the learning model to achieve robust band selection with the \( \ell_{2,1} \)-norm learning scheme. Thus, RDGSR specializes in selecting informative bands.

D. Statistical Test
To better illustrate the performance of the proposed RDGSR, we use a two-tailed \( t \)-test with 95% confidence level.
to understand if the differences between RDGSR and other compared methods are statistically significantly important. The $t$-test results with respect to clustering accuracy are shown in Table VI. When the $p$-value of the two compared methods is less than 0.05, we consider that there is a significant difference between them. According to Tables II and VI, we found that RDGSR demonstrates better performance with a significantly important difference than other methods on the Indian Pines dataset, except for EGCSR. For the KSC dataset, RDGSR also shows a significantly important difference compared to other methods except for RSR and ONR. For the remaining three datasets, RDGSR still exhibits a significantly important difference compared to some other methods. In summary, the proposed RDGSR shows its statistically significant performance in most cases.

### E. Efficiency Analysis

To better illustrate the efficiency performance, we first analyze the computational complexity and then record the CPU running time for all compared methods. The time complexity comparison is provided in Table VII. In HSI analysis, the number of spectral bands $B$ is always much smaller than the number of pixels $N$. Therefore, MVPCA, E_FDPC, EGCSR, ONR and FNGBS are much more efficient than LapScore, RSR, SOP-SRL, and RDGSR. Compared to LapScore, RSR and RDGSR are less efficient due to the required iterative learning. For SOP-SRL, the meaning of the notations in the equation of the time complexity can be referred to $[47]$. For DSNMF, $C$ denotes the number of clusters.

We conduct the experimental studies in the same computing environment (MATLAB R2016a on a workstation with Intel Core 3.00 GHZ CPU and 64 GB RAM) and record the execution time of all compared methods. The running time results to achieve band selection are presented in Table VIII.

From Table VIII, we can see that MVPCA, E_FDPC, EGCSR, ONR, and FNGBS are much more efficient due to the non-iterative learning for band selection. By contrast, LapScore, SOP-SRL, and our RDGSR require slightly more computing resources to select informative spectral bands. It is clear that RSR and DSNMF is the most time-consuming method for band selection. The reason requiring more running time for these methods is the construction of the similarity graph matrix at pixel level or the iterative learning. As can be observed, the running time of RDGSR on the five datasets is less than 20 s. It is negligible and acceptable in practice. In summary, RDGSR demonstrates superior performance with a good balance of computational efficiency.

### F. Robustness Study

To evaluate the robustness of the proposed RDGSR, we investigate the performance variation at different noise levels. We corrupt the Indian Pines dataset with noise levels of $[5\%, 10\%, 20\%]$, respectively. This means $[5\%, 10\%, 20\%]$ data are randomly chosen and reallocated random values to be noisy data in the experiment. The optimal clustering accuracy results of different methods at different noise levels on the Indian Pines dataset are reported in Fig. 5. Compare to the results without noises in Table III, it is natural that the performance is degraded when the noise increases on the dataset. For RDGSR, the performance is decreased from 0.3741 to 0.3599, 0.3495, and 0.3360 when different levels of noise are introduced. Compared to other methods, RDGSR still obtains the best performance when conducting band selection on noisy data. It can be concluded that RDGSR is robust in a noisy environment.

### G. Model Analysis

1) Ablation Study: The proposed RDGSR is realized by jointly exploiting structural information from both segmentation regions and bands in a robust $\ell_{2,1}$-norm manner. To test the effectiveness of these components, we develop three variants of RDGSR for ablation study. RDGSR-P is the variant that utilizes only the structural graph with spatial consistency from the segmented pixels. RDGSR-B denotes the variant that utilizes only the structural graph from spectral bands. Dual graph self-representation (DGSR) indicates the variant with dual graph fusion that utilizes the Frobenius norm for the loss function. Besides, RSR is also introduced to test the power of structural information. The experimental results with
different numbers of selected bands \( r \in \{10, 20, 30, 40, 50\} \) are depicted in Fig. 6. As shown in Fig. 6, the values of the parameters \( S \) and \( \lambda \) to achieve the best results for different numbers of selected bands can be also found from Table V. RDGSR is superior to other methods in most cases. Compared to RSR, RDGSR-P, RDGSR-B, and DGSR, RDGSR achieves 6.05\%, 1.52\%, 3.84\%, and 2.19\% average improvements, respectively. This demonstrates the effectiveness of our overall model RDGSR with the capability of jointly exploiting structural information with spatial consistency among superpixels and structural information among spectral bands in a robust \( \ell_{2,1}\)-norm manner.

2) Parameter Sensitivity Study: There are two important parameters in RDGSR needed to be set in advance, i.e., the number of segmentation \( S \) and the regularization parameter \( \lambda \). To study the parameter sensitivity of RDGSR, we fix the number of selected bands as 20 and get the performance variation by varying \( S \) from \( \{3, 7, 15, 25, 50, 100, 200, 500, 1000\} \) and \( \lambda \) from \( \{10^{-3}, 10^{-2}, \ldots, 10^{3}\} \). The experimental results are reported in Fig. 7. As can be observed, different combinations of \( S \) and \( \lambda \) lead to different performances. The performance of RDGSR is inferior when the value of \( S \) is too large or too small. It is clear that a moderate value of \( \lambda \) is more favorable to the experiment. Given a moderate value of \( \lambda \), the performance of RDGSR varies according to the setting of \( S \). When fixing \( S \) as a moderate value, the performance of RDGSR first increases and then becomes stable or even decreases. Superior performance with a proper value of \( S \) means that potential spatial information inherent in HSIs can be accurately exploited by superpixel segmentation technique. According to the above discovery, it is necessary to allocate appropriate values for \( S \) and \( \lambda \) in RDGSR.

V. DISCUSSIONS

In this section, we discuss the relationships and differences between the proposed RDGSR and the most related methods, including RSR [41], EGCSR [28] and DSNMF [56].

A. Connection With RSR [41]

As robust learning methods, RSR and our RDGSR both adopt the \( \ell_{2,1}\)-norm to characterize the loss function that measures the discrepancy between two entries. However, the knowledge used in the two learning methods is different. RSR only uses spectral information to conduct band selection, while our RDGSR jointly utilizes spectral, spatial, and geometric information to promote unsupervised band selection. Specifically, our RDGSR extracts local spatial information from segmentation regions and records structural information in dual similarity graphs. Compared to RSR, our RDGSR is more capable of identifying discriminative bands by jointly using spectral and spatial information in the non-Euclidean domain.

B. Connection With EGCSR [28]

As graph-based band selection methods, EGCSR and our RDGSR utilize both spectral and structural information to select discriminative bands. However, their graphs to record structural information are different. EGCSR only uses the graph constructed from bands to exploit the relationships between adjacent spectral bands. Apart from the band-based similarity graph, our RDGSR also utilizes the graph constructed from multiple homogenous segmentations to explore the relationship between neighboring pixels in the same superpixel. Besides, our RDGSR adopts the \( \ell_{2,1}\)-norm to characterize the loss function that is less sensitive to outliers and noises than the Frobenius norm used in the loss function of EGCSR. Compared to EGCSR, our RDGSR is competent to realize high-quality band selection due to the joint exploration of structural graphs for region-wise pixels and spectral bands and the \( \ell_{2,1}\)-norm regularization.

C. Connection With DSNMF [56]

As dual-graph learning methods, DSNMF and our RDGSR utilize dual graphs constructed from both pixels and bands.
to steer band selection. However, their graph integration strategies are different. DSNMF employs a graph Laplacian regularization strategy to explore structural information in the learning model, while our RDGSR adopts a graph convolution strategy to embed structural information in the learning model. Besides, our RDGSR uses superpixel segmentation to extract more accurate spatial information from homogeneous regions in the exploration of robust band selection. Compared to DSNMF, our RDGSR is promising to select representative spectral bands with potential spatial information in a robust $\ell_{2,1}$-norm manner.

VI. CONCLUSION

In this article, we proposed an RDGSR method to achieve unsupervised hyperspectral band selection. RDGSR constructs the superpixel-based similarity graph and band-based similarity graph from the segmented HSI to explore spatial and structural information among both pixels and spectral bands in a robust manner. Using the dual structural graphs constructed from superpixels and spectral bands, RDGSR extends the self-representation model to non-Euclidean domain with the help of graph convolution. Introducing the $\ell_{2,1}$-norm in the loss function and regularization term, RDGSR is robust and good at band selection to eliminate the noise at row level. To solve RDGSR, we design an effective optimization algorithm as the solution. Extensive experiments on five HSI datasets have demonstrated the superiority of RDGSR over state-of-the-art methods. In our future work, we will investigate effective band selection methods with an adaptive parameter learning scheme that can avoid manual parameter determination. Besides, we will also investigate efficient and scalable graph-based methods for large-scale HSI datasets in the future.

REFERENCES


