

Article

Dimensionality Reduction of Hyperspectral Image Using Spatial-Spectral Regularized Sparse Hypergraph Embedding

Hong Huang * , Meili Chen and Yule Duan

Key Laboratory of Optoelectronic Technology and Systems of the Education Ministry of China, Chongqing University, Chongqing 400044, China; chenmeili@cqu.edu.cn (M.C.); duanyule@cqu.edu.cn (Y.D.)

* Correspondence: hhuang@cqu.edu.cn; Tel.: +86-23-6510-6109

Received: 28 March 2019; Accepted: 29 April 2019; Published: 1 May 2019



Abstract: Many graph embedding methods are developed for dimensionality reduction (DR) of hyperspectral image (HSI), which only use spectral features to reflect a point-to-point intrinsic relation and ignore complex spatial-spectral structure in HSI. A new DR method termed spatial-spectral regularized sparse hypergraph embedding (SSRHE) is proposed for the HSI classification. SSRHE explores sparse coefficients to adaptively select neighbors for constructing the dual sparse hypergraph. Based on the spatial coherence property of HSI, a local spatial neighborhood scatter is computed to preserve local structure, and a total scatter is computed to represent the global structure of HSI. Then, an optimal discriminant projection is obtained by possessing better intraclass compactness and interclass separability, which is beneficial for classification. Experiments on Indian Pines and PaviaU hyperspectral datasets illustrated that SSRHE effectively develops a better classification performance compared with the traditional spectral DR algorithms.

Keywords: hyperspectral image; dimensionality reduction; spatial-spectral feature; hypergraph embedding; sparse representation

1. Introduction

With spectral sampling from visible to short-wave infrared region, hyperspectral image (HSI) can provide a spatial scene in hundreds of narrow contiguous spectral channels [1,2]. HSI data with high spectral resolution can provide fine spectral details for different ground objects, and they have been widely applied in many fields such as geological survey, environmental monitoring, precision agriculture, and mineral exploration [3,4]. Classification of each pixel in HSI plays a crucial role in these real applications. However, the high dimensional characteristic of HSI poses a huge challenge to the traditional classification methods, and the Hughes effect may occur if only limited training samples are available [5,6].

In general, dimensionality reduction (DR) is an effective way to reduce the volume of high-dimensional data with minimum loss of useful information by feature extraction or band selection, and it brings benefits for classification by achieving discriminating embedding features [7–9]. Many DR methods based on feature extraction have been proposed to reduce the dimension of high-dimensional data. Principal component analysis (PCA) [10] and linear discriminant analysis (LDA) [11] are the most popular subspace methods, but the two linear methods cannot discover the underlying manifold structure embedding in the original high-dimensional space. Many manifold learning-based DR methods are introduced to reveal nonlinear structure in high-dimensional data, such as locally linear embedding (LLE) [12], Laplacian eigenmap (LE) [13], isometric mapping (ISOMAP) [14], neighborhood preserving embedding (NPE) [15], and locality preserving projection (LPP) [16]. The above methods

can be unified under the graph embedding framework (GE), and the difference between them is how to define the similarity matrix of an intrinsic graph and the constraint matrix of a penalty graph [17–20]. On the basis of GE, some supervised DR methods are designed to exploit the prior knowledge of training samples for improving classification performance, such as marginal Fisher analysis (MFA) [21], local Fisher discriminant analysis (LFDA) [22], and regularized local discriminant embedding (RLDE) [23]. However, these direct graph-based DR methods only consider the pairwise relationship between data points, while HSI data usually possess complex relationships such as one sample versus multiple samples (different classes) or one class versus multiple samples. Therefore, the pairwise relation cannot discover complex relations in HSI, which limit the discriminability of embedding features for classification [24,25].

To explore the multiple adjacency relationships in high-dimensional data, hypergraph learning has been introduced to discover the complex geometric structure between HSI pixels [26,27]. In [28], discriminant hyper-Laplacian projection (DHLP) was proposed using the hypergraph Laplacian for exploring the high-order geometric relationship of samples. Semi-supervised hypergraph embedding (SHGE) learns the discriminant structure from both labeled and unlabeled data, and it reveals the complex relationships of HSI pixels by building a semi-supervised hypergraph [29]. For analyzing the intrinsic properties of HSI pixels, a hypergraph Laplacian sparse coding method was constituted to capture the similarity among data points within the same hyperedge [30]. In addition, the heterogeneous network is explored to measure the relatedness of heterogeneous objects. Pio et al. [31] introduced heterogeneous networks with an arbitrary structure to evaluate its performance for both clustering and classification tasks. Serafino et al. [32] proposed an ensemble learning approach to classify objects of different classes, which is based on the heterogeneous networks for extracting both correlation and autocorrelation that involve the observed objects.

The aforementioned methods are designed as spectral-based DR methods, in which the spatial relationship between a pixel and its spatial neighborhood is not taken into consideration for DR. Recent investigations show that incorporating spatial information into traditional spectral-based DR methods can further improve the performance of HSI classification [33–37]. Wu et al. presented a spatially adaptive model to extract the spectral and spatial-contextual information, which significantly enhances the land cover classification performance in both accuracy and computational efficiency [38]. Local pixel NPE (LPNPE) [23] and spatial consistency LLE [39] were proposed to reveal the local manifold structure in HSI data by using the distance between different spatial blocks instead Euclidean distance between pixels. As an extension of LPNPE, spatial and spectral information-based RLDE (SSRLDE) tries to maximize the ratio between local spatial-spectral data scatter and global spatial-spectral data scatter for enhancing the representation ability of embedding features [23]. The spatial-spectral coordination embedding (SSCE) method defines a spatial-spectral coordination distance for neighbor selection, and it can reduce the probability that heterogeneous objects are selected as nearest neighbors [40]. Discriminative spectral-spatial margin (DSSM) exploits spatial-spectral neighbors to obtain the low-dimensional embedding via preserving the local spatial-spectral relationship of HSI data [41]. These spatial-spectral DR methods have difficulty discovering the complex relationships in HSI data due to their pairwise nature.

Recently, the spatial information in HSI has been explored to construct a spatial-spectral hypergraph model [42–44]. Sun et al. proposed an adaptive hyperedge weight estimation scheme to preserve the prominent hyperedges, which is better for improving the classification accuracy [45]. Yuan et al. introduced a hypergraph embedding model for feature extraction, which can represent higher order relationships [46]. However, these spatial-spectral hypergraph methods are unsupervised and do not use prior information in HSI data, which is not conducive to extract discriminant features for enhancing the classification performance.

Motivated by the above limitations, a new hypergraph embedding method termed spatial-spectral joint regularized sparse hypergraph embedding (SSRHE) is proposed for DR of HSI data. SSRHE explores sparse coefficients and label information of pixels to adaptively select neighbors for constructing

a regularized sparse intraclass hypergraph and a regularized sparse interclass hypergraph, which can effectively represent the complex relationships in HSI data. Then, a local spatial neighborhood preserving scatter matrix and a total-scatter matrix are computed to preserve the neighborhood structure in spatial domain and the global structure in spectral domain, respectively. Finally, an optimal objective function is designed to extract spatial-spectral discriminant features, which not only preserves the local spatial structure of HSI data, but also compacts the samples belonging to the same class and separates the samples from different classes simultaneously. Therefore, embedding features achieve a good discriminative power for HSI classification.

The rest of this paper is organized as follows. In Section 2, some related works are briefly introduced. Section 3 gives a detailed description of the proposed SSRHE method. In Section 4, experimental results on two real HSI datasets are reported to demonstrate the effectiveness of SSRHE. Finally, Section 5 summarizes this paper and provides some suggestions for future work.

2. Related Works

In this section, we provide a brief review of GE framework and hypergraph model. For convenience, supposed a HSI dataset $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\} \in R^{D \times N}$, where D is the number of spectral bands and N is the number of pixels. The label of \mathbf{x}_i denotes $l_i \in \{1, 2, \dots, c\}$ and c is the number of land cover classes. The goal of dimensionality reduction is to map $(\mathbf{x}_i \in R^D) \rightarrow (\mathbf{y}_i \in R^d)$, where $d \ll D$. For the linear DR methods, \mathbf{Y} can be obtained by $\mathbf{Y} = \mathbf{P}^T \mathbf{X}$ with a projection matrix $\mathbf{P} \in R^{D \times d}$.

2.1. Graph Embedding

The graph embedding (GE) framework offers a unified view for understanding and explaining many popular DR algorithms such as PCA, LDA, ISOMAP, LLE, LE, NPE, and LPP. In GE, an intrinsic graph $\mathbf{G}^I(X, \mathbf{W}^I)$ represents a certain desired statistical or geometrical properties of data, and a penalty graph $\mathbf{G}^P(X, \mathbf{W}^P)$ describes some characteristics or relationships that should be avoided. \mathbf{W}^I and \mathbf{W}^P are the weight matrices of undirected graphs \mathbf{G}^I and \mathbf{G}^P . w_{ij}^I and w_{ij}^P describe the similarity and dissimilarity characteristics between vertices \mathbf{x}_i and \mathbf{x}_j in \mathbf{G}^I and \mathbf{G}^P , respectively.

The purpose of GE is to map each vertex of graph into a low-dimensional space that preserves similarities between the vertex pairs. The low-dimensional embedding can be obtained by solving the following objective function:

$$\min_{\mathbf{Y}^T \mathbf{H} \mathbf{Y} = \mathbf{C}} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \|\mathbf{y}_i - \mathbf{y}_j\|^2 w_{ij}^I = \min_{\mathbf{Y}^T \mathbf{H} \mathbf{Y} = \mathbf{C}} \text{tr}(\mathbf{Y}^T (\mathbf{D}^I - \mathbf{W}^I) \mathbf{Y}) = \min_{\mathbf{Y}^T \mathbf{H} \mathbf{Y} = \mathbf{C}} \text{tr}(\mathbf{Y}^T \mathbf{L}^I \mathbf{Y}) \quad (1)$$

where \mathbf{D}^I is a diagonal matrix with $\mathbf{D}_{ii}^I = \sum_j w_{ij}^I$, $\mathbf{L}^I = \mathbf{D}^I - \mathbf{W}^I$ is the Laplacian matrix of \mathbf{G}^I , \mathbf{C} is a constant, \mathbf{H} typically is a diagonal matrix for scale normalization, and it may be the Laplacian matrix of a penalty graph \mathbf{G}^P . That is, $\mathbf{H} = \mathbf{L}^P = \mathbf{D}^P - \mathbf{W}^P$, where $\mathbf{D}_{ii}^P = \sum_j w_{ij}^P$.

2.2. Hypergraph Model

The hypergraph can represent the complex relations between high-dimensional data, and every hyperedge connects multiple vertexes. A hypergraph $\mathbf{G} = (\mathbf{V}_H, \mathbf{E}_H, \mathbf{W}_H)$ can be constructed to formulate the relationships among data samples, where \mathbf{V}_H is a set of vertices, \mathbf{E}_H is a set of hyperedges, and \mathbf{W}_H is a diagonal matrix where each diagonal element denotes the weight of hyperedge. Each hyperedge $e_i \in \mathbf{E}_H$ is vested by a weight $w(e_i) \in \mathbf{W}_H$.

To represent the relationship of \mathbf{G} , an incidence matrix $\mathbf{H} = [H_{mn} : h(e_m, v_n)] \in R^{|\mathbf{E}_H| \times |\mathbf{V}_H|}$ is denoted as follows:

$$H_{mn} = h(e_m, v_n) = \begin{cases} 1, & \text{if } v_n \in e_m \\ 0, & \text{if } v_n \notin e_m \end{cases} \quad (2)$$

Furthermore, the degree of hyper-edge e_m and the degree of vertex \mathbf{v}_n can be defined as

$$d(e_m) = |e_m| = \sum_{\mathbf{v}_n \in \mathbf{V}_H} H_{mn} \quad (3)$$

$$d(\mathbf{v}_n) = \sum_{e_m \in \mathbf{E}_H} w(e_m) H_{mn} \quad (4)$$

There is an example to illustrate the hypergraph in Figure 1. A vertex is denoted as a circle (such as $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_5$). As shown in Figure 1a, the simple graph only holds two vertices per edge, which just describes a single one-to-one relationship (such as \mathbf{v}_1 and \mathbf{v}_2 , and \mathbf{v}_1 and \mathbf{v}_3). In Figure 1b, each hyper-edge is represented by a curve (such as a hypergraph e_1 is constructed by $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$), which represents complex multiple relations among pixels. That is, a hypergraph can describe the local neighborhood structure well and preserve the complex relationships within the neighborhood. This hypergraph consists of six vertices and four hyper-edges, and the corresponding incidence matrix is shown in Figure 1c. The incidence matrix intuitively represents affinity relationships between vertices and hyper-edges, the non-zero element in each row indicates that a hyper-edge is associated with the vertex, otherwise the vertex and the hyper-edge are not interrelated.

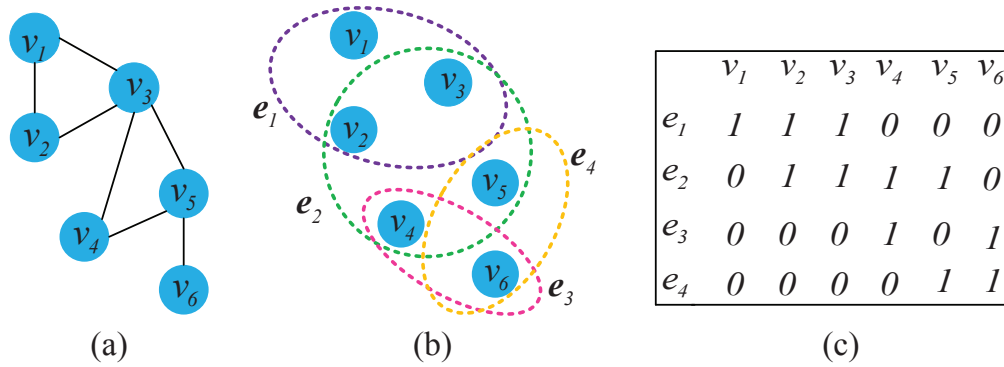


Figure 1. An example for hypergraph: (a) simple graph; (b) hypergraph; and (c) incidence matrix H .

However, the hypergraph model only represents complex higher-order relations between pixels in spectral domain, and it fails to consider spatial information in hyperspectral image that limits the discriminant ability of embedding features for land cover classification.

3. SSRHE

To reveal the complex structure in HSI, we propose a new hypergraph learning method called spatial-spectral joint regularized sparse hypergraph embedding (SSRHE) for dimensionality reduction of HSI data. At first, SSRHE constructs a regularized sparse intraclass hypergraph and a regularized sparse interclass hypergraph by exploring sparse representation (SR) and prior knowledge. After that, it exploits the spatial consistency and global structure of HSI by computing a local spatial neighborhood preserving scatter and a total scatter, which brings benefits for combining spatial structure and spectral information for DR. Finally, an optimal objective function is designed to learn a spatial-spectral discriminant projection by minimizing the regularized sparse intraclass hypergraph scatter and the local spatial neighborhood preserving scatter, while maximizing the regularized sparse interclass hypergraph scatter and the total scatter of samples simultaneously. The flowchart of the proposed SSRHE method is shown in Figure 2.

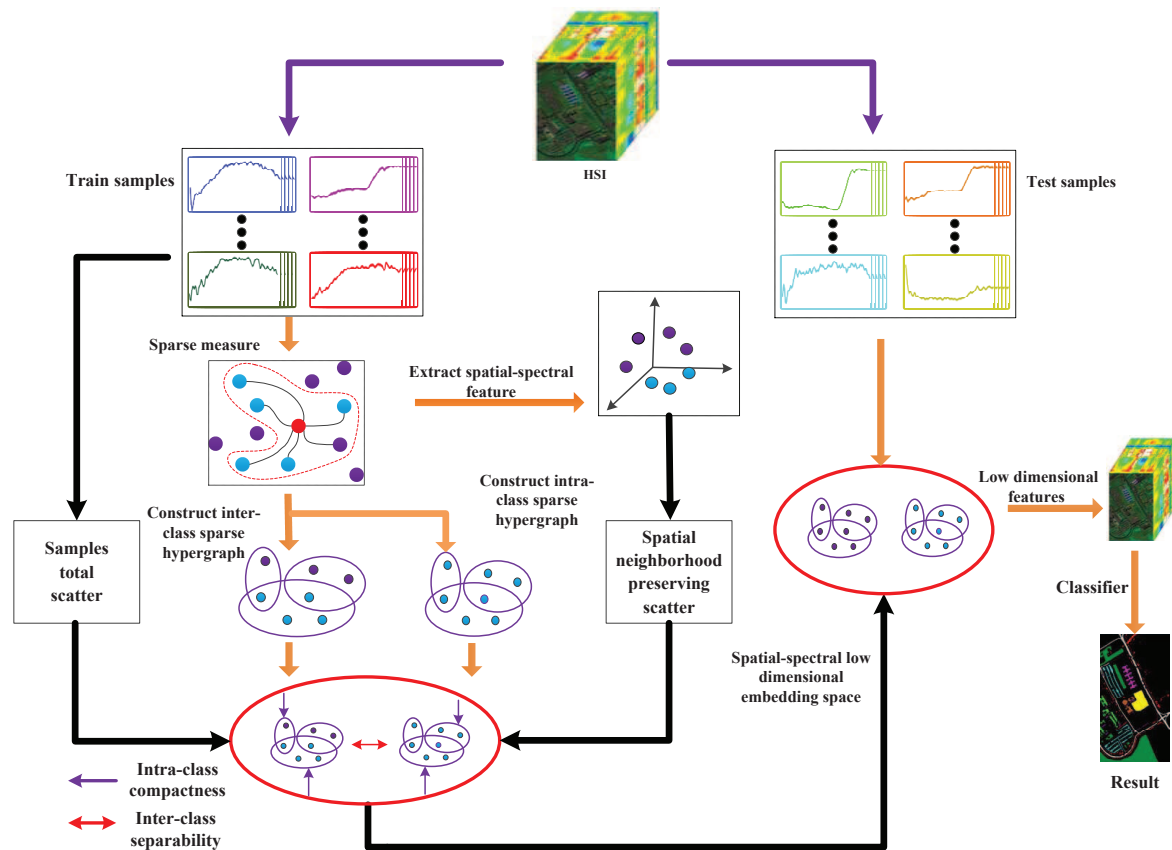


Figure 2. Flowchart of the proposed SSRHE method.

3.1. The Regularized Sparse Hypergraph Model

To discover the complex structure of HSI data, a hypergraph model is exploited to reveal the intrinsic relations between pixels. However, it remains difficult to choose a proper neighborhood size for constructing hypergraphs. Since sparse representation has natural discriminating power to adaptively reveal the inherent relationship of data [47–49], a sparse hypergraph model is designed based on SR theory.

Inspired by the observation that the most compact expression of a certain sample is generally given from similar samples, sparse coefficients are explored to find neighbors adaptively. Suppose $\mathbf{S} \in \mathbb{R}^{N \times N}$ is the sparse coefficient matrix of data samples. The sparse coefficients can be calculated as follows:

$$\min_{\mathbf{S}} \|\mathbf{s}_i\|_1 \quad s.t. \quad \|\mathbf{x}_i - \mathbf{X}\mathbf{s}_i\| \leq \varepsilon, \mathbf{s}_i > 0 \quad (5)$$

where ε denotes the sparse error, $\mathbf{s}_i = [\mathbf{s}_{i,1}, \mathbf{s}_{i,2}, \dots, \mathbf{s}_{i,i-1}, 0, \mathbf{s}_{i,i+1}, \dots, \mathbf{s}_{i,N}]^T$ represents the sparse coefficients of pixel \mathbf{x}_i , which can be optimized by using Alternating Direction Method of Multipliers (ADMM) framework [50,51]. With the sparse coefficient matrix \mathbf{S} , a hypergraph can be constructed with the criterion that nodes \mathbf{x}_i and \mathbf{x}_j are connected in a hyper-edge if sparse coefficient \mathbf{s}_{ij} is not equal to 0. Since sparse coefficients can reflect the similarity between data, non-zero correspondence coefficient indicates the correlation between pixels, a large value indicates a high similarity. Compared with Euclidean metric, sparse coefficients can more effectively select neighbors of HSI data.

According to sparse coefficients and label information of samples, we construct a within-class sparse hypergraph $\mathbf{G}^w(\mathbf{X}, \mathbf{E}_H^w, \mathbf{W}_H^w)$ and a between-class sparse hypergraph $\mathbf{G}^b(\mathbf{X}, \mathbf{E}_H^b, \mathbf{W}_H^b)$ to characterize the intrinsic structure of HSI data. \mathbf{X} is denoted as the vertex set, \mathbf{E}_H^w and \mathbf{E}_H^b are the sets of intraclass hyper-edge and interclass hyper-edge. \mathbf{W}_H^w and \mathbf{W}_H^b are the weight matrices of hyper-edges of \mathbf{G}^w and \mathbf{G}^b , respectively.

In \mathbf{G}^w , the intraclass hyper-edge e^w is formed by connecting pixel \mathbf{x}_i with its corresponding neighbors whose sparse coefficients are non-zero. The subedge weights of e^w and e^b can be defined as

$$w_{ij}^w = \begin{cases} \varphi * s_{ij}, & \text{if } s_{ij} \neq 0 \text{ and } l_i = l_j \\ 0, & \text{others} \end{cases} \quad (6)$$

$$w_{ij}^b = \begin{cases} s_{ij}, & \text{if } s_{ij} \neq 0 \text{ and } l_i \neq l_j \\ 0, & \text{others} \end{cases} \quad (7)$$

in which s_{ij} is the sparse coefficient between pixel \mathbf{x}_i and \mathbf{x}_j and parameter φ ($\varphi > 1$) is used to enhance the contribution of samples from the same class for improving the discriminant power. Figure 3 illustrates the construction of sparse hypergraph on a simple classification model. As shown in Figure 3, a simple graph considers only pairwise relation between two observed samples, while the sparse hypergraph through sparse representation can select neighbors of samples adaptively and represent complex multiple relations among HSI pixels.

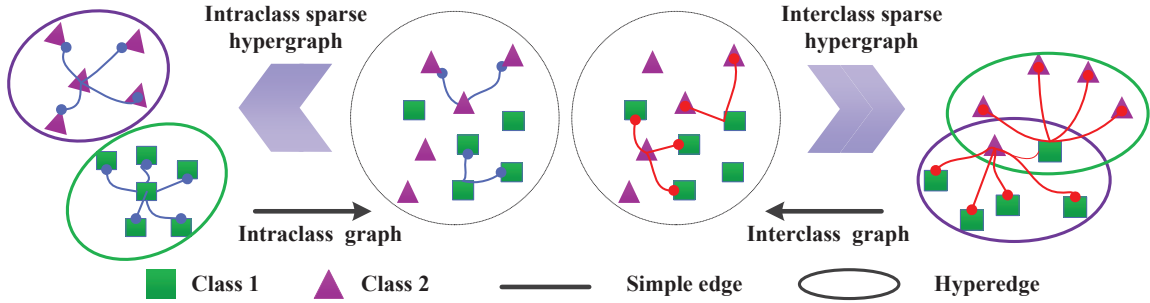


Figure 3. A simple two classification model to explain the construction of sparse hypergraph.

The weight matrices corresponding to hyper-edge regions are defined as

$$\omega_i^w = w(e_i^w) = \sum_{\mathbf{x}_j \in e_i^w} w_{ij}^w \quad (8)$$

$$\omega_i^b = w(e_i^b) = \sum_{\mathbf{x}_j \in e_i^b} w_{ij}^b \quad (9)$$

The incidence matrix $\mathbf{H}^w = [H_{ij}^w : h(e_j^w, \mathbf{x}_i)] \in R^{|\mathbf{E}^w| \times |\mathbf{X}|}$ of intraclass hypergraph is calculated by

$$H_{ij}^w = \begin{cases} \exp\left(-\frac{(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T}{2t^2}\right), & \text{if } \mathbf{x}_i, \mathbf{x}_j \in e_j^w \\ 0, & \text{others} \end{cases} \quad (10)$$

where $t = \frac{1}{\mathcal{N}(e_i^w)^2} \sum_j \sqrt{(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T}$ is the heat-kernel parameter, that is the mean value of pixels in one hyper-edge, $\mathcal{N}(e_i)$ denotes the number of vertices in each hyper-edge region.

According to \mathbf{H}^w and $w(e^w)$, the degree of vertex $\mathbf{x}_i \in \mathbf{X}$ and the degree of intraclass hyper-edge $e_i^w \in \mathbf{E}^w$ are computed by

$$\theta_i^w = d(\mathbf{x}_i, e_i^w) = \sum_j w(e_j^w) h(e_j^w, \mathbf{x}_i) = \sum_j w(e_j^w) H_{ij}^w \quad (11)$$

$$\vartheta_i^w = d(e_i^w) = \sum_j h(e_i^w, \mathbf{x}_j) = \sum_j H_{ij}^w \quad (12)$$

For the between-class hypergraph \mathbf{G}_H^b , the between-class incidence matrix $\mathbf{H}^b = [H_{ij}^b : h(e_j^b, \mathbf{x}_i)] \in R^{|\mathbf{E}^b| \times |\mathbf{X}|}$ is defined as

$$H_{ij}^b = \begin{cases} \exp(-\frac{(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T}{2t^2}), & \text{if } \mathbf{x}_i, \mathbf{x}_j \in e_j^b \\ 0, & \text{others} \end{cases} \quad (13)$$

Based on \mathbf{H}^b and $w(e^b)$, the degree of vertex \mathbf{x}_i and interclass hyper-edge e_i^b can be defined as

$$\theta_i^b = d(\mathbf{x}_i, e_j^b) = \sum_j w(e_j^b) h(e_j^b, \mathbf{x}_i) = \sum_j w(e_j^b) H_{ij}^b \quad (14)$$

$$\vartheta_i^b = d(e_i^b) = \sum_j h(e_i^b, \mathbf{x}_j) = \sum_j H_{ij}^b \quad (15)$$

In low-dimensional embedding space, the pixels from the same class should be as compact as possible for the intraclass sparse hypergraph, whereas pixels from different classes should be as separated as possible with the interclass sparse hypergraph. Therefore, the objective functions with intraclass and interclass sparse hypergraph constraint are constructed as

$$\begin{aligned} & \arg \min_P \frac{1}{2} \sum_{e_k^w \in E^w} \sum_{\mathbf{x}_i, \mathbf{x}_j \in e_k^w} \frac{w(e_k^w) h(e_k^w, \mathbf{x}_i) h(e_k^w, \mathbf{x}_j)}{d(e_k^w)} \|P^T \mathbf{x}_i - P^T \mathbf{x}_j\|^2 \\ &= \sum_{k=1}^n \sum_{i,j=1}^n \frac{w(e_k^w) H_{ik}^w H_{jk}^w}{\vartheta_k^w} \text{tr}(P^T \mathbf{x}_i \mathbf{x}_i^T P - P^T \mathbf{x}_j \mathbf{x}_j^T P) \\ &= \left\{ \sum_{k=1}^n \sum_{i=1}^n \frac{w(e_k^w) H_{ik}^w}{\vartheta_i^w} \text{tr}(P^T \mathbf{x}_i \mathbf{x}_i^T P) \sum_{j=1}^n \frac{H_{jk}^w \vartheta_j^w}{\vartheta_k^w} \right\} - \left\{ \sum_{k=1}^n \sum_{i,j=1}^n \frac{\text{tr}[P^T \mathbf{x}_j w(e_k^w) H_{ik}^w H_{jk}^w \mathbf{x}_i^T P]}{\vartheta_k^w} \right\} \\ &= \text{tr}(P^T X D_v^w X^T P) - \text{tr}[P^T X H^w W^w (D_e^w)^{-1} (H^w)^T X^T P] \\ &= \text{tr}\{P^T X [D_v^w - H^w W^w (D_e^w)^{-1} (H^w)^T] X^T P\} \\ &= \text{tr}(P^T X L^w X^T P) \end{aligned} \quad (16)$$

$$\begin{aligned} & \arg \max_P \frac{1}{2} \sum_{e_k^b \in E^b} \sum_{\mathbf{x}_i, \mathbf{x}_j \in e_k^b} \frac{w(e_k^b) h(e_k^b, \mathbf{x}_i) h(e_k^b, \mathbf{x}_j)}{d(e_k^b)} \|P^T \mathbf{x}_i - P^T \mathbf{x}_j\|^2 \\ &= \sum_{k=1}^n \sum_{i,j=1}^n \frac{w(e_k^b) H_{ik}^b H_{jk}^b}{\vartheta_k^b} \text{tr}(P^T \mathbf{x}_i \mathbf{x}_i^T P - P^T \mathbf{x}_j \mathbf{x}_j^T P) \\ &= \left\{ \sum_{k=1}^n \sum_{i=1}^n \frac{w(e_k^b) H_{ik}^b}{\vartheta_i^b} \text{tr}(P^T \mathbf{x}_i \mathbf{x}_i^T P) \sum_{j=1}^n \frac{H_{jk}^b \vartheta_j^b}{\vartheta_k^b} \right\} - \left\{ \sum_{k=1}^n \sum_{i,j=1}^n \frac{\text{tr}[P^T \mathbf{x}_j w(e_k^b) H_{ik}^b H_{jk}^b \mathbf{x}_i^T P]}{\vartheta_k^b} \right\} \\ &= \text{tr}(P^T X D_v^b X^T P) - \text{tr}[P^T X H^b W^b (D_e^b)^{-1} (H^b)^T X^T P] \\ &= \text{tr}\{P^T X [D_v^b - H^b W^b (D_e^b)^{-1} (H^b)^T] X^T P\} \\ &= \text{tr}(P^T X L^b X^T P) \end{aligned} \quad (17)$$

where $\mathbf{L}^w = \mathbf{D}_v^w - \mathbf{H}^w \mathbf{W}^w (\mathbf{D}_e^w)^{-1} (\mathbf{H}^w)^T$ is the intraclass sparse hypergraph Laplacian matrix and $\mathbf{L}^b = \mathbf{D}_v^b - \mathbf{H}^b \mathbf{W}^b (\mathbf{D}_e^b)^{-1} (\mathbf{H}^b)^T$ denotes the interclass sparse hypergraph Laplacian matrix. $\mathbf{D}_v^w, \mathbf{D}_v^b, \mathbf{D}_e^w, \mathbf{D}_e^b, \mathbf{W}^w, \mathbf{W}^b$ with respect to $\theta^w, \theta^b, \vartheta^w, \vartheta^b, w^w, w^b$ can be indicated as

$$\mathbf{D}_v^w = \text{diag}([\theta_1^w, \theta_2^w, \dots, \theta_{N-1}^w, \theta_N^w]) \quad (18)$$

$$\mathbf{D}_v^b = \text{diag}([\theta_1^b, \theta_2^b, \dots, \theta_{N-1}^b, \theta_N^b]) \quad (19)$$

$$\mathbf{D}_e^w = \text{diag}([\vartheta_1^w, \vartheta_2^w, \dots, \vartheta_{N-1}^w, \vartheta_N^w]) \quad (20)$$

$$\mathbf{D}_e^b = \text{diag}([\vartheta_1^b, \vartheta_2^b, \dots, \vartheta_{N-1}^b, \vartheta_N^b]) \quad (21)$$

$$\mathbf{W}^w = \text{diag}([\omega_1^w, \omega_2^w, \dots, \omega_{N-1}^w, \omega_N^w]) \quad (22)$$

$$\mathbf{W}^b = \text{diag}([\omega_1^b, \omega_2^b, \dots, \omega_{N-1}^b, \omega_N^b]) \quad (23)$$

Let $\mathbf{M}^b = \mathbf{X} \mathbf{L}^b \mathbf{X}^T$ represent the between-class scatter of \mathbf{G}^b . $\mathbf{M}^w = \mathbf{X} \mathbf{L}^w \mathbf{X}^T$ is the within-class scatter of \mathbf{G}^w . Then, the mapping matrix \mathbf{P} can be obtained through dealing with the following optimization problem:

$$\arg \max_{\mathbf{P}} \frac{\text{tr}(\mathbf{P}^T \mathbf{M}^b \mathbf{P})}{\text{tr}(\mathbf{P}^T \mathbf{M}^w \mathbf{P})} \quad (24)$$

In real applications, to avoid the singularity in the case of small samples, the above optimal objection can be further extended as

$$\arg \max_{\mathbf{P}} \frac{\text{tr}\{\mathbf{P}^T [(1 - \beta) \mathbf{M}^b + \beta \mathbf{X} \mathbf{X}^T] \mathbf{P}\}}{\text{tr}\{\mathbf{P}^T [(1 - \beta) \mathbf{M}^w + \beta (\text{diag}(\text{diag}(\mathbf{M}^w)))] \mathbf{P}\}} \quad (25)$$

where β is a tradeoff parameter. The regularization term $\mathbf{X} \mathbf{X}^T$ is the maximal data variance, which is used to ensure that the diversity of HSI pixels. The diagonal regularization \mathbf{M}^w is introduced for overcoming the singularity problem when the number of training samples is small. With regularization, Equation (25) becomes more stable to effectively reserve the useful discriminant information.

3.2. Spatial-Spectral Hypergraph Embedding

Due to the spatial consistency of HSI, the pixels in HSI are usually spatially related, which means the pixels within a small neighborhood usually possess the spatial distribution consistency of ground objects. Therefore, neighborhood pixels can be utilized to learn spatial-spectral combined features.

Suppose that pixel \mathbf{x}_i is the center pixel, and T is the spatial neighborhood window. (u_i, v_i) is the spatial coordinate of pixel \mathbf{x}_i in the image, and the spatial neighborhood set $\Omega(\mathbf{x}_i)$ with the window size T (positive odd) can be recorded as

$$\Omega(\mathbf{x}_i) = \left\{ \mathbf{x}_{im} : (u_m, v_m) \left| \begin{array}{l} u_m \in (u - z, u + z) \\ v_m \in (v - z, v + z) \end{array} \right. \right\} \quad (26)$$

where $z = (T - 1)/2$, $\mathbf{x}_{im} : (u_m, v_m)$ corresponds to the m -th pixel in the spatial neighborhood. $\Omega(\mathbf{x}_i)$ has a total of $T \times T$ pixels. Thus, the distance measure in the spatial neighborhood block can be defined as

$$a_i = \sum_{m=1}^{T^2} \|\mathbf{x}_i - \mathbf{x}_{im}\|^2 v_m \quad (27)$$

where $v_m = \exp \left\{ -\|\mathbf{x}_i - \mathbf{x}_{im}\|^2 / 2 \left(\frac{1}{T \times T} \sum_{m=1}^{T^2} \|\mathbf{x}_{im}\|^2 \right) \right\}$, which measures similarity between central pixel \mathbf{x}_i and its spectral-spatial neighbors. For all training samples in HSI data, the local spatial neighborhood preserving scatter matrix is calculated as follows:

$$\mathbf{S}^w = \sum_i a_i = \sum_i \sum_{m=1}^{T^2} \|\mathbf{x}_i - \mathbf{x}_{im}\|^2 v_m = \sum_{i=1}^N \sum_{m=1}^{T^2} (\mathbf{x}_i - \mathbf{x}_{im})(\mathbf{x}_i - \mathbf{x}_{im})^T v_m \quad (28)$$

To preserve the global structure of samples, a total scatter matrix is defined as

$$\mathbf{S}^b = \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \quad (29)$$

where $\bar{\mathbf{x}}$ is the mean of all samples.

To extract low-dimensional spatial-spectral joint features, a objective function should be designed to preserve the local neighborhood as well as compact the samples with interclass hypergraph and separate the samples with interclass hypergraph simultaneously. Therefore, Equations (25), (28), and (29) are transformed into the following optimization function:

$$\begin{cases} \arg \max_{\mathbf{P}} \text{tr} \left\{ \alpha \mathbf{P}^T [(1 - \beta) \mathbf{M}^b + \beta \mathbf{X} \mathbf{X}^T] \mathbf{P} + (1 - \alpha) \mathbf{P}^T \mathbf{S}^b \mathbf{P} \right\} \\ \arg \min_{\mathbf{P}} \text{tr} \left\{ \alpha \mathbf{P}^T [(1 - \beta) \mathbf{M}^w + \beta \text{diag}(\text{diag}(\mathbf{M}^w))] \mathbf{P} + (1 - \alpha) \mathbf{P}^T \mathbf{S}^w \mathbf{P} \right\} \end{cases} \quad (30)$$

The above optimal function can be further simplified as

$$\arg \max_{\mathbf{P}} \frac{\text{tr} \left\{ \alpha \mathbf{P}^T \left[(1 - \beta) \mathbf{M}^b + \beta \mathbf{X} \mathbf{X}^T \right] \mathbf{P} + (1 - \alpha) \mathbf{P}^T \mathbf{S}^b \mathbf{P} \right\}}{\text{tr} \left\{ \alpha \mathbf{P}^T \left[(1 - \beta) \mathbf{M}^w + \beta \text{diag}(\text{diag}(\mathbf{M}^w)) \right] \mathbf{P} + (1 - \alpha) \mathbf{P}^T \mathbf{S}^w \mathbf{P} \right\}} \quad (31)$$

According to Lagrange multiplier method, Equation (31) can be solved by

$$\left\{ \alpha \left[(1 - \beta) \mathbf{M}^b + \beta \mathbf{X} \mathbf{X}^T \right] + (1 - \alpha) \mathbf{S}^b \right\} \mathbf{P} = \lambda \left\{ \alpha \left[(1 - \beta) \mathbf{M}^w + \beta \text{diag}(\text{diag}(\mathbf{M}^w)) \right] + (1 - \alpha) \mathbf{S}^w \right\} \mathbf{P} \quad (32)$$

where λ is an eigenvalue set. With the d largest eigenvalues corresponding eigenvector, the optimal projection matrix can be denoted as $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_d]$. In the low-dimensional embedded space, the spatial-spectral embedding of test sample \mathbf{x}_t can be given as follows:

$$\mathbf{y}_t = \mathbf{P}^T \mathbf{x}_t \quad (33)$$

In summary, SSRHE compacts the samples from the same class and spatial neighborhood while separates interclass samples, and the embedding features possess stronger discriminative power that ensures good classification performance. The steps of the proposed SSRHE method are shown in Algorithm 1.

Algorithm 1 SSRHE.

Input: HSI dataset $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in R^{D \times N}$, corresponding class label set \mathbf{L} , tradeoff parameters

α, β , weighted coefficient φ , spatial neighborhood size T , reduced dimensionality d .

- 1: According to ADMM, compute sparse coefficient matrix \mathbf{S} ;
- 2: Compute the weights of intraclass and interclass hyperedge by Equations (8) and (9);
- 3: Compute the intraclass hyper-Laplacian matrix: $\mathbf{L}^w = \mathbf{D}_v^w - \mathbf{H}^w \mathbf{W}^w (\mathbf{D}_e^w)^{-1} (\mathbf{H}^w)^T$;
- 4: Compute the interclass hyper-Laplacian matrix: $\mathbf{L}^b = \mathbf{D}_v^b - \mathbf{H}^b \mathbf{W}^b (\mathbf{D}_e^b)^{-1} (\mathbf{H}^b)^T$;
- 5: Compute the local spatial neighborhood preserving scatter matrix:
- 6: $\mathbf{S}^w = \sum_i a_i = \sum_i \sum_{m=1}^{T^2} \|\mathbf{x}_i - \mathbf{x}_{im}\|^2 v_m = \sum_{i=1}^N \sum_{m=1}^{T^2} (\mathbf{x}_i - \mathbf{x}_{im})(\mathbf{x}_i - \mathbf{x}_{im})^T v_m$
- 7: Compute the extended total scatter matrix: $\mathbf{S}^b = \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$;
- 8: Solve the generalized eigenvalue problem:
- 9: $\left\{ \alpha \left[(1 - \beta) \mathbf{M}^b + \beta \mathbf{X} \mathbf{X}^T \right] + (1 - \alpha) \mathbf{S}^b \right\} \mathbf{P} = \lambda \left\{ \alpha \left[(1 - \beta) \mathbf{M}^w + \beta \text{diag}(\text{diag}(\mathbf{M}^w)) \right] + (1 - \alpha) \mathbf{S}^w \right\} \mathbf{P}$
- 10: Obtain the optimal projection matrix with largest eigenvalues corresponding eigenvectors:
- 11: $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_d]$

Output:

- 12: Low-dimensional spatial-spectral embedding features of test samples: $\mathbf{Y} = \mathbf{P}^T \mathbf{X}$
-

In this paper, we adopt big O notation to analyze the computational complexity of SSRHE. Spatial neighborhood size and the number of sparse iterations are denoted as T and t , respectively. The sparse coefficient matrix \mathbf{S} is computed with the cost of $O(Nt)$. The weight of intraclass hyperedge and the weight of interclass hyperedge both take $O(N^2)$. The intraclass incidence matrix \mathbf{H}^w and the interclass incidence matrix \mathbf{H}^b are both calculated with $O(N^2)$. The cost of diagonal matrices \mathbf{W}^w , \mathbf{W}^b , \mathbf{D}_v^w , \mathbf{D}_v^b , \mathbf{D}_e^w , and \mathbf{D}_e^b are $O(8N)$. The intraclass hyper-Laplacian matrix \mathbf{L}^w and the interclass hyper-Laplacian matrix \mathbf{L}^b both cost $O(N^3)$. The local spatial neighborhood preserving scatter matrix

\mathbf{S}^w takes $O(DNT^2)$. The extended total scatter matrix \mathbf{S}^b costs $O(DN)$. The costs of matrix \mathbf{M}^w and \mathbf{M}^b are both $O(DN^2)$. \mathbf{XX}^T costs $O(DN)$. It takes $O(D^3)$ to deal with the generalized problem of Equation (32). The total time complexity of SSRHE is $O(D^3 + DN^2 + N^3)$.

4. Experimental Results and Discussion

Some experiments were performed on two real HSI datasets to verify the effectiveness of the proposed method, and several state-of-the-art DR methods were compared with SSRHE in the experiments.

4.1. HSI Datasets

Indian Pines Dataset: This hyperspectral image is a scene of northwest Indiana collected by an airborne imaging spectrometer sensor in 1992. It consists of 145×145 pixels and 220 spectral bands that cover the wavelength from 400 to 2450 nm. After removing the noise and water absorption bands, 200 spectral bands remained for use in the experiments. This image contains total 16 land cover types. The dataset in false color and its corresponding ground truth map are shown in Figure 4.

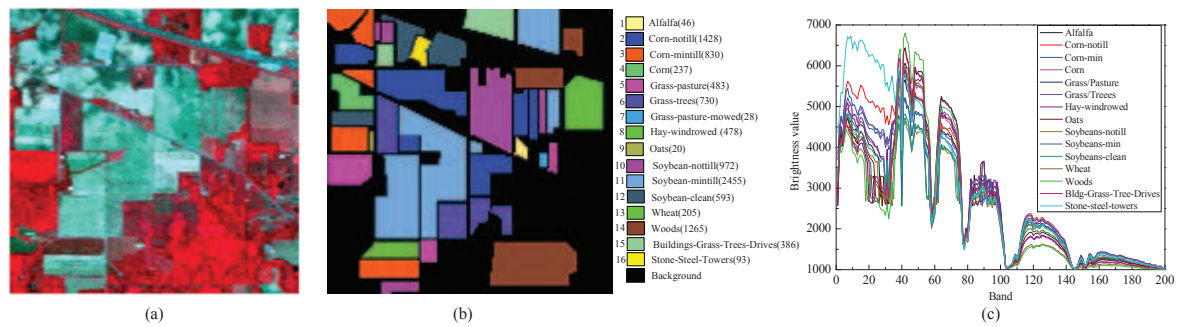


Figure 4. Indian Pines hyperspectral image: (a) HSI in false color; (b) ground truth; and (c) spectral curve. (Note that the number of samples for each class is shown in brackets.)

University of Pavia (PaviaU) Dataset: This hyperspectral image covering the area of Pavia University in northern Italy was collected by ROSIS sensor in 2002. The spatial size of this image is 640×340 pixels and the number of spectral band is 115. Due to the atmospheric affection, 12 bands were discarded and the remaining 103 spectral bands were used for experiment. The false color and ground truth map for this scene are shown in Figure 5.

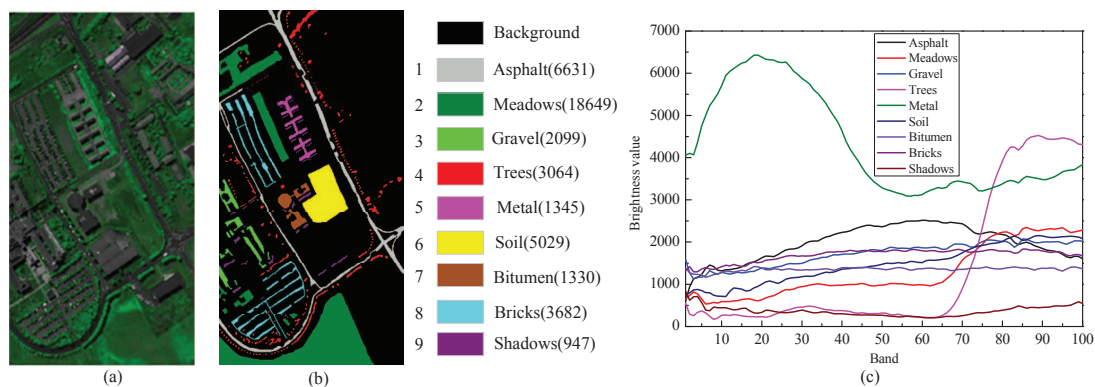


Figure 5. PaviaU hyperspectral image: (a) HSI in false color; (b) ground truth; and (c) spectral curve. (Note that the number of samples for each class is shown in brackets.)

4.2. Experimental Setup

In experiments, each HSI dataset was randomly divided into training and test sets. The training samples were utilized to learn a feature extracted model to obtain their low-dimensional embedding

features. After that, the classifier was applied to obtain the class labels of test samples. The overall classification accuracy (OA), average classification accuracy of each class (AA), and kappa coefficient (KC) were employed to evaluate classification results.

To demonstrate the effectiveness of the proposed method, we compared SSRHE with four spectral-based approaches, LPP, LDA, MFA, and RLDE; two spatial-spectral combined approaches, LPNPE and SSCE; and a hypergraph method, DHLHP. For all approaches, we tuned the parameters by cross-validation to achieve good results. In LPNPE, the size of spatial-spectral window was set to 15 for the PaviaU data and 11 for the Indian Pines data. In SSCE, the window scale and the local neighbor size were both set to 5 on both datasets. For RLDE and MFA, the intraclass neighbor size was set to 3 and 8, while the interclass neighbor size was set to 5 and 60, respectively. The number of nearest neighbor in DHLHP was selected as 9.

The nearest neighbor (1-NN) classifier was applied for classification. Each experiment was randomly repeated ten times. All experiments were performed on a personal computer with i5-4590 central processing unit, 8-G memory, and 64-bit Windows 7 using MATLAB 2014a.

4.3. Parameters Selection

For the proposed SSRHE algorithm, there are four important parameters: spatial neighborhood size T , weighted coefficient φ , tradeoff parameters α and β . To select the optimal parameters, we performed some parameters selection experiments on the Indian Pines and PaviaU datasets. In each experiment, we randomly selected five samples from each class as training set and the remaining samples were used for testing.

To explore the effect of tradeoff parameters α and β , we turned them with a set of $\{0, 0.01, 0.05, 0.1, 0.2, \dots, 0.9, 1\}$. Figures 6 and 7 show the classification accuracies with respect to α and β on two datasets. As shown in Figure 6, an increasing α produced subtle change of OAs under a fixed β . The OAs first varied slightly and then declined significantly with the increase of α on Indian Pines dataset, since a too large α led to the algorithm losing the property of spatial-domain in HSI and the obtained features might fail to represent the intrinsic structure of hyperspectral image. For PaviaU dataset, the OAs first increased with increasing α and β , and then tended to decline. To balance the influence of spectral and spatial information for classification, we set parameters $\alpha = 0.3$ and $\beta = 0.7$ for the Indian Pines dataset, and $\alpha = 0.2$ and $\beta = 0.5$ for PaviaU dataset based on the results presented in Figures 6 and 7.

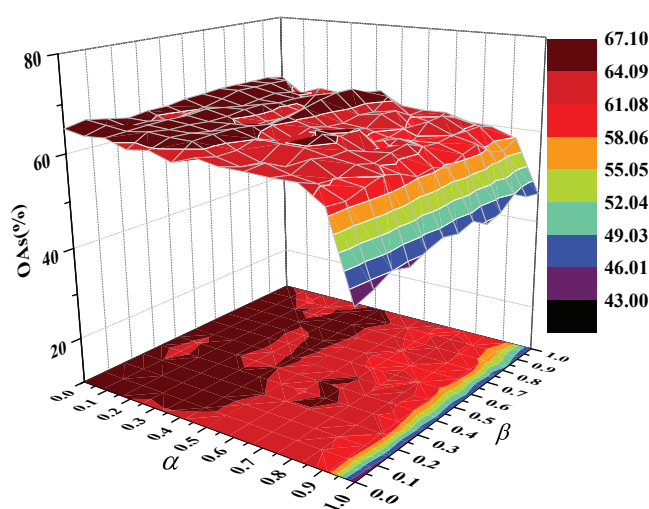


Figure 6. The OAs of SSRHE with different α and β on Indian Pines dataset.

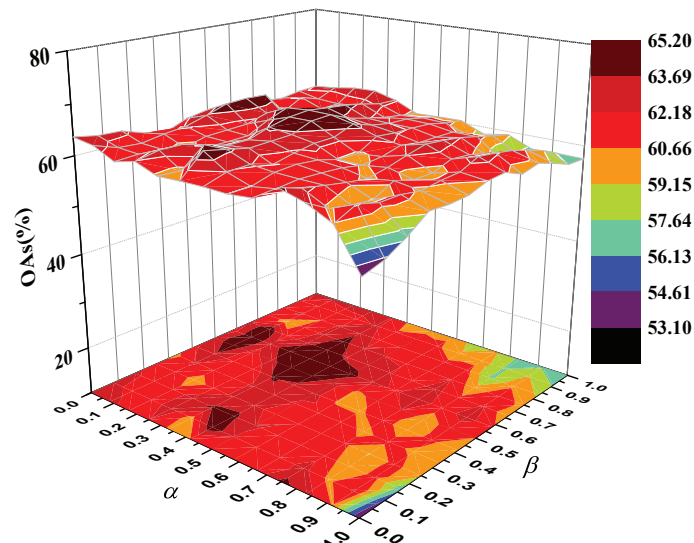


Figure 7. The OAs of SSRHE with different α and β on PaviaU dataset.

The parameters T and φ have a significant influence on the discrimination of SSRHE. The former is defined to determine the size of spatial neighborhood and the latter is exploited to adjust the compactness of the interclass neighbors. To analyze the influence of two parameters for SSRHE, we performed two experiments with respect to T and φ , and the corresponding experimental results on two HSI datasets are shown in Figure 8. In Figure 8a, it is clear that the OAs first quickly increased with the growth of T on two dataset because a larger spatial neighborhood was beneficial to preserving more useful spatial-domain information in HSI. However, if the value of T was too large, it might bring pixels from other classes into training model, as well as greatly increase computational complexity. According to the results in Figure 8a, the size of spatial neighborhood T was set to 7 for Indian Pines dataset and 15 for PaviaU dataset. As shown in Figure 8b, with the increasing of φ , the OAs improved until reaching a peak value due to further enhancing the compactness of intraclass samples in the embedding space. While φ kept increasing, the OAs declined since too large φ weakened the ability to extract between-class features. Therefore, we set $\varphi = 50$ for the Indian Pines dataset and $\varphi = 60$ for the PaviaU dataset.

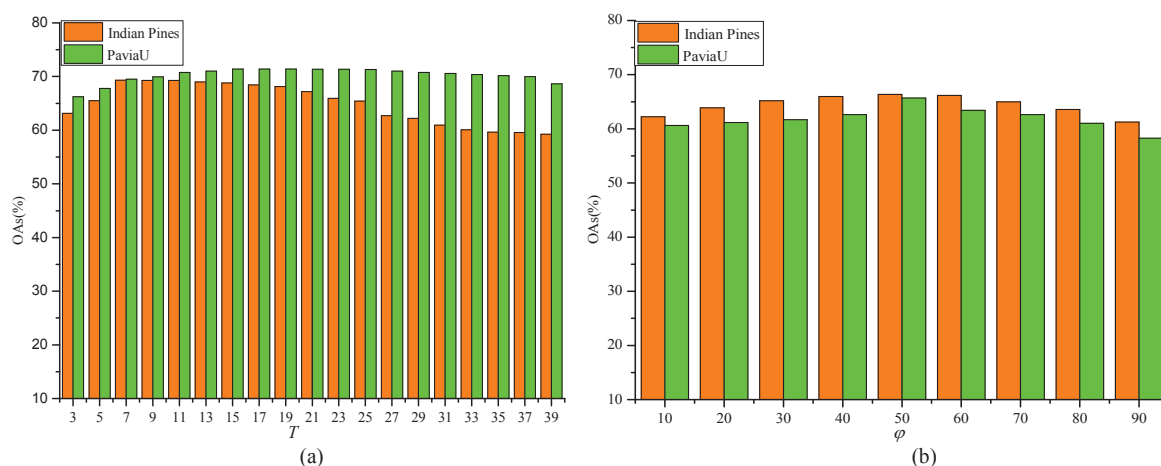


Figure 8. OAs with respect to parameters T and φ on different datasets: (a) the OAs of T ; and (b) the OAs of φ .

4.4. Investigation of Embedding Dimension

To explore how the embedding dimension d affects the performance of the proposed SSRHE method, thirty samples were randomly selected from each class for training, and d was tuned from 5 to 50 with an interval of 5. Figure 9 shows the classification accuracy of SSRHE and other DR methods with different embedding dimensions on Indian Pines and PaviaU datasets. According to Figure 9, the OAs firstly rapidly rose and then tended to stabilize with an increasing d , because embedding features with high dimension contained more useful information for training, while too large dimension might lead to information saturation. The classification accuracies of the most DR methods were better than the RAW method since these methods can remove the redundant information in HSI. The proposed SSRHE achieved the best classification performance in comparison with other methods on two HSI datasets. The reason is that SSRHE discovered the complex relationships between samples by hypergraph model and fused the useful spatial information for extracting discriminant features for classification.

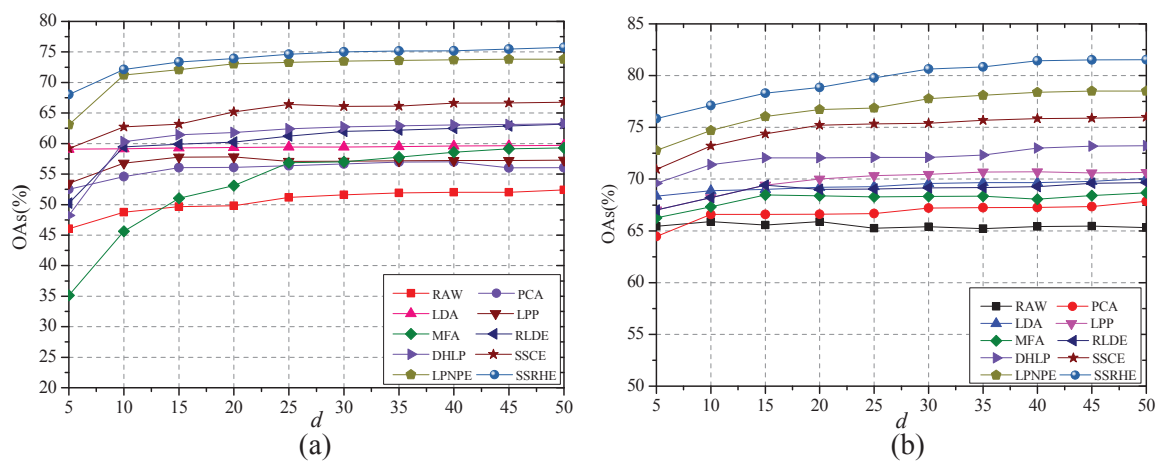


Figure 9. OAs with different embedding d : (a) Indian Pines dataset; and (b) Pavia U dataset.

4.5. Investigation of Classification Performance

To evaluate the performance of the proposed method under different training conditions, we selected n_i ($n_i = 5, 20, 50, 100, 200$) samples per class for training, and 1-NN classifier was utilized to classify the remaining samples. Each experiment was repeated ten times, the averaged OA with standard deviations (STDs) and kappa coefficient (KC) of all DR methods on Indian Pines and PaviaU datasets are shown in Tables 1 and 2.

Table 1. Classification results using different methods with different classifiers for the Indian pines dataset. (OA \pm STDs (%) (KC)).

Method	5	20	50	100	200
RAW	43.6 \pm 2.8 (0.372)	54.9 \pm 1.7 (0.495)	60.1 \pm 1.4 (0.552)	63.6 \pm 0.9 (0.588)	66.9 \pm 0.6 (0.622)
PCA	43.4 \pm 2.7 (0.370)	54.9 \pm 1.6 (0.495)	60.2 \pm 1.2 (0.553)	63.9 \pm 0.8 (0.591)	67.0 \pm 0.6 (0.622)
LDA	32.5 \pm 4.8 (0.253)	51.6 \pm 1.9 (0.459)	64.4 \pm 1.2 (0.599)	71.0 \pm 0.5 (0.672)	74.4 \pm 0.7 (0.706)
LPP	43.6 \pm 3.7 (0.372)	54.5 \pm 1.8 (0.491)	59.7 \pm 1.2 (0.546)	62.7 \pm 1.0 (0.578)	65.8 \pm 0.5 (0.609)
MFA	44.1 \pm 4.0 (0.377)	57.1 \pm 1.6 (0.520)	66.8 \pm 1.9 (0.625)	70.8 \pm 1.1 (0.669)	72.0 \pm 1.0 (0.680)
RLDE	41.7 \pm 1.1 (0.622)	60.9 \pm 1.5 (0.561)	69.8 \pm 1.4 (0.659)	74.6 \pm 0.7 (0.711)	78.4 \pm 0.6 (0.751)
DHLP	44.1 \pm 3.8 (0.377)	57.2 \pm 2.1 (0.522)	68.9 \pm 1.2 (0.649)	73.8 \pm 0.8 (0.702)	77.6 \pm 0.7 (0.741)
SSCE	30.2 \pm 4.5 (0.230)	69.7 \pm 1.0 (0.658)	76.3 \pm 0.9 (0.730)	79.1 \pm 0.5 (0.760)	82.9 \pm 0.6 (0.801)
LPNPE	60.2 \pm 3.5 (0.594)	74.0 \pm 1.4 (0.706)	79.3 \pm 0.7 (0.759)	81.6 \pm 0.6 (0.791)	84.2 \pm 0.6 (0.817)
SSRHE	65.6 \pm 2.3 (0.615)	74.8 \pm 1.2 (0.711)	80.0 \pm 1.0 (0.765)	82.9 \pm 1.0 (0.803)	86.7 \pm 1.0 (0.829)

Note that the best results of a row are marked in bold.

Table 2. Classification results using different methods with different classifiers for the PaviaU dataset. (OA \pm STDs (%) (KC)).

Method	5	20	50	100	200
RAW	60.5 \pm 4.2 (0.512)	66.4 \pm 2.4 (0.583)	73.5 \pm 1.6 (0.663)	76.4 \pm 0.8 (0.698)	78.8 \pm 0.8 (0.724)
PCA	60.5 \pm 4.2 (0.512)	66.5 \pm 2.2 (0.583)	73.4 \pm 1.6 (0.662)	76.4 \pm 0.8 (0.697)	78.7 \pm 0.6 (0.724)
LDA	46.7 \pm 6.4 (0.351)	59.6 \pm 1.8 (0.495)	73.5 \pm 1.4 (0.662)	78.9 \pm 0.9 (0.727)	83.4 \pm 0.6 (0.782)
LPP	47.0 \pm 5.6 (0.354)	59.3 \pm 2.6 (0.500)	72.8 \pm 2.3 (0.654)	78.3 \pm 1.3 (0.722)	82.2 \pm 1.2 (0.768)
MFA	64.5 \pm 4.3 (0.555)	69.2 \pm 4.5 (0.613)	76.4 \pm 2.0 (0.699)	78.1 \pm 2.4 (0.715)	79.1 \pm 2.2 (0.730)
RLDE	64.4 \pm 3.2 (0.555)	74.6 \pm 2.7 (0.677)	77.9 \pm 2.2 (0.718)	82.1 \pm 1.0 (0.770)	84.8 \pm 1.0 (0.802)
DHLP	56.8 \pm 8.0 (0.471)	62.2 \pm 3.6 (0.530)	70.8 \pm 2.1 (0.629)	77.5 \pm 2.7 (0.711)	80.2 \pm 1.5 (0.742)
SSCE	42.3 \pm 5.3 (0.309)	63.3 \pm 2.9 (0.543)	75.8 \pm 1.7 (0.692)	82.7 \pm 1.2 (0.804)	87.0 \pm 0.8 (0.828)
LPNPE	68.0 \pm 4.2 (0.606)	80.0 \pm 2.2 (0.747)	86.3 \pm 1.3 (0.822)	87.9 \pm 0.9 (0.842)	89.9 \pm 0.6 (0.877)
SSRHE	71.6 \pm 2.7 (0.646)	82.6 \pm 2.3 (0.776)	87.5 \pm 1.1 (0.837)	90.0 \pm 1.5 (0.882)	92.2 \pm 0.2 (0.908)

Note that the best results of a row are marked in bold.

As shown in Tables 1 and 2, for all methods, the OA and KC greatly improved and the STDs decreased with the increasing of training samples, because a large number of training samples usually provide more useful information for training. The spatial-spectral combined methods, SSCE, LPNPE and SSRHE, produced better classification results than the spectral-based methods, because they exploit the spatial and spectral information in HSI to improve the representation ability of extracted features. Among these spectral-based methods, DHLP presented better classification performance under most training conditions, since it applies the hypergraph learning model to discover the intrinsic complex relationship between samples. Compared with all the other methods, SSRHE had better discriminant performance with different proportions of training samples on two HSI datasets, especially with small size of training set. SSRHE utilizes the hypergraph framework to discover the complex multivariate relationships between interclass samples and intraclass samples, and computes two spatial neighborhood scatters to reveal the spatial correlation between each pixels in HSI, which further enhances the discriminating power of low-dimensional features.

To explore the classification accuracy of the proposed method for each class, we randomly selected a percentage of samples from each class as the training set while the remaining samples were used for testing. In experiments, we set the training percent to 3% for the Indian Pines dataset and 5% for PaviaU dataset. Tables 3 and 4 show the classification results of different DR methods on Indian Pines and PaviaU datasets, and the corresponding classification maps are displayed in Figures 10 and 11.

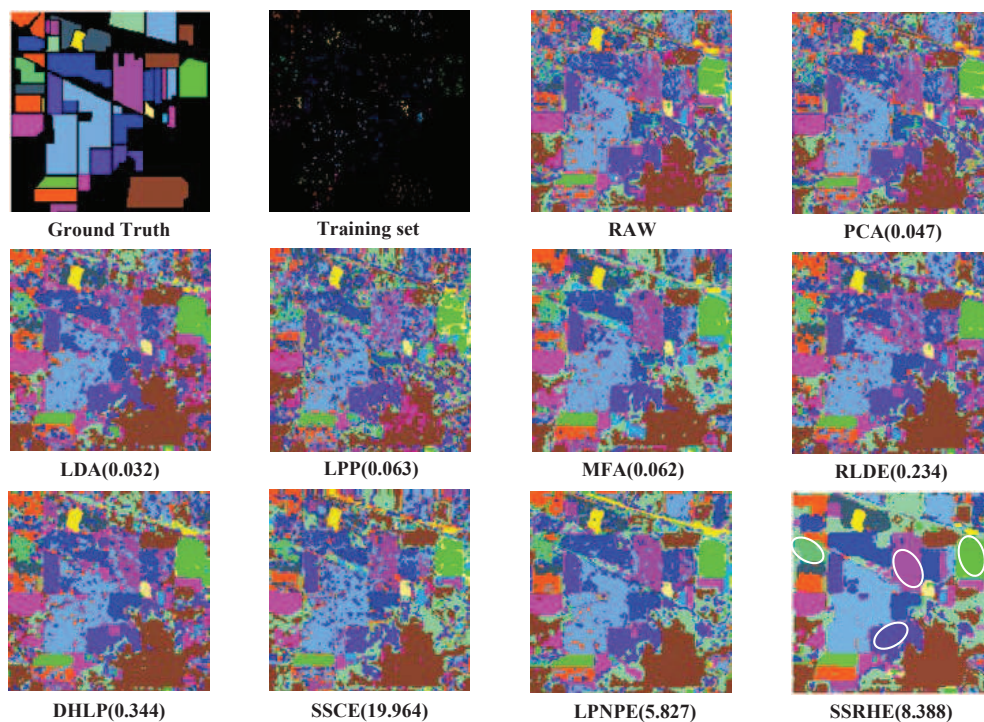
Table 3. Classification results of different classifiers on the Indian Pines dataset.

Class	Train	Test	RAW	PCA	LDA	LPP	MFA	RLDE	DHLP	SSCE	LPNPE	SSRHE
1	10	36	41.67	41.67	77.78	36.11	50.00	61.11	63.89	55.56	77.78	94.44
2	143	1285	53.39	52.45	64.12	51.05	56.03	72.14	66.69	60.31	80.78	88.17
3	83	747	57.30	55.29	57.70	47.12	50.07	61.58	61.58	63.45	74.30	80.46
4	24	213	41.78	44.60	52.58	43.19	21.60	58.69	59.15	51.17	77.00	84.04
5	48	435	78.85	78.62	89.43	77.93	78.85	86.90	87.82	80.46	91.72	96.55
6	73	657	90.26	89.50	95.74	91.02	94.67	95.28	95.89	94.52	96.04	97.02
7	10	18	77.78	88.89	100	88.89	77.78	94.44	94.44	100	94.44	100
8	48	430	95.58	95.58	99.53	93.95	93.26	99.30	99.77	92.56	99.53	98.60
9	10	10	70.00	70.00	60.00	50.00	70.00	80.00	90.00	90.00	100	80.00
10	97	875	61.03	60.46	60.91	57.49	42.06	68.91	63.20	72.11	82.74	83.89
11	246	2209	69.76	69.85	71.89	69.62	58.85	79.36	79.22	74.02	85.92	89.50
12	59	534	39.33	37.45	65.36	32.02	47.38	67.42	62.73	50.56	87.83	83.71
13	21	184	88.04	88.04	97.83	88.04	94.57	97.28	98.37	94.57	98.91	100
14	127	1138	94.02	93.94	94.11	92.88	90.69	96.66	95.61	93.94	96.10	95.52
15	39	347	31.12	30.55	54.18	25.07	42.65	40.92	48.41	55.04	71.47	83.57
16	10	83	91.57	91.57	90.36	85.54	85.54	90.36	92.77	84.34	92.77	97.59
OA (%)			68.33	67.88	74.44	65.91	64.03	78.27	77.00	74.06	87.65	89.78
AA (%)			67.59	68.03	76.97	64.37	65.87	78.15	78.72	75.79	88.02	90.88
KC			0.638	0.633	0.707	0.609	0.589	0.751	0.736	0.704	0.858	0.884

Table 4. Classification results of different classifiers on the PaviaU dataset.

Class	Train	Test	RAW	PCA	LDA	LPP	MFA	RLDE	DHLP	SSCE	LPNPE	SSRHE
1	332	6299	85.62	85.62	87.68	87.82	82.76	90.19	63.89	89.73	90.20	91.19
2	933	17,716	94.65	94.57	94.88	94.76	93.90	97.73	66.69	96.70	97.53	98.12
3	105	1994	65.15	64.64	63.34	67.00	61.84	74.77	61.58	72.37	77.28	78.6
4	154	2910	77.22	77.36	81.79	79.01	77.02	84.13	59.15	84.78	87.83	89.26
5	68	1277	98.83	98.83	98.84	99.30	99.77	99.53	87.82	99.37	99.77	99.77
6	252	4777	60.26	60.32	65.17	65.47	69.72	70.36	95.89	73.86	89.68	85.22
7	67	1263	75.30	75.30	66.67	75.69	71.26	80.36	94.4	88.60	86.06	90.18
8	185	3497	80.27	80.27	74.39	81.42	77.36	84.79	99.77	82.85	84.68	79.33
9	48	899	100	100	99.44	100	99.67	100	90.00	99.78	99.89	100
OA (%)			84.92	84.88	85.40	86.27	84.73	89.70	78.00	89.60	91.30	92.59
AA (%)			81.92	81.88	81.47	83.38	81.48	86.87	77.72	87.56	89.53	90.55
KC			0.797	0.796	0.804	0.815	0.796	0.861	0.736	0.861	0.883	0.902

According to Tables 3 and 4, SSRHE obtained better classification results than other methods in most classes, and it achieved the best OA, AA and KC on two HSI datasets. This indicates that SSRHE possessed stronger ability to reveal the intrinsic complex geometric relations between samples and extract more useful the spatial-spectral combined features for classification. As shown in Figures 10 and 11, the spatial-spectral methods generally produced smoother classification maps than traditional spectral-based methods, which demonstrates that the spatial information in HSI is effective for improving classification performance. Moreover, the classification maps of SSRHE possessed fewer misclassified pixels, especially in the areas *Grass/Pasture*, *Wheat*, and *Stone-steel towers* for Indian Pines dataset and *Asphalt*, *Meadows*, and *Shadows* for PaviaU dataset. In terms of running time, SSRHE cost more time than traditional spectral-based methods because it needs to build models in both spatial domain and spectral domain of HSI data. Compared with other spectral-spatial algorithms, the running time of SSRHE did not increase significantly and was far less than SSCE. Thus, SSRHE was more effective than other DR methods to weaken the influence of noisy points and develop the discrimination power of embedding features.

**Figure 10.** Classification maps of different DR methods with 1-NN on the Indian Pines dataset. Note that the time cost of corresponding DR algorithms is marked in bracket.

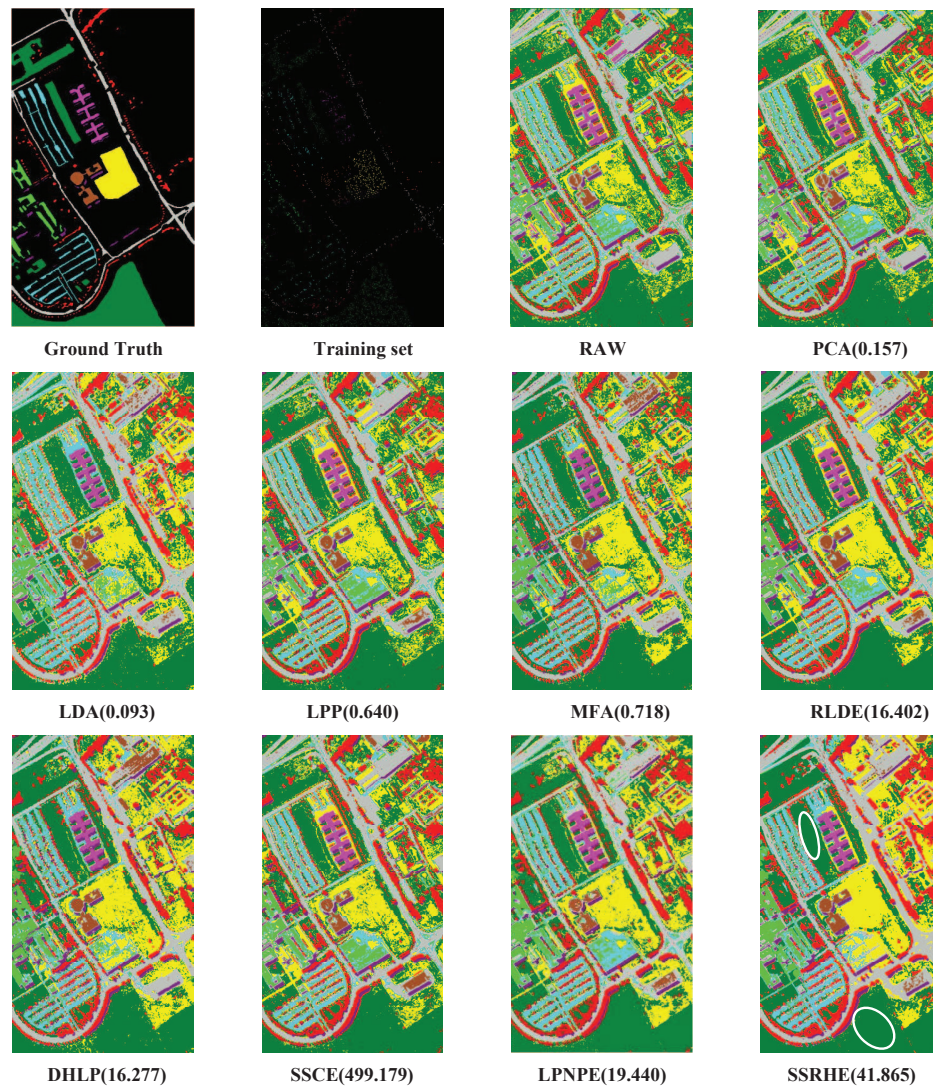


Figure 11. Classification maps of different DR methods with 1-NN on the PaviaU dataset. Note that the time cost of corresponding DR algorithms is marked in bracket.

5. Conclusions

In this paper, a new dimensionality reduction method SSRHE is proposed based on sparse hypergraph learning and spatial-spectral information for HSI. SSRHE computes two spatial neighborhood scatters to reveal the spatial correlation between each pixels in HSI, and it also constructs two discriminant hypergraphs by sparse representation to discover the complex multivariate relationships between interclass samples and intraclass samples, respectively. Based on the spatial neighborhood scatters and the Laplacian scatters of hypergraphs, a spatial-spectral combined objective function is designed to obtain an optimal projection matrix mapping original HSI data to low-dimensional embedding space where the intrinsic spatial and spectral properties are well preserved. Some experiments were conducted on Indian Pines and PaviaU datasets to demonstrate that the proposed method possesses better performance compared with some existing state-of-the-art DR methods. In the future, we will optimize the proposed algorithm to reduce the running time and extend it to other related fields such as multi-spectral images and very-high resolution images, and the sparse hypergraph model in this paper can be used in some application domains with high-dimensional data such as face images, gene expression and radiomics features. Furthermore, It would also be interesting to study integrating hypergraph learning and heterogeneous information networks when exploring high-dimensional data with multiple adjacency relationships.

Author Contributions: H.H. was primarily responsible for mathematical modeling and manuscript writing. M.C. contributed to the experimental design and experimental analysis. Y.D. provided important suggestions for improving the paper.

Funding: This work was supported in part by The Basic and Frontier Research Programmes of Chongqing under Grants cstc2018jcyjAX0093 and cstc2018jcyjAX0633, the Chongqing University Postgraduates Innovation Project under Grants CYB18048 and CYS18035, and the National Science Foundation of China under Grant 41371338.

Acknowledgments: The authors would like to thank the anonymous reviewers and associate editor for their valuable comments and suggestions to improve the quality of the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liao, D.; Qian, Y.; Tang, Y.Y. Constrained manifold learning for hyperspectral imagery visualization. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *1*, 1213–1226. [[CrossRef](#)]
2. Yuan, Q.Q.; Zhang, Q.; Li, J.; Li, J.; Zhang, L.P. Hyperspectral image denoising employing a spatial-spectral deep residual convolutional neural network. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 1205–1218. [[CrossRef](#)]
3. Liu, Q.; Sun, Y.; Hang, R.; Song, H. Spatial-spectral locality-constrained low-rank representation with semi-supervised hypergraph learning for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 4171–4182. [[CrossRef](#)]
4. Zhang, X.R.; Gao, Z.Y.; Jiao, L.C.; Zhou, H.Y. Multifeature hyperspectral image classification with local and nonlocal spatial information via markov random field in semantic space. *IEEE Trans. Geosci. Remote Sens.* **2016**, *56*, 1409–1424. [[CrossRef](#)]
5. Xu, Y.H.; Zhang, L.P.; Du, B.; Zhang, F. Spectral-spatial unified networks for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 5893–5909. [[CrossRef](#)]
6. Kumar, B.; Dikshit, O. Spectral contextual classification of hyperspectral imagery with probabilistic relaxation labeling. *IEEE Trans. Cybern.* **2017**, *47*, 4380–4391. [[CrossRef](#)] [[PubMed](#)]
7. Hang, R.; Liu, Q.; Sun, Y.; Yuan, X.; Pei, H.; Plaza, J.; Plaza, A. Robust matrix discriminative analysis for feature extraction from hyperspectral images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *10*, 2002–2011. [[CrossRef](#)]
8. Luo, F.L.; Huang, H.; Ma, Z.Z.; Liu, J.M. Semi-supervised sparse manifold discriminative analysis for feature extraction of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6197–6221. [[CrossRef](#)]
9. Xu, J.; Yang, G.; Yin, Y.F.; Man, H.; He, H.B. Sparse representation based classification with structure preserving dimension reduction. *Cogn. Comput.* **2014**, *6*, 608–621. [[CrossRef](#)]
10. Zabalza, J.; Ren, J.C.; Yang, M.Q.; Zhang, Y.; Wang, J.; Marshall, S.; Han, J.W. Novel folded-PCA for improved feature extraction and data reduction with hyperspectral imaging and SAR in remote sensing. *ISPRS J. Photogramm. Remote Sens.* **2014**, *93*, 112–122. [[CrossRef](#)]
11. Li, W.; Prasad, S.; Fowler, J.E. Noise-adjusted subspace discriminant analysis for hyperspectral imagery classification. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1374–1378. [[CrossRef](#)]
12. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [[CrossRef](#)]
13. Chen, C.; Zhang, L.; Bu, J.; Wang, C.; Chen, W. Constrained laplacian eigenmap for dimensionality reduction. *Neurocomputing* **2010**, *73*, 951–958. [[CrossRef](#)]
14. Bachmann, C.M.; Ainsworth, T.L.; Fusina, R.A. Exploiting manifold geometry in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 441–454. [[CrossRef](#)]
15. He, X.F.; Cai, D.; Yan, S.C.; Zhang, H.J. Neighborhood preserving embedding. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV’05), Beijing, China, 17–21 October 2005; Volume 2, pp. 1208–1213.
16. Zheng, Z.L.; Yang, F.; Tan, W.; Jia, J.; Yang, J. Gabor feature-based face recognition using supervised locality preserving projection. *Signal Process.* **2007**, *87*, 2473–2483. [[CrossRef](#)]
17. Huang, H.; Yang, M. Dimensionality reduction of hyperspectral images with sparse discriminant embedding. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5160–5169. [[CrossRef](#)]
18. Jiang, X.; Song, X.; Zhang, Y.; Jiang, J.; Gao, J.; Cai, Z. Laplacian regularized spatial-aware collaborative graph for discriminant analysis of hyperspectral imagery. *Remote Sens.* **2019**, *11*, 29. [[CrossRef](#)]

19. Hang, R.; Liu, Q. Dimensionality reduction of hyperspectral image using spatial regularized local graph discriminant embedding. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3262–3271. [[CrossRef](#)]
20. Feng, F.B.; Li, W.; Du, Q.; Zhang, B. Dimensionality reduction of hyperspectral image with graph-based discriminant analysis considering spectral similarity. *Remote Sens.* **2017**, *9*, 323–336. [[CrossRef](#)]
21. Luo, F.L.; Huang, H.; Duan, Y.L.; Liu, J.M.; Liao, Y.H. Local geometric structure feature for dimensionality reduction of hyperspectral imagery. *Remote Sens.* **2017**, *9*, 790. [[CrossRef](#)]
22. Sugiyama, M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *J. Mach. Learn. Res.* **2007**, *8*, 1027–1061.
23. Zhou, Y.C.; Peng, J.T.; Philip Chen, L.C. Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1082–1095. [[CrossRef](#)]
24. Huang, Y.; Liu, Q.; Zhang, S.; Metaxas, D. Image retrieval via probabilistic hypergraph ranking. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; pp. 3367–3383.
25. Liu, Q.; Sun, Y.; Wang, C.; Liu, T.; Tao, D. Elastic net hypergraph learning for image clustering and semi-supervised classification. *IEEE Trans. Image Process.* **2017**, *26*, 452–463. [[CrossRef](#)] [[PubMed](#)]
26. Yu, J.; Tao, D.; Wang, M. Adaptive hypergraph learning and its application in image classification. *IEEE Trans. Image Process.* **2012**, *21*, 3262–3272.
27. Wang, W.H.; Qian, Y.T.; Tang, Y.Y. Hypergraph-regularized sparse NMF for hyperspectral unmixing. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 681–694. [[CrossRef](#)]
28. Huang, S.; Yang, D.; Ge, Y.X.; Zhang, X.H. Discriminant hyper-Laplacian projection and its scalable extension for dimensionality reduction. *Neurocomputing* **2016**, *173*, 145–153. [[CrossRef](#)]
29. Du, W.B.; Qiang, W.W.; Lv, M.; Hou, Q.L.; Zhen, L.; Jing, L. Semi-supervised dimension reduction based on hypergraph embedding for hyperspectral images. *Int. J. Remote Sens.* **2018**, *39*, 1696–1712. [[CrossRef](#)]
30. Gao, S.H.; Tsang, I.W.H.; Chia, L.T. Laplacian sparse coding, hypergraph Laplacian sparse coding, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 92–104. [[CrossRef](#)] [[PubMed](#)]
31. Pio, G.; Serafino, F.; Malerba, D.; Ceci, M. Multi-type clustering and classification from heterogeneous networks. *Inf. Sci.* **2018**, *425*, 107–126. [[CrossRef](#)]
32. Serafino, F.; Pio, G.; Ceci, M. Ensemble learning for multi-type classification in heterogeneous networks. *IEEE Trans. Knowl. Data Eng.* **2018**, *30*, 2326–2339. [[CrossRef](#)]
33. Liao, W.Z.; Mura, M.D.; Chanussot, J.; Pizurica, A. Fusion of spectral and spatial information for classification of hyperspectral remote sensed imagery by local graph. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *9*, 583–594. [[CrossRef](#)]
34. Zhao, J.; Zhong, Y.F.; Shu, H.; Zhang, L.P. High-resolution image classification integrating spectral-spatial-location cues by conditional random fields. *IEEE Trans. Image Process.* **2016**, *25*, 4033–4045. [[CrossRef](#)]
35. Cao, J.; Wang, B. Embedding learning on spectral-spatial graph for semi-supervised hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1805–1809. [[CrossRef](#)]
36. Zhao, W.; Du, S. Spectral-spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4544–4554. [[CrossRef](#)]
37. Li, Y.; Zhang, H.; Shen, Q. Spectral-spatial classification of hyperspectral imagery with 3D convolutional neural network. *Remote Sens.* **2017**, *9*, 67. [[CrossRef](#)]
38. Wu, Z.B.; Shi, L.L.; Li, J.; Wang, Q.C.; Sun, L.; Wei, Z.H.; Plaza, J.; Plaza, A.J. GPU parallel implementation of spatially adaptive hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 1131–1143. [[CrossRef](#)]
39. Mohan, A.; Sapion, G.; Bosch, E. Spatially coherent nonlinear dimensionality reduction and segmentation of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2007**, *4*, 206–210. [[CrossRef](#)]
40. Huang, H.; Zheng, X.L. Hyperspectral image land cover classification algorithm based on spatial-spectral coordination embedding. *Acta Geod. Cartogr. Sin.* **2016**, *4*, 964–972.
41. Feng, Z.X.; Yang, S.Y.; Wang, S.G.; Jiao, L.C. Discriminative spectral-spatial margin-based semisupervised dimensionality reduction of hyperspectral data. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 224–228. [[CrossRef](#)]
42. Ji, R.R.; Gao, Y.; Hong, R.C.; Liu, Q.; Tao, D.C.; Li, X.L. Spectral-spatial constraint hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 1811–1824.

43. Luo, F.L.; Du, B.; Zhang, L.P.; Zhang, L.F.; Tao, D.C. Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image. *IEEE Trans. Cybern.* **2019**, *49*, 2406–2419. [[CrossRef](#)]
44. Tang, Y.Y.; Lu, Y.; Yuan, H. Hyperspectral image classification based on three-dimensional scattering wavelet transform. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2467–2480. [[CrossRef](#)]
45. Sun, Y.B.; Wang, S.J.; Liu, Q.S.; Hang, R.L.; Liu, G.C. Hypergraph embedding for spatial-spectral joint feature extraction in hyperspectral images. *Remote Sens.* **2017**, *9*, 506–520.
46. Yuan, H.L.; Tang, Y.Y. Learning with hypergraph for hyperspectral image feature extraction. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1695–1699. [[CrossRef](#)]
47. Sun, Y.; Hang, R.; Liu, Q.; Zhu, F.; Pei, H. Graph-rRegularized low rank representation for aerosol optical depth retrieval. *Int. J. Remote Sens.* **2016**, *37*, 5749–5762. [[CrossRef](#)]
48. Liu, J.; Xiao, Z.; Chen, Y.; Yang, J. Spatial-Spectral graph regularized kernel sparse representation for hyperspectral image classification. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 258. [[CrossRef](#)]
49. Zhang, S.Q.; Li, J.; Li, H.C.; Deng, C.Z.; Antonio, P. Spectral-spatial weighted sparse regression for hyperspectral image unmixing. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3265–3276. [[CrossRef](#)]
50. Sun, W.W.; Zhang, L.F.; Zhang, L.P.; Yenming, M.L. A dissimilarity-weighted sparse self-representation method for band selection in hyperspectral imagery classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2016**, *9*, 4374–4388. [[CrossRef](#)]
51. Luo, M.N.; Zhang, L.L.; Liu, J.; Zheng, Q.H. Distributed extreme learning machine with alternating direction method of multiplier. *Neurocomputing* **2017**, *261*, 164–170. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).