Spatial-Spectral Joint Classification of Hyperspectral Image With Locality and Edge Preserving

Hui Zhang, Wanjun Liu, and Huanhuan Lv

Abstract—Hyperspectral image (HSI) classification is an important part of its processing and application. Aiming at the problems of high data dimensionality and high spatial neighborhood correlation in HSI classification, we propose a spatial-spectral joint classification method of HSI with locality and edge preserving in this article. First, the input HSI is normalized, and the feature is extracted by principal component analysis. The first principal component image is taken as the guidance image. Second, guided filtering is used to extract the spatial features of each band separately. Then, the extracted spatial features are superimposed, and low-dimensional embedding is completed through local Fisher discriminant analysis. Finally, the obtained low-dimensional embedded features are input into a random forest classifier to get classification results. The experimental results of two HSI show that the proposed method achieves higher classification accuracy than other related methods. In the case of randomly selecting 10% and 1% samples from each class of ground object as training samples, the overall classification accuracy is improved to 99.57% and 97.79%, respectively. This method effectively uses the spatial and local information of the image in low dimensional embedding, and preserves the boundaries of the ground objects, thus improving the classification effect.

Index Terms—Guided filtering, hyperspectral remote sensing image, low-dimensional embedding, random forest.

I. INTRODUCTION

HYPERSPECTRAL image (HSI) has ultra-high spectral resolution, which can acquire hundreds of continuous spectral bands of the ground objects, thereby greatly improving the ability of distinguishing different ground objects. HSI plays an important role in national defense construction and national economy, and has been widely used in areas such as target detection [1], surface classification [2], environmental management [3], and mineral mapping [4]. The basis of many HSI applications is image classification, but the characteristics of high dimensionality, high correlation between bands, and small

Hui Zhang is with the School of Electronic and Information Engineering and the School of Software, Liaoning Technical University, Huludao 125105, China (e-mail: zhjordan45@126.com).

Wanjun Liu and Huanhuan Lv are with the School of Software, Liaoning Technical University, Huludao 125105, China (e-mail: liuwanjun@lntu.edu.cn; lvhh2010@126.com).

number of samples have brought challenges to classification, which makes classification tasks face problems such as dimensionality disaster, spatial homogeneity, and heterogeneity [5].

To solve these problems, on the one hand, many researchers engaged in HSI classification use methods of machine learning for image classification, including support vector machine (SVM) [6], Gaussian mixture model [7], random forest (RF)[8], sparse expression [9], active learning [10], etc. On the other hand, deep learning has been successfully applied in computer vision and other fields. In recent years, it is gradually expanding to HSI classification. The models include autoencoder network [11], [12], convolutional neural network [13], [14], three-dimensional convolutional neural network [15], [16] and recurrent neural network [17], [18]. The deep learning-based HSI classification method automatically extracts features from the data, which simplifies the processing of classification, but it still faces problems such as high computational complexity and small-sample learning. Therefore, compared with the high computational complexity of deep learning, the machine learning-based HSI classification method focuses on the selection of the classifier and feature extraction, which is still an important breakthrough in improving the accuracy of classification.

Feature extraction can embed high-dimensional data into a lower-dimensional space, and reduce the data dimension while maintaining the basic structural information of the original data as much as possible. Therefore, the curse of dimensionality can also be alleviated to a certain extent. So feature extraction is a common data preprocessing method in HSI classification. Principal component analysis (PCA) [19] and linear discriminant analysis (LDA) [20] are two typical subspace linear transformation approaches, but they cannot effectively reveal the nonlinear structure of the data. For this reason, researchers put forward some manifold learning methods, which can better mine potential low-dimensional manifold structures of high-dimensional data, such as local preserving projection (LPP) [21], locally linear embedding [22] and neighborhood preserving embedding [23]. The above methods can be classified into graph embedding framework. The difference lies in how to build intrinsic graph and penalty graph, but they are all unsupervised methods, and their classification performance is limited [24]. To address this issue, researchers introduce the prior knowledge of the samples into graph embedding framework to improve the performance of classification, and develop supervised learning methods such as marginal fisher analysis [25], local Fisher discriminant analysis (LFDA) [26], and regularized local discriminant embedding [27], so as to further improve the classification accuracy.

The above methods only use the spectral features in feature extraction, and ignore the spatial features of the image. In addition to spectral features, HSI also contains a wealth of spatial features, and the pixels close to each other in space are likely to belong to the same kind of ground objects [28], [29]. Only spectral features cannot fully and effectively express the structures of hyperspectral data. For this reason, researchers put forward HSI classification approaches of spatial–spectral feature fusion, and considered that spatial features can be used as a complementary of spectral features. Hang *et al.* [30] introduced matrix-based discriminant analysis (MDA) to extract the spatial and spectral features of HSI simultaneously. Then, a random sampling technique is used to produce a subspace ensemble for final SVM classification. In [31], the joint of extended morphological profiles (EMP) and spectral features was adopted to represent the features of pixels in HSI. Then, the authors construct a hypergraph to learn the low-dimensional features. Finally, the learned features were fed to SVM for classification. Huang *et al.* [32] first employed a weighted mean filter to filter the image. Then, a spatial-spectral combined distance was used to fuse the spatial and spectral information to select the neighbors of each pixel. Finally, manifold reconstruction was performed and the low-dimensional discriminative features are extracted for classification. The above methods effectively introduce the spatial features of the image, and the classification accuracy has been improved, but they only use the spatial information between the center pixel and its surrounding pixels in a specific area, or each pixel and its neighbors of the low-dimensional embedding process. The edge information of the ground objects in the image is ignored, and this information still plays an important role in improving the accuracy of classification, especially when the number of training sample is small.

To overcome the aforementioned drawbacks, we proposed a spatial-spectral joint classification method of HSI with locality and edge preserving in this article. The spatial-spectral features extracted by the proposed method make use of the spatial information, local information and preserves the boundaries of the ground objects, and perform classification with a random forest classifier, which improves the classification accuracy and reduces the computational complexity. The main characteristics of the proposed method can be concluded as follows.

1) A local linear model between the first principal component of PCA and the output image is established, and the difference function between the input and output images of each band is solved to extract the spatial features while taking into account the edge information of the ground objects.

2) In the low-dimensional embedding, the distance-based similarity weighted matrices are calculated to effectively preserve the local multimodal features of each class. The local between-class and within-class scatter matrices are obtained by the weighted matrices, which preserve neighborhood relationships and force neighboring points in the input space to remain close in the projected subspace, so that the data of different categories can be better separated in the projection space.

The remainder of this article is organized as follows: In Section II, the related works are presented. In Section III, the proposed method is discussed in detail. The experimental analysis and comparisons are presented in Section IV. Finally, the conclusion is drawn.

II. RELATED WORKS

A. Guided Filtering

Guided filtering [33] is an edge preservation filter with excellent performance that has appeared in recent years. In addition to the basic smoothing function, it can also transfer the spatial edge information in the guidance image to the output image more completely. It has been widely used in image denoising [34], image fusion [35], and other fields.

In the process of collection, acquisition and transmission of HSI, different types of noise are often introduced. As a result, the spectral characteristics of the same class of ground objects show different degrees of fluctuation, which brings great difficulties to the accurate classification of HSI under small training samples. To alleviate the problem, we use guided filtering to remove the noise in the original image and preserve the edges of the ground objects. The filter plays a smoothing role in regions where the pixels change softly. In regions where the pixels change sharply, the filter can also maintain the edges of the objects, thereby retaining the distinguishing features that are conducive to classification.

Suppose that the filter output q and the guidance image I have the following local linear relationship in a window δ_k centered at the pixel k and r as the radius

$$q_i = a_k I_i + b_k, \quad \forall i \in \delta_k, q_i \in q \tag{1}$$

where a_k and b_k are some linear coefficients assumed to be constant in δ_k.

To determine the linear coefficients, a solution to minimize the difference between q and p is sought, it can be transformed into an optimization problem according to the method of unconstrained image restoration. The cost function is

$$E(a_k, b_k) = \sum_{i \in \delta_k} \left((a_k I_i + b_k - p_i)^2 + \varepsilon a_k^2 \right) \tag{3}$$

where ε is a regularization parameter.

Least square method is used to get

$$a_k = \frac{\frac{1}{|\delta|} \sum_{i \in \delta_k} I_i p_i - \mu_k \bar{p}_k}{\sigma_k^2 + \varepsilon} \tag{4}$$

$$b_k = \bar{p}_k - a_k \mu_k \tag{5}$$

where μ_k and σ_k^2 are the mean and variance of I in δ_k, respectively. $|\delta|$ is the number of pixels in δ_k, and $\bar{p}_k = \frac{1}{|\delta|} \sum_{i \in \delta_k} p_i$ is the mean of p in δ_k.

When window operation is adopted in the whole image, the output value of a certain pixel can be obtained by means of the

average linear function of the pixel as follows:

$$q_i = \frac{1}{|\delta|} \sum_{k:i\in\delta_k} (a_k I_i + b_k)$$

$$= \bar{a}_i I_i + \bar{b}_i \qquad (6)$$

where $\bar{a}_i = \frac{1}{|\delta|}\sum_{k\in\delta_i} a_k$ and $\bar{b}_i = \frac{1}{|\delta|}\sum_{k\in\delta_i} b_k$ are the average values of a_k and b_k in all windows i.

The regularization coefficient ε and radius r are the two parameters of the filter. The proposed method analyzes and verifies the influence of their different values on the final classification results in detail.

B. Local Fisher Discriminant Analysis

Affected by illumination conditions, atmospheric effects and geometric distortions, the real class-conditional distributions of HSI usually have a complicated multimodal structure. The local multimodality of each class is preserved by using a distance-based similarity weighted matrix. In order to obtain the weighted matrix, the similarity between samples needs to be calculated. $A_{i,j} \in [0,1]$ represents the similarity between sample x_i and x_j, denoted as

$$A_{i,j} = \exp\left(-\frac{||x_i - x_j||^2}{\gamma_i \gamma_j}\right) \qquad (7)$$

where γ_i is the local scaling of x_i, defined as $\gamma_i = ||x_i - x_i^t||$, x_i^t is the neighbor of x_i, and t is the adjustment factor. The local between-class scatter matrix S^{lb} and local within-class scatter matrix S^{lw} are defined as

$$S^{lb} = \frac{1}{2}\sum_{i,j=1}^{n} W_{i,j}^{lb}(xi - xj)(xi - xj)^{T} \qquad (8)$$

and

$$S^{lw} = \frac{1}{2}\sum_{i,j=1}^{n} W_{i,j}^{lw}(xi - xj)(xi - xj)^{T} \qquad (9)$$

where two weighted matrices W^{lb} and W^{lw} are introduced to preserve local information of the data, and their $(i,j)^{th}$ elements $W_{i,j}^{lb}$ and $W_{i,j}^{lw}$ are respectively defined as

$$W_{i,j}^{lb} = \begin{cases} A_{i,j}(1/n - 1/n_l), & \text{if } y_i = y_j = l \\ 1/n, & \text{if } y_i \neq y_j \end{cases} \qquad (10)$$

and

$$W_{i,j}^{lw} = \begin{cases} A_{i,j}/n_l, & \text{if } y_i = y_j = l \\ 0, & \text{if } y_i \neq y_j \end{cases}. \qquad (11)$$

The transformation matrix WLFDA can be obtained by calculating the optimal value of the local Fisher ratio as follows:

$$W_{\text{LFDA}} = \arg\max_{w} \frac{|W^T S^{lb} W|}{|W^T S^{lw} W|}. \qquad (12)$$

The above optimization problem is equivalent to solving $S^{lb}V = \lambda S^{lw}V$, where λ is the generalized eigenvalue, and V represents the eigenvector corresponding to the Fisher discrimination direction. Assuming that the generalized eigenvalues

are $\lambda_1 \geq \lambda_2 \ldots \geq \lambda_s$ in descending order, the corresponding eigenvector $v_1, v_2, \ldots v_s$ forms the optimal projection matrix.

C. Random Forest Classifier

RF is a decision tree ensemble method based on bagging and random subspace. It can effectively solve the problems of high-dimensional data and high feature-to-instance ratio. The better generalization performance and the higher computational efficiency enable it to achieve better results in complex classification problems [36].

RF is composed of hundreds of decision-tree-based models $\{h(X, \theta k), k = 1, 2, \ldots, N\theta\}$, where θk is an independent identically distributed random variable, and $N\theta$ is the number of variables. RF trains k decision trees for k training subset sampled randomly with replacement from the original sample set. The final classification result is decided by the votes of all classification trees [37].

$$H(x) = \arg\max_{y_j} \sum_{i\in[1,2,\ldots,k]} I(h_i(x) = y_j), \quad j = 1, 2, \ldots, C$$
$$(13)$$

where $H(x)$ represents the combined model, $h_i(x), i = 1, 2, \ldots, k$ is the decision tree model generated from k training subsets, and $y_j, j = 1, 2, \ldots, C$ is the output or labels of C classes. $I(\bullet)$ is the combined strategy and defined as follows:

$$I(x) = \begin{cases} 1, & h_i(x) = y_j \\ 0, & h_i(x) \neq y_j \end{cases} \qquad (14)$$

where $h_i(x)$ is the output of the decision tree, and $y_j, j = 1, 2, \ldots, C$ is the label of class i.

III. PROPOSED METHOD

A. Procedure of Proposed Method

The flowchart of the proposed spatial-spectral joint classification method of HSI with locality and edge preserving is shown in Fig. 1. The main steps include the following:

1) The original data are preprocessed by band division and normalization.
2) PCA is performed to get the features of normalized data, and the first principal component is used as the guidance image. Guided filtering is adopted to extract spatial features of each band separately.
3) The extracted spatial features are superimposed and LFDA is applied for low-dimensional embedding.
4) The obtained low-dimensional embedded features are divided into training samples and test samples and input to the RF classifier to get classification results.

B. Preprocessing of Hyperspectral Data

The value recorded by HSI is the radiance of the ground object, not the reflectivity. Therefore, it is necessary to normalize each band of the image. Otherwise the convergence speed will be slow and may not converge to the minimum value. The
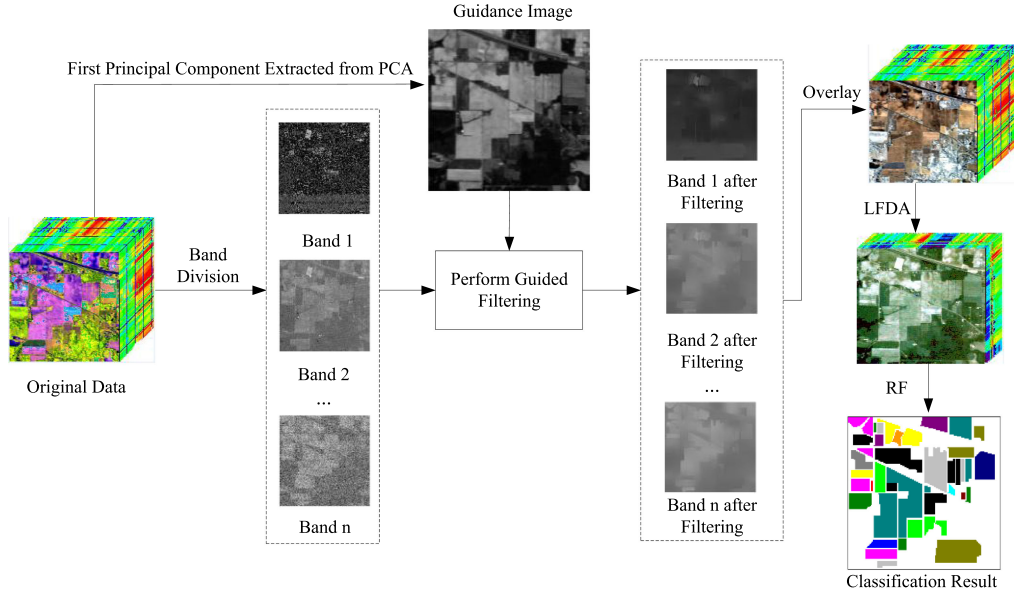
Fig. 1. Flow of the proposed classification method.

normalized calculation formula is as follows:

$$X = \frac{Y - Y_m}{Y_M - Y_m} \quad (15)$$

where Y is the input image and X is the normalized image. Y_m and Y_M represent the minimum and maximum values of each band in the image, respectively. Through the normalization, the spectral values of each band in the image are within the range of [0, 1].

C. Spatial-Spectral Feature Extraction

Let us denote X as an input HSI dataset to be classified with m rows and n columns. It can be expressed as

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{1p} \\ \cdots & \cdots & \cdots & \cdots \\ x_{s1} & x_{s2} & \cdots & x_{sp} \end{bmatrix}$$

where $s = m \times n$ is the number of pixels, and p is the number of bands, and x_{ij} represents the ith sample of the jth band.

We take the first principal component extracted by PCA as the guidance image. Specifically, because the first principal component has the largest eigenvalue, the largest component projection, and the largest amount of information, it contains the most important information of the original image. Most of the boundaries in the original image are retained, so it is used as the guidance image to filter multiple bands.

The covariance matrix of X needed to be calculated to get the guidance image. First, the mean value of all the pixels in the jth band of the image is

$$\bar{x}_j = \frac{1}{s} \sum_{i=1}^{s} x_{ij} \quad (16)$$

where $1 \le i \le s$, $1 \le j \le p$, and x_{ij} is the ith sample of the jth band. Then, the covariance matrix can be obtained by

$$C = \frac{1}{s} \sum_{i=1}^{s} (x_i - \bar{x}_j)(x_i - \bar{x}_j)^T. \quad (17)$$

The eigenvalue decomposition of the covariance matrix is carried out to get the eigenvalue $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_p$ and the corresponding eigenvector is $A = [\alpha_1, \alpha_2, \cdots, \alpha_p]$. So the guidance image can be constructed by

$$I = \alpha_1^T X. \quad (18)$$

Let X be the input image, and I obtained above be the guidance image. Build a local linear model between the output of X and I according to (1), get the values of a_k and b_k from (4), calculate the value of each pixel by (6), and finally get the output.

The third band of the HSI Indian Pines is taken as the input image. Fig. 2 shows a comparison of the features extracted by morphological filtering and guided filtering. Morphological filtering uses the circular window, and guided filtering uses the square window. The first principal component extracted by PCA is taken as the guidance image, and the windows with radius r of 2, 4, 6, and 8 are used for filtering respectively.

Fig. 2 shows that the two filtered images with window radius of 2 and 4 can basically extract the structural features of different classes of the ground objects. When the window radius is 6 and 8, the structure of the ground objects in the image extracted by morphological filtering is relatively fuzzy, while the image extracted by guided filtering can basically retain the edge information and general outlines of the ground objects.

Then, low-dimensional embedding is realized according to (7) to (12). Spatial-spectral features obtained after low-dimensional embedding are used for classification.
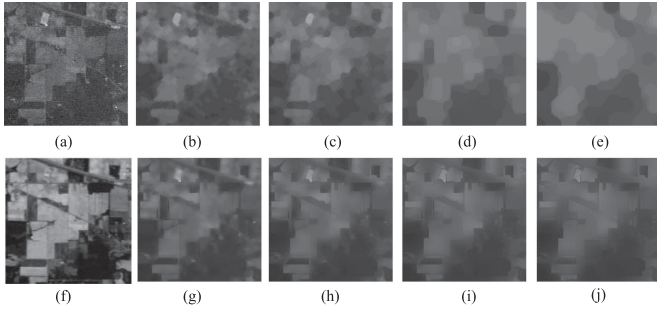
Fig. 2. Comparison of morphological filtering and guided filtering. (a) The input image. (b) The feature obtained by morphological filtering with $r = 2$. (c) The feature obtained by morphological filtering with $r = 4$. (d) The feature obtained by morphological filtering with $r = 6$. (e) The feature obtained by morphological filtering with $r = 8$. (f) The guided image. (g) The feature obtained by guided filtering with $r = 2$. (h) The feature obtained by guided filtering with $r = 4$. (i) The feature obtained by guided filtering with $r = 6$. (j) The feature obtained by guided filtering with $r = 8$.

D. Classification Based on RF

The proposed method uses RF to achieve classification. The specific steps are as follows:

1) The spatial-spectral features obtained are randomly divided into training set and test set.
2) Based on bagging sampling, N samples and M characteristic variables are randomly selected from the training samples T times to construct a new bootstrap sample dataset. T classification regression trees are generated from the new dataset. The training samples out of bag (OOB) are used to estimate the generalization ability of the classifier.
3) The root node of each classification regression tree stores the corresponding sample data. From the root node, a certain characteristic variable is selected according to the principle of minimum impurity, and child nodes are generated by splitting. The impurity index used in this article is Gini coefficient G, the Gini coefficient of characteristic variable g is calculated as

$$G(g) = 1 - \sum_{j=1}^{c} p_j^2 \qquad (19)$$

where c is the number of classes divided by g and p_i is the proportion of the ith class sample to the total sample. Then, suitable splitting points are selected for g, which can maximize the decrease of the Gini coefficient of the node before and after splitting. Assuming that the selected feature variable is f and the splitting point is z, then the decrease before and after splitting is

$$\Delta G(f) = G(g) - (1-q)G(f < z) - qG(f > z) \qquad (20)$$

where $G(f < z)$ represents the Gini coefficient of the samples whose f value of node g is less than z. q is the proportion of samples whose f value of node g is less than z. $G(f > z)$ represents the Gini coefficient of the samples whose f value of node g is greater than z.
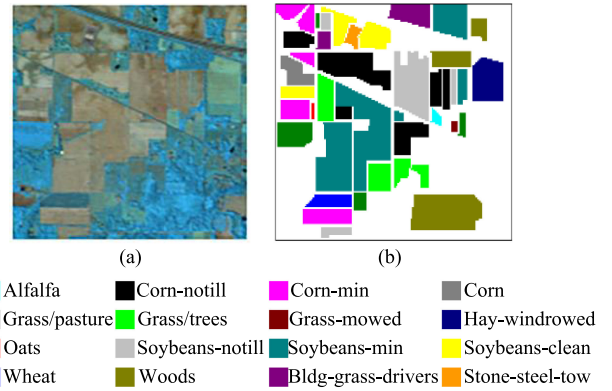

Fig. 3. False color image and the ground truth image of Indian Pines. (a) False colour image. (b) The ground truth image.

4) Recursively selects and splits nodes for each tree without clipping until the maximum decomposition depth is reached. In order to speed up the construction of the trees and reduce the risk of overfitting, an additional sample number limit condition is added to determine whether each node continues to split, that is, when the number of samples of a node is greater than the limit threshold $n_{\text{threshold}}$, the split can continue.
5) The generated T classification regression trees constitute a forest. During the test phase, each tree will vote on the test samples once, and finally assign the class with the most votes to the test samples, so as to realize the random forest classification.

IV. EXPERIMENT AND RESULT ANALYSIS

A. Experimental Data

To verify the validity of the method, representative Indian Pines and Pavia University HSI datasets are selected for classification experiments.

1) Indian Pines dataset: This dataset is derived from the AVIRIS sensor and reflects the vegetation planting in northwestern Indiana, USA. The image size is 145×145 pixels, the wavelength range is 0.4 to 2.5 μm, and the spatial resolution is 20 m. Due to the atmosphere and water absorption, the noise bands are removed and the remaining 200 bands are used for classification. The dataset contains 16 classes of ground objects and with a total of 10 249 samples. False color image of the area (synthetic bands: 140, 80, and 50) is shown in Fig. 3(a), and the ground truth image is shown in Fig. 3(b).
2) Pavia University dataset: The dataset is derived from the ROSIS sensor. The image size is 610×340 pixels, the wavelength range is 0.43 to 8.6 μm, and the spatial resolution is 1. 3m. After removing the noise bands, the remaining 103 bands are used for classification. The dataset contains 9 classes of ground objects and with a total of 42 776 samples. False color image of the area (synthetic bands: 90, 60, and 30) is shown in Fig. 4(a), and the ground truth image is shown in Fig. 4(b).
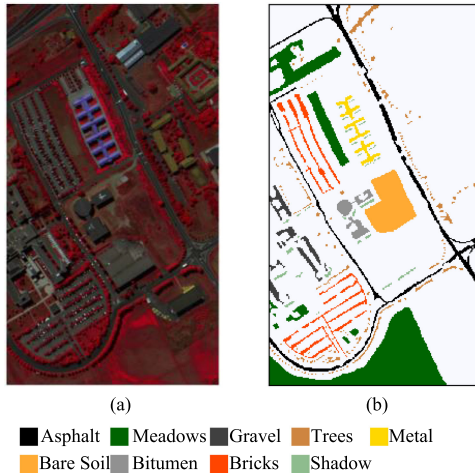
(a)   (b)

Asphalt Meadows Gravel Trees Metal
Bare Soil Bitumen Bricks Shadow

Fig. 4.   False color image and the ground truth image of Pavia University. (a) False colour image. (b) The ground truth image.
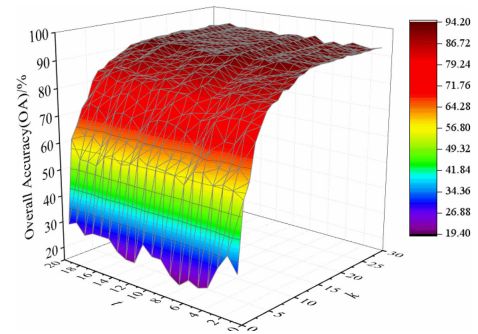

Fig. 5.   OA under different k and t of Indian Pines dataset.
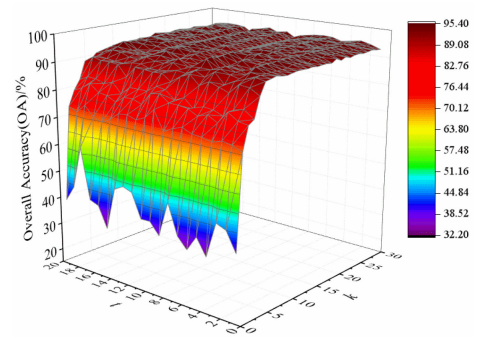
Fig. 6.   OA under different k and t of Pavia University dataset.

B. Comparison Methods and Evaluation Indicators

We use the proposed method (GF-LFDA-RF) and seven other methods to classify the datasets to test the performance. The seven methods are as follows:

1) The method uses SVM to classify raw spectral features (SP-SVM).
2) The method uses RF to classify raw spectral features (SP-RF).
3) The method combines LFDA and SVM (LFDA-SVM) [26].
4) The method combines the EMP and spectral features and adopts the KNN (K-Nearest-Neighbor) method to construct a hypergraph and uses SVM to classify (SSHG*) [31].
5) The matrix-based discriminant subspace ensemble method for HS image spatial–spectral feature fusion (PT + MDA + RS) [30].
6) The method uses guided filtering and performs RF classification (GF-RF) [38].
7) The method called spatial-spectral manifold reconstruction preserving embedding (SSMRPE) for HIS classification [32].

The evaluation indicators used in this article include overall classification accuracy (OA), average classification accuracy (AA) and Kappa coefficient (KC). In order to improve the reliability and accuracy of the experiment, the experiment is repeated ten times under each experimental condition, and the average value of the ten experimental results is taken as the final result.

C. Parameter Settings

The parameters influencing the performance of the proposed method include the embedding dimension k and neighborhood number t of LFDA, the window radius r and regularization coefficient ε of guided filtering, and the number of decision trees T and the limit threshold $n_{\mathrm{threshold}}$ of RF. According to the research results in [36], the number of samples of the split node $n_{\mathrm{threshold}}$ is set to 10. The influences of different parameter values are analyzed through experiments, and then the best parameters are selected. A total of 50 samples of each class of the ground objects are randomly selected from the two datasets (if the total number of the ground objects in a certain class is less than 50, half is selected) as training samples, and the remaining are test samples. Figs. 5 and 6 show the impact of different k and t when $r = 2$, $\varepsilon = 0.1$, and $T = 100$. As can be seen from Fig. 5, for the Indian Pines, when $k = 20$ and $t = 18$, the classification accuracy obtained is the highest, with a value of 94.02%. With the increase of k, the classification accuracy continues to increase and gradually reaches the highest value, and then no longer improves. This is because the more the embedding dimension is, the richer the discrimination information is and the classification accuracy is correspondingly improved. However, when there is enough embedded information for classification, the increase of dimensions will have a limited effect on the improvement. When t is different, the classification accuracy fluctuates, but the difference is not obvious. As shown in Fig. 6, for the Pavia University, when $k = 24$ and $t = 13$, the classification accuracy is the highest, and its value is 95.36%. Figs. 7 and 8 show the effects of different r and ε when the values of $T = 100$, k and t are optimal. As can be seen from Fig. 7, for the Indian Pines, when $r = 7$ and $\varepsilon = 0.0001$, the classification accuracy is the highest, with a value of 98.50%. For Pavia University dataset, when $r = 8$ and $\varepsilon = 0.1$, the classification accuracy is the highest, with a value of 98.04%. With the increase of r, the classification accuracy gradually reaches the maximum value,
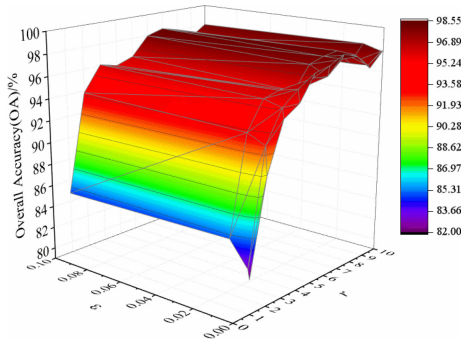
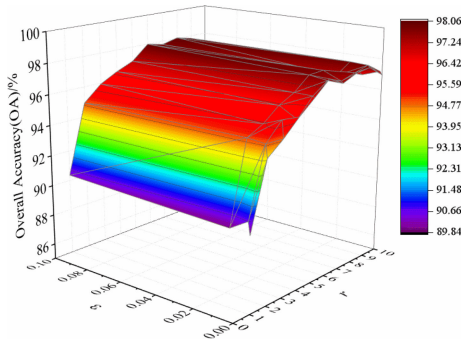Fig. 7.   OA under different r and ε of Indian Pines dataset.

Fig. 8.   OA under different r and ε of Pavia University dataset.

Fig. 9.   OA under different T.

Fig. 10.   OOB under different T.

Fig. 11.   Comparison of OA under different numbers of training samples in Indian Pines dataset.
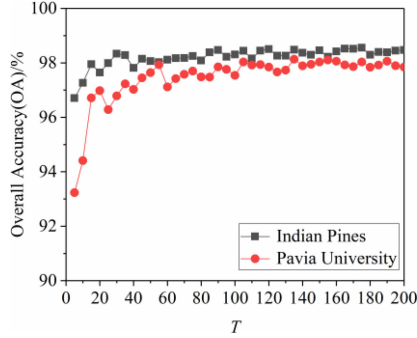
and then no longer improves or even decreases, which is related to the distribution of the ground objects in the HSI itself. When r becomes higher, the spatial information obtained is richer, but when enough spatial information is obtained, noise may be introduced in increasing the window. In addition, experiments show that ε values of 0.0001, 0.001, 0.01, and 0.1 have little effect on the classification results.

In order to evaluate the influence of T in the RF on the performance of the classifier, the T values of the two images are dynamically changed by step size 5 within the range of (5,200). Figs. 9 and 10 show the influence of different T and OOB error when parameters k, t, r, and ε are optimal. For the Indian Pines dataset, OA increases with the increase of T, while OOB error decreases with the increase of T and gradually becomes stable. When $T = 175$, OOB error is the smallest of 0.0029, and OA
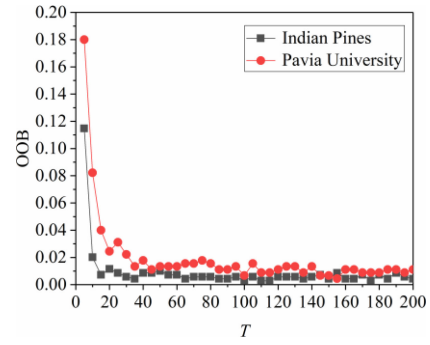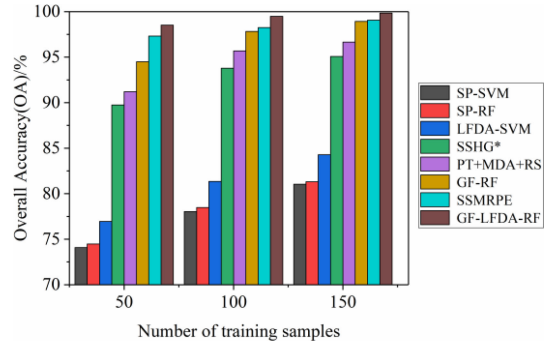
is the highest of 98.56%. Similarly, in Pavia University, when $T = 155$, OOB error value is the smallest, which is 0.0044, and OA is the highest, with a value of 98.13%.

D. Experimental Results and Evaluation of Indian Pines

In the experiment, 50, 100, and 150 samples are randomly selected from each class of ground objects (when the number of ground objects is insufficient, half is selected) as training samples, and the remaining are test samples. Fig. 11 shows the comparison of OA of each method under different numbers of training samples. It can be seen that as the number of training samples increases, OA of each method improves. This is because the more training samples, the richer the discrimination information, and the higher the classification accuracy. The classification results of SP-SVM and SP-RF are similar, indicating that the generalization capabilities of SVM and RF classifiers are similar. LFDA-SVM utilizes the local information between pixels to extract the discriminant features that are helpful for classification, which improves the accuracy compared to using only the original spectral features. However, due to only using the spectral features of the image, the improvement of results is limited. SSHG∗, PT + MDA + RS, GF-RF, SSMRPE, and GF-LFDA-RF introduce spatial information of images, and the classification results have been significantly improved. When different numbers of training samples are selected, the OA of the proposed method is always the highest, indicating that the method can extract the spatial features of the image based on

TABLE I
CLASSIFICATION RESULTS OF EACH METHOD ON INDIAN PINES (%)

| Class	SP-SVM	SP-RF	LFDA-SVM	SSHG*	PT+MDA+RS	GF-RF	SSMRPE	GF-LFDA-RF
Alfalfa	85.71	87.09	**100**	64.70	70.96	**100**	91.66	**100**
Corn-notill	72.75	76.93	75.28	92.50	95.58	97.88	97.58	**99.23**
Corn-min	73.40	77.15	70.19	90.70	92.90	93.22	96.72	**98.02**
Corn	70.06	65.35	93.37	94.83	95.16	99.04	97.57	**99.06**
Grass/pasture	88.46	85.01	98.96	96.86	97.47	99.52	96.85	**99.76**
Grass/trees	88.22	88.44	96.92	98.33	98.63	97.18	**100**	**100**
Grass-mowed	70.37	70.83	**100**	87.50	**100**	**100**	**100**	**100**
Hay-windrowed	91.59	94.64	**100**	**100**	**100**	99.77	**100**	**100**
Oats	90	71.43	71.42	76.92	**100**	85.71	71.42	**100**
Soybeans-notill	73.90	74.26	81.14	91.41	96.03	98.78	97.28	**99.77**
Soybeans-min	79.23	80.18	83.14	92.62	95.95	97.56	99.22	**100**
Soybeans-clean	80.83	71.45	60.74	92.75	92.20	**99.81**	97.89	98.16
Wheat	88.67	88.94	66.82	**100**	**100**	99.46	**100**	**100**
Woods	93.18	95.34	97.68	99.47	99.64	99.74	99.38	**100**
Bldg-grass-drives	62.18	64.78	93.29	96.56	94.15	98.57	95.76	**100**
Stone-steel-towers	97.43	98.78	48	**100**	**100**	**100**	**100**	**100**
OA	80.43	81.20	83.93	94.41	96.27	98.05	98.31	**99.57**
AA	81.62	80.66	83.56	92.20	95.54	97.89	96.33	**99.62**
Kappa	77.65	78.55	81.55	93.61	95.74	97.77	98.07	**99.51**
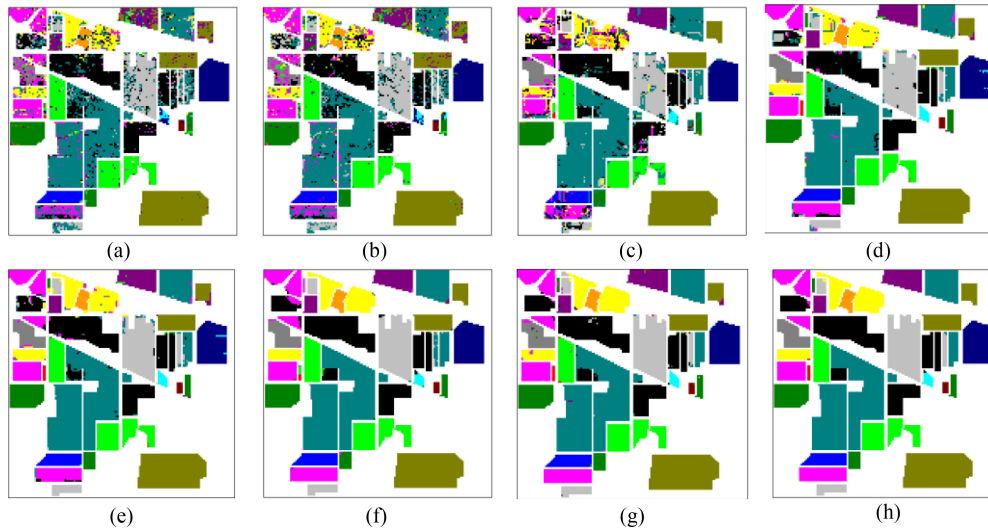
Fig. 12. Classification result and OA of each method on Indian Pines. (a) SP-SVM(OA = 80.21%). (b) SP-RF(OA = 82.04%). (c) LFDA-SVM(OA = 83.88%). (d) SSHG∗(OA = 94.05%). (e) PT + MDA + RS(OA = 95.87%). (f) GF-RF(OA = 98.04%). (g) SSMRPE(OA = 98.47%). (h) GF-LFDA-RF(OA = 99.63%).

the characteristics of HSI data such as high resolution and multimodality. At the same time, considering the edge information of the features and the local information of the pixels, the classification results are improved.

In order to verify the influence of different methods on the classification results of the ground objects under the balanced training set, 10% samples are randomly selected from each class of the ground object as training samples and the remaining are taken as test samples. Table I shows the classification accuracy of each class, OA, AA, KC, and running time of each method. Fig. 12 is the classification result of one experiment randomly selected from the ten experiments. According to Table I, OA, AA, and KC of the proposed method are about 99.57%, 99.62%, and 99.51%, respectively, which are all higher than the results of SP-SVM, SP-RF, LFDA-SVM, SSHG∗, PT + MDA + RS,

GF-RF, and SSMRPE. Among them, OA improved by about 19.14%, 18.37%, 15.64%, 5.16%, 3.3%, 1.52%, and 1.26%, respectively. AA improved by about 18%, 18.96%, 16.06%, 7.42%, 4.08%, 1.73%, and 3.29%, respectively. KC improved by about 21.86%, 20.96%, 17.96%, 5.9%, 3.77%, 1.74%, and 1.44%, respectively. At the same time, 15 of the 16 classes have obtained the best classification accuracy in proposed method. Although the generalization capabilities of SVM and RF classifiers are similar, RF consumes less time than SVM. When randomly selecting 10% of the training samples to train the model, there will be small samples such as 2, 3, and 5, but the proposed method still obtains good classification results, which shows that the method can better deal with the small sample learning problem of HSI classification. As can be seen from Fig. 8, since the SP-SVM, SP-RF, and LFDA-SVM only

TABLE II
CLASSIFICATION RESULTS OF EACH METHOD ON PAVIA UNIVERSITY (%)

| Class	SP-SVM	SP-RF	LFDA-SVM	SSHG*	PT+MDA+RS	GF-RF	SSMRPE	GF-LFDA-RF
Asphalt	75.94	78.59	80.28	87.41	93.63	91.49	92.96	**96.39**
Meadows	82.78	84.03	85.60	98.25	97.83	99.57	97.07	**99.68**
Gravel	80.60	65.50	74.91	84.93	88.01	90.02	95.14	**97.91**
Trees	93.04	98.41	93.07	95.69	99.54	**99.65**	98.15	97.66
Painted metalsheets	96.63	97.92	95.93	90.55	86.82	90.18	92.88	**98.08**
Bare Soil	93.89	92.26	95.64	97.30	97.96	95.41	97.93	**98.27**
Bitumen	64.37	80.69	84.30	90.58	97.30	90.19	**99.90**	95.47
Self-Blocking Bricks	72.45	70.34	77.67	88.26	84.53	**96.22**	94.38	92.26
Shadows	**100**	**100**	**100**	92.37	99.45	97.74	99.20	93.97
OA	82.28	83.07	85.25	94.15	95.26	96.39	96.42	**97.79**
AA	77.26	80.66	87.48	92.21	94.01	94.53	94.67	**96.63**
Kappa	75.46	78.55	79.78	91.70	93.68	95.22	95.24	**97.07**
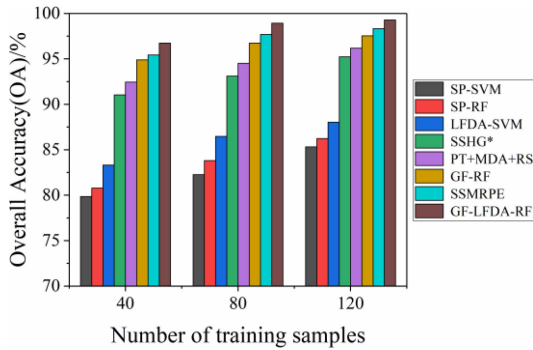
Fig. 13. Comparison of OA under different numbers of training samples in Pavia University dataset.

use spectral features, a serious "salt and pepper phenomenon" appears in the results of them. SSHG*, PT + MDA + RS, and SSMRPE M combine spectral features with spatial features, and the classification results are significantly improved, but there are many misclassification results at the edges of the features. The results of GF-RF and GF-LFDA-RF show that the boundaries of the ground objects are clear, which indicates that guided filtering can better maintain the edge information of the ground objects while achieving smoothing. The proposed GF-LFDA-RF method introduces local information between pixels on the basis of extracting spatial features by guided filtering, and obtains embedded features conducive to classification, which can better reflect the real distribution of the ground objects.

E. Experimental Results and Evaluation of Pavia University

In the experiment, 40, 80, and 120 samples are randomly selected from various ground objects as training samples, and the remaining samples are test samples. Fig. 13 shows OA of each method under different training sample numbers. It can be seen that OA of the proposed method is higher than other methods.

To verify the classification results of the ground objects by different methods under the balanced training set, 1% samples are randomly selected from each ground object as training samples and the remaining samples as test samples. In Table II, the comparison results are listed. Fig. 14 is the classification result of

one experiment randomly selected from the ten experiments. As shown in the results, the proposed method also gets the best OA, AA, and KC. At the same time, five classes have obtained the best classification accuracy. It can be seen from Fig. 14 that the classification results obtained by GF-LFDA-SVM have fewer noise points and a smoother distribution of the ground objects (especially the edges of the ground objects), further illustrating the effectiveness of the method proposed in this article.

F. Computational Complexity

In order to analysis the computational complexity of proposed method, suppose a HSI has p bands, and the number of pixels of each band is denoted as m, then the image filtering takes $O(p \times m)$. The computational complexity of low-dimensional embedding is determined by two aspects. One is the calculation of neighbors, and the other is the solution of generalized eigenvectors. The similarity distance between any two samples is calculated with the cost of $O(pn^2)$, n is the number of samples. The finding of k neighbors costs $O(kn^2)$. For ease of understanding, it is assumed that both the within-class and between-class scatter matrices are $D \times D$, and solving the generalized feature vector takes $O(D^3)$. Therefore, the computational complexity of the dimensionality reduction is $O(pn^2 + kn^2 + D^3)$. The cost of RF classifier training n samples is $O(n \log(n) \times q \times T)$, where q is the dimension after dimensionality reduction, T is the number of decision trees. As a result, the final computational complexity of proposed method is $O(s \times p + pn^2 + kn^2 + D^3 + n \log(n) \times q \times T)$, and it mainly depends on the size of training samples, pixel number, band number, and decision tree number.

To quantitatively compare the complexity of each method, the computational time of each method is shown in Table III. The time of each method is the sum of training time and testing time. All of the results were obtained on a personal computer, which has a CPU of Intel Core i5@2.7 GHz and 8-GB memory. The software implementation was performed using MATLAB 2017 and python 3.5. As shown in Table III, the proposed method is faster than other methods on Indian Pines dataset, but is slower than SP-SVM, SP-RF, and LFDA-SVM on Pavia University dataset. The reason is that Pavia University dataset has more pixels than Indian Pines dataset, so it takes more time in the filtering process.
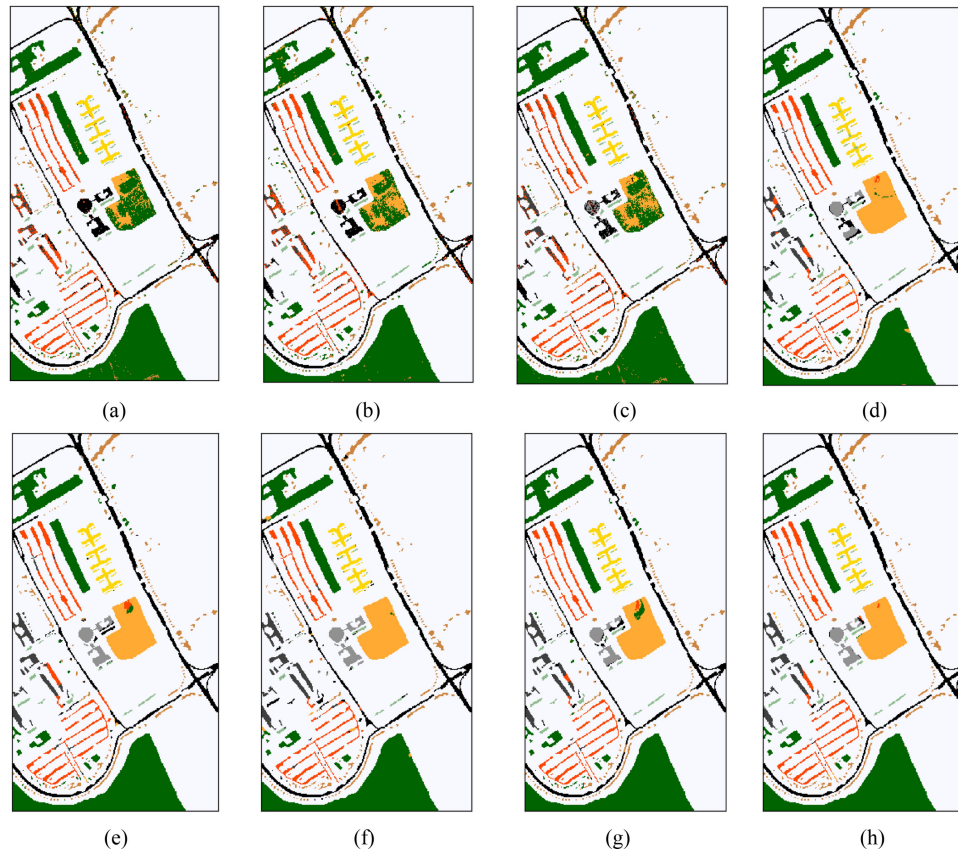
Fig. 14. Classification result and OA of each method on Pavia University. (a) SP-SVM(OA = 82.36%). (b) SP-RF(OA = 82.89%). (c) LFDA-SVM(OA = 85.61%). (d) SSHG∗(OA = 94.33%). (e) PT + MDA + RS(OA = 95.62%). (f) GF-RF(OA = 96.04%). (g) SSMRPE(OA = 96.75%). (h) GF-LFDA-RF(OA = 98.11%).

TABLE III
COMPUTIONAL TIME (IN SECONDS) OF DIFFERENT METHODS ON INDIAN PINES AND PAVIA UNIVERSITY DATASETS

Data	SP-SVM	SP-RF	LFDA-SVM	SSHG*	PT+MDA+RS	GF-RF	SSMRPE	GF-LFDA-RF
Indian Pines	24.2	13.3	23.7	26.4	45.4	14	23.5	12.8
Pavia University	10.2	3.9	8.7	22.43	395.2	24.5	43.6	22.4

V. CONCLUSION

In this article, a HSI classification method is proposed, which makes full use of spatial and local features and keeps the edges of the ground objects. In the proposed method, the local linear model between the first principal component of PCA and the output image is built to filter each band of the input image, so as to achieve the goal of removing the noise in the input image and preserving the edges of the ground objects. Then, in the low-dimensional embedding, the local scatter matrices are calculated according to the local neighborhood relationships between the samples, and the random forest classifier is used to carry out the classification. The experimental results on the Indian Pines and Pavia University hyperspectral datasets show that the proposed algorithm performs much better than some state-of-the-art methods in terms of classification accuracy and KC. The limitation of proposed method is that the parameters of the model need to be manually adjusted during classification. Therefore, our future work will focus on how to realize the adaptive parameters to further improve the classification performance of the proposed method.

REFERENCES

[1] M. Imani, "Manifold structure preservative for hyperspectral target detection," *Adv. Space. Res.*, vol. 61, no. 10, pp. 2510–2520, Mar. 2018.

[2] H. Yang, Q. Du, and G. Chen, "Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 5, no. 2, pp. 544–554, Apr. 2012.

[3] M. R. Corson, R. L. Lucke, C. O. Davis, J. H. Bowles, and W. A. Snyde, "The hyperspectral imager for the coastal ocean (HICO) environmental littoral imaging from the international space station," *IEEE Geosci. Remote Sens. Symp.*, vol. 4, no. 6, pp. 3752–3755, Aug. 2010.

[4] X. Z. Shi, I. C. Lau, and M. Aspandiar, "Comparison of PLSR modelling and indicative mineral mapping of airborne hyperspectral imagery for acid sulphate soil assessment," *Int. J. Remote Sens.*, vol. 35, no. 4, pp. 1309–1330, Feb. 2014.

[5] R. Hang *et al.*, "Robust matrix discriminative analysis for feature extraction from hyperspectral images," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 10, no. 5, pp. 2002–2011, May 2017.

[6] L. Li, H. Ge, and J. Gao, "A spectral-spatial kernel-based method for hyperspectral imagery classification," *Adv. Space Res.*, vol. 59, no. 4, pp. 954–967, Nov. 2016.

[7] Y. Zhou, A. Rangatajan, and P. D. Gader, "A Gaussian mixture model representation of endmember variability in hyperspectral unmixing," *IEEE Trans.. Image. Process.*, vol. 27, no. 5, pp. 2242–2256, Feb. 2018.

[8] Z. Liu, B. Tang, X. He, Q. Qiu, and F. Liu, "Class-specific random forest with cross-correlation constraints for spectral–spatial hyperspectral image classification," *IEEE Geosci. Remote Sens.*, vol. 14, no. 2, pp. 257–261, Jan. 2017.

[9] L. GAN, J. Xia, P. Du, and C. Jocelyn, "Class-oriented weighted kernel sparse representation with region-level kernel for hyperspectral imagery classification," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 11, no. 4, pp. 1118–1130, Apr. 2018.

[10] Z. Wang, B. Du, L. Zhang, and L. Zhang, "A novel semisupervised active-learning algorithm for hyperspectral image classification," *IEEE Trans. Geosci. Remote*, vol. 55, no. 6, pp. 3071–3083, Jun. 2017.

[11] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2017.

[12] X. Ma, H. Wang, and J. Geng, "Spectral–spatial classification of hyperspectral image based on deep auto-encoder," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 9, no. 9, pp. 4073–4085, Feb. 2016.

[13] S. Mei, J. Ji, J. Hou, X. Li, and Q. Du, "Learning sensor-specific spatial-spectral features of hyperspectral images via convolutional neural networks," *IEEE Trans. Geosci. Remote*, vol. 55, no. 8, pp. 4520–4533, May 2017.

[14] P. Ghamisi, B. Hfle, and X. X. Zhu, "Hyperspectral and LiDAR data fusion using extinction profiles and deep convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ.*, vol. 10, no. 6, pp. 3011–3024, Jun. 2017.

[15] Y. Li, H. Zhang, and Q. Shen, "Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network," *Remote Sens.*, vol. 9, no. 1, pp. 67, Jan. 2017.

[16] F. Palsson, J. R. Sveinsson, and M. O. Ulfarsson, "Multispectral and hyperspectral image fusion using a 3-d-convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 639–643, Jun. 2017.

[17] B. Liu, X. Yu, A. Yu, P. Zhang, and G. Wan, "Spectral-spatial classification of hyperspectral imagery based on recurrent neural networks," *Remote Sens. Lett.*, vol. 9, no. 12, pp. 1118–1127, Dec. 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, Apr. 2017.

[19] Q. Du and J. E. Fowler, "Hyperspectral image compression using JPEG2000 and principal component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 201–205, May 2007.

[20] W. Liao, A. Pizurica, P. Scheunders, and S. Paul, "Semisupervised local discriminant analysis for feature extraction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 184–198, Jan. 2013.

[21] Y. Deng, H. Li, L. Pan, L. Shao, Q. Du, and W. J. Emery, "Modified tensor locality preserving projection for dimensionality reduction of hyperspectral images," *IEEE Geosci. Remote Sens.*, vol. 15, no. 2, pp. 277–281, Jan. 2018.

[22] L. Zhang and C. Zhao, "Sparsity divergence index based on locally linear embedding for hyperspectral anomaly detection," *J. Appl. Remote Sens.*, vol. 10, no. 2, Jun. 2016, Art. no. 025026.

[23] M. Lv, X. Zhao, L. Liu, and L. Jing, "Discriminant collaborative neighborhood preserving embedding for hyperspectral imagery," *J. Appl. Remote Sens.*, vol. 11, no. 4, Oct. 2017, Art. no. 046004.

[24] G. Taskin, H. Kaya, and L. Bruzzone, "Feature selection based on high dimensional model representation for hyperspectral images," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2918–2928, Jun. 2017.

[25] F. Luo, H. Huang, Y, Duan, J. Liu, and Y. Liao, "Local geometric structure feature for dimensionality reduction of hyperspectral imagery," *Remote Sens.*, vol. 9, no. 8, pp. 790, Aug. 2017.

[26] W. Li, S. Prasad, J.E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 40, pp. 1185–1198, Apr. 2012.

[27] R. Hang and Q. Liu, "Dimensionality reduction of hyperspectral image using spatial regularized local graph discriminant embedding," *IEEE J-Star.*, vol. 11, no. 9, pp. 3262–3271, Sep. 2018.

[28] H. Yuan and Y. Y. Tang, "Spectral-spatial shared linear regression for hyperspectral image classification," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1–12, Mar. 2016.

[29] Y. Xu, Z. Wu, and Z. Wei, "Spectral–spatial classification of hyperspectral image based on low-rank decomposition," *IEEE J-Star.*, vol. 5, no. 6, pp. 1–11, Jun. 2015.

[30] R. Hang, Q. Liu, H. Song, and Y. Sun, "Matrix-based discriminant subspace ensemble for hyperspectral image spatial–spectral feature fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 2, pp. 1–12, Sep. 2015.

[31] Y. Sun, S. Wang, Q. Liu, R. Hang, and G. Liu, "Hypergraph embedding for spatial-spectral joint feature extraction in hyperspectral images," *Remote Sens.*, vol. 9, no. 5, May 2017, Art. no. 506.

[32] H. Huang, G. Shi, H. He, Y. Duan, F. Luo, "Dimensionality reduction of hyperspectral imagery based on spatial-spectral manifold learning," *IEEE Trans. Cybern.*, vol. 50, no. 6, pp. 2604–2616, Jun. 2020.

[33] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Softw. Eng.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013.

[34] P. S. Kumar, G. Supratim, and P. K. Sahu, "Curvelet-based multiscale denoising using non-local means & guided image filter," *IET Image Process.*, vol. 12, no. 6, pp. 909–918, Jan. 2018.

[35] J. Zhu, W. Jin, L. Li, and Z. Han, "Multiscale infrared and visible image fusion using gradient domain guided image filtering," *Infrared Phys. Techn.*, vol. 89, pp. 8–19, Dec. 2017.

[36] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[37] Z. Liu, B. Tang, X. He, Q. Qiu, and F. Liu, "Class-specific random forest with cross-correlation constraints for spectral–spatial hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 257–261, Feb. 2017.

[38] H. Ma, W. Feng, X. Cao, and L. Wang, "Classification of hyperspectral data based on guided filtering and random forest," *Int. Arch. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. XLII-2/W7, pp. 821–824, Sep. 2017.

Hui Zhang received the B.E. degree in software engineering from Liaoning Technical University, Huludao, Liaoning, China, in 2007, and the M.Sc. degree in computer software and theory from Liaoning Technical University, Huludao, Liaoning, in 2012. He is currently working toward the Ph.D. degree with the School of Electronic and Information Engineering, Liaoning Technical University, Fuxin, China.

His current research interests include machine learning, deep learning, computer vision, and image processing.
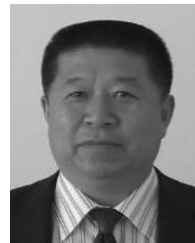
Wanjun Liu was born in Jinzhou, Liaoning, China in 1959. He received the B.E. degree in computer and application from Yanshan University, Hebei, China, in 1982, and the M.S. degree in electrical transmission and automation from Liaoning Technical University, Fuxin, China, in 1991.

From 2002 to 2018, he was the Director with the School of Software, Liaoning Technical University. He is currently a Professor with the School of Software, Liaoning Technical University, Huludao, China. His research interests include software architecture technology, computer vision and image processing, pattern recognition and artificial intelligence.

Huanhuan Lv received the Ph.D. degree from the School of Surveying and Geographic Sciences, Liaoning Technical University, Fuxin, China, in 2014.

She is currently an Associate Professor with the Software School, Liaoning Technical University, Fuxin, Liaoning, China. In the past five years, she has hosted or participated in many projects, including two National Natural Science Foundation Projects, two Science and Technology Project in Liaoning Province, and five Enterprise Cooperation Projects. She has authored or coauthored 12 journal or conference papers. Her research topics include computer vision, image processing, machine learning, and deep learning.