

Curvature Filters-Based Multiscale Feature Extraction for Hyperspectral Image Classification

Qiaobo Hao^{ID}, Student Member, IEEE, Bin Sun^{ID}, Member, IEEE, Shutao Li^{ID}, Fellow, IEEE,
Melba M. Crawford^{ID}, Life Fellow, IEEE, and Xudong Kang^{ID}, Senior Member, IEEE

Abstract—Exploring fast and effective spectral-spatial feature extraction algorithms for hyperspectral image (HSI) classification is one of the most focus problems in current hyperspectral remote-sensing research. Generally, the size of homogeneous regions in HSIs is not consistent in real scenario and real scenario usually consist of ground objects of different scales. Multiscale strategy starts to be used to construct discriminative features at different scales for HSI classification in recent years. To efficiently characterize the multiscale spectral-spatial features of HSIs, a curvature filters-based multiscale feature extraction method with multiscale superpixel segmentation constraint is proposed. The proposed algorithm is composed of the following major stages. First, global multiscale spectral-spatial features are efficiently extracted via progressively curvature filtering and downsampling operations, which can be regarded as an image pyramid decomposition method. Next, a multiscale superpixel segmentation strategy is applied on the first layer of the image pyramid, and a weighted mean operation is applied within and among superpixels to extract the local multiscale spatial features (LMSFs). Finally, the global multiscale curvature features (GMCFs) and the superpixel segmentation-based LMSFs are fused to form the final multiscale spectral-spatial features for classification purposes. To verify the capabilities of the proposed method, comprehensive experiments are performed on five real hyperspectral datasets. Experimental results demonstrate that the proposed method can significantly improve the classification accuracies compared to several standard HSI feature extraction and classification methods, especially when the number of samples for training is limited.

Index Terms—Curvature filter (CF), dimensionality reduction, hyperspectral image (HSI) classification, image segmentation, multiscale feature extraction.

I. INTRODUCTION

HYPERSPECTRAL images (HSIs), which have more than hundreds of spectral bands for each pixel, can provide abundant and helpful spectral information regarding the

Manuscript received February 25, 2021; revised May 28, 2021; accepted June 20, 2021. Date of publication July 5, 2021; date of current version January 14, 2022. This work was supported in part by the National Natural Science Fund of China under Grant 61890962, Grant 61801178, and Grant 61871179; in part by the Science and Technology Talents Program of Hunan Association for Science and Technology under Grant 2017TJ-Q09; and in part by the Scientific Research Project of Hunan Education Department under Grant 19B105. (*Corresponding author: Xudong Kang*)

Qiaobo Hao, Bin Sun, Shutao Li, and Xudong Kang are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China, and also with the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province, Changsha 410082, China (e-mail: haoqiaobo@hnu.edu.cn; sunbinxs@126.com; shutao_li@hnu.edu.cn; xudong_kang@163.com).

Melba M. Crawford is with the School of Civil Engineering and Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA (e-mail: mcrawford@purdue.edu).

Digital Object Identifier 10.1109/TGRS.2021.3091860

physical characteristic of different materials of interest. This characteristic offers a unique ability to distinguish different landscapes. Owing to this advantage, hyperspectral remote sensing has been extensively applied in many applications, such as anomaly detection [1], target detection [2], and land covers classification [3]. Especially, HSI classification has attracted increasing attention because of its importance in environment surveillance, urban investigation, and precise agriculture. The specific task of HSI classification is to classify each object represented by each pixel in the image.

The high-dimensional characteristic of HSIs increases the difficulty of classification because of the Hughes phenomenon, and also increases the computational burden of the classifiers. Therefore, exploring fast and effective HSI dimensionality reduction and feature extraction algorithms is still the focus of current hyperspectral remote-sensing research. Its main purpose is to reduce the dimensionality of data while the separable information in the data can be retained. For example, principal component analysis (PCA) [4], linear discriminant analysis (LDA) [5], and independent component analysis (ICA) [6] methods have been developed. For these methods, only the spectral information that is easily affected by noise and other factors is considered, which may lead to unsatisfactory classification performance.

To overcome this issue, many advanced spectral-spatial HSI feature extraction methods have been developed, which can combine the spatial information with the spectral features of surface materials. For example, in [7], an adaptive multiview (MV)-based active learning (AL) method is introduced for HSI classification, which incorporating the spectral information and spatial information derived from the segmentation maps into each view. In [8], an effective hierarchical deep spatial feature-based method is proposed by exploring the power of off-the-shelf convolutional neural network models, which has better representation capability. A unified metric learning-based framework is developed to alternately learn discriminative spectral and spatial features. Inspired by the success of deep learning technologies, in [9], an effective supervised version of graph convolutional networks (GCNs), i.e., miniGCNs, is introduced for HSI classification, which allows to train large-scale GCNs in a mini-batch fashion. To effectively exploit the labeled and unlabeled samples in the feature extraction process, in [10], a novel regression-based semi-supervised feature extraction method, i.e., iterative multi-task regression (IMR), is proposed to learn a low-dimensional subspace by simultaneously considering the labeled and unlabeled samples. In [11], a joint and progressive subspace

analysis (JPSA) method is proposed for semi-supervised hyperspectral feature extraction, which is a novel linearized subspace analysis technique with spatial-spectral manifold alignment. Furthermore, in [12], the authors provide a technical overview of the state-of-the-art feature extraction techniques and introduce their applications on HSI classification. Whether for classification methods based on spectral information or spatial-spectral information, feature extraction is essential. Classification accuracies of the classifiers largely depend on the extracted features.

The size of homogeneous regions is not consistent in HSIs in real scenario. To construct discriminative spatial features at different scales, multiscale strategy starts to be used in some spectral-spatial HSI classification methods. For example, in [13], a Gaussian pyramid decomposition-based multiscale spectral-spatial feature extraction method (SPCA-GPs) is introduced for the classification purposes. In [14], a novel multiscale spatial information fusion (MSIF) method is proposed for multiscale feature extraction and fusion, which can effectively extract the intrinsic spatial information contained in different sizes of homogeneous regions. In [15], a collaborative representation-based multiscale superpixel fusion (CRMSF) method is developed, which can extract superpixel segmentation maps at multiple scales from the extended morphological attribute profiles to regularize the collaborative representation-based classification results. Many studies have demonstrated that the multiscale features have a large positive impact on classification performance.

In this article, to efficiently integrate the spectral and spatial features at different scales, a curvature filters (CFs)-based multiscale feature extraction method with superpixel constraint is proposed for spectral-spatial HSI classification. The proposed method is composed of the following major components. First, the CFs-based image pyramid decomposition method is employed to extract the multiscale spectral features and the global multiscale spatial features via a series of curvature filtering and downsampling operations (see Fig. 3). Then, the multiscale superpixel segmentation method is applied on the first layer of the pyramid and the weighted mean operations are developed to extract the local multiscale spatial information within and among each superpixel (see Fig. 4). Finally, the multiscale spectral features, global multiscale spatial features, and local multiscale spatial features (LMSFs) are combined as the final features for classification. Experimental results on five real HSI datasets show that the proposed method has competitive performance compared with several classical and outstanding HSI feature extraction and classification methods, especially when there are only a few samples for training.

To sum up, the main contributions of this article are threefold.

- 1) This work proposes a CFs-based multiscale feature extraction method for HSI classification. Two different CFs, i.e., Gaussian curvature filter (GCF) and mean curvature filter (MCF), have been employed. The performance of different CFs in feature extraction is analyzed for the first time.

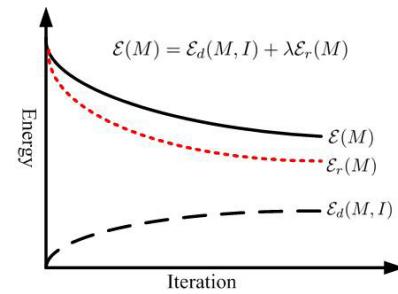


Fig. 1. Evolution process of the total energy, the data-fitting energy, and the regularization energy in the energy functional.

- 2) This work can efficiently exploit the multiscale spectral features, global multiscale spatial features via the CFs-based image pyramid method. The LMSFs can be extracted via the multiscale superpixel segmentation and the weighted mean operations. These multiscale spectral-spatial features can be fully used by combining them.
- 3) The proposed MMCF and MGCF methods can greatly improve the classification accuracy of classic classifiers. To validate the generalization of the proposed methods, experiments are conducted on five real hyperspectral datasets. Experimental results show that the proposed methods outperform various state-of-the-art HSI feature extraction and classification approaches on HSIs with different spatial resolutions. The CFs-based image pyramid decomposition technique and the multiscale superpixel segmentation-based local curvature feature extraction technique both contribute significantly to the proposed methods. Furthermore, the proposed methods will not yield over-smoothed classification maps when the number of training samples is limited.

The rest of the work is organized as follows. Section II reviews the CFs and the superpixel segmentation algorithms. In Section III, the details of the proposed multiscale feature extraction method are introduced. In Section IV, the experimental results are presented and analyzed. Finally, Section V draws the conclusions.

II. RELATED WORKS

A. Curvature Filters (CFs)

Recently, CF has been developed to implicitly minimize curvature regularization for image-processing problems such as smoothing and denoising. In every variational formulation, a minimizing function to an energy functional need to be found

$$\mathcal{E}(M) = \mathcal{E}_d(M, I) + \lambda \mathcal{E}_r(M) \quad (1)$$

where $\mathcal{E}(M)$ refers to the total energy, which is composed of two parts: $\mathcal{E}_d(M, I)$ and $\mathcal{E}_r(M)$. $\mathcal{E}_d(M, I)$ denotes the data-fitting energy, which measures the degree of fit of M to the image data I . λ is a scalar regularization coefficient, which can weight the two contributions. $\lambda \mathcal{E}_r(M)$ denotes the regularization energy, which formalizes prior knowledge about M . As Fig. 1 illustrates, the data-fitting term $\mathcal{E}_d(M, I)$

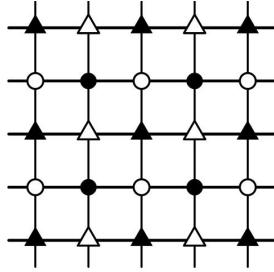


Fig. 2. Illustration of the domain decomposition. The adjacent pixels in a four-connected neighborhood are in different subsets: black circles B_C , black triangles B_T , white circles W_C , and white triangles W_T .

initially increases or remains constant when reducing the regularizer $\mathcal{E}_r(M)$. The reduction in the total energy $\mathcal{E}(M)$ initially comes from the regularizer, which indicates that the optimization of $\mathcal{E}_r(M)$ is the main part. In [16], CF is designed to minimize the regularizer. The filter algorithm is designed as described below.

1) *Theoretical Foundation:* Here, we introduce a theorem in differential geometry, which can guarantee that the proposed filter reduces total absolute Gaussian curvature. For any developable surface \mathcal{S} , We can use its tangent plane $T\mathcal{S}$ to obtain a local approximation. The following theorem can hold:

$$\begin{aligned} \forall \vec{x} \in \mathcal{S}, \quad \forall \xi > 0, \quad \exists \vec{x}_0 \in \mathcal{S} \\ \text{s.t. } 0 < |\vec{x} - \vec{x}_0| < \xi \text{ and } \vec{x} \in T\mathcal{S}(\vec{x}_0). \end{aligned} \quad (2)$$

Gaussian curvature of a 2-D surface is defined as $k_1 k_2$ in differential geometry, where k_1 and k_2 are the principal curvatures. For any developable surface, (2) requires that one of its principal curvatures must be zero. Therefore, the principal curvature with a smaller absolute value can be minimized as follows:

$$k_1 k_2 = 0 \Leftrightarrow \min\{|k_1|, |k_2|\} = 0. \quad (3)$$

This operation is equivalent to minimizing Gaussian curvature.

2) *Domain Decomposition:* The dependencies between neighboring pixels will hamper locally minimizing the smaller absolute principal curvature [17], and thus, a domain decomposition algorithm is proposed to circumvent these dependencies. As Fig. 2 illustrates, the adjacent pixels in a four-connected neighborhood are in different subsets. Therefore, dependencies between neighboring pixels can be removed. Because of the independence, convergence can be guaranteed (proof can be found in [16]). Furthermore, within a 3×3 window, we can enumerate all tangent planes.

3) *Projection to Tangent Plane:* Tangent planes can be represented by the adjacent pixels in a variety of ways, such as triangles or rectangles. Here, triangles are used because of their easy projection. All possible triangles in the 3×3 pixel neighborhood around pixel x that do not contain x as a vertex are enumerated to find the tangent plane with the minimum distance d_i .

4) *Minimal Projection Operator:* For a pixel x located at (i, j) , projection operator \mathcal{P} , i.e., projection operator \mathcal{P}_g or \mathcal{P}_m , is iterated over all pixels in each of black circles B_C ,

Algorithm 1 Curvature Filter

Input: $M(i, j)$
Output: $\hat{M}(i, j)$

- 1: $\forall x \in B_C, \mathcal{P}M(i, j)$
- 2: $\forall x \in B_T, \mathcal{P}M(i, j)$
- 3: $\forall x \in W_C, \mathcal{P}M(i, j)$
- 4: $\forall x \in W_T, \mathcal{P}M(i, j)$

Algorithm 2 Projection Operator \mathcal{P}_g of Gaussian Curvature

Input: $M(i, j)$
Output: $\hat{M}(i, j) = M(i, j) + d_m$

- 1: $d_1 = (M(i-1, j) + M(i+1, j))/2 - M(i, j)$
- 2: $d_2 = (M(i, j-1) + M(i, j+1))/2 - M(i, j)$
- 3: $d_3 = (M(i-1, j-1) + M(i+1, j+1))/2 - M(i, j)$
- 4: $d_4 = (M(i-1, j+1) + M(i+1, j-1))/2 - M(i, j)$
- 5: $d_5 = M(i-1, j) + M(i, j-1) - M(i-1, j-1) + M(i, j)$
- 6: $d_6 = M(i-1, j) + M(i, j+1) - M(i-1, j+1) + M(i, j)$
- 7: $d_7 = M(i, j-1) + M(i+1, j) - M(i+1, j-1) + M(i, j)$
- 8: $d_8 = M(i, j+1) + M(i+1, j) - M(i+1, j+1) + M(i, j)$
- 9: $|d_m| = \min\{|d_i|, i = 1, 2, \dots, 8\}$

Algorithm 3 Projection Operator \mathcal{P}_m of Mean Curvature

Input: $M(i, j)$
Output: $\hat{M}(i, j) = M(i, j) + d_m$

- 1: $\hat{d}_1 = (\frac{5}{16}(M(i-1, j) + M(i+1, j)) + \frac{5}{8}M(i, j+1) - \frac{1}{8}(M(i-1, j+1) + M(i+1, j+1)) - M(i, j)$
- 2: $\hat{d}_2 = (\frac{5}{16}(M(i-1, j) + M(i+1, j)) + \frac{5}{8}M(i, j-1) - \frac{1}{8}(M(i-1, j-1) + M(i+1, j-1)) - M(i, j)$
- 3: $\hat{d}_3 = (\frac{5}{16}(M(i, j-1) + M(i, j+1)) + \frac{5}{8}M(i-1, j) - \frac{1}{8}(M(i-1, j-1) + M(i-1, j+1)) - M(i, j)$
- 4: $\hat{d}_4 = (\frac{5}{16}(M(i, j-1) + M(i, j+1)) + \frac{5}{8}M(i+1, j) - \frac{1}{8}(M(i+1, j-1) + M(i+1, j+1)) - M(i, j)$
- 5: $|d_m| = \min\{|\hat{d}_i|, i = 1, 2, \dots, 4\}$

black triangles B_T , white circles W_C , and white triangles W_T (see Fig. 2). The CF is generated as summarized in Algorithm 1

$$\text{CF}(x) = \mathcal{P}(M(x)), \quad x \in \{B_T, B_C, W_T, W_C\}. \quad (4)$$

For GCF, the minimal projection operator \mathcal{P}_g is summarized in Algorithm 2. The distance d_i from the pixel x to the eight smallest triangular tangent planes can be found and the resulting pixel-local is updated by \mathcal{P}_g . The smallest absolute distance $|d_m|$ is used as the minimum projection of the current intensity. GCF can exactly preserve developable surfaces. Compared with other edge-preserving filters, the GCF is parameter-free.

For MCF, half-window regression is applied to preserve edges. The projection distances \hat{d}_i of MCF, which are different from the directional curvatures d_i , are given in Algorithm 3.

In this work, we refer the GCF with $\text{GCF}(\mathbf{I})$, and refer the MCF with $\text{MCF}(\mathbf{I})$, in which \mathbf{I} represents the input image.

The CFs optimize certain regular term, and the known surface of differential geometry is implicitly used in the

filtering process. In the original work [16] and [18], we can find a complete description of the CFs. Here, four major advantages of the CFs are presented as follows.

- 1) The CFs do not need to calculate the gradient of the energy, as long as they can be evaluated point-wise, the generic data-fitting terms can be handled. Therefore, they can deal with arbitrary imaging models.
- 2) The known surfaces can be used to approximate the image because the surfaces corresponding to the regularization are known. Taking the GCF as an example, its regularization term is Gaussian curvature, which corresponds to a developable surface. There are only three types of developable surfaces: cylinders, cones, and tangent developable. We can choose the most suitable one of these three surfaces to approximate the image. The filters do not assume differentiability of the signal, and thus, edges can be preserved.
- 3) The CFs do not require the calculation of curvature and have linear algorithmic complexity in the number of image pixels, therefore the computing performance is excellent. The CFs are easy to parallelize and do not require large memory to be performed, which is very meaningful when the image is very large. It can be implemented in 40 lines of MATLAB code or less than 100 lines of C++ code.
- 4) The CFs are parameter-free, and do not need guidance images.

B. Superpixel Segmentation

Superpixel segmentation was first introduced in [19], in which an image is over-segmented into superpixels in a preprocessing stage. Superpixel segmentation groups atomic regions of pixels that are perceptually similar to provide visually meaningful entities, and the pixels within each superpixel are expected to belong to the same class. The spatial structures determine the shape and size of superpixel adaptively. Compared to rigid pixel representation, using superpixels can greatly reduce the number of primitives for subsequent processing steps, and thus can reduce the computation cost. Superpixel segmentation has become an important module for many applications in the field of computer vision, including object tracking [20], image segmentation [21], depth recovery [22], and so on. Two categories, i.e., the graph-based methods and the gradient-based methods, are the most common superpixel segmentation algorithms.

1) Graph-Based Methods: For these kinds of methods, given a base image, the researches construct an undirected graph $G = (V, E)$ on the base image. V represents the set of vertex corresponding to pixels of the base image. E is the set of edge, which represents the pairwise similarities between adjacent pixels. After that, by selecting a subset of edges $A \subseteq E$, the researches partition the graph into some disjoint subgraphs. Finally, by minimizing the cost function, the superpixel segmentation results can be obtained. The main graph-based methods are described in the following categories.

a) Graph-cut-based: Algorithms in this category achieve superpixel segmentation via minimizing the cost of cuts in the

graph. For example, normalized cuts (Ncuts) algorithm [23], in which a novel global criterion, the normalized cut, is proposed for segmenting the graph. Ncuts algorithm can yield superpixels with regular shapes and similar sizes, but has extremely high computational complexity. Entropy rate superpixel (ERS) algorithm [24] presents a novel graph construction for images, in which develops a clustering objective function composting of two components: the entropy rate and a balancing term.

b) Boundary evolution: Superpixels can be produced through seeking the optimal boundaries of superpixels. For example, superpixel lattices (Lattices) [25] is a greedy superpixel algorithm that preserves the topology of regular lattice. Although Lattices adds topological information constraints, it maintains good performance in speed and segmentation accuracy.

c) Energy optimization: These superpixel segmentation algorithms obtain their optimum segmentation results by iteratively optimizing formulated energies. Initial superpixel segmentation can be achieved by partitioning image into a regular grid, and pixels are exchanged between adjacent superpixels according to the energy. For example, superpixels extracted via energy-derived sampling (SEEDS) algorithm [26] can continuously refine the superpixels by modifying the boundaries through a hill-climbing optimization.

2) Gradient-Based Methods: For the Gradient-based methods, the researches obtain an initial coarse clustering result and the clusters are continuously refined until some convergence criterions are met. The main gradient-based methods are introduced in the following categories.

a) Density-based: Quick shift (QS) [27] and mean shift [28] are two popular density-based algorithms. The basic idea of these methods is performing mode-seeking in a computed density image. The drawback is that they cannot control the number of superpixels and their compactness.

b) Contour evolution: Superpixels are represented as evolving contours starting from initial seed pixels. TurboPixel (TP) [29] method produces superpixels with regular shapes by describing a geometric-flow-based method. For this method, a data-driven curve evolution process is combined with a set of skeletal-based external constraints to segment the image into grid-like superpixels.

c) Clustering-based: These methods are inspired by clustering algorithms and can produce superpixels with regular shapes and similar sizes. Simple linear iterative clustering (SLIC) [30], linear spectral clustering (LSC) [31], and content-adaptive superpixel segmentation (CAS) [32] are clustering-based superpixel segmentation methods. They use color, spatial, and additional information.

III. PROPOSED METHOD

First, as Fig. 3 shows, feature images of different resolutions are obtained by using the CFs-based image pyramid decomposition method. The feature images include corresponding scale information. Then, as Fig. 4 shows, the LMSFs of the first layer in the pyramid are extracted via multiscale superpixel segmentation and the weighted mean operations. Finally, the multiscale spectral features, the global multiscale

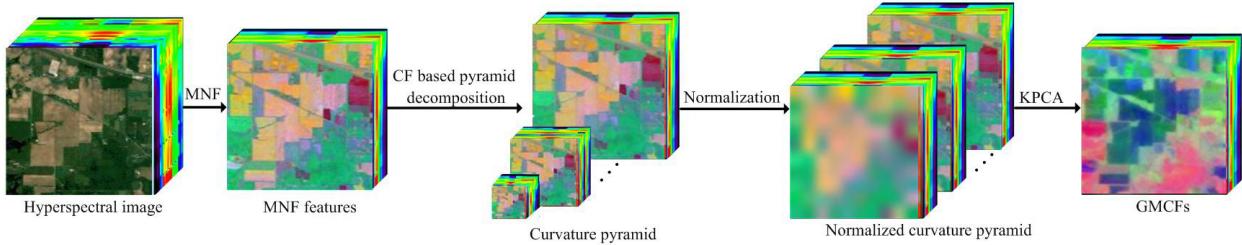


Fig. 3. Schematic of curvature filters-based image pyramid decomposition.

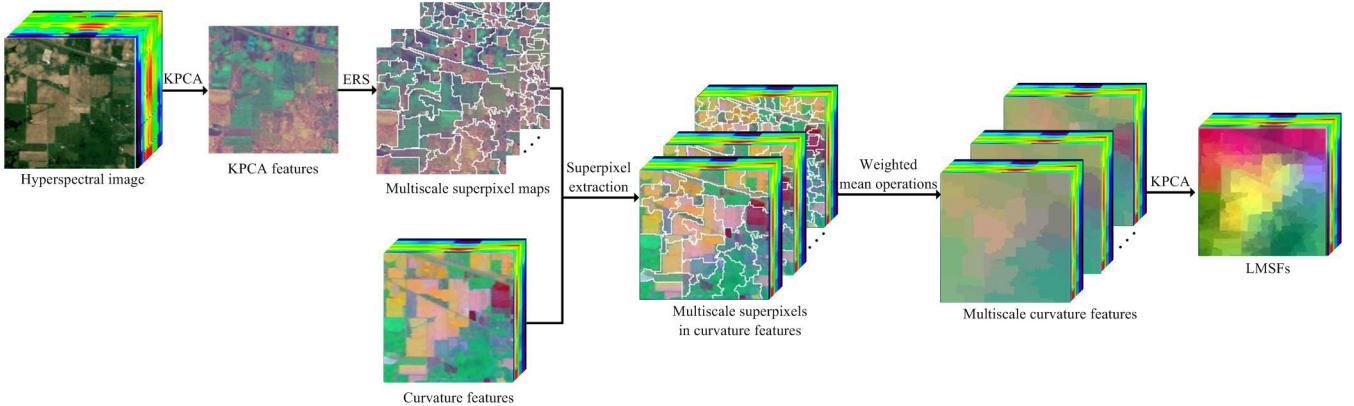


Fig. 4. Schematic of superpixel segmentation-based local multiscale spatial feature extraction.

spatial features, and the LMSFs are combined as the input of the support vector machine (SVM) classifier. The details are described as below.

A. CFs-Based Pyramid Decomposition

HSIs are high-dimensional data, and thus, the maximum noise fraction (MNF) method [33] is applied first on the raw HSIs to reduce the noise and improve the computational efficiency. Here, we fix the number of bands of HSIs after dimensionality reduction to 20, since it has been empirically found in [33] that the first 20 principal components can comprise the most useful information of the considered HSI datasets. Then, two CFs, i.e., GCF and MCF, are used to extract the features of the HSIs after dimensionality reduction. These two CFs do not assume differentiability of the signal and they smooth the images with edge protection.

In [34], the GCF has been employed to extract the spatial texture features from the first principal component of HSIs for classification. However, this GCF-based single-scale feature extraction method does not consider the multiscale information in HSIs, and thus is not effective for describing the structures' multiscale variability. Pyramid representation is a type of classical and efficient multiscale signal representation in computer vision. A signal or an image is subject to progressively filtering and subsampling operations. The repeated smoothing operations can extract multiscale spectral features and the subsampling operations can extract global multiscale spatial features. Therefore, we propose a CFs-based pyramid decomposition method to extract features at different scales.

Here, we take GCF as an example, and the process is the same when using MCF. The used GCF can exactly preserve the developable surfaces, and this property makes it expected to be an effective feature extraction tool. Compared with other edge preserving filters, the present GCF is parameter-free, and thus, is easy to implement and parallelize. As Fig. 3 shows, first, we perform the GCF on the HSIs after dimensionality reduction, and let the obtained feature map be \mathbf{H}_1^c , which is regarded as the first layer of the image pyramid. Then, the second layer \mathbf{H}_2^c is obtained by filtering \mathbf{H}_1^c with a Gaussian curvature kernel. Next, the even rows and columns in \mathbf{H}_2^c are removed. The GCF adjusts the spectral values and the subsampling operation adjusts the spatial resolution of the HSIs. If we remove the pixels in even rows and columns of the HSIs, the GCF-based pyramid can be generated via progressively Gaussian curvature filtering and subsampling operations as follows:

$$\mathbf{H}_l^c(i, j) = \text{GCF}(\mathbf{H}_{l-1}^c)(2i - 1, 2j - 1) \quad (5)$$

where $\mathbf{H}_l^c(i, j)$ is the spectral value of the pixel at spatial location (i, j) in the l th layer feature map, $l \in \{2, 3, \dots, L\}$. The value of L depends on the spatial size, which is the number of layers in the pyramid. GCF refers to the GCF operation. A series of images $\{\mathbf{H}_1^c, \mathbf{H}_2^c, \dots, \mathbf{H}_L^c\}$ produced from the above filtering and downsampling operations constitute the Gaussian curvature pyramid.

Since the layers of the pyramid have different sizes, we enlarge $\mathbf{H}_2^c, \mathbf{H}_3^c, \dots, \mathbf{H}_L^c$ to be as large as \mathbf{H}_1^c via a nearest-neighbor interpolation method. The first layer and enlarged images of other layers are fused together to take full

advantage of the complementary information in each layer. The stacked features $\{\mathbf{H}_1^c, \mathbf{H}_2^c, \dots, \mathbf{H}_L^c\}$ have high spectral dimension and contain a lot of redundant information, which easily lead to Hughes phenomenon and increase computational burden. To tackle these problems, a kernel principle component analysis (KPCA) [35] method is used to generate the global multiscale curvature features (GMCFs). Furthermore, in [36], the KPCA was found that it has the ability to improve the spectral separability of pixels. Therefore, it is suitable for fusing the filtered feature images.

B. Superpixel-Based Multiscale Segmentation

To obtain the base images for superpixel segmentation, the KPCA is used to reduce dimension of the HSIs to 3. We use the first three principle components \mathbf{H}^{kPCA} as the base images. In [24], ERS segmentation is introduced as an undirected graph-based superpixel segmentation method. By optimizing the objective function in regard to A , the clustering can be obtained as follows:

$$\max_A \{H(A) + \lambda B(A)\} \quad \text{s.t. } A \subseteq E \quad (6)$$

where $H(\cdot)$ refers to the entropy rate of the random walk on the graph. $B(\cdot)$ refers to the balance item on the clustering distribution. $H(\cdot)$ and $B(\cdot)$ constitute the clustering objective function to form the homogeneous and compact clusters. The weight of balancing term $\lambda \geq 0$ controls the contribution of $H(\cdot)$ and $B(\cdot)$. A greedy optimization scheme [37] is used to efficiently optimize the objective function in (6). As Fig. 4 shows, the multiscale superpixel maps can be generated by performing ERS segmentation algorithm [24] on the base images as follows:

$$\mathbf{H}^{kPCA} = \bigcup_{k=0}^{S_n} \chi_k^n, \quad \text{s.t. } \chi_k^n \cap \chi_g^n = \emptyset \quad (7)$$

where \mathbf{H}^{kPCA} refers to the base images, and χ_k^n is the k th superpixel in the n th segmentation result. S_n represents how many superpixels are included in the n th segmentation result. For data containing many homogeneous regions, S_n can be calculated as follows:

$$S_n = S - (n - 1) \left\lfloor \frac{S}{N} \right\rfloor, \quad n = 1, 2, \dots, N \quad (8)$$

where N refers to the total number of multiscale segmentation results, $\lfloor (S/N) \rfloor$ represents the maximum integer value that is not greater than (S/N) , which is a floor operation. n represents the serial number of the multiscale segmentation results. The value of N is empirically selected in this work. Designing a homogeneity index that can measure the degree of image homogeneity to automatically choose the value of N will be our future work. S indicates how many superpixels are included in a fine-scale superpixel map, which is selected based on the complexity of the base images. Here, we use a common edge detector, the Canny operator, to measure the complexity. The edge ratio is obtained by calculating the proportion of nonzero pixels in the detected results accounting

for the total number of elements in the base images, which represents the texture complexity of the base images. According to the obtained the edge ratio, S is selected as follows:

$$S = S_{base} \frac{E_F}{E_I} \quad (9)$$

where E_F denotes the number of nonzero pixels in the detected results, E_I refers to the number of elements in the base images, S_{base} is a predefined number.

Based on the multiscale segmentation results, we can obtain the position indexes of pixels in each superpixel. We apply the position indexes on the first layer of pyramid, and the 3-D superpixels can be obtained. Then, we apply a mean operation on the pixels within each 3-D superpixel to exploit the spatial features. We assign all elements in each superpixel to the mean value Y_i , which is the same as mean filtering. Using this operation, the interferences in each superpixel can be reduced. A mean feature image \mathbf{I}_n^{Mean} is composed of all filtered superpixels. Next, we conduct the mean operations on the neighboring superpixels $Y_{i,j}$, $j = \{1, \dots, J\}$, to extract the inter object level features among the current superpixel and the neighboring superpixels. Compared with the inter-object-level features, Y_i is more important and representative feature, and thus, we use a weighted mean operation as follows:

$$Y_i^w = a * Y_i + (1 - a) * \sum_{j=1}^J \frac{Y_{i,j}}{z} \quad (10)$$

where a is a weight scale parameter, which can control the weight of the mean pixel and the inter-object-level features, z is the number of neighboring superpixels. We assign Y_i^w to all pixels in each superpixel to compose \mathbf{I}_n^W , which is a weighted average feature image. After performing the weighted mean operations on each generated 3-D multiscale segmentation feature map, we can obtain the LMSFs $\{\mathbf{I}_1^W, \mathbf{I}_2^W, \dots, \mathbf{I}_N^W\}$. We stack the obtained feature maps of different scales to take full advantage of the complementary information. However, the stacked feature maps have high dimensionality and increase the computational cost. The KPCA is applied on the stacked LMSFs to reduce the redundant information and avoid the Hughes phenomenon.

C. Multiscale Features Fusion for Classification

To efficiently use the complementary information in both the global multiscale spatial features and the LMSFs, these two multiscale features are stacked together and used as the input of the classifier. We utilize the SVM to evaluate the proposed feature extraction method. The SVM classifier is one of the most widely used pixel-wise classifiers [38], [39]. We choose the SVM classifier since it performs quick classification for HSIs. Furthermore, if more effective classifiers were combined with the proposed method, the classification accuracies could be certainly improved. However, it would be difficult to prove whether the improvement of the final results should be attributed to the proposed method or the employed classifiers.

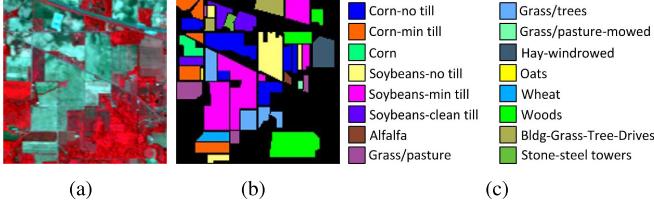


Fig. 5. Indian Pines data. (a) False color composition. (b) Reference data. (c) Class names.

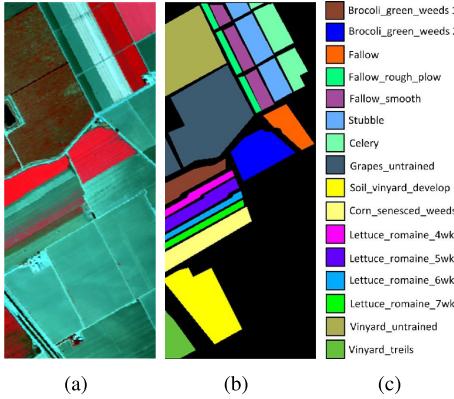


Fig. 6. Salinas data. (a) False color composition. (b) Reference data. (c) Class names.

IV. EXPERIMENTS

A. Datasets

Five real hyperspectral datasets are used to validate the proposed feature extraction method. Brief description of them is provided below.

Indian Pines Dataset: These data were captured by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) instrument at around the Indian Pines test site in North-western Indiana. These data consist of 224 spectral channels ranging from 0.4 to 2.5 μm and the spatial size is 145 \times 145 pixels. The data cover a mixed vegetation site and contains different forest, crops, and other natural perennial vegetation. For these data, water absorption bands (nos. 104–108, 150–163, and 220) are discarded and 200 spectral bands are used in the following experiments. Fig. 5(a) and (b) presents the RGB composite and ground truth.

Salinas Dataset: These data were acquired by the AVIRIS instrument over an agricultural area of Salinas Valley, CA, USA. These data consist of 224 spectral channels and the spatial size is 512 \times 217 pixels. The data contain 16 different classes which are bare soils, vegetables, and vineyard fields. For these data, water absorption bands (nos. 108–112, 154–167, and 224) are removed and 204 spectral bands are remained in the experiments. Fig. 6(a) and (b) presents the RGB composite and ground truth.

KSC Dataset: These data were acquired by the AVIRIS sensor over the Kennedy Space Center (KSC), FL, USA. These data consist of 224 bands with center wavelengths from 0.4 to 2.5 μm and has a spatial resolution of 18 m. For these data, water absorption and low SNR bands are removed and 176 bands are remained for the analysis. For certain vegetation

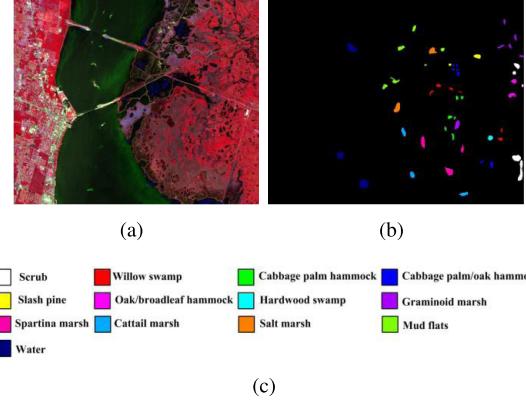


Fig. 7. KSC data. (a) False color composition. (b) Reference data. (c) Class names.

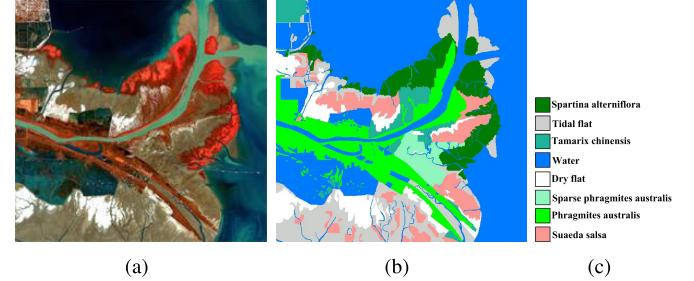


Fig. 8. Yellow River Delta data. (a) False color composition. (b) Reference data. (c) Class names.

types, the spectral signatures are very similar, which increases the difficulty of classification. Fig. 7(a) and (b) presents the RGB composite and ground truth.

Yellow River Delta Dataset: These data were acquired by the advanced hyperspectral imager (AHSI) loaded on GaoFen-5 satellite over modern Yellow River Delta (118°20'E–119°20'E, 37°16'N–38°16'N) on November 1, 2018. The data consist of 330 bands with center wavelengths from 0.39 to 2.51 μm and has a spatial resolution of 30 m. The processing steps include atmospheric correction, geometric correction, invalid bands removal, and image cropping. For these data, water absorption and low SNR bands are removed and 296 spectral bands are remained for the analysis in the experiments. Fig. 8(a) and (b) presents the RGB composite and ground truth. For the experimental area, eight classes of typical objects are annotated based on visual interpretation and field investigation.

Tiaoma Town Dataset: This HSI was captured by a Headwall Nano-Hyperspectral push-broom scanner on-board a UAV platform at around Tiaoma, Yuhua, Changsha, Hunan, China, on 2nd of August 2020. The flight height of the UAV is 200 m and the flight speed is 8 m/s. The image has 270 spectral bands. Its spectral resolution is 2.2 nm and spatial resolution is 8.7 cm. The spatial size of the image in pixels is 1424 \times 640. This scene is composed of 12 different objects, in which most classes are forest trees. Fig. 9 shows the RGB composite of hyperspectral data and the ground truth.

B. Parameter and Component Analysis

First, the first three hyperspectral data are used to analyze the influence of the number of layers l on the classification

TABLE I

INFLUENCE OF EACH COMPONENT IN THE MULTISCALE FEATURES EXTRACTED BY THE WEIGHTED MEAN OPERATIONS ON THE PERFORMANCE OF THE PROPOSED METHOD. \mathbf{I}_n^W REFERS TO THE n TH WEIGHTED AVERAGE FEATURE IMAGE.
THE OA, AA, AND KAPPA COEFFICIENT ARE RECORDED IN %

Metrics	\mathbf{I}_6^W	\mathbf{I}_5^W	\mathbf{I}_4^W	\mathbf{I}_3^W	\mathbf{I}_2^W	\mathbf{I}_1^W
OA	96.05 (0.76)	96.35 (0.58)	95.66 (0.96)	96.56 (0.84)	96.09 (0.67)	96.21 (0.74)
AA	96.35 (1.56)	94.91 (1.44)	94.96 (1.97)	94.96 (1.88)	96.09 (2.06)	95.11 (2.50)
Kappa	95.49 (0.88)	95.84 (0.66)	95.05 (1.09)	96.07 (0.96)	95.53 (0.76)	95.68 (0.84)
Metrics	$\{\mathbf{I}_6^W\}$	$\{\mathbf{I}_6^W, \mathbf{I}_5^W\}$	$\{\mathbf{I}_6^W, \mathbf{I}_5^W, \mathbf{I}_4^W\}$	$\{\mathbf{I}_6^W, \mathbf{I}_5^W, \dots, \mathbf{I}_3^W\}$	$\{\mathbf{I}_6^W, \mathbf{I}_5^W, \dots, \mathbf{I}_2^W\}$	$\{\mathbf{I}_6^W, \mathbf{I}_5^W, \dots, \mathbf{I}_1^W\}$
OA	96.05 (0.76)	96.73 (0.69)	96.47 (0.76)	96.87 (0.68)	97.10 (0.53)	97.15 (0.01)
AA	96.35 (1.56)	96.77 (1.45)	96.80 (1.54)	96.65 (1.78)	97.06 (1.41)	96.77 (0.01)
Kappa	95.49 (0.88)	96.26 (0.79)	95.97 (0.87)	96.43 (0.78)	96.82 (0.60)	96.75 (0.01)

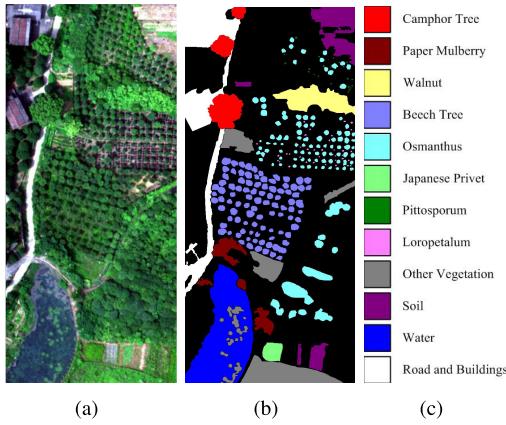


Fig. 9. Tiaoma Town data. (a) False color composition. (b) Reference data. (c) Class names.

performance. A library of SVM (LIBSVM) [40] is implemented to perform the SVM classifier and we use a radial basis function (RBF) kernel in the experiments. The numbers of training samples selected from the ground truth of the three hyperspectral data are 2% of the labeled samples, ten per class, and ten per class, respectively. The Yellow River Delta data and Tiaoma Town data are not used in this experiment because their sizes are very large and that will take a lot of time. As Fig. 10 shows, the accuracies generally increase as l increases, which means that the proposed CF-based multiscale feature extraction method is effective. According to the experimental results, for the Indian Pines and Salinas datasets, l is set to 7 in the following experiments; and for the KSC, Yellow River Delta and Tiaoma Town datasets, we set l to 5.

Then, to illustrate the effectiveness of each component in the multiscale features extracted by the weighted mean operations, we perform an experiment on the Indian Pines image. In Table I, \mathbf{I}_n^W refers to the n th weighted average feature image, which is the local multiscale spatial features introduced in Section III-B. As Table I shows, different weighted average feature images can obtain improvements of different degrees. The highest classification accuracy can be obtained by fusing the multiscale weighted average feature images. The two experiments demonstrate that these two major components

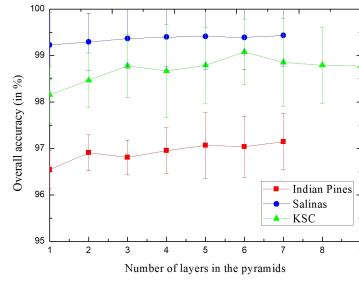


Fig. 10. Analysis of the influence of the number of layers l .

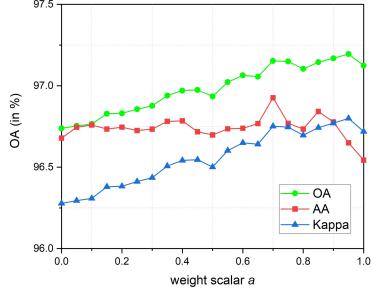


Fig. 11. Analysis of the influence of the parameter a .

mentioned above can effectively improve the classification accuracies. In the following experiments, considering both accuracies and computational complexity, we set n to 6.

Next, we perform a sensitivity analysis of an important parameter, i.e., a , on the Indian Pines dataset. We use 2% of the labeled samples to train the SVM classifier and the rest samples are used for test. a is a predefined weight scalar controlling the weight of the mean pixel and the inter object level features. Compared with the neighboring superpixels, the superpixel containing the sample is more important. When a is smaller, the neighboring superpixels will be given more power, which means the classification accuracy of large homogeneous regions improves and the classification accuracy of heterogeneous regions deceases. As Fig. 11 shows, we set the parameter a to 0.75, which can obtain satisfactory classification performance in the experiments.

Furthermore, the influences of different superpixel segmentation methods on the performance of the proposed method

TABLE II

PERFORMANCE OF THE PROPOSED METHOD WHEN USING DIFFERENT SUPERPIXEL SEGMENTATION APPROACHES. THE OA, AA, AND KAPPA COEFFICIENT ARE RECORDED IN %

Methods	Ncut	TP	CAS	SLIC	LSC	Lattices	ERS
OA	94.84	95.14	95.18	95.59	95.69	95.82	97.15
AA	94.60	95.04	95.65	94.71	95.44	96.76	96.77
Kappa	94.11	94.46	94.51	95.59	95.08	95.22	96.75

TABLE III

CLASSIFICATION ACCURACIES OF THE SVM ON THE RAW DATA, THE GCFs, THE GMCFs, THE LMCFs, AND THE PROPOSED MGCF METHOD

Metrics	raw	GCFs	GMCFs	LMCFs	MGCF
OA	69.08	86.90	92.72	94.35	97.15
AA	70.90	90.70	92.82	86.30	97.15
Kappa	64.46	85.05	91.69	93.55	96.75

are analyzed to verify the generalization ability of the proposed method for different superpixel segmentation methods. Experimental results are shown in Table II. Seven widely used superpixel segmentation methods which have been introduced in Section II-B, i.e., Ncut [23], TP [29], CAS [32], SLIC [30], LSC [31], Lattices [25], and ERS [24] are used for comparison. These methods are used because their codes are available and are easy to implement. Furthermore, the number of superpixels can be controlled for a fair comparison. As Table II shows, most of these superpixel-based multiscale segmentation methods can give satisfactory results. The adopted ERS method is more efficient and can obtain the best classification performance.

Finally, an experiment is performed to analyze the influence of three major features, i.e., Gaussian curvature features (GCFs), GMCFs, and LMCFs, on the classification performance of the proposed MGCF method. The experiment is performed on the Indian Pines dataset, and 2% labeled samples are used for training. GCFs refer to the features obtained by the GCF, which corresponds to the first layer of the Gaussian curvature pyramid. GMCFs refer to the features obtained by the CFs-based image pyramid decomposition technique, which are the global multiscale curvature features. LMCFs refer to the local multiscale curvature features obtained via the multiscale superpixel segmentation technique and the weighted mean operations. As Table III shows, three major features are all able to improve the classification accuracies in varying degrees. The classification accuracies can be further improved by combining these features. The proposed MGCF method shows the best classification performance in terms of OA, AA, and Kappa via combining the GMCFs and LMCFs. The experimental result demonstrates that the CFs-based image pyramid decomposition technique and the multiscale superpixel segmentation with weighted mean operations-based feature extraction technique both contribute significantly to the proposed MGCF method.

C. Comparison of Different Feature Extraction Methods

In the field of HSI classification, researchers have proposed many spectral-spatial feature extraction methods to

TABLE IV

COMPARISON OF THE CLASSIFICATION ACCURACIES WHEN USING DIFFERENT FEATURE EXTRACTION METHODS ON THE INDIAN PINES DATASET

Methods	OA	AA	Kappa
IAPs [41]	83.03(0.01)	86.76(0.02)	80.59(0.01)
EMAP [42]	84.69(0.01)	86.34(0.03)	82.47(0.02)
OTVCA [43]	90.26(0.01)	80.74(0.02)	88.90(0.01)
IFRF [44]	92.12(0.01)	88.60(0.02)	91.02(0.01)
SPCA-GPs [13]	94.73(1.05)	93.99(1.20)	94.56(1.65)
MSTV [36]	95.32(0.01)	95.84(0.01)	94.65(0.01)
MMCF	96.93(0.01)	96.48(0.02)	96.50(0.01)
MGCF	97.15(0.01)	96.77(0.01)	96.75(0.01)

exploit informative features suitable for data representation. To illustrate the effectiveness of the proposed multiscale feature extraction methods, six state-of-the-art feature extraction approaches, including the invariant attribute profiles (IAPs) [41], the extended morphological attribute profiles (EMAP) [42], the orthogonal total variation component analysis (OTVCA) [43], the image fusion and recursive filtering (IFRF) [44], the segmented principal component analysis and Gaussian pyramid decomposition based multiscale feature extraction method (SPCA-GPs) [13], and the multiscale total variation (MSTV) [36] approaches, are adopted for comparison. The EMAP, MSTV, and SPCA-GPs are multiscale feature extraction methods for HSI classification. For the EMAP method, four principal components are used to exploit the features and the parameters given in [42] are used. For other features, we choose the default parameters settings used in the corresponding researches. The SVM classifier is adopted to evaluate the performance of different features for a fair comparison. Table IV presents the classification accuracies obtained by these feature extraction methods. As Table IV shows, the proposed feature extraction methods show competitive performance and can achieve the highest accuracies.

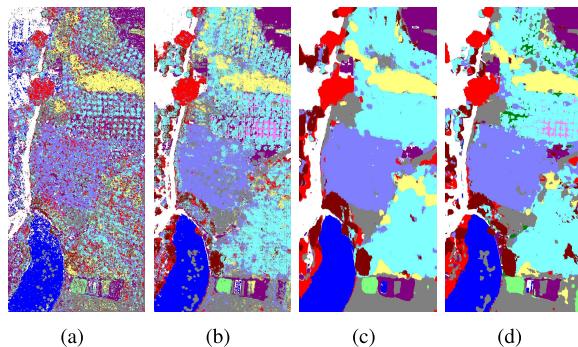
In addition, to demonstrate the effectiveness of the proposed feature extraction methods on HSIs with high spatial resolution, an experiment is conducted on the Tiaoma Town dataset, which has a spatial resolution of 8.7 cm. Six state-of-the-art feature extraction methods used before are still adopted for comparison, and the SVM classifier is used to measure the classification performance. In this experiment, 10% of the labeled samples in Fig. 9(b) are randomly selected for training, and the remaining labeled samples are used for testing. To ensure fairness, the experiment is repeated ten times, and we use the average value as the final result.

Table V shows the classification accuracies obtained by using different feature extraction methods on the Tiaoma Town dataset. It can be observed that the proposed MMCF method can achieve the highest accuracies in terms of the OA, AA, and Kappa, and the proposed MGCF method also shows satisfactory performance, which can prove that the proposed methods are still effective when applied to high-resolution HSIs. Furthermore, For the two classes of Pittosporum and Loropetalum, the performance of some comparative feature extraction methods is not satisfactory, while the performance of the proposed methods is very competitive. Fig. 12 presents

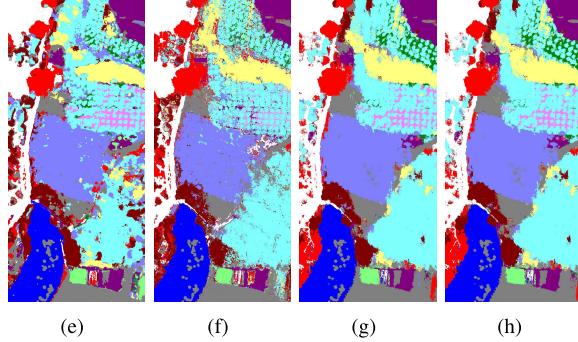
TABLE V

COMPARISON OF THE CLASSIFICATION ACCURACIES WHEN USING DIFFERENT FEATURE EXTRACTION METHODS ON THE TIAOMA TOWN DATASET.
THE STANDARD VARIANCES OF THE RESULTS OF 10 REPEATED EXPERIMENTS ARE PRESENTED IN THE PARENTHESES

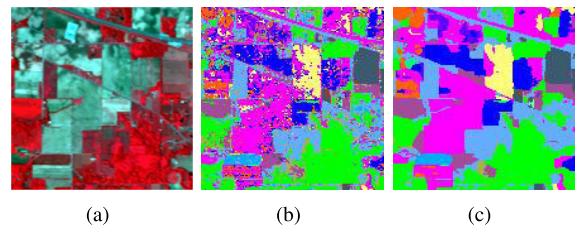
	IAPs	OTVCA	IFRF	MSTV	SPCA-GPs	EMAP	MGCF	MMCF
Camphor Tree	53.47(0.46)	78.98(0.64)	96.90(0.26)	96.37(0.28)	97.07(0.34)	96.05(0.25)	99.04(0.10)	98.98(0.17)
Paper Mulberry	49.46(0.73)	74.73(1.21)	98.45(0.43)	97.94(0.27)	93.95(0.81)	96.85(0.29)	99.71(0.05)	99.71(0.06)
Walnut	80.76(0.35)	89.05(0.41)	100.0(0.00)	99.99(0.01)	99.32(0.23)	99.70(0.09)	99.99(0.01)	100.0(0.01)
Beech Tree	63.69(0.41)	76.84(0.16)	95.60(0.32)	94.52(0.44)	97.71(0.32)	97.77(0.22)	99.89(0.03)	99.91(0.04)
Osmanthus	65.35(0.40)	72.37(0.27)	94.63(0.13)	95.38(0.27)	94.17(0.36)	97.79(0.19)	99.67(0.11)	99.73(0.08)
Japanese Privet	71.97(0.98)	93.40(0.58)	99.72(0.13)	99.76(0.16)	99.83(0.23)	98.77(0.27)	99.62(0.10)	99.82(0.07)
Pittosporum	57.16(2.79)	10.00(2.36)	0.000(0.00)	80.89(3.37)	67.74(6.64)	80.10(0.96)	91.15(1.80)	91.44(2.42)
Loropetalum	65.64(3.09)	95.49(0.74)	0.000(0.00)	96.16(2.72)	97.69(0.99)	85.03(2.17)	97.60(0.43)	98.05(0.46)
Other Vegetation	64.10(0.44)	80.81(0.30)	98.39(0.18)	96.63(0.18)	96.23(0.10)	97.50(0.08)	98.97(0.13)	99.02(0.11)
Soil	73.17(0.41)	97.52(0.10)	99.78(0.05)	99.91(0.03)	99.91(0.05)	99.81(0.04)	99.98(0.01)	99.98(0.02)
Water	88.64(0.40)	98.08(0.14)	91.35(0.01)	97.45(0.11)	99.47(0.07)	99.05(0.08)	99.15(0.05)	99.19(0.04)
Road&Buildings	87.05(0.39)	98.20(0.13)	99.39(0.07)	99.57(0.05)	99.65(0.02)	99.54(0.06)	99.66(0.05)	99.65(0.04)
OA	73.68(0.25)	87.31(0.05)	96.49(0.02)	97.42(0.09)	97.85(0.04)	98.39(0.03)	99.49(0.02)	99.51(0.03)
AA	68.37(0.34)	80.46(1.95)	81.19(0.02)	96.21(0.44)	95.23(0.59)	95.66(0.10)	98.70(0.18)	98.79(0.22)
KAPPA	69.67(0.29)	85.40(0.06)	95.96(0.03)	97.03(0.10)	97.52(0.05)	98.15(0.03)	99.41(0.02)	99.44(0.03)



(a) (b) (c) (d)



(e) (f) (g) (h)



(a) (b) (c)



(d) (e) (f)



(g) (h) (i)

Fig. 12. Classification maps yielded using different feature extraction methods on Tiaoma Town dataset. (a) IAPs, OA = 73.68%. (b) OTVCA, OA = 87.31%. (c) IFRF, 96.49%. (d) MSTV, OA = 97.42%. (e) SPCA-GPs, OA = 97.85%. (f) EMAP, OA = 98.39%. (g) MGCF, OA = 99.49%. (h) MMCF, OA = 99.51%.

the classification maps yielded by using different feature extraction methods. The proposed MMCF and MGCF methods produce better classification maps by integrating both local multiscale spectral-spatial features and global multiscale spectral-spatial features. In the experiments, the parameters of the proposed methods are empirically selected to fixed values and not optimized for each test image. If the optimal parameters for Tiaoma Town dataset can be selected adaptively, we can get better classification maps. We will work on how to adaptively select the optimal parameters for different datasets in the future.

Fig. 13. Three-band representation of the Indian Pines dataset and Classification maps yielded by using different classification methods. (a) Three-band representation of the image. (b) SVM, OA = 69.08%. (c) EPF, OA = 79.68%. (d) WMRF, 87.64%. (e) LBP, OA = 91.34%. (f) SCMK, OA = 91.42%. (g) GTR, OA = 91.90%. (h) MMCF, OA = 96.93%. (i) MGCF, OA = 97.15%.

D. Comparison of Different Classification Methods

To validate the effectiveness of the proposed method, experiments are performed on four real HSIs. Seven classical or out-of-state hyperpectral image classification methods are compared with the proposed method, including the SVM classifier on the raw HSI [38], the edge-preserving filtering-based post-processing method (EPF) [45], the weighted Markov random fields-based method (WMRF) [46], the local binary patterns-based method (LBP) [47], the superpixel-based classification via multiple kernels method [48], and the generalized tensor regression approach (GTR) [49]. The default parameters

TABLE VI

COMPARISON OF THE CLASSIFICATION ACCURACIES OF THE SVM [38], EPF [45], WMRF [46], LBP [47], SCMK [48], GTR [49], AND THE PROPOSED METHODS ON THE INDIAN PINES DATASET. THE STANDARD VARIANCES OF THE RESULTS OF 10 REPEATED EXPERIMENTS ARE PRESENTED IN THE PARENTHESES

	SVM	EPF	WMRF	LBP	SCMK	GTR	MMCF	MGCF
Alfalfa	73.20(0.21)	20.00(0.42)	1.890(0.03)	71.82(0.28)	100.0(0.00)	100.0(0.00)	100.0(0.00)	100.0(0.00)
Corn_N	57.36(0.06)	73.44(0.11)	87.89(0.04)	94.19(0.02)	88.88(0.04)	89.14(0.03)	93.61(0.04)	94.98(0.03)
Corn_M	57.49(0.09)	83.43(0.12)	73.40(0.12)	87.06(0.09)	85.04(0.07)	87.73(0.06)	94.64(0.05)	94.74(0.04)
Corn	47.02(0.12)	82.42(0.18)	56.24(0.35)	89.10(0.03)	86.68(0.15)	95.66(0.04)	97.75(0.05)	97.79(0.05)
Grass_M	86.17(0.09)	94.88(0.08)	88.64(0.06)	92.46(0.03)	85.41(0.04)	97.65(0.02)	98.52(0.02)	98.40(0.02)
Grass_T	75.85(0.02)	71.78(0.04)	97.79(0.01)	88.75(0.04)	99.75(0.01)	95.43(0.02)	100.0(0.00)	100.0(0.00)
Grass_P	80.33(0.30)	60.00(0.52)	8.400(0.08)	60.82(0.14)	91.85(0.22)	94.84(0.07)	94.15(0.18)	97.71(0.07)
Hay_W	87.04(0.02)	87.89(0.02)	100.0(0.00)	98.95(0.00)	99.96(0.00)	93.82(0.02)	100.0(0.00)	100.0(0.00)
Oats	69.11(0.20)	0.000(0.00)	4.740(0.05)	47.38(0.35)	99.47(0.02)	82.50(0.15)	79.06(0.24)	78.64(0.24)
Soybean_N	65.87(0.06)	83.85(0.09)	74.49(0.09)	87.60(0.02)	83.18(0.08)	89.09(0.02)	94.52(0.03)	94.67(0.03)
Soybean_M	68.15(0.02)	76.06(0.05)	93.63(0.04)	93.92(0.01)	94.43(0.04)	90.41(0.02)	97.51(0.02)	97.18(0.02)
Soybean_C	49.22(0.08)	79.10(0.13)	88.87(0.09)	87.71(0.05)	82.89(0.08)	92.69(0.06)	97.15(0.06)	97.79(0.05)
Wheat	86.29(0.04)	99.75(0.01)	99.95(0.00)	91.08(0.02)	93.73(0.13)	95.02(0.03)	100.0(0.00)	100.0(0.00)
Woods	87.96(0.02)	90.18(0.02)	98.67(0.01)	96.18(0.02)	97.10(0.04)	96.78(0.02)	99.83(0.00)	99.77(0.00)
Buildings	51.17(0.12)	80.04(0.24)	76.51(0.10)	95.21(0.04)	90.61(0.09)	94.28(0.02)	98.42(0.02)	98.32(0.01)
Stone	92.16(0.24)	81.30(0.34)	61.29(0.17)	70.61(0.05)	97.03(0.05)	89.55(0.04)	98.50(0.02)	98.31(0.02)
OA	69.08(0.02)	79.68(0.02)	87.64(0.02)	91.34(0.01)	91.42(0.01)	91.90(0.01)	96.93(0.01)	97.15(0.01)
AA	70.90(0.03)	72.76(0.06)	69.53(0.02)	84.55(0.01)	92.25(0.02)	92.79(0.01)	96.48(0.02)	96.77(0.01)
Kappa	64.46(0.02)	64.46(0.02)	85.83(0.02)	90.13(0.01)	90.21(0.02)	90.74(0.01)	96.50(0.01)	96.75(0.01)

settings provided in the corresponding researches are used. Especially, for the LBP method, for the Yellow River Delta dataset, because of the memory limitation, the dimension of the raw data is reduced to three instead of the optimal parameter 30 first using the PCA method [4]. For the proposed method, we set the fundamental number of superpixel S_{base} to 1600, 1600, 3500, and 9000 for the Indian Pine, Salinas, KSC and Yellow River Delta datasets, respectively. In the experiments, 15 principal components obtained by the KPCA method are remained.

We randomly selected samples from the ground truth to constitute the training set, and the rest samples are used for test. All experiments are repeated ten times, and the mean values of ten repeated experimental results are regarded as the final results to make the comparison fair. Four common metrics including the individual class accuracy(CA), overall accuracy (OA), average accuracy (AA), and Kappa coefficient (Kappa) are employed to quantitatively assess the performances of different approaches.

We perform the first experiment on the Indian Pines dataset and 2% samples in the ground truth are randomly selected for training. In Table VI, the classification accuracies of each class, OA, AA, and Kappa coefficient obtained by these approaches are shown. As Table VI shows, the proposed methods, i.e., MMCF and MGCF, are superior than those obtained by others in terms of the OA, AA, and Kappa among all studied approaches. For example, OA increases from 68.50% to 96.93% and 97.15% using the proposed methods compared with the raw data with SVM classifier, which can demonstrate that the proposed feature extraction method is indeed more effective. The MMCF and MGCF methods are more suitable for HSIs with many homogeneous regions. For most land covers, for example, Corn, Grass, and Soybean, the proposed methods can achieve the highest accuracies. The Indian Pines scene is made up of various sizes of unbalanced classes. It can

be observed from Table VI that the classification accuracy of oats is significantly lower than others. The main reason is that there are only 20 labeled samples in the category of oats in groundtruth. In the experiments, we chose 1 labeled sample for training, and the remaining 19 labeled samples for testing. The number of training samples of this class is very limited, and thus, the classification performance of the proposed methods on this class is not as good as that on others. Furthermore, Fig. 13 presents the classification maps generated by all studied approaches. We can visually observe that the proposed methods show satisfactory performance and provide higher OAs. The edges and boundaries of the classification maps obtained by the proposed methods are more consistent with the RGB image and the proposed methods will not produce over-smoothed classification results. The SVM method tends to produce obvious misclassification in the classification results. The edges of the classification maps generated by the WMRF and GTR approaches do not match the three-band representation of image very well. The LBP method may yield over-smoothed classification maps. The reason is that the author uses a slack variable to reduce the influence of spectral changeability and noise. However, when the number of samples for training is insufficient, it may lead to over-smoothed classification map.

We perform the second and third experiments on the Salinas and the KSC datasets, in which ten samples per class are randomly selected to constitute the training sets. As Tables VII and VIII show, it is clearly to see that the proposed method still achieves the highest accuracies in terms of OA, AA, and Kappa coefficient. Compared with the Indian Pines and Salinas images, the spectral signatures of certain vegetation types in the KSC data are more similar, and thus, these data are more difficult to classify. Although some class accuracies are not the highest, for all class the proposed method can achieve classification accuracies of more than 95%.

TABLE VII

COMPARISON OF THE CLASSIFICATION ACCURACIES OF THE SVM [38], EPF [45], WMRF [46], LBP [47], SCMK [48], GTR [49], AND THE PROPOSED METHODS ON THE SALINAS DATASET. THE STANDARD VARIANCES OF THE RESULTS OF TEN REPEATED EXPERIMENTS ARE PRESENTED IN THE PARENTHESES

	SVM	EPF	WMRF	LBP	SCMK	GTR	MMCF	MGCF
Weeds_1	95.30(0.07)	98.16(0.04)	99.85(0.00)	97.26(0.02)	100.0(0.00)	98.73(0.02)	99.98(0.00)	100.0(0.00)
Weeds_2	98.83(0.01)	99.83(0.00)	99.77(0.01)	99.14(0.01)	99.73(0.01)	99.33(0.01)	100.0(0.00)	100.0(0.00)
Fallow	87.37(0.05)	91.99(0.05)	91.38(0.06)	85.08(0.12)	92.98(0.05)	99.45(0.00)	99.46(0.00)	99.45(0.00)
Fallow_P	96.82(0.01)	97.22(0.00)	99.06(0.00)	67.44(0.06)	96.29(0.05)	94.35(0.03)	98.16(0.00)	97.89(0.01)
Fallow_S	97.96(0.01)	99.84(0.00)	97.99(0.01)	93.23(0.06)	96.17(0.03)	95.49(0.03)	99.91(0.00)	99.92(0.00)
Stubble	99.96(0.00)	99.98(0.00)	100.0(0.00)	90.19(0.05)	99.83(0.00)	99.99(0.00)	99.97(0.00)	99.97(0.00)
Celery	96.17(0.03)	97.19(0.03)	99.72(0.01)	94.24(0.05)	99.92(0.00)	99.94(0.00)	99.83(0.00)	99.84(0.00)
Grapes	70.04(0.03)	81.38(0.06)	85.75(0.06)	90.70(0.06)	79.47(0.09)	94.36(0.04)	99.98(0.00)	99.98(0.00)
Soil	98.47(0.01)	99.02(0.01)	99.83(0.00)	99.33(0.01)	99.95(0.00)	99.05(0.00)	99.88(0.00)	99.74(0.00)
Corn	84.54(0.05)	91.33(0.05)	94.15(0.05)	98.67(0.01)	90.58(0.08)	94.03(0.02)	99.45(0.00)	99.49(0.00)
Lettuce_4	83.94(0.06)	91.62(0.07)	99.77(0.00)	81.39(0.10)	97.55(0.02)	94.10(0.03)	98.54(0.04)	99.25(0.02)
Lettuce_5	92.47(0.08)	96.33(0.10)	99.99(0.00)	97.84(0.02)	96.43(0.07)	97.87(0.02)	96.07(0.09)	95.83(0.09)
Lettuce_6	85.69(0.14)	91.92(0.12)	98.87(0.01)	80.02(0.05)	96.77(0.02)	90.47(0.06)	95.79(0.05)	95.71(0.07)
Lettuce_7	83.99(0.13)	93.30(0.13)	99.75(0.00)	75.99(0.07)	92.75(0.02)	96.99(0.03)	94.38(0.04)	93.42(0.05)
Vinyard_U	51.27(0.05)	66.38(0.14)	70.92(0.15)	82.51(0.09)	83.23(0.08)	77.60(0.07)	100.0(0.00)	100.0(0.00)
Vinyard_T	93.48(0.09)	99.35(0.01)	97.94(0.03)	85.25(0.09)	96.29(0.07)	100.0(0.00)	100.0(0.00)	100.0(0.00)
OA	82.67(0.02)	88.32(0.04)	92.17(0.01)	89.43(0.02)	91.81(0.01)	92.17(0.01)	99.47(0.69)	99.43(0.01)
AA	88.52(0.03)	93.43(0.03)	95.92(0.01)	88.64(0.02)	94.87(0.01)	95.14(0.01)	98.84(1.33)	98.78(0.01)
Kappa	80.78(0.03)	87.04(0.04)	91.27(0.02)	88.27(0.02)	90.90(0.01)	91.34(0.01)	99.41(0.76)	99.37(0.01)

TABLE VIII

COMPARISON OF THE CLASSIFICATION ACCURACIES OF THE SVM [38], EPF [45], WMRF [46], LBP [47], SCMK [48], GTR [49], AND THE PROPOSED METHODS ON THE KSC DATASET. THE STANDARD VARIANCES OF THE RESULTS OF TEN REPEATED EXPERIMENTS ARE PRESENTED IN THE PARENTHESES

	SVM	EPF	WMRF	LBP	SCMK	GTR	MMCF	MGCF
Scrub	95.64(0.02)	99.46(0.01)	98.48(0.03)	97.29(0.08)	88.85(0.03)	88.00(0.05)	99.43(0.01)	99.61(0.00)
Willow_S	82.90(0.06)	99.76(0.00)	90.47(0.10)	99.34(0.02)	92.32(0.10)	49.30(0.12)	100.0(0.00)	100.0(0.00)
Cabbage_P	85.76(0.10)	96.42(0.08)	96.87(0.03)	99.88(0.00)	92.20(0.05)	82.56(0.28)	98.60(0.04)	97.56(0.05)
Cabbage_O	54.20(0.09)	86.34(0.09)	82.48(0.14)	98.82(0.03)	80.33(0.08)	61.32(0.26)	95.56(0.06)	95.91(0.05)
Slash_P	50.59(0.13)	78.30(0.22)	95.43(0.06)	93.30(0.06)	86.36(0.04)	32.87(0.28)	96.89(0.02)	97.92(0.01)
Oak_H	48.69(0.16)	79.97(0.17)	97.26(0.04)	84.52(0.16)	92.51(0.06)	45.48(0.15)	94.62(0.06)	95.68(0.08)
Hardwood_S	69.46(0.04)	78.62(0.09)	98.00(0.06)	100.0(0.00)	100.0(0.00)	32.11(0.07)	98.64(0.04)	99.69(0.01)
Graminoid_M	69.25(0.09)	87.59(0.14)	98.67(0.03)	93.58(0.03)	95.82(0.02)	94.70(0.06)	99.43(0.01)	99.43(0.01)
Spartina_M	83.05(0.05)	93.92(0.05)	99.96(0.00)	95.78(0.04)	96.71(0.03)	98.78(0.01)	98.85(0.02)	98.24(0.03)
Cattail_M	88.42(0.07)	97.08(0.05)	100.0(0.00)	87.00(0.08)	94.09(0.05)	79.64(0.08)	100.0(0.00)	100.0(0.00)
Salt_M	90.77(0.07)	96.56(0.04)	98.07(0.03)	100.0(0.00)	97.38(0.01)	48.59(0.05)	99.90(0.00)	100.0(0.00)
Mud_F	91.21(0.06)	97.87(0.03)	92.39(0.05)	99.83(0.00)	92.74(0.06)	65.62(0.14)	100.0(0.00)	99.94(0.00)
Water	99.98(0.00)	100.0(0.00)	100.0(0.00)	98.11(0.03)	100.0(0.00)	71.68(0.03)	100.0(0.00)	100.0(0.00)
OA	82.63(0.02)	93.89(0.03)	97.05(0.01)	95.54(0.02)	93.91(0.01)	66.40(0.04)	99.05(0.01)	98.85(0.01)
AA	77.69(0.03)	91.68(0.03)	96.01(0.01)	95.96(0.01)	93.02(0.01)	65.43(0.04)	98.61(0.01)	98.48(0.01)
Kappa	80.69(0.02)	93.20(0.03)	96.72(0.01)	95.03(0.02)	93.22(0.01)	62.41(0.04)	98.94(0.01)	98.72(0.01)

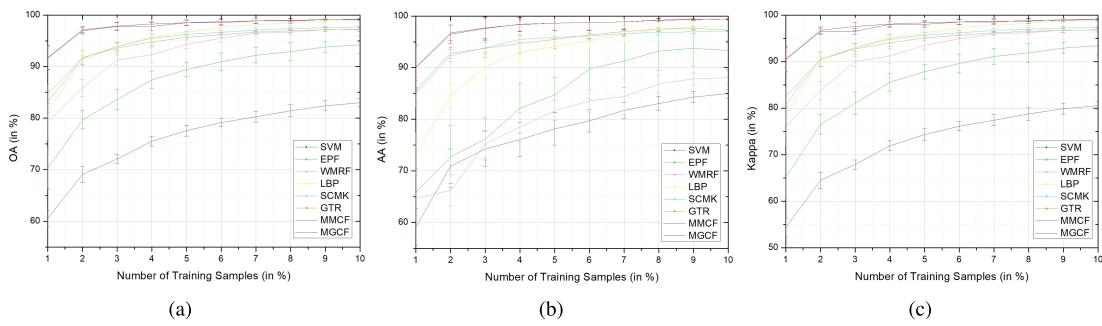


Fig. 14. Analysis of the influence of the number of samples for training on the Indian Pines dataset. (a)–(c) Variation trends of OA, AA, and Kappa.

Furthermore, the influence of the number of samples for training on the classification accuracies is analyzed. Different numbers of samples are randomly selected from the ground truth to constitute the training sets. Figs. 14–16 present the

variation tendencies of the values of the OA, AA, and Kappa coefficient on three HSIs, i.e., Indian Pines, Salinas, and KSC datasets. The size of the Yellow River Delta dataset is too large, and it takes too much time to conduct this

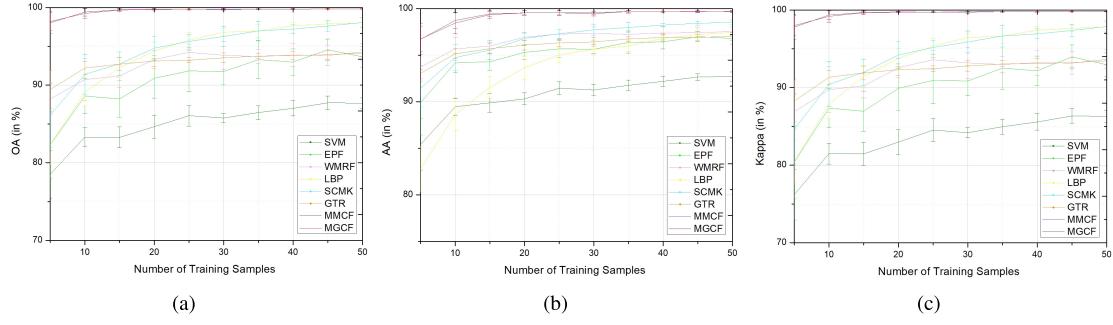


Fig. 15. Analysis of the influence of the number of samples for training on the Salinas dataset. (a)–(c) Variation trends of OA, AA, and Kappa.

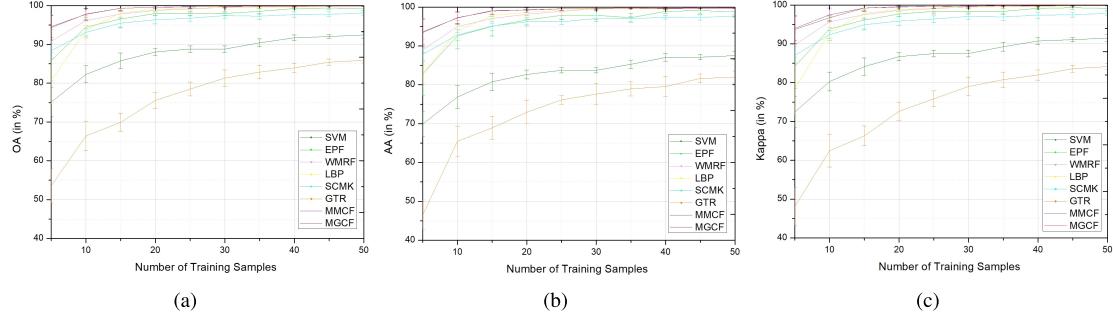


Fig. 16. Analysis of the influence of the number of samples for training on the KSC dataset. (a)–(c) Variation trends of OA, AA, and Kappa.

TABLE IX

COMPARISON OF THE CLASSIFICATION ACCURACIES OF THE SVM [38], EPF [45], WMRF [46], LBP [47], SCMK [48], GTR [49], AND THE PROPOSED METHODS ON THE YELLOW RIVER DELTA DATASET. THE STANDARD VARIANCES OF THE RESULTS OF TEN REPEATED EXPERIMENTS ARE PRESENTED IN THE PARENTHESES

	SVM	EPF	WMRF	LBP	SCMK	GTR	MMCF	MGCF
S. alterniflora	89.48(0.70)	90.84(0.46)	93.64(0.73)	61.13(2.18)	93.88(0.59)	90.43(0.00)	91.09(0.15)	92.73(0.50)
Tidal Flat	66.45(0.74)	69.73(1.06)	86.69(0.76)	50.77(1.80)	87.85(0.95)	79.04(0.01)	86.90(0.66)	86.64(0.60)
Tamarix Chinensis	63.53(1.43)	74.91(1.81)	75.54(1.39)	85.08(2.91)	80.41(1.56)	86.39(0.01)	86.32(1.57)	85.98(1.36)
Water	97.24(0.25)	97.06(0.20)	96.28(0.17)	81.09(2.06)	96.84(0.21)	96.25(0.00)	97.45(0.23)	97.46(0.21)
Dry Flat	80.82(1.06)	81.95(1.17)	87.27(0.94)	63.17(3.74)	87.18(1.23)	89.50(0.01)	92.50(0.45)	92.35(0.46)
S. Phragmites Australis	68.01(2.06)	84.37(2.60)	88.63(2.30)	78.18(2.49)	87.91(2.07)	85.07(0.01)	90.86(1.54)	90.63(1.30)
Phragmites Australis	78.68(0.53)	79.63(0.60)	90.23(0.49)	46.78(2.73)	91.16(0.56)	87.19(0.01)	93.58(0.44)	93.38(0.57)
Suaeda Salsa	60.67(0.96)	68.78(1.49)	74.12(1.61)	56.81(5.29)	78.64(1.04)	77.09(0.01)	86.51(1.01)	86.00(1.06)
OA	84.29(0.10)	86.31(0.12)	90.85(0.15)	67.89(0.34)	91.88(0.13)	89.85(0.00)	93.44(0.11)	93.33(0.11)
AA	75.61(0.31)	80.91(0.39)	86.55(0.40)	65.38(1.01)	87.98(0.41)	86.37(0.00)	90.84(0.24)	90.65(0.19)
Kappa	78.62(0.14)	81.34(0.17)	87.52(0.22)	54.62(0.75)	88.98(0.19)	86.18(0.00)	91.09(0.15)	90.95(0.16)

experiment, so we do not use this data for analysis. The number of training samples for the Indian Pines dataset varies from 1% to 10% of the ground truth. For the Salinas and KSC images, the number of samples for training varies from 5 to 50 per class. One can observe that, the increase of the number of samples for training has a beneficial effect on the classification performance. More importantly, the proposed MMCF and MGCF methods perform outstanding especially when the number of samples for training is quite limited. Taking the Indian Pines image for example, when training with 1% labeled samples, the OA obtained by the SVM classifier is about 60% while the OAs obtained by the MMCF and MGCF methods are near 93%. Analyzing Figs. 15 and 16, similar conclusions can be obtained.

We perform the fourth experiment on the Yellow River Delta dataset, in which 1% samples are randomly selected

for training. Table IX exhibits the CA, OA, AA, and Kappa coefficient obtained by different methods. As Table IX shows, the proposed MMCF method achieves the highest accuracies in five classes, OA, AA, and Kappa coefficient. Fig. 17 shows the ground truth and classification maps of all studied approaches along with the OAs. Especially, for the LBP method, three principle components (default parameter is 30) are remained to create the LBP features because of the memory limitation, and thus, the classification result is unsatisfactory. If the optimal parameter is adopted, the performance is expected to be better. Misclassification of Spartina alterniflora mainly appears in its adjacent position with other plant types. For example, the SVM classifier with raw data wrongly classifies some Spartina alterniflora areas into Phragmites australis areas. We can find that the proposed methods show satisfactory performance.

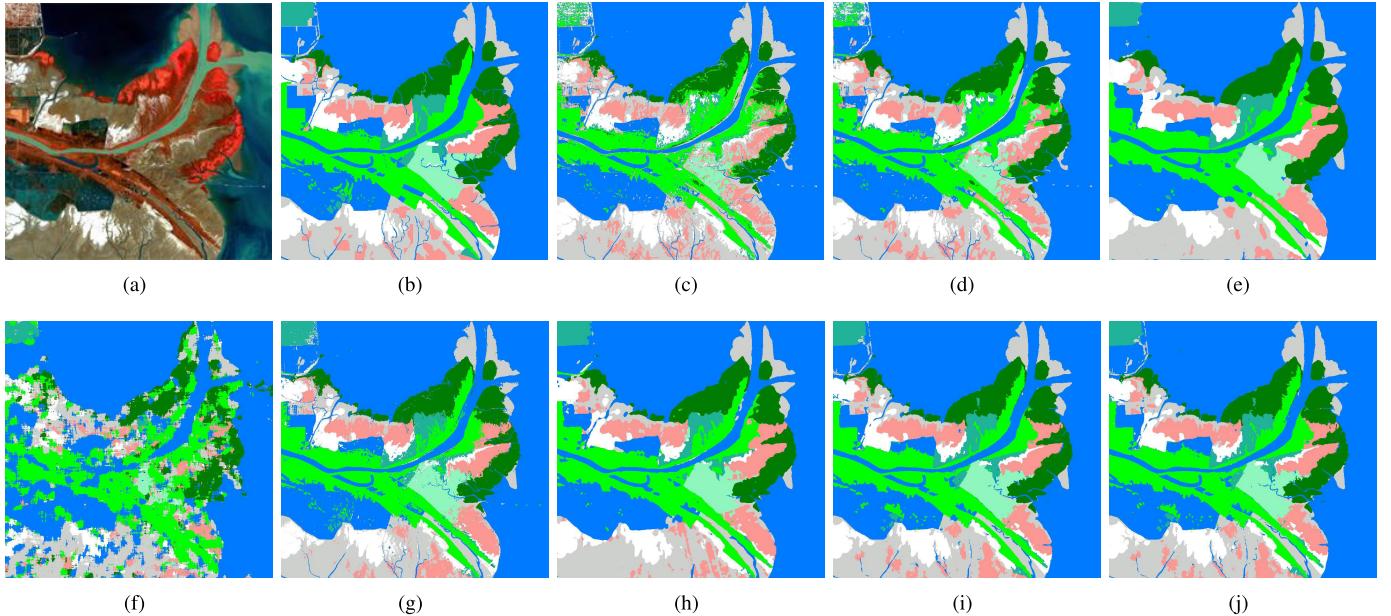


Fig. 17. Three-band representation of the Yellow River Delta dataset and classification maps yielded by using different classification methods. (a) Three-band representation of the image. (b) Ground truth. (c) SVM, OA = 84.29%. (d) EPF, OA = 86.31%. (e) WMRF, 90.85%. (f) LBP, OA = 67.89%. (g) SCMK, OA = 91.88%. (h) GTR, OA = 89.85%. (i) MMCF, OA = 93.44%. (j) MGCF, OA = 93.33%.

TABLE X
COMPUTATION TIME (IN SECONDS) OF THE COMPARED METHODS AND THE PROPOSED METHODS ON THE
INDIAN PINES, SALINAS, AND KSC DATASETS

	SVM	EPF	WMRF	LBP	SCMK	GTR	MMCF	MGCF
Indian Pines	8.68	8.88	37.16	69.33	4.10	5.96	12.58	14.16
Salinas	6.20	7.68	83.87	301.68	7.42	22.36	115.87	108.24
KSC	21.97	25.89	277.89	1656.72	10.99	54.46	415.52	408.02

E. Computation Complexity of the Proposed Method

The computation complexity of the CFs is $O(N)$, and thus, the CF-based image pyramid decomposition method has a computational complexity of $O(LN)$, where N is the number of pixels and L is the number of layers in the pyramid. The complexity of the ERS algorithm approximates $O(|V| \log |V|)$, where V is the vertex set. In our experiments, the computation cost of the proposed methods is measured on a laptop with 2.80 GHz CPU and 8 GB memory. The experiments are implemented in MATLAB. Table X reports the computational time of the compared methods and the proposed methods on the Indian Pines, Salinas, and KSC datasets, respectively. It can be observed that the running time increases as the spectral dimension and spatial size of the data increase. The computing performance of the proposed methods is acceptable in real applications. How to efficiently reduce the running time is an interesting topic? For example, C++ programming is expected to be used to further shorten the computing time of the proposed methods and the general-purpose graphics processing unit (GPU) can be adopted to greatly accelerate the process.

V. CONCLUSION

In this work, a CFs-based multiscale feature extraction method with multiscale superpixel segmentation constraint is

proposed for HSI classification. The proposed method consists the following major components: CFs-based pyramid decomposition, ERS-based multiscale superpixel segmentation, and multiscale feature fusion for classification. Two different CFs, i.e., GCF and MCF, are employed to extract the features with edge preserving. By developing the CFs-based image pyramid decomposition, multiscale spectral features, and global multiscale spatial features are extracted. Multiscale superpixel segmentation is used for spatial constraint and the weighted mean operations can efficiently extract the LMSFs within and among superpixels. The multiscale spectral features, global multiscale spatial features, and LMSFs are combined to increase the classification accuracies. By performing experiments on five real HSIs, experimental results demonstrate that the classification performance of the proposed method is competitive, especially when the number of training samples is limited. In the future, how to automatically select the optimal parameters and how to better preserve details of the HSIs will be investigated.

REFERENCES

- [1] Z. Huang, X. Kang, S. Li, and Q. Hao, "Game theory-based hyperspectral anomaly detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2965–2976, Apr. 2020.
- [2] S. Yang, Z. Shi, and W. Tang, "Robust hyperspectral image target detection using an inequality constraint," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3389–3404, Jun. 2015.

- [3] Q. Hao, S. Li, and X. Kang, "Multilabel sample augmentation-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4263–4278, Jun. 2020.
- [4] S. Prasad and L. M. Bruce, "Limitations of principal components analysis for hyperspectral target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 625–629, Oct. 2008.
- [5] A. J. Izenman, "Linear discriminant analysis," in *Modern Multivariate Statistical Techniques*. Springer, 2013, pp. 237–280.
- [6] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.
- [7] Z. Zhang, E. Pasolli, and M. M. Crawford, "An adaptive multiview active learning approach for spectral-spatial classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2557–2570, Apr. 2020.
- [8] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, "Exploring hierarchical convolutional features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6712–6722, Nov. 2018.
- [9] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, early access, Aug. 18, 2020, doi: 10.1109/TGRS.2020.3015157.
- [10] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. X. Zhu, "Learning to propagate labels on graphs: An iterative multitask regression framework for semi-supervised hyperspectral dimensionality reduction," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 35–49, Dec. 2019.
- [11] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. X. Zhu, "Joint and progressive subspace analysis (JPSA) with spatial-spectral manifold alignment for semisupervised hyperspectral dimensionality reduction," *IEEE Trans. Cybern.*, early access, Nov. 11, 2020, doi: 10.1109/TCYB.2020.3028931.
- [12] 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.
- [13] S. Li, Q. Hao, X. Kang, and J. A. Benediktsson, "Gaussian pyramid based multiscale feature fusion for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 9, pp. 3312–3324, Sep. 2018.
- [14] H. Li, Y. Song, and C. L. P. Chen, "Hyperspectral image classification based on multiscale spatial information fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 9, pp. 5302–5312, Sep. 2017.
- [15] 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.
- [16] Y. Gong and I. F. Sbalzarini, "Curvature filters efficiently reduce certain variational energies," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1786–1798, Apr. 2017.
- [17] Y. Gong and I. F. Sbalzarini, "Local weighted Gaussian curvature for image processing," in *Proc. 20th IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 534–538.
- [18] Y. Gong, "Spectrally regularized surfaces," Ph.D. dissertation, MOSAIC Group, ETH Zürich, Zürich, Switzerland, Mar. 2015.
- [19] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 10–17.
- [20] F. Yang, H. Lu, and M.-H. Yang, "Robust superpixel tracking," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1639–1651, Apr. 2014.
- [21] X. Wang, Y. Tang, S. Masnou, and L. Chen, "A global/local affinity graph for image segmentation," *IEEE Trans. Image Process.*, vol. 24, no. 4, pp. 1399–1411, Apr. 2015.
- [22] M. Van den Bergh, G. Roig, X. Boix, S. Manen, and L. V. Gool, "Online video SEEDS for temporal window objectness," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 377–384.
- [23] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [24] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2011, pp. 2097–2104.
- [25] 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.
- [26] G. B. M. Van den Bergh, X. Boix, and L. Van Gool, "SEEDS: Superpixels extracted via energy-driven sampling," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 13–26.
- [27] A. Vedaldi and S. Soatto, "Quick shift and kernel methods for mode seeking," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 705–718.
- [28] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 17, no. 8, pp. 790–799, Aug. 1995.
- [29] A. Levinshtein, 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.
- [30] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2282, Nov. 2012.
- [31] Z. Li and J. Chen, "Superpixel segmentation using linear spectral clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1356–1363.
- [32] X. Xiao, Y. Zhou, and Y.-J. Gong, "Content-adaptive superpixel segmentation," *IEEE Trans. Image Process.*, vol. 27, no. 6, pp. 2883–2896, Jun. 2018.
- [33] L. Fang, N. He, S. Li, A. J. Plaza, and J. Plaza, "A new spatial-spectral feature extraction method for hyperspectral images using local covariance matrix representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3534–3546, Jun. 2018.
- [34] J. Liao and L. Wang, "Hyperspectral image classification based on fusion of curvature filter and domain transform recursive filter," *Remote Sens.*, vol. 11, no. 7, pp. 833–856, Apr. 2019.
- [35] M. Fauvel, J. Chanussot, and J. Benediktsson, "Kernel principal component analysis for feature reduction in hyperspectral images analysis," in *Proc. 7th Signal Process. Symp.*, Jun. 2006, pp. 238–241.
- [36] P. Duan, X. Kang, S. Li, and P. Ghamisi, "Noise-robust hyperspectral image classification via multi-scale total variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 6, pp. 1948–1962, Jun. 2019.
- [37] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions—I," *Math. Program.*, vol. 14, no. 1, pp. 265–294, Dec. 1978.
- [38] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
- [39] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson, "Customizing kernel functions for SVM-based hyperspectral image classification," *IEEE Trans. Image Process.*, vol. 17, no. 4, pp. 622–629, Apr. 2008.
- [40] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:2, 2011. [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [41] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya, and X. X. Zhu, "Invariant attribute profiles: A spatial-frequency joint feature extractor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791–3808, Jun. 2020.
- [42] R. Bao, J. Xia, M. Dalla Mura, P. Du, J. Chanussot, and J. Ren, "Combining morphological attribute profiles via an ensemble method for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 359–363, Mar. 2016.
- [43] B. Rasti, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral feature extraction using total variation component analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 6976–6985, Dec. 2016.
- [44] X. Kang, S. Li, and J. A. Benediktsson, "Feature extraction of hyperspectral images with image fusion and recursive filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3742–3752, Jun. 2014.
- [45] X. Kang, S. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.
- [46] L. Sun, Z. Wu, J. Liu, L. Xiao, and Z. Wei, "Supervised spectral-spatial hyperspectral image classification with weighted Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1490–1503, Mar. 2015.
- [47] W. Li, C. Chen, H. Su, and Q. Du, "Local binary patterns and extreme learning machine for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, Jul. 2015.
- [48] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.
- [49] J. Liu, Z. Wu, L. Xiao, J. Sun, and H. Yan, "Generalized tensor regression for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 2, pp. 1244–1258, Feb. 2020.



Qiaobo Hao (Student Member, IEEE) received the B.S. degree from Hunan University, Changsha, China, in 2016, where she is pursuing the Ph.D. degree with the Key Laboratory of Visual Perception and Artificial Intelligence of Hunan Province.

From October 2019 to October 2020, she was a Visiting Ph.D. Student with the Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, IN, USA, supported by the China Scholarship Council. Her current research interests include hyperspectral image feature extraction, classification, and object detection.



Bin Sun (Member, IEEE) received the B.S. and Ph.D. degrees in control science and engineering from Hunan University, Changsha, China, in 2010 and 2016, respectively.

From 2017 to 2019, he was a Post-Doctoral Researcher in electrical engineering with the College of Electrical and Information Engineering, Hunan University, where he is an Associate Professor. His research interests include computer vision and human–robot natural interaction.



Melba M. Crawford (Life Fellow, IEEE) received the B.S. and M.S. degrees in civil engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1970 and 1973, respectively, and the Ph.D. degree in systems engineering from The Ohio State University, Columbus, OH, USA, in 1981.

She is the Nancy Uridil and Francis Bossu Professor of civil engineering with the Department of Agronomy, the School of Civil Engineering, and the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA. She is also co-leading the joint initiatives between the Purdue Colleges of Agriculture and Engineering in the development of advanced sensing technologies, data acquisition and analysis, and predictive modeling for plant breeding pipelines and for forest health. Her research interests include advanced methods for image analysis, including dimensionality reduction, AL, and deep learning for classification and prediction, and applications of these methods to hyperspectral and LiDAR data.

Dr. Crawford was a member of the NASA EO-1 Science Validation Team and served on the NASA Earth System Science and Applications Advisory Committee and the Advisory Committee to the NASA Socioeconomic Applications and Data Center (SEDAC). She is the Past President of the IEEE Geoscience and Remote Sensing Society, a past IEEE GRSS Distinguished Lecturer, and the Treasurer of the IEEE Technical Activities Board.



Shutao Li (Fellow, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

In 2001, he joined the College of Electrical and Information Engineering, Hunan University, where he is a Full Professor. From May 2001 to October 2001, he was a Research Associate with the Department of Computer Science, The Hong Kong University of Science and Technology, Hong Kong, where he was also a Visiting Professor from April 2005 to June 2005. From November 2002 to November 2003, he was a Post-Doctoral Fellow with the Royal Holloway College, University of London, Egham, U.K. He has authored or coauthored over 200 refereed articles. He received two 2nd-Grade State Scientific and Technological Progress Awards of China in 2004 and 2006. His current research interests include image processing, pattern recognition, and artificial intelligence.

Dr. Li is a member of the Editorial Board of the *Information Fusion* and the *Sensing and Imaging*. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.



Xudong Kang (Senior Member, IEEE) received the B.Sc. degree from Northeastern University, Shenyang, China, in 2007, and the Ph.D. degree from Hunan University, Changsha, China, in 2015.

In 2015, he joined the College of Electrical Engineering, Hunan University. His research interests include hyperspectral feature extraction, image classification, image fusion, and anomaly detection.

Dr. Kang received the National Nature Science Award of China (Second Class and Rank as Third) and the Second Prize in the Student Paper Competition in IGARSS 2014. He was also selected as the Best Reviewer of the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING from 2018 to 2019. He serves as an Associate Editor for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS and the IEEE JOURNAL ON MINIATURIZATION FOR AIR AND SPACE SYSTEMS.