An efficient semi-supervised classification approach for hyperspectral imagery

Kun Tan [a], Erzhu Li [a], Qian Du [b], Peijun Du [c,*]

[a] *Jiangsu Key Laboratory of Resources and Environment Information Engineering, China University of Mining and Technology, China*
[b] *Department of Electrical and Computer Engineering, Mississippi State University, USA*
[c] *Key Laboratory for Satellite Mapping Technology and Applications of State Administration of Surveying, Mapping and Geoinformation of China, Nanjing University, China*

A B S T R A C T

In this paper, an efficient semi-supervised support vector machine (SVM) with segmentation-based ensemble (S^{2}SVMSE) algorithm is proposed for hyperspectral image classification. The algorithm utilizes spatial information extracted by a segmentation algorithm for unlabeled sample selection. The unlabeled samples that are the most similar to the labeled ones are found and the candidate set of unlabeled samples to be chosen is enlarged to the corresponding image segments. To ensure the finally selected unlabeled samples be spatially widely distributed and less correlated, random selection is conducted with the flexibility of the number of unlabeled samples actually participating in semi-supervised learning. Classification is also refined through a spectral–spatial feature ensemble technique. The proposed method with very limited labeled training samples is evaluated via experiments with two real hyperspectral images, where it outperforms the fully supervised SVM and the semi-supervised version without spectral–spatial ensemble.

© 2014 International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). Published by Elsevier B.V. All rights reserved.

1. Introduction

Hyperspectral sensors provide abundant spectral information with hundreds of spectrally continuous bands, and the collected hyperspectral imagery can be used to conduct finer classification that could not be achieved by traditional multispectral imagery (Bioucas-Dias et al., 2013; Du et al., 2012; Mountrakis et al., 2011; Tan and Du, 2008, 2011; Tan et al., 2013; van der Meer et al., 2012). Over the past several decades, novel pattern recognition methods have been widely used to remote sensing image processing tasks, such as supervised classification (Li et al., 2013b; Zhong and Zhang, 2012), unsupervised classification (Villa et al., 2013; Yang et al., 2010), feature extraction (Du et al., 2008; Du and Yang, 2008; Jia et al., 2013), target identification (Du and Zhang, 2014; Manolakis et al., 2001), and change detection (Bioucas-Dias and Nascimento, 2008; Camps-Valls et al., 2008). support vector machines (SVM), as one of the research focuses of machine learning, has attracted more attention in remote sensing (Camps-Valls et al., 2006; Fauvel et al., 2008; Melgani and Bruzzone, 2004; Plaza et al., 2009). SVM has advantages such as less

rigid requirements for prior knowledge and training samples, fitness to high-dimensional data, and more robustness to noise (Du et al., 2010; Jain et al., 2000). However, it is often difficult for a traditional SVM classifier to offer satisfactory performance in hyperspectral image classification. Specifically, for a high-dimensional hyperspectral image with a limited number of training samples, classification accuracy usually is significantly decreased due to the Hughes phenomenon (i.e., for a limited number of training samples, classification accuracy is decreased with the feature dimension being increased) (Hughes, 1968). Moreover, hyperspectral imagery is usually short of training sets, because sample collection generally involves extensive and time-consuming fieldwork (Dopido et al., 2013; Li et al., 2013a).

A relevant advanced solution to this classification challenge is the introduction of semi-supervised learning (SSL) techniques (Camps-Valls et al., 2007; Chapelle et al., 2008). SSL utilizes a large number of unlabeled data together with the available labeled data, to build classifiers that may be stronger than usual. SSL also requires less human efforts in sample collection, so it is of great importance in practical applications. The research on SLL has experienced a quick development in the past few years, and some popular methods have been proposed, such as generative mixture models (Krithara et al., 2011), self-learning models (Dopido et al.,

* Corresponding author. Tel.: +86 15905159291.
   E-mail address: dupjrs@126.com (P. Du).

2013), multiview learning models (Culp et al., 2009; Sun and Shawe-Taylor, 2010), transductive support vector machines SVMs (TSVMs) (Bruzzone et al., 2006; Ratle et al., 2010), and graph-based methods (Bai et al., 2013; Camps-Valls et al., 2007).

In particular, semi-supervised SVM (S^3VM) has been successfully applied to remote sensing image classification. Marconcini proposed a novel composite S^3VM for the spectral–spatial classification of hyperspectral images, which can significantly improve classification accuracy compared to both supervised SVMs and progressive S^3VM (Marconcini et al., 2009). Munoz-Mari proposed an algorithm to modify the OC-SVM kernel by modeling the data marginal distribution with a data graph Laplacian (Munoz-Mari et al., 2010). Gu successfully used Laplacian SVM for semi-supervised hyperspectral image classification (Gu and Feng, 2013). A novel approach for semi-supervised learning that adapts available active learning methods to a self-learning framework was presented by Dopido et al. (2013). The aforementioned SSL methods improve classification performance by different strategies. However, as the number of training samples is increased, it may be unbalanced and unbearable for a classifier to correctly exploit all the proper training samples due to computational issues. Furthermore, if a mislabeled sample is added for training, it may degrade classification accuracy. Thus, a semi-supervised algorithm should be designed in a computationally efficient fashion with the objective that classification performance can be truly improved with the use of additional samples.

In this paper, we evaluate the feasibility of adapting spatial information into an SSL paradigm, in which the S^2SVMSE itself selects accurate unlabeled samples for training. It makes use of spatial information in selecting new unlabeled training samples and generating classifier ensemble. It is assumed that samples in a homogeneous spatial segment mostly likely belong to the same class, which helps generate a smooth classification map; meanwhile, samples in the same segment also exhibit within-class variations, which can be useful for classifier training. Up to now, few studies that employ spatial information to select unlabeled samples are reported in the literature. Shi et al. proposed a semi-supervised dimension reduction algorithm, where unlabeled samples were selected based on multilevel segmentation results and t spectrally nearest neighbors similarity measure, and it notes that the number of the neighbors t controls the accuracy of similarity measurement, so they recommended to set a larger t to relax the constraint (Shi et al., 2013). However, it could not perform perfectly with limited or few labeled samples, so α-one-nearest neighbor is proposed to take the action of similarity measurement when labeled samples are very limited. In this work, the mean shift algorithm is used for image segmentation and spatial feature extraction, and a strongly constrained algorithm called α-one-nearest neighbor is applied for integrating spatial information when selecting new training samples. As a result, our proposed approach combines the spectral–spatial information in the SSL strategy.

The paper is organized as follows. In Section 2, the mean shift based segmentation is briefly reviewed, and then the proposed approach is introduced. Section 3 demonstrates the experimental results. Finally, Section 4 concludes with some remarks.

2. Proposed method

Let $X = (x_1, x_2, \cdots, x_l, x_{l+1}, \cdots, x_n) \in R^d$ denote a hyperspectral image with n pixels and d bands. Let $\kappa = \{1, \cdots, k\}$ be a set of k class labels, $y = (y_1, y_2, \cdots, y_l)$ be the labels of l labeled samples $X_L = \{x_i\}_{i=1}^l$, and $X_U = \{x_i\}_{i=l+1}^n$ be $(n-l)$ unlabeled samples. To perform the proposed semi-supervised approach, a hyperspectral imagery is first segmented using the mean shift method, and then unlabeled samples are selected for training based on the proposed

strategy in Section 2.2, finally, classification result is generated based on a spectral–spatial feature ensemble method in Section 2.3. Note that during the processes of unlabeled sample selection and classification result ensemble, spatial information in image segments is considered.

2.1. Mean shift

In this paper, spatial information is used during the selection of unlabeled samples and feature ensemble for classification. Here, the image segmentation method of mean shift (MS) is employed to extract spatial information. MS method is a nonparametric clustering technique without requiring prior knowledge of the number and shape for each cluster. It defines an empirical probability density function, and the modes of densest regions in the space can be estimated by finding the local maxima of the probability density function that is estimated by kernel density estimation method (Comaniciu and Meer, 2002). Once the location of modes is determined, the cluster associated with it is delineated based on the local structure in the feature space. Here, we briefly review the basics of mean shift segmentation (Comaniciu and Meer, 2002).

With the notation defined, given n pixels x_i, $i = 1, \cdots, n$ in the d-dimensional space R^d, the kernel density estimator computed in pixel x is given by,

$$f(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \tag{1}$$

where h is the bandwidth. Here, the kernel $K(x)$ is given by $K(x) = c_{k,d} k(\|x\|^2)$, where the function $k(x)$ is called kernel profile, and the normalization constant $c_{k,d}$ makes $K(x)$ have an integral of one. Defining the function $g(x) = -k'(x)$ with the assumption that the derivative of the kernel profile $k(x)$ exists for all $x \in [0, \infty)$. Using $g(x)$ as the kernel profile, the kernel $G(x)$ is defined as $G(x) = c_{g,d} g(\|x\|^2)$, where $c_{g,d}$ is a normalization constant. The goal of the defined density is to find the modes of densest regions in the space. The modes are located among the zeros of gradient $\nabla f(x) = 0$. Thus, the mean shift vector is defined as:

$$m_{h,G}(x) = \frac{1}{2} h^2 c \frac{\nabla f_{h,K}(x)}{f_{h,G}(x)} \tag{2}$$

$$m_{h,G}(x) = \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x - x_i}{h}\right\|^2\right)} - x \tag{3}$$

For a hyperspectral image, it is typically represented as a two-dimensional lattice of d-dimensional vectors. Taking its spatial and spectral information into the mean shift procedure, the spatial and spectral information are represented in the spatial and range domains, respectively, and the multivariate kernel is defined as the product of these two radially symmetric kernels as:

$$K_{h_s, h_r}(x) = \frac{C}{h_s^2 h_r^d} k\left(\left\|\frac{x^s}{h_s}\right\|^2\right) k\left(\left\|\frac{x^r}{h_r}\right\|^2\right) \tag{4}$$

where x^s is the spatial part; x^r is the range part of a feature vector; h_s and h_r are the kernel bandwidths for spatial and range domains, respectively, and c is the normalization constant. MS cannot be used for high-dimensional data (Comaniciu and Meer, 2002; Georgescu et al., 2003), so dimensionality reduction has to be applied first. Here, principal component analysis is used for this purpose.

2.2. Selection of unlabeled samples

In contrast to supervised classification algorithms, semi-supervised methods generally assume that enlarging the training

set with unlabeled samples will improve the performance. However, several requirements have to be considered. Firstly, the selected samples should be labeled without significant error. Secondly, the number of the selected unlabeled training samples should be proper in order not to dramatically increase computational cost. Therefore, it is important to efficiently identify the most informative unlabeled samples for training so that improvement can be achieved without using a large number of unlabeled samples.

In this work, we employ one of the strongest constrained algorithms, called α-one-nearest neighbor, to identify the most informative unlabeled samples, and combine spatial information to enlarge the candidate set of unlabeled samples, then randomly selection was employed to select widely distributed unlabeled samples for training. The details are described as follows:

Let R_j, $j = 1, \cdots, m$ donate the j-th segments generated by the MS segmentation procedure, and m is the number of total segments. A similarity metric to measure the distance between the samples in R_j and the labeled samples X_L is defined as:

$$D(R_j, X_L) = \min\{d(x, x_i) | x \in R_j, \ x_i \in X_L\} \quad (5)$$

where $d(x, x_i)$ is the distance of the two samples (x, x_i). Here, spectral angle is used as similarity metric. In order to find the most similar samples in R_j, $D(R_j, X_L)$ must meet the following criterion,

$$D(R_j, X_L) < \alpha \quad (6)$$

where α is the threshold. If the distance between R_j and X_L meets the above criterion when the labeled sample is x_i, we label all the samples in R_j with the same label as $C(x_i)$ which is the given label of x_i. Each segment experiences this procedure, yielding a candidate set of unlabeled samples that may be used for training. However, there are a large number of unlabeled samples in this candidate set; if all the unlabeled samples from this candidate set are used for classification, the complexity of the learning process is dramatically increased. Also, many samples in this candidate set may be spatially clustered together and highly correlated. Thus, random selection is applied to select a smaller number of unlabeled samples. Of course, the number of samples in the candidate set can also be controlled by α in Eq. (6). Then the SVM is trained by both labeled and selected unlabeled samples.

2.3. Spectral–spatial feature ensemble

To take advantage of spatial smoothness in segments for further classification improvement, a spectral–spatial feature ensemble algorithm, called classification voting, is applied to generate final classification result. The class label of the samples in each densest region is assigned via majority voting as:

$$C(R_j) = \arg \max_{\kappa = \{1, \cdots, k\}} \frac{V_{R_j}(k)}{V_{R_j}} > r \quad (7)$$

where $C(R_j)$ is the final class label of the densest region R_j, $V_{R_j}(k)$ is the number of samples labeled as class k within R_j, V_{R_j} is the number of all samples within R_j, and r is a constraint parameter. For a segment that cannot satisfy the inequality in Eq. (7), the class labels of all samples in such a region remain unchanged.

2.4. Proposed semi-supervised approach

in this paper, segmentation results are used to exploit spatial information with the assumption that pixels in the same region may belong to the same class with high possibility. If an unlabeled sample is very similar to a labeled sample, then all the pixels in its segment are assigned to the class of the labeled sample. However, it may be computationally expensive if a large number of resulting

Table 1
The proposed semi-supervised algorithm (S^{2}SVMSE).

Input:
• A set of labeled training samples X_L
• A set of unlabeled samples X_u
• The kernel bandwidths for spatial and range h_s and h_r applied in MS process
• The threshold of Eq. (5) α, and the constraint parameter of Eq. (6) r
Output: the ensemble classification result for all samples $X_L \cup X_U$
Steps:
1: Extract spatial information based on the mean shift method given by Eqs. (1)–(4)
2: Extract the unlabeled samples candidate set SL' based on the strategy described by Eqs. (5) and (6)
3: Randomly select a smaller number of unlabeled samples from SL' as the actually added unlabeled samples SL for training
4: Construct the training samples set $X_L \cup SL$
5: Train the SVM classifier h with training samples set $X_L \cup SL$
6: Generate the classification result of all samples $X_L \cup X_U$ using h
7: Refine the classification result by Eq. (7)

unlabeled samples are used to train the classifier. Also, the resulting unlabeled samples may be spatially clustered together and highly correlated. To decrease computational complexity and to use widely distributed unlabeled samples, random selection is performed to generate a smaller subset for semi-supervised learning. The procedure of the proposed algorithm is illustrated in Table 1.

3. Experiments

In this section, two real hyperspectral images are used to evaluate the proposed approach. In order to illustrate the performance of the proposed approach, we utilized a very small labeled training set on purpose, such as 5, 10 or 15 per class. In order to analyze the effect on using different number of unlabeled samples, the unlabeled samples are randomly selected in proportions (e.g., 5%, 10%, 20%, etc.) for training. In all cases, overall accuracy (OA) are obtained by the proposed semi-supervised SVM without segmentation-based ensemble (S^{2}SVM), and the results of proposed semi-supervised SVM with segmentation-based ensemble (S^{2}SVMSE) with different constraint parameter r (e.g., 1/2, 2/3, 3/4 and 4/5).

3.1. Data used in the experiments

The first data was collected by the ROSIS sensor in 2003, with spectral coverage ranging from 0.43 to 0.86 μm and with 610×340 pixels. The data is about University of Pavia and was atmospherically corrected. It has spatial resolution of 1.3 m with 115 spectral bands, and 103 bands were used in the experiment after low signal-to-noise ratio (SNR) and water absorption bands being removed. The image scene in pseudocolor is shown in Fig. 1(a). The classification problem involves nine land cover types. The first four principal components of ROSIS data set was used for the MS based segmentation process with the given kernel bandwidths ($h_s = 8$, $h_r = 15$), and the segmentation result is shown in Fig. 1(b). During the selection of unlabeled samples, the parameter α is set to 0.01. The labeled training samples (5, 10 or 15 for each class) were randomly selected from the original training set, and the map of testing samples was shown in Fig. 1(c).

The second hyperspectral image was acquired in 1992 over the Indian Pines region in Northwestern Indiana by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor. It comprises 202 spectral channels in the wave-length range from 0.4 to 2.5 μm, and 195 channels were used in the experiment after low SNR and water absorption bands being removed. The spatial resolution is 20 m. The image scene in pseudocolor is shown in Fig. 2(a). The first four principal components of AVIRIS data set was selected for segmentation with the given kernel bandwidths
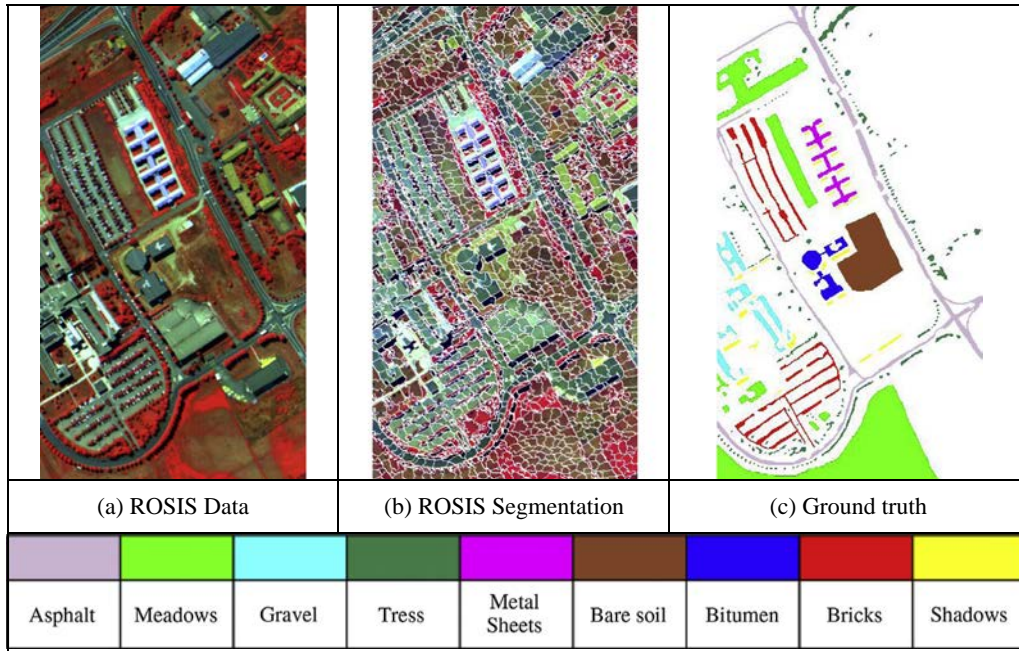
| (a) ROSIS Data | (b) ROSIS Segmentation | (c) Ground truth |

| Asphalt | Meadows | Gravel | Tress | Metal Sheets | Bare soil | Bitumen | Bricks | Shadows |

Fig. 1. Details about the ROSIS data set.

| (a) AVIRIS Indian Pines Data | (b) AVIRIS Segmentation | (c) Ground truth |

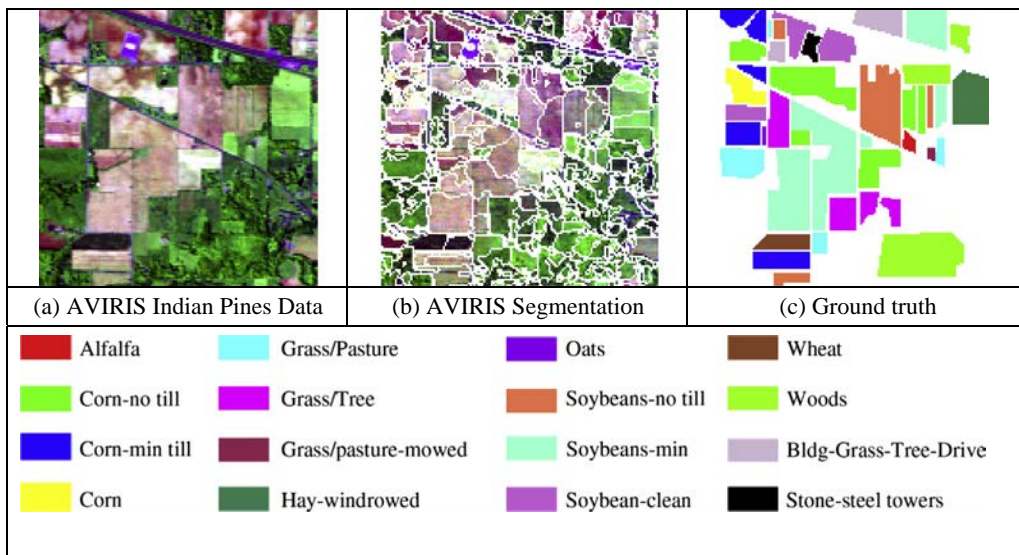| Alfalfa | Grass/Pasture | Oats | Wheat |
Corn-no till	Grass/Tree	Soybeans-no till	Woods
Corn-min till	Grass/pasture-mowed	Soybeans-min	Bldg-Grass-Tree-Drive
Corn	Hay-windrowed	Soybean-clean	Stone-steel towers

Fig. 2. Details about the AVIRIS Indian Pines data set.

(h_s = 10, h_r = 17), and the segmentation result is shown in Fig. 2(b). During the selection of unlabeled samples, the parameter α is set to 0.008. The labeled training samples (5, 10 or 15 for each class) were randomly selected from the testing set, and the map of testing samples was shown as Fig. 2(c).

3.2. Results for ROSIS data

Tables 2–4 report the classification scores achieved by the S^{2}SVM and S^{2}SVMSE algorithms for the full ROSIS data, where the values of overall accuracy (OA) are displayed for different numbers of labeled samples (5, 10, 15 per class) and unlabeled samples selected in varied proportions.

As shown in Fig. 3, the OA values are increased when using more unlabeled samples. It indicates that the proposed semi-supervised

approach can significantly improve classification accuracy, in comparison with the supervised cases. In this experiment, the semi-supervised methods (i.e., S^{2}SVM and S^{2}SVMSE) perform similarly, and the S^{2}SVMSE method is slightly better than S^{2}SVM. It demonstrates the positive effect of using spatial information for classification process. Finally, it can also be observed that the number of labeled samples influences the best accuracy that can be achieved by the proposed algorithms. As the experimental results shown, the more labeled samples used initially, the better results achieved finally. However, the proposed approach can provide the improvement with very few labeled samples (i.e., 5 samples per class).

In order to show the classification results in more details, Table 2 lists the OA values in percentage by the proposed approach (S^{2}SVM and S^{2}SVMSE) with 5 labeled samples per class, and Tables 3 and 4 shows the results with 10 and 15 labeled

Table 2
OA results for ROSIS data using proposed algorithms with 5 labeled samples per class. The best OA results of row was marked bold. The best OA results of each table was marked bold italics.

Percentage of candidate set		0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
S^2SVM	OA	28.72	61.61	65.97	68.29	71.76	72.92	75.26	76.22	**78.49**	75.71	76
	Kappa	0.197	0.524	0.573	0.602	0.643	0.657	0.684	0.693	**0.723**	0.691	0.697
$r(1/2)$	OA	30.35	67.93	71.71	72.39	77.88	79.51	80.55	81.81	**85.34**	81.44	82.96
	Kappa	0.214	0.597	0.641	0.648	0.717	0.735	0.748	0.763	**0.808**	0.762	0.782
$r(2/3)$	OA	30.37	67.78	71.78	72.43	77.99	79.61	80.66	81.87	**85.50**	81.48	83.00
	Kappa	0.215	0.595	0.642	0.648	0.719	0.736	0.750	0.763	**0.810**	0.762	0.782
$r(3/4)$	OA	30.40	67.58	71.99	72.44	78.22	79.68	80.81	81.99	***85.64***	81.59	82.79
	Kappa	0.215	0.593	0.644	0.648	0.721	0.737	0.752	0.765	***0.812***	0.764	0.779
$r(4/5)$	OA	30.40	67.68	71.97	72.60	78.19	79.56	80.84	82.00	**85.56**	81.54	82.86
	Kappa	0.215	0.594	0.644	0.651	0.721	0.736	0.752	0.765	**0.811**	0.763	0.780

Table 3
OA results for ROSIS data using proposed algorithms with 10 labeled samples per class. The best OA results of row was marked bold. The best OA results of each table was marked bold italics.

Percentage of candidate set		0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
S^2SVM	OA	63.02	76.34	82.09	82.88	83.49	83.72	84.66	84.35	**85.40**	84.95	83.90
	Kappa	0.538	0.687	0.761	0.774	0.782	0.786	0.796	0.795	**0.807**	0.802	0.788
$r(1/2)$	OA	68.78	81.69	89.41	89.02	89.90	88.89	89.98	89.95	89.69	**90.86**	89.81
	Kappa	0.603	0.752	0.857	0.854	0.865	0.852	0.865	0.866	0.862	**0.877**	0.864
$r(2/3)$	OA	68.60	81.74	89.40	89.05	89.94	88.82	89.94	89.97	89.74	**90.86**	89.88
	Kappa	0.601	0.753	0.857	0.853	0.865	0.851	0.865	0.867	0.863	**0.877**	0.865
$r(3/4)$	OA	68.23	81.70	89.30	88.89	89.84	88.91	89.91	89.93	89.73	**90.93**	89.84
	Kappa	0.597	0.752	0.856	0.851	0.864	0.853	0.864	0.866	0.863	**0.878**	0.865
$r(4/5)$	OA	67.83	81.77	89.27	88.96	89.84	88.93	89.97	89.99	89.77	***90.94***	89.86
	Kappa	0.592	0.753	0.856	0.852	0.864	0.853	0.865	0.867	0.863	***0.878***	0.865

Table 4
OA results for ROSIS data using proposed algorithms with 15 labeled samples per class. The best OA results of row was marked bold. The best OA results of each table was marked bold italics.

Percentage of candidate set		0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
S^2SVM	OA	72.51	84.48	84.55	86.41	87.30	86.74	86.12	87.35	87.45	**88.02**	87.74
	Kappa	0.651	0.796	0.798	0.821	0.833	0.826	0.819	0.834	0.836	**0.843**	0.839
$r(1/2)$	OA	84.20	94.03	94.65	94.59	**96.62**	94.46	93.34	95.89	95.64	95.34	95.61
	Kappa	0.794	0.920	0.929	0.928	**0.955**	0.926	0.912	0.945	0.942	0.938	0.942
$r(2/3)$	OA	84.15	94.00	94.64	94.55	***96.64***	94.58	93.16	95.83	95.81	95.37	95.65
	Kappa	0.793	0.920	0.929	0.927	***0.955***	0.928	0.910	0.945	0.944	0.938	0.942
$r(3/4)$	OA	83.70	93.94	94.63	94.58	**96.52**	94.59	93.08	95.79	95.53	95.37	95.50
	Kappa	0.787	0.919	0.929	0.928	**0.954**	0.928	0.909	0.944	0.941	0.939	0.940
$r(4/5)$	OA	83.50	93.73	94.20	94.63	**96.51**	94.54	93.06	95.67	95.35	95.31	95.61
	Kappa	0.785	0.916	0.923	0.929	**0.954**	0.927	0.909	0.942	0.938	0.938	0.942

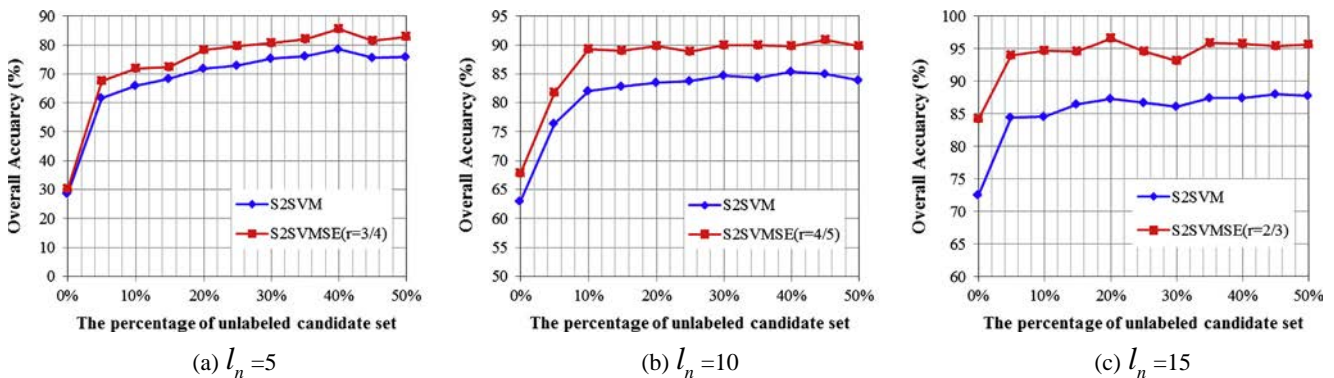(a) $l_n = 5$ · (b) $l_n = 10$ · (c) $l_n = 15$

Fig. 3. Overall classification accuracies obtained for the ROSIS data set using S^2SVM and S^2SVMSE semi-supervised methods by using 5, 10 or 15 labeled samples per class.

samples per class, respectively. In the tables, 0% means the classification results are obtained by only using labeled samples. It is noticeable that, by including unlabeled samples and spatial information, the classification results are greatly improved in all cases. For example, with $l_n = 10$, the supervised classification accuracy is only 63.02% as shown in Table 3, while the highest
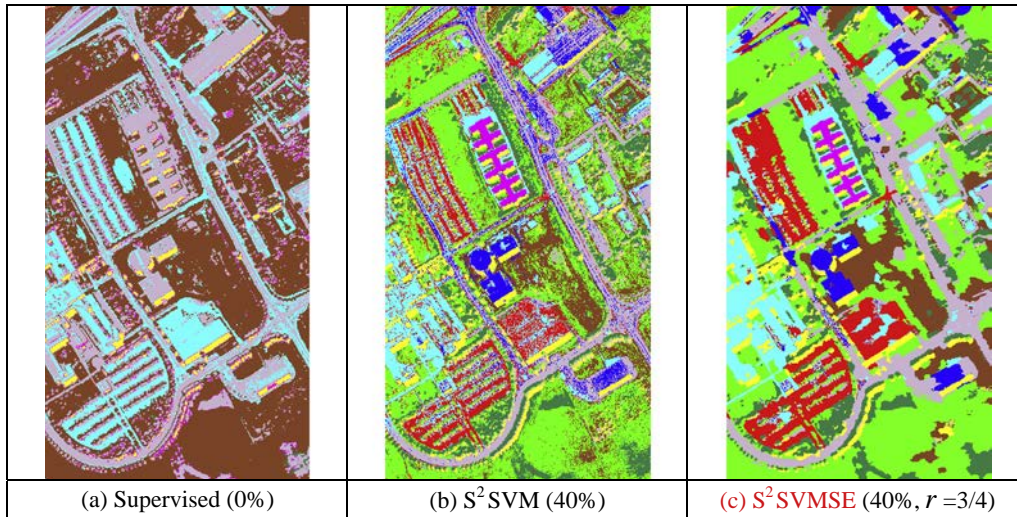
(a) Supervised (0%)   (b) S^{2}SVM (40%)   (c) S^{2}SVMSE (40%, r =3/4)

Fig. 4. Classification maps for all the methods with the ROSIS data set using 5 labeled samples per class.

(a) Supervised (0%)   (b) S^{2}SVM (40%)   (c) S^{2}SVMSE (45%, r =4/5)
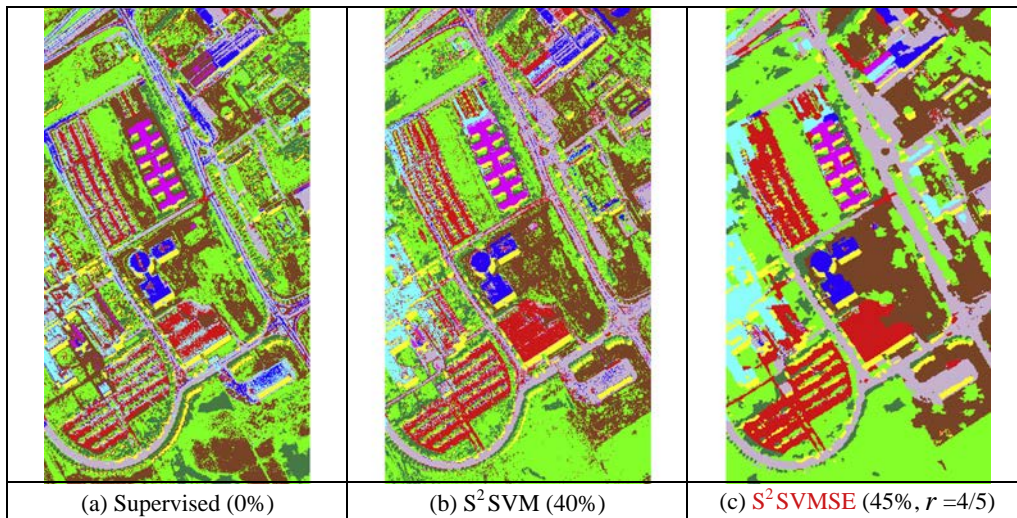
Fig. 5. Classification maps for all the methods with the ROSIS data set using 10 labeled samples per class.

accuracy (85.40%) is achieved by the semi-supervised approach, and the better result (90.94%) is obtained by further including spatial information. With consideration of spectral–spatial ensemble parameter (r), it can be found that a stricter requirement could overcome the over-segmentation problem in comparison with majority voting based only.

Figs. 4–6 show the classification maps obtained by all methods for ROSIS data set. It is noticeable that the classification maps obtained by semi-supervised methods have better performance in the confusing areas, especially in the regions dominated by mixed classes of meadows and bare soil.

3.3. Results for AVIRIS Indian Pines data

In this experiment, we evaluated the performance of the proposed algorithm with AVIRIS Indian Pines data. Similarly, the varieties of OAs obtained by the proposed methods are shown in Fig. 7. Given the initial labeled samples (5, 10 or 15 per class), the proposed semi-supervised algorithms greatly improved the accuracy in comparison with the supervised one. Tables 5–7 summarize the accuracies produced by the proposed approach (S^{2}SVM

and S^{2}SVMSE). With l_n = 10, the OA obtained by the supervised method is 60.61%, it reaches 83.08% when including 40% samples of unlabeled candidate set, and the highest accuracy is achieved by integrating the spatial information. Again, Tables 6 and 7 show the similar results. First of all, the proposed semi-supervised method can achieve more competitive results compared to the supervised one with limited labeled samples. Furthermore, the spectral–spatial-based method clearly obtained the best results in all cases. Figs. 8–10 show the classification maps. It can be seen that the proposed semi-supervised method outperforms, providing smoother classification maps than others.

3.4. Parameter analysis

The AVIRIS Indian Pines data set was used to assess the influence of the parameters in the proposed algorithm. According to the framework described in Section 2, there are four tuning parameters: h_s and h_r are the basic parameters in the mean shift method, α is the threshold in unlabeled sample selection process, and parameter r adjusts the influence of over-segmentation in classification voting.

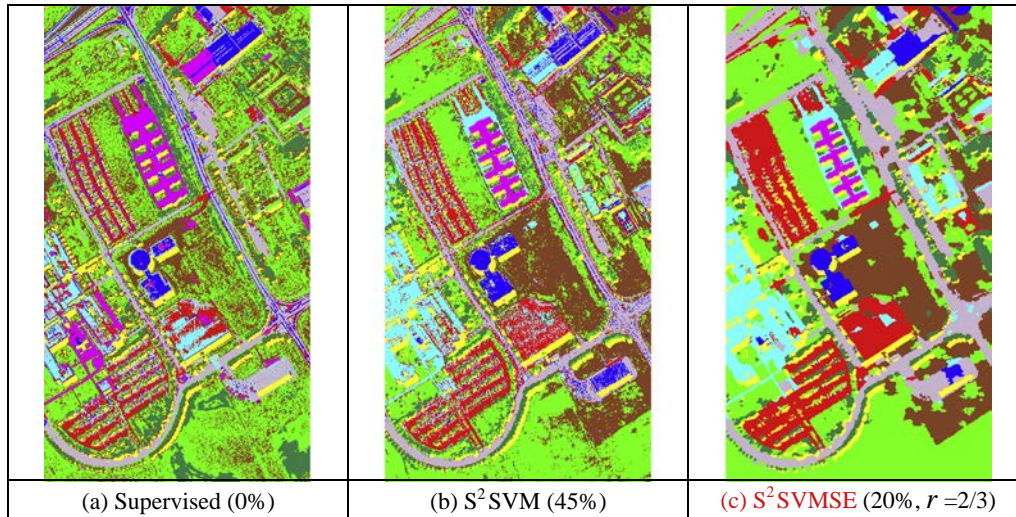| (a) Supervised (0%) | (b) S^{2}SVM (45%) | (c) S^{2}SVMSE (20%, r =2/3) |

Fig. 6. Classification maps for all the methods with the ROSIS data set using 15 labeled samples per class.

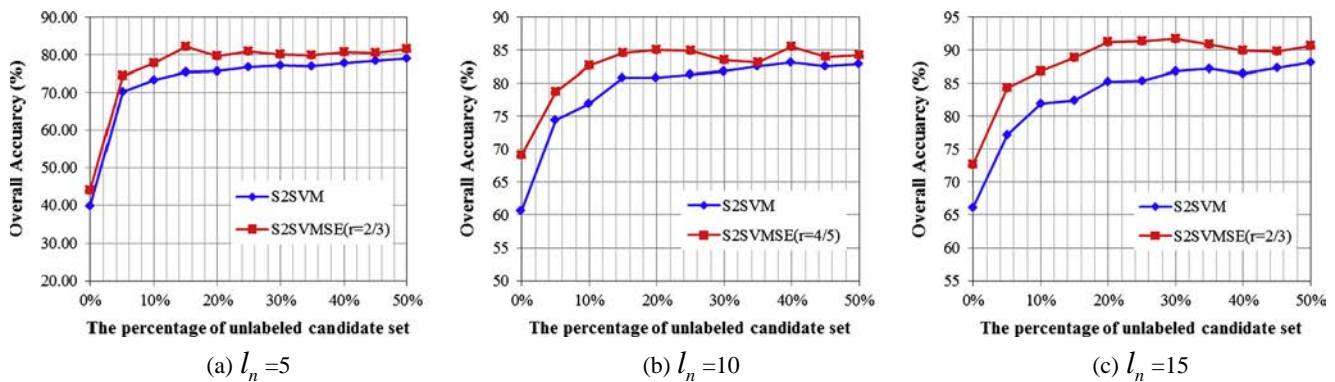(a) l_n =5　　　　　　　　　(b) l_n =10　　　　　　　　(c) l_n =15

Fig. 7. Overall classification accuracies obtained for the AVIRIS Indian Pines data set using S^{2}SVM and S^{2}SVMSE semi-supervised methods by using 5, 10 or 15 labeled samples per class.

Table 5
OA results for AVIRIS Indian Pines data using proposed algorithms with 5 labeled samples per class. The best OA results of row was marked bold. The best OA results of each table was marked bold italics.

Percentage of candidate set		0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
S^{2}SVM	OA	39.79	70.17	73.29	75.55	75.66	76.68	77.16	76.91	77.73	78.35	**78.96**
	Kappa	0.328	0.662	0.697	0.723	0.724	0.735	0.741	0.739	0.748	0.755	**0.762**
r(1/2)	OA	43.64	76.02	78.37	**82.21**	80.58	80.33	81.20	80.71	80.36	80.06	81.29
	Kappa	0.369	0.727	0.755	**0.798**	0.779	0.777	0.787	0.781	0.778	0.775	0.788
r(2/3)	OA	43.93	74.59	78.23	***82.33***	80.73	80.44	80.47	79.87	80.32	80.08	81.23
	Kappa	0.370	0.711	0.753	***0.800***	0.781	0.778	0.779	0.772	0.777	0.775	0.787
r(3/4)	OA	44.42	74.39	77.60	**82.17**	80.32	80.94	80.19	79.81	80.33	80.07	81.24
	Kappa	0.376	0.709	0.746	**0.798**	0.776	0.783	0.776	0.771	0.777	0.775	0.787
r(4/5)	OA	44.05	74.51	77.68	**82.17**	79.68	80.89	80.04	79.84	80.63	80.54	81.45
	Kappa	0.371	0.710	0.747	**0.798**	0.769	0.783	0.774	0.772	0.781	0.780	0.790

The parameter α was defined by assessing the similarity between different classes based on labeled samples, and its value should be less than the minimum of pair-wise distances between classes. In the experiments, α was 0.008. Based on previous research on mean shift, the optimal choice of h_s and h_r is still an open question. So we empirically changed h_s within [4, 12] and h_r within [13, 21] to find sub-optimal values. The classification rates with respect to h_s and h_r with 45% of selected unlabeled samples are shown in Fig. 11.

As shown in Fig. 11, when h_s is small, the segmentation result with obvious over-segmentation may increase the number of mislabeled samples. Therefore, the OA is lower, but a larger h_r can slightly improve the accuracy. With the increase of h_s, the over-segmentation phenomenon decreases, and the accuracy is increased because of less mislabeled samples being selected for training. However, when h_s reaches 10, there is a peak in the accuracy curve. As h_s continues to increase, the accuracy begins to fall off slowly. In this case, the number of selected unlabeled samples may be reduced. For the parameter h_r, the accuracy is increasing with increase of h_r when h_s is small. However, the influence due to the change of h_r is weakened when h_s reaches a larger value.
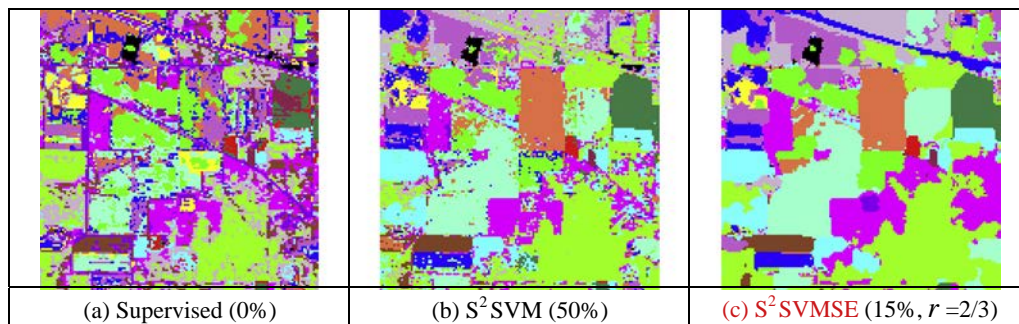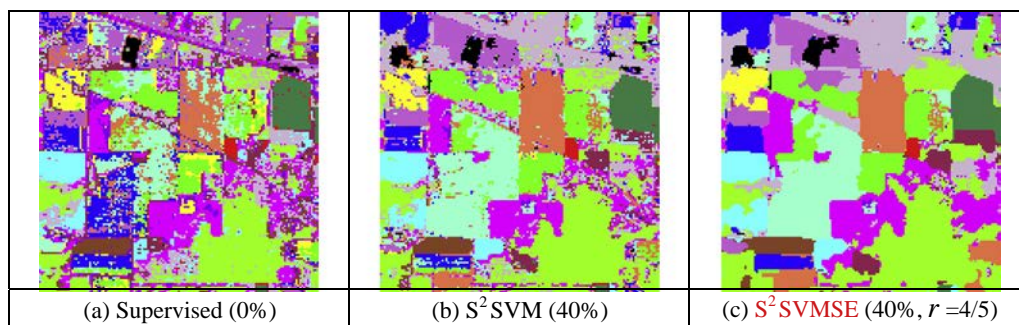
Table 6
OA results for AVIRIS Indian Pines data using proposed algorithms with 10 labeled samples per class. The best OA results of row was marked bold. The best OA results of each table was marked bold italics.

Percentage of candidate set		0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
S^2SVM	OA	60.61	74.45	76.97	80.87	80.83	81.34	81.82	82.53	**83.08**	82.55	82.94
	Kappa	0.563	0.711	0.740	0.784	0.784	0.790	0.795	0.803	**0.809**	0.803	0.808
$r(1/2)$	OA	69.46	78.63	82.06	84.19	85.54	84.35	83.30	82.79	**85.41**	83.97	83.83
	Kappa	0.660	0.757	0.797	0.821	0.837	0.824	0.812	0.806	**0.836**	0.818	0.818
$r(2/3)$	OA	69.13	78.64	82.06	84.80	85.57	84.47	83.25	82.70	**85.44**	84.34	84.17
	Kappa	0.657	0.757	0.797	0.828	0.837	0.825	0.811	0.805	**0.836**	0.823	0.822
$r(3/4)$	OA	69.33	78.39	82.23	84.46	84.96	84.86	83.47	83.14	**85.47**	83.90	84.19
	Kappa	0.659	0.754	0.799	0.824	0.831	0.829	0.813	0.810	**0.836**	0.819	0.822
$r(4/5)$	OA	69.06	78.78	82.60	84.51	85.02	84.86	83.54	83.16	***85.47***	84.01	84.22
	Kappa	0.656	0.759	0.803	0.825	0.831	0.829	0.814	0.810	***0.836***	0.820	0.822

Table 7
OA results for AVIRIS Indian Pines data using proposed algorithms with 15 labeled samples per class. The best OA results of row was marked bold. The best OA results of each table was marked bold italics.

Percentage of candidate set		0%	5%	10%	15%	20%	25%	30%	35%	40%	45%	50%
S^2SVM	OA	66.02	77.20	81.99	82.46	85.28	85.43	86.86	87.12	86.50	87.31	**88.14**
	Kappa	0.622	0.741	0.796	0.802	0.833	0.835	0.851	0.854	0.847	0.856	0.866
$r(1/2)$	OA	72.05	84.22	86.87	88.76	91.07	91.12	**91.62**	90.86	89.88	89.77	90.57
	Kappa	0.689	0.822	0.850	0.873	0.899	0.899	**0.905**	0.896	0.885	0.884	0.893
$r(2/3)$	OA	72.59	84.34	86.87	88.83	91.22	91.31	***91.72***	90.93	89.95	89.78	90.66
	Kappa	0.695	0.823	0.850	0.874	0.900	0.901	***0.906***	0.897	0.886	0.884	0.894
$r(3/4)$	OA	72.55	84.12	86.97	88.87	91.35	91.32	**91.53**	91.20	90.08	89.86	90.68
	Kappa	0.694	0.816	0.851	0.874	0.902	0.901	**0.904**	0.900	0.888	0.885	0.894
$r(4/5)$	OA	72.54	83.92	86.15	88.84	91.33	91.42	**91.41**	91.20	90.21	89.99	90.73
	Kappa	0.694	0.808	0.842	0.874	0.902	0.903	**0.903**	0.900	0.889	0.887	0.895

| (a) Supervised (0%) | (b) S^2SVM (50%) | (c) S^2SVMSE (15%, r =2/3) |

Fig. 8. Classification maps for all the methods with the AVIRIS Indian Pines data set using 5 per class labeled samples.

| (a) Supervised (0%) | (b) S^2SVM (40%) | (c) S^2SVMSE (40%, r =4/5) |

Fig. 9. Classification maps for all the methods with the AVIRIS Indian Pines data set using 10 per class labeled samples.

The value of α is used to control the similarity between unlabeled and labeled samples. To assess sensitivity of α, we set $h_s = 10$ and $h_r = 17$ to achieve the best segmentation result, and then randomly select 45% of unlabeled samples in the candidate set. When the value of α is small, there are a small quantity of unlabeled samples to be selected for training, the classification cannot reach the highest accuracy. However, when α is large, some mislabeled samples may be selected, so the accuracy will decrease.

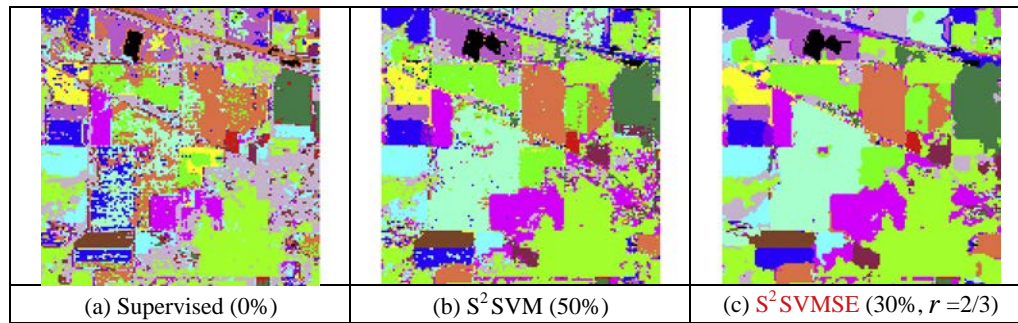| (a) Supervised (0%) | (b) S^{2}SVM (50%) | (c) S^{2}SVMSE (30%, r =2/3) |

Fig. 10. Classification maps for all the methods with the AVIRIS Indian Pines data set using 15 labeled samples per class.
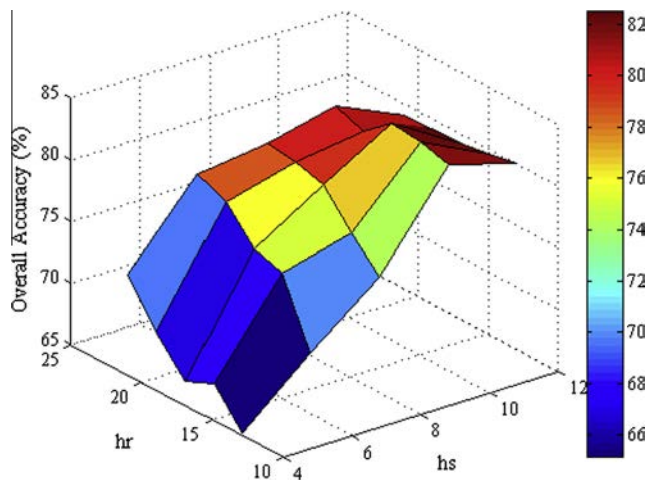

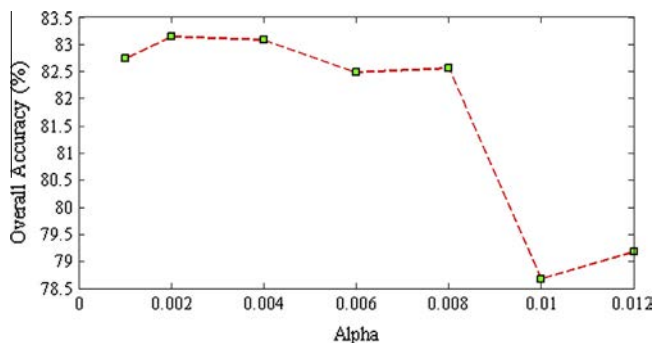**Fig. 11.** Classification rates with respect to parameter h_s and h_r.

Fig. 12. Sensitivity analysis of α on the AVIRIS Indian Pines data set.

Fig. 12 shows that the accuracy rises slowly at first and then keeps stable; but when the value of α exceeds 0.008, it begins to descend quickly.

The value of r is used to adjust the influence of over-segmentation in classification voting. From the results shown in Tables 2–7, it has an effect on the final classification results. Compared with the traditional majority voting method, our improved majority voting method aims to alleviate the negative influence of serious over-segmentation. From the results based on different initial labeled samples, the best classification accuracies are all achieved when r is larger than 1/2, which indicates its role against over-segmentation.

4. Conclusion

In this paper, we have introduced an efficient semi-supervised classification approach for remotely sensed hyperspectral image classification. The unlabeled samples that are the most similar to the labeled ones are found by the α-one-nearest neighbor criterion, and then the candidate set of unlabeled samples is enlarged by utilizing the MS-based segmentation result. To ensure the finally selected unlabeled samples be spatially widely distributed and less correlated, random selection is conducted with the flexibility choice of the number of unlabeled samples actually participating in semi-supervised learning. After training the SVM with both labeled and selected unlabeled samples, the spectral–spatial ensemble method is applied to achieve better classification. Experimental results demonstrate that the proposed method offers a better performance in terms of classification accuracy and a very small numbers of labeled samples required. It should be noted that segmentation algorithms that can provide excellent performance for hyperspectral imagery may be employed to play the same role as Mean Shift in the proposed S^{2}SVMSE.

References

Bai, J., Xiang, S.M., Pan, C.H., 2013. A graph-based classification method for hyperspectral images. IEEE Trans. Geosci. Remote Sens. 51 (2), 803–817.

Bioucas-Dias, J., Plaza, A., Camps-Valls, G., Scheunders, P., Nasrabadi, N., Chanussot, J., 2013. Hyperspectral remote sensing data analysis and future challenges. Geosci. Remote Sens. Mag., IEEE 1 (2), 6–36.

Bioucas-Dias, J.M., Nascimento, J.M.P., 2008. Hyperspectral subspace identification. IEEE Trans. Geosci. Remote Sens. 46 (8), 2435–2445.

Bruzzone, L., Chi, M.M., Marconcini, M., 2006. A novel transductive SVM for semisupervised classification of remote-sensing images. IEEE Trans. Geosci. Remote Sens. 44 (11), 3363–3373.

Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Rojo-Alvarez, J.L., Martinez-Ramon, M., 2008. Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. IEEE Trans. Geosci. Remote Sens. 46 (6), 1822–1835.

Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Vila-Frances, J., Calpe-Maravilla, J., 2006. Composite kernels for hyperspectral image classification. IEEE Geosci. Remote Sens. Lett. 3 (1), 93–97.

Camps-Valls, G., Marsheva, T.V.B., Zhou, D.Y., 2007. Semi-supervised graph-based hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 45 (10), 3044–3054.

Chapelle, O., Sindhwani, V., Keerthi, S.S., 2008. Optimization techniques for semi-supervised support vector machines. J. Mach. Learn. Res. 9, 203–233.

Comaniciu, D., Meer, P., 2002. Mean shift: a robust approach toward feature space analysis. Pattern Anal. Mach. Intell., IEEE Trans. 24 (5), 603–619.

Culp, M., Michailidis, G., Johnson, K., 2009. On multi-view learning with additive models. Ann. Appl. Stat. 3 (1), 292–318.

Dopido, I., Li, J., Marpu, P.R., Plaza, A., Dias, J.M.B., Benediktsson, J.A., 2013. Semisupervised self-learning for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 51 (7), 4032–4044.

Du, B., Zhang, L.P., 2014. Target detection based on a dynamic subspace. Pattern Recognit. 47 (1), 344–358.

Du, P.J., Tan, K., Xing, X.S., 2010. Wavelet SVM in reproducing kernel hilbert space for hyperspectral remote sensing image classification. Opt. Commun. 283 (24), 4978–4984.

Du, P.J., Xia, J.S., Zhang, W., Tan, K., Liu, Y., Liu, S.C., 2012. Multiple classifier system for remote sensing image classification: a review. Sensors 12 (4), 4764–4792.

Du, Q., Raksuntorn, N., Younan, N.H., King, R.L., 2008. End-member extraction for hyperspectral image analysis. Appl. Opt. 47 (28), F77–F84.

Du, Q., Yang, H., 2008. Similarity-based unsupervised band selection for hyperspectral image analysis. IEEE Geosci. Remote Sens. Lett. 5 (4), 564–568.

Fauvel, M., Benediktsson, J.A., Chanussot, J., Sveinsson, J.R., 2008. Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles. IEEE Trans. Geosci. Remote Sens. 46 (11), 3804–3814.

Georgescu, B., Shimshoni, I., Meer, P., 2003. Mean shift based clustering in high dimensions: a texture classification example, Computer Vision, 2003. In: Proceedings, Ninth IEEE International Conference on. IEEE, pp. 456–463.

Gu, Y.F., Feng, K., 2013. Optimized Laplacian SVM with distance metric learning for hyperspectral image classification. IEEE J. Sel. Top. Appl. Earth Observations Remote Sens. 6 (3), 1109–1117.

Hughes, G., 1968. On the mean accuracy of statistical pattern recognizers. IEEE Trans. Inf. Theory 14 (1), 55–63.

Jain, A.K., Duin, R.P.W., Mao, J.C., 2000. Statistical pattern recognition: a review. IEEE Trans. Pattern Anal. Mach. Intell. 22 (1), 4–37.

Jia, X.P., Kuo, B.C., Crawford, M.M., 2013. Feature mining for hyperspectral image classification. Proc. IEEE 101 (3), 676–697.

Krithara, A., Amini, M.R., Goutte, C., Renders, J.M., 2011. Learning aspect models with partially labeled data. Pattern Recognit. Lett. 32 (2), 297–304.

Li, J., Bioucas-Dias, J.M., Plaza, A., 2013a. Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression. IEEE Geosci. Remote Sens. Lett. 10 (2), 318–322.

Li, J., Marpu, P.R., Plaza, A., Bioucas-Dias, J.M., Benediktsson, J.A., 2013b. Generalized composite kernel framework for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 51 (9), 4816–4829.

Manolakis, D., Siracusa, C., Shaw, G., 2001. Hyperspectral subpixel target detection using the linear mixing model. IEEE Trans. Geosci. Remote Sens. 39 (7), 1392–1409.

Marconcini, M., Camps-Valls, G., Bruzzone, L., 2009. A composite semisupervised SVM for classification of hyperspectral images. IEEE Geosci. Remote Sens. Lett. 6 (2), 234–238.

Melgani, F., Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans. Geosci. Remote Sens. 42 (8), 1778–1790.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: a review. ISPRS J. Photogramm. Remote Sens. 66 (3), 247–259.

Munoz-Mari, J., Bovolo, F., Gomez-Chova, L., Bruzzone, L., Camps-Valls, G., 2010. Semisupervised one-class support vector machines for classification of remote sensing data. IEEE Trans. Geosci. Remote Sens. 48 (8), 3188–3197.

Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J.C., Trianni, G., 2009. Recent advances in techniques for hyperspectral image processing. Remote Sens. Environ. 113, S110–S122.

Ratle, F., Camps-Valls, G., Weston, J., 2010. Semisupervised neural networks for efficient hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 48 (5), 2271–2282.

Shi, Q., Zhang, L., Du, B., 2013. Semi-supervised discriminative locally enhanced alignment for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 51 (9), 4800–4815.

Sun, S.L., Shawe-Taylor, J., 2010. Sparse semi-supervised learning using conjugate functions. J. Mach. Learn. Res. 11, 2423–2455.

Tan, K., Du, P.J., 2008. Hyperspectral remote sensing image classification based on support vector machine. J. Infrared Millimeter Waves 27 (2), 123–128.

Tan, K., Du, P.J., 2011. Combined multi-kernel support vector machine and wavelet analysis for hyperspectral remote sensing image classification. Chin. Opt. Lett. 9 (1), 011003–011006.

Tan, K., Li, E., Du, Q., Du, P., 2013. Hyperspectral image classification using band selection and morphological profiles. Sel. Top. Appl. Earth Observations Remote Sens., IEEE J. 7 (1), 40–48.

van der Meer, F.D., van der Werff, H., van Ruitenbeek, F.J., Hecker, C.A., Bakker, W.H., Noomen, M.F., van der Meijde, M., Carranza, E.J.M., Smeth, J., Woldai, T., 2012. Multi-and hyperspectral geologic remote sensing: a review. Int. J. Appl. Earth Observation Geoinf. 14 (1), 112–128.

Villa, A., Chanussot, J., Benediktsson, J.A., Jutten, C., Dambreville, R., 2013. Unsupervised methods for the classification of hyperspectral images with low spatial resolution. Pattern Recognit. 46 (6), 1556–1568.

Yang, H., Du, Q.A., Ma, B., 2010. Decision fusion on supervised and unsupervised classifiers for hyperspectral imagery. IEEE Geosci. Remote Sens. Lett. 7 (4), 875–879.

Zhong, Y.F., Zhang, L.P., 2012. An adaptive artificial immune network for supervised classification of multi-/hyperspectral remote sensing imagery. IEEE Trans. Geosci. Remote Sens. 50 (3), 894–909.