On Combining Multiple Features for Hyperspectral Remote Sensing Image Classification

Lefei Zhang, *Student Member, IEEE*, Liangpei Zhang, *Senior Member, IEEE*, Dacheng Tao, *Member, IEEE*, and Xin Huang

Abstract—In hyperspectral remote sensing image classification, multiple features, e.g., spectral, texture, and shape features, are employed to represent pixels from different perspectives. It has been widely acknowledged that properly combining multiple features always results in good classification performance. In this paper, we introduce the patch alignment framework to linearly combine multiple features in the optimal way and obtain a unified low-dimensional representation of these multiple features for subsequent classification. Each feature has its particular contribution to the unified representation determined by simultaneously optimizing the weights in the objective function. This scheme considers the specific statistical properties of each feature to achieve a physically meaningful unified low-dimensional representation of multiple features. Experiments on the classification of the hyperspectral digital imagery collection experiment and reflective optics system imaging spectrometer hyperspectral data sets suggest that this scheme is effective.

Index Terms—Classification, dimensional reduction, hyperspectral, multiple features.

I. INTRODUCTION

HYPERSPECTRAL imaging (HSI), airborne or space-borne, can acquire data for ground objects, with a larger spectral range and a much higher spectral resolution [1]. In hyperspectral remote sensing image analysis, it is common to perform feature extraction [2]–[9] before classification. It is also important to employ multiple features of different types to represent a pixel's information, e.g., spectral, texture, and shape features. To well characterize a pixel, for effective and accurate classification, it is essential to find a way to encode these spatial and spectral features into a low-dimensional vector. Jimenez *et al.* [10] presented a method of unsupervised enhancement of pixels' homogeneity in a local neighborhood by integrating the spectral and spatial information, producing results that are more meaningful to the human analyst. Stathakis *et al.* [11] explored a genetic algorithm that is used to select a near-optimal subset

of input dimensions using a feedforward multilayer perceptron. Bruzzone *et al.* [12] presented a novel approach to feature selection for the classification, which aims at selecting a subset of the original set of features that exhibits, at the same time, high capability to discriminate between the considered classes and high invariance in the spatial domain of the investigated scene. Jimenez *et al.* [13] also presented an analysis and a comparison of different linear unsupervised feature extraction methods applied to high-dimensional data and their impact on classification. Another much simpler approach was suggested in [14], where the authors proposed a set of best-bases feature extraction techniques that intelligently combine subsets of adjacent bands into a smaller number of features; this approach is highly effective for classification of hyperspectral data. In addition, Rellier *et al.* [15] performed a joint texture analysis in both discrete spaces by a probabilistic vector texture model using a Gauss–Markov random field (MRF) with parameters to allow the characterization of different hyperspectral textures.

In this paper, the proposed multiple feature combining (MFC) is based on manifold learning [16], [17] and a patch alignment framework [18]. The manifold-learning-based dimensional reduction algorithms seek a low-dimensional representation, which preserves certain local geometric properties of the original feature, e.g., locally linear embedding (LLE) [19] finds an optimal linear reconstruction in a small neighborhood, local tangent space alignment (LTSA) computes a local-geometry-preserving representation of the original data [20], [21], isometric feature mapping (ISOMAP) [22] tries to preserve the geodesic distances, and Laplacian eigenmaps (LE) [23] preserve proximity relationships by manipulation of an undirected weighted graph. In remote sensing, Ma *et al.* [24] investigated an approach to combine local manifold learning and the k-nearest neighbor classifier for hyperspectral remote sensing image classification. However, the existing methods assume that samples are distributed in only one feature space, and thus, they could not deal with directly combining multiple features. To overcome this point, we focus on encoding different features that have particular statistical properties into a low-dimensional vector by an MFC framework, which can achieve a physically meaningful combination. Experiments on hyperspectral remote sensing image classification demonstrate the effectiveness of the proposed approach.

The remainder of this paper is organized as follows. In Section II, we provide the proposed MFC framework in detail, including the extraction and combination of multiple features. Then, the experiments on two publicly available hyperspectral data sets are reported in Section III, followed by the conclusion.

Fig. 1. Flowchart of the proposed MFC framework.

II. MFC FRAMEWORK

The proposed MFC framework can be divided into two main components, as shown in Fig. 1. In the first step, three kinds of features of HSI are introduced. Then, the MFC algorithm, which finds the particular contribution of each feature to the unified representation, is employed to obtain the final low-dimensional representation.

A. Multiple Feature Extraction

In this framework, three kinds of features are introduced to the MFC. Each feature is represented as a single vector $v \in R^L$.

1) Spectral Feature: The spectral feature of a pixel is obtained by arranging its digital number (DN) in all of the l bands

$$v_{\text{Spectral}} = [v_1, v_2, \ldots, v_l]^{\text{T}} \in R^l \tag{1}$$

in which v_k denotes the DN in band k.

2) Texture Feature: Various methods have been reported to extract the texture feature from the image region. A gray-level co-occurrence matrix [25] starts from a pixel in a given position and then provides a measure of the probability of occurrence of two gray levels separated by a given distance in a given direction. A new model for extracting spectral/spatial information, based on 3-D Gabor filters, which captures the specific orientation, scale, and wavelength-dependent properties of hyperspectral image data, could provide an efficient means of texture feature representation [26], [27]. Consider that the conventional 2-D Gabor wavelet has already been proved superior for representing the texture features in the natural scene and aerial photographs [28]–[31]. In this paper, we performed a 2-D Gabor wavelet on a top principal component (PC) image, which is given by implementing the PC [32] transformation of hyperspectral images, to extract the texture feature of HSI.

A Gabor wavelet is a filter whose impulse response is defined by an elliptical Gaussian envelope and a complex plane wave. The generalized 2-D Gabor function can be defined as [33]

$$G_{s,d}(x,y) = G_{\vec{\kappa}}(\overline{x}) = \frac{\|\vec{\kappa}\|}{\delta^2} \cdot e^{-\frac{\|\vec{\kappa}\|^2 \cdot \|\overline{x}\|^2}{2\delta^2}} \cdot \left[e^{i \cdot \vec{\kappa} \cdot \overline{x}} - e^{-\frac{\delta^2}{2}} \right] \tag{2}$$

where $\overline{x} = (x, y)$ is the spatial domain variable and the frequency vector $\vec{\kappa} = (\pi/2 f^s) \cdot e^{i \cdot (\pi d/8)}$, in which $f = 2$, $s = 0, 1, \ldots, 4$, and $d = 0, 1, \ldots, 11$, determines the 5 scales and 12 directions of the Gabor function. The number of oscillations under the Gaussian envelope is determined by $\delta = 2\pi$. The Gabor texture feature image in a specific scale and direction is the magnitude part of convolving the image $I(x, y)$ with the Gabor function of corresponding parameters s and d

$$F_{s,d}(x,y) = G_{s,d}(x,y) * I(x,y). \tag{3}$$

The texture feature of a pixel (x, y) is obtained by

$$v_{\text{Texture}} = [F_{1,1}(x,y), \ldots, F_{s,d}(x,y)]^{\text{T}} \in R^{sd}. \tag{4}$$

3) Shape Feature: Several effective shape features have been proposed, e.g., [34] and [35] proposed an effective length–width extraction algorithm to extract the length and width of spectrally similar connected groups of pixels, and [36]–[38] presented a series of morphological profiles (MPs) and differential MPs (DMPs), which have been proven to be effective in extracting multiscale structural information from HSIs. In this paper, based on our previous work, the pixel shape index method (PSI) [39] is adopted to describe the shape feature in a local area surrounding a certain pixel.

Pixel shape feature extraction of a specific pixel consists of the following two steps.

Step 1) Extension of direction lines: We define the pixel homogeneity of the ith direction by

$$PH_i = \|v^c - v^s\|^2 \tag{5}$$

where v^c and v^s are the spectral features of the central and surrounding pixels, respectively. The ith direction line is extended from the central pixel if the following statements are true: 1) PH_i is less than T_1, and 2) the total number of pixels in this direction is less than T_2. The values T_1 and T_2 can be determined through experiments because they are related to the shape and spatial arrangement of objects in the image. T_1 is the threshold for homogeneity and pertains to the spectral variability in a local area. T_2 is related to the average size of a shape area.

Step 2) Length of direction line: The PSI in the ith direction is calculated by the length of the direction line d_i. Then, the shape feature is achieved by

$$v_{\text{Shape}} = [d_1, d_2, \ldots, d_p]^{\text{T}} \in R^p \tag{6}$$

in which p is the total number of all directions.

B. MFC

In this framework, the proposed MFC algorithm finds a low-dimensional representation $Y = [y_1, y_2, \ldots, y_N] \in R^{d \times N}$ of features $\{V_{(i)} = [v_{(i)1}, v_{(i)2}, \cdots, v_{(i)N}] \in R^{L_i \times N}\}_{i=1}^{m}$, in which m is the number of features ($m = 3$; e.g., spectral, texture, and shape features) and N is the number of samples. First, we introduce the traditional single feature-based dimensional reduction, which can be unified to a patch alignment framework. Then, we theoretically show that multiple features can be integrated into a uniformed objective function by simultaneously optimizing the combining weights. Finally, we further extend the MFC to its linear version by linear regression to solve the out-of-sample problem in HSI classification.

1) Single Feature-Based Dimensional Reduction: Based on our previous work [18], [40], [41], representative single feature manifold-learning-based dimensional reduction algorithms, e.g., LLE, LTSA, ISOMAP, and LE, can be unified to a patch alignment framework. Here, we consider the method proposed in [23]. Given N samples $v_i \in R^L (i = 1, \ldots, N)$ in an arbitrary feature, supposing that the target dimensional reduction vectors are $y_i \in R^d (i = 1, \ldots, N)$, the weighted matrix which measured the similarity between the ith and jth samples is defined by

$$W(i, j) = \exp\left(-\|v_i - v_j\|^2 / t\right) \tag{7}$$

if v_i and v_j are k-nearest neighbors, and $W(i, j) = 0$ otherwise, where t is a radius parameter. Patch alignment first builds N patches for all samples, and then, part optimization is performed to obtain the optimal low-dimensional representations for each patch. Afterward, all low-dimensional representations from different patches are unified by a global coordinate alignment to get the objective function

$$\arg\min_Y f = \text{tr}(YMY^{\text{T}})$$

$$\text{s.t. } YY^{\text{T}} = I \tag{8}$$

where $M \in R^{N \times N}$ is the alignment matrix of input samples, which could be computed by

$$M = D - W \tag{9}$$

where D is a diagonal matrix, with its (i, i) element equal to the sum of the ith row of the weighted matrix

$$D_{ii} = \sum_{j=1}^{N} W(i, j). \tag{10}$$

2) Formulation of MFC: Part optimization and global coordinate alignment (8) find a sufficiently smooth low-dimensional combining Y by preserving the intrinsic structure of an arbitrary feature. However, because of the complementary properties of multiple features, different views definitely have different contributions to the final low-dimensional combination [42], [43]. In order to thoroughly explore the complementary properties of multiple view features, the most straightforward way is to impose a nonnegative weight $\omega = [\omega_1, \omega_2, \ldots, \omega_m]$ with conditions that $\omega_i > 0$ and $\sum_{i=1}^{m} \omega_i = 1$ on (8). The larger ω_i is, the more important is the role of the ith feature in learning to obtain the low-dimensional combining. By summing up all of the features, we have

$$\arg\min_Y f = \sum_{i=1}^{m} \omega_i \text{tr}\left(YM_{(i)}Y^{\text{T}}\right)$$

$$\text{s.t. } YY^{\text{T}} = I \tag{11}$$

in which $M_{(i)} \in R^{N \times N}$ is the alignment matrix of the ith feature. By introducing ω into the objective function, each feature has its particular contribution to the unified representation Y. However, the decision of ω is not considered in (11), which means that we might have to determine ω_i by cross-validation. This is crucial for MFC since the discriminative abilities of different features often vary significantly. In addition, the complementary properties among the multiple features should also be carefully considered. When m is larger than two, the time taken for cross-validation increases dramatically, thus a more efficient strategy is required to obtain ω_i.

In this paper, we propose to regard ω as variables in (11) and then optimize the objective function with respect to both Y and ω simultaneously. By putting ω into variables in (11), we have

$$\arg\min_{Y, \omega} f = \sum_{i=1}^{m} \omega_i \text{tr}\left(YM_{(i)}Y^{\text{T}}\right)$$

$$\text{s.t. } YY^{\text{T}} = I \quad \omega_i > 0 \quad \sum_{i=1}^{m} \omega_i = 1. \tag{12}$$

However, from (12), we can see that the current objective function $f(Y, \omega)$ is a linear programming (LP) with respect to ω. Since the optimal solution of LP will be always at the vertex of the linear feasible region, the solution of ω must be $\omega_i = 1$, corresponding to the ith feature with minimum $\text{tr}(YM_{(i)}Y^{\text{T}})$, and $\omega_j = 0$ otherwise. This result means that all of the discriminative information will be totally discarded, except for the most discriminative one, e.g., the best feature.

Therefore, this solution does not meet our goal of combining multiple features to improve the classification performance. To avoid this problem, we make a relaxation by set ω_i^r instead of ω_i, with $r > 1$. Under this current objective function and constraint condition, the optimization $f(Y, \omega)$ is no longer an LP with respect to ω. Therefore, based on the adopted alternating optimizations a) and b), which we will soon introduce, the adopted trick actually makes each feature have a particular weight for the final low-dimensional representation. If r is close to one, the combination coefficients ω_i will be very sparse, and only those features corresponding to a smooth graph will be selected. If r is increased to infinity, different features will share the same weights for the subsequent graph combination. The optimal choice of r should be based on the complementary property of the features for data representation. If the available features are complementary to each other, a larger r is preferred to guarantee that all features properly contribute to the subsequent classification; otherwise, we can choose a small r. In practice, we can use cross-validation or Bayesian model selection to estimate an optimal r for MFC. According to the aforementioned discussions, we can obtain the full formulation of MFC by

$$\arg\min_{Y,\omega} f = \sum_{i=1}^{m} \omega_i^r \mathrm{tr}\left(Y M_{(i)} Y^T\right)$$

$$\text{s.t. } YY^T = I \quad \omega_i > 0 \quad \sum_{i=1}^{m} \omega_i = 1. \tag{13}$$

The proposed optimization of MFC (13) is a nonlinearly constrained nonconvex optimization problem, which cannot find the global optimal solution to the best of the authors' knowledge. In this paper, an alternating optimization is adopted to acquire a local optimal solution by iteratively updating Y and ω.

a) Fix ω to update Y. Optimization (13) is equivalent to

$$\arg\min_{Y} \mathrm{tr}(YMY^T) \text{ s.t. } YY^T = I \tag{14}$$

in which

$$M = \sum_{i=1}^{m} \omega_i^r M_{(i)}. \tag{15}$$

Based on [44], (14) has a global optimal solution Y, given as the eigenvectors associated with the smallest d eigenvalues of M, in which d is the predefined size of low-dimensional combining Y.

b) Fix Y to update ω. The Lagrangian function for optimization (13) is

$$L(\omega, \lambda) = \sum_{i=1}^{m} \omega_i^r \mathrm{tr}\left(Y M_{(i)} Y^T\right) - \lambda \left(\sum_{i=1}^{m} \omega_i - 1\right) \tag{16}$$

with multiplier λ. Then, we obtain the partial derivative of $L(\omega, \lambda)$

$$\partial L / \partial \omega_i = 0 \rightarrow r\omega_i^{r-1} \mathrm{tr}\left(Y M_{(i)} Y^T\right) - \lambda = 0$$

$$\partial L / \partial \lambda = 0 \rightarrow \sum_{i=1}^{m} \omega_i - 1 = 0. \tag{17}$$

Then, we find the global optimal ω by the solution of (17)

$$\omega_i = \frac{\left(1/\mathrm{tr}\left(Y M_{(i)} Y^T\right)\right)^{1/(r-1)}}{\sum_{i=1}^{m} \left(1/\mathrm{tr}\left(Y M_{(i)} Y^T\right)\right)^{1/(r-1)}}. \tag{18}$$

3) Linearization: The alternating optimization tries to train an optimal subspace for original multiple features. It is worth emphasizing that the feature mapping $\{V_{(i)} \in R^{L_i \times N}\}_{i=1}^{m} \rightarrow Y \in R^{d \times N}$ from the multiple high-dimensional feature space to the low-dimensional subspace can be nonlinear and implicit. In remote sensing image classification applications, training a low-dimensional subspace is a time-consuming work because the size of the alignment matrix M scales with the number of samples, i.e., the number of pixels in HSI. When N is increased to 10^4, MFC requires that a long time is spent on computing the alignment matrix $M_{(i)}$ for each feature, as well as the eigenvalue decomposition step in (14). In such a situation, the proposed MFC framework suffers from the out-of-sample problem. In this paper, we adopt a linear approximation to deal with such an out-of-sample problem. A group of explicit linear projection matrices $\{U_i\}_{i=1}^{m}$, trained by a subset of samples $(\{X_{(i)} \in R^{L_i \times n}\}_{i=1}^{m}, n \ll N)$, is applied to approximately construct the low-dimensional representation

$$Y = \sum_{i=1}^{m} U_i^T X_{(i)}. \tag{19}$$

We expect the difference between Y and $\sum_{i=1}^{m} U_i^T X_{(i)}$ to be minimized. Based on this subset of samples $X = \{X_{(i)}\}_{i=1}^{m}$ and their corresponding low-dimensional representation $Y \in R^{d \times n}$, an objective function is designed to obtain $\{U_i\}_{i=1}^{m}$, i.e.,

$$\arg\min_{U} \|Y - U^T X\|^2 \tag{20}$$

where U is formulated as $U = [U_1^T, \ldots, U_m^T]^T$ and X is formulated as $X = [X_1^T, \ldots, X_m^T]^T$. The optimization problem in (20) can be solved easily by using the linear regression method

$$U = (X^T X)^{-1} X^T Y. \tag{21}$$

According to the aforementioned descriptions, we have the following MFC algorithm:

Algorithm 1. The procedure of the MFC framework.

Input: A hyperspectral remote sensing image.

Method:
   a) Extract multiple features $V = \{V_{(i)} \in R^{L_i \times N}\}_{i=1}^{m}$ from HSI by (1), (4), and (6).
   b) Construct the feature matrix $X = \{X_{(i)} \in R^{L_i \times n}\}_{i=1}^{m}$ using a subset of samples.
   c) Calculate an alignment matrix $M_{(i)}$ for each feature by (9) based on $X_{(i)}$.
   d) Initialize $\omega = [1/m, 1/m, \ldots, 1/m]$.
   e) Repeat
      Compute Y by optimization (14);
      Compute ω by equation (18);

Until convergence.

f) Compute linear projection matrix U by (21).

g) Compute a low-dimensional feature representation of HSI: $V_{MFC} = U^T V$.

Output: A low-dimensional MFC of the input HSI.

C. Computational Complexity of MFC

Based on the discussion in part B, the computational complexity of the MFC algorithm contains three parts, as follows. First, a patch alignment framework is introduced to compute alignment matrices $M_{(i)}$ for each feature, and the computational complexity of which is $O(n^2)$. Then, the alternating optimization is repeated: 1) fix ω to update Y, which is an eigenvalue decomposition task on a $n \times n$ matrix with a computational complexity of $O(n^3)$, and b) fix Y to update ω, which has the computational complexity of $O(n^2)$. The entire complexity of an alternating optimization is $O(n^3) \times T$, in which T is the iteration number that we will discuss in the experimental part. Finally, linearization of MFC, which is a linear regression issue, has a computational complexity of $O(n^3)$. Therefore, the computational complexity of the proposed approach is $O(n^3)$. In general, it is time consuming to use the embedding for dimensionality reduction of hyperspectral imagery. However, the adopted linearization of MFC is effective in achieving the accuracy of the manifold learning and, at the same time, in reducing the computational cost.

III. EXPERIMENTS

A. Data Set Description

The experiments were conducted on images from two hyperspectral data sets. One is the hyperspectral digital imagery collection experiment (HYDICE) airborne data over a Mall in Washington DC. A total of 210 bands were collected in the 0.4–2.4-μm region of the visible and infrared spectra. The water absorption bands were then deleted, resulting in 191 channels. The whole data set contains 1280 scan lines, with 307 pixels in each scan line. The spatial resolution of the Washington DC data set is 2 m per pixel. Data set 1 used in our experiment is a subset of the whole set, with a size of 280×307 pixels.

In the second experiment, analysis was performed on an airborne image data set provided by the Data Fusion Technical Committee of the IEEE Geoscience and Remote Sensing Society. This data set was acquired by the reflective optics system imaging spectrometer (ROSIS-03) at the urban test area of Pavia, Northern Italy. The whole data set size is 1400×512 pixels, and the spatial resolution of the Pavia data set is 1.3 m per pixel. Some channels were removed due to noise, and the remaining 102 spectral dimensions from 0.43 to 0.83 μm were processed.

B. Complementary Properties Among Multiple Features

Data set 1 and the reference data are shown in Fig. 2(a) and (b). The desired information classes are roof, road, trail, grass, shadow, and tree. This data set is a challenging one
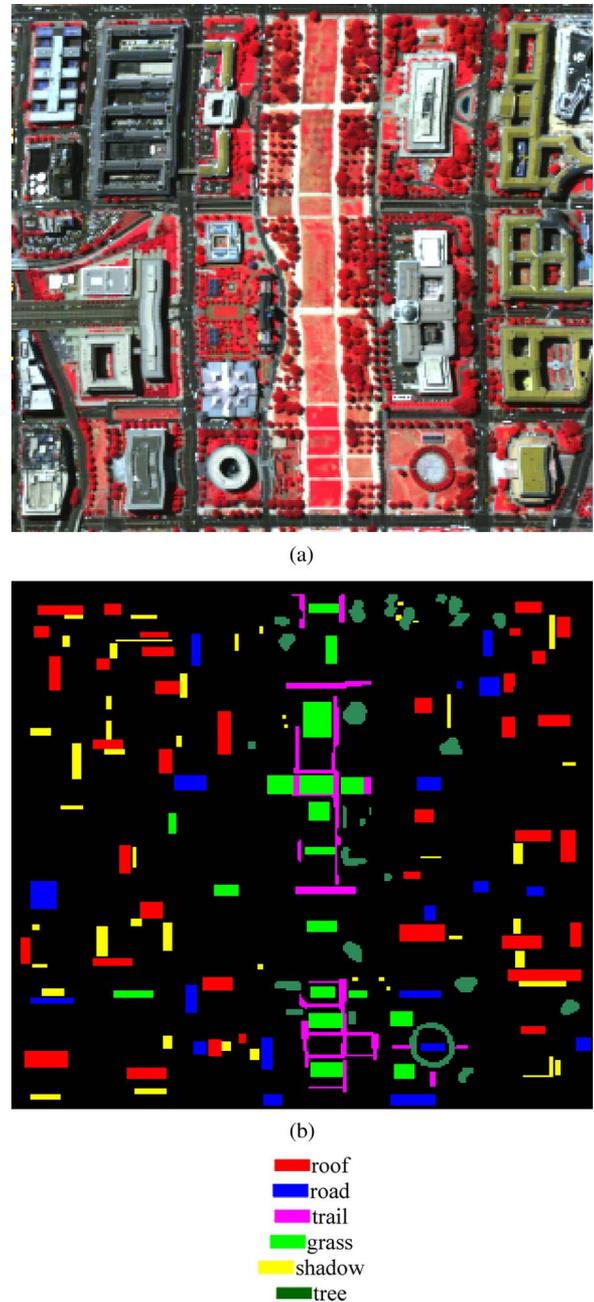
Fig. 2. (a) RGB composites of the DC data set (channels 65, 52, and 36 for RGB) and (b) reference data.

to analyze mainly because roof, road, and trail are spectrally similar in that they may be made of similar materials and there is no single spectral response representative of roof [45]. In our experiments, the shape feature was extracted based on the PSI method in which $p = 20$, we also experimentally set T_1 to a sum of standard deviations in all bands, and T_2 is fixed as ten. Based on the feature extraction mentioned in Section II, we have a spectral feature vector $v_{\text{spectral}} \in R^{191}$, texture feature vector $v_{\text{texture}} \in R^{60}$, and a shape feature vector $v_{\text{shape}} \in R^{20}$ for each pixel in HSI. Some of the feature images are shown in Fig. 3.

We now investigate the complementary properties of the aforementioned multiple features. Fig. 4 shows the spectral,
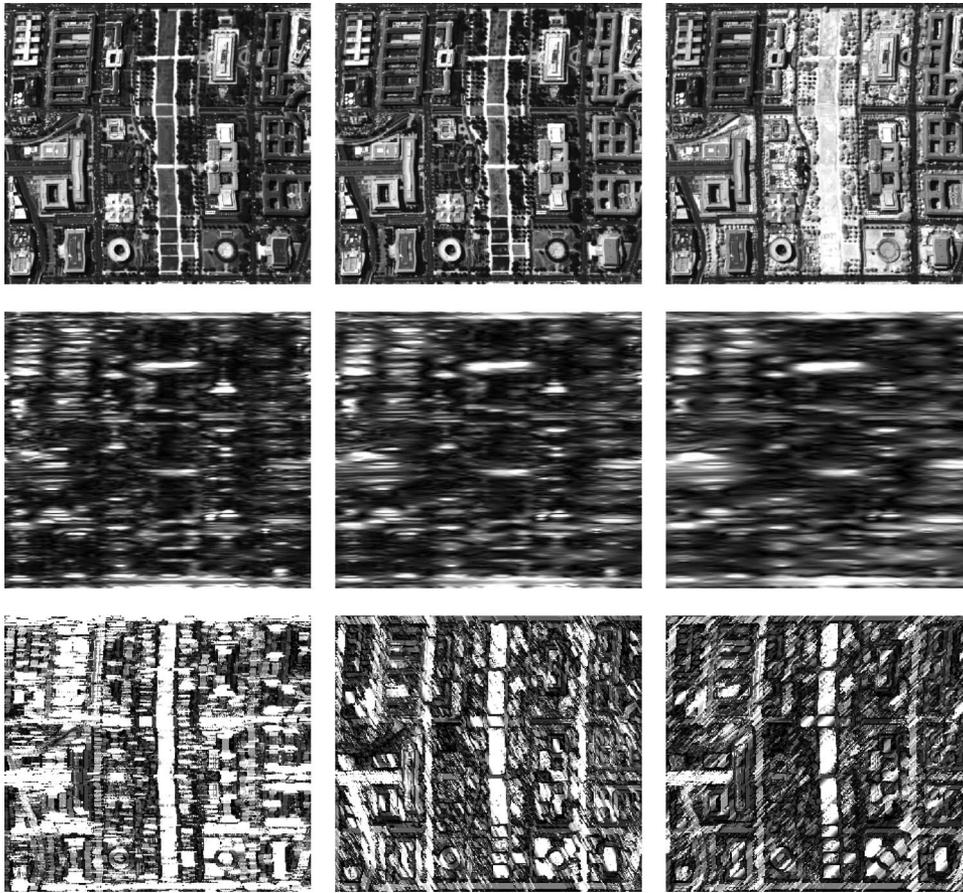
Fig. 3. Multiple features of the DC data set. (First row) Spectral feature images in band 36, 52, and 65. (Second row) Gabor texture feature images, with $d = 1$ and $s = 1$, 3, and 5, respectively. (Third row) Shape feature images in d1, d8, and d16, respectively.

texture, and shape feature curves for five different pixels, which are located at the center of each subimage. These pixels correspond to different classes, e.g., ①–③ for roof, ④ for road, and ⑤ for trail, respectively. Tables (a)–(c) present the correlation coefficients of these pixels on a single feature. In Fig. 4, we learned that, for example, ② and ④ have a very similar spectral feature curve and a high correlation coefficient, e.g., 0.99; however, we might distinguish them because they have low correlation coefficients on texture and shape features, e.g., 0.59 and 0.21, respectively. Although ⑤ has a very similar spectral feature curve to roofs ① and ③, e.g., correlation coefficients of 0.98 and 0.99, the correlation coefficients of ⑤ and ① for the shape feature are 0.36, and the correlation coefficients of ⑤ and ③ for the texture feature are 0.15, which could be useful to classify ⑤ out of the class "roof." Therefore, the complementary properties of the multiple features on data set 1 provide the information to potentially improve the classification performance. However, it could also be observed that ② and ③, which are in the same category (roof), have a not very high spectral similarity (0.86); unfortunately, their texture and shape similarities are low (0.08 and 0.02, respectively). As a result of that, the texture and shape features might deteriorate the classification of ② and ③. Therefore, it is an important and challenging task to utilize multiple features for dimensionality reduction, with the result being an improvement in the classification performance.

C. Classification Result

The proposed MFC algorithm was employed to obtain the low-dimensional feature representation of the aforementioned multiple features. $n = 8596$ samples, which were uniformly sampled from the whole data set, were used to construct the feature matrix for MFC, and the value of radius parameter t in weight matrix constructing was set by cross-validation. To compare the effectiveness of the proposed MFC with the conventional dimension reduction methods, we show the performance of the supervised classification results of the following methods: 1) best feature: the best performance of the single-view feature (in this data set, it is the spectral feature); 2) all features: the conventional multifeature concatenation method, which arranges the feature vectors together; 3) PC: implementing the PC transformation on all features concatenation; 4) MNF: executing the minimum noise fraction rotation [46] on all features concatenation; using manifold-learning-based approaches 5) LLE, 6) LTSA, and 7) LE for all features concatenation; and 8) adopting the proposed MFC. The support vector machine (SVM) classifier [47], which has been reported to be effective in the classification of hyperdimensional feature sets [48], was used to interpret the aforementioned processed feature data. Ten independent classification experiments are conducted in which the training samples were randomly selected from the reference data. For each experiment, we use all reference data as

Fig. 4. Complementary properties of multiple features for different pixels in the DC data set.

TABLE I
NUMBER OF ALL REFERENCE DATA AND TRAINING AND TEST SAMPLES FOR THE DC DATA SET

	All reference data	Training samples	Test samples
Roof	3129	50	3079
Road	1402	50	1352
Trail	1267	50	1217
Grass	1790	50	1740
Shadow	1120	50	1070
Tree	1194	50	1144
Total	9902	300	9602

test samples. The number of training and test samples is listed in Table I.

Eight different feature-based classification results from a certain group of training samples are compared in Fig. 5(a)–(h).

In Fig. 5, the proposed MFC-based classification achieved the best results in both accuracy and visual interpretation. In these maps, the most spectrally similar class pair (roof and road) is focused on. It can be seen that roof exists in the roads in all of the classification maps. This misclassification occurs because roof and road have similar spectra. However, with the help of additional information, e.g., texture and shape features, Fig. 5(h) shows fewer misclassifications than the best feature representation in Fig. 5(a), especially for roofs at the left upper side and around the plaza at the lower right. Compared to the classification results, considering all of the features shown in Fig. 5(b)–(g), the proposed MFC yields satisfactory results with few classification errors, while the conventional multiple feature concatenation method in Fig. 5(b) and five kinds of dimensional reduction approaches in Fig. 5(c)–(g) have many false predictions in the spectrally similar class pairs, including roof–road, road–shadow, and trail–roof.

In order to evaluate thoroughly the different feature representations, the averaged classification accuracies of all classes in ten independent experiments are compared in Table II. In Table II, it is shown that the MFC approach outperforms

| roof | road | trail | grass | shadow | tree |
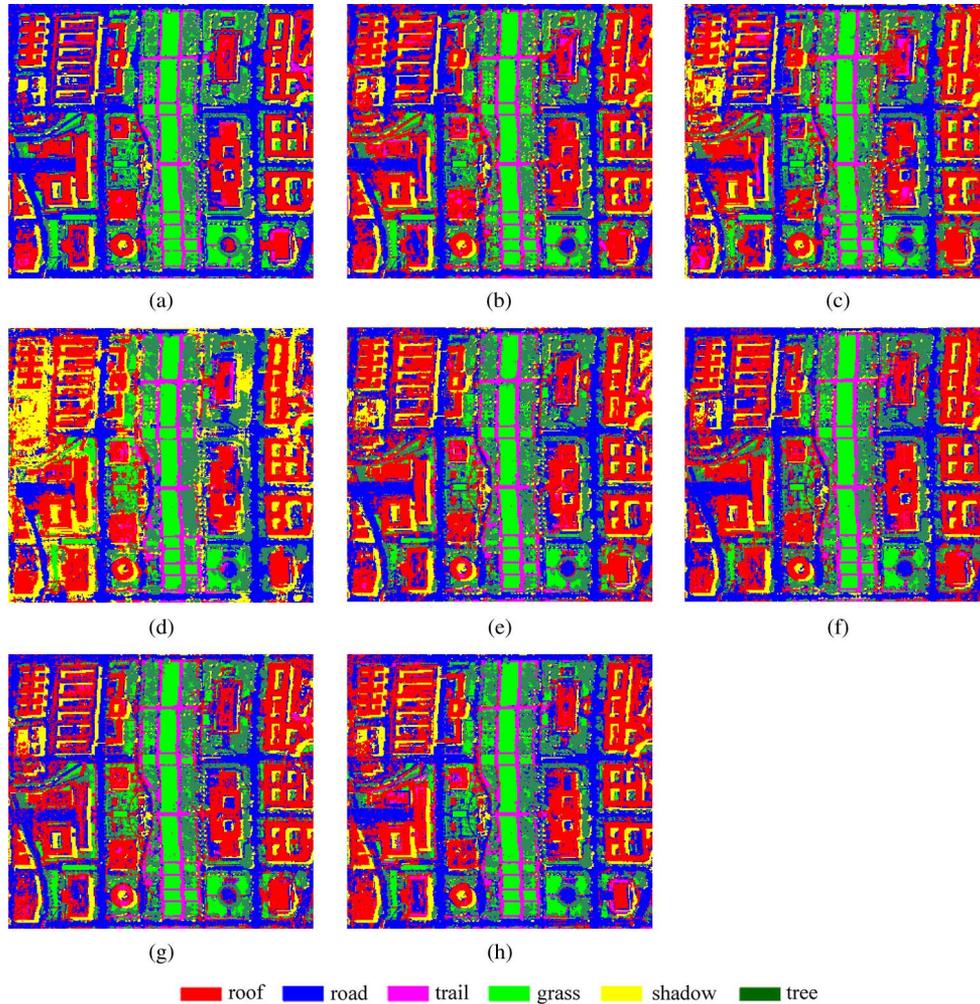
Fig. 5.   (a)–(h) Classification maps of the DC data set obtained using features of the following: best feature, all features, PC, MNF, LLE, LTSA, LE, and MFC, respectively.

TABLE II
CLASS-SPECIFIC ACCURACIES IN PERCENTAGE FOR VARIOUS FEATURES

	BF	AF	PC	MNF	LLE	LTSA	LE	MFC
Roof	82.88	83.60	83.83	84.74	89.74	90.16	86.88	91.10
Road	94.16	84.76	91.20	95.19	86.24	91.20	87.57	94.67
Trail	99.75	98.85	99.18	98.77	96.63	97.70	96.55	99.26
Grass	99.08	99.77	98.22	97.99	99.37	98.28	99.60	99.54
Shadow	97.66	98.04	94.95	98.69	97.85	98.79	98.50	98.60
Tree	96.77	95.19	96.59	97.64	97.38	97.55	97.29	97.38
OA	93.15	91.11	92.53	93.29	92.61	93.91	91.72	95.97
Kappa	0.9153	0.8901	0.9077	0.9170	0.9083	0.9244	0.8975	0.9499

the best view feature representation, the multiple feature con-catenation method, and five kinds of dimensional reduction approaches on multiple features. The improvements observed in Table II indicate that MFC obtained the top three classification

rates in all classes and achieved the highest overall accuracy (OA) and kappa coefficient.

D. Parameter Analysis

1) Effect of Parameter r: In (13), we introduced a parameter r to guarantee that each feature had a particular contribution to the final combining in MFC. Fig. 6(a) and (b) shows the performance variation of MFC with respect to r. The neighbor size k and the low-dimensional feature size d in MFC are both set to 30. Fig. 6(a) describes the relationship of r and ω_i in each feature. Features 1, 2, and 3 denote the spectral, texture, and shape views, respectively. As discussed before, the ith feature, with a larger ω_i, plays the most important role in learning to obtain the low-dimensional combination. It can therefore be seen that the spectral feature is the most discriminating feature for hyperspectral remote sensing image classification. We can also learn from Fig. 6(a) that, if r is close to one, we have ω_1 close to one, which indicates that the spectral feature corresponds to the most discriminating feature. Such a best feature has a leading weight in low-dimensional feature representation. With an increase in r, different weights will be close to each other. When r is larger than 15, the weights stay
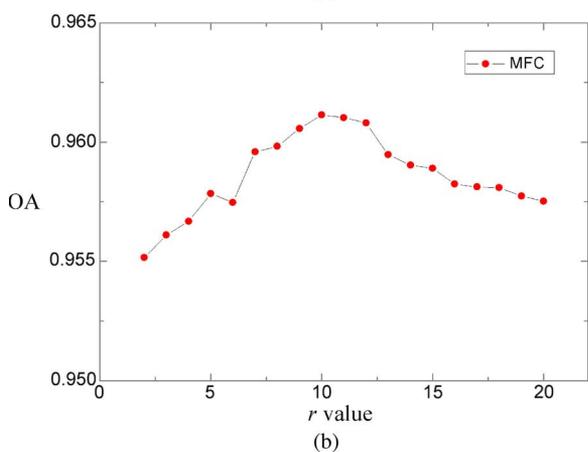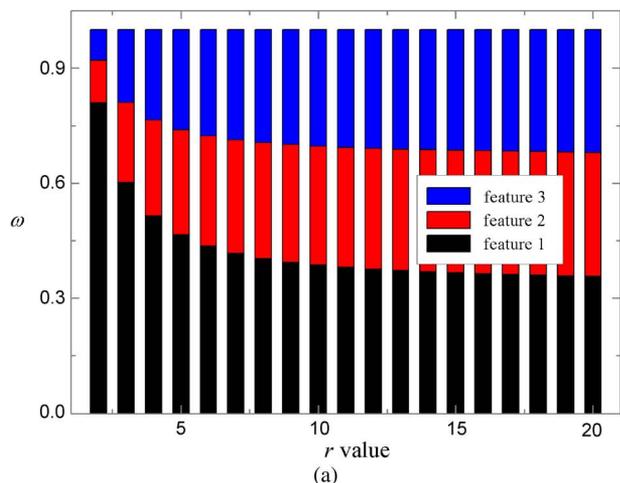
Fig. 6. Relationship of (a) parameter r and weights in each feature and (b) parameter r and OA.

Fig. 7. (a) All eigenvalues and (b) 30 smallest eigenvalues of M in the DC data set, sorted in ascending order.

at [0.357 0.323 0.320] and vary slowly with respect to r, while different features play a particular role in the final combining. It is worth emphasizing that, with r increased indefinitely to infinity, ω_i is equal to $1/m$, determined by (18), which suggests that different features would have the same weights in the final combining. Therefore, the determination of parameter r should be based on the complementary properties among all features. The richer the complementary properties of the view are, the larger r should be selected to guarantee that the weights are closer; otherwise, r should be small. Fig. 6(b) shows the relationship of r and OA of the classification result by SVM. It is known that data set 1 has rich complementary properties among all features. The r–OA curve reaches the best value around $r = 10$. When $r > 20$, the r–OA curve decreases very slowly and shows a stable tendency.

2) Effect of Parameter d: In MFC, the optimal solution Y is given as the eigenvectors associated with the smallest d eigenvalues of M. In order to investigate the effect of d on the classification performance, we first give all of the eigenvalues of M of data set 1 in Fig. 7(a) and highlight the 30 smallest eigenvalues in Fig. 7(b), sorted in ascending order. In Fig. 7(a) and (b), we can see that the eigenvalues increase slowly when the index > 50. The corresponding curve against d and OA is shown in Fig. 8.

Fig. 8. Relationship of d and OA in the DC data set for six-dimensional reduction approaches.

For a more detailed comparison of the following six-dimensional reduction approaches (PC, MNF, LLE, LTSA, LE, and MFC), the classifications are conducted using the aforementioned algorithms with an increase in d. Fig. 8 shows the OAs under the six algorithms. We set $r = 10$ in MFC to take the rich complementary properties among all of the
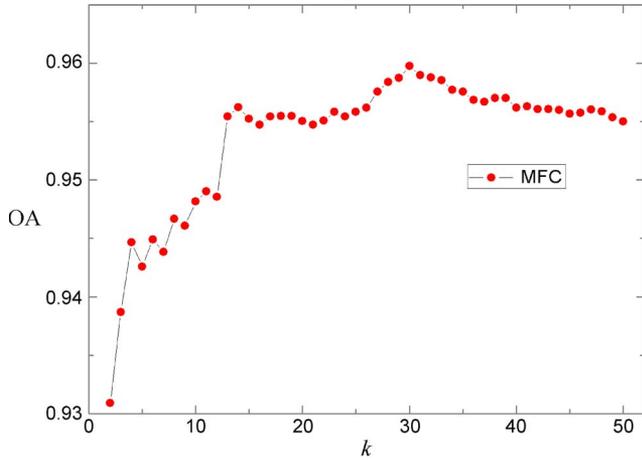
Fig. 9. Relationship of k and OA in MFC of the DC data set.
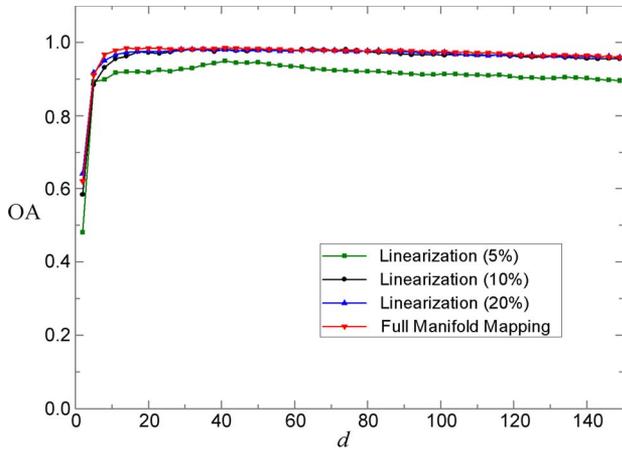
Fig. 10. Classification OA of complete manifold mapping and linearization for MFC.

TABLE III
TIME TAKEN FOR COMPLETE MANIFOLD MAPPING AND LINEARIZATION FOR MFC

Samples	5%	10%	20%	Full
Time taken (s)	0.44	1.73	7.14	201.30

Fig. 11. (a) RGB composite of the Pavia city data set (channels 102, 56, and 31 for RGB) and (b) reference data.

TABLE IV
NUMBER OF ALL REFERENCE DATA AND TRAINING AND TEST SAMPLES FOR THE PAVIA CITY DATA SET

	All reference data	Training samples	Test samples
Roof	1970	30	1940
River	1471	30	1441
Road	1701	30	1671
Grass	1788	30	1758
Shadow	1930	30	1900
Tree	1618	30	1588
Total	10478	180	10298

features into account. As shown in Fig. 8, the MFC performs better than the other five algorithms when $d > 10$ and achieves the best performance in Fig. 8 at $d = 30$. Comparing LLE, LTSA, and LE with PC and MNF, the manifold-learning-based approaches also give some improvements over the traditional methods. When $d > 60$, the performance of MFC decreased since the eigenvalues in Fig. 7 are slowly increasing when the index > 60, which means that the features in $d > 60$ are highly noisy with little discriminating information. When $d > 150$, the classification OA of the six algorithms stabilizes at some specific value. Based on the aforementioned analysis, the optimal value of d for MFC is ought to be set experimentally by considering the eigenvalues of M.

3) Effect of Parameter k: The neighbor size k is another important parameter in the proposed MFC, which has a great influence on the weighted matrix W and the part optimization

(8). Fig. 9 shows the effects of k on the classification OA from 3 to 40 based on data set 1. This OA curve shows a slow change when $k > 15$, and there is a peak at $k = 30$. The classification performance decreases when $k > 40$. Considering that the total number of samples that we selected to build the feature matrix is $n = 8596$, this OA curve suggests a small optimal neighbor size of $k = 30$ for MFC to represent the local geometry property. With this suggestion, we do not need to traverse all possible values of k to find the best one for MFC because the optimal value is usually small to preserve the local geometry property. When a large k is set, the proposed feature combining approach performs poorly for classification because a global structure is considered rather than the local geometry and the discriminative information.

roof **road** **trail** **grass** **shadow** **tree**
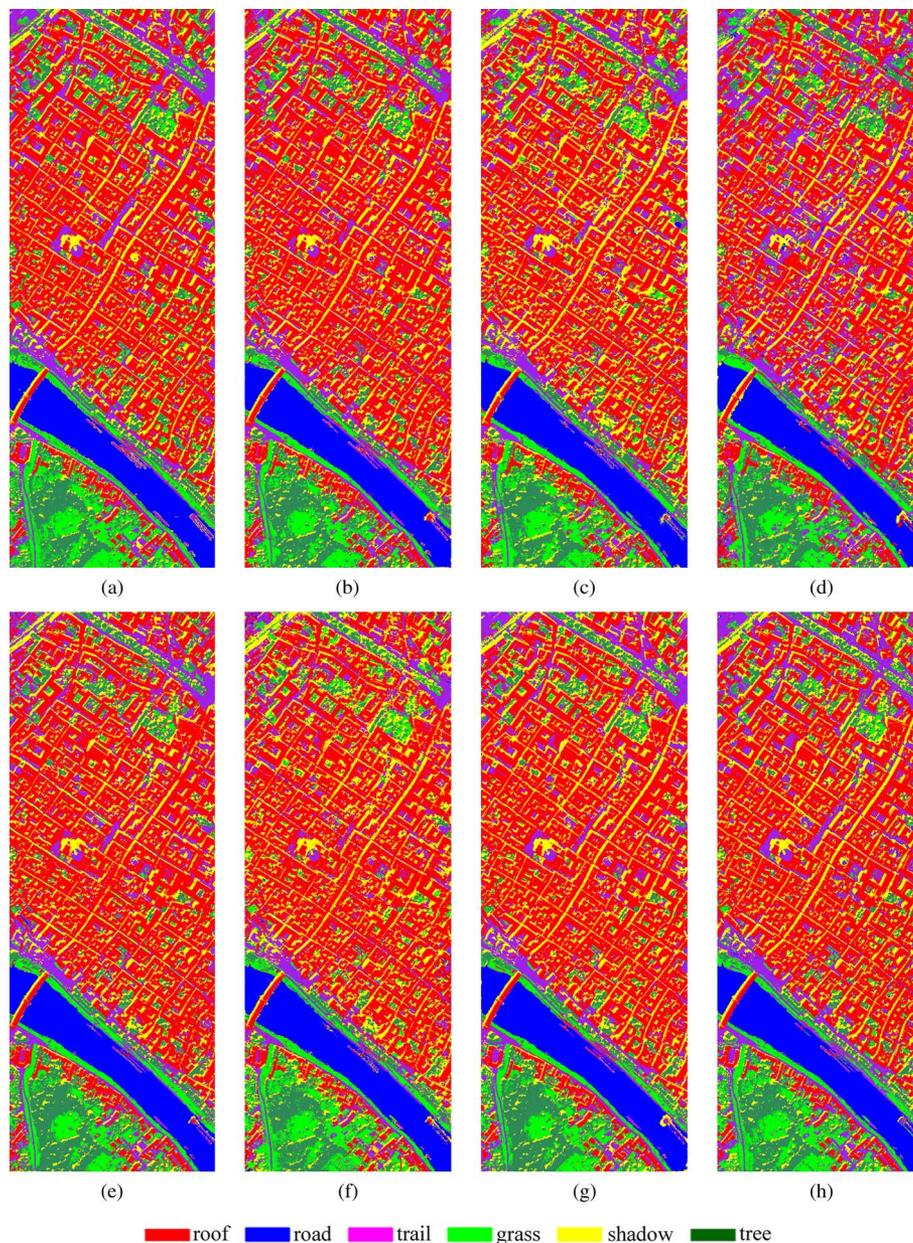
Fig. 12. (a)–(h) Classification maps of the Pavia city data set. (a) Best feature. (b) All features. (c) PC. (d) MNF. (e) LLE. (f) LTSA. (g) LE. (h) MFC.

E. Complete Manifold Mapping Test

In the linearization step of MFC, in order to overcome the out-of-sample problem, we proposed a group of explicit linear projection matrices U_i to approximately represent the MFC $X \rightarrow Y$. Here, we compare the classification rates of complete manifold mapping (without linearization) to our proposed linearization for MFC. Because it requires a long time for complete manifold mapping when the number of samples is increased to 10^4, we have only used a subset of the DC data set from location (99, 1) to (178, 80) for comparison. In this test, 5%, 10%, and 20% of the samples were randomly sampled from the test data set and were used to train linear projection matrices U_i. The corresponding classification rates and time taken are compared in Fig. 10 and Table III.

In Fig. 10 and Table III, we can see that the approximate representation of $X \rightarrow Y$ achieved an acceptable classification rate when the number of samples for training is greater than 10%, i.e., $n = 640$ for this test. However, as shown in Table III, the time taken for MFC increased exponentially with an increase in n, as we theoretically analyzed previously. Therefore, we do not need to use all of the samples to undertake complete manifold mapping for MFC because only a small number of random samples are enough to train a group of linear projection matrices U_i by the proposed MFC linearization for an excellent classification result.

F. Classification on the Pavia City Data Set

Fig. 11(a) shows an RGB composite of a ROSIS image from the Pavia city data set. Six classes of interest are considered, i.e., roof, river, road, grass, shadow, and tree. The reference data are given in Fig. 11(b) and Table IV.
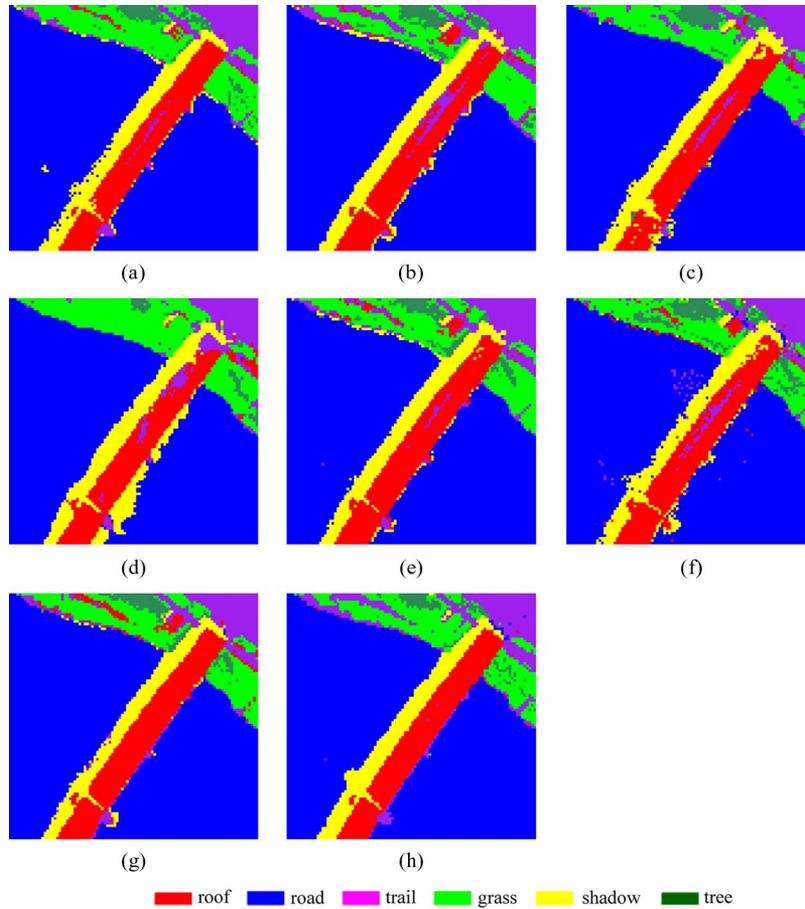
Fig. 13.   (a)–(h) Classification maps of a local region at the bridge. (a) Best feature. (b) All features. (c) PC. (d) MNF. (e) LLE. (f) LTSA. (g) LE. (h) MFC.
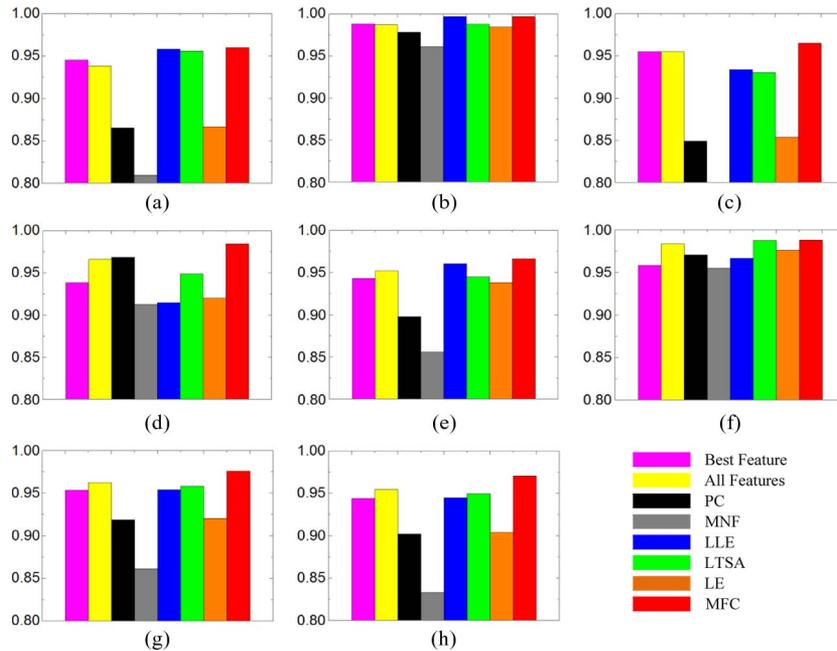
Fig. 14.   Classification accuracies of (a)–(f) all classes, (g) OA, and (h) kappa for eight different feature-based classification results.

The considered multiple features are the following: $v_{\text{spectral}} \in R^{102}$, $v_{\text{texture}} \in R^{60}$, and $v_{\text{shape}} \in R^{20}$. The total number of samples in the data set is $N = 1400 \times 512$ pixels, and $n = 3584$ samples were uniformly sampled from N and

were used in MFC. Based on the aforementioned analysis, the neighbor size k is set to ten, and the weight parameter $r = 10$. We experimentally set $d = 6$ in the Pavia city data set. Eight different feature-based classification results are compared
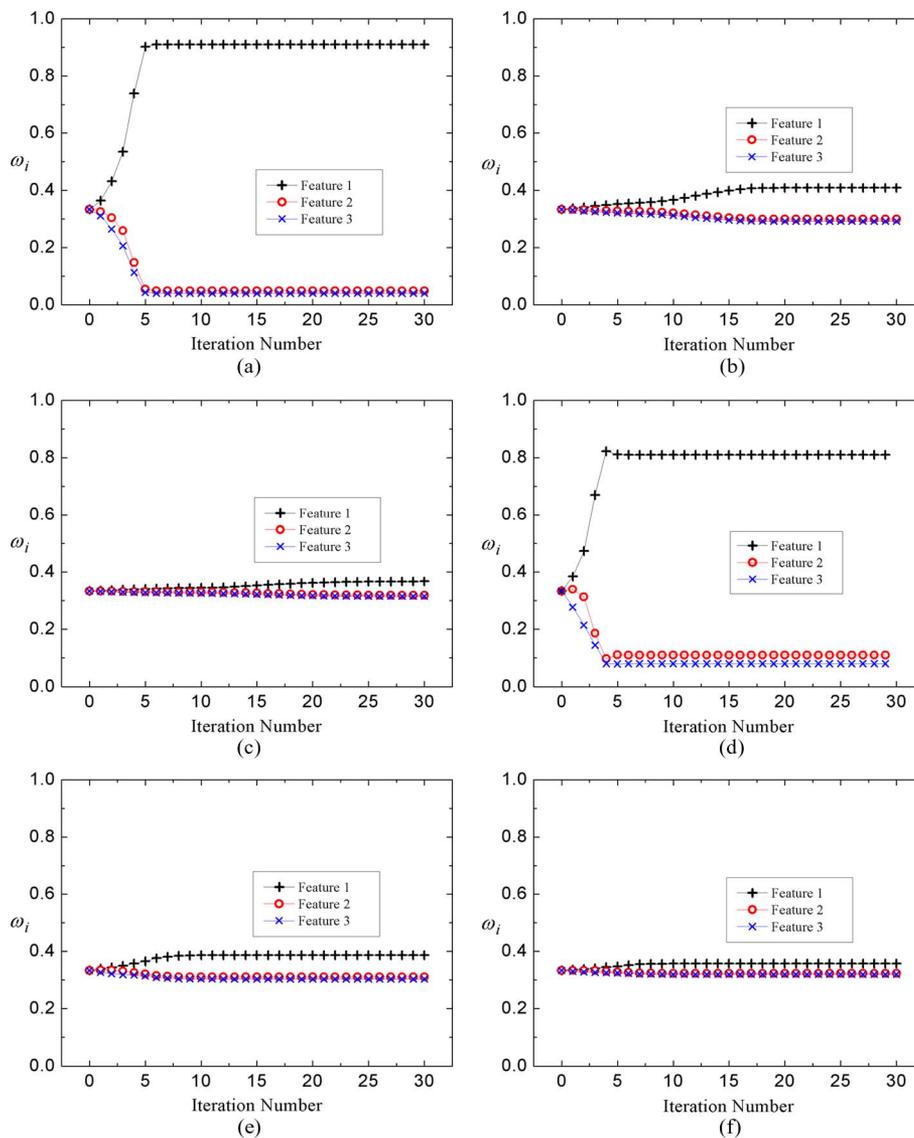
Fig. 15. Alternating optimization converges effectively. (a)–(c) Weights with respect to iteration number using $r = 2$, $r = 10$, and $r = 20$, respectively, for the DC data set. (d)–(f) Weights with respect to iteration number using $r = 2$, $r = 10$, and $r = 20$, respectively, for the Pavia city data set.

in Fig. 12(a)–(h), respectively. The classification results are similar to the aforementioned reports: the proposed MFC-based classification achieved the best performance, especially at the bridge over the river (distinguishing the pair of roof–shadow) and along the river (distinguishing the pairs of grass–tree and grass–road). For example, Fig. 13(a)–(h) shows the classification of a local region 100×100 at the bridge over the river from the subimage location (889, 14) to (988, 113).

The averaged classification accuracies of all classes, based on ten independent groups of samples of six classes, are also compared in Fig. 14. In Fig. 14, improvements can be observed, and MFC obtains the top classification rate in five classes and achieves the top OA and kappa coefficient.

G. Convergence of MFC

Fig. 15 shows that MFC converges efficiently by the alternating optimization. Usually, MFC achieves convergence in about 20 iterations with different r values in the DC data set and converges at stable values in about ten iterations in the Pavia city data set. The reason is that there are fewer samples being selected to construct the feature matrix in the Pavia city data set (we selected $n = 8596$ samples in the DC data set, while $n = 3584$ samples were employed in the Pavia city data set). We also found that fewer iterations are required when r is smaller, and a greater number of iterations are required otherwise. Considering that the r value is usually set at around ten, about 15 iterations are enough for MFC to reach convergence. Fig. 15(a)–(f) shows that MFC converges efficiently by the proposed alternating optimization.

IV. CONCLUSION AND FURTHER WORK

In this paper, we have further explored the complementary properties of multiple features, e.g., spectral, texture, and shape features, to improve the classification performance. In the proposed MFC approach, we have first introduced the patch alignment framework to construct the feature matrix for

each single feature, and then, we have integrated the objective functions of each single feature into a unified one by simultaneously optimizing the combining weights. Lastly, we have further extended the MFC to its linear version to solve the out-of-sample problem in HSI classification. Experiments on the classification of HYDICE and ROSIS hyperspectral data sets suggest that this scheme is effective for classification. Some of the advantages of our work are as follows. First, MFC considers the spectral, texture, and shape features of a pixel to achieve a physically meaningful low-dimensional representation for an effective and accurate classification. Second, the weights for each feature are optimized in the objective function of MFC simultaneously without using cross-validation. This step suggests that the discriminative abilities of different features and the complementary properties among the multiple features have been fully considered in the proposed optimization of MFC, thereby achieving the optimal classification performance.

However, the proposed manifold-learning-based MFC framework can still be revised in some aspects. For instance, how to set the value of radius parameter t in weight matrix constructing (7) has not been considered in the current framework yet, which is also a key aspect in graph-based methods [49]. Therefore, our future work will focus on how to select the optimal radius parameter for each feature, respectively, to obtain the best low-dimensional representation for MFC.

References

[1] C.-I. Chang, *Hyperspectral Data Exploitation: Theory and Applications*. Chichester, U.K.: Wiley-Blackwell, 2007.

[2] G. Licciardi, F. Pacifici, D. Tuia, S. Prasad, T. West, F. Giacco, C. Thiel, J. Inglada, E. Christophe, J. Chanussot, and P. Gamba, "Decision fusion for the classification of hyperspectral data: Outcome of the 2008 GRS-S data fusion contest," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3857–3865, Nov. 2009.

[3] B.-C. Kuo, C.-H. Li, and J.-M. Yang, "Kernel nonparametric weighted feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1139–1155, Apr. 2009.

[4] H. Shen, L. Zhang, B. Huang, and P. Li, "A MAP approach for joint motion estimation, segmentation, and super resolution," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 479–490, Feb. 2007.

[5] S. B. Serpico and L. Bruzzone, "A new search algorithm for feature selection in hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1360–1367, Jul. 2001.

[6] A. Plaza, P. Martinez, J. Plaza, and R. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, Mar. 2005.

[7] X. Huang, L. Zhang, and P. Li, "An adaptive multiscale information fusion approach for feature extraction and classification of IKONOS multispectral imagery over urban areas," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 654–658, Oct. 2007.

[8] L. Zhang, Y. Zhong, B. Huang, J. Gong, and P. Li, "Dimensionality reduction based on clonal selection for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 4172–4186, Dec. 2007.

[9] L. Zhang, B. Du, and Y. Zhong, "Hybrid detectors based on selective end-members," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 6, pp. 2633–2646, Jun. 2010.

[10] L. O. Jimenez, J. L. Rivera-Medina, E. Rodriguez-Diaz, E. Arzuaga-Cruz, and M. Ramirez-Velez, "Integration of spatial and spectral information by means of unsupervised extraction and classification for homogenous objects applied to multispectral and hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 844–851, Apr. 2005.

[11] D. Stathakis and K. Perakis, "Feature evolution for classification of remotely sensed data," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 3, pp. 354–358, Jul. 2007.

[12] L. Bruzzone and C. Persello, "A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization capability," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 9, pp. 3180–3191, Sep. 2009.

[13] L. O. Jimenez-Rodriguez, E. Arzuaga-Cruz, and M. Velez-Reyes, "Unsupervised linear feature-extraction methods and their effects in the classification of high-dimensional data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 2, pp. 469–483, Feb. 2007.

[14] S. Kumar, J. Ghosh, and M. M. Crawford, "Best-bases feature extraction algorithms for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1368–1379, Jul. 2001.

[15] G. Rellier, X. Descombes, F. Falzon, and J. Zerubia, "Texture feature analysis using a Gauss–Markov model in hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 7, pp. 1543–1551, Jul. 2004.

[16] M. H. C. Law and A. K. Jain, "Incremental nonlinear dimensionality reduction by manifold learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 3, pp. 377–391, Mar. 2006.

[17] J. Verbeek, "Learning nonlinear image manifolds by global alignment of local linear models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 8, pp. 1236–1250, Aug. 2006.

[18] T. Zhang, D. Tao, X. Li, and J. Yang, "Patch alignment for dimensionality reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 9, pp. 1299–1313, Sep. 2009.

[19] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 22, pp. 2323–2326, Dec. 2000.

[20] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimension reduction via local tangent space alignment," *SIAM J. Sci. Comput.*, vol. 26, no. 1, pp. 313–338, Jan. 2004.

[21] L. Ma, M. M. Crawford, and J. Tian, "Generalised supervised local tangent space alignment for hyperspectral image classification," *Electron. Lett.*, vol. 46, no. 7, pp. 497–498, Jul. 2010.

[22] J. B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 22, pp. 2319–2323, Dec. 2000.

[23] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, vol. 14, pp. 585–592.

[24] L. Ma, M. M. Crawford, and J. Tian, "Local manifold learning-based k-nearest-neighbor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4099–4109, Nov. 2010.

[25] P. M. Mather, *Computer Processing of Remotely-Sensed Images*. Chichester, U.K.: Wiley, 2004.

[26] T. C. Bau, S. Sarkar, and G. Healey, "Hyperspectral region classification using a three-dimensional Gabor filterbank," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 9, pp. 3457–3464, Sep. 2010.

[27] S. Sarkar and G. Healey, "Hyperspectral texture synthesis using histogram and power spectral density matching," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2261–2270, May 2010.

[28] L. Jiang and R. M. Narayanan, "Integrated spectral and spatial information mining in remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 3, pp. 673–685, Mar. 2004.

[29] B. S. Manjunath and W. Y. Ma, "Texture features for browsing and retrieval of image data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 837–842, Aug. 1996.

[30] L. Zhang, L. Zhang, D. Tao, and X. Huang, "A multifeature tensor for remote-sensing target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 374–378, Mar. 2011.

[31] X. Huang, L. Zhang, and P. Li, "A multiscale feature fusion approach for classification of very high resolution satellite imagery based on wavelet transform," *Int. J. Remote Sens.*, vol. 29, no. 20, pp. 5923–5941, Oct. 2008.

[32] I. Jolliffe, *Principal Component Analysis*. New York: Springer-Verlag, 2002.

[33] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.

[34] A. K. Shackelford and C. H. Davis, "A hierarchical fuzzy classification approach for high-resolution multispectral data over urban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1920–1932, Sep. 2003.

[35] S. Wenbo, J. M. Keller, T. L. Haithcoat, and C. H. Davis, "Automated geospatial conflation of vector road maps to high resolution imagery," *IEEE Trans. Image Process.*, vol. 18, no. 2, pp. 388–400, Feb. 2009.

[36] 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.

[37] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Dec. 2008.

[38] X. Huang and L. Zhang, "A multidirectional and multiscale morphological index for automatic building extraction from mutispectral GeoEye-1 imagery," *Photogramm. Eng. Remote Sens.*, vol. 77, no. 7, pp. 721–732, Jul. 2011.

[39] L. Zhang, X. Huang, B. Huang, and P. Li, "A pixel shape index coupled with spectral information for classification of high spatial resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 10, pp. 2950–2961, Oct. 2006.

[40] T. Zhou, D. Tao, and X. Wu, "Manifold elastic net: A unified framework for sparse dimension reduction," *Data Min. Knowl. Disc.*, vol. 22, no. 3, pp. 340–371, Jul. 2010.

[41] W. Bian and D. Tao, "Max–min distance analysis by using sequential SDP relaxation for dimension reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 5, pp. 1037–1050, May 2011.

[42] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview spectral embedding," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 60, no. 6, pp. 1438–1446, Dec. 2010.

[43] B. Xie, Y. Mu, D. Tao, and K. Huang, "m-SNE: Multiview stochastic neighbor embedding," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 41, no. 4, pp. 1088–1096, Aug. 2011.

[44] R. Bhatia, *Matrix Analysis*. New York: Springer-Verlag, 1997.

[45] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.

[46] U. Amato, R. M. Cavalli, A. Palombo, S. Pignatti, and F. Santini, "Experimental approach to the selection of the components in the minimum noise fraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 1, pp. 153–160, Jan. 2009.

[47] C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Disc.*, vol. 2, no. 2, pp. 121–167, Feb. 1998.

[48] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3858–3866, Dec. 2007.

[49] B. Xie, M. Wang, and D. Tao, "Toward the optimization of normalized graph Laplacian," *IEEE Trans. Neural Netw.*, vol. 22, no. 4, pp. 660–666, Apr. 2011.

Liangpei Zhang (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics, Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing, Wuhan University. He is also a "Chang-Jiang Scholar" Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has more than 200 research papers. He is the holder of five patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is a Fellow of the IEE, Executive Member (Board of Governor) of the China National Committee of International Geosphere–Biosphere Programme, Executive Member of the China Society of Image and Graphics, etc. He regularly serves as a Cochair of the series SPIE Conferences on Multispectral Image Processing and Pattern Recognition, Conference on Asia Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and Geoinformatics Symposiums. He also serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence*, *International Journal of Image and Graphics*, *International Journal of Digital Multimedia Broadcasting*, *Journal of Geo-spatial Information Science*, and *Journal of Remote Sensing*.

Dacheng Tao (M'07) received the B.Eng. degree from the University of Science and Technology of China, Hefei, China, the M.Phil. degree from The Chinese University of Hong Kong, Hong Kong, and the Ph.D. degree from the University of London, London, U.K.

He is a Professor of computer science with the Centre for Quantum Computation and Information Systems and the Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia. He mainly applies statistics and mathematics for data analysis problems in data mining, computer vision, machine learning, multimedia, and video surveillance. He has authored and coauthored more than 100 scientific articles at top venues, including IEEE T-PAMI, T-KDE, T-IP, NIPS, ICML, UAI, AISTATS, ICDM, IJCAI, AAAI, CVPR, ECCV, ACM T-KDD, Multimedia, and KDD.

Prof. Tao was the recipient of the best theory/algorithm paper runner up award in IEEE ICDM'07.

Xin Huang received the Ph.D. degree in photogrammetry and remote sensing from the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS), Wuhan University, Wuhan, China, in 2009.

He is currently an Associate Professor with the LIESMARS, Wuhan University. He has published more than 20 peer-reviewed articles in international journals such as IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, *Photogrammetric Engineering and Remote Sensing*, and *International Journal of Remote Sensing*. His research interests include hyperspectral data analysis, high-resolution image processing, pattern recognition, and remote sensing applications.

Dr. Huang has served as a reviewer for most of the international journals for remote sensing. He was the recipient of the Top-10 Academic Star of Wuhan University in 2009. In 2010, he received the Boeing Award for the best paper in image analysis and interpretation from the American Society for Photogrammetry and Remote Sensing.

Lefei Zhang (S'11) received the B.S. degree in sciences and techniques of remote sensing from Wuhan University, Wuhan, China, in 2008, where he is currently working toward the Ph.D. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing.

His research interests include hyperspectral data analysis, high-resolution image processing, and pattern recognition in remote sensing images.