



Article Dimensionality Reduction of Hyperspectral Image Based on Local Constrained Manifold Structure Collaborative Preserving Embedding

Guangyao Shi¹, Fulin Luo^{2,*}, Yiming Tang³ and Yuan Li³

- School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; shi_gy@126.com
- ² State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China
- ³ The Key Laboratory on Opto-electronic Technique and Systems, Ministry of Education, Chongqing University, Chongqing 400044, China; ymtang@cqu.edu.cn (Y.T.); yuan_li@cqu.edu.cn (Y.L.)
- Correspondence: luoflyn@163.com

Abstract: Graph learning is an effective dimensionality reduction (DR) manner to analyze the intrinsic properties of high dimensional data, it has been widely used in the fields of DR for hyperspectral image (HSI) data, but they ignore the collaborative relationship between sample pairs. In this paper, a novel supervised spectral DR method called local constrained manifold structure collaborative preserving embedding (LMSCPE) was proposed for HSI classification. At first, a novel local constrained collaborative representation (CR) model is designed based on the CR theory, which can obtain more effective collaborative coefficients to characterize the relationship between samples pairs. Then, an intraclass collaborative graph and an interclass collaborative graph are constructed to enhance the intraclass compactness and the interclass separability, and a local neighborhood graph is constructed to preserve the local neighborhood structure of HSI. Finally, an optimal objective function is designed to obtain a discriminant projection matrix, and the discriminative features of various land cover types can be obtained. LMSCPE can characterize the collaborative relationship between sample pairs and explore the intrinsic geometric structure in HSI. Experiments on three benchmark HSI data sets show that the proposed LMSCPE method is superior to the state-of-the-art DR methods for HSI classification.

Keywords: hyperspectral image; graph learning; dimensionality reduction; collaborative representation; local neighborhood structure

1. Introduction

Hyperspectral imagery (HSI) captures reflectance values over a wide range of electromagnetic spectra for each pixel, it can distinguish more subtle differences between land cover types than traditional multispectral imagery (MSI) [1–4]. Due to the detailed spatial structure and spectral information in HSI, it has been widely used in the fields of urban planning, environment monitoring, precision agriculture, and land-cover classification [5–7]. However, the improvement of the spectral and spatial resolution in hyperspectral sensors has led to high-dimensional data sets, while high-dimensionality data processing requires huge computational resources and storage capacity [8–10]. Besides, the classification performance often deteriorates as the dimensionality increases (Hughes phenomenon) [11]. Therefore, it is of great importance to perform dimensionality reduction (DR) for HSI data while preserving the useful feature information [12,13].

Serving as a good tool for DR, graph learning methods have attracted increasing attention of researchers by mapping the high-dimensional data into a lower-dimensional embedding space [14–16]. Based on this theory, a lot of graph learning algorithms and its variants have been proposed to reveal the intrinsic geometric structure of high-dimensional



Citation: Shi, G.; Luo, F.; Tang, Y.; Li, Y. Dimensionality Reduction of Hyperspectral Image Based on Local Constrained Manifold Structure Collaborative Preserving Embedding. *Remote Sens.* 2021, *13*, 1363. https://doi.org/10.3390/rs13071363

Academic Editors: Lefei Zhang, Liangpei Zhang, Qian Shi and Yanni Dong

Received: 10 February 2021 Accepted: 29 March 2021 Published: 2 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 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 (https:// creativecommons.org/licenses/by/ 4.0/). data [17–19], such as Laplacian eigenmaps (LE) [20], local linear embedding (LLE) [21], and isometric feature mapping (ISOMAP) [22]. LE aims to find low dimensional representations of the high dimensional data by preserving the local geometry between them [23]. LLE computes the low-dimensional features that best preserve the local geometry of each locally linear patch, and seeks a lower-dimensional projection of the data that preserves distances within local neighborhoods [24]. ISOMAP aims at preserving geodesic distances of all similarity pairs for delivering highly nonlinear manifolds, and it approximates the geodesic distance between two points by measuring shortest path between these points [25]. However, as they are nonlinear algorithms, the DR process can not obtain an explicit project matrix, which can map the test samples into low-dimensional space [26]. Therefore, it is difficult for the nonlinear algorithms to process new samples.

To overcome this drawback, researchers proposed a series of linear graph learning methods [27,28], including locality preserving projection (LPP) [29], discriminative supervised neighborhood preserving embedding (DSNPE) [30], and multi-manifold discriminant analysis (MMDA) [31]. LPP constructs an adjacency matrix to weight the distance between each pair of sample points for learning a projection that can preserve the local manifold structures of data. DSNPE finds the optimal projection direction by pulling the neighboring points with the same class label as near as possible, while simultaneously pushing the neighboring points with different labels as far as possible. MMDA designs the intraclass graph and interclass graph to characterize the intraclass compactness and the interclass separability, and it seeks for the discriminant matrix by maximizing the interclass scatter and minimizing the intraclass scatter simultaneously. In order to unify these algorithms, a graph embedding (GE) framework has been designed to analyze the graph learning methods on the basis of statistics or geometry theory [32–34]. Based on this framework, discriminant analysis with graph learning (DAGL) method was proposed by pulling the within-class similar samples together while pushing the between-class similar samples far away [35]. However, the above graph learning methods construct adjacency graph mainly depend on pairwise Euclidean distance, the classification performance of them is sensitive to data noise, which may result in suboptimal graph representation [36–38].

Recently, the graph learning methods based on sparse representation have achieved good classification performance [39–41]. The reason is that sparse coefficients can characterize the relationship between sample pairs more accurately, and they can preserve more valuable information of HSI data for DR. Among these methods, sparse graph based discriminant analysis (SGDA) exploits the discriminant capability from sparse representation [42], and enhances the discriminant power with the labeled information. Sparse manifold embedding (SME) utilizes the sparse coefficients of affine subspace to construct a similarity graph and preserves this sparse similarity in embedding space [43]. Discriminative learning by sparse representation projections (DLSP) incorporates the merits of both local geometry structure and global sparse property, it has better discrimination performance for different classes [44]. Although sparse graph-based methods are superior to traditional graph learning methods, they suffer from the problem of over-sparseness issue when there are few samples. Besides, these methods solve the sparse coefficients with l_1 -norm, which is an iterative procedure and may lead to a higher computational cost [45,46]. Therefore, collaborative representation is introduced to graph learning for avoiding the above-mentioned problems [47,48]. Collaborative graph-based discriminant analysis (CGDA) exploits the discriminant capability with collaborative representation, it can effectively characterize the relationship between sample pairs and improve the classification performance [49]. Manifold aware discriminant collaborative graph embedding (MADCGE) constructs an adjacent graph with the collaborative representation coefficients, it can preserve the linear reconstructive relationships between samples, and sufficiently utilize the merits of label information and nonlinear manifold structure to further improve the discriminative ability [50]. Collaborative representation based local discriminant projection (CRLDP) utilizes collaborative representation relationships among samples to construct the within-class graph and the between-class graph, which can characterize

the compactness and separability of samples, then it seeks to find an optimal projection matrix by maximizing the ratio of the between-class scatter to the within-class scatter [51]. However, these methods only consider global geometry information of high dimensional data, and ignore the local neighborhood information of within-class samples, which will limit the discriminant ability of collaborative representation.

To characterize the collaborative relationship and intrinsic structure of the HSI data, we proposed a novel supervised spectral DR method, termed local constrained manifold structure collaborative preserving embedding (LMSCPE) for HSI classification. The LMSCPE method makes full use of collaborative relationship and local neighborhood information of HSI to extract discriminant features for classification. The main contributions of this paper are listed as below: (1) Based on the collaborative representation theory, we proposed a novel local constrained CR model, which can obtain more effective collaborative coefficients to characterize the relationship between samples pairs. (2) According to the graph embedding frame, an intraclass collaborative graph and an interclass collaborative graph are constructed to enhance the intraclass compactness and the interclass separability. (3) To preserve the local neighborhood structure of HSI, a local neighborhood graph is constructed by its *k*-nearest neighbors, and the aggregation of HSI data can be improved. (4) An optimal objective function is designed to obtain a discriminant projection matrix, it can extract the discriminant features and subsequently improve the classification performance of HSI.

The paper organization is as follows: Section 2 gives a brief description of GE and collaborative representation (CR). Section 3 describes the details of the proposed LMSCPE method. Section 4 gives the parameter analysis of LMSCPE to achieve the best classification performance. Section 5 provides some analysis and discussion about the experiment results. Finally, we summarize this paper and give recommendations for future work in Section 6.

2. Related Works

For convenience, let us denote a HSI data set with *D* bands by $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$, where *N* is the pixels number of HSI. Suppose the class label of *i*-th pixel can be represented as $l_i \in \{1, 2, \dots, c\}$, where *c* is the number of total classes in HSI. For DR methods, the goal is to map $\mathbf{X} \in \mathbb{R}^{D \times N}$ into $\mathbf{Y} \in \mathbb{R}^{d \times N}$, where $d \ll D$ is the embedding dimensionality. For the linear DR methods, the low-dimensional embedding features $\mathbf{Y} \in \mathbb{R}^{d \times N}$ can be computed as $\mathbf{Y} = \mathbf{V}^T \mathbf{X}$, where $\mathbf{V} \in \mathbb{R}^{D \times d}$ is the projection matrix.

2.1. Graph Embedding

tr

The graph embedding (GE) framework helps to redefine most DR algorithms in a unified framework, it characterizes the statistical and geometric properties of data by intrinsic graph $\mathbf{G} = \{\mathbf{X}, \mathbf{W}\}$ and penalty graph $\mathbf{G}^P = \{\mathbf{X}, \mathbf{W}^P\}$, 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 weight matrices [52,53]. In GE, intrinsic graph \mathbf{G} is adopted to describe similarities relationship of different vertexes, and penalty graph \mathbf{G}^P is employed to reveal dissimilarities properties between vertex pairs.

The GE framework aims to represent a graph in a low dimensional space which preserves as much graph property information as possible. The optimal objective function can be given by a graph preserving criterion:

$$\min_{(\mathbf{Y}^T \mathbf{H} \mathbf{Y})=h} \frac{1}{2} \sum_{i \neq j} \left\| \mathbf{y}_i - \mathbf{y}_j \right\|^2 w_{ij} = \min_{\mathrm{tr}(\mathbf{Y}^T \mathbf{H} \mathbf{Y})=h} \mathrm{tr}(\mathbf{Y}^T \mathbf{L} \mathbf{Y})$$
(1)

in which **L** is the Laplacian matrix of intrinsic graph **G**, **H** is the constraint matrix for avoiding a trivial solution, and it is typically set as a diagonal matrix or the Laplacian matrix of penalty graph \mathbf{G}^{P} . Then, **L** and \mathbf{L}^{P} can be given by

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \quad \mathbf{D}_{ii} = \sum_{j \neq i}^{N} w_{ij}, \quad \mathbf{W} = \left[w_{ij}\right]_{i,j=1}^{N}$$
(2)

$$\mathbf{L}^{P} = \mathbf{D}^{P} - \mathbf{W}^{P}, \quad \mathbf{D}_{ii}^{P} = \sum_{j \neq i}^{N} w_{ij}^{P}, \quad \mathbf{W}^{P} = \left[w_{ij}^{P}\right]_{i,j=1}^{N}$$
(3)

2.2. Collaborative Representation

In the hyperspectral imagery community, many representative graph learning methods have been proposed for preserving intrinsic structure information of HSI data, these methods often employ sparse representation coefficients to characterize the relationship of different samples. However, sparse representation is an iterative procedure, which usually suffers from computational expensiveness and the solution is often sub-optimal. Therefore, the problem of collaborative representation (CR) has received considerable attentions recently [54–56].

The basic idea of CR is that a query sample can be reconstructed by a set of training samples, and the reconstructive coefficients achieved by the l_2 -norm optimization can be treated as the affinities between the query sample and other training samples. For each pixel \mathbf{x}_i in HSI, the typical collaborative representation can be formulated as the following l_2 -norm optimization problem:

$$\min_{\mathbf{x}_i} \|\mathbf{\alpha}_i\|_2 \text{ such that } \|\mathbf{x}_i - \mathbf{Z}_i \mathbf{\alpha}_i\|_2 \le \epsilon$$
(4)

in which \mathbf{Z}_i is the $d \times (N-1)$ dimensional dictionary excluding \mathbf{x}_i itself, $\epsilon > 0$ is a small tolerance, and $\boldsymbol{\alpha}_i = [\boldsymbol{\alpha}_{i,1}, \dots, \boldsymbol{\alpha}_{i,N-1}]^T \in \mathbf{R}^{N-1}$ is the collaborative representation coefficients with size of (N-1). Then, the objective function of CR can be reformulated as

$$\min_{\boldsymbol{\alpha}_i} \|\mathbf{x}_i - \mathbf{Z}_i \boldsymbol{\alpha}_i\|_2^2 + \gamma \|\boldsymbol{\alpha}_i\|_2^2$$
(5)

where γ is a tuning parameter to balance the residual term $\|\mathbf{x}_i - \mathbf{Z}_i \boldsymbol{\alpha}_i\|^2$ and regularization term $\|\boldsymbol{\alpha}_i\|^2$ of the collaborative representation.

With some mathematical operations, the collaborative representation vector α_j can be calculated as

$$\boldsymbol{\alpha}_{i} = \left(\mathbf{Z}_{i}^{\mathrm{T}}\mathbf{Z}_{i} + \gamma \mathbf{I}\right)^{-1} \mathbf{Z}_{i}^{\mathrm{T}} \mathbf{x}_{i}$$
(6)

in which **I** is an identity matrix of size $(N - 1) \times (N - 1)$.

3. Local Constrained Manifold Structure Collaborative Preserving Embedding

To characterize the discriminative properties and intrinsic structure of the HSI data, a local constrained manifold structure collaborative preserving embedding (LMSCPE) method was proposed for DR. LMSCPE first designs a novel CR model to discover the collaborative relationship between different samples that belong to the same class. Based on the collaborative representation coefficients, it constructs an intraclass collaborative graph and an interclass collaborative graph to characterize the intraclass compactness and the interclass separability. Then, it selects neighbors to construct a local neighborhood graph, which can effectively preserve the local geometric structure of HSI. After that, the collaborative graph and local neighborhood graph are incorporated to learn an effective projection matrix. LMSCPE can preserve the local neighborhood structure in HSI and enhance the discrimination power of embedding features. The flowchart of the proposed LMSCPE method is shown in Figure 1.



Figure 1. The flowchart of the proposed local constrained manifold structure collaborative preserving embedding (LMSCPE) method.

3.1. Local Constrained Collaborative Graph Analysis Model

Due to the nonlinear optical effects of spectrum transmitting in the atmosphere, such as reflection, absorption, and dispersion, the spectral curves of pixels of the same category usually emerge subtle differences, which will affect the final classification performance of HSI [57–59]. Existing graph learning algorithms usually employ a distance-based weight matrix to reflect the similarity relationship between sample pairs, and construct a graph structure to map the high-dimensional data into a low-dimensional space for DR [60,61]. However, the subtle differences in spectral curves often cause the weight matrix inaccurate [62]. Therefore, according to the CR theory, we design an LCCR model to obtain more accurate weight coefficients.

In the proposed LCCR model, we incorporate locality-constrained terms into collaborative representation, and the minimization problem can be formulated as

$$\arg\min_{\boldsymbol{\alpha}_{i}} \|\mathbf{x}_{i} - \mathbf{Z}_{i}\boldsymbol{\alpha}_{i}\|_{2}^{2} + \gamma \|\boldsymbol{\Gamma}_{i}\boldsymbol{\alpha}_{i}\|_{2}^{2} + \delta \|\mathbf{x}_{i} - \hat{\mathbf{Z}}_{i}\boldsymbol{\alpha}_{i}\|_{2}^{2}$$
(7)

in which γ and δ are two regularization parameters that balance the minimization between regularization terms and the residual part. $\hat{\mathbf{Z}}_i \in \mathbf{R}^{d \times N}$ consists of the *k*-nearest neighbors of \mathbf{x}_i and (N - k) zero vectors $\mathbf{0} \in \mathbf{R}^{d \times (N-k)}$, Γ_i is a biasing Tikhonov matrix, it can be given by

$$\boldsymbol{\Gamma}_{i} = \begin{bmatrix} \|\mathbf{x}_{1} - \mathbf{Z}_{i}\|_{2} & 0 \\ & \ddots & \\ 0 & \|\mathbf{x}_{N-1} - \mathbf{Z}_{i}\|_{2} \end{bmatrix}$$
(8)

where $\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{N-1}$ are the columns of matrix \mathbf{Z}_i .

In (7), the introduction of Γ_i in the first regularization term can adjust the collaborative coefficient adaptively based on the distance between sample pairs, and the second regularization term is designed to explore the local geometric structure of HSI, which can effectively enhance the aggregation of HSI data.

According to LCCR model, we construct the intraclass collaborative graph and the interclass collaborative graph to characterize the intraclass compactness and the interclass separability. For intraclass collaborative graph, the weight matrix is set as the collaborative coefficients between each point and the points from the same class. Denote the samples of the *i*th class as $\mathbf{X}_i = [\mathbf{x}_i^1, \dots, \mathbf{x}_i^{N_i}]$, where N_i is the samples number of the *i*th class. Then, the intraclass dictionary \mathbf{Z}_i^{jw} of \mathbf{x}_i^j can be given by

$$\mathbf{Z}_{i}^{jw} = \mathbf{X}_{i} / \mathbf{x}_{i}^{j} = \left[\mathbf{x}_{i}^{1}, \dots, \mathbf{x}_{i}^{j-1}, \mathbf{x}_{i}^{j+1}, \dots, \mathbf{x}_{i}^{N_{i}}\right] \in \mathbf{R}^{D \times (N_{i}-1)}$$
(9)

Therefore, the intraclass collaborative representation coefficients of x_i^j can be solved by

$$\boldsymbol{\alpha}_{i}^{jw} = \operatorname*{arg\,min}_{\boldsymbol{\alpha}_{i}^{jw}} \left\{ \left\| \mathbf{x}_{i}^{j} - \mathbf{Z}_{i}^{jw} \boldsymbol{\alpha}_{i}^{jw} \right\|^{2} + \gamma \left\| \Gamma_{i}^{w} \boldsymbol{\alpha}_{i}^{jw} \right\|^{2} + \delta \left\| \mathbf{x}_{i}^{j} - \hat{\mathbf{Z}}_{i}^{jw} \boldsymbol{\alpha}_{i}^{jw} \right\|_{2}^{2} \right\}$$
(10)

With some mathematical operations, the collaborative representation coefficients α_i^{jw} can be calculated as

$$\boldsymbol{\alpha}_{i}^{jw} = [(\mathbf{Z}_{i}^{jw})^{T} \mathbf{Z}_{i}^{jw} + \gamma (\mathbf{\Gamma}_{i}^{w})^{T} \mathbf{\Gamma}_{i}^{w} + \delta (\mathbf{\hat{Z}}_{i}^{jw})^{T} \mathbf{\hat{Z}}_{i}^{jw}]^{-1} [\delta (\mathbf{\hat{Z}}_{i}^{jw})^{T} + (\mathbf{Z}_{i}^{jw})^{T}] \mathbf{x}_{i}^{j}$$
(11)

in which Γ_i^w is the Tikhonov matrix of \mathbf{x}_i^j , it can be given by

$$\mathbf{\Gamma}_{i} = \begin{bmatrix} \left\| \mathbf{x}_{i}^{1} - \mathbf{Z}_{i}^{jw} \right\|_{2} & 0 \\ & \ddots & \\ 0 & \left\| \mathbf{x}_{i}^{N_{i}-1} - \mathbf{Z}_{i}^{jw} \right\|_{2} \end{bmatrix}$$
(12)

where $\mathbf{x}_{i}^{1}, \mathbf{x}_{i}^{2}, ..., \mathbf{x}_{i}^{N_{i}-1}$ are the columns of matrix \mathbf{Z}_{i}^{jw} .

After obtaining the collaborative coefficients of each class, the intraclass weight matrix \mathbf{W}_s can be constructed as

$$\mathbf{W}_{s} = \begin{bmatrix} \mathbf{W}_{s1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{s2} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{W}_{sc} \end{bmatrix}$$
(13)

in which **0** is the zero matrix and \mathbf{W}_{si}^{jk} is defined as

$$\mathbf{W}_{si}^{jk} = \begin{cases} 0, & k = j \\ \mathbf{\alpha}_{i}^{jkw}, & k < j \\ \mathbf{\alpha}_{i}^{j(k-1)w}, & k > j. \end{cases}$$
(14)

where α_i^{jkw} is the *k*th element of α_i^{jw} , and \mathbf{W}_{si}^{jk} is the *j*th row and *k*th column of matrix \mathbf{W}_{si} .

For interclass collaborative graph, the interclass dictionary \mathbf{Z}_i^{jb} of \mathbf{x}_i^j is designed as

$$\mathbf{Z}_{i}^{jb} = \mathbf{X}/\mathbf{X}_{i} = \left[\mathbf{x}_{1}^{1}, \dots, \mathbf{x}_{i-1}^{N_{i-1}}, \mathbf{x}_{i+1}^{1}, \dots, \mathbf{x}_{c}^{N_{c}}\right] \in \mathbf{R}^{D \times (N-N_{i})}$$
(15)

where $\mathbf{x}_{c}^{N_{c}}$ is the N_{c} th sample in the c class.

Then, the interclass collaborative representation coefficients of \mathbf{x}_{i}^{j} can be given as

$$\boldsymbol{\alpha}_{i}^{jb} = \arg\min_{\boldsymbol{\alpha}_{i}^{jb}} \left\{ \left\| \mathbf{x}_{i}^{j} - \mathbf{Z}_{i}^{jb} \boldsymbol{\alpha}_{i}^{jb} \right\|^{2} + \gamma \left\| \boldsymbol{\Gamma}_{i}^{b} \boldsymbol{\alpha}_{i}^{jb} \right\|^{2} + \delta \left\| \mathbf{x}_{i}^{j} - \hat{\mathbf{Z}}_{i}^{jb} \boldsymbol{\alpha}_{i}^{jb} \right\|_{2}^{2} \right\}$$
(16)

With some mathematical operations, the collaborative representation coefficients α_i^{jb} can be calculated as

$$\boldsymbol{\alpha}_{i}^{jb} = [(\mathbf{Z}_{i}^{jb})^{T} \mathbf{Z}_{i}^{jb} + \gamma (\mathbf{\Gamma}_{i}^{b})^{T} \mathbf{\Gamma}_{i}^{b} + \delta (\hat{\mathbf{Z}}_{i}^{jb})^{T} \hat{\mathbf{Z}}_{i}^{jb}]^{-1} [\delta (\hat{\mathbf{Z}}_{i}^{jb})^{T} + (\mathbf{Z}_{i}^{jb})^{T}] \mathbf{x}_{i}^{j}$$
(17)

where Γ_i^b has the similar form with Γ_i^w .

Considering that the CR coefficients can effectively characterize the similarity relationship between sample pairs, we adopt the coefficients to construct the interclass weight matrix \mathbf{W}_b in the following way:

$$\mathbf{W}_{b} = \left[(\mathbf{W}_{b1})^{T}, (\mathbf{W}_{b2})^{T}, \dots, (\mathbf{W}_{bc})^{T} \right]^{T}$$
(18)

where

$$\mathbf{W}_{bi}^{jk} = \begin{cases} \left| \mathbf{\alpha}_{i}^{jkb} \right|, & 0 < k \leq \sum_{t=1}^{i-1} N_{t} \\ 0, & \sum_{t=1}^{i-1} N_{t} < k \leq \sum_{t=1}^{i} N_{t} \\ \left| \mathbf{\alpha}_{i}^{j(k-N_{i})b} \right|, & \sum_{t=1}^{i} N_{t} < k \leq \sum_{t=1}^{c} N_{t}. \end{cases}$$
(19)

i = 1

in which α_i^{jkb} is the *k*th element of α_i^{jb} , and \mathbf{W}_{bi}^{jk} is the *j*th row and *k*th column of matrix \mathbf{W}_{bi} . Therefore, based on the intraclass collaborative graph $\mathbf{G}^w(\mathbf{X}, \mathbf{W}_s)$, the intraclass com-

pactness in the reduced subspace can be defined as

$$J^{w} = \sum_{i=1}^{N} \left\| \mathbf{V}^{T} \mathbf{x}_{i} - \sum_{j=1}^{N} (\mathbf{W}_{s})_{ij} \mathbf{V}^{T} \mathbf{x}_{j} \right\|^{2}$$

$$= \operatorname{tr} \left(\mathbf{V}^{T} \left(\sum_{i=1}^{N} \left(\mathbf{x}_{i} - \sum_{j=1}^{N} (\mathbf{W}_{s})_{ij} \mathbf{x}_{j} \right) \times \left(\mathbf{x}_{i} - \sum_{j=1}^{N} (\mathbf{W}_{s})_{ij} \mathbf{x}_{j} \right)^{T} \right) \mathbf{V} \right)$$

$$= \operatorname{tr} \left(\mathbf{V}^{T} \left(\mathbf{X} (\mathbf{I} - \mathbf{W}_{s}) (\mathbf{I} - \mathbf{W}_{s})^{T} \mathbf{X}^{T} \right) \mathbf{V} \right)$$

$$= \operatorname{tr} \left(\mathbf{V}^{T} \mathbf{X} \mathbf{M}^{w} \mathbf{X}^{T} \mathbf{V} \right)$$
(20)

where $\mathbf{M}^w = \mathbf{I} - \mathbf{W}_s - \mathbf{W}_s^T + \mathbf{W}_s \mathbf{W}_s^T$ and \mathbf{I} is identity matrix of size $N \times N$.

$$J^{b} = \sum_{i=1}^{N} \left\| \mathbf{V}^{T} \mathbf{x}_{i} - \sum_{j=1}^{N} (\mathbf{W}_{b})_{ij} \mathbf{V}^{T} \mathbf{x}_{j} \right\|^{2}$$

$$= \operatorname{tr} \left(\mathbf{V}^{T} \left(\sum_{i=1}^{N} \left(\mathbf{x}_{i} - \sum_{j=1}^{N} (\mathbf{W}_{b})_{ij} \mathbf{x}_{j} \right) \times \left(\mathbf{x}_{i} - \sum_{j=1}^{N} (\mathbf{W}_{b})_{ij} \mathbf{x}_{j} \right)^{T} \right) \mathbf{V} \right)$$

$$= \operatorname{tr} \left(\mathbf{V}^{T} \left(\mathbf{X} (\mathbf{I} - \mathbf{W}_{b}) (\mathbf{I} - \mathbf{W}_{b})^{T} \mathbf{X}^{T} \right) \mathbf{V} \right)$$

$$= \operatorname{tr} \left(\mathbf{V}^{T} \mathbf{X} \mathbf{M}^{b} \mathbf{X}^{T} \mathbf{V} \right)$$
(21)

where $\mathbf{M}^b = \mathbf{I} - \mathbf{W}_b - \mathbf{W}_b^T + \mathbf{W}_b \mathbf{W}_b^T$.

For the purpose of seeking an optimal projection, it is natural to minimize the intraclass compactness and maximize the interclass variance simultaneously, that is

$$\begin{pmatrix} \min_{\mathbf{V}} \operatorname{tr} \left(\mathbf{V}^{T} \mathbf{X} \mathbf{M}^{w} \mathbf{X}^{T} \mathbf{V} \right) \\ \max_{\mathbf{V}} \operatorname{tr} \left(\mathbf{V}^{T} \mathbf{X} \mathbf{M}^{b} \mathbf{X}^{T} \mathbf{V} \right)
\end{cases}$$
(22)

3.2. Local Neighborhood Graph Analysis Model

Due to the fact that the spectral curves of HSI pixels are easily affected by external environment and imaging equipment, the actually obtained spectral curves of each category exhibits a certain degree of difference, which will lead to a degraded classification performance [63–65]. To eliminate this impact, we construct a local neighborhood graph by considering the spectral similarity of pixels, and it explores the projection relationships from high dimensional space to a lower dimensional space by aggregating the local graph structure and separating the total data.

In local neighborhood graph G_L , *i*th point \mathbf{x}_i is connected with its intraclass neighbor points that are from the same class. Denote the neighbor set of \mathbf{x}_i as $C(\mathbf{x}_i) = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{ik}]$, where *k* is the neighbor number. Then, the similarity weight w_{ij}^w between \mathbf{x}_i and \mathbf{x}_j can be given as following:

$$w_{ij}^{w} = \begin{cases} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2t_i^2}\right), \text{ if } \mathbf{x}_i \in C(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in C(\mathbf{x}_i) \text{ and } l_i = l_j \\ 0, \text{ otherwise} \end{cases}$$
(23)

where $t_i = \frac{1}{k} \sum_{j=1}^{k} || \mathbf{x}_i - \mathbf{x}_j ||$.

To enhance the aggregation of data on local neighborhood structure, each point and its neighbor points are adopted to formulate the optimization problem in low-dimensional embedding space

$$J_{1}(\mathbf{V}) = \arg\min\sum_{i=1}^{N}\sum_{j=1}^{k} \left\| \mathbf{y}_{i} - \mathbf{y}_{j} \right\|^{2} w_{ij}^{w}$$

$$= \sum_{i=1}^{N}\sum_{j=1}^{k} \left\| \mathbf{V}^{T} \mathbf{x}_{i} - \mathbf{V}^{T} \mathbf{x}_{j} \right\|^{2} w_{ij}^{w}$$

$$= \operatorname{tr} \left(\mathbf{V}^{T} \left[\sum_{i=1}^{N}\sum_{j=1}^{k} (\mathbf{x}_{i} - \mathbf{x}_{j}) (\mathbf{x}_{i} - \mathbf{x}_{j})^{T} w_{ij}^{w} \right] \mathbf{V} \right)$$

$$= \operatorname{tr} \left(\mathbf{V}^{T} \mathbf{S} \mathbf{V} \right)$$
(24)

in which $\mathbf{S} = \sum_{i=1}^{N} \sum_{j=1}^{k} (\mathbf{x}_i - \mathbf{x}_j) (\mathbf{x}_i - \mathbf{x}_j)^T w_{ij}^w$ is the local neighborhood scatter matrix of training set.

In addition, aiming to separate the local graph structure of each pixel as far as possible, the optimization problem between different classes can be designed as follows.

$$J_{2}(\mathbf{V}) = \arg \max \sum_{i=1}^{N} \left\| \mathbf{V}^{T} \mathbf{x}_{i} - \mathbf{V}^{T} \overline{\mathbf{x}} \right\|^{2} c_{i}$$

$$= \sum_{i=1}^{N} \mathbf{V}^{T} (\mathbf{x}_{i} - \overline{\mathbf{x}}) c_{i} (\mathbf{x}_{i} - \overline{\mathbf{x}})^{T} \mathbf{V}$$

$$= \operatorname{tr} \left(\mathbf{V}^{T} \left[\sum_{i=1}^{N} (\mathbf{x}_{i} - \overline{\mathbf{x}}) c_{i} (\mathbf{x}_{i} - \overline{\mathbf{x}})^{T} \right] \mathbf{V} \right)$$

$$= \operatorname{tr} \left(\mathbf{V}^{T} \mathbf{H} \mathbf{V} \right)$$
(25)

where $\overline{\mathbf{x}}$ is the mean of training set, $c_i = \exp\left(-\|\mathbf{x}_i - \mathbf{x}\|^2/2t_{ik}^2\right)$ is the weight between \mathbf{x}_i and $\overline{\mathbf{x}}$, $\mathbf{H} = \sum_{i=1}^{N} (\mathbf{x}_i - \overline{\mathbf{x}})c_i (\mathbf{x}_i - \overline{\mathbf{x}})^T$ is the total scatter matrix.

To explore the local neighborhood structure in low-dimensional embedding space, we minimize the local neighborhood scatter matrix and maximize the total scatter matrix. Therefore, the optimal projection matrix \mathbf{V} should satisfy the following two optimization criteria:

$$\begin{cases} \arg\min_{\mathbf{V}} \mathbf{V}^T \mathbf{S} \mathbf{V} \\ \arg\max_{\mathbf{V}} \mathbf{V}^T \mathbf{H} \mathbf{V} \end{cases}$$
(26)

For high dimensional data, the collaborative relationship and local neighborhood structure between pixel pairs should be discovered simultaneously. Therefore, based on optimization problem of (22) and (26), we propose a local constrained manifold structure collaborative preserving embedding (LMSCPE) method for DR of HSI data. This method can reveal the collaborative relationship and the local neighborhood structure for learning a more effective projection, and the optimal objective function of LMSCPE can be designed as

$$J(\mathbf{V}) = \min \frac{\operatorname{tr}\{\mathbf{V}^{T}[a\mathbf{X}\mathbf{M}^{w}\mathbf{X}^{T} + (1-a)\mathbf{S}]\mathbf{V}\}}{\operatorname{tr}\{\mathbf{V}^{T}[a\mathbf{X}\mathbf{M}^{b}\mathbf{X}^{T} + (1-a)\mathbf{H}]\mathbf{V}\}}$$
(27)

where $a \in [0, 1]$ is a trade-off parameter that balances the contributions of collaborative graph and local neighborhood graph in embedding space. Then, the optimization problem of (27) can be transformed as:

$$\begin{cases} \min_{\mathbf{V}} \mathbf{V}^{T} [a\mathbf{X}\mathbf{M}^{w}\mathbf{X}^{T} + (1-a)\mathbf{S}]\mathbf{V} \\ \text{s.t.} \quad \mathbf{V}^{T} [a\mathbf{X}\mathbf{M}^{b}\mathbf{X}^{T} + (1-a)\mathbf{H}]\mathbf{V} = \mathbf{B} \end{cases}$$
(28)

where **B** is a constant matrix.

With the Lagrangian multiplier method, the solution of (28) can be obtained through the following generalized eigenvalue problem:

$$\left[a\mathbf{X}\mathbf{M}^{w}\mathbf{X}^{T} + (1-a)\mathbf{S}\right]\mathbf{v}_{i} = \lambda_{i}\left[a\mathbf{X}\mathbf{M}^{b}\mathbf{X}^{T} + (1-a)\mathbf{H}\right]\mathbf{v}_{i}$$
(29)

in which λ_i is the *i*th eigenvalue and \mathbf{v}_i is the corresponding eigenvector, the eigenvectors \mathbf{v}_1 , $\mathbf{v}_2, \ldots, \mathbf{v}_d$ corresponding to the first *d* eigenvalues construct the optimal projection matrix $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \dots \mathbf{v}_d] \in \mathbf{R}^{D \times d}$. Then, the low-dimensional embedding \mathbf{Y} can be represented as $\mathbf{Y} = \mathbf{V}^T \mathbf{X}$.

4. Experimental Setup and Parameters Discussion

In this section, three public HSI data sets are adopted to demonstrate the effectiveness of LMSCPE by comparing it with some state-of-the-art DR algorithms.

4.1. Data Set Description

PaviaU data set: This data set was acquired by the ROSIS-03 sensor over the University of Pavia, Italy. The spatial size of the scene is 610×340 with the spatial resolution of 1.3 m/pixel, and each pixel contains 115 spectral bands ranging from 0.43 to 0.86 μ m. Due to 12 bands suffering from water absorption, we adopted the remaining 103 bands for experiments. Figure 2 shows the false color image, ground truth (GT) and spectral curves of PaviaU data set.



Figure 2. PaviaU hyperspectral image. (**a**) False color image. (**b**) Ground truth. (**c**) Spectral curves. (Note that the number of samples for each class is shown in brackets).

LongKou data set: This data set was captured by Headwall Nano-Hyperspec imaging sensor on a UAV platform in Longkou Town, Hubei province, China. The full scene contains 550×400 pixels, each pixel contains 270 spectral bands ranging from 400 to 1000 nm. The data set possesses 9 different land cover types, and the spatial resolution is about 0.463 m. Figure 3 shows the false color image, ground truth (GT) and spectral curves of LongKou data set.



Figure 3. LongKou hyperspectral image. (**a**) False color image. (**b**) Ground truth. (**c**) Spectral curves. (Note that the number of samples for each class is shown in brackets).

MUUFL data set: This data set was captured by the ITRES CASI-1500 sensor over the University of Southern Mississippis-Gulfpark Campus. The full scene contains 325×220 pixels with a spatial resolution of 1 m, and the number of spectral bands in this dataset is 72. After removing 8 bands affected by noise, we adopt the remaining 64 spectral bands for classification. Figure 4 shows the false color image, ground truth (GT) and spectral curves of the MUUFL data set.



Figure 4. MUUFL hyperspectral image. (**a**) False color image. (**b**) Ground truth. (**c**) Spectral curves. (Note that the number of samples for each class is shown in brackets).

4.2. Experimental Setup

In this section, the HSI data were randomly divided into training and test sets. The training set was adopted to learn a DR model, while the test set was used to verify the validity of the model. Then, we employed the nearest neighbor (NN) classifier for classification. After that, four performance indicators were used to evaluate the effectiveness of the algorithms, such as classification accuracy of each class (CA), overall classification accuracy (OA), average classification accuracy (AA), and kappa coefficient (KC) [66–68]. For robustness, all experiments were repeated 10 times under the same conditions.

In experiments, we compared LMSCPE with some state-of-the-art DR algorithms, including principal component analysis (PCA), locality preserving projection (LPP), linear discriminant analysis (LDA), local Fisher Discriminant Analysis (LFDA), local geometric structure Fisher analysis (LGSFA), sparse graph based discriminant analysis (SGDA), collaborative graph-based discriminant analysis (CGDA), and low-rank graph-based discriminant analysis (LGDA). Among them, PCA and LPP are unsupervised DR algorithms, they only consider the spectrum similarity between different pixels. The later six algorithms are supervised DR algorithms, which simultaneously consider the spectrum similarity and category information of pixels. Besides, the RAW method was added to compare with the DR algorithms, it indicates that the test set was classified by the NN classifier directly without DR process.

To demonstrate the effectiveness of the proposed LMSCPE algorithm, the parameters of all the DR models were optimized to achieve higher classification accuracies. For LPP and LFDA, the neighbors number was set to 7. For LGSFA, the numbers of intraclass neighbor and interclass neighbor were set as $k_1 = 7$ and $k_2 = 7$, respectively.

4.3. Analysis of Neighbors Number k

To analyze the influence of different neighbor number k on classification performance, we randomly selected 50 labeled samples from each class to form the training set, and the remaining samples construct the testing set. The corresponding experimental results on these three different HSI data sets are given in Figure 5, in which parameter k varies from 1 to 15 with an interval of 1.

As shown in Figure 5, parameter k has similar change trends on three different HSI data sets. That is, as the neighbor number k increases, the OAs first improve and then decrease with the increase of k. The reason is that a larger number of neighbor points possesses richer spectral information, and the neighbors can be utilized to construct a more effective graph learning model, so it has more powerful discrimination capabilities for complex pixels in HSIs. However, when the neighbors number k is too large, it will produce too much useless information for adjacency graph construction, which will degrade the final classification performance of HSI. Therefore, the classification accuracies of LMSCPE will remain stable or even slightly decrease. Besides, a larger spectral neighbor set will lead to high computational cost, which requires huge computational resources and storage

capacity. To balance the running time and classification performance, we set k = 5 for the PaviaU data set and k = 7 for both the LongKou data set and MUUFL data set in the following experiments.



Figure 5. Classification results with different neighbors number of *k* on three hyperspectral image (HSI) data sets.

4.4. Analysis of Regularization Parameters γ and δ

To evaluate the classification performance of the proposed LMSCPE method under different regularization parameters of γ and δ , 50 labeled samples per class were selected as the training set, and the remaining samples were used as the testing set. Parameters γ and δ were tuned with a set of {0, 10, 20, ..., 110, 120} and a set of {0, 1, 2, ..., 9, 10}, respectively. Figure 6 shows the classification results with different regularization parameters γ and δ .



Figure 6. Classification results with different regularization parameters γ and δ . (a) PaviaU. (b) LongKou. (c) MUUFL.

From Figure 6, we can see that parameter γ has a strong influence on the classification performance of LMSCPE, which indicates that Tikhonov regularization helps the CR model to obtain more appropriate collaborative coefficients for constructing adjacency graphs. That is, Tikhonov matrix is an effective distance-based measurement between sample pairs, it can adjust the collaborative coefficients of each pixel adaptively. When two pixels in HSI have a similar spectrum, then the distance between them will be small, the introduction of the Tikhonov regularization will cause the corresponding weight coefficient to get bigger, thereby enhancing the effectiveness of adjacency graph structure and improving the discrimination ability of the DR model. Besides, an increasing δ leads to a subtle improvement on classification accuracy, because it can explore the local neighborhood

structure and enhance the aggregation of HSI data. To achieve the best classification accuracies, we set γ and δ to 60 and 4 for the PaviaU data set, 70 and 3 for the LongKou and MUUFL data sets in the experiments.

4.5. Analysis of Trade-Off Parameter a

In the experiments, the trade-off parameter *a* will also affect the classification performance of LMSCPE, it is mainly employed to balance the contributions between the collaborative graph and local neighborhood graph. To investigate the classification performance with different trade-off parameter *a*, we randomly selected 50 labeled samples per class for training, and the remaining were for testing. To obtain the optimal classification performance, parameter *a* was chosen within a set of {0, 0.1, 0.2, 0.3, ..., 0.9, 1}. Figure 7 illustrates the classification results of LMSCPE with different trade-off parameter *a*.

According to Figure 7, the OAs first improve with the increase of *a* and then slightly decrease at the peak value, which indicates that both collaborative graph and local neighborhood graph contribute to enhance the classification performance. Aiming to achieve the best classification performance, we set *a* to 0.7 for the PaviaU and MUFFL data sets and 0.8 for the LongKou data set in the following experiments.





4.6. Investigation of Embedding Dimension d

For the above-mentioned DR methods, the value of the embedding dimension d determines the effectiveness of the low-dimensional embedding features, and has an important impact on the final classification performance. Therefore, we analyzed the change of OAs with respect to different embedding dimensions d in Figure 8. In the experiments, 50 labeled samples from each type of land covers were randomly selected as training samples, and the remaining were used as test samples.

From Figure 8, we can draw some similar conclusions on PaviaU, LongKou, and MUUFL data sets. As d varies from 1 to 40, the OAs of each DR method first increase and then remain stable, because a larger embedding dimension d will preserve richer discriminant features in the low dimensional embedding space, which can be utilized to enhance the classification performance, especially when there are few labeled training samples available in complex scenes. However, when the number of training samples is fixed, the available valuable information contained in the samples is limited. Therefore, the increase of d will not lead to the continuous increase of the OAs, but remain stable after it reaches a peak value. Besides, most DR methods outperform RAW in terms of OA, it indicates that these methods can effectively extract the discriminant features of HSI data, thereby enhancing the classification performance. considering that all the DR methods reach the peak value when d is larger than 20, we set d = 30 for all DR methods except



LDA in the experiments, and the dimension of LDA is *c*-1, *c* is the class number of the HSI data set.

Figure 8. Classification results with different embedding dimensions d. (a) PaviaU. (b) LongKou. (c) MUUFL.

5. Experimental Results and Discussion

5.1. Analysis of Training Sample Size

To analyze the influence of training sample size on the classification performance, n_i ($n_i = 20, 30, 40, 50, 60$) samples were randomly selected from each type of land covers as train set, and the remaining were used as test set. After that, the NN classifier was employed to predict the labels of test set, and each experiment was repeated ten times to reduce the random errors. Tables 1–3 present the quantitative results of different training sample size on the PaviaU, LongKou, and MUUFL data sets.

Table 1. Classification results of each method with different training sample sizes on the PaviaU dataset [OA \pm STD (%)(KC)]. (OA—overall classification accuracy, STD—standard deviation, KC—kappa coefficient, and the best results of a column are marked in bold).

Method	20	30	40	50	60
RAW	$68.77 \pm 1.74 \ (0.607)$	$70.48 \pm 1.93 \ (0.628)$	$71.21 \pm 2.04 \ (0.637)$	$72.79 \pm 1.03 \ (0.655)$	$73.29 \pm 0.76 \ (0.661)$
PCA	$68.75 \pm 1.73 \ (0.607)$	$70.44 \pm 1.94 \ (0.628)$	$71.23 \pm 1.97 \ (0.637)$	$72.77 \pm 1.01 \ (0.655)$	$73.28 \pm 0.79 \ (0.661)$
LPP	$66.12 \pm 0.91 \ (0.578)$	$70.91 \pm 2.90 \ (0.634)$	$72.56 \pm 1.48 \ (0.654)$	$74.58 \pm 0.89 \ (0.677)$	$76.13 \pm 1.05 \ (0.695)$
LDA	$67.28 \pm 4.54 \ (0.589)$	$72.53 \pm 2.21 \ (0.652)$	$75.38 \pm 0.85 (0.683)$	$77.36 \pm 1.10 \ (0.709)$	$77.57 \pm 1.56 \ (0.711)$
LFDA	$61.88 \pm 2.44 \ (0.525)$	$69.95 \pm 2.84 \ (0.622)$	$74.03 \pm 1.80 \ (0.670)$	$75.39 \pm 1.02 \ (0.687)$	$77.42 \pm 1.89 \ (0.711)$
LGSFA	$65.08 \pm 1.96 \ (0.560)$	$71.02 \pm 1.71 \ (0.634)$	$73.01 \pm 1.51 \ (0.657)$	$75.18 \pm 3.02 \ (0.683)$	75.60 ± 0.98 (0.686)
SGDA	$73.04 \pm 1.19 \ (0.659)$	$75.62 \pm 1.77 \ (0.690)$	76.13 ± 2.33 (0.696)	$78.43 \pm 1.29 \ (0.723)$	$79.06 \pm 1.15 \ (0.730)$
LGDA	$71.55 \pm 1.90 \ (0.642)$	$73.85 \pm 1.60 \ (0.669)$	$75.30 \pm 2.44 \ (0.682)$	$75.65 \pm 0.93 \ (0.690)$	$77.25 \pm 0.81 \ (0.709)$
CGDA	$72.22 \pm 2.60 \ (0.651)$	$75.20 \pm 1.03 \ (0.686)$	$76.03 \pm 2.62 \ (0.696)$	$77.20 \pm 1.44 \ (0.709)$	$78.95 \pm 1.47 \ (0.730)$
LMSCPE	76.81 \pm 1.55 (0.705)	$\textbf{78.11} \pm 1.53 \; \textbf{(0.722)}$	78.91 \pm 2.50 (0.732)	79.89 \pm 0.64 (0.742)	${\bf 81.62} \pm 2.19 \ {\bf (0.763)}$

column are marked in bold).												
Method	20	30	40	50	60							
RAW	$79.39 \pm 1.52 \ (0.740)$	81.39 ± 0.68 (0.765)	82.20 ± 0.84 (0.775)	82.49 ± 0.78 (0.778)	$83.53 \pm 0.89 \ (0.791)$							
PCA	$79.38 \pm 1.53 \ (0.740)$	$81.38 \pm 0.68 \ (0.764)$	$82.18 \pm 0.84 \ (0.774)$	$82.48 \pm 0.80 \ (0.778)$	$83.52 \pm 0.91 \ (0.791)$							
LPP	$66.70 \pm 1.81 \ (0.592)$	$73.26 \pm 0.39 \ (0.668)$	$76.66 \pm 1.38 \ (0.708)$	$80.06 \pm 1.69 \ (0.749)$	$82.45 \pm 0.67 \ (0.777)$							
LDA	$83.94 \pm 1.93 \ (0.795)$	$85.60 \pm 0.74 \ (0.816)$	$87.73 \pm 0.68 \ (0.843)$	$89.74 \pm 1.06 \ (0.868)$	$91.16 \pm 0.63 \ (0.886)$							
LFDA	$77.76 \pm 2.43 \ (0.719)$	$83.43 \pm 2.33 \ (0.789)$	$83.02 \pm 0.55 \ (0.784)$	$89.47 \pm 0.57 \ (0.865)$	$91.68 \pm 0.47 \ (0.893)$							
LGSFA	$83.46 \pm 0.66 \ (0.789)$	$83.54 \pm 2.16 \ (0.790)$	$85.78 \pm 1.13 \ (0.818)$	$90.12 \pm 0.36 \ (0.873)$	$91.79 \pm 0.40 \ (0.894)$							
SGDA	$87.47 \pm 2.04 \ (0.839)$	$89.05 \pm 0.65 \ (0.859)$	$89.93 \pm 0.91 \ (0.870)$	$90.59 \pm 0.65 \ (0.879)$	$91.62 \pm 0.44 \ (0.892)$							
LGDA	$84.73 \pm 2.62 \; (0.805)$	$85.02 \pm 0.62 \ (0.809)$	$85.50 \pm 1.23 \ (0.815)$	$85.79 \pm 0.50 \; (0.819)$	$86.77 \pm 0.53 \ (0.831)$							
CGDA	$88.03 \pm 2.82 \ (0.847)$	$89.61 \pm 0.94 \ (0.866)$	$89.92 \pm 1.01 \ (0.870)$	$90.17 \pm 0.72 \; (0.874)$	$90.99 \pm 0.69 \; (0.884)$							
LMSCPE	$\textbf{90.74} \pm 1.31 \ \textbf{(0.881)}$	91.28 ± 0.98 (0.887)	$91.57 \pm 0.73 \ \mathbf{(0.891)}$	$91.85 \pm 0.82 \ (0.895)$	$92.06 \pm 0.94 \ \mathbf{(0.897)}$							

Table 2. Classification results of each method with different training sample sizes on the LongKou dataset [OA \pm STD (%)(KC)]. (OA—overall classification accuracy, STD—standard deviation, KC—kappa coefficient, and the best results of a column are marked in bold).

As shown in Tables 1-3, the OAs of each algorithm significantly improve with the increase of training samples. The reason is that a larger number of training samples will produce more abundant feature information, which is helpful to distinguish the land covers with similar spectrum in complex scenes. In addition, the supervised DR methods of LDA, LFDA, LGSFA, SGDA, LGDA, CGDA, and LMSCPE perform better than unsupervised ones in most cases, which indicates that the category information of land covers contributes to improve the classification performance of the DR models. Among all the supervised DR algorithms, representation-based DR methods, SGDA, LGDA, CGDA, and LMSCPE, are superior to the other methods, because the sparse or collaborative coefficients of these methods can effectively characterize the sparse or collaborative relationship between sample pairs, which can be utilized as the weight coefficients of adjacency graphs, so they have great advantages over the traditional DR methods whose structure depends on pairwise Euclidean distance. For all the above-mentioned DR methods, LMSCPE achieves the best OA and KC in all cases, because it simultaneously considers the collaborative relationship and the local neighborhood structure of HSI, which can further improve the classification accuracy of HSI.

Table 3. Classification results of each method with different training sample sizes on the MUUFL dataset [OA \pm STD (%)(KC)]. (OA—overall classification accuracy, STD—standard deviation, KC—kappa coefficient, and the best results of a column are marked in bold).

Method	20	30	40	50	60
RAW	68.72 ± 3.60 (0.610)	$70.55 \pm 1.02 \ (0.632)$	$71.41 \pm 1.19 (0.641)$	$72.58 \pm 1.36 \; (0.655)$	$72.70 \pm 0.40 \ (0.656)$
PCA	$68.71 \pm 3.58 \ (0.610)$	70.55 ± 1.02 (0.632)	$71.42 \pm 1.17 \ (0.642)$	$72.57 \pm 1.32 \ (0.654)$	$72.74 \pm 0.40 \ (0.657)$
LPP	$63.08 \pm 4.46 \ (0.545)$	$67.32 \pm 1.70 \ (0.594)$	$68.86 \pm 2.25 \ (0.612)$	$70.90 \pm 1.29 \ (0.635)$	$72.16 \pm 1.01 \; (0.649)$
LDA	$62.23 \pm 4.26 \ (0.532)$	$65.65 \pm 1.70 \ (0.570)$	$65.76 \pm 2.23 \ (0.571)$	$66.89 \pm 0.77 \ (0.583)$	$67.04 \pm 1.18 \ (0.586)$
LFDA	$62.65 \pm 4.03 \ (0.537)$	$65.68 \pm 2.88 \ (0.574)$	$68.45 \pm 2.09 \ (0.606)$	$70.89 \pm 1.25 \ (0.636)$	$71.56 \pm 0.74 \ (0.643)$
LGSFA	$63.04 \pm 4.68 \ (0.544)$	$67.32 \pm 1.38 \ (0.593)$	$68.24 \pm 2.50 \ (0.605)$	$69.97 \pm 2.23 \ (0.625)$	$70.42 \pm 0.73 \ (0.630)$
SGDA	$70.03 \pm 3.78 \ (0.625)$	$72.61 \pm 1.10 \ (0.656)$	$73.32 \pm 0.95 (0.664)$	$73.87 \pm 1.36 \ (0.671)$	$74.89 \pm 0.42 \ (0.682)$
LGDA	$71.33 \pm 3.75 \ (0.640)$	$72.71 \pm 0.72 \ (0.658)$	$73.54 \pm 0.86 \ (0.667)$	$74.11 \pm 0.20 \ (0.673)$	$74.34 \pm 1.19 \ (0.676)$
CGDA	$71.12 \pm 3.52 \ (0.638)$	72.40 ± 0.98 (0.654)	$73.04 \pm 1.08 \ (0.661)$	$74.62 \pm 1.33 \ (0.679)$	74.71 ± 0.44 (0.680)
LMSCPE	72.43 \pm 4.31 (0.654)	73.96 \pm 1.62 (0.672)	$\textbf{74.54} \pm 1.83~\textbf{(0.679)}$	76.46 \pm 1.39 (0.702)	76.57 \pm 0.93 (0.708)

5.2. Analysis of Classification Results

Considering that the land cover types of HSI have the problem of sample imbalance in practical scenes, we constructed the training set by selecting a certain proportion of training samples from each class, and the rest samples were adopted as test set. In the experiments, we set the proportion to 1% for the PaviaU data set, 0.2% for the LongKou data set, and 1%

for the MUUFL data set. For convenience of comparison, we listed the CAs, OAs, AAs, and KCs of different algorithms in Tables 4–6, respectively, and Figures 9–11 present the corresponding classification maps for whole scene on these three different HSI data sets.

Table 4. Classification results (%) of each type of land covers (t = 1%) with the nearest neighbor (NN) classifier on the PaviaU data set. (AA—average classification accuracy, OA—overall classification accuracy, KC—kappa coefficient, and the best results of a row are marked in bold).

Class	Land Covers	Training	Test	RAW	PCA	LPP	LDA	LFDA	LGSFA	SGDA	LGDA	CGDA	LMSCPE
1	Asphalt	66	6565	84.54	84.42	84.74	84.49	69.72	84.04	85.35	84.78	88.26	91.36
2	Meadows	186	18,463	89.87	89.84	89.43	88.93	77.56	96.91	89.60	90.88	92.83	95.54
3	Gravel	21	2078	49.09	48.89	41.29	38.69	48.85	31.57	52.17	56.59	60.59	64.87
4	Trees	31	3033	75.70	75.77	79.59	86.25	92.65	88.16	75.37	78.37	78.57	85.23
5	Metal	13	1332	98.57	98.57	99.17	99.62	99.02	99.70	99.10	98.57	99.32	99.55
6	Soil	50	4979	54.25	54.35	53.99	55.79	66.98	38.24	56.92	61.54	66.94	69.65
7	Bitumen	13	1317	64.54	64.77	51.86	20.12	51.94	32.73	65.60	63.86	72.74	63.33
8	Bricks	37	3645	72.98	72.62	65.24	62.09	64.20	73.39	73.77	73.20	78.38	76.27
9	Shadows	10	937	99.89	99.89	99.79	88.26	99.89	98.40	99.89	99.15	99.36	99.79
	AA			76.60	76.57	73.90	69.36	74.53	71.46	77.53	78.55	81.89	82.84
	OA			80.09	80.05	78.76	77.56	73.99	80.28	80.66	81.97	84.95	87.16
	KC			73.29	73.23	71.55	70.02	66.59	72.92	74.10	75.87	79.84	82.76

From Tables 4–6, we can see that the proposed LMSCPE method possesses best OAs, AAs, and KCs for the three different HSI data sets, and it has the higher classification accuracies in most classes than the other algorithms. Besides, the classification maps of LMSCPE are smoother and have less Salt&Pepper noise on the homogeneous regions, especially in the classes of *Asphalt*, *Gravel*, and *Soil* for the PaviaU data set, *Cotton*, *Sesame*, and *Mixed weed* for the LongKou data set, and *Trees*, *Mixed ground surface*, and *Building shadow* for the MUUFL data set. The reason is that LMSCPE not only explores the pixel-wise collaborative relationship of HSI data but also reveals the local neighborhood structure, so it has a stronger ability to extract the discriminant features of HSI data and enhance the classification performance.

Table 5. Classification results (%) of each type of land covers (t = 0.2%) with the NN classifier on the LongKou data set. (AA—average classification accuracy, OA—overall classification accuracy, KC—kappa coefficient, and the best results of a row are marked in bold).

Class	Land Covers	Training	Test	RAW	PCA	LPP	LDA	LFDA	LGSFA	SGDA	LGDA	CGDA	LMSCPE
1	Corn	69	34,442	93.36	93.30	82.20	93.03	94.56	96.90	95.46	96.85	97.77	98.81
2	Cotton	17	8357	47.24	47.24	34.79	47.79	52.02	56.19	53.05	45.14	54.83	66.02
3	Sesame	10	3021	37.80	37.54	15.33	33.40	32.54	65.18	40.19	56.04	64.71	68.42
4	Broad-leaf soybean	126	63,086	87.33	87.32	80.98	88.89	90.57	95.36	89.74	90.09	92.22	96.67
5	Narrow-leaf soybean	10	4141	53.39	53.34	27.58	33.30	35.69	57.38	54.94	74.50	76.33	76.45
6	Rice	24	11,830	90.51	90.45	78.88	98.08	97.98	95.58	91.62	96.13	99.82	99.80
7	Water	134	66,922	99.93	99.93	99.95	99.92	99.91	99.98	99.93	99.94	99.91	99.91
8	Roads and houses	14	7110	75.09	75.08	60.03	62.57	67.33	79.54	76.17	83.36	85.54	85.23
9	Mixed weed	10	5219	28.68	28.66	29.70	60.80	57.14	58.59	34.68	50.89	59.11	65.97
	AA			68.15	68.10	56.60	68.64	69.75	78.30	70.64	76.99	81.14	84.14
	OA			87.67	87.65	81.30	88.47	89.52	92.84	89.33	90.91	92.78	95.01
	KC			83.80	83.77	75.47	84.78	86.13	90.52	85.96	88.03	90.49	93.42

Table 6. Classification results (%) of each type of land covers (t = 1%) with the NN classifier on the MUUFL data set.
(AA-average classification accuracy, OA-overall classification accuracy, KC-kappa coefficient, and the best results of a
row are marked in bold).

Class	Land Covers	Training	Test	RAW	PCA	LPP	LDA	LFDA	LGSFA	SGDA	LGDA	CGDA	LMSCPE
1	Trees	465	22,781	88.98	89.02	89.41	90.06	81.71	89.83	89.06	89.65	89.81	90.95
2	Mostly grass	85	4185	67.78	67.68	65.65	60.82	67.64	65.27	67.14	70.95	69.43	70.43
3	Mixed ground surface	138	6744	62.64	62.70	65.14	69.03	53.00	64.60	63.09	65.49	66.84	71.64
4	Dirt/sand	37	1789	58.13	58.08	64.88	62.39	80.53	72.84	57.52	62.94	67.87	71.24
5	Road	134	6553	86.60	86.66	83.56	72.01	66.89	77.92	86.27	88.17	87.34	87.70
6	Water	10	456	77.19	76.97	73.90	27.41	53.51	49.78	76.75	80.92	78.29	76.75
7	Building shadow	45	2188	61.28	60.83	57.49	30.08	52.92	41.75	60.83	62.78	64.36	65.31
8	Buildings	125	6115	77.78	77.70	75.17	69.96	64.45	76.21	77.97	82.03	83.20	82.10
9	Sidewalk	28	1357	43.25	43.25	44.42	32.82	50.69	41.06	43.76	45.95	43.03	41.79
10	Yellow curb	10	173	47.40	46.82	43.35	78.03	82.66	82.08	47.98	57.23	53.76	70.52
11	Cloth panels	10	259	88.80	88.80	89.96	94.98	94.59	94.98	88.42	89.19	87.64	89.58
	AA			69.08	68.96	68.45	62.51	68.05	68.76	68.98	72.30	71.96	74.36
	OA			78.70	78.69	78.42	74.98	70.84	77.39	78.69	80.65	80.93	82.21
	KC			72.06	72.04	71.65	66.61	62.44	70.05	72.04	74.64	74.97	76.60



Figure 9. Classification maps for the whole scene on the PaviaU data set. (a) False color image; (b) ground truth; (c) RAW; (d) PCA; (e) LPP; (f) LDA; (g) LFDA; (h) LGSFA; (i) SGDA; (j) LGDA; (k) CGDA; (l) LMSCPE.



Figure 10. Classification maps for the whole scene on the LongKou data set. (a) False color image; (b) ground truth; (c) RAW; (d) PCA; (e) LPP; (f) LDA; (g) LFDA; (h) LGSFA; (i) SGDA; (j) LGDA; (k) CGDA; (l) LMSCPE.



Figure 11. Classification maps for the whole scene on the MUUFL data set. (a) False color image; (b) ground truth; (c) RAW; (d) PCA; (e) LPP; (f) LDA; (g) LFDA; (h) LGSFA; (i) SGDA; (j) LGDA; (k) CGDA; (l) LMSCPE.

5.3. Analysis of Computational Efficiency

For the classification task, both classification accuracy and running time are important performance evaluation indicators. To analyze the computational complexity of LMSCPE, denote the neighbor number of each sample as k, the intraclass weight matrix \mathbf{W}_s and the

interclass weight matrix \mathbf{W}_b both take $O(N^2)$. The calculation of $\mathbf{X}\mathbf{M}^w\mathbf{X}^T$ and $\mathbf{X}\mathbf{M}^b\mathbf{X}^T$ are both $O(DN^2)$. The local neighborhood scatter matrix \mathbf{S} and the total scatter matrix \mathbf{H} are calculated with $O(Nk^2)$ and O(N), respectively. It takes $O(D^3)$ to solve the generalized eigenvalue problem of (29). Therefore, the final computational complexity of LMSCPE is $O(N^2 + Nk^2 + DN^2 + D^3)$, and it mainly depends on the size of the training samples, band number, and neighbor number.

Aiming to quantitatively evaluate the classification performance of the above-mentioned DR methods, we detailed the running time of all DR algorithms in Table 7, in which 50 labeled samples were randomly selected from each type of land covers for training, and the rest samples were for testing. All the experimental results were obtained on a personal computer, which had an i7-7800X CPU and 32-G memory, and the version of Windows system and MATLAB are 64-bit Windows 10 and 2016b, respectively.

 Table 7. Computational time (in seconds) of the different algorithms on the PaviaU, LongKou, and MUUFL data sets.

Dataset	PCA	LPP	LDA	LFDA	LGSFA	SGDA	LGDA	CGDA	LMSCPE
PaviaU	0.025	0.031	0.013	0.037	0.292	0.659	3.009	0.315	2.650
LongKou	0.014	0.124	0.023	0.040	0.616	0.308	1.655	0.112	0.887
MUUFL	0.031	0.042	0.012	0.019	0.356	0.287	2.038	0.125	1.038

As shown in Table 7, LMSCPE costs more time than the most DR methods, because LMSCPE simultaneously considers the pixel-wise collaborative relationship and the local neighborhood structure of HSI data, the whole process will slightly increase the computational complexity. However, the slight increase in running time and computational complexity is acceptable relative to the improvement for classification performance.

6. Conclusions

Hyperspectral images (HSI) contain abundant spectral information that can accurately distinguish the subtle differences between different pixels. However, the high dimensionality of HSIs has brought huge challenge to land cover classification. Traditional graph learning algorithms cannot effectively characterize the collaborative relationship between sample pairs, which will lead to a degraded classification performance for HSI. In this paper, we designed a supervised spectral DR method termed local constrained manifold structure collaborative preserving embedding (LMSCPE) for HSI classification. LMSCPE first adopts a novel local constrained CR model to obtain more effective collaborative coefficients between samples pairs. Then, two collaborative graphs are constructed to enhance the intraclass compactness and the interclass separability, and a local neighborhood graph is constructed to explore the local neighborhood structure of HSI. After that, an optimal objective function is designed to obtain a discriminant projection matrix, and the embedding features of different land cover types can be obtained. Experiments on PaviaU, LongKou, and MUUFL hyperspectral data sets demonstrate that LMSCPE is superior to some state-of-the-art methods. However, spectral-based DR methods only consider the spectral information in HSI, which limit the final classification performance. Therefore, in the future, we consider fusing the spatial information into the DR model to further improve the classification accuracy of HSI.

Author Contributions: All the authors made significant contributions to the manuscript. G.S. designed the DR algorithm, completed the corresponding experiments, and finished the first manuscript draft. F.L. analyzed the results and performed the validation work. Y.T. and Y.L. edited the first manuscript and finalized the manuscript for communication. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61801336 and Grant 62071340, in part by the Fundamental Research Funds for the

Central Universities under Grant 2042020kf0013, and in part by the China Postdoctoral Science Foundation under Grant 2019M662717 and Grant 2020T130480.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

- Chen, P.H.; Jiao, L.C.; Liu, F.; Zhao, J.Q.; Zhao, Z.Q. Dimensionality reduction for hyperspectral image classification based on multiview graphs ensemble. J. Appl. Remote Sens. 2016, 10, 030501. [CrossRef]
- Tao, C.N.; Zhu, H.Z.; Sun, P.; Wu, R.M.; Zheng, Z.R. Hyperspectral image recovery based on fusion of coded aperture snapshot spectral imaging and RGB images by guided filtering. *Opt. Commun.* 2020, 458, 124804. [CrossRef]
- Xue, Z.H.; Du, P.J.; Li, J.; Su, H.J. Simultaneous Sparse Graph Embedding for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* 2015, 53, 6114–6133. [CrossRef]
- 4. Jia, S.; Hu, J.; Xie, Y.; Shen, L.L.; Jia, Q.Q. Gabor Cube Selection Based Multitask Joint Sparse Representation for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3174–3187. [CrossRef]
- 5. Sun, Y.B.; Wang, S.J.; Liu, Q.S.; Hang, R.L.; Liu, G.C. Hypergraph embedding for spatial-spectral joint feature extraction in hyperspectral images. *Remote Sens.* 2017, *9*, 506.
- Ren, J.S.; Wang, R.X.; Liu, G.; Feng, R.Y.; Wang, Y.N.; Wu, W. Partitioned Relief-F Method for Dimensionality Reduction of Hyperspectral Images. *Remote Sens.* 2020, 12, 1104. [CrossRef]
- Liu, B.; Yu, X.C.; Zhang, P.Q.; Tan, X.; Wang, R.R.; Zhi, L. Spectral-spatial classification of hyperspectral image using threedimensional convolution network. J. Appl. Remote Sens. 2018, 12, 016005.
- 8. Huang, X.; Zhang, L.P. An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *IEEE Trans. Geosci. Remote Sens.* **2012**, *51*, 257–272. [CrossRef]
- 9. Zhang, L.F.; Song, L.C.; Du, B.; Zhang, Y.P. Nonlocal low-rank tensor completion for visual data. *IEEE Trans. Cybern.* 2021, 51, 673–685. [CrossRef]
- Lan, M.; Zhang, Y.P.; Zhang, L.F.; Du, B. Global Context Based Automatic Road Segmentation Via Dilated Convolutional Neural Network. *Inf. Sci.* 2020, 535, 156–171. [CrossRef]
- 11. Zhang, Q.; Tian, Y.; Yang, Y.P.; Pan, C.H. Automatic spatial-spectral feature selection for hyperspectral image via discriminative sparse multimodal learning. *IEEE Trans. Geosci. Remote Sens.* **2014**, *53*, 261–279. [CrossRef]
- 12. Zhou, F.; Hang, R.L.; Liu, Q.S.; Yuan, X.T. Hyperspectral image classification using spectral-spatial LSTMs. *Neurocomputing* **2019**, 328, 39–47. [CrossRef]
- 13. Yi, C.; Zhao, Y.Q.; Chan, J.C.W. Spectral super-resolution for multispectral image based on spectral improvement strategy and spatial preservation strategy. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 9010–9024. [CrossRef]
- 14. 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]
- 15. Xia, J.S.; Falco, N.; Benediktsson, J.A.; Du, P.J.; Chanussot, J. Hyperspectral Image Classification with Rotation Random Forest Via KPCA. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2017**, *10*, 1601–1609. [CrossRef]
- 16. Shao, Y.; Lan, J.H. A Spectral Unmixing Method by Maximum Margin Criterion and Derivative Weights to Address Spectral Variability in Hyperspectral Imagery. *Remote Sens.* **2019**, *11*, 1045. [CrossRef]
- 17. 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]
- 18. Li, W. Prasad, S.; Fowler, J. E.; Bruce, L.M. Locality-preserving dimensionality reduction and classification for hyperspectral image analysis. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 1185–1198. [CrossRef]
- 19. Zhang, Z.; Chow, T.; Zhao, M. M-Isomap: Orthogonal constrained marginal isomap for nonlinear dimensionality reduction. *IEEE Trans. Cybern.* **2013**, *43*, 1292–1303. [CrossRef] [PubMed]
- 20. Shi, G.Y.; Huang, H.; Wang, L.H. Unsupervised Dimensionality Reduction for Hyperspectral Imagery via Local Geometric Structure Feature Learning. *IEEE Geosci. Remote Sens. Lett.* **2020**, *17*, 1425–1429. [CrossRef]
- Shi, G.Y.; Huang, H.; Liu, J.M.; Li, Z.Y.; Wang, L.H. Spatial-Spectral Multiple Manifold Discriminant Analysis for Dimensionality Reduction of Hyperspectral Imagery. *Remote Sens.* 2019, 11, 2414. [CrossRef]
- 22. Zhang, Y.; Zhang, Z.; Jie, Q.; Zhang, L.; Li, B; Li, F.Z. Semi-supervised local multi-manifold Isomap by linear embedding for feature extraction. *Pattern Recognit.* 2018, *76*, 662–678. [CrossRef]

- 23. Li, Q.; Ji, H.B. Multimodality image registration using local linear embedding and hybrid entropy. *Neurocomputing* **2013**, *111*, 34–42. [CrossRef]
- 24. 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]
- 25. 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]
- 26. Pu, H.Y.; Chen, Z.; Wang, B.; Jiang, G.M. A novel spatial-spectral similarity measure for dimensionality reduction and classification of hyperspectral image. *IEEE Trans. Geosci. Remote Sens.* 2014, *52*, 7008–7022.
- 27. Fang, L.Y.; Li, S.T.; Kang, X.D.; Benediktssonet, J.A. Spectral-spatial classification of hyperspectral images with a superpixel-based discriminative sparse model. *IEEE Trans. Geosci. Remote Sens.* 2015, 53, 4186–4201. [CrossRef]
- 28. Yuan, H.L.; Tang, Y.Y. Learning with hypergraph for hyperspectral image feature extraction. *IEEE Geosci. Remote Sens. Lett.* 2015, 12, 1695–1699. [CrossRef]
- Zhou, Y.C.; Wei, Y.T. Learning Hierarchical Spectral-Spatial Features for Hyperspectral Image Classification. *IEEE Trans. Cybern.* 2016, 46, 1667–1678. [CrossRef] [PubMed]
- Luo, R.B.; Liao, W.Z.; Pi, Y.G. Discriminative Supervised Neighborhood Preserving Embedding Feature Extraction for Hyperspectral-image Classification. *Telkomnika* 2012, 10, 1051–1056. [CrossRef]
- Yang, W.K.; Sun, C.Y.; Zhang, L. A multi-manifold discriminant analysis method for image feature extraction. *Pattern Recognit.* 2012, 10, 1051–1056. [CrossRef]
- 32. 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] [PubMed]
- 33. Zhang, G.Y.; Jia, X.P.; Hu, J.K. Superpixel-based graphical model for remote sensing image mapping. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 5861–5871. [CrossRef]
- Gao, Y.; Ji, R.R.; Cui, P.; Dai, Q.H.; Hua, G. Hyperspectral image classification through bilayer graph-based learning. *IEEE Trans. Image Process.* 2014, 23, 2769–2778. [CrossRef]
- Chen, M.L.; Wang, Q.; Li, X.L. Discriminant Analysis with Graph Learning for Hyperspectral Image Classification. *Remote Sens.* 2018, 10, 836. [CrossRef]
- 36. Zhang, C.J.; Li, G.D.; Du, S.H.; Tan, W.Z.; Gao, F. Three-dimensional densely connected convolutional network for hyperspectral remote sensing image classification. *J. Appl. Remote Sens.* **2019**, *13*, 016519. [CrossRef]
- 37. Pan, B.; Shi, Z.W.; Zhang, N.; Xie, S.B. Hyperspectral image classification based on nonlinear spectral-spatial network. *IEEE Geosci. Remote Sens. Lett.* 2016, 13, 1782–1786. [CrossRef]
- Zhang, L.P.; Zhong, Y.F.; Huang, B.; Gong, J.Y.; Li, P.X. Dimensionality reduction based on clonal selection for hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* 2007, 45, 4172–4186. [CrossRef]
- Tang, Y.Y.; Yuan, H.L.; Li, L.Q. Manifold-based sparse representation for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 2014, 52, 7606–7618. [CrossRef]
- 40. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral image classification using dictionary-based sparse representation. *IEEE Trans. Geosci. Remote Sens.* 2011, 49, 3973–3985. [CrossRef]
- 41. 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]
- 42. Li, W.; Liu, J.B.; Du, Q. Sparse and low-rank graph for discriminant analysis of hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4094–4105. [CrossRef]
- 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]
- 44. Zang, F.; Zhang, J.S. Discriminative learning by sparse representation for classification. *Neurocomputing* **2011**, *74*, 2176–2183. [CrossRef]
- 45. 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]
- 46. Zhong, Y.F.; Wang, X.Y.; Zhao, L.; Feng, R.Y.; Zhang, L.P.; Xu, Y.Y. Blind spectral unmixing based on sparse component analysis for hyperspectral remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 49–63. [CrossRef]
- 47. Ye, Z.; Dong, R.; Bai, L.; Nian, Y.J. Adaptive collaborative graph for discriminant analysis of hyperspectral imagery. *Eur. J. Remote Sens.* **2020**, *53*, 91–103. [CrossRef]
- 48. Lv, M.; Hou, Q.L.; Deng, N.Y.; Jing, L. Collaborative Discriminative Manifold Embedding for Hyperspectral Imagery. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 272–281. [CrossRef]
- 49. Ly, N.H.; Du, Q.; Fowler, J.E. Collaborative Graph-Based Discriminant Analysis for Hyperspectral Imagery. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 2014, 7, 2688–2696. [CrossRef]
- Lou, S.J.; Ma, Y.H.; Zhao, X.M. Manifold aware discriminant collaborative graph embedding for face recognition. In Proceedings
 of the Tenth International Conference on Digital Image Processing, Chengdu, China, 12–14 December 2018.
- 51. Huang, P.; Li, T.; Gao, G.W.; Yao, Y.Z.; Yang, G. Collaborative representation based local discriminant projection for feature extraction. *Digit. Signal Prog.* **2018**, *76*, 84–93. [CrossRef]

- 52. 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]
- 53. Wong, W.K.; Zhao, H.T. Supervised optimal locality preserving projection. Pattern Recognit. 2012, 45, 186–197. [CrossRef]
- 54. Li, W.; Du, Q. Collaborative Representation for Hyperspectral Anomaly Detection. *IEEE Trans. Geosci. Remote Sens.* 2014, 53, 1463–1474. [CrossRef]
- Li, W.; Du, Q.; Zhang, B. Combined sparse and collaborative representation for hyperspectral target detection. *Pattern Recognit.* 2015, 48, 3904–3916. [CrossRef]
- 56. Zhu, P.F.; Zuo, W.M.; Zhang, L.; Shiu, S.C.; Zhang, D. Image Set-Based Collaborative Representation for Face Recognition. *IEEE Trans. Inf. Forensic Secur.* 2014, *9*, 1120–1132.
- 57. Zhang, L.F.; Zhang, L.P.; Tao, D.C.; Huang, X. A modified stochastic neighbor embedding for multi-feature dimension reduction of remote sensing images. *ISPRS J. Photogramm. Remote Sens.* **2013**, *83*, 30–39. [CrossRef]
- Shi, Q.; Zhang, L.P.; Du, B. Semisupervised discriminative locally enhanced alignment for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* 2013, 51, 4800–4815. [CrossRef]
- Dong, Y.N.; Du, B.; Zhang, L.P.; Zhang, L.F. Exploring locally adaptive dimensionality reduction for hyperspectral image classification: A maximum margin metric learning aspect. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* 2017, 10, 1136–1150. [CrossRef]
- Dong, Y.N.; Du, B.; Zhang, L.P.; Zhang, L.F. Dimensionality reduction and classication of hyperspectral images using ensemble discriminative local metric learning. *IEEE Trans. Geosci. Remote Sens.* 2017, 55, 2509–2524. [CrossRef]
- Sun, W.W.; Halevy, A.; Benedetto, J.J.; Czaja, W.; Liu, C.; Wu, H.B.; Shi, B.Q.; Li, Q.Y. UL-Isomap based nonlinear dimensionality reduction for hyperspectral imagery classification. *ISPRS J. Photogramm. Remote Sens.* 2014, *89*, 25–36. [CrossRef]
- 62. Sun, W.W.; Yang, G.; Du, B.; Zhang, L.P.; 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]
- 63. 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]
- 64. Huang, H.; Li, Z.Y.; Pan, Y.S. Multi-Feature Manifold Discriminant Analysis for Hyperspectral Image Classification. *Remote Sens.* **2019**, *11*, 651. [CrossRef]
- 65. 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**, *66*, 1646–1657. [CrossRef]
- 66. 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]
- 67. 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]
- 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]