

Article

Multi-Feature Manifold Discriminant Analysis for Hyperspectral Image Classification

Hong Huang , Zhengying Li and Yinsong Pan

The Key Laboratory on Opto-Electronic Technique and Systems, Ministry of Education, Chongqing University, Chongqing 400044, China; zhengying_li@cqu.edu.cn (Z.L.); panys@cqu.edu.cn (Y.P.)

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

Received: 18 January 2019; Accepted: 13 March 2019; Published: 17 March 2019



Abstract: Hyperspectral image (HSI) provides both spatial structure and spectral information for classification, but many traditional methods simply concatenate spatial features and spectral features together that usually lead to the curse-of-dimensionality and unbalanced representation of different features. To address this issue, a new dimensionality reduction (DR) method, termed multi-feature manifold discriminant analysis (MFMDA), was proposed in this paper. At first, MFMDA explores local binary patterns (LBP) operator to extract textural features for encoding the spatial information in HSI. Then, under graph embedding framework, the intrinsic and penalty graphs of LBP and spectral features are constructed to explore the discriminant manifold structure in both spatial and spectral domains, respectively. After that, a new spatial-spectral DR model for multi-feature fusion is built to extract discriminant spatial-spectral combined features, and it not only preserves the similarity relationship between spectral features and LBP features but also possesses strong discriminating ability in the low-dimensional embedding space. Experiments on Indian Pines, Heihe and Pavia University (PaviaU) hyperspectral data sets demonstrate that the proposed MFMDA method performs significantly better than some state-of-the-art methods using only single feature or simply stacking spectral features and spatial features together, and the classification accuracies of it can reach 95.43%, 97.19% and 96.60%, respectively.

Keywords: hyperspectral image; multi-feature classification; dimensionality reduction; graph embedding; spatial-spectral features

1. Introduction

Hyperspectral imagery (HSI) provides hundreds of narrow and continuous adjacent bands through dense spectral sampling from visible to short-wave infrared regions [1–8]. A fine-spectral-resolution HSI provides useful information for classifying different types of ground objects, and it has a variety of applications in many fields such as mineral exploration, environmental monitoring, precision agriculture, and target recognition [9–13]. Classification of each pixel in HSI plays a crucial role in these real applications, but complex spectral characteristics within HSI data pose huge challenges to the traditional spectral feature-based HSI classification [14–19].

Recent investigations have demonstrated that combining spatial and spectral information is beneficial to the feature extraction and classification of HSI data [20–28]. In recent years, many effective spatial-based features have been proposed by concerning structure, shape, texture, geometric, etc. Li et al. [29] extracted textural features of HSI using LBP operator and then classified them through extreme learning machine (ELM). Mauro et al. [30] designed an extended multi-attribute profiles (EMAP) algorithm to explore morphological features from HSI, and the extracted features were classified by a random forest classifier. Li et al. [31] introduced generalized composite kernel machines to explore spatial information through EMAP, then used the multinomial logistic regression for

classification. However, in real applications, it is impossible to find a single feature that is suitable for different image scenes due to the variety and irregular distribution of ground objects. The conventional method for addressing this issue is to explore a feature stacking (FS) approach for the combination of different types of features. Li et al. [32] tried to get combined features by fusing spectral features and EMAP features which improved the classification accuracy of HSI. Song et al. [33] used LBP operator for extracting textural features, and then stacked spectral features and textural features for classification. However, the feature stacking method commonly poses the problem of the curse-of-dimensionality for the increase in the dimension of stacked features, and thus such methods do not necessarily ensure better performance for HSI classification. Therefore, an urgent challenge in multi-feature classification of HSI data is how to reduce the dimension of spatial and spectral combined features largely with some valuable intrinsic information preserved [34].

To solve this problem, many DR methods have been proposed to reduce the number of bands and obtain some desired information in HSI [35–38]. Principal component analysis (PCA) and Linear Discriminant Analysis (LDA) are two classical DR methods [39,40]. However, the two subspace methods cannot analyze the data that lies on or near a submanifold embedded in the original space. Therefore, the graph-based manifold learning methods have attracted wide attention recently [41]. Such methods include isometric mapping (Isomap), Laplacian eigenmaps (LE), locality preserving projections (LPP), locally linear embedding (LLE), neighborhood preserving embedding (NPE), and local tangent space alignment (LTSA) [42–47]. These graph embedding (GE) methods are unsupervised learning methods without using the discriminant information in training samples. Some supervised learning methods were designed to explore the label information of training data to enhance the discriminating ability for classification, such as marginal Fisher analysis (MFA), locality sensitive discriminant analysis (LSDA), coupled discriminant multi-manifold analysis (CDMMA), and local geometric structure Fisher analysis (LGSFA) [48–51]. However, the above DR methods only make use of spectral features in HSI, while it is commonly accepted that exploiting multiple features, spectral, texture and shape features, brings significant benefits in terms of improving the classification performance.

To explore DR of multiple features for HSI classification, Fauvel et al. [52] used PCA to reduce the dimension of EMAP features and stacked them with spectral features to form fused feature vectors. Huo et al. [53] selected the first three PCs of HSI to extract Gabor textures, then concatenated Gabor textures and spectral features from the same pixel to form the combined feature for classification. However, the above multi-feature-based methods simply stacked the reduced spectral and spatial features together after applying DR on the different types of features, respectively. The embedding features are obtained in different subspaces that cannot ensure global optimization. Furthermore, the direct stacking strategy may lead to unbalanced representation of different features.

To overcome above drawbacks, we propose a novel DR algorithm termed multi-feature manifold discriminant analysis for HSI data. The MFMDA method first exploits the spatial information in HSI by extracting LBP textural features. Then it constructs the intrinsic graphs and penalty graphs of spectral features and LBP features within GE framework, which can effectively discover the manifold structure of spatial features and spectral features. After that, MFMDA learns low-dimensional embedding space from original spectral features as well as LBP features for compacting the intramanifold samples while separating intermanifold samples, which will increase the margins between different manifolds. As a result, the spatial-spectral embedding features possess stronger discriminating ability for HSI classification. Experimental results on three real hyperspectral data sets show that the proposed MFMDA algorithm can significantly improve the classification accuracy compared with some state-of-art DR methods, especially in the case of limited training samples are available.

The remainder of this paper is organized as follows. In Section 2, we briefly introduce the spectral features, textural features, and the GE framework. The details of our algorithm are introduced in Section 3. Section 4 gives experimental results to demonstrate the effectiveness of our algorithm. We give some concluding remarks and suggestions for further work in Section 5.

2. Related Works

2.1. Spectral and LBP Features of HSI

Spectral and textural information are the fundamental properties of hyperspectral imagery. Spectral information provides densely sampled reflectance values over a wide range of the electro-magnetic spectrum to distinguish similar materials, while texture is a typical spatial feature which gives a description of the homogeneity of an image using the texture element as the fundamental unit. Recent studies show that combining spatial context into pixel-based spectral classification can substantially improve the classification performance of HSI [54].

Local binary pattern is a discriminative and computationally local texture descriptor that has shown promising performance in classification. The original LBP operator represents the pixels of an image with binary numbers called LBP codes, which encode the local structure around each pixel, and then the codes are used for further analysis [29]. The procedure of it is shown in Figure 1, where the 10th band of PaviaU hyperspectral image is used to extract LBP features. As in Figure 1, for a given center pixel in a 3×3 window, the neighbor pixels are assigned with binary labels ("0" or "1"), depending on whether the gray value of center pixel is larger or not. An 8-digit binary number can be obtained by concatenating all these binary codes in a clockwise direction starting from the top-left one, and the derived binary numbers are referred to as LBP code.

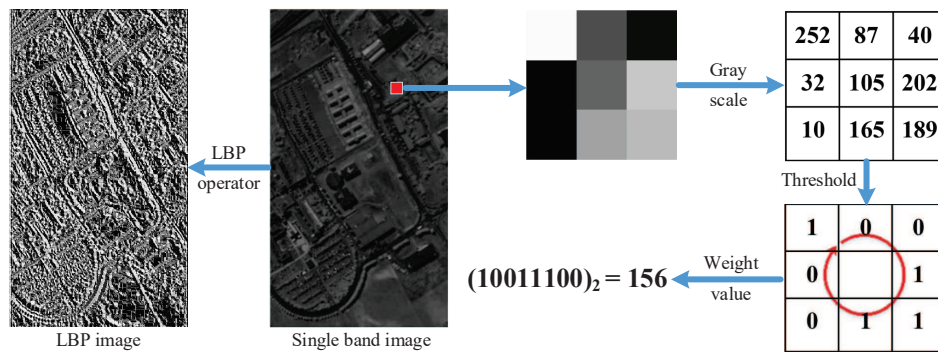


Figure 1. The procedure of local binary patterns (LBP) operator on the PaviaU image.

According to the aforementioned analysis, spectral features and LBP features represent the information in HSI from different perspectives. Spectral features provide continuous spectral measurement across the entire electromagnetic spectrum, while LBP features present a better expression of detailed local spatial features, such as edges, corners, and knots. Thus, it is promising to apply LBP features as a supplement to spectral features that lack the consideration of spatial relations between pixels in HSI. However, both spectral features and LBP features are characterized by high dimensionality. A common approach to address the problem is to explore DR methods which will reduce the dimension of high-dimensional features largely without loss of information.

2.2. Graph Embedding

The GE framework is explored to unify many classical DR algorithms such as PCA, LDA, ISOMAP, LLE, LE, LPP and NPE. In GE, an intrinsic graph is constructed to characterize the statistical or geometrical properties that need to be preserved, and a penalty graph is explored to describe some properties which should be avoided. The intrinsic graph $G_I(X, W^w)$ and the penalty graph $G_P(X, W^b)$ are both undirected weighted graphs, where X is the vertex set of graph, $W^w \in \mathbb{R}^{n \times n}$ and $W^b \in \mathbb{R}^{n \times n}$ are the weight matrices of G_I and G_P , respectively. w_{ij}^w indicates the similarity between vertices x_i and x_j in G_I , while w_{ij}^b measures the dissimilarity of vertices x_i and x_j in G_P . Under this framework, MFA has been proposed for dimensionality reduction of high-dimensional data. In MFA, G_I connects each point with its neighbors from the same class to represent intraclass compactness, and G_P connects the

neighbor points which from different classes to represent the interclass separability. In low-dimensional embedding space, the intraclass compactness and interclass separability should be enhanced. Therefore, the optimal projection matrix V can be obtained with the following optimization problem:

$$J(V) = \arg \min \frac{\sum_{i,j} \|V^T x_i - V^T x_j\|^2 w_{ij}^w}{\sum_{i,j} \|V^T x_i - V^T x_j\|^2 w_{ij}^b} = \arg \min \frac{\text{tr}(V^T X L X^T V)}{\text{tr}(V^T X L^P X^T V)} \quad (1)$$

where $L = D^w - W^w$ is the Laplacian matrix of graph G_L , $W^w = [w_{ij}^w]_{i,j=1}^n$, D^w is a diagonal matrix, $D^w = [D_{ij}^w]_{i,j=1}^n$, $D_{ii}^w = \sum_{j=1}^n w_{ij}^w$, and $L^P = D^b - W^b$ is the Laplacian matrix of graph G_P , $W^b = [w_{ij}^b]_{i,j=1}^n$, $D^b = [D_{ij}^b]_{i,j=1}^n$, $D_{ii}^b = \sum_{j=1}^n w_{ij}^b$.

3. Proposed Approach

Let us suppose that a hyperspectral data set $X = [x_1, x_2, x_3, \dots, x_N] \in \mathbb{R}^{D \times N}$, where D is the number of bands and N indicates the number of pixels in HSI data. $X^S = \{x_i^s\}_{i=1}^N$ and $X^L = \{x_i^l\}_{i=1}^N$ denote the spectral features and LBP features of X , respectively. The class label of x_i is indicated by $\ell(x_i) = \{1, 2, \dots, c\}$, where c is the number of classes. The purpose of DR is to find a low-dimensional embedding space $Y = [y_1, y_2, y_3, \dots, y_N] \in \mathbb{R}^{d \times N}$, where d ($d \ll D$) is the embedding dimensionality of extracted features.

3.1. Motivation

Since different types of features represent HSI data from different perspectives, multiple feature fusion will bring benefits to enhance the discrimination capability for classification. The most common way to combine these features is to simply concatenate different types of features together, and then a classifier is employed to classify the stacked features. However, such stack-based methods have witnessed limited performance due to the simple strategy, and they may even perform worse than using a single feature in HSI data. The reasons for this phenomenon are summarized as follows:

- Simply stacking spatial and spectral features may yield redundant information, and it remains difficult to achieve an optimal combination for different kinds of features;
- The spatial information and spectral information is not equally represented by simply stacking;
- The stacked features greatly increase the dimensionality of spatial-spectral combined features, this will make HSI classification fairly challenging for the curse-of-dimensionality problem, especially when only limited training samples are available.

Many DR methods have been explored to reduce the dimension of stacked features. However, different types of features usually lie on different manifolds. Performing dimensionality reduction directly on the simply stacked features cannot reveal the manifold structure of different features in HSI. As a result, the discriminant information contained by different features is not effectively represented, which will restrict their discriminant capability for classification.

To overcome the shortcomings as discussed above, a new DR method called MFMDA is introduced in next section. By exploring the manifold structures of different features, it can effectively extract the spatial-spectral combined features and subsequently improve the classification performance of HSI.

3.2. MFMDA

The goal of the proposed MFMDA method is to find an optimized projection matrix which can couple dimensionality reduction and data fusion of original features (from HSI data) and spatial features (LBP features generated from HSI) based on GE framework. MFMDA simultaneously learns a low-dimensional embedding space from original spectral features as well as LBP features for

compacting the intramanifold samples while separating intermanifold samples, which will increase the margins between different manifolds. As a result, the obtained embedding features possess stronger discriminating ability that helps to subsequent classification. The flowchart of MFMDA is shown in Figure 2.

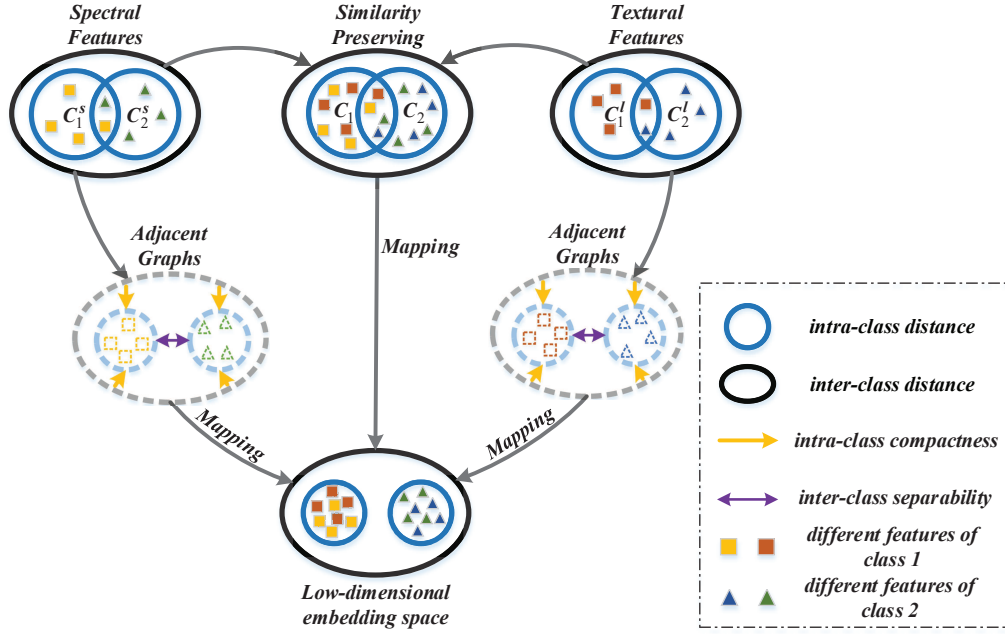


Figure 2. The flowchart of multi-feature manifold discriminant analysis.

As illustrated in Figure 2, due to the fact that the similarity relationship between spectral features and LBP features from the same pixel should be preserved in the low-dimensional embedding space. Let us assume $A_S \in \mathbb{R}^{D \times d}$ and $A_L \in \mathbb{R}^{D \times d}$ are the corresponding projection matrices of spectral features and LBP features, respectively. A_S and A_L should be explored to minimize the distance between the two embedding features from the same pixel, and the objective function can be defined as follows:

$$J_1(A_S, A_L) = \min \sum_{i=1}^N \|A_S^T x_i^s - A_L^T x_i^l\|^2 \quad (2)$$

With some mathematical operations, Equation (2) can be reduced as:

$$\begin{aligned} J_1(A_S, A_L) &= \min \sum_{i=1}^N \|A_S^T x_i^s - A_L^T x_i^l\|^2 = \text{tr} \left(\sum_{i=1}^N (A_S^T x_i^s - A_L^T x_i^l)(A_S^T x_i^s - A_L^T x_i^l)^T \right) \\ &= \text{tr} \left(\sum_{i=1}^N (A_S^T x_i^s (x_i^s)^T A_S - A_L^T x_i^l (x_i^l)^T A_S - A_S^T x_i^s (x_i^l)^T A_L + A_L^T x_i^l (x_i^l)^T A_L) \right) \\ &= \text{tr} \left(\begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}^T \begin{bmatrix} (X^S)^T & 0 \\ 0 & (X^L)^T \end{bmatrix} \begin{bmatrix} X^S & 0 \\ 0 & X^L \end{bmatrix} L_1 \begin{bmatrix} (X^S)^T & 0 \\ 0 & (X^L)^T \end{bmatrix} \begin{bmatrix} X^S & 0 \\ 0 & X^L \end{bmatrix} \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix} \right) \\ &= \text{tr}(A^T E L_1 E^T A) \end{aligned} \quad (3)$$

where A_S and A_L are respectively parameterized as $A_S = X^S B$ and $A_L = X^L C$, B and C are projection matrices that map spectral information and texture information in high-dimensional features to the low-dimensional embedded space, respectively. $A = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}$, $E = \begin{bmatrix} (X^S)^T X^S & 0 \\ 0 & (X^L)^T X^L \end{bmatrix}$,

$L_1 = \begin{bmatrix} I & -I \\ -I & I \end{bmatrix}$, I is the identity matrix in L .

From the view point of classification, in the low-dimensional embedding space, we expect that the samples are as close as possible if they belong to the same manifold, while samples are as far as possible if they are from different manifolds. To achieve this goal, we define the objective function as follows:

$$J_2(A_S, A_L) = \min \left(\sum_{i=1}^N \sum_{j=1}^N \left\| A_S^T x_i^s - A_S^T x_j^s \right\|^2 w_{ij}^{w_s} + \sum_{i=1}^N \sum_{j=1}^N \left\| A_L^T x_i^l - A_L^T x_j^l \right\|^2 w_{ij}^{w_l} \right) \quad (4)$$

$$J_3(A_S, A_L) = \max \left(\sum_{i=1}^N \sum_{j=1}^N \left\| A_S^T x_i^s - A_S^T x_j^s \right\|^2 w_{ij}^{b_s} + \sum_{i=1}^N \sum_{j=1}^N \left\| A_L^T x_i^l - A_L^T x_j^l \right\|^2 w_{ij}^{b_l} \right) \quad (5)$$

where $w_{ij}^{w_s}$ and $w_{ij}^{b_s}$ are the affinity weights to characterize the similarity between spectral features x_i^s and x_j^s of intrinsic graph G_I^S as well as the dissimilarity between x_i^s and x_j^s of the penalty graph G_P^S , $w_{ij}^{w_l}$ and $w_{ij}^{b_l}$ are the affinity weights to characterize the similarity between LBP features x_i^l and x_j^l of the intrinsic graph G_I^L and the dissimilarity between x_i^l and x_j^l of the penalty graph G_P^L , respectively.

In the intrinsic graph G_I^S of spectral features, the vertices x_i^s and x_j^s are connected by an edge if $l(x_i) = l(x_j)$ and they are close to each other in terms of some distance. When it comes to the penalty graph G_P^S , the vertices x_i^s and x_j^s are connected by an edge if $l(x_i) \neq l(x_j)$ and x_j^s belongs the k_b nearest neighbors of x_i^s . The weights $w_{ij}^{w_s}$ and $w_{ij}^{b_s}$ in two spectral-based graphs are defined as:

$$w_{ij}^{w_s} = \begin{cases} \exp\left(-\frac{\|x_i^s - x_j^s\|^2}{2(t_i^s)^2}\right), & x_i^s \in N_{s,w}(x_j^s) \text{ or } x_j^s \in N_{s,w}(x_i^s) \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

$$w_{ij}^{b_s} = \begin{cases} \exp\left(-\frac{\|x_i^s - x_j^s\|^2}{2(t_i^s)^2}\right), & x_i^s \in N_{s,b}(x_j^s) \text{ or } x_j^s \in N_{s,b}(x_i^s) \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

where $N_{s,w}(x_i^s)$ is the n_w -intranifold neighbors of spectral feature x_i^s , $N_{s,b}(x_i^s)$ indicates the n_b -intermanifold neighbors of x_i^s , and $t_i^s = \frac{1}{n} \sum_{j=1}^n \|x_i^s - x_j^s\|$.

In the intrinsic graph G_I^L of LBP features, an edge is added between the vertices x_i^l and x_j^l if $l(x_i) = l(x_j)$ and x_j^l belongs the k_w nearest neighbors of x_i^l ; in the penalty graph G_P^L , an edge is connected by x_i^l and x_j^l if $l(x_i) \neq l(x_j)$ and x_j^l belongs the k_b nearest neighbors of x_i^l . The weights $w_{ij}^{w_l}$ and $w_{ij}^{b_l}$ in two LBP-based graphs can be set as:

$$w_{ij}^{w_l} = \begin{cases} \exp\left(-\frac{\|x_i^l - x_j^l\|^2}{2(t_i^l)^2}\right), & x_i^l \in N_{l,w}(x_j^l) \text{ or } x_j^l \in N_{l,w}(x_i^l) \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

$$w_{ij}^{b_l} = \begin{cases} \exp\left(-\frac{\|x_i^l - x_j^l\|^2}{2(t_i^l)^2}\right), & x_i^l \in N_{l,b}(x_j^l) \text{ or } x_j^l \in N_{l,b}(x_i^l) \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where $N_{l,w}(x_i^l)$ is the n_w -intranifold neighbors of spectral feature x_i^l , $N_{l,b}(x_i^l)$ indicates the n_b -intermanifold neighbors of x_i^l , and $t_i^l = \frac{1}{n} \sum_{j=1}^n \|x_i^l - x_j^l\|$.

The objective function of $J_2(A_S, A_L)$ in Equation (4) is to minimize the intramanifold distance to ensuring the samples from the same manifold should be as close as possible, and the objective function

of $J_3(A_S, A_L)$ in Equation (5) is to maximize the intermanifold distance for enlarging the manifold margins in the low-dimensional embedding space.

With some mathematical operations, Equations (4) and (5) can be reduced as:

$$\begin{aligned} & \sum_{i=1}^N \sum_{j=1}^N \|A_S^T x_i^s - A_S^T x_j^s\|^2 w_{ij}^{w_s} + \sum_{i=1}^N \sum_{j=1}^N \|A_L^T x_i^l - A_L^T x_j^l\|^2 w_{ij}^{w_l} \\ &= \text{tr} \left(\sum_{i=1}^N \sum_{j=1}^N (A_S^T x_i^s w_{ij}^{w_s} (x_i^s)^T A_S - 2A_S^T x_i^s w_{ij}^{w_s} (x_j^s)^T A_S + A_S^T x_j^s w_{ij}^{w_s} (x_i^s)^T A_S) \right. \\ & \quad \left. + \sum_{i=1}^N \sum_{j=1}^N (A_L^T x_i^l w_{ij}^{w_l} (x_i^l)^T A_L - 2A_L^T x_i^l w_{ij}^{w_l} (x_j^l)^T A_L + A_L^T x_j^l w_{ij}^{w_l} (x_i^l)^T A_L) \right) \\ &= \text{tr} \left(\begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}^T \begin{bmatrix} (X^S)^T & 0 \\ 0 & (X^L)^T \end{bmatrix} \begin{bmatrix} X^S & 0 \\ 0 & X^L \end{bmatrix} L_2 \begin{bmatrix} (X^S)^T & 0 \\ 0 & (X^L)^T \end{bmatrix} \begin{bmatrix} X^S & 0 \\ 0 & X^L \end{bmatrix} \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix} \right) \\ &= \text{tr}(A^T E L_2 E^T A) \end{aligned} \quad (10)$$

$$\begin{aligned} & \sum_{i=1}^N \sum_{j=1}^N \|A_S^T x_i^s - A_S^T x_j^s\|^2 w_{ij}^{b_s} + \sum_{i=1}^N \sum_{j=1}^N \|A_L^T x_i^l - A_L^T x_j^l\|^2 w_{ij}^{b_l} \\ &= \text{tr} \left(\sum_{i=1}^N \sum_{j=1}^N (A_S^T x_i^s w_{ij}^{b_s} (x_i^s)^T A_S - 2A_S^T x_i^s w_{ij}^{b_s} (x_j^s)^T A_S + A_S^T x_j^s w_{ij}^{b_s} (x_i^s)^T A_S) \right. \\ & \quad \left. + \sum_{i=1}^N \sum_{j=1}^N (A_L^T x_i^l w_{ij}^{b_l} (x_i^l)^T A_L - 2A_L^T x_i^l w_{ij}^{b_l} (x_j^l)^T A_L + A_L^T x_j^l w_{ij}^{b_l} (x_i^l)^T A_L) \right) \\ &= \text{tr} \left(\begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}^T \begin{bmatrix} (X^S)^T & 0 \\ 0 & (X^L)^T \end{bmatrix} \begin{bmatrix} X^S & 0 \\ 0 & X^L \end{bmatrix} L_3 \begin{bmatrix} (X^S)^T & 0 \\ 0 & (X^L)^T \end{bmatrix} \begin{bmatrix} X^S & 0 \\ 0 & X^L \end{bmatrix} \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix} \right) \\ &= \text{tr}(A^T E L_3 E^T A) \end{aligned} \quad (11)$$

where $L_2 = \begin{bmatrix} 2(D^{w_s} - W^{w_s}) & 0 \\ 0 & 2(D^{w_l} - W^{w_l}) \end{bmatrix}$, $W^{w_s} = [w_{ij}^{w_s}]_{i,j=1}^N$, $W^{w_l} = [w_{ij}^{w_l}]_{i,j=1}^N$,
 $D^{w_s} = \text{diag}([\sum_{j=1}^N w_{ij}^{w_s}]_{i=1}^N)$, $D^{w_l} = \text{diag}([\sum_{j=1}^N w_{ij}^{w_l}]_{i=1}^N)$; $L_3 = \begin{bmatrix} 2(D^{b_s} - W^{b_s}) & 0 \\ 0 & 2(D^{b_l} - W^{b_l}) \end{bmatrix}$,
 $W^{b_s} = [w_{ij}^{b_s}]_{i,j=1}^N$, $W^{b_l} = [w_{ij}^{b_l}]_{i,j=1}^N$, $D^{b_s} = \text{diag}([\sum_{j=1}^N w_{ij}^{b_s}]_{i=1}^N)$, $D^{b_l} = \text{diag}([\sum_{j=1}^N w_{ij}^{b_l}]_{i=1}^N)$.

As discussed, the MFMDA method not only preserves the similarity relationship between spectral features and LBP features but also possesses strong discriminating ability in the low-dimensional embedding space. Therefore, a reasonable criterion for choosing a good projection matrix is to optimize the following objective functions:

$$\begin{cases} \min \text{tr}(A^T E L_1 E^T A) \\ \min \text{tr}(A^T E L_2 E^T A) \\ \max \text{tr}(A^T E L_3 E^T A) \end{cases} \quad (12)$$

The multi-objective function optimization problem in Equation (12) can be equivalent to:

$$\begin{aligned} J(A_S, A_L) &= \min \{ \text{tr}(A^T E L_1 E^T A) + \alpha (\text{tr}(A^T E L_2 E^T A)) - \beta (\text{tr}(A^T E L_3 E^T A)) \} \\ &= \min \{ \text{tr}(A^T E L E^T A) \} \end{aligned} \quad (13)$$

where $\alpha, \beta > 0$ are two tradeoff parameters which can adjust intramanifold compactness and intermanifold separability, $L = L_1 + \alpha L_2 - \beta L_3$.

A constraint $A^T E E^T A = I$ is imposed to remove an arbitrary scaling factor in the projection, and the objective function can be recast as follows:

$$\begin{cases} \min (\text{tr}(A^T E L E^T A)) \\ \text{s.t. } A^T E E^T A = I \end{cases} \quad (14)$$

With the method of Lagrangian multiplier, the optimization solution is formulated as

$$\frac{\partial}{\partial A} \text{tr}(A^T E L E^T A - \lambda(A^T E E^T A - I)) = 0 \quad (15)$$

where λ is the Lagrangian multiplier. Then the optimization problem is transformed to solve a generalized eigenvalue problem, i.e.,

$$E L E^T A = \lambda E E^T A \quad (16)$$

where the optimal projection matrix $A = [a_1, a_2, \dots, a_d]$ is composed of d minimum eigenvalues of Equation (16) corresponding eigenvectors. Then the low-dimensional feature is given by:

$$\begin{aligned} y_i &= \begin{bmatrix} y_i^s \\ y_i^l \end{bmatrix} = \begin{bmatrix} A_S^T x_i^s \\ A_L^T x_i^l \end{bmatrix} = \begin{bmatrix} (X^S B)^T x_i^s \\ (X^L C)^T x_i^l \end{bmatrix} = \begin{bmatrix} (X^S B)^T & 0 \\ 0 & (X^L C)^T \end{bmatrix} \begin{bmatrix} x_i^s \\ x_i^l \end{bmatrix} \\ &= \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}^T \begin{bmatrix} (X^S)^T & 0 \\ 0 & (X^L)^T \end{bmatrix} \begin{bmatrix} x_i^s \\ x_i^l \end{bmatrix} \end{aligned} \quad (17)$$

The proposed MFMDA algorithm is summarized in Algorithm 1.

Algorithm 1 MFMDA.

Input: data set $X = [x_1, x_2, x_3, \dots, x_N] \in \mathbb{R}^{D \times N}$, corresponding class labels $\ell(x_i) = \{1, 2, \dots, c\}$,

the number of intraclass neighbor points n_w and the number of interclass neighbor points n_b ,
balance parameter α and β .

1: Get LBP features generated from the data set, $X^S = \{x_i^s\}_{i=1}^N \in \mathbb{R}^{D \times N}$ and $X^L = \{x_i^l\}_{i=1}^N \in \mathbb{R}^{D \times N}$
denote the spectral and LBP features.

2: Find the n_w intraclass neighbor points and n_b interclass neighbor points of spectral features and
LBP features, respectively.

3: Calculate the edge weights of the intrinsic and penalty graphs by Equations (6)–(9).

4: Compute the D^{w_s} , D^{w_l} , D^{b_s} and D^{b_l} by $D^{w_s} = \text{diag}([\sum_{j=1}^N w_{ij}^{w_s}]_{i=1}^N)$, $D^{w_l} = \text{diag}([\sum_{j=1}^N w_{ij}^{w_l}]_{i=1}^N)$,
 $D^{b_s} = \text{diag}([\sum_{j=1}^N w_{ij}^{b_s}]_{i=1}^N)$ and $D^{b_l} = \text{diag}([\sum_{j=1}^N w_{ij}^{b_l}]_{i=1}^N)$, respectively.

5: Obtain the Lagrangian matrix which contain manifold structure through $L = L_1 + \alpha L_2 - \beta L_3$.

6: Calculate matrix E by $E = \begin{bmatrix} (X^S)^T X^S & 0 \\ 0 & (X^L)^T X^L \end{bmatrix}$.

7: Construct the matrix $A = [a_1, a_2, \dots, a_d]$ according to Equation (16).

8: Obtain the projection matrix of spectral and LBP features by $A = \begin{bmatrix} B & 0 \\ 0 & C \end{bmatrix}$, $A_S = X^S B$ and
 $A_L = X^L C$.

9: Achieve the low-dimensional features Y through Equation (17).

Output: $Y = [y_1, y_2, y_3, \dots, y_N] \in \mathbb{R}^{d \times N}$, $d \ll D$; Projection matrix of spectral and LBP features:

$$A_S = X^S B \text{ and } A_L = X^L C.$$

4. Experimental Results and Discussion

In this section, experiments are conducted on three real HSI data sets to evaluate the effectiveness of the proposed MFMDA method.

4.1. Experiment Data Set

Indian Pines data set: This HSI data set was collected by NASA using the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor in Northwest Indiana. After removing water absorption bands, the remaining 200 bands were used in the experiments. The size of this image is 145×145 pixels with a spatial resolution of 20 m, and it contains sixteen land cover types such as Wheat, Woods and Oats. This scene in false color and its corresponding ground truth are shown in Figure 3, and the values in brackets indicate the sample size of each class.

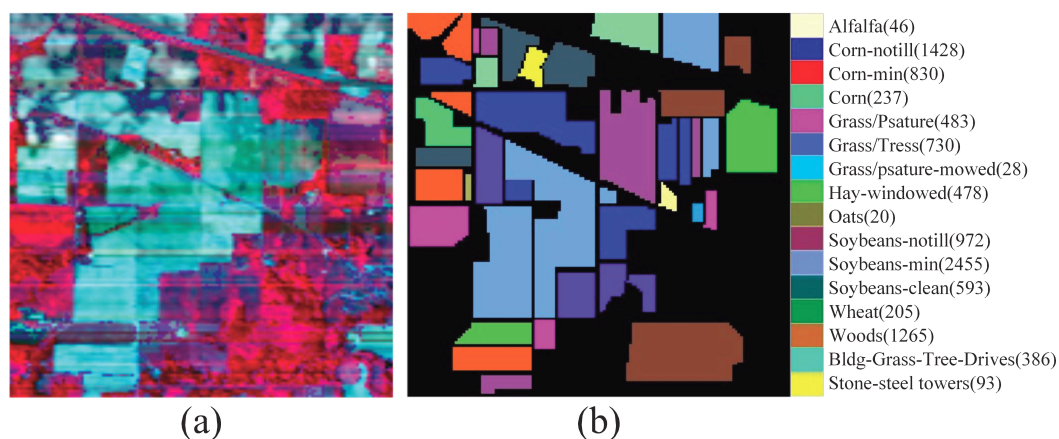


Figure 3. Indian Pines hyperspectral image. (a) HSI in false-color (bands 50, 27 and 17 for RGB); (b) Ground-truth map (please note that the number of samples is given in parentheses).

Heihe data set [55,56]: This data set 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, and it was captured by Compact Airborne Spectrographic Imager (CASI)/Shortwave Infrared Airborne Spectrographic Imager (SASI) in Zhangye basin which is located in the middle reaches of Heihe watershed, Gansu Province, China. The data possesses a spatial size of 684×453 pixels, and it has a geometric resolution of 2.4 m. Exactly 135 bands remained after removal of 14 bands which have noise and atmospheric effects. The data contains 9 different kinds of land covers. The scene in false color and its ground-truth map are shown in Figure 4.

PaviaU data set: This data set is a scene of the PaviaU collected by the reflective optics system imaging spectrometer (ROSIS) sensor. It consists of 610×340 pixels and the spatial resolution is 1.3 m. 103 spectral bands remained after the removal of some channels as a result of dense water vapor and atmospheric effects. There are nine classes of ground objects are considered in the data set such as trees, soil and meadows. This HSI in false color and its corresponding ground truth are shown in Figure 5.

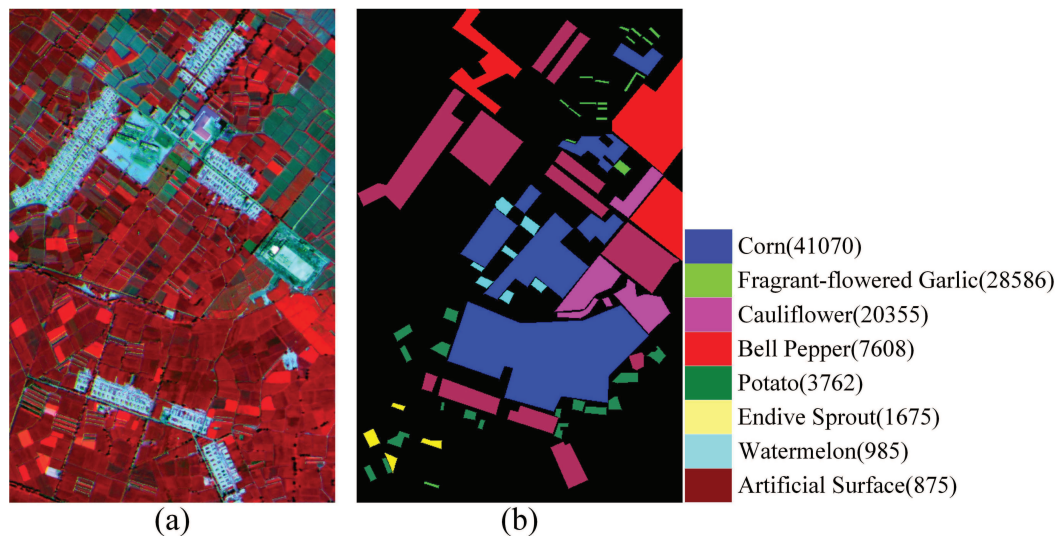


Figure 4. Heihe hyperspectral image. (a) HSI in false-color (bands 57, 19 and 80 for RGB); (b) Ground-truth map (please note that the number of samples is given in parentheses).

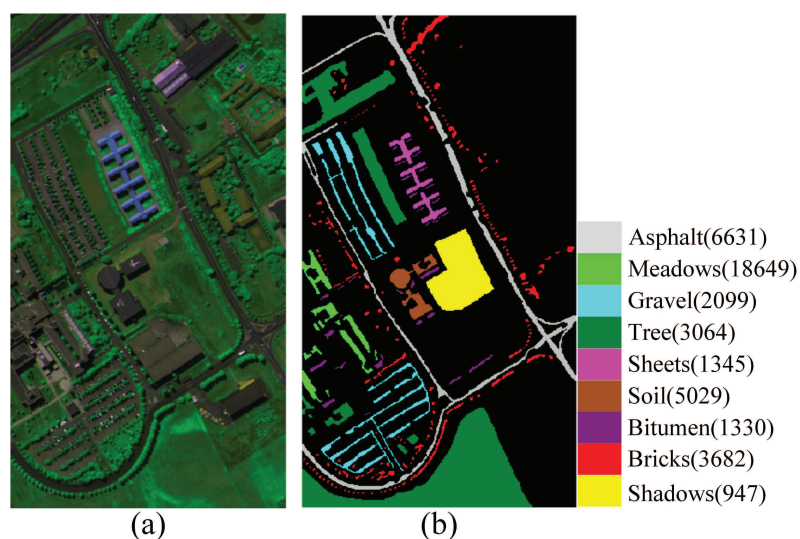


Figure 5. PaviaU hyperspectral image. (a) HSI in false-color (bands 60, 100 and 20 for RGB); (b) Ground-truth map (please note that the number of samples is given in parentheses).

4.2. Experimental Setup

In each experiment, the HSI data set was randomly divided into training and test sets. For the classes that are very small, i.e., Alfalfa, Grass/pasture-mowed, and Oats in Indian Pines data set, the number of training samples was set to 10 per class. The training samples are used to learn a low-dimensional embedding space, then all test samples are mapped into the embedding space for extracting low-dimensional features. After that, support vector machine (SVM) with the radial basis function (RBF) kernel were used to classify test samples, and the library for SVM (LibSVM) Toolbox was employed to implement SVM [57]. The parameters for SVM were optimized by a grid search. The classification accuracy for each class, overall classification accuracies (OAs), average classification accuracies (AAs), and *kappa* coefficient (*k*) are used to evaluate the classification performance of different DR methods. To robustly evaluate the performance of different methods under different

conditions, we repeated the experiments 10 times in each condition, and presented the results in the form of mean with standard deviation (STD).

The proposed MFMDA algorithm was compared with some state-of-art DR algorithms including Baseline, PCA [39], LDA [40], NPE [46], LPP [44], MFA [48] and LGSFA [51], where Baseline represents that test samples are classified directly by a classifier without dimensionality reduction. To verify the effectiveness of MFMDA, the above DR algorithms were applied to spectral features, LBP features and stacked features, respectively. Notice that LBP features are obtained by the “uniform LBP” pattern, and the ratio of neighborhood radius and the number of sampling points are set to 1 and 8, respectively [58]. The stacked features refer to stack original spectral features and LBP features after applying normalization.

For all methods, the parameters are optimized by using cross validation to achieve good results. The numbers of neighbors for NPE and LPP is set to 9. For MFA and LGSFA, the numbers of intraclass and interclass neighbors are chosen as 9 and 180, respectively. All experiments were performed on a personal computer with i7-7800X central processing unit, 32-G memory, and 64-bit Windows 10 using MATLAB 2014b.

4.3. Dimension Selection

To analyze the influence of different embedding dimensions on each DR algorithm, 40 samples were randomly selected from the stacked features of each class in three HSI data sets for training, and the remaining samples were tested. Figure 6 shows the overall classification accuracy under different embedding sizes.

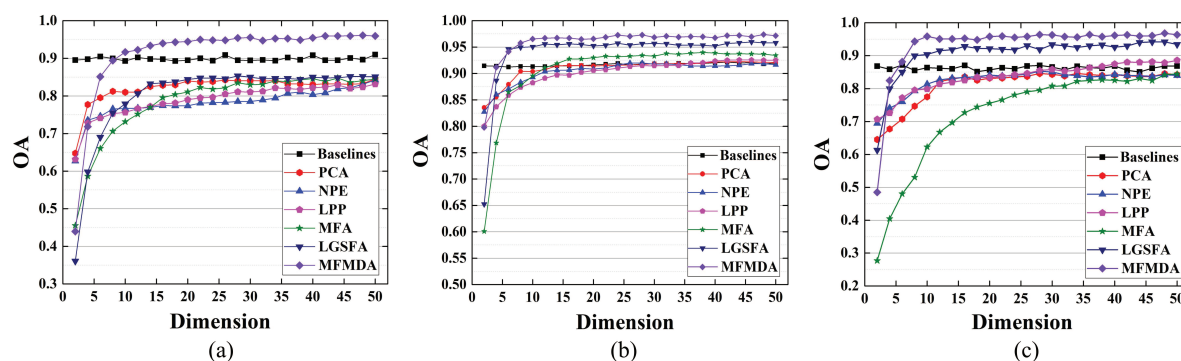


Figure 6. Classification results with different dimensions on the Indian Pines, Heihe and PaviaU data sets. (a) Indian Pines; (b) Heihe; (c) PaviaU.

As shown in Figure 6, with the increase of embedding dimension, the OAs of all methods are gradually improved. The reason for this is that the more discriminant information can be contained with the increase of embedding features, which is helpful for classification. However, when the dimension has been increased to a certain extent, the low-dimensional embedding space contains enough information for classification, and then the increase of embedding dimension has little effect on the improvement of classification performance. Meanwhile, MFMDA achieves better classification results than other methods, because the MFMDA can better characterize the intrinsic manifold structure of HSI and obtain more effective low-dimensional discriminant features. To achieve better classification performance for each algorithm, the embedding dimensions of all methods are set to 40. When it comes to LDA, the embedding dimension is set to $c - 1$, and c is the class number of the data set.

4.4. Experiments on the Indian Pines Data Set

In this section, the experiments were conducted on the Indian Pines data set to evaluate the effectiveness of the proposed algorithm. The proposed MFMDA method has different parameters, and we conducted experiments to analysis the sensitivity of parameters. In the experiments, 40 samples

per class were randomly selected for training, and the remaining ones for testing. The SVM classifier is used for classification. To investigate the classification influence on intraclass and interclass neighbors, parameters n_w and n_b are tuned with a set of $\{1, 2, 3, \dots, 9\}$ and a set of $\{2, 4, 6, \dots, 20\}$, respectively. Figure 7a shows the OAs with different values of n_w and n_b . When it comes to tradeoff parameters, parameters α and β are all tuned with the set of $\{0, 0.1, 0.2, \dots, 1\}$. The OAs with different values of α and β are displayed in Figure 7b.

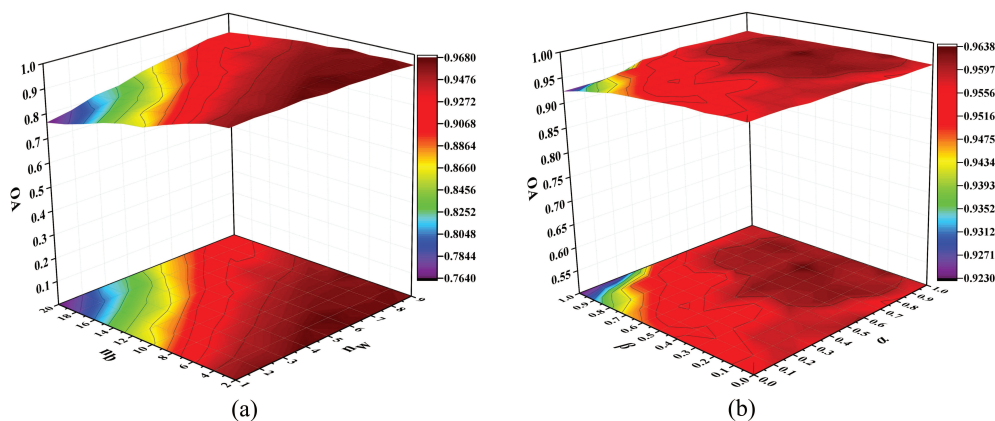


Figure 7. The experiments for parameter analysis of MFMDA on Indian Pines data set. (a) Classification results of MFMDA with different values of n_w and n_b ; (b) Classification results of MFMDA with different parameters α and β .

As can be seen in Figure 7a, with the increase of n_w , the classification accuracy improves and then tends to be stable, for the reason is that a large number of intraclass neighbors are conducive to reveal the intrinsic structure of HSI data. When the value of n_w is lower than 7, the OAs maintain a stable value with an increase of n_b , but the OAs significantly decline when the value of n_b exceeded 10. The reason is that too large values of n_b will cause the phenomenon of over-learning in the margins between interclass samples. In Figure 7b, the classification performance improves with the increase of parameter α and then slightly fluctuates. While the proposed method can achieve good results at a wide range. However, when β has a very large value, the effect of the intraclass separability will be limited. Thus, parameters α and β can balance the contribution between intraclass compactness and interclass separability. According to this figure, we set the parameters n_w and n_b to 6 and 4, α and β to 0.8 and 0.5 for achieving a satisfactory performance.

To analyze the classification performance of each algorithm under different numbers of training samples, n_i ($n_i = 5, 10, 20, 30, 40$) samples were randomly selected from each class for training, and the remaining data were used as test samples. Table 1 shows the average OAs with STD for different DR methods with different numbers of training samples.

According to Table 1, with the increase in the sample size of training set, the OAs of all methods continuously raise, for the reason is that a large training set contains more available information to learn discriminant features for classification. Furthermore, the classification results of each algorithm on LBP features are superior to that of spectral features, this is because LBP features are spatial-based features which bring benefits to classification. However, the classification performance of simply stacked features is even worse than LBP features, this may be due to the fact that spatial features and spectral features are not equally represented by simply stacking. While the proposed MFMDA algorithm produces a better classification effect than other methods in all conditions, especially when there are only a small number of labeled samples. The reason for this is that the proposed algorithm not only guarantees the similarity for spectral features and LBP features of the same pixel in the low-dimensional embedding space, but also discovers the manifold structure in the hyperspectral data by constructing intrinsic graphs and penalty graphs, and then extracts the spatial-spectral combined discriminant features to achieve the compactness for intraclass data and the separability for interclass data.

Table 1. Classification results using different methods with different classifiers for the Indian Pines data set. [Overall Accuracy $\pm$ Std (%)].

Algorithm		$n_i = 5$	$n_i = 10$	$n_i = 20$	$n_i = 30$	$n_i = 40$
Spectral Features	Baseline	42.21 $\pm$ 4.63	53.72 $\pm$ 3.97	64.05 $\pm$ 2.03	69.72 $\pm$ 1.25	69.78 $\pm$ 1.58
	PCA	42.00 $\pm$ 5.56	53.35 $\pm$ 4.70	64.04 $\pm$ 1.73	68.07 $\pm$ 1.05	68.72 $\pm$ 1.59
	LDA	40.27 $\pm$ 4.46	40.29 $\pm$ 1.76	52.93 $\pm$ 1.44	61.00 $\pm$ 0.84	63.51 $\pm$ 1.23
	NPE	34.25 $\pm$ 6.33	48.49 $\pm$ 3.36	60.61 $\pm$ 1.45	65.82 $\pm$ 1.76	67.22 $\pm$ 1.63
	LPP	35.40 $\pm$ 5.40	47.41 $\pm$ 2.77	60.14 $\pm$ 2.37	66.64 $\pm$ 1.82	69.10 $\pm$ 1.51
	MFA	42.98 $\pm$ 4.78	49.66 $\pm$ 1.71	60.38 $\pm$ 2.59	64.09 $\pm$ 1.87	66.16 $\pm$ 0.89
	LGSFA	41.26 $\pm$ 5.35	49.57 $\pm$ 3.00	60.95 $\pm$ 1.91	66.40 $\pm$ 0.65	68.81 $\pm$ 1.36
LBP Features	Baseline	68.12 $\pm$ 5.00	79.07 $\pm$ 2.97	87.58 $\pm$ 1.66	90.36 $\pm$ 2.82	93.54 $\pm$ 1.53
	PCA	65.72 $\pm$ 6.98	75.50 $\pm$ 4.09	83.89 $\pm$ 2.15	86.37 $\pm$ 3.40	89.80 $\pm$ 1.35
	LDA	71.30 $\pm$ 3.75	79.69 $\pm$ 3.24	87.62 $\pm$ 1.78	91.02 $\pm$ 2.04	93.63 $\pm$ 1.25
	NPE	66.71 $\pm$ 4.61	75.06 $\pm$ 4.32	84.60 $\pm$ 1.95	86.55 $\pm$ 1.89	89.87 $\pm$ 1.24
	LPP	65.88 $\pm$ 2.80	75.73 $\pm$ 4.51	85.17 $\pm$ 1.97	87.71 $\pm$ 1.55	89.77 $\pm$ 0.85
	MFA	65.60 $\pm$ 7.20	75.69 $\pm$ 4.70	84.68 $\pm$ 1.63	85.73 $\pm$ 2.55	90.34 $\pm$ 1.24
	LGSFA	65.59 $\pm$ 2.11	74.90 $\pm$ 5.18	84.92 $\pm$ 1.70	87.86 $\pm$ 1.67	91.34 $\pm$ 1.34
Stacked Features	Baseline	56.36 $\pm$ 5.70	73.25 $\pm$ 2.15	82.74 $\pm$ 2.33	89.46 $\pm$ 1.49	90.83 $\pm$ 1.62
	PCA	54.22 $\pm$ 5.04	68.83 $\pm$ 2.83	76.39 $\pm$ 1.66	81.34 $\pm$ 1.30	83.71 $\pm$ 1.25
	LDA	70.06 $\pm$ 4.65	84.37 $\pm$ 2.34	89.58 $\pm$ 2.52	92.93 $\pm$ 0.59	94.24 $\pm$ 1.26
	NPE	56.02 $\pm$ 6.91	66.73 $\pm$ 3.68	74.37 $\pm$ 2.51	76.28 $\pm$ 1.71	78.62 $\pm$ 0.91
	LPP	58.08 $\pm$ 4.73	70.81 $\pm$ 2.49	74.43 $\pm$ 2.54	78.59 $\pm$ 1.36	79.30 $\pm$ 0.73
	MFA	53.94 $\pm$ 2.49	69.44 $\pm$ 2.47	77.15 $\pm$ 2.12	80.57 $\pm$ 2.70	82.94 $\pm$ 1.83
	LGSFA	63.83 $\pm$ 5.17	73.35 $\pm$ 1.03	79.10 $\pm$ 0.37	83.15 $\pm$ 1.77	84.57 $\pm$ 0.57
MFMDA		74.01 $\pm$ 5.72	85.74 $\pm$ 2.42	91.79 $\pm$ 2.39	94.61 $\pm$ 1.15	96.19 $\pm$ 0.89

Notes: The bold numbers represent the maximum OA of the column.

To explore the classification performance of MFMDA on each class, 3% samples per class were randomly selected for training, and remaining samples were used for testing. We can see from Table 1, the experimental results of DR methods on LBP features are better than spectral features and stacked features, and thus LBP features were chosen for comparison in the following experiments. Table 2 lists the classification accuracy of each class, OAs, AAs, and *Kappa* coefficients of different methods, and Figure 8 shows the corresponding classification maps.

Table 2. Classification results of each class samples via different DR methods in Indian Pines data set (%).

Class	Samples		DR with SVM Classifier							
	Train	Test	Baseline	PCA	LDA	NPE	LPP	MFA	LGSFA	MFMDA
1	10	36	99.72	99.17	98.89	99.44	99.72	99.44	99.44	96.67
2	43	1385	93.91	92.94	93.86	92.25	93.23	94.09	93.10	93.33
3	25	805	94.31	92.42	93.74	92.01	91.69	91.66	92.46	92.57
4	10	227	95.37	94.85	94.98	92.73	93.66	94.98	94.23	93.17
5	14	469	84.54	82.41	83.22	83.71	84.86	82.90	84.61	90.00
6	22	708	93.29	92.97	92.19	92.13	91.44	92.20	91.91	98.18
7	10	18	100	100	100	100	100	100	100	98.89
8	14	464	98.53	98.38	98.41	97.91	98.38	97.80	97.95	99.74
9	10	10	100	100	100	100	100	100	99.00	100
10	29	943	89.72	88.23	90.38	88.29	87.41	89.00	88.09	91.56
11	74	2381	95.46	94.93	95.72	94.51	94.49	95.08	95.09	97.96
12	18	575	87.84	83.34	88.52	85.63	85.76	86.73	87.06	92.28
13	10	195	96.36	95.74	95.85	94.67	95.49	94.72	95.85	98.15
14	38	1227	96.85	96.27	97.42	95.98	96.09	96.10	95.75	99.62
15	12	374	87.86	89.12	84.47	87.54	88.34	86.47	87.99	96.52
16	10	83	95.78	96.27	92.89	96.39	96.39	96.27	95.42	99.52
OA			93.57	92.57	93.43	92.29	92.43	92.78	92.72	95.71
AA			94.35	93.56	93.78	93.33	93.56	93.59	93.62	96.14
Kappa			0.93	0.92	0.92	0.91	0.91	0.92	0.92	0.95

Notes: The bold numbers represent the maximum OA of the row.

As illustrated in Table 2, the MFMDA algorithm achieved good classification results in most classes, especially for the areas labeled as Wheat, Grass/Trees, Soybeans-min and Woods. By observing Figure 8, the classification map of MFMDA algorithm produced more homogenous regions than other methods.

The above results show that the proposed method compacts spectral features and LBP features from the same class and separates the features belong to different classes in low-dimensional embedding space, it can make better use of the manifold structure hidden in hyperspectral data.

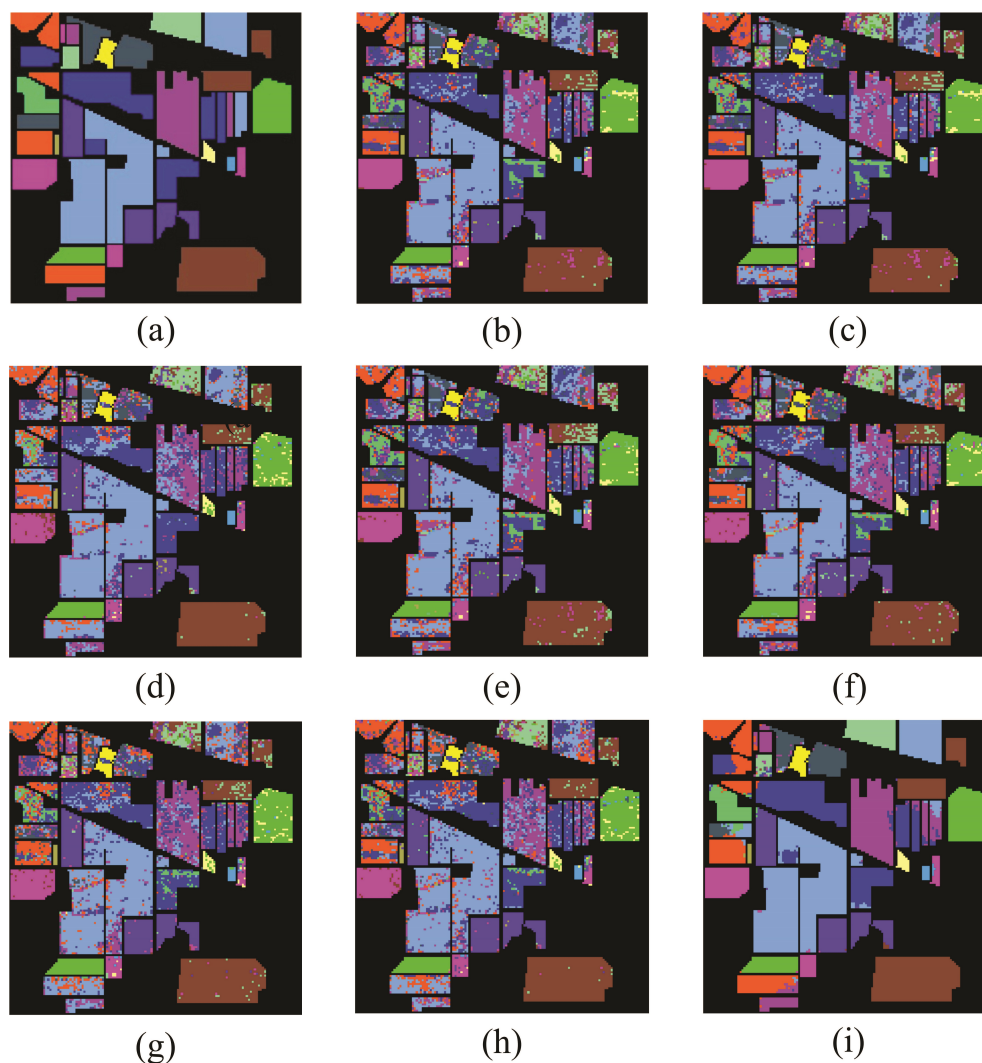


Figure 8. Classification results of different algorithms on Indian pines data set. (a) Ground truth; (b) Baseline(93.57%, 0.93); (c) PCA(92.57%, 0.92); (d) LDA(93.43%, 0.92); (e) NPE(92.29%, 0.91); (f) LPP(92.43%, 0.91); (g) MFA(92.78%, 0.92); (h) LGSFA(92.72%, 0.92); (i) MFMDA(95.71%, 0.95). Please note that OA and k coefficients are given in parentheses.

4.5. Experiments on the Heihe Data Set

In this section, the Heihe hyperspectral image was used to further evaluate the classification performance of the proposed algorithm. In the parameter sensitivity experiments, we randomly selected 40 samples from each class for training and the rest for testing. At first, we analyze the influence of different parameters on MFMDA algorithm, and the OAs with different values of parameters are displayed in Figure 9.

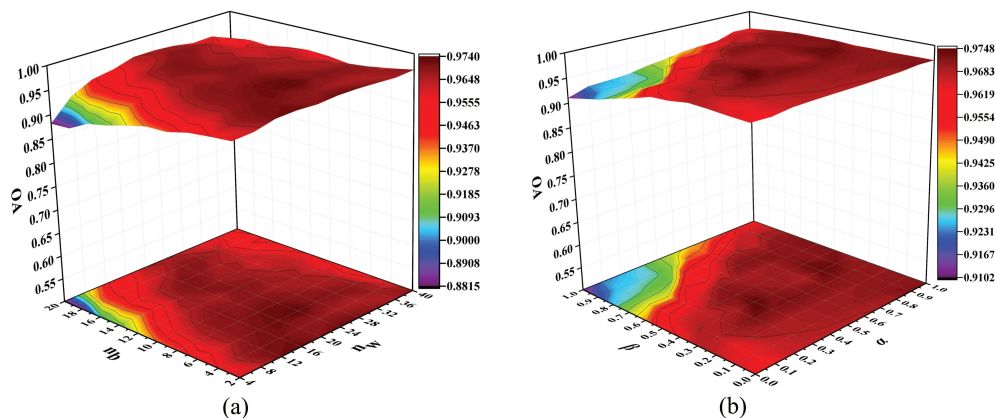


Figure 9. The experiments for parameter analysis of MFMDA on Heihe data set. (a) OAs of MFMDA with different values of n_w and n_b ; (b) OAs of MFMDA with different values of α and β .

As in Figure 9a, the OA increases and then declines with the increase of n_w , it is because a small value of n_w cannot get enough information to represent the intraclass structure, and a large value of n_w will lead to overfitting. At the same time, an appropriate size of interclass neighbor points can prevent overfitting and effectively obtain discriminant information of HSI data. In Figure 9b, it can be observed that the OAs increase and then maintain slight fluctuation with the increase of α , and a too small value of β will lead to unsatisfactory classification performance. This indicates that the suitable α and β can balance the intramanifold and intermanifold relations of spectral features and textual features. Based on the above analysis, we set the parameters n_w and n_b to 24 and 6, α and β to 0.8 and 0.4.

To compare the MFMDA algorithm with other DR methods under different numbers of training samples, we randomly selected n_i samples from each class for training, and the remaining samples were used for testing. Table 3 is the classification results of various algorithms.

Table 3. Classification results using different methods with different classifiers for the Heihe data set. [Overall Accuracy $\pm$ Std (%)].

	Algorithm	$n_i = 5$	$n_i = 10$	$n_i = 20$	$n_i = 30$	$n_i = 40$
Spectral Features	Baseline	80.88 $\pm$ 3.47	85.48 $\pm$ 2.93	90.17 $\pm$ 0.96	91.04 $\pm$ 1.00	91.95 $\pm$ 1.15
	PCA	80.88 $\pm$ 3.46	85.47 $\pm$ 2.93	90.17 $\pm$ 0.96	91.03 $\pm$ 1.00	91.95 $\pm$ 1.15
	LDA	74.72 $\pm$ 5.42	80.12 $\pm$ 3.43	89.85 $\pm$ 1.94	92.07 $\pm$ 0.77	92.99 $\pm$ 0.96
	NPE	78.33 $\pm$ 4.65	83.39 $\pm$ 3.29	88.70 $\pm$ 1.50	90.16 $\pm$ 1.36	91.36 $\pm$ 1.21
	LPP	69.60 $\pm$ 5.10	72.35 $\pm$ 11.48	91.62 $\pm$ 1.44	93.02 $\pm$ 0.84	93.34 $\pm$ 0.88
	MFA	83.27 $\pm$ 3.48	88.49 $\pm$ 3.93	92.26 $\pm$ 1.34	92.77 $\pm$ 1.01	93.14 $\pm$ 0.67
	LGSFA	80.00 $\pm$ 4.15	87.93 $\pm$ 1.87	90.47 $\pm$ 1.65	91.86 $\pm$ 0.77	93.42 $\pm$ 0.80
LBP Features	Baseline	75.82 $\pm$ 6.15	84.21 $\pm$ 2.59	89.95 $\pm$ 1.75	92.52 $\pm$ 1.08	93.58 $\pm$ 0.69
	PCA	75.79 $\pm$ 6.17	83.19 $\pm$ 3.16	89.02 $\pm$ 2.20	92.50 $\pm$ 1.05	93.74 $\pm$ 1.27
	LDA	77.91 $\pm$ 5.19	83.35 $\pm$ 4.33	89.95 $\pm$ 2.16	92.86 $\pm$ 0.91	93.86 $\pm$ 0.74
	NPE	75.18 $\pm$ 6.35	84.65 $\pm$ 3.17	91.38 $\pm$ 1.35	93.22 $\pm$ 0.89	94.62 $\pm$ 0.78
	LPP	55.31 $\pm$ 23.12	68.39 $\pm$ 12.66	87.84 $\pm$ 4.25	93.94 $\pm$ 1.00	94.57 $\pm$ 0.87
	MFA	73.51 $\pm$ 6.93	83.35 $\pm$ 3.93	91.18 $\pm$ 1.52	94.62 $\pm$ 0.65	95.11 $\pm$ 0.74
	LGSFA	76.39 $\pm$ 6.42	82.99 $\pm$ 4.48	90.04 $\pm$ 1.96	93.05 $\pm$ 0.65	94.32 $\pm$ 0.61
Stacked Features	Baseline	80.91 $\pm$ 3.44	85.68 $\pm$ 2.99	90.21 $\pm$ 0.97	91.35 $\pm$ 1.15	91.87 $\pm$ 1.17
	PCA	80.90 $\pm$ 3.44	85.67 $\pm$ 2.99	90.20 $\pm$ 0.97	91.38 $\pm$ 1.10	92.04 $\pm$ 1.15
	LDA	79.30 $\pm$ 3.29	90.75 $\pm$ 2.13	94.25 $\pm$ 0.80	94.93 $\pm$ 1.01	95.62 $\pm$ 0.54
	NPE	80.22 $\pm$ 3.87	88.19 $\pm$ 1.46	89.44 $\pm$ 1.45	90.78 $\pm$ 1.22	91.90 $\pm$ 1.02
	LPP	80.30 $\pm$ 2.64	87.52 $\pm$ 3.18	93.32 $\pm$ 1.53	94.16 $\pm$ 0.99	94.44 $\pm$ 0.65
	MFA	84.04 $\pm$ 3.22	90.05 $\pm$ 3.00	92.78 $\pm$ 1.47	93.96 $\pm$ 1.05	94.58 $\pm$ 0.57
	LGSFA	82.72 $\pm$ 3.36	91.30 $\pm$ 1.84	94.46 $\pm$ 1.43	95.68 $\pm$ 1.20	96.34 $\pm$ 0.80
MFMDA		89.32 $\pm$ 3.52	91.87 $\pm$ 2.53	95.92 $\pm$ 0.82	96.82 $\pm$ 0.54	97.41 $\pm$ 0.48

Notes: The bold numbers represent the maximum OA of the column.

According to Table 3, the classification accuracy increases as the number of training samples increases. Meanwhile, the experimental results of supervised learning methods, LDA, MFA and LGSFA, are superior to the unsupervised ones in most conditions, because the class information of data are used to enhance the discriminating capability of embedded features. The proposed method is more effective than other methods under various conditions, especially when a training set contains few samples. This shows that MFMDA can extract effective spatial-spectral joint features by exploring the inherent manifold structure of HSI data on the basis of GE, and then improve the classification accuracy.

To further show the classification results of each class, 0.1% samples were randomly selected for training, and the rest were used as test samples. The classification results of different methods on the Heihe data set is shown in Table 4, and Figure 10 shows the corresponding classification maps.

As illustrated in Table 4, it can be concluded that the proposed method achieves good classification performance on many classes, such as Endive Sprout and Artificial Surfaces. In addition, it possesses a smoother classification map, which is more conducive to practical application scenarios.

Table 4. Classification results of each class samples via different DR methods in Heihe data set (%).

Class	Samples		DR with SVM Classifier							
	Train	Test	Baseline	PCA	LDA	NPE	LPP	MFA	LGSFA	MFMDA
1	42	41029	96.23	95.87	98.49	97.33	98.07	97.64	98.56	97.86
2	29	28557	97.86	97.53	98.74	96.77	96.89	98.06	97.93	81.23
3	21	20334	95.46	95.26	97.24	95.27	95.77	96.54	95.45	95.66
4	10	7598	81.20	80.72	59.71	63.93	67.57	71.42	59.28	81.00
5	10	3752	84.81	84.21	72.62	84.03	73.25	78.11	72.79	83.31
6	10	1665	84.25	87.98	61.65	61.77	68.06	74.43	61.37	99.53
7	10	975	88.76	88.18	73.56	81.12	69.66	76.39	72.91	79.27
8	10	865	90.97	90.98	92.91	88.31	89.99	89.49	93.57	94.45
OA			94.72	94.45	93.71	93.09	93.40	94.31	93.14	95.64
AA			89.94	90.09	81.87	83.57	82.41	85.26	81.48	89.04
Kappa			0.93	0.92	0.92	0.91	0.90	0.91	0.92	0.94

Notes: The bold numbers represent the maximum OA of the row.

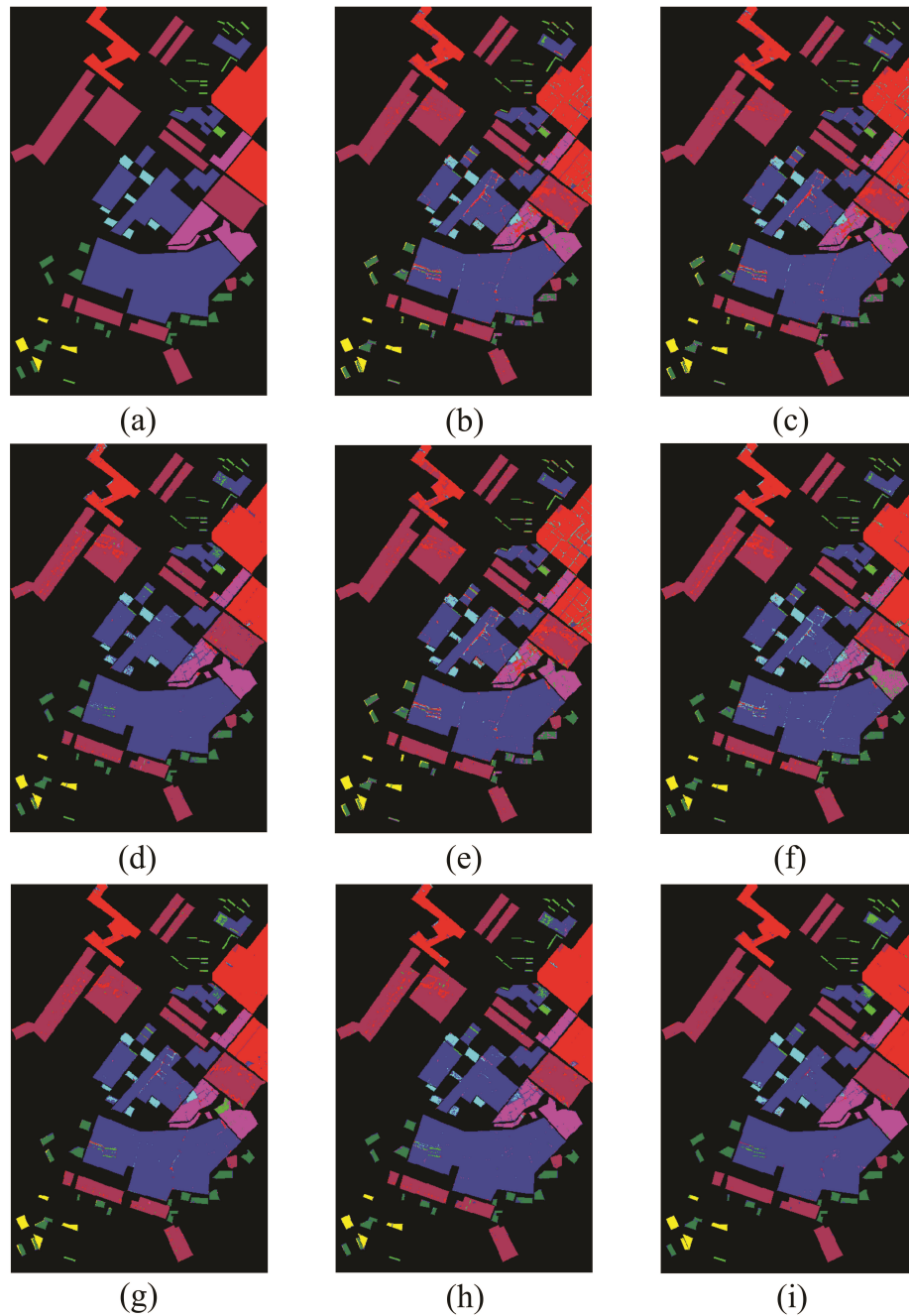


Figure 10. Classification results of different algorithms on Heihe data set. (a) Ground truth; (b) Baseline (94.72%, 0.93); (c) PCA (94.45%, 0.92); (d) LDA (93.71%, 0.92); (e) NPE (93.09%, 0.91); (f) LPP (93.40%, 0.90); (g) MFA (94.31%, 0.91); (h) LGSFA (93.14%, 0.92); (i) MFMDA (95.64%, 0.94). Please note that OA and k coefficients are given in parentheses.

4.6. Experiments on the PaviaU Data Set

In this section, we used PaviaU data set to analyze the classification performance of the proposed algorithm under different scenes. We randomly selected 40 samples per class as training set to explore OAs with respect to different parameters. The results are displayed in Figure 11.

In Figure 11a, as the increase of n_w , the OA rises first and then decreases slightly, the reason for this is that a small number of intraclass neighbor points cannot effectively explore intramanifold structure, while a large value of n_w will include redundant information and lead to a decrease in classification accuracy. At the same time, when n_b is lower than 8, the OAs can maintain a stable value. As shown in Figure 11b, the classification accuracy can fluctuate in a small range when the

values of α and β continue to increase. It shows that α and β can balance the information between the intramanifold and intermanifold structures in HSI data. To achieve good classification performance, we selected n_w , n_b , α and β as 28, 4, 0.5, 0.3, respectively.

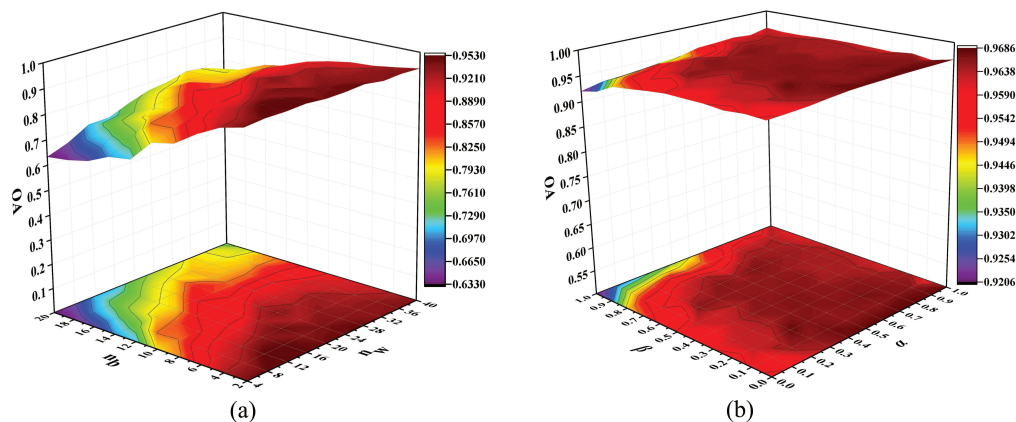


Figure 11. The experiments for parameter analysis of MFMDA on PaviaU data set. (a) Classification results of MFMDA with different parameters of n_w and n_b ; (b) Classification results of MFMDA with different parameters α and β .

To verify the effectiveness of the proposed algorithm, we randomly selected n_i ($n_i = 5, 10, 20, 30, 40$) samples from each class for training and remaining samples for testing. The average OAs with STD are given in Table 5.

Table 5. Classification results using different methods with different classifiers for the PaviaU data set. [Overall Accuracy $\pm$ Std (%)].

	Algorithm	$n_i = 5$	$n_i = 10$	$n_i = 20$	$n_i = 30$	$n_i = 40$
Spectral Features	Baseline	57.16 $\pm$ 9.94	69.72 $\pm$ 4.19	78.07 $\pm$ 2.88	81.11 $\pm$ 3.91	82.97 $\pm$ 2.33
	PCA	57.16 $\pm$ 9.94	69.72 $\pm$ 4.19	78.01 $\pm$ 2.98	81.11 $\pm$ 3.91	83.13 $\pm$ 2.35
	LDA	53.09 $\pm$ 5.85	57.87 $\pm$ 3.84	65.79 $\pm$ 3.21	70.08 $\pm$ 3.11	74.14 $\pm$ 2.15
	NPE	57.64 $\pm$ 9.81	67.02 $\pm$ 4.68	73.50 $\pm$ 4.81	79.98 $\pm$ 3.71	81.54 $\pm$ 2.47
	LPP	49.70 $\pm$ 5.35	50.66 $\pm$ 5.65	66.37 $\pm$ 2.23	73.57 $\pm$ 1.53	75.84 $\pm$ 2.25
	MFA	62.24 $\pm$ 5.56	75.16 $\pm$ 3.27	77.51 $\pm$ 2.17	80.97 $\pm$ 2.70	82.26 $\pm$ 3.68
	LGSFA	57.38 $\pm$ 4.50	62.91 $\pm$ 3.32	69.37 $\pm$ 3.36	71.11 $\pm$ 2.37	75.97 $\pm$ 1.76
LBP Features	Baseline	52.05 $\pm$ 8.27	72.01 $\pm$ 5.50	81.38 $\pm$ 2.22	86.32 $\pm$ 1.91	88.22 $\pm$ 1.21
	PCA	50.44 $\pm$ 7.55	67.93 $\pm$ 6.81	78.23 $\pm$ 3.68	81.71 $\pm$ 7.89	85.68 $\pm$ 2.42
	LDA	60.31 $\pm$ 6.87	76.02 $\pm$ 2.59	82.54 $\pm$ 1.32	86.70 $\pm$ 1.27	88.88 $\pm$ 0.56
	NPE	55.73 $\pm$ 6.77	73.68 $\pm$ 5.81	75.50 $\pm$ 3.50	85.69 $\pm$ 2.68	86.45 $\pm$ 1.41
	LPP	43.22 $\pm$ 11.11	55.85 $\pm$ 13.49	76.32 $\pm$ 4.19	84.27 $\pm$ 3.10	85.03 $\pm$ 2.62
	MFA	57.35 $\pm$ 7.51	71.79 $\pm$ 4.28	81.39 $\pm$ 1.93	85.63 $\pm$ 2.69	86.86 $\pm$ 1.43
	LGSFA	63.02 $\pm$ 6.49	73.59 $\pm$ 3.66	81.54 $\pm$ 2.42	85.81 $\pm$ 2.66	86.75 $\pm$ 2.16
Stacked Features	Baseline	59.40 $\pm$ 5.75	68.29 $\pm$ 5.54	80.06 $\pm$ 4.65	83.93 $\pm$ 2.53	85.73 $\pm$ 2.31
	PCA	57.52 $\pm$ 10.11	70.67 $\pm$ 3.63	77.52 $\pm$ 6.12	82.69 $\pm$ 3.33	85.63 $\pm$ 1.87
	LDA	59.89 $\pm$ 5.94	76.62 $\pm$ 2.74	81.85 $\pm$ 2.89	85.14 $\pm$ 3.80	87.28 $\pm$ 2.22
	NPE	57.50 $\pm$ 10.99	75.35 $\pm$ 2.34	77.89 $\pm$ 3.47	84.53 $\pm$ 3.87	85.62 $\pm$ 2.73
	LPP	64.96 $\pm$ 6.68	71.29 $\pm$ 3.56	78.12 $\pm$ 3.83	86.07 $\pm$ 1.68	87.86 $\pm$ 1.92
	MFA	61.57 $\pm$ 8.86	74.61 $\pm$ 4.66	79.07 $\pm$ 2.08	81.58 $\pm$ 2.30	82.34 $\pm$ 2.66
	LGSFA	64.45 $\pm$ 4.72	78.79 $\pm$ 2.68	88.03 $\pm$ 2.26	92.94 $\pm$ 2.02	94.15 $\pm$ 1.42
MFMDA		78.70 $\pm$ 2.70	84.60 $\pm$ 2.34	92.66 $\pm$ 2.32	95.10 $\pm$ 1.84	96.09 $\pm$ 0.97

Notes: The bold numbers represent the maximum OA of the column.

It can be seen from Table 5, the OAs of each method are improved when more samples are used for training. MFMDA achieves better results than other algorithms in most cases, the reason is that it can increase the margins between different classes, so the discriminant features are obtained for classification.

To compare the classification performance of various DR methods, we randomly selected 1% data in each class for training, and remaining data were used as test samples. As shown in Table 5, LBP features and stacked features achieve better experiment results than spectral features, so we choose stacked features compared with the MFMDA method. Table 6 gives the classification accuracies of different methods and Figure 12 shows the corresponding classification maps.

Table 6. Classification results of each class samples via different DR methods in PaviaU data set (%).

Class	Samples		DR with SVM Classifier							
	Train	Test	Baseline	PCA	LDA	NPE	LPP	MFA	LGSFA	MFMDA
1	10	6565	89.98	89.77	92.95	89.24	93.72	90.37	92.43	97.90
2	186	18463	97.93	97.41	98.04	97.63	98.03	97.46	97.27	99.79
3	21	2078	72.28	69.59	75.81	68.50	73.39	69.82	77.02	94.02
4	31	3033	85.50	85.32	89.52	85.38	87.94	86.74	89.10	84.89
5	13	1332	98.97	98.84	99.65	98.74	99.29	99.41	99.64	99.99
6	50	4979	85.45	80.95	84.26	83.02	80.26	80.16	80.90	99.14
7	13	1317	81.97	76.74	67.06	78.02	67.56	74.68	69.35	96.10
8	37	3645	85.48	84.37	86.75	83.81	89.99	84.82	88.13	98.06
9	10	937	99.73	99.68	94.26	99.75	99.86	99.79	99.61	60.91
OA			91.59	90.39	91.96	90.59	91.78	90.52	91.49	96.95
AA			88.59	86.96	87.59	87.12	87.78	87.03	88.16	92.31
Kappa			0.89	0.87	0.89	0.87	0.89	0.87	0.89	0.96

Notes: The bold numbers represent the maximum OA of the row.

As shown in Table 6, the proposed method obtained the best classification results in most classes, especially in Asphalt, Gravel, Bare Soil, Bitumen, Bricks. The reason is that the MFMDA algorithm effectively fuses the multiple features by compacting spectral features and LBP features from the same class in low-dimensional space. As displayed in Figure 12, MFMDA algorithm has fewer misclassified points and the classification map is smoother than other methods.

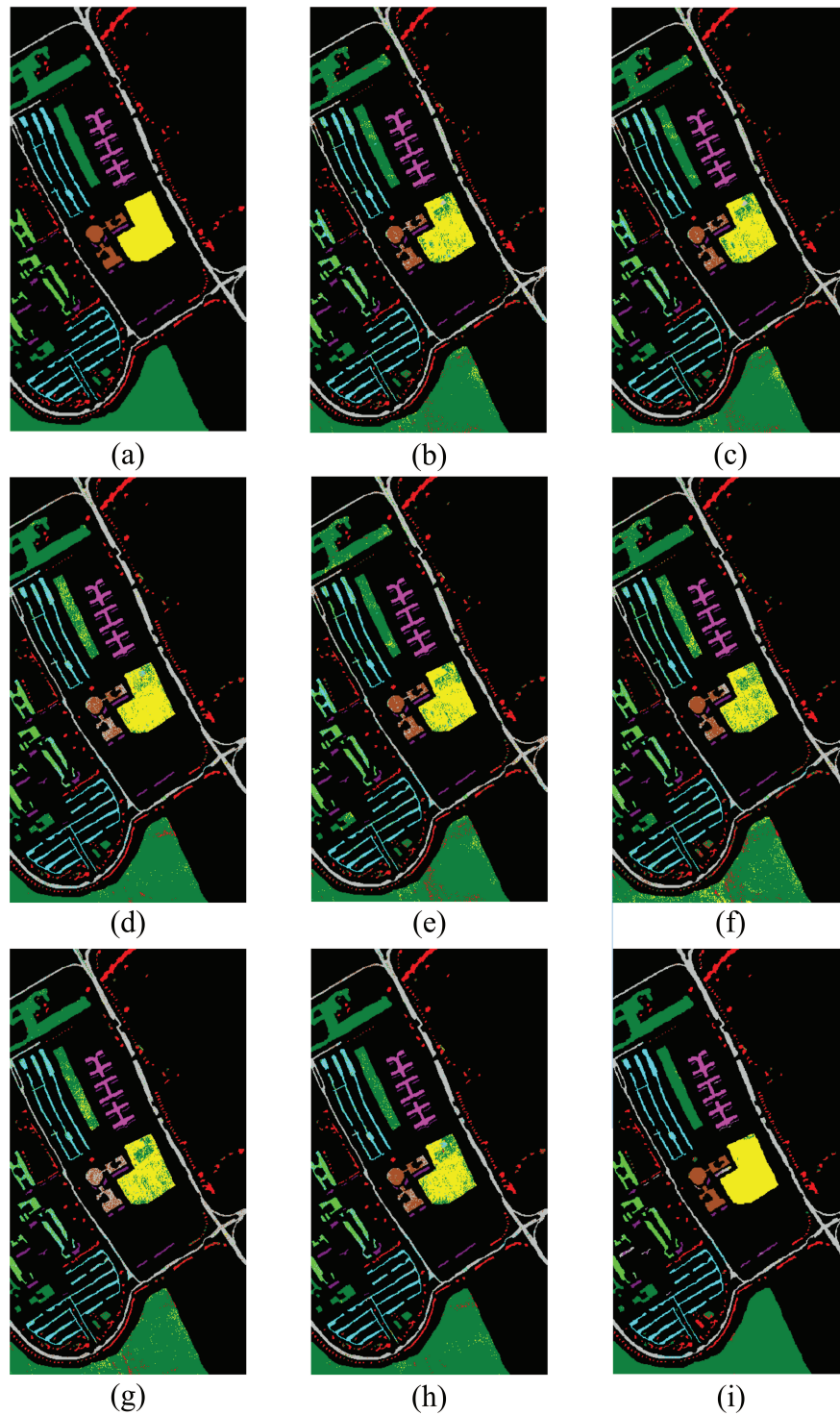


Figure 12. Classification results of different algorithms on PaviaU data set. (a) Ground truth; (b) Baseline(91.59%, 0.89); (c) PCA(90.39%, 0.87); (d) LDA(91.96%, 0.89); (e) NPE(90.59%, 0.87); (f) LPP(91.78%, 0.89); (g) MFA(90.52%, 0.87); (h) LGSFA(91.49%, 0.89); (i) MFMDA(96.95%, 0.96). Please note that OA and k coefficients are given in parentheses.

4.7. Discussion

The experiments on three HSI data sets reveal some interesting points.

- As shown in Tables 1, 3 and 5, the classification performance of simply stacked features is even worse than LBP features in most cases, for the reason that simply stacked spatial and spectral features may yield redundant information and even lead to the curse-of-dimensionality.
- From the experimental results, it is obviously that DR methods on LBP features or spectral features usually perform better than DR methods on the simply stacked features. This may be due to the fact that performing dimensionality reduction directly on the simply stacked features cannot reveal the manifold structure of different features in HSI, which will restrict their discriminant capability for classification.
- The proposed MFMDA algorithm is superior to other DR methods under different training conditions. The reason is that MFMDA constructs the intrinsic graphs and penalty graphs of spectral features and LBP features to discover the manifold structure of spatial features and spectral features, then it learns low-dimensional embedding space from original spectral features as well as LBP features for compacting the intramanifold samples while separating intermanifold samples. As a result, the spatial-spectral embedding features possess stronger discriminating ability for HSI classification.

5. Conclusions

Traditional methods explore only a single feature or simply stacked features in hyperspectral image, which will restrict their discriminant capability for classification. In this paper, we proposed a new dimensionality reduction method termed MFMDA to couple DR and fusion of spectral and textual features of HSI data. MFMDA first explores LBP operator to extract textural features for encoding the spatial information in HSI. Then, within GE framework, the intrinsic and penalty graphs of LBP and spectral features are constructed to explore the discriminant manifold structure in both spatial and spectral domains, respectively. After that, a new spatial-spectral DR model is built to extract discriminant spatial-spectral combined features which not only preserve the similarity relationship between spectral features and LBP features but also possess strong discriminating ability in the low-dimensional embedding space. Experiments on Indian Pines, Heihe and PaviaU hyperspectral data sets demonstrate that the proposed MFMDA method can significantly improve classification performance and result in smoother classification maps than some state-of-the-art methods, and with fewer training samples, the classification accuracy can reach 95.43%, 97.19% and 96.60%, respectively. In the future, we will focus on conducting a more detailed investigation of other possible features to further improve the performance of MFMDA.

Author Contributions: H.H. contributed to mathematical modeling, experiment analysis and revised the paper. Z.L. was primarily responsible for experimental design and completed the comparison with other methods. Y.P. provided important suggestions for improving the paper.

Funding: This work was supported by the National Science Foundation of China under Grant 41371338, the Basic and Frontier Research Programmes of Chongqing under Grant cstc2018jcyjAX0093, and the graduate research and innovation foundation of Chongqing under Grant 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. 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](#)]
2. Zhong, Y.F.; Wang, X.Y.; Xu, Y.; Wang, S.Y.; Jia, T.Y.; Hu, X.; Zhao, J.; Wei, L.F.; Zhang, L.P. Mini-UAV-Borne Hyperspectral Remote Sensing: From Observation and Processing to Applications. *IEEE Geosci. Remote Sens. Mag.* **2018**, *6*, 46–62. [[CrossRef](#)]
3. Chen, Y.S.; Jiang, H.L.; Li, C.Y.; Jia, X.P.; Ghamisi, P. Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 6232–6251. [[CrossRef](#)]
4. Huang, H.; Duan, Y.L.; Shi, G.Y.; Lv, Z.Y. Fusion of Weighted Mean Reconstruction and SVMCK for Hyperspectral Image Classification. *IEEE Access* **2018**, *6*, 15224–15235. [[CrossRef](#)]
5. Sun, W.W.; Yang, G.; Wu, K.; Li, W.Y.; Zhang, D.F. Pure endmember extraction using robust kernel archetypoid analysis for hyperspectral imagery. *ISPRS J. Photogramm. Remote Sens.* **2017**, *131*, 147–159. [[CrossRef](#)]
6. Jiao, C.Z.; Chen, C.; McGarvey, R.G.; Bohlman, S.; Jiao, L.C.; Zare, A. Multiple Instance Hybrid Estimator for Hyperspectral Target Characterization and Sub-pixel Target Detection. *J. Photogramm. Remote Sens.* **2018**, *146*, 235–250. [[CrossRef](#)]
7. Dian, R.W.; Li, S.T.; Guo, A.J.; Fang, L.Y. Deep Hyperspectral Image Sharpening. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, *29*, 5345–5355. [[CrossRef](#)] [[PubMed](#)]
8. Wang, X.Y.; Zhong, Y.F.; Xu, Y.; Zhang, L.P.; Xu, Y.Y. Saliency-Based Endmember Detection for Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3667–3680. [[CrossRef](#)]
9. Wang, Q.; Lin, J.Z.; Yuan, Y. Salient Band Selection for Hyperspectral Image Classification via Manifold Ranking. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *27*, 1279–1289. [[CrossRef](#)] [[PubMed](#)]
10. Qian, Y.T.; Xiong, F.C.; Zeng, S.; Zhou, J.; Tang, Y.Y. Matrix-Vector Nonnegative Tensor Factorization for Blind Unmixing of Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 1776–1792. [[CrossRef](#)]
11. Luo, F.L.; Huang, H.; Liu, J.M.; Ma, Z.Z. Fusion of Graph Embedding and Sparse Representation for Feature Extraction and Classification of Hyperspectral Imagery. *Photogramm. Eng. Remote Sens.* **2017**, *83*, 37–46. [[CrossRef](#)]
12. Kang, X.D.; Duan, P.H.; Li, S.T.; Benediktsson, J.A. Decolorization-Based Hyperspectral Image Visualization. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 4346–4360. [[CrossRef](#)]
13. Ke, W.; Xu, G.; Zhang, Y.X.; Du, B. Hyperspectral image target detection via integrated background suppression with adaptive weight selection. *Neurocomputing* **2017**, *315*, 59–67.
14. 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](#)]
15. Li, S.T.; Hao, Q.B.; Kang, X.D.; 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](#)]
16. Peng, J.T.; Li, L.Q.; Tang, Y.Y. Maximum Likelihood Estimation-Based Joint Sparse Representation for the Classification of Hyperspectral Remote Sensing Images. *IEEE Trans. Neural Netw. Learn. Syst.* **2018**, 1–13. [[CrossRef](#)]
17. Wang, Z.M.; Du, B.; Zhang, L.F.; Zhang, L.P.; Jia, X.P. A Novel Semisupervised Active-Learning Algorithm for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 3071–3083. [[CrossRef](#)]
18. Su, H.J.; Zhao, B.; Du, Q.; Sheng, Y.H. Tangent Distance-Based Collaborative Representation for Hyperspectral Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1236–1240. [[CrossRef](#)]
19. Zhang, L.F.; Zhang, L.P.; Tao, D.C.; Huang, X. On Combining Multiple Features for Hyperspectral Remote Sensing Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 879–893. [[CrossRef](#)]
20. 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](#)]
21. Su, H.J.; Zhao, B.; Du, Q.; Du, P.J.; Xue, Z.H. Multifeature Dictionary Learning for Collaborative Representation Classification of Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2467–2484. [[CrossRef](#)]

22. Fang, L.Y.; Li, S.T.; Duan, W.H.; Ren, J.C.; Benediktsson, J.A. Classification of Hyperspectral Images by Exploiting Spectral-Spatial Information of Superpixel via Multiple Kernels. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 6663–6674. [[CrossRef](#)]
23. Liang, M.M.; Jiao, L.C.; Yang, S.Y.; Liu, F.; Hou, B.; Chen, H. Deep Multiscale Spectral-Spatial Feature Fusion for Hyperspectral Images Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2018**, *11*, 2911–2924. [[CrossRef](#)]
24. Chen, C.; Li, W.; Su, H.J.; Liu, K. Spectral-Spatial Classification of Hyperspectral Image Based on Kernel Extreme Learning Machine. *Remote Sens.* **2014**, *6*, 5795–5814. [[CrossRef](#)]
25. Gu, Y.F.; Liu, T.Z.; Jia, X.P.; Benediktsson, J.A.; Chanussot, J. Nonlinear Multiple Kernel Learning With Multiple-Structure-Element Extended Morphological Profiles for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3235–3247. [[CrossRef](#)]
26. 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](#)]
27. 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.* **2018**. [[CrossRef](#)]
28. 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.* **2018**, *56*, 1409–1424. [[CrossRef](#)]
29. Li, W.; Chen, C.; Su, H.J.; Du, Q. Local Binary Patterns and Extreme Learning Machine for Hyperspectral Imagery Classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *55*, 3681–3693. [[CrossRef](#)]
30. Mauro, M.D.; Benediktsson, J.A.; Waske, B.; Bruzzone, L. Morphological Attribute Profiles for the Analysis of Very High Resolution Images. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 3747–3762.
31. Li, J.; Marpu, P.R.; Plaza, A.; Bioucas-Dias, J.M.; Benediktsson, J.A. Generalized Composite Kernel Framework for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 4816–4839. [[CrossRef](#)]
32. Li, J.; Huang, X.; Gamba, P.; Bioucas-Dias, J.M.; Zhang, L.P.; Benediktsson, J.A.; Plaza, A. Multiple Feature Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1592–1606. [[CrossRef](#)]
33. Song, C.Y.; Yang, F.J.; Li, P.J. Rotation invariant texture measured by local binary pattern for remote sensing image classification. In Proceedings of the 2010 Second International Workshop on Education Technology and Computer Science, Wuhan, China, 6–7 March 2010; Volume 3, pp. 3–6.
34. Zhong, Y.F.; Ma, A.; Ong, Y.S.; Zhu, Z.X.; Zhang, L.P. Computational intelligence in optical remote sensing image processing. *Appl. Soft Comput.* **2018**, *64*, 75–93. [[CrossRef](#)]
35. 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](#)]
36. Wang, J.; He, H.B.; Prokhorov, D.V. A Folded Neural Network Autoencoder for Dimensionality Reduction. *Proc. Comput. Sci.* **2012**, *13*, 120–127. [[CrossRef](#)]
37. 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](#)]
38. Zheng, X.T.; Yuan, Y.; Lu, X.Q. Dimensionality Reduction by Spatial-Spectral Preservation in Selected Bands. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 5185–5197. [[CrossRef](#)]
39. Bonifazi, G.; Capobianco, G.; Serranti, S. Asbestos containing materials detection and classification by the use of hyperspectral imaging. *J. Hazard. Mater.* **2018**, *13*, 981–993. [[CrossRef](#)] [[PubMed](#)]
40. Huang, X.Y.; Zhang, B.; Qiao, H.; Nie, X.L. Local Discriminant Canonical Correlation Analysis for Supervised PolSAR Image Classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 2102–2106. [[CrossRef](#)]
41. Xu, X.; Huang, Z.H.; Zuo, L.; He, H.B. Manifold-based Reinforcement Learning via Locally Linear Reconstruction. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *28*, 934–947. [[CrossRef](#)] [[PubMed](#)]
42. 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*, 1532–1539. [[CrossRef](#)]
43. Xu, X.; Yang, H.Y.; Lian, C.Q.; Liu, J.H. Self-learning control using dual heuristic programming with global Laplacian eigenmaps. *IEEE Trans. Ind. Electron.* **2017**, *64*, 9517–9526. [[CrossRef](#)]

44. Deng, Y.J.; Li, H.C.; Pan, L.; Shao, L.Y.; Du, Q.; Emery, W.J. Modified Tensor Locality Preserving Projection for Dimensionality Reduction of Hyperspectral Images. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 277–281. [[CrossRef](#)]
45. Zhang, L.L.; Zhao, C.H. Sparsity divergence index based on locally linear embedding for hyperspectral anomaly detection. *J. Appl. Remote Sens.* **2016**, *10*, 025026. [[CrossRef](#)]
46. Lu, G.F.; Jin, Z.; Zou, J. Face recognition using discriminant sparsity neighborhood preserving embedding. *Knowl. Based Syst.* **2012**, *31*, 119–127. [[CrossRef](#)]
47. Wang, J.; Sun, X.L.; Du, J.X. Local tangent space alignment via nuclear norm regularization for incomplete data. *Neurocomputing* **2018**, *273*, 141–151. [[CrossRef](#)]
48. Lu, Y.W.; Lai, Z.H.; Fan, Z.Z.; Cui, J.R.; Zhu, Q. Manifold discriminant regression learning for image classification. *Neurocomputing* **2015**, *166*, 475–486. [[CrossRef](#)]
49. Yu, H.Y.; Gao, L.R.; Li, W.; Du, Q.; Zhang, B. Locality sensitive discriminant analysis for group sparse representation-based hyperspectral imagery classification. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 1358–1362. [[CrossRef](#)]
50. 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](#)]
51. 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](#)]
52. Fauvel, M.; Benediktsson, J.A.; Chanussot, J.; Sveinsson, J.R. Spectral and Spatial Classification of Hyperspectral Data Using SVMs and Morphological Profile. *IEEE Trans. Geosci. Remote Sens.* **2018**, *9*, 3804–3814.
53. Huo L.Z.; Tang, P. Spectral and spatial classification of hyperspectral data using SVMs and Gabor textures. *Proc. Int. Geosci. Remote Sens. Symp.* **2011**, *9*, 1708–1711.
54. Huang, K.S.; Li, S.T.; Kang, X.D.; Fang, L.Y. Spectral-spatial hyperspectral image classification based on KNN. *Sens. Imaging* **2016**, *17*, 1. [[CrossRef](#)]
55. Xiao, Q.; Wen, J.G. HiWATER: Thermal-Infrared Hyperspectral Radiometer (4th, July, 2012). *Heihe Plan Sci. Data Center* **2013**. [[CrossRef](#)]
56. 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.
57. Chang, C.C.; Lin, C.J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2011**, *2*, 475–486. [[CrossRef](#)]
58. Chang, C.C.; Lin, C.J. LIBSVM: Perceptual Image Hashing Using Latent Low-Rank Representation and Uniform LBP. *Appl. Sci.* **2018**, *8*, 317.

