Superpixel-Guided Variable Gabor Phase Coding Fusion for Hyperspectral Image Classification

Shuyu Zhang, Dingding Tang, Nanying Li, Xiuping Jia, *Fellow, IEEE*, and Sen Jia, *Senior Member, IEEE*

***Abstract*— 3-D Gabor, as a typical filter, plays a critical role in extracting discriminative spectral–spatial features from hyperspectral images (HSIs). However, the performance of traditional 3-D Gabor is limited by the uniform response to each direction, which is inconsistent with the complexity of land cover distribution. It has been a continuing concern for researchers to investigate the anisotropic 3-D Gabor filters. In addition, the 3-D Gabor wavelets do not make full use of spatial distribution information, thus reducing the accuracy. This article proposes a superpixel-guided variable 3-D Gabor phase coding fusion (*Su*VGF) framework for HSI classification with limited training samples. First, the variable 3-D Gabor filters are created based on various asymmetric sinusoidal waves and spatial kernel sizes to achieve multidirectional features. Second, the local Gabor phase ternary pattern is adopted to encode the Gabor phases and improve the feature discrimination. Meanwhile, a scale map is produced by the majority voting of multiscale simple noniterative clustering (SNIC) and entropy rate superpixel (ERS) segmentation, which contains sufficient and complementary spatial distribution information. Then, geometric optimization is employed on the scale map to reduce noise disturbances. Finally, all Gabor features are modified by the filter with the guidance of a scale map and fused together as a confidence cube, and the random forest algorithm is exploited for classification. The *Su*VGF is applied to three real hyperspectral datasets to demonstrate the superiority of higher accuracy, stronger robustness, and less computational complexity in comparison with several state-of-the-art ones.**

***Index Terms*— Feature extraction, hyperspectral image (HSI) classification, superpixel segmentation, Variable Gabor (VG) filter.**

I. Introduction

HYPERSPECTRAL images (HSIs) acquire spatial information and simultaneously collect tens or even hundreds of narrowband continuous spectral information, which allows the detection and analysis of the surface material [1], [2]. The increment of spectral dimensionality makes HSI look like a data cube, and each pixel of HSI is equipped with a vector of hyperspectral profile. HSI classification is one of the leading research fields, which has been broadly applied to precision agriculture [3], environmental monitoring [4], ocean exploration [5], and so on. The high spectral resolution of HSI achieves better attribute description but raises many challenges for accurate feature extraction: 1) the high dimensionality of HSIs leads to the increased number of free parameters to be estimated in the model, which can easily cause the overfitting problems and reduce the generalization ability [6], [7] and 2) the high cost and time-consuming of manual labeling result in the relatively small training sample set, and the Hughes phenomenon is inevitable [5]. Therefore, it has become a popular topic to effectively improve the HSI classification performance with the small number of labeled samples.

Dimensionality reduction (DR) extracts more representative features and removes redundant information from the HSI, including the methods of feature selection and feature transformation. Feature selection directly selects the most distinguishable features from the original HSI to form a new low-dimension feature space for classification [8], [9]. Differently, feature transformation uses linear or nonlinear operations to map the original spectral attributes from a high-dimension space to a low-dimension subspace [10]. The principal component analysis (PCA) [11] is a widely used feature transformation method, and it is subsequently expanded to the kernel principal component analysis (KPCA) [12]. Besides, the 1-D convolutional neural network (1-D CNN) [13]–[15], the generative adversarial network (GAN) [16], [17], and the recurrent neural network (RNN) [18], [19] were designed as spectral feature extractors to explore the correlation between hyperspectral vectors. The DR based on spectral information can effectively alleviate the Hughes phenomenon and achieve better classification results. Nonetheless, the ignorance of spatial information has led to unsatisfactory accuracy in regions with complex spatial structures. In recent years, researchers have been more and more interested in exploiting the spatial semantic information of HSIs. The shape and size of various components and objects provide additional spatial information to enhance the feature ability and classification accuracy. Typical methods of spatial feature extraction include the gray-level cooccurrence matrix (GLCM) [20], the wavelet transform [21], the Gabor filter, the local binary descriptor [22], the morphological operator [23], the Markov random field [24], [25], and so on. CNN methods can learn spatial features through

Manuscript received December 2, 2021; revised January 23, 2022; accepted February 10, 2022. Date of publication February 15, 2022; date of current version March 24, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 41971300 and in part by the Key Project of the Department of Education of Guangdong Province under Grant 2020ZDZX3045. *(Corresponding author: Sen Jia.)*

Shuyu Zhang, Dingding Tang, Nanying Li, and Sen Jia are with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China, and also with the Key Laboratory for Geo-Environmental Monitoring of Coastal Zone, Ministry of Natural Resources, Shenzhen University, Shenzhen 518060, China (e-mail: shuyu-zhang@szu.edu.cn; tangdingding2019@email.szu.edu.cn; linanying2021@email.szu.edu.cn; senjia@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.2022.3151875

iterative convolution calculation of receptive fields to increase the representation capabilities [26]. For HSIs with high spatial resolution, the reduced interclass separability and increased inner class variance would lead to more misclassification of land covers, and thus, spatial features are important for accurate recognition.

Subsequently, the joint spectral–spatial feature extraction has been powerful means for HSI classification and attracted more attention [23]. There exist methods of separate processing and simultaneous processing to combine the spatial and spectral information in different ways. The separate processing methods achieve spectral and spatial features independently and use the spatial component to guide spectral features to obtain classification results. There are some typical models such as sparse representation [27], [28], low-rank representation [29], [30], extended morphological attribute profile (EMAP) [31], [32], and local binary pattern (LBP) [22], [33]. Direct stacking of spectral and spatial features will yield higher dimensionality and increase the redundancy and complexity, whereas separate spectral–spatial feature extraction alleviates this issue. Two-branch CNN architecture was developed to extract spectral and spatial characteristics, respectively, to obtain deep and comprehensive expression [34], [35]. In addition, superpixels are employed to correct the classification of spectral attributes with the neighborhood information, via spatial preprocessing or postprocessing operation [30], [36], [37]. However, separate processing methods cannot utilize the characteristics of 3-D HSI cube sufficiently, lacking the integrated extraction of intrinsic spectral–spatial features. On the contrary, simultaneous processing methods view HSI as a 3-D structure and process spectral–spatial features using 3-D descriptors to make better identification. A series of algorithms have been extended to three dimensions to obtain better feature representation, such as 3-D GLCM [38], 3-D discrete wavelet transform (3-D DWT) [39], 3-D dense LBP [40], 3-D CNN [13], 3-D GAN [41], and multiple kernel learning (MKL) [42], [43]. MKL methods enhanced flexibility and achieved superior performance over single-kernel methods since multiscale kernels can fuse comprehensive information of single kernels at different scales. Likewise, 3-D Gabor filters have been successfully applied in many HSI classification tasks [30], [44]. Shen and Jia [45] adopted 3-D Gabor with thirteen directions in four scales to obtain the magnitude features for classification and demonstrated its superiority of capturing distinguishable spectral–spatial details.

3-D Gabor plays a critical role in extracting the discriminative spectral–spatial features from HSIs. However, the performance of traditional 3-D Gabor is limited by the uniform response to each direction, which is inconsistent with the complexity of land cover distribution. That is to say, most 3-D Gabor filters are isotropic representations, whereas anisotropic filters match the spatial distribution better. It has been a continuing concern for researchers to investigate the anisotropic 3-D Gabor filters [46], [47]. Moreover, it is hard to select a suitable spatial kernel for different HSIs, and it cannot efficiently acquire multiscale details with a single kernel. Furthermore, directly convolving the original HSI of multiple directions and scales inevitably leads to high feature dimensionality, and some feature selection methods were proposed to address the redundancy [48], [49]. Traditionally, the Gabor phase has often been ignored as it is more sensitive to positional changes, whereas binary quadrant coding makes phase features usable. It codes Gabor phase features effectively to determine the sample similarity and achieves favorable classification results, which is jointed with the Hamming distance [50], [51]. On the other hand, the Gabor filter can extract texture information but fails to achieve satisfactory smoothness on class boundaries. In comparison, superpixels offer a good representation of local regions and maintain object boundaries, which is fast becoming a key instrument in regularizing preclassified results. However, different superpixel algorithms use distinct segmentation criteria and obtain various visual contents, and iterative attempts are required to determine the optimal superpixel size for each HSI. Therefore, it is necessary to explore the diverse feature extraction and superpixel segmentation to capture more visual information and improve the classification performance of 3-D Gabor filters.

In this article, a superpixel-guided variable 3-D Gabor phase coding fusion (*Su*VGF) framework is proposed to achieve higher classification accuracy with limited training samples, as shown in Fig. 1. First, we introduce two parameters to control the type of plane sinusoidal waves for each group of 3-D Gabor filters, capturing the local characteristics of sophisticated interband correlations. For each group of Gabor filters, only the filters parallel to the spectral axis are used to enhance the discriminability and reduce redundancy. Then, local Gabor ternary patterns (LGTPs) are applied in the variable 3-D Gabor phase features to describe the variation between the center pixel and its neighbors (called V-Gabor features). Furthermore, V-Gabor features are classified via the random forest (RF) classifier to produce a confidence cube. Meanwhile, two superpixel segmentation algorithms are applied to HSI, including simple noniterative clustering (SNIC) and entropy rate superpixel (ERS) algorithms at multiple scales. During multiscale superpixel segmentation, the minimum distance is calculated from each pixel to the points of boundaries. The final scale map is generated by the majority voting of all multiscale distance maps, which conducts the selection of convolution kernel size. Subsequently, the confidence cube is convolved with the Gaussian filter via the guidance of the scale map, and the *Su*-Variable Gabor (VG) cube is obtained. The scale map is beneficial to maintain the edges and ensure the spatial structure. Finally, the comprehensive classification results are achieved by linear summation of all *Su*-VG cubes to improve the classification performance and identification accuracy.

Compared with other related methods and the previous works of [40], [52], which focus on 3-D LBPs and multiple magnitude features of 3-D Gabor, respectively, our approach presents three main innovations, which are given as follows.

1) With respect to the anisotropic and nonlinear characteristics of HSI, we create a new format of the sine wave to control the grating of directed sinusoids and obtain a more robust representation. Besides, we adopt the Gabor filters of multiple kernels to track local response regions, which can effectively catch the changes in distribution patterns. The improved representation of frequency and direction and the strategy of multiple kernels

are conductive to comprehensive feature extraction and meaningful model construction.

2) The Gabor phase is a significant representation of texture information, but matching directly with phase angles will undoubtedly decrease the computational efficiency. We first integrate the local ternary pattern (LTP) with the Gabor phase to obtain features that are more robust to noise interference for HSI classification. The LTP distinguishes positive response mode and negative response mode, which can retain as much information as possible during feature extraction and preserve the feature distinguishability.
3) Two types of multiscale superpixel algorithms are designed to acquire spatial information to characterize the homogeneity degree of local regions, and a scale map is generated via majority voting. The consideration of multiscale spatial characteristics is complementary with the VG phase coding. The scale map is employed to select the convolution kernel size of the variable filter, reduce interference from noise, and further improve classification performance.

The rest of this article is organized as follows. Section II briefly introduces some basic notations of related works. Section III presents the steps of proposed *Su*VGF, respectively, in detail. Section IV shows the information of three real HSI datasets and describes the experimental setup. The experimental results are described in Section V. Finally, the conclusions are given in Section VI.

II. Related Works

In this section, we briefly introduce some basic notions of the 3-D Gabor filter and superpixel segmentation.

A. 3-D Gabor Wavelet

Over the last decades, the Gabor filter has long been a powerful texture information extractor in a wide range of fields, such as texture analysis [53], face recognition [54], and HSI classification [49]. Gabor wavelet can enhance the representation of visual properties due to its biological relevance to human vision. Traditional Gabor wavelet is composed of a plane sinusoid and a convolution kernel, constrained by the Gaussian envelope function [55]. The 3-D Gabor wavelet is defined in a general form as

$$\Phi(x, y, b) = u(x, y, b) \times \upsilon(x, y, b) \tag{1}$$

where (x, y) is the spatial coordinate and b is the spectral band for a point (x, y, b). $u(x, y, b)$ is the plane sinusoidal wave, and $\upsilon(x, y, b)$ is the Gaussian envelop function. $u(x, y, b)$ and $\upsilon(x, y, b)$ are defined as follows:

$$u(x, y, b) = \exp\big(j2\pi\left(F_x x + F_y y + F_b b\right)\big) \tag{2}$$

$$\upsilon(x, y, b) = \frac{1}{(2\pi)^{3/2}\sigma^3}\exp\left(-\frac{x^2+y^2+b^2}{2\sigma^2}\right) \tag{3}$$

where σ is the standard deviation of the Gaussian function, and it controls the shape of the Gaussian envelope. (F_x, F_y, F_b) describes the multiple frequencies of sinusoidal function, respectively. These parameters are computed as

$$\begin{aligned} F_x &= f\cos\varphi\sin\theta \\ F_y &= f\sin\varphi\sin\theta \\ F_b &= f\cos\varphi \end{aligned} \tag{4}$$

where (F_x, F_y, F_b) are usually determined by three parameters: f is the center frequency of sinusoidal wave; θ and φ are the angles of wave vector in 3-D frequency domain, which are shown in Fig. 2. θ and φ range from 0 to π, and the directions of θ and φ are changed by 45° each time. More precisely, f, θ, and φ are specifically defined as

$$\begin{aligned} f &\in \left[\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}\right] \\ (\theta, \varphi) &\in \left[0, \frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\right]. \end{aligned} \tag{5}$$

Thus, there are totally 52 Gabor filters (when $\varphi = 0$, the corresponding filter is one wavelet, not four wavelets) in four frequencies. Generally, Gaussian envelopes and sinusoids tend to be the same although they can have different orientations.

B. Superpixel Segmentation

Superpixel segmentation clusters neighboring pixels with similar texture, color, luminance, and other attributes in HSI to form regions with adjacent pixels. Ren and Malik [56] first proposed that the superpixel segmentation of spatial information produces local regions with specific semantic content, which significantly reduces the dimensionality redundancy and algorithm complexity. The superpixel segmentation algorithms are usually classified into graph-theoretic and gradient-based approaches. The corresponding classical and representative algorithms are ERS [57] and SNIC [58], respectively, and their descriptions are given in the following.

ERS divides the HSI into N nonoverlapping uniform regions, where N denotes the number of superpixels and can be calculated heuristically as [36]

$$N = \left\lfloor \frac{X \times Y}{100 \times \tau^{\sqrt{\mathrm{Res}}}} \right\rfloor \tag{6}$$

where $\lfloor\cdot\rfloor$ represents the rounding down operation. Res is the spatial resolution of HSI, determined by the imagery sensor. In particular, the value of τ should be less than 1. $\mathbf{R} \in \mathbb{R}^{X\times Y\times B}$ represents the raw HSI, $X \times Y$ corresponds to the spatial size, and B is the number of bands.

The HSI can be represented as a nondirectional graph $\mathbf{U} = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V}$ and $\mathbf{E}$ denote the set of vertices and edges in $\mathbf{U}$, respectively. ERS aims to find a suitable subset $\mathbf{A} \subseteq \mathbf{E}$ to divide $\mathbf{U}$ into N subgraphs, and each subgraph corresponds to a superpixel. The entropy rate of random walk on a subgraph $\widetilde{\mathbf{U}}_{\mathbf{A}} = (\mathbf{V}, \mathbf{A})$ is defined as $\mathcal{H}(\mathbf{A})$, which is the quantification of random process uncertainty and can be used to determine the compactness and uniformity. A balance term function $\mathcal{B}(A)$ is constructed to constrain the subgraphs and make superpixels more compact and homogeneous. Finally,

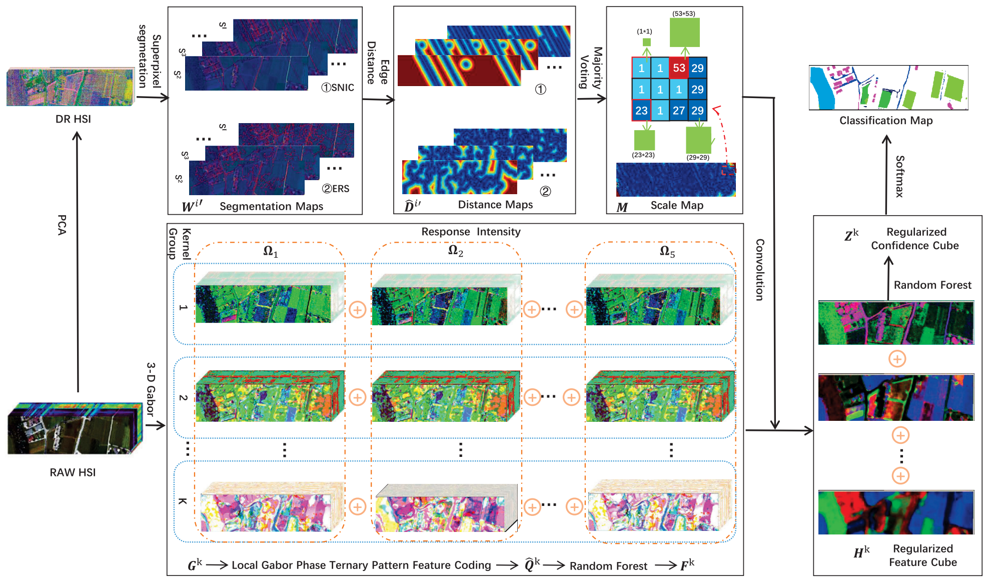


Fig. 1. Proposed *Su*VGF procedure for HSI classification.

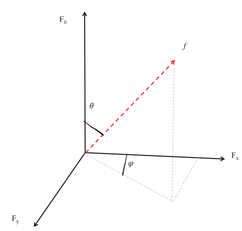


Fig. 2. 3-D frequency domain.

N connected subgraphs are obtained by optimizing the objective function, which is defined as follows:

$$F_{\text{ers}} = \max_{\mathbf{A}} \mathcal{H}(\mathbf{A}) + \omega\mathcal{B}(\mathbf{A}) \quad \text{s.t. } \mathbf{A} \subseteq \mathbf{E} \tag{7}$$

where ω represents the weight with a value greater than zero, and the appropriate selection of ω value has an essential effect on the generation of superpixels. The segmentation process is completed by maximizing the object F_{ers}, and the ERS algorithm produces relatively satisfactory superpixels through the collaborative work of $\mathcal{H}(\mathbf{A})$ and $\mathcal{B}(\mathbf{A})$.

SNIC is an improved version of simple linear iterative (SLIC), and it has the advantages of low computational complexity and good segmentation results. The distance from a pixel (x_1, y_1) to another pixel (x_2, y_2) in HSI from the CIELAB color domain $\boldsymbol{\Upsilon}$ and spatial domain $\boldsymbol{\chi}$ is computed as

$$F_{\text{snic}} = \sqrt{\frac{\left\|\boldsymbol{\chi}_{(x_1,y_1)} - \boldsymbol{\chi}_{(x_2,y_2)}\right\|^2}{\xi_1} + \frac{\left\|\boldsymbol{\Upsilon}_{(x_1,y_1)} - \boldsymbol{\Upsilon}_{(x_2,y_2)}\right\|^2}{\xi_2}} \tag{8}$$

where ξ_1 is derived from $((X \times Y)/N)^{1/2}$, and $X \times Y$ is the total number of HSI pixels. ξ_2 is the expected compactness parameter, which describes the weight of spatial coordinates. Specifically, SNIC divides the HSI into N square regions and then calculates the geometric center of each area as the initial center. A priority queue is built to record the distance from nodes to centers. Equation (8) is used to obtain the element with shortest distance in priority queue, and its eight neighboring pixels are calculated and added to the priority queue. Meanwhile, the corresponding center coordinates evolve until the queue is empty.

III. Superpixel-Guided Variable Gabor Fusion Framework

This section introduces the proposed method *Su*VGF for HSI classification in detail, and the schematic of *Su*VGF is illustrated in Fig. 1. First, the variable 3-D Gabor feature extraction is applied to the raw HSI to obtain Gabor features $\mathbf{G}^k$ of multiple kernels, with response intensity parameters $\boldsymbol{\Omega}$. Second, the LTP is designed to encode the phase features of $\mathbf{G}^k$ to produce local Gabor phase ternary pattern features $\widehat{\mathbf{Q}}^k$, which are then classified via RF to obtain initial confidence cubes $\mathbf{F}^k$. Meanwhile, we make DR via PCA on the raw HSI to get DR HSI. Then, the scale map $\mathbf{M}$ is constructed from various multiscale superpixel maps $\mathbf{W}^{i'}$ and distance maps $\widehat{\mathbf{D}}^{i'}$ through the edge distance calculation and majority voting. The scale map guides the Gaussian smoothing on $\mathbf{F}^k$ to get regularized feature cubes $\mathbf{H}^k$, which are then classified via RF to obtain regularized confidence cubes $\mathbf{Z}^k$. Finally, the strategy of multiple kernel fusion is utilized to generate the classification results.

A. Variable 3-D Gabor Feature Extraction

In HSI processing and analysis, the Gabor wavelet is a classical method with a fantastic performance for texture extraction. Commonly, most methods directly use 52 3-D Gabor filters of different frequencies and orientations to obtain an effective representation. However, this would cause massive redundancy of data and exacerbate the Hughes phenomenon. Jia *et al.* [50] found that the 3-D Gabor filters oriented parallel to the spectral axis are more representative than the others, so all the 3-D Gabor filters that we applied are parallel to the spectral axis. The specific parameters of the four obtained 3-D Gabor filters are listed as

$f \in [1/2, 1/4, 1/8, 1/16]$, $(\theta, \varphi) \in [0]$. Collectively, we make enhancements to traditional 3-D Gabor filters in terms of the response intensity parameters of sine waves and the convolution kernel size to create variable 3-D Gabor filters.

On the one hand, we create a more general format for the response intensity of sinusoids. The sinusoids of the traditional 3-D Gabor filter are defined as (2), showing a uniform response in all directions. However, there are complicated spatial distribution and diverse spectral attributes within HSI, which are hard to represent using the uniform response [52], [59]. In contrast, anisotropic Gabor wavelets fit the internal structure of HSI better with a diverse representation of information. For this reason, we create a new format of sine wave $u_{\boldsymbol{\lambda}}(x, y, b)$ to explore the different structural components of HSI, which is denoted as follows:

$$u_{\boldsymbol{\lambda}}(x, y, b) = \exp\left(j2\pi\left(F_x x^{\lambda_1} + F_y y^{\lambda_1} + F_b b^{\lambda_2}\right)^{\frac{1}{\lambda_3}}\right) \quad (9)$$

where (x, y, b) represents a point situated at the spatial coordinate (x, y) and the bth band in HSI. Here, $u_{\boldsymbol{\lambda}}$ represents the isotropic sine wave when $\boldsymbol{\lambda} = (1, 1, 1)$ and represents the anisotropic sine wave when $\boldsymbol{\lambda}$ equals to other values. The parameter $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \lambda_3)$ is a set of response intensity parameters, controlling the grating degree of oriented sinusoids. Within this, λ_1, λ_2, and λ_3 describe the response intensities in the spatial domain, the spectral domain, and the overall structure, respectively. Their relationship is defined as

$$\lambda_3 = \frac{2\lambda_1 + \lambda_2}{3} \quad (10)$$

where the value of λ_3 is rounded to the nearest integer. Varying the response intensity parameters creates the variable 3-D Gabor filters with different properties and extracts a more diverse collection of spectral–spatial information.

Admittedly, there are many possible values of $\boldsymbol{\lambda}$, and the integrated consideration of isotropic and anisotropic sine waves is beneficial to extract more discriminative features. Nonetheless, too much choices of $\boldsymbol{\lambda}$ setting would lead to the problems of high feature redundancy and low calculation efficiency, reducing the classification performance. Here, we design and retain five specific choices to balance the feature representation and model complexity, which are shown as

$$\boldsymbol{\Omega} = [(1, 1, 1), (1, 2, 1), (1, 3, 2), (2, 1, 2), (2, 2, 2)] \quad (11)$$

where $\boldsymbol{\Omega}$ is the entire collection of $\boldsymbol{\lambda}$. It is worth mentioning that $\boldsymbol{\lambda} = (1, 1, 1)$ describes the intensity parameter of traditional 3-D Gabor filter, and the others represent the parameters of anisotropic 3-D Gabor filters. A series of 3-D Gabor filters with different response intensity parameters can be expressed as

$$\Phi_{\boldsymbol{\lambda}}(x, y, b) = u_{\boldsymbol{\lambda}}(x, y, b) \times \upsilon(x, y, b) \quad (12)$$

where $\Phi_{\boldsymbol{\lambda}}$ captures the uneven 3-D spectral–spatial information and ensures the integrity of representation. $u_{\boldsymbol{\lambda}}$ denotes the isotropic or anisotropic sine wave, and υ denotes the Gaussian envelope function. Subsequently, five types of Gabor filters are stacked along the direction of spectral axis. Such a new module of multiple response intensity parameters reduces the data redundancy and model complexity.

On the other hand, the Gaussian function of the Gabor filter tends to use a fixed kernel parameter, which would lead to the underlearning when the spatial distribution is not identical. Moreover, the parameter mostly needs to be chosen artificially through extensive experiments, not only reducing the generality but also increasing the workload. In this case, we introduce a learning strategy of multiple kernels to solve this problem and enhance feature extraction. Gabor wavelets with multiple kernels can fully incorporate the various heterogeneous elements to efficiently represent the structure of HSI. Considering the traditional perspective that the center pixel is most correlated with its neighborhood in a 3×3 region, the lower threshold of kernel size is set to three. For HSI with high spatial resolution, there are obvious detailed information and large intraclass differences, and it is suitable to use a smaller window for the Gabor feature extractor to avoid introducing interference. Alternatively, for HSI with low spatial resolution, there are more smooth details, and it is desirable to employ a larger window to improve the accuracy of feature extraction. Moreover, HSI data with more categories contain more contextual information, and a larger window is required to capture the spatial relationships. Therefore, the upper threshold J of kernel size is defined as

$$J = \begin{cases} \dfrac{T-3}{2} + \left\lfloor C \times \left(\sqrt{\mathrm{Res}}\right)\right\rfloor, & \text{if } J \le T \\ T, & \text{otherwise} \end{cases} \quad (13)$$

where C and Res denote the number of land cover classes and spatial resolution (i.e., meters per pixel), respectively. T is a user-defined parameter that controls the upper threshold. The maximum value of J is set to T because the spatial regions contain different materials and noise when the scale continuously expands. At the same time, the minimum value of J is set as $((T-3)/2)$ to avoid a small upper threshold and, thus, limit Gabor feature extraction. We select K groups of kernel size parameters for variable 3-D Gabor filters, whose values range from three to J. The maximum and minimum settings of J make K change within a proper range related to HSI. The features generated with different types of sinusoidal plane wave for each group are stacked along the spectral dimension $\Phi^k = \{\Phi_1^k, \ldots, \Phi_5^k\}, k = 1, \ldots, K$.

Naturally, the characteristics obtained via variable 3-D Gabor filters can be defined as

$$\mathbf{G}^k = \Phi^k \otimes \mathbf{R}, \quad k = 1, \ldots, K \quad (14)$$

where $\otimes$ is the convolution operation and $\mathbf{R}$ is the HSI data cube. In summary, the obtained variable 3-D Gabor features for each group are defined as $\mathbf{G}^k$, where $k = 1, \ldots, K$.

B. Local Gabor Phase Ternary Pattern Feature Coding

The traditional 3-D Gabor wavelet is extended to obtain abundant Gabor features by optimizing the response intensity parameters and spatial kernel dimensions. The coefficient of a point (x, y, b) is a complex number, which is obtained by Gabor wavelet Φ^k, including the magnitude and phase of convolutional results. Among them, the phases catch variation

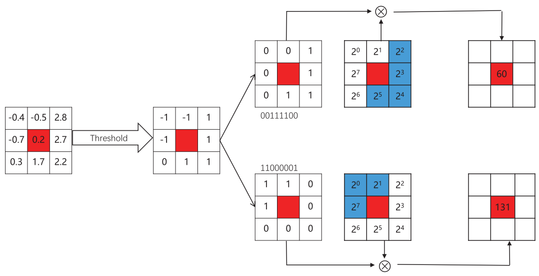


Fig. 3. Calculation of LTP operators.

in the surface texture, which are computed as

$$P^k(x, y, b) = \arctan \frac{\mathrm{Im}\big(\mathbf{G}^k(x, y, b)\big)}{\mathrm{Re}\big(\mathbf{G}^k(x, y, b)\big)} \tag{15}$$

where Re means the real part and Im represents the imaginary of complex coefficient [60]. The real and imaginary parts are combined to produce a coordinate axis, as shown in Fig. 3.

It has been revealed that phase characterization can improve the feature stability and discrimination via appropriate encoding methods [61]. Compared with LBP, LTP retains the thresholds and expands the codes into -1, 0, and 1 instead of 0 and 1, which makes it more hierarchical than the binary basis. Specifically, LBP treats the positive response and negative response as the same, but the positive and negative edge responses have obvious visual differences. Thus, LBP is sensitive to the nonuniform illumination and random noise, which are poorly discriminated. In contrast, LTP distinguishes the positive response and the negative response, and will not mistake different modes with huge differences, retaining as much information as possible during feature extraction. Therefore, we adopt local Gabor phase ternary pattern to encode the phases to produce more robust features. The LTP takes a pixel (x, y, b) as the center and makes the comparison between other pixel values $\widehat{P}^k(x, y, b)$ and center pixel value $P^k(x, y, b)$ in a 3×3 region. The feature coding is assigned to 1, -1, and 0 if the absolute value of comparison result is greater than threshold ρ, less than $-\rho$, and otherwise, respectively, which is expressed as

$$\mathbf{Q}^k(x, y, b) = \begin{cases} 1, & \widehat{P}^k(x, y, b) \geq P^k(x, y, b) + \rho \\ 0, & |\widehat{P}^k(x, y, b) - P^k(x, y, b)| < \rho \\ -1, & \widehat{P}^k(x, y, b) \leq P^k(x, y, b) - \rho. \end{cases} \tag{16}$$

For complexity reduction, the coding of LTP is split into an upper layer and a lower layer. The upper part replaces all -1 s with 0 s and leaves the rest as it is. It assigns different weights to the comparison results at different positions, thus obtaining a weighted sum for each position. The process can be described as

$$\mathrm{up}^k(x, y, b) = \begin{cases} 0, & \text{if } \mathbf{Q}^k_{(x,y,b)} = -1 \\ \mathbf{Q}^k_{(x,y,b)}, & \text{otherwise} \end{cases} \tag{17}$$

$$\mathrm{LTP}^k_{\mathrm{up}}(x, y, b) = \sum_{n=0}^{7} 2^n \mathrm{up}^k(x, y, b). \tag{18}$$

The lower part replaces all -1 s with 1 s, replaces 1 s with 0 s, and leaves the rest as it is, respectively. Then, the weighted sum of comparison results is obtained, which can be described as

$$\mathrm{low}^k(x, y, b) = \begin{cases} 1, & \text{if } \mathbf{Q}^k_{(x,y,b)} = -1 \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

$$\mathrm{LTP}^k_{\mathrm{low}}(x, y, b) = \sum_{n=0}^{7} 2^n \mathrm{low}^k(x, y, b). \tag{20}$$

The eigenvalues are concatenated as the local Gabor phase features $\widehat{\mathbf{Q}}^k = \{\mathrm{LTP}^k_{\mathrm{up}}, \mathrm{LTP}^k_{\mathrm{low}}\}$, which then are classified by the RF algorithm to obtain an initial confidence cube $\mathbf{F}^k$.

C. Multiscale Superpixel Segmentation

The variable 3-D Gabor wavelets exploit the nonlinear statistical properties in HSI from the perspective of spectral information and feature space. However, it ignores the spatial adjacency between pixels. Fortunately, superpixel segmentation is able to reflect the spatial distribution of local regions, via distance calculation and similarity comparison [62]. Recently, superpixels, which are homogeneous and uniform patches in images, have been increasingly applied in HSI classification. Hence, the spatial information of superpixels can be used to regularize and optimize the preclassification results of pixels. In this regard, there is a requirement for the richness of spatial information, and we use various multiscale superpixel maps to generate a scale map to measure spatial similarity comprehensively and adaptively.

Considering the variation of object shapes and distribution patterns, we use superpixel maps at various scales to extract the spatial information in HSI. The number of superpixels in the ith scale N_i is computed as follows:

$$N_i = \frac{X \times Y}{S_i}, \quad i = 2, 3, \ldots, I \tag{21}$$

where S_i denotes the estimated number of pixels in superpixel, which is calculated by the designed heuristic formula

$$S_i = \left\lfloor \frac{10 \times i}{\mathrm{Lg}(\mathrm{Res} + 1)} \right\rfloor \tag{22}$$

where $\mathrm{Lg}(\cdot)$ means the logarithm based on 10, and Res is the HSI spatial resolution (i.e., meters per pixel). For the ith scale, N_i is one of the input key parameters to ERS and SNIC algorithms.

We can obtain $I - 1$ different superpixel maps $\mathbf{W} = \{\mathbf{W}^2, \ldots, \mathbf{W}^I\}$ for each superpixel segmentation algorithm through the variation of scale parameter i. We use multiscale ERS algorithm to create superpixel segmentation maps $\mathbf{W}_{\mathrm{ERS}} = \{\mathbf{W}^2_{\mathrm{ERS}}, \ldots, \mathbf{W}^I_{\mathrm{ERS}}\}$ and exploit multiscale SNIC algorithm to obtain superpixel segmentation maps $\mathbf{W}_{\mathrm{SNIC}} = \{\mathbf{W}^2_{\mathrm{SNIC}}, \ldots, \mathbf{W}^I_{\mathrm{SNIC}}\}$ to take the complementary advantages of different segmentation and multiple scales. $\mathbf{W}_{\mathrm{ERS}}$ and $\mathbf{W}_{\mathrm{SNIC}}$ are combined together to get the total set of superpixel maps, and the i'th map is expressed as $\mathbf{W}^{i'}, i' \in 1, \ldots, 2(I-1)$.

At the next step, we introduce a geometric model into the various multiscale superpixel segmentations to enhance the feature expression under spatial constraints. The spatial similarity of pixels is converted into the calculation of distance scale, and the spatial distribution of pixels is determined by the minimum distance from pixels to points on superpixel edges. Specifically, for the pixels close to superpixel edges, the surrounding context is more complex, and the category distinction is not particularly obvious, so the contextual information in the smaller neighborhood with shorter distances should be considered. Hence, we choose the minimum value to generate the initial scale map $\widehat{\mathbf{D}}^{i'}(x, y)$, which is expressed as

$$\widehat{\mathbf{D}}^{i'}(x, y) = \min_{(x',y')\in\zeta}\left\{\sqrt{(x-x')^2+(y-y')^2}\right\}$$
$$(x, y) \in \mho, \quad i' \in 1, \ldots, 2(I-1) \tag{23}$$

where $\mho$ represents the set of pixel spatial coordinates and (x', y') represents any point on superpixel edges. i' ranges from 1 to the total number of multiscale ERS and SNIC superpixel maps.

Directly utilizing $\widehat{\mathbf{D}}^{i'}(x, y)$ as the convolution kernel radius size of variable filter is highly risky to oversize and interfere with other features. Geometrically, the new radius size is considered as one-half size of the square window diagonal based on original minimum distance to avoid confusing interference. Therefore, the geometric optimization process converts the distance map into a scale map, which is represented as

$$\mathbf{D}^{i'}(x, y) = 2 \times \left\lfloor \frac{\widehat{\mathbf{D}}^{i'}(x, y)}{\sqrt{2}} \right\rfloor + 1, \quad (x, y) \in \mathbf{\Omega}. \tag{24}$$

In the scale map, $\mathbf{D}^{i'}(x, y)$ is adopted as the convolution kernel diameter size of variable filter. Threshold is set to T to avoid the filter becoming too large and bringing in sample points of other categories. The threshold processing is expressed as

$$\mathbf{D}^{i'}(x, y) = \begin{cases} \mathbf{D}^{i'}(x, y), & \mathbf{D}^{i'}(x, y) \le T \\ T, & \mathbf{D}^{i'}(x, y) > T \end{cases} \tag{25}$$

where T is a user-defined parameter controlling the upper range of kernel size, which is the same parameter in (13). The value of $\mathbf{D}^{i'}(x, y)$ represents the convolution kernel size generated from superpixel map $\mathbf{W}^{i'}$, and it is odd value between one and T. Finally, the integrated scale map $\mathbf{M}$ is further generated by the majority voting of various multiscale superpixel maps, which can be shown as

$$\mathbf{M}(x, y) = \mathrm{MV}\{\mathbf{D}^1, \ldots, \mathbf{D}^{2(I-1)}\}(x, y) \tag{26}$$

where MV is the majority voting operator. MV can perfectly combine the global and local spatial information of HSI, providing a more representative scale map $\mathbf{M}$. The fusion of various multiscale superpixel maps obtains different visual content and further improves the performance.

D. Multiple Kernel Fusion

Since the distribution of various land covers is unknown in advance, the windows in changeable size may mix with anomalous data in practical situations. Fortunately, the integrated scale map $\mathbf{M}$ controls the selection of convolution kernel size and produces the optimized neighborhood window centered on each pixel. Thus, an adaptive filter is created, and the regularized feature $\mathbf{H}^k$ is defined as

$$\mathbf{H}^k(x, y) = \frac{\sum_{y'=-r_{xy}}^{r_{xy}} \sum_{x'=-r_{xy}}^{r_{xy}} \hat{\upsilon}(x', y')\mathbf{F}^k(x+x', y+y')}{\sum_{y'=-r_{xy}}^{r_{xy}} \sum_{x'=-r_{xy}}^{r_{xy}} \hat{\upsilon}(x', y')}$$
$$\hat{\upsilon}(x', y') = \frac{1}{2\pi\sigma^2} e^{-\frac{x'^2+y'^2}{2\sigma^2}} \tag{27}$$

where $\hat{\upsilon}(x, y)$ is the 2-D Gaussian function and r_{xy} equals to $((\mathbf{M}(x, y) - 1)/2)$ that is the convolution kernel radius size at (x, y) in $\mathbf{M}$. $\mathbf{F}^k$ is the initial confidence cube obtained in Gabor phase feature coding, and $\mathbf{H}^k(x, y)$ is the regularized feature cube obtained after Gaussian smoothing. After the regularizing step, the consistency of data distribution can be well preserved.

With the diversity and complexity constraints of actual HSI data, it is difficult for a single kernel function to perfectly solve the classification problems for different datasets. Therefore, the effective representation of heterogeneous features is designed to combine the multiple kernel features and exploit the discriminative structural information of samples. This article proposes a strategy of MKL to fully utilize the intrinsic features embedded in K sets of features $\mathbf{H}^k$. Then, $\mathbf{H}^k$ is adopted to the RF classifier to produce the confidence cube $\mathbf{Z}^k \in \mathbb{R}^{X\times Y\times C}$. $\mathbf{Z}^k_c(x, y)$ represents the probability of pixel (x, y) in the kth group belonging to the class c. We determine the label of each pixel through the probabilistic prediction of

Algorithm 1 *Su*VGF for HSI Classification

```
INPUT: raw HSI R ∈ ℝ^{X×Y×B} with C classes;
OUTPUT: The predicted classification map of all pixels
    Class ∈ ℝ^{X×Y};
Begin
adopt PCA to R to create R';
utilize multiscale ERS and SNIC on R' to obtain the
    superpixel maps W^{i'};
use (24), (25), and (26) to generate the scale map M;
for k = 1 to K do
  for j = 1 to 5 do
    for i = 1 to 4 do
      use (11) and (13) for (12) to create the variable 3-D
      Gabor features G^k_{i,j};
    end for
  end for
  use (15) and (16) to encode the variable 3-D Gabor
  features G^k and obtain the 3-D Gabor phase features Q^k;
  use (17)−(20) to obtain the upper and lower layer LTP
  features, and combine them to get the total features Q̂^k;
  obtain the confidence cube F^K via RF classifier applied
  to the total LTP features Q̂^k;
  use (27) to obtain the regularized feature cubes H^K,
  and then get the regularized confidence cubes Z^k via RF
  classifier;
end for
use (28) to obtain the predicted classification map Class;
End
```

Fig. 4. Ground-truth map of Indiana HSI (16 land cover classes).

TABLE I
LAND COVER CLASSES WITH THE NUMBER OF SAMPLES FOR INDIANA

Class	Land cover Type	No. of Samples
$C1$	Stone-steel-towers	95
$C2$	Hay-windrowed	489
$C3$	Corn-min till	834
$C4$	Soybean-no till	968
$C5$	Alfalfa	54
$C6$	Soybean-clean till	614
$C7$	Grass-pasture	497
$C8$	Woods	1,294
$C9$	Building-grass-tree-drives	380
$C10$	Grass-pasture-mowed	26
$C11$	Corn	234
$C12$	Oats	20
$C13$	Corn-no till	1,434
$C14$	Soybean-min till	2,468
$C15$	Grass-trees	747
$C16$	Wheat	212
	Total	10,366

features, expressed as

$$\mathbf{Class}(x, y) = \arg\max_{c=1,\ldots,C} \sum_{k=1}^{K} \mathbf{Z}_c^k(x, y). \tag{28}$$

The generalization ability and robustness of the *Su*VGF method are enhanced, and the classification performance is improved with comprehensive spectral–spatial feature extraction.

IV. EXPERIMENTAL SETUP

In this section, first, three real HSI datasets imaged in various places are introduced to demonstrate the superiority of the *Su*VGF method. Second, the contribution of variable 3-D Gabor wavelets is performed on the classification. Moreover, the analysis of kernel parameter K and threshold parameter T is achieved. Finally, ablation experiments are accomplished to validate the effectiveness of modules in *Su*VGF.

A. Hyperspectral Dataset Description

1) Indiana Pines Dataset: The first dataset used for testing is Indiana Pines HSI, which was collected by the Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor in Northwest Indiana, USA. The wavelength ranges from 0.4 to 2.6 μm, covering 224 spectral bands. Moreover, only 185 bands are retained after removing the zero bands and noise bands. The image provides a spatial resolution of 20 m per pixel, containing 145×145 pixels. All the 10 249 labeled samples are classified into 16 classes, and the specific distribution of labeled samples is shown in Fig. 4. Table I gives a description of the numbers of labeled samples corresponding to different classes.

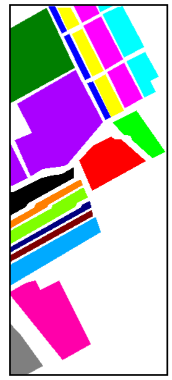


Fig. 5. Ground-truth map of Salinas HSI (16 land cover classes).

TABLE II
LAND COVER CLASSES WITH THE NUMBER OF SAMPLES FOR SALINAS

Class	Land cover Type	No. of Samples
$C1$	Brocoli green weeds 1	2,009
$C2$	Brocoli green weeds 2	3,726
$C3$	Fallow	1,976
$C4$	Fallow rough plow	1,394
$C5$	Fallow smooth	2,678
$C6$	Stubble	3,959
$C7$	Celery	3,579
$C8$	Grapes untrained	11,271
$C9$	Soil vinyard develop	6,203
$C10$	Corn senesced green weeds	3,278
$C11$	Lettuce romaine 4wk	1,068
$C12$	Lettuce romaine 5wk	1,927
$C13$	Lettuce romaine 6wk	916
$C14$	Lettuce romaine 7wk	1,070
$C15$	Vinyard untrained	7,268
$C16$	Vinyard vertical trellis	1,807
	Total	54,129

2) Salinas Dataset: The second real-world dataset is Salinas HSI, which is captured by the AVIRIS sensor over the Salinas Valley, California, USA. The image consists of 224 bands by 512×217 pixels with the spatial resolution of 3.7 m per pixel, and 204 bands are retained after removing noise bands. The dataset contains a total of 54 129 labeled samples, which can be classified into 16 classes. The distribution of different land cover classes is displayed in Fig. 5, and the numerical information for corresponding categories is shown in Table II.

3) Trento Dataset: The third dataset is Trento HSI, gathered in Trento, Italy. The image is filled with 166×600 pixels, which achieves a high spatial resolution of 1 m per pixel. There are 30 414 labeled samples that can be divided into six different classes, which occupies 30% of the total samples. In addition, the spectral dimension is 63 bands ranging from 0.4 to 0.98 μm. The detailed distribution and numerical information are presented in Fig. 6 and Table III, respectively.

B. Ablation Experiment

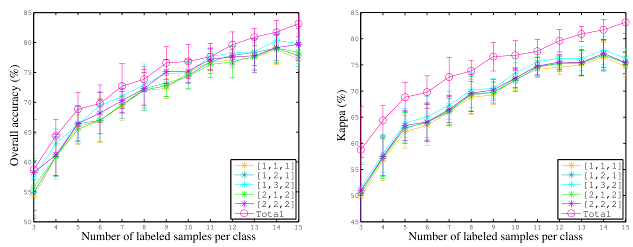
There are several standard indicators to quantify the performance of HSI classification, including overall accuracy (OA), class accuracy (CA), and kappa coefficient (κ) [63].

Fig. 6. Ground-truth map of Trento HSI (six land cover classes).

TABLE III

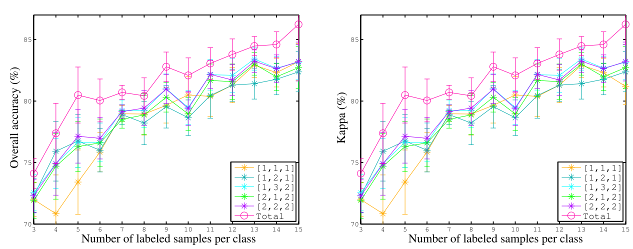
LAND COVER CLASSES WITH THE NUMBER OF SAMPLES FOR TRENTO

Class	Land cover Type	No. of Samples
$C1$	Apple trees	4034
$C2$	Buildings	2903
$C3$	Ground	479
$C4$	Wood	9123
$C5$	Vineyard	10501
$C6$	Roads	3374
	Total	30,414



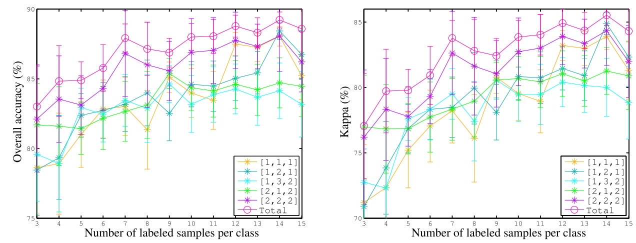
(a) (b)

Fig. 7. Indiana dataset: (a) OA and (b) κ of five kinds of sine waves and VG wavelets, respectively.



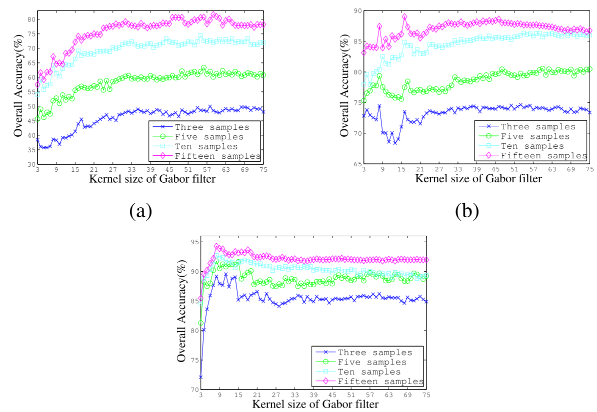
(a) (b)

Fig. 8. Salinas dataset: (a) OA and (b) κ of five kinds of sine waves and VG wavelets, respectively.



(a) (b)

Fig. 9. Trento dataset: (a) OA and (b) κ of five kinds of sine waves and VG wavelets, respectively.



(a) (b)

(c)

Fig. 10. OA versus the threshold parameter T for Gabor kernel size on (a) Indiana, (b) Salinas, and (c) Trento using different numbers of training samples per class.

Concretely, the obtained features are more identifiable when the values of the three indicators are higher. All experimental results in terms of OA, CA, and κ are the average of training for ten times to avoid the contingency. As mentioned above, we focus on solving the problems of a small sample set so that three to 15 samples per class are randomly selected to form the training set, and the rest samples are used as the testing set for each experiment.

1) Integration Necessity of Five Sine Waves: We design five kinds of sine waves for 3-D Gabor wavelets to explore the effect of various Gabor filter structures, and the comparison results are shown in Figs. 7–9, respectively. As we can see, different types of sinusoidal waves contribute unequally to the classification among three datasets, and the accuracy raises generally as the number of labeled samples increases. Concretely, the sinusoids with parameters (1, 3, 2) make the most considerable contribution to the Indiana HSI classification, but it contributes relatively little to the Salinas dataset. Generally, the sinusoids with parameters (1, 3, 2) and (2, 2, 2) obtain a little better performance than the other single sine waves. Furthermore, the total integration of five sine waves achieves the best accuracy and consistency for classification due to the comprehensive description and flexible utilization. Considering different types of sine waves show distinct performance on each dataset, it is necessary to take the five kinds of sinusoids into a unified model to improve the generalization and stability.

2) Analysis for Threshold Parameter T: The Gabor filters use different sizes of convolution kernels to enhance the acquired feature distinguishability. Fig. 10 shows the classification results achieved by selecting the convolution kernels in various sizes for each HSI dataset, using different numbers of training samples, respectively. It can be seen that the optimal convolution kernel size of a single Gabor is not the maximum value, and the trend of OA versus the kernel size is not the same on the three datasets. Specifically, the OA gradually rises with the increasing kernel size of the Gabor filter on Indiana, and it rises in fluctuation with the increasing kernel size on Salinas. In contrast, the OA ascents rapidly and descents slightly during the smaller kernel sizes on Trento. Although the optimal kernel parameter with the best performance is different for multiple sample sizes on three HSIs, the accuracy similarly tends to be stable when the kernel size is larger than the medium to high value. Therefore, comprehensively considering the overall performance, computational efficiency, and model applicability, the threshold parameter T is set to 53 to maintain stable accuracy and appropriate complexity.

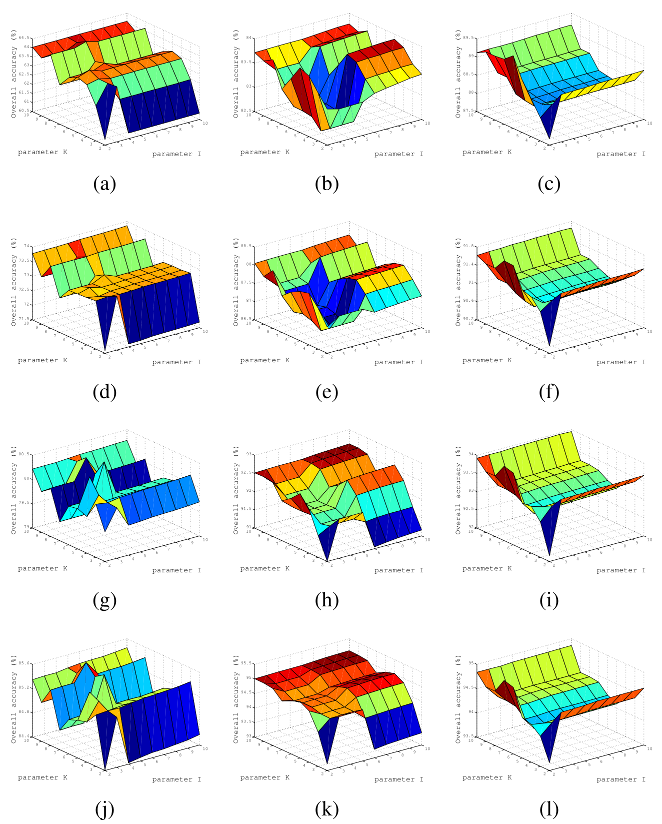


Fig. 11. OA versus the kernel parameter K for variable 3-D Gabor filers and the initial scale I for superpixel maps on Indiana using (a) three, (d) five, (g) ten, and (j) 15 training samples per class, on Salinas using (b) three, (e) five, (h) ten, and (k) 15 training samples per class, and on Trento using (c) three, (f) five, (i) ten, and (l) 15 training samples per class.

3) Analysis for Kernel Parameter K and Scale Parameter I: The fused features of five kinds of sinusoids retain the detailed content of multiscale kernel Gabors, and we choose K groups of variable 3-D Gabor wavelets for feature extraction, where K controls the selection of kernel parameters. On the one hand, selecting too many groups would result in massive redundancy, increasing the computational complexity and reducing the classification accuracy. On the other hand, selecting too few groups cannot complement each other effectively through the MKL. Thus, the selection of kernel parameter K directly influences the final classification results.

Besides, there is an assumption that adjacent pixels belong to the same class rather than separate ones. Thus, the spatial filter achieves per-pixel likelihoods based on neighborhood, and the Gaussian smoothing is applied to reduce local variations. The traditional Gaussian filters use the fixed convolution kernel size, which does not adequately guarantee the homogeneity of pixels in the neighborhood. Meanwhile, multiscale superpixel segmentation methods are excellent at extracting the abundant spatial structure features, and the initial scale I affects the number and quality of superpixels. The superpixel-based scale map guides the kernel selection for Gaussian filters to minimize the local fluctuations of per-pixel likelihoods. Therefore, the selection of the initial scale I measures the homogeneity of regions and influences the effect of Gaussian filters.

As shown in Fig. 11, we analyze the kernel parameter K and scale parameter I on classification results using different numbers of training samples, and the value range of K and I is [2, 10]. The larger K means that more Gabor kernel parameters are selected, and the larger I represents that more pixels are included in superpixels. It is shown that the OA generally rises with the increment of parameter K, demonstrating the fusion effectiveness of Gabor filters with multiple kernels. There are some declining fluctuations during the intermediate K values on Indiana and Salinas datasets, and during the lower K values on the Trento dataset. Moreover, the OA is shown to be more stable with the larger values of parameter I, considering the integration improvement of various multiscale superpixel segmentation. The OA starts to be stable from the intermediate I values on Indiana and Salinas datasets, and from the lower I values on the Trento dataset. Although it is illustrated that the HSIs with high and low spatial resolutions have different local changes on parameter analysis, they also have a common stable tendency. Therefore, in order to balance the feature redundancy, computation complexity, and classification performance, K and I are specifically set to 4 and 7, respectively, ensuring the robustness of *Su*VGF.

4) Module Effectiveness: We conduct several ablation methods on three datasets separately to demonstrate the effectiveness of modules in *Su*VGF. All ablation experiments are classified by the RF classifier, differing in the feature extraction strategies. First, the *Su*RAW method utilizes a superpixel map to regularize the RF classification result on the original HSI data to demonstrate superpixel guidance without Gabor filters. Second, the Gabor method uses traditional 3-D Gabor wavelets with 52 filters to obtain magnitude features, which are stacked along the spectral direction and input into the classifier. Third, the FCG method applies the LTP coding to phase features achieved by traditional 3-D Gabor wavelets, and only the filters parallel to the spectral axis are used. This comparison is designed to demonstrate the characterization capability of phase features versus magnitude features. Fourth, the VG method uses the 3-D VG wavelets and introduces five types of response intensity parameters to obtain phase features, with the parallel orientation to the spectral axis. It is set to further confirm the effectiveness of Gabor phase features with different sinusoids. Fifth, the *Su*VG method adopts the superpixel-based integrated scale map to guide the Gaussian smoothing on VG features to reflect the smoothing effectiveness based on multiscale superpixel segmentation and spatial structure extraction. Finally, our *Su*VGF method fuses all the variable 3-D Gabor features of multiple kernels to gain the comprehensive representation and final classification, which exhibits the complementarity of features and the need for MKL.

The comparison results of six ablation methods on three datasets are listed in Tables IV–VI, where only three labeled samples per class were selected randomly for training. The *Su*RAW method shows the worst classification performance on Indiana and Trento datasets, and the Gabor method shows the worst on the Salinas dataset. The raw HSI information on Salinas is more distinct between categories for land cover recognition than other datasets. It is illustrated that traditional

TABLE IV

CLASSIFICATION ACCURACY (%) AND KAPPA OF MODULE EFFECTIVENESS ON THE INDIANA DATASET WITH THREE LABELED SAMPLES PER CLASS AS THE TRAINING SET

Class	*Su*RAW	Gabor	FCG	VG	*Su*VG	*Su*VGF
C1	86.84	97.89	**100.00**	**100.00**	98.00	98.63
C2	43.50	74.56	90.59	92.23	91.62	**99.71**
C3	22.48	34.98	18.71	32.37	35.06	**37.22**
C4	37.64	36.83	50.41	47.00	47.68	**52.92**
C5	71.11	**98.52**	61.11	92.59	75.74	93.89
C6	19.87	41.76	12.05	36.97	34.30	**50.13**
C7	54.95	34.79	57.34	62.58	63.34	**68.21**
C8	73.71	**84.68**	77.13	78.90	75.89	84.17
C9	20.05	58.89	59.21	31.58	69.00	**71.89**
C10	88.46	99.62	**100.00**	**100.00**	94.23	**100.00**
C11	30.94	75.68	55.98	66.67	59.96	**94.40**
C12	87.00	96.50	95.00	**100.00**	97.00	99.50
C13	25.01	45.13	58.23	**60.25**	42.36	48.90
C14	35.17	32.14	39.83	33.83	45.85	**55.11**
C15	50.67	65.94	72.96	**79.38**	78.94	79.10
C16	90.47	94.10	71.23	**98.11**	94.48	96.70
OA	40.71	50.77	52.93	55.01	56.13	**63.55**
κ	0.34	0.45	0.47	0.50	0.51	**0.59**

TABLE V

CLASSIFICATION ACCURACY (%) AND KAPPA OF MODULE EFFECTIVENESS ON THE SALINAS DATASET WITH THREE LABELED SAMPLES PER CLASS AS THE TRAINING SET

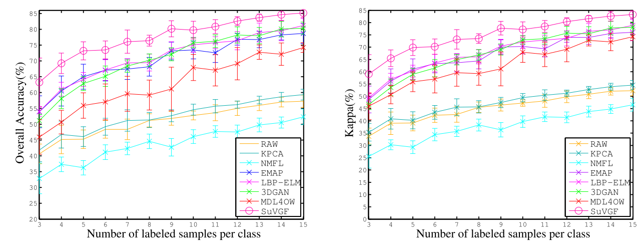
Class	*Su*RAW	Gabor	FCG	VG	*Su*VG	*Su*VGF
C1	97.92	80.86	99.00	93.19	94.64	**99.59**
C2	98.42	83.21	67.36	77.88	78.26	**99.79**
C3	67.16	68.11	85.43	78.62	78.74	**99.99**
C4	97.94	91.09	97.85	98.03	**98.41**	97.68
C5	**90.24**	71.69	86.45	83.98	83.97	81.04
C6	97.18	94.00	98.46	99.25	99.25	**99.80**
C7	98.88	68.85	99.92	**99.93**	99.92	98.16
C8	**48.61**	34.67	31.81	43.36	43.91	46.25
C9	91.63	94.13	77.54	90.40	90.54	**99.89**
C10	65.07	49.43	34.66	31.25	32.23	**83.38**
C11	78.63	85.01	74.16	69.20	70.06	**95.91**
C12	95.20	84.61	98.18	88.61	89.34	**98.94**
C13	**95.91**	89.05	87.77	88.60	91.54	95.69
C14	90.62	91.22	**99.25**	98.92	99.06	98.64
C15	57.89	66.07	58.96	58.22	59.51	**81.78**
C16	79.42	75.38	81.29	77.67	78.47	**97.14**
OA	76.84	67.86	68.68	71.40	72.48	**83.86**
κ	0.74	0.65	0.66	0.68	0.70	**0.82**

TABLE VI

CLASSIFICATION ACCURACY (%) AND KAPPA OF MODULE EFFECTIVENESS ON THE TRENTO DATASET WITH THREE LABELED SAMPLES PER CLASS AS THE TRAINING SET

Class	*Su*RAW	Gabor	FCG	VG	*Su*VG	*Su*VGF
C1	79.84	**81.72**	47.58	41.08	50.71	63.56
C2	63.44	73.31	70.39	58.04	71.52	**89.48**
C3	**96.16**	72.68	66.94	33.19	69.60	56.20
C4	92.57	87.73	90.29	95.98	92.83	**99.59**
C5	54.24	86.53	94.59	96.72	97.70	**99.92**
C6	49.86	51.36	55.60	**73.94**	59.77	51.41
OA	70.32	79.61	80.16	81.95	83.01	**88.17**
κ	0.62	0.77	0.74	0.76	0.77	**0.84**

3-D Gabor wavelets enhance the feature representation but inevitably exacerbate the Hughes phenomenon due to a large amount of feature redundancy. Hence, we only use the 3-D Gabor filters parallel to the spectral axis in the *Su*VGF method to reduce redundant information and retain significant features. The FCG method achieves higher accuracy than the Gabor method, which indicates that the phase features become more identifiable after 1 LGTP feature coding. Moreover, the VG method can capture a richer diversity of features with multiple response intensity parameters for higher accuracy, thus supporting the need to introduce a response strength parameter. Besides, we reduce one parameter to decrease the algorithm complexity. The accuracy of *Su*VG is furthermore enhanced compared to the previous methods, indicating that superpixel segmentation can effectively extract the spatial structure information and guide the adaptive Gaussian filter to suppress noise significantly. As expected, our *Su*VGF method yields the best results with the highest accuracy and consistency, considering it fuses Gabor features with multiple kernels and integrates multiscale spatial structure information, to enhance the feature discriminability.



(a) (b)

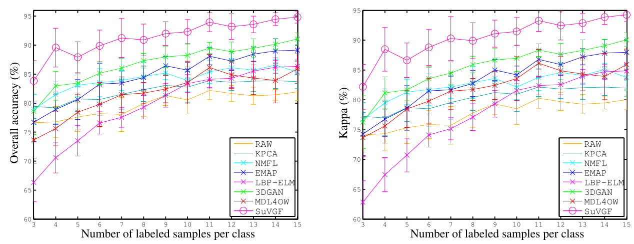
Fig. 12. Indiana HSI: (a) OA and (b) Kappa as functions of the number of labeled samples per class.

V. EXPERIMENTAL RESULT

In this section, comparison experiments are carried out and analyzed to demonstrate the superiority of the *Su*VGF method. Specifically, the raw hyperspectral image (RAW) method implements land cover classification directly using spectral values and RF classifier, without further feature extraction. The KPCA method uses a typical DR approach of kernel principal component extraction to transform the spectral features for HSI classification. The other three common feature extraction methods for comparison are nonlinear multiple feature learning-based classification (NMFL) [64], EMAP [65], and LBP with extreme learning machine (LBP-ELM) [22]. In addition, the 3-D generative adversarial network (3-DGAN) [16] and multitask deep learning in the open world (MDL4OW) [66] are well-designed deep network methods under few-shot conditions. Among all comparison methods, the RAW, KPCA, and EMAP make use of RF classifiers, and the others employ their own classifiers. We analyze the performance of experiments using three to 15 training samples per class in terms of OA, CA, and κ, with cross-validations for ten times. First, the curves of OA and kappa are displayed to analyze the performance trends with increasing training samples. Second, the evaluation results for three datasets with three labeled samples per class as training set are listed in tables on average. Third, the classification result maps are drawn for qualitative analysis and visual comparison.

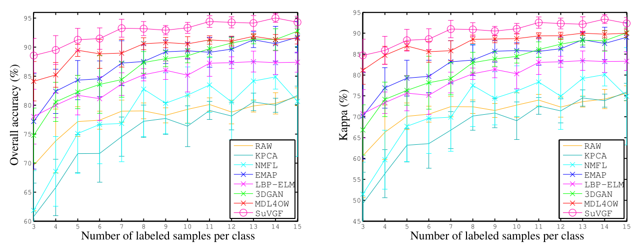
A. Classification Results

The OA and κ of Indiana, Salinas, and Trento datasets are shown in Figs. 12–14, respectively. All figures depict the



(a) (b)

Fig. 13. Salinas HSI: (a) OA and (b) Kappa as functions of the number of labeled samples per class.



(a) (b)

Fig. 14. Trento HSI: (a) OA and (b) Kappa as functions of the number of labeled samples per class.

general rising trends for all methods with increasing training samples, as more samples allow better feature extraction. The OA of the Trento dataset is higher than that of the Salinas and Indiana datasets because the Salinas and Indiana datasets possess complex spatial distribution and confusing classes with similar spectral features. In contrast, the Trento dataset has more distinguishable categories in a much simpler distribution, which is easy for classification. In general, the performance of each method is gradually improved as the number of samples increases, but not equally.

Specifically, the performance of RAW, KPCA, NMFL, and LBP-ELM methods is relatively poor. The RAW method has weak feature discriminability as no feature extraction processing was done to HSI. The KPCA method only carries out the DR from a spectral perspective, and a large amount of valid information may be lost, resulting in low classification accuracy. The NMFL method simply stacks the multiple linear and nonlinear features without fusion strategies, which lacks feature mining and complementary fusion, and the LBP-ELM method does a similar operation after DR. However, there is a dramatic improvement in the classification performance of NMFL and LBP-ELM with an increasing number of samples on Trento and Salinas datasets, respectively, showing that they are more sensitive to the sample set size. In addition, the NMFL method has the lowest accuracy on the Indiana dataset but presents good classification performance on the Salinas dataset, considering it is not a stable method for HSIs under various conditions. In contrast, the EMAP method shows a more steady growth trend than NMFL and LBP-ELM methods on different datasets since the effective single feature expression is better than the direct combination of confusing features without a proper strategy. Moreover, the MDL4OW method accomplishes the classification and reconstruction simultaneously in a multitask framework to obtain a better learning effect, and the 3-DGAN method makes full use of the 1-D hyperspectral profile and 3-D spectral–spatial information to enhance feature extraction. Both MDL4OW and 3-DGAN show stable classification performance, but deep neural networks still suffer from the problems of the small sample set and overfitting due to a large number of learnable parameters in the model and the requirement of abundant training samples. As we expect, the *Su*VGF method consistently shows the best classification performance with an increasing number of samples, especially using only three training samples per class. It demonstrates that the fusion strategy of multiple kernels is conductive to comprehensive feature representation, and superpixel guidance contributes to reducing noise interference and raising accuracy.

B. Quantitative Analysis

We present the CA, OA, and κ in Tables VII–IX with only three samples per class, which is used to analyze the method performance in addressing small sample problem. For the Indiana dataset, as Table VII reveals, C3 (corn-min till), C4 (soybean-no till), C6 (soybean-clean till), C13 (corn-no till), and C14 (soybean-min till) are particularly difficult to classify because they are corn and soybean in different stages of cultivation with similar plant characteristics. Despite this, *Su*VGF achieves the best classification results for C13 and C14, and has stable performance for C3, C4, and C6. C11 (Corn) presents a small patch with fewer labeled samples, which requires the extraction and discrimination ability of the spatial structure and object relationship for the algorithm. The differences exhibited by methods are particularly large on C11, and NMFL and *Su*VGF obtain the worst and best classification accuracies, respectively. Despite the uneven distribution of samples in Indiana and weak separability of vegetation categories, *Su*VGF achieves 100% or near 100% accuracy on the C2 (hay-windrowed), C10 (grass/pasture-mowed), and C12 (oats) classes.

For the Salinas dataset, C1 (brocoli green weeds) and C2 (brocoli green weeds 2) are weeds, and C11 (lettuce romaine 4wk), C12 (lettuce romaine 5wk), C13 (lettuce romaine 6wk), and C14 (lettuce romaine 7wk) are lettuces in various periods. These categories are relatively distinguishable due to the fast growth and obvious feature changes of weeds and lettuces, and most methods have correspondingly good recognition, except the MDL4OW on C2. Moreover, C3 (Fallow), C4 (Fallow rough plow), and C5 (Fallow smooth) are croplands in different states, and they are also relatively easy to classify due to the good regularity of distribution. In contrast, C8 (Grapes untrained) and C15 (Vinyard untrained) are similar categories with confusing properties, which are hard to identify. Consequently, *Su*VGF achieves over 95% accuracy for most classes as Table VIII displays due to the VG feature fusion and multiscale spatial structure description.

For the Trento dataset, C4 (Wood) and C5 (Vineyard) are different categories with distinguishable features, and the *Su*VGF method achieves the best CA close to 100%. C2 (Buildings) and C6 (Roads) are artificial surfaces, and C6 has a relatively lower classification accuracy in all methods due to

TABLE VII

CLASSIFICATION ACCURACY (%) AND KAPPA MEASURE FOR INDIANA ON THE TEST SET WITH THREE LABELED SAMPLES PER CLASS AS THE TRAINING SET

Class	RAW	KPCA	EMAP	LBP-ELM	NMFL	3DGAN	MDL4OW	*Su*VGF
C1	87.05	87.37	89.79	95.54	57.72	90.53	92.93	**98.63**
C2	44.60	53.01	93.15	89.96	25.21	88.90	59.24	**99.71**
C3	21.01	28.69	**40.44**	38.76	25.96	39.83	28.11	37.22
C4	36.35	38.76	44.93	**53.87**	24.30	50.40	39.63	52.92
C5	71.85	68.70	79.63	**98.04**	25.88	97.41	83.53	93.89
C6	19.69	22.28	41.87	**55.45**	18.97	40.34	32.11	50.13
C7	54.79	58.83	62.80	52.27	48.50	47.99	57.67	**68.21**
C8	73.45	61.45	79.84	80.37	37.58	75.18	67.53	**84.17**
C9	20.16	19.87	61.29	65.38	33.87	70.58	51.46	**71.89**
C10	88.46	91.92	93.85	99.57	72.61	99.62	92.17	**100.00**
C11	31.32	39.66	61.28	84.81	20.17	79.91	46.49	**94.40**
C12	87.50	83.00	**100.00**	98.24	58.24	98.50	97.65	99.50
C13	24.84	23.16	38.87	36.55	20.44	28.88	32.78	**48.90**
C14	35.55	37.09	40.59	37.36	34.09	36.37	33.44	**55.11**
C15	51.57	63.08	59.30	55.28	60.00	61.57	71.94	**79.10**
C16	90.42	89.43	**98.77**	86.22	62.82	83.40	90.48	96.70
OA	40.63	41.82	53.93	53.98	32.87	51.14	45.96	**63.55**
κ	0.34	0.35	0.49	0.49	0.26	0.47	0.40	**0.59**

TABLE VIII

CLASSIFICATION ACCURACY (%) AND KAPPA MEASURE FOR SALINAS ON THE TEST SET WITH THREE LABELED SAMPLES PER CLASS AS THE TRAINING SET

Class	RAW	KPCA	EMAP	LBP-ELM	NMFL	3DGAN	MDL4OW	*Su*VGF
C1	97.32	97.09	98.34	95.18	97.84	77.94	95.01	**99.59**
C2	98.21	90.82	88.07	77.47	97.19	75.08	38.71	**99.79**
C3	66.01	87.61	80.69	70.57	89.65	78.24	61.82	**99.99**
C4	97.82	98.66	98.55	95.99	97.65	**99.29**	95.45	97.68
C5	90.18	93.10	88.88	78.73	**96.81**	96.47	86.53	81.04
C6	97.21	96.75	98.07	66.84	97.25	83.62	99.46	**99.80**
C7	98.86	97.02	89.57	63.85	**99.64**	82.61	98.42	98.16
C8	49.82	53.45	50.87	49.56	46.37	**59.95**	43.33	46.25
C9	90.84	92.68	84.79	58.75	98.25	95.74	91.00	**99.89**
C10	64.26	79.01	63.70	70.61	80.30	**91.53**	72.13	83.38
C11	76.65	84.89	78.11	71.38	93.31	**98.46**	72.41	95.91
C12	95.33	**99.32**	97.01	81.04	90.02	91.00	98.35	98.94
C13	95.02	96.81	89.80	79.13	87.78	96.36	**99.69**	95.69
C14	89.78	89.55	88.03	69.32	79.68	98.13	92.60	**98.64**
C15	56.41	58.33	68.18	64.33	55.98	61.61	71.57	**81.78**
C16	79.95	82.71	72.72	73.50	88.01	82.33	84.14	**97.14**
OA	76.62	79.44	76.69	66.34	78.93	78.60	73.65	**83.86**
κ	0.74	0.77	0.77	0.74	0.63	0.76	0.71	**0.82**

TABLE IX

CLASSIFICATION ACCURACY (%) AND KAPPA MEASURE FOR TRENTO ON THE TEST SET WITH THREE LABELED SAMPLES PER CLASS AS THE TRAINING SET

Class	RAW	KPCA	EMAP	LBP-ELM	NMFL	3DGAN	MDL4OW	*Su*VGF
C1	77.53	72.98	75.94	**89.16**	68.38	53.79	84.26	63.56
C2	62.26	55.28	74.79	45.01	38.05	50.61	82.08	**89.48**
C3	96.43	93.72	**97.35**	56.24	74.66	56.93	90.13	56.20
C4	90.77	77.62	86.97	91.66	64.47	94.78	83.63	**99.59**
C5	55.21	45.84	77.63	86.13	66.01	77.13	93.92	**99.92**
C6	49.34	46.54	48.32	31.20	52.09	48.95	**53.89**	51.41
OA	69.64	60.80	77.19	78.01	61.85	74.78	84.12	**88.17**
κ	0.61	0.49	0.51	0.70	0.71	0.67	0.79	**0.84**

its linearly narrow shape and distribution along with other land covers. C2 is distributed discretely with a blocky appearance, which is more identifiable than C6, and the methods have higher recognition accuracy on C2 than C6. The results of all methods are different, but the *Su*VGF method still achieves the best performance overall.

C. Qualitative Analysis

The classification maps of all methods in a single experiment are shown in Figs. 15–17, which displays visual results for qualitative analysis. Compared to the classification maps of the RAW method, KPCA reduces the fragmentation of C2 and C14 on Indiana, decreases the confusion of C3 on Salinas, and enhances the continuity of C1 on Trento, respectively. It is suggested that the DR method can reduce the intraclass spectral variation and, thus, enhance the feature discrimination. The NMFL method has an obvious salt and pepper effect in the classification maps, which is particularly evident in the distribution of C2 and C8 on the Indiana dataset. LBP-ELM and EMAP methods enhance the regional continuity through spatial feature extraction, but their classification results for pixels close to boundaries are relatively poor. It is indicated

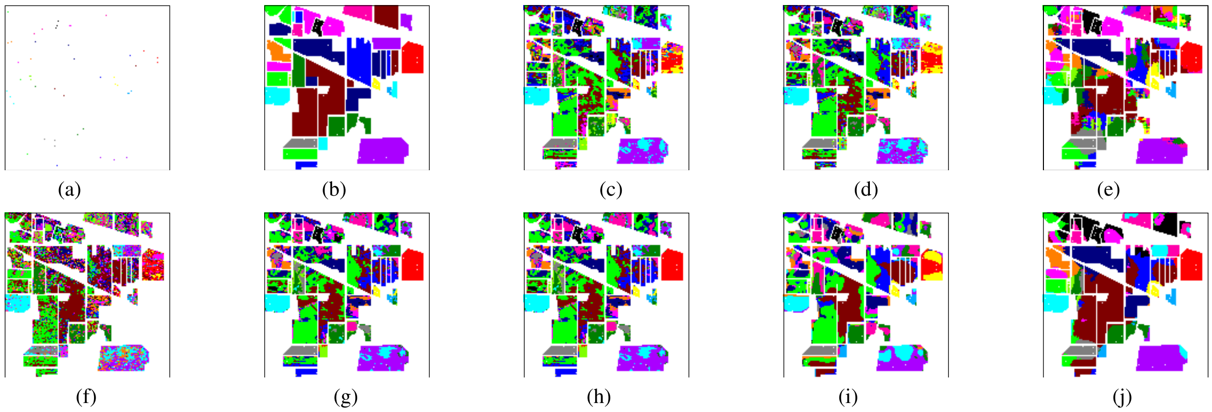


Fig. 15. Indiana classification maps (three labeled samples per class for training). (a) Training set. (b) Testing set. (c) RAW (41.73%). (d) KPCA (38.92%). (e) NMFL (41.70%). (f) EMAP (36.27%). (g) LBP-ELM (54.33%). (h) 3-DGAN (54.10%). (i) MDL4OW (42.06%). (j) *Su*VGF (67.32%).

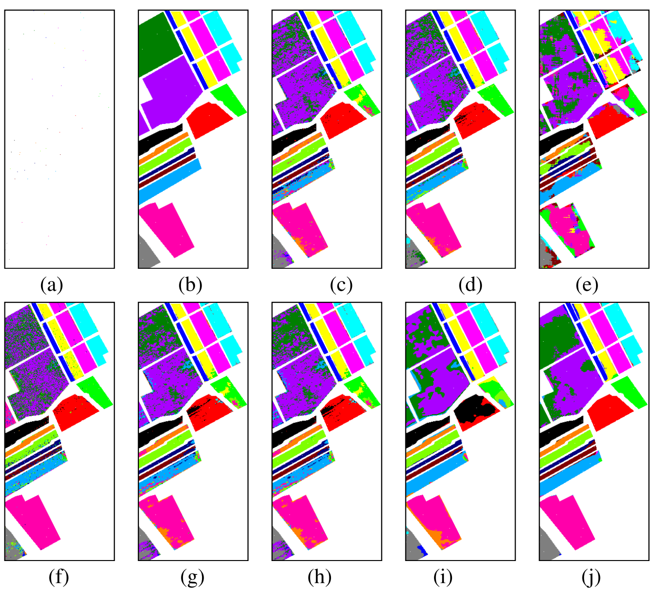


Fig. 16. Salinas classification maps (three labeled samples per class for training). (a) Training set. (b) Testing set. (c) RAW (81.49%). (d) KPCA (82.35%). (e) NMFL (79.46%). (f) EMAP (82.12%). (g) LBP-ELM (66.07%). (h) 3-DGAN (87.07%). (i) MDL4OW (76.01%). (j) *Su*VGF (88.54%).

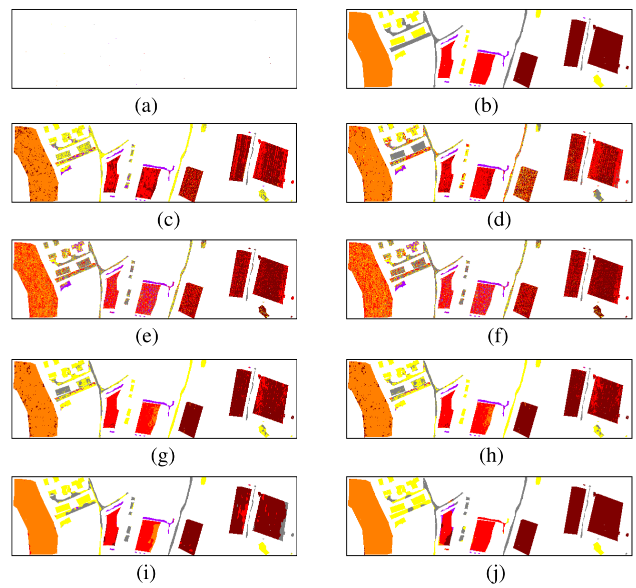


Fig. 17. Trento classification maps (three labeled samples per class for training). (a) Training set. (b) Testing set. (c) RAW (69.38%). (d) KPCA (65.64%). (e) NMFL (60.70%). (f) EMAP (83.47%). (g) LBP-ELM (79.74%). (h) 3-DGAN (79.42%). (i) MDL4OW (89.08%). (j) *Su*VGF (91.46%).

that single feature extraction does not make full use of the abundant information and reduces the identification accuracy. The class boundaries of 3-DGAN and MDL4OW methods remain relatively clear, and the confusion within boundaries is low, considering the contribution of integrated spatial–spectral feature learning. Besides, 3-DGAN and MDL4OW tend to have the mixed classification results of C1, C4, and C5 on the Trento dataset and are not well differentiated for the man-made surface categories (C2 and C6) with similar spectral attributes. In contrast, *Su*VGF achieves the best results of classification maps on three datasets. Compared to other methods, *Su*VGF has the lowest category confusion for C8 in Indiana, obtains the highest regional continuity for C6 and C15 on Salinas, and describes the clearest boundaries for C5 on Trento, respectively. It is demonstrated that the strategy of multiple kernel feature fusion and guidance of multiscale superpixels have advantages of comprehensive feature extraction and accurate land cover classification.

VI. Conclusion

This article proposes a *Su*VGF method for HIS classification to alleviate the small sample size problem. The 3-D Gabor filters parallel to the spectral axis are applied to achieve the phase feature extraction and encoding, decreasing data redundancy and enhancing feature robustness. Moreover, the designed plane sine wave of 3-D Gabor reduces one response intensity parameter, and we can obtain the rich spectral–spatial features through groups of variable 3-D Gabor filters. Meanwhile, we exploit the complementary characteristics of diverse multiscale superpixel segmentation, including ERS and SNIC algorithms, to perform the spatial structure extraction. The Gaussian smoothing guided by a superpixel-based scale map has a good effect on noise suppression. Finally, the fusion of all regularized features significantly improves the generalization ability and classification accuracy of the proposed *Su*VGF method.

The superiority of the *Su*VGF method is demonstrated with the best performance on three real-world HSI datasets, which is motivated by the insufficient exploitation of Gabor filters and inadequate use of spatial information. The *Su*VGF improves the feature expression and classification performance by making full use of the highly informative contents in HSI. The relevant parameter settings are applicable to three datasets in this study, and we believe that these settings are also adaptable to other hyperspectral data without sacrificing the generalization ability. In future work, the attention mechanism and more fusion strategies will be introduced to describe the different contributions of multiple kernels and multiscale superpixels for classification.

References

[1] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. 1, pp. 110–122, Sep. 2009.

[2] 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.

[3] M. A. Lee, Y. Huang, H. Yao, S. J. Thomson, and L. M. Bruce, "Determining the effects of storage on cotton and soybean leaf samples for hyperspectral analysis," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2562–2570, Jun. 2014.

[4] A. Brook and E. B. Dor, "Quantitative detection of settled dust over green canopy using sparse unmixing of airborne hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 884–897, Feb. 2016.

[5] 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.

[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] P. Ghamisi *et al.*, "Advances in hyperspectral image and signal processing: A comprehensive overview of the state of the art," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 37–78, Dec. 2017.

[8] 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.

[9] H. Su, H. Yang, Q. Du, and Y. Sheng, "Semisupervised band clustering for dimensionality reduction of hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 6, pp. 1135–1139, Nov. 2011.

[10] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, Mar. 2009.

[11] J. Zabalza *et al.*, "Novel two-dimensional singular spectrum analysis for effective feature extraction and data classification in hyperspectral imaging," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4418–4433, Aug. 2015.

[12] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for the classification of hyperspectral remote sensing data over urban areas," *EURASIP J. Adv. Signal Process.*, vol. 2009, no. 1, pp. 1–14, Dec. 2009.

[13] Y. Chen, H. Jiang, C. Li, X. Jia, and P. Ghamisi, "Deep feature extraction and classification of hyperspectral images based on convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6232–6251, Oct. 2016.

[14] J. M. Haut, M. E. Paoletti, J. Plaza, J. Li, and A. Plaza, "Active learning with convolutional neural networks for hyperspectral image classification using a new Bayesian approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 11, pp. 6440–6461, Nov. 2018.

[15] X. Yang, Y. Ye, X. Li, R. Y. K. Lau, X. Zhang, and X. Huang, "Hyperspectral image classification with deep learning models," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5408–5423, Sep. 2018.

[16] L. Zhu, Y. Chen, P. Ghamisi, and J. A. Benediktsson, "Generative adversarial networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5046–5063, Sep. 2018.

[17] Y. Zhan, D. Hu, Y. Wang, and X. Yu, "Semisupervised hyperspectral image classification based on generative adversarial networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 2, pp. 212–216, Feb. 2018.

[18] L. Mou, P. Ghamisi, and X. X. Zhu, "Deep recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 7, pp. 3639–3655, Jul. 2016.

[19] H. Wu and S. Prasad, "Convolutional recurrent neural networks for hyperspectral data classification," *Remote Sens.*, vol. 9, no. 3, p. 298, Mar. 2017.

[20] X. Huang, X., Han, L. Zhang, J. Gong, W. Liao, and J. A. Benediktsson, "Generalized differential morphological profiles for remote sensing image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 4, pp. 1736–1751, Apr. 2016.

[21] Y. Y. Tang, Y. Lu, and H. Yuan, "Hyperspectral image classification based on three-dimensional scattering wavelet transform," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2467–2480, May 2015.

[22] 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.

[23] P. Ghamisi, M. D. Mura, and J. A. Benediktsson, "A survey on spectral–spatial classification techniques based on attribute profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2335–2353, May 2015.

[24] P. Ghamisi, J. A. Benediktsson, and M. O. Ulfarsson, "Spectral–spatial classification of hyperspectral images based on hidden Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2565–2574, May 2014.

[25] J. Li, J. Bioucas-Dias, and A. Plaza, "Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Aug. 2012.

[26] 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, Jun. 2018.

[27] W. Fu, S. Li, L. Fang, X. Kang, and J. A. Benediktsson, "Hyperspectral image classification via shape-adaptive joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 556–567, Feb. 2016.

[28] S. Jia, J. Hu, Y. Xie, L. Shen, X. Jia, and Q. Li, "Gabor cube selection based multitask joint sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3174–3187, Jun. 2016.

[29] S. Mei, J. Hou, J. Chen, L.-P. Chau, and Q. Du, "Simultaneous spatial and spectral low-rank representation of hyperspectral images for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 5, pp. 2872–2886, May 2018.

[30] S. Jia, B. Deng, J. Zhu, X. Jia, and Q. Li, "Superpixel-based multitask learning framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 5, pp. 2575–2588, May 2017.

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

[32] P. Ghamisi, J. N. A. Benediktsson, and J. R. Sveinsson, "Automatic spectral–spatial classification framework based on attribute profiles and supervised feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5771–5782, Sep. 2014.

[33] 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.

[34] J. Yang, Y.-Q. Zhao, and J. C.-W. Chan, "Learning and transferring deep joint spectral–spatial features for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4729–4742, Aug. 2017.

[35] X. Xu, W. Li, Q. Ran, Q. Du, L. Gao, and B. Zhang, "Multisource remote sensing data classification based on convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 937–949, Feb. 2018.

[36] 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.

[37] S. Li, T. Lu, L. Fang, X. Jia, and J. A. Benediktsson, "Probabilistic fusion of pixel-level and superpixel-level hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7416–7430, Dec. 2016.

[38] F. Tsai, C. K. Chang, and G. R. Liu, "Texture analysis for three dimensional remote sensing data by 3D GLCM," in *Proc. Asian Conf. Remote Sens. (ACRS)*, 2006, pp. 430–435.

[39] X. Guo, X. Huang, and L. Zhang, "Three-dimensional wavelet texture feature extraction and classification for multi/hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 12, pp. 2183–2187, Dec. 2014.

[40] 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.

[41] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[42] Y.-R. Yeh, T.-C. Lin, Y.-Y. Chung, and Y.-C. F. Wang, "A novel multiple kernel learning framework for heterogeneous feature fusion and variable selection," *IEEE Trans. Multimedia*, vol. 14, no. 3, pp. 563–574, Jun. 2012.

[43] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple-kernel classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3780–3791, Oct. 2010.

[44] W. Li and Q. Du, "Gabor-filtering-based nearest regularized subspace for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1012–1022, Apr. 2014.

[45] L. Shen and S. Jia, "Three-dimensional Gabor wavelets for pixel-based hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 5039–5046, Dec. 2011.

[46] J.-M. Geusebroek, A. W. M. Smeulders, and J. van de Weijer, "Fast anisotropic Gauss filtering," *IEEE Trans. Image Process.*, vol. 12, no. 8, pp. 938–943, Aug. 2003.

[47] W. V. Hecke, A. Leemans, S. D. Backer, B. Jeurissen, P. M. Parizel, and J. Sijbers, "Comparing isotropic and anisotropic smoothing for voxel-based DTI analyses: A simulation study," *Hum. Brain Mapping*, vol. 31, no. 1, pp. 98–114, 2010.

[48] S. Jia, Z. Zhu, L. Shen, and Q. Li, "A two-stage feature selection framework for hyperspectral image classification using few labeled samples," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1023–1035, Apr. 2014.

[49] Z. Zhu, S. Jia, S. He, Y. Sun, Z. Ji, and L. Shen, "Three-dimensional Gabor feature extraction for hyperspectral imagery classification using a memetic framework," *Inf. Sci.*, vol. 298, pp. 274–287, Mar. 2015.

[50] S. Jia, L. Shen, J. Zhu, and Q. Li, "A 3-D Gabor phase-based coding and matching framework for hyperspectral imagery classification," *IEEE Trans. Cybern.*, vol. 48, no. 4, pp. 1176–1188, Apr. 2018.

[51] 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.

[52] 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.

[53] F. Riaz, A. Hassan, S. Rehman, and U. Qamar, "Texture classification using rotation-and scale-invariant Gabor texture features," *IEEE Signal Process. Lett.*, vol. 20, no. 6, pp. 607–610, Apr. 2013.

[54] L. Shen and S. Zheng, "Hyperspectral face recognition using 3D Gabor wavelets," in *Proc. IEEE Conf. Pattern Recognit. (ICPR)*, Nov. 2012, pp. 1574–1577.

[55] L. Shen and L. Bai, "3D Gabor wavelets for evaluating SPM normalization algorithm," *Med. Image Anal.*, vol. 12, no. 3, pp. 375–383, Jun. 2008.

[56] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, vol. 2, Oct. 2003, pp. 10–17.

[57] M.-Y. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 2097–2104.

[58] 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.

[59] F. Meng, X. Wang, F. Shao, D. Wang, and X. Hua, "Energy-efficient Gabor kernels in neural networks with genetic algorithm training method," *Electronics*, vol. 8, no. 1, p. 105, Jan. 2019.

[60] J. G. Daugman, "High confidence visual recognition of persons by a test of statistical independence," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1148–1161, Nov. 1993.

[61] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.

[62] F. Xiong, J. Chen, J. Zhou, and Y. Qian, "Superpixel-based nonnegative tensor factorization for hyperspectral unmixing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 6392–6395.

[63] J. A. Richards and J. Richards, *Remote Sensing Digital Image Analysis*, vol. 3. Berlin, Germany: Springer, 1999.

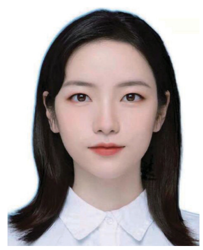
[64] J. Li *et al.*, "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, Mar. 2015.

[65] M. M. Dalla, B. J. Atli, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5975–5991, 2010.

[66] S. Liu, Q. Shi, and L. Zhang, "Few-shot hyperspectral image classification with unknown classes using multitask deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 6, pp. 5085–5102, Jun. 2021.

Shuyu Zhang received the B.E. and Ph.D. degrees from the College of Earth Sciences, Zhejiang University, Hangzhou, China, in 2015 and 2020, respectively.

She is a Post-Doctoral Researcher with the College of Computer Science and Software Engineering, Shenzhen University, Shenzhen, China. Her research interests include hyperspectral image classification and deep learning.



Dingding Tang received the B.E. degree in information technology from Southwest Forestry University, Kunming, 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.

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, 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 National Engineering Research Center for Information Technology in Agriculture, Beijing, China. She is the coauthor of the 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.

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.