Classification of Hyperspectral Data Over Urban Areas Using Directional Morphological Profiles and Semi-Supervised Feature Extraction

Wenzhi Liao, *Student Member, IEEE*, Rik Bellens, *Student Member, IEEE*, Aleksandra Pižurica, *Member, IEEE*, Wilfried Philips, *Senior Member, IEEE*, and Youguo Pi

***Abstract*—When using morphological features for the classification of high resolution hyperspectral images from urban areas, one should consider two important issues. The first one is that classical morphological openings and closings degrade the object boundaries and deform the object shapes. Morphological openings and closings by reconstruction can avoid this problem, but this process leads to some undesirable effects. Objects expected to disappear at a certain scale remain present when using morphological openings and closings by reconstruction. The second one is that the morphological profiles (MPs) with different structuring elements and a range of increasing sizes of morphological operators produce high-dimensional data. These high-dimensional data may contain redundant information and create a new challenge for conventional classification methods, especially for the classifiers which are not robust to the Hughes phenomenon. In this paper, we first investigate morphological profiles with partial reconstruction and directional MPs for the classification of high resolution hyperspectral images from urban areas. Secondly, we develop a semi-supervised feature extraction to reduce the dimensionality of the generated morphological profiles for the classification. Experimental results on real urban hyperspectral images demonstrate the efficiency of the considered techniques.**

***Index Terms*—Classification, high spatial resolution, hyperspectral data, morphological profiles, semi-supervised feature extraction.**

I. Introduction

RECENT advances in sensors technology have led to an increased availability of hyperspectral data from urban areas at very high both spatial and spectral resolutions. Many techniques are developed to explore the spatial information of the high resolution remote sensing data, in particular, mathematical morphology [1], [2] is one of the most popular methods. Pesaresi and Benediktsson [3] proposed the use of morphological transformations to build a morphological profile (MP). Bellens *et al.* [4] further explored this approach by using both disk-shaped and linear structuring elements to improve the classification of very high-resolution panchromatic urban imagery. The approach of [5] extended the method in [1] for hyperspectral data with high spatial resolution. The resulting method built the MPs on the first principal components (PCs) extracted from a hyperspectral image, leading to the definition of extended MP (EMP). The appoach of [6] performs spectral-based morphology using the full hyperspectral image without dimensionality reduction. In [7], kernel principal components are used to construct the EMP, with significant improvement in terms of classification accuracies compared with the conventional EMP built on PCs. In [8], the attribute profiles (APs) [9] were applied to the first PCs extracted from a hyperspectral image, generating an extended AP (EAP). The approach of [10] improved the classification results by constructing the EAP with the independent component analysis.

When using MPs, one should consider two important issues. The first one is that classical morphological openings and closings degrade the object boundaries and deform the object shapes, which may result in losing some crucial information and introducing fake objects in the image. To avoid this problem, one often uses morphological openings and closings by reconstruction [3], [5], [11]–[13], which can reduce some shape noise in the image. However, morphological openings and closings by reconstruction lead to some unexpected results for remote sensing images, such as over-reconstruction, as was discussed in [4]. Objects which are expected to disappear at a certain scale remain present when using morphological openings and closings by reconstruction. The approach of [4] proposed a partial reconstruction for the classification of very high-resolution panchromatic urban imagery. Morphological openings and closings by partial reconstruction can solve the problem of over-reconstruction while preserving the shape of objects as much as possible. They limit the extent of the reconstruction. The edges of simple objects are reconstruct well, but a full retrieval of complex elongated shapes might not be obtained. For simple objects like rectangles for example, the reconstruction is complete. Since, in urban remote sensing scenes, most objects are not very complex and even rectangular shaped, partial reconstruction is very well suited.

Manuscript received October 31, 2011; revised January 24, 2012; accepted February 15, 2012. Date of publication May 17, 2012; date of current version July 20, 2012. This work was supported by the FWO project "Spatial extension for classification of multispectral images" and China Scholarship Council.

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

R. Bellens, A. Pižurica, and W. Philips are with the Department of Telecommunications and Information Processing, Ghent University, 9000 Ghent, Belgium.

Y. Pi is with the School of Automation Science and Engineering, South China University of Technology, 510640 Guangzhou, China.

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.2012.2190045

The second problem is that the resulting data sets may contain redundant information, because the construction of the generated profiles is based on different structuring elements (SEs) and a range of increasing sizes of morphological operators. Furthermore, the increase in the dimensionality of the generated profiles may create a new challenge for conventional classification methods, especially for the classifiers which are not robust to the Hughes phenomenon [14] (for a limited number of training samples, the classification accuracy decreases as the dimension increases). Although some advanced classifiers, such as neural networks [5], SVM [12], [15] and random forest classifiers [15], are shown to deal efficiently with these high dimensional data sets, common statistical classifiers are often limited in this context. For this reason, feature extraction (FE), aiming at reducing the dimensionality of data while keeping as much intrinsic information as possible, is a desirable preprocessing tool to reduce the dimensionality of the generated profiles for classification. Relatively few bands can represent most information of the data, making feature extraction very useful for classification of remote sensing data [7], [16]. The effect of different FE methods on reducing the dimensionality of the generated profiles for classification of hyperspectral data from urban areas has been discussed in several studies [5], [12], [15], [17].

However, to the best of our knowledge the use of semi-supervised FE methods for the generated morphological profiles has not been investigated yet. In many real world applications, it is usually difficult, expensive and time-consuming to collect sufficient amount of labeled samples. Meanwhile, it is much easier to obtain unlabeled samples. For this reason, semi-supervised methods [18]–[20], [34]–[36], which aim at improved classification by utilizing both unlabeled and limited labeled data gained popularity in the machine learning community.

In this paper, we first investigate the effect of the morphological profiles with partial reconstruction and the effect of directional morphological profiles [4] on the classification of hyperspectral images from urban areas. Secondly, we develop a semi-supervised FE method as a preprocessing to reduce the dimensionality of the generated morphological profiles for classification.

The organization of the paper is as follows. Section II gives a brief review of the morphological profiles with partial reconstruction and the directional morphological profiles. In Section III, we present our proposed method which applies semi-supervised feature extraction to reduce the dimensionality of the generated morphological profiles for classification. The experimental results on real urban hyperspectral images are presented and discussed in Section IV. Finally, the conclusions of the paper are drawn in Section V.

II. Morphological Features

Morphological operators act on the values of the pixels according to transformations that consider the neighborhood (with a given size and shape) of the pixels. The basic operators are dilation and erosion [1]. These operators are applied to an image with a set of known shapes, called the structuring elements. In the case of erosion, a pixel takes the minimum value of all the pixels in its neighborhood, defined by the SE. By contrast, dilation takes the maximum value of all the pixels in its neighborhood. Dilation and erosion are usually employed in pairs, either dilation of an image followed by erosion of the dilated result, or erosion of an image followed by dilation of the eroded result. These combinations are known as opening and closing. An opening acts on bright objects compared with their surrounding, while closings act on dark objects. For example, an opening deletes (this means the pixels in the object take on the value of their surrounding) bright objects that are smaller than the SE. By increasing the size of the SE, more and more objects are removed. We will use the term scale of an opening or closing to refer to this size. A vector containing the pixel values in openings and closings by reconstruction of different scales is called the morphologic profile. The MPs carries information about the size and the shape of objects in the image.

Aside from deleting objects smaller than the SE, morphological openings and closings also deform the objects which are still present in the image, see Fig. 1(a) and Fig. 2(a), the corners of rectangular objects in Fig. 1(a) (square object on the top right) are rounded. To preserve the shapes of objects, morphological openings and closings by reconstruction are generally the tool of choice [11], [12]. This process reconstructs the whole object if at least one pixel of the object survives the opening or closing. We can see the results in Fig. 1(b) and Fig. 2(b), the shapes of the objects are well preserved, and some small objects disappear as the scale (here the scale is related to the size of the SE) increases. However, an MP with reconstruction will lead to some undesirable effects (such as over-reconstruction), a lot of objects that disappeared in the MP without reconstruction remain present in the MP with reconstruction. Objects which are expected to disappear in the image at a low scale, are still present at the highest scales, as shown in Fig. 1(b) (small bright road on the middle left) and Fig. 2(b) (small black road on the middle right). The approach of [4] proposed a partial reconstruction to solve the problem of over-reconstruction while preserving the shape of objects as much as possible, and made a great improvement in the classification of very high-resolution panchromatic urban imagery. In the partial reconstruction process, a pixel is only reconstructed if it is connected to a pixel that was not erased, and this second pixel lies within a certain geodesic distance $dist$ from the pixel. The geodesic distance between two pixels is the length of the shortest path between the two pixels that lie completely within the object. The parameter $dist$ sets the amount of reconstruction. For disk shaped SE, this amount can be chosen such that rectangular objects are completely reconstructed. For linear SE, the choose of a good value is more difficult. However, 10% of the lenght of the SE seems a good value [4]. Fig. 1(c) and Fig. 2(c) show the results of MP with partial reconstruction in different scales. The shapes of objects are better preserved with partial reconstruction compared to the MP without reconstruction. Some of the more complex shapes are not so well preserved as with geodesic reconstruction. On the other hand, a lot of small objects which remain present

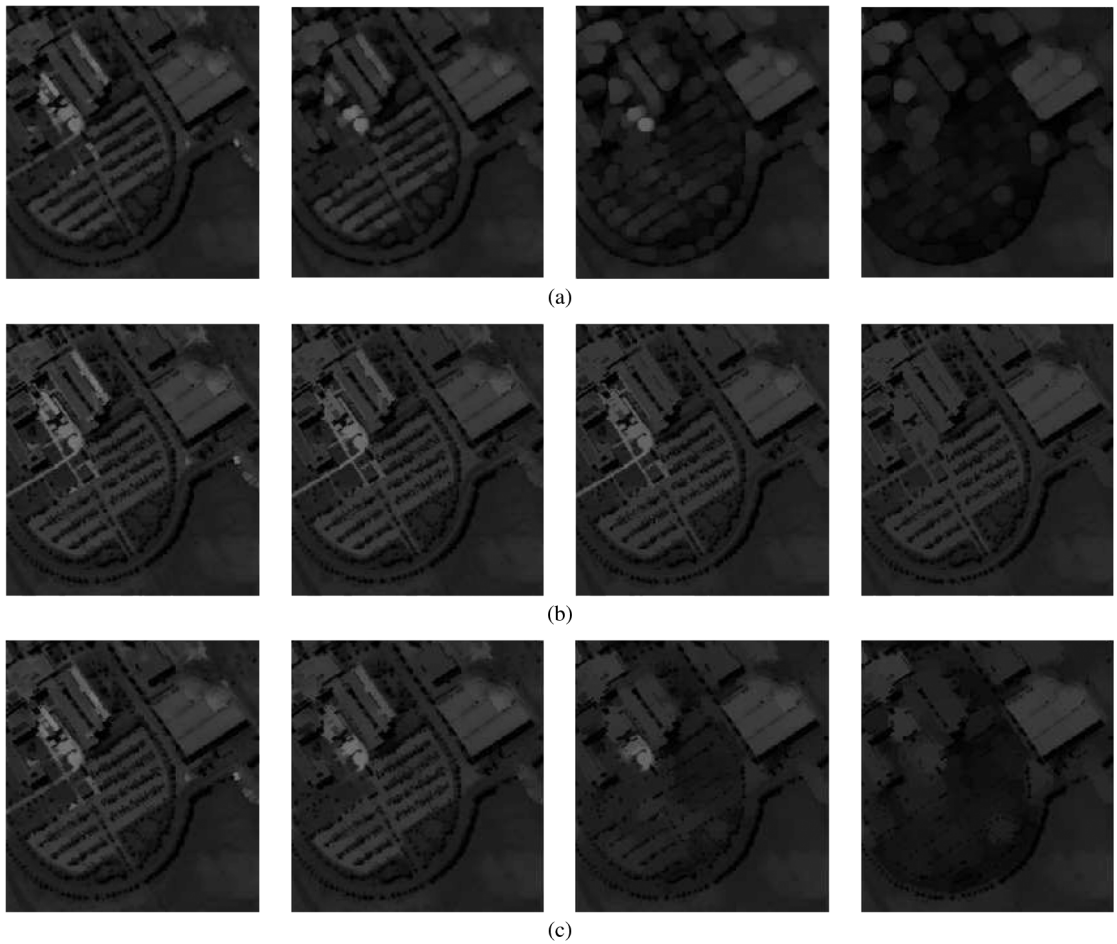


Fig. 1. Openings with disk-shaped SEs of increasing size. The scales of SEs vary from 2 to 8, with step 2. The image processed is part of the first PC extracted from University Area data set in Fig. 3(a). (a) Without reconstruction; (b) geodesic reconstruction; (c) partial reconstruction.

in the MP with reconstruction, now disappear in the case with partial reconstruction. Basically this is because in remote sensing (urban) scenes different objects lie closely together and because of noise and other effects, different objects are often connected by a sequence of pixels with similar (or more extreme) pixel values. Therefore, reconstruction considers all those connected objects as a single object and objects will only disappear when the SE does not fit the broadest (for disk shapes) or longest (for directional) part of the connected object, even though this part might be far away from the actual object. Partial reconstruction only reconstructs the immediate surrounding of the surveiving part.

Because of its isotropic character morphological openings and closings with disk-shaped SEs are the most popular methods used in current literature [5], [12], [15]. Objects where the SE does not fit are deleted from the image. For disk shaped SE this means objects where the smallest objects size (i.e. the width) is smaller than the diameter of the disk. Closings and openings with disk-shaped SEs thus act on the minimum size of objects. This results in an disk-based MP carrying information about the minimum size of objects. Fig. 1(c) shows the result of the opening transform with partial reconstruction for different-sized, disk-shaped SEs. As the size of the SE increases, more and more bright objects disappear in the dark background. The size of the SE that makes objects disappear corresponds to the minimum size of the object. In [4], directional MP was proposed to obtain an indication of the maximum size of objects. With a linear structuring element of length L and orientation θ, an opening (resp. closing) deletes bright (resp. dark) objects (or object parts) which are smaller than that length in that direction. When performing such openings (or closings) with different orientations, objects which are shorter than L will be completely removed in all of these images. The maximum (resp. minimum) over all of these openings (resp. closings) will therefore remove the short objects (or object parts) and keep the long objects. Creating multiple such maximum or minimum images for different lengths L gives you the directional MP. Thus the directional MP carries information about the maximum size of objects. This information can be used for detecting linear objects (roads), since these objects have large maximum sizes and small minimum sizes. Fig. 2(c) shows an example of the directional MP with partial reconstruction. Note that individual houses disappear at lower scales, while roads and apartment buildings with a more elongated shape have almost constant intensities. For more details on MP with partial reconstruction and directional MP, the readers should consult [4]. This paper investigates MP with par-

tial reconstruction and directional MP of [4] for the classification of high resolution hyperspectral images from urban areas.

The morphologic profiles with a certain SE produce a vector of values, each value corresponding with the feature output for a specific scale. While morphologic profiles with different SEs will then be a high-dimensional stacked vector. The resulting high-dimensional data may contain redundant information. Furthermore, if we use these high-dimensional data as an input feature for classification, this may create a challenge for conventional classification methods. Therefore, we will use feature extraction as a preprocessing to reduce the dimensionality of the generated morphological profiles before classification.

III. Proposed Semi-Supervised Feature Extraction for Morphological Profiles

A number of approaches exist for feature extraction of the generated morphological profiles [5], [12], [15], [17], ranging from unsupervised methods to supervised ones. One of the best known unsupervised methods is Principle Component Analysis (PCA) [28], which is widely used [5], [12], [13]. Green *et al.* [29] introduced the minimum noise fraction (MNF) transformation. Recently, some local methods, which preserve the properties of local neighborhoods were used to reduce the dimensionality of hyperspectral images [30]–[32], such as Locally Linear Embedding [31], Neighborhood Preserving Embedding (NPE) [33]. By considering neighborhood information around the data, these local methods can preserve local neighborhood information and detect the manifold embedded in the high-dimensional feature space.

Supervised methods rely on the existence of labeled samples to infer class separability. Two widely used supervised feature extraction methods are the Fisher Linear discriminant analysis (LDA) [22] and nonparametric weighted feature extraction (NWFE) [23]. Many extensions to these two methods have been proposed in recent years, such as modified Fisher's linear discriminant analysis [24], regularized linear discriminant analysis [25], modified nonparametric weight feature extraction using spatial and spectral information [26], and kernel nonparametric weighted feature extraction [27].

However, to the best of our knowledge, the use of semi-supervised FE methods for the generated morphological profiles has not been investigated yet. In real-world applications, labeled samples are usually very limited, while unlabeled ones are available in large quantities at very low cost. Recently, some semi-supervised feature extraction methods were proposed to reduce the dimension of hyperspectral data sets. The approach of [20] proposed a general semi-supervised dimensionality reduction framework based on pairwise constraints, which employs regularization with sparse representation. In an earlier work [21], we proposed a semi-supervised local discriminant analysis (SELD) method, which combines LDA and NPE, to extract features from the original hyperspectral data. In this paper, we propose a generalized SELD (GSELD) to extract features from the generated morphological profiles.

Let $\{\mathbf{x}_i\}_{i=1}^N$, $\mathbf{x}_i \in R^d$ denote high-dimensional data, $\{\mathbf{z}_i\}_{i=1}^N$, and $\mathbf{z}_i \in R^r$ the low-dimensional representations of the high-dimensional data $r \leq d$. In our application, d is the dimensionality of the generated profiles, and r is the dimensionality of the extracted features. The goal of linear feature extraction is to find a $d \times r$ projection matrix $\mathbf{W}$, which can map every high-dimensional data $\mathbf{x}_i$ to $\mathbf{z}_i = \mathbf{W}^T\mathbf{x}_i$ such that most information of the high-dimensional data is kept in a much lower dimensional feature space.

A. Background of Some Related Methods

1) Linear Discriminant Analysis (LDA): As a supervised method, LDA seeks directions on which the ratio of the between-class covariance to within-class covariance is maximized. The objective function of LDA is as follows:

$$\mathbf{w}_{LDA} = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^T\mathbf{S}_b\mathbf{w}}{\mathbf{w}^T\mathbf{S}_w\mathbf{w}} \tag{1}$$

$$\mathbf{S}_b = \sum_{k=1}^{C} n_k(\mathbf{u}^{(k)} - \mathbf{u})(\mathbf{u}^{(k)} - \mathbf{u})^T \tag{2}$$

$$\mathbf{S}_w = \sum_{k=1}^{C}\left(\sum_{i=1}^{n_k}\left(\mathbf{x}_i^{(k)} - \mathbf{u}^{(k)}\right)\left(\mathbf{x}_i^{(k)} - \mathbf{u}^{(k)}\right)^T\right) \tag{3}$$

where n_k is the number of samples in the kth class, $\mathbf{u}$ is the mean of the entire training samples, $\mathbf{u}^{(k)}$ is the mean of the kth class, $\mathbf{x}_i^{(k)}$ is the ith sample in the kth class, C is the number of classes. $\mathbf{S}_b$ is called the between-class scatter matrix and $\mathbf{S}_w$ the within-classes scatter matrix.

2) Neighborhood Preserving Embedding (NPE): NPE [33] seeks a projection direction on which neighborhood data points in the high-dimensional feature space are kept neighborhood in the low-dimensional feature space as well. NPE first finds k nearest neighbors (kNN) for each data point $\mathbf{x}_i$; then calculates the reconstruction weights Q_{ij} by minimizing the reconstruction error, which results from approximating $\mathbf{x}_i$ by its k nearest neighbors:

$$\min \sum_i \left\| \mathbf{x}_i - \sum_{j=1}^{k} Q_{ij}\mathbf{x}_j \right\|^2 \quad S.t. \quad \sum_{j=1}^{k} Q_{ij} = 1 \tag{4}$$

The extracted features $\mathbf{z}_i$ in the low-dimensional projected subspace that best preserve the local neighborhood information are then obtained as:

$$\min \sum_i \left\| \mathbf{z}_i - \sum_{j=1}^{k} Q_{ij}\mathbf{z}_j \right\|^2 \quad S.t. \quad \mathbf{z}_i^T\mathbf{z}_i = \mathbf{I} \tag{5}$$

The objective function of NPE is as follows:

$$\mathbf{w}_{NPE} = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^T\mathbf{X}\mathbf{X}^T\mathbf{w}}{\mathbf{w}^T\mathbf{X}\mathbf{M}\mathbf{X}^T\mathbf{w}} \tag{6}$$

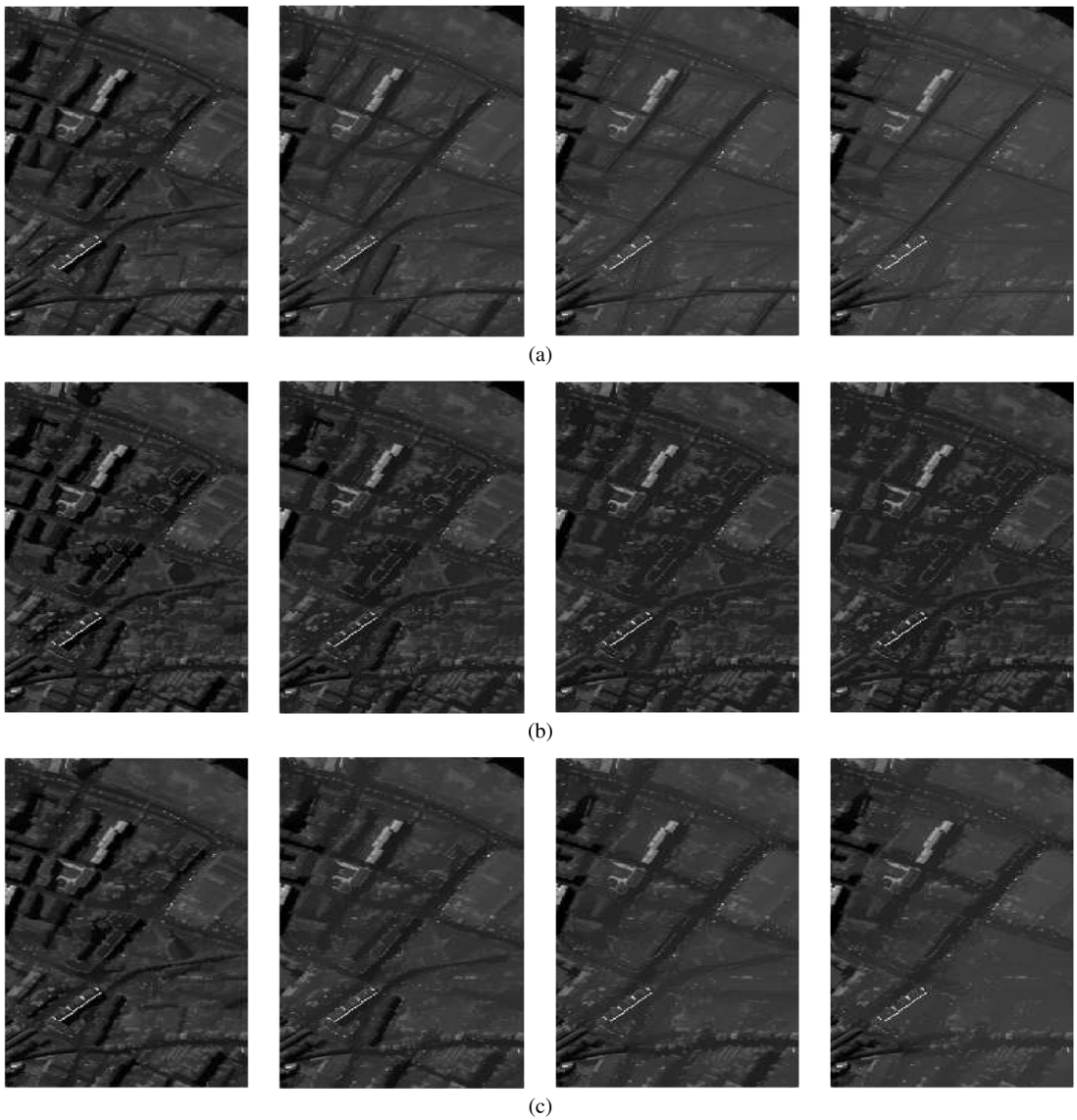


Fig. 2. Closings with linear SEs of increasing size. The scales of SEs vary from 20 to 80, with step 20. The image processed is part of the first PC extracted from Pavia Centre data set in Fig. 4(a). (a) Without reconstruction; (b) geodesic reconstruction; (c) partial reconstruction.

where $\mathbf{X}$ is the training set, $\mathbf{M} = (\mathbf{I}-\mathbf{Q})^T(\mathbf{I}-\mathbf{Q})$, $\mathbf{I}$ represents the identity matrix and $\mathbf{Q}$ is the reconstruction weights matrix [33]. The solution to (1) and (6) is equivalent to solving the following generalized eigenvalue problem:

$$\overline{\mathbf{S}}\mathbf{w} = \lambda \underline{\mathbf{S}}\mathbf{w} \tag{7}$$

For LDA method, the $\overline{\mathbf{S}} = \mathbf{S}_b$ and $\underline{\mathbf{S}} = \mathbf{S}_w$. For NPE method, $\overline{\mathbf{S}} = \mathbf{X}\mathbf{X}^T$ and $\underline{\mathbf{S}} = \mathbf{X}\mathbf{M}\mathbf{X}^T$. The projection matrix $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_r)$ is made up by the r eigenvectors of the matrix $\underline{\mathbf{S}}^{-1}\overline{\mathbf{S}}$ associated with the largest r eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r$.

B. The Proposed Method

Focusing on class discrimination, LDA is in general well suited to preprocessing for the task of classification, since the transformation improves class separation. However, when only a small number of labeled samples are available, LDA tends to perform poorly due to overfitting. Moreover, as the rank of the between-class scatter matrix S_b is $C-1$, the LDA can extract at most $C-1$ features, which is not always sufficient to represent essential information of the original data. NPE works directly on the data without any ground truth, and incorporates the local neighborhood information of data points in its feature extraction process. In our earlier work [21], we combined LDA and NPE in a new framework, and proposed a semi-supervised local discriminant analysis (SELD) method to extract features from the original hyperspectral data. SELD magnified the advantages of LDA and NPE, and compensated for disadvantages of the two at the same time. In this paper, we propose a new semi-supervised method to extract features from the generated morphological profiles. The proposed method extends SELD [21] with a tunable parameter and we abbreviate this generalized SELD method as GSELD.

Suppose a training data set is made up of the labeled set $\mathbf{X}_{labeled} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{n}$, $y_i \in \{1, 2, \cdots, C\}$, C is the number

of classes, and unlabeled set $\mathbf{X}_{unlabeled} = \{\mathbf{x}_i\}_{i=n+1}^{N}$. The kth class has n_k samples with $\sum_{k=1}^{C} n_k = n$. Without loss of generality, we center the data points by subtracting the mean vector from all the sample vectors, and assume that the labeled samples in $\mathbf{X}_{labeled} = \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n\}$ are ordered according to their labels, with data matrix of the kth class $\mathbf{X}^{(k)} = [\mathbf{x}_1^{(k)}, \mathbf{x}_2^{(k)}, \cdots, \mathbf{x}_{n_k}^{(k)}]$ where $\mathbf{x}_i^{(k)}$ is the ith sample in the kth class. Then the labeled set can be expressed as $\mathbf{X}_{labeled} = [\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \cdots, \mathbf{X}^{(n)}]$, all training set $\mathbf{X} = [\mathbf{X}_{labeled}, \mathbf{X}_{unlabeled}]$. The optimization problem of the proposed GSELD is:

$$\mathbf{w}_{GSELD} = \arg\max_{\mathbf{w}} \frac{\mathbf{w}^T \mathbf{X}(\alpha \mathbf{P} + \underline{\mathbf{I}})\mathbf{X}^T \mathbf{w}}{\mathbf{w}^T \mathbf{X}\left(\alpha(\overline{\mathbf{I}} - \mathbf{P}) + \mathbf{M}\right)\mathbf{X}^T \mathbf{w}} \tag{8}$$

where the matrices $\mathbf{P}$ and $(\overline{\mathbf{I}} - \mathbf{P})$ are from the reformulation of LDA part, and the matrices $\underline{\mathbf{I}}$ and $\mathbf{M}$ are from the reformulation of NPE part, for more details, the readers should consult [21]. α is the tunable parameter. When α is set to zero, (8) reduces to (6). When the parameter α is set to 1, the proposed method reduces to SELD [21]. Let $\overline{\mathbf{S}}_{GSELD} = \mathbf{X}(\alpha \mathbf{P} + \underline{\mathbf{I}})\mathbf{X}^T$ and $\underline{\mathbf{S}}_{GSELD} = \mathbf{X}(\alpha(\overline{\mathbf{I}} - \mathbf{P}) + \mathbf{M})\mathbf{X}^T$, we can solve the generalized eigenvalue problem of GSELD as (7), and get the projection matrix $\mathbf{W}$.

The algorithmic procedure of the proposed method which uses GSELD to extract features from the generated MPs is formally stated below:

1) Use PCA to extract the most p significant principal components (usually with cumulative variance near to 99%) from the original hyperspectral data sets.
2) Build the MPs on the p extracted PCs. Actually, the MP are defined the same way as [4], [5]. An MP consist of the original image (one of the PC features) and M openings with SE of increasing size (all applied on the original image) and M closings with the same SE. Then, an Extended Morphological Profile (EMP) is obtained with $d = p \times (2M + 1)$ dimension.
3) Divide the training samples into two subsets. Suppose that the labeled samples in $\mathbf{X}_{labeled} = [\mathbf{x}_1, \cdots, \mathbf{x}_n]$ are ordered according to their labels, with data matrix of the kth class $\mathbf{X}^{(k)} = [\mathbf{x}_1^{(k)}, \cdots, \mathbf{x}_{n_k}^{(k)}]$ where $\mathbf{x}_i^{(k)}$ is the ith sample in the kth class, then the labeled set can be expressed as $\mathbf{X}_{labeled} = [\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \cdots, \mathbf{X}^{(C)}]$. The unlabeled set is denoted as $\mathbf{X}_{unlabeled} = \{\mathbf{x}_i\}_{i=n+1}^{N}$.
4) Construct the matrices $\mathbf{P}$ and $\overline{\mathbf{I}}$ from the labeled samples, and construct the matrix $\underline{\mathbf{I}}$ and $\mathbf{M}$ from the unlabeled samples.
5) Compute the eigenvectors and eigenvalues for the generalized eigenvector problem in (6). The projection matrix $\mathbf{W}_{GSELD} = (\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_r)$ is made up by the r eigenvectors of the matrix $\underline{\mathbf{S}}_{GSELD}^{-1} \overline{\mathbf{S}}_{GSELD}$ associated with the largest r eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r$.
6) Project the high dimensional generated morphological profiles ($\mathbf{x}_i \in R^d$) into a lower dimensional subspace ($\mathbf{z}_i \in R^r$) by

$$\mathbf{x} \rightarrow \mathbf{z} = \mathbf{W}_{GSELD}^T \mathbf{x}$$

7) Use these extracted features $\mathbf{Z}$ in the lower dimensional subspace as an input to do classification.

IV. Experimental Results

A. Hyperspectral Image Data Sets

Experiments were run on three data sets, namely the '*Pavia Center*', '*University Area*' and '*Indian Pine*'.The first two data sets 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 very fine spatial resolution of 1.3 meters by pixel.

Pavia Center: The data with 1096 × 492 pixels was collected over Pavia city center, Italy. It contains 102 spectral channels after removal of noisy bands (see Fig. 3(a) for a color composite). Nine groundtruth classes were considered in experiments, see Table I. Note that the color in the cell denotes different classes in the classification maps (Fig. 3 and Fig. 4).*University Area*: The data with 610 × 340 pixels was collected over the University of Pavia, Italy. It contains 103 spectral channels after removal of noisy bands (see Fig. 4(a) for a color composite). The data also includes 9 land cover/use classes, see Table I.

Indian Pine: The data set was captured by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over northwestern Indiana in June 1992, with 220 spectral bands in the wavelength range from 0.4 to 2.5 μm and low spatial resolution of 20 meters by pixel. The calibrated data are available online (along with detailed ground-truth information) from http://cobweb.ecn.purdue.edu/~biehl/. The whole scene, consisting of the full 145 × 145 pixels, which contains 16 classes, ranging in size from 20 to 2468 pixels. 9 classes were selected for the experiments, 70% of labeled samples are randomly selected as training set, the rest 30% of the labeled samples are assigned to the test set.

B. Experimental Setup

To apply the morphological profiles with partial reconstruction and directional morphological profiles of [4] from panchromatic imagery to hyperspectral images, principal component analysis (PCA) was first applied to the original hyperspectral data set, and the first 3 principal components (PCs) were selected (representing 99% of the cumulative variance) to construct the MPs. For disk-shaped structuring elements, morphological profiles with 15 openings and closings (ranging from 1 to 15 with step size increment of 1) were then computed for each PC. For linear structuring elements, morphological profiles with

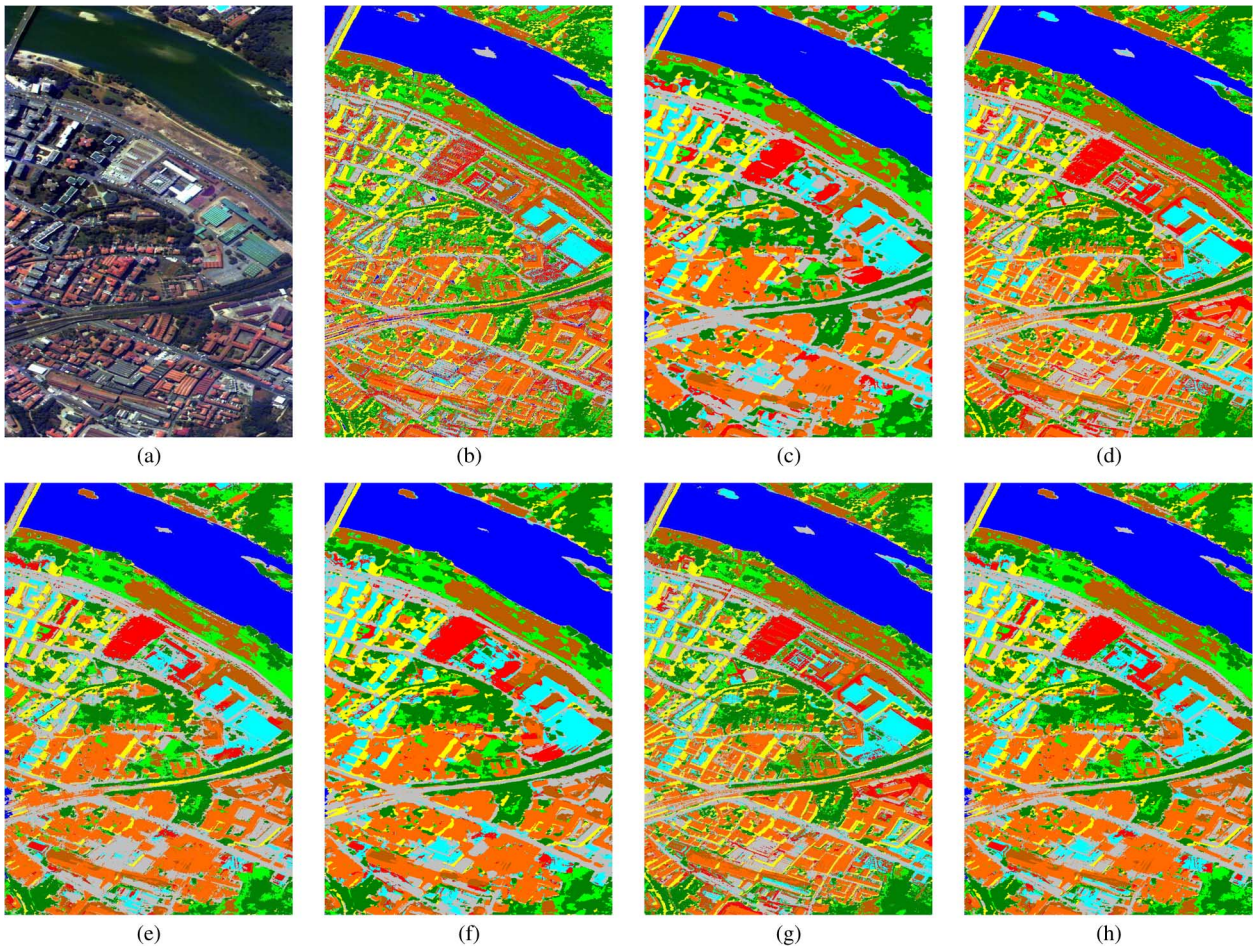


Fig. 3. Classification maps for Pavia Center with best classification accuracy over ten runs, 20 training samples per class with SVM classifier were used. (a) False color image, and thematic map using (b) Spectral Only, (c) Disk-based MP without reconstruction, (d) Disk-based MP with reconstruction, (e) Disk-based MP with partial reconstruction, (f) Disk- and linear-based MP without reconstruction, (g) Disk- and linear-based MP with reconstruction, and (h) Disk- and linear-based MP with partial reconstruction.

TABLE I
TRAINING AND TEST SAMPLES FOR DATA SETS USED IN THE EXPERIMENTS

Pavia Center			University Area			Indian Pine		
Class Name	# Training set	# Test set	Class Name	# Training set	# Test set	Class Name	# Training set	# Test set
Water	745	65278	Asphalt	548	6641	Corn-notill	1004	430
Trees	785	6508	Meadows	540	18649	Corn-min	584	250
Meadows	797	2905	Gravel	392	2099	Grass/Pasture	348	149
Bricks	485	2140	Trees	524	3064	Grass/Trees	523	224
Soil	820	6549	Metal Sheets	265	1345	Hay-windrowed	343	146
Asphalt	678	7585	Soil	532	5029	Soybeans-notill	678	290
Bitumen	808	7287	Bitumen	375	1330	Soybeans-min	1728	740
Tiles	223	3122	Bricks	514	3682	Soybeans-clean	430	184
Shadows	195	2165	Shadows	231	947	Woods	906	388

only 15 closings (ranging from 10 to 150 with step size increment of 10) were constructed for each PC, since objects like roads in the extracted PCs proved to be mostly dark compared to the background, we only made use of closing transforms. As a result, each disk-based profile was made up of 31 bands and the final disk-based MPs, constructed using three principal components, consisted of 93 bands. The final MPs based on both disk and linear SEs were 138 bands.

We used three common classifiers: 1-nearest neighbor (1 NN), linear discriminant classifier (LDC) [37], and support vector machines (SVM) [38]. The SVM classifier with radial basis function (RBF) kernels in Matlab SVM Toolbox, LIBSVM [39], is applied in our experiments. SVM with RBF kernels has two parameters: the penalty factor C and the RBF kernel width γ. We apply a grid-search on C and γ using 5-fold cross-validation to find the best C within the given set

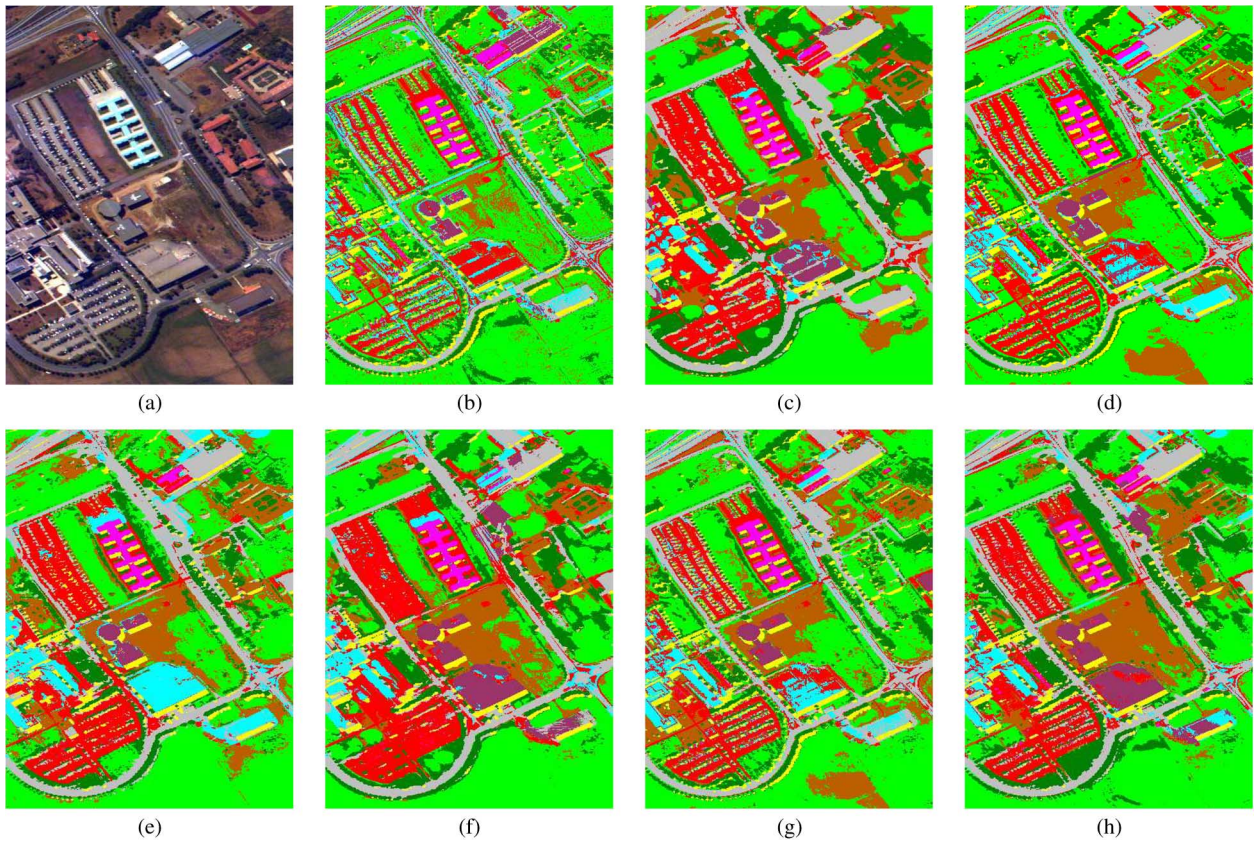


Fig. 4. Classification maps for University Area with best classification accuracy over ten runs, 20 training samples per class with SVM classifier were used. (a) False color image, and thematic maps using (b) Spectral Only, (c) Disk-based MP without reconstruction, (d) Disk-based MP with reconstruction, (e) Disk-based MP with partial reconstruction, (f) Disk- and linear-based MP without reconstruction, (g) Disk- and linear-based MP with reconstruction, and (h) Disk- and linear-based MP with partial reconstruction.

TABLE II
OVERALL ACCURACY IN A CLASSIFICATION WITH SPECTRAL ONLY COMPARED TO CLASSIFICATIONS WITH DISK-BASED MPS WITHOUT RECONSTRUCTION, WITH RECONSTRUCTION, AND WITH PARTIAL RECONSTRUCTION

Dataset	Methods	Classifer	Training Set Size				
			10	20	40	80	160
Pavia Center	Spectral Only	1NN	0.941	0.948	0.952	0.956	0.961
		SVM	0.935	0.942	0.946	0.947	0.948
	No Re.	1NN	0.943	0.956	0.96	0.964	0.965
		SVM	0.961	0.968	0.974	0.977	0.979
	Re.	1NN	0.96	0.97	**0.976**	**0.981**	**0.983**
		SVM	**0.966**	0.968	0.974	0.979	0.981
	Partial Re.	1NN	0.949	0.963	0.967	0.971	0.973
		SVM	0.963	**0.971**	**0.976**	0.98	0.981
University Area	Spectral Only	1NN	0.626	0.637	0.644	0.678	0.69
		SVM	0.653	0.729	0.725	0.734	0.787
	No Re.	1NN	0.818	0.826	0.833	0.841	0.837
		SVM	0.825	0.884	0.886	0.896	0.894
	Re.	1NN	0.75	0.782	0.786	0.823	0.825
		SVM	0.709	0.766	0.799	0.797	0.802
	Partial Re.	1NN	0.806	0.806	0.809	0.829	0.821
		SVM	**0.835**	**0.894**	**0.909**	**0.916**	**0.917**

$\{10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}\}$ and the best γ within the given set $\{10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}\}$.

In order to investigate the influences of the training samples size in more detail, the training data sets were then randomly subsampled to create samples whose sizes corresponded to five distinct cases: 10, 20, 40, 80 and 160 samples per class, respectively. All classifiers were evaluated against the testing sets, the results were averaged over five runs. The word 'Reconstruction' in the tables is shortened as 'Re.'.

C. Results Using Morphological Profiles With Partial Reconstruction and Directional MPs

We compared the MPs with reconstruction, without reconstruction, and with partial reconstruction in both two data sets. We also compared the results with the directional MPs. Since Gaussian Classifier LDC is not efficient to deal with high-dimensional data, we use 1 NN and SVM classifiers in this experiment. The resulting accuracies are shown in Table II, Table III.

TABLE III
OVERALL ACCURACY COMPARISON IN A CLASSIFICATION AMONG DISK-& LINEAR-BASED MPS WITHOUT RECONSTRUCTION, WITH RECONSTRUCTION, AND WITH PARTIAL RECONSTRUCTION

Dataset	Methods	Classifer	Training Set Size				
			10	20	40	80	160
Pavia Center	No Re.	1NN	0.953	0.962	0.968	0.971	0.973
		SVM	0.966	0.972	0.976	0.981	0.981
	Re.	1NN	0.961	0.971	0.976	0.981	**0.983**
		SVM	0.967	0.972	0.975	0.979	0.981
	Partial Re.	1NN	0.957	0.968	0.974	0.976	0.978
		SVM	**0.968**	**0.975**	**0.979**	**0.982**	0.982
University Area	No Re.	1NN	0.826	0.834	0.841	0.85	0.85
		SVM	0.854	0.907	0.912	0.918	0.916
	Re.	1NN	0.757	0.776	0.779	0.814	0.816
		SVM	0.733	0.763	0.815	0.799	0.81
	Partial Re.	1NN	0.83	0.829	0.839	0.857	0.848
		SVM	**0.884**	**0.924**	**0.941**	**0.947**	**0.95**

TABLE IV
PAVIA CENTER: BEST CLASSIFICATION ACCURACY (%) OVER TEN RUNS FOR CLASSIFICATION MAPS IN FIG. 3, 20 TRAINING SAMPLES PER CLASS WERE USED

	Spectral Only	Disk-based MP			Disk- and Linear-based MP		
		No Re.	Re.	Partial Re.	No Re.	Re.	Partial Re.
OA	95.9	97.9	98.2	98	**98.3**	98.1	**98.3**
AA	91	94.2	96.8	96.2	96.2	96.6	**97**
κ	93.4	96.4	96.9	96.6	**97.1**	96.8	**97.1**
std	0.92	0.43	0.55	0.36	0.41	0.44	**0.27**
Water	98.9	**100**	99.3	99.7	99.9	99.3	99.5
Trees	87.2	93	93.1	**93.8**	92.7	93.2	93.2
Meadows	**94.6**	85.5	85	91	90	85.6	93.5
Bricks	62.6	84.1	**99.8**	98.1	96.4	99.6	99.3
Soil	94.8	95.8	96	97	96.7	95.7	**97.8**
Asphalt	94.5	96.2	**98.8**	97.7	98.2	97.8	98.4
Bitumen	86.6	95	97.3	90.2	93.7	**97.9**	92.7
Tiles	99.6	**100**	99.9	**100**	**100**	99.9	**100**
Shadows	**100**	98.7	99.9	98.7	98	**100**	98.9

The best overall accuracy (OA) of each data set in each training sample size is highlighted (in column) in bold font.

From these tables, we have the following findings:

1) The results confirm that the MPs (without reconstruction, with reconstruction, and with partial reconstruction) can improve the classification performance on hyperspectral images. By building the extended morphological profiles on the first 3 principal components, the results can be improved a lot. Compared to the situation with only spectral bands in each training sample size, the OA of Pavia Center and University Area data sets with MPs have 0.2%–2.6% and 12.4%–20% improvements for the 1 NN classifier, respectively. For SVM classifier, these improvements are 2%–3.3% and 1.5%–25.5%, respectively.
2) As the number of training samples increases, the OA will increase. Especially for SVM classifier, in Pavia Center data set, the OA of spectral only has 2% improvements from 10 training samples per class to 160 training samples, this also happens on MPs with nearly 2% improvements; in University Area data set, the OA of spectral only increases from 65.3% to 78.7% when the number of training samples per class changes from 10 to 160, while MPs with almost 7% improvements.
3) The results can be improved by adding the directional MPs. There is a substantial improvement of the overall accuracy over the classification with only disk-based MPs. However, when using MPs with reconstruction, the classification accuracies by adding the directional MPs improves very little and is comparatively much less than those without reconstruction and with partial reconstruction. This is because the disk-based MPs and linear-based MPs with reconstruction contain much the same information.
4) It is better not to use MPs with reconstruction in some cases. This is in particular the case in University Area data set, where the MPs with reconstruction perform even worse than MPs without reconstruction. The MPs with partial reconstruction and SVM classifier almost gets the best results all the time, this is obvious in University Area data set.

In order to compare the classification results visually, we randomly select 20 training samples per class for training, and use all the samples for testing. The SVM classifier was used. The best results over ten runs are shown in Fig. 3, Fig. 4 and Table IV, Table V. The Z tests [40] were reported Table VI, Table VII.

1) The MPs (without reconstruction, with reconstruction, and with partial reconstruction) can preserve well spatial information on hyperspectral images. The classification maps with MPs produce much smoother homogeneous regions than that of spectral only, which is particularly significant when using MPs with no reconstruction and with partial reconstruction, see Table VI, Table VII. The statistical difference of accuracy $|Z| > 1.96$ clearly demonstrates the benefit of using the MPs with no reconstruction and with partial reconstruction rather than the spectral only.
2) The classification maps using the MPs with reconstruction look much noisier because of the over reconstruction problems. The MPs with no reconstruction deform the objects, see Fig. 4(c) and Fig. 4(f), the borders of some objects are deformed. While small objects might be fused together (e.g., the buildings and shadows in the bottom part of the

TABLE V
UNIVERSITY AREA: BEST CLASSIFICATION ACCURACY (%) OVER TEN RUNS FOR CLASSIFICATION MAPS IN FIG. 4, 20 TRAINING SAMPLES PER CLASS WERE USED

	Spectral Only	Disk-based MP			Disk- and Linear-based MP		
		No Re.	Re.	Partial Re.	No Re.	Re.	Partial Re.
OA	78.1	89.1	80.3	90.5	92.6	81.7	**93.8**
AA	78.7	85.6	86.7	89.2	90.2	88	**93.7**
κ	71	85.3	74.5	87.2	90.1	76.3	**91.9**
std	4.23	0.77	2.42	**0.73**	1.21	4.15	1.14
Asphalt	59.3	88.6	80.7	86.5	82.9	87.9	**89.6**
Meadows	89.6	98.1	77.8	98.4	**99.5**	77.5	94.6
Gravel	53.2	43.9	74.2	70.7	66.4	**77.1**	69
Trees	91.2	95.7	92.8	86.7	93.6	97.7	**97.8**
Metal Sheets	98.9	**99.9**	99.6	99.6	99.6	99.2	99.5
Soil	48.8	60.9	56.9	65.6	84.5	58.6	**97.5**
Bitumen	86.2	90.4	99.6	97.4	98	99.1	**99.9**
Bricks	81	96.2	**98.6**	98.2	95.7	95.1	98.4
Shadows	**100**	97.2	99.9	99.8	91.9	99.9	97.6

TABLE VI
UNIVERSITY AREA: STATISTICAL SIGNIFICANCE OF DIFFERENCES IN CLASSIFICATION (Z) OVER TEN RUNS. EACH CASE OF THE TABLE REPRESENTS Z_{rc} WHERE R IS THE ROW AND C IS THE COLUMN, 20 TRAINING SAMPLES PER CLASS WITH SVM CLASSIFIER WERE USED

Z_{rc}		Spectral Only	Disk-based MP			Disk- and Linear-based MP		
			No Re.	Re.	Partial Re.	No Re.	Re.	Partial Re.
Spectral Only		0	-4.0065	-1.1304	-4.5819	-4.7267	-0.8276	-5.3427
Disk-based MP	No Re.	4.0065	0	4.1963	-1.8734	-2.1650	2.6714	-3.7365
	Re.	1.1304	-4.1963	0	-5.0917	-5.1235	0.1011	-6.0577
	Partial Re.	4.5819	1.8734	5.0917	0	-0.8098	3.2139	-2.3441
Disk- and Linear-based MP	No Re.	4.7267	2.1650	5.1235	0.8098	0	3.4092	-1.2275
	Re.	0.8276	-2.6714	-0.1011	-3.2139	-3.4092	0	-3.9860
	Partial Re.	5.3427	3.7365	6.0577	2.3441	1.2275	3.9860	0

TABLE VII
PAVIA CENTER: STATISTICAL SIGNIFICANCE OF DIFFERENCES IN CLASSIFICATION (Z) OVER TEN RUNS. EACH CASE OF THE TABLE REPRESENTS Z_{rc} WHERE R IS THE ROW AND C IS THE COLUMN, 20 TRAINING SAMPLES PER CLASS WITH SVM CLASSIFIER WERE USED

Z_{rc}		Spectral Only	Disk-based MP			Disk- and Linear-based MP		
			No Re.	Re.	Partial Re.	No Re.	Re.	Partial Re.
Spectral Only		0	-2.8358	-2.5593	-2.9667	-3.1620	-2.7805	-3.3847
Disk-based MP	No Re.	2.8358	0	0.1722	-0.0756	-0.4919	0.0670	-0.6849
	Re.	2.5593	-0.1722	0	-0.2473	-0.5993	-0.1123	-0.7600
	Partial Re.	2.9667	0.0756	0.2473	0	-0.4601	0.1472	-0.6827
Disk- and Linear -based MP	No Re.	3.1620	0.4919	0.5993	0.4601	0	0.5541	-0.1174
	Re.	2.7805	-0.0670	0.1123	-0.1472	-0.5541	0	-0.7521
	Partial Re.	3.3847	0.6849	0.7600	0.6827	0.1174	0.7521	0

Pavia center image) when using the MPs with partial reconstruction and no reconstruction, in this case, the MPs with full reconstruction perform better.

3) When using both disk-based and directional MPs with partial reconstruction, we get the best OA, AA and $Kappa$ for both data sets, and relative lower standard deviation (std). For University Area data set, the difference is statistically significant. For Pavia Center data set, the difference is not statistically significant with $|Z| < 1.96$.

D. Results Using Semi-Supervised Feature Extraction to Reduce the Dimensionality of the Generated MPs

We compare the resulting classification accuracies using the proposed GSELD method to extract features from the generated morphological profiles with those resulting from the following methods: PCA [28]; LDA [22]; NPE [33]; NWFE [23]. The data sets of *University Area* and *Indian Pine* are used. For *Indian Pine* data set, we compared the results with partial reconstruction based on only disk-based MPs. In our experiments, $u = 1500$ unlabeled samples are randomly selected for training the proposed GSELD, the parameter α in (8) is set as $\alpha = u/n$ (n is the number of labeled training samples), which can change automatically according to the ratio of the number of unlabeled and labeled samples while increasing the class separability. 20 features (except only for the $C-1$ features in LDA) are extracted, then, the testing accuracies of each employed number of features are calculated respectively. The highest OA with three classifiers in different samples size are shown in Fig. 5–Fig. 7, the number of extracted features changed from 1 to 20, each experiment was repeated 5 times, the average was acquired.

1) The results confirm that feature extraction can improve the classification performance. Especially for conventional classifiers (such as LDC classifier), FE makes the classification possible. For 1 NN classifier, the results of University Area dat set can be improved a lot by using FE as a preprocessing.
2) SVM classifier is more efficient to deal with the high dimensional data, this is obvious in University Area data set, see Fig. 5(c) and Fig. 7(c). In some cases, it can achieve even better performances than those using FE as a preprocessing, see Fig. 5(c). When using only the

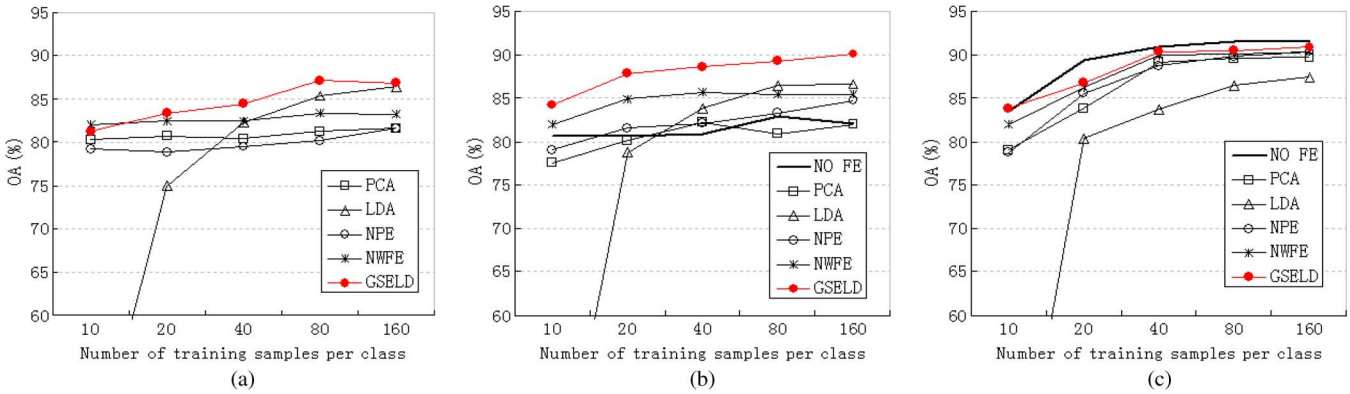


Fig. 5. Highest OA of University Area in different samples size with partial reconstruction based on only disk-based MPs, the number of extracted features changed from 1 to 20, each experiment was repeated 5 times, the average was acquired. (a) LDC classifier; (b) 1 NN classifier; (c) SVM classifier.

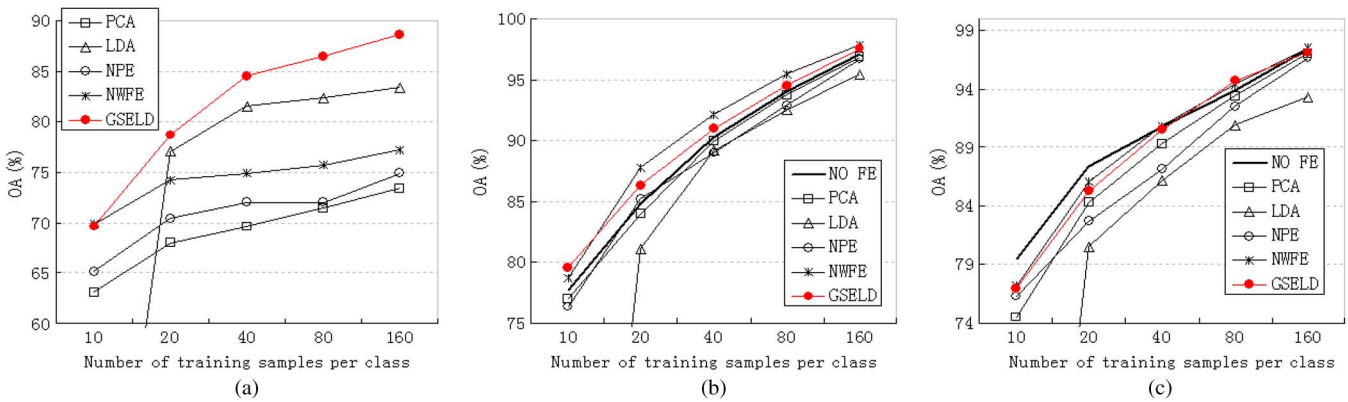


Fig. 6. Highest OA of Indian Pine in different samples size with partial reconstruction based on only disk-based MPs, the number of extracted features changed from 1 to 20, each experiment was repeated 5 times, the average was acquired. (a) LDC classifier; (b) 1 NN classifier; (c) SVM classifier.

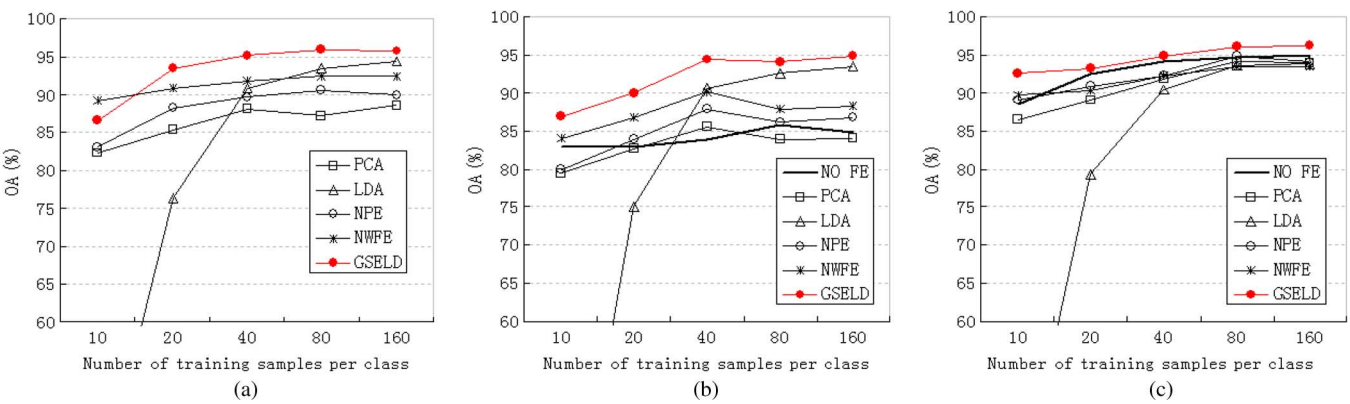


Fig. 7. Highest OA of University Area in different samples size with partial reconstruction based on both disk-based and linear-based MPs, the number of extracted features changed from 1 to 20, each experiment was repeated 5 times, the average was acquired. (a) LDC classifier; 1 NN classifier; (c) SVM classifier.

disk-based MPs, SVM classifier with no FE outperforms those with FE as a preprocessing.

3) For the LDC and the SVM classifiers, as the number of training samples per class increases, the OA of each method will increase. This is particular for the supervised LDA method, when the number of training samples per class is 10, the OA is much lower than 60%. When the number of training samples per class is more than 80, the OA of LDA increases above 80%.
4) For low resolution Indian Pine data set, when the training sample size increases, the proposed GSELD with LDC classifier outperforms the other methods with LDC classifier. While NWFE with KNN classifier performs a little bit better than GSELD with KNN classifier. When using SVM classifier, the OA of GSELD is similar with that of NWFE.
5) For high resolution urban data set (University Area), when using both the disk-based and linear-based morphological

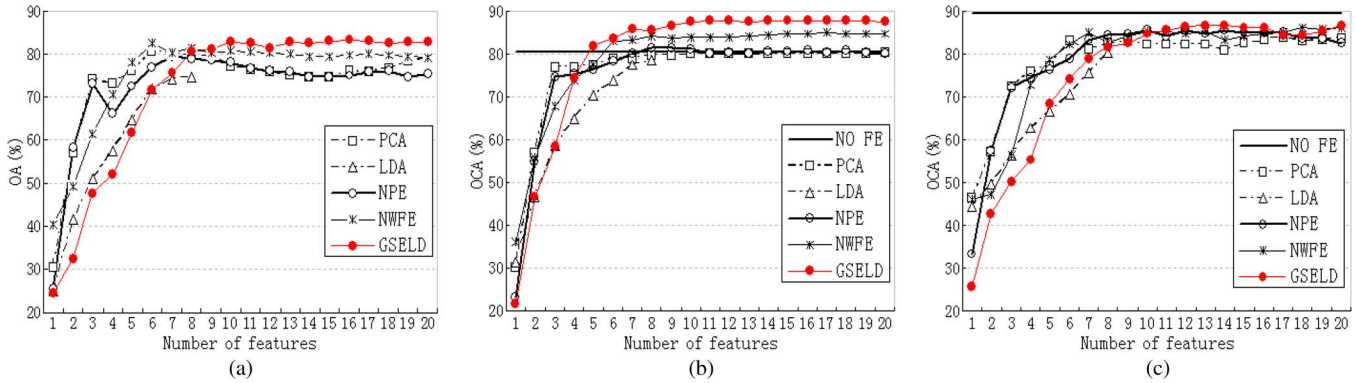


Fig. 8. Performance of each feature extraction method using 20 training samples per class for University Area data set, the MPs are based on only disk SE with Partial Reconstruction. Each experiment was repeated 5 times, the average was acquired. (a) LDC classifier; (b) 1 NN classifier; (c) SVM classifier.

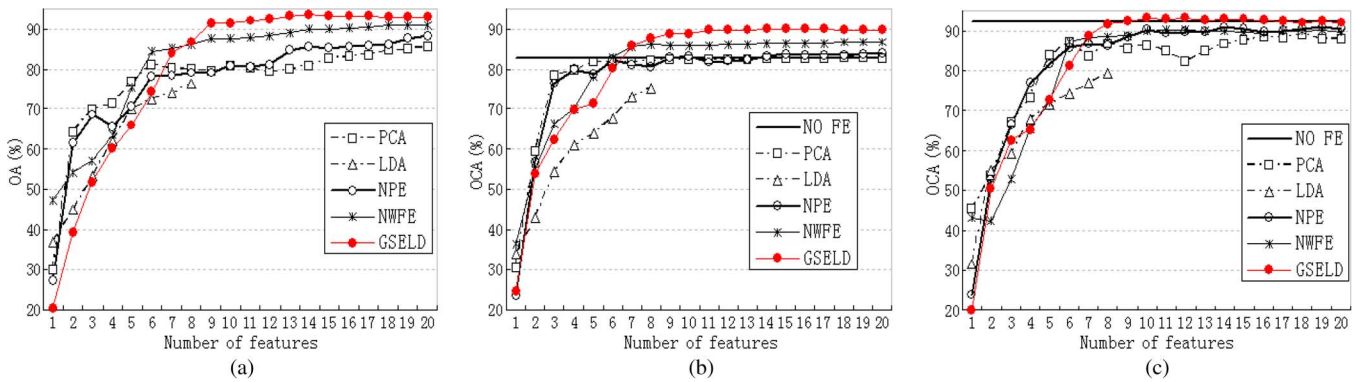


Fig. 9. Performance of each feature extraction method using 20 training samples per class for University Area data set, the MPs are based on both disk and linear SEs with Partial Reconstruction. Each experiment was repeated 5 times, the average was acquired. (a) LDC classifier; (b) 1 NN classifier; (c) SVM classifier.

features, the proposed GSELD gets the highest OA in different samples size. The highest OA for training samples size with 10, 20, 40, 80 and 160 are 92.5% (GSELD with SVM classifier), 93.4% (GSELD with LDC classifier), 95.1% (GSELD with LDC classifier), 96% (GSELD with SVM classifier) and 96.2% (GSELD with SVM classifier), respectively.

The experiments were carried out on 64-b, 2.67 GHz Intel i7 920 (8 core) CPU computer with 12 GB memory, the time was only consumed in the process of feature extraction for MPs based on both disk and linear SEs with Partial Reconstruction. When the training sample size of University Area data set changes from 80 to 160, the consumed time of NWFE increases from 14.8 seconds to 113.9 seconds, while for the proposed GSELD, the consumed time increases from 2.4 seconds to 5.3 seconds. Fig. 8 and Fig. 9 show the performances with different number of extracted features when 20 training samples per class are used as training set. The Z tests using MPs based on both disk and linear SEs with Partial Reconstruction were reported in Table VIII–Table X. The results confirm some findings in Fig. 5 and Fig. 6, moreover we find the following:

1) Most information of the generated MPs can be preserved even with a few extracted features. For 1 NN classifier, when the number of extracted features is more than 7, the results of NWFE and GSELD are better than that without FE. When using both disk-based and linear-based MPs, the difference is statistically significant with $|Z| > 1.96$. For the SVM classifier using both disk-based and linear-based MPs, the proposed GSELD gets better result even with 9 extracted features.
2) Using only $C - 1$ features may not be enough in some situation, which is one limitation of LDA. PCA and NPE can improve their performances by using more extracted features, as shown in Fig. 7 and Fig. 8. When more features are used, the overall classification accuracy can be improved with statistical significance ($|Z| > 1.96$).
3) The proposed GSELD outperforms the other feature extraction methods with all these three classifiers, with $Z > 0$. When using both the disk-based and linear-based morphological features, the proposed GSELD gets the highest

TABLE VIII
LDC CLASSIFIER: STATISTICAL SIGNIFICANCE OF DIFFERENCES IN CLASSIFICATION (Z) OVER FIVE RUNS. EACH CASE OF THE TABLE REPRESENTS Z_{rc} WHERE R IS THE ROW AND C IS THE COLUMN, 20 TRAINING SAMPLES PER CLASS WERE USED. THE MPS ARE BASED ON BOTH DISK AND LINEAR SES WITH PARTIAL RECONSTRUCTION

Z_{rc}	PCA	LDA	NPE	NWFE	GSELD
PCA	0	2.6330	-1.2986	-1.4557	-2.7927
LDA	-2.6330	0	-3.8497	-3.9939	-5.4819
NPE	1.2986	3.8497	0	-0.1588	-1.2281
NWFE	1.4557	3.9939	0.1588	0	-1.0275
GSELD	2.7927	5.4819	1.2281	1.0275	0

TABLE IX
1 NN CLASSIFIER: STATISTICAL SIGNIFICANCE OF DIFFERENCES IN CLASSIFICATION (Z) OVER FIVE RUNS. EACH CASE OF THE TABLE REPRESENTS Z_{rc} WHERE R IS THE ROW AND C IS THE COLUMN, 20 TRAINING SAMPLES PER CLASS WERE USED. THE MPS ARE BASED ON BOTH DISK AND LINEAR SES WITH PARTIAL RECONSTRUCTION

Z_{rc}	NO FE	PCA	LDA	NPE	NWFE	GSELD
NO FE	0	0.0368	1.3690	-0.3187	-1.0036	-2.7382
PCA	-0.0368	0	1.3503	-0.3581	-1.0452	-2.8113
LDA	-1.3690	-1.3503	0	-1.5848	-2.0409	-3.2841
NPE	0.3187	0.3581	1.5848	0	-0.7027	-2.3453
NWFE	1.0036	1.0452	2.0409	0.7027	0	-1.3659
GSELD	2.7382	2.8113	3.2841	2.3453	1.3659	0

TABLE X
SVM CLASSIFIER: STATISTICAL SIGNIFICANCE OF DIFFERENCES IN CLASSIFICATION (Z) OVER FIVE RUNS. EACH CASE OF THE TABLE REPRESENTS Z_{rc} WHERE R IS THE ROW AND C IS THE COLUMN, 20 TRAINING SAMPLES PER CLASS WERE USED. THE MPS ARE BASED ON BOTH DISK AND LINEAR SES WITH PARTIAL RECONSTRUCTION

Z_{rc}	NO FE	PCA	LDA	NPE	NWFE	GSELD
NO FE	0	0.6960	3.5715	0.3081	0.6967	-0.1704
PCA	-0.6960	0	3.3101	-0.3829	0.1476	-1.1095
LDA	-3.5715	-3.3101	0	-3.4251	-2.8042	-4.1037
NPE	-0.3081	0.3829	3.4251	0	0.4424	-0.5633
NWFE	-0.6967	-0.1476	2.8042	-0.4424	0	-0.9627
GSELD	0.1704	1.1095	4.1037	0.5633	0.9627	0

OA for all these three classifiers. The highest OA for LDC classifier, 1 NN classifier and SVM classifier are 93.4% (GSELD with 14 extracted features), 90% (GSELD with 14 extracted features) and 93.2% (GSELD with 10 extracted features), respectively.

V. CONCLUSION

In this study, we first investigated the morphological profiles with partial reconstruction and directional morphological profiles for the classification of high resolution hyperspectral images from urban areas. We showed on two real urban hyperspectral data sets that the MPs with partial reconstruction are more competitive than those with no reconstruction and with reconstruction, and some classes like road are classified better with the directional morphological features. Secondly, we developed a semi-supervised feature extraction as a preprocessing tool to reduce the dimensionality of the generated morphological profiles for classification. The results show that feature extraction can improve significantly the performance for some classifiers, and the proposed semi-supervised method compares favorably with conventional feature extraction methods as preprocessing approaches for the morphological profiles generated on high resolution hyperspectral data from the urban area.

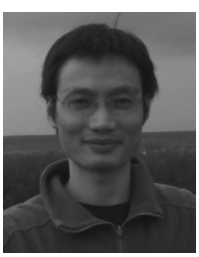
ACKNOWLEDGMENT

The authors would like to thank Prof. Paolo Gamba from the University of Pavia, Italy, for kindly providing both the Centre and University data sets, Prof. Kuo for providing NWFE codes, and the anonymous reviewers for their careful reading and valuable remarks, which have contributed in improving the quality of the paper.

REFERENCES

[1] P. Soille, *Morphological Image Analysis, Principles and Applications*, 2nd ed. Berlin, Germany: Springer-Verlag, 2003.
[2] P. Soille and M. Pesaresi, "Advances in mathematical morphology applied to geoscience and remote sensing," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 9, pp. 2042–2055, Sep. 2002.
[3] M. Pesaresi and J. A. Benediktsson, "A new approach for the morphological segmentation of high-resolution satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 2, pp. 309–320, Feb. 2001.
[4] R. Bellens, S. Gautama, L. Martinez-Fonte, W. Philips, J. C.-W. Chan, and F. Canters, "Improved classification of VHR images of urban areas using directional morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 2803–2812, Oct. 2008.
[5] J. A. Benediktsson, J. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
[6] A. Plaza, P. Martinez, J. Plaza, and R. M. Perez, "Dimensionality reduction and classification of hyperspectral image data using sequences of extended morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 466–479, March 2005.
[7] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernel principal component analysis for the classification of hyperspectral remote-sensing data over urban areas," *EURASIP J. Advances in Signal Processing*, vol. 2009, February 2009, 14 pages.
[8] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5975–5991, Nov. 2010.
[9] M. Dalla Mura, J. A. Benediktsson, B. Waske, and L. Bruzzone, "Morphological attribute profiles for the analysis of very high resolution images," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 10, pp. 3747–3762, Oct. 2010.
[10] M. Dalla Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 541–545, May 2011.
[11] J. Crespo, J. Serra, and R. Shafer, "Theoretical aspects of morphological filters by reconstruction," *Signal Process.*, vol. 47, no. 2, pp. 201–225, Nov. 1995.
[12] 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.
[13] J. A. Benediktsson, M. Pesaresi, and K. Arnason, "Classification and feature extraction for remote sensing images from urban areas based on morphological transformations," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 1940–1949, Sep. 2003.
[14] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Information Theory*, vol. 14, no. 1, pp. 55–63, 1968.
[15] T. Castaings, B. Waske, J. A. Benediktsson, and J. Chanussot, "On the influence of feature reduction for the classification of hyperspectral images based on the extended morphological profile," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5921–5939, July 2010.
[16] L. Bruzzone and C. Persello, "A novel approach to the selection of spatially invariant features for the classification of hyperspectral images with improved generalization capability," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 9, pp. 3180–3191, Sep. 2009.
[17] M. Dalla Mura, J. Benediktsson, and L. Bruzzone, "Classification of hyperspectral images with extended attribute profiles and feature extraction techniques," in *Proc. IGARSS*, July 2010, pp. 76–79.
[18] X. Zhu, "Semi-supervised learning literature survey," Dept. Comput. Sci., Univ. Wisconsin Madison. Madison, WI. Comput. Sci. TR 1530, Jul. 19, 2008.
[19] 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.
[20] S. G. Chen and D. Q. Zhang, "Semisupervised dimensionality reduction with pairwise constraints for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett*, vol. 8, no. 2, pp. 369–373, March 2011.
[21] W. Z. Liao, A. Pizurica, W. Philips, and Y. G. Pi, "Feature extraction for hyperspectral image based on semi-supervised local discriminant analysis," in *Proc. IEEE Joint Urban Remote Sensing Event (JURSE 2011)*, Apr. 2011, pp. 401–404.
[22] C.-I. Chang and H. Ren, "An experiment-based quantitative and comparative analysis of target detection and image classification algorithms for hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 38, no. 2, pp. 1044–1063, March 2000.

[23] B. C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 5, pp. 1096–1105, May 2004.
[24] 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.
[25] T. V. Bandos, L. Bruzzone, and G. Camps-Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 3, pp. 862–873, March 2009.
[26] B. C. Kuo, C. W. Chang, C. C. Hung, and H. P. Wang, "A modified nonparametric weight feature extraction using spatial and spectral information," in *Proc. Int. Geoscience and Remote Sensing Symp.*, 2006, pp. 172–175.
[27] B. C. Kuo, C. H. Li, and J. M. Yang, "Kernel nonparametric weighted feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 4, pp. 1139–1155, April 2009.
[28] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educational Psychology*, vol. 24, pp. 417–441, 1933.
[29] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A transformation for ordering multispectral data in terms of image quality with implications for noise removal," *IEEE Trans. Geosci. Remote Sens.*, vol. 26, no. 1, pp. 65–74, Jan. 1988.
[30] 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.
[31] G. Chen and S.-E. Qian, "Dimensionality reduction of hyperspectral imagery using improved locally linear embedding," *J. Appl. Remote Sens.*, vol. 1, pp. 1–10, March 2007.
[32] H. Y. Huang and B. C. Kuo, "Double nearest proportion feature extraction for hyperspectral-image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4034–4046, Nov. 2010.
[33] X. F. He, D. Cai, S. C. Yan, and H. J. Zhang, "Neighborhood preserving embedding," in *Proc. 10th IEEE Int. Conf. Computer Vision 2005*, 2005, vol. 2, pp. 1208–1213.
[34] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. 11th Annual Conf. Computational Learning Theory*, 1998, pp. 92–100.
[35] V. N. Vapnik, *Statistical Learning Theory*. New York: Wiley, 1998.
[36] G. Camps-Valls, T. Bandos, and D. Zhou, "Semisupervised graph-based hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 10, pp. 3044–3054, Oct. 2007.
[37] R. P. W. Duin, P. Juszczak, D. de Ridder, P. Paclik, E. Pekalska, and D. M. J. Tax, PRTools, a Matlab Toolbox for Pattern Recognition, 2004 [Online]. Available: http://www.prtools.org
[38] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Mining and Knowledge Discovery*, vol. 2, pp. 121–167, 1998.
[39] C. C. Chang and C. J. Lin, LIBSVM: A Library for Support Vector Machines, 2001 [Online]. Available: http://www.csie.ntu.edu.tw/~cjlin/libsvm
[40] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogramm. Eng. Remote Sens.*, vol. 70, no. 5, pp. 627–633, 2004.



Wenzhi Liao received the B.S. degree in mathematics from Hainan Normal University, HaiKou, China, in 2006, the M.S. degree in mathematics from South China University of Technology (SCUT), GuangZhou, China, in 2008. Currently, he is working toward the Ph.D. degree both in the School of Automation Science and Enginerring, SCUT, GuangZhou, China, and Department of Telecommunications and Information Processing, Ghent University, Ghent, Belgium.

His current research interests include pattern recognition, remote sensing, and image processing.

Rik Bellens (M'07) received the diploma in computer science engineering in 2004 from Ghent University, Belgium, where he is currently working toward the Ph.D. degree in the Department of Telecommunications and Information Processing.

His main research interests are pattern recognition, remote sensing, image processing, mobility and crowd behaviour.

Aleksandra Pižurica received the Diploma Degree in Electical Engineering from the University of Novi Sad (Serbia) in 1994, Master of Science Degree in Telecommunications from The University of Belgrade (Serbia) in 1997 and the PhD Degree in Engineering from Ghent University (Belgium) in 2002. From 1994 till 1997 she was working as a research and teaching assistant at the Department of Telecommunications of the University of Novi Sad and in 1997, she joined the Department of Telecommunications and Information Systems at Ghent University.

Since 2005 she is a postdoctoral research fellow of The Research Foundation—Flanders (FWO) and since 2009 she is also as a part time lecturer at Ghent University. Aleksandra Pižurica is author/co-author of more than 30 publications in international journals and book chapters and nearly 150 publications in the proceedings of international conferences. Her research interests include statistical image modeling, multiresolution and sparse signal representations, Bayesian estimation and applications in video analysis, remote sensing and medical imaging.

Wilfried Philips (S'90–M'93) was born in Aalst, Belgium on October 19, 1966. In 1989, he received the Diploma degree in electrical engineering and in 1993 the Ph.D. degree in applied sciences, both from Ghent University, Belgium. From October 1989 until October 1997 he worked at the Department of Electronics and Information Systems of Ghent University for the Flemish Fund for Scientific Research (FWO-Vlaanderen), first as a research assistant and later as a post-doctoral research fellow.

Since November 1997 he is with the Department of Telecommunications and Information Processing of Ghent University, where he is currently a full-time professor and is heading the research group "Image Processing and Interpretation," which has recently become part of the virtual Flemish ICT research institute IBBT.

Some of the recent research activities in the group include image and video restoration and analysis and the modelling of image reproduction systems. Important application areas targeted by the group include remote sensing, surveillance and industrial inspection.

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

From July 1998 to June 2002, he was with the Information Technology Department, Automation Engineering Center, Academy Guangdong Province. Since July 2002, He has been with the College of Automation Science and Engineering, South China University of Technology, where he is currently a full-time Professor. Some of the recent research activities in the group including image processing and patter recognition and motion control. Important application areas targeted by the group include Intelligent Chinese character formation and servo system.