Feature Extraction of Hyperspectral Images With Semisupervised Graph Learning

Renbo Luo, *Student Member, IEEE*, Wenzhi Liao, *Member, IEEE*, Xin Huang, *Senior Member, IEEE*, Youguo Pi, and Wilfried Philips, *Senior Member, IEEE*

***Abstract*—We propose a semisupervised graph learning (SEGL) method for feature extraction of hyperspectral remote sensing imagery in this paper. The proposed SEGL method aims to build a semisupervised graph that can maximize the class discrimination and preserve the local neighborhood information by combining labeled and unlabeled samples. In our semisupervised graph, we connect labeled samples according to their label information and unlabeled samples by their nearest neighborhood information. By sorting the mean distance between a unlabeled sample and labeled samples of each class, we connect the unlabeled sample with all labeled samples belonging to its nearest neighborhood class. Moreover, the proposed SEGL better models the actual differences and similarities between samples, by setting different weights to the edges of connected samples. Experimental results on four real hyperspectral images (HSIs) demonstrate the advantages of our method compared to some related feature extraction methods.**

***Index Terms*—Classification, feature extraction, graph, hyperspectral images (HSIs), semisupervised.**

I. Introduction

With advanced sensor technology, hyperspectral images (HSIs), offering much richer spectral information than regular RGB and multispectral images, are easily accessible nowadays. However, the large number of spectral bands (hundreds to thousands) implies the high dimensionality of HSI, thus causes challenges to HSI analysis [1]. Conventional classification methods perform poorly due to the curse of dimensionality (i.e., the Hughes phenomenon [2]. One way to overcome this problem is to adopt a proper feature extraction method in the classification framework [3]. An effective and helpful feature extraction method can explore the intrinsic structure of HSI, reduce the data redundancy, and decrease the operation complexity, which is useful for improving the classification accuracy [4].

There are a number of existing approaches for feature extraction of HSIs [5]–[7], ranging from unsupervised methods to supervised ones. One of the best known unsupervised methods is principle component analysis (PCA) [8]. Some local methods, which preserve the properties of local neighborhoods, were also proposed to extract the features of HSIs [9], such as neighborhood preserving embedding (NPE) [10], locality preserving projection (LPP) [11], and linear local tangent space alignment (LLTSA) [12]. By considering neighborhood information around the data points, these local methods can preserve local neighborhood information and detect the manifold embedded in the high-dimensional feature space. As supervised feature extraction methods, linear discriminant analysis (LDA) [13] and nonparametric weighted feature extraction (NWFE) [14], are the two widely used methods, which extract features by optimizing the fisher criterion. Many extensions to these two methods have been proposed, such as modified Fisher's LDA [15], cosine-based nonparametric feature extraction [16], and double-nearest proportion feature extraction [17].

However, in real-world applications, very limited labeled samples are usually available, because manually labeling datasets are time consuming and fairly expensive. On the other hand, unlabeled samples could be available in large quantities at very low cost. For this reason, semisupervised learning methods [18], which aim at improving classification by utilizing both unlabeled and limited labeled data, have aroused a great deal of interest in the machine learning community [19], and been successfully used in HSI classification [20]–[23]. For the semisupervised classification, the representative methods include transductive support vector machine (SVM) [21], [22], and graph-based semisupervised classification methods [23], [24]. For semisupervised feature extraction methods, Zhang *et al.* [25] propose a semisupervised dimensionality reduction (SSDR) technique by utilizing the must- and cannot-link constraints. Some semisupervised feature extraction methods add a regularization term to preserve certain potential properties of the data, e.g., semisupervised discriminant analysis (SDA) [26] adds a regularizer into the objective function of LDA, it makes use of a limited number of labeled samples to maximize the class discrimination and employs both labeled and unlabeled samples to preserve the local properties of the data. Some semisupervised feature extraction methods combine

Manuscript received July 07, 2015; revised January 12, 2016; accepted January 22, 2016. Date of publication February 17, 2016; date of current version September 30, 2016. This work was supported in part by China Scholarship Council and in part by the FWO project G037115N: Data fusion for image analysis in remote sensing.

R. Luo is with the School of Automation Science and Engineering, South China University of Technology, 510640 Guangzhou, China, and also with the Department of Telecommunications and Information Processing, Ghent University, 9000 Ghent, Belgium (e-mail: reluo@telin.ugent.be).

W. Liao and W. Philips are with the Department of Telecommunications and Information Processing, Ghent University, 9000 Ghent, Belgium (e-mail: wliao@telin.ugent.be; philips@telin.ugent.be).

X. Huang is with the School of Remote Sensing and Information Engineering, Wuhan University, 430079 Wuhan, China (e-mail: huang_whu@163.com).

Y. Pi is with the School of Automation Science and Engineering, South China University of Technology, 510640 Guangzhou, China (e-mail: auygpi@scut.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/JSTARS.2016.2522564

supervised methods with unsupervised ones using a tradeoff parameter, such as semisupervised local fisher discriminant analysis (SELF) [27]. Other semisupervised feature extraction methods are driven by certain graphs constructed by both labeled and unlabeled data [28], [29], such as in the approach of [28], two similarity matrices were built first, of which one matrix based on labeled information and the other based on local neighborhoods' information, then these two matrices were fused by a parameter. However, it may not be easy to specify the optimal parameter values in these semisupervised methods, as mentioned in [30].

Recently, Liao *et al.* [31] proposed a semisupervised local discriminant analysis (SELD) for feature extraction of HSI without parameters. Their method divided the dataset into two sets: labeled and unlabeled sets, and employed the labeled samples through supervised method (LDA) only to maximize the class discrimination and the unlabeled samples through unsupervised local linear feature extraction methods (such as LPP, NPE, and LLSTA) only to preserve the local neighborhood information. However, the correlations between labeled and unlabeled samples are not well exploited in SELD; furthermore, in the supervised part (LDA), it tends to give undesired results if samples in a class form several separate clusters (as multimodal) [32]. An improved semisupervised local discriminant analysis (ISELD) method [33] was proposed in our previous work, in which we added a similarity matrix to model the correlations between labeled and unlabeled samples. However, the similarities between samples are not well modeled, as ISELD [33] sets the same edges to same class.

In this paper, we propose a novel semisupervised graph leaning method (named SEGL) for feature extraction of hyperspectral imagery. Our proposed method aims to build a semisupervised graph that can maximize the class discrimination and preserve the local neighborhood information by combining labeled and unlabeled samples. In our semisupervised graph, we connect samples according to either label information (labeled samples) or their k-nearest neighbors (unlabeled samples). We connect a unlabeled sample with labeled samples in a class by minimizing the mean distance of this unlabeled samples to labeled samples of each class. The main contributions of this paper are: the proposed SEGL does not combine supervised and unsupervised methods in a framework as SELD [31] and ISELD [33], but proposed a novel framework to utilize labeled and unlabeled samples in a semisupervised graph. Moreover, the proposed SEGL method does not set the same weights to the edges of the same class or samples within their k-nearest neighbors, as [28], [31], [33], but employs weighted edges (with weights corresponding to the distance between samples). This way, we proposed a more general framework to build a semisupervised graph, where the actual differences and similarities between samples are better modeled. The experimental results on four real hyperspectral remote sensing images demonstrate that the proposed SEGL enables better performances on classification.

The organization of this paper is as follows. In Section II, the proposed semisupervised graph learning (SEGL) method for feature extraction of HSI is analyzed in detail. Section III presents the experimental results. Finally, the conclusion of this paper is drawn in Section IV.

II. Feature Extraction With the Proposed SEGL

This section details our proposed SEGL method for feature extraction of HSI. Let us define $\mathbf{X}_{labeled} = \{\mathbf{x}_{L_i}\}_{i=1}^{n}$, $\mathbf{Y} = \{y_i\}_{i=1}^{n}$, $y_i \in \{1, 2, \ldots, C\}$, here $\mathbf{x}_{L_i}$ denotes the ith sample in the labeled samples set, y_i is the label of $\mathbf{x}_{L_i}$, C is the number of classes, $\mathbf{X}_{unlabeled} = \{\mathbf{x}_{U_{n+1}}, \mathbf{x}_{U_{n+2}}, \ldots, \mathbf{x}_{U_{n+m}}\}$, $\mathbf{x}_{U_j}$ denotes the jth sample in the unlabeled samples set, n is the number of labeled training samples, and m is the number of unlabeled training samples. Assume that the labeled samples set $\mathbf{X}_{labeled} = \{\mathbf{x}_{L_1}, \mathbf{x}_{L_2}, \ldots, \mathbf{x}_{L_n}\}$, are ordered according to their labels y_i, with the data matrix of the cth class $\mathbf{X}^{(c)} = \{\mathbf{x}_1^{(c)}, \mathbf{x}_2^{(c)}, \ldots, \mathbf{x}_{n_c}^{(c)}\}$, where $\mathbf{x}_i^{(c)}$ is ith sample in cth class. Then, the labeled set can be expressed as $\mathbf{X}_{labeled} = \{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(C)}\}$.

We exploits the label information and local neighborhood information through our proposed semisupervised graph, and our proposed semisupervised graph is defined as $\mathbf{G} = (\mathbf{X}, \mathbf{A})$, $\mathbf{X} = \{\mathbf{X}_{labeled}, \mathbf{X}_{unlabeled}\} = \{\mathbf{x}_{L_1}, \mathbf{x}_{L_2}, \ldots, \mathbf{x}_{L_n}, \mathbf{x}_{U_{n+1}}, \mathbf{x}_{U_{n+2}}, \ldots, \mathbf{x}_{U_{n+m}}\}$ is a set of nodes that are connected by a set of edges $\mathbf{A}$. The edge connecting nodes $\mathbf{x}_i$ and $\mathbf{x}_j$, has an associated weight A_{ij}. The basic goal of our proposed method is to find a transformation matrix $\mathbf{W} \in \mathbb{R}^{D \times d}$, which can transfer the data $\mathbf{x}_i$ in high-dimensional feature space into low-dimensional sample $\mathbf{z}_i$ in a way of $\mathbf{z}_i = \mathbf{W}^T\mathbf{x}_i$. The transformation matrix $\mathbf{W}$ can be optimized as follows:

$$\arg \min_{\mathbf{W} \in \mathbb{R}^{D \times d}} \left(\sum_{i,j=1}^{n+m} ||\mathbf{W}^T\mathbf{x}_i - \mathbf{W}^T\mathbf{x}_j||^2 A_{ij} \right). \quad (1)$$

In many applications, labeled samples are typically used to enhance class discrimination, but are always very limited. Unlabeled samples, on the other hand, are much easier accessible. The idea behind semisupervised feature extraction methods [31] and [33] is to infer class discrimination from labeled samples, as well as the local neighborhood information from unlabeled samples. Motivated by [31] and [33], we define our proposed semisupervised graph to model different correlations between samples as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{LL} & \mathbf{A}_{LU} \\ \mathbf{A}_{UL} & \mathbf{A}_{UU} \end{bmatrix} \quad (2)$$

where $\mathbf{A}$ is the adjacency matrix, two nodes are adjacent if they have connections (i.e., $A_{ij} \neq 0$), the adjacency matrix $\mathbf{A}_{LL}$ is a $n \times n$ matrix that models the correlations between labeled samples, if two samples belong to the same class, we set an edge between them. The adjacency matrix $\mathbf{A}_{UU}$, which is a $m \times m$ matrix, models the correlations between unlabeled samples. We connect two unlabeled samples, if two unlabeled samples are within the k-nearest neighbors of each other, we define $\mathbf{A}_{LL}$ and $\mathbf{A}_{UU}$, respectively, as

$$A_{L_iL_j} = \begin{cases} 1, & y_i = y_j \\ 0, & \text{others} \end{cases} \tag{3}$$

$$A_{U_iU_j} = \begin{cases} 1, & \mathbf{x}_{U_j} \in kNN(\mathbf{x}_{U_i}) \\ & \text{or } \mathbf{x}_{U_i} \in kNN(\mathbf{x}_{U_j}) \\ 0, & \text{others} \end{cases} \tag{4}$$

where $kNN(\mathbf{x}_{U_i})$ denotes a set of unlabeled samples that are within the k nearest neighbors of $\mathbf{x}_{U_i}$.

The adjacency matrices $\mathbf{A}_{LU}$ and $\mathbf{A}_{UL}$ contain the connection between labeled and unlabeled samples, $\mathbf{A}_{UL} = (\mathbf{A}_{LU})^T$, as $\mathbf{A}$ is a symmetric matrix. Suppose the labeled sample $\mathbf{x}_{L_i}$ belong to class c_j, the $n \times m$ adjacency matrix $\mathbf{A}_{LU}$ is defined as

$$A_{L_iU_j} = \begin{cases} 1, & \mathbf{x}_{L_i} \in \mathbf{X}^{(c_j)} \\ 0, & \text{others} \end{cases} \tag{5}$$

where c_j represents the class that $\mathbf{x}_{U_j}$ is closest to, and $\mathbf{X}^{(c_j)}$ is a set including all labeled samples in class c_j, $c_j \in \{1, \ldots, C\}$. Here, we assume that the distribution of each class is unimodal [34]

$$c_j = \arg\min_c m_c(\mathbf{x}_{U_j}), \quad c = 1, 2, \ldots, C \tag{6}$$

where $m_c(\mathbf{x}_{U_j})$ denotes the mean distance of an unlabeled sample $\mathbf{x}_{U_j}$ to all labeled samples in class c, and we define it as follows:

$$m_c(\mathbf{x}_{U_j}) = \frac{1}{n_c} \sum_{t=1}^{n_c} d\left(\mathbf{x}_{U_j}, \mathbf{x}_t^{(c)}\right) \tag{7}$$

where $d(\mathbf{x}_i, \mathbf{x}_j)$ is the Euclidean distance between $\mathbf{x}_i$ and $\mathbf{x}_j$. $\mathbf{x}_i$ and $\mathbf{x}_j$ are closer to each other if $d(\mathbf{x}_i, \mathbf{x}_j)$ is smaller. Therefore, if $m_c(\mathbf{x}_{U_j})$ is smaller, it means $\mathbf{x}_{U_j}$ is more similar to labeled samples from class c and more likely to belong to class c, we set edges between an unlabeled sample $\mathbf{x}_{U_j}$ and all labeled samples $\mathbf{x}_{L_i}$ in class c.

In the above definition of adjacency matrix $\mathbf{A}$, if two nodes are connected, the weights of their edges are set to 1. If we set same weights for samples with same label or within their k-nearest neighbors, the connected nodes would make same contribution to the definition of the optical projection matrix (1) . The actual differences and similarities between samples are not well modeled. Therefore, we set an weighted edge between connected nodes by combing the affinity matrix $\mathbf{F}$ with adjacency matrix $\mathbf{A}$, to make nearest connected samples be closest, and far apart connected samples be far away when projecting them into low-dimensional feature space. The final similarity matrix $\mathbf{A}$ in our proposed semisupervised graph $\mathbf{G} = (\mathbf{X}, \mathbf{A})$ is defined as follows:

$$\mathbf{A} = \mathbf{F} \odot \mathbf{A} \tag{8}$$

where "$\odot$" denoting the element-wise multiplication and $F_{i,j}$ is the affinity between $\mathbf{x}_i$ and $\mathbf{x}_j$. In this paper, we use the local scaling heuristic [35] as the definition of affinity matrix $\mathbf{F}$

$$F_{i,j} = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\delta_i \delta_j}\right). \tag{9}$$

The parameter δ_i represents the local scaling around $\mathbf{x}_i$ defined by $\delta_i = \|\mathbf{x}_i - \mathbf{x}_i^k\|$ and $\delta_j = \|\mathbf{x}_j - \mathbf{x}_j^k\|$, where $\mathbf{x}_i^k$ is the kth nearest neighbor of $\mathbf{x}_i$. A heuristic choice of $k = 7$ was shown to be useful through experiments [35]. $F_{i,j}$ is large if $\mathbf{x}_i$ and $\mathbf{x}_j$ are "close," and $F_{i,j}$ is small if $\mathbf{x}_i$ and $\mathbf{x}_j$ are "far apart." Fig. 1 shows the graph constructed by our proposed semisupervised method. The graph was constructed by selecting five labeled samples per class and 100 unlabeled samples randomly from the *University of Pavia* dataset. Without F, the edges of all connected samples are equally set to 1, as shown in Fig. 1(a). However, in real applications, samples even from the same class have spectral differences, as shown in Fig. 1(b), if we set all the weights of the samples from the same labeled class to the same value as SELD [31] did, the differences and similarities cannot be well modeled. With our proposed semisupervised graph, the differences and similarities are much better modeled, this means two labeled samples $\mathbf{x}_i$ and $\mathbf{x}_j$, if they are closer to each other, their connection weight A_{ij} would be larger. On the contrary, if they are far away from each other or mislinked, their connection weight A_{ij} would be smaller, thus can reduce the negative influence of mislinking, as shown in Fig. 1(b).

In order to avoid degeneracy, we use the constraint as follows:

$$\mathbf{W}^T\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{W} = \mathbf{I} \tag{10}$$

where $\mathbf{D}$ is a diagonal matrix with $D_{i,i} = \sum_{j=1}^{m+n} A_{i,j}$ and $\mathbf{I}$ is the identity matrix, we can obtain the transformation matrix $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r)$, which is made up by r eigenvectors associated with the least r eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_r$ of the following generalized eigenvalue problem:

$$\mathbf{X}\mathbf{L}\mathbf{X}^T\mathbf{w} = \lambda\mathbf{X}\mathbf{D}\mathbf{X}^T\mathbf{w} \tag{11}$$

where $\mathbf{L} = \mathbf{D} - \mathbf{A}$ is a Laplacian matrix.

III. Experiments

A. HSI Datasets

Four real hyperspectral datasets are used in our experiments: *University of Pavia*, *Pavia Center*, *Botswana*, and *Kennedy Space Center*. Table I shows the number of labeled samples in each class for all the datasets. Note that the color in the cell denotes different classes in the classification maps (Figs. 3 and 4). The datasets *University of Pavia* and *Pavia Center* are from urban areas in the city of Pavia, Italy. The data were collected by the ROSIS (reflective optics system imaging spectrometer) sensor, with 115 spectral bands in the wavelength range from 0.43 to 0.86 μm, and a very fine spatial resolution of 1.3 m by pixel.

University of Pavia (Upavia): The data with 610×340 pixels were collected over the University of Pavia, Italy. It contains 103 spectral channels after removal of 12 noisy bands. The data include nine land cover/use classes (see Table I).

Pavia Center (PCenter): The data with 1096×715 pixels were collected over Pavia, Italy. It contains 102 spectral channels after removal of 13 noisy bands. Nine groundtruth classes were considered in the experiments (see Table I).

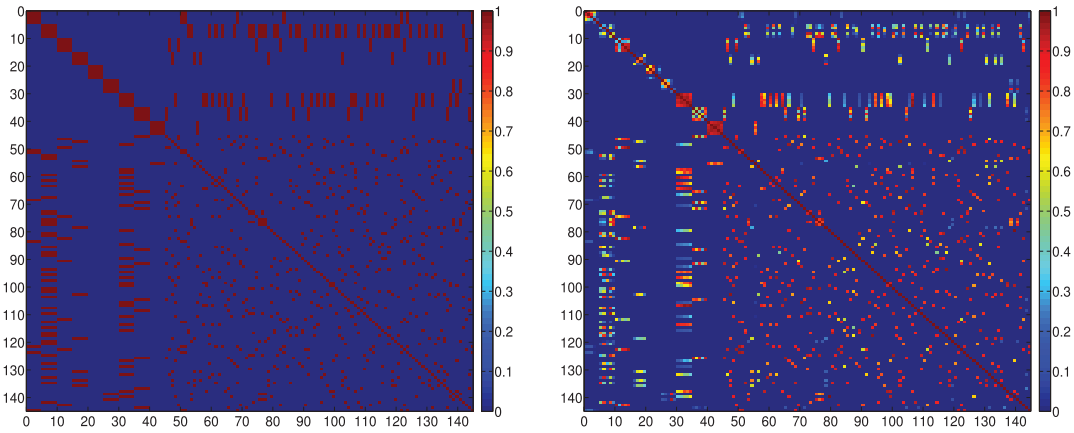


Fig. 1. Semisupervised graph. (a) Graph with same weights. (b) Proposed graph with different weights.

TABLE I
DATASETS USED IN THE EXPERIMENTS

No	University of Pavia		KSC		Botswana		Pavia Centre	
	Class name	Samples	Class name	Samples	Class name	Samples	Class name	Samples
1	Asphalt	6631	Scrub	761	Water	270	Water	65 971
2	Meadows	18 649	Willow swamp	243	Hippo grass	101	Trees	7598
3	Gravel	2099	Cabbage palm hammock	256	Floodplain Grass1	251	Meadows	3090
4	Trees	3064	Cabbage palm/oak hammock	252	Floodplain Grass2	215	Bricks	2685
5	Metal sheets	1345	Slash Pine	161	Reeds	269	Soil	6584
6	Soil	5029	Oak/broadleaf hammock	229	Riparian	269	Asphalt	9248
7	Bitumen	1330	Hardwood swamp	105	firescar2	259	Bitumen	7287
8	Bricks	3682	Graminoid marse	431	Island interior	203	Tiles	42 826
9	Shadow	947	Spartan marse	520	Acacia woodlands	341	Shadow	2863
10			Cattail marse	404	Acacia shrublands	248		
11			Salt marse	419	Acacia grasslands	305		
12			Mud flats	503	Short mopane	181		
13			Water	927	Mixed mopane	268		
14					Exposed soils	95		
Total		42 686		5211		3248		148 155

Kennedy Space Center (KSC): The dataset was acquired by NASA AVIRIS instrument over the KSC, Florida, in 1996 and consists of 224 bands of 10-nm width with center wavelengths from 0.4 to 2.5 μm. The data, acquired from an altitude of approximately 20 km, have a spatial resolution of 18 m/pixel. Several spectral bands were removed from the data due to noise and water absorption phenomena, leaving a total of 176 bands to be used for the analysis. For classification purposes, 13 classes representing the various land cover types that occur in this environment were defined for the site (see Table I). For more information, see website http://www.csr.utexas.edu/hyperspectral/.

Botswana: The dataset was acquired over the Okavango Delta, Botswana, in May 31, 2001 by the NASA EO-1 satellite, with 30-m/pixel resolution over a 7.7-km strip in 242 bands covering the 0.4–2.5 μm portion of the spectrum in 10-nm windows. Uncalibrated and noisy bands that cover water absorption features were removed, leaving a total of 145 radiance channels to be used in the experiments. The data consist of observations from 14 identified classes intended to reflect the impact of flooding on vegetation (see Table I). For more information, see http://www.csr.utexas.edu/hyperspectral/.

B. Experimental Setup

The training set $\mathbf{X}$ is composed of labeled subset $\mathbf{X}_{labeled}$ and unlabeled subset $\mathbf{X}_{unlabeled}$ (such that $\mathbf{X} = \mathbf{X}_{labeled} \cup \mathbf{X}_{unlabeled}$ and $\mathbf{X}_{labeled} \cup \mathbf{X}_{unlabeled} = \phi$). In order to analyze the influence of the size of labeled samples on classification accuracy, a number of unlabeled samples $u = 2000$ were randomly selected from the image parts with no labels to compose $\mathrm{X}_{unlabeled}$, and labeled subset $\mathrm{X}_{labeled}$ was made of labeled training samples which was randomly selected from the labeled data with the samples size corresponding to different cases: 20, 40, and 80 samples per class, respectively. The training of the classifiers was carried out using the labeled subset $\mathrm{X}_{labeled}$. The remaining labeled samples were used as the test set. We compare the classification accuracies using the proposed SEGL method with results from the following methods: PCA [8]; LPP [11]; NWFE [14]; and SDA [26], where the parameter α is

optimized with fivefold cross validation within the given set $\{0.1, 0.5, 2.5, 12.5, 62.5\}$; SELF [27], of which the parameter β is chosen from $\{0, 0.1, 0.2, \ldots, 0.9, 1\}$ by fivefold cross validation; SLPPCE [28]; SELD [31], and ISELD [33].

We used three common classifiers: 1-nearest neighbor (1NN), support vector machines (SVMs), and random forest (RF). The SVM classifier with RBF kernels has two parameters: the penalty factor C and the RBF kernel widths γ, we optimized C within the given set $\{10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$ and γ within the given set $\{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}\}$ by fivefold cross validation, the RF classifier with 200 trees. All classifiers were evaluated against the test set. Meanwhile, we use overall classification accuracy (OA) to evaluate the feature extraction results. The results were averaged over 10 runs on different number of extracted features from 1 to 20, and the averaged OA was recorded for each method. The number of nearest neighbors was set to 8.

C. Results on Different Number of Labeled Training Samples

Tables II–IV display the classification accuracies of testing data with different distinct labeled samples size: 20, 40, and 80 per class, respectively. The best averaged accuracy of each dataset (in column) is highlighted in bold. From these tables, we have the following findings.

1) The results confirm that most semisupervised feature extraction methods can achieve better results in the classification of HSIs, comparing the unsupervised or supervised feature extraction methods. The classification accuracy will be higher as the number of labeled training samples increasing. Especially, for the proposed SELG method, on the *University of Pavia* dataset, when the labeled training size is small (20 labeled training samples per class), the best averaged OA is 79.40%, and if we choose 80 labeled training samples from each class, the best averaged OA reaches to 83.57%, which has more than 4% improvements.
2) The semisupervised feature extraction methods SELD [31], ISELD [33], and the proposed method SEGL, which divide the samples into two sets (labeled and unlabeled) first, infer class discrimination from labeled samples and keep local neighborhood information from unlabeled samples, perform better than other semisupervised methods (SELF [27] and SLPPCE [28]). Thus, suggests that dividing the samples into two sets first and achieving different goals from different sets is an effective way in semisupervised learning.
3) By connecting unlabeled and labeled samples in the semisupervised graph and employing weighted edges between samples, SEGL outperforms than other semisupervised methods. Compared with SEGL method SLPPCE, our proposed SEGL method has about 10% improvements on the *University of Pavia* dataset, more than 3% improvements on the *KSC* dataset.
4) SVM classifier is more efficient to deal with the datasets *KSC* and *Pavia Centre*; however, 1NN classifier gets the highest averaged accuracies in most cases for other two datasets. When the preprocessing features were extracted by unsupervised methods PCA and LPP, or supervised methods NWFE, RF, and SVM classifier perform much better than 1NN classifier. On the other hand, SVM classifier needs more time for classification than RF and 1NN classifier from the experiments.
5) The proposed SEGL method gets best averaged OA on the four datasets from all the results of experiments. In Table II, the best averaged OA in *University of Pavia* and *KSC* are 79.40% (SEGL with 1NN classifier) and 92.25% (SEGL with SVM classifier), respectively, which have at least 1% improvements than others. In Table III, just the results of the proposed methods SEGL with 1NN classifier have reached above 81% in *University of Pavia*. For the other three datasets, SEGL still have best results, 93.19%, 95.47%, and 97.29%, respectively. In Table IV, SEGL with SVM classifier has the best results in *KSC* and *Pavia Center* datasets, 95.37% and 97.64%, respectively. SEGL with RF classifier gets the highest OA 97.13% in *Botswana* dataset, and SEGL with 1NN classifier gets the highest OA 83.57% in *University of Pavia* dataset.

TABLE II
OA% (OPTIMAL NUMBER OF FEATURES) BY USING FEATURE EXTRACTION APPROACHES WITH LABELED TRAINING SAMPLE SIZE 20

Feature extraction	Classifier	Data set			
		UPavia	*KSC*	*Botswana*	*PCentre*
PCA	1NN	67.38 (11)	77.90 (14)	86.18 (14)	94.77 (11)
	SVM	70.21 (8)	85.66 (18)	92.72 (8)	96.08 (12)
	RF	73.46 (11)	86.40 (12)	92.16 (7)	95.63 (11)
LPP	1NN	66.97 (18)	77.93 (19)	85.82 (13)	94.78 (13)
	SVM	71.80 (9)	86.12 (17)	92.15 (12)	96.14 (11)
	RF	72.46 (18)	85.46 (19)	92.13 (9)	95.97 (11)
NWFE	1NN	71.34 (9)	85.63 (17)	90.07 (13)	95.92 (11)
	SVM	72.29 (8)	89.64 (13)	92.15 (10)	96.47 (14)
	RF	75.84 (9)	89.20 (15)	92.41 (12)	96.09 (7)
SDA	1NN	52.67 (7)	73.16 (8)	72.89 (11)	79.86 (14)
	SVM	51.62 (8)	73.87 (10)	72.98 (11)	79.55 (12)
	RF	55.72 (9)	70.56 (8)	71.97 (12)	79.33 (8)
SELF	1NN	61.42 (18)	78.79 (18)	83.84 (18)	93.64 (13)
	SVM	63.93 (19)	85.98 (17)	79.20 (14)	93.38 (10)
	RF	67.98 (18)	85.33 (15)	82.64 (12)	94.34 (6)
SLPPCE	1NN	68.96 (8)	87.18 (13)	89.26 (16)	93.81 (8)
	SVM	67.40 (8)	86.14 (19)	84.04 (12)	92.43 (10)
	RF	66.83 (9)	83.63 (12)	86.33 (11)	92.28 (6)
SELD	1NN	77.27 (17)	88.82 (19)	93.91 (16)	95.44 (12)
	SVM	75.71 (11)	89.44 (16)	91.20 (10)	95.42 (12)
	RF	75.53 (8)	88.78 (13)	91.85 (6)	95.59 (8)
ISELD	1NN	78.66 (11)	90.74 (18)	93.94 (18)	96.60 (9)
	SVM	76.20 (9)	91.09 (15)	93.71 (12)	96.52 (7)
	RF	75.66 (8)	89.54 (15)	93.41 (12)	96.04 (7)
SEGL	1NN	**79.40 (9)**	90.81 (15)	**94.58 (12)**	96.54 (7)
	SVM	77.57 (8)	**92.25 (12)**	94.14 (13)	**96.70 (8)**
	RF	76.67 (8)	89.52 (13)	93.74 (13)	96.29 (6)

Fig. 2 shows the averaged OA of different semisupervised learning methods changing with the increasing number of extracted features. It can be seen that the proposed SEGL with SVM classifier has better performance than other methods for *University of Pavia* and *KSC* datasets. With the number of extracted features increasing, the OA of SEGL will improve first and then keep stable or have a little decrease. The optimal number of the features extracted by our proposed method is 8 and 12 with SVM classifier for *University of Pavia* and *KSC*, respectively. However, automatic selection of the optimal number of features is still very challenging for most methods

TABLE III
OA% (OPTIMAL NUMBER OF FEATURES) BY USING FEATURE EXTRACTION APPROACHES WITH LABELED TRAINING SAMPLE SIZE 40

Feature extraction	Classifier	Data set			
		UPavia	*KSC*	*Botswana*	*PCentre*
PCA	1NN	69.46 (12)	81.39 (19)	87.98 (9)	95.33 (12)
	SVM	74.77 (8)	88.53 (9)	93.30 (8)	96.78 (11)
	RF	73.88 (11)	88.87 (18)	93.79 (11)	96.13 (11)
LPP	1NN	69.43 (18)	81.20 (18)	87.59 (9)	95.33 (13)
	SVM	77.14 (12)	89.13 (14)	92.42 (11)	96.92 (10)
	RF	74.20 (19)	88.91 (9)	93.26 (8)	96.23 (12)
NWFE	1NN	73.34 (9)	87.76 (14)	90.44 (8)	96.32 (14)
	SVM	76.62 (8)	91.83 (12)	93.72 (8)	96.87 (9)
	RF	77.03 (11)	91.15 (13)	93.98 (12)	96.64 (8)
SDA	1NN	62.64 (8)	84.98 (9)	85.12 (12)	91.43 (13)
	SVM	63.64 (9)	86.75 (10)	85.45 (13)	89.99 (14)
	RF	66.97 (9)	84.98 (8)	84.87 (12)	90.68 (12)
SELF	1NN	68.75 (18)	81.86 (17)	88.58 (9)	95.54 (11)
	SVM	75.85 (19)	89.76 (15)	90.54 (12)	96.75 (6)
	RF	75.97 (12)	89.87 (11)	92.79 (10)	96.86 (9)
SLPPCE	1NN	73.34 (8)	89.77 (13)	91.93 (13)	95.31 (9)
	SVM	68.16 (10)	89.55 (18)	88.70 (10)	94.73 (7)
	RF	67.56 (11)	89.98 (10)	91.34 (12)	94.73 (8)
SELD	1NN	79.03 (10)	91.12 (18)	94.69 (12)	96.10 (9)
	SVM	76.20 (10)	92.03 (15)	93.26 (11)	96.68 (8)
	RF	77.50 (9)	91.94 (12)	93.69 (12)	96.45 (7)
ISELD	1NN	79.79 (10)	92.65 (17)	94.52 (18)	97.07 (9)
	SVM	76.26 (9)	92.78 (15)	94.31 (12)	96.90 (6)
	RF	76.79 (8)	91.91 (9)	94.89 (12)	96.49 (7)
SEGL	1NN	**81.22 (9)**	93.02 (12)	**95.47 (13)**	97.19 (7)
	SVM	78.16 (8)	**93.19 (13)**	95.39 (12)	**97.29 (7)**
	RF	79.14 (8)	92.35 (12)	95.33 (12)	96.59 (6)

TABLE IV
OA% (OPTIMAL NUMBER OF FEATURES) BY USING FEATURE EXTRACTION APPROACHES WITH LABELED TRAINING SAMPLE SIZE 80

Feature extraction	Classifier	Data set			
		UPavia	*KSC*	*Botswana*	*PCentre*
PCA	1NN	70.23 (11)	83.31 (19)	89.71 (9)	95.97 (12)
	SVM	76.81 (9)	90.50 (18)	94.45 (17)	96.61 (11)
	RF	76.13 (11)	91.13 (14)	95.32 (13)	96.60 (8)
LPP	1NN	70.17 (12)	84.77 (13)	89.4 (12)	95.94 (13)
	SVM	77.95 (18)	89.82 (8)	94.02 (15)	95.52 (9)
	RF	75.82 (19)	90.90 (12)	95.33 (11)	96.56 (8)
NWFE	1NN	74.42 (12)	90.03 (17)	90.91 (12)	96.37 (7)
	SVM	77.18 (9)	93.42 (9)	94.90 (11)	96.39 (8)
	RF	78.74 (8)	92.57 (19)	95.53 (11)	96.00 (6)
SDA	1NN	70.76 (8)	91.74 (12)	91.98 (10)	95.45 (13)
	SVM	70.96 (10)	91.57 (10)	91.87 (19)	96.89 (9)
	RF	71.34 (9)	91.38 (8)	92.57 (13)	95.43 (12)
SELF	1NN	69.56 (16)	90.50 (12)	90.13 (11)	95.89 (9)
	SVM	82.74 (18)	92.88 (11)	94.12 (14)	96.45 (12)
	RF	80.56 (12)	92.89 (13)	95.86 (12)	96.59 (10)
SLPPCE	1NN	74.36 (10)	92.87 (18)	94.41 (9)	96.40 (9)
	SVM	72.91 (10)	92.58 (14)	93.22 (13)	96.23 (9)
	RF	78.53 (11)	92.26 (15)	95.60 (13)	96.22 (11)
SELD	1NN	81.95 (11)	93.43 (18)	95.14 (12)	97.25 (8)
	SVM	82.03 (11)	93.71 (13)	94.97 (15)	97.20 (8)
	RF	82.26 (9)	93.80 (12)	95.60 (11)	96.92 (9)
ISELD	1NN	82.47 (9)	94.05 (12)	94.99 (14)	97.16 (11)
	SVM	79.40 (10)	94.22 (15)	95.75 (12)	97.20 (8)
	RF	79.75 (9)	94.09 (10)	95.66 (12)	96.86 (7)
SEGL	1NN	**83.57 (10)**	94.69 (13)	96.44 (15)	97.18 (8)
	SVM	82.26 (8)	**95.37 (12)**	96.02 (12)	**97.64 (7)**
	RF	83.51 (9)	93.65 (12)	**97.13 (13)**	96.95 (7)

[36]. It depends on the distribution of the datasets, the training samples and the classifiers (see Tables II–IV). Many approaches selected the optimal number of features according to the cumulative variance [36]. However, these approaches do not always work well, as was discussed in [36] and [37].

In order to compare the classification results visually, we randomly select 40 labeled training samples per class from *University of Pavia* and *KSC* datasets. With SVM classifier, the best classification maps of each methods are shown in Figs. 3 and 4, respectively. It can be seen that the classification maps of the proposed SEGL looks smooth on *University of Pavia* dataset, and this is specially clear for classes *Meadows* and *Soil*. In the classification maps of *KSC*, the proposed method SEGL also yields a good classification result, and outperforms than other feature extraction methods in the *Water* region near to the coastline, also in the *Salt marsh* parts located in the center of *Water* region.

D. Results on Different Number of Unlabeled Training Samples

This experiment investigates the influence of the unlabeled sample size on the classification performances. The choice of unlabeled samples is also a very important step in the semisupervised methods. Large number of unlabeled training samples increases computational complexity, while a small number of unlabeled samples is not sufficient to exploit the local neighborhood information. We choose 20 labeled training samples from each class to compose the labeled subset $\mathrm{X}_{labeled}$, the number of unlabeled subset $\mathrm{X}_{unlabeled}$ was evaluated from 500 to 5000 with a step 500. Fig. 5 shows the tendency of averaged OA with the number of unlabeled training samples increasing, using several different feature extraction methods with SVM classifier for *University of Pavia* and *KSC* datasets. As can be seen, the classification accuracy will improve first and then keep stable as more and more unlabeled samples are used. The averaged OA of the proposed SEGL method just improved about 2% when the number of unlabeled training samples increased from 500 to 5000 on these two datasets. As on the *KSC* dataset, the averaged OA of ISELD method just reached to 89% when only 500 unlabeled training samples were chosen, while OA reaches 92.86% by using 4500 unlabeled samples.

E. Results on Different Number of Nearest Neighbors

The number of nearest neighbors (k) is an important parameter in our proposed semisupervised graph. On the one hand, when k is too small, the local information may not be properly modeled. On the other hand, large k (with same unlabeled samples) will lead to mislabeling problems. To investigate the effect of the number of nearest neighbors and unlabeled samples on the classification accuracy, we take *University of Pavia* and *KSC* datasets as examples in our experiments. Forty labeled training samples were selected from each class with fivefold cross validation to compose the labeled subset $X_{labeled}$, the number of unlabeled subset $X_{unlabeled}$ was evaluated from 500 to 7000 with a step 500, the number of nearest neighbors was changed from 4 to 30 with a step 2. Fig. 6 shows the surface of the OA as a function of nearest neighbors (k) and unlabeled samples (u), with SVM classifier. As can be seen, when u was set to 500, the averaged OA increased at first and then decreased as the number of nearest neighbors changed from 4 to 30. This indicates that the increase of k, with a fixed number of u, will mislabel many unlabeled samples, leading to poor classification performances.

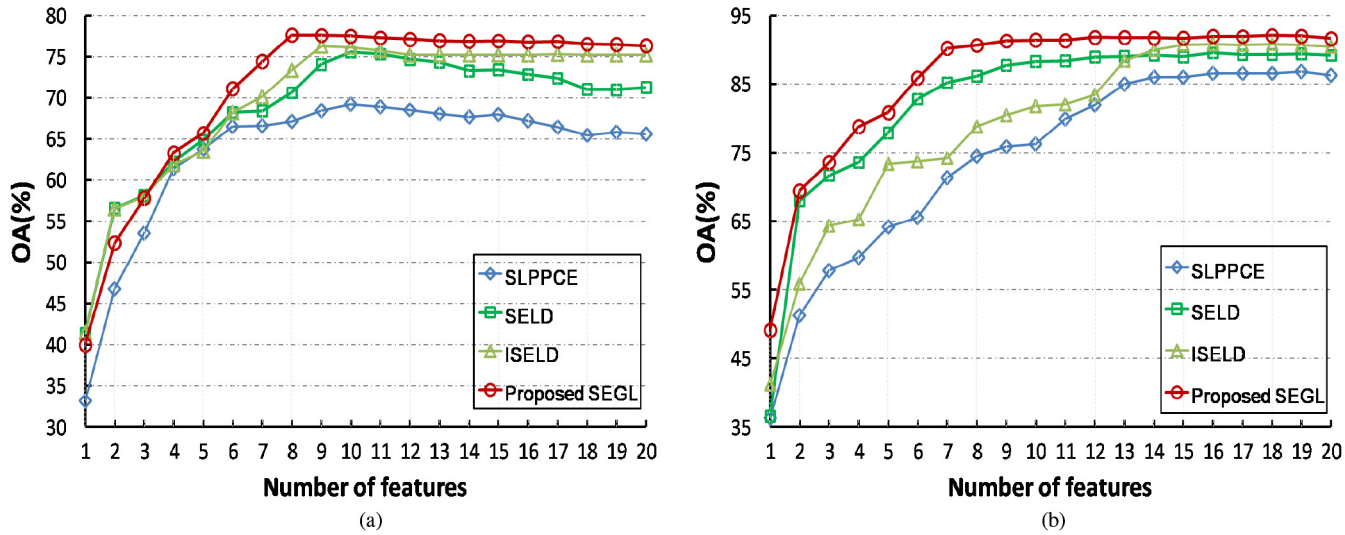


Fig. 2. Averaged OA (%) with the number of extracted features increasing for different semisupervised feature extraction method with SVM classifier. Forty labeled training samples are chosen randomly from each class. (a) *University of Pavia*. (b) *KSC*.

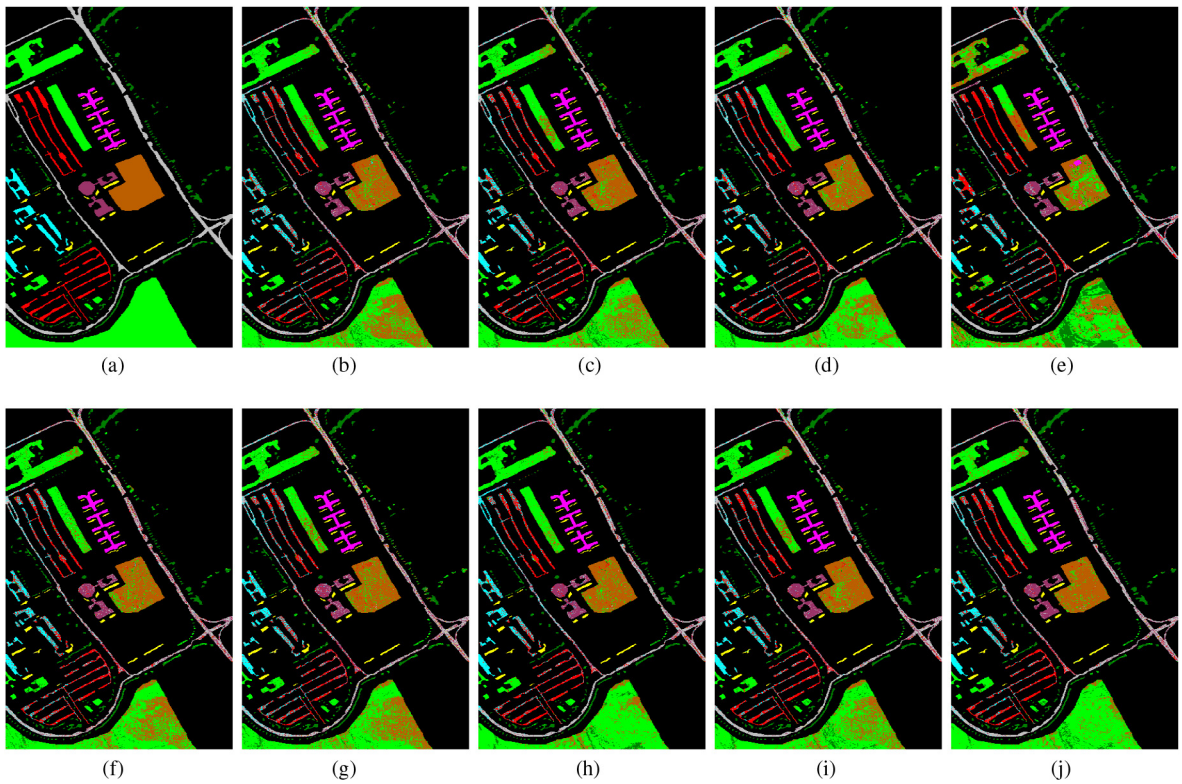


Fig. 3. Classification maps of the different methods with SVM classifier for *University of Pavia*. Forty labeled samples per class were randomly selected from the training set. (a) Groundtruth. (b) PCA. (c) LPP. (d) NWFE. (e) SDA. (f) SELF. (g) SLPPCE. (h) SELD. (i) ISELD. (j) Proposed SEGL.

When we keep k stable, the OA will first increase then fall down as the number of unlabeled samples changes. This means if k (or u) is set to a large value, it will increase the possibilities of wrong linked, i.e., some unlabeled samples that belong to different classes in reality would be connected, degrading the performances of the proposed method. We can also see that when the number of unlabeled samples is less than 2000, the classification results are not stable as k changes. This is because the distribution of nearest neighborhood unlabeled samples is sparse, the change of k has a big effect on the averaged radius (i.e., the averaged distance between a sample and its kth nearest neighbors). Consequently, if k is fixed, the effect of k on classification decreases as the number of unlabeled data increases. Therefore, in our proposed method, the number of nearest neighbors (k) should be changed in accordance with the number of unlabeled samples (u). Furthermore, the results show that selecting training samples with cross validation can improve the classification performances of our proposed method.

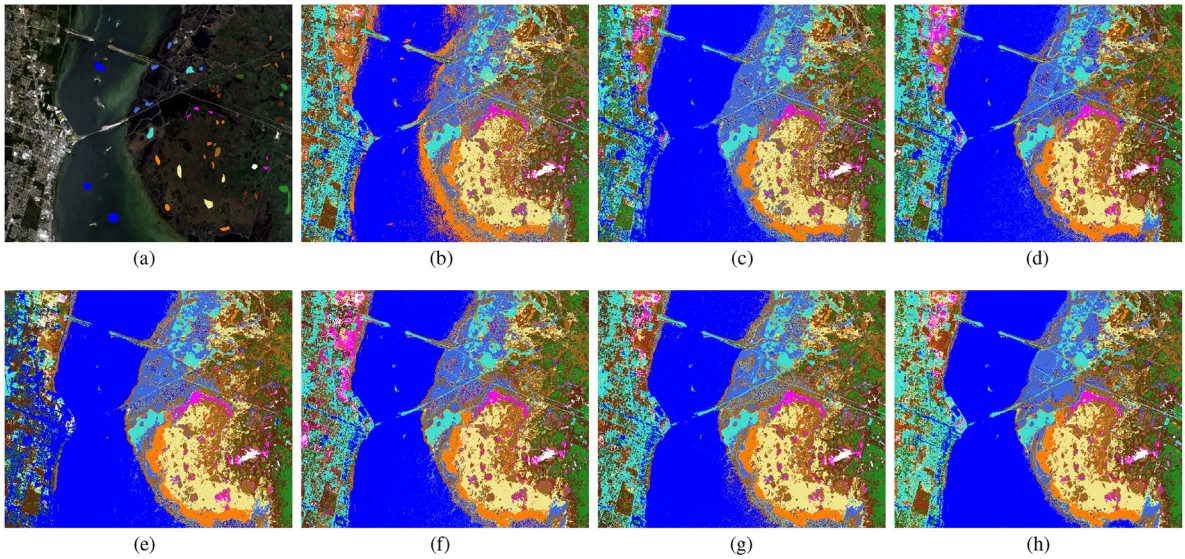


Fig. 4. Classification maps of the different methods with SVM classifier for *KSC*. Forty labeled samples per class were randomly selected from the training set. (a) RGB with Groundtruth. (b) PCA and SELF. (c) LPP. (d) NWFE. (e) SLPPCE. (f) SELD. (g) ISELD. (h) Proposed SEGL.

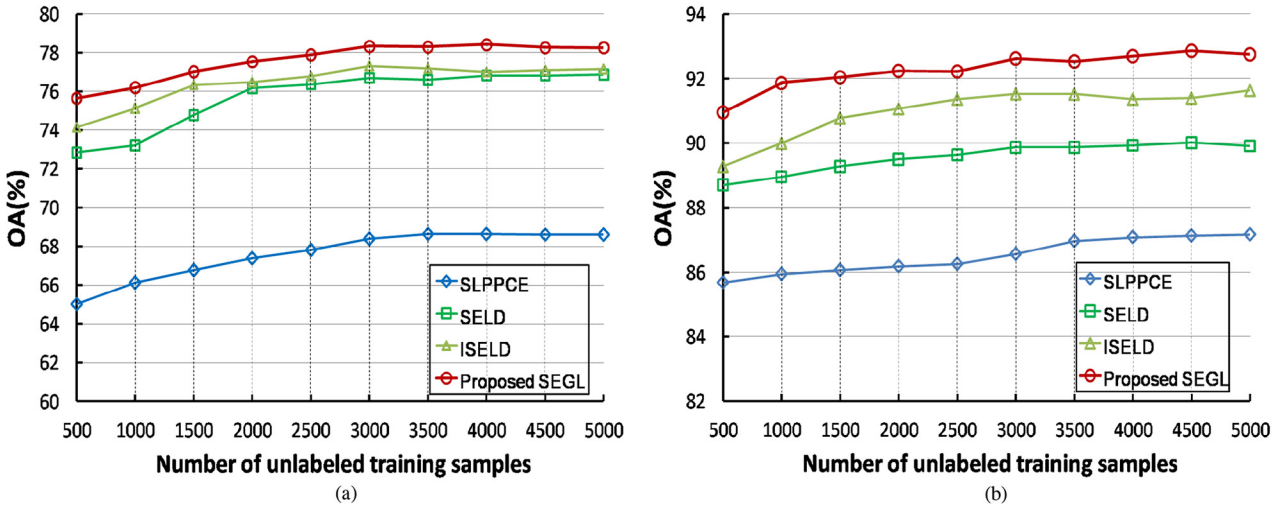


Fig. 5. Averaged OA (%) with the number of unlabeled training samples increasing for different semisupervised feature extraction method with SVM classifier. (a) *University of Pavia*. (b) *KSC*.

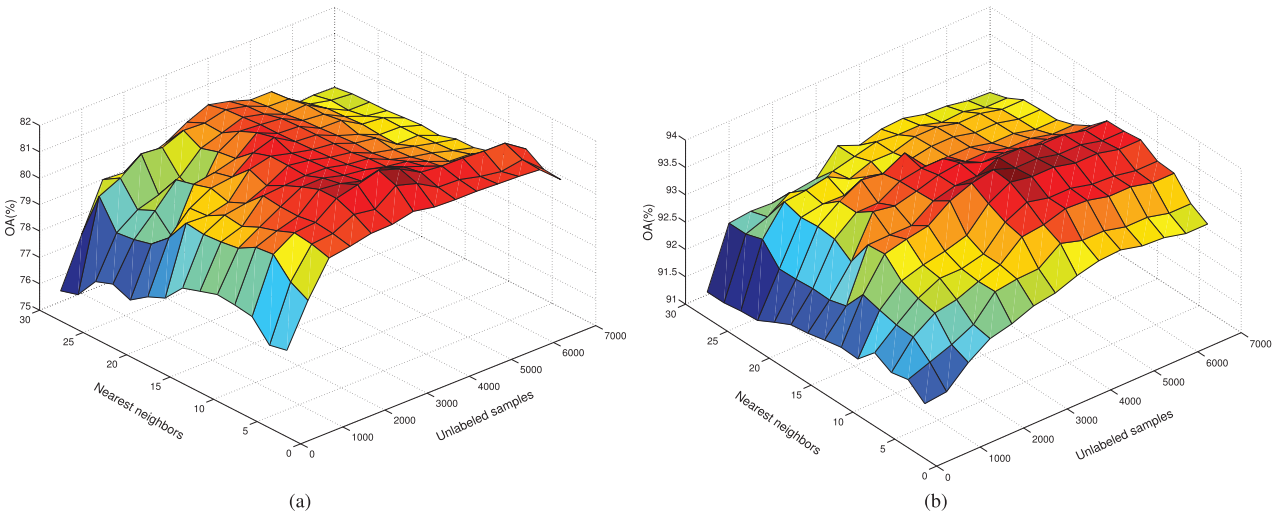
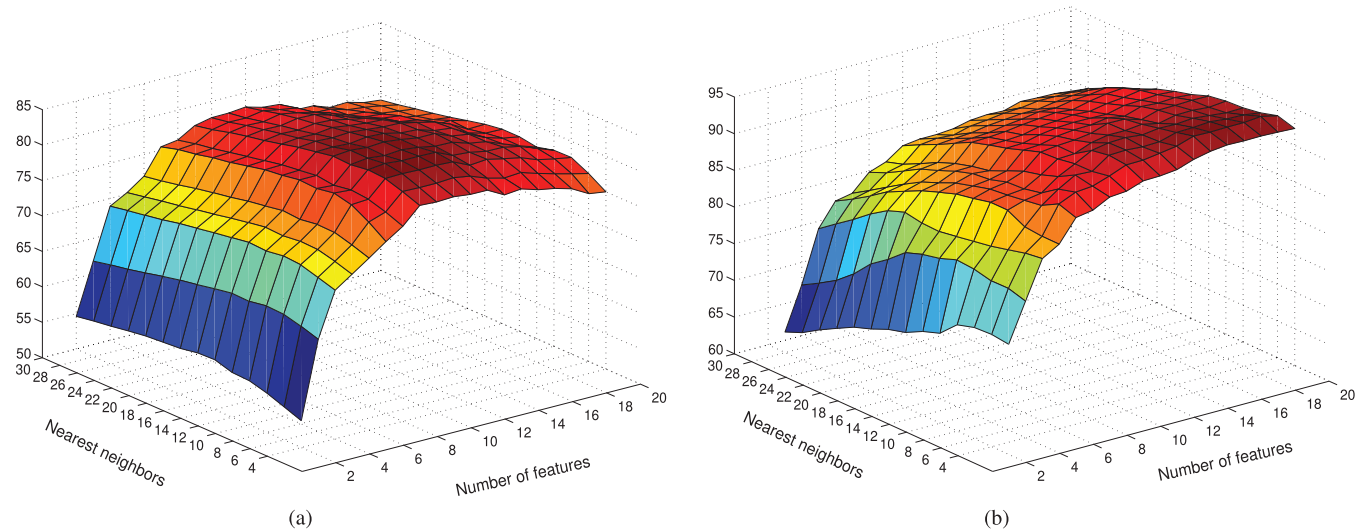


Fig. 6. OA (%) as a function of the number of nearest neighbors and unlabeled samples, with 40 labeled samples per class and 8 extracted features. (a) *University of Pavia*. (b) *KSC*.



(a) (b)

Fig. 7. OA (%) as a function of the number of nearest neighbors and number of features, with 40 labeled samples per class and 2000 unlabeled samples. (a) *University of Pavia*. (b) *KSC*.

Fig. 7 shows the surface of the OA as a function of nearest neighbors and number of features, with SVM classifier. The number of nearest neighbors was changed from 4 to 30 with a step 2, the number of extracted feature was evaluated from 2 to 20 with a step 1. It could be noted from the results that larger nearest neighbors (with a fixed number of the extracted features) will increase the possibility of mislabeling, leading to poor classification performances. What is more, the optimal numbers of features will increase as the raise of nearest neighbors.

IV. Conclusion

In this paper, we present a new feature extraction method with SEGL, and applied it to the classification of HSIs. The proposed method first divides the samples into labeled and unlabeled sets, and we connected labeled samples according to their label information and unlabeled samples by their k-nearest neighbors. For the connection between labeled and unlabeled samples, we sorted the mean distance from a unlabeled sample to each class, and connected the unlabeled sample with labeled samples belonging to the nearest class. Last but not least, the proposed SEGL method sets the weighted edges to connected samples by utilizing distance information between samples. This way our proposed SEGL method better modeled the connections between samples through a general semisupervised graph. Comparing with some related feature extraction methods on four hyperspectral data sets, our proposed SEGL has better performance and higher classification accuracies. Recently, many researchers have combined active learning with semisupervised learning, such as [38]–[40]. In our future research work, we will focus on integrating the active learning into our proposed SEGL method, and unlabeled data sampling (e.g., searching nearest neighborhood samples within a fixed radius, as well as investigating the relationship between the averaged radius and unlabeled samples selection). What is more, in order to better cope with the problems of multimodality, we will explore more criterion to connect unlabled samples with labeled samples instead of using the mean distance.

Acknowledgment

The authors would like to thank P. Gamba for providing both the *Centre Pavia* and *University of Pavia* datasets, Crawford for providing *KSC* and *Botswana* datasets, Cai for providing SDA source code, Sugiyama for providing SELF source code, and Kuo for providing NWFE source code. The authors would also like to thank the members of Image Processing and Interpretation (IPI) research group for supporting this research work.

References

[1] D. Tuia, N. Courty, and R. Flamary, "Feature learning for hyperspectral image classification: Sparse and hierarchical solutions," *ISPRS J. Photogramm. Remote Sens.*, vol. 105, pp. 272–285, 2015.

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

[3] D. A. Landgrebe, "Information extraction principles and methods for multispectral and hyperspectral image data," *in Information Processing for Remote Sensing*, C. H. Chen, Ed. Singapore: World Scientific, 1999, ch. 1.

[4] J. Li *et al.*, "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592–1606, Mar. 2015.

[5] D. Tuia, E. Merenyi, X. Jia, and M. Grana-Romay, "Foreword to the special issue on machine learning for remote sensing data processing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1007–1011, Apr. 2014.

[6] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification," *IEEE Signal Proc. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.

[7] X. Jia, B. Kuo, and M. Crawford, "Feature mining for hyperspectral image classification," *Proc. IEEE*, vol. 101, no. 3, pp. 676–697, Mar. 2013.

[8] J. Schott, "Remote Sensing: The Image Chain Approach. London, U.K.: Oxford Univ. Press, 1996.

[9] L. Ma, M. M. Crawford, and J. Tian, "Local manifold learning-based k-nearest-neighbor for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4099–4109, Nov. 2010.

[10] X. He, C. Deng, S. C. Yan, and H. J. Zhang, "Neighborhood preserving embedding," *in Proc. 10th IEEE Int. Conf. Comput. Vis.*, 2005, vol. 2, pp. 1208–1213.

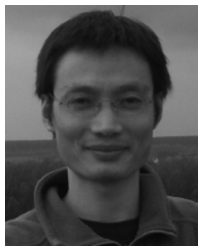
[11] X. He and P. Niyogi, "Locality preserving projections," *in Proc. Adv. Neural Inf. Process. Syst.*, 2004, vol. 16, pp. 153–160.

[12] T. Zhang, J. Yang, D. Zhao, and X. Ge, "Linear local tangent space alignment and application to face recognition," *Neurocomput. Lett.*, vol. 70, pp. 1547–1553, Mar. 2007.

[13] S. Ji and J. Ye, "Generalized linear discriminant analysis: A unified framework and efficient model selection," *IEEE Trans. Neural Netw.*, vol. 19, no. 10, pp. 1768–1782, Oct. 2008.
[14] B. Kuo and D. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 5, pp. 1096–1105, May 2004.
[15] Q. Du, "Modified Fisher¡'s linear discriminant analysis for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 503–507, Oct. 2007.
[16] J. Yang, P. Yu, and B. Kuo, "A nonparametric feature extraction and its application to nearest neighbor classification for hyperspectral image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 3, pp. 1279–1293, Mar. 2010.
[17] H. Huang and B. Kuo, "Double nearest proportion feature extraction for hyperspectral-image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4034–4046, Nov. 2010.
[18] B. Kuo, C. Chang, T. Sheu, and C. Hung, "Feature extractions using labeled and unlabeled data," *in Proc. Int. Geosci. Remote Sens. Symp. (IGARSS)*, 2005, pp. 1257–1260.
[19] J. Li, P. Gamba, and A. Plaza, "A novel semi-supervised method for obtaining fine resolution urban extents exploiting coarser resolution maps," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 10, pp. 4276–4287, Oct. 2014.
[20] W. Liao, R. Bellens, A. Pizurica, W. Philips, and Y. Pi, "Classification of hyperspectral data over urban areas using directional morphological profiles and semi-supervised feature extraction," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 3, pp. 1164–1176, Aug. 2012.
[21] V. N. Vapnik, *Statistical Learning Theory*. Hoboken, NJ, USA: Wiley, 1998.
[22] 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.
[23] G. Camps-Valls, T. Bandos, and D. Zhou, "Semi-supervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.
[24] X. Zhu, "Semi-supervised learning literature," *Dept. Comput. Sci., Univ. Wisconsin-Madison*, Madison, WI, Tech. Rep. 1530, 2008.
[25] D. Zhang, Z. Zhou, and S. Chen. "Semi-supervised dimensionality reduction," *in Proc. 7th SIAM Int. Conf. Data Min.*, 2007, pp. 629–634.
[26] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," *in Proc. IEEE 11th Int. Conf. Comput. Vis. (ICCV07)*, 2007, pp. 1–7.
[27] M. Sugiyama, T. Ide, S. Nakajima, and J. Sese, "Semi-supervised local Fisher discriminant analysis for dimensionality reduction," *Mach. Learn.*, vol. 78, pp. 35–61, 2010.
[28] S. Zhang and G. Yu, "Semi-supervised locality preserving projections with compactness enhancement," *in Proc. Int. Conf. Educ. Inf. Technol. (ICEIT)*, 2010, pp. 460–464.
[29] Z. Ji, W. pang, Y. He, and H. Zhang, "Semi-supervised LPP algorithms for learning-to-rank-based visual search reranking," *Inf. Sci.*, vol. 302, no. 1, pp. 83–93, 2015.
[30] S. G. Chen and D. Q. Zhang, "Semi-supervised dimensionality reduction with pairwise constraints for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 369–373, Mar. 2011.
[31] W. Liao, A. Pizurica, and etc. "Semisupervised local discriminant analysis for feature extraction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 184–198, Jan. 2013.
[32] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed. New York, NY, USA: Academic, 1990.
[33] R. Luo, W. Liao, W. Philips, and Y. Pi, "An improved semi-supervised local discriminant analysis for feature extraction of hyperspectal image," *in Proc. Joint Urban Remote Sens. Event (JURSE)*, 2015, pp. 1–4.
[34] M. Sugiyama *et al.*, "Dimensionality Reduction of Multimodal Labeled Data by Local Fisher Discriminant Analysis," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, 2007.
[35] L. Zelnik and P. Perona. "Self-tuning spectral clustering," *in Advances in Neural Information Processing Systems 17*. Cambridge, MA, USA: MIT Press, 2005, pp. 1601–1608.
[36] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profile," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.
[37] W. Liao, M. Dalla Mura, J. Chanussot, and A. Pizurica, "Fusion of spectral and spatial information for classification of hyperspectral remote sensed imagery by local graph, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, pp. 1–12, 2015
[38] I. Dopido, J. Li, P. R. Marpu, A. Plaza, J. B. Dias, and J. A. Benediktsson, "Semisupervised self-learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 4032–4044, Jul. 2013.
[39] 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.
[40] L. Wan, K. Tang, M. Li, Y. Zhong, and A. K. Qin, "Collaborative active and semisupervised learning for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2384–2396, May 2015.

Renbo Luo (S'16) received the B.S. degree in automation from South China University of Technology (SCUT), Guangzhou, China, in 2010. Currently, he is pursuing the Ph.D. degree both at the School of Automation Science and Enginerring, SCUT, and the Department of Telecommunications and Information Processing, Ghent University, Ghent, Belgium.

His research interests include image processing, object recognition, data fusion, and remote sensing applications.



Wenzhi Liao (S'10–M'14) received the B.S. degree in mathematics from Hainan Normal University, HaiKou, China, in 2006, the Ph.D. degree both in engineering from South China University of Technology, Guangzhou, China, in 2012, and in computer science engineering from Ghent University, Ghent, Belgium, in 2012.

Since 2012, he was a Postdoc with Ghent University. His research interests include pattern recognition, remote sensing, image processing, mathematical morphology, multi-task feature learning, multi-sensor data fusion, and hyperspectral image restoration.

Mr. Liao is a member of the *Geoscience and Remote Sensing Society* (GRSS) and IEEE GRSS Data Fusion Technical Committee (DFTC). He was the recipient of the Best Paper Challenge Awards on both 2013 IEEE GRSS Data Fusion Contest and 2014 IEEE GRSS Data Fusion Contest.

Xin Huang (M'13–SM'14) received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2009 and is working with the State Key Laboratory of Information Engineering in Surveying, Mapping, and Remote Sensing (LIESMARS).

He is currently a Full Professor with Wuhan University, Wuhan, China, where he teaches remote sensing, photogrammetry, image interpretation, etc. He is the Founder and Director of the Institute of Remote Sensing Information Processing (IRSIP), School of Remote Sensing and Information Engineering, Wuhan University. He has authored more than 60 peer-reviewed articles in the international journals. His research interests include hyperspectral data analysis, high-resolution image processing, pattern recognition, and remote sensing applications.

Prof. Huang was the recipient of the Top-Ten Academic Star of Wuhan University in 2009, the Boeing Award for the Best Paper in Image Analysis and Interpretation from the American Society for Photogrammetry and Remote Sensing (ASPRS) in 2010, the New Century Excellent Talents in University from the Ministry of Education of China in 2011, the National Excellent Doctoral Dissertation Award of China in 2012, and the China National Science Fund for Excellent Young Scholars in 2015. In 2011, he was recognized by the IEEE Geoscience and Remote Sensing Society (GRSS) as the Best Reviewer of *IEEE Geoscience and Remote Sensing Letters*. He was the winner of the IEEE GRSS 2014 Data Fusion Contest. He was the lead guest editor of the special issue on information extraction from high-spatial-resolution optical remotely sensed imagery for the *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (vol. 8, no.5, May 2015). Since 2014, he serves as an Associate Editor of the *IEEE Geoscience and Remote Sensing Letters*.

Youguo Pi received the Diploma degree in automation engineering from Chongqing University, Chongqing, China, and the Ph.D. degree in mechanical engineering from South China University of Technology, Guangzhou, China, in 1982 and 1998, respectively.

From July 1998 to June 2002, he was with the Department of Information Technology, Automation Engineering Center, Academy Guangdong Province. Since July 2002, he has been a Full-Time Professor with the College of Automation Science and Engineering, South China University of Technology. His research interests include image processing and patter recognition, motion control, Intelligent Chinese character formation, and servo system.

Wilfried Philips (S'90–M'93–SM'10) was born in Aalst, Belgium, on October 19, 1966. He received the Diploma degree in electrical engineering and the Ph.D. degree in applied sciences from Ghent University, Ghent, Belgium, in 1989 and 1993, respectively.

From October 1989 to October 1997, he was with the Department of Electronics and Information Systems, Ghent University, for the Flemish Fund for Scientific Research (FWO-Vlaanderen), first as a Research Assistant and later as a Postdoctoral Research Fellow. Since November 1997, he has been a Full-Time Professor with the Department of Elecommunications and Information Processing, Ghent University, and is the Head of the Research Group "Image Processing and Interpretation," which is part of the virtual Flemish ICT Research Institute iMinds. His research interests include image and video restoration, analysis and the modeling of image reproduction systems, remote sensing, surveillance, and industrial inspection.