

A Multiscale Superpixel-Level Group Clustering Framework for Hyperspectral Band Selection

Sen Jia¹, Senior Member, IEEE, Yue Yuan, Nanying Li, Jianhui Liao, Qiang Huang²,
 Xiuping Jia³, Fellow, IEEE, and Meng Xu⁴, Member, IEEE

Abstract—Hyperspectral imagery (HSI) contains hundreds of bands, which provide a wealth of spectral information and enable better characterization of features. However, the excessive dimensionality also poses a dimensional disaster for subsequent processing. Fortunately, band selection (BS) gives a straightforward and effective way to pick out a subset of bands with rich information and low correlation. Although many hyperspectral BS methods, especially clustering-based ones, have been proposed by researchers in recent years, the contextual information of adjacent bands and the spatial structural information of materials are not well investigated. Therefore, in this article, a multiscale superpixel-level group-clustering framework (MSGCF) has been proposed for hyperspectral BS. Different from previous, a new superpixel-level distance measure is elaborately utilized to group and cluster the spectral bands, which jointly considers the spectral context and spatial structure information. Concretely, to preserve the spatial structural information of HSI, multiple superpixel segmentation is first performed to generate superpixel maps in multiscales, which enables complementarity of multiple superpixel segmentation algorithms and adaptation to diverse scales of land cover types. Second, the grouping and clustering paradigm is introduced to conduct the contextual information among bands. Here the maximum points of superpixel-level KL- ℓ_1 distance of adjacent bands are adopted as partition points to separate bands into groups, which encourages adjacent bands with strong correlation to be divided into the same group. Third, a superpixel-level fast density-based clustering method (SuFDPC) with superpixel-level $\ell_{2,1}$ distance is developed to select representative bands in every group. Finally, BS results are achieved with a ranking-based voting strategy by concerning information entropy and frequency of occurrence in a unified scheme. A series of ablation analyses and experimental comparisons on four real HSI datasets have been conducted, as well as similarity comparisons for the selected bands. The experimental results consistently demonstrated the effectiveness of our MSGCF approach. The codes of this work will be available at <http://jiasen.tech/papers/> for the sake of reproducibility.

Manuscript received December 10, 2021; revised January 16, 2022; accepted February 5, 2022. Date of publication February 9, 2022; date of current version March 23, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 41971300 and Grant 61901278, in part by the Key Project of Department of Education of Guangdong Province under Grant 2020ZDZX3045, and in part by the Natural Science Foundation of Guangdong Province under Grant 2021A1515011413. (Corresponding author: Meng Xu.)

Sen Jia, Yue Yuan, Nanying Li, Jianhui Liao, Qiang Huang, and Meng Xu are with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China (e-mail: senjia@szu.edu.cn; 1900271007@email.szu.edu.cn; linanying2021@email.szu.edu.cn; liaojianhui2019@email.szu.edu.cn; jameshq@szu.edu.cn; m.xu@szu.edu.cn).

Xiuping Jia is with the School of Engineering and Information Technology, University of New South Wales, Canberra, ACT 2612, Australia (e-mail: x.jia@adfa.edu.au).

Digital Object Identifier 10.1109/TGRS.2022.3150361

Index Terms—Band selection (BS), hyperspectral imagery (HSI), superpixel.

NOMENCLATURE

I	A hyperspectral image cube.
X, Y, B	Height, width, and number of bands of hyperspectral imagery.
T	Spatial resolution of hyperspectral imagery.
K, N_k	Number of superpixel scales and number of superpixels at the k th scale.
G	Superpixel cube.
g, p	Value and histogram statistic of the pixel.
W_n	Number of pixels in the n th superpixel.
$D^{(KL-\ell_1)}, D^{(\ell_{2,1})}$	KL- ℓ_1 and $\ell_{2,1}$ distance matrices between bands.
M, M_S	KL- ℓ_1 distance of adjacent bands before and after Gaussian smoothing.
e	Sequence of partition points.
V	Voting metric of the band.
$\mathcal{O}, \mathcal{O}'$	Results of BS before and after voting.

I. INTRODUCTION

HYPERSPECTRAL imagery (HSI) captured by hyperspectral sensors usually contains hundreds of spectral bands ranging from the visible region (0.4–0.7 μm) to the shortwave infrared region (nearly 2.4 μm). They provide a tremendous amount of spatial and spectral information of the objects. With the acquired abundant information, HSI-oriented processing has become one of the most promising techniques in numerous fields, including geological exploration [1], ecological sciences [2], disaster prediction [3], food industry [4], [5], environmental monitoring [6], [7], astronomy [8], ocean monitoring [9], etc. Nevertheless, the high dimensionality and redundant band information not only increases computational complexity, but also results in the so-called Hughes phenomenon, making practical processing of HSI data still challenging [10]–[12]. Therefore, dimensionality reduction is widely adopted in HSI preprocessing, which aims to eliminate irrelevant and redundant information of HSI data while retaining as much material discriminative information as possible [13], [14].

In practice, feature extraction and band selection (BS) are the most commonly used methods for the dimensionality reduction of HSI [15], [16]. Concretely, feature extraction for HSI aims to build the projections between the original high-dimensional space and the low-dimensional subspace via different manners (e.g., high-low, high-higher-low, etc), which usually presents good classification performance, yet totally changes the physical characteristics of the original HSI data [17]–[20]. A number of feature extraction methods have been proposed in the literature [21]–[23]. For instance, principal component analysis (PCA) was introduced as a linear dimensionality reduction technique for HSI [24]–[29], which projects the high-dimensional data to a low-dimensional subspace with principal components maximizing the variance of the projected data [30]. Independent component analysis (ICA) also conducts the projection from the high-dimensional space to a low-dimensional space [31]. Differently, it performs a linear transformation on the original data, intending to identify the independent components that make up the feature space. Furthermore, to deal with nonlinearly separable high-dimensional data, kernel PCA (KPCA) first introduces kernel methods to map the original space to a higher-dimensional linearly separable reproducing kernel Hilbert space (RKHS), and then performs PCA on the kernel matrix to extract principal components, with which projecting RKHS to a low-dimensional space [32]. Besides, nonparametric weighted feature extraction (NWFE) uses the nonparametric scatter matrix to reduce the impact of singularity problems [33], while decision boundary feature extraction (DBFE) extracts the most informative features based on the decision boundary feature matrix [34]. Apart from these, a number of wavelet-based feature extraction methods have also been developed [35]–[39].

Alternatively, BS identifies a subset of bands with rich discriminative information and weak correlation from the original bands, which tries to preserve most of the inherent characteristics of HSI data [40], [41]. In particular, in many hyperspectral application scenarios (e.g., food quality assessment [42] and crop disease detection [43]), fast and nondestructive representation is desirable for dimensionality reduction algorithms, and direct analysis of the selected spectra without changing the original features is a better choice to reserve the physical meaning. In general, BS algorithms can be divided into supervised, unsupervised, and semisupervised methods depending on whether the training set is utilized [44]. Supervised BS approaches use prior information to select the most discriminative bands [45], [46], while the importance of a band obtained by unsupervised methods is evaluated by various statistical measures or clustering quality assessment [47]. Unsupervised BS methods are preferable due to the lack of labeled samples, which could well ensure the generalization of the selection results. The semisupervised approach uses the association between labeled and unlabeled sample information to measure the representativeness of the bands [48]. BS using improved classification (BSIC) [49] utilizes edge-preserving filtering to improve the pixel classification map. BS based on dynamic classifier selection (BSDCS) [50] filters the classification map to provide a reference before defining local regions to select bands with

good classification performance. In particular, unsupervised BS algorithms can be roughly divided into two categories: ranking-based and clustering-based [40].

Ranking-based BS methods quantify the importance of each band according to the quantified metrics, and then select the bands with the highest scores in the ranking sequence. Some single-band metrics, such as information entropy, signal-to-noise ratio (SNR), and information divergence (ID), have been introduced [51]. Maximum variance principal component analysis (MVPCA) [52] calculates the ranking score of each band as a weighted sum of the transformed basis vector. However, the selected bands are usually highly correlated with each other. Several deep neural network techniques have also been applied to HSI BS. Attention-based convolutional neural networks (ACNNs) [53] develop an end-to-end network for HSI BS and classification by detecting outliers in the attention map of the band scores. Furthermore, based on the assumption that the complete set of bands can be reconstructed from its informative subsets, BS Network (BSNet) [54] uses a framework consisting of a band attention module (BAM) and a reconstruction network (RecNet) to learn the band weight vector, which is then ranked to select bands. Based on the same assumption, a double attention reconstruction network (DAREcNet) [55] jointly uses a position attention module (PAM) and a channel attention module (CAM) to calibrate the band weights and then ranks the entropy rate of each band in the reconstructed band set to select bands. With respect to clustering-based approaches, they divide the original bands into clusters and select the central band from each cluster to form the final subset. A hierarchical clustering algorithm for BS has been proposed in [56], which minimizes the intracluster variance and maximizes the intercluster variance. Besides, a number of clustering algorithms based on K-means [57], [58], affinity propagation (AP) [59], [60], and density clustering [61], [62] have been proposed. Concretely, K-means divide the bands into K clusters by iteration, while AP treats all data points as potential clustering centers and clusters them according to the correlation between bands. Moreover, concerning the density-based clustering algorithms, the density of the region where a data point is located is measured by the number of neighbors in the neighborhood.

In general, ranking-based BS only considers the scoring criteria of individual bands without considering the correlation information between bands, which makes the selected bands be stable enough, but the correlation between bands is relatively high. In contrast, clustering-based BS methods favor the selection of bands with low correlation and less redundant information, but the selected bands present a lack of stability (refer to the consistency of the bands when different numbers of bands are selected, i.e., the number of bands that change when adjacent numbers of bands are selected), and the algorithm is of high time complexity. Therefore, it is desirable to integrate both methods in a single scheme. Recently, several algorithms combining ranking-based and clustering-based strategies have been proposed, including enhanced density-peak-based clustering algorithm (EFDPC) [63], fast and robust principal component analysis on Laplacian graph (FRPCALG) [64], importance-assisted column subset BS (iCSBS) [65], optimal

clustering framework (OCF) [66], and local potential-based clustering (LPC) algorithm [67]. Among them, EFDPC extracts the clusters by ranking a density-based measure, which selects the most reliable bands by weighting the ranking scores of each band and progressively narrowing the cutoff threshold. OCF develops an OCF to search for the best clustering results on the whole band set, then uses the ranking on clusters strategy (RCS) method to evaluate the bands in each cluster, and finally sorts them and selects the top bands.

However, the spectral bands are usually considered to be disordered when performing clustering procedure, which may result in the ignorance of the contextual information across the spectral bands. Therefore, grouping-based BS approaches have been proposed, such as adaptive subspace partition strategy (ASPS) [68] and fast neighborhood grouping method (FNGBS) [69]. ASPS first divides the HSI cube into several subcubes by maximizing the ratio of interclass distance to intraclass distance and then selects the least noisy band in each cube. FNGBS, on the other hand, uses a coarse-fine strategy to spatially divide the HSI data into several groups before acquiring a relevant and informative subset of bands based on the local density and information entropy. However, the key parameter (i.e., the number of grouping) should be preset, which is less applicable for HSI data with various number of spectral bands. Meanwhile, the spatial structure information of each band is not fully exploited.

Fortunately, superpixel segmentation partitions image into multiple semantic subregions with similar spatial structure. The contextual similarity allows regions with homogeneity to be aggregated together. Superpixel extraction has been extensively studied for HSI classification through preprocessing [70], [71] or post-processing [72], [73]. But superpixel segmentation for HSI BS has not been well exploited.

Motivated by the idea of adaptive grouping and superpixel-based distance measure, we propose a multiscale superpixel-level group clustering framework (MSGCF) for HSI BS. First, to preserve the spatial structural information of HSI, multiple superpixel segmentation is performed to generate superpixel maps in multiscales, which enables complementarity of multiple superpixel segmentation algorithms and adaptation to diverse scales of land cover types. Second, the grouping and clustering paradigm is introduced to conduct the contextual information among bands. Here the maximum points of superpixel-level Kullback–Leibler (KL)- ℓ_1 distance of adjacent bands are adopted as partition points to separate bands into groups, which encourages adjacent bands with strong correlation to be divided into the same group. In this way, the following grouping oriented clustering procedure is possible to select bands with stronger discriminability. Third, a superpixel-level fast density-based clustering method (SuFDPC) with superpixel-level $\ell_{2,1}$ distance measure is developed to perform group-wise clustering. Finally, BS results are achieved with a ranking-based voting strategy by concerning information entropy and frequency of occurrence in a unified scheme. Fig. 1 shows the schematic of the MSGCF approach to make the method more easily to be understood.

Three main contributions of our MSGCF method are summarized as follows.

- 1) First, superpixel-level distance measure is comprehensively utilized to measure the similarity between bands instead of pixel-based, which jointly considers the spectral context and spatial structure information. Taking advantage of the homogeneity of superpixel regions, spatial structure information is integrated into the distance metric. In addition, due to the complementarity of different superpixel segmentation methods and the variability of scales for different land cover types, we combine different superpixel segmentation algorithms and different scales of superpixels to leverage the structural information and the spatial consistency of HSI data.
- 2) Second, we adopt a group-clustering framework to divide the HSI cube into multiple groups, and then cluster the bands within each group, which effectively utilizes the numerous contextual information of the bands and makes the final selected bands more dispersed and less correlated. Besides, the bands are grouped by using the superpixel-level KL- ℓ_1 distances of adjacent bands after Gaussian smoothing, and the extreme value points are used as segmentation points to better measure the correlation of adjacent bands. Since the grouping procedure is carried out in an adaptive way rather than presetting the number of groups by hand, the generalization of the proposed method can be ensured.
- 3) Finally, a SuFDPC with superpixel-level $\ell_{2,1}$ distance measure is proposed to select representative bands in every group. In order to obtain the final result of BS, joint information entropy and band appearance frequency are introduced as the scoring standard, and the final selected bands are determined by metric ranking-based voting. The codes of this work will be available at <http://jiasen.tech/papers/> for the sake of reproducibility.

The descriptions of all the important mathematical notations in this article are summarized in Nomenclature. The rest of this article is organized as follows. Section II briefly reviews the superpixel segmentation and ranking-based clustering algorithms. Section III presents the proposed MSGCF approach in detail. Experiments on four real HSI datasets are given in Section IV. Finally, Section V concludes this article with some remarks.

II. PRELIMINARIES

In this section, we briefly review three different superpixel segmentation methods and density peak clustering methods that are relevant to the proposed method.

A. Superpixel Segmentation

It can be found that the surface materials are generally continuous in spatial distribution, and the neighboring pixels are highly correlated by observing the real land cover distribution of the HSI datasets, which means that the land cover types in the same region may have the same class labels. Therefore, it is desirable to extract homogeneous regions in HSI by the consistency within the superpixels to characterize the structural information of materials. In natural image processing, superpixel segmentation algorithms can be broadly classified into

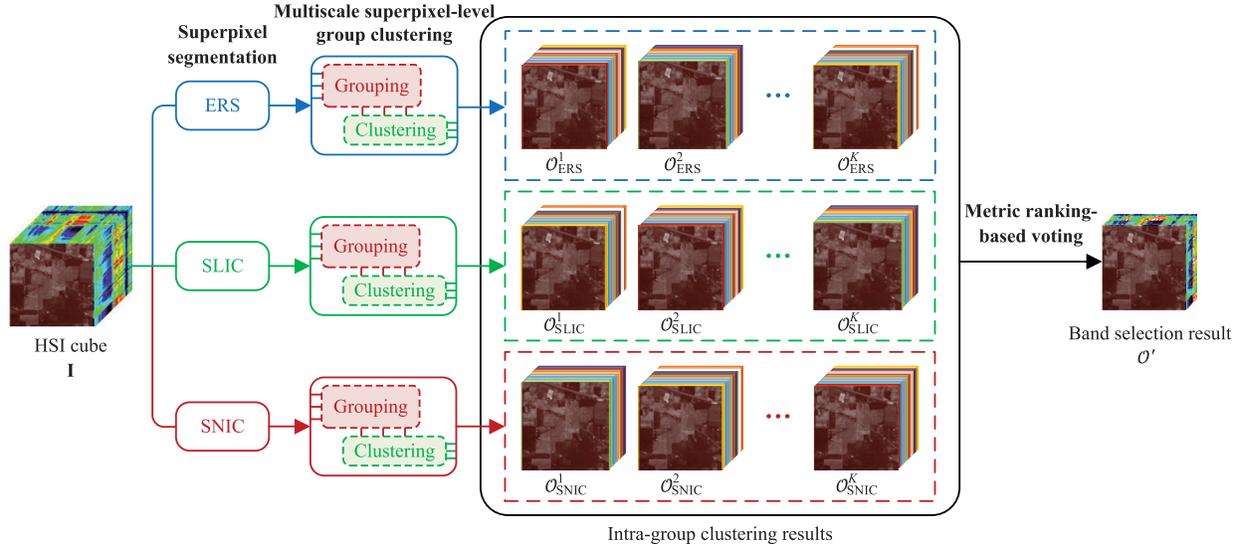


Fig. 1. Overall framework of the proposed MSGCF approach for HSI BS. First, three superpixel segmentation algorithms, i.e., ERS, SLIC, and SNIC, are applied to generate superpixel representations of HSI data at K different scales, respectively. The data streams corresponding to the three segmentation algorithms are denoted as three branches with three colors (blue, green, red). Second, the group-clustering paradigm is utilized, and superpixel representations of HSI at $3K$ scales are grouped and clustered in corresponding branches, generating a total of $3K$ intragroup clustering results. At last, the final BS results are obtained by metric ranking-based voting strategy, performed on all the $3K$ intragroup clustering results.

two categories: graph-based methods (normalized cut [74], Felzenswalb and Huttenlocher [75], superpixel lattice [76], constant intensity superpixels [77], entropy rate superpixels (ERS) [78]) and gradient-based methods (watershed [79], mean shift [80], quick shift [81], turbopixels (TP) [82], simple linear iterative clustering (SLIC) [83], simple noniterative clustering (SNIC) [84]). Among them, the ERS [36], [70], [85], [86], SLIC [87]–[90] and SNIC [7], [21] algorithms are widely used and have been proven to be effective in HSI processing.

1) *Entropy Rate Superpixel (ERS)*: ERS is an efficient superpixel segmentation algorithm that maps an image into an undirected graph $\mathcal{G} = (P, E)$, in which each pixel is considered as a vertex, where P denotes the set of vertices and E denotes the set of edges. Then the topology of the graph is maximized, and the objective function is optimized by a greedy algorithm to obtain N homogeneous superpixels of similar size. The objective function consists of two parts: the entropy rate of random walk on graph $\mathcal{H}(A)$ and the balancing term $\mathcal{T}(A)$, where A is the selected edge. Then, the objective function of ERS is defined as

$$\max_A \mathcal{H}(A) + \lambda \mathcal{T}(A) \quad \text{s.t. } A \subseteq E \quad (1)$$

where $\lambda \geq 0$ is the adjustable weighting factor between two terms. The entropy rate is conducive to the formation of compact and homogeneous clusters, while the balancing term is designed to constrain the size of superpixels and reduces the number of superpixels.

2) *Simple Linear Iterative Clustering (SLIC)*: SLIC is a local iterative clustering algorithm based on the relationship of color similarity and spatial distance, which can be considered as an application of the K-means clustering algorithm to generate superpixels. It is a process of transforming an image into a feature vector in CIELAB color space and spatial

coordinates, and then locally clustering image pixels to obtain regular superpixels by constructing a distance measure for feature vector [91]. First, the image needs to be initialized into N superpixel clustering centers of the same size, and the distance D is computed within a $2Q \times 2Q$ block region around the superpixel centers, where $Q = \sqrt{W/N}$, and W is the number of pixels in the image. Then, the distance measure between the clustering center i and the pixel j within the block is given by

$$D_{i,j} = \sqrt{\left(\frac{D_{i,j}^{(S)}}{Q}\right)^2 + \left(\frac{D_{i,j}^{(C)}}{\omega}\right)^2} \quad (2)$$

where $D^{(S)}$ and $D^{(C)}$ are the spatial distance and the color distance, respectively, while ω is the maximum color distance within a given cluster. With spatial position $[x, y]$ and CIELAB color $[a, b, c]$, the spatial distance and the color distance can be calculated by

$$D_{i,j}^{(S)} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (3)$$

$$D_{i,j}^{(C)} = \sqrt{(a_i - a_j)^2 + (b_i - b_j)^2 + (c_i - c_j)^2}. \quad (4)$$

After the initial clustering, the clustering centers are iteratively updated in accordance with the mean values of the distance measures in the corresponding clustering blocks until the clustering centers of each pixel no longer change.

3) *Simple Noniterative Clustering (SNIC)*: SNIC is an improved version of the SLIC superpixel segmentation with the use of the same distance measure as defined in (2). However, unlike SLIC, SNIC clusters pixels without using the K-means iterations while explicitly enforcing connectivity from the beginning. Starting from the initial centers, it uses a priority queue to choose the next pixel to add to a cluster, then selects the pixel with the smallest distance from the center

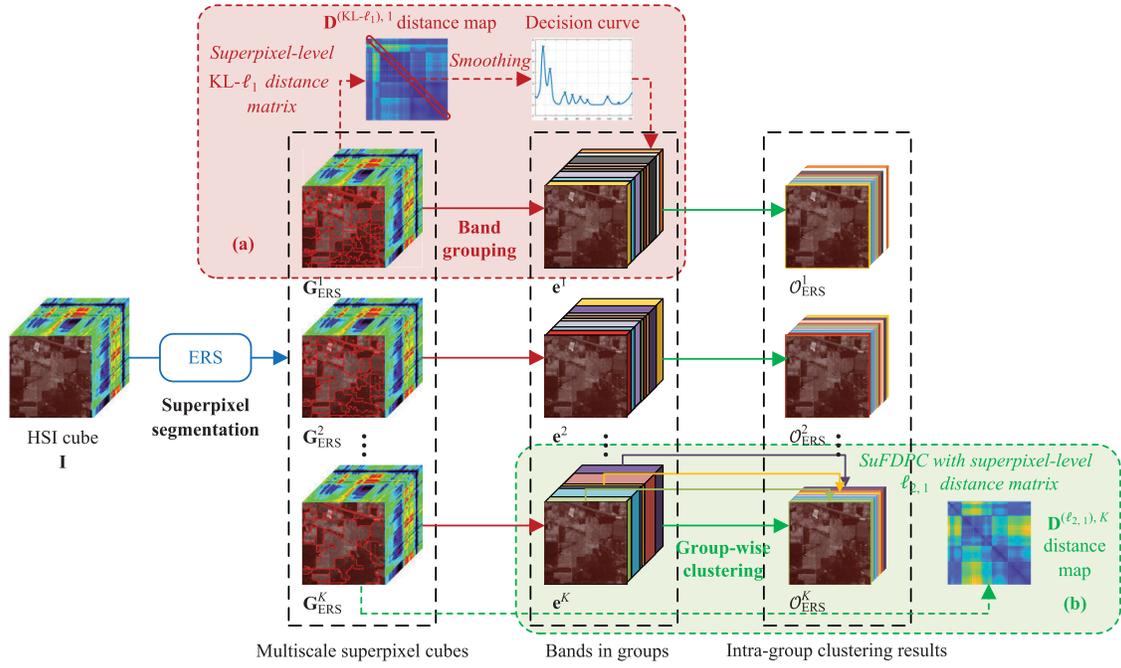


Fig. 2. Schematic of the multiscale superpixel-level group clustering procedure (taking the ERS method as an example). First, The $KL-\ell_1$ and $\ell_{2,1}$ distance maps between all bands are calculated using the superpixel cubes segmented by ERS at K scales, separately. (a) All bands are adaptively separated into groups with the maximum points of $KL-\ell_1$ distance between adjacent bands as partition points. (b) Band clustering using SuFDPC with $\ell_{2,1}$ norm is performed to select the most informative bands from each group.

as a candidate. Finally, the element-adjacent pixel is labeled and pushed into the priority queue, and this process is repeated until the priority queue is empty.

B. Density Peak-Based Fast Clustering

The density peak-based fast clustering algorithm (FDPC) is an effective algorithm to find cluster centers based on local density ρ_i and distance δ_i from points of higher density [92]. Points with relatively high ρ_i and δ_i tend to be the cluster centers according to FDPC. Respectively, ρ_i and δ_i can be calculated by

$$\rho_i = \sum_j \chi(\mathbf{D}_{i,j} - d_c) \quad (5)$$

$$\delta_i = \begin{cases} \max_j (\mathbf{D}_{i,j}), & \text{if } \rho_i = \max(\rho) \\ \min_{j:\rho_j > \rho_i} (\mathbf{D}_{i,j}), & \text{otherwise} \end{cases} \quad (6)$$

where d_c is the cutoff distance, $\mathbf{D}_{i,j}$ is the distance between the i th and the j th band, $\chi(x) = 1$ if $x < 0$, and $\chi(x) = 0$ otherwise. With spectral bands as points to be selected, FDPC has been introduced to hyperspectral BS. Jia *et al.* [63] proposed an enhanced version of FDPC (EFDPC), which calculates the weight of each band by weighting the normalized local density and the distance within the cluster, making the bands have different weight, and outperforms the FDPC for hyperspectral BS. The new calculation formula of local density ρ_i is defined as

$$\rho_i = \sum_{i=1, i \neq j}^B \exp\left(-\left(\frac{\mathbf{D}_{i,j}}{d_c}\right)^2\right) \quad (7)$$

where B is the number of bands, and ρ_i and δ_i are min-max normalized by

$$\rho = (\rho - \rho_{\min}) ./ (\rho_{\max} - \rho_{\min}) \quad (8)$$

$$\delta = (\delta - \delta_{\min}) ./ (\delta_{\max} - \delta_{\min}) \quad (9)$$

where $./$ is the element-wise division operator. The ranking score ψ_i for any band i is finally obtained by

$$\psi_i = \rho_i \times \delta_i^2. \quad (10)$$

The larger ψ_i is, the more likely the i th band is to be selected as a cluster center.

III. MULTISCALE SUPERPIXEL-LEVEL GROUP CLUSTERING FRAMEWORK (MSGCF)

Fig. 1 shows the overall illustration of our proposed MSGCF for hyperspectral BS, which is mainly composed of three steps. First, three superpixel segmentation algorithms, including ERS, SLIC, and SNIC, are applied to generate the superpixel representation of HSI data in multiscales. Second, the group-clustering paradigm is developed, which is capable of exploiting local contextual information more effectively. Meanwhile, the schematic of multiscale superpixel-level group clustering procedure (taking the ERS method as an example) is displayed in Fig. 2. In this procedure, superpixel cubes are utilized to compute the superpixel-wise KL divergence and ℓ_2 norm between bands, after which the ℓ_1 norms of the superpixel-wise KL divergence and ℓ_2 norm are calculated to acquire the $KL-\ell_1$ and $\ell_{2,1}$ distance maps between all spectral bands. Third, all bands are adaptively separated into groups with the maximum points of $KL-\ell_1$ distance between

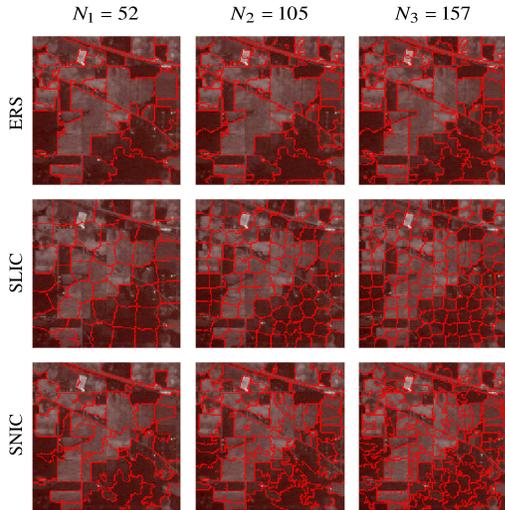


Fig. 3. Segmentation maps with different number of superpixels (N_1 , N_2 , and N_3) by ERS (first row), SLIC (second row), and SNIC (third row) on the IP HSI dataset.

adjacent bands as partition points, and then band clustering using SuFDPC with $\ell_{2,1}$ norm is performed to select the most informative bands from each group. Finally, the selected bands at various scales are merged by metric ranking-based voting to obtain the final BS results. With all these carefully designed strategies, MSGCF has the ability to pick out the most representative bands with strong discrimination.

A. Multiple and Multiscale Superpixel Segmentation

It is meaningful to incorporate superpixel segmentation into HSI band distance computation since the homogeneity within superpixels allows the introduction of spatial structural information of materials. Since three superpixel segmentation algorithms are originally designed for RGB images, PCA processing is applied on HSI before performing superpixel segmentation, downscaling them to three dimensions. Specifically, the first three principal components were used to map the raw hyperspectral data into 3-D space, which preserved the primary hyperspectral information and generated the base image for superpixel segmentation.

To maximize the effect of superpixel, the generated superpixel representations need to have properties such as boundary adhesion, compactness, and regularity. However, all these semantic descriptions cannot be characterized by a single superpixel segmentation algorithm. Specifically, ERS has good boundary adhesion because of the segmented superpixels with shape adaptability, while SLIC and SNIC exhibit good compactness and regularity as the algorithms enforce connectivity from the start. Fig. 3 shows the segmentation maps with different number of superpixels by ERS, SLIC and SNIC on the widely used Indian Pines (IP) HSI dataset. It can be observed that these three superpixel segmentation algorithms usually result in different styles of superpixel shapes. Furthermore, various HSI datasets can have different feature classes, different feature classes contain different spectral features, and the features after dimensionality reduction are also different, so the results of superpixel segmentation methods acting on

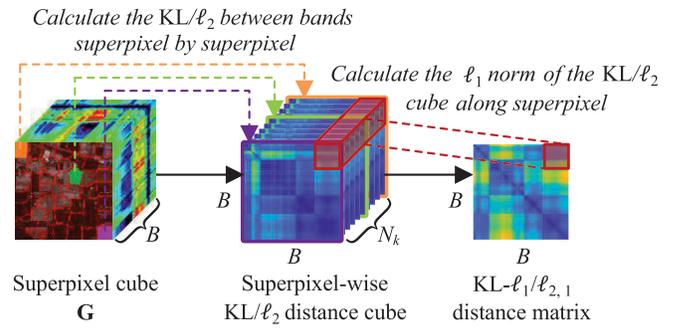


Fig. 4. Computation process of the superpixel-level $\text{KL-}\ell_1$ and $\ell_{2,1}$ distance measure. The superpixel-based $\text{KL-}\ell_1$ and $\ell_{2,1}$ distances first consider each homogeneous region of the superpixel segmentation as a whole, and then calculate the KL divergence or ℓ_2 distance between all bands in each superpixel to obtain a distance matrix (2-D, $B \times B$). These matrices with each superpixel region are stacked together to obtain the 3-D distance cube $B \times B \times N_k$. Finally, ℓ_1 norm of the 3-D distance cube is derived to obtain the distance matrix (2-D, $B \times B$) between the entire image bands.

different datasets are not the same. There are still not a general rule to decide which superpixel segmentation methods to use on a certain dataset. Therefore, in this study, a combination of multiple superpixel segmentation methods is used to cope with different types of datasets to obtain superpixel segmentation maps with richer structure information.

Likewise, since the scales of various land cover types in spatial domain are usually totally different from each other, a single scale of superpixel is clearly not sufficient to accommodate diverse ground distributions. Therefore, multiscale superpixel segmentation is introduced by performing superpixel segmentation procedure with different number of superpixels, as displayed in Fig. 3. As for the inconsistent acquisition specifications of different HSI datasets, the number of superpixels extracted at each scale should vary according to the spatial dimension and spatial resolution of HSI. Concretely, for an HSI cube $\mathbf{I} \in \mathbb{R}^{X \times Y \times B}$, where X , Y represent the spatial dimensions, and B is the number of bands, the number of superpixels at the k th scale can be calculated by

$$N_k = \left\lfloor \frac{k \times X \times Y}{T^2} \right\rfloor, \quad k = 1, \dots, K \quad (11)$$

where T represents the spatial resolution of HSI, and parameter K denotes the total number of superpixel scales. In the experiments, K is set as 3 to make a trade-off between classification accuracy and computational complexity. Fig. 3 shows the segmentation maps on the IP HSI dataset at three different superpixel scales, N_1 , N_2 , and N_3 .

Let $\mathbf{G} \in \mathbb{R}^{X \times Y \times B}$ denote the achieved superpixel cube (as shown in Fig. 4, the dimension of \mathbf{G} is the same as \mathbf{I} that contains the structural information of materials), and the corresponding superpixel cubes generated by ERS, SLIC, and SNIC at the k th scale are, respectively, denoted as $\mathbf{G}_{\text{ERS}}^k$, $\mathbf{G}_{\text{SLIC}}^k$ and $\mathbf{G}_{\text{SNIC}}^k$ ($k = 1, \dots, K$), which are used in Fig. 1 and 2. Since the subsequent processing of superpixel cube of three methods is basically similar to each other, for the sake of brevity, we omit the subscripts and use \mathbf{G}^k in the following description directly. Besides, the n th superpixel at the k th scale

is formulated as $\mathbf{G}_n^k = [g_{n,1}^k, \dots, g_{n,w}^k, \dots, g_{n,W_n^k}^k] \in \mathbb{R}^{W_n^k \times B}$, where $g_{n,w}^k \in \mathbb{R}^B$ is the w th pixel of the n th superpixel, and W_n^k is the total number of pixels in the n th superpixel.

KL divergence and ℓ_2 norm are two widely used distance measures between bands. Normally, they are calculated in pixel level. Alternatively, in the proposed MSGCF method, KL divergence and ℓ_2 norm measures between bands are computed in superpixel level. And then, ℓ_1 norm is adopted to integrate the superpixel-wise KL divergence and ℓ_2 norm, obtaining the KL- ℓ_1 and the $\ell_{2,1}$ distance matrix between bands, which introduces the structural information within superpixel homogeneous regions. The calculation procedure of KL- ℓ_1 and $\ell_{2,1}$ distance matrix is formulated as

$$\mathbf{D}_{i,j}^{(\text{KL-}\ell_1),k} = \sum_{n=1}^{N_k} \sum_{w=1}^{W_n^k} \left| p_{n,w,i}^k \ln \frac{p_{n,w,i}^k}{p_{n,w,j}^k} \right| \quad (12)$$

$$\mathbf{D}_{i,j}^{(\ell_{2,1}),k} = \sum_{n=1}^{N_k} \sqrt{\sum_{w=1}^{W_n^k} (g_{n,w,i}^k - g_{n,w,j}^k)^2} \quad (13)$$

where $\mathbf{D}_{i,j}^{(\text{KL-}\ell_1),k}$ and $\mathbf{D}_{i,j}^{(\ell_{2,1}),k}$, respectively, represent the KL- ℓ_1 and $\ell_{2,1}$ distance between the i th and the j th bands of the n th superpixel at the k th scale, and $p_{n,w,i}^k$ and $p_{n,w,j}^k$ are the histogram statistics of the w th pixel in the n th superpixel of the i th and the j th bands. Fig. 4 visually illustrates the calculation process of the superpixel-level KL- ℓ_1 and $\ell_{2,1}$ distance measure. Evidently, superpixel-wise calculation process is capable of preserving the spatial structure of HSI data, which helps to better represent the correlation between bands.

B. Adaptive Band Grouping

High similarity usually exists among adjacent bands in HSI. Here, KL- ℓ_1 defined in (12) is utilized to divide the spectral bands into multiple groups. The larger the KL- ℓ_1 values, the greater the irrelevance between these two bands. As displayed in Fig. 2(a), the maximum points of KL- ℓ_1 correlation of adjacent bands are regarded as the partition points, which is utilized to divide the bands into appropriate groups after performing a Gaussian smooth filtering step on the KL- ℓ_1 decision curve. More precisely, the KL- ℓ_1 distance measure of the k th scale is expressed as

$$\mathbf{M}^k(i) = \mathbf{D}_{i,i+1}^{(\text{KL-}\ell_1),k}, \quad i = 1, 2, \dots, B-1 \quad (14)$$

where \mathbf{M}^k represents the KL- ℓ_1 distance of adjacent bands, and $\mathbf{M}^k(i)$ denotes the i th value of \mathbf{M}^k . After applying Gaussian smoothing operation on \mathbf{M}^k , the smoothed curve \mathbf{M}_S^k is achieved as

$$\mathbf{M}_S^k = \mathbf{M}^k \otimes \mathbf{S} \quad (15)$$

where \otimes is the convolution operation. \mathbf{S} is the Gaussian kernel, which is formulated as

$$\mathbf{S}(i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{i^2}{2\sigma^2}} \quad (16)$$

where parameter σ is set as the variance of the \mathbf{M}^k . Fig. 5 illustrates the KL- ℓ_1 distance measure before (orange curve) and after (blue curve) Gaussian smoothing. The partition

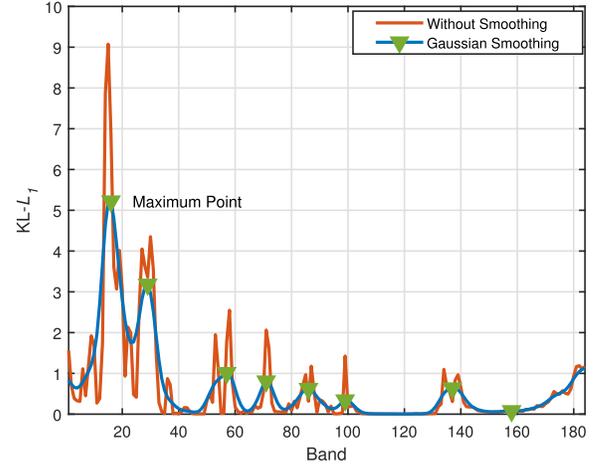


Fig. 5. Decision curves of band grouping. Before band grouping, Gaussian smoothing is performed on the superpixel-level distance KL- ℓ_1 of adjacent bands to ensure reasonable grouping while alleviating the influence of noise. The figure illustrates the KL- ℓ_1 distance measure before (orange curve) and after (blue curve) Gaussian smoothing, and the partition points are marked by green inverted triangles.

points are marked by green inverted triangles, and the partition points at the k th scale can be determined by

$$\mathbf{e}^k(u) = \begin{cases} 0, & \text{if } u = 1 \\ i, & \text{if } u > 1, \mathbf{M}_S^k(i) > \mathbf{M}_S^k(i-1) \text{ and} \\ & \mathbf{M}_S^k(i) > \mathbf{M}_S^k(i+1), i \in \mathbb{Z}, \\ & \mathbf{e}^k(u-1) < i < B \end{cases} \quad (17)$$

$$U = \text{length}(\mathbf{e}^k) \quad (18)$$

$$\mathbf{e}^k(U+1) = B \quad (19)$$

where U is the number of groups after band grouping. It can be easily found from (17) that the number of groups is adaptive to various HSI data rather than determined by hand, and thus the generalization of the proposed band grouping method is ensured. Meanwhile, the discriminative ability of following clustering procedure can also be enhanced.

C. Superpixel-Level Fast Density Peak Clustering (SuFDPC)

After band neighborhood grouping, all bands are separated into groups, as shown in Fig. 2(b). To pick out the most representative and informative bands from each group, we proposed a superpixel-level fast density peak clustering (SuFDPC) algorithm, which is more suitable than EFDPC for group clustering with limited bands. More precisely, SuFDPC has three enhancements over EFDPC. First, traditional Euclidean distance is replaced with superpixel-level $\ell_{2,1}$ distance measure [as computed in (13)], which could well represent the spatial homogeneity of materials. It has the ability to better evaluate the intraclass similarity and interclass variability, and the local density ρ_i in the u th group is calculated by

$$\rho_i = \sum_{i,j=\mathbf{e}^k(u)+1, i \neq j}^{\mathbf{e}^k(u+1)} \exp\left(-\left(\frac{\mathbf{D}_{i,j}^{(\ell_{2,1}),k}}{d_e}\right)^2\right) \quad (20)$$

where d_e is the cutoff threshold. Second, it can be easily found that the cutoff distance d_e has a critical impact on ρ_i , which should change with the group situation. Let l_e denotes the length of the group, and the length of the u th group can be directly computed by

$$l_e(u) = \mathbf{e}^k(u+1) - \mathbf{e}^k(u), \quad u = 1, 2, \dots, U. \quad (21)$$

If the value of l_e is higher, more bands should be selected, and the value of d_e should be correspondingly smaller. Therefore, a heuristic manner for d_e is proposed as follows:

$$d_e = \frac{d_{\text{init}}}{\exp(l_e/B)} \quad (22)$$

where d_{init} represents the initial value of the cutoff distance, which is empirically defined as the value of position at $2\% \times l_e \times (l_e - 1)$ of the sorted sequence from low to high in $\mathbf{D}^{(\ell_{2,1}),k}$ (excluding elements of zero). Third, although EFDPC can select bands with weak correlation, the selected bands are not informative enough. To alleviate the above problem, we introduce information entropy H_i as a factor to update the clustering ranking score ψ_i , which considers both correlation and representability of bands. Concretely, the ranking score ψ_i and information entropy of the i th band H_i are, respectively, defined by

$$\psi_i = \rho_i \times \delta_i \times H_i \quad (23)$$

$$H_i = \sum_{i=1}^B p_i \log p_i \quad (24)$$

where p_i is the histogram statistic of the gray level of the i th band. Similarly, ρ_i , δ_i , and H_i are normalized to $[0, 1]$ as (8) and (9). After applying the proposed SuFDPC on all the U groups, the intragroup clustering results is represented as \mathcal{O} . As shown in Fig. 1 and 2, the intragroup clustering results of ERS, SLIC, and SNIC at the k th scale are, respectively, denoted as $\mathcal{O}_{\text{ERS}}^k$, $\mathcal{O}_{\text{SLIC}}^k$, and $\mathcal{O}_{\text{SNIC}}^k$.

D. Metric Ranking-Based Voting

With respect to the BS procedure of our MSGCF, the superpixel-level distance matrices between adjacent bands at multiple scales are usually different, resulting in different band grouping results. Furthermore, the bands selected by SuFDPC at each group are also different from each other. To obtain the final BS results, we introduce a voting strategy by joint considering information entropy H_i [defined in (24)] and frequency of occurrence f_i , which encourages more informative bands to be selected. The final selected bands are determined by a metric ranking strategy, and the voting metric V_i is defined as

$$V_i = f_i \times H_i. \quad (25)$$

As shown in Fig. 1, the intragroup clustering results have a total number of $3K$ sets of band results. We select the required number of bands according to the value of V_i in descending order to achieve the final result \mathcal{O}' .

IV. EXPERIMENTS

To estimate the performance and verify the validity of our proposed MSGCF, four real-world HSI datasets with different spatial resolutions, i.e., IP dataset, Kennedy Space Center (KSC) dataset, Houston 2013 (HU) dataset, and Salinas Valley (SV) dataset were employed to conduct the experiments.

Several popular BS techniques, including two ranking-based methods (ID [51] and MVPCA [52]), two clustering-based methods (K-centers [93] and AP [59]), three methods combining ranking-based and clustering-based strategies (FDPC [92], EFDPC [63] and OCF [66]), two grouping-clustering method (ASPS [68] and FNGBS [69]), uniform BS (UBS [52]), in which the bands are selected uniformly, and three deep learning-based methods (ACNN [53], BSNet [54], and DAREcNet [55]) were selected as comparison methods. Regarding the classification manner, we used three common classifiers to evaluate the representativeness of the selected bands, i.e., K-Nearest Neighbor (KNN), Random Forest (RF), and Support Vector Machine (SVM) classifiers. Since classifiers are only used to justify the representativeness and importance of selected bands, we believe that different parameters of classifiers have the same impact on the various compared BS methods. Hence, the number of neighbors in KNN is fixed as 3, the number of decision trees in RF is fixed as 500 in the revision, while the parameters of the SVM were set as the default of library for support vector machine (LIBSVM) [94].

We verified the validity and robustness of the compared methods from two perspectives, i.e., different number of bands and different number of training samples. For experiments with different numbers of bands, we randomly selected ten labeled samples per class according to Tables I–IV as training set and the remainder as test samples. With respect to experiments with different number of training samples, ten bands were fixedly selected from each dataset. The experimental results were averaged by ten independent runs to reduce the randomness. Before presenting the detailed results, the number of ten bands chosen by different methods on the four datasets is listed in Table V, which are utilized to conduct the following comparison. It is worth pointing out that the chosen bands are kept unchanged during the classification procedure since all the BS methods are in an unsupervised manner and unrelated to the training samples.

In terms of the criteria metrics, overall accuracy (OA) and Kappa coefficient with standard deviation were used as measures of accuracy in the experiments. Specifically, OA is the ratio between the correct prediction of the model on all test sets and the total number, and the Kappa coefficient is a statistic widely used to measure the agreement of classification results [95], which represents the ratio of error reduction between classification and completely random classification.

A. Indian Pines Dataset (IP)

The first dataset we used in the experiments was captured by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the agricultural area of northwestern Indiana in 1992. The dataset covers a scene of 145×145 pixels with

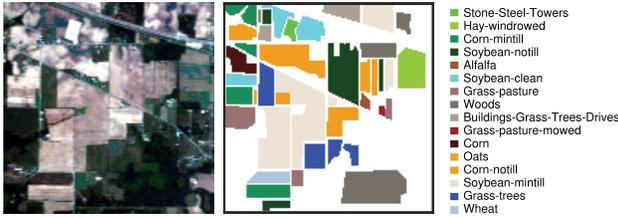


Fig. 6. False-color map and ground-truth map of the IP dataset (16 land cover types).

TABLE I
LAND COVER TYPES WITH NUMBER OF SAMPLES PER CLASS FOR THE IP HSI

Class	Land Cover Type	No. of Samples
C1	Stone-steel towers	95
C2	Hay-windrowed	489
C3	Corn-min	834
C4	Soybeans-notill	968
C5	Alfalfa	54
C6	Soybean-clean Till	614
C7	Grass/Pasture	497
C8	Woods	1294
C9	Bldg-Grass-Tree-Drives	380
C10	Grass/Pasture-mowed	26
C11	Corn	234
C12	Oats	20
C13	Corn-notill	1434
C14	Soybeans-min	2468
C15	Grass/Trees	747
C16	Wheat	212
	Total	10366

a spatial resolution of 20 m per pixel. The imaging wavelength range of the AVIRIS imaging spectrometer is 400–2500 nm, and it sequentially images ground objects in 220 continuous bands. After discarding 20 bands that could not be reflected by water, 200 bands were retained for experiments. Fig. 6 and Table I provide a detailed display of the dataset containing 10366 labeled pixels and 16 ground-truth classes, most of which are different types of crops, including one-third of forest and two-thirds of agriculture or other natural perennial vegetation.

Fig. 7 reports the classification accuracy of all the compared methods by selecting 5–50 bands on the IP dataset. Meanwhile, the accuracy with all bands is also given, as illustrated in the dotted line with black color. In this figure, it can be easily seen that the ranking-based methods (ID and MVPCA) always had lower classification accuracy than the clustering-based methods, since the bands selected by the ranking-based methods are usually highly correlated. Similarly, the ACNN algorithm performs not well when the number of bands is small, which is much improved when the number of selected bands becomes large. With respect to the UBS method, the classification performance is not satisfactory since the internal spectral characteristics of various materials have not been well exploited. In addition, the grouping-based methods (FNGBS) performed better than three deep learning-based methods (ACNN, BSNet, and DARecNet). However, the increase in the number of bands does not necessarily lead to an increase in classification accuracy, due to the redundancy and noise introduced by the additional bands. Nevertheless, our MSGCF

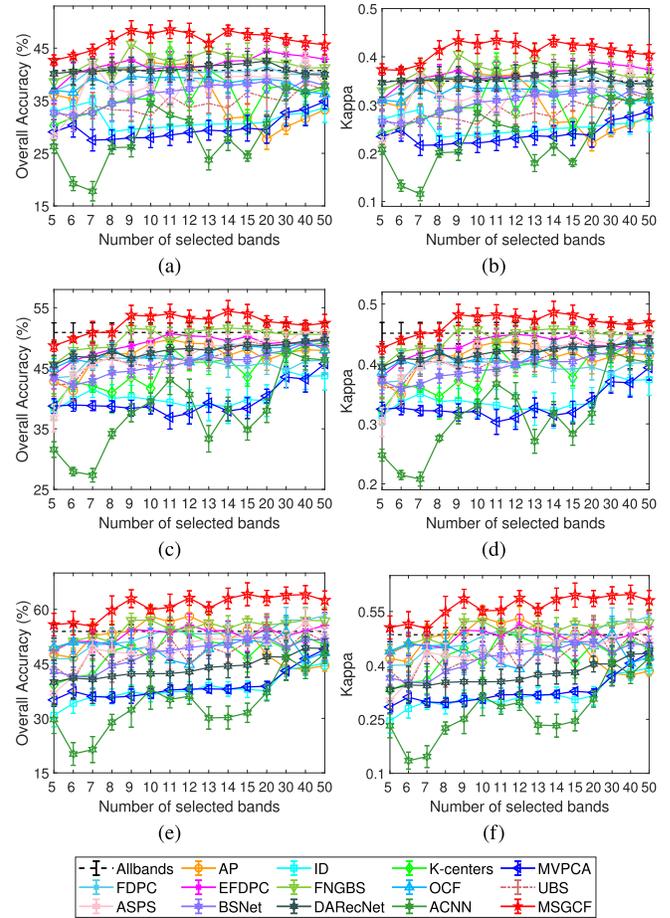


Fig. 7. Performance versus the number of selected bands of compared methods on the IP dataset with (a) OA and (b) Kappa using KNN classifier, (c) OA and (d) Kappa using RF classifier, and (e) OA and (f) Kappa using SVM classifier.

always provides the best in most cases on the dataset, validating the effectiveness of the proposed approach.

To verify the robustness of the proposed MSGCF method, Fig. 8 displays the effect of the training samples per class ranging from 3 to 50 on the classification accuracy of the IP dataset. The rising curves prove that the larger the number of training samples, the greater the classification accuracy. Similar to the above, it can be clearly found from these figures that the grouping-based strategy plays an important role in hyperspectral BS in most cases, and our MSGCF constantly maintains the highest classification accuracy compared with the other methods.

Moreover, band correlation is adopted to measure the representativeness of the selected bands. The correlation between the selected bands is computed by [63]

$$\text{Cor} = 1 - \frac{2}{b(b-1)} \sum_{i,j=1,i < j}^b \mathbf{D}_{C_i, C_j} \quad (26)$$

where b is the number of selected bands, and C are the indexes of the selected bands. The smaller the value, the weaker the correlation, indicating the effectiveness of the BS method. The results are demonstrated in Fig. 9, in most cases, the bands selected by our MSGCF present low correlations. In contrast, MVPCA and ID methods select bands with high correlations,

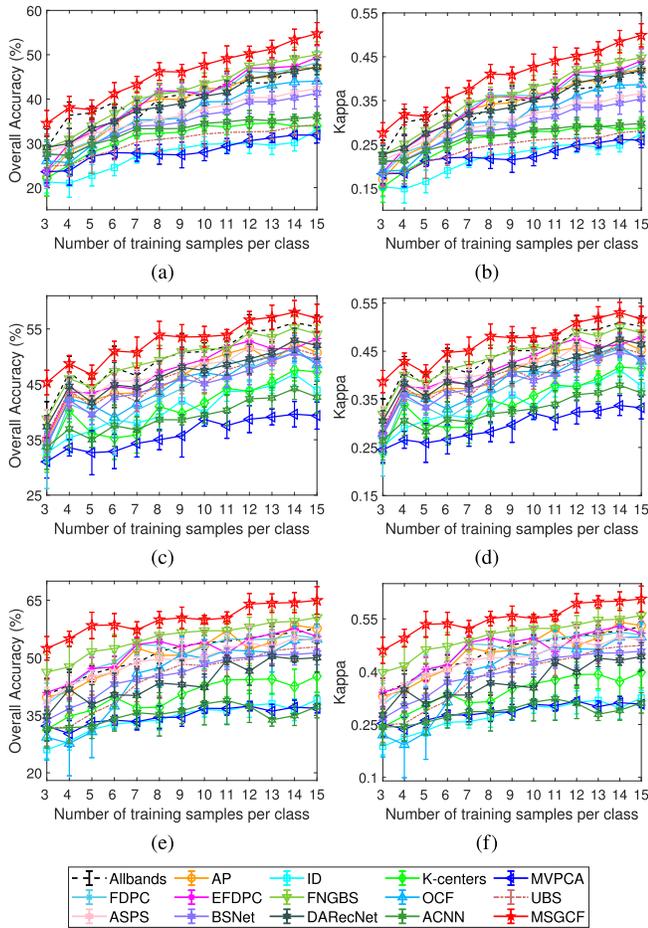


Fig. 8. Performance versus the number of training samples per class of compared methods on the IP dataset with (a) OA and (b) Kappa using KNN classifier, (c) OA and (d) Kappa using RF classifier, and (e) OA and (f) Kappa using SVM classifier.

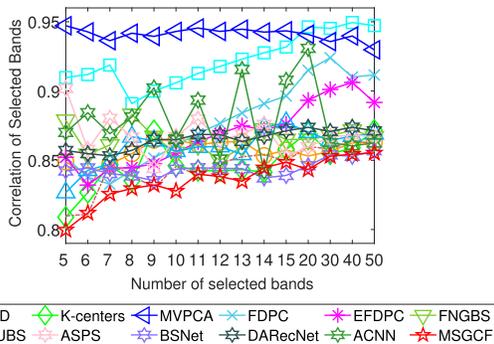


Fig. 9. Correlation measures of different numbers of selected bands by compared methods on the IP dataset.

which is consistent with the basic characteristics that they ignore correlations in the selected bands.

B. Kennedy Space Center Dataset (KSC)

The second dataset was captured from an altitude of approximately 20 km by the AVIRIS sensor at the KSC in Florida on March 23, 1996 [96]. The spectral range of the KSC dataset is 400–2500 nm with 224 bands and a geometric resolution of 18 m per pixel. After removing water absorption and low



Fig. 10. False-color map and ground-truth map of the KSC dataset (13 land cover types).

TABLE II
LAND COVER TYPES WITH NUMBER OF SAMPLES
PER CLASS FOR THE KSC HSI

Class	Land Cover Type	No. of Samples
C1	Cabbage Palm Hammock	256
C2	Cabbage Palm-oak Hammock	252
C3	Cattail Marsh	404
C4	Graminoid Marsh	431
C5	Hardwood Swamp	105
C6	Mud Flats	503
C7	Oak-broadleaf Hammock	229
C8	Salt Marsh	419
C9	Scrub	761
C10	Slash Pine	161
C11	Spartina Marsh	520
C12	Water	927
C13	Willow Swamp	243
	Total	5211

SNR bands, 176 bands were used for the analysis in the experiments. The dataset contains 421×444 pixels, and there are 5211 labeled samples from 13 different land cover types in the image (see Fig. 10). The number of samples in each class is explicitly listed in Table II.

The classification accuracy increases with the number of bands and training samples on the KSC dataset, which can be found in Fig. 11 (as the classification accuracy results of Kappa are similar to OA, the remaining three datasets only show the classification accuracy results of OA). Of these, our MSGCF method consistently holds the best results, with the two grouping-based methods (ASPS and FNGBS) coming in just behind it. The deep learning-based method BSNet performed inconsistently, with good and bad results, while the DARecNet and ACNN methods showed unsatisfactory results on this dataset. In addition, Fig. 12 displays the band correlations for various band selected methods on the KSC dataset. It can be seen that ID, MVPCA, and DARecNet methods have the strongest correlations, while MSGCF method has a weaker and stable correlation, indicating the representativeness of bands selected by our proposed method.

C. Houston 2013 Dataset (HU)

The third dataset was acquired by the ITRES Compact Airborne Spectrographic Imager (CASI)-1500 sensor, provided by the 2013 IEEE Geoscience and Remote Sensing Society (GRSS) Data Fusion Competition [97]. The image scene was captured over the University of Houston campus and its adjacent areas. It covers a spectral range of 380–1050 nm, with 144 bands. The spatial size of the dataset is 349×1905 , and the ground-sampling distance

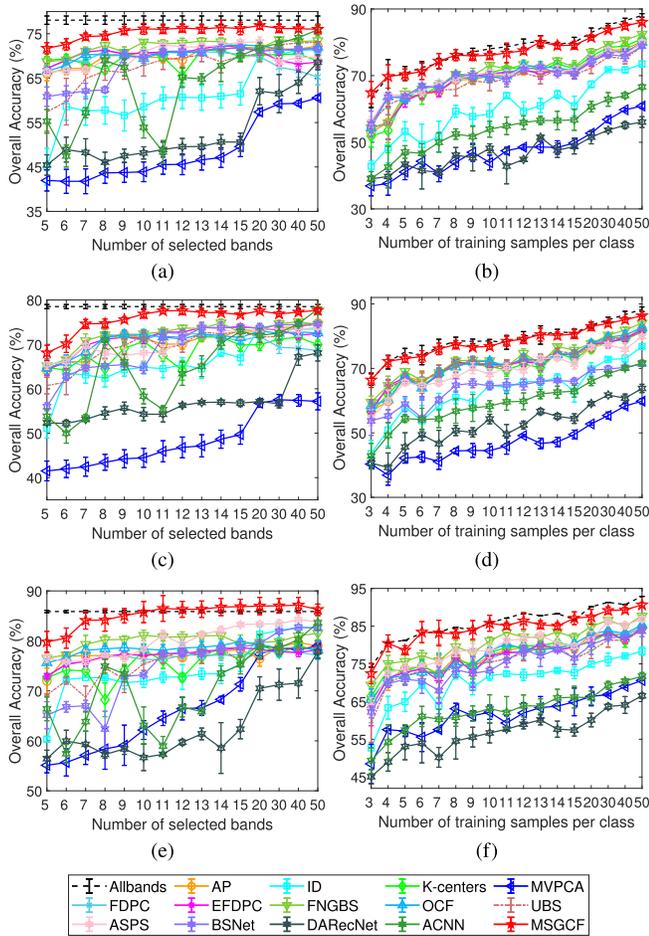


Fig. 11. OA (%) results of compared methods versus the number of selected bands using (a) KNN, (c) RF, and (e) SVM classifier, and versus the number of training samples per class using (b) KNN, (d) RF, and (f) SVM classifier on the KSC dataset.

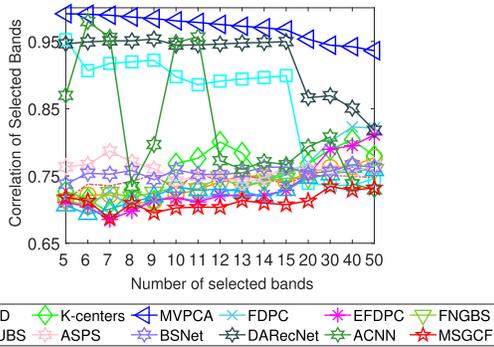


Fig. 12. Correlation measures of different numbers of selected bands by compared methods on the KSC dataset.

is 2.5 m per pixel. A total number of 15029 samples from 15 land covers were labeled (see Fig. 13). The number of samples per class is clearly listed in Table III.

Fig. 14 separately reports the OA of all the compared methods by selecting different numbers of bands and training samples per class on the HU dataset. For most methods, it is logical that the classification accuracy increases as the number of samples increases. As with the previous datasets, ID and MVPCA methods show the worst results, while the curve of

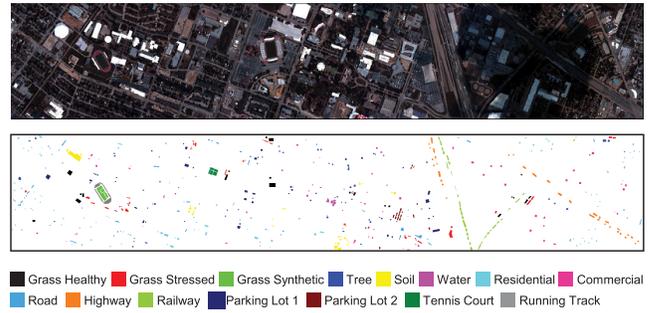


Fig. 13. False-color map and ground-truth map of the HU dataset (15 land cover types).

TABLE III
LAND COVER TYPES WITH NUMBER OF SAMPLES PER CLASS FOR THE HU HSI

Class	Land Cover Type	No. of Samples
C1	Healthy Grass	1251
C2	Stressed Grass	1254
C3	Synthetic Grass	697
C4	Trees	1244
C5	Soil	1242
C6	Water	325
C7	Residential	1268
C8	Commercial	1244
C9	Roads	1252
C10	Highways	1227
C11	Railways	1235
C12	Parking Lot 1	1233
C13	Parking Lot 2	469
C14	Tennis Court	428
C15	Running Track	660
	Total	15 029

ACNN method is very unstable. Evidently, the classification results of most clustering-involved methods are relatively close, but they have different degrees of fluctuation, whereas MSGCF provides the most stable and accurate performance, indicating that our method is a valid option for HSI BS.

D. Salinas Valley Dataset (SV)

The last dataset is an image of the SV in California, USA, which was also acquired by the AVIRIS sensor. The coverage area of the SV dataset consists of 512×217 samples, with 224 spectral bands and a spatial resolution of 3.7 m per pixel. Likewise, we generally use images from 204 bands after removing the 20 bands that cannot be reflected by water. In the image, 54 129 pixels were labeled and the ground truth map of all labeled samples is shown in Fig. 15, with 16 land cover categories provided, details of which are given in Table IV.

The classification accuracy results for the SV dataset are shown in Fig. 16. We can find that a few bands and training samples can also give decent classification accuracy results for most of the methods, mainly due to the high spatial resolution of the SV dataset. Likewise, the MSGCF method remains a better performer in most cases.

Table VI quantitatively lists the classification accuracy of various compared methods on four HSI datasets, where ten bands are selected. Evidently, the proposed MSGCF remains highly advantageous over the compared ones. Since more

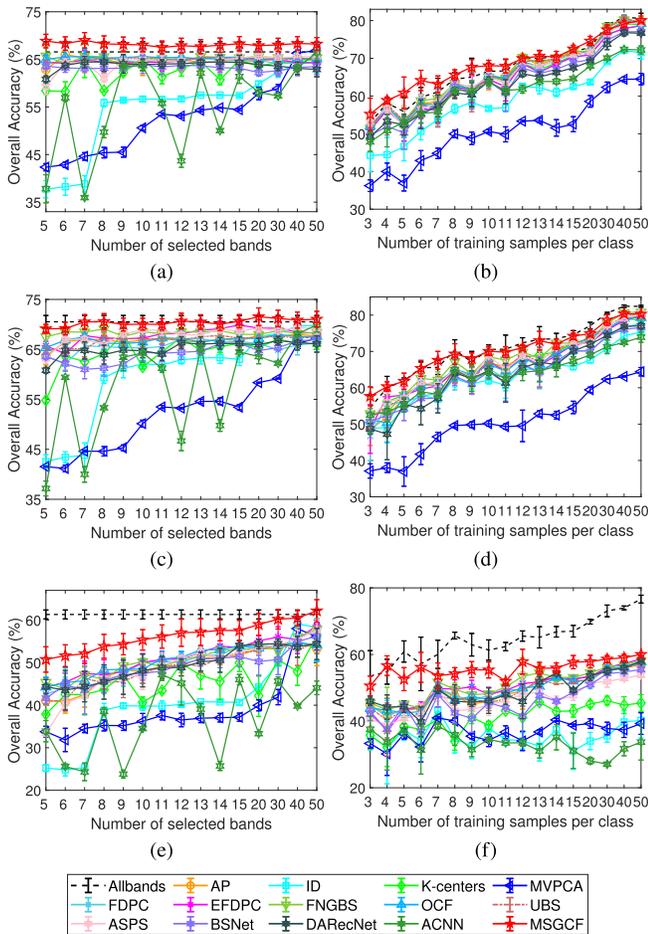


Fig. 14. OA (%) results of compared methods versus the number of selected bands using (a) KNN, (c) RF, and (e) SVM classifier, and versus the number of training samples per class using (b) KNN, (d) RF, and (f) SVM classifier on the HU dataset.

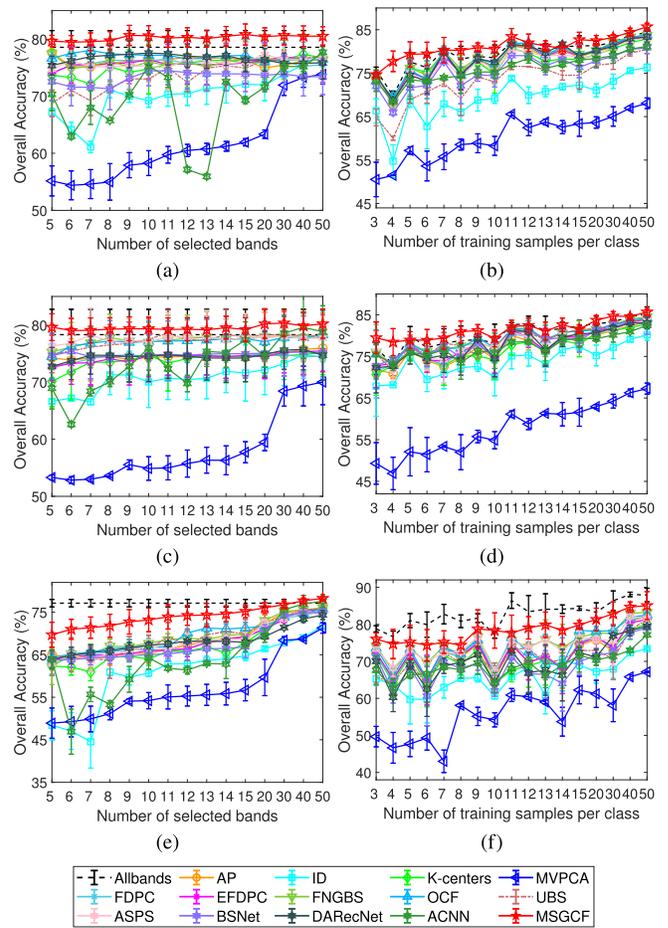


Fig. 16. OA (%) results of compared methods versus the number of selected bands using (a) KNN, (c) RF, and (e) SVM classifier, and versus the number of training samples per class using (b) KNN, (d) RF, and (f) SVM classifier on the SV dataset.

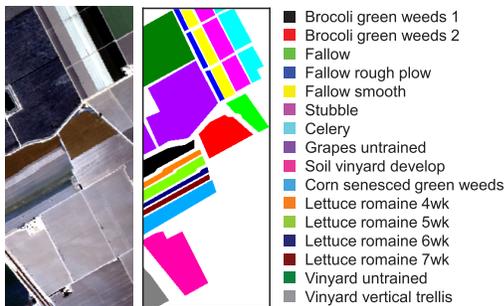


Fig. 15. False-color map and ground-truth map of the SV dataset (16 land cover types).

structural description and contextual information is embedded in the selected bands by the proposed method, these advantages allow MSGCF to select bands with strong representativeness and low redundancy.

E. Ablation Studies Toward Multiscale Superpixel Segmentation

To verify the rationality of fusing three different superpixel segmentation methods, the proposed MSGCF was compared with the versions of a single superpixel segmentation method

TABLE IV
LAND COVER TYPES WITH NUMBER OF SAMPLES PER CLASS FOR THE SV HSI

Class	Land Cover Type	No. of Samples
C1	Broccoli green weeds 1	2009
C2	Broccoli green weeds 2	3726
C3	Fallow	1976
C4	Fallow rough plow	1394
C5	Fallow smooth	2678
C6	Stubble	3959
C7	Celery	3579
C8	Grapes untrained	11 271
C9	Soil vinyard develop	6203
C10	Corn senesced green weeds	3278
C11	Lettuce romaine 4wk	1068
C12	Lettuce romaine 5wk	1927
C13	Lettuce romaine 6wk	916
C14	Lettuce romaine 7wk	1070
C15	Vinyard untrained	7268
C16	Vinyard vertical trellis	1807
	Total	54 129

and the versions combining two superpixel segmentation methods in the same framework (as shown in Fig. 2).

Figs. 17–20 show in detail the results of the evaluation metrics with various number of selected bands and number of training samples per class on the four datasets, separately,

TABLE V
TEN SELECTED BANDS USING DIFFERENT BS METHODS
ON FOUR DIFFERENT DATASETS

Dataset	Method	Selected Bands
IP	AP	15/26/48/67/95/113/124/136/163/191
	ID	88/199/35/142/151/156/164/1/170/155
	K-centers	6/18/32/40/60/83/86/89/117/189
	MVPCA	108/110/112/31/29/30/106/111/28/109
	FDPC	7/48/67/98/124/153/163/173/176/186
	EFDPC	163/48/124/98/67/71/15/186/28/152
	FNGBS	9/28/49/66/92/108/129/148/168/181
	ASPS	30/37/70/77/105/115/122/126/140/182
	OCF	42/48/29/17/54/117/89/160/166/176
	UBS	1/23/45/67/89/111/133/155/177/199
	BSNet	145/152/26/81/69/78/195/6/51/93
	DARecNet	78/37/122/84/25/174/64/14/86/168
	ACNN	34/35/66/67/68/69/70/71/198/199
MSGCF	28/125/163/50/186/136/8/48/98/65	
KSC	AP	8/22/41/58/69/83/90/113/145/152
	ID	5/28/29/30/8/156/27/6/4/134
	K-centers	6/23/33/49/71/105/113/114/131/149
	MVPCA	73/72/74/71/75/70/69/76/68/67
	FDPC	122/82/145/52/27/114/124/128/142/151
	EFDPC	122/82/145/27/151/52/71/8/92/65
	FNGBS	6/24/42/48/70/84/101/116/129/153
	ASPS	28/49/66/75/89/110/123/145/158/168
	OCF	75/97/65/44/118/107/153/161/32/12
	UBS	1/20/39/58/77/96/115/134/153/172
	BSNet	64/77/133/124/175/96/163/169/22/152
	DARecNet	172/136/102/142/175/158/143/173/137/134
	ACNN	1/2/3/9/10/11/12/13/14/15
MSGCF	83/71/52/151/27/142/120/11/125/24	
HU	AP	7/13/26/43/60/83/103/119/125/136
	ID	3/2/9/14/1/5/10/129/117/13
	K-centers	4/17/34/47/49/54/89/102/133/143
	MVPCA	52/47/54/55/53/56/46/59/48/65
	FDPC	26/40/54/64/88/92/108/116/125/138
	EFDPC	26/40/54/64/88/92/108/116/125/138
	FNGBS	5/15/22/40/61/68/89/108/124/134
	ASPS	4/22/46/57/71/79/113/117/125/132
	OCF	88/106/52/47/74/32/144/132/18/8
	UBS	1/16/31/46/61/76/91/106/121/136
	BSNet	54/78/128/37/58/130/76/120/72/21
	DARecNet	113/70/49/46/60/8/24/18/56/53
	ACNN	42/43/44/45/46/47/127/128/129/130
MSGCF	88/66/40/108/116/58/137/25/125/28	
SV	AP	23/51/70/91/109/112/127/141/170/190
	ID	111/146/107/149/155/204/202/56/55/54
	K-centers	20/76/86/107/112/114/151/167/175/187
	MVPCA	44/43/45/49/47/46/48/50/52/51
	FDPC	12/30/46/73/91/109/113/138/165/178
	EFDPC	12/30/46/73/91/109/113/127/138/165
	FNGBS	10/31/51/67/88/113/132/142/168/193
	ASPS	28/51/59/76/90/105/148/166/178/203
	OCF	45/34/58/25/15/8/93/120/125/167
	UBS	1/23/45/67/89/111/133/155/177/199
	BSNet	98/88/31/154/158/93/27/85/40/41
	DARecNet	123/92/122/31/48/103/10/50/20/85
	ACNN	13/29/30/102/103/104/109/110/111/112
MSGCF	165/31/74/184/138/14/128/92/177/45	

where RF was adopted as the classifier. As can be seen from these curves, the classification results of MSGCF, i.e., the fusion of three superpixel segmentation methods, have higher accuracy and stability than other results containing only a single or two segmentation methods. Furthermore, it can be seen from these figures that three different superpixel segmentation methods have their own advantages on different datasets. The SLIC method is better than the ERS and SNIC methods on the KSC dataset, while the ERS method can achieve better results on the HU and SV datasets, but the SNIC method only performs best on the IP dataset. This suggests that the single superpixel segmentation method is not applicable to all hyperspectral datasets. Experimental results consistently

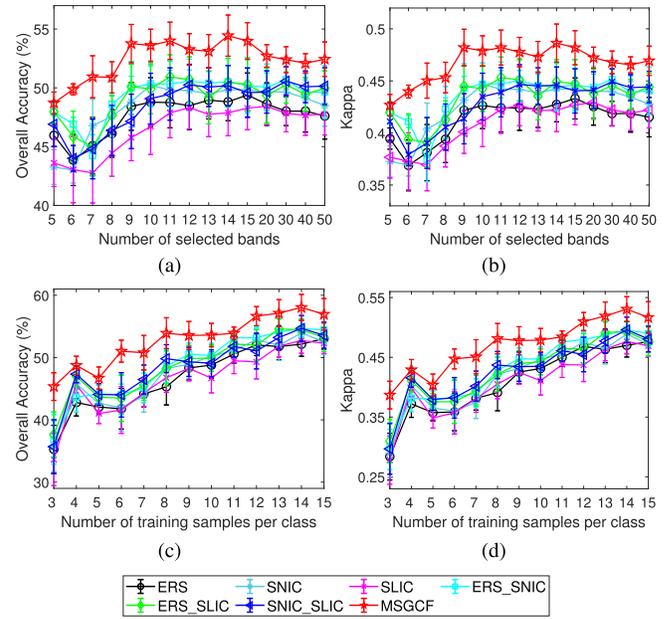


Fig. 17. Ablation results of multiscale superpixel segmentation versus the number of selected bands with (a) OA and (b) Kappa, and versus the number of training samples with (c) OA and (d) Kappa on the IP dataset using RF classifier.

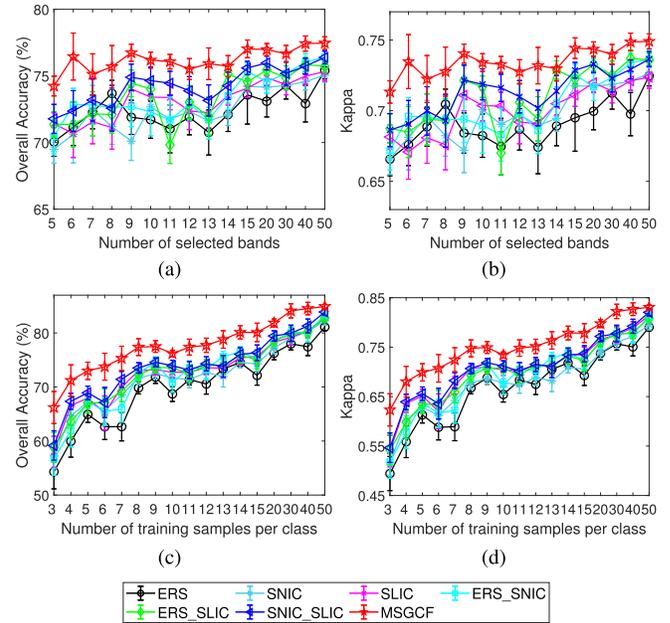


Fig. 18. Ablation results of multiscale superpixel segmentation versus the number of selected bands with (a) OA and (b) Kappa, and versus the number of training samples with (c) OA and (d) Kappa on the KSC dataset using RF classifier.

show that the fusion of the three segmentation methods is the best choice, validating the rationality of the combination of multiscale superpixel segmentation strategy.

F. Ablation Studies Toward the $KL-\ell_1$ and $\ell_{2,1}$ Superpixel-Level Distances

It is one of the main contributions of our proposed MSGCF to utilize superpixel-level distances for band grouping and

TABLE VI
CLASSIFICATION ACCURACY OF VARIOUS COMPARED METHODS ON FOUR HSI DATASETS (TEN SELECTED BANDS)

Dataset	Classifier (Measure)	AP	ID	K-centers	MVPCA	FDPC	EFDPC	FNGBS	ASPS	OCF	UBS	BSNet	DARecNet	ACNN	MSGCF
IP	KNN(OA)	42.30±1.58	29.75±1.72	35.32±1.53	27.98±1.91	41.69±1.87	40.98±2.99	43.49±2.39	37.85±0.52	39.35±1.40	32.09±2.75	36.45±2.57	40.62±2.16	34.53±1.67	47.72±1.71
	KNN(Kappa)	0.36±0.02	0.24±0.02	0.29±0.02	0.22±0.02	0.36±0.02	0.35±0.03	0.38±0.03	0.32±0.01	0.33±0.02	0.26±0.02	0.31±0.03	0.35±0.02	0.28±0.02	0.43±0.02
	RF(OA)	49.12±0.88	39.86±2.31	41.68±1.87	38.67±1.24	46.75±1.75	49.5±1.26	51.43±1.57	48.60±1.14	47.11±0.74	45.29±1.99	45.11±1.11	47.51±0.94	39.47±1.46	53.61±1.37
	RF(Kappa)	0.43±0.01	0.33±0.02	0.36±0.02	0.32±0.01	0.40±0.02	0.44±0.01	0.46±0.02	0.43±0.01	0.41±0.01	0.39±0.02	0.39±0.01	0.42±0.01	0.33±0.01	0.47±0.01
	SVM(OA)	58.03±3.35	36.53±1.98	46.59±4.33	36.82±1.39	52.99±2.69	54.09±3.10	57.04±3.71	50.13±3.86	49.86±4.87	47.96±1.32	48.24±3.30	42.28±3.46	37.97±5.47	59.93±1.15
KSC	KNN(OA)	69.85±0.84	58.55±2.99	70.91±0.94	43.91±2.21	70.56±0.51	70.56±0.55	73.15±0.46	70.94±0.50	69.39±1.50	67.88±2.42	69.81±1.43	48.16±1.38	53.91±2.38	76.09±1.23
	KNN(Kappa)	0.67±0.01	0.54±0.03	0.68±0.01	0.38±0.02	0.68±0.01	0.68±0.01	0.70±0.01	0.68±0.01	0.67±0.02	0.65±0.03	0.66±0.02	0.42±0.02	0.49±0.02	0.73±0.01
	RF(OA)	71.38±1.21	64.85±1.32	70.63±1.04	44.46±2.13	70.86±1.52	71.36±1.22	72.36±1.00	68.12±1.77	72.20±1.09	71.97±1.03	64.59±1.19	54.30±1.49	58.30±1.69	76.84±0.91
	RF(Kappa)	0.69±0.01	0.61±0.01	0.68±0.01	0.39±0.02	0.68±0.01	0.68±0.01	0.69±0.01	0.65±0.02	0.69±0.01	0.69±0.01	0.61±0.01	0.49±0.02	0.54±0.02	0.74±0.01
	SVM(OA)	77.29±0.43	71.91±1.26	78.07±1.34	62.16±1.88	76.28±0.93	77.07±0.29	81.11±1.04	78.35±1.81	78.37±0.88	75.42±1.34	73.23±0.08	56.67±2.63	63.11±3.68	85.77±2.11
HU	KNN(OA)	64.49±2.87	56.68±0.45	65.17±1.27	50.61±0.49	64.41±0.98	64.41±0.98	65.04±2.45	65.06±1.29	65.59±1.60	65.41±1.42	63.93±1.87	63.90±1.17	63.67±1.08	68.04±1.43
	KNN(Kappa)	0.62±0.03	0.53±0.01	0.62±0.01	0.46±0.01	0.61±0.01	0.61±0.01	0.62±0.03	0.62±0.01	0.63±0.02	0.63±0.02	0.61±0.01	0.61±0.01	0.61±0.01	0.66±0.02
	RF(OA)	66.86±1.56	62.09±1.05	67.45±0.84	50.14±0.70	65.26±1.09	67.26±1.09	68.54±1.50	68.86±0.35	66.45±1.70	67.15±1.60	64.42±1.70	64.72±1.34	64.28±1.86	70.05±1.18
	RF(Kappa)	0.64±0.02	0.59±0.01	0.65±0.01	0.46±0.01	0.63±0.01	0.65±0.01	0.66±0.01	0.66±0.01	0.64±0.02	0.65±0.02	0.62±0.02	0.62±0.01	0.61±0.02	0.68±0.01
	SVM(OA)	45.90±2.70	37.79±1.02	38.71±1.24	34.21±1.85	47.93±0.01	48.05±2.07	48.20±0.91	44.78±3.10	46.49±2.13	45.61±2.46	45.52±1.62	46.15±3.84	34.52±0.41	55.35±2.40
SV	KNN(OA)	76.13±3.51	69.20±1.14	74.75±4.35	58.27±2.18	76.25±3.51	76.39±3.44	76.00±3.56	75.26±2.79	77.16±2.65	73.01±0.06	74.41±2.59	77.54±2.18	75.25±1.47	80.58±1.76
	KNN(Kappa)	0.74±0.04	0.66±0.01	0.72±0.05	0.55±0.02	0.74±0.04	0.74±0.04	0.73±0.04	0.73±0.03	0.75±0.03	0.70±0.01	0.72±0.03	0.75±0.02	0.73±0.02	0.78±0.02
	RF(OA)	74.34±3.85	70.04±4.48	70.68±3.73	54.88±2.06	74.54±3.72	74.72±3.72	77.00±3.43	77.01±4.53	77.12±3.53	77.52±3.54	74.63±4.71	74.53±4.39	75.98±3.88	79.34±2.02
	RF(Kappa)	0.72±0.04	0.67±0.05	0.720±0.04	0.51±0.02	0.72±0.04	0.72±0.04	0.75±0.04	0.75±0.05	0.75±0.04	0.75±0.04	0.72±0.05	0.72±0.05	0.73±0.04	0.77±0.02
	SVM(OA)	65.34±1.92	60.69±0.83	64.03±1.53	54.20±1.90	66.20±1.65	65.94±1.37	67.68±1.89	67.18±1.35	66.78±1.99	66.31±3.28	65.05±1.48	67.26±0.76	64.28±2.67	73.18±2.78
SV	KNN(OA)	65.34±1.92	60.69±0.83	64.03±1.53	54.20±1.90	66.20±1.65	65.94±1.37	67.68±1.89	67.18±1.35	66.78±1.99	66.31±3.28	65.05±1.48	67.26±0.76	64.28±2.67	73.18±2.78
	KNN(Kappa)	0.62±0.02	0.57±0.01	0.61±0.02	0.51±0.02	0.63±0.02	0.63±0.02	0.65±0.02	0.64±0.01	0.64±0.02	0.63±0.03	0.62±0.02	0.64±0.01	0.61±0.03	0.70±0.03

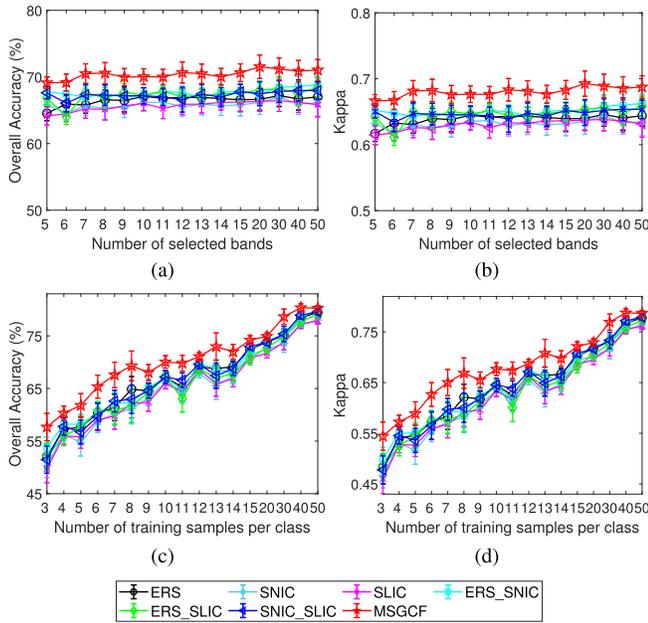


Fig. 19. Ablation results of multiscale superpixel segmentation versus the number of selected bands with (a) OA and (b) Kappa, and versus the number of training samples with (c) OA and (d) Kappa on the HU dataset using RF classifier.

clustering. To demonstrate the necessity of grouping and the effectiveness of superpixel-level distances, we also performed ablation experiments on four different datasets. Specifically, the first comparison method was used to verify the significance of grouping, i.e., no grouping was performed, the length of the group l_e was equal to B , and all bands were clustered using the superpixel-level $\ell_{2,1}$ distance. This method was abbreviated as $MSGCF_{NG}$. The second comparison method used superpixel-level KL- ℓ_1 distance to group bands and then performed intragroup clustering by the EFDPC method. Likewise, this method was named $MSGCF_{CE}$, which aimed to investigate the effectiveness of superpixel-level distances for clustering. The last one was called as $MSGCF_{GE}$, which first adopted traditional Euclidean distance for grouping

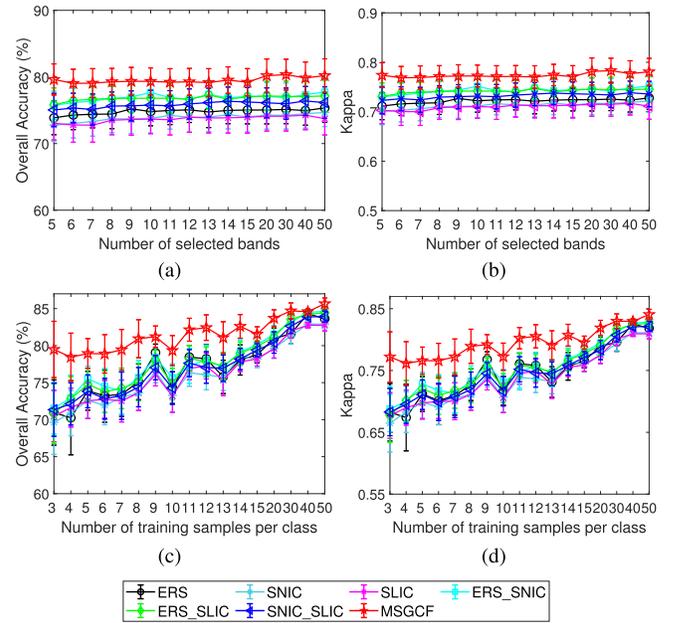


Fig. 20. Ablation results of multiscale superpixel segmentation versus the number of selected bands with (a) OA and (b) Kappa, and versus the number of training samples with (c) OA and (d) Kappa on the SV dataset using RF classifier.

and then performed intragroup clustering by superpixel-level $\ell_{2,1}$ distance. Evidently, $MSGCF_{GE}$ was used to reveal the validity of the superpixel-level distance for band grouping.

Tables VII–X, respectively, show the classification results of the RF classifier using different numbers of selected bands on the IP, KSC, HU, and SV datasets. It can be observed from the four tables that our proposed $MSGCF$ maintains the best performance on four different datasets. Especially, the advantages of $MSGCF$ are more obvious when the number of chosen bands is small. Furthermore, the classification accuracy of $MSGCF_{NG}$ is decreasing as the number of selected bands increases, which may be due to the large amount of redundant information contained in the additional bands. This phenomenon can be mitigated by the $MSGCF$ on KSC and

TABLE VII
ABLATION RESULTS OF SUPERPIXEL-LEVEL BAND GROUPING
AND CLUSTERING ON THE IP DATASET WITH DIFFERENT
NUMBER OF SELECTED BANDS

No. of selected bands (Measure)	MSGCF _{NG}	MSGCF _{CE}	MSGCF _{GE}	MSGCF
10 (OA)	50.27±0.02	50.81±0.01	51.57±0.02	53.61±1.37
10 (Kappa)	0.44±0.02	0.45±0.01	0.45±0.02	0.48±0.01
20 (OA)	49.30±0.02	50.96±0.01	51.18±0.02	52.71±0.95
20 (Kappa)	0.43±0.02	0.45±0.01	0.45±0.02	0.47±0.01
50 (OA)	49.14±0.03	51.32±0.01	51.65±0.02	52.44±1.46
50 (Kappa)	0.43±0.03	0.45±0.01	0.45±0.02	0.47±0.01

TABLE VIII
ABLATION RESULTS OF SUPERPIXEL-LEVEL BAND GROUPING
AND CLUSTERING ON THE KSC DATASET WITH
DIFFERENT NUMBER OF SELECTED BANDS

No. of selected bands (Measure)	MSGCF _{NG}	MSGCF _{CE}	MSGCF _{GE}	MSGCF
10 (OA)	72.81±0.01	73.90±0.01	74.59±0.01	76.84±0.91
10 (Kappa)	0.71±0.01	0.72±0.01	0.72±0.01	0.74±0.01
20 (OA)	72.78±0.01	74.26±0.01	75.28±0.01	77.54±0.47
20 (Kappa)	0.70±0.01	0.72±0.01	0.72±0.01	0.75±0.01
50 (OA)	71.82±0.01	76.09±0.01	76.40±0.01	77.68±0.49
50 (Kappa)	0.70±0.01	0.73±0.01	0.73±0.01	0.75±0.01

TABLE IX
ABLATION RESULTS OF SUPERPIXEL-LEVEL BAND GROUPING
AND CLUSTERING ON THE HU DATASET WITH
DIFFERENT NUMBER OF SELECTED BANDS

No. of selected bands (Measure)	MSGCF _{NG}	MSGCF _{CE}	MSGCF _{GE}	MSGCF
10 (OA)	66.63±0.01	67.11±0.01	69.16±0.01	70.05±1.18
10 (Kappa)	0.63±0.01	0.64±0.01	0.66±0.01	0.68±0.01
20 (OA)	66.91±0.01	67.61±0.02	69.76±0.01	71.53±1.76
20 (Kappa)	0.64±0.01	0.65±0.02	0.67±0.02	0.69±0.02
50 (OA)	67.26±0.01	68.02±0.01	70.02±0.02	71.05±1.57
50 (Kappa)	0.65±0.01	0.65±0.01	0.67±0.02	0.69±0.02

HU datasets. And the classification accuracy of MSGCF is much higher than that of MSGCF_{NG}, indicating that band grouping plays a vital role in the proposed framework. Besides, the classification results of MSGCF_{CE} are all lower than those of MSGCF_{GE}, which shows that superpixel-level $\ell_{2,1}$ distance is more significant in clustering than superpixel-level KL- ℓ_1 distance in band grouping. Furthermore, we can also see that MSGCF_{CE} is higher than MSGCF_{NG}, validating the role of superpixel-level KL- ℓ_1 distance in band grouping. In summary, both the superpixel-level distance measures, including KL- ℓ_1 and $\ell_{2,1}$, provide a positive effect on the hyperspectral BS.

G. Theoretical Analysis of Time and Space Complexity

Finally, the time complexity of our proposed MSGCF method is theoretically analyzed, which can be roughly divided into three parts, namely preprocessing part (including superpixel segmentation and superpixel-level distance computation), group clustering part and voting part. Specifically, the time complexity of the three superpixel segmentation methods (ERS, SLIC, SNIC) is $O(XY \log(XY))$, $O(XY)$ and $O(XY)$,

TABLE X
ABLATION RESULTS OF SUPERPIXEL-LEVEL BAND GROUPING
AND CLUSTERING ON THE SV DATASET WITH DIFFERENT
NUMBER OF SELECTED BANDS

No. of selected bands (Measure)	MSGCF _{NG}	MSGCF _{CE}	MSGCF _{GE}	MSGCF
10 (OA)	76.16±0.02	77.45±0.02	77.81±0.02	79.34±2.02
10 (Kappa)	0.75±0.02	0.75±0.02	0.75±0.02	0.77±0.02
20 (OA)	76.94±0.02	77.46±0.02	78.27±0.02	80.19±2.50
20 (Kappa)	0.75±0.02	0.76±0.02	0.76±0.03	0.78±0.03
50 (OA)	76.91±0.02	77.23±0.02	78.34±0.02	80.20±2.52
50 (Kappa)	0.75±0.02	0.75±0.03	0.76±0.03	0.78±0.03

TABLE XI
ADVANTAGES AND DISADVANTAGES OF MSGCF
FOR HSI BS

Advantages	Disadvantages
<ul style="list-style-type: none"> • Superpixel-level distances jointly consider the spectral context and spatial structure information, and can better evaluate the intra-class similarity and inter-class variability. • Multiple and multiscale superpixel segmentations are used to cope with different types of land covers and different types of datasets. • The group-clustering framework effectively utilizes the numerous contextual information of the bands and makes the final selected bands more dispersed and less correlated. • The grouping procedure is carried out in an adaptive way. 	<ul style="list-style-type: none"> • The time complexity is higher than the comparative methods.

respectively. Because multiscale segmentation is performed, the time complexity of superpixel segmentation process is $O(kXY \log(XY))$, $O(kXY)$ and $O(kXY)$, respectively. The complexity of computing the superpixel-level distance is $O(N_k B^2)$. Besides, the time complexity of group clustering is $O(kB^2)$, while that of voting is $O(1)$. It is worth noting that although the time complexity of our method is higher, the preprocessing step is totally unsupervised, which can be carried out in advance. Moreover, all branches can be carried out in parallel to further compress the time cost. Similarly, the space complexity of MSGCF is correspondingly divided into three parts, where that of three superpixel segmentation methods (ERS, SLIC, SNIC) is $O(XY)$, and that of superpixel-level distance computation is $O(N_k B^2)$. The space complexity of group clustering is $O(kB^2)$ and the space complexity of voting is $O(kB)$. At last, the advantages and disadvantages of our approach are summarized in Table XI.

V. CONCLUSION

In this article, a MSGCF is proposed for hyperspectral BS, abbreviated as MSGCF. Motivated by the complementary properties of three different superpixel segmentation methods (ERS, SLIC and SNIC), a multiscale superpixel segmentation fusion scheme is performed to accommodate characteristics of various land cover types, and a series of multiscale superpixel maps is thus generated, which provides rich spatial structural information for measuring the correlation of bands.

Subsequently, the grouping and clustering paradigm is introduced to conduct the contextual information among bands. The maximum points of superpixel-level $KL-\ell_1$ distance of adjacent bands are adopted as partition points to separate bands into groups, and the grouping procedure is carried out in an adaptive way rather than presetting the number of groups by hand, hence the generalization of the proposed method can be ensured. We also proposed a SuFDPC to cluster the bands within the group, which is applicable to the limited band clustering situation. Finally, BS results are achieved with a ranking-based voting strategy by concerning information entropy and frequency of occurrence in a unified scheme. With these steps, MSGCF is able to select the most informative bands with strong discrimination.

A number of ablation analysis has been carried out on four real-world HSI datasets to validate the necessity of multiscale superpixel segmentation fusion and the effectiveness of the superpixel-level band grouping and clustering modules. Meanwhile, the classification performance and band similarity (which is one of the important metrics for measuring the representativeness of the selected bands) of MSGCF is compared with several state-of-the-art methods, i.e., AP, ID, K-centers, MVPCA, FDPC, EFDPC, FNGBS, OCF, UBS, ASPS, ACNN, BSNet, and DARecNet. The experimental results always exhibit the superiority of the proposed MSGCF approach.

REFERENCES

- [1] F. D. van der Meer *et al.*, "Multi- and hyperspectral geologic remote sensing: A review," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 14, no. 1, pp. 112–128, Feb. 2012.
- [2] J. Pontius, M. Martin, L. Plourde, and R. Hallett, "Ash decline assessment in emerald ash borer-infested regions: A test of tree-level, hyperspectral technologies," *Remote Sens. Environ.*, vol. 112, no. 5, pp. 2665–2676, May 2008.
- [3] Q. Tong, Y. Xue, and L. Zhang, "Progress in hyperspectral remote sensing science and technology in China over the past three decades," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 70–91, Jan. 2014.
- [4] J. Qiao, M. O. Ngadi, N. Wang, C. Gariépy, and S. O. Prasher, "Pork quality and marbling level assessment using a hyperspectral imaging system," *J. Food Eng.*, vol. 83, no. 1, pp. 10–16, 2007.
- [5] H. Yao, Z. Hruska, R. Kincaid, R. L. Brown, D. Bhatnagar, and T. E. Cleveland, "Detecting maize inoculated with toxigenic and atoxigenic fungal strains with fluorescence hyperspectral imagery," *Biosyst. Eng.*, vol. 115, no. 2, pp. 125–135, 2013.
- [6] J. M. Bioucas-Dias, A. Plaza, G. Camps-Valls, P. Scheunders, N. M. Nasrabadi, and J. Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 2, pp. 6–36, Jun. 2013.
- [7] S. Jia *et al.*, "Multiple feature-based superpixel-level decision fusion for hyperspectral and LiDAR data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 2, pp. 1437–1452, Feb. 2021.
- [8] E. K. Hege, D. O'Connell, W. Johnson, S. Bastly, and E. L. Dereniak, "Hyperspectral imaging for astronomy and space surveillance," in *Imaging Spectrometry*, vol. 5159. Bellingham, WA, USA: SPIE, 2004, pp. 380–391.
- [9] F. Müller-Karger *et al.*, "Satellite remote sensing in support of an integrated ocean observing system," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 4, pp. 8–18, Dec. 2013.
- [10] H. Pu, Z. Chen, B. Wang, and G. Jiang, "A novel spatial-spectral similarity measure for dimensionality reduction and classification of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 7008–7022, Nov. 2014.
- [11] S. S. Sawant and M. Prabukumar, "A review on graph-based semi-supervised learning methods for hyperspectral image classification," *Egyptian J. Remote Sens. Space Sci.*, vol. 23, no. 2, pp. 243–248, Aug. 2020.
- [12] S. Jia, S. Jiang, Z. Lin, N. Li, M. Xu, and S. Yu, "A survey: Deep learning for hyperspectral image classification with few labeled samples," *Neurocomputing*, vol. 448, pp. 179–204, Aug. 2021.
- [13] J. Khodr and R. Younes, "Dimensionality reduction on hyperspectral images: A comparative review based on artificial datas," in *Proc. 4th Int. Congr. Image Signal Process.*, vol. 4, Oct. 2011, pp. 1875–1883.
- [14] C. Zhao, X. Gao, Y. Wang, and J. Li, "Efficient multiple-feature learning-based hyperspectral image classification with limited training samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4052–4062, Jul. 2016.
- [15] 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.
- [16] S. Jia *et al.*, "3-D Gaussian-Gabor feature extraction and selection for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 11, pp. 8813–8826, Nov. 2019.
- [17] D. Lunga, S. Prasad, M. M. Crawford, and O. Ersoy, "Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 55–66, Jan. 2014.
- [18] W. Zhao 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, Aug. 2016.
- [19] S. Jia, J. Hu, J. Zhu, X. Jia, and Q. Li, "Three-dimensional local binary patterns for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2399–2413, Apr. 2017.
- [20] B. Rasti *et al.*, "Feature extraction for hyperspectral imagery: The evolution from shallow to deep: Overview and toolbox," *IEEE Geosci. Remote Sens. Mag.*, vol. 8, no. 4, pp. 60–88, Dec. 2020.
- [21] S. Jia, Z. Zhan, and M. Xu, "Shearlet-based structure-aware filtering for hyperspectral and LiDAR data classification," *J. Remote Sens.*, vol. 2021, pp. 1–25, May 2021.
- [22] T. Liu, Y. Gu, J. Chanussot, and M. D. Mura, "Multimorphological superpixel model for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 6950–6963, Dec. 2017.
- [23] S. Jia *et al.*, "Gradient feature-oriented 3-D domain adaptation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–17, 2022.
- [24] X. Jia and J. A. Richards, "Segmented principal components transformation for efficient hyperspectral remote-sensing image display and classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 1, pp. 538–542, Jan. 1999.
- [25] F. Tsai, E.-K. Lin, and K. Yoshino, "Spectrally segmented principal component analysis of hyperspectral imagery for mapping invasive plant species," *Int. J. Remote Sens.*, vol. 28, no. 5, pp. 1023–1039, Mar. 2007.
- [26] J. Zabalza *et al.*, "Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing," *ISPRS J. Photogramm. Remote Sens.*, vol. 93, pp. 112–122, Jul. 2014.
- [27] M. P. Uddin, M. A. Mamun, and M. A. Hossain, "Effective feature extraction through segmentation-based folded-PCA for hyperspectral image classification," *Int. J. Remote Sens.*, vol. 40, no. 18, pp. 7190–7220, Sep. 2019.
- [28] M. P. Uddin, M. A. Mamun, and M. A. Hossain, "PCA-based feature reduction for hyperspectral remote sensing image classification," *IETE Tech. Rev.*, vol. 38, no. 4, pp. 377–396, Jul. 2021.
- [29] M. P. Uddin, M. A. Mamun, M. I. Afjal, and M. A. Hossain, "Information-theoretic feature selection with segmentation-based folded principal component analysis (PCA) for hyperspectral image classification," *Int. J. Remote Sens.*, vol. 42, no. 1, pp. 286–321, Jan. 2021.
- [30] A. Agarwal, T. El-Ghazawi, H. El-Askary, and J. Le-Moigne, "Efficient hierarchical-PCA dimension reduction for hyperspectral imagery," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2007, pp. 353–356.
- [31] J. Wang and C.-I. Chang, "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1586–1600, Jun. 2006.
- [32] A. Datta, S. Ghosh, and A. Ghosh, "PCA, kernel PCA and dimensionality reduction in hyperspectral images," in *Advances in Principal Component Analysis*. Singapore: Springer, 2018, pp. 19–46.
- [33] B.-C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 5, pp. 1096–1105, May 2004.

- [34] M. Dutta, "Assessment of feature extraction techniques for hyperspectral image classification," in *Proc. Int. Conf. Adv. Comput. Eng. Appl. (ICACEA)*, Mar. 2015, pp. 499–502.
- [35] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2013.
- [36] S. Jia, Z. Lin, B. Deng, J. Zhu, and Q. Li, "Cascade superpixel regularized Gabor feature fusion for hyperspectral image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 5, pp. 1638–1652, May 2019.
- [37] L. He, J. Li, A. Plaza, and Y. Li, "Discriminative low-rank Gabor filtering for spectral-spatial hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1381–1395, Mar. 2017.
- [38] S. Jia, K. Wu, J. Zhu, and X. Jia, "Spectral-spatial Gabor surface feature fusion approach for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1142–1154, Feb. 2019.
- [39] S. Jia *et al.*, "3-D Gabor convolutional neural network for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2022.
- [40] W. Sun and Q. Du, "Hyperspectral band selection: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 7, no. 2, pp. 118–139, Jun. 2019.
- [41] H. Yang, Q. Du, and G. Chen, "Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 544–554, Apr. 2012.
- [42] Q. Dai, J.-H. Cheng, D.-W. Sun, and X.-A. Zeng, "Advances in feature selection methods for hyperspectral image processing in food industry applications: A review," *Crit. Rev. Food Sci. Nutrition*, vol. 55, no. 10, pp. 1368–1382, Aug. 2015.
- [43] B.-H. Chen *et al.*, "Fusarium wilt inspection for *Phalaenopsis* using uniform interval hyperspectral band selection techniques," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Sep. 2020, pp. 2831–2834.
- [44] S. S. Sawant, M. Prabhukumar, and S. Samiappan, "A band selection method for hyperspectral image classification based on cuckoo search algorithm with correlation based initialization," in *Proc. 10th Workshop Hyperspectral Imag. Signal Process., Evol. Remote Sens. (WHISPERS)*, Sep. 2019, pp. 1–4.
- [45] H. Yang, Q. Du, H. Su, and Y. Sheng, "An efficient method for supervised hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 1, pp. 138–142, Jan. 2011.
- [46] B. Guo, R. I. Damper, S. R. Gunn, and J. D. B. Nelson, "Improving hyperspectral band selection by constructing an estimated reference map," *J. Appl. Remote Sens.*, vol. 8, no. 1, Jan. 2014, Art. no. 083692.
- [47] Y. Qian, F. Yao, and S. Jia, "Band selection for hyperspectral imagery using affinity propagation," *IET Comput. Vis.*, vol. 3, no. 4, pp. 213–222, Dec. 2009.
- [48] Z. Guo, H. Yang, X. Bai, Z. Zhang, and J. Zhou, "Semi-supervised hyperspectral band selection via sparse linear regression and hypergraph models," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2013, pp. 1474–1477.
- [49] X. Cao, C. Wei, J. Han, and L. Jiao, "Hyperspectral band selection using improved classification map," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2147–2151, Nov. 2017.
- [50] X. Cao, C. Wei, Y. Ge, J. Feng, J. Zhao, and L. Jiao, "Semi-supervised hyperspectral band selection based on dynamic classifier selection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 4, pp. 1289–1298, Apr. 2019.
- [51] C.-I. Chang and S. Wang, "Constrained band selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 6, pp. 1575–1585, Jun. 2006.
- [52] 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.
- [53] 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.
- [54] Y. Cai, X. Liu, and Z. Cai, "BS-Nets: An end-to-end framework for band selection of hyperspectral image," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 3, pp. 1969–1984, Mar. 2020.
- [55] 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.*, vol. 18, no. 12, pp. 2152–2156, Dec. 2021.
- [56] A. Martínez-Usó, F. Pla, J. M. Sotoca, and P. García-Sevilla, "Clustering-based hyperspectral band selection using information measures," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4158–4171, Dec. 2007.
- [57] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010.
- [58] C. Cariou, K. Chehdi, and S. L. Moan, "BandClust: An unsupervised band reduction method for hyperspectral remote sensing," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 565–569, May 2011.
- [59] B. J. Frey and D. Dueck, "Clustering by passing messages between data points," *Science*, vol. 315, no. 5814, pp. 972–976, Feb. 2007.
- [60] S. Jia, Z. Ji, Y. Qian, and L. Shen, "Unsupervised band selection for hyperspectral imagery classification without manual band removal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 531–543, Apr. 2012.
- [61] X. Cao, B. Wu, D. Tao, and L. Jiao, "Automatic band selection using spatial-structure information and classifier-based clustering," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4352–4360, Sep. 2016.
- [62] M. Zhang, J. Ma, and M. Gong, "Unsupervised hyperspectral band selection by fuzzy clustering with particle swarm optimization," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 773–777, May 2017.
- [63] 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.
- [64] 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.
- [65] X. Luo, Z. Shen, R. Xue, and H. Wan, "Unsupervised band selection method based on importance-assisted column subset selection," *IEEE Access*, vol. 7, pp. 517–527, 2019.
- [66] 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.
- [67] Z. Li, L. Huang, J. He, C. Liu, X. Shi, and D. Zhang, "A local potential-based clustering algorithm for unsupervised hyperspectral band selection," *IEEE Access*, vol. 7, pp. 69027–69041, 2019.
- [68] Q. Wang, Q. Li, and X. Li, "Hyperspectral band selection via adaptive subspace partition strategy," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 12, pp. 4940–4950, Dec. 2019.
- [69] Q. Wang, Q. Li, and X. Li, "A fast neighborhood grouping method for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5028–5039, Jun. 2021.
- [70] S. Jia *et al.*, "Flexible Gabor-based superpixel-level unsupervised LDA for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10394–10409, Dec. 2021.
- [71] F. Yang, Q. Sun, H. Jin, and Z. Zhou, "Superpixel segmentation with fully convolutional networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 13964–13973.
- [72] A. E. Ilesanmi, O. P. Idowu, and S. S. Makhnov, "Multiscale superpixel method for segmentation of breast ultrasound," *Comput. Biol. Med.*, vol. 125, Oct. 2020, Art. no. 103879.
- [73] S. Jia, X. Deng, J. Zhu, M. Xu, J. Zhou, and X. Jia, "Collaborative representation-based multiscale superpixel fusion for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7770–7784, Oct. 2019.
- [74] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, vol. 2. Washington, DC, USA: IEEE Computer Society, 2003, p. 10.
- [75] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. 2, pp. 167–181, Sep. 2004.
- [76] A. P. Moore, S. J. D. Prince, J. Warrell, U. Mohammed, and G. Jones, "Superpixel lattices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2008, pp. 1–8.
- [77] O. Veksler, Y. Boykov, and P. Mheirani, "Superpixels and supervoxels in an energy optimization framework," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Berlin, Germany: Springer, 2010, pp. 211–224.
- [78] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. CVPR*, Jun. 2011, pp. 2097–2104.
- [79] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.
- [80] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.
- [81] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Berlin, Germany: Springer, 2008, pp. 705–718.

- [82] A. Levinshstein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "TurboPixels: Fast superpixels using geometric flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2290–2297, Dec. 2009.
- [83] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels," *École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, Tech. Rep.* 149300, 2010.
- [84] R. Achanta and S. Susstrunk, "Superpixels and polygons using simple non-iterative clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4651–4660.
- [85] S. Jia, B. Deng, J. Zhu, X. Jia, and Q. Li, "Local binary pattern-based hyperspectral image classification with superpixel guidance," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 749–759, Feb. 2018.
- [86] S. Jia, X. Deng, M. Xu, J. Zhou, and X. Jia, "Superpixel-level weighted label propagation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 7, pp. 5077–5091, Jul. 2020.
- [87] Y. Zhang, K. Liu, Y. Dong, K. Wu, and X. Hu, "Semisupervised classification based on SLIC segmentation for hyperspectral image," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 8, pp. 1440–1444, Aug. 2020.
- [88] Y. Ding, Y. Guo, Y. Chong, S. Pan, and J. Feng, "Global consistent graph convolutional network for hyperspectral image classification," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–16, 2021.
- [89] X. Xu, J. Li, C. Wu, and A. Plaza, "Regional clustering-based spatial preprocessing for hyperspectral unmixing," *Remote Sens. Environ.*, vol. 204, pp. 333–346, Jan. 2018.
- [90] M. Wang, B. Zhang, X. Pan, and S. Yang, "Group low-rank nonnegative matrix factorization with semantic regularizer for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1022–1029, Apr. 2018.
- [91] R. Achanta *et al.*, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, May 2012.
- [92] A. Rodriguez and A. Laio, "Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.
- [93] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, vol. 344. Hoboken, NJ, USA: Wiley, 2009.
- [94] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, Apr. 2011.
- [95] T. O. Kvalseth, "A coefficient of agreement for nominal scales: An asymmetric version of Kappa," *Educ. Psychol. Meas.*, vol. 51, no. 1, pp. 95–101, Mar. 1991.
- [96] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.
- [97] C. Debes *et al.*, "Hyperspectral and LiDAR data fusion: Outcome of the 2013 GRSS data fusion contest," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2405–2418, Jun. 2014.



Sen Jia (Senior Member, IEEE) received the B.E. and Ph.D. degrees from the College of Computer Science, Zhejiang University, Hangzhou, China, in 2002 and 2007, respectively.

Since 2008, he has been with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China, where he is a Full Professor. His research interests include hyperspectral image processing, signal and image processing, and machine learning.



Yue Yuan received the B.E. degree in computer science and technology from Henan University, Kaifeng, China, in 2019. She is pursuing the M.E. degree with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China.

Her research interests include hyperspectral image processing and machine learning.



Nanying Li received the M.S. degree in information and communication engineering from the Hunan Institute of Science and Technology, Yueyang, China, in 2021. She is pursuing the Ph.D. degree in computer science and technology from Shenzhen University, Shenzhen, China.

Her research interests include hyperspectral image classification and anomaly detection.



Jianhui Liao received the B.S. degree from the College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen, China, in 2019, where he is pursuing the M.E. degree with the College of Computer Science and Software Engineering.

His research interests include hyperspectral image processing, machine learning, and pattern recognition.



Qiang Huang received the B.S. degree in electrical engineering from Shenzhen University, Shenzhen, China, in 1999, and the Ph.D. degree in electrical engineering and electronics from the University of Liverpool, Liverpool, U.K., in 2004.

He worked as a Research Associate with the University of Leicester, Leicester, U.K., from 2003 to 2004. He was teaching as Lecturer, an Associate Professor and a Professor with Shenzhen University, Shenzhen, China. His research interests include distributed embedded real-time systems, digital image processing, super computer systems, and cloud computing. He has published more than 36 research articles.



Xiuping Jia (Fellow, IEEE) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982, and the Ph.D. degree in electrical engineering from the University of New South Wales, Canberra, ACT, Australia, in 1996.

Since 1988, she has been with the School of Information Technology and Electrical Engineering, University of New South Wales, where she is a Senior Lecturer. She is also a Guest Professor with Harbin Engineering University, Harbin, China, and an Adjunct Researcher with the China National Engineering Research Center for Information Technology in Agriculture. She has coauthored remote sensing textbook titled *Remote Sensing Digital Image Analysis* [Springer-Verlag, third (1999) and fourth edition (2006)]. Her research interests include remote sensing, image processing, and spatial data analysis.

Dr. Jia is a Subject Editor for the *Journal of Soils and Sediments* and an Associate Editor of the *IEEE Transactions on Geoscience and Remote Sensing*.



Meng Xu (Member, IEEE) received the B.S. and M.E. degrees in electrical engineering from the Ocean University of China, Qingdao, China, in 2011 and 2013, respectively, and the Ph.D. degree from the University of New South Wales, Canberra, ACT, Australia, in 2017.

She is an Associate Research Fellow with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. Her research interests include cloud removal and remote sensing image processing.