Dimension Reduction Using Spatial and Spectral Regularized Local Discriminant Embedding for Hyperspectral Image Classification

Yicong Zhou, *Member, IEEE*, Jiangtao Peng, and C. L. Philip Chen, *Fellow, IEEE*

***Abstract*—Dimension reduction (DR) is a necessary and helpful preprocessing for hyperspectral image (HSI) classification. In this paper, we propose a spatial and spectral regularized local discriminant embedding (SSRLDE) method for DR of hyperspectral data. In SSRLDE, hyperspectral pixels are first smoothed by the multiscale spatial weighted mean filtering. Then, the local similarity information is described by integrating a spectral-domain regularized local preserving scatter matrix and a spatial-domain local pixel neighborhood preserving scatter matrix. Finally, the optimal discriminative projection is learned by minimizing a local spatial–spectral scatter and maximizing a modified total data scatter. Experimental results on benchmark hyperspectral data sets show that the proposed SSRLDE significantly outperforms the state-of-the-art DR methods for HSI classification.**

***Index Terms*—Dimension reduction (DR), hyperspectral image (HSI), local pixel neighborhood preserving embedding (LPNPE), regularized local discriminant embedding (RLDE).**

I. Introduction

HYPERSPECTRAL images (HSIs) are widely used in environmental mapping, geological research, crop analysis, and mineral identification [1]. These applications often require to classify each pixel in the scene. Because there are a huge number of features (or spectral bands) with only limited training samples available, HSI classification becomes a challenging task. A large number of spectral bands provide rich information for classifying various materials in the scene. However, with limited training samples, the performance of classifiers deteriorates as the dimensionality increases [2] (Hughes phenomenon [3]). High-dimensionality data processing also requires huge computational resources and storage capacity [4]. Meanwhile, the spectral bands are often correlated, and not all of them are useful for the specific classification task. Therefore, to achieve an excellent classification performance, a dimension reduction (DR) procedure is required before training the classifiers.

As a preprocessing method for classification, DR seeks a low-dimensional representation for high-dimensional data that may contain crucial information. DR helps to ameliorate statistical ill-posed problem caused by the small-sample-size and to improve the HSI classification performance [5]–[10]. There are many DR methods that can be classified into the unsupervised, supervised, and semisupervised ones. Unsupervised DR methods reveal the low-dimensional data structure without using any label information, e.g., principal component analysis (PCA), which finds an orthogonal projection to maximize the global data variance, locality preserving projection (LPP) [11] and neighborhood preserving embedding (NPE) [12], which intend to keep the local structure of the data manifold. Supervised DR methods use labeled samples to learn the discriminative projections, such as linear discriminant analysis (LDA) [13], [14], nonparametric weighted feature extraction (NWFE) [15], local Fisher discriminant analysis (LFDA) [16], local discriminant embedding (LDE) [17] (or marginal Fisher analysis [18]), and so on. LDA seeks the best projection to maximize the between-class distance while minimizing the within-class distance. NWFE extends LDA by integrating nonparametric scatter matrices with training samples around the decision boundary [15]. LFDA combines the discriminative ability of LDA with the local preserving ability of LPP [16]. LDE extends the global LDA by performing the local discriminant in a graph embedding framework [17]. Semisupervised DR methods learn the discriminant projections from both a limited number of labeled data and a large amount of unlabeled data while preserving a certain potential data structure [8]–[10], [19], [20]. Semisupervised discriminant analysis (SDA) learns the discriminant structure from the labeled samples while inferring the intrinsic geometrical structure from both the labeled and unlabeled samples [19]. Semisupervised local discriminant analysis (SELD) combines the supervised and unsupervised DR methods by using the supervised LDA to maximize the class discriminations and the unsupervised LPP or NPE to preserve the local data structures [9].

The aforementioned DR methods are spectral-based methods. They measure the similarity between samples using the spectral-domain Euclidean distance. However, for HSIs, the spectral-domain similarity is insufficient to reveal the intrinsic relationships between different samples. Two samples with a small spectral distance may have a large spatial pixel distance.

Manuscript received December 16, 2013; revised March 9, 2014 and May 8, 2014; accepted June 19, 2014. Date of publication July 15, 2014; date of current version August 12, 2014. This work was supported in part by the Macau Science and Technology Development Fund under Grant FDCT/106/2013/A3; by the Research Committee at University of Macau under Grant MYRG2014-00003-FST, Grant RG017/ZYC/2014/FST, Grant MYRG113(Y1-L3)-FST12-ZYC, and Grant MRG001/ZYC/2013/FST; and by the National Natural Science Foundation of China under Grant 11371007. *(Corresponding author: Jiangtao Peng.)*

Y. Zhou and C. L. P. Chen are with the Department of Computer and Information Science, University of Macau, Taipa, Macau, China (e-mail: yicongzhou@umac.mo; Philip.Chen@ieee.org).

J. Peng is with the Faculty of Mathematics and Statistics, Hubei University, Wuhan 430062, China, and also with the Department of Computer and Information Science, University of Macau, Taipa, Macau, China (e-mail: pengjt1982@126.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2014.2333539

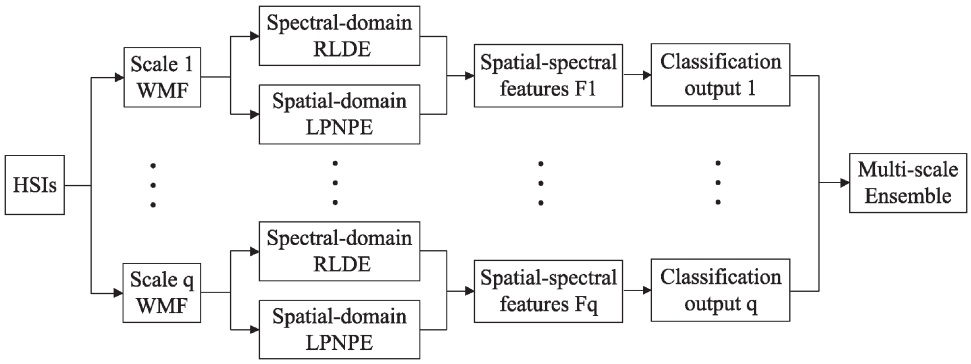


Fig. 1. Flowchart of the proposed SSRLDE-based HSI classification system.

The projection and classification based on only the spectral similarity metric may lead to underclassification or overclassification. Therefore, the spatial interpixel correlations should be considered in measuring the sample similarity and learning the discriminant projections.

Spatial contexture information has been proven to be useful to improve the HSI data representation and to increase classification accuracy [2], [21]. The methods using the spatial information can be classified into the spatial filtering methods and feature extraction techniques. The spatial filtering methods include the spatial filtering preprocessing methods [22]–[25], which filter spatial homogeneous regions prior to classification, and the spatial filtering postprocessing methods [26], [27], which smooth the pixelwise classification map. The feature extraction techniques contain the mean-shift-based object extraction [28] and structural feature set method [29], morphological transformation methods [30]–[32] which extract the morphological profiles from the principle components of hyperspectral data for classification, Markov random fields techniques [33]–[36], which incorporate spatial information into a probabilistic framework by modeling the spatial neighboring pixel correlations, support vector machine with composite kernels [37], which represents spatial features as the mean or standard derivation vectors of neighboring pixels and uses them in the SVM kernels, and multifeature models [38], [39] which construct an SVM ensemble integrating multiple spectral, structural, and semantic features at both the pixel and object levels.

Instead of using the spatial information in classification, we incorporate the spatial information into the DR process of HSIs and propose a spatial and spectral information-based RLDE (SSRLDE) method in this paper. In SSRLDE, the spatial information is used in two ways: spatial filtering and spatial discriminant analysis. Particularly, a multiscale spatial weighted mean filtering (WMF) is used to preprocess the HSI pixels. Then, the multiscale spatial–spectral local discriminant features in the filtered data are extracted by a combination of the spectral-domain regularization LDE (RLDE) and spatial-domain local pixel NPE (LPNPE). Finally, a majority voting method is used to fuse the complementary classification results of each individual scale.

The proposed SSRLDE performs the LDE by using both the spectral and spatial information. It has at least the following characteristics.

1) The spatial filtering increases the neighboring pixel consistency. The weighted method emphasizes the most similar neighboring pixels and is effective and robust to resist the noisy and background points.
2) The regularization strategy overcomes the singularity caused by the small-sized samples. The regularized local preserving scatter matrices well reveal the neighborhood class discriminant relations and data diversity.
3) The multiscale method can describe different spatial structures of HSIs. The rich complementary information in the multiscale filtered data helps to improve the performance of the subsequent DR and classification.
4) Incorporating the local pixel consistency, multiscale complementarity, spectral data diversity, and local spatial–spectral similarity information, SSRLDE preserves the spectral-domain local neighborhood relations and spatial-domain local pixel neighborhood structures simultaneously, achieving significantly discriminative projections and high classification accuracy.

The rest of this paper is organized as follows. In Section II, the proposed SSRLDE is described in detail. The experimental results and analysis are provided in Section III. Finally, Section IV gives a summary of this paper.

II. Proposed Approach

The flowchart of the proposed SSRLDE-based HSI classification system is shown in Fig. 1. The HSI pixels are preprocessed by the multiscale spatial WMF. The discriminant information in the filtered data is extracted by using a spectral-domain RLDE and a spatial-domain LPNPE. The spatial–spectral discriminant information is then incorporated to produce multiscale spatial–spectral features. Finally, the classification results of each individual scale are combined by a majority voting.

A. Spatial WMF

To reduce noise and smoothen the homogeneous regions, we use a spatial WMF to preprocess the HSI pixels. Although there are many filter techniques, such as edge-preserving filtering [27] and anisotropic diffusion filtering [23], the WMF is relatively simple and fast.

Denoting the pixel coordinate of sample $\mathbf{x}_i$ as (p_i, q_i), we define the local pixel neighborhood centered at $\mathbf{x}_i$ as

$$N(\mathbf{x}_i) = \left\{ \mathbf{x} \triangleq (p,q) | p \in [p_i - a, p_i + a], q \in [q_i - a, q_i + a] \right\} \quad (1)$$

where $a = (w-1)/2$, the odd number w is the width of the neighborhood window (called the scale). The pixels in the spatial neighborhood $N(\mathbf{x}_i)$ are denoted as $\mathbf{x}_i, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{is}$, where $s = w^2 - 1$ is the number of neighbors of $\mathbf{x}_i$.

The spatial WMF of a labeled pixel $\mathbf{x}_i$ is

$$\hat{\mathbf{x}}_i = \frac{\sum_{\mathbf{x}_j \in N(\mathbf{x}_i)} \nu_j \mathbf{x}_j}{\sum_{\mathbf{x}_j \in N(\mathbf{x}_i)} \nu_j} = \frac{\mathbf{x}_i + \sum_{k=1}^{w^2-1} \nu_k \mathbf{x}_{ik}}{1 + \sum_{k=1}^{w^2-1} \nu_k} \tag{2}$$

where weight $\nu_k = \exp\{-\gamma_0 \|\mathbf{x}_i - \mathbf{x}_{ik}\|^2\}$ measures spectral distance of the neighboring pixels to the central pixel. Parameter γ_0 reflects the degree of filtering, and is empirically set to be 0.2 in the experiments. Considering that HSIs may contain small and large homogenous regions simultaneously, multiscale WMF is used to capture different spatial structures of HSIs.

B. Spectral-Domain RLDE

To extract the spectral-domain local similarity and diversity information, we propose a regularized LDE (RLDE) method. It inherits the local discriminant ability of LDE [17], preserves the data diversity, and overcomes the singularity in the case of limited training samples.

Suppose there are m training samples $\mathbf{x}_i \in \mathcal{R}^d$, $i = 1, 2, \ldots, m$, with labels $\{y_i\}_{i=1}^m$. Denote the data matrix as $X = [\mathbf{x}_1 \ \mathbf{x}_2 \ \cdots \ \mathbf{x}_m] \in \mathcal{R}^{d \times m}$. In order to discover both the geometrical and discriminant structures of the data manifold, a within-class graph G_w and a between-class graph G_b are constructed [17]. To construct G_w, for each pair of points $\mathbf{x}_i$ and $\mathbf{x}_j$ from the same class, we add an edge between $\mathbf{x}_i$ and $\mathbf{x}_j$ with weight $w_{w,ij}$ if $\mathbf{x}_j$ is one of $\mathbf{x}_i$'s k_1-nearest neighbors (NN). For G_b, considering each pair $\mathbf{x}_i$ and $\mathbf{x}_j$ with $y_i \neq y_j$, we connect $\mathbf{x}_i$ and $\mathbf{x}_j$ with weight $w_{b,ij}$ if $\mathbf{x}_j$ is one of $\mathbf{x}_i$'s k_2-NNs. The heat kernel weight is used

$$w_{ij} = \exp\left(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / t\right)$$

where t is the heat kernel parameter. The corresponding affinity matrices W_w and W_b are used to generate the within-class and between-class graph Laplacian matrices: $L_w = D_w - W_w$ and $L_b = D_b - W_b$, where $D_w(D_b)$ is a diagonal matrix whose entries are column sums of $W_w(W_b)$.

RLDE performs the spectral-domain local discriminant in a graph embedding framework, and it finds the optimal projection V from the following optimization problem:

$$\max_V \frac{\mathrm{Tr}\left\{V^T \left[(1-\alpha) X L_b X^T + \alpha X X^T\right] V\right\}}{\mathrm{Tr}\left\{V^T \left[(1-\alpha) X L_w X^T + \alpha \mathrm{diag}(X L_w X^T)\right] V\right\}} \tag{3}$$

where $\alpha \in [0\ 1]$ is a regularization parameter, $\mathrm{Tr}\{A\}$ represents the trace of matrix A, and $\mathrm{diag}(A)$ means the diagonal parts of matrix A.

Our contribution on RLDE is the regularization strategy. The regularization term XX^T in the numerator is used to preserve the maximal data variance. The diagonal regularization in the denominator improves the stability of the solution without impacting the local intraclass neighborhood preserving ability. It only decreases the off-diagonal elements of XL_wX^T while keeping the diagonal entries unchanged. Although the trace is unchanged, the diagonal regularization allows the large eigenvalues of XL_wX^T to decrease and the small or zero ones to increase. Thus, the regularized local preserving scatter matrix defined in (3) is positive definite. RLDE is suitable for the small-sample-size HSI classification problem.

When $\alpha = 0$, RLDE reverts back to LDE defined in [17]. When $\alpha = 1$ and the $\mathrm{diag}(XL_wX^T)$ is replaced with an identity matrix, RLDE becomes PCA. RLDE can utilize labeled samples to learn the discriminative projections and can use regularization terms to incorporate the data diversity and avoid the overfitting problem. Thus, RLDE is able to surpass LDE and PCA for the HSI DR.

The optimal RLDE projection $V = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_\ell]$ is obtained by finding the generalized eigenvectors corresponding to the ℓ largest eigenvalues in

$$\left[(1-\alpha) X L_b X^T + \alpha X X^T\right] \mathbf{v} = \lambda \left[(1-\alpha) X L_w X^T + \alpha \mathrm{diag}(X L_w X^T)\right] \mathbf{v}. \tag{4}$$

C. Spatial-Domain LPNPE

Due to the fact that the neighboring pixels in a spatial local homogeneous region consist of the same materials and belong to the same class, we propose a spatial-domain LPNPE method. Different from the traditional spectral-based DR methods, LPNPE uses the spatial information to learn the discriminant projections. It preserves the spatial local pixel neighborhood structures, that is, neighborhood pixels in the original spatial space remain neighbors in the LPNPE-embedded space, and vice versa.

A training sample $\mathbf{x}_i$ and its pixel neighbors in $N(\mathbf{x}_i)$ form a local pixel patch: $P_i = \{\mathbf{x}_i, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{is}\}$. The distance metric (scatter) in the local pixel patch can be defined as

$$h_i = \sum_{k=1}^{s} \frac{\nu_k}{\sum_{j=1}^{s} \nu_j} (\mathbf{x}_i - \mathbf{x}_{ik})(\mathbf{x}_i - \mathbf{x}_{ik})^T$$

where weight ν_k defined in (2) measures the spectral similarity between the neighboring pixels to the central pixel.

Consider all training samples in the HSI data, the local pixel neighborhood preserving scatter matrix is

$$H = \sum_{i=1}^{m} h_i = \sum_{i=1}^{m} \sum_{k=1}^{s} \frac{\nu_k}{\sum_{j=1}^{s} \nu_j} (\mathbf{x}_i - \mathbf{x}_{ik})(\mathbf{x}_i - \mathbf{x}_{ik})^T.$$

Denote the total scatter matrix as

$$S = \sum_{i=1}^{m} (\mathbf{x}_i - \mathbf{m})(\mathbf{x}_i - \mathbf{m})^T = \overline{XX}^T$$

where $\mathbf{m}$ is the mean of training samples.

LPNPE seeks a linear projection matrix such that the local pixel neighborhood preserving scatter is minimized, whereas the total scatter is maximized in the projected space. The optimal projection $V = [\mathbf{v}_1 \ \mathbf{v}_2 \ \cdots \ \mathbf{v}_\ell]$ can be obtained by solving the generalized eigenvalue problem

$$S\mathbf{v} = \lambda H \mathbf{v}. \tag{5}$$

D. Spatial–Spectral Feature Extraction and Classification

RLDE uses only the spectral information to learn the discriminant projections, and it aims to preserve the local neighborhood class relations measured by the spectral-domain Euclidean distance. Due to the distribution characteristics of HSIs: large homogeneous regions and multimodal data structures, neighboring samples in the same class may have large pixel distances and even belong to different subregions. If it only preserves the spectral neighborhood relations, the RLDE projections might attract unwanted neighbors of different subregions and destroy the spatial local pixel neighborhood structures. Assume additional spatial information is available, it can reduce the spectral variations in the local regions and preserve the detailed image structures.

On the other hand, LPNPE preserves the spatial local pixel neighborhood structures by using only spatial information. It does not use the labeled spectral information. It is actually an unsupervised DR method. If two labeled pixels (spectral neighbors) in a homogeneous region have a large pixel distance, LPNPE will not connect them directly. In this case, the labeled information enforces the connection of spectral neighbors, which will improve the discrimination.

The spectral information and spatial information are complementary to each other. In learning the discriminant projections, not only the spectral labeled information but also the spatial interpixel correlations are required to be considered. Thus, we propose a spatial–spectral information-based RLDE (SSRLDE) method. It preserves not only the spectral-domain local Euclidean neighborhood class relations but also the spatial-domain local pixel neighborhood structures.

Combining RLDE and LPNPE, we obtain the spatial–spectral local neighborhood preserving scatter matrix

$$R_w = \beta\left[(1-\alpha)S_w + \alpha \mathrm{diag}(S_w)\right] + (1-\beta)H$$

and a modified total data scatter matrix

$$\begin{aligned} R_b &= \beta\left[(1-\alpha)S_b + \alpha XX^T\right] + (1-\beta)S \\ &= \beta(1-\alpha)S_b + (1-\beta(1-\alpha))\,S \end{aligned}$$

where $0 \leqslant \alpha, \beta \leqslant 1$, $S_w = XL_wX^T$, $S_b = XL_bX^T$, and $S = XX^T$ (assuming the data are mean centered).

For sample $\mathbf{x}_i$, the final spatial–spectral discriminant feature is

$$\mathbf{z} = V^T\mathbf{x}_i \tag{6}$$

where the projection matrix $V = [\mathbf{v}_1\ \mathbf{v}_2\ \cdots\ \mathbf{v}_\ell]$ is obtained by finding the generalized eigenvectors corresponding to the ℓ largest eigenvalues in

$$R_b\mathbf{v} = \lambda R_w\mathbf{v}. \tag{7}$$

When the multiscale spatial–spectral discriminant features are obtained, we can perform the classification on the spatial–spectral features of each scale individually and finally fuse the classification results in different scales by a majority voting. More about the classifier fusion strategy can be referred to [38] and [39]. The ensemble of the multiscale complementary information ensures excellent classification performance of the proposed SSRLDE.

E. Discussion

The proposed SSRLDE has the following properties.

1) Local pixel consistency. The spatial WMF used in SSRLDE can smoothen the homogeneous areas and increase the local neighborhood pixel consistency. After filtering, samples in the same homogeneous areas are similar and consistent. This conforms to the characteristics of HSIs: the local homogeneous distribution.
2) Multiscale complementarity. The multiscale method can solve the scale selection problem and provide rich complementary information. The scale or window should cover the local homogeneous regions as accurately as possible such that the filtered data have better local pixel consistency. Because different HSI data have different spatial contextual structures and even the same data may contain small and large homogenous regions simultaneously, it is difficult to pre-define an optimal scale. In the multiscale framework, it is free of the scale selection. Different scales provide different views of a local homogeneous region. The complementary information in different scales helps to accurately describe the local homogeneous regions and improve the classification performance.
3) Regularization stability and data diversity. Regularization in RLDE solves the singularity (stability) problem without discarding the discriminant features. From the viewpoint of eigenspectrum, when the training samples are limited, eigenvalues decay rapidly and finally reach zero [40]. Small and zero eigenvalues are unstable, and the corresponding null spaces lose the discriminative information. In RLDE, regularization can counteract the bias estimation of small eigenvalues based on limited training samples [13]. After regularization, the large eigenvalues of the local preserving scatter matrix decrease, whereas the small ones increase. The decay of eigenvalues slows down. The zero eigenvalues are no longer zeros. The useful discriminative information retains. Hence, RLDE is more stable.

   In addition, RLDE preserves the data diversity via the regularization terms. The data diversity has been proven to be effective in improving the classification performance [41]. In RLDE, the covariance regularization part corresponds to the data variance. By maximizing the data variance, the data diversity can be well preserved [42]. Moreover, the diagonal regularization involves the diversity of data in the same class and was known to be beneficial for classification [15], [43]. In RLDE in (3), the diagonal regularization decreases the cross terms of the local preserving scatter distances. Relatively, it emphasizes more on the diagonal terms which reflect the intraclass variation at a certain extent.
4) Algorithm flexibility. When $\beta = 0$, (7) is reduced to (5). SSRLDE reverts back to LPNPE. LPNPE does not use the spectral-domain labeled information but uses only

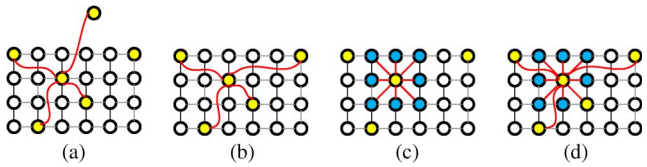


Fig. 2. Local neighbors. (a) Spectral-domain neighbors. (b) Spectral-domain neighbors after spatial filtering. (c) Spatial-domain local pixel neighbors. (d) Spatial–spectral local neighbors.

the spatial-domain local pixel neighbors of each training sample. When a large amount of training samples (labeled or unlabeled) are available, LPNPE can be significantly improved.

When $\beta = 1$, (7) is reduced to (4). SSRLDE becomes RLDE, which does not use the spatial information. When the spatial local pixel neighbors and pixel positions for each training sample are not available, RLDE can still be applicable. We will show later that RLDE outperforms the state-of-the-art spectral-based DR methods.

For different problems, parameter β needs to be tuned accordingly to balance the spatial and spectral information.

5) Local spatial–spectral similarity. SSRLDE integrates both the local spectral and spatial similarity discriminant information. In the spectral domain, RLDE preserves the local Euclidean neighborhood of the same class. Considering four NNs, the spectral-domain within-class neighborhood relation is shown in Fig. 2(a), where a labeled sample is connected with its four NNs in the same class (yellow points). Because HSIs have multimodal data structures, the data of a specific class usually contain several subregions. Therefore, the nearest within-class neighbors of a given sample may distribute in different spatial subregions. In Fig. 2(a), the upper isolated sample just belongs to another subregion and has large pixel (or location) distances to other labeled samples. If it forces to preserve the Euclidean neighborhood relations and thus connects different subregions, the projection might destroy the spatial structures of HSIs. The spatial filtering can alleviate this problem at a certain extent. It increases the similarity between neighboring pixels, reduces spectral variations in a local region, and thus improves the spectral neighboring relations. After the spatial filtering, the samples in a local pixel neighborhood are similar, and the four nearest labeled spectral-domain neighbors are changed, as shown in Fig. 2(b). In the spatial domain, LPNPE preserves the local pixel neighborhood relations shown in Fig. 2(c), where the window of 3×3 is used and the labeled sample is connected with its eight pixel neighbors (blue points). SSRLDE incorporates RLDE and LPNPE. It aims to preserve both the local spatial–spectral neighbors, as shown in Fig. 2(d).

F. Projected Data Distribution

Here, we visualize the projected data distribution of SSRLDE. For this purpose, we choose three classes Reeds1, Riparian, and Firescar2 from the Botswana data set [shown in Fig. 6(a)] and select 60 samples from each class to train the model. We compare the projection results of the proposed RLDE and SSRLDE with those of three spectral-based DR methods: LDA, LFDA [16], and LDE [17]. In SSRLDE, we use a moderate scale of 9×9 for simplicity. The data are projected into two dimensions. Fig. 3 shows the 2-D projected testing data distributions.

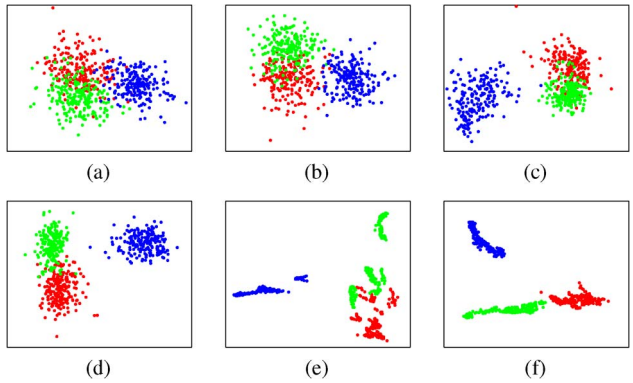


Fig. 3. Two-dimensional projections for different DR methods. The red, green, and blue points represent the Reeds1, Riparian, and Firescar2, respectively. (a) LDA. (b) LFDA. (c) LDE. (d) RLDE. (e) SSRLDE ($\alpha = 0.5,\ \beta = 0.1$); (f) SSRLDE ($\alpha = 0.5,\ \beta = 0.5$).

LDA and LFDA show lower interclass separability. For LDE shown in Fig. 3(c), the Firescar2 class in the blue color is visibly separated, whereas the other two classes in the red and green colors are largely overlapped. Compared with three existing spectral DR methods, RLDE provides a relatively better discriminant result. These four spectral domain DR methods provide large intraclass scatters. By using the spatial information, the projected data of SSRLDE are more compact, as shown in Fig. 3(e) and (f). In Fig. 3(e), $\beta = 0.1$ implies that the spatial-domain LPNPE is dominant; therefore, the projection mainly preserves the local pixel neighborhood. From the ground-truth map of the original data, as shown in Fig. 6(e), we can see that the three classes Reeds1, Riparian, and Firescar2 have about 6, 6, and 3 subregions, respectively. LPNPE preserves the spatial local neighborhood structures, that is, spatial nearby pixels in the original space are mapped to nearby pixels in the low-dimensional feature space. Thus, the projected data in Fig. 3(e) also have several clusters in each class. In Fig. 3(f), a moderate value of β is used and the corresponding SSRLDE preserves the spatial local pixel neighborhood while connecting different subregions of the same class using the spectral labeled samples. It realizes a tradeoff of the local neighborhood preserving between the spectral and spatial domains. The projected data are concentrated within the same class and dispersed between different classes.

Furthermore, Fig. 4 plots the 2-D projected distributions of SSRLDE for all 14 classes in the testing data of Botswana. Although there are several overlaps, the projected data show that SSRLDE has good class separability considering that only two features are used.

G. Comparison

Here, we investigate the effectiveness of the regularization, spatial filtering, spatial–spectral discriminant, and multiscale

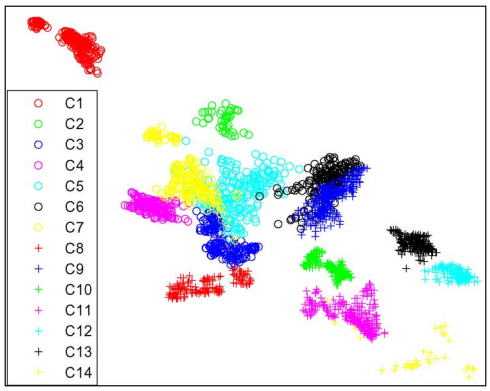


Fig. 4. Two-dimensional data projections of 14 classes for SSRLDE.

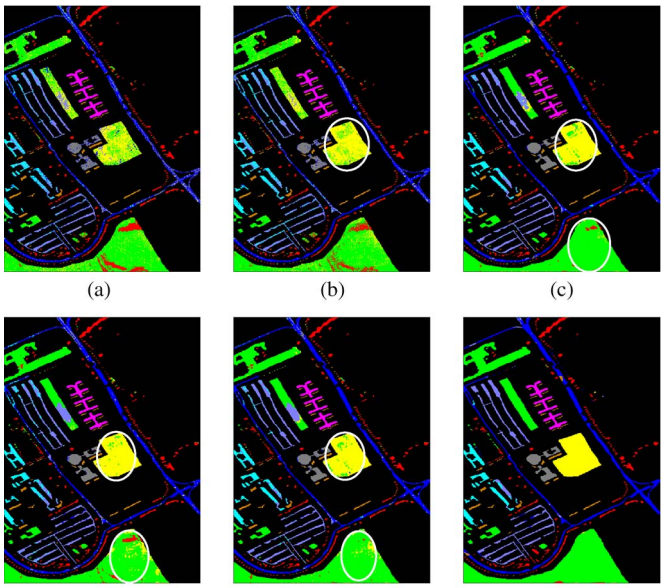


Fig. 5. Classification maps of the University of Pavia data set. (a) LDE. (b) RLDE. (c) RLDE on the 3×3 filtered data. (d) LPNPE on the 3×3 filtered data. (e) SSRLDE on the 3×3 filtered data. (f) SSRLDE on the multiscale filtered data.

in SSRLDE from the classification map containing only the training and testing samples of the University of Pavia data set. The classification maps of LDE and RLDE on the original spectral data, RLDE, LPNPE, and SSRLDE on the 3×3 filtered data, and SSRLDE on the multiscale filtered data are shown in Fig. 5, where 30 samples in each class are chosen as the training samples.

From Fig. 5, we can find that:

1) Comparing Fig. 5(b) with (a), regularization in RLDE improves LDE, particularly the central circled region.
2) Comparing Fig. 5(c) with (b), the spatial filtering preprocessing increases the separability of homogenous regions. The spatial filtering removes most noise in large homogeneous areas, as shown in two circled regions in Fig. 5(c).
3) Investigating Fig. 5(d) and (c), there exists complementarity between the spectral and spatial information. In the central circled region, LPNPE shows better results. However, in the bottom circled region, RLDE performs much better.
4) SSRLDE takes both advantages of RLDE and LPNPE. The joint spatial–spectral discriminant in Fig. 5(e) is much better than either the spectral discriminant in Fig. 5(c) or the spatial discriminant in Fig. 5(d).
5) The multiscale method exploits the rich complementary information. The corresponding SSRLDE provides satisfactory classification map in Fig. 5(f).

III. Experimental Results

Here, we perform SSRLDE on four HSI data sets. The effectiveness of SSRLDE is evaluated by the classification overall accuracy (OA) of the 1-NN and SVM [44] classifiers on the projected data.[1]

A. Hyperspectral Data Sets

1) Okavango Delta, Botswana: The data set was acquired by the NASA EO-1 satellite over the Okavango Delta, Botswana on May 31, 2001 [45]. The image scene has the size of 1476×256 pixels. After discarding water absorption and noisy bands, 145 bands are retained. The data contain 3428 samples from 14 identified classes [45].[2] The false color image and the ground-truth map are shown in Fig. 6(a) and (e).
2) Kennedy Space Center (KSC): The data set was acquired by the NASA AVIRIS instrument over the KSC, Florida, on March 23, 1996 [45].[3] The image scene has the size of 512×614 pixels and 224 spectral channels. After discarding water absorption and noisy bands, 176 bands are retained. It contains 13 ground-truth classes. The total number of samples is 5211 ranging from 105 to 927 in each class. The false color composition of bands 31, 21, and 11 and the ground-truth map are shown in Fig. 6(b) and (f).
3) University of Pavia: The data set was acquired in 2001 by the ROSIS instrument over the city of Pavia, Italy [36].[4] This image scene corresponds to the University of Pavia and has the size of 610×340 pixels and 115 spectral bands. After discarding noisy and water absorption bands, 103 bands are retained. The data contain 9 ground-truth classes. The false color composition of bands 60, 30, and 2 and the ground-truth map are shown in Fig. 6(c) and (g).
4) Indian Pines: The data set was acquired by the AVIRIS sensor in 1992.[4] The image scene contains 145×145 pixels and 220 spectral bands, where 20 channels were discarded because of the atmospheric affection. There are 16 classes in the data set. The total number of samples is 10249 ranging from 20 to 2455 in each class. The false color composition of bands 50, 27, and 17 and the ground-truth map are shown in Fig. 6(d) and (h).

[1] Available online: http://www.csie.ntu.edu.tw/~cjlin/libsvm
[2] Available online: http://www.csr.utexas.edu/hyperspectral/data/Botswana/
[3] Available online: http://www.csr.utexas.edu/hyperspectral/data/KSC/
[4] Available online: http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes

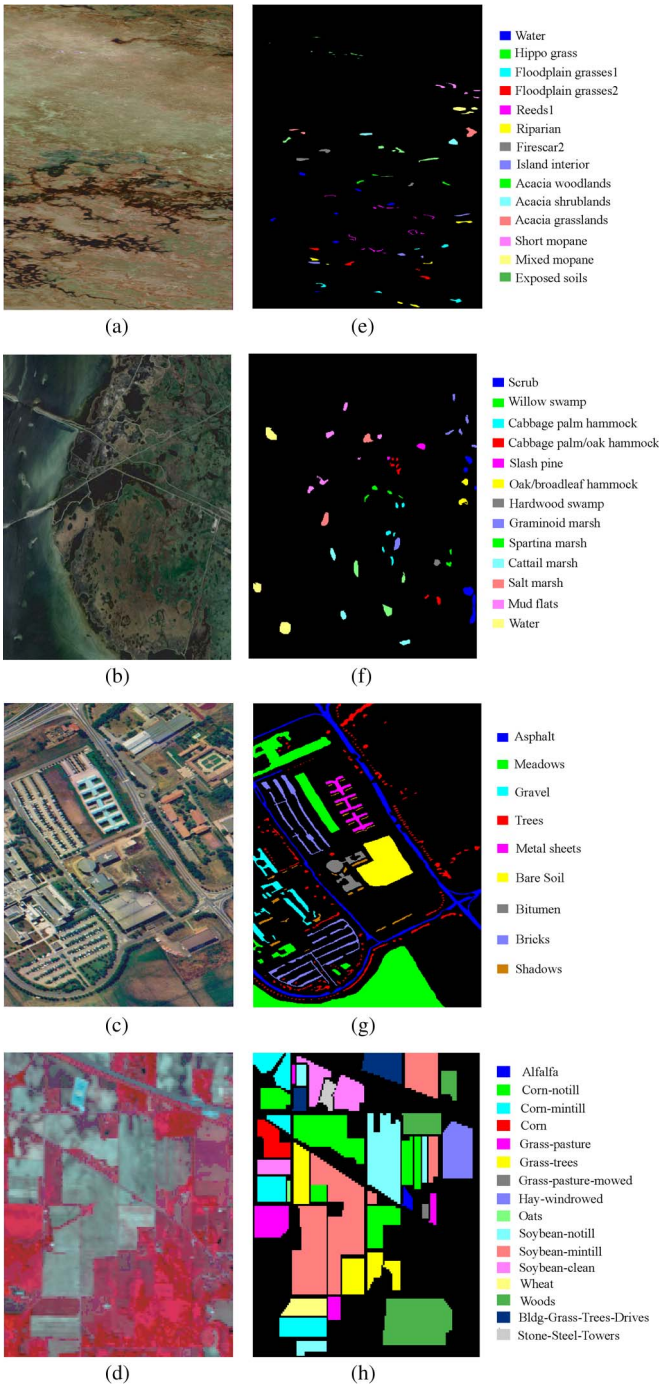


Fig. 6. RGB composite images and ground-truth maps of four data sets. (a) Botswana image of bands 51, 149, and 31. (b) KSC image of bands 31, 21, and 11. (c) University of Pavia image of bands 60, 30, and 2. (d) Indian Pines image of bands 50, 27, and 17. (e)–(h) Corresponding ground-truth maps.

B. Parameter Selection

To construct the spectral-domain local preserving scatter matrix, SSRLDE needs to determine three basic RLDE graph parameters: the number of intraclass and interclass neighbors k_1 and k_2, and the heat kernel parameter t. In the experiments, we empirically set $k_1 = k_2 = 5$ and $t = 0.5$. Then, we tune the regularization parameters α and β to make a compromise between the spectral-domain and spatial-domain local neighborhood preserving. The parameters α and β are chosen in the set $\{0, 0.1, 0.2, \ldots, 0.9\}$. A fixed 3×3 neighborhood window is used in the spatial WMF. Fifteen samples in each class are selected as the training set for the parameter selection. We find the optimum parameter values in the region to ensure that the classification OA of the SSRLDE method combined with the NN classifier is maximized.

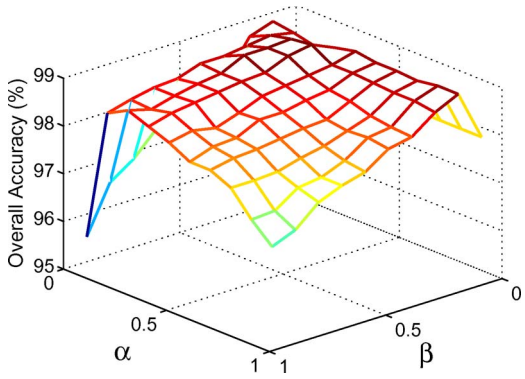


Fig. 7. OA versus α and β on the Botswana data set.

We take the Botswana data set as an example to show the parameter selection. Fig. 7 shows the OA of SSRLDE as a function of α and β. It can be clearly seen that the overall accuracies are relatively small in three boundaries. When $\alpha = 0$ and $\beta = 1$, SSRLDE is reduced to LDE. When $\alpha = 1$ and $\beta = 1$, SSRLDE can be considered as a modified PCA. When $\beta = 0$, SSRLDE reverts back to LPNPE. In these three cases, SSRLDE provides the worse results. The parameter pair $(\alpha, \beta) = (0.3, 0.2)$ achieves the maximum OA and is then used in the following experiments.

Similarly, we can find the optimal parameters for other data sets. Namely, for the KSC data set, $\alpha = 0.1$ and $\beta = 0.4$; for the University of Pavia data set, $\alpha = 0.2$ and $\beta = 0.3$; and for the Indian Pines data set, $\alpha = 0.1$ and $\beta = 0.1$.

C. Investigation of the Local Preserving Term

SSRLDE contains two local preserving terms: the spectral-domain and spatial-domain local neighborhood preserving terms. When considering only one of them, SSRLDE reverts back to RLDE or LPNPE, as discussed in Section II-E. Here, we investigate the effectiveness of these two local preserving terms separately.

We randomly select 15 labeled samples from each class to form the training set. The remaining samples construct the testing set. The reduced dimensionality varies from 2 to 30. The NN classifier is used. The averaged classification OA of over ten runs is measured. The classification results of RLDE, LPNPE, and SSRLDE are shown in Fig. 8.

As can be seen from the results, for the Botswana and University of Pavia data sets, RLDE performs better than LPNPE. For the KSC and Indian Pines data sets, LPNPE is better than RLDE. Neither of these two methods (i.e., spectral based and spatial based) is intrinsically better than the other. However, in all cases, SSRLDE is superior to RLDE and LPNPE. This demonstrates that the spectral information and spatial information are essential and complementary. By integrating the two local neighborhood preserving terms, SSRLDE achieves higher

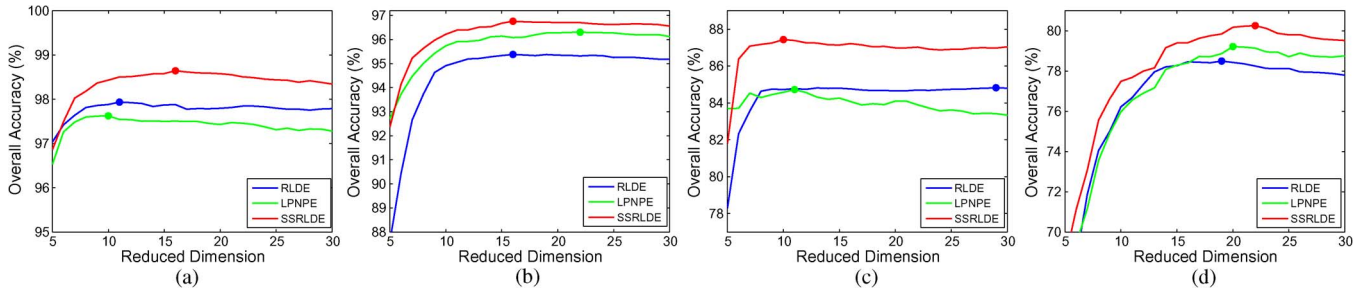


Fig. 8. Comparison of the local preserving terms on four data sets. (a) Botswana. (b) KSC. (c) University of Pavia. (d) Indian Pines. (The highest OAs are marked as solid circles.).

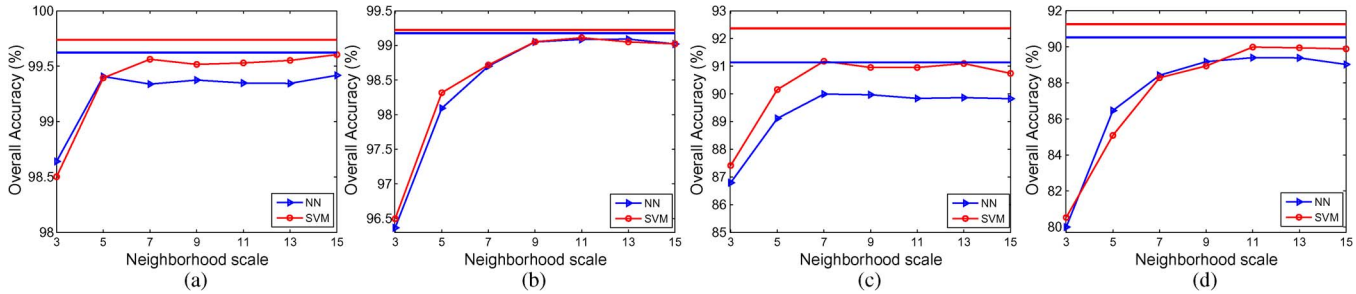


Fig. 9. Effects of the neighborhood scale on four data sets. (a) Botswana. (b) KSC. (c) University of Pavia. (d) Indian Pines. (The horizonal lines correspond to the ensemble results of NN and SVM.).

classification accuracies than that of the individual discriminant projection.

D. Investigation of the Neighborhood Scale

The neighborhood scale (window size) affects the filtering results. To show its impact, we provide seven different scales, i.e., 3×3, $5 \times 5, \ldots, 15 \times 15$. Fifteen samples in each class are selected as the training set, and the remaining samples are set as testing samples. The reduced dimension is fixed as 15. Fig. 9 shows the changes of overall accuracies as a function of scales for four data sets, where the NN and SVM classification results are reported and the corresponding ensemble results using the majority voting are indicated by the horizonal lines.

From Fig. 9, we can see that the best OAs for the Botswana, KSC, University of Pavia, and Indian Pines data sets are achieved at the neighborhood scale 15, 11, 7, 11, respectively. Roughly, the large scale benefits HSI classification. This may be because of the use of WMF and the presence of large homogenous regions. The WMF endows each pixel in the window with a weight according to its spectral similarity to the central pixel. By adjusting the weights, the effect of unrelated pixels in a large window decreases, whereas the effect of the similar pixels is emphasized.

For four data sets, the multiscale outputs outperform the optimal results in each single scale. The multiscale windows can provide an accurate description of various HSI homogeneous regions. The ensemble of complementary information in different scales achieves higher OAs.

E. Contribution of Each Step in SSRLDE

Here, we analyze the effect and contribution of each processing step of SSRLDE to its final classification results. We compare the algorithms with and without the weighted mean filter (WMF)-based spatial preprocessing (No WMF versus WMF), dimensionality reduction (No DR versus DR), regularization (LDE versus RLDE), joint spatial–spectral discriminant analysis (RLDE, LPNPE versus SSRLDE1), and multiscale voting (SSRLDE1 versus SSRLDE). The comparison results in terms of the classification overall accuracies in the case of 15 labeled samples for four data sets are shown in Table I, where SSRLDE1 refers to the combination of RLDE and LPNPE. In the case of "WMF," a single neighborhood window 3×3 is used for different algorithms except for SSRLDEopt and SSRLDE, where SSRLDEopt refers to SSRLDE with the optimal window selected in Section III-D and SSRLDE is the ensemble of seven different scales.

Throughout different processing steps of SSRLDE in Table I, the WMF-based spatial preprocessing dramatically enhances OAs on the original spectral data by increasing the similarity and consistency of neighboring pixels, then the dimensionality reduction method of LDE improves the classification results on the whole spectral bands by discarding redundant features. The regularization method used in RLDE plays an important role and achieves higher OAs than LDE. Later, SSRLDE1 takes advantages of both the spatial and spectral discriminant information and provides the better classification performance than either the spectral-based RLDE or spatial-based LPNPE. Finally, a multiscale ensemble strategy further improves the classification results by exploiting the rich complementary information in different scales.

F. Comparison With Other DR Methods

For evaluating the performance of the proposed SSRLDE, we randomly choose $N = 5, 10, 15, 20, 25, 30$ samples from

TABLE I
CLASSIFICATION ACCURACIES (IN PERCENT) ON FOUR DATA SETS IN THE CASE OF 15 LABELED SAMPLES TO SHOW DIFFERENT CONTRIBUTIONS OF EACH PROCESSING STEP OF SSRLDE TO ITS FINAL CLASSIFICATION RESULTS

			No DR	LDE	RLDE	LPNPE	SSRLDE1	SSRLDEopt	SSRLDE
Bot	No WMF	NN	84.19±0.86	89.01(11)±1.20	91.15(18)±0.79	93.68(13)±0.72	94.37(13)±0.68	—	—
		SVM	89.64±0.61	90.86(12)±0.50	91.82(26)±0.77	94.16(15)±1.11	94.82(29)±0.73	—	—
	WMF	NN	93.03±0.59	96.32(6)±0.37	98.41(26)±0.31	98.06(11)±0.95	98.70(15)±0.31	99.54(7)±0.56	99.70(8)±0.31
		SVM	96.17±0.95	96.57(6)±0.62	98.56(26)±0.31	98.57(9) ±0.36	98.77(30)±0.56	99.50(18)±0.56	99.67(27)±0.29
KSC	No WMF	NN	82.07±1.36	87.49(8)±1.02	88.96(11)±1.12	93.21(21)±0.73	93.33(15)±1.11	—	—
		SVM	88.38±1.73	87.94(8)±1.41	89.56(25)±1.04	93.17(17)±1.11	93.43(20)±1.11	—	—
	WMF	NN	88.74±1.62	93.26(7)±1.43	94.65(28)±1.34	96.64(16)±0.68	97.43(17)±0.85	99.03(30)±0.66	99.19(27)±0.64
		SVM	93.39±1.19	93.13(7)±1.36	95.00(29)±1.25	96.80(27)±0.81	97.17(22)±0.82	98.87(18)±0.75	99.11(23)±0.68
Pavia	No WMF	NN	65.99±2.65	73.10(18)±2.94	75.87(16)±2.10	77.72(7)±2.57	80.43(7)±2.24	—	—
		SVM	74.08±2.57	76.10(26)±2.50	78.64(17)±1.86	79.17(9)±2.07	81.59(9)±2.43	—	—
	WMF	NN	69.88±2.76	83.44(7)±1.92	83.83(8)±2.32	83.62(9)±3.34	86.03(9)±2.25	89.64(30)±1.99	91.04(30)±2.04
		SVM	80.17±3.41	84.33(7)±2.64	86.51(8)±2.24	85.74(24)±2.41	88.18(27)±1.70	90.87(30)±2.40	92.26(30)±2.17
Indian	No WMF	NN	51.45±2.28	59.34(25)±2.06	65.56(13)±1.77	71.18(19)±1.60	71.97(19)±1.71	—	—
		SVM	60.05±1.65	61.18(27)±2.15	64.58(19)±1.44	71.69(19)±2.09	72.78(24)±1.89	—	—
	WMF	NN	65.64±2.64	73.33(20)±1.90	78.21(22)±1.94	79.08(28)±1.91	80.02(24)±2.07	90.16(30)±1.86	91.11(30)±1.95
		SVM	74.89±2.20	75.45(20)±1.70	77.81(22)±1.60	80.04(29)±2.30	81.13(30)±2.72	90.67(28)±2.07	91.90(28)±2.19

TABLE II
HIGHEST OVERALL ACCURACIES (IN PERCENT) ON THE BOTSWANA DATA SET

N	Classfier	RAW	PCA	LDA	LFDA	NWFE	LDE	RLDE	LPNPE	SDA	SELDLPP	SELDNPE	SSRLDE
5	NN	86.59 ±2.17	86.57(30) ±2.16	90.75(10) ±1.30	92.33(27) ±1.51	92.16(30) ±2.31	91.77(5) ±1.60	94.53(29) ±0.96	90.40(9) ±2.11	94.98(29) ±1.27	94.77(28) ±1.37	95.52(21) ±1.00	**96.29(7) ±1.59**
	SVM	89.82 ±2.61	89.77(27) ±2.61	90.72(10) ±1.33	92.66(29) ±1.19	92.78(30) ±2.04	91.73(5) ±2.37	94.61(24) ±1.04	92.06(26) ±2.84	94.84(24) ±0.98	94.95(18) ±1.62	95.26(22) ±1.34	**96.42(9) ±2.11**
10	NN	90.96 ±1.11	90.81(29) ±1.09	94.28(13) ±0.42	93.69(27) ±1.17	94.60(30) ±1.09	94.36(5) ±1.39	96.95(11) ±0.56	96.40(9) ±0.85	97.30(21) ±0.75	97.21(28) ±0.66	97.65(22) ±0.89	**98.65(8) ±0.93**
	SVM	94.17 ±1.53	93.99(30) ±1.47	94.33(13) ±0.51	94.00(30) ±0.62	95.94(28) ±0.79	93.97(5) ±1.46	97.28(25) ±0.39	97.01(21) ±1.07	97.00(30) ±1.13	97.78(30) ±0.53	97.61(30) ±1.11	**98.68(14) ±0.92**
15	NN	93.03 ±0.59	92.94(30) ±0.60	96.71(13) ±0.62	96.88(20) ±0.76	95.87(29) ±0.82	96.32(6) ±0.37	98.41(26) ±0.31	98.06(11) ±0.95	98.71(11) ±0.35	98.53(28) ±0.43	99.24(29) ±0.16	**99.70(8) ±0.31**
	SVM	96.17 ±0.95	95.97(28) ±0.91	96.50(13) ±0.70	97.10(30) ±0.58	97.81(28) ±0.66	96.57(6) ±0.62	98.56(26) ±0.31	98.57(9) ±0.36	98.58(21) ±0.51	98.60(26) ±0.24	99.05(22) ±0.48	**99.67(27) ±0.29**
20	NN	93.72 ±0.67	93.54(30) ±0.75	97.42(13) ±0.56	97.51(16) ±0.64	96.26(29) ±0.59	96.73(6) ±0.47	98.46(18) ±0.64	98.85(13) ±0.34	98.79(11) ±0.49	98.76(29) ±0.43	99.23(18) ±0.39	**99.88(14) ±0.12**
	SVM	96.83 ±0.77	96.79(30) ±0.83	97.26(13) ±0.64	97.63(19) ±0.58	97.80(27) ±0.63	96.91(6) ±0.60	98.57(21) ±0.74	99.10(26) ±0.36	98.54(28) ±0.83	98.81(30) ±0.85	99.05(18) ±0.48	**99.93(21) ±0.12**
25	NN	94.78 ±0.79	94.62(30) ±0.77	98.14(13) ±0.44	98.08(20) ±0.59	96.96(29) ±0.75	96.86(6) ±0.52	98.93(14) ±0.34	99.10(11) ±0.27	99.08(11) ±0.29	98.93(29) ±0.28	99.42(16) ±0.26	**99.96(22) ±0.08**
	SVM	97.92 ±0.52	97.85(29) ±0.50	97.65(13) ±0.46	98.18(21) ±0.47	98.45(30) ±0.49	97.41(6) ±0.49	99.04(29) ±0.33	99.16(26) ±0.28	98.97(11) ±0.56	99.14(24) ±0.23	99.27(30) ±0.33	**99.90(21) ±0.10**
30	NN	95.00 ±0.76	94.83(30) ±0.74	98.65(11) ±0.41	98.28(16) ±0.57	97.05(30) ±0.62	97.01(6) ±0.55	99.00(13) ±0.50	99.31(11) ±0.19	99.17(11) ±0.37	98.96(27) ±0.42	99.49(15) ±0.25	**99.99(23) ±0.03**
	SVM	97.89 ±0.55	97.71(28) ±0.62	98.31(13) ±0.65	98.67(21) ±0.50	98.66(26) ±0.43	97.54(6) ±0.33	99.32(28) ±0.42	99.40(30) ±0.37	99.06(11) ±0.53	99.27(30) ±0.49	99.46(16) ±0.36	**99.98(20) ±0.09**

each class to form the training set, respectively (For the Indian Pines data set, at a maximum half of total samples in Grass-pasture-mowed and Oats classes are chosen). The remaining samples are set as the testing set. In each case, the experiment is repeated ten times with randomly chosen training samples. Finally, the results of ten runs are averaged. In SSRLDE, seven scales from 3×3 to 15×15 are used. For a fixed scale, HSI pixels are preprocessed by the spatial WMF defined in (2). Performing the joint spatial and spectral local discriminant on the filtered data, the spatial–spectral features can be extracted, as shown in (6). A classifier is then performed on the extracted features to obtain the classification results (the prediction vector contains the predicted labels for all samples). For multiple scales, we repeat the aforementioned processes for each scale to obtain the multiple prediction vectors. Fusing the multiple prediction results with a majority voting, we can obtain the final classification label for each sample.

The proposed SSRLDE is compared with other commonly used supervised DR methods, including PCA, LDA, LFDA [16], NWFE [15], and LDE [17], and semisupervised DR methods, including SDA [19], SELDLPP, and SELDNPE [9]. For the semisupervised DR methods, the unlabeled samples are chosen as follows: for the Botswana and KSC data sets, all testing samples are set as unlabeled samples, for the University of Pavia and Indian Pines data sets, 300 samples of each class in the testing set are chosen as unlabeled samples (For the class less than 300 testing samples, all testing samples are used). These spectral-based DR methods are performed on the filtered data using a 3×3 weighted mean filter. The classification results on the filtered data without dimensionality reduction

TABLE III
HIGHEST OVERALL ACCURACIES (IN PERCENT) ON THE KSC DATA SET

N	Classfier	RAW	PCA	LDA	LFDA	NWFE	LDE	RLDE	LPNPE	SDA	SELDLPP	SELDNPE	SSRLDE
5	NN	81.58	81.58(21)	88.39(12)	88.55(19)	84.54(20)	86.34(6)	88.39(29)	89.17(21)	92.00(26)	89.93(30)	90.93(28)	**96.09(17)**
		±1.82	±1.83	±1.61	±1.57	±3.47	±2.63	±3.04	±3.20	±2.39	±2.25	±2.31	**±2.34**
	SVM	84.43	84.85(21)	88.38(12)	88.56(15)	84.63(29)	86.71(6)	88.92(22)	89.32(30)	91.74(21)	90.65(25)	91.11(29)	**95.92(25)**
		±3.15	±3.12	±1.59	±1.57	±3.84	±2.73	±3.38	±3.19	±2.35	±2.34	±2.53	**±2.74**
10	NN	85.94	85.92(24)	91.02(12)	92.26(30)	88.97(23)	91.35(7)	93.14(21)	95.49(30)	95.36(28)	94.22(30)	95.40(30)	**99.06(19)**
		±1.99	±1.98	±1.21	±1.27	±2.77	±1.90	±1.42	±1.23	±1.87	±2.06	±1.48	**±0.89**
	SVM	90.41	90.38(20)	91.05(12)	92.15(19)	90.22(21)	91.47(7)	93.59(22)	95.49(30)	95.08(11)	94.98(30)	95.61(29)	**98.94(23)**
		±3.04	±3.05	±1.19	±1.19	±2.56	±1.98	±1.75	±1.38	±1.52	±1.61	±1.43	**±0.98**
15	NN	88.74	88.73(29)	92.84(12)	94.80(22)	91.90(30)	93.26(7)	94.65(28)	96.64(16)	96.68(30)	95.90(30)	96.55(29)	**99.19(27)**
		±1.82	±1.64	±1.12	±0.99	±1.10	±1.43	±1.34	±0.68	±0.99	±1.22	±0.95	**±0.64**
	SVM	93.39	93.32(25)	92.73(12)	94.64(29)	93.49(26)	93.13(7)	95.00(29)	96.80(27)	96.66(27)	96.39(28)	96.53(30)	**99.11(23)**
		±1.19	±1.27	±1.20	±1.10	±1.18	±1.36	±1.25	±0.81	±0.96	±0.83	±0.66	**±0.68**
20	NN	89.72	89.72(29)	94.71(12)	96.18(19)	93.22(27)	94.34(7)	95.69(22)	97.12(27)	97.15(30)	96.73(30)	97.20(29)	**99.41(30)**
		±1.04	±1.04	±0.92	±0.68	±0.87	±1.01	±0.85	±0.96	±0.64	±0.49	±0.54	**±0.41**
	SVM	94.65	94.60(19)	94.36(12)	95.94(29)	94.01(21)	93.97(7)	96.21(26)	97.22(20)	97.11(27)	97.05(30)	97.05(30)	**99.53(11)**
		±0.96	±0.87	±0.97	±0.67	±1.39	±1.50	±0.59	±0.63	±0.86	±0.86	±0.81	**±0.30**
25	NN	91.07	91.06(29)	95.98(12)	97.14(22)	93.68(24)	94.66(7)	96.77(30)	97.71(27)	97.65(27)	97.24(29)	97.74(30)	**99.45(12)**
		±1.01	±1.02	±0.57	±0.49	±1.19	±0.96	±0.47	±0.57	±0.57	±0.69	±0.64	**±0.36**
	SVM	95.01	95.07(27)	95.67(12)	96.87(30)	95.16(22)	94.55(7)	96.85(24)	97.65(14)	97.49(25)	97.35(29)	97.45(26)	**99.55(9)**
		±1.09	±1.11	±0.59	±0.38	±0.85	±1.03	±0.66	±0.67	±0.50	±0.58	±0.60	**±0.31**
30	NN	92.10	92.09(28)	96.66(12)	97.83(24)	94.38(30)	95.54(7)	96.92(25)	98.47(29)	98.16(30)	98.00(30)	98.34(30)	**99.82(12)**
		±0.74	±0.74	±0.41	±0.56	±0.80	±0.69	±0.67	±0.50	±0.28	±0.24	±0.25	**±0.14**
	SVM	96.49	96.31(26)	96.06(12)	97.90(27)	95.88(26)	95.51(7)	97.42(27)	98.38(16)	98.06(27)	98.24(29)	98.05(30)	**99.85(23)**
		±0.58	±0.63	±0.45	±0.37	±0.68	±0.60	±0.53	±0.28	±0.42	±0.25	±0.53	**±0.05**

TABLE IV
HIGHEST OVERALL ACCURACIES (IN PERCENT) ON THE UNIVERSITY OF PAVIA DATA SET

N	Classfier	RAW	PCA	LDA	LFDA	NWFE	LDE	RLDE	LPNPE	SDA	SELDLPP	SELDNPE	SSRLDE
5	NN	60.03	60.03(24)	72.38(8)	73.50(16)	64.89(26)	73.03(8)	75.14(30)	73.90(8)	70.44(21)	68.78(21)	70.39(25)	**80.35(8)**
		±4.43	±4.43	±9.00	±8.55	±5.80	±7.40	±6.79	±4.45	±5.67	±4.24	±5.51	**±5.64**
	SVM	64.36	64.39(14)	72.43(8)	74.67(19)	72.01(25)	73.89(6)	75.53(14)	76.12(14)	75.62(11)	73.37(15)	73.53(25)	**81.36(25)**
		±5.21	±5.13	±8.65	±9.15	±6.45	±8.28	±10.0	±5.96	±6.74	±4.81	±6.16	**±5.74**
10	NN	63.98	63.97(29)	81.24(7)	78.95(8)	71.29(19)	80.84(8)	81.96(11)	82.18(8)	78.33(27)	76.28(16)	78.61(20)	**88.84(10)**
		±4.48	±4.48	±3.25	±2.77	±4.54	±4.40	±4.94	±3.77	±5.10	±4.49	±4.94	**±3.07**
	SVM	76.35	77.61(16)	80.98(8)	80.57(25)	80.34(30)	80.19(9)	83.68(20)	82.55(21)	82.75(30)	83.71(14)	83.04(11)	**90.46(28)**
		±5.56	±4.93	±3.38	±4.00	±4.85	±7.24	±4.21	±5.28	±5.56	±5.24	±4.75	**±4.40**
15	NN	69.88	69.88(28)	83.37(6)	82.29(9)	74.79(22)	83.44(7)	83.83(8)	83.62(9)	81.81(17)	80.54(16)	82.90(14)	**91.04(30)**
		±2.76	±2.76	±2.19	±3.64	±2.53	±1.92	±2.32	±3.34	±2.52	±2.76	±2.60	**±2.04**
	SVM	80.17	80.96(12)	82.13(8)	83.54(29)	83.48(20)	84.33(7)	86.51(8)	85.74(24)	85.49(21)	85.26(23)	86.79(21)	**92.26(30)**
		±3.41	±2.94	±2.10	±2.76	±3.57	±2.64	±2.24	±2.41	±1.96	±1.94	±2.00	**±2.17**
20	NN	69.27	69.27(30)	85.95(7)	85.97(10)	75.97(20)	85.34(10)	87.06(9)	86.44(9)	85.89(21)	82.83(16)	87.01(18)	**93.17(29)**
		±1.32	±1.33	±1.12	±2.30	±1.35	±2.00	±1.49	±2.38	±2.02	±2.01	±1.91	**±1.47**
	SVM	86.16	86.15(29)	85.00(6)	86.98(23)	86.95(25)	86.35(9)	89.77(29)	88.56(18)	87.82(29)	88.39(15)	89.72(20)	**94.13(30)**
		±3.04	±3.02	±1.73	±2.43	±3.78	±2.15	±1.99	±1.92	±1.56	±2.25	±1.39	**±1.50**
25	NN	71.58	71.57(30)	86.63(7)	86.47(10)	78.05(30)	86.32(7)	88.17(12)	87.44(9)	86.96(30)	84.98(15)	87.88(16)	**95.03(30)**
		±1.20	±1.20	±1.07	±1.50	±1.72	±2.95	±1.30	±1.69	±1.87	±1.72	±1.84	**±0.68**
	SVM	88.02	88.35(30)	86.27(8)	89.70(30)	89.60(19)	89.57(10)	91.59(24)	90.66(19)	87.77(21)	89.84(16)	90.82(29)	**96.33(30)**
		±2.32	±1.92	±1.34	±1.82	±2.44	±1.29	±1.11	±1.59	±3.43	±2.25	±1.73	**±0.90**
30	NN	73.24	73.24(30)	88.16(8)	87.97(8)	79.05(30)	87.82(7)	89.47(9)	88.54(8)	88.22(30)	86.63(16)	89.59(16)	**95.67(30)**
		±2.81	±2.81	±1.25	±1.66	±2.36	±1.60	±1.78	±1.28	±1.97	±2.32	±1.88	**±0.67**
	SVM	89.85	89.79(29)	87.91(8)	90.19(9)	90.11(26)	89.47(9)	92.03(24)	90.56(25)	90.05(27)	91.32(20)	91.94(20)	**96.77(27)**
		±1.90	±1.86	±1.30	±0.81	±1.32	±1.74	±1.27	±0.94	±1.67	±1.87	±1.94	**±0.71**

("RAW"), the RLDE classification results without using the spatial discriminant information, and the LPNPE classification results without using the spectral information are also included for comparisons. In LDA, a regularization is used to alleviate the singularity problem by shrinking the within-class scatter matrix toward the identity matrix [13]. In LDE, a PCA preprocessing is employed to overcome the singularity of the local preserving scatter matrix [17]. The NN and SVM classifiers are used. The three-fold cross validation is used to select the optimal penalty parameters C and RBF kernel parameter γ in SVM. The reduced dimensionality varies from 2 to 30, and the best results are presented.

The comparison results on four HSI data sets are shown in Tables II–V, respectively. The results in the tables are the highest overall accuracies (%) among the first 30 features. Each number in the brackets corresponds to the optimal number of extracted features. From these results, we can obtain the following conclusions.

1) DR can improve the HSI classification performance using only few extracted features. This is because most features are redundant and the intrinsic discriminant information is within few features.
2) For all methods, the OA improves as the number of training samples increases.
3) RLDE greatly improves LDE and shows better overall performance than other supervised spectral-based DR methods. This demonstrates the regularization is effective in ameliorating the small-sample-size singularity problem and in improving the classification performance.
4) Using both the labeled and unlabeled data, the semisupervised DR methods show certain improvements over

TABLE V
HIGHEST OVERALL ACCURACIES (IN PERCENT) ON THE INDIAN PINES DATA SET

N	Classfier	RAW	PCA	LDA	LFDA	NWFE	LDE	RLDE	LPNPE	SDA	SELDLPP	SELDNPE	SSRLDE
5	NN	51.88	51.61(30)	59.75(15)	59.05(24)	62.26(16)	58.93(13)	62.61(19)	66.33(26)	60.40(30)	58.58(30)	61.47(29)	**78.18(15)**
		±3.50	±3.44	±3.86	±4.11	±3.40	±3.06	±3.70	±3.63	±3.45	±3.88	±3.40	**±3.60**
	SVM	59.54	58.58(26)	59.95(15)	59.62(25)	64.28(17)	59.58(16)	63.19(29)	67.25(29)	60.76(30)	61.02(30)	63.68(25)	**78.16(27)**
		±4.89	±4.35	±3.72	±4.32	±4.15	±2.58	±3.23	±3.93	±3.55	±3.27	±3.36	**±4.16**
10	NN	60.12	59.71(29)	69.30(15)	64.91(29)	71.19(16)	69.00(14)	73.44(15)	74.78(29)	70.37(16)	67.73(30)	71.18(29)	**86.77(30)**
		±1.92	±1.92	±2.48	±2.45	±2.69	±2.37	±1.96	±3.33	±2.91	±1.90	±2.34	**±2.79**
	SVM	70.07	69.16(29)	69.68(15)	64.47(28)	73.60(18)	69.57(19)	72.78(27)	76.45(27)	72.05(21)	72.06(30)	73.42(27)	**87.44(29)**
		±3.01	±2.62	±2.40	±2.01	±2.60	±2.21	±2.21	±2.52	±2.72	±3.22	±3.44	**±2.33**
15	NN	65.64	65.14(30)	72.36(15)	68.16(30)	75.56(16)	73.33(20)	78.21(22)	79.08(28)	73.72(17)	72.23(30)	75.91(30)	**91.11(30)**
		±2.64	±2.72	±1.86	±1.73	±2.60	±1.90	±1.94	±1.91	±2.13	±1.84	±2.01	**±1.95**
	SVM	74.89	73.91(29)	72.49(15)	67.36(30)	78.12(15)	75.45(20)	77.81(22)	80.04(29)	74.39(15)	77.02(30)	78.98(28)	**91.90(28)**
		±2.20	±2.15	±2.11	±2.10	±2.01	±1.70	±1.60	±2.30	±2.59	±2.21	±1.95	**±2.19**
20	NN	69.02	68.45(30)	76.56(15)	74.01(28)	78.17(18)	76.54(12)	82.13(19)	82.28(19)	77.51(17)	75.41(30)	78.95(29)	**93.84(11)**
		±1.42	±1.44	±1.48	±0.69	±1.77	±1.89	±1.85	±1.70	±1.50	±1.49	±1.81	**±1.20**
	SVM	79.27	78.08(28)	76.59(15)	74.26(30)	81.99(16)	79.41(20)	82.10(25)	83.51(30)	79.31(12)	80.04(30)	81.67(30)	**94.08(30)**
		±1.59	±1.24	±1.30	±1.07	±1.26	±2.09	±1.49	±2.01	±2.73	±1.96	±1.67	**±1.12**
25	NN	70.87	70.18(30)	78.72(15)	77.51(30)	79.82(21)	78.06(15)	83.69(20)	83.82(23)	78.71(16)	77.36(30)	81.38(30)	**94.00(30)**
		±1.81	±1.79	±1.44	±1.63	±1.60	±1.54	±1.27	±1.20	±1.54	±1.24	±1.38	**±1.57**
	SVM	80.49	79.12(28)	78.28(15)	78.04(28)	83.78(20)	81.27(20)	84.17(30)	85.24(18)	80.95(12)	82.50(30)	83.70(29)	**94.51(28)**
		±1.78	±1.57	±1.93	±1.59	±1.95	±2.45	±1.71	±2.08	±3.00	±1.49	±1.44	**±1.68**
30	NN	73.04	72.27(30)	80.14(15)	79.56(30)	81.45(30)	80.15(12)	85.14(19)	86.02(20)	81.31(24)	79.63(30)	83.17(30)	**95.44(30)**
		±0.94	±0.90	±1.07	±1.15	±0.87	±1.00	±0.55	±1.43	±1.00	±0.79	±0.70	**±0.81**
	SVM	82.34	80.95(26)	80.10(15)	79.51(27)	84.47(30)	82.34(20)	85.57(22)	87.47(20)	82.95(14)	83.78(30)	83.69(30)	**95.97(29)**
		±1.34	±1.31	±1.26	±1.81	±1.67	±1.68	±1.21	±1.64	±1.64	±1.33	±2.03	**±1.27**

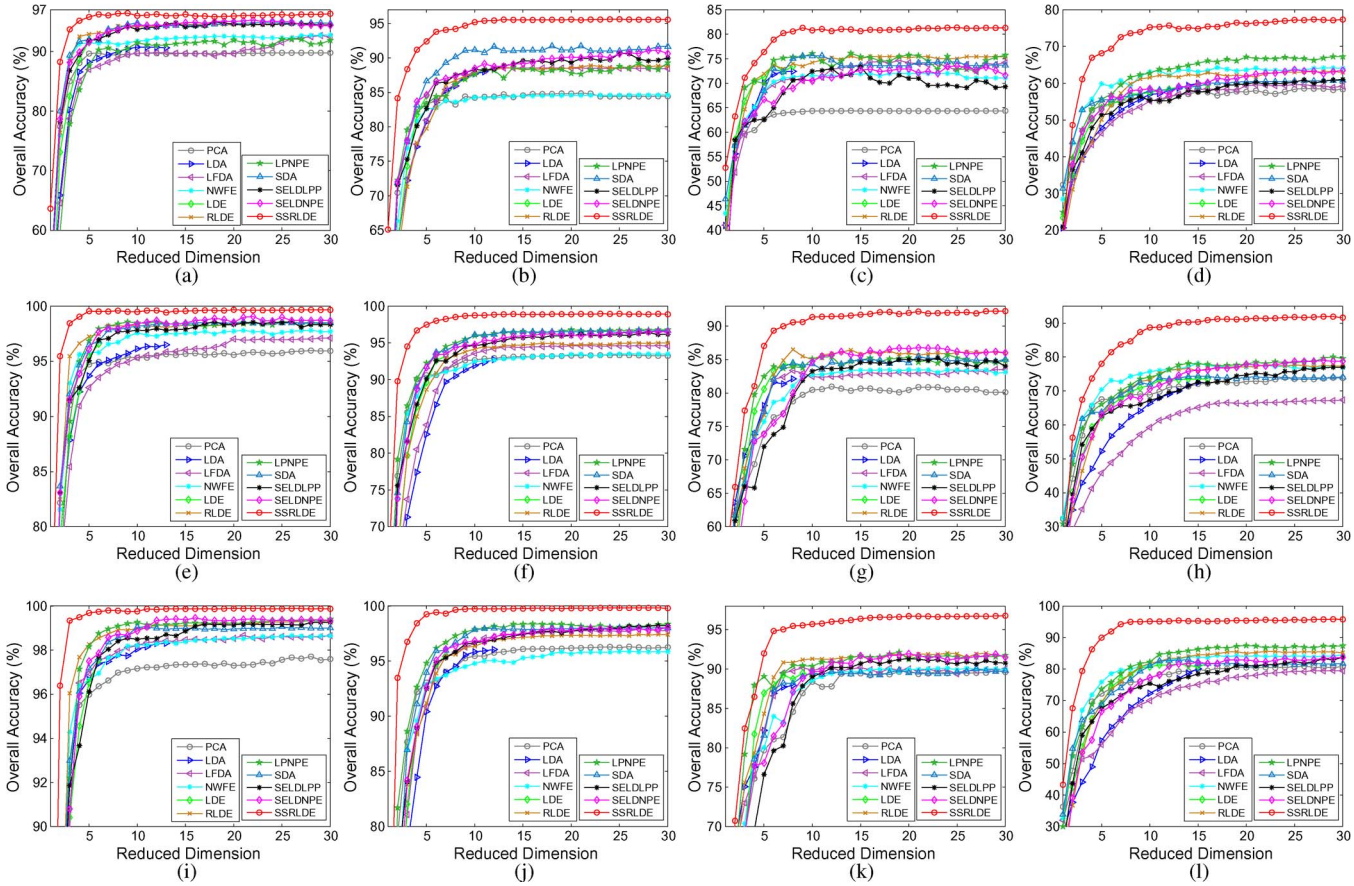


Fig. 10. Effects of the reduced dimensions on four data sets. From left to right, Botswana, KSC, University of Pavia, Indian Pines. (a)–(d) Five labeled samples. (e)–(h) Fifteen labeled samples. (i)–(l) Thirty labeled samples.

the supervised DR methods. Different semisupervised methods are suitable for different data sets.

5) Incorporating the spatial and spectral information, SSRLDE dramatically outperforms the spectral-based supervised and semisupervised DR methods. Although both SSRLDE and semisupervised DR methods use additional unlabeled samples, the unlabeled spatial pixels in SSRLDE contain the interpixel structure information, whereas the potential structure of the unlabeled spectral samples in SDA or SELD are unknown.

Fig. 10 shows the variations of OAs with the reduced dimensions for four data sets, where 5, 15 and 30 samples per class are labeled. We can see that SSRLDE outperforms other

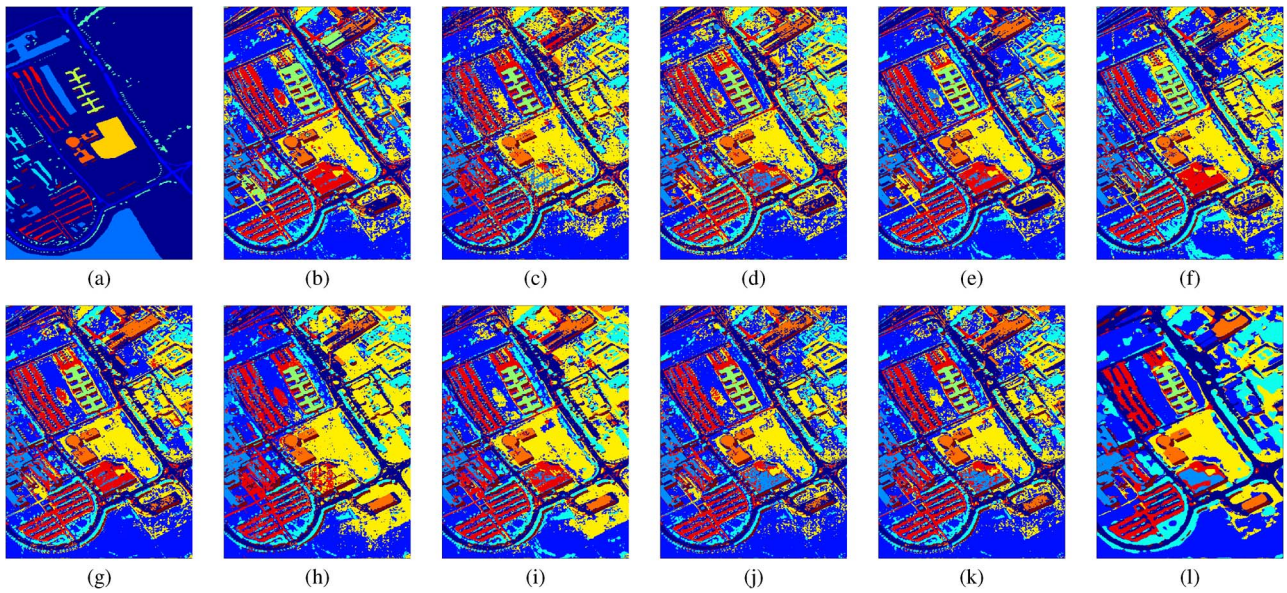


Fig. 11. Classification maps for the University of Pavia data set using the SVM classifier. (a) Ground truth. (b) PCA. (c) LDA. (d) LFDA. (e) NWFE. (f) LDE. (g) RLDE. (h) LPNPE. (i) SDA. (j) SELDLPP. (k) SELDNPE. (l) SSRLDE.

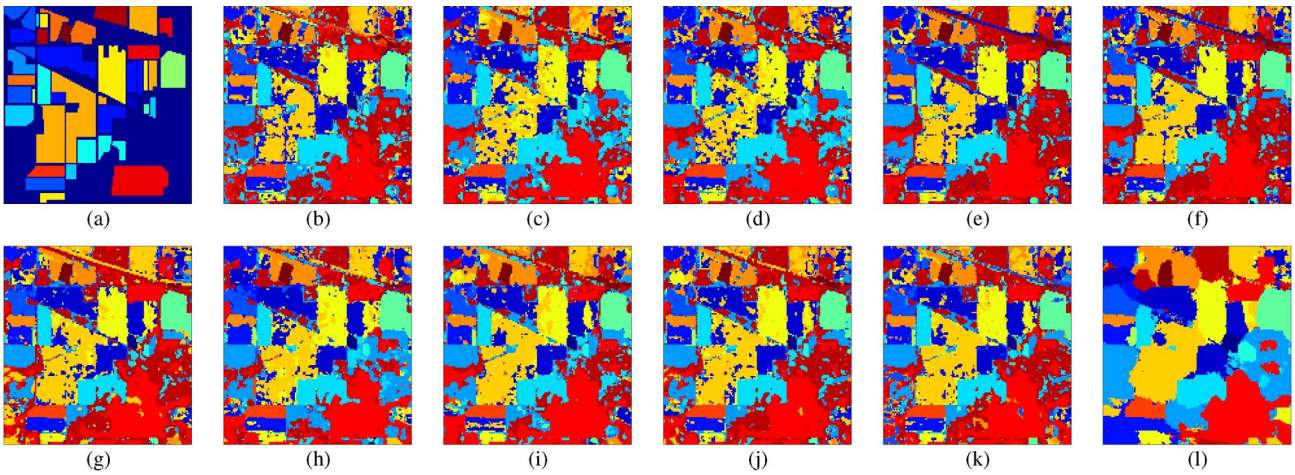


Fig. 12. Classification maps for the Indian Pines data set using the SVM classifier. (a) Ground-truth. (b) PCA. (c) LDA. (d) LFDA. (e) NWFE. (f) LDE. (g) RLDE. (h) LPNPE. (i) SDA. (j) SELDLPP. (k) SELDNPE. (l) SSRLDE.

DR methods in terms of the consistent classification results over a wide range of extracted features. The proposed SSRLDE achieves stable classification results around 8, 12, 15 and 20 features for the Botswana, KSC, University of Pavia, and Indian Pines data sets, respectively.

Figs. 11 and 12 show the classification maps on the whole image for the University of Pavia and Indian Pines data sets. The maps correspond to the classification results using the SVM classifier with different DR methods, including PCA, LDA, LFDA, NWFE, LDE, RLDE, LPNPE, SDA, SELDLPP, SELDNPE, and SSRLDE. As can be seen from Figs. 11 and 12, the spectral-based DR methods produce the classification results with more noise. SSRLDE significantly removes the Salt&Pepper noise on homogeneous regions and leads to a much smoother classification map than the pixelwise algorithms. It also notes that the spatial information used in SSRLDE causes oversmoothing over neighboring classes. Due to the spatial filtering preprocessing, neighboring pixels are forced to be similar. This increases the separability of classes but blurs the edges or boundaries between adjacent objects, particularly when a large filtering window is being used. It can be seen from the maps, the spatial aggregation makes large homogeneous regions moderately growing, and benefits to the classification at both boundary and inner homogeneous regions with the risk of losing detailed structures. It needs to balance the tradeoff between the degree of smoothness and the preservation of boundaries in real applications.

IV. Conclusion

In this paper, we have proposed an SSRLDE method for DR of the HSI data. The proposed method has the following characteristics: 1) the multiscale spatial WMF increases the neighboring pixel consistency and provides the rich and robust complementary information; 2) the regularization strategy overcomes the singularity, embodies the data diversity, and

improves the classification performance; 3) incorporating the spatial and spectral discriminant information leads to excellent discriminative projections. The proposed method has been compared with other DR methods on four HSI data sets in the cases of limited labeled samples. Experimental results have demonstrated that our proposed SSRLDE has better performance and higher classification accuracies.

ACKNOWLEDGMENT

The authors would like to thank Prof. M. Crawford for providing the Botswana and KSC data sets, Prof. P. Gamba for providing the University of Pavia data set, Prof. D. Landgrebe for providing the Indian Pines data set, Prof. M. Sugiyama for sharing LFDA source code, Prof. B. Kuo for sharing NWFE source code, Prof. D. Cai for sharing LPP and NPE source codes, and Prof. C. Lin for providing the LIBSVM toolbox. The authors would also like to thank the handling editor and three anonymous reviewers for their detailed and constructive comments and suggestions, which greatly helped us to improve the quality of our manuscript.

REFERENCES

[1] J. C. Harsanyi and C. I. Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.

[2] A. Plaza *et al.*, "Recent advances in techniques for hyperspectral image processing," *Remote Sens. Environ.*, vol. 113, no. S1, pp. S110–S122, Sep. 2009.

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

[4] B. Guo, S. R. Gunn, R. I. Damper, and J. D. B. Nelson, "Band selection for hyperspectral image classification using mutual information," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 4, pp. 522–526, Oct. 2006.

[5] C. Lee and D. A. Landgrebe, "Analyzing high-dimensional multispectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 31, no. 4, pp. 792–800, Jul. 1993.

[6] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.

[7] S. Chen and D. Zhang, "Semisupervised dimensionality reduction with pairwise constraints for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 2, pp. 369–373, Mar. 2011.

[8] Q. Shi, L. P. Zhang, and B. Du, "Semi-supervised discriminative locally enhanced alignment for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 9, pp. 4800–4815, Sep. 2013.

[9] W. Liao, A. Pižurica, P. Scheunders, W. Philips, and Y. Pi, "Semisupervised local discriminant analysis for feature extraction in hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 184–198, Jan. 2013.

[10] S. Yang *et al.*, "Semisupervised dual-geometric subspace projection for dimensionality reduction of hyperspectral image data," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3587–3593, Jun. 2014.

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

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

[13] J. Friedman, "Regularized discriminant analysis," *J. Amer. Stat. Assoc.*, vol. 84, no. 405, pp. 165–175, Mar. 1989.

[14] 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, Mar. 2009.

[15] 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.

[16] M. Sugiyama, "Dimensionality reduction of multimodal labeled data by local Fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 8, pp. 1027–1061, May 2007.

[17] H. T. Chen, H. W. Chang, and T. L. Liu, "Local discriminant embedding and its variants," in *Proc. Int. Conf. Comput. Vis. Pattern Recognit.*, 2005, pp. 846–853.

[18] S. C. Yan *et al.*, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[19] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," in *Proc. 11th IEEE Int. Conf. Comput. Vis.*, 2007, pp. 1–7.

[20] M. Sugiyama, T. Idé, S. Nakajima, and J. Sese, "Semi-supervised local Fisher discriminant analysis for dimensionality reduction," *Mach. Learn.*, vol. 78, no. 1/2, pp. 36–61, 2010.

[21] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, "Advances in spectral-Spatial classification of hyperspectral images," *Proc. IEEE*, vol. 101, no. 3, pp. 652–675, Mar. 2013.

[22] S. Velasco-Forero and V. Manian, "Improving hyperspectral image classification using spatial preprocessing," *IEEE Geosci. Remote Sens. Lett.*, vol. 6, no. 2, pp. 297–301, Apr. 2009.

[23] M. Lennon, G. Mercier, and L. Hubert-Moy, "Nonlinear filtering of hyperspectral images with anisotropic diffusion," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2002, pp. 2477–2479.

[24] M. Lennon, G. Mercier, and L. Hubert-Moy, "Classification of hyperspectral images with nonlinear filtering and support vector machines," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2002, pp. 1670–1672.

[25] R. D. Phillips, C. E. Blinn, L. T. Watson, and R. H. Wynne, "An adaptive noise-filtering algorithm for AVIRIS data with implications for classification accuracy," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 9, pp. 3168–3179, Sep. 2009.

[26] I. Tomas, "Spatial postprocessing of spectrally classified landsat data," *Photogramm. Eng. Remote Sens.*, vol. 46, no. 9, pp. 1201–1206, 1980.

[27] X. Kang, S. Li, and J. A. Benediktsson, "Spectral-Spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.

[28] X. Huang and L. Zhang, "An adaptive mean-shift analysis approach for object extraction and classification from urban hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 12, pp. 4173–4185, Dec. 2008.

[29] X. Huang, L. Zhang, and P. Li, "Classification and extraction of spatial features in urban areas using high-resolution multispectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 2, pp. 260–264, Apr. 2007.

[30] 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.

[31] A. Plaza, P. Martínez, J. Plaza, and R. Pérez, "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, Mar. 2005.

[32] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.

[33] Q. Jackson and D. Landgrebe, "Adaptive bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.

[34] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736–740, Oct. 2010.

[35] G. Moser and S. B. Serpico, "Combining support vector machines and Markov random fields in an integrated framework for contextual image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 5, pp. 2734–2752, May 2013.

[36] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral-Spatial hyperspectral image segmentation using subspace multinomial logistic regression and Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 3, pp. 809–823, Mar. 2012.

[37] G. Camps-Valls, L. Gomez-Chova, J. Muñoz Maré, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, Jan. 2006.

[38] X. Huang and L. Zhang, "An SVM ensemble approach combining spectral, structural, semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 257–272, Jan. 2013.

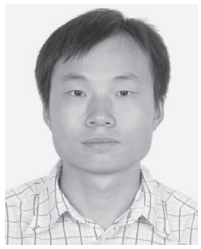
[39] X. Huang and L. Zhang, "Comparison of vector stacking, multi-SVMs fuzzy output, multi-SVMs voting methods for multiscale VHR urban mapping," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 2, pp. 261–265, Apr. 2010.

[40] X. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 3, pp. 383–394, Mar. 2008.
[41] Q. Gao, J. Ma, H. Zhang, X. Gao, and Y. Liu, "Stable orthogonal local discriminant embedding for linear dimensionality reduction," *IEEE Trans. Image Process.*, vol. 22, no. 7, pp. 2521–2531, Jul. 2013.
[42] K. Q. Weinberger and L. K. Saul, "An introduction to nonlinear dimensionality reduction by maximum variance unfolding," in *Proc. 21st Assoc. Adv. Artif. Intell.*, 2006, pp. 1683–1686.
[43] B. C. Kuo and D. A. Landgrebe, "A covariance estimator for small sample size classification problems and its application to feature extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 4, pp. 814–819, Apr. 2002.
[44] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.
[45] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 492–501, Mar. 2005.

Yicong Zhou (M'07) received the B.S. degree from Hunan University, Changsha, China, and the M.S. and Ph.D. degrees from Tufts University, Medford, MA, USA, all in electrical engineering.

He is currently an Assistant Professor with the Department of Computer and Information Science, University of Macau, Taipa, Macau, China. His research interests focus on multimedia security, image/signal processing, pattern recognition, and medical imaging.

Dr. Zhou is a member of the SPIE (International Society for Photo-Optical Instrumentations Engineers).



Jiangtao Peng received the B.S. and M.S. degrees from Hubei University, Wuhan, China, in 2005 and 2008, respectively, and the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2011.

He is currently a Lecturer with the Faculty of Mathematics and Statistics, Hubei University. His research interests include machine learning and hyperspectral image processing.

C. L. Philip Chen (S'88–M'88–SM'94–F'07) received the M.S. degree from the University of Michigan, Ann Arbor, MI, USA, in 1985, and the Ph.D. degree from Purdue University, West Lafayette, IN, USA, in 1988, all in electrical engineering.

After having worked in the U.S. for 23 years as a tenured Professor, Department Head, and Associate Dean in two different universities, he is currently the Dean of the Faculty of Science and Technology and a Chair Professor of the Department of Computer and Information Science, University of Macau, Taipa, Macau, China.

Dr. Chen is a Fellow of the American Association for the Advancement of Science and a Fellow of Hong Kong Institution of Engineers. Currently, he is the Junior Past President of the IEEE Systems, Man, and Cybernetics Society and the Editor-in-Chief of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS.