

Probabilistic Fusion of Pixel-Level and Superpixel-Level Hyperspectral Image Classification

Shutao Li, *Senior Member, IEEE*, Ting Lu, *Student Member, IEEE*, Leyuan Fang, *Member, IEEE*, Xiuping Jia, *Senior Member, IEEE*, and Jón Atli Benediktsson, *Fellow, IEEE*

Abstract—A novel hyperspectral image (HSI) classification method by the probabilistic fusion of pixel-level and superpixel-level classifiers is proposed. Generally, pixel-level classifiers based on spectral information only may generate “salt and pepper” result in the classification map since spatial correlation is not considered. By incorporating spatial information in homogeneous regions, the superpixel-level classifiers can effectively eliminate the noisy appearance. However, the classification accuracy will be deteriorated if undersegmentation cannot be fully avoided in superpixel-based approaches. Therefore, it is proposed to adaptively combine both the pixel-level and superpixel-level classifiers, to improve the classification performance in both homogenous and structural areas. In the proposed method, a support vector machine classifier is first applied to estimate the pixel-level class probabilities. Then, superpixel-level class probabilities are estimated based on a joint sparse representation. Finally, the two levels of class probabilities are adaptively combined in a maximum *a posteriori* estimation model, and the classification map is obtained by solving the maximum optimization problem. Experimental results on real HSI images demonstrate the superiority of the proposed method over several well-known classification approaches in terms of classification accuracy.

Index Terms—Hyperspectral image (HSI), probabilistic fusion, spectral-spatial classification, superpixel segmentation.

I. INTRODUCTION

HYPERSPECTRAL image (HSI) can provide rich spatial and spectral information, which has attracted much attention in different civil and military applications, e.g., land use analysis, precision agriculture, and pollution monitoring.

Manuscript received April 4, 2016; revised August 6, 2016; accepted August 18, 2016. This work was supported in part by the National Natural Science Fund of China for Distinguished Young Scholars under Grant 61325007, by the National Natural Science Fund of China for International Cooperation and Exchanges under Grant 61520106001, and by the Science and Technology Plan Project Fund of Hunan Province under Grant 2015WK3001.

S. Li, T. Lu, and L. Fang are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: shutao_li@hnu.edu.cn; tingluhnu@gmail.com; fangleyuan@gmail.com).

X. Jia is with the School of Engineering and Information Technology, University of New South Wales, Canberra, BC 2610, Australia (e-mail: x.jia@adfa.edu.au).

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

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.2016.2603190

Over the past decades, a large number of techniques have been developed for HSI classification. Among these methods, the support vector machine (SVM) [1] and multinomial logistic regression (MLR) [2], as two classical classifiers, have been proven effective in classifying HSI. In addition, extensive works have been proposed to develop more advanced pixel-level classifiers, e.g., random forests [3], [4], neural networks [5], [6], and active learning [7], [8]. Generally, pixel-level classification methods mainly exploit discriminant spectral information in HSI to distinguish varying objects in the scene. However, due to the existence of noise and mixed spectral pixels in HSI, the pixel-level methods are likely to generate noisy appearance in classification maps.

In addition, many researchers focus on designing spectral-spatial classifiers to improve classification performance, by jointly integrating the spectral and spatial information in HSI. In some works, effective feature extraction algorithms [9]–[11] and multifeature fusion [12]–[14] techniques have been developed to better characterize the spectral-spatial feature of different materials in the image scene. For example, a semantic allocation level multifeature fusion scheme is introduced in [14] to jointly integrate the spectral, textural, and structural features. Moreover, some other works [15]–[17] attempt to enhance the discriminative capabilities of classifiers by extending the classical classifiers. In [15], a generalized composite kernel is introduced to extend the MLR classifier, which can improve the classification accuracy by flexibly combining spectral and spatial information. In addition, as one powerful image processing tool, the sparse representation (SR) model [18], [19] has attracted much attention for HSI classification [20]–[23]. In the SR-based classification methods, the similar pixels in a neighborhood are jointly represented with a few training samples that are selected from a constructed dictionary. By involving the local spatial information and exploiting the sparsity in HSI, the SR-based methods can provide promising classification performance. The main drawback of these spectral-spatial classification methods is the use of a fixed window to extract local spatial information from HSI, which cannot adaptively capture the structural characters of varying sizes and shapes. Furthermore, active learning [24]–[26] and probabilistic modeling [27]–[30] also have been successfully extended to develop effective spectral-spatial classifiers in the HSI domain.

Apart from the HSI classification methods previously described, another interesting trend is to investigate image-segmentation-based approaches, aiming at better exploiting the

spatially structural information to improve classification performance. Typically, hyperspectral segmentation techniques (e.g., watershed-based segmentation [31], [32] and stochastic minimum spanning forest [33]) are first used to divide spatial regions into nonoverlapping homogeneous patches. Then, based on the initial classification results by pixel-level classifiers, decision fusion rules are used to determine the final classification results for pixels within each segmented region. For example, by using the majority voting, all pixels in the same region are assigned to the most frequent class within this region. More recently, some works [34]–[36] focus on developing segmentation-based HSI classifiers with superpixel techniques [37]–[39], to jointly combine both the spectral–spatial similarity and discrimination to improve classification results. Specifically, each superpixel in HSI corresponds to a set of spatially connected and spectrally similar pixels, and the common class label for each superpixel is determined according to a distance metric measured between each superpixel and each class of training samples. For example, the works in [34] and [35] construct the distance metric based on extended SR models, i.e., the discriminative sparse model [34] and the multitask joint sparse model [35]. However, an accurate oversegmentation map is difficult to obtain. Once spectral pixels within a superpixel should belong to different classes, a wrong estimation cannot be avoided because of the “one label for one superpixel” manner.

Considering the merits and drawbacks brought by both pixel-level and superpixel-level classifiers, a novel probabilistic fusion framework for HSI classification is proposed, to adaptively combine both levels of classification results to improve the overall accuracy (OA). In the proposed method, the probabilistic SVM algorithm is first used to classify the HSI. This way, the discriminative spectral information can be used to estimate the pixel-level class probability for each individual pixel. Then, by sparsely coding the pixels within each superpixel with a joint SR (JSR) model, the spatially similar information can be utilized for the estimation of the superpixel-level class probabilities. Based on the spatial distribution of pixels in each superpixel, both levels of class probabilities are adaptively combined, resulting in the generation of joint class probabilities. Finally, the joint class probabilities are incorporated into a maximum *a posteriori* (MAP) estimation model, and the classification map is obtained via solving the MAP-based optimization problem. By jointly integrating pixel-based spectral information and superpixel-based shape-adaptive spatial information in the proposed method, the noisy classification results generated by an individual pixel-wise classifier can be eliminated. Moreover, the misclassification results caused by the inaccurate segmentation map in a superpixel-level classifier can also be suppressed. As a result, better balance in classifying both homogenous and structural regions can be achieved.

The rest of this paper is structured as follows. Section II briefly reviews related works, including the SVM classifier, the JSR classifier, and the MAP estimation model. Section III introduces the proposed probabilistic fusion method for HSI classification. Then, experimental results and analysis are reported in Section IV. Finally, conclusions and future works are discussed in Section V.

II. OVERVIEW OF RELATED MODELS

A. Support Vector Machine

Let $\mathbf{X} = \{\mathbf{x}_i\}$, $\mathbf{X} \in R^{B \times N_t}$, represent all the test samples, where B is the feature dimension of each test sample, and N_t is the total number of test samples. The class labels for all test samples are represented by $Y = \{y_i\}$, $Y \in R^{N_t}$, where y_i is the label for the i th test sample \mathbf{x}_i . In the context of supervised classification, $\mathbf{Z} = \{\mathbf{z}_n\}$ ($n = 1, 2, \dots, N_{tr}$) represent all the training samples. For each training sample \mathbf{z}_n , y_n represents the corresponding label, where N_{tr} is the total number of training samples.

Generally, the SVM tries to determine the optimal separating hyperplane that maximizes the margin between the closest training samples, and then, the separating hyperplane uses the boundary pixels to create a decision surface. The SVM classifier can be defined as [1]

$$f(\mathbf{x}_i) = \text{sign} \left(\sum_{n=1}^{N_{tr}} a_n y_n \Phi(\mathbf{x}_i, \mathbf{z}_n) + b \right) \quad (1)$$

where a_n is the Lagrange multiplier, b is the estimation bias, and $\Phi(\mathbf{x}_i, \mathbf{z}_n)$ is a linear or nonlinear function, e.g., Gaussian kernel function. Since the classification of HSI often involves the discrimination of more than two classes, different multi-class classification strategies have been widely developed in the literature [40]–[42]. One typical strategy is the so-called “one against one” approach, which combines the binary SVM classifier with voting-based decision fusion to estimate the final labels. Let C represent the number of classes, the first step is to construct $C(C-1)/2$ binary SVM classifiers, and each one of them are trained from two of the classes. Then, the classification result by each binary classification is considered to be a voting, and the final classification label for each sample is determined according to the maximum number of votes.

B. Joint Sparse Representation

Generally, the JSR model assumes that all the neighboring pixels around the central test pixel should share a common sparsity pattern with different sets of coefficients. This is reasonable for an HSI because the neighboring pixels in a local patch always correspond to similar materials. Based on this assumption, both the central test pixel and the neighboring pixels are stacked in a matrix and sparsely represented by a row-sparse coefficient matrix. This way, the locally spatial information can be incorporated to decide on the label for the central pixel. Let C represent the number of classes in a test image. Then, each subdictionary \mathbf{D}^j ($j \in [1, C]$) that contains all the training samples belonging to the j th class can be constructed. Based on all the subdictionaries, one global dictionary \mathbf{D} is formed, i.e., $\mathbf{D} = \{\mathbf{D}^j\}$. Let \mathbf{X}_i be a set of samples similar to the test sample \mathbf{x}_i , the JSR model aims at selecting very few atoms from \mathbf{D} (denoted as \mathbf{D}_i) to adaptively represent \mathbf{X}_i . Generally, the JSR model [20] can be denoted as

$$\begin{aligned} \hat{\mathbf{A}}_i &= \arg \min(\mathbf{X}_i - \mathbf{D}\mathbf{A}_i) \\ \text{s.t. } \|\mathbf{A}_i\|_{\text{row},0} &\leq L \end{aligned} \quad (2)$$

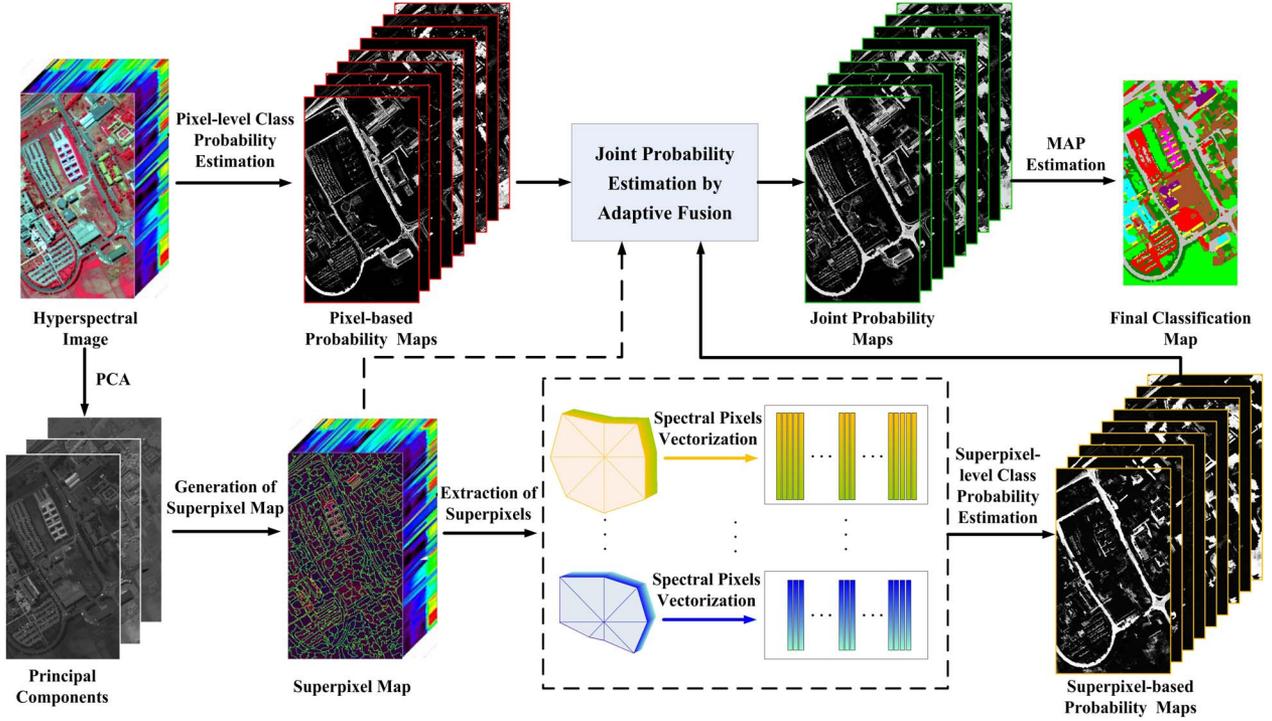


Fig. 1. Framework of the proposed PSPFC method.

where \mathbf{A}_i is the matrix of sparse coefficients corresponding to \mathbf{X}_i based on the dictionary \mathbf{D} , and L is the sparsity level. $\|\mathbf{A}_i\|_{\text{row},0} \leq L$ is the joint sparse restriction norm, which is used to restrict the maximum number of nonzero rows in \mathbf{A}_i to L . Typically, the simultaneous orthogonal matching pursuit (SOMP) algorithm [43] can solve the optimization problem in (2) in an iteration way. At each iteration, the atom that yields the best joint sparse approximation to all of the residual vectors is selected from the dictionary \mathbf{D} . The SOMP algorithm terminates when the predefined level of sparsity (controlled by the number of iterations) is satisfied. In the context of classification, the reconstruction residual errors with regard to different subdictionaries are used to indicate the class label of test sample \mathbf{x}_i . Let $\mathbf{A}_{i,j}$ represent the vectors of sparse coefficients that correspond to the atoms in \mathbf{D}^j . Then, the estimated label for \mathbf{x}_i is determined based on the minimal representation error criteria, denoted as

$$\begin{aligned} \hat{y}_i &= \arg \min_j \{ \|\mathbf{X}_i - \mathbf{D}^j \cdot \mathbf{A}_{i,j}\|_2 \} \\ \text{s.t. } j &= 1, 2, \dots, C. \end{aligned} \quad (3)$$

By (3), the label of \mathbf{x}_i is estimated to be \hat{y}_i if the condition of $\|\mathbf{X}_i - \mathbf{D}^{\hat{y}_i} \cdot \mathbf{A}_{i,\hat{y}_i}\|_2 = \min_j \{ \|\mathbf{X}_i - \mathbf{D}^j \cdot \mathbf{A}_{i,j}\|_2 \}$ is satisfied.

C. Maximum a Posteriori Estimation

The MAP model tries to maximize a posterior $P(Y|\mathbf{X})$, which can be represented as

$$\begin{aligned} \arg \max P(Y|\mathbf{X}) &= \arg \max P(\mathbf{X}|Y)P(Y) \\ &= \arg \max \prod_{i=1}^{N_t} \frac{P(y_i|\mathbf{x}_i)}{P(y_i)} P(Y) \end{aligned} \quad (4)$$

where the densities $P(y_i)$ are generally assumed to be equally distributed [44], [45]. By taking a logarithm operator, i.e., $\log(\cdot)$, (4) can be transformed into

$$\hat{Y} = \arg \max \left\{ \sum_{i=1}^{N_t} \log P(y_i|\mathbf{x}_i) + \log P(Y) \right\}. \quad (5)$$

In (5), the posterior probabilities, i.e., $P(y_i|\mathbf{x}_i)$, can be estimated by probabilistic pixel-wise classifiers, e.g., SVM [1]. The prior probability, i.e., $P(Y)$, is generally modeled with an isotropic pairwise multilevel logistic (MLL) model [45], which is proportional to $\exp\{-\beta \sum_{i \in N_i} (1 - \delta(y_i, y_{i+}))\}$. Here, β is a predefined smoothing coefficient, N_i denotes the neighborhood of central pixel \mathbf{x}_i , and $\delta(\cdot, \cdot)$ is the Kronecker delta function ($\delta(a, b) = 1$ if $a = b$, and $\delta(a, b) = 0$ otherwise). The expression in (5) is a combination of unary and pairwise interaction terms, and the minimization problem can be solved by different optimization algorithms in the literature, e.g., loopy belief propagation [46], tree-reweighted message passing [47], and the graph-cut [48]. Among them, one typical option resorts to the α -expansion-based graph-cut algorithm, which has been widely demonstrated to yield good approximations to the MAP estimation problem and have high efficiency [48]–[50]. Thus, this algorithm is adopted to solve the optimization problem in the proposed method. Additional details of the α -expansion-based graph-cut algorithm are given in the following sections.

III. PROPOSED METHOD

In Fig. 1, the framework of the proposed pixel-level and superpixel-level classifier-based probabilistic fusion classification (PSPFC) method is shown. The proposed method

consists of three main stages: pixel-level class probability estimation, superpixel-level class probability estimation, and the probabilistic-fusion-based labeling result. The detailed description and analysis for each stage is given in the following.

A. Pixel-Level Probability Estimation

As defined in (1), the standard SVM classifier does not provide probability estimates for individual class. Here, pairwise coupling of binary probabilistic estimation [51] is applied to generate these probability estimates for individual class. By the one-against-one (i.e., pairwise) approach for multiclass classification [40], the pairwise class probabilities are first estimated, i.e., $r_{j,k} \approx p(y_i = j | y_i = j \text{ or } y_i = k, \mathbf{x}_i)$. Let p_j ($j = 1, 2, \dots, C$) represent the class probability of assigning testing sample \mathbf{x}_i with the j th label, a $C \times 1$ vector of multi-class probability estimates, i.e., $\mathbf{p} = [p_1, p_2, \dots, p_C]^T$, can be estimated by solving

$$\begin{aligned} \min_{\mathbf{p}} \quad & \frac{1}{2} \sum_{j=1}^C \sum_{k:k \neq j}^C (r_{k,j} p_j - r_{j,k} p_k)^2 \\ \text{s.t.} \quad & p_j \geq 0, \sum_{j=1}^C p_j = 1. \end{aligned} \quad (6)$$

By introducing a semidefinite matrix \mathbf{Q} , $\mathbf{Q}_{j,k} = \sum_{s:s \neq j} r_{s,j}^2$ if $j = k$, and $\mathbf{Q}_{j,k} = -r_{k,j} r_{j,k}$ if $j \neq k$, it is proved in [51] that the optimal solution of (6) can be estimated via solving a simple linear function, defined as

$$\begin{bmatrix} \mathbf{Q} & \mathbf{e} \\ \mathbf{e}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{p} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix} \quad (7)$$

where \mathbf{b} is the Lagrangian multipliers of the equality constraint, \mathbf{e} is the $C \times 1$ vector of all ones, and $\mathbf{0}$ is the $C \times 1$ vector of all zeros. Here, instead of solving (7) by a direct method, e.g., Gaussian elimination, a simpler and more efficient iterative algorithm derived in [51] is adopted, which is implemented with the toolbox in the LIBSVM library¹ [42]. Finally, based on the optimal solution of (7), a set of estimated class probabilities, i.e., $\{p_j\}$ ($j = 1, 2, \dots, C$), can be obtained. Assuming one pixel \mathbf{x}_i is assigned to the j th class, the estimated class probability is

$$P_{\text{pix}}(y_i | \mathbf{x}_i) = P_{\text{pix}}(y_i = j | \mathbf{x}_i) = p_j. \quad (8)$$

B. Superpixel-Level Probability Estimation

1) *Generation of Superpixel Map*: Here, superpixels are generated from HSI based on the entropy rate superpixel (ERS) [37] algorithm, which consists of three steps.

Step 1. Generation of base image. The base image is generated via principal component analysis (PCA). Specifically, the PCA is operated on the HSI to obtain the significant principal components (PCs), i.e., the first three PCs, which remain the majority of the structural spatial information in HSI. Then, the extracted first three PCs are regarded as the base image.

Step 2. Generation of superpixel map from base image. With the ERS algorithm, a subset of edges \hat{E}_s is selected via solving a graph optimization problem with an iterative greedy algorithm [37], such that the resulting graph contains exactly K connected subgraphs. The pixels in each subgraph correspond to a homogeneous region. This way, a superpixel map is constructed, which divides the overall spatial area into K nonoverlapping homogeneous regions.

Step 3. Generation of superpixels from HSI. Based on the superpixel map, i.e., S , the pixels within each homogeneous patch are extracted from HSI and form each superpixel. This way, each superpixel contains multiple spatially connected and spectrally similar pixels.

Particularly, the ERS-algorithm-based generation of a superpixel map from a base image in Step 2 is described as follows. The base image is first mapped to a graph, i.e., $G = (V, E)$, where vertices (V) denote the pixels, and edge weights (E) denote the pairwise similarities given in the form of a similarity matrix. Then, the task of generating a superpixel map is transformed into solving a graph optimization problem, which is represented as

$$\begin{aligned} \max_{\hat{E}_s} \quad & \{H(E_s) + \lambda F(E_s)\} \\ \text{s.t.} \quad & E_s \in E \end{aligned} \quad (9)$$

where $H(\cdot)$ is the entropy rate term of the random walk on graph G , and $F(\cdot)$ is a balancing term encouraging clusters with similar sizes and reducing the number of unbalanced superpixels. The problem in (9) can be solved with the iterative greedy algorithm introduced in [37]. Initially, the algorithm starts with an empty set ($E_s = \emptyset$). Then, it selects the edge that yields the largest gain in the objective function [defined in (9)] and adds the edge to E_s in each iteration. Along with the updated set of edges, the number of connected subgraphs is increased after each iteration. The iterative process is terminated when the number of connected subgraphs ($N_{\hat{E}_s}$) reaches a preset number, i.e., $N_{\hat{E}_s} = K$, resulting in the generation of the final estimated set of edges (\hat{E}_s). Finally, one superpixel map, i.e., S , can be generated according to the estimated edges (\hat{E}_s), which oversegments the spatial area into multiple homogeneous patches.

2) *Estimation of Superpixel-Level Class Probability*: The pixels in each superpixel are first adaptively represented with the JSR model [defined in (2)]. Then, the class probability for each pixel in the same superpixel is estimated according to the reconstructed residuals. Here, the residual caused by reconstructing \mathbf{x}_i with atoms from the j th subdictionary is defined as

$$r^j(\mathbf{x}_i) = \|\mathbf{x}_i - \mathbf{D}^j \cdot \boldsymbol{\alpha}_i\|_2 \quad (10)$$

where \mathbf{x}_i represents one of the pixels in \mathbf{X}_i , and $\boldsymbol{\alpha}_i$ is the corresponding sparse coefficient vector. Generally, a small residual means that the test sample can be well represented by the selected atoms from one specific subdictionary. Thus, a larger class probability should be assigned. Then, the superpixel-level class probability is defined as

$$P_{\text{sup}}(y_i = j | \mathbf{x}_i) = \frac{1}{Z} \exp \left(-\frac{(r^j(\mathbf{x}_i))^2}{\sigma^2} \right) \quad (11)$$

¹<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

where Z is the normalization factor, which is calculated by $Z = \sum_{j=1}^C \exp(-((r^j(\mathbf{x})/\sigma^2))$. In (11), σ^2 is a bias, which is estimated based on the deviation of pixels within each superpixel. Assuming that one pixel \mathbf{x}_i is assigned to the j th class, the estimated superpixel-level class probability is

$$P_{\text{sup}}(y_i|\mathbf{x}_i) = P_{\text{sup}}(y_i = j|\mathbf{x}_i). \quad (12)$$

C. Probabilistic-Fusion-Based Labeling

The probabilistic-fusion-based labeling result aims at adaptively combining both the pixel-level and superpixel-level class probabilities to generate the final class estimation. Here, a weight matrix W ($W \in \mathbb{R}^{N_t \times C}$) is introduced to guide the fusion process, where each weight in W , i.e., $W_{i,j}$, is defined as

$$W_{i,j} = \frac{1}{M_i} \sum_{i_+ \in S_i} \delta(j, \hat{y}_{i_+}^s) \quad (13)$$

where S_i is the superpixel that contains pixel \mathbf{x}_i , M_i is the total number of pixels within superpixel S_i , and $\delta(\cdot, \cdot)$ is the Kronecker delta function. $\hat{y}_{i_+}^s$ is the initial class estimation for \mathbf{x}_i by the superpixel-level classifier, which is determined by

$$\begin{aligned} \hat{y}_{i_+}^s &= \arg \max_j P_{\text{sup}}(y_{i_+} = j|\mathbf{x}_{i_+}) \\ \text{s.t. } j &= 1, 2, \dots, C. \end{aligned} \quad (14)$$

Specifically, if the weight is close to 1, the spatial area that corresponds to the current superpixel S_i will be determined to have a high homogeneity level. That is, there is a high probability that pixels in S_i should belong to one class. Therefore, a higher confidence should be laid on the superpixel-level classification results for pixels in S_i . Otherwise, with a lower value of weight, the pixels in superpixel S_i are more likely to belong to two or more classes. Correspondingly, a lower confidence is given to the superpixel-level label estimation. This way, the $P_{\text{pix}}(y_i|\mathbf{x}_i)$ and $P_{\text{sup}}(y_i|\mathbf{x}_i)$ can be adaptively combined to generate a more accurate joint posterior probability. Let $y_i = j$ denote the class label of pixel \mathbf{x}_i and w_i denote the corresponding weight calculated by (13), i.e., $w_i = W_{i,j} = W_{i,y_i}$. Then, the joint posterior probability, i.e., $P(y_i|\mathbf{x}_i)$, is defined as

$$P(y_i|\mathbf{x}_i) = (1 - w_i)P_{\text{pix}}(y_i|\mathbf{x}_i) + w_iP_{\text{sup}}(y_i|\mathbf{x}_i). \quad (15)$$

By integrating the joint probabilities defined in (15), the maximization problem in (5) is rewritten as

$$\begin{aligned} \hat{Y} &= \arg \max_Y \left\{ \sum_{i=1}^N \log P(y_i|\mathbf{x}_i) + \log P(Y) \right\} \\ &= \arg \max_Y \left\{ \sum_{i=1}^N \log \{ (1 - w_i)P_{\text{pix}}(y_i|\mathbf{x}_i) + w_iP_{\text{sup}}(y_i|\mathbf{x}_i) \} \right. \\ &\quad \left. - \beta \sum_{i \in N_i} \{ 1 - \delta(y_i, y_{i_+}) \} \right\}. \end{aligned} \quad (16)$$

Algorithm 1 Probabilistic Fusion for Labeling

Input: C , number of classes; N_t , number of test samples;
 For $i = 1, 2, \dots, N_t$, \mathbf{x}_i , test sample;
 $P_{\text{pix}}(y_i|\mathbf{x}_i)$, pixel-level class probabilities;
 $P_{\text{sup}}(y_i|\mathbf{x}_i)$, superpixel-level class probabilities;
 S_i , indexes of pixels in the superpixel that contains \mathbf{x}_i ;
 M_i , number of pixels in S_i .

Output: \hat{Y} , $\hat{Y} = \{\hat{y}_i\}$, $i = 1, 2, \dots, N_t$, estimated labeling result.

1. Calculate the adaptive weight matrix W

$$\begin{aligned} \hat{y}_{i_+}^s &= \arg \max_j P_{\text{sup}}(y_i = j|\mathbf{x}_i), \quad \text{s.t. } j = 1, 2, \dots, C \\ W_{i,j} &= \frac{1}{M_i} \sum_{i_+ \in S_i} \delta(j, \hat{y}_{i_+}^s). \end{aligned}$$

2. Generate the joint posterior probabilities

$$\begin{aligned} P(y_i|\mathbf{x}_i) &= (1 - w_i)P_{\text{pix}}(y_i|\mathbf{x}_i) \\ &\quad + w_iP_{\text{sup}}(y_i|\mathbf{x}_i) \quad \text{s.t. } w_i = W_{i,y_i}. \end{aligned}$$

3. Construct a probabilistic fusion model by integrating the introduced joint posterior probabilities with the MLL prior

$$\begin{aligned} \hat{Y} &= \arg \max_Y \left\{ \sum_{i=1}^N \log P(y_i|\mathbf{x}_i) + \log P(Y) \right\} \\ &= \arg \max_Y \left\{ \sum_{i=1}^N \log \{ (1 - w_i)P_{\text{pix}}(y_i|\mathbf{x}_i) + w_iP_{\text{sup}}(y_i|\mathbf{x}_i) \} \right. \\ &\quad \left. - \beta \sum_{i \in N_i} \{ 1 - \delta(y_i, y_{i_+}) \} \right\}. \end{aligned}$$

4. Solve the probabilistic fusion model by the α -expansion algorithm.

(i) Set the iteration time t to be 0, $t = 0$.

Start with an arbitrary labeling result Y^0 , $Y^0 = \{y_i^0\}$,

Initiate energy function by:

$$E_f(Y^0) = - \sum_{i=1}^N \log P(y_i^0|\mathbf{x}_i) - \log P(Y^0).$$

(ii) Set the value of termination Flag: Flag = 0.

(iii) For each label α , $\alpha \in (1, C)$,

Update the iteration time: $t = t + 1$.

Find $Y' = \arg \min_Y E_f(Y^{(t-1)})$ in one α -expansion move;

If $E_f(Y') < E_f(Y^{(t-1)})$, set $Y^t = Y'$, and Flag = 1.

If $E_f(Y') \geq E_f(Y^{(t-1)})$, set $Y^t = Y^{(t-1)}$, and Flag = 0.

(iv) If Flag = 1, go to (ii).

If Flag = 0, generate the final estimated labeling result, $\hat{Y} = Y^t$.

Then, by defining an energy cost function as $E_f(Y) = - \sum_{i=1}^N \log P(y_i|\mathbf{x}_i) - \log P(Y)$, the maximization optimization in (16) can be transformed into an energy minimization

TABLE I
NUMBER OF TRAINING AND TEST SAMPLES OF DIFFERENT CLASSES IN THE FOUR TEST DATA SETS

Indian Pines				Salinas			University of Pavia			Washington DC		
Class	Name	Train	Test	Name	Train	Test	Name	Train	Test	Name	Train	Test
1	Alfalfa	10	36	Weeds	20	1989	Asphalt	100	6531	Roof	125	3004
2	Corn-no till	143	1285	Weed	37	3689	Meadows	100	18549	Grass	72	1718
3	Corn-min till	83	747	Fallow	20	1956	Gravel	100	1999	Road	56	1346
4	Corn	24	213	Fallow plow	14	1380	Trees	100	2964	Train	50	1214
5	Grass/pasture	48	435	Fallow smooth	27	2651	Metal sheet	100	1245	Tree	48	1146
6	Grass/trees	73	657	Tubble	40	3919	Bare Soil	100	4929	Shadow	45	1075
7	Grass/pasture-mowed	10	18	Celery	36	3543	Bitumen	100	1230			
8	Hay-windrowed	48	430	Grapes	113	11158	Bricks	100	3582			
9	Oats	10	10	Soil	62	6141	Shadows	100	847			
10	Soybean-no till	97	875	Corn	33	3245						
11	Soybean-min till	246	2209	Lettuce 4wk	11	1057						
12	Soybean-clean till	59	534	Lettuce 5wk	19	1908						
13	Wheat	21	184	Lettuce 6wk	9	907						
14	Woods	127	1138	Lettuce 7wk	11	1059						
15	Blgd-grass-tree-drives	39	347	Vinyard untrained	73	7195						
16	Stone-steel towers	10	83	Vinyard trellis	18	1789						
	Total	1048	9201	Total	543	53586	Total	900	41876	Total	396	9503

problem, denoted as $E_f(\hat{Y}) = \arg \min_Y E_f(Y)$. Here, the α -expansion algorithm² is used to solve this energy optimization problem. Generally, given the labeling result Y and arbitrary label α ($\alpha = 1, 2, \dots, C$), an α -expansion move allows any set of image pixels to change their labels to α while the rest of the pixels remain their labels unchanged. This α -expansion algorithm mainly contains two loops: outer loop (called as “cycle”) and inner loop (called as “iteration”), as summarized in Step 4 of Algorithm 1. In each “iteration” process, an optimal set of pixels determined by the minimum cut is allowed to modify their labels to be α within one α -expansion move. This way, an updated labeling result, Y' , is generated, resulting in the minimization of energy cost function, $E_f(Y') < E_f(Y)$. In each “cycle” process, the value of α is varied between 1 and C , and the “iteration” process is repeated. The overall process of two-stage loops is terminated if any α -expansion move will no longer lead to further minimization of the energy cost function, leading to the final estimation of labeling result.

IV. EXPERIMENTAL RESULTS

A. Experimental Setup

In this section, four HSIs,³ including the “AVIRIS Indian Pines,” “AVIRIS Salinas,” “ROSIS-03 University of Pavia,” and “Washington DC” are used to test the performance of the proposed classification method. The “AVIRIS Indian Pines image,” which covers the agricultural Indian Pines test site in Northwestern Indiana, was collected by the Visible Infrared Imaging Spectrometer (AVIRIS) sensor. This image is of size $145 \times 145 \times 220$, which has a spatial resolution of 20 m and a spectral range from 0.2 to 2.4 μm . Before the classification, 20 spectral bands (i.e., the 104th–108th, 150th–163rd, and 220th bands) are discarded due to the water absorption. The “AVIRIS Salinas” image was also captured by the AVIRIS sensor over the area of Salinas Valley, California. The data set is of size $512 \times 217 \times 224$ and has a spatial resolution of

3.7 m per pixel. Before the classification, 20 spectral bands corrupted by water absorption are removed. The “University of Pavia” image was acquired by the ROSIS-03 sensor over the campus at the University of Pavia, Italy. This image contains 103 spectral bands after discarding the seriously noisy bands, and each band is of size 610×340 . The spatial resolution of this image is 1.3 m, and the spectral coverage ranges from 0.43 to 0.86 μm . The “Washington DC” image was captured by the Hyperspectral Digital Image Collection Experiment (HYDICE) sensor over the region of Washington DC Mall. The image is of size $280 \times 307 \times 210$, with a spectral coverage ranging from 0.2 to 2.4 μm . The reference classes and the number of training/test samples for each test image are shown in Table I.

The classification results of the proposed PSPFC method are visually and quantitatively compared with some widely used classification methods, i.e., SVM [1], edge-preserving filtering (EPF) [16], JSR classifier (JSRC) [20], shape-aware dictionary learning (SADL) [23], and maximum *a posteriori* marginal-loopy belief propagation (MPM-LBP) [30]. In addition, the individual superpixel-level classifier is also compared to demonstrate the superiority of the proposed fusion method over the superpixel-based approach. Based on the work in [34], the superpixel-level classifier is achieved by applying the SOMP algorithm to jointly estimate the common label for each superpixel, which is denoted as the SOMP-SUP method. Among these compared methods, the SVM is a pixel-level classifier without considering the spatial context, while the rest of the classifiers are spectral-spatial classifiers that jointly integrate both spectral and spatial information in HSI. To objectively evaluate the classification results, three metrics of OA, average accuracy (AA), and Kappa coefficient (k) are used.

B. Parameter Setting

The parameters used in the SVM classifier are estimated by fivefold cross validation. For the EPF and JSRC, the window sizes for the four test images are tuned, respectively, to reach their best results in terms of the classification accuracy. For the SADL, the maximum iteration time is set to be 100, the reconstruction tolerance is set to be 10^{-3} , and the patch size for

²<http://www.wisdom.weizmann.ac.il/~bagon>.

³http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

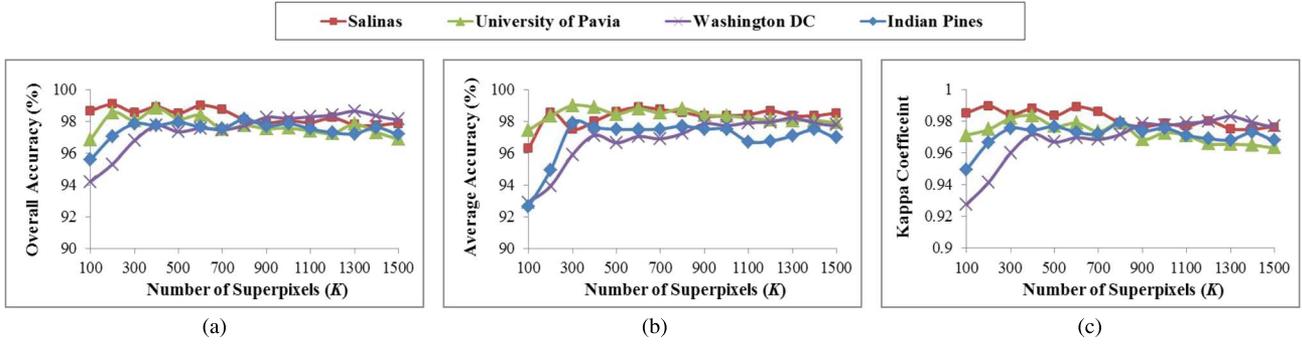


Fig. 2. Effect of the superpixel number on the proposed PSPFC method for the four HSI images. (a) OA. (b) AA. (c) κ .

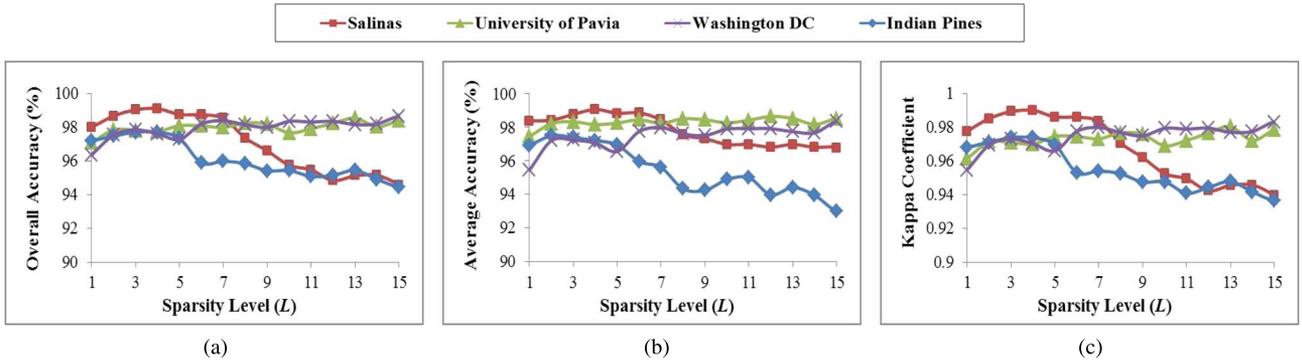


Fig. 3. Effect of the sparsity level on the proposed PSPFC method for the four HSI images. (a) OA. (b) AA. (c) κ .

different test images is tuned variously as suggested in [23]. The parameters needed in the MPM-LBP classifier are set as default values given in [30]. For the SOMP-SUP, the generation of superpixels is accomplished in the same way as in the proposed method, and the numbers of superpixels for the four test images are also set as the same values as in the proposed method for a fair comparison. In the proposed method, the parameter β in (16) is empirically set to be 2. In addition, the other two important parameters, i.e., the number of superpixels K and the sparsity level L , will be discussed in detail below.

1) *Effect of the Superpixel Number*: Fig. 2 shows the influence of parameter K on the classification accuracy, with the values ranging between 100 and 1500. Overall, it can be observed that there is a dramatically upward tendency by increasing K from 100 to 400 in terms of the three objective metrics, followed by much less obvious changes after K grows higher than 400. By taking a closer look, the changing trend for different test images varies when K is increased after 400. For the Indian Pines image and the Salinas image, relatively stable and high overall classification accuracy can be obtained when $K \in [400, 1000]$, whereas an unstable trend and drop in the AA-related curves can be observed when K is larger than 1000. For the University of Pavia image, there is a very slight decrease but more stable trend in the curves, when the value of K is larger than 1000. For the Washington DC image, there is an increase by growing K from 700 to 1000. This different changing trend is caused by the various image characteristics in different test images. For the Indian Pines image and the Salinas image, there exist multiple homogeneous regions of large size. Therefore, good classification results can be obtained for these two images, by setting the number of superpixels in more

relaxed ranges. However, for the University of Pavia image and the Washington DC image, different textures and structures of much smaller size are contained in the scenes. For these two test images, larger values for K are needed to prevent that different classes of pixels are forced into the same superpixel. In addition, the running time will increase as K is growing due to the superpixel-by-superpixel classification process. Taking both the performance and computational cost into consideration, K is set to be 600 for the Indian Pines image and the Salinas image and 1000 for the University of Pavia image and the Washington DC image, in the following experiments.

2) *Effect of the Sparsity Level*: The influence of the sparsity level, i.e., L , on the classification results for the four test images is tested, with L being varied from 1 to 15. As shown in Fig. 3, it can be observed that the OA of the proposed method clearly increases for all the test images, as the value of L rises from 1 to 3. After the value of L grows larger than 3, the changing trends of the classification accuracy for the University of Pavia image and the Washington DC image are more similar, whereas the Indian Pines image and the Salinas image change in a different way. For the Indian Pines image and the Salinas image, the classification accuracy drops dramatically as the value of L increases above 6. For the University of Pavia image and the Washington DC image, the OA of the proposed method demonstrates an increase by further increasing the sparsity level from 4 to 8, whereas there is only a slight increase or decrease with L varying between 8 and 15. The different change trends in terms of classification accuracy for different images can be explained as below. Because the numbers of dictionary atoms for some classes are small in the Indian Pines image and the Salinas image, e.g., $L \leq 10$, much higher sparsity levels will

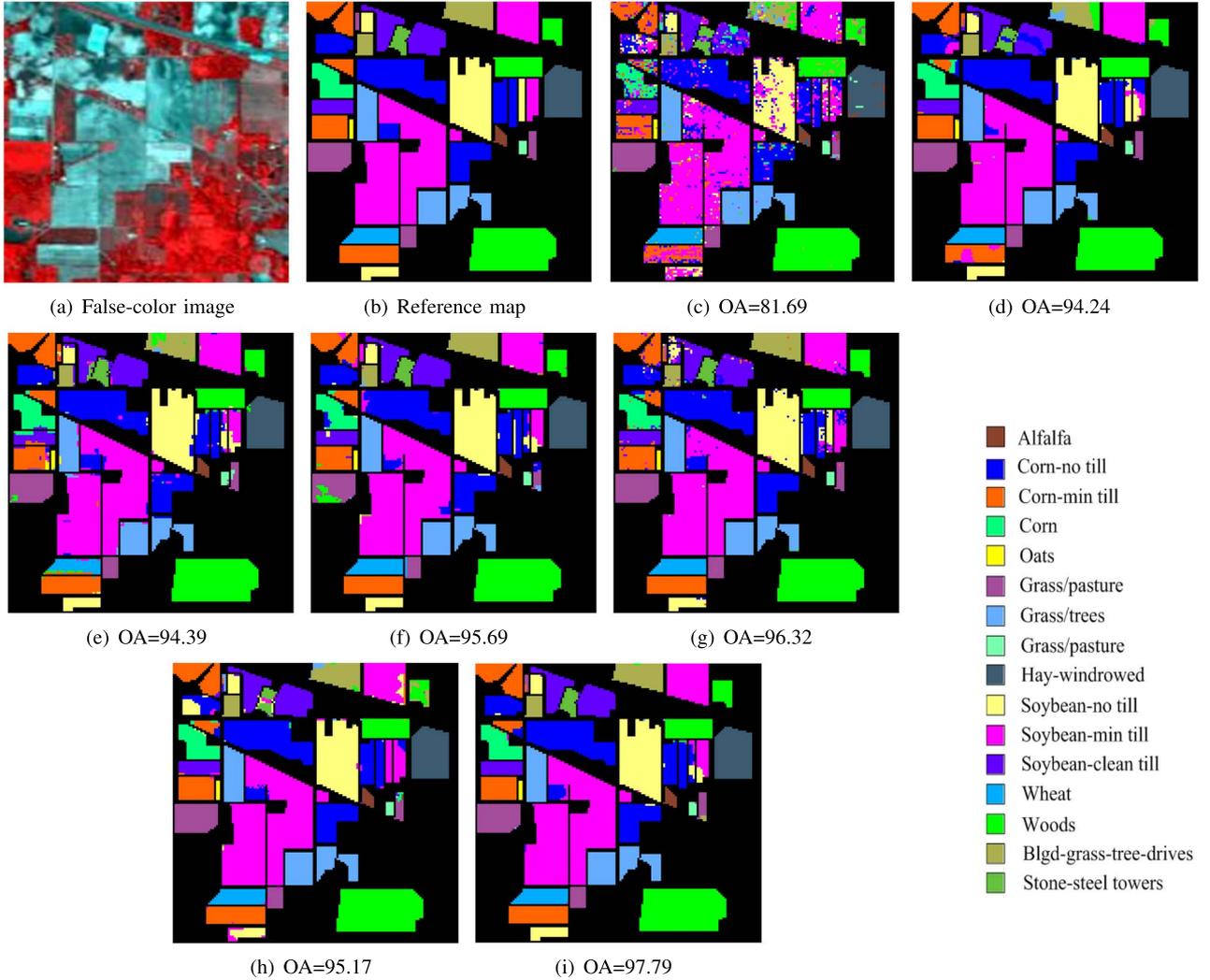


Fig. 4. Reference map and classification results for the Indian Pines image. (a) False-color image, composed of three spectral bands in the Indian Pines image. (b) Reference map. (c)–(i) Classification maps generated by different compared methods: SVM [1], EPF [16], JSRC [20], SOMP-SUP [34], SADL [23], MPM-LBP [30], and the proposed PSPFC methods. The numerical results refer to the overall classification accuracy values in percent.

enforce the proposed method to select more atoms from other incorrect classes. As a result, the classification accuracy will be negatively affected. However, for the University of Pavia image and the Washington DC image, the number of dictionary atoms from each class is larger than 40. Thus, there are less possibilities that incorrect atoms are selected for each class. Moreover, the increase of L will lead to higher computational time. Based on the above analysis, it is suggested that the value of L should be set between 3 and 6. In the following experiments, $L = 3$ for the Indian Pines image and the Salinas image, and $L = 6$ for the University of Pavia image and the Washington DC image, to generate the optimal classification results.

C. Comparison of Results

The first two experiments are tested on the AVIRIS Indian Pines image and the Salinas image, which mainly consists of large-sized homogeneous regions. The detailed numbers of training and test samples are shown in Table I, and the classification maps obtained by different classification methods are shown in Figs. 4 and 5. Besides, the quantitative results

of these two test images in terms of classification accuracies are shown in Tables II and III. For the AVIRIS Indian Pines image, 10% of the labeled samples were randomly selected as the training samples, and the remaining 90% are used as test samples. It can be observed that the SVM classifier generates a very noisy estimation in the classification map, exposing the drawbacks of the pixel-level classifiers without incorporating a spatial prior. In contrast, the other spectral-spatial methods can yield smoother classification maps. However, taking a closer look at Fig. 4, an obvious spot-like phenomenon can be seen in the SADL-based result, while the EPF, JSRC, SOMP-SUP, and MPM-LBP classifiers are likely to generate misclassification results in the areas around corners and edges. Overall, it can be seen that the classification map obtained by the proposed method looks more similar to the reference map in Fig. 4(b). In addition, better balance in classifying flat and structural regions can be obtained by the proposed method when compared with individual pixel-level and superpixel-level classifiers, demonstrating the effectiveness of the proposed probabilistic fusion scheme. For the test image of Salinas, only 1% of the labeled reference data were randomly

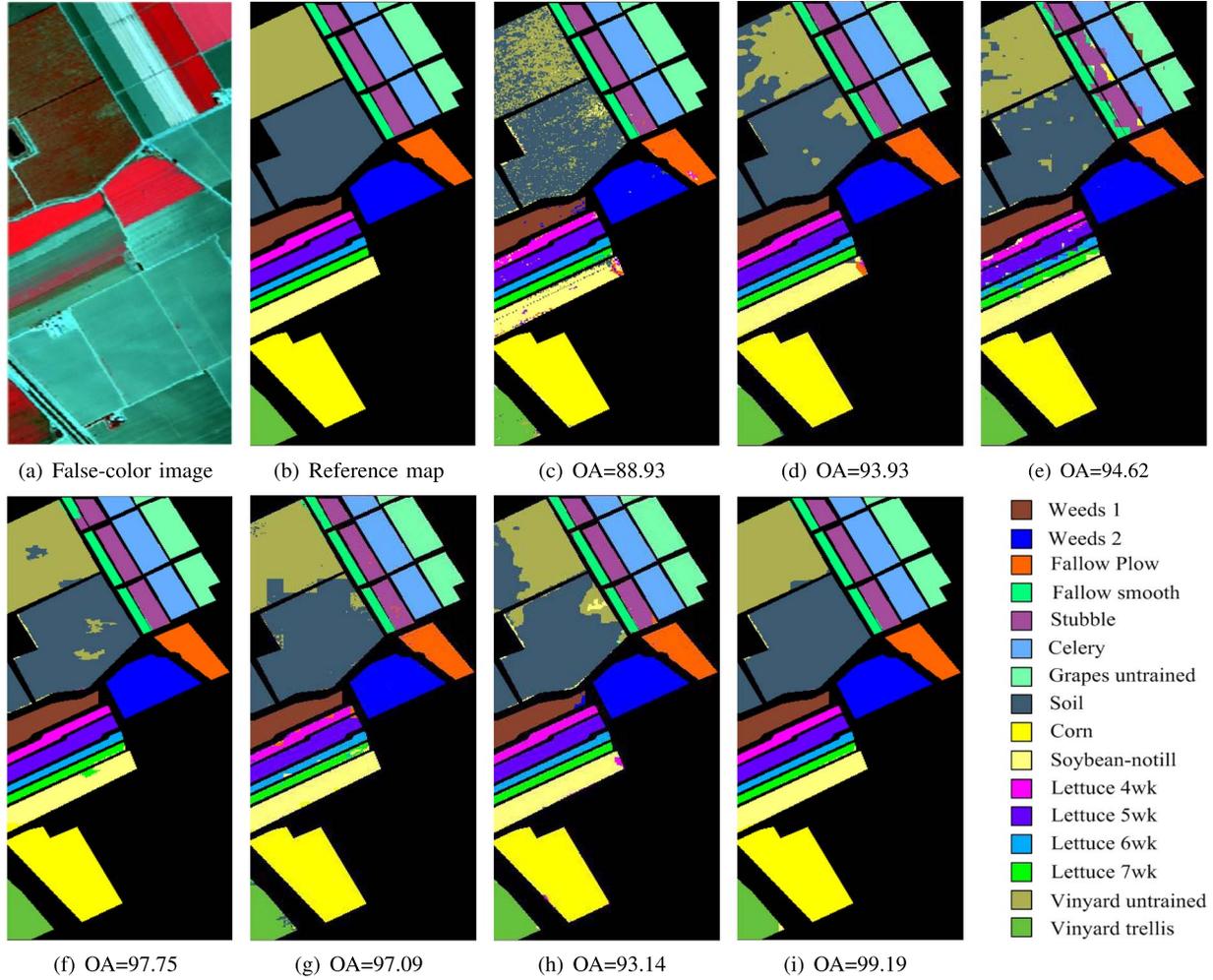


Fig. 5. Reference map and classification results for the Salinas image. (a) False-color image, composed of three spectral bands in the Salinas image. (b) Reference map. (c)–(i) Classification maps generated by different compared methods: SVM [1], EPF [16], JSRC [20], SOMP-SUP [34], SADL [23], MPM-LBP[30], and the proposed PSPFC methods. The numerical results refer to the overall classification accuracy values in percent.

TABLE II
CLASSIFICATION ACCURACY (%) OF THE INDIAN PINES IMAGE
OBTAINED BY SVM [1], EPF [16], JSRC [20], SOMP-SUP [34],
SADL [23], MPM-LBP [30], AND THE PROPOSED PSPFC METHOD

Class	SVM	EPF	JSRC	SOMP-SUP	SADL	MPM-LBP	PSPFC
1	55.28	92.22	98.06	97.50	93.73	96.67	97.92
2	74.09	90.43	93.99	94.35	94.17	91.61	95.05
3	64.61	88.17	92.42	94.04	93.86	91.41	97.64
4	48.50	83.52	93.24	93.05	93.64	90.05	94.95
5	88.07	94.76	90.92	93.40	99.42	90.71	97.04
6	95.18	99.67	93.58	99.98	99.33	99.94	100
7	87.78	95.56	99.44	96.67	100	95.56	97.22
8	96.60	100	99.79	100	100	99.81	100
9	91.00	99.00	78.00	100	73.96	100	97.50
10	74.46	90.41	92.29	92.66	94.93	93.19	94.40
11	85.11	97.95	96.41	96.36	97.59	96.81	98.45
12	67.23	93.43	89.91	93.02	84.51	97.04	97.59
13	94.73	99.67	91.36	99.57	99.92	99.46	99.46
14	94.92	98.66	99.28	99.42	99.47	97.80	99.90
15	54.50	82.19	89.19	91.24	95.84	91.30	97.01
16	82.41	91.81	88.80	92.05	98.62	86.99	96.99
OA	80.52	94.31	94.51	95.78	96.10	95.11	97.64
AA	78.40	93.59	92.92	95.83	94.94	94.90	97.67
κ	0.777	0.935	0.937	0.952	0.956	0.944	0.973

TABLE III
CLASSIFICATION ACCURACY (%) OF THE SALINAS IMAGE OBTAINED BY
SVM [1], EPF [16], JSRC [20], SOMP-SUP [34], SADL [23],
MPM-LBP [30], AND THE PROPOSED PSPFC METHOD

Class	SVM	EPF	JSRC	SOMP-SUP	SADL	MPM-LBP	PSPFC
1	98.42	99.83	100	100	99.66	99.04	100
2	99.30	99.98	99.75	100	99.92	99.76	100
3	95.49	99.25	97.83	99.51	93.81	98.72	100
4	98.22	99.54	85.87	98.20	95.90	99.00	97.36
5	96.85	98.56	86.20	98.74	98.14	97.88	99.06
6	99.43	100	98.30	99.85	99.98	99.71	99.85
7	99.27	99.96	97.28	99.92	99.47	99.81	99.87
8	85.34	93.83	94.56	97.46	94.69	91.32	99.01
9	99.31	99.91	100	99.79	99.29	99.79	100
10	87.69	95.29	94.11	95.91	97.41	93.89	97.65
11	84.83	98.63	90.10	100	98.46	95.52	99.81
12	97.89	100	90.73	97.53	96.94	100	99.65
13	97.88	98.96	80.88	98.54	93.08	98.37	92.49
14	88.09	97.88	85.40	87.32	91.52	94.41	95.79
15	58.7	67.83	91.49	93.84	89.79	69.16	98.34
16	92.86	95.01	99.35	99.49	99.46	97.59	98.60
OA	89.13	93.73	94.76	97.85	96.28	93.09	99.02
AA	92.47	96.53	93.24	97.88	96.72	95.87	98.59
κ	0.879	0.930	0.942	0.976	0.959	0.923	0.989

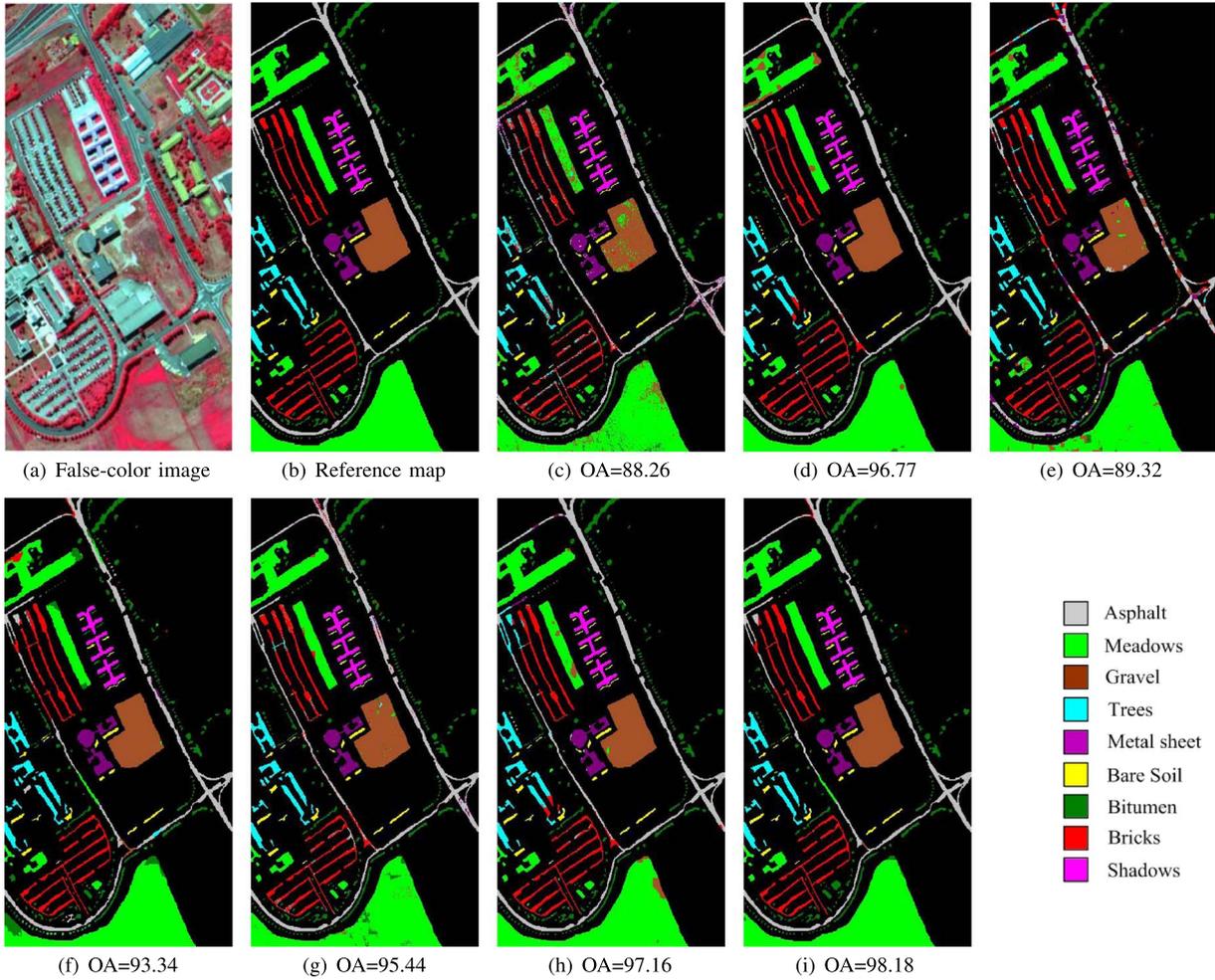


Fig. 6. Reference map and classification results for the University of Pavia image. (a) False-color image, composed of three spectral bands in the University of Pavia image. (b) Reference map. (c)–(i) Classification maps generated by different compared methods: SVM [1], EPF [16], JSRC [20], SOMP-SUP [34], SADL [23], MPM-LBP [30], and the proposed PSPFC method. The numerical results refer to the overall classification accuracy values in percent.

selected as the training samples, and the remaining 99% of data were selected as the test set for the comparison experiment. It is clear that an obvious advantage of the proposed method over the other compared methods can be achieved in classifying the 8th (grapes) and 15th (vineyard untrained) classes. Because of spectral similarity and mixture of these two classes, the compared methods fail to distinguish between them. As a result, obvious misclassification for some homogenous regions can be observed in Fig. 5(c)–(h). In contrast, the PSPFC method shows a much better discriminative capability in classifying these two classes, due to the joint use of both pixel-level spectral information and superpixel-level spatial context. The better discriminative capability of the proposed method can also be confirmed by the quantitative comparisons in Table III. Specifically, the classification accuracy values by the proposed method for classifying these two classes are, respectively, 99.01% and 98.34%, whereas the highest accuracy values by the rest of the compared methods are, respectively, 97.46% and 93.84%. Furthermore, the overall classification accuracy in terms of the three objective metrics listed in Table III also indicates that the proposed method outperformed the other compared methods, with $OA = 99.02\%$, $AA = 98.59\%$, and $\kappa = 0.989$, respectively.

The other two experiments are conducted on the University of Pavia image and the Washington DC image, which consist of more detailed structures. The visual classification maps obtained by various classifiers on this image are shown in Figs. 6 and 7, and the quantitative results in terms of classification accuracy are shown in Tables IV and V. For the University of Pavia, a fixed number, i.e., 100, of training samples for each class is adopted. It is shown in Fig. 6 that various degrees of misclassification results in the large area of meadows (the 2nd class, laid on the bottom of the image) are generated by the SVM, EPF, JSRC, SOMP-SUP, SADL, and MPM-LBP methods. In contrast, the SOMP-SUP and the proposed method can obtain smoother results in this meadows-related area, as shown in Fig. 6(f) and (i), which can benefit from the exploitation of shape-adaptive spatial information. However, the SOMP-SUP fails to accurately classify pixels around edges, which is caused by the inaccurate segmentation map. By integrating both pixel-level and superpixel-level classification estimates with the proposed probabilistic fusion framework, a promising performance in accurately classifying both the homogeneous and structural regions can be obtained. For the Washington DC image, 4% of the total number of labeled samples were randomly selected as the training samples, and the

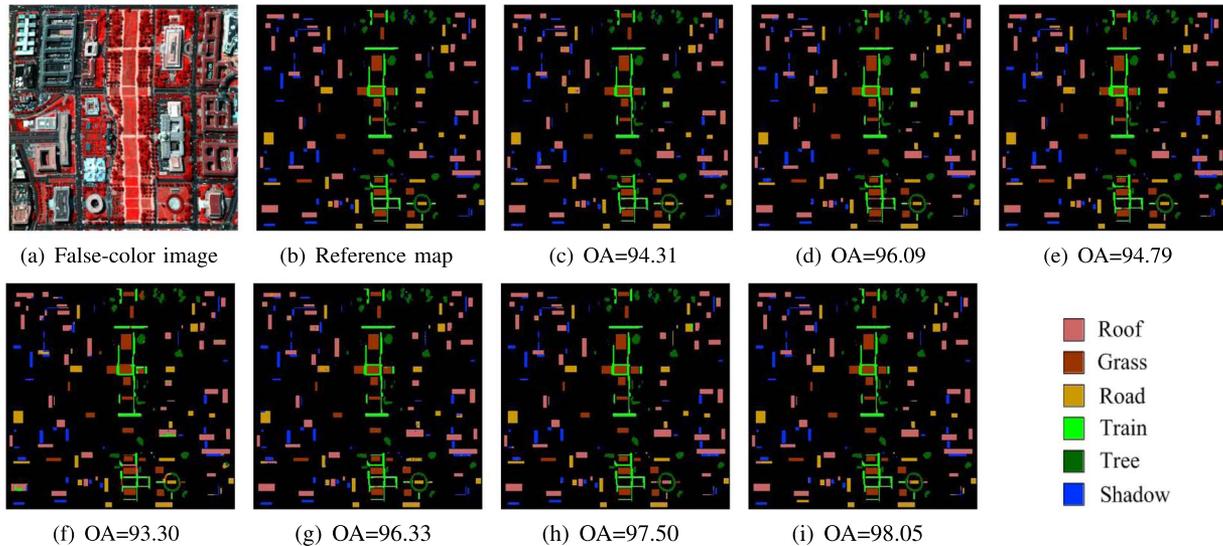


Fig. 7. Reference map and classification results for the Washington DC image. (a) False-color image, composed of three spectral bands in the Washington DC image. (b) Reference map. (c)–(i) Classification maps generated by different compared methods: SVM [1], EPF [16], JSRC [20], SOMP-SUP [34], SADL [23], MPM-LBP [30], and the proposed PSPFC method. The numerical results refer to the overall classification accuracy values in percent.

TABLE IV
CLASSIFICATION ACCURACY (%) OF THE UNIVERSITY OF PAVIA IMAGE OBTAINED BY SVM [1], EPF [16], JSRC [20], SOMP-SUP [34], SADL [23], MPM-LBP [30], AND THE PROPOSED PSPFC METHOD

Class	SVM	EPF	JSRC	SOMP-SUP	SADL	MPM-LBP	PSPFC
1	81.31	96.70	66.59	84.86	97.18	93.83	93.94
2	88.73	96.35	92.40	94.17	99.44	96.93	98.76
3	84.68	94.03	96.57	98.48	88.11	89.76	97.91
4	94.95	93.38	92.97	89.44	92.55	95.80	97.24
5	99.39	100	96.86	98.77	99.99	99.92	99.76
6	89.50	99.24	94.99	97.57	92.40	99.41	100
7	93.88	99.92	98.94	99.59	86.60	98.29	99.55
8	84.32	97.08	92.07	96.82	88.87	96.33	99.24
9	99.96	99.94	66.43	86.03	99.71	99.62	99.54
OA	88.23	96.77	88.69	93.35	95.68	96.45	98.11
AA	90.75	97.40	88.65	93.97	93.87	95.28	98.44
κ	0.846	0.957	0.852	0.912	0.943	0.967	0.975

TABLE V
CLASSIFICATION ACCURACY (%) OF THE WASHINGTON DC IMAGE OBTAINED BY SVM [1], EPF [16], JSRC [20], SOMP-SUP [34], SADL [23], MPM-LBP [30], AND THE PROPOSED PSPFC METHOD

Class	SVM	EPF	JSRC	SOMP-SUP	SADL	MPM-LBP	PSPFC
1	94.91	96.93	94.49	93.08	96.83	98.00	99.84
2	97.95	99.00	96.61	96.15	98.61	99.77	99.88
3	83.79	93.57	93.55	93.61	89.31	94.26	94.93
4	96.51	94.74	93.84	97.73	98.51	98.47	97.46
5	96.33	98.05	92.60	86.79	97.25	97.42	97.26
6	95.88	90.61	95.71	88.64	98.30	96.65	95.68
OA	94.37	95.97	94.96	93.74	96.44	97.63	98.05
AA	94.23	95.49	95.01	93.63	96.47	97.43	97.55
κ	0.930	0.950	0.922	0.896	0.956	0.970	0.976

rest of the labeled samples are used as test samples. In Table V, the overall classification accuracy values for the pixel-level classifier (SVM) and the superpixel-level classifier (SOMP-SUP) are 94.37% and 93.74%, respectively, whereas the result by the proposed fusion scheme is 98.05%. The obvious increase in classification accuracy can demonstrate the effectiveness of the proposed fusion scheme in improving classification performance. Consistent with the comparison results for the

Indian Pines image and the Salinas image, these two experiments again show that the PSPFC method works better than the other compared spectral–spatial methods in terms of classification accuracy.

D. Classification Results With Different Numbers of Training Samples

Here, it is tested how the number of training samples influences the performance for various classification methods on four test images, including the images of Indian Pines, Salinas, University of Pavia, and Washington DC. For each test image, a balanced training set was constructed by randomly choosing varying pixels, with the number of training samples for each class (M_{tr}) varying between a range of [10, 30, 50, 80, 100]. The overall classification accuracy values are shown in Fig. 8. It can be seen from Table I that there are some classes of very few samples in the Indian Pines image, e.g., the 1st, 7th, and 9th classes. Thus, to test the influence of M_{tr} on classification accuracy by varying M_{tr} from 10 to 100, the ten-class reference [35] for this image is used instead of the 16-class reference shown in Table I. It is clear from Fig. 8 that the classification accuracy values for all the compared classifiers are increased as the number of training samples grows. In addition, constant higher overall classification accuracy values are obtained by the proposed method with varying numbers of training samples, demonstrating that the proposed method outperforms the other compared methods.

Furthermore, the classification results of different methods in the case of $M_{tr} = 10$ (ten samples for each class are used for training) are shown in Table VI, to numerically compare the classification performance when a small size of training samples is available. In terms of both the OA- and κ -based objective metrics, it is clear that the proposed PSPFC classifier always ranks the first among the compared methods. Although the MPM-LBP method has a competitive performance in

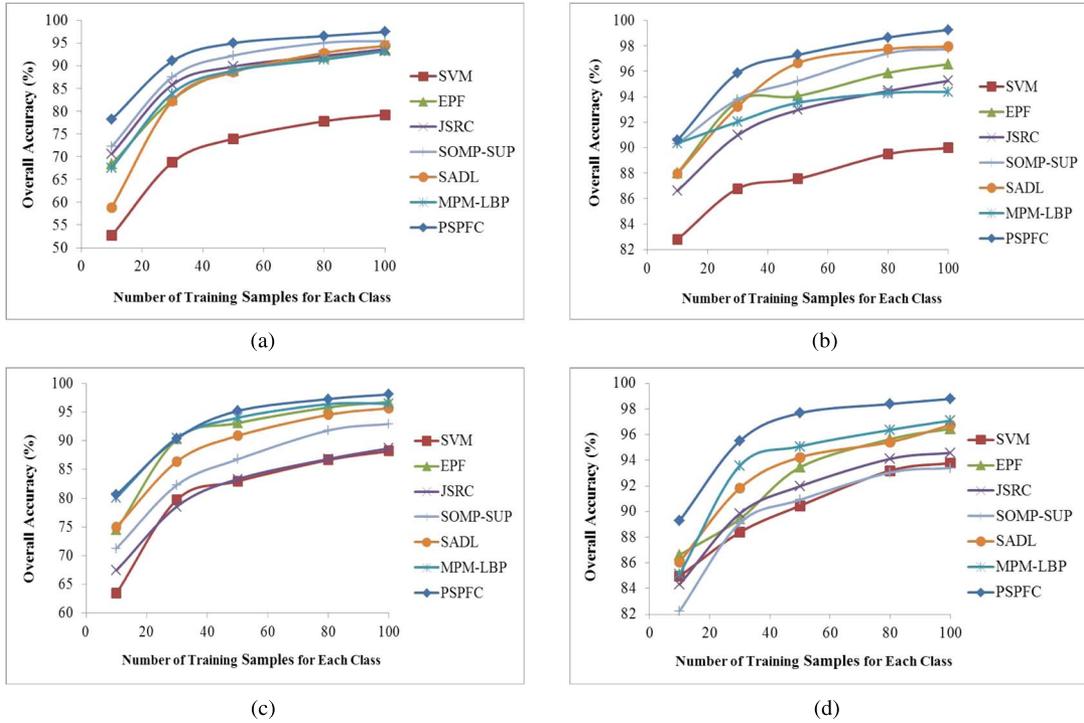


Fig. 8. Classification results for four test images with varying numbers of training samples by SVM [1], EPF [16], JSRC [20], SOMP-SUP [34], SADL [23], MPM-LBP [30], and the proposed PSPFC method. (a) Indian Pines. (b) Salinas. (c) University of Pavia. (d) Washington DC.

TABLE VI
CLASSIFICATION RESULTS BY SELECTING TEN TRAINING SAMPLES FOR EACH CLASS IN THE FOUR TEST DATA SETS

Methods	Indian Pines			Salinas			University of Pavia			Washington DC		
	OA	AA	κ	OA	AA	κ	OA	AA	κ	OA	AA	κ
SVM [1]	52.68	58.40	0.465	82.79	90.60	0.809	63.43	76.28	0.555	84.96	86.68	0.813
EPF [16]	68.37	71.96	0.637	88.02	94.46	0.867	74.48	81.84	0.676	86.62	88.34	0.834
JSRC [20]	70.64	75.16	0.664	86.62	90.15	0.851	67.43	70.35	0.588	84.32	88.52	0.809
SOMP-SUP [34]	72.25	75.57	0.681	90.31	95.32	0.890	71.25	74.61	0.638	82.25	84.01	0.782
SADL [23]	58.88	63.15	0.534	87.97	92.59	0.867	74.96	74.38	0.684	86.08	87.58	0.830
MPM-LBP [30]	67.59	63.09	0.739	90.35	94.54	0.893	80.10	87.08	0.748	85.01	89.12	0.817
PSPFC	78.24	82.52	0.750	90.58	95.83	0.895	80.58	85.81	0.750	89.29	91.10	0.872

TABLE VII
COMPUTING TIME (SECONDS) FOR THE CLASSIFICATION OF FOUR TEST IMAGES OBTAINED BY SVM [1], EPF [16], JSRC [20], SOMP-SUP [34], SADL [23], MPM-LBP [30], AND THE PROPOSED PSPFC METHOD

Image	SVM	EPF	JSRC	SOMP-SUP	SADL	MPM-LBP	PSPFC	PSPFC+Parallel Process
Indian Pines	9.72	10.97	134.19	4.13	54.64	174.03	15.15	10.31
Salinas	17.64	19.23	1236.18	12.97	331.25	864.97	33.35	20.16
University of Pavia	19.48	22.16	560.87	15.89	205.80	1585.03	38.96	23.19
Washington DC	5.99	6.70	27.52	5.34	73.58	577.79	12.17	6.84

classifying the University of Pavia image, it is inferior to the proposed method in terms of classification accuracy for both the Indian Pines image and the Washington DC image. Overall, it can be concluded that the better classification performance is obtained by the proposed method when limited training samples are available. The superiority of the proposed PSPFC classifier is due to the effective fusion of pixel-level and superpixel-level classifiers. With the introduced fusion scheme, the pixel-level classifier will benefit from the classification of structured regions, and the superpixel-level method can effectively exploit the spatial similarity in the neighborhood to improve the discriminative capability in homogeneous regions.

E. Comparison of Computational Time

In this section, the running times for both the proposed and compared methods to classify the four test HSIs are reported in Table VII, to quantitatively compare the efficiency of different classifiers. All of the programs are executed on a computer with an Intel Core E5-2603 and CPU 1.60 GHz, and the software platform is MATLAB R2011a. For the SVM classifier, the time cost by fivefold cross validation for parameter selection is excluded in Tables VII and VIII. It is clear from Table VII that the JSRC, SADL, and MPM-LBP classifiers take much longer time than SVM, EPF, SOMP-SUP, and the proposed method. For the JSRC method, the joint sparse coding of all

TABLE VIII

COMPUTING TIME (SECONDS) FOR EACH STEP IN THE PROPOSED METHOD. STAGE 1: PIXEL-LEVEL CLASS PROBABILITY ESTIMATION. STAGE 2: SUPERPIXEL-LEVEL CLASS PROBABILITY ESTIMATION. STAGE 3: PROBABILISTIC FUSION FOR LABELING

Image	Stage 1	Stage 2	Stage 3
Indian Pines	9.72	4.84	0.59
Salinas	17.64	13.19	2.52
University of Pavia	19.48	15.77	3.71
Washington DC	5.99	5.33	0.85

the pixels in each overlapped patch is the main cause for the increase in computational cost. For the MPM-LBP classifier, the loopy belief propagation occupies the main computational cost because of the multilevel-loop-based optimization process. For the SADL classifier, the stage of the optimal SADL is time consuming due to the two-step strategy of alternative iterative sparse coding and dictionary update until a predefined condition of local convergence is satisfied. Among all the compared methods, the high efficiency in terms of running time by the SOMP-SUP method benefits from the superpixel-by-superpixel processing, instead of a pixel-by-pixel manner by the other classifiers. In contrast, SVM, EPF, and the proposed method are comparable but slightly slower than the SOMP-SUP classifier. In addition, it should be noted that both the EPF classifier and the proposed method include the process of SVM-based classification. Thus, longer time is needed than the individual SVM classifier.

In addition, the time cost by each step of the proposed method is shown in Table VIII, to more clearly illustrate the computational cost and the probable acceleration approach. Specifically, the proposed PSPFC method mainly contains three steps: pixel-level class probability estimation, superpixel-level class probability estimation, and probabilistic-fusion-based labeling result. From Table VIII, it can be observed that Stages 1 and 2 occupy the main computational cost in classifying HSI. However, it should also be noted that the first two stages are independent. Thus, they can be processed in a parallel way (denoted as ‘‘PSPFC+Parallel Process’’), accelerating the overall classification process. The reduced time by the ‘‘PSPFC+Parallel Process’’ is shown in Table VII. In addition, it is also easy to find that the time cost by Stage 3 can be ignored when compared with the overall running time. This is because the main calculation in Stage 3 is to solve the MAP model via the α -expansion algorithm, which is a linear optimization problem with complexity of $O(n)$.

V. CONCLUSION

In this paper, a novel spectral-spatial HSI classification method by the probabilistic fusion of pixel-level and superpixel-level class estimates has been proposed. First, the probabilistic SVM classifier uses discriminative spectral information to generate pixel-level class probability estimation. Second, the shape-adaptive spatial information is exploited by a JSR, resulting in the estimation of superpixel-level class probabilities. Finally, both levels of class probabilities are adaptively integrated in a MAP model for the estimation of the classification result. Due to the incorporation of both pixel-level spectral information and superpixel-level spatial information,

obvious improvement in classification results can be obtained when compared with individual pixel-level and superpixel-level classifiers. This demonstrates the effectiveness of the proposed probabilistic fusion scheme. In addition, better classification accuracy over the other compared state-of-the-art classifiers indicates the superiority of the proposed PSPFC method. In our future work, multifeature (i.e., the pixel-level and superpixel-level features) fusion schemes that can jointly use the complementary information will be investigated. Furthermore, extending this supervised classification work to semisupervised and unsupervised frameworks is another emphasis of future research.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe from Purdue University and the NASA Jet Propulsion Laboratory for providing the free downloads of the hyperspectral data sets. They would also like to thank Dr. J. Li and Dr. A. Soltani-Farani for providing the executive codes of the MPM-LBP and SADL classifiers, respectively.

REFERENCES

- [1] 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.
- [2] J. Li, J. Bioucas-Dias, and A. Plaza, ‘‘Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning,’’ *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] S. Rajan, J. Ghosh, and M. M. Crawford, ‘‘An active learning approach to hyperspectral data classification,’’ *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1231–1242, Apr. 2008.
- [8] D. Tuia, F. Ratle, F. Pacifici, M. F. Kanevski, and W. J. Emery, ‘‘Active learning methods for remote sensing image classification,’’ *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 7, pp. 2218–2232, Jul. 2009.
- [9] J. A. 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.
- [10] X. Kang, S. Li, and J. A. Benediktsson, ‘‘Feature extraction of hyperspectral images with image fusion and recursive filtering,’’ *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3742–3752, Jun. 2014.
- [11] J. Xia, J. Chanussot, P. Du, and X. He, ‘‘Spectral-spatial classification for hyperspectral data using rotation forests with local feature extraction and Markov random fields,’’ *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2532–2546, May 2015.
- [12] 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.
- [13] W. Li, C. Chen, H. Su, and Q. Du, ‘‘Local binary patterns and extreme learning machine for hyperspectral imagery classification,’’ *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 7, pp. 3681–3693, Jul. 2015.
- [14] Y. Zhong, Q. Zhu, and L. Zhang, ‘‘Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery,’’ *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 6207–6222, Nov. 2015.

- [15] J. Li, P. Marpu, A. Plaza, J. M. Bioucas-Dias, and J. A. Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4816–4829, Sep. 2013.
- [16] 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.
- [17] Y. Gu, T. Liu, X. Jia, J. A. Benediktsson, and J. Chanussot, "Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3235–3247, Jun. 2016.
- [18] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, Feb. 2009.
- [19] R. Rubinfeld, A. M. Bruckstein, and M. Elad, "Dictionaries for sparse representation modeling," *Proc. IEEE*, vol. 98, no. 6, pp. 1045–1057, Jun. 2010.
- [20] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
- [21] J. Li, H. Zhang, Y. Huang, and L. Zhang, "Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3707–3719, Jun. 2014.
- [22] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738–7749, Dec. 2014.
- [23] A. Soltani-Farani, H. R. Rabiee, and S. A. Hosseini, "Spatial-aware dictionary learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 527–541, Jan. 2015.
- [24] 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.
- [25] C. Persello, A. Boularias, M. Dalponte, T. Gobakken, E. Naeset, and B. Scholkopf, "Cost-sensitive active learning with lookahead: Optimizing field surveys for remote sensing data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6652–6664, Oct. 2014.
- [26] S. Sun, P. Zhong, H. Xiao, and R. Wang, "Active learning with Gaussian process classifier for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 1746–1760, Apr. 2015.
- [27] B. B. Damodaran, R. R. Nidamanuri, and Y. Tarabalka, "Dynamic ensemble selection approach for hyperspectral image classification with joint spectral and spatial information," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 2405–2417, Jun. 2015.
- [28] 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.
- [29] X. Kang, S. Li, L. Fang, M. Li, and J. A. Benediktsson, "Extended random walker-based classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 144–153, Jan. 2015.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] K. Bernard, Y. Tarabalka, J. Angulo, J. Chanussot, and J. A. Benediktsson, "Spectral–spatial classification of hyperspectral data based on a stochastic minimum spanning forest approach," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2008–2021, Apr. 2012.
- [34] L. Fang, S. Li, X. Kang, and J. A. Benediktsson, "Spatial classification of hyperspectral images with a superpixel-based discriminative sparse model," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4186–4201, Aug. 2015.
- [35] J. Li, H. Zhang, and L. Zhang, "Efficient superpixel-level multitask joint sparse representation for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 10, pp. 5338–5351, Oct. 2015.
- [36] G. Zhang, X. Jia, and J. Hu, "Superpixel-based graphical model for remote sensing image mapping," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 11, pp. 5861–5871, Nov. 2015.
- [37] M. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 2097–2104. [Online]. Available: <http://mingyuliu.net>
- [38] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2274–2281, Nov. 2012.
- [39] J. Shen, Y. Du, W. Wang, and X. Li, "Lazy random walks for superpixel segmentation," *IEEE Trans. Image Process.*, vol. 23, no. 4, pp. 1451–1462, Apr. 2014.
- [40] C. W. Hsu and C. J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.
- [41] B. Liu, Z. Hao, and E. C. C. Tsang, "Nesting one-against-one algorithm based on SVMs for pattern classification," *IEEE Trans. Neural Netw.*, vol. 19, no. 12, pp. 2044–2052, Dec. 2008.
- [42] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 27, pp. 1–27, Apr. 2011.
- [43] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.
- [44] D. Landgrebe, *Signal Theory Methods in Multispectral Remote Sensing*. Hoboken, NJ, USA: Wiley, 2003.
- [45] S. Z. Li and S. Singh, *Markov Random Field Modeling in Image Analysis*. Berlin, Germany: Springer-Verlag, 2009.
- [46] J. Yedidia, W. Freeman, and Y. Weiss, "Constructing free energy approximations and generalized belief propagation algorithms," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2282–2312, Jul. 2005.
- [47] V. Kolmogorov, "Convergent tree-reweighted message passing for energy minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 10, pp. 1568–1583, Oct. 2006.
- [48] Y. Boykov, O. Veksler, and R. Zabih, "Efficient approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 12, pp. 1222–1239, Nov. 2001.
- [49] J. Li, J. 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, Feb. 2012.
- [50] M. Golipour, H. Ghassemian, and S. F. Mirzapour, "Integrating hierarchical segmentation maps with MRF prior for classification of hyperspectral images in a Bayesian framework," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 54, no. 2, pp. 805–816, Feb. 2016.
- [51] T. F. Wu, C. J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *J. Mach. Learn. Res.*, vol. 5, no. 4, pp. 975–1005, Aug. 2004.



Shutao Li (M'07–SM'15) received the B.S., M.S., and Ph.D. degrees from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively, all in electrical engineering.

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, Surrey, U.K. From April 2005 to June 2005, he was a Visiting Professor with the Department of Computer Science, Hong Kong University of Science and Technology. In 2013, he was Granted the National Science Fund for Distinguished Young Scholars in China. He is currently a Full Professor with the College of Electrical and Information Engineering, Hunan University. He is also a Chang-Jiang Scholar Professor appointed by the Ministry of Education of China. He has authored/coauthored more than 180 refereed papers. His professional interests include compressive sensing, sparse representation, image processing, and pattern recognition.

Dr. Li is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and a member of the Editorial Board of *Information Fusion* and *Sensing and Imaging*. He was a recipient of two 2nd-Grade National Awards at the Science and Technology Progress of China in 2004 and 2006.



Ting Lu (S'16) received the B.Sc. degree from Hunan University, Changsha, China, in 2011, where she is currently working toward the Ph.D. degree.

From September 2014 to September 2015, she was a visiting Ph.D. student with the Department of Information Engineering and Computer Science, University of Trento, Trento, Italy, supported by the China Scholarship Council. Her research interests include sparse representation and remote sensing image processing.



Leyuan Fang (S'10–M'14) received the Ph.D. degree in electrical engineering from Hunan University, Changsha, China, in 2015.

From September 2011 to September 2012, he was a visiting Ph.D. student with the Department of Ophthalmology, Duke University, Durham, NC, USA, supported by the China Scholarship Council. Since June 2013, he has been an Assistant Professor with the College of Electrical and Information Engineering, Hunan University. His research interests include sparse representation and multiresolution analysis in

remote sensing and medical image processing.

Dr. Fang was a recipient of the Scholarship Award for Excellent Doctoral Student Granted by the Chinese Ministry of Education in 2011.



Xiuping Jia (M'93–SM'03) received the B.Eng. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 1982 and the Ph.D. degree in electrical engineering from The University of New South Wales, Sydney, Australia, in 1996.

Since 1988, she has been with the School of Information Technology and Electrical Engineering, The University of New South Wales, Canberra Campus, Australia, where she is currently a Senior Lecturer. She is also a Guest Professor with China University

of Petroleum, Beijing and Harbin Engineering University, Harbin, China, and an Adjunct Researcher with the China National Engineering Research Center for Information Technology in Agriculture. She has authored or coauthored more than 160 referred papers, including 65 journal papers. She is the coauthor of the remote sensing textbook titled *Remote Sensing Digital Image Analysis* [Springer-Verlag, 3rd ed. (1999) and 4th ed. (2006)]. Her research interests include remote sensing, image processing, and spatial data analysis.

Dr. Jia is a Subject Editor of the *Journal of Soils and Sediments* and an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.



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

From 2009 to 2015, he was the Pro Rector of Science and Academic Affairs and a Professor of electrical and computer engineering with the University of Iceland. On July 1, 2015, he became the Rector of the University of Iceland. He has published

extensively in his fields of interest. His research interests include remote sensing, image analysis, pattern recognition, biomedical analysis of signals, and signal processing.

Dr. Benediktsson was the 2011–2012 President of the IEEE Geoscience and Remote Sensing Society (GRSS) and has been on the GRSS Administrative Committee since 2000. He was the Editor-in-Chief of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING (TGRS) from 2003 to 2008 and has served as an Associate Editor of TGRS since 1999, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003, and IEEE ACCESS since 2013. He is on the Editorial Board of the PROCEEDINGS OF THE IEEE and the International Editorial Board of the *International Journal of Image and Data Fusion* and was the Chairman of the Steering Committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING during 2007–2010. He is a co-founder of the biomedical startup company Oxymap. He is a Fellow of the International Society for Optics and Photonics (SPIE). He is a member of the 2014 IEEE Fellow Committee. He received the Stevan J. Kristof Award from Purdue University in 1991 as outstanding graduate student in remote sensing. In 1997, he was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award; in 2000, he was Granted the IEEE Third Millennium Medal; in 2004, he was a corecipient of the University of Iceland's Technology Innovation Award; in 2006, he received the yearly research award from the Engineering Research Institute of the University of Iceland; and in 2007, he received the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society. He was also a corecipient of the 2012 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Paper Award, and in 2013, he was a corecipient of the IEEE GRSS Highest Impact Paper Award. In 2013, he received the IEEE/VFI Electrical Engineer of the Year Award. In 2014, he was a corecipient of the International Journal of Image and Data Fusion Best Paper Award. He is a member of the Association of Chartered Engineers in Iceland (VFI), Societas Scientiarum Islandica, and Tau Beta Pi.