Marginalized 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 is an essential step in preprocessing hyperspectral images (HSIs) to select informative bands. Most existing methods exploit the spatial information from the entire HSI while ignoring the difference between diverse homogeneous regions. Moreover, traditional methods utilize the limited size of data for model training that may result in degraded generalization performance. In this article, we propose a marginalized graph self-representation (MGSR) method for unsupervised hyperspectral band selection. To explore the spatial information from diverse homogenous regions, MGSR generates the segmentations of an HSI by superpixel segmentation and records the relationships between adjacent pixels of the same segmentation in a structural graph. Meanwhile, to improve the generalization and robustness, infinite corrupted samples are obtained from the original pixels by introducing noises in spectral bands for model training. To solve the proposed formulation, we design an alternating optimization algorithm to marginalize out the corruption and search for the optimal solution. Experimental studies on HSI datasets demonstrate the effectiveness of the proposed MGSR and the superiority over the state-of-the-art methods. The source code is available at https://github.com/ZhangYongshan/MGSR.**

***Index Terms*— Graph convolution, hyperspectral band selection, marginalized corruption, self-representation (SR), unsupervised learning.**

I. Introduction

With advances in spectral imaging technology, hyperspectral image (HSI) data are captured by different spaceborne or airborne sensors. Using hundreds of thousands of spectral wavelengths, HSIs typically record a scene from a specified observation area and provide abundant information to distinguish the subtle differences among land-cover materials. Due to the rich spectral signatures, HSIs become prevalent and provide a new perspective to solve earth observation tasks [1]–[5]. With the increasing number of spectral channels, HSI data are represented by high-dimensional spectral bands. The high dimensionality of HSI data leads to two challenges in learning tasks, i.e., the heavy burden of computation and storage and the curse of dimensionality problem. To alleviate these problems, it is necessary to conduct dimensionality reduction to preprocess HSI data [6]–[8].

In general, dimensionality reduction of HSI data can be divided into two categories: feature extraction [9]–[11] and band selection (i.e., feature selection) [12]–[15]. The former one endeavors to learn a low-dimensional feature representation via an optimal transformation of the original spectral bands, while the latter one aims to discriminate an appropriate subset of informative bands from the original ones. Feature extraction generates a discriminative representation from a new feature space that differs from the entire spectral bands, while band selection selects the existing bands from the original band set without any transformation. In this article, our study focuses on band selection because the intrinsic data characteristic can be preserved from the selected bands.

According to the availability of label information, band selection can be realized in supervised or unsupervised manner [16]. Considering the difficulty of obtaining sufficient labels, unsupervised band selection is more flexible and applicable for HSIs without using any prior knowledge. In recent decades, researchers have proposed a number of unsupervised band selection methods. Representative methods are maximum-variance principal component analysis (MVPCA) [17], sparse-based band selection (SpaBS) [18], improved sparse subspace clustering (ISSC) [19], and fast density-peak-based clustering (FDPC) [20]. As a popular technique, self-representation (SR) utilizes the self-expressive property of data and various regularization terms to fulfill band selection [21], [22]. Representative methods include dissimilarity-weighted sparse self-representation (DWSSR) [23], fast and robust self-representation (FRSR) [24], and efficient graph convolutional self-representation (EGCSR) [25]. Deep learning is also a well-suited technique to fulfill band selection. In [26], an attention-based convolutional neural network is coupled with anomaly detection to select informative bands. To achieve stable results, a concrete autoencoder identifies discriminative bands with the help of Gumbel-Softmax trick and entropy constraint [27].

Manuscript received July 20, 2021; revised September 28, 2021; accepted October 17, 2021. Date of publication October 21, 2021; date of current version February 8, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62106241, Grant 61703355, and Grant 62073300; in part by the Natural Science Foundation of Hubei Province of China under Grant 2020CFB328; in part by the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) under Grant CUG200116; and in part by the University of Macau (File no. MYRG2018-00136-FST). *(Corresponding author: Yicong Zhou.)*

Yongshan Zhang is with the School of Computer Science, China University of Geosciences, Wuhan 430074, China, and also with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: yszhang.cug@gmail.com).

Xinxin Wang and Yicong Zhou are with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: yicongzhou@um.edu.mo).

Xinwei Jiang is with the School of Computer Science, China University of Geosciences, Wuhan 430074, China (e-mail: ysjxw@hotmail.com).

Digital Object Identifier 10.1109/TGRS.2021.3121671

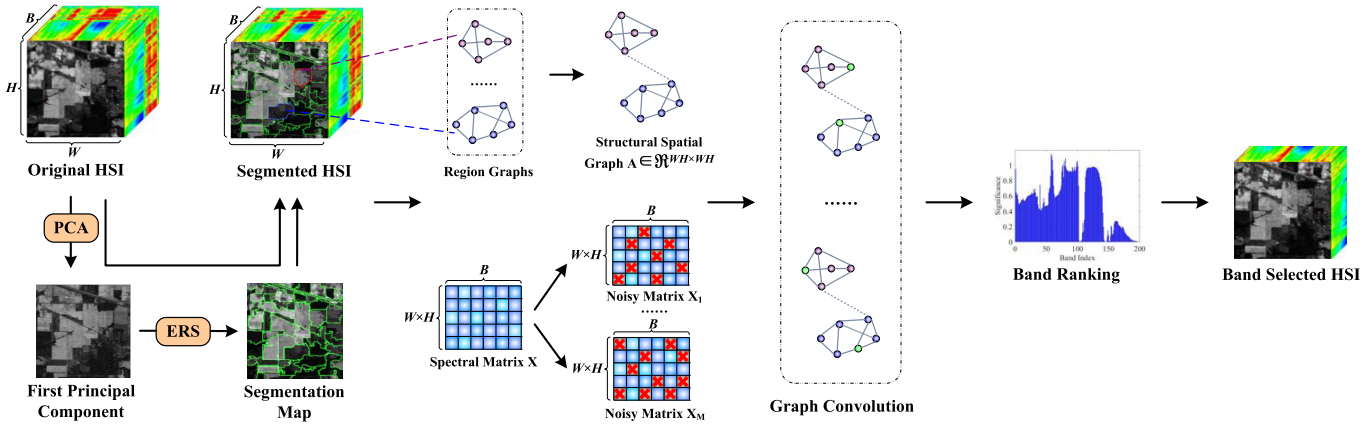

Fig. 1. Flowchart of the proposed MGSR method. The original HSI is first preprocessed by PCA to get the first principal component, and then, the ERS segmentation method localizes the homogeneity of an HSI to generate segmentation map for homogenous regions. With the segmentation information, multiple graphs are constructed from different regions to record the spatial information and constitute an entire structural graph. After that, the spectral matrix is corrupted by noises multiple times to generate infinite samples. Taking the structural spatial graph and corrupted and uncorrupted spectral matrices as inputs, band selection is achieved accompanied with graph convolution. Finally, top-ranked bands are selected from the original band set by measuring the significance of bands.

In addition, subspace clustering [28], dual attention mechanism [29], and reinforcement learning [30] are also considered into deep learning-based band selection.

Although the above-mentioned methods gain success in HSI analysis, they utilize only the spectral signatures while neglecting the spatial information that can further improve the performance of learning models. Apart from the spectral signatures, the rich spatial information is also inherent in HSIs. In [13], the Laplacian graph recording the spatial information is incorporated into the robust principal components analysis. To promote band selection, the spatial information is integrated into the hypergraph to measure the similarity between the adjacent pixels [31]. For simultaneous feature selection and extraction, the method in [32] learns a spectral–spatial representation by the transformation of the most significant original bands. In [33], multiple spectral–spatial features are utilized as the complementarity via multimodal learning for better band selection. It has been proven that the spatial information is beneficial to accurate identification of HSIs [34]–[37].

A. Motivation and Contributions

In spectral–spatial methods, there are still at least two limitations that hinder the performance of band selection. First, most exiting methods take an HSI as a whole to extract the spatial information to facilitate band selection [13], [31], [38]. In reality, the spatially neighboring pixels always show similar spectral signatures and have the same land cover with a large probability in a homogenous region. The homogenous regions are mainly constructed by numerous pixels with the same land cover and show different reflection characteristics of objects. However, the difference of homogeneous regions in an HSI is usually ignored. A reasonable solution is to jointly extract the spatial information from different regions by superpixel segmentation [9], [10], [39]. Second, various conventional models are generally developed on the original inputs and trained with the limited size of data. When using the limited data to train learning models, it is inevitable that the generalization capability to unknown data will be compromised. To improve the generalization and robustness of a model, infinite corrupted pixels with introduced noises should be taken into consideration [40]–[42].

Motivated by these, in this article, we propose a marginalized graph self-representation (MGSR) method to select discriminative bands in an unsupervised fashion. The flowchart of the proposed MGSR method is shown in Fig. 1. Specifically, MGSR employs superpixel segmentation to generate diverse homogeneous regions of an HSI and record the corresponding spatial information of homogeneous regions in a structural graph. The original pixels are then corrupted to generate infinite samples by introducing noises in spectral bands for model training. MGSR is effective and robust for unsupervised band selection by simultaneously using the structural spatial information from the non-Euclidean domain and infinite corrupted pixels. Accordingly, we present an efficient alternating optimization algorithm for the proposed MGSR to marginalize out the corruption and find the optimal solution. Substantial experiments on HSI datasets demonstrate the superiority of the proposed MGSR.

In general, the contributions of this article are summarized as follows.

1) We proposed an MGSR method to perform unsupervised band selection directly using the structural spatial information from the non-Euclidean domain and infinite corrupted pixels.
2) MGSR extracts the spatial information of homogeneous regions in a structural graph by superpixel segmentation and marginalizes out the corruption without introducing explicit noises.
3) We present an efficient alternating iterative algorithm to solve the optimization problem of the proposed MGSR and achieve the optimal solution.
4) Experimental studies on HSI datasets demonstrate the effectiveness of the proposed MGSR and the superiority over the state-of-the-art methods.

B. Organization

The rest of this article is structured as follows. Section II reviews the related work about SR and marginalized corruption. Section III introduces the proposed MGSR method for unsupervised hyperspectral band selection and presents an optimization algorithm as a solution. The discussion of the proposed method is provided in Section IV. Experimental studies are conducted in Section V. Finally, Section VI concludes this article.

II. Related Work

A. Self-Representation

SR makes good use of the self-expressive property to model the data itself [43]–[45]. The interpretation of SR is that each feature can be represented by linearly weighting all features. Given the data matrix $\mathbf{X} = \{\mathbf{x}_1; \mathbf{x}_2, \ldots, \mathbf{x}_N\} \in \mathbb{R}^{N\times B}$ with N samples and B features, the SR can be formulated as

$$\min_{\mathbf{W}} \mathcal{L}(\mathbf{X} - \mathbf{X}\mathbf{W}) + \lambda \mathcal{R}(\mathbf{W}) \tag{1}$$

where $\mathbf{W} \in \mathbb{R}^{B\times B}$ is the feature weight matrix, $\mathcal{L}(\mathbf{X}-\mathbf{X}\mathbf{W})$ is the loss function, and $\mathcal{R}(\mathbf{W})$ is the regularization term using parameter λ to trade off. The Frobenius norm is commonly used in the loss function to measure the minimum reconstruction error. The regularization term is used to achieve an explicit purpose, such as the Frobenius norm for regular purpose [46], $\ell_{2,1}$-norm for row sparsity [47], and low-rank constraint for low-rank representation [48]. By imposing the $\ell_{2,1}$-norm regularization on $\mathbf{W}$, feature selection can be achieved with the row sparsity information.

In the recent work, Zhu *et al.* [43] presented a regularized SR model with an $\ell_{2,1}$-norm loss function to characterize the reconstruction error and select informative features. To perform robust feature selection, Hu *et al.* [49] integrated the graph regularization term into the SR for stable performance and easy interpretation. Based on the work in [43], Tang *et al.* [45] extended the SR with the dual concept to take advantage of both the sample and feature self-expressive property for feature selection. To capture the high-order relationship of data, He *et al.* [44] proposed an advanced SR model integrated with the hypergraph regularizer and low-rank constraint. Although the above-mentioned methods achieve promising results, they are only applicable to the Euclidean domain or use the graph information as an additional regularization term.

Different from the previous work, our method extends the SR to the non-Euclidean domain. Instead of treating the graph information as a regularization term, our method provides a natural formulation for unsupervised band selection by directly exploring the relationships and interactions between the spectral bands and structural graph with spatial consistency.

B. Marginalized Corruption

Denoising autoencoders take the randomly corrupted samples as inputs to recover the original undistorted ones by minimizing their reconstruction error [40]–[42]. A robust representation can be achieved by stacking multiple denoising autoencoders in a layer-wise fashion. The denoising process enables learning models more stable and robust. Generally, more corrupted samples always yield robust and better performance. It is inevitable that the computational cost is a burden. Marginalized corruption is an efficient way to mimic the infinite data corruption without introducing explicit corrupted samples [50]. It has shown a great potential in unsupervised learning tasks [40]–[42], [50].

Recently, to develop a robust predictor, Maaten *et al.* [41] presented a marginalized corrupted features learning model with infinitely corrupted samples by minimizing their expected value under the corrupting distribution. For graph clustering, Wang *et al.* [50] proposed a marginalized graph autoencoder by exploiting both the content and structure information to discover community structures from the networked data. To achieve robust multiview clustering, Tao *et al.* [42] presented a marginalized multiview ensemble clustering method to seek the consensus partitions from the basic partitions of multiple views. Marginalized corruption in these methods avoids introducing explicit noises and yields a robust and promising result. However, they only focus on unsupervised domain adaption and clustering.

In this work, we introduce marginalized corruption into the SR model in the non-Euclidean domain for unsupervised band selection. The integration of marginalized corruption and SR in the non-Euclidean domain fully exploits the interactions between the spectral bands and structural graph with spatial consistency. It is beneficial to select the most discriminant bands from the original ones.

III. Methodology

Given a hyperspectral data cube $\mathcal{X} \in \mathbb{R}^{W\times H\times B}$ recording a scene with $W \times H$ pixels and B spectral bands, the goal of unsupervised band selection is to distinguish a subset of informative bands without supervision. In this section, we will elaborate on the proposed method, called MGSR, for unsupervised band selection.

A. Region-Wise Similarity Graph Construction

For a hyperspectral data cube, the spectral information is contained in bands, while the spatial information is included in pixels. As can be seen from the false-color images of HSIs in Figs. 3–5(a), it is notable that different regions show different characteristics of different land covers. This also indicates that the spatial information is different in diverse regions. In general, the adjacent pixels with a high probability to the same land cover often show similar spectral characteristics. This situation is more evident if the adjacent pixels are in the same homogenous region. However, such spatial information may be ignored in most existing methods.

In light of this, we adopt the entropy rate superpixel (ERS) segmentation method [51] to localize the homogeneity of an HSI for the promising performance in both efficacy and efficiency. Thus, the spatial information with homogeneous consistency can be effectively extracted by ERS. Specifically, we first reshape the 3-D hyperspectral data to a 2-D spectral matrix $\mathbf{X} \in \mathbb{R}^{N\times B}(N = W \times H)$ and then perform a principal

component analysis (PCA) [9], [10] to obtain the first principal component $I_f \in \mathbb{R}^{W\times H}$. This operation captures the primary knowledge of an HSI and lessens the computational burden of superpixel segmentation. To generate S homogeneous segmentations, we conduct ERS on I_f as follows:

$$I_f = \bigcup_{s=1}^{S} \mathcal{H}_s, \quad \text{s.t. } \mathcal{H}_s \cap \mathcal{H}_g = \varnothing, \ (s \neq g) \tag{2}$$

where $\mathcal{H}_s$ is the sth segmentation. ERS is originally designed for RGB image segmentation. During the segmentation, ERS first converts RGB images to grayscale images and then conducts segmentation on the grayscale images, that is to say, segmentation using ERS is conducted on one component instead of more components. When there are more components in an image, it is necessary to get the only one component before segmentation. Therefore, we only use the first principal component with the primary knowledge of HSI for segmentation. This operation is common in the HSI analysis [9], [10].

To record the spatial information in a graph structure $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, we treat each pixel as a node of the graph $\upsilon_i \in \mathcal{V}$ and relationship between pair-wise pixels $\mathbf{x}_i$ and $\mathbf{x}_j$ as an edge of the graph $e_{ij} \in \mathcal{E}$. Therefore, the graph of an HSI can be represented by the spectral matrix $\mathbf{X}$ to denote the spectral information and the similarity graph $\mathbf{A} \in \mathbb{R}^{N\times N}$ to represent the structural spatial information. Using the segmentation information as in (2), the similarity graph $\mathbf{A}$ of an HSI [52] is defined as

$$A_{ij} = \begin{cases} \exp\left(-\dfrac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), & \mathbf{x}_i, \mathbf{x}_j \in \mathcal{H}_s, \\ 0, & \mathbf{x}_i \in \mathcal{H}_s \text{ and } \mathbf{x}_j \in \mathcal{H}_g. \end{cases} \tag{3}$$

In detail, the similarity of pair-wise pixels within the same segmentation is calculated by $\exp(-(\|\mathbf{x}_i - \mathbf{x}_j\|^2/2\sigma^2))$, while the similarity of pair-wise pixels from different homogeneous segmentations is set to be zero. The similarity graph $\mathbf{A}$ records the affinities between pair-wise pixels that represent the structural information with spatial consistency. With the spectral matrix $\mathbf{X}$ and similarity graph $\mathbf{A}$, the band selection problem is extended into the non-Euclidean domain.

B. Marginalized Graph Self-Representation

SR makes use of the self-expressive property to reconstruct the original input data [23], [24]. Most existing methods were generally designed for Euclidean domain. They can effectively handle the problems with Euclidean data but not for non-Euclidean data. To handle the non-Euclidean hyperspectral data with the spectral matrix $\mathbf{X}$ and similarity graph $\mathbf{A}$, we formulate the graph SR model by jointly integrating the spectral and spatial information as

$$\min_{\mathbf{W}} \|\mathbf{X} - \hat{\mathbf{A}}\mathbf{X}\mathbf{W}\|_F^2 + \lambda\mathcal{R}(\mathbf{W}) \tag{4}$$

where $\hat{\mathbf{A}} = \widetilde{\mathbf{D}}^{-(1/2)}\widetilde{\mathbf{A}}\widetilde{\mathbf{D}}^{-(1/2)}$ is the normalized similarity graph matrix, $\mathbf{W} \in \mathbb{R}^{B\times B}$ is a trainable weight matrix, and $\mathcal{R}(\cdot)$ is the regularization term with parameter λ to trade off. Specifically, $\widetilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ is the similarity graph matrix with self-loops and $\widetilde{\mathbf{D}}$ is the degree matrix with diagonal entries $\widetilde{D}_{ii} = \sum_{j=1}^{N} \widetilde{A}_{ij}$. Instead of simply using the spectral information, the graph SR of (4) provides a natural formulation to effectively and directly integrate both the spectral and the spatial information.

Machine learning is aimed at developing learning models that generalize well to unknown data. This can be only achieved in an ideal situation with infinite or very large training data. However, learning on such data is computationally costly. The more realistic situation is that learning problems are described by finite training data [40], [41]. In this case, a nontrivial solution is to artificially extend the training data. To this end, we corrupt the existing hyperspectral data with noises. By reconstructing noisy inputs to the original ones, the variant of SR can also achieve great success as a denoising autoencoder. More corrupted inputs yield more stable and better performance. Therefore, training on additional corrupted inputs leads to substantially more robust learning models. Suppose that the spectral matrix $\mathbf{X}$ is corrupted by M times, and we can get the corrupted matrices $[\widetilde{\mathbf{X}}_1, \widetilde{\mathbf{X}}_2, \ldots, \widetilde{\mathbf{X}}_M]$ by random feature removal, meaning that each feature is set to 0 with a probability p $(0 < p < 1)$. The size of $\{\widetilde{\mathbf{X}}_m\}_{m=1}^{M}$ and $\mathbf{W}$ is still $N \times B$ and $B \times B$ that is unchangeable during learning. Taking the original spectral matrix and its corruptions as inputs, we can rewrite (4) as

$$\min_{\mathbf{W}} \frac{1}{M}\sum_{m=1}^{M} \|\mathbf{X} - \hat{\mathbf{A}}\widetilde{\mathbf{X}}_m\mathbf{W}\|_F^2 + \lambda\mathcal{R}(\mathbf{W}). \tag{5}$$

The loss function in (5) represents the MGSR to reconstruct the spectral bands from both the random corruptions and structural spatial graph. Since the random corruptions come after superpixel segmentation, the introduced noises have no influence on the generation of homogenous regions.

To fulfill band selection, an $\ell_{2,1}$-norm regularization [47] is required to impose on the weight matrix $\mathbf{W}$. This will result in row sparsity of $\mathbf{W}$ and make it suitable for band selection. Thus, the formulation of the proposed MGSR is finally represented as

$$\min_{\mathbf{W}} \frac{1}{M}\sum_{m=1}^{M} \|\mathbf{X} - \hat{\mathbf{A}}\widetilde{\mathbf{X}}_m\mathbf{W}\|_F^2 + \lambda\|\mathbf{W}\|_{2,1}. \tag{6}$$

The harmonious integration of the corrupted spectral matrix and similarity graph in a unified model benefits the exploitation of the interactions between the spectral and spatial information. It is significantly conducive to unsupervised band selection. With the learned weight matrix $\mathbf{W}$, band selection can be achieved by ranking bands according to $\|\mathbf{w}_i\|_2$ $(i = 1, \ldots, B)$ in a descending order.

C. Solution to MGSR

It is difficult to solve the optimization problem of MGSR in (6) due to the random corruptions and $\ell_{2,1}$ regularization term. For ease of representation, we define an auxiliary matrix $\widetilde{\mathbf{Z}} = (1/M)\sum_{m=1}^{M} \hat{\mathbf{A}}\widetilde{\mathbf{X}}_m$ and relax $\|\mathbf{W}\|_{2,1}$ by $\mathrm{Tr}(\mathbf{W}^T\mathbf{C}\mathbf{W})$. The matrix $\mathbf{C} \in \mathbb{R}^{B\times B}$ is a diagonal matrix with elements $C_{ii} = (1/2\|\mathbf{w}_i\|_2)$. With the above assistance, the optimization of MGSR in (6) can be rewritten as

$$\begin{aligned} \mathcal{J}(\mathbf{W}) &= \|\mathbf{X} - \widetilde{\mathbf{Z}}\mathbf{W}\|_F^2 + \lambda\,\mathrm{Tr}(\mathbf{W}^T\mathbf{C}\mathbf{W}) \\ &= \mathrm{Tr}((\mathbf{X} - \widetilde{\mathbf{Z}}\mathbf{W})^T(\mathbf{X} - \widetilde{\mathbf{Z}}\mathbf{W})) + \lambda\,\mathrm{Tr}(\mathbf{W}^T\mathbf{C}\mathbf{W}). \end{aligned} \tag{7}$$

Algorithm 1 MGSR

Input: Hyperspectral data $\mathbf{X}$, regularized parameter λ, noise probability p, number of segmentation S;
Output: Top r ranked spectral bands;

```
Apply PCA to obtain the first PC of an HSI;
Apply ERS on the first PC to generate S partitions;
Construct the graph matrix A by (3);
Compute the normalized graph matrix Â = D̃^{-1/2} Ã D̃^{-1/2};
Calculate the uncorrupted matrices S = X^T Â^T X and U = X^T Â^T Â X;
Compute the matrices P and Q by (11);
repeat
    Update the weight matrix W by (10);
    Update the matrix C as C_ii = 1/(2||w_i||_2);
until Convergence
Calculate the importance of spectral bands according to ||w_i||_2 (i = 1, ..., B);
```

To update $\mathbf{W}$, we first calculate the derivative of (7) with respect to $\mathbf{W}$ as

$$\frac{\partial \mathcal{J}}{\partial \mathbf{W}} = -2\widetilde{\mathbf{Z}}^T(\mathbf{X} - \widetilde{\mathbf{Z}}\mathbf{W}) + 2\lambda \mathbf{C}\mathbf{W}. \tag{8}$$

Setting the derivative $(\partial \mathcal{L}/\partial \mathbf{W}) = 0$, the update rule of $\mathbf{W}$ is given as

$$\mathbf{W} = (\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{Z}} + \lambda \mathbf{C})^{-1}\widetilde{\mathbf{Z}}^T\mathbf{X}. \tag{9}$$

The update rule of $\mathbf{W}$ in (9) requires the explicit corruptions. To avoid using the explicit corruptions, we concentrate on the limit case $M \to \infty$ so that (9) can be rewritten as its expectation by the weak law of large numbers [40]. Thus, the update rule of $\mathbf{W}$ is reformulated as

$$\mathbf{W} = (\mathbf{Q} + \lambda \mathbf{C})^{-1}\mathbf{P} \tag{10}$$

where $\mathbf{P} = \mathbb{E}[\widetilde{\mathbf{Z}}^T\mathbf{X}]$ and $\mathbf{Q} = \mathbb{E}[\widetilde{\mathbf{Z}}^T\widetilde{\mathbf{Z}}]$ is the expectation with their entries

$$P_{ij} = S_{ij}(1-p) \quad \text{and} \quad Q_{ij} = \begin{cases} U_{ij}(1-p)^2, & \text{if } i \neq j \\ U_{ij}(1-p), & \text{if } i = j. \end{cases} \tag{11}$$

$\mathbf{S} = \mathbf{X}^T\hat{\mathbf{A}}^T\mathbf{X}$ is the uncorrupted matrix of $\widetilde{\mathbf{X}}^T\hat{\mathbf{A}}^T\mathbf{X}$, while $\mathbf{U} = \mathbf{X}^T\hat{\mathbf{A}}^T\hat{\mathbf{A}}\mathbf{X}$ is the uncorrupted matrix of $\widetilde{\mathbf{X}}^T\hat{\mathbf{A}}^T\hat{\mathbf{A}}\widetilde{\mathbf{X}}$. The calculation of $\mathbf{P}$ and $\mathbf{Q}$ is achieved by $\mathbf{S}$ and $\mathbf{U}$ with the survival ratios of the bands. It is clear that the entries of $\mathbf{P}$ share the same surviving probability $(1-p)$ from $\mathbf{x}_i$ and $\widetilde{\mathbf{x}}_i$, while the surviving probability is $(1-p)$ for the diagonal entries of $\mathbf{Q}$ from $\mathbf{x}_i$ and $\widetilde{\mathbf{x}}_i$ and $(1-p)^2$ for other entries of $\mathbf{Q}$ both from $\widetilde{\mathbf{x}}_i$.

Equation (10) provides a closed-form solution to update $\mathbf{W}$. This solution is efficient and easy to implement. Besides, the calculation of $\mathbf{C}$ is also dependent on $\mathbf{W}$. It is necessary to alternately update $\mathbf{W}$ and $\mathbf{C}$ in an iterative learning process. Finally, the objective function value of (6) monotonically decreases due to the optimization solution to MGSR. The pseudocode of MGSR is presented in Algorithm 1.

D. Time Complexity Analysis

The time complexity of the proposed MGSR in Algorithm 1 is analyzed here. The main contribution of the time complexity of MGSR is to learn the optimal weight matrix $\mathbf{W}$. To update $\mathbf{W}$, it takes $\mathcal{O}(B^3)$ for the calculation of matrix inverse and $\mathcal{O}(B^2N)$ for the computation of matrix multiplication. Given the number of iterations T, the overall complexity of MGSR is $\mathcal{O}((B^2N + B^3)T)$.

IV. Discussion

In this section, we provide a deep insight into the proposed MGSR method and analyze the relationships between the proposed MGSR and EGCSR [25].

1) Remark on Our MGSR: Most SR methods for HSIs only use spectral bands to guide band selection. These methods developed for the Euclidean domain neglect the underlying spatial information. In addition, there are some advanced SR methods that utilize the structural information to be a graph regularization term. Different from the graph regularized methods, our MGSR takes full advantage of the spectral and structural information by graph convolution. The direct integration of the spectral bands and structural spatial graph is beneficial to band selection. Similar to EGCSR [25], our MGSR is also developed for unsupervised band selection in the non-Euclidean domain.

2) Comparison With EGCSR [25]: As unsupervised graph-based methods, EGCSR and our MGSR utilize both the spectral bands and structural graph to achieve band selection. However, their structural graphs are different. EGCSR constructs the graph to exploit the relationship between the adjacent spectral bands, while our MGSR constructs the graph to explore the relationship between the adjacent pixels in the same homogenous region. It takes full advantage of the spatial information from homogenous regions. Besides, EGCSR was proposed to select informative bands from hyperspectral data with a limited size, while our MGSR is designed to select important bands from infinite hyperspectral data by dynamically introducing noises at a certain level. This may result in a more robust and effective result.

V. Experiments

In this section, we conduct the experimental studies on three HSI datasets to validate the performance of the proposed MGSR in terms of four evaluation metrics.

A. Hyperspectral Datasets

We adopt three public HSI datasets to evaluate our MGSR in the experiments, including Indian Pines, Kennedy Space Center (KSC), and Botswana. The information is summarized in Table I.

The Indian Pines dataset was captured by the AVIRIS sensor to record a scene from Northwest Indiana in 1992. The size of this scene is 145 × 145 pixels and each pixel is described by 200 valid spectral bands after removing the water-absorption and noise bands. There are 16 different types of land covers with regular geometry recorded in this scene.

TABLE I
SUMMARY OF THE THREE HSI DATASETS

Datasets	Indian Pines	KSC	Botswana
Pixels	145×145	512×614	1476×256
Bands	200	176	145
Clusters	16	13	14
Sensor	AVIRIS	AVIRIS	Hyperion

The KSC dataset was collected by the NASA AVIRIS instrument to record a scene from KSC, FL, USA, in 1996. This scene consists of 512×614 pixels with 13 different categories of land covers. After the removal of water-absorption and low SNR bands, there are 176 available spectral bands to describe each pixel.

The Botswana datasets were gathered by the NASA EO-1 satellite with Hyperion sensor to record a scene from the Okavango Delta, Botswana, in 2001. This image provides a natural scene containing 256×1476 pixels of 14 different classes of land covers. Each pixel is represented by 145 useful spectral bands after removing water-absorption and low SNR bands.

B. Experimental Setup

1) Compared Methods: To verify the superiority of the proposed MGSR, we consider the following unsupervised band selection methods for comparison. **AllBands** is a baseline to adopt all original spectral bands for experiment. **LapScore** [53] relies on the Laplacian matrix to select a subset of spectral bands with better local manifold structure information. **MVPCA** [17] uses a maximum-variance principal components analysis as criteria to select informative spectral bands. **E_FDPC** [20] selects valuable spectral bands according to the weighting strategies of the normalized local density and intracluster distance. **SOP-SRL** [22] dynamically selects informative spectral bands from the hyperspectral data stream in an adaptive fashion. **EGCSR** [25] selects useful spectral bands with the help of graph convolution using the graph constructed from spectral bands. **MGSR*** represents a degeneration of the proposed method without corruptions and using the graph calculating by a k-nearest neighbor graph as in EGCSR. **MGSR** is the proposed MGSR method for effective unsupervised band selection. Consistent with unsupervised learning scheme, k-means clustering is adopted to measure the quality of the selected bands.

2) Evaluation Metrics: We utilize four quantitative metrics to measure the quality of the selected bands, including clustering accuracy (ACC), normalized mutual information (NMI) [46], purity [54], and Kappa coefficient [55]. It is clear that for all these evaluation metrics, the larger values mean the better quality of the selected bands.

1) ACC calculates the percentage of correctly predicted clusters to pixels. Assume that r_i and l_i are the clustering result and ground truth label, and the ACC is defined as

$$\mathrm{ACC} = \frac{\sum_{i=1}^{N} \delta(\mathrm{map}(r_i), l_i)}{N} \tag{12}$$

where N is the number of predicted pixels, map$(\cdot)$ is the permutation mapping function that maps each cluster label to a true class label, and $\delta(x, y)$ is the delta function that equals one if $x = y$ and zero otherwise.

2) NMI measures the quality of predicted clusters. Assume that there are K clusters, n_i denotes the number of pixels in the ith cluster $\mathcal{C}_i$, $\widehat{n}_j$ represents the number of pixels in the jth ground truth class $\mathcal{G}_j$, and $n_{i,j}$ is the number of overlap between $\mathcal{C}_i$ and $\mathcal{G}_j$. The NMI is defined as

$$\mathrm{NMI} = \frac{\sum_{i=1}^{K}\sum_{j=1}^{K} n_{i,j} \log \frac{n_{i,j}}{n_i \widehat{n}_j}}{\sqrt{\left(\sum_{i=1}^{K} n_i \log \frac{n_i}{n}\right)\left(\sum_{i=1}^{K} \widehat{n}_j \log \frac{\widehat{n}_j}{n}\right)}}. \tag{13}$$

3) Purity is also used to estimate the quality of predicted clusters. Given the predicted clusters $\{\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_K\}$ and ground truth class $\{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_L\}$, the purity is defined as

$$\mathrm{Purity} = \frac{1}{N}\sum_{i=1}^{K} \max_j \left|\mathcal{C}_i \cap \mathcal{G}_j\right|. \tag{14}$$

4) Kappa coefficient statistically evaluates the degree of clustering agreement. Assume that p_o is the relatively observed agreement and p_e is the hypothetical probability of chance agreement, and the Kappa is defined as

$$\mathrm{Kappa} = \frac{p_o - p_e}{1 - p_e}. \tag{15}$$

3) Parameter Setting: In MGSR, the number of segmentation S is chosen from $\{1, 3, 7, 15, 25, 50, 100, 200, 500, 1000\}$. The parameter λ is tuned from $\{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$. The corruption ratio p varies within different noise levels of $\{0.1, 0.2, \ldots, 0.9\}$. The number of selected bands is varying from $\{5, 10, 15, \ldots, 50\}$. For other compared methods, the hyperparameters are also tuned to achieve the best performance. The average results with optimal parameters are reported from ten repeated trials.

C. Experimental Results

We first conduct the experiments to select the top 25 bands by different unsupervised band selection methods and evaluate the quality of the selected bands. The detailed results of the four evaluation metrics are summarized in Table II. Accordingly, we list the wavelengths of top 25 bands selected by all compared methods in Fig. 2 and present the corresponding clustering maps in Figs. 3–5. From the results in Table II, we can find that the proposed MGSR is generally superior to all compared methods on the three datasets, except the KSC dataset in terms of purity. Besides, we can also gain the following observations.

First, we note that LapScore achieves the lowest results among all compared band selection methods. LapScore is also inferior to the baseline using all bands on the three datasets. This indicates that LapScore is insufficient to distinguish informative bands. Second, compared to MVPCA, E_FDPC, and SOP-SRL, our MGSR achieves at least 10.93%, 10.63%, and 3.96% enhancements on the Indian Pines dataset.

TABLE II
CLUSTERING PERFORMANCE COMPARISON OF DIFFERENT METHODS WITH 25 SELECTED BANDS FROM HSI DATASETS

Dataset	Measure	AllBands	LapScore	MVPCA	E_FDPC	SOP-SRL	EGCSR	MGSR*	MGSR
Indian Pines	ACC	0.3527	0.3181	0.3197	0.3211	0.3399	0.3551	0.3411	**0.3645**
	NMI	0.4246	0.4038	0.3978	0.3933	0.4187	0.4162	0.4195	**0.4243**
	Purity	0.5109	0.5017	0.4921	0.4924	0.5145	0.5126	0.5132	**0.5159**
	Kappa	0.2917	0.2550	0.2574	0.2615	0.2844	0.2932	0.2829	**0.3038**
KSC	ACC	0.5134	0.4771	0.5064	0.5024	0.5127	0.5148	0.5119	**0.5282**
	NMI	0.5536	0.5185	0.5576	0.5659	0.5584	0.5607	0.5582	**0.5674**
	Purity	0.5623	0.5591	0.5756	**0.6281**	0.6056	0.5849	0.5756	0.6004
	Kappa	0.4517	0.4186	0.4456	0.4530	0.4526	0.4562	0.4519	**0.4702**
Botswana	ACC	0.5558	0.5166	0.5655	0.5636	0.5764	0.5653	0.5738	**0.5901**
	NMI	0.6569	0.5985	0.6380	0.6423	0.6495	0.6503	0.6556	**0.6805**
	Purity	0.6271	0.5733	0.6165	0.6093	0.6212	0.6232	0.6251	**0.6418**
	Kappa	0.5198	0.4775	0.5301	0.5279	0.5319	0.5299	0.5390	**0.5568**

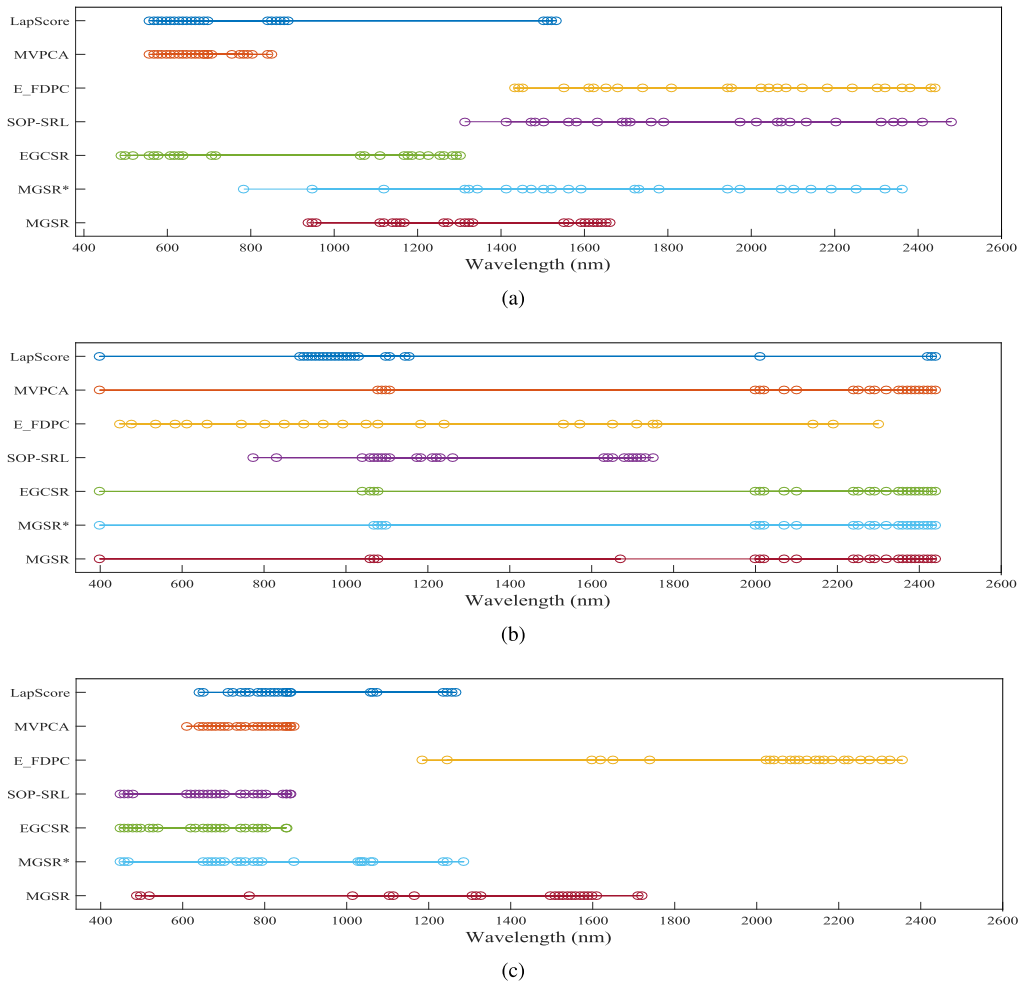

Fig. 2. Wavelengths of top 25 bands selected by different methods from HSI datasets. (a) Indian Pines. (b) KSC. (c) Botswana.

The superiority of our MGSR can also be found from the results of the other two datasets. Third, we find that EGCSR and MGSR* without introducing random corruptions show similar better performance over LapScore, MVPCA, E_FDPC, and SOP-SRL. This signifies that the methods developed for the non-Euclidean domain attain better performance over those methods for the Euclidean domain. For band selection, there is no difference between the performance of EGCSR and MGSR* that use a k-nearest neighbor graph to construct the graph from bands or pixels. Fourth, compared to EGCSR and MGSR*, our MGSR obtains 2.97% and 3.49% average improvements on the three datasets. This demonstrates that the random corruption and segmentation information introduced in MGSR are beneficial to more effective band selection. From the clustering maps in Figs. 3–5, we can visually describe the quality of the selected bands provided by different methods as in Fig. 2. The superiority of MGSR is consistent with the above observations. There are 200 valid spectral bands

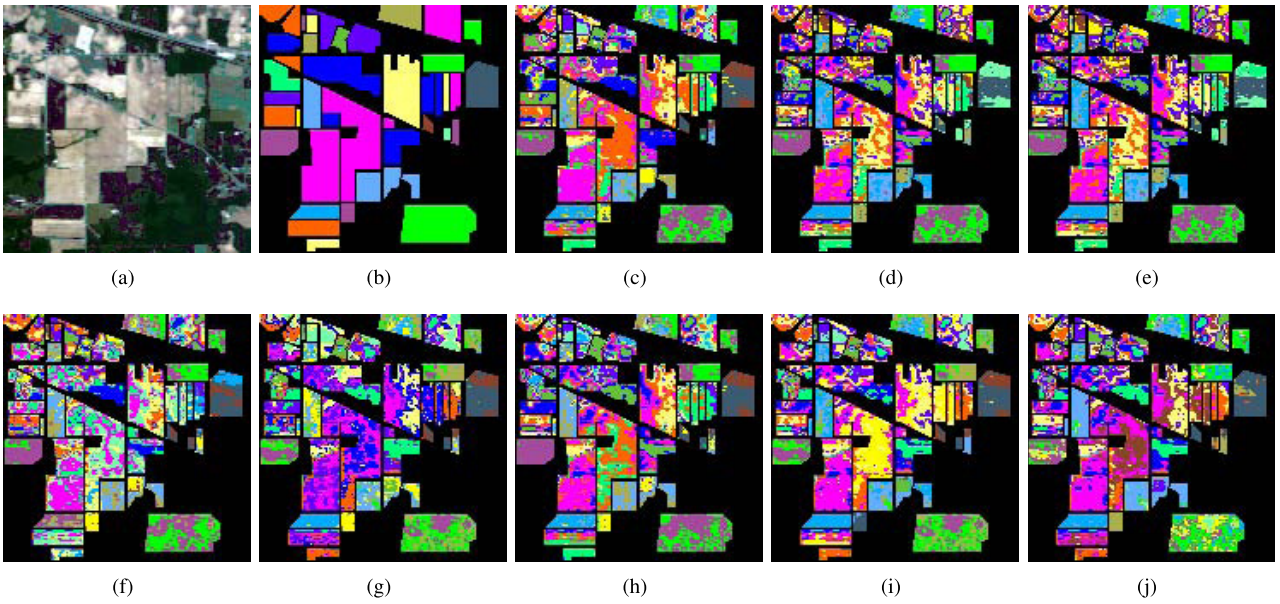

Fig. 3. Clustering maps of the Indian Pines dataset obtained by different methods. (a) False-color image. (b) Ground truth. (c) AllBands. (d) LapScore. (e) MVPCA. (f) E_FDPC. (g) SOPSRL. (h) EGCSR. (i) MGSR*. (j) MGSR.

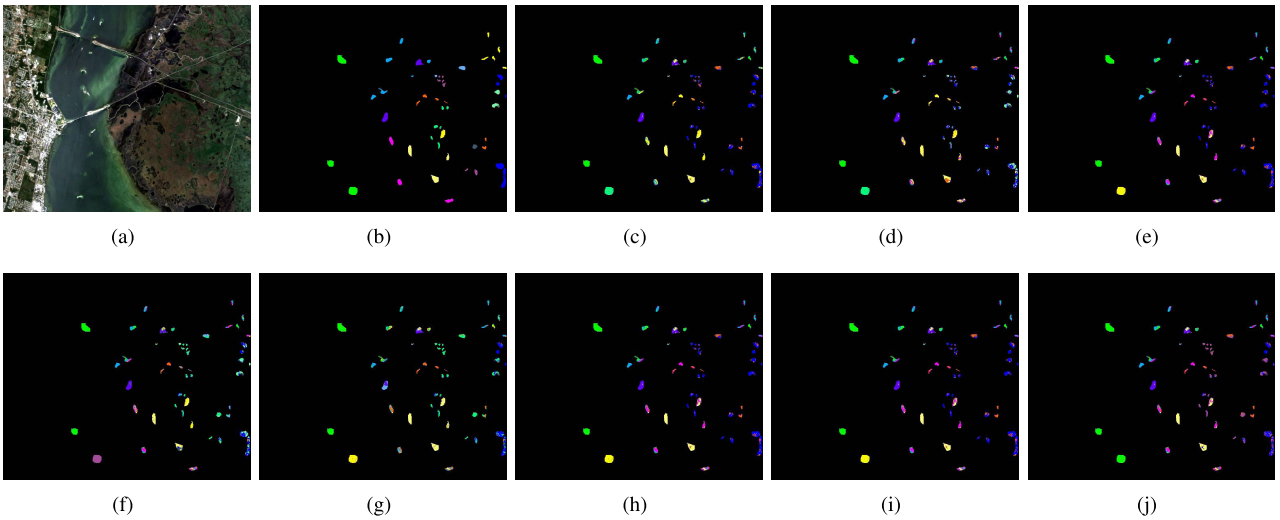

Fig. 4. Clustering maps of the KSC dataset obtained by different methods. (a) False-color image. (b) Ground truth. (c) AllBands. (d) LapScore. (e) MVPCA. (f) E_FDPC. (g) SOPSRL. (h) EGCSR. (i) MGSR*. (j) MGSR.

used in the Indian Pines dataset. The previous 41 spectral bands are obtained by visible spectrum in the wavelength range 400–780 nm, while the latter 159 spectral bands are acquired by near-infrared (NIR) spectrum in the wavelength range 780–2500 nm. From this information and Fig. 2, it is easy to find that the top 25 bands selected by MGSR are all in NIR spectral regions.

Besides, we also conduct the experiments to assess the quality of different numbers of bands selected by different compared methods. The detailed results are presented in Fig. 6. With an increasing number of selected bands, all compared methods first show the increasing performance and then present the stable or even degraded performance. From the results in Fig. 6, we can find that our MGSR obtains the highest accuracy in most cases. Specifically, MGSR always achieves better results with different numbers of selected bands $r \in \{5, 10, \ldots, 50\}$ and the best result is achieved with $r = 15$ on the Botswana dataset. For the Indian Pines dataset, MGSR gains better results when $r \geq 10$ and the optimal result is obtained with $r = 20$. For the KSC dataset, MGSR yields better performance when $r \geq 20$ and the best results are gained with $r = 15$. The superiority of MGSR is mainly attributed to jointly introducing the structural graph with spatial consistency and marginalized corruption. Therefore, MGSR is good at distinguishing informative bands.

D. Running Time Comparison

In this section, we conduct the experimental study to compare the execution time of all compared methods. The compared methods are all developed in MATLAB R2016a on

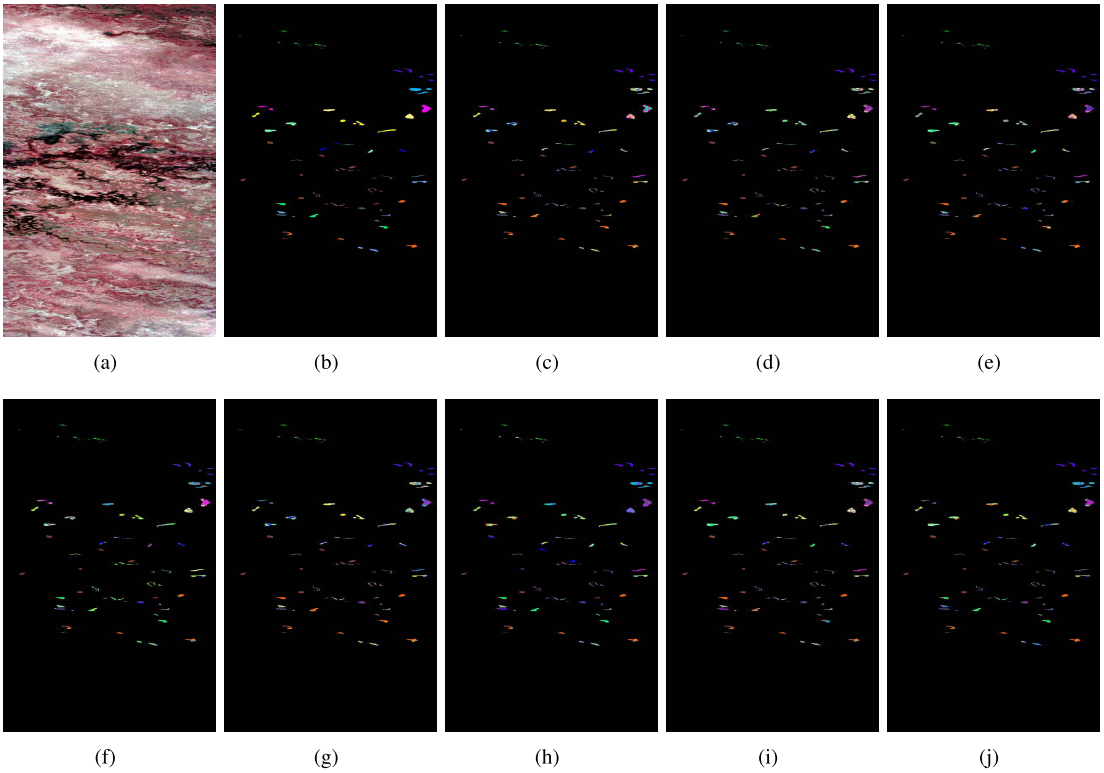

Fig. 5. Clustering maps of the Botswana dataset obtained by different methods. (a) False-color image. (b) Ground truth. (c) AllBands. (d) LapScore. (e) MVPCA. (f) E_FDPC. (g) SOPSRL. (h) EGCSR. (i) MGSR*. (j) MGSR.

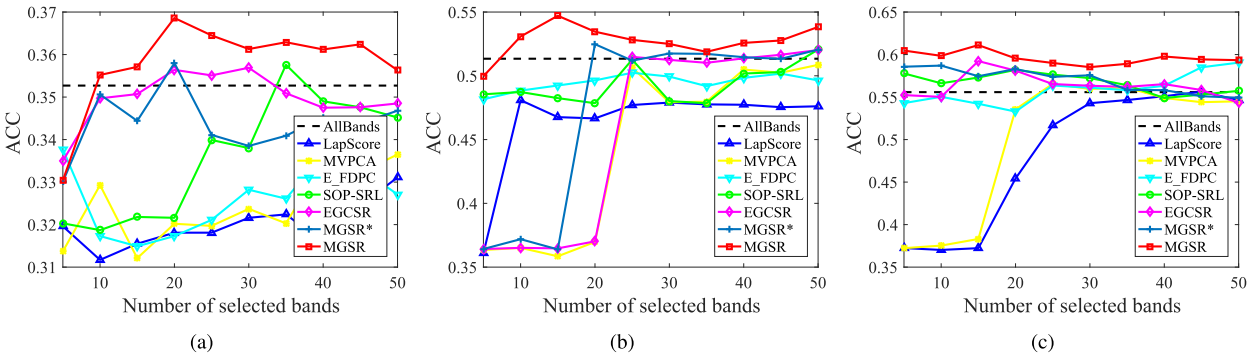

Fig. 6. Clustering performance comparison of different methods with respect to different numbers of selected bands. (a) Indian Pines. (b) KSC. (c) Botswana.

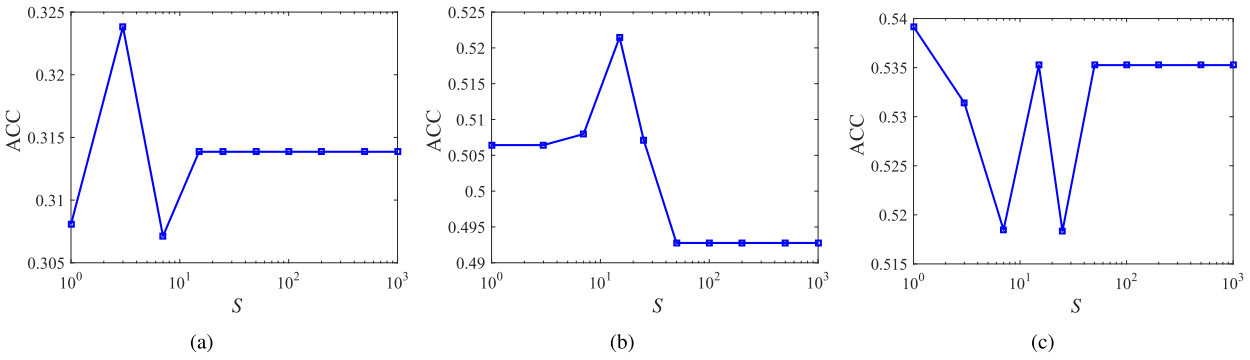

Fig. 7. Performance variation of MGSR with respect to different numbers of segmentation S. (a) Indian Pines. (b) KSC. (c) Botswana.

a workstation (Intel Core 2.70-GHz CPU and 8-GB RAM). Table III presents the running time of all compared methods to conduct band selection. From the results in Table III, we notice that MVPCA, E_FDPC, and EGCSR yield very small running

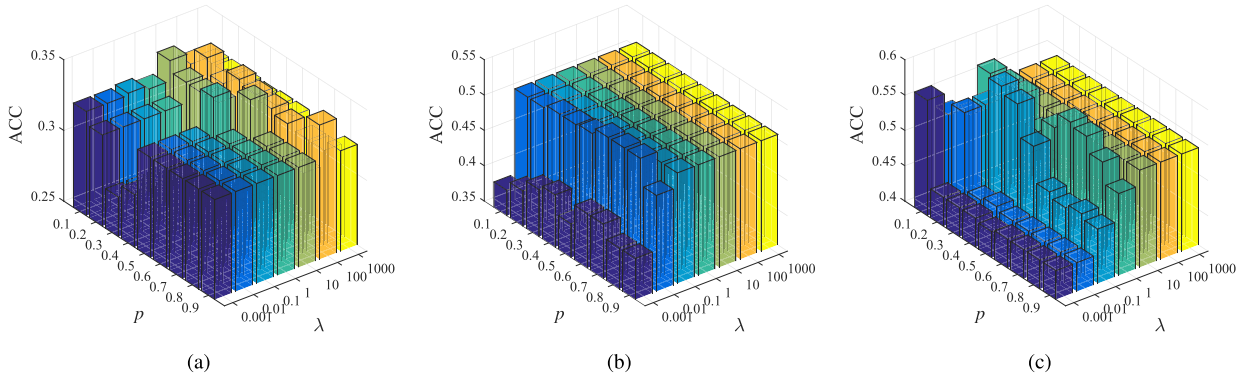

Fig. 8. Performance variation of MGSR with respect to different noise levels p and different values of parameter λ. (a) Indian Pines. (b) KSC. (c) Botswana.

TABLE III
RUNNING TIME COMPARISON ON HSI DATASETS (IN SECOND)

Method	Indian Pines	KSC	Botswana
LapScore	2.622	1.049	0.544
MVPCA	0.058	0.039	0.029
E_FDPC	0.079	0.035	0.023
SOP-SRL	3.325	1.326	0.697
EGCSR	0.187	0.144	0.079
MGSR*	3.327	1.285	0.713
MGSR	3.466	1.325	0.764

TABLE IV
ABLATION STUDY OF THE PROPOSED MGSR ON HSI DATASETS

Dataset	#Bands	SR	MGSR-N	MGSR-G	MGSR
Indian Pines	15	0.3420	0.3569	0.3568	**0.3571**
	25	0.3439	0.3576	0.3578	**0.3645**
	35	0.3422	0.3622	0.3594	**0.3629**
	45	0.3455	0.3486	0.3597	**0.3624**
KSC	15	0.5194	0.5247	0.5373	**0.5471**
	25	0.5225	0.5236	0.5242	**0.5282**
	35	0.5156	0.5175	**0.5192**	0.5188
	45	0.5111	0.5153	0.5261	**0.5277**
Botswana	15	0.5819	0.6042	0.5927	**0.6111**
	25	0.5663	0.5756	0.5808	**0.5901**
	35	0.5621	0.5728	0.5678	**0.5891**
	45	0.5543	0.5929	0.5611	**0.5943**

time on the three datasets. By contrast, LapScore, SOP-SRL, MGSR*, and our MGSR take relatively more running time to achieve band selection. This is mainly attributed to the construction of the graph matrix or the iterative learning process. In practice, their running time of less than 5 s is trivial and negligible.

E. Ablation Study

The proposed MGSR is achieved with the joint utilization of the corrupted pixels and structural graph obtained from the segmented HSI. To investigate how the performance can be improved by our method, we create two variants of MGSR. MGSR-N represents the variant that conducts linear reconstruction of original pixels from infinite noisy ones, whereas MGSR-G denotes the variant that performs linear reconstruction of original pixels from the product of the structural graph and original pixels. Besides, the basic model of SR conducting linear reconstruction of original pixels without using the corrupted pixels and structural graph is also introduced in the ablation study to test the effectiveness of the two components. Using the previous parameter setting, we report the best results with different numbers of selected bands $r \in \{15, 25, 35, 45\}$ in Table IV. With a reduced number of bands, MGSR-N occasionally outperforms MGSR-G. The rationale behind this behavior may be the inherent characteristic of the given datasets and the information discrepancy between infinite noisy pixels and the product of the structural graph and original pixels. MGSR provides the best clustering accuracy on the three HSI datasets, except the KSC dataset using 35 selected bands. To be specific, compared to MGSR-N, MGSR yields 3.96% improvement on the Indian Pines dataset using 45 selected bands and achieves 4.27% enhancement on the KSC dataset using 15 selected bands. This indicates the effectiveness of training model with the corrupted pixels. Compared to MGSR-G, MGSR obtains 5.92% enhancement on the Botswana dataset using 45 selected bands. This signifies the effectiveness of training model with the structural graph from the segmented HSI. Compared to SR, MGSR gains much better results than the comparison with MGSR-N and MGSR-G. The superior results demonstrate the effectiveness of the two components simultaneously used in MGSR.

F. Parameter Sensitivity Study

In the proposed MGSR, three parameters need to be manually determined by users. They are the number of segmentation S, noise level p, and coefficient parameter λ. In this section, we conduct the parameter sensitivity study to analyze how these three factors influence the performance of MGSR. We first study the performance variation of MGSR by varying S from $\{1, 3, 7, 15, 25, 50, 100, 200, 500, 1000\}$ while fixing $p = 0.1$ and $\lambda = 0.01$. The clustering results with top 25 selected bands are shown in Fig. 7. As can be observed, MGSR achieves the best performance when $S = 3$ for the Indian Pines dataset and $S = 15$ for the KSC dataset. This demonstrates that it is suitable to determine a small number of segmentations for MGSR.

Next, we study the performance variation of MGSR with respective to different settings of λ and p. Specifically, fixing $S = 3$, p is varying from $\{0.1, 0.2, \ldots, 0.9\}$, while λ is tuned

within $\{10^{-3}, 10^{-2}, \ldots, 10^{3}\}$. The clustering accuracy results with top 25 ranked bands are reported in Fig. 8. From the results in Fig. 8, it is clear that different values for the two factors result in different performance. For the Indian Pines dataset, clustering accuracy is relatively stable when $p \geq 0.4$ and the optimal p and λ are 0.5 and 1, respectively. For the KSC dataset, clustering performance is poor when $\lambda = 0.001$ and the optimal p and λ are 0.6 and 0.01, respectively. For the Botswana dataset, clustering accuracy is superior when $\lambda \geq 1$ and the optimal p and λ are 0.1 and 0.01, respectively. Therefore, it is important to determine appropriate values for λ and p in MGSR to achieve unsupervised band selection.

VI. Conclusion

In this article, we proposed an MGSR method for unsupervised hyperspectral band selection. MGSR is a general extension of SR for non-Euclidean domain. In MGSR, the pixel graph is constructed to retain the spatial and structural information from the segmented HSI and the original spectral bands are reconstructed from the infinite random corruptions and structural spatial graph. Band selection is achieved with an $\ell_{2,1}$-norm regularization in the joint exploration of the spectral bands and structural graph. To solve MGSR, we provided an effective optimization algorithm with the marginalized denoising process as a solution. Extensive experiments and comparisons on three HSI datasets have demonstrated that MGSR specializes in unsupervised band selection and yields better performance over the compared methods. In our future work, we will investigate the sensitivity of the quality of the superpixels obtained from superpixel segmentation methods in the HSI analysis.

References

[1] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[2] L. Zhang, L. Zhang, and B. Du, "Deep learning for remote sensing data: A technical tutorial on the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 4, no. 2, pp. 22–40, Jun. 2016.

[3] Y. Gu, J. Chanussot, X. Jia, and J. A. Benediktsson, "Multiple kernel learning for hyperspectral image classification: A review," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6547–6565, Nov. 2017.

[4] G. Shaw and D. Manolakis, "Signal processing for hyperspectral image exploitation," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 12–16, Jan. 2002.

[5] Y. Xu, Z. Wu, J. Chanussot, and Z. Wei, "Hyperspectral computational imaging via collaborative Tucker3 tensor decomposition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 98–111, Jan. 2021.

[6] Y. Zhou, J. Peng, and C. L. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, Feb. 2015.

[7] W. Shao and S. Du, "Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Oct. 2016.

[8] L. M. Bruce, C. H. Koger, and J. Li, "Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 10, pp. 2331–2338, Oct. 2002.

[9] J. Jiang, J. Ma, C. Chen, Z. Wang, Z. Cai, and L. Wang, "SuperPCA: A superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4581–4593, Aug. 2018.

[10] X. Zhang, X. Jiang, J. Jiang, Y. Zhang, X. Liu, and Z. Cai, "Spectral-spatial and superpixelwise PCA for unsupervised feature extraction of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, early access, Feb. 26, 2021, doi: 10.1109/TGRS.2021.3057701.

[11] X. Jiang, X. Fang, Z. Chen, J. Gao, J. Jiang, and Z. Cai, "Supervised Gaussian process latent variable model for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1760–1764, Oct. 2017.

[12] F. He, F. Nie, R. Wang, W. Jia, F. Zhang, and X. Li, "Semisupervised band selection with graph optimization for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, early access, Nov. 30, 2020, doi: 10.1109/TGRS.2020.3037746.

[13] W. Sun and Q. Du, "Graph-regularized fast and robust principal component analysis for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3185–3195, Jun. 2018.

[14] Q. Wang, F. Zhang, and X. Li, "Optimal clustering framework for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 10, pp. 5910–5922, Oct. 2018.

[15] X. Wang, Z. Wang, Y. Zhang, X. Jiang, and Z. Cai, "Latent representation learning based autoencoder for unsupervised feature selection in hyperspectral imagery," *Multimedia Tools Appl.*, pp. 1–15, Feb. 2021, doi: 10.1007/s11042-020-10474-8.

[16] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 118–139, Jun. 2019.

[17] C.-I. Chang, Q. Du, T.-L. Sun, and M. L. G. Althouse, "A joint band prioritization and band-decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 6, pp. 2631–2641, Nov. 1999.

[18] S. Li and H. Qi, "Sparse representation based band selection for hyperspectral images," in *Proc. 18th IEEE Int. Conf. Image Process.*, Sep. 2011, pp. 2693–2696.

[19] W. Sun, L. Zhang, B. Du, W. Li, and Y. M. Lai, "Band selection using improved sparse subspace clustering for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2784–2797, Jun. 2015.

[20] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 88–102, Jan. 2016.

[21] P. Hu, X. Liu, Y. Cai, and Z. Cai, "Band selection of hyperspectral images using multiobjective optimization-based sparse self-representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 3, pp. 452–456, Mar. 2019.

[22] X. Wei, W. Zhu, B. Liao, and L. Cai, "Scalable one-pass self-representation learning for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 7, pp. 4360–4374, Jul. 2019.

[23] W. Sun, L. Zhang, L. Zhang, and Y. M. Lai, "A dissimilarity-weighted sparse self-representation method for band selection in hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4374–4388, Sep. 2016.

[24] W. Sun, L. Tian, Y. Xu, D. Zhang, and Q. Du, "Fast and robust self-representation method for hyperspectral band selection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 11, pp. 5087–5098, Nov. 2017.

[25] Y. Cai, Z. Zhang, X. Liu, and Z. Cai, "Efficient graph convolutional self-representation for band selection of hyperspectral image," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 4869–4880, 2020.

[26] P. Ribalta Lorenzo, L. Tulczyjew, M. Marcinkiewicz, and J. Nalepa, "Hyperspectral band selection using attention-based convolutional neural networks," *IEEE Access*, vol. 8, pp. 42384–42403, 2020.

[27] H. Sun, J. Ren, H. Zhao, P. Yuen, and J. Tschannerl, "Novel Gumbel–Softmax trick enabled concrete autoencoder with entropy constraints for unsupervised hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, early access, Jun. 4, 2021, doi: 10.1109/TGRS.2021.3075663.

[28] M. Zeng, Y. Cai, Z. Cai, X. Liu, P. Hu, and J. Ku, "Unsupervised hyperspectral image band selection based on deep subspace clustering," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 12, pp. 1889–1893, Dec. 2019.

[29] S. K. Roy, S. Das, T. Song, and B. Chanda, "DARecNet-BS: Unsupervised dual-attention reconstruction network for hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, early access, Aug. 11, 2020, doi: 10.1109/LGRS.2020.3013235.

[30] L. Mou, S. Saha, Y. Hua, F. Bovolo, L. Bruzzone, and X. X. Zhu, "Deep reinforcement learning for band selection in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, early access, Mar. 26, 2021, doi: 10.1109/TGRS.2021.3067096.

[31] X. Bai, Z. Guo, Y. Wang, Z. Zhang, and J. Zhou, "Semisupervised hyperspectral band selection via spectral–spatial hypergraph model," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2774–2783, Jun. 2015.
[32] L. Zhang, Q. Zhang, B. Du, X. Huang, Y. Y. Tang, and D. Tao, "Simultaneous spectral-spatial feature selection and extraction for hyperspectral images," *IEEE Trans. Cybern.*, vol. 48, no. 1, pp. 16–28, Jan. 2018.
[33] Q. Zhang, Y. Tian, Y. Yang, and C. Pan, "Automatic spatial–spectral feature selection for hyperspectral image via discriminative sparse multimodal learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 261–279, Jan. 2015.
[34] L. Sun *et al.*, "Low rank component induced spatial–spectral kernel method for hyperspectral image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 10, pp. 3829–3842, Oct. 2020.
[35] J. Fan, T. Chen, and S. Lu, "Superpixel guided deep-sparse-representation learning for hyperspectral image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 11, pp. 3163–3173, Nov. 2018.
[36] X. Jiang, X. Song, Y. Zhang, J. Jiang, J. Gao, and Z. Cai, "Laplacian regularized spatial-aware collaborative graph for discriminant analysis of hyperspectral imagery," *Remote Sens.*, vol. 11, no. 1, p. 29, 2019.
[37] Y. Zhang, X. Jiang, X. Wang, and Z. Cai, "Spectral-spatial hyperspectral image classification with superpixel pattern and extreme learning machine," *Remote Sens.*, vol. 11, no. 17, p. 1983, Aug. 2019.
[38] X. Lu, Y. Wang, and Y. Yuan, "Graph-regularized low-rank representation for destriping of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4009–4018, Jul. 2013.
[39] F. Fan, Y. Ma, C. Li, X. Mei, J. Huang, and J. Ma, "Hyperspectral image denoising with superpixel segmentation and low-rank representation," *Inf. Sci.*, vols. 397–398, pp. 48–68, Aug. 2017.
[40] M. Chen, Z. Xu, K. Weinberger, and F. Sha, "Marginalized denoising autoencoders for domain adaptation," in *Proc. 29th Int. Conf. Mach. Learn.*, 2012, pp. 1627–1634.
[41] L. Maaten, M. Chen, S. Tyree, and K. Weinberger, "Learning with marginalized corrupted features," in *Proc. 30th Int. Conf. Mach. Learn.*, 2013, pp. 410–418.
[42] Z. Tao, H. Liu, S. Li, Z. Ding, and Y. Fu, "Marginalized multiview ensemble clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 2, pp. 600–611, Feb. 2020.
[43] P. Zhu, W. Zuo, L. Zhang, Q. Hu, and S. C. K. Shiu, "Unsupervised feature selection by regularized self-representation," *Pattern Recognit.*, vol. 48, no. 2, pp. 438–446, 2015.
[44] W. He, X. Cheng, R. Hu, Y. Zhu, and G. Wen, "Feature self-representation based hypergraph unsupervised feature selection via low-rank representation," *Neurocomputing*, vol. 253, pp. 127–134, Aug. 2017.
[45] T. Chang *et al.*, "Robust unsupervised feature selection via dual self-representation and manifold regularization," *Knowl.-Based Syst.*, vol. 145, pp. 109–120, Apr. 2018.
[46] Y. Zhang, J. Wu, C. Zhou, Z. Cai, J. Yang, and P. S. Yu, "Multi-view fusion with extreme learning machine for clustering," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 5, pp. 1–23, Nov. 2019.
[47] Y. Zhang, J. Wu, Z. Cai, and P. S. Yu, "Multi-view multi-label learning with sparse feature selection for image annotation," *IEEE Trans. Multimedia*, vol. 22, no. 11, pp. 2844–2857, Nov. 2020.
[48] X. Xiao, Y. Chen, Y.-J. Gong, and Y. Zhou, "Low-rank preserving t-linear projection for robust image feature extraction," *IEEE Trans. Image Process.*, vol. 30, pp. 108–120, 2021.
[49] R. Hu *et al.*, "Graph self-representation method for unsupervised feature selection," *Neurocomputing*, vol. 220, pp. 130–137, Jan. 2017.
[50] C. Wang, S. Pan, G. Long, X. Zhu, and J. Jiang, "MGAE: Marginalized graph autoencoder for graph clustering," in *Proc. ACM Conf. Inf. Knowl. Manage.*, Nov. 2017, pp. 889–898.
[51] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. 24th IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 2097–2104.
[52] J. Jiang, J. Ma, Z. Wang, C. Chen, and X. Liu, "Hyperspectral image classification in the presence of noisy labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 851–865, Feb. 2018.
[53] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in *Proc. 25th Annu. Conf. Neural Inf. Process. Syst.*, 2005, pp. 507–514.
[54] Y. Zhang, X. Wang, Z. Cai, Y. Zhou, and P. S. Yu, "Tensor-based unsupervised multi-view feature selection for image recognition," in *Proc. IEEE Int. Conf. Multimedia Expo.*, Jul. 2021, pp. 1–6.
[55] N. Yokoya, C. Grohnfeldt, and J. Chanussot, "Hyperspectral and multispectral data fusion: A comparative review of the recent literature," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 2, pp. 29–56, Jun. 2017.

Yongshan Zhang (Member, IEEE) received the B.Eng. and Ph.D. degrees in computer science from the China University of Geosciences, Wuhan, China, in 2014 and 2019, respectively.

She joined the BDSC Laboratory, University of Illinois at Chicago, Chicago, IL, USA, from September 2017 to September 2018. She is currently an Associate Professor with the School of Computer Science, China University of Geosciences. Her research interests include machine learning and hyperspectral image analysis.

Xinxin Wang received the B.Eng. and M.Eng. degrees in computer science from the China University of Geosciences, Wuhan, China, in 2013 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Department of Computer and Information Science, University of Macau, Macau, China.

His research interests include data mining and machine learning.

Xinwei Jiang (Member, IEEE) received the Ph.D. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2012.

He is an Associate Professor with the China University of Geosciences, Wuhan. His research interests include nonparametric statistical models, dimensionality reduction, and hyperspectral image processing.

Yicong Zhou (Senior Member, IEEE) received the B.S. degree in electrical engineering from Hunan University, Changsha, China, in 1992, and the M.S. and Ph.D. degrees in electrical engineering from Tufts University, Medford, MA, USA, in 2008 and 2010, respectively.

He is a Professor with the Department of Computer and Information Science, University of Macau, Macau, China. His research interests include image processing, computer vision, machine learning, and multimedia security.

Dr. Zhou is a fellow of the Society of Photo-Optical Instrumentation Engineers (SPIE) and was recognized as one of the World's Top 2% Scientists and one of the Highly Cited Researchers in 2020. He received the Third Price of Macao Natural Science Award as a sole winner in 2020 and was a co-recipient in 2014. He has been a leading Co-Chair of the Technical Committee on Cognitive Computing of the IEEE Systems, Man, and Cybernetics Society since 2015. He serves as an Associate Editor for IEEE TRANSACTIONS ON NEUTRAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and four other journals.