Spectral–Spatial Hyperspectral Image Classification With Edge-Preserving Filtering

Xudong Kang, *Student Member, IEEE*, Shutao Li, *Member, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

Abstract—The integration of spatial context in the classification of hyperspectral images is known to be an effective way in improving classification accuracy. In this paper, a novel spectral–spatial classification framework based on edge-preserving filtering is proposed. The proposed framework consists of the following three steps. First, the hyperspectral image is classified using a pixelwise classifier, e.g., the support vector machine classifier. Then, the resulting classification map is represented as multiple probability maps, and edge-preserving filtering is conducted on each probability map, with the first principal component or the first three principal components of the hyperspectral image serving as the gray or color guidance image. Finally, according to the filtered probability maps, the class of each pixel is selected based on the maximum probability. Experimental results demonstrate that the proposed edge-preserving filtering based classification method can improve the classification accuracy significantly in a very short time. Thus, it can be easily applied in real applications.

Index Terms—Classification, edge-preserving filters (EPFs), hyperspectral data, spatial context.

I. Introduction

SATELLITE images captured by hyperspectral sensors often have more than 100 spectral bands for each pixel. Therefore, a hyperspectral image can provide very informative spectral information regarding the physical nature of the different materials which can be utilized to distinguish objects in the image scene. However, the high dimensionality of hyperspectral data may produce the Hughes phenomenon [1] and thus may influence the performance of supervised classification methods.

During the last decade, a large number of feature extraction, feature reduction, and combination techniques [2]–[5] have been proposed to address the high-dimensionality problem. Furthermore, intensive work has been proposed to build accurate pixelwise classifiers for the analysis of hyperspectral

X. Kang is with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China, and also with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavík, Iceland (e-mail: xudong_kang@163.com).

S. Li is with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: shutao_li@hnu.edu.cn).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavk, Iceland (e-mail: benedikt@hi.is).

images [6], e.g., random forests [7], [8], neural networks [9], [10], AdaBoost [11], support vector machines (SVMs) [12], sparse representation [13], [14], and active learning [15]–[17] methods. Among these methods, the SVM classifier has, in particular, shown a good performance in terms of classification accuracy [12] since it has two major advantages: First, it requires relatively few training samples to obtain good classification accuracies; second, it is robust to the spectral dimension of hyperspectral images [12].

To improve the classification performance further, many researchers have worked on spectral–spatial classification which can incorporate the spatial contextual information into the pixelwise classifiers [18]. For example, extended morphological profiles (EMPs) have been proposed for constructing spectral–spatial features [19] which are adaptive definitions of the neighborhood of pixels. Furthermore, spectral–spatial kernels, e.g., composite [20], morphological [21], and graphic [22] kernels, have also been proposed for the improvement of the SVM classifier. The morphological and kernel-based methods have given good results in terms of accuracies for classifying hyperspectral images [20]–[25].

Another family of spectral–spatial classification methods is based on image segmentation [18], [26]. A hyperspectral image is first segmented into different regions based on the homogeneity of either intensity or texture [27] so that all the pixels within the same region can be considered as a spatial neighborhood. Different hyperspectral segmentation techniques such as partitional clustering [28], watershed [29], hierarchical segmentation [30], and minimum spanning forest [31] have been proposed for this objective. Then, majority voting is applied for combining the pixelwise classification results obtained by a pixelwise classifier with the segmentation map obtained by image segmentation. Specifically, all the pixels in the same region are assigned to the most frequent class within this region. Furthermore, multiple spectral–spatial classification methods can be combined together to improve the classification accuracy further [30]. However, segmentation-based methods rely on the performance of the segmentation techniques. In order to get accurate segmentation results, advanced segmentation methods are usually required, but those methods may be time-consuming.

Different from the traditional segmentation-based spectral–spatial classification framework described previously, Li *et al.* proposed that the spectral–spatial classification problem can be solved in a Bayesian framework [17], [32]. The final classification result is obtained from a posterior distribution built on the class distributions which are learned from multinomial logistic regression and multilevel logistic (MLL) prior.

Li *et al.*'s method has given better performances in terms of classification accuracies when compared with several widely used spectral–spatial classification methods [17], [32].

Recently, edge-preserving filtering [33]–[37] has become a very active research topic of image processing and has been applied in many applications such as high dynamic imaging [36], stereo matching [38], image fusion [39], [40], enhancing [41], dehazing [42], and denoising [43]. In hyperspectral image analysis, edge-preserving filtering has also been successfully applied for hyperspectral image visualization [44].

In this paper, an edge-preserving filtering based spectral–spatial classification framework is proposed. The pixelwise classification map obtained by the SVM classifier is first represented as multiple probability maps (the probability that a pixel belongs to a specified class) and then processed with edge-preserving filtering. Finally, the class of each pixel is selected based on the maximum probability. Taking neighborhood information into account, edge-preserving filtering can smooth the probabilities while ensuring that the smoothed probabilities are aligned with real object boundaries. Experiments show that the proposed method can improve the classification accuracy of SVM significantly with much less computer resources needed. Thus, the proposed method will be quite useful in real applications.

The rest of this paper is organized as follows. Section II introduces two widely used edge-preserving filters (EPFs). Section III describes the proposed spectral–spatial classification method. Section IV gives the results and provides a discussion. Finally, the conclusion is given in Section V.

II. EDGE-PRESERVING FILTERING

During the last decade, many different EPFs, e.g., the joint bilateral filter [33], the weighted least-squares (WLS) filter [36], the guided filter [34], the domain transform filter [45], the local linear Stein's unbiased risk estimate filter [46], and the L_0 gradient filter [47], have been proposed. Most of these EPFs have a similar property, i.e., they can be used for joint filtering, where the content of one image is smoothed based on the edge information from a guidance image. It means that the spatial information of the guidance image is able to be considered in the filtering process. In this section, the two most widely used EPFs, i.e., the joint bilateral filter and the guided filter, are illustrated.

A. Joint Bilateral Filter

The joint bilateral filter is based on the widely used Gaussian filter considering the distance in the image plane (the spatial domain) and the distance in the intensity axis (the range domain). The spatial and range distances are defined using two Gaussian decreasing functions, i.e., $G_{\delta_s}(\|i-j\|) = \exp(-(\|i-j\|)/\delta_s^2)$ and $G_{\delta_r}(|I_i - I_j|) = \exp(-|I_i - I_j|^2/\delta_r^2)$. Specifically, the filtering output O_i of the input pixel P_i can be represented as a weighted average of its neighborhood pixels P_j as follows:

$$O_i = \frac{1}{K_i^b} \sum_{j \in \omega_i} G_{\delta_s}(\|i-j\|) G_{\delta_r}(|I_i - I_j|) P_j$$

with

$$K_i^b = \sum_{j \in \omega_i} G_{\delta_s}(\|i-j\|) G_{\delta_r}(|I_i - I_j|) \tag{1}$$

where i and j represent the ith and jth pixels, ω_i is a local window of size $(2\delta_s + 1) \times (2\delta_s + 1)$ around pixel i, K^b is a normalizing term of the joint bilateral filter, P and I are the input image and the guidance image, respectively, δ_s controls the size of the local window used to filter a pixel, and δ_r defines how much the weight of a pixel decreases because of the intensity difference between the reference pixels, i.e., I_i and I_j.

Based on (1), it is easy to imagine that, if the neighborhood pixels of pixel i in the guidance image have similar intensities or colors, i.e., $I_i \approx I_j$, the weight of pixel j, represented by neighboring pixel j, will be quite large, especially when it is very close to i, i.e., $\|i-j\|$ is very small. In contrast, if the neighboring pixels have quite different intensities in the guidance image, the situation will be the opposite. This means that those adjacent input pixels which have similar intensities or colors in the guidance image tend to have similar outputs.

B. Guided Filter

The guided filter is based on a local linear model which assumes that the filtering output O can be represented as a linear transform of the guidance image I in a local window ω of size $(2r + 1) \times (2r + 1)$ as follows:

$$O_i = a_j I_i + b_j, \quad \forall i \in \omega_j. \tag{2}$$

This model ensures $\nabla O \approx a \nabla I$, which means that the filtering output O will have an edge only if the guidance image I has an edge. To determine the coefficients a_j and b_j, an energy function [see (3)] is constructed as follows:

$$E(a_j, b_j) = \sum_{i \in \omega_j} \left((a_j I_i + b_j - P_i)^2 + \epsilon a_j^2 \right) \tag{3}$$

where ϵ is a regularization parameter deciding the degree of blurring for the guided filter. The energy function is based on two goals. First, the filtering output, i.e., $(a_j I_i + b_j)$, should be as close as possible to the input image P. Second, the local linear model should be maintained in the energy function. By solving the energy function, abrupt intensity changes in the guidance image I can be mostly preserved in the filtering output O.

C. Comparison of the Two EPFs

Given the aforementioned simple description of the two EPFs, the properties of the EPFs are studied in Fig. 1. Fig. 1(a) and (c) shows the input image P and the guidance image I, respectively. Fig. 1(b) and (d) shows the filtering outputs obtained by different EPFs with different parameter settings.

Regarding the joint bilateral filter (see Fig. 1(b), from left to right), the two parameters, i.e., δ_s and δ_r, are set as follows: $\delta_s = 2$, $\delta_r = 0.01$; $\delta_s = 2$, $\delta_r = 0.1$; $\delta_s = 4$, $\delta_r = 0.1$;

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

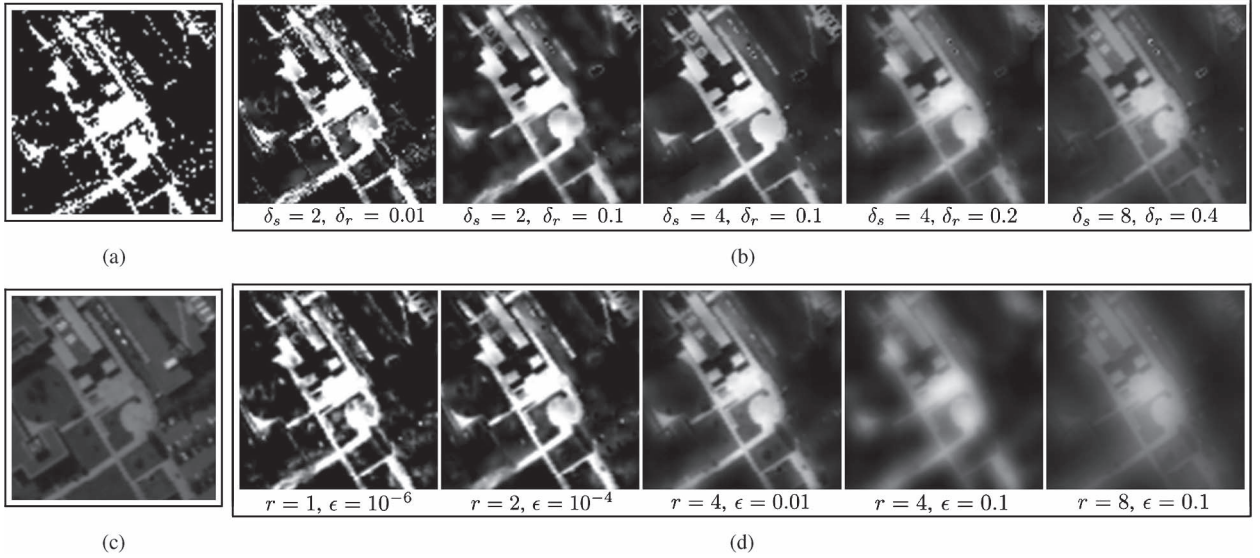KANG *et al.*: SPECTRAL–SPATIAL HYPERSPECTRAL IMAGE CLASSIFICATION

3

Fig. 1. (a) Input image. (b) and (d) Filtered images obtained by different EPFs with different parameter settings. (c) Guidance image.
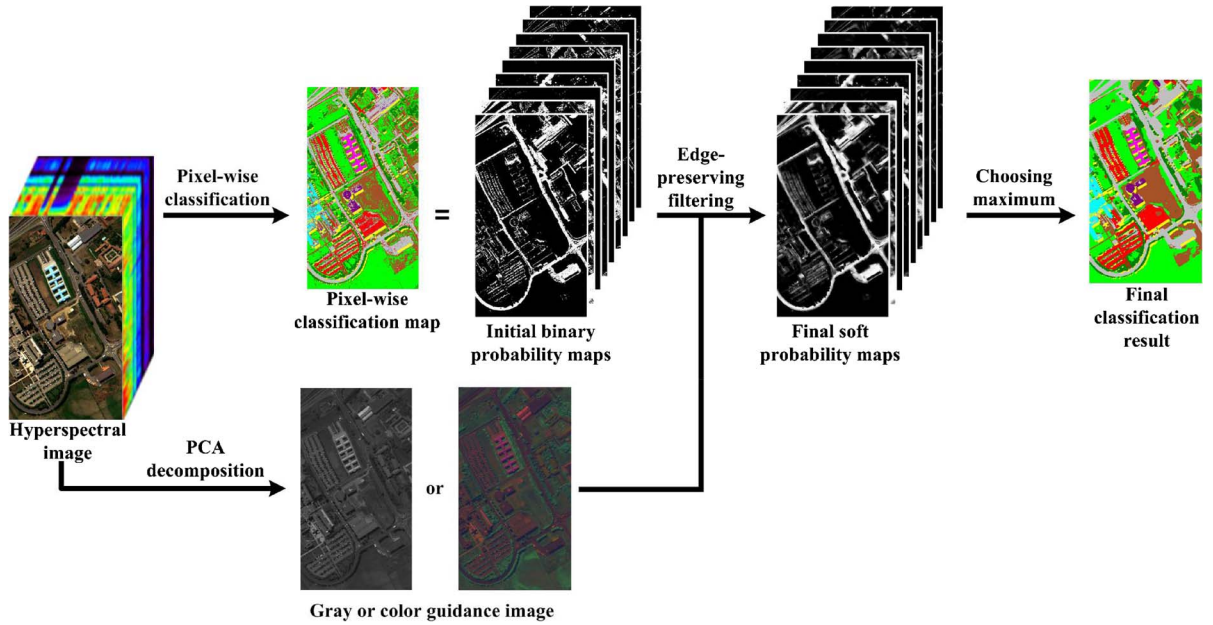

Fig. 2. Schematic of the proposed edge-preserving spectral–spatial classification method.

$\delta_s = 4, \delta_r = 0.2$; and $\delta_s = 8, \delta_r = 0.4$, respectively. Regarding the guided filter (see Fig. 1(d), from left to right), the two parameters, i.e., r and ϵ, are set as follows: $r = 1, \epsilon = 10^{-6}$; $r = 2, \epsilon = 10^{-4}$; $r = 4, \epsilon = 0.01$; $r = 4, \epsilon = 0.1$; and $r = 8$, $\epsilon = 0.1$, respectively. The figure shows that, as the filter size (δ_s and r) and the degree of blurring (δ_r and ϵ) increase, a more obvious smoothing will be produced on the filtering outputs. More importantly, compared with the input image, the pixels in the filtering outputs tend to be aligned with edges in the guidance image. In other words, adjacent input pixels with similar intensities in the guidance image tend to get similar filtering outputs. This property is quite correlated with image segmentation which aims at dividing an image into different regions based on the homogeneity of intensity or color.

III. SPECTRAL–SPATIAL CLASSIFICATION WITH EDGE-PRESERVING FILTERING

Here, the spectral–spatial classification problem is considered as a probability optimization process, as illustrated in Fig. 2. The initial probability that a pixel belongs to a specified class is estimated based on a widely used pixelwise classifier, i.e., the SVM classifier. Taking spatial contextual information into account, the final probabilities are obtained by performing edge-preserving filtering on the initial probability maps, with the principal components of the hyperspectral image serving as the guidance image. Specifically, if the nth probability map has quite a large intensity in location i which refers to a higher probability, then the ith pixel of the hyperspectral image tends to belong to the nth class.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

4   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

A. Problem Formulation

The general supervised hyperspectral classification problem can be formulated as follows: Let $\mathcal{S} \equiv \{1, \ldots, i\}$ denote the set of pixels of the hyperspectral image, $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_i) \in \mathbb{R}^{d \times i}$ denote an image of d-dimensional feature vectors, $\mathcal{L} \equiv \{1, \ldots, n\}$ be a set of labels, and $\mathbf{c} = (c_1, \ldots, c_i)$ be the classification map of labels. Given a training set $T_\tau \equiv \{(\mathbf{x}_1, c_1), \ldots, (\mathbf{x}_\tau, c_\tau)\} \in (\mathbb{R}^d \times \mathcal{L})^\tau$, where τ is the total number of training samples, the goal of classification is to obtain a classification map, i.e., $\mathbf{c}$, which assigns a label $c_i \in \mathcal{L}$ to each pixel $i \in \mathcal{S}$.

B. Proposed Approach

The proposed approach consists of three steps: 1) construction of the initial probability maps; 2) filtering of the probability maps; and 3) classification based on the maximum probability.

1) Construction of the Initial Probability Maps: It is known that an initial classification map $\mathbf{c}$ can be obtained by a pixelwise classifier. In this paper, the pixelwise classification map $\mathbf{c}$ is represented using probability maps, i.e., $\mathbf{p} = (\mathbf{p}_1, \ldots, \mathbf{p}_n)$, in which $\mathbf{p}_{i,n} \in [0,1]$ is the initial probability that a pixel i belongs to the nth class. Specifically, the probability $\mathbf{p}_{i,n}$ is defined as follows:

$$\mathbf{p}_{i,n} = \begin{cases} 1 & \text{if } \mathbf{c}_i = n \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

The SVM classifier is adopted for pixelwise classification since it is one of the most widely used pixelwise classifiers and has been successfully applied for many other spectral–spatial classification methods [6], [18], [48].

2) Filtering of the Probability Maps: Initially, spatial information is not considered. All probabilities are valued at either 0 or 1. Therefore, the probability maps appear noisy and not aligned with real object boundaries. To solve this problem, the probability maps are optimized by edge-preserving filtering. Specifically, the optimized probabilities are modeled as a weighted average of its neighborhood probabilities

$$\acute{\mathbf{p}}_{i,n} = \sum_j W_{i,j}(I)\mathbf{p}_{j,n} \quad (5)$$

where i and j represent the ith and jth pixels and the filtering weight W is chosen such that the filter preserves edges of a specified guidance image I. Therefore, this step has two major problems: 1) how to choose an EPF and 2) how to choose a guidance image. Different filters and guidance images produce different filtering weights $W_{i,j}$ for (5).

For the choice of EPF, two widely used EPFs are adopted in this paper. The weights $W_{i,j}$ obtained by the two filters are reviewed as follows.

1) Filtering weight for the joint bilateral filter: the weight of the joint bilateral filter is already presented in (1), i.e., $W_{i,j}(I) = 1/K_i^b \sum_{j \in \omega_i} G_{\delta_s}(\|i - j\|)G_{\delta_r}(|I_i - I_j|)$. Based on the corresponding description in Section II, it is easy to know that those adjacent input probabilities which have similar intensities or colors in the guidance image will have similar outputs after filtering.
2) Filtering weight for the guided filter: as described in [37], (2) can be represented in a weighted average form as (5),
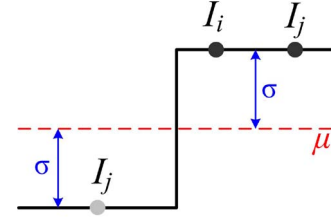

Fig. 3. Example of 1-D step edge. Here, μ and σ are shown for a filtering kernel centered exactly at an edge.

and the filtering weight $W_{i,j}(I)$ of the guided filter can be expressed as follows:

$$W_{i,j}(I) = \frac{1}{|\omega|^2} \sum_{k \in \omega_i, k \in \omega_j} \left(1 + \frac{(I_i - \mu_k)(I_j - \mu_k)}{\sigma_k^2 + \epsilon}\right) \quad (6)$$

where ω_i and ω_j are local windows around pixel i and j, respectively, μ_k and σ_k are the mean and variance of I in ω_k, and $|\omega|$ is the number of pixels in ω_k. A 1-D step edge example is presented in Fig. 3 to demonstrate the edge-preserving property of the filtering weight for the guided filter. As shown in the figure, if I_i and I_j are on the same side of an edge, the term $(I_i - \mu_k)(I_j - \mu_k)$ in (6) will have a positive sign. However, if I_j is located on the other side of the edge, the term will have a negative sign. Thus, the filtering weight becomes larger for pixel pairs on the same side of the edge but small otherwise. Hence, those probabilities on the same side of an edge in the guidance image I tend to have similar filtering outputs.

Regarding the choice of the guidance image I, principal component analysis (PCA) is adopted because it gives an optimal representation of the image in the mean squared sense. Here, two options are given as follows.

1) Gray-scale guidance image: the PCA decomposition is conducted on the original hyperspectral image, and the first principal component which contains most of the edge information is adopted as the guidance image (see Fig. 2).
2) Color guidance image: instead of guiding the filtering with a gray-scale image, the first three principal components are used as the color guidance image of the EPFs (see Fig. 2).

Fig. 4 shows an example of probability filtering. It shows that the initial probabilities look noisy and are not aligned with real object boundaries. Probability optimization with edge-preserving filtering has two major advantages in this example [see Fig. 4(b)]. First, noise probabilities that appear as scattered points or lines can be effectively smoothed. Second, the refined probabilities are always aligned with real object boundaries. The two advantages demonstrate that the spatial contextual information of the guidance image is well utilized in the edge-preserving filtering process.

3) Classification Based on the Maximum Probability: According to (7), once the probability maps are filtered, the label at pixel i can be simply chosen in a maximization manner as follows:

$$\acute{\mathbf{c}}_i = \arg \max_n \acute{\mathbf{p}}_{i,n}. \quad (7)$$

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

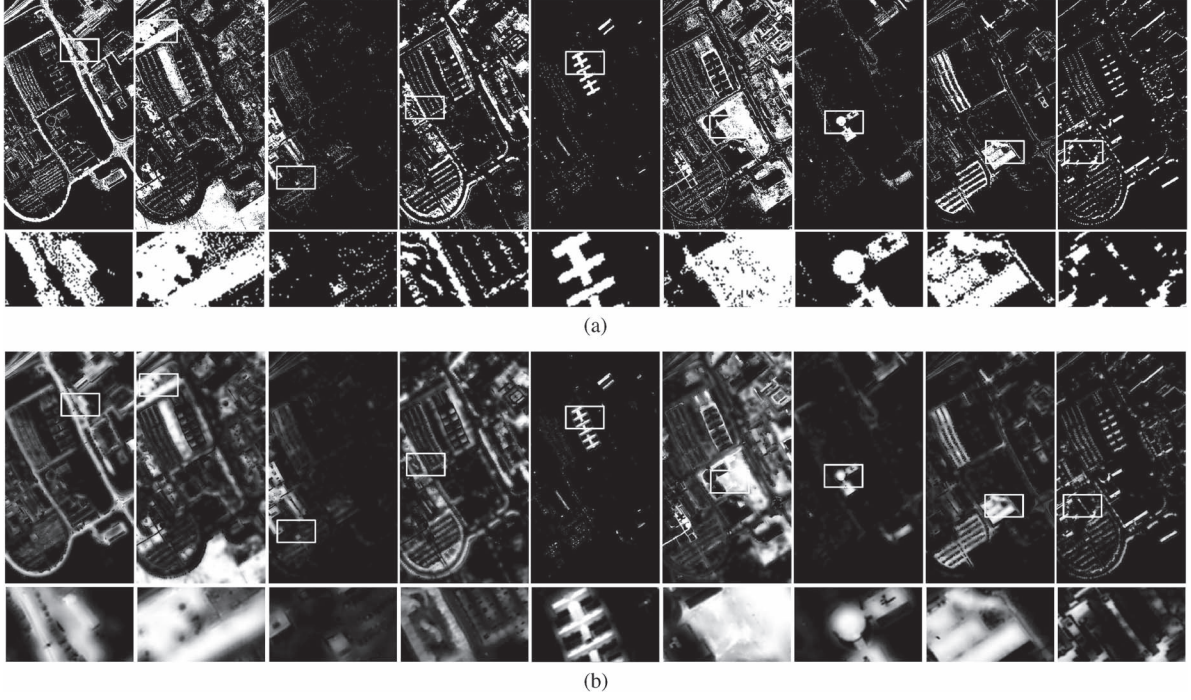KANG *et al.*: SPECTRAL–SPATIAL HYPERSPECTRAL IMAGE CLASSIFICATION

5

Fig. 4. (a) Initial probability maps for the University of Pavia data set. (b) Final probability maps obtained by edge-preserving filtering. The lower figures in (a) and (b) show the close-up views of the probability maps denoted by boxes in the upper figures.

This step aims at transforming the probability maps $\acute{\mathbf{p}}_n$ into the final classification result $\acute{c}$.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Experimental Setup

1) Data Sets: The proposed method is performed on three hyperspectral data sets, i.e., the Indian Pines image, the University of Pavia, and the Salinas image. The Indian Pines image capturing the agricultural Indian Pine test site of North-western Indiana was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. The image has 220 bands of size 145×145 with a spatial resolution of 20 m per pixel and a spectral coverage ranging from 0.4 to 2.5 μm (20 water absorption bands no. 104–108, 150–163, and 220 were removed before experiments). Fig. 5 shows the color composite of the Indian Pines image and the corresponding ground truth data.

The University of Pavia image capturing an urban area surrounding the University of Pavia was recorded by the ROSIS-03 satellite sensor. The image has 115 bands of size 610×340 with a spatial resolution of 1.3 m per pixel and a spectral coverage ranging from 0.43 to 0.86 μm (12 most noisy channels were removed before experiments). Nine classes of interest are considered for this image. Fig. 6 shows the color composite of the University of Pavia image and the corresponding ground truth data.

The Salinas image was captured by the AVIRIS sensor over Salinas Valley, CA, USA, and with a spatial resolution of 3.7 m per pixel. The image has 224 bands of size 512×217. As with the Indian Pines and University of Pavia scenes, 20 water absorption bands no. 108–112, 154–167, and 224 were discarded. Fig. 7 shows the color composite of the Salinas image and the corresponding ground truth data. For the three images,
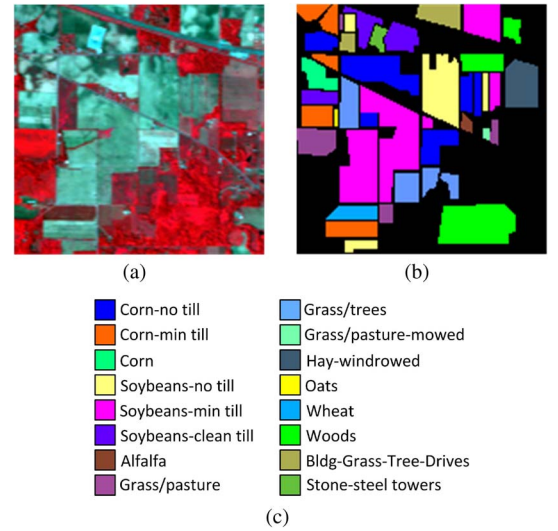

Fig. 5. (a) Three-band color composite of the Indian Pines image. (b) and (c) Ground truth data of the Indian Pines image.

Legend:
- Corn-no till
- Corn-min till
- Corn
- Soybeans-no till
- Soybeans-min till
- Soybeans-clean till
- Alfalfa
- Grass/pasture
- Grass/trees
- Grass/pasture-mowed
- Hay-windrowed
- Oats
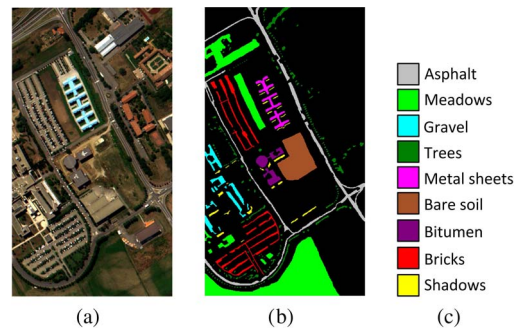- Wheat
- Woods
- Bldg-Grass-Tree-Drives
- Stone-steel towers

Fig. 6. (a) Three-band color composite of the University of Pavia image. (b) and (c) Ground truth data of the University of Pavia image.

Legend:
- Asphalt
- Meadows
- Gravel
- Trees
- Metal sheets
- Bare soil
- Bitumen
- Bricks
- Shadows

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

6   IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING
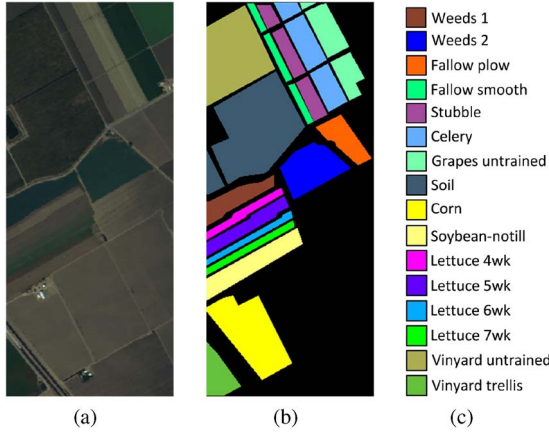

Fig. 7. (a) Three-band color composite of the Salinas image. (b) and (c) Ground truth data of the Salinas image.

the number of training and test samples for each class is detailed in Tables II–IV, respectively.

2) Quality Indexes: Three widely used quality indexes, i.e., the overall accuracy (OA), the average accuracy (AA), and the kappa coefficient, are adopted to evaluate the performance of the proposed method. OA is the percentage of correctly classified pixels. AA is the mean of the percentage of correctly classified pixels for each class. The kappa coefficient gives the percentage of correctly classified pixels corrected by the number of agreements that would be expected purely by chance.

B. Classification Results

1) Analysis of the Influence of Parameters: For the proposed joint bilateral filtering-based technique, the parameters δ_s and δ_r determine the filtering size and blur degree, respectively. Similarly, the parameters r and ϵ denote the filtering size and blur degree of the guided filter, respectively. The influence of these parameters on the classification performance is analyzed in Figs. 8 and 9 (experiment is performed on the Indian Pines image). In the experiment, the training set which accounts for 10% of the ground truth was chosen randomly (see Table II). The OA, AA, and kappa of the proposed method are measured with different parameter settings. When the influence of δ_s is analyzed, δ_r is fixed to be 0.2. Similarly, for the guided filter, when the influence of r is analyzed, ϵ is fixed to be 0.01. Furthermore, δ_r and ϵ can be analyzed in the same way with δ_s and r fixed at 3.

When the first principal component is used as the guidance image for the EPFs, it can be seen from Fig. 8 that, if the filtering size and blur degree, i.e., δ_s, δ_r, r, and ϵ, are too large, the average classification accuracy may decrease dramatically. The reason is that a large filtering size and blur degree may over-smooth the probability maps, and thus, those small-scale objects may be misclassified. For example, although the OA obtained with $\delta_s = 4$ is similar to the accuracy obtained with $\delta_s = 5$ [see Fig. 8(a)], the AA decreases dramatically when $\delta_s = 5$ because one small-scale class which contains 20 pixels is totally misclassified when the filtering size is too large. Similarly, a very small filtering size or blur degree is also not good for the proposed method because it means that only very
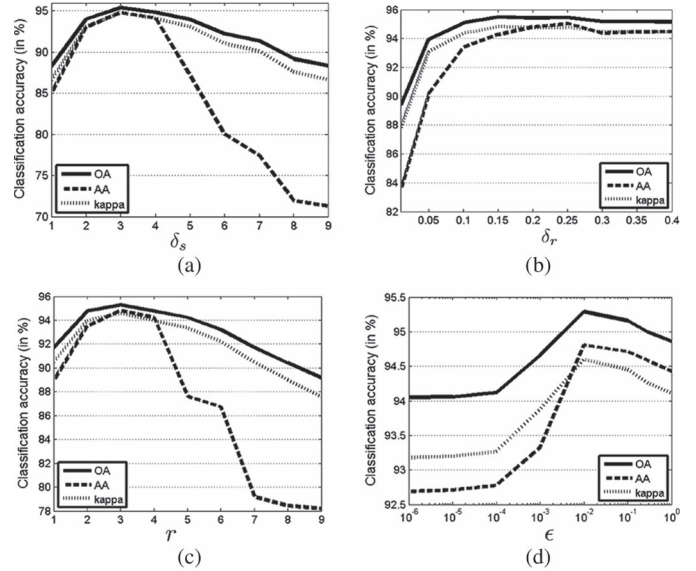

Fig. 8. Indian Pines image: analysis of the influence of the parameters δ_s, δ_r, r, and ϵ when the first principal component is selected as the guidance image.
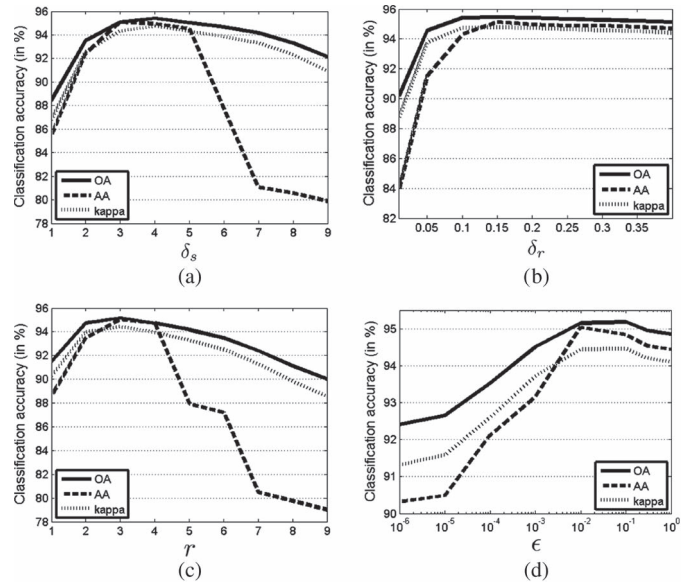

Fig. 9. Indian Pines image: analysis of the influence of the parameters δ_s, δ_r, r, and ϵ when the color composite of the first three principal components is selected as the guidance image.

limited local spatial information is considered in the filtering process.

Furthermore, Fig. 9 shows the influence of the parameters when the color composite of the first three principal components serves as the guidance image of the EPFs. From this figure, a similar conclusion can be obtained that the filtering size and blur degree cannot be too small or large. In this paper, the default parameter setting of the proposed method is given as follows: When the guidance image of the EPF is a gray-scale image, $\delta_s = 3$, $\delta_r = 0.2$, $r = 3$, and $\epsilon = 0.01$ are set to be the default parameters; when the guidance image of the EPF is a color image, $\delta_s = 4$, $\delta_r = 0.2$, $r = 4$, and $\epsilon = 0.01$ are set to be the default parameters. In the following experiments, it is

TABLE I
CLASSIFICATION ACCURACY (IN PERCENT) OF THE PROPOSED
METHOD WITH DIFFERENT POSTPROCESSING TECHNIQUES. THE
STATISTICS-BASED METHOD IS APPLIED WITHIN A 7×7 WINDOW.
THE PARAMETERS OF THE WLS FILTER ARE SET TO BE $\alpha = 1.4$
AND $\lambda = 0.3$. THE PARAMETERS OF THE NC FILTER ARE SET
TO BE $\delta_s = 3$ AND $\delta_r = 0.2$

Methods	Statistics	WLS	NC	BF	GF
OA	89.33	94.93	95.20	**95.42**	95.29
AA	83.33	94.48	94.32	**94.81**	**94.81**
Kappa	87.81	94.17	94.49	**94.75**	94.60

shown that, with the provided parameter setting, we are able to obtain good classification accuracies for different images.

2) Analysis of the Influence of Different Postprocessing Techniques: For the proposed spectral–spatial classification framework, different postprocessing techniques, e.g., other types of edge-preserving smoothing methods or statistics-based method, can be used to refine the binary probability maps. For example, the statistics-based method [50] assigns the label of a pixel to the most frequent class within its neighborhood pixels. The WLS method [36] aims at solving a WLS optimization problem which ensures that the input pixels are as smooth as possible, except across edges in the guidance image. The normalized convolution (NC) filter [45] is a recently proposed joint EPF which has a similar property as the joint bilateral filter. Table I compares the performance of the proposed method with different postprocessing techniques. It can be seen that other types of joint edge-preserving smoothing methods such as the WLS method and the NC filter are also able to obtain satisfactory overall classification accuracy, i.e., 94.93% and 95.20%, respectively. The reason is that the different joint edge-preserving smoothing methods are all able to obtain spatially smooth and edge-aligned probability maps. By contrast, since the edge information of the guidance image is not considered in the statistics-based method, that method can only ensure the spatial smoothness of the resulting classification map, thus leading to an unsatisfactory OA of 89.33%.

3) Comparison of Different Classification Methods: In this section, the proposed edge-preserving filtering based methods (EPF-B-g, EPF-B-c, EPF-G-g, and EPF-G-c) are compared with several widely used classification methods including the method based on SVM [12], EMPs [25], automatic extended attribute profiles (AEAPs) [49], logistic regression, and multilevel logistic (L-MLL) [17]. The SVM algorithm is implemented in the LIBSVM library [51]. Furthermore, the Gaussian kernel with fivefold cross validation is adopted for the classifier. For the EMP method, the morphological profiles are constructed with the first three principal components, a circular structural element, a step size increment of two, and four openings and closings for each principal component. In order to construct the AEAPs, we use the Profattran software which is kindly provided by the author of [49]. For the L-MLL method, the code is available on Dr. Li's homepage.[1] For the proposed EPF-B-g, EPF-B-c, EPF-G-g, and EPF-G-c methods, the abbreviations B and G represent that the joint bilateral filter or the guided filter is adopted for edge-preserving filtering. The abbreviations g or c denote that either the first

(a) OA=79.81 (b) OA=89.7
(c) OA=87.67 (d) OA=91.31
(e) OA=95.42 (f) OA=95.41
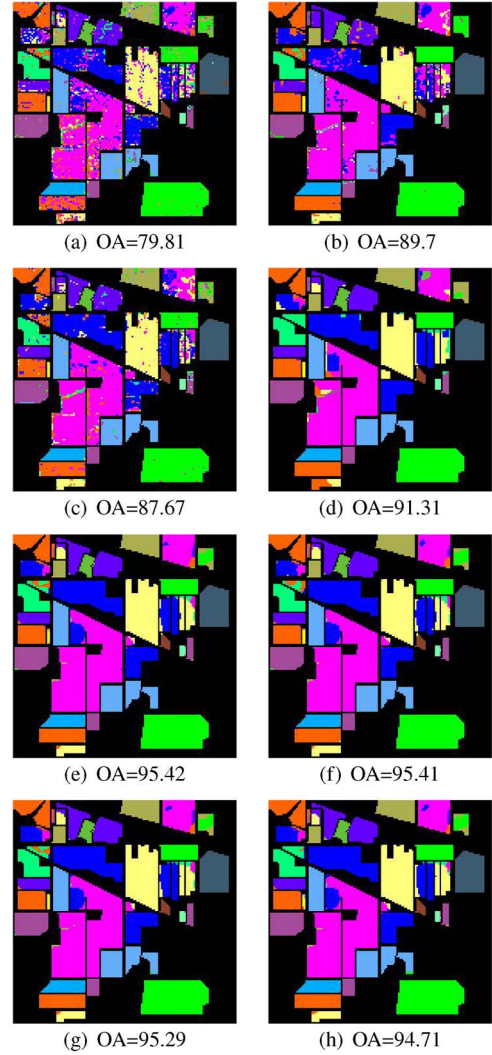(g) OA=95.29 (h) OA=94.71

Fig. 10. Classification results (Indian Pines image) obtained by the (a) SVM method, (b) EMP method, (c) AEAP method, (d) L-MLL method, (e) EPF-B-g method, (f) EPF-B-c method, (g) EPF-G-g method, and (h) EPF-G-c method. The value of OA is given in percent.

principal component or the color composite of the first three components, respectively, is used as the guidance image. The default parameters given in the first section are adopted for all images, although an adjustment of the parameters can improve the classification accuracy further for different images. In order to make the proposed method reproducible, the code and data will be made available on Mr. Kang's homepage.[2]

The first experiment is performed on the Indian Pines data set. Fig. 10 shows the classification maps obtained by different methods associated with the corresponding OA scores. From this figure, it can be seen that the classification accuracy obtained by the EMP and AEAP methods is not very satisfactory since some noisy estimations are still visible. By contrast, the L-MLL method and the proposed method perform much better in removing "noisy pixels." Specifically, the proposed method increases the OA compared to the SVM method by about 15%. Compared with the recently proposed Bayesian-based classification method (L-MLL), the proposed EPF method also

---

[1]http://www.lx.it.pt/~jun/

[2]http://xudongkang.weebly.com

TABLE II

NUMBER OF TRAINING (TRAIN) AND TEST (TEST) SAMPLES OF THE INDIAN PINES IMAGE AND CLASSIFICATION ACCURACIES (IN PERCENT)
FOR THE SVM [12], EMP [25], AEAP[49], L-MLL [17], EPF-B-g, EPF-B-c, EPF-G-g, AND EPF-G-c METHODS

Class	Train	Test	SVM	EMP	AEAP	L-MLL	EPF-B-g	EPF-B-c	EPF-G-g	EPF-G-c
Alfalfa	25	21	48.6	95.0	**100**	90.5	**100**	**100**	**100**	**100**
Corn-N	83	1345	74.7	88.9	82.4	89.4	**95.3**	94.6	95.1	94.4
Corn-M	78	752	65.6	82.7	78.5	90.6	94.0	94.6	**94.8**	94.4
Corn	68	169	45.8	80.4	64.3	**99.4**	73.2	74.4	72.2	69.8
Grass-M	79	404	83.9	84.1	85.6	98.5	98.7	97.8	**99.0**	**99.0**
Grass-T	78	652	96.7	96.4	95.9	**100**	99.5	99.5	99.7	99.2
Grass-P-M	14	14	66.7	36.8	60.9	**100**	**100**	**100**	**100**	**100**
Hay-W	66	412	99.0	99.5	**100**	**100**	**100**	**100**	**100**	99.8
Oats	10	10	50.0	76.9	55.6	**100**	**100**	**100**	**100**	**100**
Soybean-N	81	891	73.2	83.6	80.1	88.3	**88.4**	88.1	**88.4**	86.7
Soybean-M	99	2356	83.7	88.5	93.4	84.5	**98.4**	98.3	98.2	98.0
Soybean-C	73	520	75.0	85.6	85.4	**98.5**	94.3	94.7	92.9	91.0
Wheat	70	135	95.7	**100**	98.5	**100**	**100**	**100**	**100**	**100**
Woods	90	1175	95.2	99.2	98.4	91.6	**99.6**	99.5	99.4	99.0
Buildings-G-T-D	65	321	64.0	97.5	75.8	**99.4**	85.1	87.5	85.1	86.1
Stone-S-T	46	47	90.2	97.9	90.4	**100**	90.4	90.4	92.2	97.9
OA	-	-	79.81	89.70	87.67	91.31	**95.42**	95.41	95.29	94.71
AA	-	-	75.50	87.06	84.09	**95.66**	94.81	94.96	94.81	94.70
Kappa	-	-	76.99	88.16	85.92	90.07	**94.75**	94.74	94.60	93.93

(a) OA=91.57 (b) OA=97.07 (c) OA=93.42 (d) OA=97.54

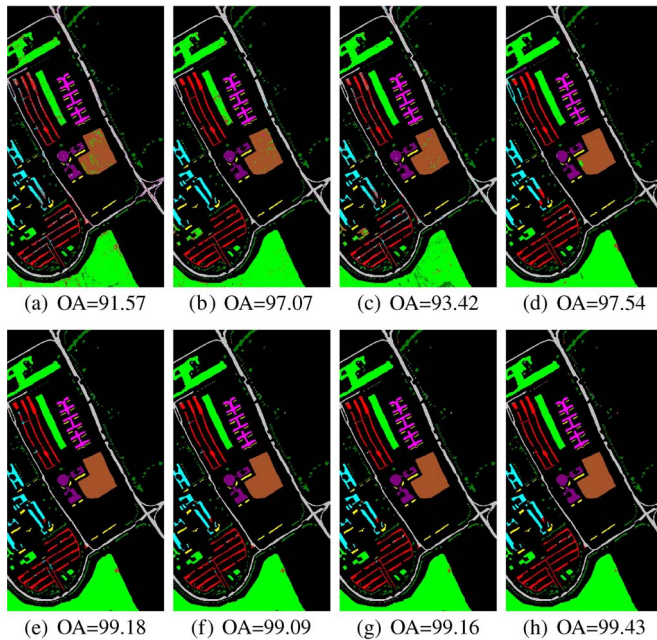(e) OA=99.18 (f) OA=99.09 (g) OA=99.16 (h) OA=99.43

Fig. 11. Classification results (University of Pavia image) obtained by the (a) SVM method, (b) EMP method, (c) AEAP method, (d) L-MLL method, (e) EPF-B-g method, (f) EPF-B-c method, (g) EPF-G-g method, and (h) EPF-G-c method. The value of OA is given in percent.

(a) OA=89.21 (b) OA=95.65 (c) OA=92.75 (d) OA=94.65

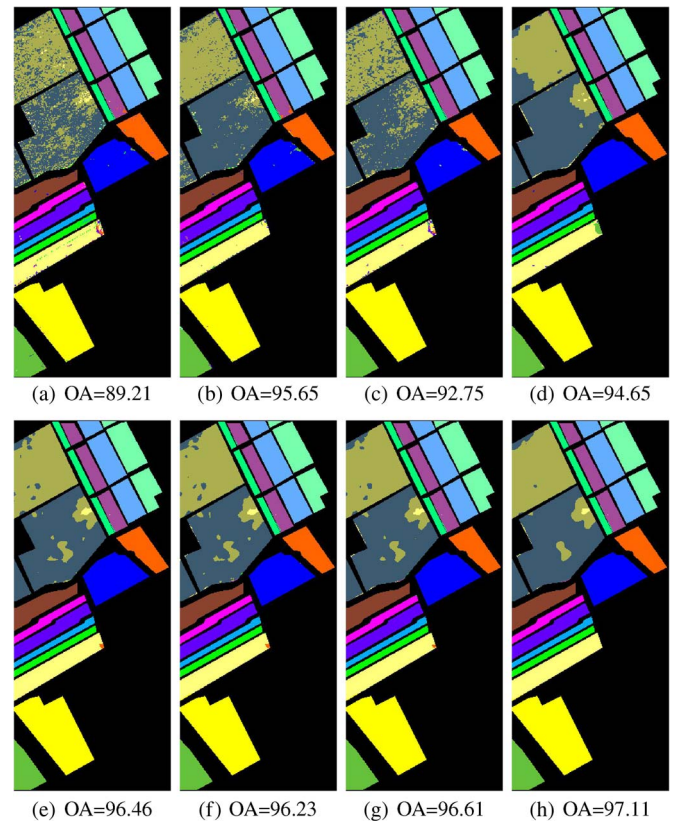(e) OA=96.46 (f) OA=96.23 (g) OA=96.61 (h) OA=97.11

Fig. 12. Classification results (Salinas image) obtained by the (a) SVM method, (b) EMP method, (c) AEAP method, (d) L-MLL method, (e) EPF-B-g method, (f) EPF-B-c method, (g) EPF-G-g method, and (h) EPF-G-c method. The value of OA is given in percent.

gives a higher classification accuracy. Table II presents the number of training and test samples (the training set which accounts for 10% of the ground truth was chosen randomly) and the classification accuracies for different methods. From this table, it can be observed that, by using the proposed EPF method, the AA of SVM is increased from 75% to 95% and the kappa accuracy can be also increased significantly. The proposed EPF-B-g method gives the best performance in terms of OA and kappa but not for the AA.

The second and third experiments were performed on the University of Pavia and Salinas images, respectively. Figs. 11 and 12 show the classification maps obtained by different methods associated with the corresponding OAs. Tables III

and IV present the number of training and test samples (for the University of Pavia image and the Salinas image, the training sets which, respectively, account for 6% and 2% of the ground truth were chosen randomly) and the classification accuracies for different methods. From the two examples, it can be seen that the proposed method always outperforms the EMP, AEAP, and L-MLL methods in terms of OA, AA, and kappa.

TABLE III
NUMBER OF TRAINING (TRAIN) AND TEST (TEST) SAMPLES OF THE UNIVERSITY OF PAVIA IMAGE AND CLASSIFICATION ACCURACIES (IN PERCENT) FOR THE SVM [12], EMP [25], AEAP[49], L-MLL [17], EPF-B-g, EPF-B-c, EPF-G-g, AND EPF-G-c METHODS

Class	Train	Test	SVM	EMP	AEAP	L-MLL	EPF-B-g	EPF-B-c	EPF-G-g	EPF-G-c
Asphalt	286	6345	96.9	**99.6**	98.1	94.8	99.2	98.9	98.7	99.1
Meadows	286	18363	97.7	99.0	98.9	99.0	99.8	99.7	99.7	**99.8**
Gravel	285	1814	75.0	96.5	73.0	90.1	94.9	97.5	98.6	**99.4**
Trees	285	2779	87.0	98.8	75.7	98.8	99.2	98.9	99.1	**99.5**
Metal sheets	285	1060	96.3	96.9	96.8	**100**	99.4	99.2	99.0	99.2
Bare soil	285	4744	81.4	86.0	90.1	98.5	96.4	98.2	98.6	**99.0**
Bitumen	285	1045	67.5	95.8	96.0	98.9	**100**	**100**	**100**	**100**
Bricks	285	3397	88.8	**98.5**	90.5	94.7	95.2	97.8	97.8	98.2
Shadows	285	662	**100**	**100**	**100**	99.2	99.8	99.7	**100**	**100**
OA	-	-	91.57	97.07	93.42	97.54	99.18	99.09	99.16	**99.43**
AA	-	-	87.85	96.79	91.02	96.67	99.05	98.88	99.05	**99.36**
Kappa	-	-	88.72	96.05	91.21	97.12	98.89	98.77	98.87	**99.22**

TABLE IV
NUMBER OF TRAINING (TRAIN) AND TEST (TEST) SAMPLES OF THE SALINAS IMAGE AND CLASSIFICATION ACCURACIES (IN PERCENT) FOR THE SVM [12], EMP [25], AEAP [49], L-MLL [17], EPF-B-g, EPF-B-c, EPF-G-g, AND EPF-G-c METHODS

Class	Train	Test	SVM	EMP	AEAP	L-MLL	EPF-B-g	EPF-B-c	EPF-G-g	EPF-G-c
Weeds_1	67	1942	99.9	99.9	**100**	99.9	**100**	**100**	**100**	**100**
Weeds_2	67	3659	99.6	99.8	**100**	99.6	**100**	**100**	**100**	**100**
Fallow	67	1909	95.7	96.7	98.3	95.7	98.7	98.4	98.9	**100**
Fallow_P	69	1325	97.5	97.7	**97.8**	97.5	97.4	97.7	97.4	97.6
Fallow_S	67	2611	98.3	98.1	98.7	98.3	99.2	**99.4**	**99.4**	99.3
Stubble	67	3892	100	99.9	99.9	**100**	**100**	**100**	**100**	**100**
Celery	68	3511	99.0	99.7	99.9	99.0	**100**	**100**	**100**	**100**
Grapes	69	11202	81.9	92.1	85.6	81.9	95.8	95.4	96.1	**97.4**
Soil	68	6135	99.5	99.6	99.3	99.5	99.6	99.6	99.7	**99.8**
Corn	68	3210	87.3	94.1	96.4	87.3	96.0	95.7	95.5	**96.7**
Lettuce_4wk	68	1000	96.6	**100**	99.4	96.6	**100**	**100**	**100**	**100**
Lettuce_5wk	67	1860	96.9	96.4	94.5	96.9	**100**	**100**	**100**	**100**
Lettuce_6wk	67	849	96.9	96.8	99.9	96.9	**100**	**100**	**100**	**100**
Lettuce_7wk	67	1003	89.2	96.4	96.9	89.2	99.1	99.1	99.3	**99.8**
Vinyard_U	70	7198	64.2	**88.8**	75.3	64.2	84.9	84.1	85.5	86.4
Vinyard_T	67	1740	95.0	93.7	97.5	95.0	99.9	99.8	**100**	**100**
OA	-	-	89.21	95.65	92.75	94.65	96.46	96.23	96.61	**97.11**
AA	-	-	93.59	96.86	96.21	97.44	98,17	98.08	98.24	**98.56**
Kappa	-	-	88.01	95.15	91.92	94.04	96.05	95.80	96.23	**96.78**

Compared with the SVM method, the proposed method can improve the classification accuracies significantly. For example, in Table III, the classification accuracy of the Bitumen class increases from 67.5% to 100%. Similar improvements can be found in the experimental results of the Salinas example. The two sets of results further demonstrate the accuracy of the proposed method.

4) Classification Results With Different Training and Test Sets: In this section, the influence of different training and test sets to the performance of the proposed method is analyzed. Experiments are performed on three images, i.e., the Indian Pines image, the University of Pavia image, and the Salinas image. Only the classification result obtained by the EPF-B-g method is presented because different edge-preserving filtering methods tend to obtain similar classification results (see Tables II–IV). Fig. 13 shows the classification results of the EPF-B-g method with the number of training samples (in percent) increased from 1% to 10% for the Indian Pines image, 1% to 6% for the University of Pavia image, and 1% to 2% for the Salinas image. From this figure, it can be seen that the proposed method can always improve the classification accuracy significantly with a different number of training samples. For example, regarding the Indian Pines image, when the OA of SVM is about 72% (4% ground truth samples are used as training samples), the EPF-B-g method can obtain a classifi-
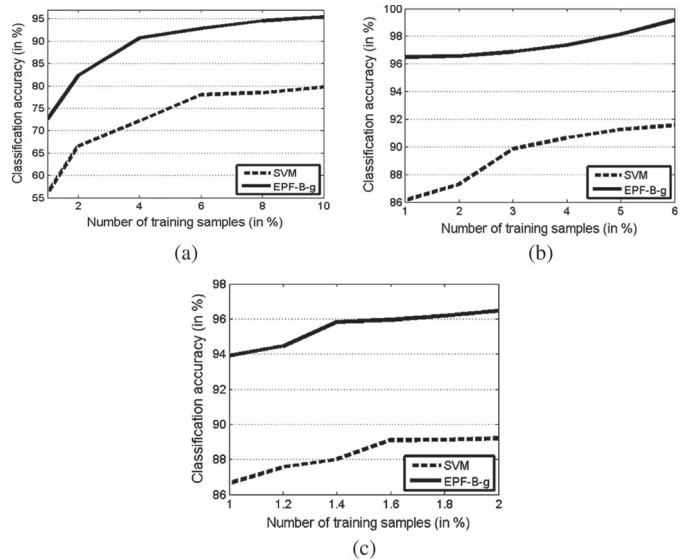

Fig. 13. OA of the proposed EPF-B-g method with different numbers of training samples on different images. (a) Indian Pines. (b) University of Pavia. (c) Salinas.

cation accuracy near 92%. For the University of Pavia image, it can be seen that, with relatively limited training samples (1% of the ground truth), the proposed method can obtain an OA near

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

10

IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING

TABLE V
COMPUTING TIME (IN SECONDS) OF THE PROPOSED ALGORITHMS.
THE NUMBERS OUTSIDE AND INSIDE THE PARENTHESES
SHOW THE COMPUTING TIMES OF THE MATLAB AND
C++ IMPLEMENTATIONS, RESPECTIVELY

Input data sets	EPF-G-g	EPF-G-c	EPF-B-g	EPF-B-c
Indian Pines	**0.27(0.17)**	6.414(0.26)	0.45(0.21)	8.07(0.32)
University	**2.17(0.46)**	38.27(0.72)	2.18(0.69)	44.18(1.32)
Salinas	**1.60(0.57)**	35.03(0.87)	2.15(0.90)	42.89(1.57)

96.5%. A similar conclusion can be obtained when analyzing the experimental results of the Salinas image.

5) Computational Complexity: Here, experiments are performed using MATLAB and C++ on a Laptop with 2.5-GHz CPU and 4-GB memory. Table V shows the computing time of the proposed methods, i.e., EPF-G-g, EPF-G-c, EPF-B-g, and EPF-B-c. From this table, it can be seen that the C++ implementations of the proposed methods are all very fast (the EPF-G-g method takes only 0.17 s for the Indian Pines image). The reason is that the proposed method only requires edge-preserving filtering to be done several times and PCA decomposition to be done once. The complexity of the two EPFs is $O(N)$, and thus, the probability optimization step of the proposed method has a computing complexity of $O(KN)$, where N is the number of pixels and K is the number of classes. Considering that the PCA decomposition step also has fast implementations, the proposed spectral–spatial method is indeed computationally efficient. For the spectral part of the proposed method, the MATLAB implementation of the SVM classifier is the most time-consuming part. Specifically, it takes about 8.68 s for classifying the Indian Pines image. Therefore, the algorithm can be accelerated further by using more efficient pixelwise classifiers or more efficient implementations of the SVM classifier.

V. CONCLUSION

A simple yet powerful filtering approach has been proposed for spectral–spatial hyperspectral image classification. The proposed method aims at optimizing the pixelwise classification maps in a local filtering framework. The filtering operation achieves a local optimization of the probabilities, which is in contrast to some segmentation methods that aim at optimizing the classification map globally. This paper has shown that "local smoothing" is also able to achieve a high classification accuracy, i.e., 99.43% for the University of Pavia image. One advantage of the proposed method is the dominance of the pixelwise spectral information with respect to the spatial contextual information in spectral–spatial classification. An observation on the proposed method is that the optimization of probability maps will not cause a big difference in the overall appearance of the initial pixelwise probability maps, which means that the pixelwise information is also well considered in the filtering process. Furthermore, a major advantage of the proposed method is that it is computationally efficient, and thus, it will be quite useful in real applications. In the future, a further improvement may be achieved by adaptively adjusting the filtering size and the blur degree of the EPF.

REFERENCES

[1] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.

[2] L. Zhang, L. Zhang, D. Tao, and X. Huang, "Tensor discriminative locality alignment for hyperspectral image spectral–spatial feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 242–256, Jan. 2013.

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

[4] 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.

[5] 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.

[6] A. Plaza, J. A. Benediktsson, J. W. Boardman, J. Brazile, L. Bruzzone, G. Camps-Valls, J. Chanussot, M. Fauvel, P. Gamba, A. Gualtieri, M. Marconcini, J. C. Tilton, and G. Trianni, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. Suppl. 1, pp. S110–S122, Sep. 2009.

[7] J. Ham, Y. Chen, M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

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

[9] F. Ratle, G. Camps-Valls, and J. Weston, "Semisupervised neural networks for efficient hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2271–2282, May 2010.

[10] Y. Zhong and L. Zhang, "An adaptive artificial immune network for supervised classification of multi-/hyperspectral remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 894–909, Mar. 2012.

[11] S. Kawaguchi and R. Nishii, "Hyperspectral image classification by bootstrap AdaBoost with random decision stumps," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3845–3851, Nov. 2007.

[12] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.

[13] Y. Chen, N. M. Nasrabadi, and T. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.

[14] A. Castrodad, Z. Xing, J. B. Greer, E. Bosch, L. Carin, and G. Sapiro, "Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4263–4281, Nov. 2011.

[15] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 2, pp. 844–856, Feb. 2013.

[16] W. Di and M. M. Crawford, "View generation for multiview maximum disagreement based active learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 5, pp. 1942–1954, May 2012.

[17] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral image segmentation using a new Bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947–3960, Oct. 2011.

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

[19] J. Benediktsson, M. Pesaresi, and K. Amason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.

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

[21] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "A spatial–spectral kernel-based approach for the classification of remote-sensing images," *Pattern Recognit.*, vol. 45, no. 1, pp. 381–392, Jan. 2012.

[22] G. Camps-Valls, N. Shervashidze, and K. M. Borgwardt, "Spatio-spectral remote sensing image classification with graph kernels," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 741–745, Oct. 2010.

[23] 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.

[24] D. M. Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542–546, May 2011.

[25] 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.

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

[27] 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.

[28] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral–spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.

[29] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367–2379, Jul. 2010.

[30] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Multiple spectral–spatial classification approach for hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4122–4132, Nov. 2010.

[31] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and classification of hyperspectral images using minimum spanning forest grown from automatically selected markers," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 5, pp. 1267–1279, Oct. 2010.

[32] 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.

[33] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. Int. Conf. Comput. Vis.*, Jan. 1998, pp. 839–846.

[34] K. He, J. Sun, and X. Tang, "Guided image filtering," in *Proc. Eur. Conf. Comput. Vision*, Heraklion, Greece, Sep. 2010, pp. 1–14.

[35] S. Paris and F. Durand, "A fast approximation of the bilateral filter using a signal processing approach," *Int. J. Comput. Vis.*, vol. 81, no. 1, pp. 24–52, Jan. 2009.

[36] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 67:1–67:10, Aug. 2008.

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

[38] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 35, no. 2, pp. 504–511, Feb. 2013.

[39] S. Li, X. Kang, and J. Hu, "Image fusion with guided filtering," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2864–2875, Jul. 2013.

[40] S. Li and X. Kang, "Fast multi-exposure image fusion with median filter and recursive filter," *IEEE Trans. Consum. Electron.*, vol. 58, no. 2, pp. 626–632, May 2012.

[41] B. Zhang and J. P. Allebach, "Adaptive bilateral filter for sharpness enhancement and noise removal," *IEEE Trans. Image Process.*, vol. 17, no. 5, pp. 664–678, May 2008.

[42] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.

[43] C. H. Lin, J. S. Tsai, and C. T. Chiu, "Switching bilateral filter with a texture/noise detector for universal noise removal," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2307–2320, Sep. 2010.

[44] K. Kotwal and S. Chaudhuri, "Visualization of hyperspectral images using bilateral filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2308–2316, May 2010.

[45] E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 69:1–69:12, Jul. 2011.

[46] T. Qiu, A. Wang, N. Yu, and A. Song, "LLSURE: Local linear sure-based edge-preserving image filtering," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 80–90, Jan. 2013.

[47] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L_0 gradient minimization," *ACM Trans. Graph.*, vol. 30, no. 6, pp. 174:1–174:12, Dec. 2011.

[48] 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.

[49] P. R. Marpu, M. Pedergnana, M. D. Mura, J. A. Benediktsson, and L. Bruzzone, "Automatic generation of standard deviation attribute profiles for spectral–spatial classification of remote sensing data," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 2, pp. 293–297, Mar. 2013.

[50] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 2nd ed. Boston, MA, USA: Addison-Wesley, 2001.

[51] C. C. Chang and C. J. Lin (2011, Apr.). LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* [Online]. *2(3)*, pp. 27:1–27:27. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm

Xudong Kang (S'13) received the B.Sc. degree from Northeast University, Shenyang, China, in 2007. He is currently working toward the Ph.D. degree in electrical engineering in Hunan University, Changsha, China.

He is currently a Visiting Ph.D. student in electrical engineering at the University of Iceland, Reykjavik, Iceland. He is engaged in image fusion, image superresolution, pansharpening, and hyperspectral image classification.

Shutao Li (M'07) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

In 2001, he joined the College of Electrical and Information Engineering, Hunan University. From May 2001 to October 2001, he was a Research Associate with the Department of Computer Science, Hong Kong University of Science and Technology, Kowloon, Hong Kong. From November 2002 to November 2003, he was a Postdoctoral Fellow with the Royal Holloway College, University of London, Egham, U.K., working with Prof. J.-S. Taylor. From April 2005 to June 2005, he was a Visiting Professor with the Department of Computer Science, Hong Kong University of Science and Technology. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He has authored or coauthored more than 160 refereed papers. His professional interests are information fusion, pattern recognition, and image processing.

Jón Atli Benediktsson (S'84–M'90–SM'99–F'04) received the Cand.Sci. degree in electrical engineering from the University of Iceland, Reykjavik, Iceland, in 1984 and the M.S.E.E. and Ph.D. degrees in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

He is currently the Pro Rector of Academic Affairs and a Professor of electrical and computer engineering with the University of Iceland. He is a cofounder of the biomedical startup company Oxymap. His research interests are in remote sensing, image analysis, pattern recognition, biomedical analysis of signals, and signal processing, and he has published extensively in these fields.

Prof. Benediktsson was the 2011–2012 President of the IEEE Geoscience and Remote Sensing Society (GRSS), and he has been on the GRSS Administrative Committee since 2000. He is a Fellow of SPIE. He is a member of Societas Scientiarum Islandica and Tau Beta Pi. He was the Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2003 to 2008, and he has served as an Associate Editor of TGRS since 1999 and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003. He received the Stevan J. Kristof Award from Purdue University in 1991 as an outstanding graduate student in remote sensing. In 1997, he was the recipient of the Icelandic Research Councils Outstanding Young Researcher Award; in 2000, he was granted the IEEE Third Millennium Medal; in 2004, he was a corecipient of the University of Icelands Technology Innovation Award; in 2006, he received the yearly research award from the Engineering Research Institute, University of Iceland; and in 2007, he received the Outstanding Service Award from the GRSS. He was the corecipient of the 2012 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Best Paper Award.