Local Binary Patterns and Extreme Learning Machine for Hyperspectral Imagery Classification

Wei Li, *Member, IEEE*, Chen Chen, *Student Member, IEEE*,
Hongjun Su, *Member, IEEE*, and Qian Du, *Senior Member, IEEE*

***Abstract*—It is of great interest in exploiting texture information for classification of hyperspectral imagery (HSI) at high spatial resolution. In this paper, a classification paradigm to exploit rich texture information of HSI is proposed. The proposed framework employs local binary patterns (LBPs) to extract local image features, such as edges, corners, and spots. Two levels of fusion (i.e., feature-level fusion and decision-level fusion) are applied to the extracted LBP features along with global Gabor features and original spectral features, where feature-level fusion involves concatenation of multiple features before the pattern classification process while decision-level fusion performs on probability outputs of each individual classification pipeline and soft-decision fusion rule is adopted to merge results from the classifier ensemble. Moreover, the efficient extreme learning machine with a very simple structure is employed as the classifier. Experimental results on several HSI data sets demonstrate that the proposed framework is superior to some traditional alternatives.**

***Index Terms*—Decision fusion, extreme learning machine (ELM), Gabor filter, hyperspectral imagery (HSI), local binary patterns (LBPs), pattern classification.**

I. Introduction

With the advance of sensor technology, hyperspectral imagery (HSI) with high spatial resolution has been continually becoming more available. In conventional HSI classification systems, classifiers [1]–[4] solely consider spectral signatures while ignoring spatial information at neighboring locations. During the last decade, it is of great interest in exploiting spatial features to improve HSI classification performance. For example, composite kernels (CK) for combination of both spectral and spatial information were employed by a support vector machine (SVM) classifier, referred to as SVM-CK [5]. Another model to incorporate spatial-context information is Markov random field (MRF) [6], which was successfully applied in HSI. In [7], MRF was utilized as a postprocessing stage based on a segmentation map obtained by a pixelwise SVM classifier, namely, SVM-MRF. Similarly, MRF was combined with Bayesian framework, such as Gaussian mixture model (GMM) classifier and subspace multinomial logistic regression (SubMLR), denoted as GMM-MRF [8] and SubMLR-MRF [9], respectively.

In addition to the CK- or MRF-based framework, many researchers have worked with other spatial features. For instance, morphological profile (MP) generated by certain morphological operators (e.g., opening and closing), which is widely used for modeling structural information, was introduced in [10]. In [11], a spectral–spatial preprocessing method was proposed for noise-robust HSI classification by employing a multihypothesis prediction strategy which was originally developed for compressed sensing image reconstruction [12] and superresolution [13]. In [14], an edge preserving filter was employed to smooth a probability map produced by SVM while taking neighborhood information into account. In [15], a multiple conditional random field ensemble model, including gray-level cooccurrence matrix (GLCM), Gabor texture features, and gradient orientation features, was proposed. Meanwhile, a multifeature model, including GLCM, different MP, and urban complexity index, was presented in [16]. Two-dimensional Gabor features extracted from selected bands were investigated in [17], and a patch alignment framework was introduced to linearly combine multiple features (e.g., spectral, texture, shape, etc.) in [18].

Recently, local binary pattern (LBP) operator [19] has been presented for rotation-invariant texture classification. LBP is a simple yet efficient advanced operator to describe local spatial pattern. In the original LBP, a texture feature is extracted on a pixel level along with a local neighborhood. To be brief, the operator labels the pixels of a local region by binary thresholding with the center pixel value. It can be seen that the operator does not make any assumption on the distribution of a local region. LBP has also been introduced in remote sensing community. In [20], LBP and local phase quantization were investigated for texture characterization of land-cover classification of optical remote sensing image data. In [21], LBP, histogram of oriented gradient (HoG), and mean-variance descriptor were extracted

Manuscript received May 9, 2014; revised September 22, 2014 and December 2, 2014; accepted December 10, 2014. This work was supported in part by the National Natural Science Foundation of China under Grants NSFC-61302164 and 41201341 and in part by the Fundamental Research Funds for the Central Universities under Grant YS1404.

W. Li is with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China (e-mail: liwei089@ieee.org).

C. Chen is with the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX 75080 USA (e-mail: chenchen870713@gmail.com).

H. Su is with the School of Earth Sciences and Engineering, Hohai University, Nanjing 210098, China (e-mail: hjsurs@163.com).

Q. Du is with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS 39762 USA (e-mail: du@ece.msstate.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2014.2381602

in long-wave infrared imagery for anomaly detection. In [22], LBP and HoG features were employed for tracking of ship traffic in TerraSAR-X images. In [23], LBP features were calculated on a single band chosen from an HSI cube for colon biopsy classification.

In this paper, a texture-based classification paradigm is proposed. It employs LBP and global Gabor filter in a set of selected bands to produce a comprehensive description of spatial texture information. To our best knowledge, [23] is currently the only publication to employ LBP for HSI classification; unfortunately, the work merely used a single spectral band and purely viewed it as a gray-scale image while ignoring the discriminative information in other spectral bands. In this paper, the criterion of linear prediction error (LPE) [24], based on band dissimilarity, is used for unsupervised band selection, and LBP will be extracted from the selected bands. According to the investigation in [24], its performance is better than several popular band selection methods in the literature. Dimensionality reduction can also be achieved by a transform-based method, such as principal component analysis (PCA); the experimental results in [17] and [25] suggest that using selected bands may offer a slightly better performance than using principal components (PCs) because fine structures tend to be present in minor PCs instead of major PCs. In this paper, band-selection-based dimensionality reduction is adopted due to its relative simplicity and overall robust performance, although more advanced transform-based methods exist (e.g., maximum orthogonal complements algorithm [26]).

In the proposed scheme, feature-level fusion and decision-level fusion are investigated on the extracted multiple features. Feature-level fusion combines different feature vectors together into a single feature vector. Decision-level fusion performs on probability outputs of each individual classification pipeline and combines the distinct decisions into a final one. The decision fusion process can occur as either "hard" fusion (e.g., majority voting rule [27]) at the class-label level or "soft" fusion (e.g., linear opinion pool or logarithm opinion pool (LOGP) rule [28], [29]) at the probability level. In this paper, LOGP is chosen because it is a weighted product of different probability distributions and its decision from multiple classifiers is treated independently in the fusion process.

Extreme learning machine (ELM) [30], [31] classifier is employed to provide probability classification outputs using the extracted features. The choice of ELM classifier is due to its efficient computation and its even better classification performance than SVM. ELM has a very simple structure: one hidden layer and one linear output layer. Compared to traditional neural networks and SVM, ELM can be trained much faster since its input weights are randomly generated and the output weights are analytically computed with a least squares solution. ELM for HSI classification is recently paid great attention. In [32], kernel-based ELM was considered for land-cover classification and provided comparable performance to SVM. In [33], ELM was employed to obtain accurate thematic maps of soybean crops in precision agriculture. Reference [34] developed an automatic-solution-based differential evolution to optimize the parameters of ELM. Furthermore, Bagging-based ELM and AdaBoost-based ELM were introduced in [35]. It would be interesting to study its performance when using multiple features (i.e., spectral bands, local spatial feature, and global spatial feature) as presented in this work.

There are two primary contributions in this research. As the first contribution, an effective texture feature extraction approach that is more suitable to HSI is provided. Specifically, an LPE-based band selection is first employed to find a set of distinctive and informative bands; for each band, the LBP code is computed for every pixel in the entire image to form an LBP code image, and then for each local patch centered at a pixel of interest, the LBP histogram is generated. The second contribution is in the effective fusion of extracted local LBP features, global Gabor features, and original spectral features, wherein LOGP plays a critical role in merging probability outputs of the multiple texture and spectral features. It is worth mentioning that a Gabor filter can be viewed as a global operator to capture the global texture features (e.g., orientation and scale). On the other hand, LBP applied to each pixel belongs to a local operator that can characterize the local spatial textures such as edges, corners, and knots. Therefore, Gabor features and LBP represent texture information from different perspectives, which is our motivation of combining them for classification improvement.

The remainder of this paper is organized as follows. Section II introduces spatial feature extraction using a Gabor filter and an LBP operator. Section III briefly reviews the SVM and ELM classifiers. Section IV provides a detailed description of the proposed classification framework, including band selection and two fusion strategies. Section V presents the experimental results with three real hyperspectral data sets. Finally, Section VI makes several concluding remarks.

II. Gabor Filter and LBP

A. Gabor Filter

A Gabor filter [36], [37] can be viewed as an orientation-dependent bandpass filter, which is orientation-sensitive and rotation-variant. In order to achieve rotation-invariance, a circularly symmetric Gabor filter is commonly employed, in which all directions for each pass band are considered. The magnitudes of each Gabor-filtered image reflect signal power in the corresponding filter pass band and are used as Gabor features [17], [18]. In a 2-D (a, b) coordinate system, a Gabor filter, including a real component and an imaginary term, can be represented as

$$G_{\delta,\theta,\psi,\sigma,\gamma}(a,b) = \exp\left(-\frac{a'^2+\gamma^2 b'^2}{2\sigma^2}\right)\exp\left(\mathrm{j}\left(2\pi\frac{a'}{\delta}+\psi\right)\right) \tag{1}$$

where

$$a' = a\cos\theta + b\sin\theta \tag{2}$$

$$b' = -a\sin\theta + b\cos\theta. \tag{3}$$

Here, δ represents the wavelength of the sinusoidal factor, θ represents the orientation separation angle (e.g., $\pi/8$, $\pi/4$, $\pi/2$, etc.) of Gabor kernels, ψ is the phase offset, σ is the standard derivation of the Gaussian envelope, and γ is the spatial aspect

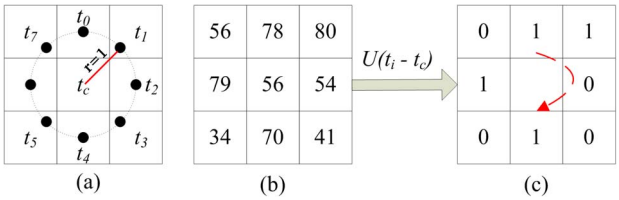

Fig. 1. Example of LBP binary thresholding. (a) Center pixel t_c and its eight circular neighbors $\{t_i\}_{i=0}^{7}$ with radius $r = 1$. (b) 3×3 sample block. (c) Binary labels of eight neighbors.

ratio (the default value is 0.5 in [36] and [37]) specifying the ellipticity of the support of the Gabor function. $\psi = 0$ and $\psi = \pi/2$ return the real and imaginary parts of the Gabor filter, respectively. Parameter σ is determined by δ and spatial frequency bandwidth bw as

$$\sigma = \frac{\delta}{\pi}\sqrt{\frac{\ln 2}{2}}\frac{2^{bw}+1}{2^{bw}-1}. \tag{4}$$

B. LBP

LBP [19] belongs to a gray-scale and rotation-invariant texture operator. Given a center pixel (scalar value) t_c, each neighbor of a local region is assigned with a binary label, which can be either "0" or "1," depending on whether the center pixel has a larger intensity value or not. The neighboring pixels are from a set of equally spaced samples over a circle of radius r centered at the center pixel. Radius r determines how far the neighboring pixels can be located away from the center pixel. Along with selected m neighbors $\{t_i\}_{i=0}^{m-1}$, the LBP code for the center pixel t_c is given by

$$\text{LBP}_{m,r}(t_c) = \sum_{i=0}^{m-1} U(t_i - t_c)2^i \tag{5}$$

where $U(t_i - t_c) = 1$ if $t_i > t_c$ and $U(t_i - t_c) = 0$ if $t_i \leq t_c$. Fig. 1 illustrates an example of the binary thresholding process of eight $((m, r) = (8, 1))$ circular neighbors given the center pixel t_c. The LBP code is then calculated in a clockwise direction, i.e., the binary label sequence "11001010" $= 83$. Supposing that the coordinate of t_c is (0, 0), each neighbor t_i has a coordinate of $(r\sin(2\pi i/m), r\cos(2\pi i/m))$. In practice, parameter set (m, r) may change, such as (4, 1), (8, 2), etc. The locations of circular neighbors that do not fall exactly on image grids are estimated by bilinear interpolation [38]. The output of the LBP operator in (5) indicates that the binary labels in a neighborhood, represented as an m-bit binary number (including 2^m distinct values), reflect texture orientation and smoothness in a local region. After obtaining the LBP code, an occurrence histogram, as a nonparametric statistical estimate, is computed over a local patch. A binning procedure is required to guarantee that the histogram features have the same dimension.

C. Comparison of LBP and Gabor Filter

Given the aforementioned brief description of the Gabor filter and LBP, it can be noticed that the former belongs to a global operator while the latter is a local one. As a consequence, Gabor features and LBP represent texture information from different perspectives. Fig. 2 illustrates an example of comparison between LBP and Gabor features in a natural image (namely, *boat*) of size 256×256. Fig. 2(b) shows the LBP-coded image obtained using (5) with $(m, r) = (8, 1)$, and Fig. 2(c)–(d) illustrates the filtered images obtained by the Gabor filter with different θ (i.e., 0, $\pi/4$, $\pi/2$, and $3\pi/4$). In Fig. 2, the Gabor features produced by the average magnitude response for each Gabor-filtered image reflect the global signal power, while the LBP-coded image is a better expression of detailed local spatial features, such as edges, corners, and knots. Thus, it is promising to apply the global Gabor filter as a supplement to the local LBP operator that lacks the consideration of distant pixel interactions, which is also the motivation of this work.

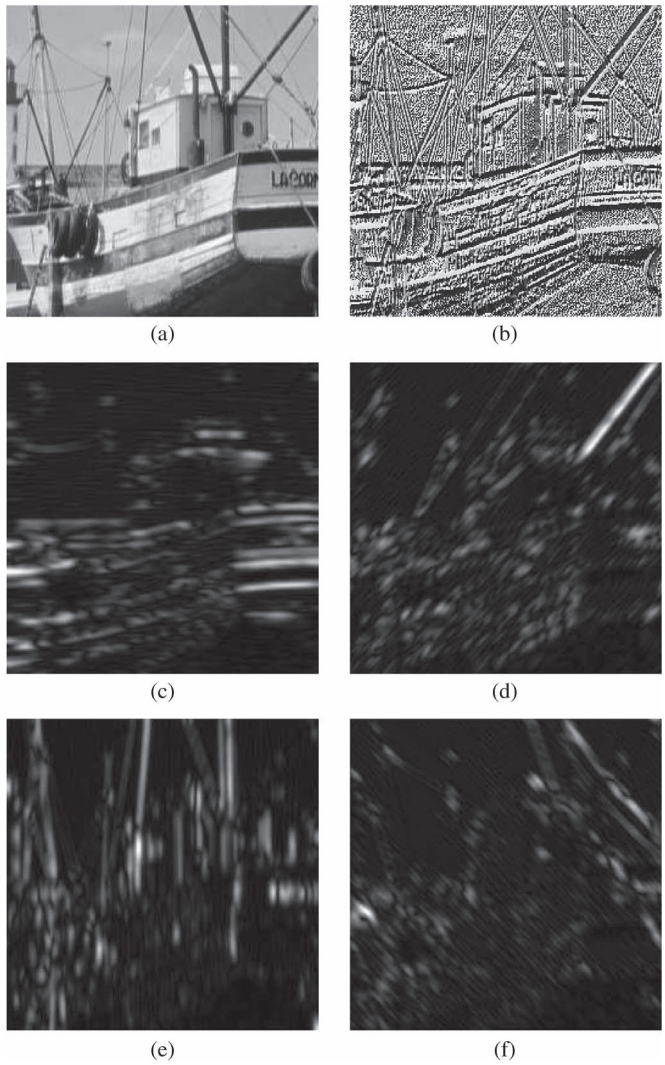

Fig. 2. Example of LBP versus Gabor filter. (a) Input image. (b) LBP-coded image (different intensities representing different codes). (c)–(f) Filtered images obtained by the Gabor filter with different θ values. (a) Input image. (b) LBP-coded image. (c) Gabor feature image, $\theta = 0$. (d) Gabor feature image, $\theta = \pi/4$. (e) Gabor feature image, $\theta = \pi/2$. (f) Gabor feature image, $\theta = 3\pi/4$.

As we stated earlier, the Gabor filter can capture the global texture information of an image, while LBP represents the

local texture information. It is known that an HSI data usually contains homogeneous regions where pixels fall into the same class. Gabor features are able to reflect such global texture information because the Gabor filter can effectively capture the orientation and scale of physical structures in the scene. Therefore, combining Gabor and LBP features can achieve better classification performance than using only LBP features.

III. SVM AND ELM

A. SVM

SVM [3], [5] has been proved to be powerful in HSI classification. The key idea behind a kernel version of SVM is to map the data from its original input space into a high-dimensional kernel-induced feature space where classes may become more separable. For training samples $\{\mathbf{x}_i\}_{i=1}^n$ in $\mathbb{R}^d$ with class labels $y_i \in \{1, -1\}$ and a nonlinear kernel mapping $\phi(\cdot)$, an SVM [5] separates binary classes by determining an optimal hyperplane in the kernel-induced space by solving

$$\min_{\boldsymbol{\omega},\xi_i,p} \left\{ \frac{1}{2}\|\boldsymbol{\omega}\|^2 + \varsigma \sum_{i=1}^{n} \xi_i \right\} \tag{6}$$

subject to the constraints

$$y_i \left(\langle \phi(\boldsymbol{\omega}, \mathbf{x}_i) \rangle + p \right) \geq 1 - \xi_i \tag{7}$$

for $\xi_i \geq 0$ and $i = 1, \ldots, n$, where $\boldsymbol{\omega}$ is normal to the optimal decision hyperplane (i.e., $\langle \boldsymbol{\omega}, \phi(\mathbf{x}) \rangle + p = 0$), n denotes the number of samples, p is the bias term, ς is the regularization parameter which controls the generalization capacity of SVM, and ξ_i is the positive slack variable allowing us to accommodate permitted errors appropriately. The aforementioned problem is solved by maximizing its Lagrangian dual form

$$\max_{\boldsymbol{\alpha}} \left\{ \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{n} \alpha_i \alpha_i y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \right\} \tag{8}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are nonzero Lagrange multipliers constrained to $0 \leq \alpha_i \leq \varsigma$, and $\Sigma_i \alpha_i y_i = 0$, for $i = 1, \ldots, n$. Some commonly implemented kernel functions include the polynomial kernel and the RBF kernel. In this paper, RBF is considered, which is represented as

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma'^2} \right) \tag{9}$$

where σ' is a width parameter. Finally, the decision function is represented as

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{n} y_i \alpha_i K(\mathbf{x}_i, \mathbf{x}) + p \right). \tag{10}$$

B. ELM

ELM [30], [31] is a neural network with only one hidden layer and one linear output layer. The weights between the input and the hidden layers are randomly assigned, and the weights of the output layer are computed using a least squares method. Therefore, the computational cost is much lower than any other neural-network-based methods. For C classes, let the class labels be defined as $y_k \in \{1, -1\}$ $(1 \leq k \leq C)$. Thus, a constructed row vector $\mathbf{y} = [y_1, \ldots, y_k, \ldots, y_C]$ indicates the class to which a sample belongs. For example, if $y_k = 1$ and the other elements in $\mathbf{y}$ are -1, then the sample belongs to the kth class. Thus, the training samples and corresponding labels are represented as $\{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^n$, where $\mathbf{x}_i \in \mathbb{R}^d$ and $\mathbf{y}_i \in \mathbb{R}^C$, the output function of an ELM with L hidden nodes can be expressed as

$$f_L(\mathbf{x}_i) = \sum_{j=1}^{L} \boldsymbol{\beta}_j h(\mathbf{w}_j \cdot \mathbf{x}_i + b_j) = \mathbf{y}_i, \quad i = 1, \ldots, n \tag{11}$$

where $h(\cdot)$ is a nonlinear activation function (e.g., sigmoid function), $\boldsymbol{\beta}_j \in \mathbb{R}^C$ denotes the weight vector connecting the jth hidden node to the output nodes, $\mathbf{w}_j \in \mathbb{R}^d$ denotes the weight vector connecting the jth hidden node to the input nodes, and b_j is the bias of the jth hidden node. The term $\mathbf{w}_j \cdot \mathbf{x}_i$ denotes the inner product of $\mathbf{w}_j$ and $\mathbf{x}_i$. If a value of 1 is padded to $\mathbf{x}_i$ to make it a $(d+1)$-dimensional vector, then the bias can be considered as an element of the weight vector, which is also randomly assigned. For n equations, (11) can be written as

$$\mathbf{H}\boldsymbol{\beta} = \mathbf{Y} \tag{12}$$

where $\mathbf{Y} = [\mathbf{y}_1; \mathbf{y}_2; \ldots; \mathbf{y}_n] \in \mathbb{R}^{n \times C}$, $\boldsymbol{\beta} = [\boldsymbol{\beta}_1; \boldsymbol{\beta}_2; \ldots; \boldsymbol{\beta}_n] \in \mathbb{R}^{L \times C}$, and $\mathbf{H}$ is the hidden layer output matrix of the neural network expressed as

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}(\mathbf{x}_1) \\ \vdots \\ \mathbf{h}(\mathbf{x}_n) \end{bmatrix} = \begin{bmatrix} h(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & h(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & \ddots & \vdots \\ h(\mathbf{w}_1 \cdot \mathbf{x}_n + b_1) & \cdots & h(\mathbf{w}_L \cdot \mathbf{x}_n + b_L) \end{bmatrix}. \tag{13}$$

In (13), $\mathbf{h}(\mathbf{x}_i) = [h(\mathbf{w}_1 \cdot \mathbf{x}_i + b_1), \ldots, h(\mathbf{w}_L \cdot \mathbf{x}_i + b_L)]$ is the output of the hidden nodes in response to the input $\mathbf{x}_i$, which maps the data from d-dimensional input space to L-dimensional feature space. In most cases, the number of hidden neurons is much smaller than the number of training samples, i.e., $L \ll n$, the least squares solution of (12) described in [39] can be used

$$\boldsymbol{\beta}' = \mathbf{H}^{\dagger}\mathbf{Y} \tag{14}$$

where $\mathbf{H}^{\dagger}$ is the Moore–Penrose generalized inverse of matrix $\mathbf{H}$ and $\mathbf{H}^{\dagger} = \mathbf{H}^T(\mathbf{H}\mathbf{H}^T)^{-1}$. For better stability and generalization, a positive value $1/\rho$ is normally added to each diagonal element of $\mathbf{H}\mathbf{H}^T$. As a result, the output function of the ELM classifier is expressed as

$$f_L(\mathbf{x}_i) = \mathbf{h}(\mathbf{x}_i)\boldsymbol{\beta} = \mathbf{h}(\mathbf{x}_i)\mathbf{H}^T \left(\frac{\mathbf{I}}{\rho} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{Y}. \tag{15}$$

In ELM, the feature mapping $\mathbf{h}(\mathbf{x}_i)$ is assumed to be known. Recently, kernel-based ELM [31] has been proposed by extending explicit activation functions in ELM to implicit mapping functions, which have exhibited a better generalization

capability. If the feature mapping is unknown, a kernel matrix of ELM can be considered as

$$\Omega_{\text{ELM}} = \mathbf{H}\mathbf{H}^T : \Omega_{\text{ELM}_{i,j}} = \mathbf{h}(\mathbf{x}_i) \cdot \mathbf{h}(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j). \quad (16)$$

Hence, the output function of KELM is given by

$$f_L(\mathbf{x}_i) = \begin{bmatrix} K(\mathbf{x}_i, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}_i, \mathbf{x}_n) \end{bmatrix}^T \left(\frac{\mathbf{I}}{\rho} + \Omega_{ELM} \right)^{-1} \mathbf{Y}. \quad (17)$$

The input data label is finally determined according to the index of the output node with the largest value. In the experiments, the kernel version of ELM is implemented.

Compared to the standard SVM, which needs to solve a large constrained optimization problem, the training of ELM has only one analytical step. In the experiments, it will be demonstrated that ELM can provide a classification accuracy that is similar to or even better than that of SVM.

IV. Proposed Classification Framework

A. Band Selection

Hyperspectral images consist of a large number of spectral bands, but many of which contain redundant information. Band selection, such as LPE [24], reduces the dimensionality by selecting a subset of spectral bands with distinctive and informative features. Linear projections, such as PCA, can also transform the high-dimensional data into a lower dimensional subspace. In our previous study [17], [25], we investigated both LPE and PCA for spatial-feature-based hyperspectral image classification and found that the classification performance of LPE was superior to that of PCA. The reason may be that fine spatial structures tend to be present in minor PCs rather than in major PCs. Thus, band selection (i.e., LPE) is employed in this research.

LPE [24] is a simple while efficient band selection method based on band similarity measurement. Assume that there are two initial bands B_1 and B_2. For every other band B, an approximation can be expressed as $B' = a_0 + a_1 B_1 + a_2 B_2$, where a_0, a_1, a_2 are the parameters to minimize the LPE: $e = \|B - B'\|_2$. Let the parameter vector be $\mathbf{a} = [a_1, a_2, a_3]^T$. A least squares solution is employed to obtain $\mathbf{a} = (\mathbf{X}_{B_1B_2}^T \mathbf{X}_{B_1B_2})^{-1} \mathbf{X}_{B_1B_2}^T \mathbf{x}_B$, where $\mathbf{X}_{B_1B_2}$ is an $N \times 3$ matrix whose first column is with all **1**s, second column is the B_1-band, and third column is the B_2-band. Here, N is the total number of pixels, and $\mathbf{x}_B$ is the B-spectral band. The band that produces the maximum error e is considered as the most dissimilar band to B_1 and B_2, and it will be selected. Then, using these three bands, a fourth band can be found via the similar strategy and so on. More implementation details can be found in [24].

B. Spatial Feature Extraction

After band selection, the LBP feature extraction process or Gabor filtering is applied to each selected band image. Fig. 3 illustrates the implementation of LBP feature extraction. The input image is from the 63th band of the University of Pavia data to be introduced in Section V. In Fig. 3, the LBP code is first calculated for the entire image to form an LBP image, and the LBP features are then generated for the pixel of interest in its corresponding local LBP image patch. Note that patch size is a user-defined parameter, and classification performance with various patch sizes will be examined in Section V.

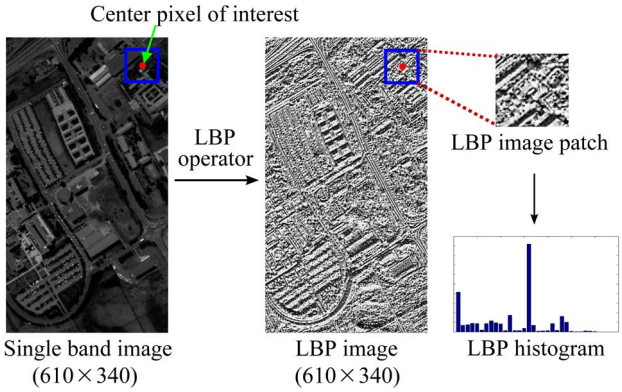

Fig. 3. Implementation of LBP feature extraction.

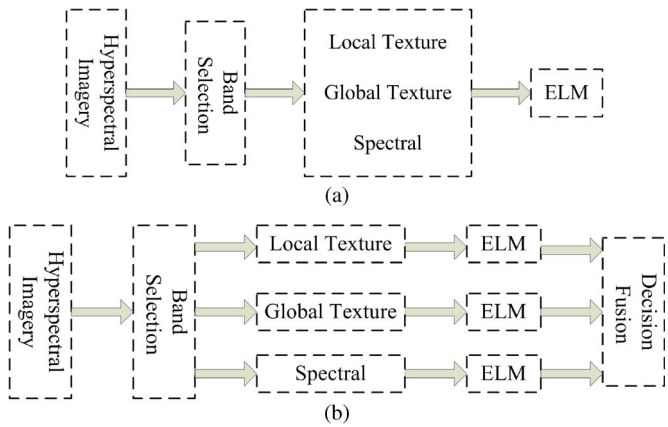

Fig. 4. Flowchart of the proposed classification framework. (a) Feature-level fusion. (b) Decision-level fusion.

TABLE I
Class Labels and Train-Test Distribution of Samples for the University of Pavia Data Set

#	*Class*	*Train*	*Test*
1	Asphalt	30	6601
2	Meadows	30	18619
3	Gravel	30	2069
4	Trees	30	3034
5	Painted Metal Sheets	30	1315
6	Bare Soil	30	4999
7	Bitumen	30	1300
8	Self-Blocking Bricks	30	3652
9	Shadows	30	917
	Total	270	42506

C. Feature-Level Fusion

In this paper, the most common feature-level fusion is first employed in the proposed classification framework, as shown in Fig. 4(a). Each feature reflects various properties and has its special meaning, such as the Gabor feature provides spatial localization and orientation selectivity, the LBP feature reveals

TABLE II
CLASS LABELS AND TRAIN-TEST DISTRIBUTION OF SAMPLES FOR THE INDIAN PINES DATA SET

#	Class	Train	Test
1	Alfalfa	6	48
2	Corn-notill	30	1404
3	Corn-mintill	30	804
4	Corn	24	210
5	Grass-pasture	30	467
6	Grass-trees	30	717
7	Grass-pasture-mowed	3	23
8	Hay-windrowed	30	459
9	Oats	2	18
10	Soybean-notill	30	938
11	Soybean-mintill	30	2438
12	Soybean-clean	30	584
13	Wheat	22	190
14	Woods	30	1264
15	Build-Grass-Trees-Drives	30	350
16	Stone-Steel-Towers	10	85
	Total	367	9999

TABLE III
CLASS LABELS AND TRAIN-TEST DISTRIBUTION OF SAMPLES FOR THE SALINAS DATA SET

#	Class	Train	Test
1	Brocoli-green-weeds-1	30	1979
2	Brocoli-green-weeds-2	30	3696
3	Fallow	30	1946
4	Fallow-rough-plow	30	1364
5	Fallow-smooth	30	2648
6	Stubble	30	3929
7	Celery	30	3549
8	Grapes-untrained	30	11241
9	Soil-vinyard-develop	30	6173
10	Corn-senesced-green-weeds	30	3248
11	Lettuce-romaine-4wk	30	1038
12	Lettuce-romaine-5wk	30	1897
13	Lettuce-romaine-6wk	30	886
14	Lettuce-romaine-7wk	30	1040
15	Vinyard-untrained	30	7238
16	Vinyard-vertical-trellis	30	1777
	Total	480	53649

the local image texture (e.g., edges, corners, etc.), and the spectral feature represents the correlation among bands. For various classification tasks, these features have their own advantages and disadvantages, and it is difficult to determine which one is always optimal [16]. Thus, it is straightforward to stack multiple features into a composite one. In this fusion strategy, feature normalization before feature stacking is a necessary preprocessing step to modify the scale of feature values. A simple treatment is to perform a linear transformation on these data and preserves the relationships among the values. For instance, a min–max technique maps all of the values into the range of [0, 1].

Here, three aforementioned features, i.e., LBP features (local texture), Gabor features (global texture), and selected bands (spectral features), and their combinations, such as LBP features + Gabor features + spectral features, LBP features + spectral features, Gabor features + spectral features, etc., will be discussed. Note that there are at least two potential disadvantages of feature-level fusion: 1) multiple feature sets to be stacked may be incompatible, which causes the induced feature space to be highly nonlinear, and 2) the induced feature space has a much larger dimensionality, which may deteriorate classification accuracy and processing efficiency.

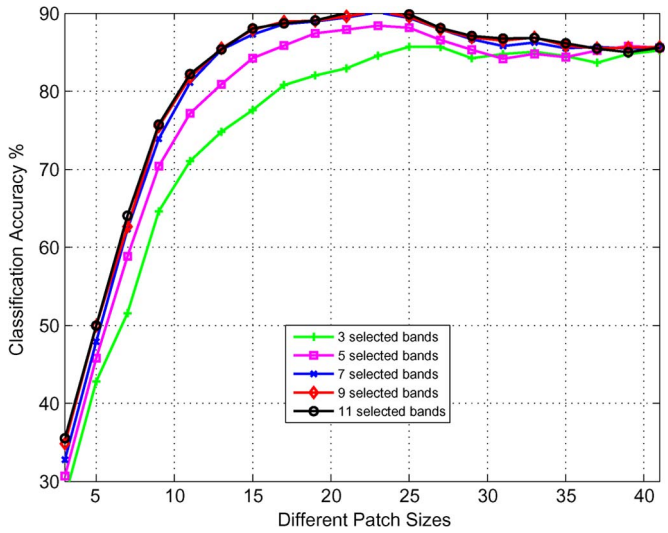

Fig. 5. LBP-ELM using the University of Pavia data: classification performance versus different patch sizes and numbers of selected bands.

TABLE IV
CLASSIFICATION ACCURACY (IN PERCENT) OF LBP-ELM WITH DIFFERENT PARAMETERS (m, r) OF THE LBP OPERATOR USING THE UNIVERSITY OF PAVIA DATA

m	4	6	8	10
r=1	86.14	87.99	88.53	88.84
r=2	81.37	86.57	**89.49**	89.31
r=3	84.59	87.10	88.11	87.55
Dim.	15	33	59	93
Cost	56.53	59.88	65.60	86.43

TABLE V
OPTIMAL PARAMETERS OF THE LBP OPERATOR AND GABOR FILTER FOR THE ELM CLASSIFIER USING THREE EXPERIMENTAL DATA

	# of Selected Bands	Patch Size	bw
University of Pavia			
LBP	7	21×21	–
Gabor	10	–	5
Indian Pines			
LBP	7	17×17	–
Gabor	7	–	1
Salinas			
LBP	8	25×25	–
Gabor	8	–	5

D. Decision-Level Fusion

Different from feature-level fusion, decision-level fusion [27], [29] is to merge results from a classifier ensemble of multiple features as shown in Fig. 4(b). This mechanism combines distinct classification results into a final decision, improving the accuracy of a single classifier that uses a certain type of features.

The objective here is to utilize the information of each type of features, compute the probability outputs by ELM, and then combine them with the soft LOGP for final decision. Since

TABLE VI
OVERALL CLASSIFICATION ACCURACY (IN PERCENT) AND KAPPA COEFFICIENT (κ) OF DIFFERENT TECHNIQUES FOR THE UNIVERSITY OF PAVIA DATA

Class	Spec-SVM	SVM-MRF	Gabor-SVM	LBP-SVM	Gabor-Spec-SVM	LBP-Spec-SVM	FF-SVM	DF-SVM	Spec-ELM	Gabor-ELM	LBP-ELM	Gabor-Spec-ELM	LBP-Spec-ELM	FF-ELM	DF-ELM
1	77.79	94.16	79.16	81.12	85.02	91.98	91.22	95.57	69.52	79.49	77.35	86.10	93.12	92.88	98.55
2	77.31	83.56	89.82	92.65	84.50	94.47	98.42	98.54	77.93	92.13	92.19	92.29	97.33	98.98	99.73
3	79.37	89.33	89.90	91.69	86.71	95.71	97.33	98.86	79.32	90.23	88.50	87.99	98.48	99.57	99.71
4	92.89	92.59	90.31	69.84	94.55	95.92	96.15	98.63	91.12	86.91	75.51	92.69	98.07	98.04	98.69
5	98.88	99.33	99.78	93.38	99.78	99.41	99.70	99.93	97.84	99.93	94.52	98.96	99.85	99.93	100
6	76.34	93.12	88.67	97.28	86.46	94.41	96.38	97.24	81.37	95.72	99.14	94.65	99.66	99.40	99.96
7	87.97	94.89	90.90	96.31	92.63	98.80	98.87	98.57	91.35	94.44	96.85	94.14	99.92	100	100
8	77.51	90.20	76.83	93.73	82.62	98.34	94.60	96.58	75.94	91.44	96.65	90.85	98.75	98.70	99.73
9	99.79	99.79	93.03	70.34	100	100	100	99.89	99.89	91.34	69.25	99.47	100	100	99.89
OA	80.01	89.04	87.37	89.47	86.56	94.99	96.61	97.85	79.40	90.37	89.43	91.73	97.40	98.11	**99.25**
κ	0.7440	0.8586	0.8354	0.8635	0.8261	0.9341	0.9552	0.9716	0.7375	0.8742	0.8633	0.8918	0.9658	0.9750	**0.9836**

TABLE VII
OVERALL CLASSIFICATION ACCURACY (IN PERCENT) AND KAPPA COEFFICIENT (κ) OF DIFFERENT TECHNIQUES FOR THE INDIAN PINES DATA

Class	Spec-SVM	SVM-MRF	Gabor-SVM	LBP-SVM	Gabor-Spec-SVM	LBP-Spec-SVM	FF-SVM	DF-SVM	Spec-ELM	Gabor-ELM	LBP-ELM	Gabor-Spec-ELM	LBP-Spec-ELM	FF-ELM	DF-ELM
1	88.89	71.69	98.15	98.15	100	100	100	100	75.93	98.15	100	100	100	100	100
2	76.50	80.47	83.61	88.49	89.05	93.03	92.12	91.84	71.34	83.96	88.35	90.03	91.35	91.03	91.28
3	78.18	90.05	91.85	96.28	94.84	98.08	97.96	95.56	74.58	93.53	97.00	96.88	97.12	98.44	97.96
4	80.34	97.86	100	97.44	99.15	96.58	98.72	100	82.91	100	100	100	100	100	100
5	94.16	94.97	97.59	98.79	97.79	97.99	98.59	97.99	95.17	97.99	98.59	98.39	98.39	99.20	98.59
6	93.78	99.46	95.58	98.13	94.78	97.86	97.19	97.99	94.91	95.98	98.93	98.53	98.53	98.93	99.06
7	76.92	63.57	88.46	92.13	96.15	96.15	96.15	96.15	69.23	88.46	100	96.15	96.15	96.15	100
8	98.77	99.80	100	100	100	100	100	100	99.39	100	100	100	100	100	100
9	50.00	0	100	100	100	100	100	100	55.00	100	100	100	100	100	100
10	69.73	92.87	86.26	86.16	86.67	87.09	88.12	88.84	64.26	86.98	90.19	87.91	89.46	89.36	89.98
11	57.29	77.84	74.35	82.62	76.01	77.55	79.29	83.51	51.09	77.39	85.21	78.97	85.85	85.36	88.49
12	79.15	98.05	79.32	82.08	79.64	84.85	81.43	81.27	86.32	80.29	79.32	79.32	80.29	85.78	79.80
13	99.53	100	99.53	99.53	99.53	99.53	99.06	99.13	99.53	99.53	100	99.53	100	100	99.53
14	72.33	76.51	91.58	95.05	99.07	98.15	99.46	99.61	78.98	93.43	97.76	98.61	99.85	99.00	99.69
15	80.79	100	98.42	98.68	97.63	98.68	99.74	100	87.63	99.74	100	100	100	100	100
16	96.84	98.95	94.74	96.84	100	100	100	100	90.53	94.74	96.84	97.89	100	100	100
OA	75.14	86.20	86.82	90.63	89.19	90.89	91.21	92.21	73.72	88.18	92.03	90.61	92.76	92.93	**93.58**
κ	0.7215	0.8438	0.8514	0.8939	0.8777	0.8959	0.9006	0.9122	0.7062	0.8667	0.9098	0.8939	0.9180	0.9271	**0.9199**

TABLE VIII
OVERALL CLASSIFICATION ACCURACY (IN PERCENT) AND KAPPA COEFFICIENT (κ) OF DIFFERENT TECHNIQUES FOR THE SALINAS DATA

Class	Spec-SVM	SVM-MRF	Gabor-SVM	LBP-SVM	Gabor-Spec-SVM	LBP-Spec-SVM	FF-SVM	DF-SVM	Spec-ELM	Gabor-ELM	LBP-ELM	Gabor-Spec-ELM	LBP-Spec-ELM	FF-ELM	DF-ELM
1	98.31	99.35	99.20	99.10	99.15	99.90	100	100	99.50	99.60	100	99.65	100	100	100
2	91.92	94.52	95.76	95.28	93.53	99.09	99.49	99.57	96.05	98.50	97.07	98.82	99.95	99.97	100
3	95.34	94.84	98.48	99.95	96.46	100	100	100	90.23	97.22	100	99.49	100	100	100
4	99.57	98.92	97.06	98.42	98.92	99.86	99.50	98.42	99.14	98.21	99.00	99.07	99.93	99.86	99.86
5	96.15	97.91	98.10	98.43	97.98	98.02	98.69	98.24	98.21	98.92	97.68	97.24	98.84	99.07	99.51
6	99.44	99.70	98.56	96.54	99.55	98.71	99.85	99.92	99.80	99.90	97.85	99.77	99.65	99.92	100
7	99.53	99.69	99.69	99.02	99.69	99.50	99.97	99.94	99.58	99.89	98.32	99.69	100	99.97	100
8	80.53	83.49	79.43	94.94	80.91	95.42	94.31	94.99	81.09	82.42	96.50	85.29	96.48	96.96	98.94
9	99.74	99.76	98.68	99.92	99.05	100	99.98	99.98	99.87	98.94	99.94	99.73	100	100	100
10	90.30	93.96	96.89	99.18	93.08	98.72	99.15	99.97	92.95	96.92	98.84	95.24	99.21	99.66	100
11	97.19	99.53	98.31	100	98.31	99.72	99.53	100	96.35	98.97	100	99.34	100	99.81	100
12	99.95	100	100	97.61	100	100	100	99.95	100	99.79	97.87	100	97.61	100	100
13	98.69	99.02	99.02	95.52	99.02	99.02	99.02	99.02	99.02	100	94.43	99.02	99.13	99.45	99.89
14	89.07	91.40	93.18	94.86	89.53	91.42	95.33	99.72	88.97	99.81	99.72	91.50	98.41	97.66	100
15	66.94	78.03	81.05	97.76	81.10	97.73	98.36	99.32	69.19	88.99	98.89	84.12	95.79	99.19	99.12
16	98.51	98.56	98.51	99.94	98.12	98.89	99.61	100	98.51	99.17	100	98.62	100	100	100
OA	89.48	92.21	91.93	97.53	91.85	98.14	98.29	98.67	90.33	94.16	98.26	93.93	98.43	99.12	**99.63**
κ	0.8830	0.9134	0.9104	0.9725	0.9095	0.9793	0.9809	0.9852	0.8924	0.9351	0.9806	0.9325	0.9826	0.9902	**0.9959**

the output function [i.e., (11)] of ELM estimates the accuracy of the predicted label and reflects the classifier confidence, the conditional class probability from the decision function is attempted to achieve. As noted by Platt [40], the probability should be higher for a larger output of the decision function. Platt's empirical analysis using scaling functions of the

following form is adopted

$$\mathbf{p}_q(y_k|\mathbf{x}) = \frac{1}{1+\exp\left(A_k f_L(\mathbf{x}) + B_k\right)} \tag{18}$$

where $\mathbf{p}_q(y_k|\mathbf{x})$ means the conditional class probability of the qth classifier, $f_L(\mathbf{x})$ is the output decision function of each ELM, and (A_k, B_k) are parameters estimated for ELM in class k $(1 \le k \le C)$. The parameters A_k and B_k are found by minimization of the cross-entropy error over the validation data. Note that A_k is negative.

In the proposed framework, LOGP [28], [29] uses the conditional class probabilities to estimate a global membership function $\mathbf{P}(y_k|\mathbf{x})$—a weighted product of these output probabilities. The final class label y is given according to

$$y = \arg \max_{k=1,\ldots,C} \mathbf{P}(y_k|\mathbf{x}) \tag{19}$$

where the global membership function is

$$\mathbf{P}(y_k|\mathbf{x}) = \prod_{q=1}^{Q} \mathbf{p}_q(y_k|\mathbf{x})^{\alpha_q} \tag{20}$$

or

$$\log \mathbf{P}(y_k|\mathbf{x}) = \sum_{q=1}^{Q} \alpha_q \mathbf{p}_q(y_k|\mathbf{x}) \tag{21}$$

with $\{\alpha_q\}_{q=1}^{Q}$ being the classifier weights uniformly distributed over all of the classifiers and Q being the number of pipelines (classifiers) in Fig. 4(b).

V. Experiments

A. Experimental Data

The first experimental data set was collected by the Reflective Optics System Imaging Spectrometer sensor. The image scene, with a spatial coverage of 610 × 340 pixels covering the city of Pavia, Italy, was collected under the HySens project managed by DLR (the German Aerospace Agency) [41]. The data set has 103 spectral bands prior to water-band removal. It has a spectral coverage from 0.43- to 0.86-μm and a spatial resolution of 1.3 m. Approximately 42 776 labeled pixels with nine classes are from the ground truth map. More detailed information of the number of training and testing samples is summarized in Table I.

The second data employed were acquired using National Aeronautics and Space Administration's Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor and was collected over northwest Indiana's Indian Pine test site in June 1992.[1] The image represents a classification scenario with 145 × 145 pixels and 220 bands in 0.4- to 2.45-μm region of visible and infrared spectrum with a spatial resolution of 20 m. The scenario contains two-thirds agriculture and one-third forest. In this paper, a total of 202 bands are used after removal of water absorption bands. There are 16 different land-cover classes, but not all are mutually exclusive in the designated ground truth map. The number of training and testing samples is shown in Table II.

TABLE IX
Statistical Significance From the Standardized McNemar's Test About the Difference Between Methods

University of Pavia	Indian Pines	Salinas
Z/significant?	Z/significant?	Z/significant?
DF-ELM *vs* DF-SVM		
23.78/yes	6.31/yes	19.76/yes
FF-ELM *vs* FF-SVM		
19.28/yes	9.48/yes	9.36/yes
LBP-Spec-ELM *vs* LBP-Spec-SVM		
23.52/yes	8.43/yes	4.76/yes
Gabor-Spec-ELM *vs* Gabor-Spec-SVM		
35.02/yes	6.97/yes	22.75/yes
LBP-ELM *vs* LBP-SVM		
-0.43/no	6.23/yes	12.78/yes
Gabor-ELM *vs* Gabor-SVM		
19.35/yes	5.88/yes	25.70/yes
FF-ELM *vs* LBP-Spec-ELM		
10.99/yes	1.14/no	15.05/yes

The third data were also collected by the AVIRIS sensor, capturing an area over Salinas Valley, CA, USA, with a spatial resolution of 3.7 m. The image comprises 512 × 217 pixels with 204 bands after 20 water absorption bands are removed. It mainly contains vegetables, bare soils, and vineyard fields.[2] There are also 16 different classes, and the number of training and testing samples is listed in Table III.

B. Parameter Tuning

In the proposed classification framework, some parameters, such as the number of selected bands of LPE [24], bw of Gabor filter, (m, r) of the LBP operator, and Gaussian kernel of ELM, are important; moreover, the patch size of the LBP operator affects the number of spatial LBP codes as shown in Fig. 3. Here, a fivefold cross validation based on available training samples is considered to tune the Gaussian kernel of ELM.[3]

For either feature or decision-level fusion presented in Section IV, the optimal parameters of the LBP features followed by ELM (namely, LBP-ELM) need to be estimated first. Take the University of Pavia data for example, (m, r) is fixed to be (8, 1), and then the impacts from different patch sizes and numbers of selected bands are investigated as shown in Fig. 5. Cross validation strategy is employed for tuning these parameters. It can be seen that the accuracy tends to be maximum with 7 or more selected bands and with 21 × 21 patch size. Note that, for each selected band, the dimensionality (i.e., number of bins) of the LBP features is $m(m-1)+3$ mentioned in [19] and [42]. Therefore, more selected bands will increase the dimensionality of the LBP features and computational complexity. The influence of parameter set (m, r) is then studied. In LBP, radius r defines the region for selecting circular neighbors, and m determines the dimensionality of the LBP histogram.

[1] https://engineering.purdue.edu/biehl/MultiSpec/hyperspectral.html

[2] http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

[3] http://www.ntu.edu.sg/home/egbhuang/elm_codes.html

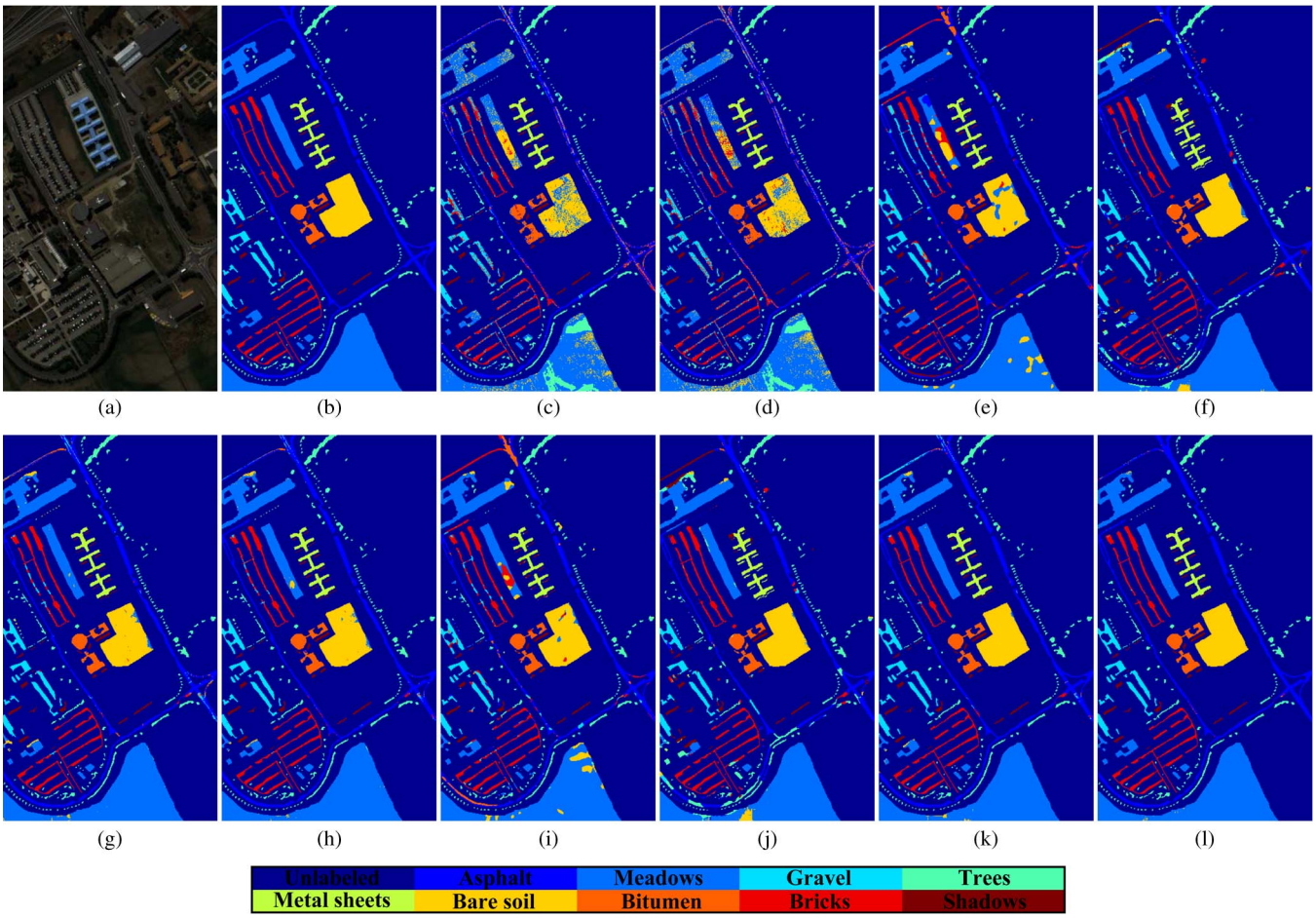

Fig. 6. Thematic maps resulting from classification for the University of Pavia data set with nine classes. (a) Pseudocolor image. (b) Ground truth map. (c) SVM: 80.01%. (d) ELM: 79.40%. (e) Gabor-SVM: 87.37%. (f) LBP-SVM: 89.54%. (g) FF-SVM: 96.61%. (h) DF-SVM: 97.85%. (i) Gabor-ELM: 90.37%. (j) LBP-ELM: 89.49%. (k) FF-ELM: 98.11%. (l) DF-ELM: 99.25%.

Table IV shows the classification accuracies of LBP-ELM with different (m, r) values. The accuracy tends to be stable when m is 8 or larger and insensitive to various r values. Due to the fact that spatially adjacent pixels probably belong to the same material, the radius r of the LBP operator should be kept small. *Dim* represents the dimensionality of the LBP features per band that is nearly the square of m; *Cost* (in *seconds*) indicates the computational time of LBP-ELM for different m values. Both *Dim* and *Cost* increase when m becomes greater. Based on the aforementioned observation, $(m, r) = (8, 2)$ is considered to be optimal in terms of classification accuracy and computational complexity. Similar tuning experiments using the Indian Pines data and Salinas data are performed, and relevant results are obtained. For the Gabor filter, a detailed discussion on its parameter tuning can be found in [17]. Here, choices of the number of selected bands and spatial frequency bandwidth bw are further evaluated for the ELM classifier. The optimal parameters for LBP and Gabor are summarized in Table V.

C. Classification Results

The performance of the proposed classification methods is shown in Tables VI–VIII for the three experimental data with different features and feature combinations, where FF-ELM indicates feature-level fusion based ELM using Gabor features, LBP features, and spectral feature concatenation, and DF-ELM indicates decision-level fusion based ELM that merges the probability outputs from a classifier ensemble of these three types of features.

From the results of each individual classifier, with LBP features, the performance is much better than that with the original spectral signatures only; for example, in Table VI, LBP-SVM offers over 9% higher accuracy than Spec-SVM, and LBP-ELM yields 10% higher accuracy than Spec-ELM. Moreover, LBP-feature-based classification methods (i.e., LBP-SVM and LBP-ELM) achieve much higher classification accuracies than Gabor-feature-based methods (i.e., Gabor-SVM and Gabor-ELM) for all three experimental data except LBP-ELM for the University of Pavia data. This clearly demonstrates that LBP is a highly discriminative spatial operator. We also include the experimental results of SVM-MRF [7] as comparison. SVM-MRF belongs to classification postprocessing strategy by a refinement of the labeling in a classified map in order to enhance its initial accuracy, which is different from the proposed scheme. From the results, SVM-MRF is superior to Spec-SVM but inferior to LBP-SVM. In particular, for classes with small sample size (e.g., *Oats*), SVM-MRF even deteriorates the classification accuracy from using original spectral bands only.

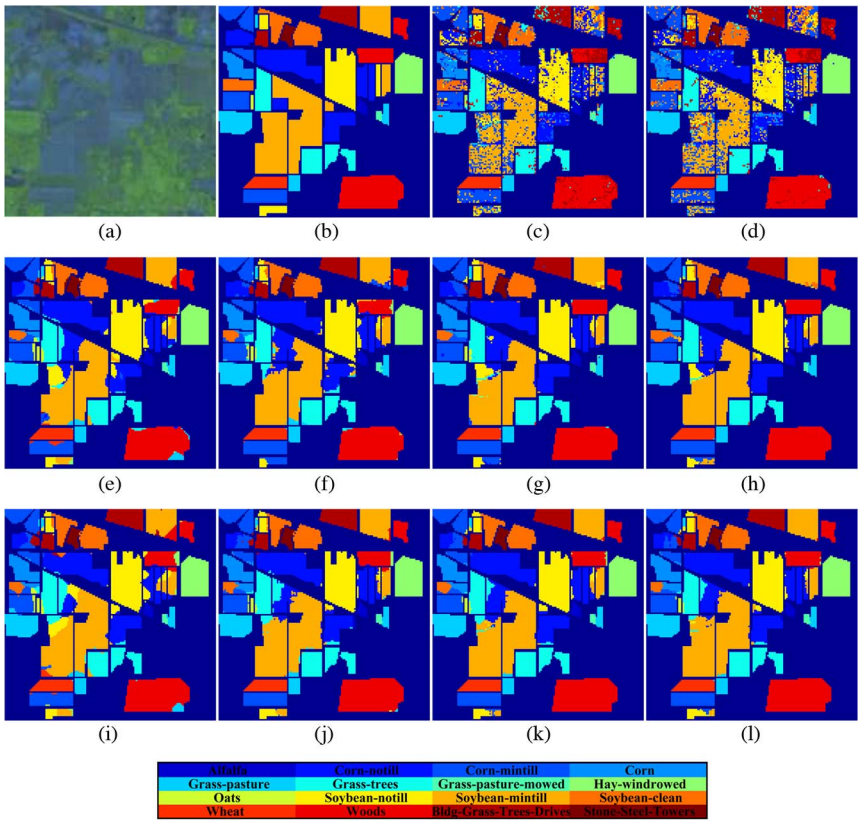

Fig. 7. Thematic maps resulting from classification for the Indian Pines data set with 16 classes. (a) Pseudocolor image. (b) Ground truth map. (c) SVM: 75.14%. (d) ELM: 73.72%. (e) Gabor-SVM: 86.82%. (f) LBP-SVM: 90.63%. (g) FF-SVM: 91.21%. (h) DF-SVM: 92.21%. (i) Gabor-ELM: 88.18%. (j) LBP-ELM: 92.03%. (k) FF-ELM: 92.93%. (l) DF-ELM: 93.58%.

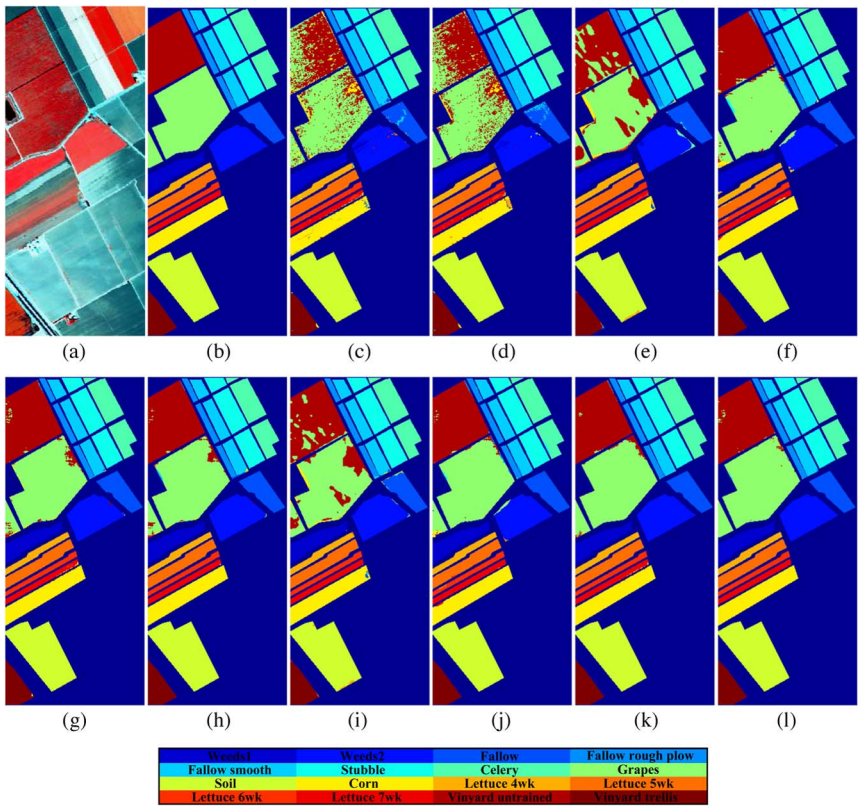

Fig. 8. Thematic maps resulting from classification for the Salinas data set with 16 classes. (a) Pseudocolor image. (b) Ground truth map. (c) SVM: 89.48%. (d) ELM: 90.33%. (e) Gabor-SVM: 91.93%. (f) LBP-SVM: 97.53%. (g) FF-SVM: 98.29%. (h) DF-SVM: 98.67%. (i) Gabor-ELM: 94.16%. (j) LBP-ELM: 98.26%. (k) FF-ELM: 99.12%. (l) DF-ELM: 99.63%.

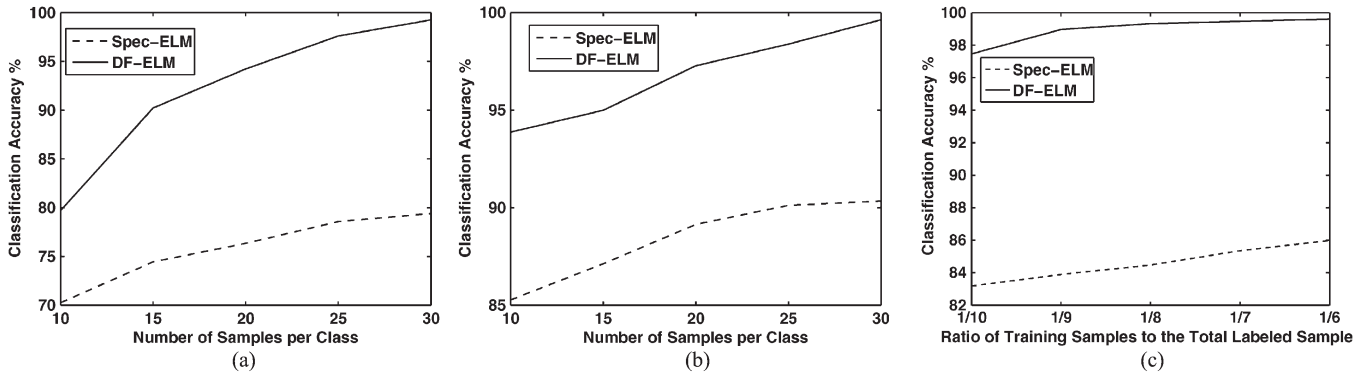

Fig. 9. Classification performance of Spec-ELM and DF-ELM with different numbers of training sample sizes for three experimental data. (a) University of Pavia. (b) Salinas. (c) Indian Pines.

As for feature combination, LBP features are found to be more compatible with spectral features than Gabor features. Specifically, by feature-level fusion of Gabor and spectral features (i.e., Gabor-Spec-SVM and Gabor-Spec-ELM), classification accuracy is only increased a little or even decreased as compared with that of using Gabor features only (i.e., Gabor-SVM and Gabor-ELM). For example, in Table VI, Gabor-SVM has an overall classification accuracy of 87.37% for the University of Pavia data. However, the accuracy of Gabor-Spec-SVM is decreased to 86.56%. On the other hand, the classification accuracy of LBP-Spec-SVM is a little higher than LBP-SVM. Gabor-Spec-ELM offers only 1.4% higher accuracy than Gabor-ELM, while LBP-Spec-ELM has about 8% higher accuracy than LBP-ELM.

Comparing feature-level fusion with decision-level fusion (i.e., FF-based methods and DF-based methods), DF-based methods yield superior performance than FF-based methods in all three experimental data. This is because feature-level fusion based methods may not be able to take advantage of the discriminative power of each individual features. As stated earlier, feature-level fusion has potential disadvantages, such as incompatibility of multiple feature sets and much larger dimensionality. Furthermore, from Tables VI–VIII, ELM generally has a better performance than SVM. Therefore, ELM-based classification methods turn out to be very effective for HSI classification under the small sample size condition. The standardized McNemar's test [2] is employed to verify the statistical significance in accuracy improvement of the proposed methods. As listed in Table IX, the Z values of the McNemar's test larger than 1.96 and 2.58 mean that two results are statistically different at the 95% and 99% confidence levels, respectively. The sign of Z indicates whether classifier 1 statistically outperforms classifier 2 ($Z > 0$) or vice versa. It further confirms performance improvement from the proposed classification framework with multiple features, ELM, and decision fusion.

Comparing LBP-Spec-ELM and FF-ELM, the latter outperforms the former for the University of Pavia data and Salinas data (also according to the Z values). However, for the Indian Pines data, the improvement is not statistically significant. From the tables, quantitative improvement from FF-ELM (which means LBP features + Gabor features + spectral features with ELM classifier) is not very large, compared to LBP-Spec-ELM, and the same phenomenon is for FF-SVM and LBP-Spec-SVM. This means that, if spatial features from LBP have been considered, then adding Gabor features may not be very critical (but performance is not degraded). However, if comparing Gabor-ELM with FF-ELM (or Gabor-SVM with FF-SVM), the improvement is always very significant. This means that local spatial features presented in LBP are more important. As a consequence, we believe that the overall excellent classification performance of the proposed framework is mainly due to the use of LBP and ELM.

Figs. 6–8 provide the classification maps of the aforementioned classifiers. These maps are consistent with the results shown in Tables VI–VIII. Maps generated from classification using spatial features (e.g., Gabor or LBP) are less noisy and more accurate than those from using spectral features. Moreover, LBP-based methods yield cleaner and smoother maps than Gabor-based methods. For example, the classification map of LBP-SVM [Fig. 8(f)] is more accurate than the map of Gabor-SVM [Fig. 8(e)]. The misclassification of Gabor-SVM mostly occurred between the class of *vinyard untrained* and the class of *grapes untrained*. This is probably because Gabor features are global ones that could not well capture local structures. The classification map of LBP-SVM exhibits spatial smoothness within every labeled class area.

Fig. 9 shows the influence of different training sample sizes. For the Indian Pines data, the training size is changed from 1/10 to 1/6 (note that 1/10 is the ratio of the number of training samples to the total labeled data), while for the University of Pavia data and the Salinas data, the training sample size is changed from 10 samples per class to 30 samples. It is obvious that the classification performance of DF-ELM is consistently much better than that of Spec-ELM. When ten samples per class are used for training, DF-ELM even reaches 94% accuracy for the Salinas data. Therefore, the conclusion is that classification accuracy can be greatly improved by integrating two complementary features, i.e., global Gabor features and local LBP features.

Finally, the computational complexity of the aforementioned classification methods is reported in Table X. All experiments were carried out using MATLAB on an Intel Core 2 Duo CPU machine with 4 GB of RAM. It should be noted that SVM is implemented in the libsvm package which uses MEX function to call C program in MATLAB, while ELM is

TABLE X
EXECUTION TIME (IN SECONDS) IN THE THREE EXPERIMENTAL DATA SETS

	University of Pavia	Indian Pines	Salinas
SVM	1.95	1.15	3.43
Gabor-SVM	33.26	3.75	21.37
LBP-SVM	60.97	8.43	167.56
FF-SVM	79.07	12.95	106.45
DF-SVM	99.20	20.31	157.77
ELM	1.13	0.73	4.37
Gabor-ELM	31.67	2.94	20.42
LBP-ELM	55.54	5.74	77.69
FF-ELM	71.05	8.73	82.14
DF-ELM	92.24	16.73	116.32

implemented purely in MATLAB. The computational cost of texture-feature-based methods is higher than spectral-feature-based methods due to the fact that they carry the burden of texture feature extraction. It is worth mentioning that extractions of Gabor features and LBP features are performed independently on each selected band, which means that the feature extraction procedure can potentially go parallel. Thus, the speed of texture feature extraction can be further improved.

VI. CONCLUSION

In this paper, a framework based on LBP has been proposed to extract local image features for classification of HSI. Specifically, LBP was implemented to a subset of original bands selected by the LPE method. Two types of fusion levels (i.e., feature and decision levels) were investigated on the extracted LBP features along with the Gabor features and the selected spectral bands. A soft-decision fusion process of ELM utilizing LOGP was also proposed to merge the probability outputs of multiple texture and spectral features. The experimental results have demonstrated that local LBP representations are effective in HSI spatial feature extraction, as they encode the information of image texture configuration while providing local structure patterns. Moreover, the decision-level fusion of kernel ELM can provide effective classification and is superior to SVM-based methods. Currently, feature-level fusion simply concatenates a pair of different features (i.e., Gabor features, LBP features, and spectral features) in the feature space. In our future work, more sophisticated feature selection approaches will be investigated.

REFERENCES

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

[2] 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.

[3] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosc. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.

[4] W. Li, S. Prasad, and J. E. Fowler, "Noise-adjusted subspace discriminant analysis for hyperspectral imagery classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1374–1378, Nov. 2013.

[5] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[6] G. Moser and S. B. Serpico, "Combining support vector machines and Markov random fields in an integrated framework for contextual image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2734–2752, May 2013.

[7] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method of accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.

[8] W. Li, S. Prasad, and J. E. Fowler, "Hyperspectral image classification using Gaussian mixture model and Markov random field," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 153–157, Jan. 2014.

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

[10] 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, Nov. 2008.

[11] C. Chen *et al.*, "Multihypothesis prediction for noise-robust hyperspectral image classification," *IEEE J. Sel. Topics Applies Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1047–1059, Apr. 2014.

[12] C. Chen, W. Li, E. W. Tramel, and J. E. Fowler, "Reconstruction of hyperspectral imagery from random projections using multihypothesis prediction," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 365–374, Jan. 2014.

[13] C. Chen and J. E. Fowler, "Single-image super-resolution using multihypothesis prediction," in *Proc. 46th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, USA, Nov. 2012, pp. 608–612.

[14] 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.

[15] P. Zhong and R. Wang, "A multiple conditional random fields ensemble model for urban area detection in remote sensing optical images," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 12, pp. 3978–3988, Dec. 2007.

[16] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 257–272, Jan. 2013.

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

[18] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.

[19] T. Ojala, M. Pietikainen, and T. T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary pattern," *IEEE Trans. Pattern Analysis Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[20] M. Musci, R. O. Feitosa, G. A. O. P. Costa, and M. L. F. Velloso, "Assessment of binary coding techniques for texture characterization in remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 6, pp. 1607–1611, Nov. 2013.

[21] D. T. Anderson, K. E. Stone, J. M. Keller, and C. J. Spain, "Combination of anomaly algorithms and image features for explosive hazard detection in forward looking infrared imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 1, pp. 313–323, Feb. 2013.

[22] M. Teutsch and G. Saur, "Segmentation and classification of man-made maritime objects in TerraSAR-X images," in *Proc. Int. Geosci. Remote Sens. Symp.*, Vancouver, BC, Canada, Jul. 2011, pp. 2657–2660.

[23] K. Masood and N. Rajpoot, "Texture based classification of hyperspectral colon biopsy samples using CLBP," in *IEEE Int. Symp. Biomed. Imag.*, Boston, MA, USA, Jul. 2009, pp. 1011–1014.

[24] Q. Du and H. Yang, "Similarity-based unsupervised band selection for hyperspectral image analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 4, pp. 564–568, Oct. 2008.

[25] K. Tan, E. Li, Q. Du, and P. Du, "Hyperspectral image classification using band selection and morphological profiles," *IEEE J. Sel. Topics Applies Earth Observ. Remote Sens.*, vol. 7, no. 1, pp. 40–48, Jan. 2014.

[26] O. Kuybeda, D. Malah, and M. Barzohar, "Rank estimation and redundancy reduction of high-dimensional noisy signals with preservation of rare vectors," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5579–5592, Dec. 2007.

[27] S. Prasad and L. M. Bruce, "Decision fusion with confidence-based weight assignment for hyperspectral target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1448–1456, May 2008.

[28] S. Prasad, W. Li, J. E. Fowler, and L. M. Bruce, "Information fusion in the redundant-wavelet-transform domain for noise-robust hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 9, pp. 3474–3486, Sep. 2012.
[29] W. Li, S. Prasad, and J. E. Fowler, "Decision fusion in kernel-induced spaces for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3399–3411, Jun. 2014.
[30] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern., Part B: Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.
[31] C. Chen, W. Li, H. Su, and K. Liu, "spectral–spatial classification of hyperspectral image based on kernel extreme learning machine," *Remote Sens.*, vol. 6, no. 6, pp. 5795–5814, Jun. 2014.
[32] M. Pal, A. E. Maxwell, and T. A. Warner, "Kernel-based extreme learning machine for remote sensing image classification," *Remote Sens. Lett.*, vol. 9, no. 4, pp. 852–862, Jun. 2013.
[33] R. Moreno, F. Corona, A. Lendasse, M. Grana, and L. S. Galvao, "Extreme learning machines for soybean classification in remote sensing hyperspectral images," *Neurocomputing*, vol. 128, no. 27, pp. 207–216, Mar. 2014.
[34] Y. Bazi *et al.*, "Differential evolution extreme learning machine for the classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 6, pp. 1066–1070, Jun. 2014.
[35] A. Samat, P. Du, S. Liu, J. Li, and L. Cheng, "E2LMs: Ensemble extreme learning machines for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1060–1069, Apr. 2014.
[36] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer. A*, vol. 2, no. 7, pp. 1160–1169, Jun. 1985.
[37] D. A. Clausi and M. E. Jernigan, "Designing Gabor filters for optimal texture separability," *Pattern Recogn.*, vol. 33, no. 11, pp. 1835–1849, Nov. 2000.
[38] S. Liao, M. W. K. Law, and A. C. S. Chung, "Dominant local binary patterns for texture classification," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 1107–1118, May 2009.
[39] G. B. Huang, Q. Y. Zhu, and C. K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, no. 3, pp. 489–501, Dec. 2006.
[40] J. Platt, "Advances in large margin classifiers," in *Probabilistic Outputs for Support Vector Machines and Comparison to Regularized Likelihood Methods*, A. Smola, ed. Cambridge, MA, USA: MIT Press, 1999.
[41] X. Huang and L. Zhang, "A comparative study of spatial approaches for urban mapping using hyperspectral ROSIS images over Pavia City, northern Italy," *Int. J. Remote Sens.*, vol. 30, no. 12, pp. 3205–3221, Jan. 2009.
[42] Z. Guo, L. Zhang, and D. Zhang, "Rotation invariant texture classification using LBP variance (LBPV) with global matching," *Pattern Recogn.*, vol. 43, no. 3, pp. 706–719, Mar. 2010.

Wei Li (S'11–M'13) received the B.E. degree in telecommunications engineering from Xidian University, Xi'an, China, in 2007, the M.S. degree in information science and technology from Sun Yat-Sen University, Guangzhou, China, in 2009, and the Ph.D. degree in electrical and computer engineering from Mississippi State University, Starkville, MS, USA, in 2012.

Subsequently, he spent 1 year as a Postdoctoral Researcher with the University of California, Davis, CA, USA. He is currently with the College of Information Science and Technology, Beijing University of Chemical Technology, Beijing, China. His research interests include statistical pattern recognition, hyperspectral image analysis, and data compression.

Dr. Li is an active Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE REMOTE SENSING LETTERS, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Chen Chen (S'10) received the B.E. degree in automation from Beijing Forestry University, Beijing, China, in 2009, the M.S. degree in electrical engineering from Mississippi State University, Starkville, MS, USA, in 2012. He is currently working toward the Ph.D. degree in the Department of Electrical Engineering, University of Texas at Dallas, Richardson, TX, USA.

His research interests include compressed sensing, signal and image processing, pattern recognition, computer vision, and hyperspectral image analysis.

Hongjun Su (S'09–M'11) received the B.E. degree in geography information system from the China University of Mining and Technology, Xuzhou, China, in 2006, and the Ph.D. degree in cartography and geography information system from the Key Laboratory of Virtual Geographic Environment (Ministry of Education), Nanjing Normal University, China, in 2011.

From 2009 to 2010, he was a Visiting Student with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. He is now with the School of Earth Sciences and Engineering, Hohai University, Nanjing. He is a Postdoctoral Fellow with the Department of Geographical Information Science, Nanjing University. His main research interests include hyperspectral remote sensing dimensionality reduction, classification, and spectral unmxing.

Dr. Su is an active Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE REMOTE SENSING LETTERS, the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, the *International Journal of Remote Sensing*, and *Optics Express*.

Qian Du (S'98–M'00–SM'05) received the Ph.D. degree in electrical engineering from the University of Maryland Baltimore County, Baltimore, MD, USA, in 2000.

Currently, she is a Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Mississippi State, MS, USA. Her research interests include hyperspectral remote sensing image analysis, pattern classification, data compression, and neural networks.

Dr. Du has served as Cochair for the Data Fusion Technical Committee of IEEE Geoscience and Remote Sensing Society during 2009–2013 and during 2010–2014 as the Chair for the Remote Sensing and Mapping Technical Committee of the International Association for Pattern Recognition. She also serves as the Associate Editor for IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE SIGNAL PROCESSING LETTERS, and the *Journal of Applied Remote Sensing*. She was the recipient of the 2010 Best Reviewer award from IEEE Geoscience and Remote Sensing Society. She is the General Chair for the Fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS) in Shanghai, China, in 2012. She is a Fellow of SPIE.