

# Set-to-Set Distance-Based Spectral–Spatial Classification of Hyperspectral Images

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

**Abstract**—A novel set-to-set distance-based spectral–spatial classification method for hyperspectral images (HSIs) is proposed. In HSIs, the spatially connected and spectrally similar pixels within each homogeneous region can be considered as one set of test samples, i.e., a test set, which should belong to the same class. In addition, each class of labeled pixels can be regarded as one set of training samples, i.e., a training set. Therefore, it is a natural consideration in the proposed method to measure the similarity between test and training sets via specific set-based distance criteria and then decide the classification label for each test set, accordingly. Specifically, the superpixel algorithm-based oversegmentation technique jointly exploits both the spatial similarity and structural information to first divide the HSI into multiple perceptually uniform regions. As a result, each segmented region corresponds to one test set. Then, each test/training set is represented with an affine hull (AH) model, which utilizes both the similarity and variance of pixels within each set to adaptively characterize the set. Finally, the class label for each test set is determined based on the closest geometry distance between test and training AHs. Experimental results on real HSI data sets demonstrate the superiority of the proposed algorithm over several well-known classification approaches, in terms of classification accuracy and computational speed.

**Index Terms**—Affine hull (AH) representation, hyperspectral image (HSI), set-to-set distance, spectral–spatial classification.

## I. INTRODUCTION

**H**YPERSPECTRAL image (HSI) classification plays a very important role in different applications, such as land use analysis, pollution monitoring, wide-area reconnaissance, and field surveillance. In HSIs, each pixel is a high-dimensional vector, whose entries denote the spectral responses of multiple spectral bands. Based on the discriminant spectral information

Manuscript received February 21, 2016; revised June 17, 2016; accepted July 9, 2016. This paper was supported 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.

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

L. Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy (e-mail: lorenzo.bruzzone@ing.unitn.it).

J. A. Benediktsson is with the Faculty of Electrical and Computer Engineering, University of Iceland, 101 Reykjavik, 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.2016.2596260

provided in HSIs, each pixel can be assigned a unique label, indicating the belonging class.

In the literature, the development of effective HSI classification methods is an active research area of remote sensing. Among these methods, the support vector machine (SVM) classifier [1], which aims to find an optimal separating hyperplane to solve the binary classification problem, has proven to be a powerful tool for HSI classification. Another widely used discriminative hyperspectral classifier is multinomial logistic regression (MLR) [2], which generates classification results under a probabilistic framework. In addition, other supervised classification techniques based on pixelwise classifiers have been developed for general purposes, e.g., random forests [3], [4], neural networks [5], [6], active learning [7], [8], and sparse representation [9], [10]. These pixelwise classifiers regard each pixel as the basic classification unit, which assign the classified label according to the spectral similarity between each single test pixel and each class of training samples. However, without considering the spatial context, pixel-based classification methods are likely to be negatively affected by noisy and mixed pixel characteristics.

In addition, some image-segmentation-based classification methods [11]–[13] have been developed, aiming at improving the classification performance by adaptively incorporating spatial contextual information. Specifically, segmentation-based methods first segment the spatial regions into nonoverlapping homogeneous patches with some hyperspectral segmentation techniques (e.g., watershed-based segmentation [13] and stochastic minimum spanning forest [14]). Then, with the initial classification results by pixelwise classifiers, the classification results for each segmented region can be obtained via some decision fusion. For example, with the majority voting, all pixels in the same region are assigned to the most frequent class within this region. It can be noticed that these segmentation-based classification methods are still pixelwise since they cannot jointly exploit both the spectral and spatial information in the process of classification.

Recently, many researchers have focused on designing spectral–spatial classifiers for HSI, in order to simultaneously exploit both spectral and spatial information [15]. For example, different feature extraction algorithms [16], [17], and multiple feature fusion techniques [18], [19] have been introduced to better characterize the spectral–spatial feature of different materials. In [20]–[23], by designing different kernel functions (e.g., the morphological kernel [20], composite kernel [21], [22], and region kernel [23]), both spectral information and spatial information are integrated to extend the traditional classifiers

(e.g., SVM and MLR), in order to enhance the discriminative capability. The aforementioned spectral–spatial classification methods have been demonstrated to be effective in improving classification accuracy. However, the fixed-size square-window-based spatial information cannot adaptively capture the spatial structure. For example, the square-window-based region for a test pixel along an edge may contain different classes, which will negatively affect the classification result for the test pixel. In addition, as one powerful image processing tool, the sparse representation (SR) model has gained much attention for classification of HSI [24]–[27]. Generally, SR-model-based methods simultaneously represent the samples in a neighborhood with a few common training samples selected from a learnt dictionary. By involving the neighboring spatial information, the SR-based methods can provide promising classification accuracy. However, repeatedly sparse-coding pixels within overlapped regions always leads to an increase in computational burden. In addition, other types of spectral–spatial classification methods, such as active-learning-based approaches [28]–[30] and probabilistic-modeling-based methods [31]–[33], have also been successfully applied to HSI classification.

In this paper, a novel set-based spectral–spatial HSI classification method is proposed. By regarding each class of training samples as one training set and each homogeneous region containing similar spectral pixels as test sets, the proposed setwise method aims at classifying each test set according to set-based distance. Specifically, test sets are generated via the superpixel-based oversegmentation. By exploiting both spatial similarity and structural information in images, the superpixel technique can divide an image into shape-adaptive (SA) homogeneous regions [34]–[36]. In the context of HSI, each segmented SA region corresponds to a test set that contains spatially connected and spectrally similar pixels. Then, each test and training set is adaptively represented by an affine hull (AH) model, which utilizes both the similarity and the variance of pixels within each set to adaptively characterize the set [37]–[39]. Based on the AH-model-based representation for each set, the similarity between test and training sets can be transformed to the geometric distance between the corresponding AHs. Finally, the label for each test set is determined according to the minimum distance between the test and training AHs. By combining both spatial and spectral information to classify the HSI in a set-by-set manner, both the classification accuracy and the efficiency can be improved greatly with the proposed method, particularly when compared to commonly developed pixelwise classification methods.

The remainder of this paper is structured as follows. First, Section II briefly reviews the AH model for set representation. Second, Section III introduces the proposed set-to-set distance-based spectral–spatial HSI classification method. Then, experimental results and analysis are reported in Section IV. Finally, conclusions and future work are discussed in Section V.

## II. AH MODEL

For a set of points in the 2-D plane, the AH is a finite subset of the plane, which should be the smallest convex polygon

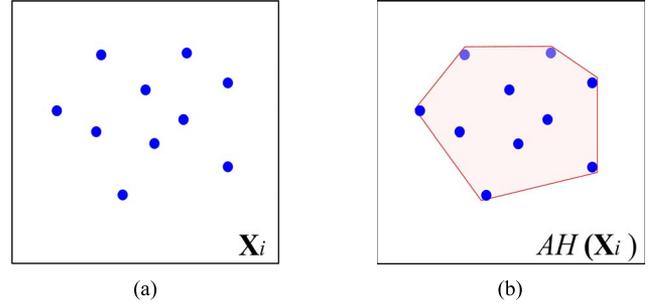


Fig. 1. AH representation for a set of points/feature points. (a) Set of points/feature points, denoted by  $\mathbf{X}_i$ . (b) AH of the set  $\mathbf{X}_i$ , denoted by  $AH(\mathbf{X}_i)$ .

containing all points [37]. As illustrated in Fig. 1, the AH for representing the set of points in Fig. 1(a) is marked with a convex polygon in Fig. 1(b). The AH possesses attractive properties that make it suitable for data/image representation and analysis [38]: 1) uniqueness, which means that the convex region corresponding to a set should be specific; 2) local controllability, which means that, when some feature points are either added to or removed from the original set, the convex hull is locally affected; 3) high computational efficiency, i.e., the upper bound of the computational complexity for finding the AH for one data set is  $O(N \log_2 N)$ . When each point within each set corresponds to one 1-D vector or 2-D matrix, the AH-model-based representation should be explained in feature subspace. Let  $\mathbf{X}_i$  represent a set consisting of multiple samples, the AH model regards each sample within the set as one feature point in a linear or affine feature space and characterizes each set with a convex geometric region spanned by its feature points [37]–[39]. Mathematically, the AH model can be represented as

$$AH(\mathbf{X}_i) = \left\{ \mathbf{x}_i = \sum \alpha_{ik} \mathbf{x}_{ik} \mid \sum \alpha_{ik} = 1 \right\} \\ \text{s.t. } \alpha_{ik} \in (0, 1) \quad (1)$$

where  $\mathbf{x}_{ik}$  is the  $k$ th sample in data set  $\mathbf{X}_i$ , and  $\alpha_{ik}$  is the corresponding coefficient. The AH of samples  $\mathbf{x}_i$  contains all points with the form of  $\sum \alpha_{ik} \mathbf{x}_{ik}$ , where  $\sum \alpha_{ik} = 1$ . By using any point  $\boldsymbol{\mu}_i$  (e.g., one of the samples in  $\mathbf{X}_i$  or the mean of the samples in  $\mathbf{X}_i$ ) on the AH as a reference to represent each set, the AH can also be rewritten with another parametric form [39], [40], which is denoted by

$$AH(\mathbf{X}_i) = \{ \mathbf{x}_i = \boldsymbol{\mu}_i + \mathbf{U}_i \mathbf{v}_i \}. \quad (2)$$

Here,  $\mathbf{U}_i$  is an orthonormal basis for the directions spanned by the affine subspace, which can be generated via both  $l_1$ - and  $l_2$ -norm-based fitting procedures [41], [42].  $\mathbf{v}_i$  is a vector of free parameters that provides reduced coordinates for the points within the subspace, which is expressed with respect to  $\mathbf{U}_i$ . With this model, any affine combinations of samples in the set are considered even when the combinations do not appear in the existing samples of the set.

## III. PROPOSED METHOD

As illustrated in Fig. 2, there are two main steps in the proposed set-to-set distance-based spectral–spatial classification method (SD-SSC): (i) the generation of training and test sets

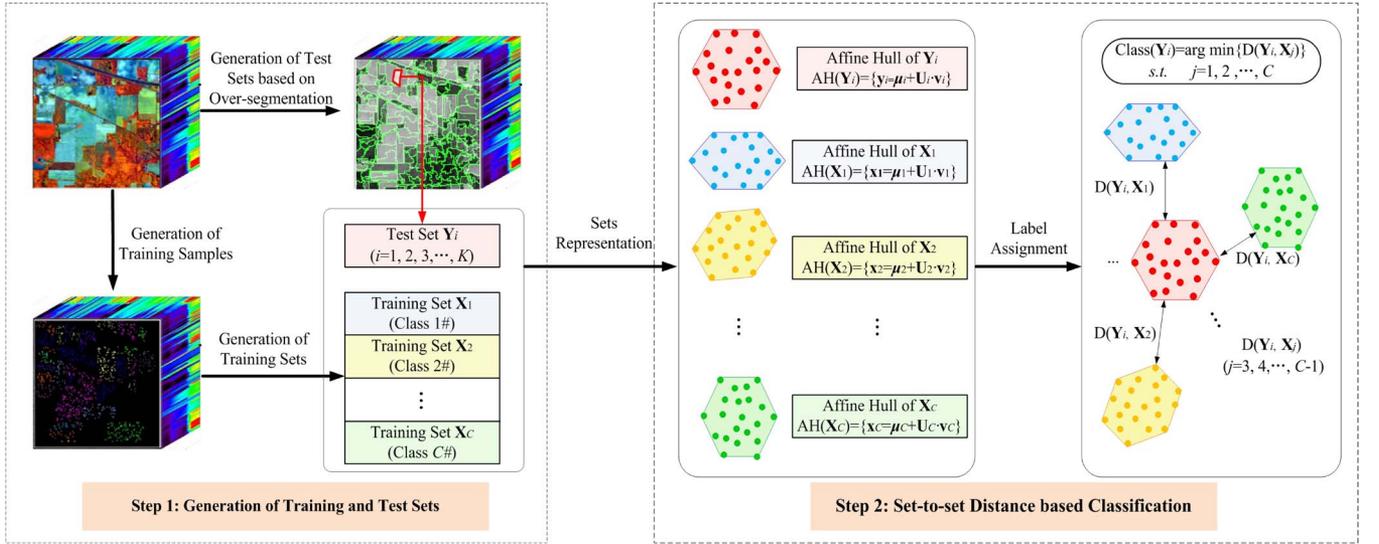


Fig. 2. Illustration of the proposed SD-SSC method for HSI classification.

and (ii) set-to-set distance-based classification. Specifically, each training set can be obtained directly by grouping each class of labeled training samples. For generating the test sets, the HSI is first divided into multiple homogeneous regions without overlaps by the oversegmentation method. Here, the superpixel technique [34] is used, in order to effectively exploit both spatial context and structural information for the generation of oversegmentation result. Then, each superpixel is considered as one test set, which contains spatially connected and spectrally similar pixels that are assumed to have the same label. Step (ii) can be divided into two subtasks: representation of sets and label assignment. Considering both similarity and variance existing in spectral pixels within each set, the AH model is used to adaptively represent the characteristics of each set. Then, the geometric distance between AHs corresponding to both training and test sets is compared, and the minimum distance determines the class of the test set. The final classification result for HSI is by repeating step (ii) for  $K$  times, where  $K$  represents the number of test sets to be classified.

### A. Generation of Training and Test Sets

The illustration of generating both training and test sets is shown in Fig. 3. Specifically, each class of labeled training samples is chosen to be one training set, and all the training sets can be denoted by  $\{Y_1, Y_2, \dots, Y_C\}$ . In addition, each test set  $X_i$  can be generated via a superpixel-based oversegmentation. Generally, superpixels are defined as perceptually uniform regions in 2-D images [34]–[36], which can more adaptively and sufficiently explore the spatial information for different types of spatial structures than the fixed rectangular image patches. In the field of HSI, each superpixel contains multiple spatially connected and spectrally similar pixels, which are assumed to have the same label. Then, each superpixel in HSI is regarded as one test set. The details of how to generate superpixels from HSI are given as follows.

1) *Estimation of Superpixel Number*: Generally, the number of superpixels  $K$  was manually selected according to observations and experience [34], [35]. Here, an automatic estima-

tion strategy for the value of  $K$  based on edge detection is introduced. This is reasonable due to the general fact that the more complex the image structure, the larger the number of superpixel should be set. The introduced automatic parameter estimation strategy contains the following four steps.

- Step 1: Generation of fused image  $Z_f$ . Following our previous work [44], a fused image ( $Z_f$ ) is generated by performing a weighted average operator on all the spectral bands, i.e.,  $Z_f = \sum w_i Z_i / s_w$ , where  $s_w$  is the sum of all the weights. For the  $i$ th spectral band  $Z_i$ , the weight  $w_i$  is defined as  $w_i = e^{-(\sigma_i^2/h^2)}$ , where  $\sigma_i$  is the noise deviation estimated by the median of absolute deviation, and  $h$  is a scaling factor.
- Step 2: Generation of binary edge map  $E_m$ . For the fused image  $Z_f$ , the canny detector is used to obtain the 2-D binary edge map ( $E_m$ ).
- Step 3: Detection of continuous edges CE. The 8-connected component labeling algorithm [45] using a two-pass approach is applied to detect continuous edges CE in  $E_m$ . For the first pass, the preliminary label of each point  $p_{x,y}$  in  $E_m$  is assigned according to the labels of points in the 8-connected neighborhood. The points with the same preliminary labels are grouped into the same label set, and the label equivalences between label sets are recorded. For the second pass, all the label sets with a label equivalence relationship are merged and assigned one updated label, resulting in the final labels  $\hat{L}_p(p_{x,y})$  for each point  $p_{x,y}$  in  $E_m$ . Then, continuous edges CE =  $\{CE_q\}$  can be generated, where each continuous edge  $CE_q$  consists of points with the same label in  $\hat{L}_p$ , i.e.,  $CE_q = \{p_{x,y} | \hat{L}_p(p_{x,y}) = q\}$ . By this way, it can be guaranteed that any point in  $CE_q$  has at least one 8-connected point in the same continuous edge. Meanwhile, there are no 8-connected point pairs in any two continuous edges  $CE_{q1}$  and  $CE_{q2}$ , i.e.,  $q1 \neq q2$ .

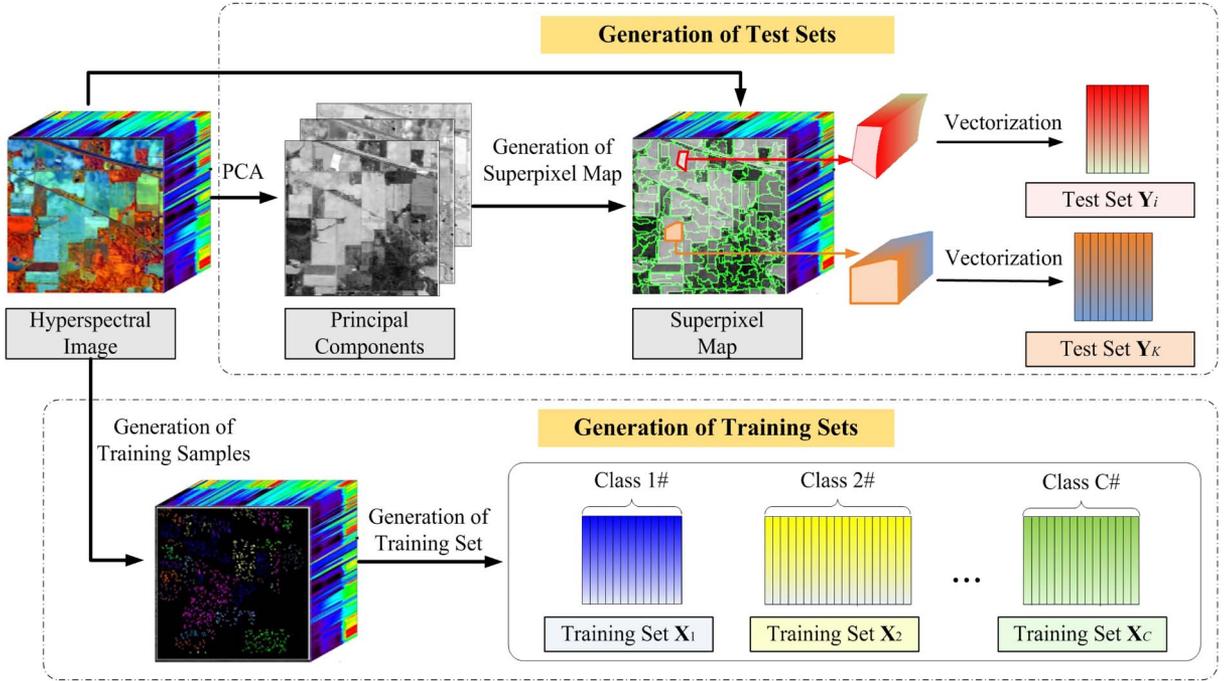


Fig. 3. Illustration of generating training and test sets in HSI.

Step 4: Estimation of parameter  $K$ . By calculating the number of detected continuous edges (CE), which is denoted by  $N_{CE}$ , the estimated value of  $K$  is set to  $N_{CE}$ .

The influence of  $K$  on classification performance by the proposed method is discussed in the experimental section, where the effectiveness of the introduced automatic parameter selection for the value of  $K$  is also demonstrated.

2) *Generation of Superpixels in HSI*: As one classic oversegmentation method, the entropy rate superpixel (ERS) [34] is adopted in the proposed method to effectively generate a superpixel map. First, the principal component analysis (PCA) [43] is used to decompose the HSI. Since the first few principal components (e.g., the first three principal components) always contain the most important information in the HSI, they are utilized as the base image for the oversegmentation. Then, the ERS divides the spatial region of the base image into multiple nonoverlapping homogeneous regions via solving a graph optimization problem. With the given number of superpixels (denoted by  $K$ ), the ERS maps an image to a graph  $G = (V, E)$ , where  $V$  denotes the pixels in image, and  $E$  is the edge weights representing the pairwise similarities between pixels. By optimizing the objective function in (3), a subset of edges  $A \in E$  can be generated, resulting in  $K$  connected subgraphs. Each subgraph corresponds to a homogeneous region, containing spatially connected and spectrally similar pixels. Thus

$$\begin{aligned} & \max_A \{H(A) + \lambda B(A)\} \\ & \text{s.t. } A \in E \end{aligned} \quad (3)$$

where  $H$  is the entropy rate of the random walk on graph  $G$ , and  $B$  is a balancing term encouraging clusters with similar sizes, reducing the number of unbalanced superpixels.

### B. Set-to-Set Distance-Based Classification

After all the training and test sets have been generated, the set-to-set distance is measured to determine the final classification result. There are two main steps: AH-model-based representation of sets and set-to-set distance-based label assignment. Considering that the number of samples within some sets may be large, a smaller subset is selected to replace the original large-size sets before the set representation, in order to improve the computational efficiency. Since high redundancy exists in each processing set, the use of subsets of smaller size still can provide sufficient spectral information for effective classification. Here,  $N$  point pairs  $\{\mathbf{x}_i, \mathbf{y}_i\}$  ( $i = 1, 2, \dots, N$ ) with high spectral similarity are selected to generate the subsets from both training and test sets, where the spectral angle distance (SAD) is used to measure the spectral similarity between spectral pixels, i.e.,

$$\text{SAD}(\mathbf{x}_i, \mathbf{y}_i) = \cos^{-1} \left( \frac{\mathbf{x}_i^T \cdot \mathbf{y}_i}{\|\mathbf{x}_i\| \cdot \|\mathbf{y}_i\|} \right) \quad (4)$$

where  $\mathbf{x}_i$  and  $\mathbf{y}_i$  represent pixels in each training set and test set, respectively. The introduced subset selection is easy to operate and time efficient. In the experimental section, it can be demonstrated that this process is effective in accelerating the distance calculation without decreasing the classification accuracy. After the subset selection, the process of set-to-set distance-based classification is illustrated in detail as follows.

1) *AH-Model-Based Representation of Sets*: In HSIs, the pixels in each spectral set not only share similar spectral characteristics but also provide a varying appearance because of the existence of noise and pixel mixture. In HSIs, the pixels within each training/test set have both similarity and variance. On the one hand, since the pixels in each set belong to the same

class, they have a common reflectance corresponding to the same material. Therefore, they should have similar spatial and spectral characteristics, which can be used to distinguish between different categories of objects. On the other hand, due to the spectral mixture in HSI, some pixels that belong to the same class may present different spectral and spatial appearances and vice versa. Therefore, the variance of pixels in one set should also be considered for a better classification result. Based on the aforementioned analysis, the AH model that can exploit information about both the similarity and the variability of the samples within each set is used to adaptively characterize each set of pixels. Here, the singular value decomposition (SVD), as a standard  $l_2$ -norm-based estimator [46], is used for the estimation of the orthonormal basis ( $\mathbf{U}_i$ ) in (2). The SVD-based decomposition operated on  $\overline{\mathbf{X}}_i = [\mathbf{x}_{i,1} - \boldsymbol{\mu}_i, \mathbf{x}_{i,2} - \boldsymbol{\mu}_i, \dots, \mathbf{x}_{i,p} - \boldsymbol{\mu}_i]$  can be represented as

$$\overline{\mathbf{X}}_i = \mathbf{U}_i \boldsymbol{\Sigma}_i \mathbf{V}_i^T. \quad (5)$$

Here,  $\mathbf{V}_i$  and  $\mathbf{U}_i$  are unitary matrices, which represent two different orthonormal bases, and  $\boldsymbol{\Sigma}_i$  describes how much the vectors in  $\mathbf{V}_i$  are stretched to give the vectors in  $\mathbf{U}_i$ . In addition to the attractive properties that make the AH model suitable for image representation and analysis that are discussed in Section II, more benefits introduced by the AH model for HSI classification are analyzed here. Specifically, when the number of samples in a training/test set is small, the limited spectral information that is provided by the original samples may lead to a poor classification result. With the AH model, more meaningful samples can be synthesized by the linear combinations of existing samples in the set. By this way, both the original samples and the synthesized samples can offer more complementary information to generate better classification results.

2) *Set-to-Set Distance-Based Label Assignment*: With all the training and test sets being adaptively represented by AHs, the similarity between sets can be measured by the geometric distance between AHs [39], [46]. Let  $\mathbf{X}_i$  and  $\mathbf{Y}_j$  represent each training set and test set, respectively, the distance between them, i.e.,  $D(\mathbf{X}_i, \mathbf{Y}_j)$ , is denoted by

$$\begin{aligned} D(\mathbf{X}_i, \mathbf{Y}_j) &= D(\text{AH}(\mathbf{X}_i), \text{AH}(\mathbf{Y}_j)) \\ &= \min_{\mathbf{v}_i, \mathbf{v}_j} \left\| (\mathbf{U}_i \mathbf{v}_i + \boldsymbol{\mu}_i) - (\mathbf{U}_j \mathbf{v}_j + \boldsymbol{\mu}_j) \right\|^2 \\ &= \min_{\mathbf{v}_i, \mathbf{v}_j} \left\| (\mathbf{U}_i \mathbf{v}_i - \mathbf{U}_j \mathbf{v}_j) - (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i) \right\|^2 \\ &= \min_{\mathbf{v}_i, \mathbf{v}_j} \left\| [\mathbf{U}_i - \mathbf{U}_j] [\mathbf{v}_i \ \mathbf{v}_j]^T - (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i) \right\|^2. \quad (6) \end{aligned}$$

Let  $\mathbf{U}_{ij} = [\mathbf{U}_i \ -\mathbf{U}_j]$  and  $\mathbf{v}_{ij} = [\mathbf{v}_i \ \mathbf{v}_j]^T$ , then (6) can be transformed to

$$\begin{aligned} D(\mathbf{X}_i, \mathbf{Y}_j) &= D(\text{AH}(\mathbf{X}_i), \text{AH}(\mathbf{Y}_j)) \\ &= \min_{\mathbf{v}_{ij}} \left\| \mathbf{U}_{ij} \mathbf{v}_{ij} - (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i) \right\|^2 \quad (7) \end{aligned}$$

which is a standard least square problem and has a closed-form solution, which is represented as

$$\mathbf{v}_{ij}^{\text{opt}} = (\mathbf{U}_{ij}^T \mathbf{U}_{ij})^{-1} \mathbf{U}_{ij}^T (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i). \quad (8)$$

The distance between  $\mathbf{X}_i$  and  $\mathbf{Y}_j$ , which is denoted by  $D(\mathbf{X}_i, \mathbf{Y}_j)$ , is estimated by applying the optimal estimation of  $\mathbf{v}_{ij}^{\text{opt}}$  in (8). Then, the label for the spectral pixels in test set  $\mathbf{Y}_j$  will be assigned according to

$$\begin{aligned} \text{Class}(\mathbf{Y}_j) &= \arg \min_{i=1}^C \{D(\mathbf{X}_i, \mathbf{Y}_j)\} \\ \text{s.t. } &i = 1, 2, \dots, C. \quad (9) \end{aligned}$$

## IV. EXPERIMENTAL RESULTS

### A. Experimental Setup

Here, three hyperspectral data sets,<sup>1</sup> including the ‘‘AVIRIS Indian Pines,’’ ‘‘AVIRIS Salinas,’’ and ‘‘ROSIS-03 University of Pavia image,’’ are used to test the performance of the proposed classification method. The Indian Pines image, which covers the agricultural Indian Pines test site in Northwest Indiana, was collected by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. The data set of ‘‘AVIRIS Indian Pines’’ 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  $\mu\text{m}$ . Before the classification, 20 spectral bands (i.e., 104th–108th, 150th–163rd, and 220th) are discarded because of the water absorption. This image contains 16 classes, e.g., corns, soybeans, and wheat. The data set of ‘‘Salinas’’ 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 it has a spatial resolution of 3.7 m per pixel. Before the classification, 20 water absorption spectral bands are removed. The data set of ‘‘University of Pavia’’ was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS-03) sensor over the campus at the University of Pavia, Italy. This data set contains 103 spectral bands after the seriously noise-corrupted bands being discarded, and each band is of size  $610 \times 340$ . The spatial resolution of this data set is 1.3 m, and the spectral coverage ranges from 0.43 to 0.86  $\mu\text{m}$ . The reference classes and the number of training and test samples are shown in Table I.

The classification results of the proposed SD-SSC method are visually and quantitatively compared with some widely used classification methods, i.e., SVM [1], SVM-CK [21], extended morphological profile (EMP) [16], joint-sparse-representation-based classification (JSRC) [24], logistic regression via splitting and augmented Lagrangian-multilevel logistic (L-MLL) [28], and maximum *a posteriori* marginal loopy belief propagation (MPM-LBP) [47]. The SVM-CK, L-MLL, JSRC, and MPM-LBP methods take advantage of the spatial information for HSI classification, whereas the SVM is a pixelwise classification method without considering the spatial information. To objectively evaluate the classification results, metrics of the overall accuracy (OA), average accuracy (AA), and the Kappa coefficient ( $k$ ) are used.

<sup>1</sup>[http://www.ehu.es/ccwintco/index.php/Hyperspectral\\_Remote\\_Sensing\\_Scenes](http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes)

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

Indian Pines				Salinas				University of Pavia			
Class	Name	Train	Test	Class	Name	Train	Test	Class	Name	Train	Test
1	Alfalfa	6	40	1	Weeds_1	20	1989	1	Asphalt	548	6083
2	Corn-no till	144	1285	2	Weed_2	37	3689	2	Meadows	540	18109
3	Corn-min till	84	746	3	Fallow	20	1956	3	Gravel	392	1707
4	Corn	24	213	4	Fallow plow	14	1380	4	Trees	524	2540
5	Grass/pasture-mowed	50	433	5	Fallow smooth	27	2651	5	Metal sheets	265	1080
6	Grass/trees	75	655	6	Tubble	40	3919	6	Bare soil	532	4497
7	Grass/pasture	3	25	7	Celery	36	3543	7	Bitumen	375	955
8	Hay-windrowed	49	429	8	Grapes	113	11158	8	Bricks	514	3168
9	Oats	2	18	9	Soil	62	6141	9	Shadows	231	716
10	Soybean-no till	97	875	10	Corn	33	3245				
11	Soybean-min till	247	2208	11	Lettuce 4wk	11	1057				
12	Soybean-clean till	62	531	12	Lettuce 5wk	19	1908				
13	Wheat	22	183	13	Lettuce 6wk	9	907				
14	Woods	130	1135	14	Lettuce 7wk	11	1059				
15	Bldg-Grass-Tree-Drives	38	348	15	Vinyard untrained	73	7195				
16	Stone-steel towers	10	83	16	Vinyard trellis	18	1789				
Total		1043	9206	Total		543	53586	Total		3921	38855

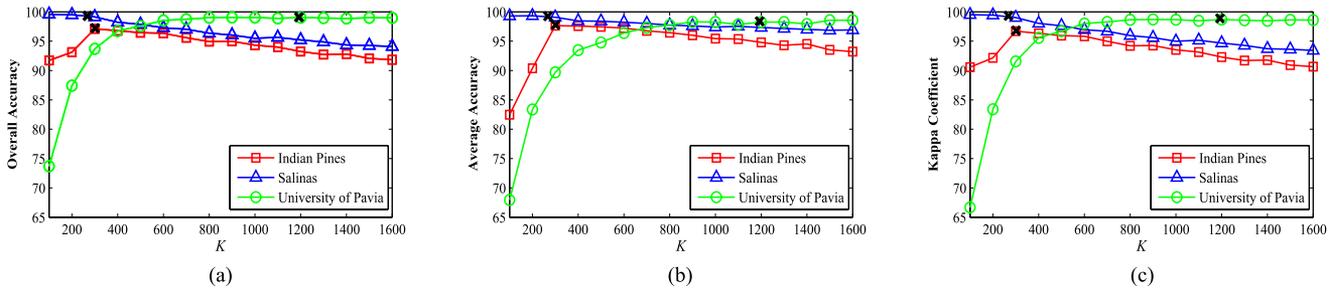


Fig. 4. Effect of the superpixel number  $K$  on classification accuracy, when applying the proposed SD-SSC algorithm to the three test images. The classification performance is measured based on three objective metrics, including (a) OA, (b) AA, and (c) Kappa coefficient. The black marks indicate the estimated values of  $K$  by the proposed automatic estimation strategy, i.e.,  $K = 301$ ,  $K = 268$ , and  $K = 1193$  for the Indian Pines, Salinas, and University of Pavia images, respectively.

## B. Parameter Setting

For the SVM method, the parameters  $C$  and  $\sigma$  are by fivefold cross-validation, which is also the default setting in the SVM-CK method. In addition, the standard deviation and the average value are regarded as the textural features, and the original spectral data are used as the spectral feature in the SVM-CK method. In order to composite the summation kernel, the weight of the textural content is 0.4, and the weight of the spectral content is 0.6, which are set empirically to generate the best classification accuracy. For the JSRC method, the window sizes for the Indian Pines image, the Salinas image, and the University of Pavia image are respectively tuned to reach their best results, in terms of the classification accuracy. The parameters for the EMP, L-MLL, and MPM-LBP methods are set to the default values as in [16], [28], and [47], respectively. In the proposed method, there are two important parameters, i.e., the number of superpixels  $K$  and the number of subsets  $N$ , which will be discussed as follows.

1) *Effect of the Superpixel Number*: The influence of parameter  $K$ , varying from 100 to 1600, on the classification performance, in terms of OA, AA, and  $\kappa$  coefficients, is shown in Fig. 4. It can be observed that the optimal selection of  $K$  for different data sets varies largely. For the University of Pavia, a larger value than 1000 can lead to relatively stable and high classification accuracy. This can be explained by the fact that

there is more detailed texture and structure in the University of Pavia data set, which corresponds to more superpixels of smaller size. For both the Indian Pines and Salinas data sets, large homogeneous regions can be observed. Therefore, a much lower value for  $K$  is preferred in these two data sets. Specifically, it can be found that the performance of the proposed SD-SSC method improves as the superpixel number  $K$  is set to values between 300 and 600, for the Indian Pines image, and between 100 and 300, for the Salinas image. Based on the aforementioned observation, parameter  $K$  can be estimated according to the structure complexity in the spatial domain. In the proposed method, the introduced automatic parameter estimation based on edge detection is used to determine the value of  $K$ , instead of using a manual setting. By this way, a data set with more structure and texture information can be separated into a larger number of superpixels, whereas a data set with higher homogeneity can correspond to fewer superpixels. For the remaining experiments, the value of parameter  $K$  for each specific test image is set as the automatic estimated value, i.e.,  $K = 301$  for the Indian Pines image,  $K = 268$  for the Salinas image, and  $K = 1193$  for the University of Pavia image. It is obvious from the black marks in Fig. 4 that the adaptively estimated value of  $K$  can result in the relatively high classification accuracy for each test image, particularly for the Indian Pines and the University of Pavia images. However, for

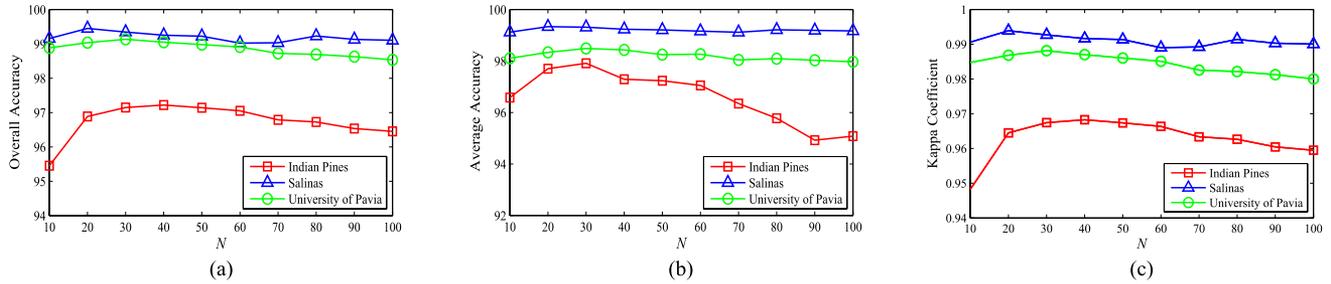


Fig. 5. Effect of the pixel number within each subset  $N$  on classification accuracy, when applying the proposed SD-SSC algorithm to the three test images. The classification performance is measured based on four objective metrics, including (a) OA, (b) AA, and (c) Kappa coefficient.

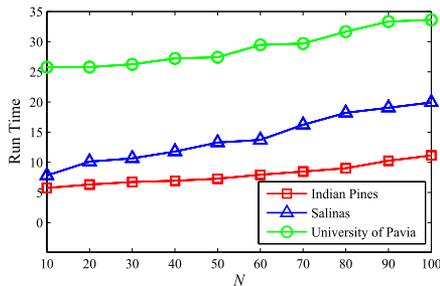


Fig. 6. Effect of the pixel number within each subset  $N$  on run time (in seconds) when applying the proposed SD-SSC algorithm to the three test images.

the Salinas image, a slight increase in classification accuracy can be found in Fig. 4, when  $K$  is manually set to the optimal values. Overall, the introduced adaptive parameter selection strategy offers an efficient way to estimate the value of  $K$  to generate favorable classification results, which can be better applied in practical applications than manual parameter setting for each specific test image.

2) *Effect of the Pixel Number in Subsets*: In the calculation of the set-based distance, subsets are chosen according to the SAD-based metric, in order to improve the efficiency of the proposed method. To examine the effect of  $N$  on the classification accuracy and efficiency, the parameter  $N$  is varied from 10 to 100, as shown in Figs. 5 and 6. Meanwhile, another parameter  $K$  for each specific test image is set as the estimated value (as marked in black color in Fig. 4), which is determined by the introduced automatic parameter estimation strategy for  $K$ . It can be found that the classification accuracy (e.g., OA, AA, and  $\kappa$ ) is negatively affected by very small values of  $N$ , i.e.,  $N \leq 20$ , particularly in the data set of Indian Pines. This is because, if  $N$  has a much smaller value, not enough spectral–spatial information in both test and training sets can be exploited, resulting in a poor classification performance. In addition, it is clear in Fig. 6 that the run time grows by increasing the value of  $N$ . Therefore, considering both classification accuracy and run time,  $N$  is set to 30 to generate favorable classification results in the following experiments.

In addition, to parametrically represent the AH, the reference vector defined in (2), i.e.,  $\mu_i$ , is calculated as the mean vector of a few samples within one set. Due to the noisy and mixed characteristics of HSI, spectral pixels that belong to the same

set may have variety in spectral appearance. Therefore, a poor estimation of the reference  $\mu_i$  may be achieved by averaging all the samples in one set. Here, instead of using all the samples to calculate the mean, only a certain portion of samples (i.e., 10%) in each set that have the closest SAD to the compared set are averaged to generate the reference vector  $\mu_i$ .

### C. Comparison Results of Classification Accuracy

For the image of “AVIRIS Indian Pines,” 10% of the labeled samples were randomly selected as the training samples, and the remaining 90% were used as test samples, as shown in Table I. The classification maps by various classification methods are shown in Fig. 7. It can be observed that the SVM, which only utilizes the spectral information, presents a very noisy estimation in the classification maps. By incorporating the local spatial context in HSI, the SVM-CK, EMP, L-MLL, and MPM-LBP can yield a smoother visual effect. However, these approaches still fail to identify the details and the near-edge regions. In contrast, the proposed SD-SSC approach not only generates a smoother appearance in homogeneous regions but also achieves a more accurate classification result. The quantitative results (averaged over ten runs) in terms of the three important objective metrics (OA, AA, and  $\kappa$ ) are presented in Table II. As one can observe, the proposed method performs better than the other compared methods, in terms of the three objective metrics. Although the proposed method cannot guarantee the best result in classifying every specific class, the better classification performance in classifying 11 out of 16 classes indicates the overall advantage of the proposed method over the state-of-the-art methods. In addition, advantage can be observed in classifying the categories with small number of training samples  $T_{\text{num}}$ . For example, the numbers of training samples for the first, seventh, and ninth classes in the Indian Pines image are 6, 3, and 2, respectively. Due to the lack of available training samples, the highest classification accuracy values for these three categories by the other compared methods are 94.72%, 89.77%, and 86.41%, respectively. In contrast, the proposed AH-model-based classifier can achieve very high classification accuracy values that are near 100%. This benefit is introduced since the AH-model-based representation can exploit the variance appearance of samples in each set to synthesize more meaningful samples. As a result, sufficient spectral information can be extended for better classification results.

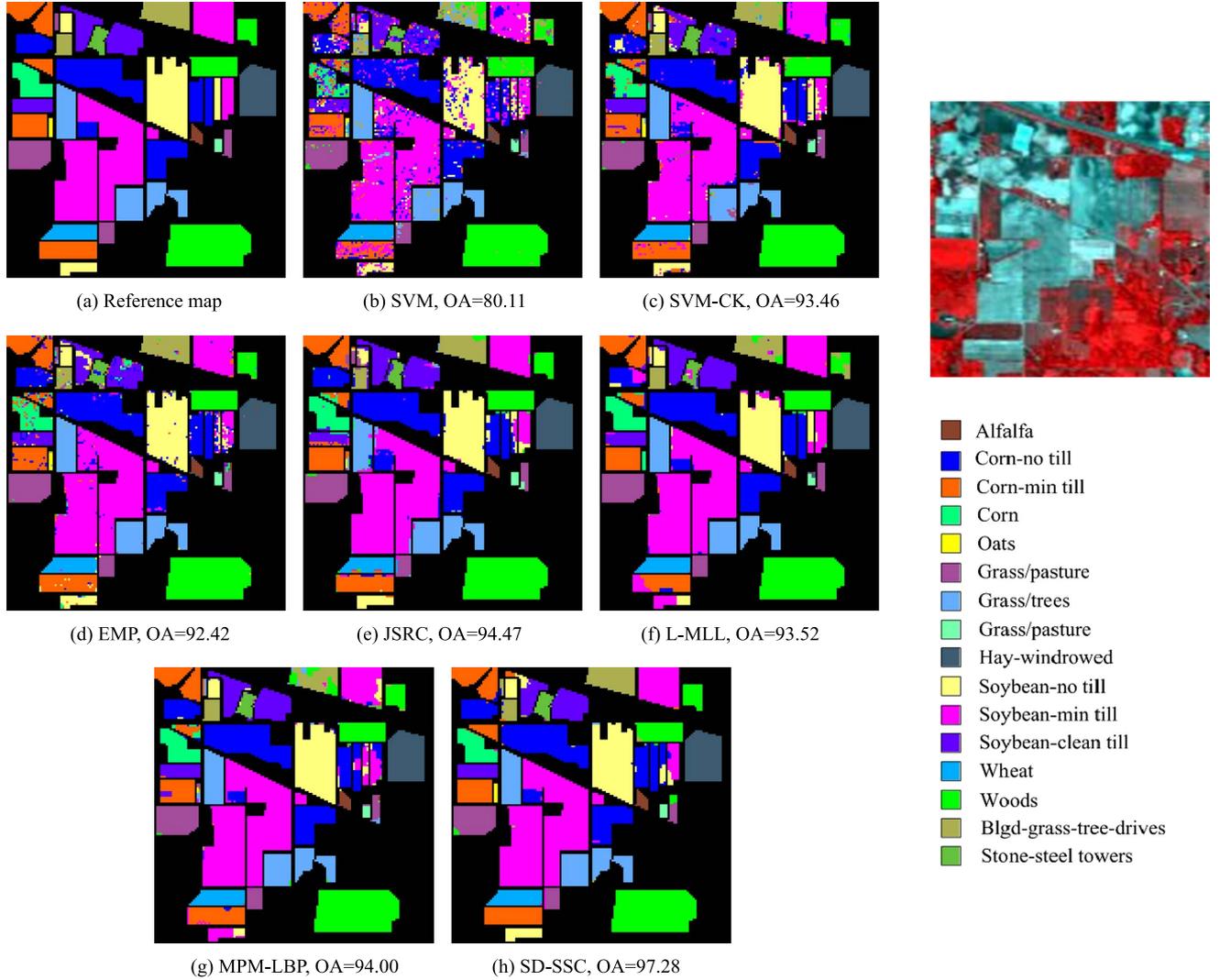


Fig. 7. Reference map and classification results for the Indian Pines image. (a) Reference map and (b)–(h) classification maps generated by different compared methods: SVM [1], SVM-CK [21], EMP [16], JSRC [24], L-MLL [28], MPM-LBP [47], and the proposed SD-SSC. The numerical results refer to the overall classification accuracy in percent.

TABLE II  
CLASSIFICATION ACCURACY (%) OF THE INDIAN PINES IMAGE  
OBTAINED BY SVM [1], SVM-CK [21], EMP [16], JSRC [24], L-MLL  
[28], MPM-LBP [47], AND THE PROPOSED SD-SSC METHODS

Class	SVM	SVM-CK	EMP	JSRC	L-MLL	MPM-LBP	SD-SSC
1	64.84	85.25	92.92	94.72	83.00	79.50	<b>100</b>
2	71.99	89.14	91.18	93.82	94.20	93.75	<b>96.33</b>
3	74.72	90.48	88.79	93.61	90.66	91.90	<b>96.32</b>
4	59.50	92.25	79.00	91.03	96.67	88.69	<b>100</b>
5	89.59	94.36	94.37	94.20	95.06	91.78	<b>99.12</b>
6	86.80	99.16	96.61	93.35	<b>99.88</b>	99.76	99.33
7	89.07	88.80	89.77	76.00	53.60	84.80	<b>99.62</b>
8	91.39	99.77	99.40	99.56	<b>100</b>	<b>100</b>	<b>100</b>
9	72.08	71.67	86.41	28.40	15.00	56.11	<b>100</b>
10	76.00	88.87	90.52	92.13	88.86	91.11	<b>93.86</b>
11	78.94	95.65	94.09	96.30	96.44	95.95	<b>98.56</b>
12	68.37	86.53	87.73	89.58	<b>97.36</b>	95.78	93.14
13	94.47	98.96	99.02	88.71	99.62	99.51	<b>100</b>
14	91.22	98.24	99.24	<b>99.44</b>	96.76	97.92	97.90
15	74.29	86.98	<b>97.39</b>	92.02	92.44	85.34	92.87
16	<b>98.95</b>	98.07	97.87	93.98	82.29	88.19	98.33
OA	79.90	93.42	93.47	94.54	94.83	94.60	<b>97.15</b>
AA	80.14	91.51	92.77	88.56	86.36	90.01	<b>97.84</b>
$\kappa$	0.770	0.925	0.925	0.937	0.941	0.938	<b>0.968</b>

The second and third experiments are conducted on the Salinas and University of Pavia images, respectively. Following some related papers [19], [47], [48], only 1% of the labeled reference data were randomly selected as the training samples and the remaining 99% of data as the test set for the comparison experiment on the Salinas image. In addition, 10% of total number of labeled samples were randomly selected from University of Pavia as the training samples and the rest as the test samples. The detailed number of samples for each class in both test images is shown in Table I. The visual classification maps by various classifiers on the Salinas and University of Pavia images are shown in Figs. 8 and 9, respectively. It is shown that the classification maps by the SD-SSC method look more similar to the reference map shown in Figs. 8(a) and 9(a) than those maps by the other classical and state-of-the-art methods. In addition, the quantitative results in terms of classification accuracy (averaged over ten runs) are shown in Tables III and IV. In general, it can be observed that the proposed method shows the best performance in terms of the OA-, AA-, and  $\kappa$ -based objective metrics, with OA = 99.35%, AA = 99.26%,  $\kappa$  = 0.993

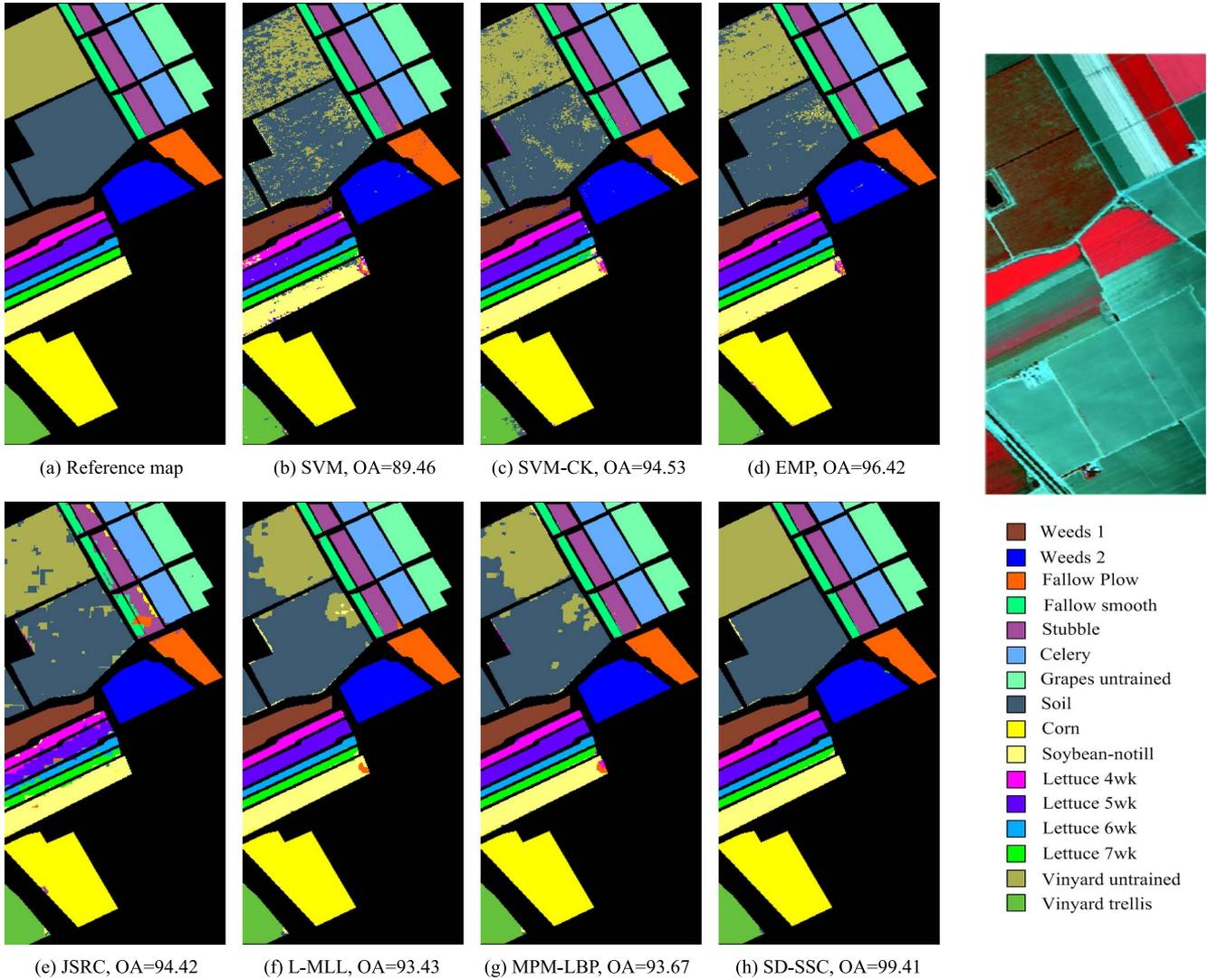


Fig. 8. Reference map and classification results for the Salinas image. (a) Reference map and (b)–(h) classification maps generated by different compared methods: SVM [1], SVM-CK [21], EMP [16], JSRC [24], L-MLL [28], MPM-LBP [47], and the proposed SD-SSC. The numerical results refer to the overall classification accuracy in percent.

for the Salinas image and  $OA = 99.17\%$ ,  $AA = 98.59\%$ ,  $\kappa = 0.989$  for the University of Pavia image. Although the SVM-CK, EMP, JSRC, L-MLL, and MPM-LBP methods can generate better classification results than the SVM method because they incorporate both spectral and spatial information, they are still inferior to the proposed SD-SSC method, in terms of classification accuracy. In addition, it should also be noticed that the proposed method gives the highest classification accuracy for most classes. The comparison results for these two experiments again demonstrate that the SD-SSC method works better than the other compared spectral–spatial methods, in terms of classification accuracy.

#### D. Comparison Results of Computational Time

In order to compare the efficiency of different classification methods for HSI, the running times of the proposed SD-SSC

method and other compared methods for classifying the three HSI images are reported in Table V. The execution time for the proposed method reported in Table V includes the time spent in the processes of set generation, AH-based set representation, and set-to-set distance-based classification. All the programs are executed on a computer with an Intel Core i7-4790 CPU at 3.60 GHz and 4 GB of RAM, and the software platform is MATLAB R2011a. In Table V, it is shown that the proposed algorithm requires much less running time, when compared with the SVM, SVM-CK, EMP, JSRC, L-MLL, and MPM-LBP methods. Particularly, it takes less than 30 s, whereas the compared methods take more than 100 s, for the University of Pavia image. This is due to the fact that the set-to-set distance calculation in (6) is based on solving a standard least square problem with a closed-form solution. Overall, the comparison of computational time shown in Table V demonstrates the high efficiency of the proposed setwise classification strategy.

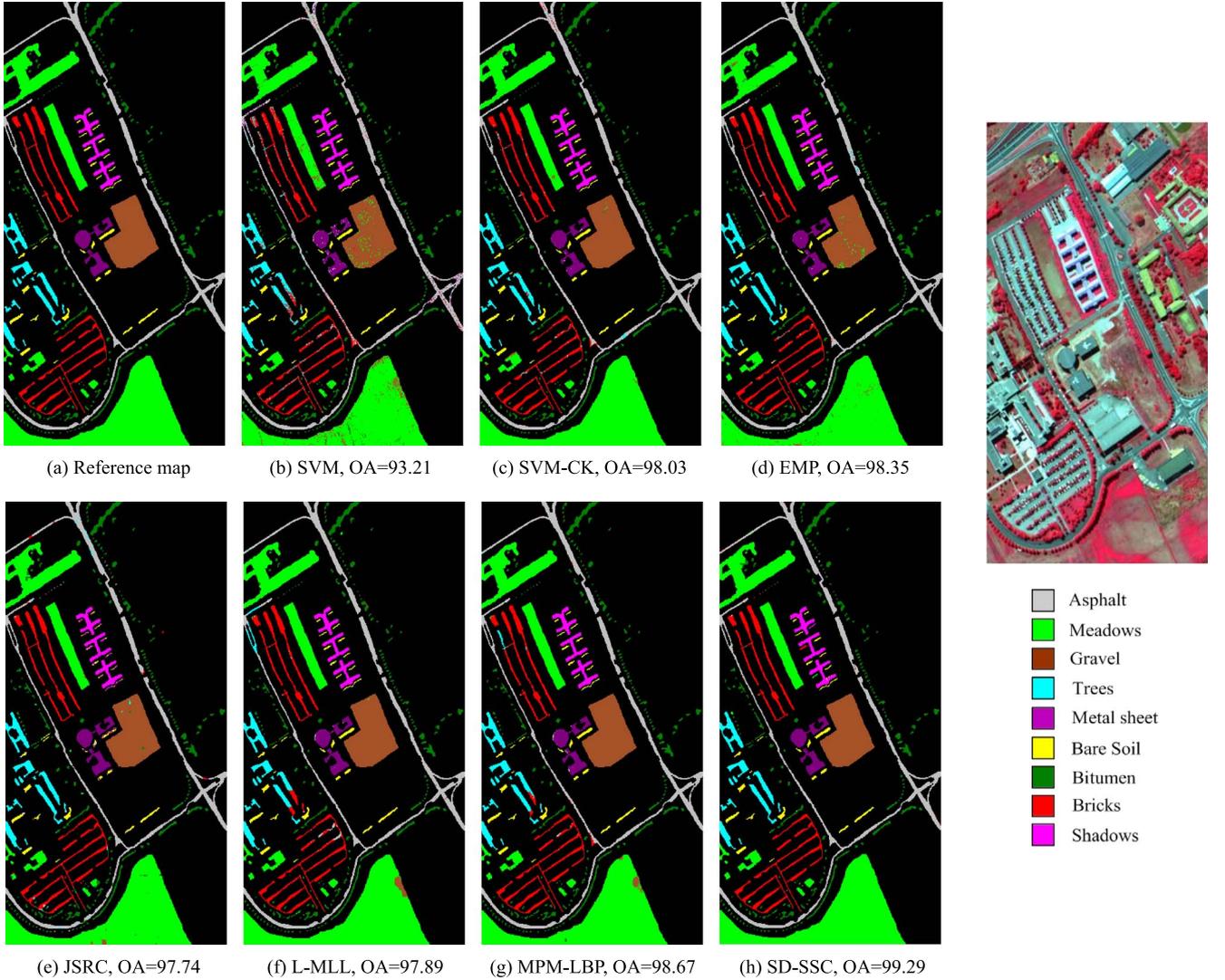


Fig. 9. Reference map and classification results for the University of Pavia image. (a) Reference map and (b)–(h) classification maps generated by different compared methods: SVM [1], SVM-CK [21], EMP [16], JSRC [24], L-MLL [28], MPM-LBP [47], and the proposed SD-SSC. The numerical results refer to the overall classification accuracy in percent.

TABLE III  
CLASSIFICATION ACCURACY (%) OF THE SALINAS IMAGE OBTAINED BY SVM [1], SVM-CK [21], EMP [16], JSRC [24], L-MLL [28], MPM-LBP [47], AND THE PROPOSED SD-SSC METHODS

Class	SVM	SVM-CK	EMP	JSRC	L-MLL	MPM-LBP	SD-SSC
1	99.58	99.06	99.95	94.72	99.36	99.56	<b>99.97</b>
2	98.90	99.53	98.96	93.82	99.91	99.91	<b>99.96</b>
3	91.65	95.35	91.97	93.61	98.38	98.76	<b>99.88</b>
4	98.37	<b>99.28</b>	98.08	91.03	98.95	97.74	98.93
5	97.95	96.99	97.69	94.20	98.68	98.88	<b>99.06</b>
6	<b>99.95</b>	99.86	99.92	93.35	99.82	99.79	99.84
7	98.59	99.51	99.72	76.00	99.55	99.71	<b>99.94</b>
8	75.80	87.53	94.19	<b>99.56</b>	89.66	93.82	<b>99.56</b>
9	98.88	99.11	99.20	28.40	99.78	<b>99.79</b>	99.14
10	88.66	94.77	95.50	92.13	93.76	94.32	<b>96.45</b>
11	92.42	95.53	94.98	96.30	92.79	96.43	<b>98.82</b>
12	96.44	99.40	97.35	89.58	<b>100</b>	<b>100</b>	99.28
13	92.67	97.06	96.36	88.71	98.37	98.73	<b>100</b>
14	96.13	90.54	98.84	<b>99.44</b>	95.08	94.95	97.68
15	74.69	85.64	90.70	92.02	71.76	71.42	<b>99.76</b>
16	97.96	93.06	99.70	93.98	97.27	98.04	<b>99.90</b>
OA	89.54	94.03	96.28	94.54	93.07	94.03	<b>99.35</b>
AA	93.66	95.76	97.07	88.56	95.82	96.37	<b>99.26</b>
$\kappa$	0.883	0.934	0.959	0.912	0.923	0.9334	<b>0.993</b>

TABLE IV  
CLASSIFICATION ACCURACY (%) OF THE PAVIA IMAGE OBTAINED BY SVM [1], SVM-CK [21], EMP [16], JSRC [24], L-MLL [28], MPM-LBP [47], AND THE PROPOSED SD-SSC METHODS

Class	SVM	SVM-CK	EMP	JSRC	L-MLL	MPM-LBP	SD-SSC
1	97.09	95.88	99.15	93.37	96.71	98.59	<b>99.77</b>
2	98.33	99.14	99.49	97.65	98.65	98.51	<b>99.72</b>
3	82.94	94.93	97.15	99.43	88.93	93.19	<b>98.76</b>
4	89.92	<b>99.54</b>	97.78	97.67	98.19	98.97	95.70
5	99.11	<b>99.99</b>	99.84	99.63	99.91	99.79	98.10
6	83.30	99.39	94.09	98.90	99.91	<b>99.98</b>	99.00
7	74.75	98.68	96.39	<b>99.83</b>	98.85	98.40	99.34
8	86.70	96.04	98.75	<b>98.90</b>	95.96	97.95	98.77
9	99.96	<b>100</b>	<b>100</b>	88.84	99.86	99.89	98.13
OA	93.23	98.28	98.45	97.25	97.88	98.50	<b>99.17</b>
AA	90.23	98.18	98.07	97.14	97.44	98.36	<b>98.59</b>
$\kappa$	0.908	0.977	0.979	0.963	0.971	0.980	<b>0.989</b>

## V. CONCLUSION

In this paper, a novel set-based spectral–spatial HSI classification method has been proposed. First, each class of training

TABLE V  
COMPUTING TIME (IN SECONDS) FOR THE CLASSIFICATION OF THREE TEST IMAGES OBTAINED BY SVM [1], SVM-CK [21], EMP [16], JSRC [24], L-MLL [28], MPM-LBP [47], AND THE PROPOSED SD-SSC METHODS

Image	SVM	SVM-CK	EMP	JSRC	L-MLL	MPM-LBP	SD-SSC
Indian Pines	127.6	201.6	51.9	86.2	24.8	60.9	<b>6.7</b>
Salinas	36.5	90.4	14.4	1262.1	30.4	299.4	<b>10.7</b>
University of Pavia	309.1	699.7	150.8	1764.3	100.2	628.6	<b>26.2</b>

pixels is regarded as one test set; meanwhile, spatially connected and spectrally similar pixels within each homogeneous region are grouped as one test set. Then, the label for each test set is determined according to the set-to-set distance. In terms of classification accuracy, the proposed method presents a superior performance over the compared methods. This superiority is due to 1) the incorporation of the spatial context (including spatial similarity and structure information) and spectral characteristics in HSI for effective set generation, 2) the combination of spectral similarity and variance within each processing set for adaptive set representation, and 3) the use of more discriminant spectral information carried by each set instead of the individual pixel for effective set-to-set distance-based label assignment. In terms of computation efficiency, the computational burden can be decreased dramatically by classifying the HSI in a set-by-set manner with the proposed method, when compared to the pixelwise classification methods. In addition, the set-to-set distance calculation is based on solving a standard least square problem with a closed-form solution, which also contributes to the efficient computation. Since the proposed setwise approach works well for supervised HSI classification, a topic of future research is to investigate whether the set-based method can be integrated into a semisupervised and unsupervised classification framework.

#### 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, and Dr. J. Li, for providing the software for the L-MLL and MPM-LBP methods.

#### 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] 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.
- [10] 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.
- [11] 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.
- [12] Y. Tarabalka, J. A. Benediktsson, and J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitioning clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973–2987, Aug. 2009.
- [13] 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.
- [14] 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.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.
- [19] J. Li, X. Huang, P. Gamba, J. M. Bioucas-Dias, L. Zhang, J. A. Benediktsson, and A. Plaza, "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, Mar. 2015.
- [20] 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.
- [21] 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.
- [22] 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.
- [23] J. Peng, Y. Zhou, and C. Chen, "Region-kernel-based support vector machines for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 4810–4824, Sep. 2015.
- [24] 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.

- [25] U. Srinivas, Y. Chen, V. Monga, N. M. Nasrabadi, and T. D. Tran, "Exploiting sparsity in hyperspectral image classification via graphical models," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 505–509, May 2013.
- [26] 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.
- [27] W. Fu, S. Li, L. Fang, X. Kang, and J. A. Benediktsson, "Hyperspectral image classification via shape-adaptive joint sparse representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 556–567, Feb. 2016.
- [28] 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.
- [29] 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.
- [30] 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.
- [31] 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.
- [32] 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.
- [33] 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.
- [34] 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.
- [35] 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.
- [36] 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.
- [37] F. P. Preparata and M. I. Shamos, *Computational Geometry: An Introduction*. Berlin, Germany: Springer-Verlag, 1985.
- [38] Z. Yang and F. S. Cohen, "Image registration and object recognition using affine invariants and convex hulls," *IEEE Trans. Image Process.*, vol. 8, no. 7, pp. 934–946, Jul. 1999.
- [39] H. Cevikalp and B. Triggs, "Face recognition based on image sets," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2010, pp. 2567–2573.
- [40] Y. Hu, A. S. Mian, and R. Owens, "Sparse approximated nearest points for image set classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 121–128.
- [41] C. Ding, D. Zhou, X. He, and H. Zha, "R1-PCA: Rotational invariant  $l_1$ -norm principal component analysis for robust subspace factorization," in *Proc. Int. Conf. Mach. Learn.*, 2006, pp. 281–288.
- [42] Q. Ke and T. Kanade, "Robust  $l_1$  norm factorization in the presence of outliers and missing data by alternative convex programming," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 592–599.
- [43] I. Jolliffe, *Principal Component Analysis*. Hoboken, NJ, USA: Wiley, 2005.
- [44] T. Lu, S. Li, L. Fang, Y. Ma, and J. Benediktsson, "Spectral-spatial adaptive sparse representation for hyperspectral image denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 373–385, Jan. 2016.
- [45] R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, vol. 1. Reading, MA, USA: Addison-Wesley, 1992, pp. 28–48.
- [46] H. Cevikalp and B. Triggs, "Large margin classifiers based on convex class models," in *Proc. Int. Conf. Comput. Vis.*, 2009, pp. 101–108.
- [47] 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.
- [48] A. Soltani-Farani, H. Rabiee, and S. Hosseini, "Spatial-aware dictionary learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 527–541, Jan. 2015.



**Ting Lu** (S'16) received the B.Sc. degree in 2011 from Hunan University, Changsha, China, 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.



**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, The 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, The Hong Kong University of Science and Technology. 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 Second-Grade National Awards at the Science and Technology Progress of China in 2004 and 2006. In 2013, he was granted the National Science Fund for Distinguished Young Scholars in China.



**Leyuan Fang** (S'10–M'14) received 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 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.



**Lorenzo Bruzzone** (S'95–M'98–SM'03–F'10) received the Laurea (M.S.) degree in electronic engineering (*summa cum laude*) and the Ph.D. degree in telecommunications from the University of Genoa, Genova, Italy, in 1993 and 1998, respectively.

He is currently a Full Professor in telecommunications with the University of Trento, Trento, Italy, where he teaches remote sensing, radar, pattern recognition, and electrical communications. He is the Founder and the Director of the Remote Sensing Laboratory in the Department of Information Engineering and Computer Science, University of Trento. His current research interests are in the areas of remote sensing, radar and synthetic aperture radar, signal processing, and pattern recognition. He promotes and supervises research on these topics, within the frameworks of many national and international projects. Among the others, he is the Principal Investigator of the Radar for icy Moon exploration (RIME) instrument in the framework of the Jupiter ICy moons Explorer (JUICE) mission of the European Space Agency. He is the author (or coauthor) of 161 scientific publications in referred international journals (111 in IEEE journals), more than 220 papers in conference proceedings, and 17 book chapters. He is an editor/coeditor of 16 books/conference proceedings and one scientific book. His papers are highly cited, as proven from the total number of citations (more than 15 300) and the value of the h-index (63) (source: Google Scholar).

Dr. Bruzzone ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (Seattle, July 1998). Since that time, he has been a recipient of many international and national honors and awards. He was a Guest Coeditor of different special issues of international journals. He was also invited as a Keynote Speaker in 30 international conferences and workshops. He is the Cofounder of the IEEE International Workshop on the Analysis of Multitemporal Remote Sensing Images (MultiTemp) series and is currently a member of the Permanent Steering Committee of this series of workshops. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. Since 2009, he has been a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society. Since 2012, he has been a Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society. Since 2013, he has been the Founder/Editor-in-Chief of the IEEE GEOSCIENCE AND REMOTE SENSING MAGAZINE. He currently serves as an Associate Editor for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the *Canadian Journal of Remote Sensing*.



**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 and computer engineering from Purdue University, West Lafayette, IN, USA, in 1987 and 1990, respectively.

He is currently the Rector and a Professor in electrical and computer engineering with the University of Iceland. He is a Cofounder of the biomedical startup company Oxymap. His research interests include remote sensing, biomedical analysis of signals, pattern recognition, image processing, and signal processing, and he has published extensively in those 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 the Association of Chartered Engineers in Iceland (VFI), the Societas Scinetiarum Islandica, and Tau Beta Pi. He was an 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, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS since 2003, and IEEE ACCESS since 2013. He is on the International Editorial Board of the *International Journal of Image and Data Fusion*, and he was the Chairman of the Steering Committee of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (J-STARS) in 2007–2010. He received the Stevan J. Kristof Award from Purdue University in 1991 as an outstanding graduate student in remote sensing. He was the recipient of the Icelandic Research Council's Outstanding Young Researcher Award in 1997, he was granted the IEEE Third Millennium Medal in 2000, he was a corecipient of the University of Iceland's Technology Innovation Award in 2004, he received the yearly research award from the Engineering Research Institute of the University of Iceland in 2006, and he received the Outstanding Service Award from the IEEE Geoscience and Remote Sensing Society in 2007. He was the corecipient of the 2012 IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING Paper Award. He received the 2013 IEEE/VFI Electrical Engineer of the Year Award, and in 2013, he was a corecipient of the IEEE GRSS Highest Impact Paper Award.