ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/301791521

A Novel Synergetic Classification Approach for Hyperspectral and Panchromatic Images Based on Self-Learning

Article *in* IEEE Transactions on Geoscience and Remote Sensing · May 2016

DOI: 10.1109/TGRS.2016.2553047

CITATIONS
16

READS
141

4 authors, including:

Xiaochen Lu
Donghua University
31 PUBLICATIONS **254** CITATIONS
SEE PROFILE

Junping Zhang
Harbin Institute of Technology
149 PUBLICATIONS **1,374** CITATIONS
SEE PROFILE

Tong Li
Harbin Institute of Technology
22 PUBLICATIONS **418** CITATIONS
SEE PROFILE

All content following this page was uploaded by Xiaochen Lu on 17 November 2017.

The user has requested enhancement of the downloaded file.

A Novel Synergetic Classification Approach for Hyperspectral and Panchromatic Images Based on Self-Learning

Xiaochen Lu, *Student Member, IEEE*, Junping Zhang, *Member, IEEE*, Tong Li, and Ye Zhang, *Member, IEEE*

***Abstract*—In this paper, we propose a self-learning approach for remote sensing image classification. The main work of this paper aims at providing a new framework of semisupervised learning technique for multiple-source synergetic classification, thereby improving the classification accuracy under the condition of small samples. Considering the high spectral resolution of a hyperspectral (HS) image and the high spatial resolution of a panchromatic (PAN) image, the proposed approach combines image segmentation with an active learning algorithm and adopts a standard active learning method for a self-learning strategy, in which the learning algorithm automatically selects informative unlabeled samples by itself according to their collaborative spatial–spectral features and the predicted information of a spectral-based classifier. This way, no extra cost of human expertise is required for labeling the selected pixels when compared with conventional active learning methods. Experiments on three data sets, including HS and PAN images, indicate that our proposed approach has a great enhancement on overall classification accuracy compared with classical supervised algorithms and turns out to be a promising strategy in synergetic classification of HS and PAN images.**

***Index Terms*—Active learning, hyperspectral (HS), self-learning, semisupervised, synergic classification.**

I. Introduction

OVER the past decades, land-cover classification has become a main subject in hyperspectral (HS) remote sensing applications. HS imaging enables a detailed separation of similar ground materials due to its high spectral resolution. However, the characteristics of large amount of data, interband spectral redundancy, and mixed pixels present challenging problems in analyzing HS data. During the development of HS image classification, most studies were conducted through supervised approaches, which require the availability of labeled samples for training the model of classification. Nevertheless, in many real-world problems, the available training samples are not sufficient in quantity and are not adequate in quality for properly training the classifier; meanwhile, the collection of labeled samples is generally difficult, expensive, and time-consuming [1]. A possible approach to address this problem is to exploit unlabeled samples in the learning process of the classification algorithm. Thus, to enrich the information given in input to the supervised learning algorithm and to improve the classification accuracy, semisupervised learning (SSL) techniques have been adopted for jointly leveraging the information of both labeled and unlabeled samples in the training of the classifier [2].

Semisupervised [3]–[6] algorithms incorporate the unlabeled data into the classifier training phase to find more precise decision rules. Many semisupervised classification techniques have been proposed in the literature so far, which can be grouped into several main categories: 1) self-training [7]; 2) cotraining; 3) generative probabilistic models; 4) semisupervised support vector machine (SVM) [8]; and 5) graph-based SSL. Self-training is one of the earliest ideas about using unlabeled data in the classification. The main idea is to repeatedly add the most confident unlabeled samples, together with their predicted labels, into the training set in order to avoid introducing possibly wrong semilabeled samples. However, the most confident samples sometimes can hardly influence the decision boundary, which leads to a high number of iterations. To reduce the number of iterations to reach convergence, the diversity criterion should be considered to minimize the redundancy among the selected semilabeled samples [2]. Other SSL algorithms have been also successfully applied to remote sensing image classification.

The other approach to both enrich the information given as input to the supervised classifier and improve the statistics of the classes is known in the machine learning community as active learning (AL). In AL, the learning process repeatedly queries unlabeled samples whose classification outcomes are the most uncertain [9]. After accurate labeling by the supervisor, who attributes the labels to the selected unlabeled samples, pixels are included into the training set to reinforce the model. This way, unnecessary and redundant samples are not included in the training set, thus greatly reducing both the labeling and computational costs [10]. The main challenge in AL is how to evaluate the information content of the unlabeled pixels. Generally, the selecting criteria can be grouped into three different families [11]. First, the committee-based heuristics, such as query-by-bagging, entropy query-by-bagging (EQB) [12], and

Manuscript received February 4, 2015; revised June 12, 2015, December 7, 2015, and March 3, 2016; accepted April 8, 2016. Date of publication May 2, 2016; date of current version June 1, 2016. This work was supported by the National Natural Science Foundation of China under Grants 61271348 and 61471148.

The authors are with the School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China (e-mail: lxchen09@163.com; zhangjp@hit.edu.cn; ltong@hit.edu.cn; zhye@hit.edu.cn).

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.2016.2553047

0196-2892 © 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

normalized EQB (nEQB) [13], search for the pixels that show maximal disagreement between the different classification models. Di and Crawford [14] have presented a multiview AL method based on adaptive weight disagreement measure. In [15], the random forest algorithm is adopted to establish the committee followed by region-based query techniques. Second, the large margin-based heuristics such as margin sampling (MS) [16], [17], are more suitable for SVM classifications since they take advantage of the geometrical features of SVMs. Ferecatu and Boujemaa proposed to add a constraint of orthogonality to the MS, resulting in maximal distance between the chosen examples [18]. The multiclass-level uncertainty (MCLU) with enhanced clustering-based diversity combines MCLU with a diversity criterion based on kernel k-means clustering [19], [20]. The third class, for example, maximizing Kullback–Leibler divergence [21] and breaking ties (BTs) [22], uses the estimation of posterior probabilities of class membership to evaluate the uncertainty of candidates. Li *et al.* presented the modified BT (MBT) technique to promote diversity in the selection process and guarantee unbiased samplings among the classes [23].

Recently, another family of AL heuristics, the cluster-based heuristics, has been proposed in [24]. Persello *et al.* [25] proposed a cost-sensitive AL method, which formalizes the problem of sample selection as a Markov decision process, to minimize the annotation cost. Pasolli *et al.* [26] took advantage of spatial information in conjunction with the spectral features to solve problems in different remote sensing contexts, leading to good improvement of the learning process on high-resolution remote sensing images.

From the overview of the literature, it can be seen that AL requires the intervention of the user/supervisor for labeling the selected samples. In addition, most of the previous studies focused on the selection of informative unlabeled samples at each iteration by assessing their uncertainty and then human experts would be asked for providing their class labels. This is sometimes difficult and time-consuming for supervisors when there is a dense distribution of ground objects, particularly in high-resolution remote sensing images. To avoid the effort required for labeling the selected samples, in [27], standard AL methods are adopted into a two-step learning scenario: inferring a candidate set using a semisupervised self-learning strategy based on a spatial neighborhood criterion and selecting informative samples from the obtained candidate pool. The main advantage of this framework is to use machine–machine interaction instead of human supervision.

However, the semisupervised self-learning strategy based on neighborhood criterion (neighbor-based self-learning (NBSL) hereafter)has some limitations in constructing the candidate set since the neighbor pixels are much fewer than the unlabeled samples, which are generally used in AL. Although the number of candidates increases with the iteration progress, the improvement of learning effect becomes slighter after several iterations. Another deficiency of this strategy is that the neighbor samples usually have high spectral similarity, which leads to a lack of diversity in candidate sets.

In this paper, a novel self-learning approach based on image segmentation and SSL is proposed for synergetic classification of HS and panchromatic (PAN) images to overcome the limitation in constructing the candidate set. Compared with HS images, PAN images have relatively higher spatial resolution with more details of ground objects. Hence, combining spatial features of PAN images is believed to offer strengthened capabilities for land-cover classification. For the synergetic classification task of HS and PAN images, the work is mostly focused on three basic techniques: 1) classification based on pixel-level fusion; 2) classification based on multiple feature extraction and combination; and 3) classification fusion based on decision level, which is also called multiclassifier fusion.

The proposed learning algorithm is comprised of two steps. In the first step, the learning algorithm itself predicts the candidates by utilizing the output of a spectral-based classifier and the collaborative spatial–spectral information of labeled samples through image segmentation technique. The second step is a standard AL algorithm, i.e., the learning algorithm automatically selects the samples with the highest uncertainty from the candidate set. This way, the learning machine can select the most useful samples as new training samples automatically and get their class labels by itself at the same time. As a result, no extra cost is required for labeling the selected samples, which realizes the goal of "self-learning" [27]. In particular, the main contributions of this paper are as follows: 1) proposing a new framework of synergetic processing for HS and PAN images rather than the conventional fusion techniques; 2) adapting a standard active learning method for a self-learning strategy, in which the step of human interaction of labeling selected samples is omitted; and 3) enhancing the abundance and diversity of samples by considering more candidates, thereby surpassing the corresponding supervised classification and collaborative spatial–spectral classification approaches. The performance of the algorithm has been proved by theoretical analyses and massive experiments.

The remainder of this paper is organized as follows. Section II elaborates the proposed approach for synergetic classification. We focus on the construction of candidate set based on the self-learning strategy. Section III reports classification results using three HS data sets. The detailed analysis is also presented here. Finally, conclusions are given in Section IV.

II. Methodology of Proposed Self-Learning

Here, we elaborate the proposed synergetic classification approach. The proposed approach consists of two prevalent techniques, including image segmentation and active learning. Fig. 1 shows the main framework of the algorithm. First, image segmentation is conducted on the PAN image to obtain the objects where labeled samples locate. Then, a spectral-based classifier is employed to conduct classification for HS images. By applying the self-learning strategy, the candidate set will be constructed. In addition, some unlabeled samples will be selected according to standard AL methods. The procedure will be iterated until a stop criterion is met. In the following, we mainly describe the construction of a candidate set by applying the self-learning strategy.

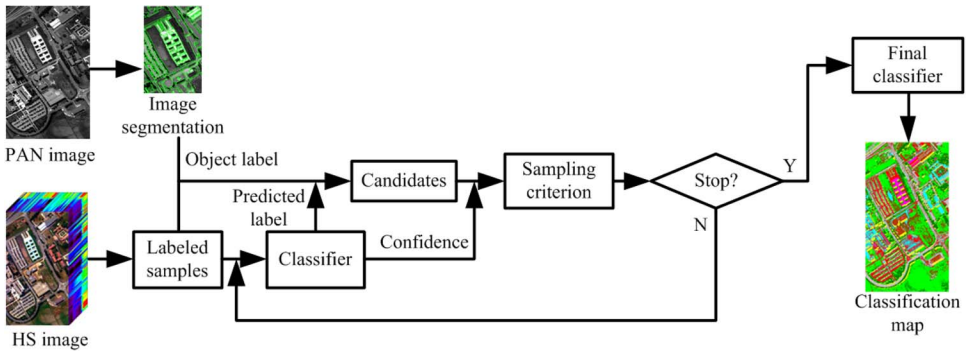

Fig. 1. Main procedures of the proposed synergetic classification approach.

A. *Self-Learning Based on Image Segmentation*

High-resolution images present more ground detail, thereby making the extraction of finer spatial information possible. Nevertheless, the problem of excessive heterogeneity occurs that the ground objects of same category may reflect different spectral features. This noisy phenomenon appears as many isolated pixels and a lack of spatial coherency (speckles and holes in classified areas). Image segmentation makes use of both spatial information and spectral features, and seeks to divide an image into spatially continuous disjoint and nonoverlapping homogeneous regions. As a result, the entire image consists of numerous objects. Each object consists of thousands of pixels that are considered to represent a single ground object [28], [29]. Therefore, the salt-and-pepper noise or spots usually caused by "the same object but in a different spectrum" will be eliminated [30].

During the past few decades, various segmentation techniques have been proposed, e.g., optimal edge detector [31], watershed algorithm [32], [33], region growing [34], etc., which can be classified as boundary-based, region-based, and hybrid approaches. The boundary-based approaches gather the edge detection techniques. These methods do not lead directly to a segmentation of the image because contours obtained are seldom closed. Segmentation by regions might obtain closed ring but imprecise segment boundaries [35]. Therefore, many studies have been dedicated to the combination of edge and region information. Kermad and Chehdi [36] oversegmented images using edge detection and region merging methods and then iteratively processed the segmentations with gradually relaxed constraints until both methods reached dependable coherent results. Li *et al.* [37] extracted image edges using an edge detector embedded with confidence and then used the edges in marker-based watershed segmentation. Experiments showed that these hybrid approaches worked well on high-resolution remote sensing images. Hence, in this paper, edge detection and region merging have been conducted on PAN images to acquire an optimal segmentation result. This segmentation technique comprises two steps. First, boundary detection and region segmentation have been applied to generate a preliminary segmentation result. The optimal segmentation scale can be selected manually. The second step, which refers to region merging, is conducted to eliminate oversegmentation. By defining a merging cost for each pair of adjacent regions, the full λ-schedule algorithm is conducted to merge those similar neighbor regions [38].

The proposed self-learning strategy is based on the assumption that under the optimal segmentation scale, all pixels belonging to a same object have an identical class label. Thus, for a labeled sample, the pixels that share the same object can be labeled with high confidence as belonging to the same class with this labeled sample. Here, we call it the object label. The proposed strategy considers the consistency between the object labels and predicted labels of the classifier, thereby determining which pixels ought to be new training samples in the learning procedure. This way, the self-learning strategy integrates the spatial information and spectral features of labeled samples to construct a new candidate set without human interaction.

Suppose $X^L = \{(x_1^L, y_1), (x_2^L, y_2), \ldots, (x_n^L, y_n)\}$, $y_k \in \{1, 2, \ldots, C\}$, denotes a set of n labeled samples, in which every labeled sample x_k^L, $k \in \{1, 2, \ldots, n\}$, belongs to an individual object. C is the number of classes. O_k^U denotes the set of unlabeled samples inside the object where x_k^L locates. $O^U = \bigcup_{k=1}^{n} O_k^U$ is the set of all unlabeled samples locating in those objects. Therefore

$$\text{If } \forall\, x_i^U \in O_k^U, \text{ then } x_i^U :\rightarrow y_k \tag{1}$$

in which x_i^U is the ith unlabeled sample of O^U, and y_k is the label of the kth object.

Suppose $\forall\, x_i^U \in O_k^U$ is predicted $x_i^U :\rightarrow \tilde{y}_i$ by the classifier, which has been trained by X^L. $\tilde{y}$ denotes the predicted labels. Thus

$$\text{If } \tilde{y}_i = y_k, \text{ then } \left(x_i^U, y_k\right) \rightarrow X^C \tag{2}$$

where X^C is the candidate set

$$X^C = \left\{x_i^U | \tilde{y}_i = y_k\right\}. \tag{3}$$

In other words, for each iteration of the self-learning procedure, the candidate set is updated by selecting the samples that have the identical predicted labels with object labels. After each iteration, the training set and unlabeled set will be also updated

$$X^L = X^L \cup X^I \tag{4}$$

$$O^U = O^U \backslash X^I \tag{5}$$

$$X^C = \varnothing \tag{6}$$

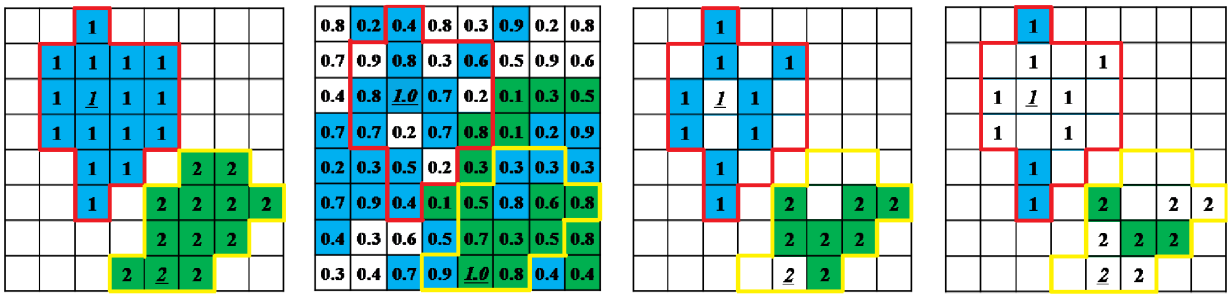

Fig. 2. Main idea of self-learning based on image segmentation (Suppose the blue pixels represent first class and green pixels second class, whereas the white pixels represent the other classes.) (a) Objects where training samples locate. (The pixels with underlined italic labels denote the labeled samples.) (b) Estimated confidence of classifier. (c) Candidates. (d) Selected uncertain samples.

in which $X^I \subseteq X^C$ denotes the selected informative samples. The algorithm of self-learning based on region segmentation is described in the following.

Algorithm 1 Procedure of self-learning

Input: Initial training set X^L, unlabeled sets O_k^U, $k \in \{1, 2, \ldots, n\}$;

1. Assign the corresponding labels y_k to the samples of O_k^U;

Repeat

2. Train the SVM model using X^L;
3. Classify the unlabeled set O^U to obtain the predicted labels using the trained SVM model;
4. Construct the candidate set X^C according to (2) and (3), and optimize the candidate set X^C;
5. Select the informative samples X^I from X^C by applying standard AL strategy;
6. Update training set and candidate set according to (4) and (6), remove X^I from O^U according to (5);

Until: A stopping criterion is satisfied;

Output: SVM classifier retrained by the final training set X^L.

Fig. 2 shows the main idea of constructing the candidate set by using the proposed self-learning strategy. Fig. 2(a) represents two objects belonging to different classes, which is obtained by the presented segmentation method. Fig. 2(b) shows the predicted probabilities (i.e., the confidence) of the spectral-based classifier. Then, the candidate set can be constructed in Fig. 2(c), in which the blue and green cells comprise the candidate set. In addition, the new training samples in Fig. 2(d) can be selected through standard AL strategy. In some cases, there exist a few mislabeled samples, which are probably caused by the misclassification of SVM and suboptimal segmentation. A simple approach to reduce these samples is to assign a confidence threshold for the predicted probabilities of the SVM. Another effective way is to optimize the candidate set (as the second half of Step 4 in Algorithm 1 suggested), which will be described in the succeeding section.

B. Optimization of Candidate Set Using Spectral Similarity

An HS image usually has relatively lower spatial resolution compared with a PAN image, which probably causes mixed pixels on the edge of objects. These mixed pixels usually have relatively lower confidence and dissimilar spectral features, whereas they are meaningless for improving the classification accuracy since they may introduce significant divergence or even reduce the classification accuracy. Therefore, here, we elaborate the impact on the self-learning procedures of mixed pixels and present an approach to exclude those mixed pixels from the candidate set, thereby optimizing the construction of candidate set.

For this purpose, we take the classification algorithm of the SVM, for example. First, let us review the algorithm of SVM briefly. Given a training set made up of some labeled samples, the goal of a binary SVM is to find out an optimal hyperplane that separates the feature space into two classes [39]–[41]. In practical applications, when the training data are nonlinearly separable, then an effective way is to project the data onto another higher dimensional space through a positive definite kernel [42]. The binary SVM can be easily extended for multiple-class classification scenarios through several approaches [43], [44].

During the iterative procedure of self-learning, the candidate set will be constructed according to (2) and (3). Some of them will be selected by a standard AL algorithm and added to the training set to retrain the separating hyperplane. Then, those mixed pixels, which usually have relatively lower confidence and dissimilar spectral features with labeled samples, may cause a significant variation of the hyperplane, thereby leading to an unsatisfactory result that more points may be involved into the margin (the space between $w \cdot x + b = \pm 1$). While some points even may be misclassified, e.g. the points $(Su1)$s in Fig. 3(b). Hence, conventional AL strategy sometimes leads to an obvious decline in classification accuracy.

In order to avoid this problem, the distribution constraint of the candidates should be taken into account. As mentioned earlier, the mixed pixels usually have quite different spectral features with labeled samples. Then, we introduce a Euclidean distance d_{ij} to measure the spectral similarity between each candidate and the support vectors acquired in the previous iteration, i.e.,

$$d_{ij} = \sqrt{\sum_m \left(x_{im}^U - x_{jm}^{\mathrm{SV}}\right)^2} \tag{7}$$

in which x_i^U denotes the ith candidates with m features, $i \in \{1, 2, \ldots, N\}$, and N is the number of aforementioned candidates acquired at each iteration in Section II-B. x_j^{SV} is the jth

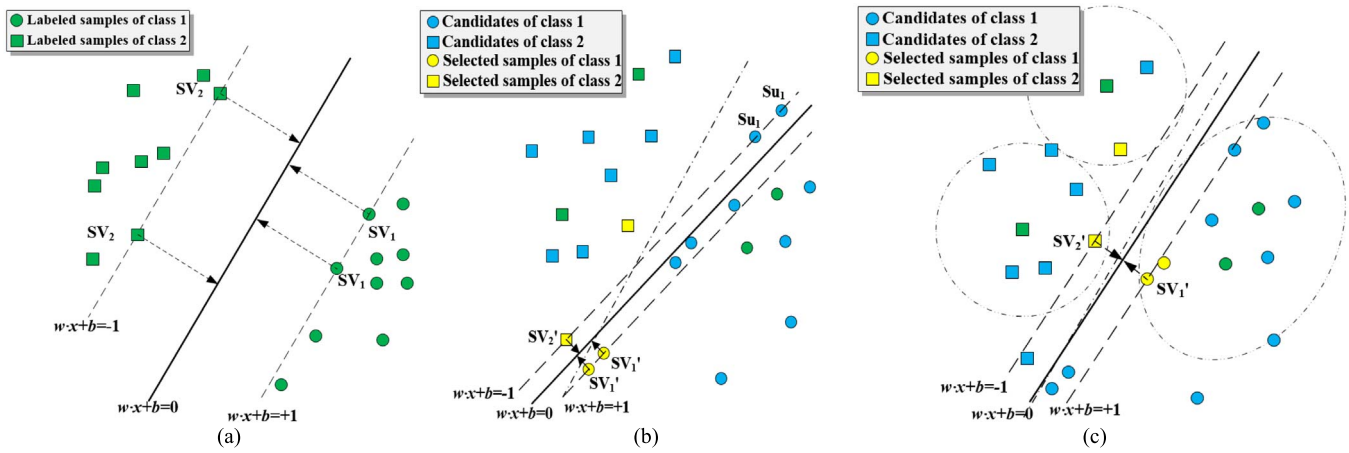

Fig. 3. SVM hyperplanes obtained by training set (the labeled samples) and selected unknown samples. (a) Hyperplane obtained by the training set (Suppose the samples in (a) are the initial labeled samples.) (b) Retrained hyperplane after the first iteration of AL. Four candidates have been selected by MS. (c) The retrained hyperplane after first iteration using the optimized candidate set. The samples in (b) and (c) are the candidates acquired after first iteration. The dot-dash lines in (b) and (c) denote the hyperplane in (a).

support vector, $j \in \{1, 2, \ldots, M\}$, and M is the number of support vectors that have the same class label with x_i^C. By defining a threshold δ, the candidates that have a large distance with the corresponding nearest support vectors will be abandoned, i.e.,

$$\hat{X}^C = X^C \cap \left\{ x_i^U | d_{ij}^{\min} \leq \delta \right\} \tag{8}$$

$$d_{ij}^{\min} = \arg\min_j d_{ij}. \tag{9}$$

Fig. 3(c) shows the distribution constraint of spectral similarity, where the samples out of the ellipses will not be taken into consideration during the learning procedure. By applying AL rules, the new separating hyperplane obtains a much more correct result since none of the samples has been misclassified. The similar rule can be applied to other classification algorithms, in which the support vectors can be substituted with some training samples. Algorithm 2 summarizes the optimization procedure of the candidate set.

Algorithm 2 Optimization procedure of candidate set

Input: Candidate set X^C, support vector set X^{SV};
For $i = 1$ **to** N
1. Find the support vectors $\{x_j^{\mathrm{SV}}\}_{j=1}^{M}$ that has the same label y_i with candidate x_i^U;
2. Compute the Euclidean distance d_{ij} between x_i and x_j^{SV};
3. Find the minimum d_{ij} according to (9);

End
4. Remain the samples which satisfy $d_{ij}^{\min} < \delta$, and remove the others from X^C according to (8);

Output: An optimized candidate set $\hat{X}^C$.

C. AL Strategy

In this paper, four different AL strategies will be adopted to evaluate our proposed approach: 1) MS; 2) BTs; 3) MBTs; and 4) modified MS (MMS). As an SVM-specific AL strategy, MS heuristic samples the candidates lying within the margin by computing their distance to the separating hyperplane [45]–[47]. The classical MS method has not considered the distribution of the candidates in the feature space and has the drawback of biased sampling among the classes [12]. The BT strategy is equal to the MCLU technique for probabilistic SVM classifiers since they both aim at finding the samples with minimum difference between two most possible classes. The MBT technique has the ability to promote diversity in the selection process [23]. MMS combines the idea of MBT and MS strategies, and aims at dividing the candidate set into several groups according to their hypothetical labels and then applying MS to select the uncertain samples that are closest to the separating hyperplane.

The AL procedure is iterated until an automatic or manual stop criterion is met, for instance, a maximal iteration is met, or the decision boundary remains unchanged. Generally, the within-class variance keeps invariant when the decision boundary is almost not changed. Hence, in this paper, we introduce the within-class variance to define the stop criterion of a self-learning procedure.

III. Experimental Results and Analysis

In this paper, three groups of HS and PAN images have been used to test the proposed approach. The first HS image acquired by the ROSIS optical sensor over the University of Pavia in Italy consists of 103 spectral bands (with spectral range from 0.43 to 0.86 μm) of size (610 $\times$ 340) pixels with a spatial resolution of 1.3 m per pixel [i.e., Fig. 4(a)]. For the sake of avoiding the registration error caused by different platform, we use bands 1–65 (within the spectrum range of about 0.43–0.7 μm) of the HS image to synthesize a PAN image. In addition, the segmentation process will be conducted on this PAN image. The ground-truth image includes nine classes with the same resolution of the HS image, i.e., Fig. 4(b). The second HS data sample is a low altitude AVIRIS image of a portion of the

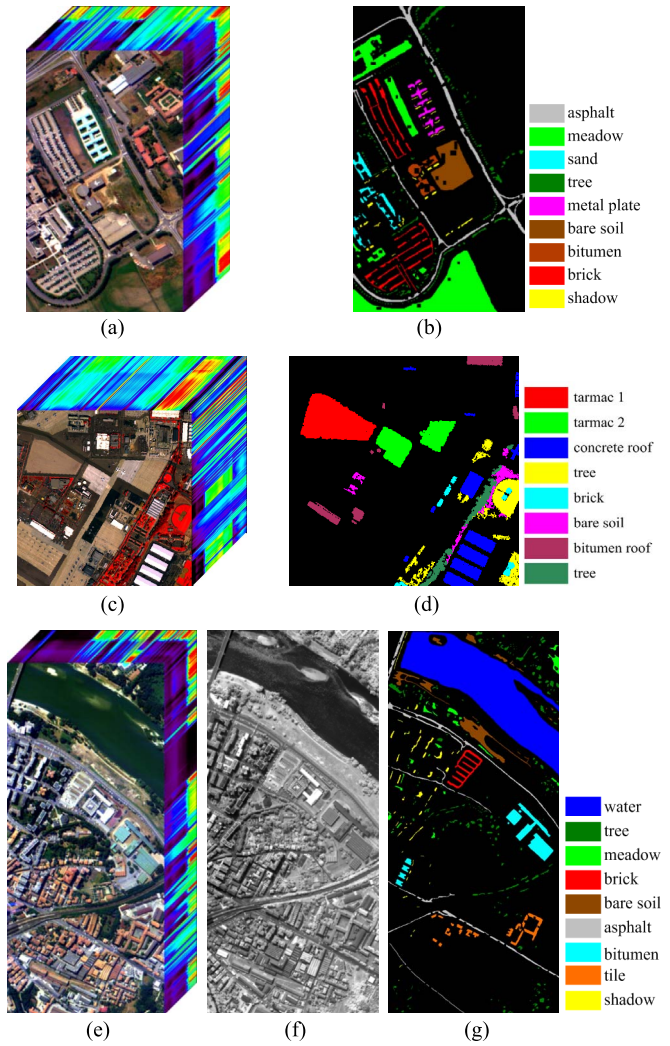

Fig. 4. Experimental data sets. (a) HS image of University of Pavia. (b) Ground-truth image with nine classes. (c) HS image of San Diego. (d) Ground-truth image with eight classes. (e) HS image of Pavia City Centre. (f) PAN image of Pavia City Centre. (g) Ground-truth image with nine classes.

North Island of the U.S. Naval Air Station in San Diego, CA, USA. It consists of 126 bands of size (400 $\times$ 400) pixels with spatial resolution of 3.5 m per pixel [i.e., Fig. 4(c)]. For better comparison, we use bands 6–36 to synthesize a PAN image. Then, the HS image is subsampled to a lower scale by a factor of 4 (i.e., a resolution of 14 m). The ground-truth image has the same resolution with the synthesized PAN image with eight classes inside [i.e., Fig. 4(d)]. The last scene is a pair of real HS and PAN images collected over the Pavia City Centre. The HS image that is also acquired by the ROSIS optical sensor and consists of 102 spectral bands of size 1096 $\times$ 490 pixels is shown in Fig. 4(e). The PAN image was collected by QuickBird satellite (size of 2375 $\times$ 1062 pixels) with a spatial resolution of 0.6 m per pixel [see Fig. 4(f)] and has been registered to the ground-truth image. The corresponding ground-truth image has a spatial resolution of 0.6 m with nine classes inside, i.e., Fig. 4(g). A training map is also provided along with the ground-truth image, including a subset of labeled samples. All images have been radiometrically calibrated.

It is important to emphasize that the proposed self-learning strategy is applicable to various probabilistic classifiers [48]. Although in our experiments, we mainly adopt SVM algorithm. To demonstrate the performance of our proposed approach, we use very small labeled training sets, i.e., only 5, 10, or 15 samples per class will be selected randomly at most for training purposes. Obviously, with the increment of labeled samples, more unlabeled samples (i.e., the candidates) are available both in the NBSL approach and our proposed approach [segmentation-based self-learning (SBSL)]. Four standard AL strategies, i.e., MS, BT, MMS, and MBT, have been utilized to test our proposed approach. In all cases, we conduct ten independent Monte Carlo runs with respect to the labeled training set from the ground-truth images, where the remaining labeled samples are used for testing purposes. All the graphics display the average values of ten experiments. Thus, our results and conclusions should not depend on any particular choice of labeled data. It is also worth noting that defining the optimal scale for image segmentation remains problematic as no standard method currently exists for setting the scale parameter in segmentation algorithms. Here, an oversegmentation would be preferred because it is necessary to ensure that the given samples with different class labels should not locate in a single object. It is the precondition of the next step. The parameters of the SVM model, i.e., the penalty factor C (controls the penalty assigned to errors) and σ (controls the spread of the Gaussian radial basis function kernel), are chosen by fivefold cross validation and will be updated at each iteration of the AL procedure. The selection of distance threshold δ depends on the amount of selected unlabeled samples per iteration and the amount of labeled samples. The maximum iteration is 20.

A. *Comparison of Segmentation Results*

Here, we first describe the impact of segmentation scale on the classification results. The comparison is conducted on the University of Pavia data set. The segmentation parameters contain a scale level and a merge level, which decide the edge intensity and merge cost, respectively. Three groups of segmentation parameters have been selected to show the results corresponding to three different scales, respectively. The comparisons are conducted under different amounts of training samples and AL approaches. Fig. 5(a)–(c) shows the segmentation results under different scale and merge levels. Fig. 5(a) represents a typical case of oversegmentation, in which several objects have been segmented into much more undesired small regions. In this case, the amount of unlabeled samples is limited since only part of the pixels near the labeled samples can be considered candidates. While Fig. 5(c) shows a typical case of undersegmentation, in which segments representing different land covers are incorrectly merged together. Then, in such a case, the classification result may be unstable due to the possibly mislabeled samples. Fig. 5(b) displays the segmented image under relatively appropriate parameters, where most segment boundaries in the upper right corner are coherent to real image edges.

Fig. 5(d)–(i) shows part of the results. The horizontal axis represents the number of iterations, which also represents the number of unlabeled samples selected during the learning

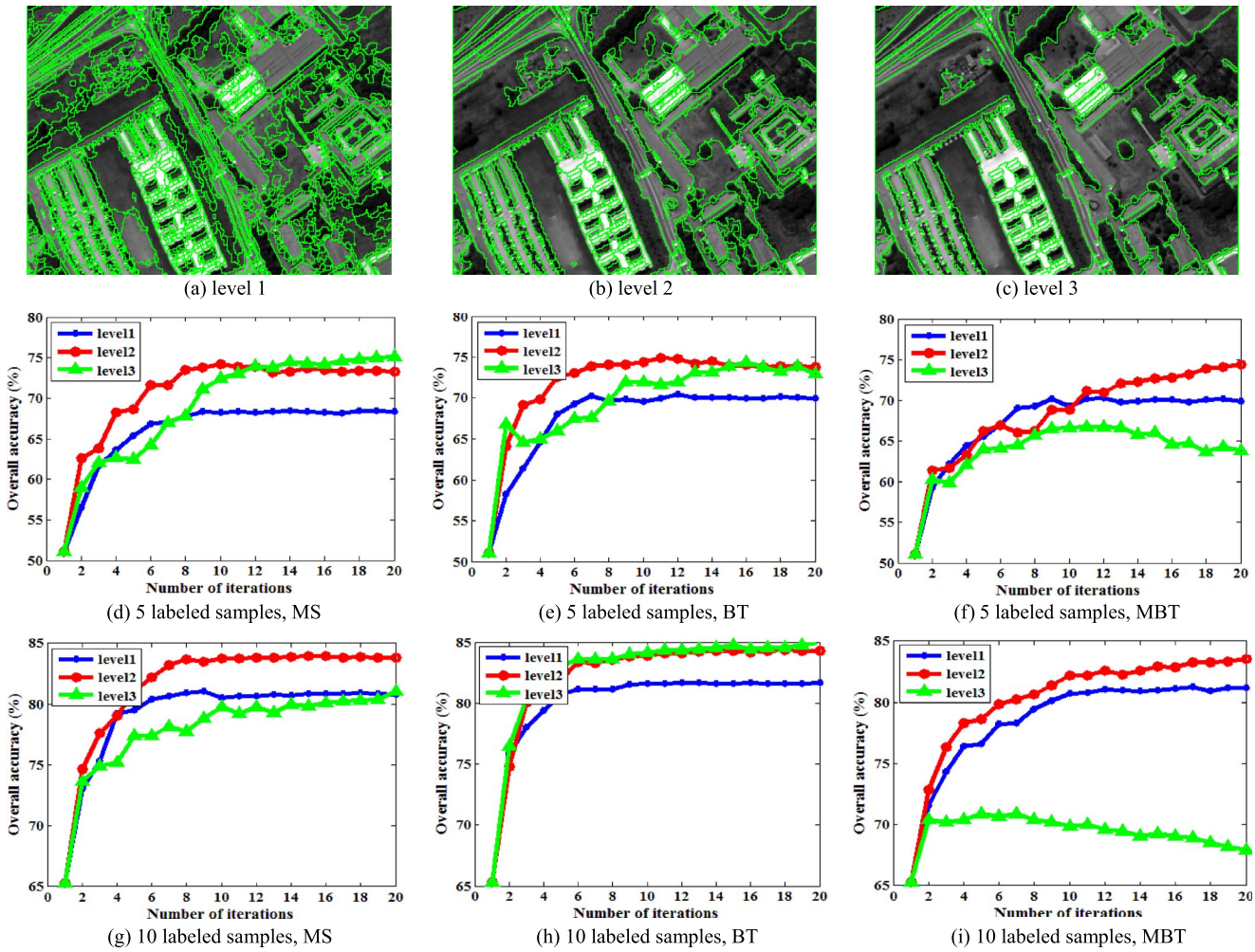

Fig. 5. Impact of segmentation scale on overall classification accuracy (as a function of iteration of AL) obtained for the University of Pavia data set. The segmentation parameters of three levels are as follows: (a) scale level = 20, merge level = 60; (b) scale level = 40, merge level = 90; and (c) scale level = 60, merge level = 90. The second row represents the OAs obtained by AL sampling strategies, (i.e., MS, BT, and MBT) under five initial training samples per class with 90 selected samples per iteration. In addition, the bottom row represents the OAs under ten initial training samples per class with 180 selected samples per iteration.

procedure. In addition, the vertical axis represents the overall accuracy values (OA). The first point on each curve represents the overall accuracy obtained by using SVM algorithm alone, which is generally affected by the labeled samples (50.08% and 65.29% corresponding to five and ten labeled samples per class). From these figures, it can be seen that the proposed self-learning algorithm can incorporate unlabeled samples in a few objects into the training set, thereby significantly improving the classification result, for instance, nearly 75%/80% after ten iterations when using five/ten labeled samples per class, respectively. However, these figures also reveal the impact of segmentation results. We can observe that, in most cases, level 2 (i.e., when the segmentation levels are selected appropriately) performs better than the other two levels, since it has relatively stable improvement of overall accuracy. Level 1 has lower accuracy values compared with level 2 in these figures, which is mainly because the insufficient unlabeled samples and relatively less information of those samples under the condition of oversegmentation. On the other side, level 3 has quite unsteady results that sometimes it has excellent improvement, e.g., Fig. 5(h) but sometimes it seems to be meaningless, e.g., Fig. 5(i). This is possibly because the mislabeled samples in different objects will decrease the performance of self-learning in the case of undersegmentation. A more intuitive comparison of overall accuracies (OAs) with respect to segmentation levels is given in Fig. 6, in which the horizontal axis represents

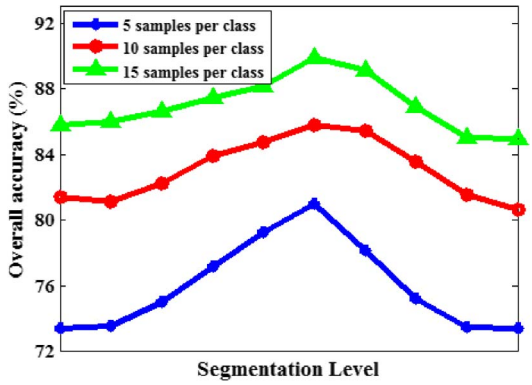

Fig. 6. Comparison of OAs with respect to segmentation level. (Ten groups of segmentation parameters from finer scales to coarser scales have been tested by the proposed self-learning approach using BT sampling strategy.)

TABLE I
OAs OF COLLABORATIVE SPATIAL–SPECTRAL CLASSIFICATIONS

	5 samples per class		10 samples per class		15 samples per class	
	OA (%)	Kappa	OA (%)	Kappa	OA (%)	Kappa
SPE	51.51	0.4184	66.69	0.5874	72.39	0.6534
SFC	57.86	0.4971	77.08	0.7070	80.83	0.7544
MFC	59.19	0.5096	78.35	0.7206	80.54	0.7493
MKL	66.91	0.5855	77.86	0.7154	81.07	0.7551
NBSL	72.50	0.6541	83.11	0.7812	87.25	0.8325
SBSL	**76.54**	**0.6969**	**86.21**	**0.8156**	**89.01**	**0.8524**

ten segmentation levels from finer scales to coarser scales, and the vertical axis shows the final OAs after the learning processes. The results reveal that the classification results will depend on the segmentation results to a great extent. Therefore, to exploit the information of unlabeled samples properly, the segmentation scale should be selected appropriately to reduce undesired oversegmentation or undersegmentation.

B. Comparison of Spatial–Spectral Features Classifications

To demonstrate the feasibility and superiority of our proposed algorithm in combining spatial information, here, we compare our self-learning algorithm with several widely utilized classification approaches that combine collaborative spatial–spectral features. For this reason, we have employed the Gabor wavelet [49] to extract 32 texture features (i.e., four scales and eight directions) from a corresponding PAN image. Meanwhile, the pixel shape index and the structural feature set have been employed to describe the shape features in a local area surrounding a pixel, which measures the gray similarity distance in every direction [50]. The spectral features and spatial features (i.e., texture features and shape features) will be integrated by multiple feature combining (MFC) [51] and multiple kernel learning (MKL) [52], [53], as well as a serial feature combination (SFC) method [54], and followed by SVM classifiers. In addition, the spectral features (SPE) will be used to conduct supervised classification (i.e., SVM) and semisupervised classification (i.e., NBSL and SBSL, the self-learning algorithms using BT sampling strategy). Table I shows the classification accuracy values obtained for the University of Pavia data set, from which it can be seen that when fewer samples are used in this experiment, it is very difficult for supervised algorithms to provide good classification results as very little information is available about the class distribution. For instance, when only five samples per class are available in Table I, the results turn out to be very poor (only 51.51% on average).

Although various techniques of extracting spatial features have been utilized to raise the performance of supervised classifiers, they are still sometimes unsatisfactory. While both the proposed SBSL algorithm and the NBSL algorithm exhibit clear advantages of using unlabeled samples since the OAs have an impressive enhancement compared with these feature combination approaches (nearly 76%/73% when five labeled samples per class have been used by SBSL and NBSL, respectively).

Fig. 7 displays some of the classification maps obtained by spatial–spectral synergic classification and self-learning approaches. These maps correspond to one of the ten experiments. Aside from the OA values, the effect of our proposed approach is apparent in the classification maps. For example, the meadow in the lower left corner is much more homogeneous and continuous than that of the top row.

C. Experiments With San Diego Data Set

The experimental results of San Diego Data set are described here. These data sets comprise of a subsampled AVIRIS HS image and a synthesized PAN image. We compared our proposed SBSL algorithm with the classical SVM algorithm and the NBSL algorithm. The within-class variance criterion is adopted to terminate the iterations in this experiment. While for better comparison, we plot the OAs within 20 iterations in Fig. 8. The first point on each curve also represents the OA obtained by using the SVM algorithm alone. From each figure, it can be seen that both the NBSL and the proposed SBSL approaches show great enhancement of OAs, for instance, nearly 88%/90% for NBSL and SBSL, respectively, after a few iterations when using 5 labeled samples per class, which is much higher than supervised classifications. In all cases, the proposed strategy and the NBSL strategy outperform the corresponding supervised algorithms significantly and the increases in performance are more relevant as the number of unlabeled samples increases.

By comparing the two curves in each figure, it is obvious that our proposed approach has much higher OAs in most cases. This is because, compared with NBSL, the proposed approach aims at selecting those most informative samples in a single object (usually have lower spectral similarity with the known sample). While NBSL selects the uncertain samples locating in the neighborhood of the labeled samples (usually having greater spectral similarity). Therefore, when fewer labeled samples are used, for instance, when only five samples per class are available in Fig. 8(a)–(d), the improvement of our self-learning approach will be much more evident. While with the increment of labeled samples, the unlabeled samples make fewer contributions to the accuracy compared with the labeled samples. Fig. 8 also reveals that different AL sampling strategies always lead to discrepant results. MMS and MBT approaches seem to have better stability and faster convergence than MS and BT when fewer samples are available. When more samples are available, the MS approach can obtain higher accuracy values than MMS and MBT. The BT approach seems to be not a preferable choice for this data set.

D. Experiments With Pavia City Centre Data Set

The experimental results of Pavia City Centre data set are described here. These data sets comprise of real PAN and HS images. Before the classification experiments, we perform principal component analysis (PCA) on the HS image and extract the first three components (contain more than 99% of the total energy of the data) for subsequent classifications.

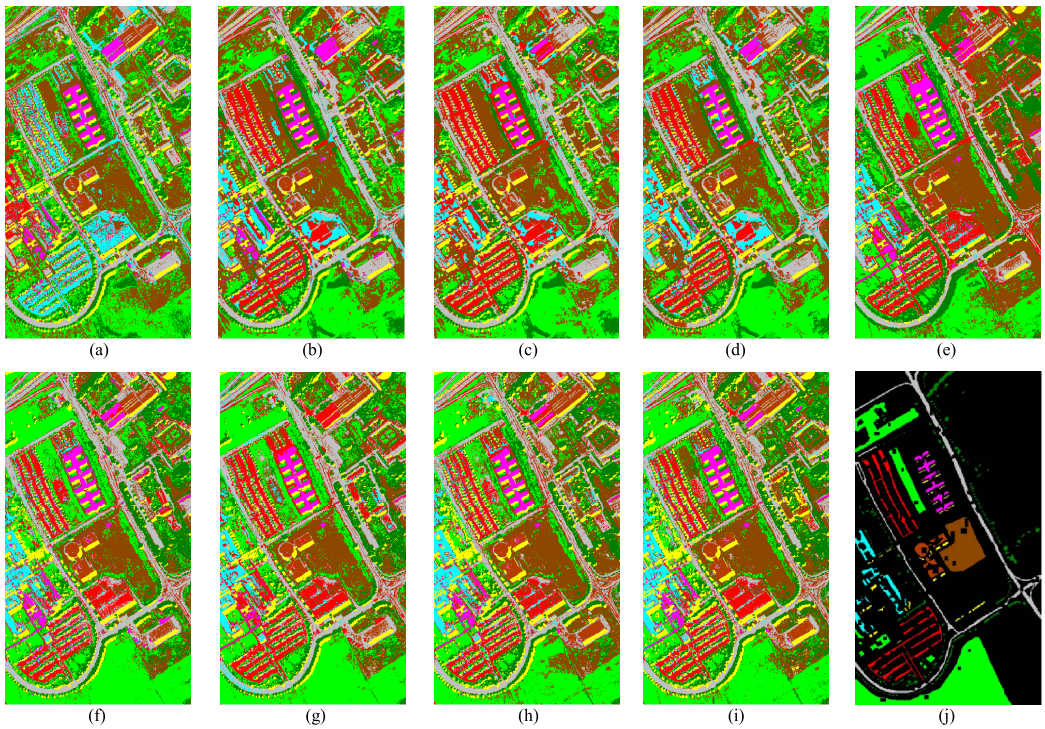

Fig. 7. Classification maps and OAs obtained for the University of Pavia data set after applying different spatial–spectral feature synergic classification and self-learning approaches. (a) SPE (60.92%). (b) SFC (77.81%). (c) MFC (74.42%). (d) MKL (78.62%). (e) NBSL+BT (83.47%). (f) SBSL+MS (89.17%). (g) SBSL+BT (88.09%). (h) SBSL+MMS (83.70%). (i) SBSL+MBT (86.01%). (j) Ground-truth image.

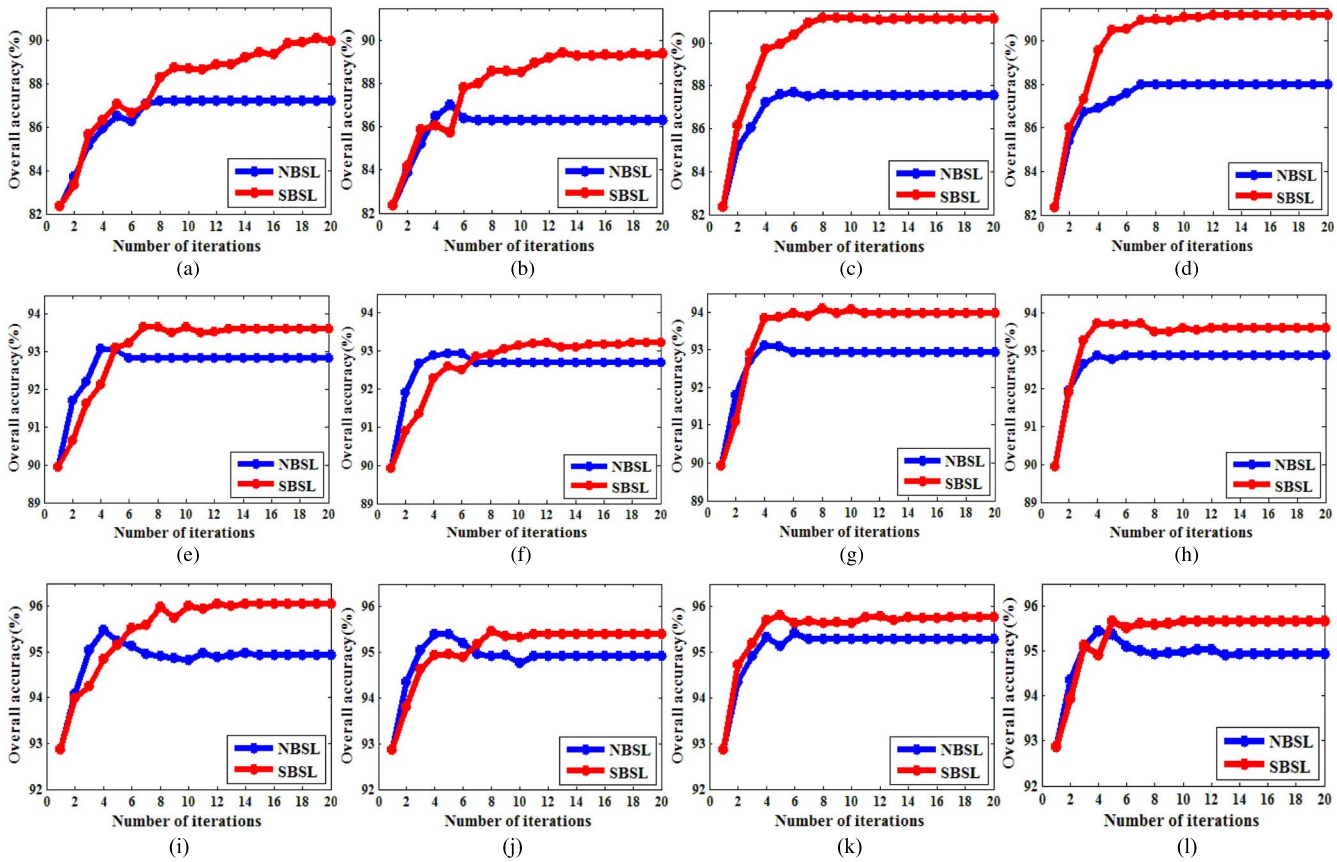

Fig. 8. Overall classification accuracy values (as a function of iteration of AL) obtained for the data set of San Diego. The four figures in each row represent the OAs using different AL sampling strategies, i.e., MS, BT, MMS, and MBT. While the three figures in each column represent the OAs using different amounts of samples, i.e., 5 initial training samples per class with 30 selected samples per iteration, 10 initial training samples per class with 60 selected samples per iteration, and 15 initial training samples per class with 90 selected samples per iteration. (a) Five labeled samples, MS. (b) Five labeled samples, BT. (c) Five labeled samples, MMS. (d) Five labeled samples, MBT. (e) Ten labeled samples, MS. (f) Ten labeled samples, BT. (g) Ten labeled samples, MMS. (h) Ten labeled samples, MBT. (i) Fifteen labeled samples, MS. (j) Fifteen labeled samples, BT. (k) Fifteen labeled samples, MMS. (l) Fifteen labeled samples, MBT.

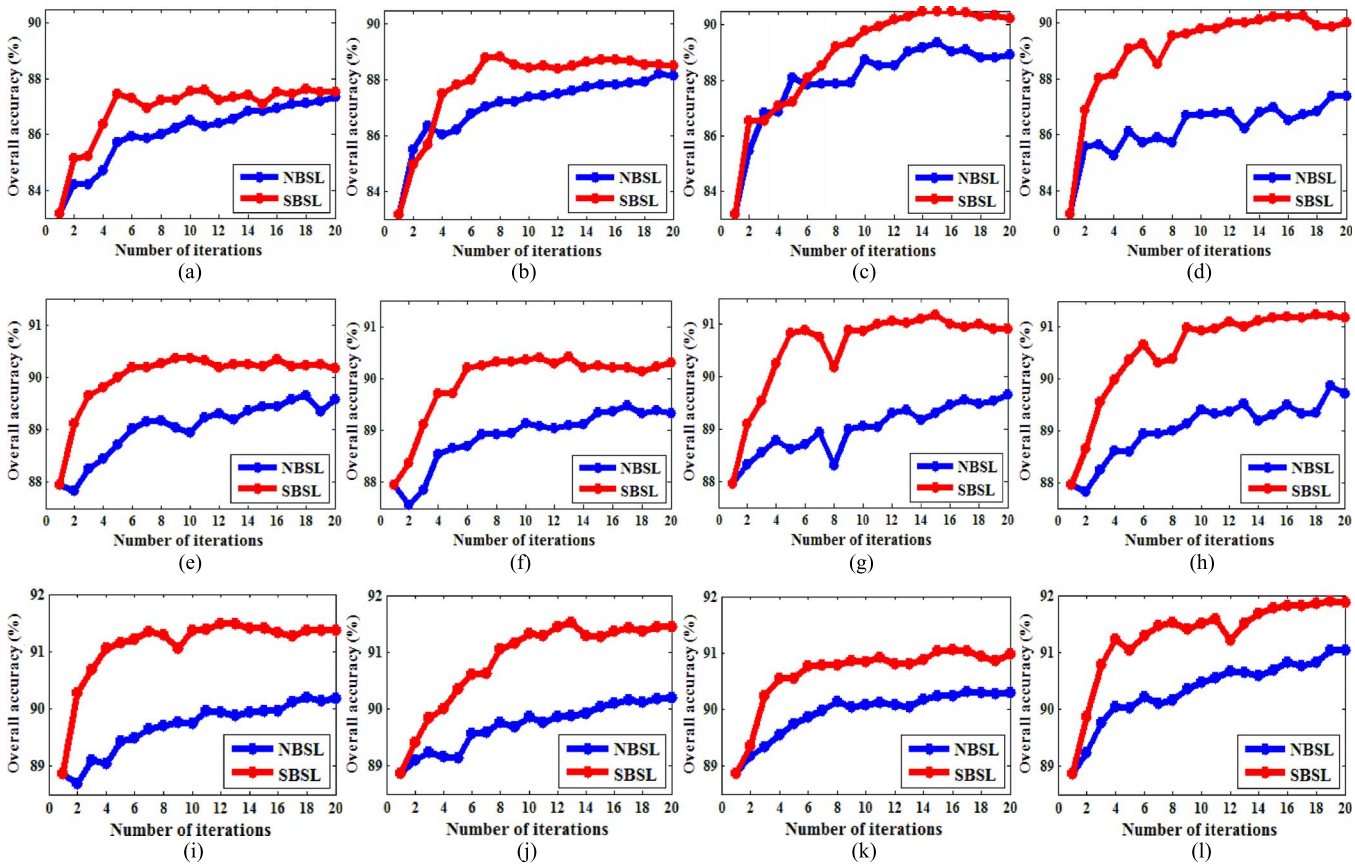

Fig. 9. Overall classification accuracy values (as a function of iteration of AL) obtained for the data set of Pavia City Centre. The four figures in each row represent the OAs using AL sampling strategies, i.e., MS, BT, MMS, and MBT. While the three figures in each column represent the OAs using different amounts of samples, i.e., 5 initial training samples per class with 45 selected samples per iteration, 10 initial training samples per class with 90 selected samples per iteration, and 15 initial training samples per class with 135 selected samples per iteration. (a) Five labeled samples, MS. (b) Five labeled samples, BT. (c) Five labeled samples, MMS. (d) Five labeled samples, MBT. (e) Ten labeled samples, MS. (f) Ten labeled samples, BT. (g) Ten labeled samples, MMS. (h) Ten labeled samples, MBT. (i) Fifteen labeled samples, MS. (j) Fifteen labeled samples, BT. (k) Fifteen labeled samples, MMS. (l) Fifteen labeled samples, MBT.

Fig. 9 gives the detailed accuracy comparison within 20 iterations. From these figures, we can observe a similar phenomenon on the whole with Fig. 8. With the increment of unlabeled samples, all curves show an uptrend of classification accuracy. In most cases, the proposed approach has higher OAs than that of NBSL with faster convergence as well. The MS and BT have lower accuracy values when fewer unlabeled samples are used (only 87.51% and 88.50% when selecting 45 unlabeled samples per iteration), whereas MMS and MBT have relatively higher accuracy values (90.02% and 90.22%). This is mainly caused by the biased sampling among the classes. This problem can be alleviated by adding more unlabeled samples per iteration, for example, Fig. 9(e) and (i). In addition, the proposed approach is more robust particularly when applying MMS and MBT strategies since there are only a few fluctuations during the AL iterations.

To sum up, the self-learning strategies can take advantage of both spectral features and spatial information, and enable a great enhancement of classification accuracy values by adding unlabeled samples into the training set. In addition, the distribution of unlabeled samples in feature space has also quite different influences on the learning procedure.

IV. Conclusion

In this paper, we have proposed a novel self-learning algorithm based on image segmentation for the synergetic classification of HS and PAN images. Considering the spectral and spatial characteristics of HS/PAN images, the proposed algorithm utilizes the concept of object label and predicted label to produce more available training samples automatically, thereby reducing the labeling cost of human beings. It is worth noting that our proposed approach is not only applicable to multisource classification but can also be used to deal with single data source (e.g., airborne HS images with high spatial resolution). The experiments conducted on HS and PAN images have demonstrated that the proposed approach can take good advantage of both labeled and unlabeled samples to achieve much better results compared with supervised classification algorithm and conventional feature combination approaches, particularly when the supervised method is not able to provide satisfactory classification results. Although, in this paper, we focus on the SVM classifier, the proposed approach can be also applied to other types of probabilistic classifiers. It is also worth noting that the proposed approach and the neighbor-based self-learning strategy are equivalent when the number of iterations

is large enough since the NBSL approach is equal to the region growing-based image segmentation, whereas the proposed approach has faster convergence. Nevertheless, in our proposed approach, the threshold of spectral similarity and the amount of unlabeled samples are manually determined. Then in future work, we consider designing an adaptive method to adjust the parameters automatically during the learning procedure.

ACKNOWLEDGMENT

The authors would like to thank Prof. P. Gamba for providing the ROSIS data over Pavia, Italy.

REFERENCES

[1] Q. Sami ul Haq, L. Tao, F. Sun, and S. Yang, "Fast and robust sparse approach for hyperspectral data classification using a few labeled samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2287–2302, Jun. 2012.

[2] C. Persello and L. Bruzzone, "Active and semisupervised learning for the classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 11, pp. 6937–6956, Nov. 2014.

[3] S. Patra and L. Bruzzone, "A fast cluster-assumption based active-learning technique for classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 5, pp. 1617–1626, May 2011.

[4] G. Camps-Valls, T. Bandos Marsheva, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.

[5] L. Bruzzone, M. Chi, and M. Marconcini, "A novel transductive SVM for semisupervised classification of remote-sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 11, pp. 3363–3373, Nov. 2006.

[6] Y. Wang, S. Chen, and Z. Zhou, "New semi-supervised classification method based on modified cluster assumption," *IEEE Trans. Neural. Netw. Learn Syst.*, vol. 23, no. 5, pp. 689–702, May 2012.

[7] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. Workshop Comput. Learn. Theory*, 1998, pp. 92–100.

[8] M. Chi and L. Bruzzone, "Semisupervised classification of hyperspectral images by SVMs optimized in the primal," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 6, pp. 1870–1880, Jun. 2007.

[9] L. Zhang *et al.*, "Active learning based on locally linear reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no 10, pp. 2026–2038, Oct. 2011.

[10] D. Lewis and J. Catlett, "Heterogeneous uncertainty sampling for supervised learning," in *Proc. ICML*, 1994, pp. 148–156.

[11] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari, "A survey of active learning algorithms for supervised remote sensing image classification," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 3, pp. 606–617, Jun. 2011.

[12] D. Tuia, F. Ratle, F. Pacifici, M. 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.

[13] L. Copa, D. Tuia, M. Volpi, and M. Kanevski, "Unbiased query-by-bagging active learning for VHR image classification," in *Proc. SPIE Remote Sens. Conf.*, Toulouse, France, 2010, pp. 1–8.

[14] W. Di and M. Crawford, "Multi-view adaptive disagreement based active learning for hyperspectral image classification," in *Proc. IEEE IGARSS*, Honolulu, HI, USA, 2010, pp. 1374–1377.

[15] A. Stumpf, N. Lachiche, J. Malet, N. Kerle, and A. Puissant, "Active learning in the spatial domain for remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2492–2507, May 2014.

[16] S. Zomer, M. N. Sànchez, R. G. Brereton, and J. L. Pérez-Pavón, "Active learning support vector machines for optimal sample selection in classification," *J. Chemometrics*, vol. 18, no. 6, pp. 294–305, 2004.

[17] C. Campbell, N. Cristianini, and A. J. Smola, "Query learning with large margin classifiers," in *Proc. ICML*, 2000, pp. 111–118.

[18] M. Ferecatu and N. Boujemaa, "Interactive remote-sensing image retrieval using active relevance feedback," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 4, pp. 818–826, Apr. 2007.

[19] B. Demir, C. Persello, and L. Bruzzone, "Batch mode active learning methods for the interactive classification of remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 3, pp. 1014–1031, Mar. 2011.

[20] C. Persello, "Interactive domain adaptation for the classification of remote sensing images using active learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 4, pp. 736–740, Jul. 2013.

[21] S. Rajan, J. Ghosh, and M. Crawford, "An active learning approach to hyperspectral data classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1231–1242, Apr. 2008.

[22] D. Tuia, E. Pasolli, and W. J. Emery, "Using active learning to adapt remote sensing image classifiers," *Remote Sens. Environ.*, vol. 115, no. 9, pp. 2232–2242, 2011.

[23] 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.

[24] D. Tuia, J. Muñoz-Marí, M. Kanevski, and G. Camps-Valls, "Cluster based active learning for compact image classification," in *Proc. IEEE IGARSS*, Honolulu, HI, USA, 2010, pp. 2827–2827.

[25] C. Persello *et al.*, "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] E. Pasolli, F. Melgani, D. Tuia, F. Pacifici, and W. J. Emery, "SVM active learning approach for image classification using spatial information," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 2217–2233, Apr. 2014.

[27] I. Dopido *et al.*, "Semisupervised self-learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4032–4044, Jul. 2013.

[28] K. Haris, S. N. Efstratiadis, N. Maglaveras, and A. K. Katsaggelos, "Hybrid image segmentation using watersheds and fast region merging," *IEEE Trans. Image Process.*, vol. 7, no. 12, pp. 1684–1699, Dec. 1998.

[29] M. Bouziani, K. Goita, and D. He, "Rule-based classification of a very high resolution image in an urban environment using multispectral segmentation guided by Cartographic data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 8, pp. 3198–3211, Aug. 2010.

[30] P. Ghamisi, M. S. Couceiro, M. Fauvel, and J. Atli Benediktsson, "Integration of segmentation techniques for classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 342–346, Jan. 2014.

[31] R. J. Qian and T. S. Huang, "Optimal edge detection in two-dimensional images," *IEEE Trans. Image Process.*, vol. 5, no. 7, pp. 1215–1220, Jul. 1996.

[32] I. Levner and H. Zhang, "Classification-driven watershed segmentation," *IEEE Trans. Image Process.*, vol. 16, no. 5, pp. 1437–1445, May 2007.

[33] Y. Tarabalka, J. Chanussot, J. A. Benediktsson, J. Angulo, and M. Fauvel, "Segmentation and classification of hyperspectral images using watershed transformation," in *Proc. IEEE IGARSS*, Jul. 2008, vol. 3, pp. III-652–III-655.

[34] Y. Gao, J. F. Mas, N. Kerle, and J. A. N. Pacheco, "Optimal region growing segmentation and its effect on classification accuracy," *Int. J. Remote Sens.*, vol. 32, no. 13, pp. 3747–3763, Jul. 2011.

[35] M. Wang and R. Li, "Segmentation of high spatial resolution remote sensing imagery based on hard-boundary constraint and two-stage merging," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5712–5725, Sep. 2014.

[36] C. D. Kermad and K. Chehdi, "Automatic image segmentation system through iterative edge-region co-operation," *Image Vis. Comput.*, vol. 20, no. 8, pp. 541–555, Jun. 2002.

[37] D. Li, G. Zhang, Z. Wu, and L. Yi, "An edge embedded marker-based watershed algorithm for high spatial resolution remote sensing image segmentation," *IEEE Trans. Image Process.*, vol. 19, no. 10, pp. 2781–2787, Oct. 2010.

[38] D. Crisp, "Improved data structures for fast region merging segmentation using A Mumford-Shah energy functional," in *Proc. DICTA*, Dec. 2008, pp. 586–592.

[39] S. Cheng and F. Y. Shih, "An improved incremental training algorithm for support vector machines using active query," *Pattern Recognit.*, vol. 40, no. 3, pp. 964–971, 2007.

[40] G. M. Foody and M. Ajay, "A relative evaluation of multiclass image classification by support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 6, pp. 1335–1343, Jun. 2004.

[41] 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.

[42] B. Wohlberg, D. M. Tartakovsky, and A. Guadagnini, "Subsurface characterization with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 1, pp. 47–57, Jan. 2006.

[43] A. Mathur and G. M. Foody, "Multiclass and binary SVM classification: Implications for training and classification users," *IEEE Geosci. Remote Sens. Lett.*, vol. 5, no. 2, pp. 241–245, Apr. 2008.

[44] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 66, no. 3, pp. 247–259, May 2011.

[45] P. Mitra, B. U. Shankar, and S. K. Pal, "Segmentation of multispectral remote sensing images using active support vector machines," *Pattern Recognit. Lett.*, vol. 25, no. 9, pp. 1067–1074, Jul. 2004.
[46] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," *J. Mach. Learn. Res.*, vol. 2, no. 1, pp. 45–66, Mar. 2002.
[47] G. Schohn and D. Cohn, "Less is more: Active learning with support vectors machines," in *Proc. 17th ICML*, Stanford, CA, USA, 2000, pp. 839–846.
[48] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.
[49] D. Tao, X. Li, X. Wu, and S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 10, pp. 1700–1715, Oct. 2007.
[50] X. Huang, L. Zhang, and P. Li, "Classification and extraction of spatial features in urban areas using high-resolution multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 260–264, Apr. 2007.
[51] L. Zhang, L. Zhang, D. Tao, and X. Huang, "On combining multiple features for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 879–893, Mar. 2012.
[52] D. Tuia, G. Camps-Valls, G. Matasci, and M. Kanevski, "Learning relevant image features with multiple-kernel classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3780–3791, Oct. 2010.
[53] 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.
[54] J. Yang, J. Yang, D. Zhang, and J. Lua, "Feature fusion: Parallel strategy vs. serial strategy," *Pattern Recognit.*, vol. 36, no. 6, pp. 1369–1381, Jun. 2003.

Xiaochen Lu (S'14) received the B.S. degree in automation from the Hefei University of Technology, Hefei, China, in 2010, and the M.S. degree in control engineering from the Nanjing University of Science and Technology, Nanjing, China, in 2013. He is currently working toward the Ph.D. degree at the Harbin Institute of Technology, Harbin, China.

His research interests include multi/hyperspectral and high-resolution remote sensing image processing and multisource information fusion, and their applications.

Junping Zhang (M'05) received the B.S. degree in biomedical engineering and instruments from Harbin Engineering University, Harbin, China, and Harbin Medical University, Harbin, in 1993, and the M.S. and Ph.D. degrees in signal and information processing from the Harbin Institute of Technology (HIT), Harbin, China, in 1998 and 2002, respectively.

Currently, she is a Professor with the Department of Information Engineering, School of Electronics and Information Engineering, HIT. Her research interests include hyperspectral data analysis and image processing, multisource information fusion, pattern recognition, and classification.

Tong Li received the B.S. degree in information countermeasure technology, and the M.S. degree in electronics and communication engineering from the Harbin Institute of Technology, Harbin, China, in 2012 and 2014, respectively. Currently, she is working toward the Ph.D. degree in information and communication engineering at Harbin Institute of Technology.

Her research interests include hyperspectral image processing and multisource information fusion.

Ye Zhang (M'10) received the B.S. degree in communication engineering and the M.S. and Ph.D. degrees in communication and electronic systems from the Harbin Institute of Technology (HIT), Harbin, China, in 1982, 1985, and 1996, respectively.

Since 1985, he has joined HIT as a Teacher. Between 1998 and 1999, he was a Visiting Scholar with the University of Texas at San Antonio, San Antonio, TX, USA. He is currently a Professor and Doctoral Supervisor in information and communication engineering. He is the Director of the Institute of Image and Information Technology with the School of Electronic and Information Engineering, HIT. His research interests include remote sensing hyperspectral image analysis and processing, image video compression, and transmission, and multisource information collaboration processing and applications.

View publication stats