IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 17, 2019, accepted July 2, 2019, date of publication July 10, 2019, date of current version August 16, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2927786

Multiple Feature Learning Based on Edge-Preserving Features for Hyperspectral Image Classification

WEI TIAN[1,2], LIZHONG XU[1], ZHE CHEN[1], AND AIYE SHI[1]
[1]Computer and Information College, Hohai University, Nanjing 211100, China
[2]Anhui Technical College of Mechanical and Electrical Engineering, Wuhu 241002, China

Corresponding author: Lizhong Xu (hhucomputer@aliyun.com)

ABSTRACT The classification of hyperspectral images is the basis and hotspot in the research of hyperspectral images. In this paper, a classification algorithm of hyperspectral image based on multiple edge-preserving features and multiple feature learning (MFL) is proposed. First, aiming to eliminate the high correlation between adjacent bands and to remove the noise in the image, a new band clustering algorithm is employed to reduce the dimensions, where the dimension-reduced image can be used as spectral information features to extract linearly separable classes. Then, spatial information features are obtained by applying the multiple edge-preserving filter on the reduced-dimensional image. This filter is used to acquire more comprehensive spatial information features of the image for extraction of nonlinearly separable classes. Following that, the locality preserving projections method is applied to retain the representative spatial information from the extracted spatial information for classification accuracy. Finally, the spectral information features and spatial information features are combined for classification using the MFL. The experiments are conducted to verify the validity of the proposed algorithm on three universally adopted hyperspectral datasets.

INDEX TERMS Band clustering, multiple edge-preserving features, multiple feature learning, locality preserving projections.

I. INTRODUCTION

Hyperspectral images provide high spectral resolution and robust classification ability by recording hundreds of spectral bands for each pixel. Therefore, remote sensing images are widely used to identify substances [1], such as minerals and rock types, to distinguish the composition of various pollutants in the environment [2], [3] and the types of crops and forests [4]. The primary objective of the remote sensing technology is to identify the object classes and its distribution according to the radiant attributes of the ground object in the remote sensing images [5], [6]. Therefore, the classification of hyperspectral images is considered a research hotspot in this field [7], [8]. Classical hyperspectral classification methods, such as the support vector machines (SVMs) [9], random forest [10], and the artificial neural networks [11] consider only the original spectral information of the hyperspectral images [12]. Therefore, the classification accuracy values in these methods are not high due to "the curse of the dimension," especially in the case of limited training sets.

To prevent the Hughes phenomenon, classical dimensionality reduction algorithms, such as the principal component analysis (PCA) [13], [14], independent component correlation algorithm [15], and the generalized discriminant analysis [16], have been proposed. However, they are not ideal for improving classification accuracy. Ghorbanian and Mohammadzadeh [17] proposed a k-means clustering method based on band correlation to achieve the dimensionality reduction by calculating the average value of each cluster. However, this method suffers the disadvantage of slow convergence and different initial cluster center selection, which can lead to unstable results. Kang *et al.* [18] employed the mean fusion of adjacent bands to reduce the dimension, which performed a uniform partitioning on the group fusion. There was a possibility of losing some critical information with this method.

Aiming to enhance the accuracy of the image classification, experts in the field mostly focus on extracting the spatial information, such as the morphological attribute profiles (MAPs) [19], [20]. Such a method overcomes the limitation

The associate editor coordinating the review of this manuscript and approving it for publication was Huimin Lu.

of using a pixel or a fixed neighborhood as the processing unit in traditional methods and expands the processing unit to all the pixels with similar attributes to obtain various attribute information of the hyperspectral images. Note that when applying the MAPs, it is usually necessary to calculate multiple attributes of the image and multiple threshold features so that the obtained attribute features demonstrate a high dimensionality. The edge-preserving filter (EPF) method [21]–[23] can extract the spatial detail information of the image, eliminate the noise, and effectively alleviate the phenomenon of ''the same object with different spectrum'' to achieve smoothing and edge preservation. However, the drawback of this method is that the standard single parameter setting of the EPF is not capable of adequately representing the spatial information of the image. Chen *et al.* [24] employed the convolutional neural networks to extract the depth features of the image, so that the classification accuracy is improved. However, this method is time-consuming and involves considerable costs of computational complexity [25].

In recent years, several studies have combined the original spectral information with various spatial information features extracted from the image (for example, texture features, shape features, and spatial relationship features) that describe the features of the pixel from different angles to represent the characteristics of the hyperspectral image in entirety. The method of constructing the composite kernel (CK) functions [26], [27] combines the spatial information extracted by MAPs with the spectral information of the image. After being mapped to the high-dimensional kernel space, the nonlinearly separable features of the original feature space turn linearly separable. The application of the CK function overcomes the limitations of the single kernel function in the SVMs model. The multiple kernel learning method [28], [29] adopts different kernels for various features (e.g., spectral information features and spatial information features), composes multiple kernels with different parameters, and then optimizes the weight of each kernel to consider the best combination of kernel functions for classification. The two aforementioned methods generally require the combination of kernel functions to be convex, which poses difficulties in optimizing some parameters during the learning process. The emergence of the generalized composite kernel (GCK) method [30] has successfully overcome this limitation. GCK based on multinomial logistic regression (MLR) [31] can flexibly combine various features of the images linearly without setting the weight parameters for each of the features. This learning method-based kernel function maps various features of the image to the nonlinear kernel space, which leads to redundancy or loss of the physical meaning of the feature. The multiple feature learning (MFL) method [32] utilizes not only the original spectral information and spatial information of the image but also the kernel transformation of the spectral information and spatial information considering both linear and nonlinear features, which improves the classification accuracy to a certain extent. In multiple-feature-based adaptive sparse representation method [33], four different features (spectral value feature, extended morphological profile, Gabor texture and differential morphological profiles) are extracted from the original hyperspectral images. The adaptive sparse representation algorithm is introduced to determine the class label of each pixel. However, the feature dictionary comprises four huge features, which increases the computational complexity.

In summary, this article presents a hyperspectral classification based on multiple feature learning and edge-preserving features (MFL-EPFs). Firstly, a new band clustering method is applied to divide the high correlation bands into groups. Then the resulting clustering groups are averaged to obtain the dimensionality reduction image, which retains the useful spectral information in the original image and plays vital role in the subsequent classification. Next, the multiple edge-preserving filter is used to extract useful spatial information. The outstanding performance of the multiple edge-preserving filter in the hyperspectral classification has been confirmed in the literature [34]. Then LPP is performed on the multiple edge-preserving features to extract the features that are highly useful for classification [35]. Finally, the MFL method is utilized to assign a unique label to each pixel by combining the dimension-reduced spectral features with the dimension-reduced multiple edge-preserving features (MEPFs).

Following a brief overview of the multiple MEPFs, LPP and MFL in section 2, the descriptions are mainly provided for readers with no foundation knowledge in this field. Section 3 is dedicated to introducing the proposed method MFL-EPFs in detail. Section 4 includes an evaluation of the performance of MFL-EPFs compared to other mainstream methods and analyzes the experimental results. Possible objectives for future research are discussed and conclusions are presented in section 5.

II. BACKGROUND KNOWLEDGE

A. MULTIPLE EDGE-PRESERVING FEATURES

During the process of acquisition and transmission of hyperspectral remote sensing images, different kinds of noises are often introduced, which results in the fluctuation of spectral characteristics of the same objects. It is complicated to classify the hyperspectral remote sensing images accurately under small training samples. Edge-preserving filtering is one of the most critical approaches used to eliminate the influence of noise. The three typical edge-preserving filtering methods include the bilateral filtering [36], guided filtering [37], and the domain transform [18]. The respective advantages and disadvantages of these methods have been elaborated in the literature [34]. Domain-transform recursive filtering is adopted in our study, as it provides the advantages of real-time processing and high efficiency.

Domain transform recursive filtering applies the approximate distance-preserving transformation to the one-dimensional signal, i.e., the Euclidean distance between two adjacent samples in the transform domain is equal to L1-norm distance between two adjacent samples in the original domain. For the one-dimensional signal $\boldsymbol{I}$,

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

FIGURE 1. Influence of the two parameters, i.e., δ_s and δ_r on the performance of the filter. (a) Input hyperspectral band. (b)–(d) Filtered bands with different parameter settings.

we obtained

$$U_i = I_0 + \sum_{j=1}^{i} \left(1 + \frac{\delta_s}{\delta_r}\left|I_j - I_j - 1\right|\right), \tag{1}$$

where U is the domain transform signal. δ_s and δ_r are the spatial standard deviation and the range standard deviation, respectively, which are used to control the window size and ambiguity of the filter, respectively.

Then the domain-transformed signal is transmitted to the recursive filtering for processing, and the output of the recursive filter is related to both the input and the output, as shown in the following equation:

$$J_i = \left(1 - a^b\right) I_i + a^b J_{i-1}, \tag{2}$$

where J_i is the output of the ith pixel, $a = \exp(\sqrt{2}/\delta_s) \in [0, 1]$ represents the feedback coefficient, and b refers to the distance between two adjacent samples in the transform domain. As a^b head towards zero, the edge information of the signal is preserved and vice versa.

When the spatial standard deviation δ_s and the range standard deviation δ_r are provided different values, the obtained images are different, as shown in Fig. 1.

Different parameter settings produce different filtering results that represent the features of the hyperspectral images on multiple scales and have their peculiar advantages in classifying those objects or features of different scales [34] Therefore, a combination of the edge features of different parameters can extract additional spatial information from the images and improve the classification accuracy.

The steps to obtain the MEPFs for the k-dimensional hyperspectral images I_k are as follows:

1) First, each dimension image is filtered by the EPF with different parameters, therefore,

$$F_{k^x} = MEPF(I_k, \delta_s^x, \delta_r^x), \tag{3}$$

where $k = 1, \ldots, K, x = 1, \ldots, X$, δ_s^x and δ_r^x are the xth parameter settings adopted by the EPF,

$$(\delta_s^x, \delta_r^x) \subset \left\{\left(\delta_s^1, \delta_r^1\right), \ldots, \left(\delta_s^X, \delta_r^X\right)\right\}. \tag{4}$$

2) The resulting features are stacked together in a series as equation (5):

$$F = \left\{F^1, \ldots, F^X\right\}. \tag{5}$$

B. LOCALITY PRESERVING PROJECTIONS

PCA occupies a dominant position in the unsupervised dimension reduction methods because of its simple calculation and high efficiency [38]. However, it has a poor effect on the non-Gaussian data, and only linear subspace can be discovered. Therefore, the data distributed on the nonlinear manifold cannot be processed appropriately, which results in data distortion and loss of some useful information after the dimension reduction [39].

LPP can be used to resolve the problem mentioned above. It introduces linearization into the manifold learning and establishes a graph for the data set with the neighborhood information [40]. Using the Laplace concept of the graph, a projection matrix is calculated to map the data on to a subspace. This linear transformation maintains the local neighborhood information optimally to some extent. Therefore, the nearest neighbor search in the low-dimensional space yields results similar to that in the high-dimensional space. LPP can provide highly useful information. LPP is linear and therefore, enables faster processing. LPP possesses the capability to learn out-of-sample points so that the new data points can be mapped to a low-dimensional space through a linear transformation [41]. In summary, LPP can be used as a natural alternative to PCA [35].

Let $X = (X_1, X_2, \ldots, X_n)$ be the sample data set in the original space R^D, n be the number of samples, D be the data set, and x_i and x_j be the i-th and j-th samples in X. Then the similarity of the samples [42] is defined as follows:

$$s_{ij} = \exp(-\frac{\left\|x_i - x_j\right\|^2}{t}). \tag{6}$$

In (6), the parameter $t > 0$ is a constant that is set empirically. If w is the projection vector, then the objective function of LPP can be defined as follows [35]:

$$\begin{aligned} w^* &= \arg\min_w \sum_{i \neq j}^{N} s_{ij} \left\|w^T x_i - w^T x_j\right\|^2 \\ &= \arg\min_w w^T X(D - S)X^T w \\ &= \arg\min_w w^T XLX^T w, \end{aligned} \tag{7}$$

where D is a diagonal matrix, in which the elements on the diagonal are the sum of the rows corresponding to the matrix S and $L = D\text{-}S$ is the Laplacian matrix. Then, as per the constraint $w^T XDX^T w = 1$, the optimal projection vector is the eigenvector corresponding to the solution of the minimum eigenvalue.

$$XLX^T w = \lambda XDX^T w, \tag{8}$$

where $w = (w_1, w_2, \ldots, w_d)$ is the projection matrix, $w_1, w_2, \ldots, w_d$ is the eigenvector corresponding to the smallest d eigenvalues in (8), which holds most of the useful information.

C. MULTIPLE FEATURE LEARNING

Combining various features of an image is beneficial in improving the separability of the data, and the multiple feature learning method can combine different kinds of features using a linear combination to improve the classification accuracy. This method does not require any regularization parameters to control the weight of each feature. Therefore, it effectively utilizes different features and flexibly combines them [32].

The linear combination of various features is realized using the MLR. After linearly combining the independent variables (which can be multiple features) and the corresponding parameters, a probabilistic model is employed to calculate and predict the probabilities of the different possible outcomes of a categorically distributed dependent variable. The calculation formula employed is as follows [31]:

$$p\left(y_i = k|x_i, w\right) = \frac{\exp\left(w^{(k)^T} h\left(x_i\right)\right)}{\sum_{k=1}^{K} \exp\left(w^{(k)^T} h\left(x_i\right)\right)}, \tag{9}$$

where $k = (1, \ldots, K)$ is a set of K class labels, $x = (x_1, \ldots, x_n) \in R^d$ is the hyperspectral image composed of d-dimensional feature vectors, and n is the band number of the reduced-dimensional hyperspectral image.w is the regression coefficient, and $w = \left[w^{(1)^T}, \ldots, w^{(K-1)^T}\right]^T$, $w^{(K)} = 0, y = \left(y_1, \ldots, y_n\right)$ denotes the class label of the image. $h(x_i)$ is a set of all kinds of input features, which can be linear or nonlinear. It is defined as follows [32]:

$$h(x_i) = [1, h_1(x_i)^T, \ldots, h_l(x_i)^T]^T, \tag{10}$$

where l is the total number of classes with various features, and $h_j(x_i)$ can be

$$h_j\left(x_i\right) = [x_{i1}, \ldots, x_{id}]^T, \tag{11}$$

where x_{ij} denotes the jth component of x_i, which represents the original spectral information of the hyperspectral image and is used to extract linear features. $h_j(x_i)$can also be

$$h_j(x_i) = [f_1(x_i), \ldots, f_{l2}(x_i)]^T, \tag{12}$$

where $f(\bullet)$ is the nonlinear feature extraction transformation on the original image (for example, the MEPFs in [34]), and l_2 is the number of elements in $h_j(x_i)$. $h_j(x_i)$ can also be the result of the extracted linear features and nonlinear features mapped to a high-dimensional space by the kernel function.

$$h_j(x_i) = [K(x_i, x_j), \ldots, K(x_i, x_l)]^T, \tag{13}$$

where

$$K\left(x_i, x_j\right) = \left\langle \phi\left(x_i\right) \phi\left(x_j\right)\right\rangle, \tag{14}$$

and $\varphi(\bullet)$ is a nonlinear mapping function.

To obtain the probability p in (6), it is necessary to estimate the logistic regression coefficient w of the input feature by calculating the maximum a posteriori estimate [42].

$$\hat{w} = \arg\max_{w} l(w) + \log p(w), \tag{15}$$

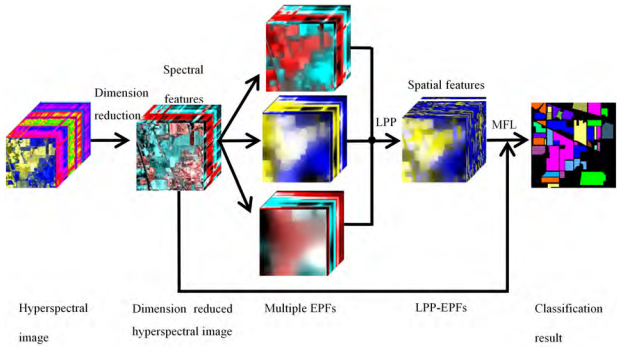


FIGURE 2. Schematic of the proposed MFL-EPFs algorithm.

where $l(w)$ is the likelihood function and the expression is defined as follows:

$$l(w) = \sum_{i=1}^{L} (h^T(x_i) w^{(y_i)} - \log \sum_{k=1}^{K} \exp(h^T(x_i) w^{(k)})), \tag{16}$$

and $logp\ (w)$ is a prior estimate over w that is independent of x. In addition, w is a random vector with Laplace distribution. Therefore,

$$p(w) \propto \exp(-\lambda \|w\|_1), \tag{17}$$

where λ is the regularization parameter that controls the degree of sparsity [43].

III. PROPOSED ALGORITHM

The algorithm flow chart of this study is shown in Fig. 2. Firstly, the spectral features beneficial for the classification were obtained by performing the dimensionality reduction on the original hyperspectral image. Then, multiple features were extracted from the reduced-dimensional hyperspectral image using the EPF of varied parameters, and these features were combined in series. The next step was to reduce the dimensionality of the multiple features using the LPP to obtain the spatial features of the image. Finally, the MFL method was used for classification by combining spectral features and spatial features.

A. HYPERSPECTRAL IMAGE DIMENSIONALITY REDUCTION

High-dimensional hyperspectral images increase the computational cost and include various noises. One of the prior studies [18] has verified that calculating the mean of adjacent bands can remove the noise in the hyperspectral images and preserve the complementary information of adjacent bands. This method preserves the edges and other critical spatial structures compared to the PCA and ICA dimensionality reduction methods. However, the method of uniform grouping in the literature [18] loses some meaningful information. Literature [17] employs the k-means clustering method for grouping, aiming to retain useful information to the maximum extent. Unfortunately, this method suffers the drawbacks that are similar to those of the k-means algorithm.

Therefore, this study proposed a new dimension reduction method for band clustering. The overview of the method is described as follows:

1) From the first band, N bands at equal intervals were considered the cluster centers, denoted as C_1, $C_2, \ldots C_n$.
2) The bands between the cluster centers C_i and C_j were respectively calculated with C_i and C_j to obtain the spectral angle mapping (SAM) values [44]. The calculation formula employed is as follows:

$$SAM(s_i, s_j) = \cos^{-1}(s_i \cdot s_j / \|s_i\| \|s_j\|) \quad (18)$$

where s_i and s_j are the vectors of the i-th band and the j-th band images, respectively. Smaller the value of the SAM, more similar the two spectra. Moreover, they were more likely to belong to the same objects.
3) If the value of the SAM with the cluster center C_i was small, the corresponding band was divided into a group with C_i. This step was applicable for the center C_j, as well.
4) By analogy, the hyperspectral image was divided into N groups.
5) Finally, each group was averaged to obtain a dimensionally reduced image $SpeF$., and N was considered the dimension of the image.

This method is simple and easy to implement. By availing the high correlation between adjacent bands, the averaging method was used to remove the redundancy of the hyperspectral image, eliminate the noise, and finally achieve the dimensionality reduction. It improves the deficiency of the uniform grouping method and overcomes the disadvantage of the K-means algorithm effectively.

B. LPP DIMENSION REDUCTION

The edge preserving features extracted with multiple parameters are concatenated together, resulting in a high dimensionality and a large amount of redundant information, which will affect the separability of the pixels. Although the MEPFs preserves information such as the edge of the image, it also reduces the differences between the pixels of different classes [34]. To extract features that are favorable for image classification, this study proposed a method based on the LPP for the MEPFs. The extracted MEPFs were reduced by the LPP, and the pixel separability of the image was improved by reducing the dimension. Compared to the PCA dimension reduction method, the LPP can preserve the local structure of the image space after dimension reduction [45], to improve the classification precision.

$$SpaF = LPP(F, L) \quad (19)$$

where F is the MEPFs in the series, and L is the dimensionality of the image after the dimensionality reduction. $SpaF$ is the spatial information of the image, which can extract the pixel features of the original hyperspectral image that are nonlinearly separable.

TABLE 1. The comparative results of the PCA and LPP dimension reduction methods for the Indian Pines.

Number of training samples (%)	1	2	3	4	5
The overall accuracy for PCA (%)	81.87	87.19	91.66	93.51	95.17
The overall accuracy for LPP (%)	85.23	91.50	94.61	95.51	96.34
Number of training samples (%)	6	7	8	9	10
The overall accuracy for PCA (%)	95.30	95.80	96.21	96.58	96.83
The overall accuracy for LPP (%)	96.90	97.33	97.57	98.07	98.55

In order to verify the superiority of the LPP, experimental data is displayed in Table 1:

C. MFL FOR CLASSIFICATION

The MFL method can integrate multiple types of features and discover the most representative elements according to the corresponding weights of each feature for the training samples. Even if the input feature dimension is very high, the linear combination results of multiple features can be obtained with a limited training set. More importantly, the method determines the weights of each feature without any regularization parameters, and therefore, these features can be flexibly combined. The combination of linear and nonlinear features is the best approach to represent the essential features of the hyperspectral images. In this study, dimensionally reduced image $SpeF$ is considered the linear feature, which retains most of the spectral information of the original hyperspectral image and plays vital role in the linearly separable classes. Instead of dimensionality reduction, high-dimensional original spectral information is used in [32] and [33]. As the original hyperspectral information is highly correlated and redundant, using all of this information increases the computational burden and affects the classification accuracy to some extent. This study uses the reduced-spectrum hyperspectral information $SpeF$ to resolve the aforementioned issues. Combining the reduced-spectrum hyperspectral information with the reduced-dimensional MEPFs for classification is another innovation of our study.

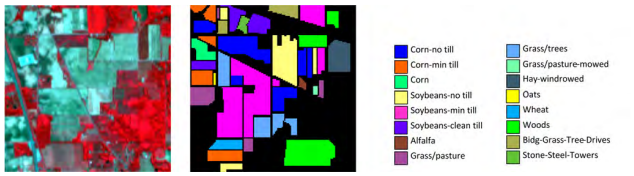
This article does not consider the characteristics of mapping to high dimensional kernel space, mainly to reduce the time consumption. Therefore, the input feature matrix is defined as follows:

$$h(x_i) = [1, h_1(x_i), h_2(x_i)], \quad (20)$$

where $h_1(x_i)$ denotes the spectral information features after the dimension reduction, which is mainly used for the linearly separable classes. $h_2(x_i)$ is the MEPFs after the dimensionality reduction, which extracts the spatial information of the image and is beneficial for the nonlinearly separable classes.

IV. EXPERIMENTAL RESULTS

In this section, we present our experiments on three public hyperspectral datasets to verify the effectiveness of the proposed algorithm. The classification results of the



(a) (b)

FIGURE 3. Indian Pines data set. (a) Three-band color composite. (b) Reference data with 16 land cover classes.

hyperspectral images adopted three commonly used accuracy evaluation indexes: overall accuracy (OA), average accuracy (AA) and Kappa coefficient. As the sample selection is random, we repeated the experiment ten times and then averaged the obtained values to achieve experimental results that are highly objective.

A. DATASET

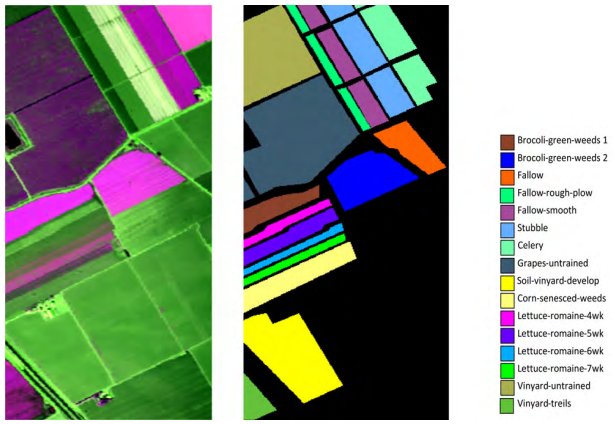
The first experimental dataset considered was the Indian Pines image, which was gathered by an airborne visible infra-red imaging spectrometer (AVIRIS) sensor over the Indian Pines test site in Northwestern Indiana. It comprised 145 × 145 pixels, and its spatial resolution was 20 m. The original image possessed 224 spectral reflectance bands, and we reduced it to 200 by removing the bands covering the region of water absorption. As the scene was captured in June, some of the crops present, such as the corn and the soybeans, were in their early stages of growth with less than 5% coverage. The ground truth available was designated into sixteen classes. Fig. 3(a) and (b) illustrate a false-color image of the Indian Pine and the ground truth information, respectively. Each color in the map of the ground truth represents a different object.

The second experimental data set was the Salinas image, captured over the Salinas Valley, California, USA by a 224-band AVIRIS sensor, which was characterized by its high spatial resolution of 3.7 m. The area covered comprised of 512 lines by 217 samples. As with the Indian Pines scene, we discarded the 20 water absorption bands. Figure 4(a) shows the false-color image of the Salinas image synthesized by three bands, and Figure 4(b) shows the corresponding ground truth with 16 different classes.

The third experimental dataset was the University of Pavia image, which was captured by a German-made reflective optical system imaging spectrometer (ROSIS-3) sensor during a flight campaign over Pavia, Northern Italy. The spatial resolution of this image was 1.3 m, which was the highest among all the three datasets. This image was also the largest, consisting of 610 × 340 pixels. After removing the 12 noisy bands, the remaining 103 bands were used as the experimental dataset. The three-band false color image and the ground truths that differentiate nine classes are presented in Fig. 5.

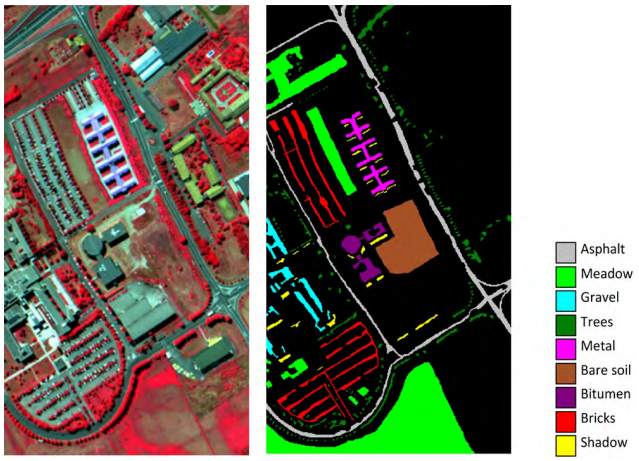
B. EXPERIMENTAL PARAMETER ANALYSIS

1) The parameter values of the edge-preserving filter were observed to be consistent with the values in [34].



(a) (b)

FIGURE 4. Salinas dataset. (a) Three-band color composite. (b) Reference data with 16 land cover classes.



(a) (b)

FIGURE 5. University of Pavia data set. (a) Three-band color composite. (b) Reference data with 16 land cover classes.

Therefore, the value of the first parameter set is defined as follows:

$$\delta_s^1 = \mathbf{30}, \quad \delta_r^1 = \mathbf{0.3}, \tag{21}$$

and the value of the second parameter set is defined as follows:

$$\delta_s^2 = \mathbf{115}, \quad \delta_r^2 = \mathbf{0.6}. \tag{22}$$

The value of the last parameter set is defined as follows:

$$\delta_s^3 = \mathbf{200}, \quad \delta_r^3 = \mathbf{0.9}. \tag{23}$$

These three sets of parameters were used together to extract useful spatial information features from the original image.

2) The values of the parameters N (band number of the dimension-reduced original hyperspectral image) and L (dimension of the dimension-reduced MEPFs) can

TABLE 2. The comparative results of the PCA and LPP dimension reduction methods for the Indian Pines Classification accuracies of the SVM, BCC, EPF, GCK, MFL, PCA-EPFs, and the MFL-EPFs methods for the Indian Pines dataset with 1% training samples.

Class	Train/Test	SVM	BCC	EPF	GCK	MFL	PCA-EPFs	MFL-EPFs
Alfalfa	6/40	27.98(1.34)	81.00(11.7)	48.87(30.4)	95.23(4.84)	95.74(1.22)	**99.52(1.51)**	98.25(1.20)
Corn_N	7/1421	47.19(8.50)	36.93(13.3)	62.06(16.4)	65.27(10.4)	63.29(6.81)	**74.95(10.5)**	71.30(6.73)
Corn_M	6/824	40.81(7.47)	42.43(8.55)	61.73(14.9)	79.41(9.75)	66.02(6.35)	72.34(12.6)	**83.03(12.1)**
Corn	6/231	26.49(6.65)	46.97(9.06)	40.41(13.8)	85.27(6.41)	73.02(12.9)	76.78(12.8)	**93.85(9.50)**
Grass_M	6/477	59.66(17.9)	66.17(8.48)	79.83(22.0)	82.51(6.12)	77.14(9.16)	**89.66(12.7)**	89.46(9.12)
Grass_T	6/724	78.89(4.53)	67.35(8.90)	80.94(7.78)	96.53(2.22)	92.33(5.67)	92.75(5.67)	**97.27(3.25)**
Grass_P	6/22	26.49(15.1)	84.55(11.8)	59.71(32.8)	97.50(7.91)	96.82(2.20)	84.30(26.7)	**100.0(0.00)**
Hay_W	7/471	93.66(4.77)	72.56(11.1)	98.34(5.24)	99.19(0.48)	98.54(2.71)	99.98(0.07)	**100.0(0.00)**
Oats	6/14	10.20(5.07)	82.14(16.2)	32.80(27.7)	100.0(0.00)	100.0(0.00)	86.86(29.3)	**100.0(0.00)**
Soybean_N	7/965	43.07(7.97)	50.80(13.6)	56.30(11.6)	77.07(6.68)	75.16(10.5)	68.53(18.3)	**83.67(9.63)**
Soybean_M	8/2447	62.56(6.33)	45.51(7.90)	74.62(6.57)	72.61(6.04)	75.04(15.1)	**91.55(3.86)**	80.18(5.71)
Soybean_C	6/587	26.87(7.24)	25.30(6.70)	39.31(15.5)	79.67(9.00)	72.85(13.2)	**87.70(10.5)**	83.99(10.1)
Wheat	6/199	78.33(3.02)	89.44(10.2)	95.02(4.56)	99.36(0.41)	99.45(0.16)	**100.0(0.00)**	99.94(0.15)
oods	6/1259	84.28(5.48)	70.08(14.1)	91.13(7.49)	90.02(4.56)	86.99(6.43)	**98.65(1.51)**	93.39(8.49)
Buildings	6/380	28.94(8.30)	28.71(10.7)	55.13(19.5)	78.08(7.83)	72.50(6.03)	**94.33(8.32)**	89.79(10.9)
Stone	7/86	83.64(19.6)	88.28(6.08)	79.38(8.90)	88.94(6.23)	91.02(6.24)	**93.79(9.21)**	90.80(7.32)
OA		52.42(2.94)	51.13(3.11)	65.91(5.25)	79.91(2.15)	76.12(3.02)	83.57(3.60)	**85.22(2.37)**
AA		51.19(2.77)	61.14(2.60)	65.97(6.51)	86.67(1.50)	83.24(1.55)	88.23(1.47)	**90.93(1.82)**
Kappa		46.67(3.37)	45.26(3.15)	61.50(6.12)	77.35(2.42)	73.03(3.29)	81.41(4.00)	**83.25(2.69)**

affect the classification accuracy. However, as the randomness of the training samples caused the classification results to float, it was difficult to determine the optimal values of N and L. This will be considered one of the subsequent research directions. Therefore, by the rule of thumb, the value of N was considered 38 and that of L was 48 to achieve the desired results.

C. COMPARISON ALGORITHM

We compared our algorithm with some highly cited studies and outstanding methods in the field to establish its advantages. The methods used for comparison included the classical algorithm SVM [9], which is one of the most classic hyperspectral classification methods, an unsupervised feature extraction method based on band correlation(BCC) [17], single parameter EPF-based method [22], GCK-based method [30], MFL-based method [32], and the PCA-based edge-preserving filtering (PCA-EPFs) [34] method combining the MEPFs and the PCA. The SVM method was implemented in the LIBSVM library using the radial function kernel, and five cross-validations were performed. The parameter values of other methods were consistent with those provided in the literature. The computer used in this experiment was a 2.2 GHz dual-core processor with a memory capacity of 8 GB.

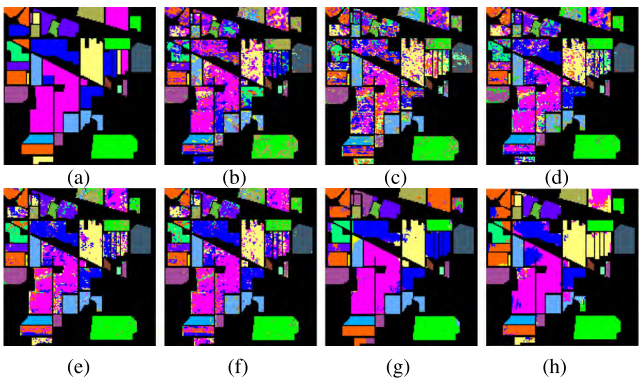


FIGURE 6. Classification maps of different methods obtained from the Indian Pines data set. (a) Original ground truth of the Indian Pines. Classification maps obtained by (b) SVM, OA = 52.40% (c) BBC, OA = 50.85% (d) EPF, OA = 65.16% (e) GCK, OA = 79.03% (f) MFL, OA = 77.42% (g) PCA-EPFs, OA = 82.49% (h) MFL-EPFs, OA = 85.82%.

First, we conducted experiments on the Indian Pines dataset. One hundredth of the ground truth data was randomly designated as the training sample for ten independent

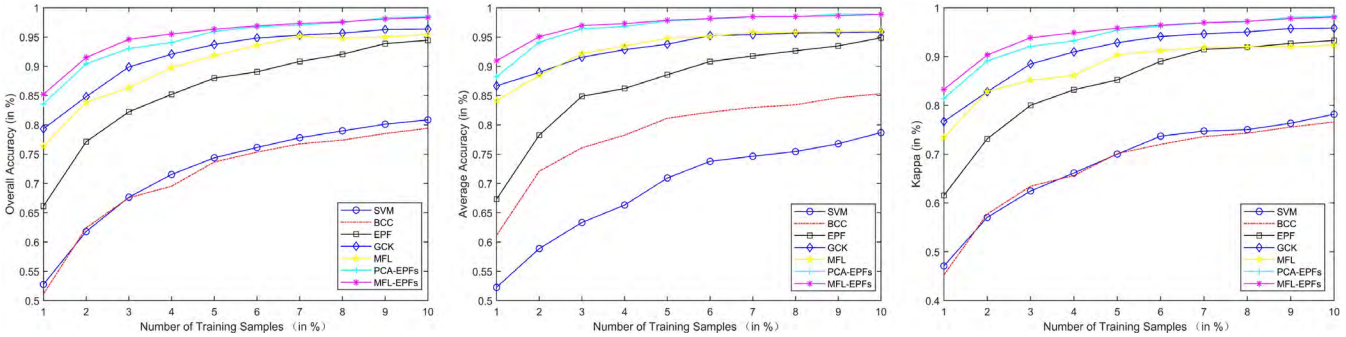
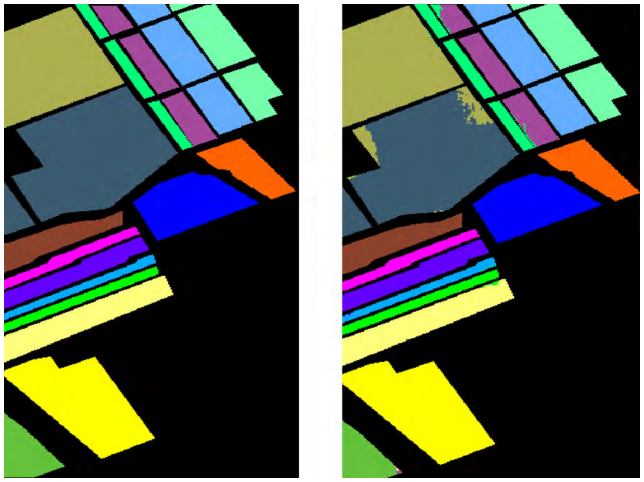


FIGURE 7. Classification accuracies of the compared methods on the Indian Pines data set against different number of training samples (variations from 1–10%).



(a) (b)

FIGURE 8. Classification accuracies of the compared methods on the Indian Pines data set against different number of training samples (variations from 1–10%).

Monte Carlo tests. Then average and the standard deviation of these ten results were obtained. These values are listed in Table 2. It is observed from the table that the algorithm proposed in this study is better than the other algorithms, as it demonstrates the highest OA, AA, and Kappa coefficient

TABLE 3. Computing time of the PCA-EPFs and MFL-EPFs methods for the Indian Pines.

Number of training samples (%)	5	6	7	8	9	10
Computing time for PCA_EPFs (s)	10.584	12.808	16.284	19.712	23.239	36.50
Computing time for MFL-EPFs (s)	5.451	5.441	5.632	5.964	5.681	5.950

when the training samples are small. Especially for Grass_P (grass/ranch), Hay_W (hay/feed), and Oats, the classification accuracy is up to 100%. Soybean_N (uncultivated soybean land) and Corn, which are generally low in classification accuracy, can achieve a classification accuracy of over 80% and 90%, respectively. The OA of each of the classification maps is reported in Fig. 6. It can be observed that the maps of the proposed algorithm are similar to the ground truth with less rate of noises, indicating that the extracted spectral information and spatial information are representative. Fig. 7 shows the variation trend of the classification accuracy values concerning the training samples. It is observed from this figure that the classification accuracy increases with an increase in the number of training samples. When the number of training samples reaches 10%, the accuracy reaches as high as 99%. It is evident from this figure that the classification

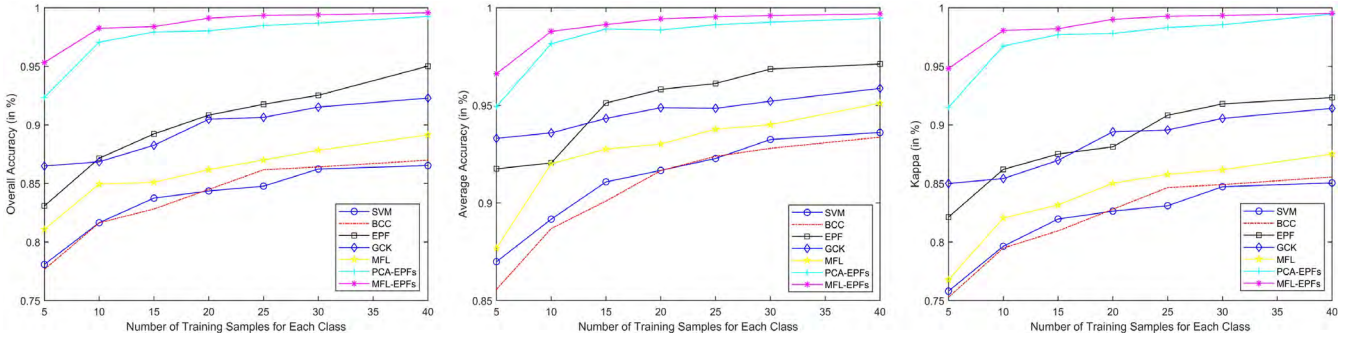


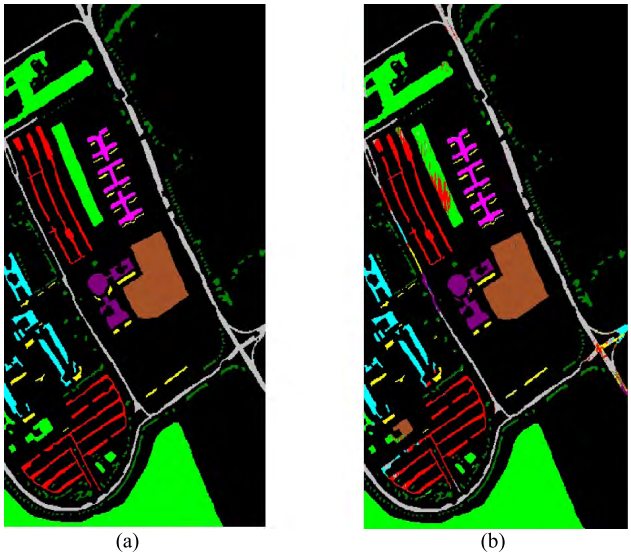
FIGURE 9. Classification accuracies of the compared methods on the Salinas dataset with the different number of labeled samples per class (Variations from 5–40).

TABLE 4. Classification accuracies of the SVM, BCC, EPF, GCK, MFL, PCA-EPFs, and the MFL-EPFs methods for the Salinas dataset with ten labeled samples per class as the training set.

Class	Train/Test	SVM	BCC	EPF	GCK	MFL	PCA-EPFs	MFL-EPFs
Weeds_1	10/1999	98.96(1.10)	98.13 (0.92)	**100.0 (0.00)**	95.23(4.84)	99.85(0.06)	**100.0 (0.00)**	99.98(0.04)
Weeds_2	10/3716	98.92(0.41)	97.10 (2.27)	99.93 (0.19)	65.27(10.4)	99.58 (0.13)	**99.79(0.26)**	99.71(0.57)
Fallow	10/1966	87.44(3.64)	82.72 (9.88)	92.51(4.15)	79.41(9.75)	99.79(0.39)	98.58.(2.74)	**100.0 (0.00)**
Fallow_P	10/1384	96.75(1.04)	98.56(0.72)	97.39(0.33)	85.27(6.41)	96.62(2.65)	95.81(1.94)	**98.65(2.07)**
Fallow_S	10/2668	**98.30(1.18)**	93.91 (3.70)	99.91(0.07)	82.51(6.12)	98.65(0.07)	99.99(0.02)	96.05 (3.83)
Stubble	10/3949	**99.98(0.03)**	98.07 (2.87)	99.98(0.02)	96.53(2.22)	99.17(0.72)	99.74(0.28)	99.66(0.30)
Celery	10/3569	96.49(2.85)	98.75 (0.87)	97.47(3.02)	97.50(7.91)	99.75(0.07)	99.50(0.91)	**99.85(0.07)**
Grapes	10/11261	68.51(5.91)	54.63(13.2)	79.19(11.1)	99.19(0.48)	61.84(12.4)	**99.43(0.57)**	96.06(2.79)
Soil	10/6193	98.70(0.86)	96.65 (1.71)	99.26(0.42)	**100.0(0.00)**	99.66(0.31)	99.57(0.40)	**100.0(0.00)**
Corn	10/3268	78.68(7.55)	75.32(8.73)	87.62(5.52)	77.07(6.68)	87.93(6.36)	**99.73(0.33)**	98.72 (1.08)
Lettuce_4	10/1058	85.43(6.45)	85.84 (4.85)	94.29(4.75)	72.61(6.04)	90.95(1.90)	99.83(0.19)	**99.98(0.06)**
Lettuce_5	10/1917	93.64(5.67)	96.89 (2.04)	97.67(5.78)	79.67(9.00)	99.78(0.11)	**99.32(1.17)**	98.96(2.21)
Lettuce_6	10/906	89.82(5.32)	98.27 (0.41)	97.49(1.92)	99.36(0.41)	99.18(0.30)	97.05(4.78)	**98.45(1.43)**
Lettuce_7	10/1060	84.07(15.5)	88.20 (3.93)	95.14(6.43)	90.02(4.56)	89.06(3.29)	96.32(5.90)	**98.80(0.71)**
Lettuce_U	10/7258	49.67(5.81)	64.21 (10.9)	61.36(14.5)	78.08(7.83)	56.24(9.81)	85.95(9.23)	**96.73 (3.63)**
Lettuce_T	10/1797	90.87(7.96)	91.51 (3.32)	97.82(4.54)	88.94 (6.23)	91.98(7.53)	**99.98(9.21)**	98.94 (0.44)
OA		82.15(2.94)	81.46(2.31)	87.58(5.21)	79.91 (2.15)	84.43(1.92)	97.06(2.19)	**98.25(0.48)**
AA		88.51(1.47)	88.69(1.53)	93.56(1.89)	86.67 (1.50)	91.88(0.87)	98.16(1.05)	**98.79(0.38)**
Kappa		80.23(3.18)	79.46(2.51)	86.25(5.72)	77.35 (2.42)	82.69(2.09)	96.74(2.42)	**98.06(0.54)**

accuracy of the presented algorithm is comparable to that of the PCA-EPFs algorithm, when the number of training samples is large. The reason is that the advantage of this algorithm on Grass_P (grass/pasture) and Oats disappears with an increase in the number of samples. When the number of training samples increases to a boundary, the training samples of these two objects are 100%. Therefore, the accuracy values of these two compared algorithms tend to be consistent. However, the algorithm proposed in this study demonstrates an advantage over the PCA-EPFs algorithm in terms of time consumption. It is observed from Table 3 that the PCA-EPFs algorithm consumes a considerable amount of time, especially when there is an increase in the number of training samples. In contrast, the proposed algorithm does not demonstrate such a drawback, which provides a significant advantage.

Next, we verified the effectiveness of the proposed algorithm in the Salinas image dataset. Only 10 samples of each class were used as the training set. The results obtained are shown in Table 4. Although the number of training samples is limited, the classification accuracy is observed to be over 98%. Figure 8 presents the classification maps on the Salinas dataset. It is evident from this figure that the data obtained by the proposed algorithm is only marginally different from the original ground truth data. The variance of this algorithm is the lowest among all the algorithms used for comparison, indicating that the algorithm demonstrates better stability, when the number of training samples is limited. In particular, the classification accuracy of the Vinyard_U (unproductive vineyard) far exceeds those of



(a) (b)

FIGURE 10. Classification maps on the University of Pavia dataset. (a) Original ground truth of the University of Pavia (b) obtained by MFL-EFPs, OA = 94.19%.

TABLE 5. Classification accuracies of the SVM, BCC, EPF, GCK, MFL, PCA-EPFs, and the MFL-EPFs methods for the University of Pavia dataset with twenty labeled samples per class as the training set.

Class	Train/Test	SVM	BCC	EPF	GCK	MFL	PCA-EPFs	MFL-EPFs
Asphalt	20/6611	93.29(2.46)	71.22 (4.73)	**96.94 (1.67)**	87.38(4.73)	80.19(2.12)	88.46(5.29)	88.72(4.26)
Meadows	20/18629	88.51(1.74)	70.41 (6.33)	92.43 (3.15)	93.96(2.27)	84.49(3.27)	**94.43(4.25)**	93.94(4.31)
Gravel	20/2079	57.31(7.63)	74.18 (4.32)	81.88(13.2)	85.10 (4.15)	67.03(9.05)	95.85.(4.96)	**97.19 (2.37)**
Trees	20/3044	71.85(11.8)	87.94 (3.73)	78.61(14.6)	87.42(3.88)	94.28(2.42)	92.22(2.46)	**96.28(1.31)**
Sheets	20/1325	92.68(4.79)	99.18 (0.27)	94.37(4.14)	93.70(5.41)	98.35(1.01)	**99.97(0.05)**	99.26 (0.44)
Soil	20/5009	48.79(7.25)	73.92 (8.85)	61.77(10.2)	82.76(5.10)	76.63(4.93)	**99.47(0.61)**	98.40**(2.30)**
Bitumen	20/1310	49.39(4.60)	88.90 5.47)	71.67(9.42)	88.04(5.65)	89.83(4.18)	96.33(3.40)	**99.85(0.23)**
Bricks	20/3662	77.27(4.61)	74.68 (4.69)	86.85(6.43)	81.51(6.60)	63.28(8.71)	91.41(3.12)	**93.64(5.11)**
Shadows	20/927	**99.88(0.12)**	99.83 (0.33)	98.25(1.15)	99.33(0.38)	98.22(2.19)	91.56(5.08)	98.36(1.73)
OA		75.66(3.31)	81.46(2.31)	84.70(4.11)	89.58 (1.55)	87.38(2.17)	93.92(2.36)	**94.40(2.08)**
AA		75.44(2.39)	88.69(1.53)	84.75(3.81)	88.81 (1.02)	89.25(2.02)	94.42(1.31)	**96.19(0.88)**
Kappa		68.96(3.88)	79.46(2.51)	80.26(5.07)	86.22 (2.00)	83.59(2.68)	92.05(3.01)	**92.68(2.67)**

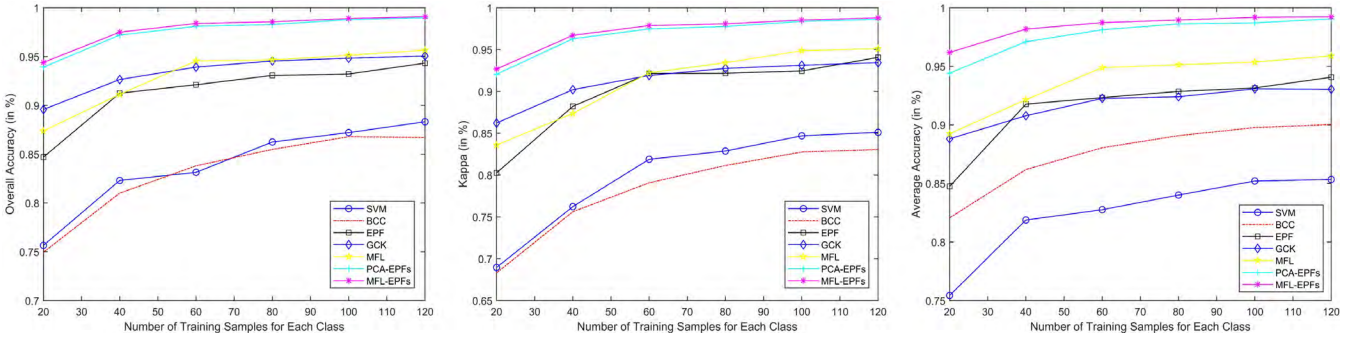


FIGURE 11. Classification accuracies of the compared methods on the University of Pavia dataset with the different number of labeled samples per class (variations from 20–120).

other comparison algorithms. Fig. 9 shows the OA, AA, and Kappa coefficient as functions of the training size for samples per class (variations 5–40).

Finally, the proposed algorithm was applied to the University of Pavia dataset. Each class used only 20 samples as the training set. The overall classification accuracy was observed to be as high as 94%. The accuracies of different methods with 20 labeled samples per class as the training set are summarized in Table 5. The comparison between the original ground truth data and the classification map of the algorithm is shown in Fig. 10. The two figures appear very similar, except for some minor details. Fig. 11 shows a plot of the classification accuracy, as the sample increases from 20 to 140. Overall, the accuracy of the proposed algorithm is observed to be superior to the other algorithms.

V. CONCLUSION

In this study, a hyperspectral classification algorithm combining multiple edge-preserving filtering features and multiple feature learning is proposed. The LPP-based multiple edge-preserving filtering method was employed to extract the spatial features of different scales, while the band clustering method was used to extract the spectral features, which represent the nonlinear and linear features of the hyperspectral images respectively. These two features determined the classification results using the MFL. Based on the aforementioned analyses, it was confirmed that the proposed algorithm demonstrated advantages in terms of classification accuracy compared to other algorithms, especially when the number of training samples was relatively limited. In addition, there was no increase in the execution time, as the number of samples was increased. However, the LPP algorithm adopted for the purpose of dimension reduction in this study could accurately obtain the local structure of the neighborhood. It was not capable of extracting the inherent structure in the original image, resulting in the loss of some useful information. Besides, the dimensionality $\boldsymbol{N}$ of the dimension-reduced original hyperspectral image and the dimensionality $\boldsymbol{L}$ of the dimension-reduced MEPFs were considered the empirical and not the optimal values for each dataset.

REFERENCES

[1] H. Lu, Y. Li, S. Mu, D. Wang, H. Kim, and S. Serikawa, "Motor anomaly detection for unmanned aerial vehicles using reinforcement learning," *IEEE Internet Things J.*, vol. 5, no. 4, pp. 2315–2322, Aug. 2018. doi: 10.1109/JIOT.2017.2737479.

[2] B. Zhang, D. Wu, L. Zhang, Q. Jiao, and Q. Li, "Application of hyperspectral remote sensing for environment monitoring in mining areas," *Environ. Earth Sci.*, vol. 65, no. 3, pp. 649–658, Feb. 2012. doi: 10.1007/s12665-011-1112-y.

[3] H. Lu, D. Wang, Y. Li, J. Li, X. Li, H. Kim, S. Serikawa, and I. Humar, "CONet: A cognitive ocean network," *IEEE Wireless Commun.*, vol. 26, no. 3, pp. 90–96, Jun. 2019.

[4] M. Dalponte, H. O. Orka, T. Gobakken, D. Gianelle, and E. Naesset, "Tree species classification in boreal forests with hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2632–2645, May 2013. doi: 10.1109/TGRS.2012.2216272.

[5] A. E. Maxwell, T. A. Warner, and F. Fang, "Implementation of machine-learning classification in remote sensing: An applied review," *Int. J. Remote Sens.*, vol. 39, no. 9, pp. 2784–2817, 2018. doi: 10.1080/01431161.2018.1433343.

[6] P. Ghamisi, J. Plaza, Y. Chen, J. Li, and A. J. Plaza, "Advanced spectral classifiers for hyperspectral images: A review," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 1, pp. 8–32, Mar. 2017. doi: 10.1109/MGRS.2016.2616418.

[7] L. He, J. Li, C. Liu, and S. Li, "Recent advances on spectral–spatial hyperspectral image classification: An overview and new guidelines," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1579–1597, Mar. 2018. doi: 10.1109/TGRS.2017.2765364.

[8] L. Wang, J. Zhang, P. Liu, K.-K. R. Choo, and F. Huang, "Spectral–spatial multi-feature-based deep learning for hyperspectral remote sensing image classification," *Soft Comput.*, vol. 21, no. 1, pp. 213–221, Jan. 2017. doi: 10.1007/s00500-016-2246-3.

[9] M. Pal, and G. M. Foody, "Evaluation of SVM, RVM and SMLR for accurate image classification with limited ground data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 5, pp. 1344–1355, Oct. 2012. doi: 10.1109/JSTARS.2012.2215310.

[10] M. Belgiu and L. Drăgut, "Random forest in remote sensing: A review of applications and future directions," *ISPRS J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, Apr. 2016. doi: 10.1016/j.isprsjprs.2016.01.011.

[11] P. M. Atkinson and A. R. L. Tatnall, "Introduction neural networks in remote sensing," *Int. J. Remote Sens.*, vol. 18, no. 4, pp. 699–709, Nov. 2010. doi: 10.1080/014311697218700.

[12] H. Lu, Y. Li, M. Chen, H. Kim, and S. Serikawa, "Brain intelligence: Go beyond artificial intelligence," *Mobile Netw. Appl.*, vol. 23, no. 2, pp. 368–375, Apr. 2018. doi: 10.1007/s11036-017-0932-8.

[13] J. Ren, J. Zabalza, S. Marshall, and J. Zheng, "Effective feature extraction and data reduction in remote sensing using hyperspectral imaging [applications corner]," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 149–154, Jul. 2014. doi: 10.1109/MSP.2014.2312071.

[14] G. Licciardi, P. R. Marpu, J. Chanussot, and J. A. Benediktsson, "Linear versus nonlinear PCA for the classification of hyperspectral data based on the extended morphological profiles," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 447–451, May 2012. doi: 10.1109/LGRS.2011.2172185.

[15] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011. doi: 10.1109/TGRS.2011.2153861.

[16] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Comput.*, vol. 12, no. 10, pp. 2385–2404, 2000. doi: 10.1162/089976600300014980.

[17] A. Ghorbanian and A. Mohammadzadeh, "An unsupervised feature extraction method based on band correlation clustering for hyperspectral image classification using limited training samples," *Remote Sens. Lett.*, vol. 9, no. 10, pp. 982–991, Aug. 2018. doi: 10.1080/2150704X.2018.1500723.

[18] X. Kang, S. Li, and J. A. Benediktsson, "Feature extraction of hyperspectral images with image fusion and recursive filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3742–3752, Jun. 2014. doi: 10.1109/TGRS.2013.2275613.

[19] M. D. Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010. doi: 10.1109/TGRS.2010.2048116.

[20] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005. doi: 10.1109/TGRS.2004.842478.

[21] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017. doi: 10.1109/TGRS.2017.2710145.

[22] X. Kang, S. Li, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014. doi: 10.1109/TGRS.2013.2264508.

[23] S. Serikawa and H. Lu, "Underwater image dehazing using joint trilateral filter," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 41–50, 2014. doi: 10.1016/j.compeleceng.2013.10.016.

[24] 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. doi: 10.1109/TGRS.2016.2584107.

[25] H. Lu, Y. Li, T. Uemura, H. Kim, and S. Serikawa, "Low illumination underwater light field images reconstruction using deep convolutional neural networks," *Future Gener. Comput. Syst.*, vol. 82, pp. 142–148, May 2018. doi: 10.1016/j.future.2018.01.001.

[26] G. Camps-Valls, L. Gomez-Chova, J. Munoz-Mari, J. Vila-Frances, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006. doi: 10.1109/LGRS.2005.857031.

[27] Y. Zhou, J. Peng, and C. L. P. Chen, "Extreme learning machine with composite kernels for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2351–2360, Jun. 2015. doi: 10.1109/JSTARS.2014.2359965.

[28] Y. Gu, C. Wang, D. You, Y. Zhang, S. Wang, and Y. Zhang, "Representative multiple kernel learning for classification in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 7, pp. 2852–2865, Jul. 2012. doi: 10.1109/TGRS.2011.2176341.

[29] A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," *J. Mach. Learn. Res.*, vol. 9, pp. 2491–2521, Nov. 2008.

[30] J. Li, P. R. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, Sep. 2013. doi: 10.1109/TGRS.2012.2230268.

[31] D. Böhning, "Multinomial logistic regression algorithm," *Ann. Inst. Stat. Math.*, vol. 44, no. 1, pp. 197–200, Mar. 1992. doi: 10.1007/BF00048682.

[32] J. Li, X. Huang, P. Gamba, J. M. Bioucas-Dias, L. Zhang, and J. A. Benediktsson, "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, Mar. 2015. doi: 10.1109/TGRS.2014.2345739.

[33] L. Fang, C. Wang, S. Li, and J. A. Benediktsson, "Hyperspectral image classification via multiple-feature-based adaptive sparse representation," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1646–1657, Jul. 2017. doi: 10.1109/TIM.2017.2664480.

[34] X. Kang, X. Xiang, S. Li, and J. A. Benediktsson, "PCA-based edge-preserving features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7140–7151, Dec. 2017. doi: 10.1109/TGRS.2017.2743102.

[35] X. He and P. Niyogi, "Locality preserving projections," in *Advances in neural information processing systems*. Cambridge, MA, USA: MIT Press, 2003, pp. 153–160.

[36] X. Li, J. Pan, Y. He, and C. Liu, "Bilateral filtering inspired locality preserving projections for hyperspectral images," *Neurocomputing*, vol. 164, pp. 300–306, Sep. 2015. doi: 10.1016/j.neucom.2015.01.021.

[37] K. He, J. Sun, and X. Tang, "Guided image filtering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 6, pp. 1397–1409, Jun. 2013. doi: 10.1109/TPAMI.2012.213.

[38] M. A. Turk and A. P. Pentland, "Face Recognition Using eigenfaces," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 1991, pp. 586–591.

[39] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005. doi: 10.1109/TPAMI.2005.55.

[40] D. Cai, X. He, and J. Han, "Document clustering using locality preserving indexing," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 12, pp. 1624–1637, Dec. 2005. doi: 10.1109/TKDE.2005.198.

[41] D. Hu, G. Feng, and Z. Zhou, "Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition," *Pattern Recognit.*, vol. 40, no. 1, pp. 339–342, 2007. doi: 10.1016/j.patcog.2006.06.022.
[42] S. Wang, D. Xie, F. Chen, and Q. Gao, "Dimensionality reduction by LPP-L21," *IET Comput. Vis.*, vol. 12, no. 5, pp. 659–665, Aug. 2018. doi: 10.1049/iet-cvi.2017.0302.
[43] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011. doi: 10.1109/TGRS.2011.2128330.
[44] B. Bigdeli, F. Samadzadegan, and P. Reinartz, "Band grouping versus band clustering in SVM ensemble classification of hyperspectral imagery," *Photogramm. Eng. Remote Sens.*, vol. 76, no. 6, pp. 523–533, Jun. 2013. doi: 10.14358/PERS.79.6.523.
[45] J.-H. Cai, X.-J. Zhao, S.-W. Sun, J.-F. Zhang, and H.-F. Yang, "Stellar spectra association rule mining method based on the weighted frequent pattern tree," *Res. Astron. Astrophys.*, vol. 13, no. 3, pp. 334–342, Mar. 2013.

WEI TIAN received the B.S. degree in electronic and information engineering from the Changsha University of Science and Technology, Changsha, China, the M.S. degree in signal and information processing from Yangzhou University, Yangzhou, China. She is currently pursuing the Ph.D. degree in information and communication engineering from Hohai University, Nanjing, China.

Her current research interests include image processing, remote sensing, and computer vision.

LIZHONG XU received the Ph.D. degree in control science and engineering from the China University of Mining and Technology, Xuzhou, China, in 1997.

He is currently a Professor with the College of Computer and Information, and the Director of the Institute of Communication and Information System Engineering, Hohai University, Nanjing, China. He is the author of five books, more than 150 articles, and more than ten inventions. His current research interests include signal processing in remote sensing and remote control, information processing system and its applications, system modeling, and system simulation.

Dr. Xu is a Senior Member of the Chinese Institute of Electronic, and China Computer Federation.

ZHE CHEN received the Ph.D. degree in computer application technology from Hohai University, Nanjing, China.

He is currently an Associate Professor with the College of Computer and Information, Hohai University, and also a Senior Researcher with the Center of Marine Monitoring Equipment and Data Processing Engineering. His research interests include optical imaging, image processing, pattern recognition, and complex systems.

AIYE SHI received the M.S. degree from the Nanjing University of Science and Technology, in 2002, and the Ph.D. degree in optical engineering and hydroinformatics from Hohai University, China, in 2009.

He is currently an Associate Professor with the College of Computer and Information Engineering, Hohai University. His current research interest includes image change detection.