Article

Spatial-Spectral Multiple Manifold Discriminant Analysis for Dimensionality Reduction of Hyperspectral Imagery

Guangyao Shi, Hong Huang ^{*} , Jiamin Liu, Zhengying Li and Lihua Wang

Key Laboratory on Opto-Electronic Technique and Systems, Ministry of Education, Chongqing University, Chongqing 400044, China; shiguangyao@cqu.edu.cn (G.S.); liujm@cqu.edu.cn (J.L.); zhengying_li@cqu.edu.cn (Z.L.); 20170802019t@cqu.edu.cn (L.W.)

^{*} Correspondence: hhuang@cqu.edu.cn

Received: 14 September 2019; Accepted: 15 October 2019; Published: 18 October 2019

Abstract: Hyperspectral images (HSI) possess abundant spectral bands and rich spatial information, which can be utilized to discriminate different types of land cover. However, the high dimensional characteristics of spatial-spectral information commonly cause the Hughes phenomena. Traditional feature learning methods can reduce the dimensionality of HSI data and preserve the useful intrinsic information but they ignore the multi-manifold structure in hyperspectral image. In this paper, a novel dimensionality reduction (DR) method called spatial-spectral multiple manifold discriminant analysis (SSMMDA) was proposed for HSI classification. At first, several subsets are obtained from HSI data according to the prior label information. Then, a spectral-domain intramanifold graph is constructed for each submanifold to preserve the local neighborhood structure, a spatial-domain intramanifold scatter matrix and a spatial-domain intermanifold scatter matrix are constructed for each sub-manifold to characterize the within-manifold compactness and the between-manifold separability, respectively. Finally, a spatial-spectral combined objective function is designed for each submanifold to obtain an optimal projection and the discriminative features on different submanifolds are fused to improve the classification performance of HSI data. SSMMDA can explore spatial-spectral combined information and reveal the intrinsic multi-manifold structure in HSI. Experiments on three public HSI data sets demonstrate that the proposed SSMMDA method can achieve better classification accuracies in comparison with many state-of-the-art methods.

Keywords: hyperspectral image; dimensionality reduction; spatial-spectral information; multi-manifold structure; discriminative features

1. Introduction

Hyperspectral image (HSI) is acquired by different imaging spectrometer sensors (e.g., EO-1 Hyperion, HyMap and AVIRIS), which possesses abundant spectral information about ground objects in hundreds of spectral bands [1–3]. Due to the advancement of imaging spectrometer technology, the spectral resolution of hyperspectral sensors has been improved significantly, which can provide richer spectral information to differentiate different ground objects [4,5]. However, the sensors with higher resolution produce a very large volume of data, it will render the traditional image processing algorithms designed for multispectral imagery ineffective [6,7]. In particular, the high dimensionality of HSI data brings about the “curse of dimensionality” problem, that is, under a fixed, small number of training samples, the classification accuracy of HSI data decreases when the dimensionality of HSI data increases [8–10]. Besides, the spectral bands are highly correlated and some spectral bands may not carry discriminant information in a specific application. The reason is that the sample point in HSI

often exhibits similar characteristics to electromagnetic waves in adjacent bands, which will produce a lot of useless information and restrict the classification performance of HSI. Therefore, to achieve an excellent classification performance, it is an urgent task to perform dimensionality reduction (DR) for HSI data while preserving the intrinsic valuable information.

Serving as a good tool for data mining, manifold learning has been the main focus of DR [11–13]. Recently, many manifold learning algorithms have been introduced to discover the intrinsic structure in high-dimensional data, such methods include local linear embedding (LLE) [14], Laplacian eigenmaps (LE) [15] and isometric feature mapping (ISOMAP) [16]. LLE assumes that the data is linear over a small locality, it preserves the local linear structure of data in low-dimensional space [17]. LE builds a graph incorporating neighborhood information of data and computes a low-dimensional representation by optimally preserving local neighborhood information [18]. ISOMAP characterizes the data distribution by geodesic distances instead of Euclidean distances, it seeks a lower-dimensional embedding which maintains geodesic distances between all points [19]. However, due to their nonlinear characteristic, they suffer from the problem of out-of-sample and cannot process the unknown samples [20–22]. To address this issue, many linear manifold learning methods were designed to obtain explicit feature mappings, which can map unknown samples into low-dimensional space [23,24]. The representative methods include neighborhood preserving embedding (NPE) [25], locality preserving projection (LPP) [26] and local scaling cut (LSC) [27]. NPE aims at preserving the local manifold structure which is constructed by the k -nearest neighbors [28]. LPP seeks an optimal linear projection while preserving the local geometry structure in original data [29]. LSC uses the spectral information to explore the intrinsic manifold structure of HSI data by constructing a pairwise dissimilarity matrix [30]. Although the motivations of the methods are different, their common objective is to derive a lower-dimensional representation and facilitate the subsequent classification task. In order to unify these methods, a graph embedding (GE) framework was proposed to describe many existing DR techniques [31]. Based on this framework, local geometric structure Fisher analysis (LGSFA) was proposed by compacting the homogeneous data while separating the heterogeneous data [32]. However, the above manifold learning methods assume that the observed data is located on a single manifold and they cannot discover the multi-manifold structure embedded in HSI. Therefore, the single manifold methods have a limited discriminating power to distinguish different land covers and their classification performance will be reduced rapidly when applied to more complex scenes.

In practical applications, the observed data can be divided into many different subsets and each subset resides on a low-dimensional submanifold [33]. To reveal the multi-manifold structure in high-dimensional data, some multiple manifold learning methods were explored for feature learning. Hettiarachchi et al. [34] proposed a multiple manifold locally linear embedding (MM-LLE) algorithm which adopts a supervised LLE form of neighborhood selection in learning individual manifolds and it uses a manifold-manifold distance (MMD) as a measure to find the optimum embedded dimension. Jiang et al. [35] proposed a coupled discriminant multi-manifold analysis (CDMMA) algorithm which explores the neighborhood information as well as local geometric structure of the multi-manifold space spanned by samples and it exploits the discriminant information by minimizing the intramanifold distance and maximizing the intermanifold distance simultaneously. Chu et al. [36] proposed a multi-feature multi-manifold learning (M^3L) method, which learns multiple discriminative feature subspaces by maximizing the multi-feature manifold margins of different classes and it recognizes the probe subjects with a multi-feature manifold-manifold distance. Shi et al. [37] proposed a supervised multi-manifold learning (SMML) algorithm, which extracts multi-manifold features through maximizing the between-class Laplacian graph and projects samples from different classes into respective submanifolds. Although the multiple manifold learning methods can enhance the classification performance of high-dimensional data, they ignore the spatial consistency property of HSI and do not utilize the abundant spatial information.

Spatial information has been proven to be useful for improving the classification performance of HSI [38,39]. Recently, many DR methods have been proposed based on the spatial distribution of

hyperspectral data and they can be divided into two main types: spatial filtering methods and spatial DR methods [40–42]. Spatial filtering methods focus on how to utilize spatial homogeneous regions to smooth the pixel-wise classification map, while spatial DR methods incorporate spatial information into the process of DR by modeling the spatial neighboring correlations [43,44]. Although spatial filtering techniques can improve the classification accuracy of HSI data, they are often used as the pre-processing techniques to process the whole image [45]. In view of this, spatial DR methods have been widely studied for HSI classification [46,47]. Zhou et al. [48] proposed a spatial-domain local pixel NPE (LPNPE) method, which seeks a linear projection by minimizing the local pixel neighborhood preserving scatter and maximizing the total scatter of data. Feng et al. [49] developed a discriminate spectral-spatial margins (DSSM) method, which reveals the local neighborhood information of HSI data and explores the global structure of labeled and unlabeled data via low rank representation. Ramanarayan et al. [50] designed a neighboring pixel local scaling cut (NPLSC) method, which combines the unlabeled spectral weight with the local graph cut-based segmentation strategy on the homogeneous HSI data to maintain the label consistency in the spatial domain. The spatial-spectral combined DR methods have improved the classification performance significantly but they do not consider the multi-manifold structure of HSI data to further enhance the classification performance.

To explore the multiple manifold structure and spatial information in HSI, we proposed a novel spatial-spectral multi-manifold DR method, called spatial-spectral multiple manifold discriminant analysis (SSMMDA) for HSI classification. Differently from the methods of spatial filtering, SSMMDA incorporates spatial information into the DR process for feature learning. The main contributions of SSMMDA can be summarized as following. (1) According to the prior label information of training samples, we divide the samples data into several different subsets, each subset is treated as a submanifold. (2) Based on the graph embedding theory, an intramanifold graph is constructed for each submanifold to preserve the local neighborhood structure in spectral domain and an intramanifold scatter matrix and an intermanifold scatter matrix are constructed for each submanifold to characterize the within-manifold compactness and the between-manifold separability in spatial domain. (3) A spatial-spectral combined objective function is designed for each submanifold and each submanifold can obtain a discriminant projection matrix by maximizing the spatial-spectral intermanifold scatter matrix and minimizing the spatial-spectral intramanifold scatter matrix simultaneously. (4) Embedding features for each submanifold can be obtained via the corresponding projection matrix and then different embedding features are fused to enhance the classification performance.

The remainder of the paper is structured as follows. Section II briefly reviews some related works including GE and LSC. Section III details our proposed SSMMDA method. Section IV reports the experimental results. Section V presents some analysis and discussion on the experiments. Finally, Section VI provides our concluding remarks and gives recommendations for future work.

2. Related Works

Suppose that a HSI data set $\mathbf{X}$ contains n samples with D bands, i -th pixel can be represented as $\mathbf{x}_i \in \mathbf{R}^D$. Denote the class label of $\mathbf{x}_i$ as $l_i \in \{1, 2, \dots, c\}$, where c is the number of land-cover types. For linear manifold learning methods, the goal is to seek a projection matrix $\mathbf{V} \in \mathbf{R}^{D \times d}$, which can map $\mathbf{X} \in \mathbf{R}^{D \times n}$ into a d -dimension embedding space, where $d \ll D$. With the projection matrix $\mathbf{V}$, we can compute low-dimensional embedding features $\mathbf{Y} \in \mathbf{R}^{d \times n}$ as $\mathbf{Y} = \mathbf{V}^T \mathbf{X}$.

2.1. Graph Embedding

The graph embedding (GE) framework embeds a graph into a vector space where the inherent properties of the graph can be preserved. In GE, an intrinsic graph $\mathbf{G} = \{\mathbf{X}, \mathbf{W}\}$ and a penalty graph $\mathbf{G}^P = \{\mathbf{X}, \mathbf{W}^P\}$ are constructed based on the statistical or geometric properties of data, where $\mathbf{X}$ is the vertex set, $\mathbf{W} \in \mathbf{R}^{n \times n}$ and $\mathbf{W}^P \in \mathbf{R}^{n \times n}$ are the similarity matrices of $\mathbf{G}$ and $\mathbf{G}^P$, respectively. Intrinsic graph $\mathbf{G}$ is used to describe desirable statistical or geometrical properties of data, while penalty graph $\mathbf{G}^P$ is adopted to characterize dissimilarity relationship between the vertex pairs.

The purpose of GE is to represent each sample as a low-dimensional vector in graphs which can preserve similarity relationship of vertex pairs. The optimal low-dimensional embedding can be given by the graph preserving criterion as Equation (1):

$$\min_{\text{tr}(\mathbf{Y}^T \mathbf{B} \mathbf{Y})=h} \frac{1}{2} \sum_{i \neq j} \|y_i - y_j\|^2 w_{ij} = \min_{\text{tr}(\mathbf{Y}^T \mathbf{B} \mathbf{Y})=h} \text{tr}(\mathbf{Y}^T \mathbf{L} \mathbf{Y}) \quad (1)$$

where h is a constant, $\mathbf{B}$ is a constraint matrix which is defined to avoid the trivial solution of Equation (1). For scale normalization, $\mathbf{B}$ is typically set as a diagonal matrix or the Laplacian matrix of $\mathbf{G}^P$. Then, the Laplacian matrices $\mathbf{L}$ and $\mathbf{L}^P$ in graphs $\mathbf{G}$ and $\mathbf{G}^P$ can be defined as Equations (2) and (3):

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \quad \mathbf{D} = \text{diag} \left(\left[\sum_{j=1}^n w_{1j}, \sum_{j=1}^n w_{2j}, \dots, \sum_{j=1}^n w_{nj} \right] \right), \quad \mathbf{W} = [w_{ij}]_{i,j=1}^n \quad (2)$$

$$\mathbf{L}^P = \mathbf{D}^P - \mathbf{W}^P, \quad \mathbf{D}^P = \text{diag} \left(\left[\sum_{j=1}^n w_{1j}^P, \sum_{j=1}^n w_{2j}^P, \dots, \sum_{j=1}^n w_{nj}^P \right] \right), \quad \mathbf{W}^P = [w_{ij}^P]_{i,j=1}^n \quad (3)$$

2.2. Local Scaling Cut

Local scaling cut (LSC) is proposed based on GE framework, it compacts the data which come from the same class and separates the data which come from different classes. LSC seeks a linear projection matrix such that the between-class dissimilarity matrix is maximized, whereas the total-class dissimilarity matrix is minimized in the projected space.

To exploit the intrinsic geometry and local manifold structure of the data, LSC constructs a localized k -nearest neighbor graph to characterize the similarity relationship between the neighboring pixels. Suppose $N^b(\mathbf{x}_i)$ represents the k_b -nearest neighbors of $\mathbf{x}_i$ from different classes and $N^w(\mathbf{x}_i)$ represents the k_w -nearest neighbors of $\mathbf{x}_i$ from the same class, then the local between-class dissimilarity matrix $\mathbf{S}_b$ and the local within-class dissimilarity matrix $\mathbf{S}_w$ can be given in Equations (4) and (5) as:

$$\mathbf{S}_b = \sum_{s=1}^c \sum_{\mathbf{x}_i \in U_s} \sum_{\mathbf{x}_j \in N^b(\mathbf{x}_i)} H_{ij}^b (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T \quad (4)$$

$$\mathbf{S}_w = \sum_{s=1}^c \sum_{\mathbf{x}_i \in U_s} \sum_{\mathbf{x}_j \in N^w(\mathbf{x}_i)} H_{ij}^w (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T \quad (5)$$

where U_s denotes all the samples in s -th class, H_{ij}^b and H_{ij}^w are the weights of $\mathbf{x}_i$ and $\mathbf{x}_j$ in $N^b(\mathbf{x}_i)$ and $N^w(\mathbf{x}_i)$, respectively. The weights H_{ij}^b and H_{ij}^w are defined in Equations (6) and (7) as:

$$H_{ij}^b = \begin{cases} 1/(n_s k_b) & \mathbf{x}_j \in N^b(\mathbf{x}_i) \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

$$H_{ij}^w = \begin{cases} 1/(n_s k_w) & \mathbf{x}_j \in N^w(\mathbf{x}_i) \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

in which n_s is the total number of elements in the s -th class. Then the local scaling cut criterion is formulated as Equation (8):

$$S_{cuts}(\mathbf{V}) = \max_{\mathbf{V}} \frac{\text{tr}(\mathbf{V}^T \mathbf{S}_b \mathbf{V})}{\text{tr}[\mathbf{V}^T (\mathbf{S}_b + \mathbf{S}_w) \mathbf{V}]} = \max_{\mathbf{V}} \frac{\text{tr}(\mathbf{V}^T \mathbf{S}_b \mathbf{V})}{\text{tr}(\mathbf{V}^T \mathbf{T} \mathbf{V})} \quad (8)$$

3. Spatial-Spectral Multi-Manifold Discriminant Analysis

To effectively reveal the intrinsic multi-manifold structure of HSI data, a spatial-spectral multi-manifold discriminant analysis (SSMMDA) method was proposed for DR. SSMMDA divides samples into different subsets based on their label information, each subset is treated as a submanifold. Then, for each submanifold, it selects spectral neighbors to construct a spectral-domain intramanifold graph to preserve the local neighborhood structure of HSI and it constructs two weighted scatter matrices in spatial domain to characterize the within-manifold compactness and the between-manifold separability. After that, a spatial-spectral combined objective function is designed for each submanifold and we can seek a discriminant projection matrix for each submanifold by maximizing the spatial-spectral intermanifold scatter matrix and minimizing the spatial-spectral intramanifold scatter matrix simultaneously. Finally, low-dimensional embedding for each submanifold can be obtained by corresponding projection matrix and embedding features of each submanifold are fused for classification. SSMMDA improves both the intramanifold compactness and intermanifold separability of HSI data, it possesses a better discriminative power to improve classification performance. The process of the SSMMDA method is shown in Figure 1.

Figure 1. The flowchart of the proposed spatial-spectral multi-manifold discriminant analysis (SSMMDA) method.

3.1. Spectral-Domain Multi-Manifold Analysis Model

In some situations, high-dimensional data do not lie on a single manifold, samples from different classes lie on different submanifolds [33,34]. Existing manifold learning algorithms usually characterize the similarity relationships of high-dimensional data in the same manifold space and they can not discover the multi-manifold structure of data. Therefore, it is a crucial issue to preserve the local structure of different submanifolds by performing multiple manifold learning. In the proposed SSMMDA method, a multi-manifold learning model in spectral domain is designed to reveal the multi-manifold structure of HSI. At first, according to the label information of training samples,

HSI data is divided into different subsets. Then, several intramanifold graphs are constructed in spectral domain for different subsets, which can effectively characterize the similarity relationship between samples belonging to the same submanifold.

For HSI data, each class can be treated as a submanifold. Therefore, a HSI data set with D bands can be represented as $\mathbf{X} = [\mathbf{M}_1, \mathbf{M}_2, \dots, \mathbf{M}_c] \in \mathbf{R}^{D \times c}$, where c is the number of land-cover types. Denote the i -th point in the r -th submanifold as $\mathbf{M}_r^i$, then the r -th submanifold can be given as $\mathbf{M}_r = [\mathbf{M}_r^1, \mathbf{M}_r^2, \mathbf{M}_r^3, \dots, \mathbf{M}_r^{n_r}]$, where n_r is the number of samples in $\mathbf{M}_r$.

To characterize the discriminative manifold structure of $\mathbf{M}_r$, a spectral domain intramanifold graph $\mathbf{G}_r(\mathbf{M}_r, \mathbf{W}_r)$ can be constructed, where $\mathbf{M}_r$ is the vertex set of the graph and $\mathbf{W}_r$ is the weight matrix. In graph $\mathbf{G}_r$, an edge is put between nodes i and j if $\mathbf{M}_r^i$ and $\mathbf{M}_r^j$ are neighbors, otherwise they will not be connected and the weight between $\mathbf{M}_r^i$ and $\mathbf{M}_r^j$ can be defined as Equation (9):

$$w_r^{ij} = \begin{cases} \exp\left(-\frac{\|\mathbf{M}_r^i - \mathbf{M}_r^j\|^2}{2t_{ri}^2}\right), & \mathbf{M}_r^j \in N_{ik}(\mathbf{M}_r^i) \text{ or } \mathbf{M}_r^i \in N_{jk}(\mathbf{M}_r^j) \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where $N_{ik}(\mathbf{M}_r^i)$ is the neighbor set of $\mathbf{M}_r^i$, k is the number of intramanifold neighbors for graph $\mathbf{G}_r$ and $t_{ri} = \frac{1}{k} \sum_{j=1}^k \|\mathbf{M}_r^i - \mathbf{M}_r^j\|$ is a heat kernel parameter.

To enhance the aggregation of data points which possess similar spectral characteristics in low-dimensional space, the objective function can be given in Equation (10):

$$\begin{aligned} & \min_{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_c} J(\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_c) \\ &= \min_{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_c} (\psi(\mathbf{V}_1) + \psi(\mathbf{V}_2) + \dots + \psi(\mathbf{V}_c)) \\ &= \min_{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_c} \left[\frac{1}{2} \sum_{i=1}^{n_1} \sum_{j=1}^k \left\| y_1^i - y_1^j \right\|^2 w_1^{ij} + \frac{1}{2} \sum_{i=1}^{n_2} \sum_{j=1}^k \left\| y_2^i - y_2^j \right\|^2 w_2^{ij} + \dots + \frac{1}{2} \sum_{i=1}^{n_c} \sum_{j=1}^k \left\| y_c^i - y_c^j \right\|^2 w_c^{ij} \right] \quad (10) \\ &= \sum_{r=1}^c \min_{\mathbf{V}_r} \left(\frac{1}{2} \sum_{i=1}^{n_r} \sum_{j=1}^k \left\| y_r^i - y_r^j \right\|^2 w_r^{ij} \right) \end{aligned}$$

in which y_r^i and y_r^j are the low-dimensional representations of $\mathbf{M}_r^i$ and $\mathbf{M}_r^j$ for submanifold $\mathbf{M}_r$, respectively.

According to Equation (10), c different projection matrices can be obtained. For the r -th submanifold in spectral domain, the optimization problem of submanifold $\mathbf{M}_r$ can be transformed as Equation (11):

$$\begin{aligned} & \min_{\mathbf{V}_r} \left(\frac{1}{2} \sum_{i=1}^{n_r} \sum_{j=1}^k \left\| y_r^i - y_r^j \right\|^2 w_r^{ij} \right) \\ &= \min_{\mathbf{V}_r} \left(\frac{1}{2} \sum_{i=1}^{n_r} \sum_{j=1}^k \left\| \mathbf{V}_r^T \mathbf{M}_r^i - \mathbf{V}_r^T \mathbf{M}_r^j \right\|^2 w_r^{ij} \right) \\ &= \min_{\mathbf{V}_r} \left(\text{tr}(\mathbf{V}_r^T \left[\frac{1}{2} \sum_{i=1}^{n_r} \sum_{j=1}^k (\mathbf{M}_r^i w_r^{ij} \mathbf{M}_r^{iT} - 2\mathbf{M}_r^i w_r^{ij} \mathbf{M}_r^{jT} + \mathbf{M}_r^j w_r^{ij} \mathbf{M}_r^{jT}) \right] \mathbf{V}_r) \right) \quad (11) \\ &= \min_{\mathbf{V}_r} (\text{tr}(\mathbf{V}_r^T \mathbf{M}_r (\mathbf{D}_r - \mathbf{W}_r) \mathbf{M}_r^T \mathbf{V}_r)) \\ &= \min_{\mathbf{V}_r} (\text{tr}(\mathbf{V}_r^T \mathbf{M}_r \mathbf{B}_r \mathbf{M}_r^T \mathbf{V}_r)) \end{aligned}$$

where the affinity matrix $\mathbf{W}_r$ is used to generate the intramanifold graph Laplacian matrix $\mathbf{B}_r = \mathbf{D}_r - \mathbf{W}_r$ and $\mathbf{D}_r$ is a diagonal matrix whose entries are column sums of $\mathbf{W}_r$.

To eliminate the impact of scale factor, a constraint $\mathbf{Y}_r \mathbf{D}_r \mathbf{Y}_r^T = \mathbf{V}_r^T \mathbf{M}_r \mathbf{D}_r \mathbf{M}_r^T \mathbf{V}_r = \mathbf{I}$ is added for Equation (11), where $\mathbf{I} = \text{diag}[1, \dots, 1]$. Therefore, the optimal function of Equation (11) equals to the following optimization problem of Equation (12):

$$\begin{cases} \min_{\mathbf{V}_r} \mathbf{V}_r^T \mathbf{M}_r \mathbf{B}_r \mathbf{M}_r^T \mathbf{V}_r \\ \text{s. t. } \mathbf{V}_r^T \mathbf{M}_r \mathbf{D}_r \mathbf{M}_r^T \mathbf{V}_r = \mathbf{I} \end{cases} \quad (12)$$

With the Lagrange multiplier method, the optimization problem in Equation (12) can be transformed into the following generalized eigenvalue problem of Equation (13):

$$\mathbf{M}_r \mathbf{B}_r \mathbf{M}_r^T \mathbf{V}_r = \lambda_r \mathbf{M}_r \mathbf{D}_r \mathbf{M}_r^T \mathbf{V}_r \quad (13)$$

where λ_r is the eigenvalue.

3.2. Spatial-Domain Multi-Manifold Analysis Model

According to the spatial distribution consistency of HSI, the pixels are generally distributed in blocks, such as *Soil*, *Water*, *Building* and *Woods* [51]. Therefore, the neighboring pixels in a spatial window are more likely belonging to the same class and they lie on the same manifold [52]. To utilize the spatial information in HSI, a spatial-domain weighted intramanifold scatter matrix is designed to characterize the similarity relationship for each submanifold and a spatial-domain weighted intermanifold scatter matrix is defined to represent the dissimilarity relationship between different submanifolds.

Suppose a pixel $\mathbf{x}_i$ in HSI lies on the r -th spectral submanifold, then it can be represented as $\mathbf{M}_r^i$. Since the pixels in the local spatial patch of $\mathbf{x}_i$ commonly lie on the same submanifold as $\mathbf{x}_i$, so the neighboring pixels can be given as $\mathbf{M}_r^{i1}, \mathbf{M}_r^{i2}, \mathbf{M}_r^{i3}, \dots, \mathbf{M}_r^{iw^2}$, where the odd number w is the size of spatial window.

From the viewpoint of classification, we aim to minimize the distances of intramanifold points and maximize the distances between different submanifolds in low-dimensional space so that the submanifold margins can be maximized for DR. Denoting y_r^i and y_r^{ij} as the low-dimensional representations of i -th sample and its j -th spatial neighbor in the r -th manifold space, the optimization problem can be formulated as Equation (14):

$$\begin{aligned} & \min_{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_c} J(\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_c) \\ &= \min_{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_c} (\psi(\mathbf{V}_1) + \psi(\mathbf{V}_2) + \dots + \psi(\mathbf{V}_c)) \\ &= \min_{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_c} \left(\frac{\sum_{i=1}^{n_1} \sum_{j=1}^{w^2} \|y_1^i - y_1^{ij}\|^2 s_1^{ij}}{\sum_{i=1}^{n_1} \|y_1^i - \bar{y}^1\|^2 s_1^i} + \frac{\sum_{i=1}^{n_2} \sum_{j=1}^{w^2} \|y_2^i - y_2^{ij}\|^2 s_2^{ij}}{\sum_{i=1}^{n_2} \|y_2^i - \bar{y}^2\|^2 s_2^i} + \dots + \frac{\sum_{i=1}^{n_c} \sum_{j=1}^{w^2} \|y_c^i - y_c^{ij}\|^2 s_c^{ij}}{\sum_{i=1}^{n_c} \|y_c^i - \bar{y}^c\|^2 s_c^i} \right) \quad (14) \\ &= \sum_{r=1}^c \min_{\mathbf{V}_r} \frac{\sum_{i=1}^{n_r} \sum_{j=1}^{w^2} \|y_r^i - y_r^{ij}\|^2 s_r^{ij}}{\sum_{i=1}^{n_r} \|y_r^i - \bar{y}^r\|^2 s_r^i} \end{aligned}$$

where y_r^i is the low-dimensional representation of i -th sample in the r -th manifold space, $\bar{y}^r$ is the mean of all training data after they are mapped to submanifold $\mathbf{M}_r$, $s_r^{ij} = \exp\left(-\|y_r^i - y_r^{ij}\|^2 / 2(t_{ij})^2\right)$ is the similarity weight between y_r^i and y_r^{ij} , $s_r^i = \exp\left(-\|y_r^i - \bar{y}^r\|^2 / 2(t_{ij})^2\right)$ is the similarity weight between y_r^i and $\bar{y}^r$ and $t_{ij} = \frac{1}{k} \sum_{j=1}^k \|y_r^i - y_r^{ij}\|$ is a heat kernel parameter.

Considering that the c projection matrices are independent of each other, we can divide Equation (14) into different parts and solve them separately. As for the submanifold $\mathbf{M}_r$, the optimization problem can be written as Equation (15):

$$\begin{aligned}
& \min_{\mathbf{V}_r} \frac{\sum_{i=1}^{n_r} \sum_{j=1}^{w^2} \left\| \mathbf{y}_r^i - \mathbf{y}_r^{ij} \right\|^2 s_r^{ij}}{\sum_{i=1}^n \left\| \mathbf{y}_r^i - \bar{\mathbf{y}}^r \right\|^2 s_r^i} \\
&= \min_{\mathbf{V}_r} \operatorname{tr} \left(\frac{\sum_{i=1}^{n_r} \sum_{j=1}^{w^2} \left\| \mathbf{V}_r \mathbf{M}_r^i - \mathbf{V}_r \mathbf{M}_r^{ij} \right\|^2 s_r^{ij}}{\sum_{i=1}^n \left\| \mathbf{V}_r \mathbf{M}_r^i - \mathbf{V}_r \bar{\mathbf{M}} \right\|^2 s_r^i} \right) \\
&= \min_{\mathbf{V}_r} \operatorname{tr} \left(\frac{\mathbf{V}_r^T \sum_{i=1}^{n_r} \sum_{j=1}^{w^2} \left\| \mathbf{M}_r^i - \mathbf{M}_r^{ij} \right\|^2 s_r^{ij} \mathbf{V}_r}{\mathbf{V}_r^T \sum_{i=1}^n \left\| \mathbf{V}_r \mathbf{M}_r^i - \mathbf{V}_r \bar{\mathbf{M}} \right\|^2 s_r^i \mathbf{V}_r} \right) \\
&= \min_{\mathbf{V}_r} \operatorname{tr} \left(\frac{\mathbf{V}_r^T \mathbf{H}_r \mathbf{V}_r}{\mathbf{V}_r^T \mathbf{S}_r \mathbf{V}_r} \right)
\end{aligned} \tag{15}$$

in which $\bar{\mathbf{M}}$ is the mean of all samples in HSI data, $\mathbf{H}_r$ and $\mathbf{S}_r$ are the intramanifold scatter matrix and the intermanifold scatter matrix for the submanifold $\mathbf{M}_r$ and they can be defined as Equations (16) and (17):

$$\mathbf{H}_r = \sum_{i=1}^{n_r} \sum_{j=1}^{w^2} \left\| \mathbf{M}_r^i - \mathbf{M}_r^{ij} \right\|^2 s_r^{ij} = \sum_{i=1}^{n_r} \sum_{j=1}^{w^2} \left(\mathbf{M}_r^i - \mathbf{M}_r^{ij} \right) \left(\mathbf{M}_r^i - \mathbf{M}_r^{ij} \right)^T s_r^{ij} \tag{16}$$

$$\mathbf{S}_r = \sum_{i=1}^n \left\| \mathbf{M}_r^i - \bar{\mathbf{M}} \right\|^2 s_r^i \tag{17}$$

After that, an optimal projection matrix $\mathbf{V}_r = [v_{r1}, v_{r2}, \dots, v_{rd}]$ can be obtained by solving the generalized eigenvalue problem of Equations (18):

$$\mathbf{H}_r \mathbf{V}_r = \lambda_r \mathbf{S}_r \mathbf{V}_r \tag{18}$$

in which λ_r is the eigenvalue of Equation (18).

3.3. Spatial-Spectral Multi-Manifold Analysis Model

In HSI, spectral information and spatial structure are complementary to each other, spectral information characterizes the spectrum reflection characteristics of land covers, while spatial information reflects the spatial relationship of objects. To learn discriminant projections, the spectral and spatial information can be combined for feature learning. Therefore, we propose a spatial-spectral multiple manifold-based DR method for HSI classification, which extracts the feature information of HSI from both spatial and spectral domains and it can effectively reveal the multi-manifold structure in HSI.

By combining the spatial information and spectral information, we can obtain a spatial-spectral intramanifold scatter matrix and a spatial-spectral intermanifold scatter matrix for submanifold $\mathbf{M}_r$ as Equations (19) and (20):

$$\mathbf{Z}_r^{intra} = \mathbf{M}_r \mathbf{B}_r \mathbf{M}_r^T + \alpha * \mathbf{H}_r \tag{19}$$

$$\mathbf{Z}_r^{inter} = \mathbf{M}_r \mathbf{D}_r \mathbf{M}_r^T + \beta * \mathbf{S}_r \tag{20}$$

in which α and β are the tradeoff parameters which are used to balance the contribution of spatial and spectral information in the process of DR.

With the Lagrange multiplier method, the optimization problem for submanifold $\mathbf{M}_r$ can be transformed into the following form of Equation (21):

$$(\mathbf{M}_r \mathbf{B}_r \mathbf{M}_r^T + \alpha * \mathbf{H}_r) \mathbf{V}_r = \lambda_r (\mathbf{M}_r \mathbf{D}_r \mathbf{M}_r^T + \beta * \mathbf{S}_r) \mathbf{V}_r \tag{21}$$

in which λ_r is the eigenvalue of Equation (21). After obtaining the eigenvectors $v_{r1}, v_{r2}, v_{r3}, \dots, v_{rd}$ corresponding to the d smallest eigenvalues, the optimal projection matrix can be given by Equation (22):

$$\mathbf{V}_r = [\mathbf{v}_{r1}, \mathbf{v}_{r2}, \dots, \mathbf{v}_{rd}] \in \mathbb{R}^{D \times d} \quad (22)$$

With the same operations, different projection matrices $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_c$ can be obtained as Equation (22), which correspond to different submanifolds embedded in HSI data. Denote the feature representations of $\mathbf{X}$ in each low-dimensional space as $\mathbf{Y}_1 = \mathbf{V}_1^T \mathbf{X}, \mathbf{Y}_2 = \mathbf{V}_2^T \mathbf{X}, \dots, \mathbf{Y}_c = \mathbf{V}_c^T \mathbf{X}$, then the embedding features can be fused to enhance the classification performance of HSI and the fused features of $\mathbf{X}$ are expressed as $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_c]$. The detailed steps of SSMDA are shown in Algorithm 1.

Algorithm 1 SSMDA

Input: HSI data set $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_n]$ and the class label $l_i \in \{1, 2, 3, \dots, c\}$, spatial window size w , neighbor number k , tradeoff parameters α and β and embedding dimension d ($d \ll D$).

- 1: Divide the HSI data into c different submanifolds of $\mathbf{M}_1, \mathbf{M}_2, \mathbf{M}_3, \dots, \mathbf{M}_c$,
- 2: for $r = 1$ to c do
- 3: Find the spectral-domain intramanifold neighbors set $N_{ik}(\mathbf{M}_r^i)$ of $\mathbf{M}_r^i$.
- 4: Calculate intramanifold graph weight by Equation (9).
- 5: Compute intramanifold graph scatter matrix $\mathbf{M}_r \mathbf{B}_r \mathbf{M}_r^T$ and constraint matrix $\mathbf{M}_r \mathbf{D}_r \mathbf{M}_r^T$.
- 6: Find the spatial-domain neighboring pixels set of $\mathbf{M}_r^i$.
- 7: Compute the intramanifold scatter matrix $\mathbf{H}_r$ and the intermanifold scatter matrix $\mathbf{S}_r$ in spatial domain.
- 8: Compute the spatial-spectral intramanifold scatter matrix $\mathbf{Z}_r^{\text{intra}}$ and the spatial-spectral intermanifold scatter matrix $\mathbf{Z}_r^{\text{inter}}$.
- 9: Solve the generalized eigenvalue problem of Equation (21).
- 10: Obtain the projection matrix $\mathbf{V}_r$ for submanifold $\mathbf{M}_r$: $\mathbf{V}_r = [v_{r1} \ v_{r2} \ \dots \ v_{rd}] \in \mathbb{R}^{D \times d}$
- 11: end for
- 12: Obtain c different projection matrices as $\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_c$.
- 13: Calculate the low dimensional features of $\mathbf{X}$ in each submanifold as $\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_c$, where $\mathbf{Y}_1 = \mathbf{V}_1^T \mathbf{X}$, $\mathbf{Y}_2 = \mathbf{V}_2^T \mathbf{X}, \dots, \mathbf{Y}_c = \mathbf{V}_c^T \mathbf{X}$

Output: The low-dimensional features are represented as $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_c]$

4. Experimental Results

In this section, three public HSI data sets are adopted to evaluate the effectiveness of SSMDA by comparing it with some state-of-art DR algorithms. Experiments on these HSI data sets demonstrate that the proposed SSMDA method can explore spatial-spectral combined information and reveal the intrinsic multi-manifold structure in HSI, which can significantly improve the classification performance of HSI.

4.1. Experiment Data Set

PaviaU data set: This data set was captured on the city of Pavia, Italy by the ROSIS-03 (Reflective Optics Spectrographic Imaging System) airborne instrument. The ROSIS-03 sensor comprises 115 data channels with a spectral coverage ranging from 0.43 to 0.86 μm . After removing water absorption bands, we adopted the remaining 103 bands for experiments. This data set was originally 610×340 pixels with a spatial resolution of 1.3 m. There are nine classes considered in the data set, that is, *asphalt*,

meadows, gravel, trees, metal sheets, soil, bitumen, bricks and shadows. Figure 2 shows its false color scene and the corresponding ground truth.

Figure 2. PaviaU hyperspectral image (HSI). (a) HSI in false-color; (b) Ground-truth map.

Heihe data set [53,54]: This data set is captured by CASI/SASI sensors in Zhangye basin, Gansu Province, China. It is provided by Heihe Plan Science Data Center which is sponsored by the integrated research on the eco-hydrological process of the Heihe River Basin of the National Natural Science Foundation of China. The data set possesses a spatial size of 684×453 pixels with a geometric resolution of 2.4 m. Considering that 14 bands are easily absorbed by water absorption, the remaining 135 spectral bands are used for experiments. The data set contains 9 different land cover types. The scene in false color and its ground truth are shown in Figure 3.

Figure 3. Heihe hyperspectral image. (a) HSI in false-color; (b) Ground-truth map.

Washington DC Mall data set: This data set is acquired by the airborne HYDICE from the mall in Washington DC. The spatial size of this data set is 250×307 pixels and the spatial resolution equals to 3 m per pixel. This data set covers a $0.4 \mu\text{m}$ to $2.4 \mu\text{m}$ spectral range with 210 bands. There are six different land cover types in the data set such as *Road*, *Building*, *Trail* and *Vegetation*. This scene in false color and its corresponding ground truth are shown in Figure 4.

Figure 4. Washington DC Mall hyperspectral image. (a) HSI in false-color; (b) Ground-truth map.

4.2. Experimental Setup

In each experiment, we randomly divided the HSI data into training and test sets. Training set is used to construct a DR model and a test set is adopted to verify the effectiveness of the DR model. Then, the nearest neighbor classifier (NN) [55] was used for classification. After that, the classification accuracy of each class (CA), overall classification accuracy (OA), average classification accuracy (AA) and kappa coefficient (KC) were adopted to evaluate the performance of different DR methods [47,52]. Among them, CA is the classification accuracy on each class of land covers, AA is a measure of the mean value of the classification accuracies of all classes, OA refers to the number of correctly classified instances divided by the total number of test samples, KC is a statistical measurement of consistency between the ground truth map and the final classification map. Suppose N_{ij} is the number of samples in the j -th class that are classified into the i -th class, then the classification accuracy of i -th class (CA_i), AA, OA and KC can be defined as Equations (23)–(26):

$$CA_i = \frac{N_{ii}}{\sum_{j=1}^c N_{ji}} \quad (23)$$

$$AA = \frac{1}{c} \sum_{i=1}^c CA_i = \frac{1}{c} \sum_{i=1}^c \frac{N_{ii}}{\sum_{j=1}^c N_{ji}} \quad (24)$$

$$OA = \frac{1}{n} \sum_{i=1}^c N_{ii} \quad (25)$$

$$KC = \frac{n \sum_{i=1}^c N_{ii} - \sum_{i=1}^c \left(\sum_{j=1}^c N_{ji} \times \sum_{j=1}^c N_{ij} \right)}{n^2 - \sum_{i=1}^c \left(\sum_{j=1}^c N_{ji} \times \sum_{j=1}^c N_{ij} \right)} \quad (26)$$

in which c is the number of land cover types of HSI data and n is the total number of HSI pixels.

To evaluate the classification performance of the proposed SSMDA algorithm, some state-of-art DR methods were selected for comparison, such methods include principal component analysis (PCA) [56], NPE [28], LPP [29], linear discriminant analysis (LDA) [57], maximum margin criterion (MMC) [58], MFA [31], nonparametric weighted feature extraction (NWFE) [59], sparsity preserving projections (SPP) [29], sparse discriminant embedding (SDE) [60], sparse manifold embedding (SME) [61], SMML [37], MMDA [33], DSSM [49], LPNPE [48] and weighted spatial-spectral and global-local discriminant analysis (WSSGLDA) [62]. RAW indicates that samples are classified without dimensionality reduction. The former twelve methods are spectral-based DR methods, they only consider the spectral information in HSI. The later three algorithms are spatial-spectral combined DR methods, which simultaneously consider the spectral information and spatial structure of HSI data. Among all the spectral-based DR algorithms, PCA, NPE, LPP, LDA, MMC and MFA are single

manifold learning methods, while SMML and MMDA assume that the high dimensional HSI data lies on a multi-manifold structure.

In the experiments, we optimized the parameters to achieve good results for each method. For NPE, LPP, SMML and MMDA, we set the number of neighbors to 5. For MFA and DSSM, the number of intraclass and interclass neighbors were set to 7. For DSSM, LPNPE and WSSGLDA, the window size was set to 15 for PaviaU and Heihe data sets and 5 for the Washington DC data set. To reduce noise in the HSI, the HSI was pre-processed with a 5×5 spatial filter. For robustness, all the experiments were repeated for 10 times in each condition.

4.3. Analysis of Window Size w and Neighbor Number k

To investigate the classification performance of SSMMDA with different window size w and neighbor number k , 20 samples were randomly selected from each class as training set, and the remaining were used as test set. In the experiment, parameters w and k were chosen within a set of $\{3, 5, 7, \dots, 25\}$ and a set of $\{1, 2, 3, \dots, 20\}$, respectively and the NN classifier were used to classify the test samples. Figure 5 shows the OAs versus different values of w and k .

Figure 5. Overall classification accuracies (OAs) with different window sizes w and neighbor number k . (a) PaviaU. (b) Heihe. (c) Washington DC Mall.

As shown in Figure 5, as the window size w increases, the OAs first improve and then maintain a stable value. The reason is that a larger spatial window possesses more spatial neighbors, the neighbors can be utilized to construct a more effective DR model. However, if w is too large, the spatial information in neighbors will be redundant for DR of HSI. Therefore, the classification accuracies of SSMMDA will tend to remain stable. Besides, the classification accuracies improve with the increase of k and then slightly fluctuate. Because a small number of neighbor points cannot effectively reveal the multi-manifold structure of HSI in spectral domain, while a large k will produce too much useless information for adjacency graph construction. To balance running time and classification performance, we set $w = 15$, $k = 5$ for PaviaU data set, $w = 15$ and $k = 8$ for Heihe data set, and $w = 17$ and $k = 8$ for Washington DC Mall data set in the following experiments.

4.4. Analysis of Tradeoff Parameters α and β

SSMMDA has two tradeoff parameters α and β that are used to balance the contribution between the spectral information and spatial structure of HSI. To verify the classification performance of different values of α and β , we randomly selected 20 samples from each type of land covers for training and the rest were for testing. Parameters α and β were both tuned with a set of $\{0, 0.1, 0.2, 0.3, \dots, 0.9, 1\}$. After that, the NN classifier was used for classification. Figure 6 shows the average OAs with different α and β .

As can be seen from Figure 6, an increasing α causes a subtle change in terms of classification accuracy, because both spectral-domain intramanifold graph and spatial-domain intramanifold scatter matrix can enhance the intramanifold compactness and a suitable α can balance the contribution of spectral information and spatial structure in HSI. For parameter β , the OAs improve with the increase

of β and then reach to a stable peak value, which indicates that the intermanifold scatter matrix can effectively enhance the intermanifold separability. To achieve better classification performance, we set α and β to 0.8 and 0.6 for the PaviaU data set, 0.9 and 0.6 for the Heihe and Washington DC Mall data sets in the experiments.

Figure 6. OAs with different tradeoff parameters α and β . (a) PaviaU. (b) Heihe. (c) Washington DC Mall.

4.5. Investigation of Embedding Dimension d

The proposed SSMDA method belongs to a dimensionality reduction method, thus we analyze the classification accuracy with different embedding dimensions d . To evaluate the influence of different embedding dimensions on classification performance, 30 samples per class were randomly chosen as training samples, the remaining were used as test samples. Figure 7 shows the classification results of different algorithms under different embedding dimensions.

Figure 7. OAs with different embedding dimensions d . (a) PaviaU. (b) Heihe. (c) Washington DC Mall.

In Figure 7, some similar conclusions can be obtained on the PaviaU, Heihe and Washington DC Mall data sets. When the value of d is lower than 15, the OAs of the DR algorithms increase significantly as the dimension increases. The reason is that a larger d means more discriminant features are utilized in the classification process, which is helpful to distinguish the land covers with similar spectrum. When d is larger than 15, the feature information contained in embedding space is close to saturation, an increasing d will not lead to the increase of the OAs. Besides, most DR methods can obtain better classification performance than RAW, which indicates that these algorithms can effectively reduce the useless information in HSI. To achieve better classification results for all algorithms, we set $d = 30$ for all DR methods on these three data sets and the dimension of LDA is $c-1$, c is the number of land cover types in each data set.

5. Analysis and Discussion

5.1. Analysis of Training Sample Size

To investigate the classification performance of the algorithms under different number of training samples, n_i ($n_i = 10, 20, 30, 40, 50, 60$) samples were randomly selected from each class for training, and the rest of the samples were used for testing. For robustness, all the experiments were repeated for 10 times in each condition. Tables 1–3 report the classification results of different DR methods with different number of training samples on the PaviaU, Heihe and Washington DC Mall data sets.

According to Tables 1–3, the OAs of the above algorithms are significantly improved as n_i increases. The reason is that a large number of training samples contain more abundant feature information, which is conducive to the feature learning of HSI data. For single manifold algorithms with spectral information, supervised methods of LDA, SME and MFA, are superior to unsupervised ones such as PCA, NPE, LPP and SPP, it indicates that the priori label information of HSI is helpful to improve the classification performance of HSI. For the single manifold-based DR algorithms, most spatial-spectral combined methods achieve higher classification accuracies than the spectral DR methods. Because the HSI possesses the spatial consistency property and pixels from the same class are commonly distributed in blocks. Therefore, the spatial information is of great significance to improve the feature representation and classification performance of HSI. For the multiple manifold-based DR algorithms, the proposed SSMMDA method achieves better classification results than SMML and MMDA in most conditions, this is because SMML and MMDA only consider the spectral information of HSI and they do not utilize the spatial structure to further enhance the classification performance. For all the DR methods, SSMMDA has the better OAs and KCs than the other algorithms, because it simultaneously considers the multi-manifold structure and the spatial-spectral combined information in HSI.

5.2. Analysis of Classification Results

To explore the classification performance of SSMMDA on each class, we randomly selected a percentage of samples per class as training set and the remaining samples were used as a test set. The training set is used to construct a DR model and the test set is adopted to verify the effectiveness of the DR model. In the experiments, we set the percentage to 0.5% for the PaviaU data set, 0.1% for the Heihe data set and 1% for the Washington DC Mall data set. The CAs, OAs, AAs and KCs of different methods on the PaviaU, Heihe, Washington DC Mall data sets are detailed in Tables 4–6, respectively, and Figures 8–10 show the corresponding classification maps of different algorithms on these three data sets.

As shown in Tables 4–6, SSMMDA achieved better classification results than other algorithms in most classes and it possessed the best OA, AA and KC on the three HSI datasets. This indicates that the proposed SSMMDA method has a stronger ability to characterize the geometric relations between samples and extract more effective spatial-spectral combined information for classification. Besides, the classification maps of SSMMDA in Figures 8–10 produced more homogenous regions than the other methods, especially in the classes of *Asphalt*, *Bitumen* and *Bricks* for the PaviaU data set, *Corn*, *Watermelon* and *Artificial Surface* for the Heihe data set and *Road*, *Water*, *Building* and *Shadow* for the Washington DC Mall data set. The reason is that SSMMDA not only explores the multiple manifold structure of HSI data but also utilizes the spatial-spectral combined information to extract the discriminant features, which is conducive to the classification of HSI data.

Table 1. Classification results with different values of n_i on the PaviaU dataset with nearest neighbor (NN) classifier. Each method has two rows, where the first row is the OA $\pm$ STD (%) and the second row is the Kappa coefficient (KC).

Algorithm	10	20	30	40	50	60
RAW	67.71 $\pm$ 3.44 (0.601)	73.53 $\pm$ 2.51 (0.669)	76.49 $\pm$ 1.62 (0.703)	78.41 $\pm$ 1.42 (0.726)	79.72 $\pm$ 1.56 (0.742)	82.25 $\pm$ 1.26 (0.772)
PCA [56]	67.71 $\pm$ 3.44 (0.601)	73.53 $\pm$ 2.51 (0.669)	76.49 $\pm$ 1.61 (0.703)	78.40 $\pm$ 1.42 (0.726)	79.71 $\pm$ 1.57 (0.742)	82.24 $\pm$ 1.26 (0.772)
NPE [28]	73.62 $\pm$ 3.84 (0.670)	80.15 $\pm$ 2.73 (0.748)	84.30 $\pm$ 1.75 (0.798)	86.34 $\pm$ 2.13 (0.824)	88.37 $\pm$ 2.05 (0.850)	90.36 $\pm$ 1.82 (0.874)
LPP [29]	75.88 $\pm$ 2.25 (0.698)	82.96 $\pm$ 1.43 (0.782)	85.27 $\pm$ 0.67 (0.810)	86.47 $\pm$ 1.66 (0.826)	87.37 $\pm$ 1.56 (0.837)	89.32 $\pm$ 1.83 (0.861)
LDA [57]	75.59 $\pm$ 3.51 (0.696)	80.60 $\pm$ 1.66 (0.754)	86.66 $\pm$ 1.75 (0.827)	88.32 $\pm$ 1.57 (0.848)	90.43 $\pm$ 1.69 (0.875)	91.68 $\pm$ 1.25 (0.891)
MMC [58]	67.03 $\pm$ 3.18 (0.593)	70.94 $\pm$ 2.79 (0.638)	73.10 $\pm$ 1.73 (0.662)	74.95 $\pm$ 1.34 (0.684)	75.94 $\pm$ 1.66 (0.697)	78.14 $\pm$ 1.50 (0.722)
MFA [31]	79.67 $\pm$ 3.54 (0.742)	86.40 $\pm$ 1.95 (0.824)	88.63 $\pm$ 1.68 (0.852)	89.31 $\pm$ 1.52 (0.861)	90.53 $\pm$ 2.11 (0.876)	92.46 $\pm$ 1.44 (0.901)
NWFE [59]	73.07 $\pm$ 3.79 (0.664)	78.89 $\pm$ 1.92 (0.733)	82.89 $\pm$ 1.85 (0.78)	83.15 $\pm$ 1.78 (0.784)	84.61 $\pm$ 1.87 (0.802)	85.35 $\pm$ 1.45 (0.811)
SPP [29]	60.83 $\pm$ 3.74 (0.519)	66.62 $\pm$ 1.91 (0.586)	74.66 $\pm$ 1.84 (0.679)	76.87 $\pm$ 2.54 (0.707)	78.29 $\pm$ 1.90 (0.725)	78.68 $\pm$ 1.60 (0.730)
SDE [60]	67.42 $\pm$ 3.29 (0.597)	73.73 $\pm$ 1.55 (0.671)	78.85 $\pm$ 1.50 (0.730)	79.94 $\pm$ 2.37 (0.744)	81.00 $\pm$ 1.74 (0.758)	83.54 $\pm$ 1.35 (0.789)
SME [61]	77.43 $\pm$ 5.10 (0.717)	84.31 $\pm$ 2.65 (0.799)	89.05 $\pm$ 1.69 (0.857)	90.14 $\pm$ 1.04 (0.872)	90.26 $\pm$ 1.50 (0.873)	91.87 $\pm$ 0.89 (0.894)
SMML [37]	67.43 $\pm$ 3.34 (0.598)	73.09 $\pm$ 2.39 (0.663)	75.90 $\pm$ 1.55 (0.696)	77.78 $\pm$ 1.34 (0.718)	79.02 $\pm$ 1.43 (0.734)	81.34 $\pm$ 1.19 (0.761)
MMDA [33]	73.00 $\pm$ 3.94 (0.659)	81.00 $\pm$ 2.89 (0.758)	85.83 $\pm$ 1.82 (0.818)	88.04 $\pm$ 1.56 (0.845)	88.28 $\pm$ 1.42 (0.849)	90.76 $\pm$ 1.20 (0.879)
DSSM [49]	67.60 $\pm$ 3.44 (0.600)	73.19 $\pm$ 3.10 (0.665)	76.39 $\pm$ 1.61 (0.702)	78.34 $\pm$ 1.42 (0.725)	79.47 $\pm$ 1.52 (0.739)	82.16 $\pm$ 1.28 (0.771)
LPNPE [48]	82.17 $\pm$ 4.12 (0.773)	86.79 $\pm$ 2.59 (0.830)	90.17 $\pm$ 1.83 (0.873)	92.41 $\pm$ 0.89 (0.901)	92.48 $\pm$ 1.07 (0.902)	93.71 $\pm$ 1.04 (0.917)
WSSGLDA [62]	82.12 $\pm$ 2.75 (0.772)	88.30 $\pm$ 2.53 (0.849)	91.40 $\pm$ 1.77 (0.888)	92.79 $\pm$ 1.15 (0.906)	93.19 $\pm$ 1.06 (0.911)	94.48 $\pm$ 1.05 (0.927)
SSMMFL	85.86 $\pm$ 4.34 (0.820)	91.00 $\pm$ 2.35 (0.884)	93.27 $\pm$ 1.74 (0.912)	94.06 $\pm$ 0.80 (0.922)	94.39 $\pm$ 0.95 (0.927)	95.76 $\pm$ 1.06 (0.944)

Table 2. Classification results with different values of n_i on the Heihe dataset with NN classifier. Each method has two rows, where the first row is the OA $\pm$ STD (%) and the second row is the KC.

Algorithm	10	20	30	40	50	60
RAW	87.60 $\pm$ 1.15 (0.833)	89.72 $\pm$ 1.60 (0.862)	91.02 $\pm$ 0.70 (0.879)	91.49 $\pm$ 1.00 (0.885)	91.86 $\pm$ 0.69 (0.890)	92.08 $\pm$ 1.35 (0.893)
PCA [56]	87.60 $\pm$ 1.15 (0.833)	89.71 $\pm$ 1.60 (0.862)	91.02 $\pm$ 0.70 (0.879)	91.48 $\pm$ 1.00 (0.885)	91.86 $\pm$ 0.69 (0.890)	92.07 $\pm$ 1.35 (0.893)
NPE [28]	87.36 $\pm$ 2.37 (0.831)	90.41 $\pm$ 1.44 (0.871)	91.39 $\pm$ 1.26 (0.884)	92.06 $\pm$ 1.08 (0.893)	92.78 $\pm$ 0.96 (0.903)	93.31 $\pm$ 1.09 (0.910)
LPP [29]	90.30 $\pm$ 1.95 (0.869)	93.42 $\pm$ 1.14 (0.911)	93.90 $\pm$ 1.01 (0.917)	94.22 $\pm$ 0.97 (0.922)	94.68 $\pm$ 0.95 (0.928)	94.91 $\pm$ 0.84 (0.931)
LDA [57]	89.37 $\pm$ 1.18 (0.857)	92.44 $\pm$ 1.76 (0.898)	93.65 $\pm$ 0.63 (0.914)	93.74 $\pm$ 0.90 (0.915)	94.49 $\pm$ 0.30 (0.925)	94.50 $\pm$ 1.03 (0.925)
MMC [58]	86.56 $\pm$ 1.13 (0.820)	89.27 $\pm$ 1.83 (0.856)	90.66 $\pm$ 0.75 (0.874)	91.10 $\pm$ 1.09 (0.880)	91.48 $\pm$ 0.84 (0.885)	91.73 $\pm$ 1.44 (0.889)
MFA [31]	88.99 $\pm$ 3.04 (0.854)	90.65 $\pm$ 4.56 (0.876)	92.35 $\pm$ 1.01 (0.897)	92.86 $\pm$ 1.48 (0.904)	93.50 $\pm$ 1.85 (0.912)	93.71 $\pm$ 1.68 (0.915)
NWFE [59]	90.33 $\pm$ 2.29 (0.871)	92.75 $\pm$ 2.49 (0.903)	93.42 $\pm$ 1.55 (0.911)	94.08 $\pm$ 1.01 (0.920)	94.13 $\pm$ 0.53 (0.921)	94.74 $\pm$ 0.20 (0.929)
SPP [29]	64.53 $\pm$ 5.61 (0.555)	80.89 $\pm$ 2.35 (0.749)	85.00 $\pm$ 2.47 (0.802)	86.54 $\pm$ 1.24 (0.821)	87.72 $\pm$ 3.28 (0.837)	90.40 $\pm$ 0.95 (0.871)
SDE [60]	85.84 $\pm$ 2.59 (0.812)	88.96 $\pm$ 1.59 (0.852)	91.15 $\pm$ 1.44 (0.881)	91.77 $\pm$ 0.71 (0.889)	91.93 $\pm$ 1.22 (0.891)	92.48 $\pm$ 0.11 (0.898)
SME [61]	89.32 $\pm$ 3.24 (0.858)	92.52 $\pm$ 2.05 (0.899)	94.12 $\pm$ 1.01 (0.921)	94.85 $\pm$ 0.82 (0.930)	94.92 $\pm$ 0.82 (0.931)	95.56 $\pm$ 0.42 (0.940)
SMML [37]	87.46 $\pm$ 1.13 (0.832)	89.29 $\pm$ 1.93 (0.856)	90.70 $\pm$ 0.65 (0.875)	91.17 $\pm$ 1.15 (0.881)	91.56 $\pm$ 0.85 (0.886)	91.72 $\pm$ 1.49 (0.888)
MMDA [33]	90.63 $\pm$ 1.64 (0.875)	93.72 $\pm$ 1.23 (0.915)	94.10 $\pm$ 0.44 (0.923)	94.55 $\pm$ 1.12 (0.930)	95.57 $\pm$ 0.25 (0.941)	96.47 $\pm$ 0.92 (0.952)
DSSM [49]	87.64 $\pm$ 1.15 (0.834)	89.70 $\pm$ 1.63 (0.862)	91.02 $\pm$ 0.70 (0.879)	91.49 $\pm$ 1.00 (0.885)	91.87 $\pm$ 0.69 (0.890)	92.08 $\pm$ 1.35 (0.893)
LPNPE [48]	86.04 $\pm$ 2.41 (0.817)	92.23 $\pm$ 2.10 (0.896)	93.07 $\pm$ 1.30 (0.907)	93.51 $\pm$ 0.91 (0.913)	95.14 $\pm$ 0.85 (0.934)	95.34 $\pm$ 0.52 (0.937)
WSSGLDA [62]	87.33 $\pm$ 2.93 (0.834)	93.24 $\pm$ 1.37 (0.907)	94.09 $\pm$ 1.38 (0.921)	94.54 $\pm$ 1.26 (0.926)	95.45 $\pm$ 1.30 (0.938)	95.92 $\pm$ 0.60 (0.945)
SSMMFL	92.79 $\pm$ 2.10 (0.903)	94.95 $\pm$ 1.32 (0.932)	95.47 $\pm$ 0.94 (0.939)	95.71 $\pm$ 1.02 (0.942)	96.60 $\pm$ 0.65 (0.954)	97.30 $\pm$ 0.89 (0.959)

Table 3. Classification results with different values of n_i on the Washington DC Mall dataset with NN classifier. Each method has two rows, where the first row is the OA $\pm$ STD (%) and the second row is the KC.

Algorithm	10	20	30	40	50	60
RAW	84.70 $\pm$ 2.86 (0.810)	87.28 $\pm$ 1.45 (0.842)	88.05 $\pm$ 1.47 (0.851)	89.39 $\pm$ 1.02 (0.868)	90.40 $\pm$ 0.60 (0.880)	90.57 $\pm$ 0.86 (0.882)
PCA [56]	84.70 $\pm$ 2.86 (0.810)	87.28 $\pm$ 1.44 (0.842)	88.05 $\pm$ 1.46 (0.851)	89.37 $\pm$ 1.03 (0.868)	90.39 $\pm$ 0.59 (0.880)	90.55 $\pm$ 0.87 (0.882)
NPE [28]	86.84 $\pm$ 2.68 (0.836)	89.49 $\pm$ 1.69 (0.869)	90.57 $\pm$ 1.30 (0.883)	91.07 $\pm$ 1.33 (0.889)	91.81 $\pm$ 0.91 (0.898)	92.28 $\pm$ 0.99 (0.903)
LPP [29]	88.69 $\pm$ 2.46 (0.859)	90.21 $\pm$ 1.79 (0.878)	91.50 $\pm$ 1.19 (0.894)	92.03 $\pm$ 0.96 (0.901)	93.34 $\pm$ 0.91 (0.917)	93.83 $\pm$ 0.62 (0.923)
LDA [57]	85.67 $\pm$ 3.13 (0.822)	89.44 $\pm$ 1.32 (0.868)	89.44 $\pm$ 2.84 (0.868)	90.64 $\pm$ 0.97 (0.883)	91.63 $\pm$ 0.85 (0.896)	92.06 $\pm$ 0.69 (0.901)
MMC [58]	83.88 $\pm$ 3.50 (0.800)	87.12 $\pm$ 1.40 (0.840)	87.86 $\pm$ 1.43 (0.849)	89.19 $\pm$ 1.02 (0.865)	90.19 $\pm$ 0.61 (0.878)	90.23 $\pm$ 0.93 (0.878)
MFA [31]	88.18 $\pm$ 2.54 (0.853)	90.48 $\pm$ 1.62 (0.881)	90.54 $\pm$ 1.03 (0.882)	91.40 $\pm$ 0.98 (0.893)	92.58 $\pm$ 1.39 (0.907)	92.95 $\pm$ 1.04 (0.912)
NWFE [59]	87.98 $\pm$ 1.88 (0.850)	91.22 $\pm$ 2.28 (0.891)	91.48 $\pm$ 1.63 (0.894)	92.38 $\pm$ 1.79 (0.905)	92.68 $\pm$ 1.28 (0.908)	92.99 $\pm$ 1.37 (0.912)
SPP [29]	83.33 $\pm$ 2.67 (0.793)	83.53 $\pm$ 1.94 (0.796)	86.19 $\pm$ 2.06 (0.828)	88.99 $\pm$ 1.27 (0.863)	90.29 $\pm$ 1.12 (0.879)	91.57 $\pm$ 0.58 (0.895)
SDE [60]	85.17 $\pm$ 1.87 (0.815)	87.21 $\pm$ 2.53 (0.841)	88.10 $\pm$ 1.52 (0.852)	89.81 $\pm$ 1.02 (0.873)	90.06 $\pm$ 0.87 (0.876)	91.56 $\pm$ 0.39 (0.894)
SME [61]	86.84 $\pm$ 2.67 (0.836)	91.29 $\pm$ 2.19 (0.891)	92.18 $\pm$ 1.34 (0.902)	93.76 $\pm$ 0.64 (0.922)	94.29 $\pm$ 0.80 (0.929)	94.47 $\pm$ 0.34 (0.931)
SMML [37]	84.42 $\pm$ 3.09 (0.807)	87.13 $\pm$ 1.32 (0.840)	87.77 $\pm$ 1.39 (0.848)	89.04 $\pm$ 1.07 (0.863)	90.07 $\pm$ 0.56 (0.876)	90.23 $\pm$ 0.93 (0.878)
MMDA [33]	86.53 $\pm$ 3.91 (0.833)	89.77 $\pm$ 2.07 (0.873)	92.17 $\pm$ 0.97 (0.902)	93.91 $\pm$ 1.01 (0.924)	94.39 $\pm$ 0.75 (0.930)	95.30 $\pm$ 0.53 (0.941)
DSSM [49]	85.68 $\pm$ 3.32 (0.822)	87.29 $\pm$ 1.43 (0.842)	88.06 $\pm$ 1.45 (0.851)	89.35 $\pm$ 1.06 (0.867)	90.38 $\pm$ 0.59 (0.880)	90.58 $\pm$ 0.86 (0.882)
LPNPE [48]	86.99 $\pm$ 2.62 (0.838)	89.39 $\pm$ 1.01 (0.868)	90.17 $\pm$ 1.38 (0.878)	91.42 $\pm$ 0.80 (0.893)	92.09 $\pm$ 1.24 (0.901)	93.61 $\pm$ 0.96 (0.921)
WSSGLDA [62]	88.22 $\pm$ 1.96 (0.853)	90.14 $\pm$ 1.08 (0.877)	90.63 $\pm$ 1.99 (0.883)	91.43 $\pm$ 1.17 (0.893)	92.72 $\pm$ 1.30 (0.909)	93.98 $\pm$ 0.89 (0.925)
SSMMFL	90.93 $\pm$ 1.42 (0.887)	93.44 $\pm$ 1.38 (0.918)	94.66 $\pm$ 1.17 (0.933)	95.72 $\pm$ 0.89 (0.947)	96.33 $\pm$ 0.79 (0.954)	96.79 $\pm$ 0.65 (0.960)

Table 4. Classification results (%) of each class ($t = 0.5\%$) with NN classifier on the PaviaU data set.

Class	1	2	3	4	5	6	7	8	9	AA	OA	KC
RAW	89.21	94.05	41.31	70.42	99.48	55.76	79.09	74.26	99.79	78.15	82.64	0.765
PCA [56]	89.21	94.04	41.31	70.42	99.48	55.76	79.02	74.24	99.79	78.14	82.63	0.765
NPE [28]	90.09	97.35	65.53	83.11	99.63	70.38	91.52	92.96	99.79	87.84	90.04	0.866
LPP [29]	91.97	96.89	63.62	78.39	99.55	76.74	94.92	80.76	99.89	86.97	89.50	0.859
LDA [57]	89.54	93.05	67.88	81.99	99.03	81.35	70.15	78.03	97.44	84.27	87.38	0.833
MMC [58]	88.85	89.30	37.87	69.14	99.48	48.04	76.52	73.61	99.79	75.84	79.21	0.719
MFA [31]	93.24	98.51	74.25	83.83	98.28	83.91	94.85	87.23	99.89	90.44	92.67	0.902
NWFE [59]	93.94	95.38	62.09	75.14	99.63	65.11	84.47	88.18	99.89	84.88	87.79	0.835
SPP [29]	82.54	86.50	45.38	67.07	99.78	41.99	59.47	62.83	99.47	71.66	75.07	0.659
SDE [60]	90.71	96.50	51.27	74.06	99.55	59.43	85.91	77.48	99.79	81.63	85.62	0.805
SME [61]	91.63	98.38	72.47	83.14	99.33	83.31	94.85	86.30	99.47	89.87	92.10	0.894
SMML [37]	89.28	93.16	41.17	69.92	99.48	54.04	78.64	74.48	99.79	77.77	82.03	0.757
MMDA [33]	94.15	95.34	71.33	83.67	100.00	70.30	85.68	88.07	99.68	87.58	89.51	0.860
DSSM [49]	89.00	94.01	41.60	70.68	99.48	55.54	79.17	74.45	99.79	78.19	82.62	0.764
LPNPE [48]	92.51	99.59	60.89	78.42	97.68	98.24	80.30	81.74	99.89	87.70	92.73	0.903
WSSGLDA [62]	95.68	99.04	62.42	84.59	99.85	97.82	91.67	78.96	99.89	89.99	93.63	0.915
SSMMDA	97.83	99.43	63.52	85.18	99.93	94.32	97.80	93.53	99.89	92.38	95.26	0.937

Table 5. Classification results (%) of each class ($t = 0.1\%$) with NN classifier on the Heihe data set.

Class	1	2	3	4	5	6	7	8	AA	OA	KC
RAW	94.85	88.73	88.51	73.99	74.15	89.43	80.31	90.29	85.03	89.44	0.857
PCA [56]	94.85	88.73	88.51	73.99	74.15	89.43	80.31	90.29	85.03	89.44	0.857
NPE [28]	94.86	90.92	88.39	82.78	74.31	86.55	75.18	91.21	85.53	90.57	0.872
LPP [29]	94.90	90.92	91.72	87.35	65.51	81.50	84.82	90.64	85.92	91.26	0.882
LDA [57]	93.88	89.39	87.62	54.41	70.42	78.20	67.59	87.28	78.60	87.19	0.827
MMC [58]	94.76	88.55	88.43	71.85	73.77	88.71	78.77	90.29	84.39	89.14	0.853
MFA [31]	94.46	92.95	91.78	82.17	93.10	89.97	89.44	91.10	90.62	92.44	0.897
NWFE [59]	94.76	95.26	90.75	84.60	96.19	91.41	81.95	93.76	91.20	93.52	0.912
SPP [29]	87.32	89.60	85.24	35.96	54.37	63.66	55.38	75.95	68.43	81.87	0.756
SDE [60]	95.06	91.10	87.92	75.01	71.83	89.73	78.15	89.48	84.78	90.02	0.864
SME [61]	94.97	90.94	89.72	85.29	90.30	93.99	84.72	93.41	90.41	91.86	0.889
SMML [37]	94.83	88.31	88.42	73.73	73.77	89.55	79.59	89.71	84.74	89.26	0.855
MMDA [33]	95.64	94.54	92.89	97.43	95.66	66.13	76.72	95.49	89.31	94.29	0.922
DSSM [49]	94.83	88.75	88.48	73.91	74.23	89.49	79.69	90.64	85.00	89.42	0.857
LPNPE [48]	94.45	90.66	97.98	96.09	70.10	76.82	89.85	97.80	89.22	93.05	0.905
WSSGLDA [62]	92.91	93.64	98.89	97.53	68.44	75.80	86.67	98.15	89.00	93.44	0.911
SSMMDA	97.38	95.39	94.30	95.72	72.89	75.38	91.18	96.65	91.76	94.83	0.929

Table 6. Classification results (%) of each class ($t = 1\%$) with NN classifier on the Washington DC Mall data set.

Class	1	2	3	4	5	6	AA	OA	KC
RAW	92.85	96.09	87.01	97.13	79.15	78.01	88.37	89.75	0.872
PCA [56]	92.76	96.09	87.03	97.13	79.15	78.10	88.38	89.74	0.872
NPE [28]	92.03	96.27	87.53	96.81	79.26	83.55	89.24	90.25	0.878
LPP [29]	94.14	98.78	91.59	97.76	89.80	88.75	93.47	93.69	0.921
LDA [57]	73.98	95.29	59.12	84.94	55.57	66.26	72.53	72.25	0.656
MMC [58]	92.76	96.09	86.75	97.10	78.04	78.05	88.13	89.57	0.869
MFA [31]	93.55	96.76	89.68	95.21	85.79	79.83	90.13	91.00	0.887
NWFE [59]	93.91	96.88	93.81	97.13	89.69	84.99	92.73	93.36	0.916
SPP [29]	92.52	94.37	84.88	98.52	80.04	83.26	88.93	90.06	0.875
SDE [60]	92.40	95.72	87.41	97.25	77.20	81.31	88.54	89.92	0.873
SME [61]	93.74	98.10	91.00	97.44	91.58	87.99	93.31	93.40	0.917
SMML [37]	92.68	96.09	86.66	97.10	78.04	78.01	88.10	89.52	0.869
MMDA [33]	95.62	98.96	87.15	96.84	93.70	88.58	93.48	93.33	0.917
DSSM [49]	92.90	96.09	86.84	97.13	79.26	77.84	88.34	89.72	0.871
LPNPE [48]	95.48	98.53	88.80	95.89	94.59	79.11	92.07	92.37	0.905
WSSGLDA [62]	92.20	96.76	89.49	96.88	83.84	83.68	90.48	91.20	0.890
SSMMDA	96.13	99.27	96.98	96.88	92.36	89.98	95.27	95.65	0.946

Figure 8. Classification maps of different methods on the PaviaU data set. (a) False color image; (b) Ground truth; (c) Training samples; (d) RAW; (e) PCA; (f) NPE; (g) LPP; (h) LDA; (i) MMC; (j) MFA; (k) NWFE; (l) SPP; (m) SDE; (n) SME; (o) SMML; (p) MMDA; (q) DSSM; (r) LPNPE; (s) WSSGLDA; (t) SSMDA.

Figure 9. Classification maps of different methods on the Heihe data set. (a) False color image; (b) Ground truth; (c) Training samples; (d) RAW; (e) PCA; (f) NPE; (g) LPP; (h) LDA; (i) MMC; (j) MFA; (k) NWFE; (l) SPP; (m) SDE; (n) SME; (o) SMML; (p) MMDA; (q) DSSM; (r) LPNPE; (s) WSSGLDA; (t) SSMDA.

Figure 10. Classification maps of different methods on the Washington DC Mall data set. (a) False color image; (b) Ground truth; (c) Training samples; (d) RAW; (e) PCA; (f) NPE; (g) LPP; (h) LDA; (i) MMC; (j) MFA; (k) NWFE; (l) SPP; (m) SDE; (n) SME; (o) SMML; (p) MMDA; (q) DSSM; (r) LPNPE; (s) WSSGLDA; (t) SSMMDA.

5.3. Analysis of Computational Efficiency

To quantitatively compare the computational efficiency of each algorithm, we show the computational time of each algorithm in Table 7, where 30 labeled samples per class were selected for training, and the remaining samples were for testing. All the results in Table 7 were obtained on a personal computer, which has the i3-7100 CPU and 12-G memory. Besides, the version of Windows system and MATLAB are 64-bit Windows 10 and 2017a, respectively.

Table 7. Computational time (in seconds) of different algorithms on PaviaU, Heihe and Washington DC Mall data sets.

Data Set	RAW	PCA [56]	NPE [28]	LPP [29]	LDA [57]	MMC [58]	MFA [31]	NWFE [59]	SPP [29]
PaviaU	0.697	0.641	0.665	0.654	0.619	0.702	0.659	0.709	0.962
Heihe	1.662	1.366	1.393	1.348	1.314	2.133	1.356	1.021	1.191
Washington DC Mall	0.241	0.198	0.237	0.215	0.227	0.676	0.202	0.400	0.560
Data Set	SDE [60]	SME [61]	SMML [37]	MMDA [33]	DSSM [49]	LPNPE [48]	WSSGLDA [62]	SSMMDA	
PaviaU	0.661	3.988	0.662	0.867	1.283	1.498	2.164	3.723	
Heihe	0.965	3.346	1.368	1.707	2.253	3.359	3.626	5.429	
Washington DC Mall	0.443	3.146	0.231	0.734	0.552	1.587	2.774	3.541	

As shown in Table 7, the proposed SSMDA method requires more time in the classification process of HSI. The reason is that SSMDA simultaneously considers the spatial and spectral information and it needs to build two models in both the spatial domain and the spectral domain. Besides, Since SSMDA has c different mapping matrices, it needs to project the samples into c low-dimensional spaces when calculating the embedding features of samples, so the computational complexity of SSMDA is slightly increased. However, the slight increase in computational complexity is acceptable relative to the improvement for classification performance.

6. Conclusions

High-dimensional data can be divided into different subsets; each subset is located on a particular submanifold. Traditional feature learning methods can not effectively discover the multi-manifold structure in hyperspectral image, which restricts their classification performance of HSI. In this paper, we proposed a new DR method termed spatial-spectral multiple manifold discriminant analysis (SSMDA) for the classification on HSI data. SSMDA first divides the HSI data into several different subsets according to the label information of training samples. Then, it constructs an intramanifold graph for each submanifold to characterize the within-manifold similarity in spectral domain and designs an intramanifold scatter matrix and an intermanifold scatter matrix for each submanifold to characterize the within-manifold similarity and the between-manifold dissimilarity in the spatial domain. After that, a new spatial-spectral combined DR model is built for each submanifold to obtain an optimal projection, and the discriminative features on different submanifolds are fused to enhance the classification performance. Experiments on PaviaU, Heihe and Washington DC Mall hyperspectral data sets demonstrate that the proposed SSMDA method can obtain effective discriminative features and significantly improve the classification performance of HSI. In our future work, we will focus on how to obtain the optimum embedding dimension for each submanifold to further improve the classification performance of HSI.

Author Contributions: G.S. carried out the experiments and finished the first version of this manuscript. H.H. and J.L. was primarily responsible for mathematical modeling and experimental design. Z.L. and L.W. provided important suggestions to improve the paper.

Funding: This work was supported by the Basic and Frontier Research Programmes of Chongqing under Grant cstc2018jcyjAX0093, the Fundamental Research Funds for the Central Universities under Grant 2019CDYGYB008, and the graduate research and innovation foundation of Chongqing under Grants CYB19039 and CYS18035.

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. Li, S.T.; Hao, Q.B.; Kang, X.; Benediktsson, J.A. Gaussian Pyramid Based Multiscale Feature Fusion for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 3312–3324. [[CrossRef](#)]
2. Zhai, H.; Zhang, H.Y.; Zhang, L.P.; Li, P.X. Laplacian-Regularized Low-Rank Subspace Clustering for Hyperspectral Image Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 1723–1740. [[CrossRef](#)]
3. Luo, F.L.; Zhang, L.P.; Zhou, X.C.; Guo, T.; Cheng, Y.X.; Yin, T.L. Sparse-Adaptive Hypergraph Discriminant Analysis for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2019**, *99*, 1–5. [[CrossRef](#)]
4. Peng J.T.; Du Q. Robust joint sparse representation based on maximum correntropy criterion for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 7152–7164. [[CrossRef](#)]
5. Li, W.; Wu, G.D.; Du, Q. Transferred Deep Learning for Anomaly Detection in Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2017**, *14*, 597–601. [[CrossRef](#)]
6. Hu, F.; Xia G.S.; Wang, Z.F.; Huang, X.; Zhang, L.P.; Sun, H. Unsupervised feature learning via spectral clustering of patches for remotely sensed scene classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *8*, 2015–2030. [[CrossRef](#)]

7. Song, W.W.; Li, S.T.; Fang, L.Y.; Lu, T. Hyperspectral Image Classification with Deep Feature Fusion Network. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4544–4554. [[CrossRef](#)]
8. Zhai, H.; Zhang, H.Y.; Zhang, L.P.; Li, P.X. Total Variation Regularized Collaborative Representation Clustering with a Locally Adaptive Dictionary for Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 166–180. [[CrossRef](#)]
9. Yuan, Y.; Zheng, X.T.; Lu, X.Q. Discovering diverse subset for unsupervised hyperspectral band selection. *IEEE Trans. Image Process.* **2017**, *26*, 51–64. [[CrossRef](#)]
10. Jia, S.; Tang, G.H.; Zhu, J.S.; Li, Q.Q. A Novel Ranking-Based Clustering Approach for Hyperspectral Band Selection. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 88–102. [[CrossRef](#)]
11. Dong, Y.N.; Du, B.; Zhang, L.P.; Zhang, L.F. Dimensionality reduction and classification of hyperspectral images using ensemble discriminative local metric learning. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2509–2524. [[CrossRef](#)]
12. Luo, Y.; Wen, Y.G.; Tao, D.C.; Gui, J.; Xu, C. Large Margin Multi-Modal Multi-Task Feature Extraction for Image Classification. *IEEE Trans. Image Process.* **2016**, *25*, 414–427. [[CrossRef](#)]
13. Hang, R.L.; Liu, Q.S.; Sun, Y.B.; Yuan, X.T.; Pei, H.C.; 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](#)]
14. Li, Q.; Ji, H.B. Multimodality image registration using local linear embedding and hybrid entropy. *Neurocomputing* **2013**, *111*, 34–42. [[CrossRef](#)]
15. Tu, S.T.; Chen, J.Y.; Yang, W.; Sun, H. Laplacian Eigenmaps-Based Polarimetric Dimensionality Reduction for SAR Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 170–179. [[CrossRef](#)]
16. Li, W.; Zhang, L.P.; Zhang, L.F.; Du, B. GPU Parallel Implementation of Isometric Mapping for Hyperspectral Classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1532–1536. [[CrossRef](#)]
17. Chen, J.; Liu, Y. Locally linear embedding: A survey. *Artif. Intell. Rev.* **2011**, *36*, 29–48. [[CrossRef](#)]
18. Belkin, M.; Niyogi, P. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Comput.* **2003**, *15*, 1373–1396. [[CrossRef](#)]
19. Feilhauer, H.; Faude, U.; Schmidtlein, S. Combining Isomap ordination and imaging spectroscopy to map continuous floristic gradients in a heterogeneous landscape. *Remote Sens. Environ.* **2011**, *115*, 2513–2524. [[CrossRef](#)]
20. Sun, W.W.; Yang, G.; Du, B.; Zhang, L.F.; Zhang, L.P. A sparse and low-rank near-isometric linear embedding method for feature extraction in hyperspectral imagery classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 4032–4046. [[CrossRef](#)]
21. Huang, H.; Li, Z.Y.; Pan, Y.S. Multi-Feature Manifold Discriminant Analysis for Hyperspectral Image Classification. *Remote Sens.* **2019**, *11*, 651. [[CrossRef](#)]
22. Fang, L.Y.; Wang, C.; Li, S.T.; Benediktsson, J.A. Hyperspectral image classification via multiple-feature-based adaptive sparse representation. *IEEE Trans. Instrum. Meas.* **2017**, *7*, 1646–1657. [[CrossRef](#)]
23. Wang, Q.; Wan, J.; Yuan, Y. Locality Constraint Distance Metric Learning for Traffic Congestion Detection. *Pattern Recognit.* **2018**, *9*, 272–281. [[CrossRef](#)]
24. 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. [[CrossRef](#)]
25. Gui, J.; Sun, Z.N.; Jia, W.; Hu, R.X.; Lei, Y.K.; Ji, S.W. Discriminant sparse neighborhood preserving embedding for face recognition. *Pattern Recognit.* **2012**, *45*, 2884–2893. [[CrossRef](#)]
26. Wong, W.K.; Zhao, H.T. Supervised optimal locality preserving projection. *Pattern Recognit.* **2012**, *45*, 186–197. [[CrossRef](#)]
27. Zhang, X.R.; He, Y.D.; Zhou, N.; Zheng, Y.G. Semisupervised Dimensionality Reduction of Hyperspectral Images via Local Scaling Cut Criterion. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1547–1551. [[CrossRef](#)]
28. Lu, G.F.; Jin, Z.; Zou, J. Face recognition using discriminant sparsity neighborhood preserving embedding. *Knowl.-Based Syst.* **2012**, *31*, 119–127. [[CrossRef](#)]
29. Qiao, L.S.; Chen, S.C.; Tan, X.Y. Sparsity preserving projections with applications to face recognition. *Pattern Recognit.* **2010**, *43*, 331–341. [[CrossRef](#)]
30. Mohanty, R.; Happy, S.L.; Routray, A. A Semisupervised Spatial Spectral Regularized Manifold Local Scaling Cut With HGF for Dimensionality Reduction of Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 3423–3435. [[CrossRef](#)]

31. Pang, Y.W.; Ji, Z.; Jing, P.G.; Li, X.L. Ranking Graph Embedding for Learning to Rerank. *IEEE Trans. Neural Netw. Learn. Syst.* **2013**, *24*, 1292–1303. [[CrossRef](#)] [[PubMed](#)]
32. 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](#)]
33. Yang, W.K.; Sun, C.Y.; Zhang, L. A multi-manifold discriminant analysis method for image feature extraction. *Pattern Recognit.* **2011**, *44*, 1649–1657. [[CrossRef](#)]
34. Hettiarachchi, R.; Peters, J.F. Multi-manifold LLE learning in pattern recognition. *Pattern Recognit.* **2015**, *48*, 2947–2960. [[CrossRef](#)]
35. Jiang, J.J.; Hu, R.M.; Wang, Z.Y.; Cai, Z.H. CDMMA: Coupled discriminant multi-manifold analysis for matching low-resolution face images. *Signal Process.* **2016**, *124*, 162–172. [[CrossRef](#)]
36. Chu, Y.J.; Zhao, L.D.; Ahmad, T. Multiple feature subspaces analysis for single sample per person face recognition. *Vis. Comput.* **2019**, *35*, 239–256. [[CrossRef](#)]
37. Shi, L.K.; Hao, J.S.; Zhang, X. Image recognition method based on supervised multi-manifold learning. *J. Intell. Fuzzy Syst.* **2017**, *32*, 2221–2232. [[CrossRef](#)]
38. Jiang, J.J.; Ma, J.Y.; Chen, C.; Wang, Z.Y.; Cai, Z.H.; Wang, L.Z. SuperPCA: A Superpixelwise PCA Approach for Unsupervised Feature Extraction of Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4581–4593. [[CrossRef](#)]
39. Tu, B.; Zhang, X.F.; Kang, X.D.; Wang, J.P.; Benediktsson, J.A. Spatial Density Peak Clustering for Hyperspectral Image Classification with Noisy Labels. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 5085–5097. [[CrossRef](#)]
40. Fang, L.Y.; Zhuo, H.J.; Li, S.T. Super-resolution of hyperspectral image via superpixel-based sparse representation. *Neurocomputing* **2018**, *273*, 171–177. [[CrossRef](#)]
41. Zhao, W.Z.; Du, S.H. 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](#)]
42. He, L.; Li, J.; Liu, C.Y.; Li, S.T. Recent advances on spectral-spatial hyperspectral image classification: An overview and new guidelines. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 1579–1597. [[CrossRef](#)]
43. Zhang, L.F.; Zhang, Q.; Du, B.; Huang, X.; Tang, Y.Y.; Tao, D.C. Simultaneous Spectral-Spatial Feature Selection and Extraction for Hyperspectral Images. *IEEE Trans. Cybern.* **2018**, *48*, 16–28. [[CrossRef](#)] [[PubMed](#)]
44. 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](#)]
45. Kang, X.D.; Li, S.T.; Benediktsson, J.A. Feature extraction of hyperspectral images with image fusion and recursive filtering. *IEEE Geosci. Remote Sens. Lett.* **2014**, *52*, 3742–3752. [[CrossRef](#)]
46. Huang, H.; Shi, G.Y.; He, H.B.; Duan, Y.L.; Luo, F.L. Dimensionality Reduction of Hyperspectral Imagery Based on Spatial-spectral Manifold Learning. *IEEE Trans. Cybern.* **2019**, 1–14. [[CrossRef](#)]
47. 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](#)]
48. Zhou, Y.C.; Peng, J.T.; Chen, C.L.P. Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1082–1095. [[CrossRef](#)]
49. 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](#)]
50. Mohanty, R.; Happy, S.L.; Routray, A. Spatial-Spectral Regularized Local Scaling Cut for Dimensionality Reduction in Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 932–936. [[CrossRef](#)]
51. Sellami, A.; Farah, I.R. High-level hyperspectral image classification based on spectro-spatial dimensionality reduction. *Spat. Stat.* **2016**, *16*, 103–117. [[CrossRef](#)]
52. Huang, H.; Chen, M.L.; Duan, Y.L. Dimensionality Reduction of Hyperspectral Image Using Spatial-Spectral Regularized Sparse Hypergraph Embedding. *Remote Sens.* **2019**, *11*, 1039. [[CrossRef](#)]
53. Xiao, Q.; Wen, J.G. HiWATER: Thermal-Infrared Hyperspectral Radiometer (4th, July, 2012). *Heihe Plan Sci. Data Center* **2013**. [[CrossRef](#)]

54. Xue, Z.H.; Su, H.J.; Du, P.J. Sparse graph regularization for robust crop mapping using hyperspectral remotely sensed imagery: A case study in Heihe Zhangye oasis. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 779–782. [\[CrossRef\]](#)
55. Li, W.; Du, Q.; Zhang, F.; Hu, W. Collaborative-representation-based nearest neighbor classifier for hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 389–393. [\[CrossRef\]](#)
56. Datta, A.; Ghosh, S.; Ghosh, A. Unsupervised band extraction for hyperspectral images using clustering and kernel principal component analysis. *Int. J. Remote Sens.* **2017**, *38*, 850–873. [\[CrossRef\]](#)
57. Liao, W.Z.; Pizurica, A.; Scheunders, P.; Philips, W.; Pi, Y.G. Semisupervised Local Discriminant Analysis for Feature Extraction in Hyperspectral Images. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 184–198. [\[CrossRef\]](#)
58. Li, H.F.; Jiang, T.; Zhang, K.S. Efficient and robust feature extraction by maximum margin criterion. *IEEE Trans. Neural Netw.* **2006**, *17*, 157–165. [\[CrossRef\]](#)
59. Kuo, B.C.; Landgrebe, D.A. Nonparametric weighted feature extraction for classification. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1096–1105. [\[CrossRef\]](#)
60. Huang, H. Classification of hyperspectral remote-sensing images based on sparse manifold learning. *J. Appl. Remote Sens.* **2013**, *7*, 073464. [\[CrossRef\]](#)
61. Huang, H.; Luo, F.L.; Liu, J.M.; Yang, Y.Q. Dimensionality reduction of hyperspectral images based on sparse discriminant manifold embedding. *ISPRS-J. Photogramm. Remote Sens.* **2015**, *106*, 42–54. [\[CrossRef\]](#)
62. Zeng, S.; Wang, Z.Y.; Gao, C.J.; Kang, Z.; Feng, D.G. Hyperspectral Image Classification With Global-Local Discriminant Analysis and Spatial-Spectral Context. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 5005–5018. [\[CrossRef\]](#)

© 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/>).