

Article

Multiscale Superpixelwise Locality Preserving Projection for Hyperspectral Image Classification

Lin He ^{1,*} , Xianjun Chen ¹, Jun Li ^{2,*} and Xiaofeng Xie ³

¹ School of Automation Science and Engineering, South China University of Technology, Guangzhou 510640, China; lxzxj_chan@163.com

² Guangdong Provincial Key Laboratory of Urbanization and Geo-Simulation, Center of Integrated Geographic Information Analysis, School of Geography and Planning, Sun Yat-sen University, Guangzhou 510275, China

³ Mechanical and Electrical Engineering College, Hainan University, Haikou 570228, China; xfxie@outlook.com

* Correspondence: helin@scut.edu.cn (L.H.); lijun48@mail.sysu.edu.cn (J.L.)

Received: 31 March 2019; Accepted: 22 May 2019; Published: 27 May 2019



Abstract: Manifold learning is a powerful dimensionality reduction tool for a hyperspectral image (HSI) classification to relieve the curse of dimensionality and to reveal the intrinsic low-dimensional manifold. However, a specific characteristic of HSIs, i.e., irregular spatial dependency, is not taken into consideration in the method design, which can yield many spatially homogenous subregions in an HSI scene. Conventional manifold learning methods, such as a locality preserving projection (LPP), pursue a unified projection on the entire HSI, while neglecting the local homogeneities on the HSI manifold caused by those spatially homogenous subregions. In this work, we propose a novel multiscale superpixelwise LPP (MSuperLPP) for HSI classification to overcome the challenge. First, we partition an HSI into homogeneous subregions with a multiscale superpixel segmentation. Then, on each scale, subregion specific LPPs and the associated preliminary classifications are performed. Finally, we aggregate the classification results from all scales using a decision fusion strategy to achieve the final result. Experimental results on three real hyperspectral data sets validate the effectiveness of our method.

Keywords: hyperspectral image manifold learning; dimensionality reduction; local homogeneity; irregular spatial dependency; multiscale superpixel segmentation; covariance feature; classification

1. Introduction

Hyperspectral image (HSI) classification has been a research hotspot over recent years, since it plays a critical role in military target detection, precision agriculture, mine exploration, and many other applications [1–3]. With abundant spectral information, HSIs show great potential in identifying different ground objects of interest. However, hundreds of spectral bands in an HSI also bring about some problems, such as heavy computation burdens and the Hughes phenomenon which means that a large amount of training HSI pixels are required to maintain statistical confidence in the HSI classification task [4,5]. Such problems usually hinder HSI classifiers from achieving excellent performance in real applications. In fact, HSI spectral bands are strongly correlated; hence, the spectral signature of each HSI pixel can be represented by only a few features [6–8]. Therefore, one often-used strategy to overcome the dimensionality dilemma mentioned above is dimensionality reduction which aims to represent the high-dimensional HSI data in a low-dimensional space without losing important information for discriminating tasks.

Numerous dimensionality reduction methods have been developed, which can be roughly categorized into two groups: feature transform and feature selection [9,10]. Feature transform projects

the original HSI data into an appropriate low-dimensional space, while feature selection chooses the most representative bands from the HSI. Feature transform gains the advantage over feature selection that it has the potential to maintain the original information in the high-dimensional data during dimensionality reduction and thus generate more discriminative features [11,12]. Principle component analysis (PCA) and linear discriminant analysis (LDA) are two typical feature transform methods. PCA attempts to map the data along the directions of maximal variance [13]. LDA tends to maximize between-class distances and minimize within-class distances in the meantime [14]. Both PCA and LDA just assume that the information contained in high-dimensional data lies on a linear low-dimensional space, whereas nonlinear structures are often exhibited in real HSI data. To cope with the nonlinearities, manifold learning methods were developed. Manifold learning assumes that the high-dimensional data actually lie on a low-dimensional manifold structure, which can be parameterized with a group of identifiable coordinates. One of the most popular manifold methods is locality preserving projection (LPP) which builds a graph to capture geometric structures of data and subsequently establishes a projection from the original data space to the low-dimensional space [15–19]. Other manifold learning approaches include locally linear embedding (LLE) which supposes that the structure represented by the linear combinations of the data's nearest neighbors is unchanged in both the high-dimensional and the associated low-dimensional spaces [20], isometric mapping (ISOMAP) which utilizes the geodesic distance to perform low-dimensional embedding [21], and local tangent space alignment (LTSA) which aims to recover the intrinsic manifold of data by aligning the local tangent space of each pixel [22].

Irregular spatial dependency is an important characteristic of HSIs, which is caused by the usual occurrences of complex irregular ground objects in HSI scenes. The dependency brings about spatially local homogeneous subregions in different shapes and sizes in an HSI. Such subregions can be detected effectively with appropriate techniques such as superpixel [23–25], where pixels in a homogeneous subregion have similar spectral properties but vary relatively significantly across different subregions. Intuitively, such homogeneous subregions would result in local homogeneities on an HSI manifold. However, conventional manifold learning-based HSI dimensionality reduction methods, such as LPP, directly apply a unified projection on the entire HSI, missing those local homogeneities on the HSI manifold. Motivated by such a consideration, we propose a multiscale superpixelwise LPP (MSuperLPP) method for HSI classification in this work. The method is able to deal with spatially local homogeneities in HSIs during dimensionality reduction, thus offering the potential to improving the subsequent classification. To the best of our knowledge, there is no similar work to ours in existing literature, which performs local homogeneity manifold-based HSI dimensionality reduction. Our methods comprise three major phases. First, we segment an HSI into many homogeneous subregions using entropy rate superpixel segmentation (ERS) with a series of scales, which can exploit fully rich spatial dependencies in different shapes and sizes. Next, on each scale, LPP is run on each subregion with spectral-spatial covariance feature, and the obtained low-dimensional features are fed into a preliminary classifier. In the final step, the results on all the scales are aggregated with a decision fusion strategy to yield the final classification result.

The remainder of this paper is organized as follows. In Section 2, we introduce some backgrounds related to the proposed approach. Section 3 gives the details of the proposed MSuperLPP. Section 4 presents the experiment results, and Section 5 provides a summary of this paper.

2. Related Work and Background

Denote an HSI with N pixels and D bands as $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in \mathbb{R}^{D \times N}$, and the connected data set in a low-dimensional space as $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N] \in \mathbb{R}^{d \times N}$, where $d \ll D$. LPP is a widely used manifold learning method for HSI dimensionality reduction [16–19,26], while the region covariance descriptor is an effective spectral-spatial feature for HSI classification [26–28]. In the following, they are briefly introduced as backgrounds of our proposed method.

2.1. Locality Preserving Projection

LPP builds a graph to obtain the connections among all HSI pixels and then intends to find an optimal projection \mathbf{A} to map the original HSI data into a low-dimensional space, while maintaining the local connection relationships [15–19]. Its objective function can be mathematically represented as

$$\min \sum_{i,j} \|\mathbf{y}_i - \mathbf{y}_j\|^2 \mathbf{W}_{ij} \tag{1}$$

and the weight \mathbf{W}_{ij} is defined as

$$\mathbf{W}_{ij} = \begin{cases} e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}}, & \text{if } \mathbf{x}_i \in \mathbf{N}(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \mathbf{N}(\mathbf{x}_i) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where σ denotes the scale and $\mathbf{N}(\mathbf{x}_i)$ stands for the neighbors of \mathbf{x}_i . Equation (2) measures the similarity or the distance between pixels \mathbf{x}_i and \mathbf{x}_j . Under the similarity, the objective in Equation (1) would lead to a heavy penalty if \mathbf{x}_i and \mathbf{x}_j are mapped to be far away from each other; thus minimizing it guarantees that a small distance between \mathbf{x}_i and \mathbf{x}_j tends to force a small distance between \mathbf{y}_i and \mathbf{y}_j . Given a constraint $\mathbf{YDY} = \mathbf{I}$, which means removing the arbitrary scaling factor, the aforementioned optimization problem can be further formulated as

$$\begin{aligned} & \arg \min tr(\mathbf{A}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{A}) \\ & s.t. \quad \mathbf{A}^T \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{A} = \mathbf{I} \end{aligned} \tag{3}$$

where \mathbf{I} is an identity matrix, \mathbf{D} is a diagonal matrix with diagonal entries $\mathbf{D}_{ii} = \sum_j \mathbf{W}_{ij}$, and $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is the graph Laplacian matrix. The problem in Equation (3) can be solved by the following generalized eigenvalue formulation:

$$\mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{a} = \lambda \mathbf{X} \mathbf{D} \mathbf{X}^T \mathbf{a} \tag{4}$$

where λ denotes the eigenvalue and eigenvectors corresponding to the d smallest eigenvalues of Equation (4) from the projection matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_d]$.

In brief, LPP aims to obtain a projection to perform an HSI dimensionality reduction while preserving the neighborhood structure of the data. Besides the standard LPP mentioned above, some considerations such as sparsity, tensor, and orthogonality can be utilized additionally to solve the projection [17–19,29].

2.2. Region Covariance Descriptor

Region covariance descriptor has been applied in the computer vision and brain computer interface problem [30–32]. Deng et al. introduced the descriptor to HSI processing [26–28]. Suppose that $\mathcal{X} \in \mathbb{R}^{W \times H \times D}$ represents the original HSI cube, $\mathcal{X}_i \in \mathbb{R}^{w \times h \times D}$ denotes the spatial region around the i th pixel, and $s = w \times h$ is size of a spatially local window, then the region covariance descriptor is as follows:

$$\mathbf{C}_i = \frac{1}{s-1} \sum_{t=1}^s (\mathbf{x}_t - \boldsymbol{\mu}_i)(\mathbf{x}_t - \boldsymbol{\mu}_i)^T \tag{5}$$

This can be defined as a spectral-spatial feature of \mathcal{X}_i , where $\mu_i = \frac{1}{s} \sum_{t=1}^s \mathbf{x}_t$. Since the covariance feature is connected to a symmetric positive definite matrix and lies on a Riemannian manifold [32], the Log-Euclidean distance metric shown below can measure the similarity between \mathbf{C}_i and \mathbf{C}_j :

$$d_{LE}(\mathbf{C}_i, \mathbf{C}_j) = \|\log(\mathbf{C}_i^{-1}\mathbf{C}_j)\|_F = \left[\sum_{k=1}^n \log^2 \lambda_k \right]^{1/2} \tag{6}$$

where λ_k is the k th eigenvalue of $\mathbf{C}_i^{-1}\mathbf{C}_j$.

Performing a region covariance descriptor can yield an excellent spectral-spatial covariance feature for HSI classification [26–28], which has the advantage of robustness to noise and spectral variabilities over traditional spectral feature (the original spectral signature) [33–35].

Following one of the most effective HSI classification routines which comprises three successive phases, i.e., feature extraction, dimensionality reduction, and classification, the combination of various features and various manifold dimensionality reduction ways can yield different HSI manifold learning schemes which can be testified with a subsequent HSI classifier.

3. Proposed Method

A remotely sensed HSI scene usually comprises irregular ground objects in various shapes and sizes, which leads to the irregular spatial dependency characteristic of an HSI, thus leading to the local homogeneities on the connected HSI manifold. In the following, we develop a new method called MSuperLPP to deal with the manifold local homogeneities. Different from a conventional LPP which applies an unified projection on the entire HSI data, our method adaptively determines homogenous subregions and thus the HSI manifold local homogeneity and then employs a divide-and-conquer strategy to perform LPP processing based on those subregions. Therefore, our method is able to fully explore local homogeneities on the HSI manifold and thus collects more useful discriminative information, which poses the potential to enhance the final classification performance.

3.1. Determination of Manifold Local Homogeneity with Multiscale Superpixel Segmentation

Taking into consideration the irregular spatial dependency characteristic of HSIs, we adopt a multiscale superpixel segmentation strategy here to determine the irregular homogeneous subregions in various shapes and sizes in an HSI. ERS can achieve a natural representation of visual scenes [36,37]. It is based on an undirected graph $\mathcal{G}(V, E)$ where V is the vertex set and E is the edge set. The vertices can be the pixels of an HSI, and then, the weight of an edge measures the similarity of the connected two pixels. Thus, the image segmentation task becomes the problem of how to partition the graph properly. ERS aims to choose a subset S of E which is in the partition of K subgraphs using the following criterion:

$$\max H(S) + \lambda B(S) \tag{7}$$

where the entropy rate term $H(S)$ prefers homogeneous clusters, the regularization term $B(S)$ forces the clusters to be of similar sizes, and the weight λ needs to be nonnegative. To solve the optimization problem in Equation (7), Liu et al. gave an efficient greedy algorithm [36]. In our method, we first use PCA to obtain the first principal component of HSI and then perform ERS on the extracted principal component for convenience.

ERS tends to divide the image into subregions of similar sizes once the number of superpixels is given. However, the homogenous subregions in the HSI are in different sizes. That is to say, ERS with a single scale cannot represent the homogeneities well. Therefore, we perform the segmentation in a multiscale way to tackle this issue, where ERS is run with several scales. Specifically, if an HSI is segmented with $2P + 1$ scales, then the number of superpixels on a scale is determined by

$$N_s = (\sqrt{2})^p N_{sf}, \quad p = 0, \pm 1, \pm 2, \dots, \pm P \tag{8}$$

where p is the index of scales and N_{sf} is the fundamental number of superpixels and given empirically. With a series of scales, the segmentation is expected to adapt homogeneities in various sizes in the scene and to offer the potential to further reduce the impact of an irregular spatial dependency.

A superpixel obtained in the HSI represents spatially homogenous subregions, while different superpixels means there are spectral signature differences to some degree. Such homogenous subregions in an HSI imply that there are local homogeneities on the associated HSI manifold. Thus, we can determine local homogeneities on the connected HSI manifold by performing the multiscale superpixel segmentation.

3.2. Divide-and-Conquer-Based LPP Classification

As local homogeneities on an HSI manifold can be reflected with superpixels, performing an LPP in each superpixel with a superpixel specific projection is expected to preserve the spatially local geometric structure of the data well.

Our method utilizes the region covariance descriptor to construct the graph for an LPP. The descriptor can yield spectral-spatial covariance features more robust to noise and spectral variabilities than the spectral feature [26–28]. More specifically, we characterize pixels with their region covariance descriptors, calculate the pairwise distances among pixels in a superpixel with Log-Euclidean distance metric, find the nearest neighbors of the pixels in a superpixel, and compute their similarities as

$$W_{ij} = \begin{cases} e^{-\frac{d_{LE}(C_i, C_j)^2}{2\sigma^2}}, & \text{if } C_i \in N(C_j) \text{ or } C_j \in N(C_i) \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

where $N(C_i)$ denotes the neighbors of C_i . Then, we run an LPP on each superpixel.

After performing the dimensionality reduction on the HSI with $2P + 1$ different segmentation scales, we classify the obtained low-dimensional data on each scale, and then majority voting decision fusion is utilized to aggregate the results from all the scales. Suppose that the classification result on the j th scale for a pixel is l_j . Then, the score for the i th class is

$$N(i) = \sum_{j=1}^{2P+1} I(l_j = i) \quad (10)$$

where $I(l_j = i) = \begin{cases} 1, & \text{if } l_j = i \\ 0, & \text{otherwise} \end{cases}$ denotes an indicator function. The final classification result is achieved by

$$l = \arg \max_{i \in \{1, 2, \dots, L\}} N(i) \quad (11)$$

where L is the number of all possible classes. The flowchart of our proposed MSuperLPP for HSI classification is shown in Figure 1, and the connected performing steps can be found in Algorithm 1.

Algorithm 1 MSuperLPP

Input: HSI $\mathcal{X} \in \mathbb{R}^{W \times H \times D}$; scale set $N_{sp} = \{N_i | i = 1, 2, \dots, S\}$ obtained by (8); window size $s = w \times h$.
 Extract spectral-spatial covariance features using (5) and perform PCA to extract the first principle component I_f .
for $i=1$ to S **do**
 Segment \mathcal{X} into N_i homogeneous subregions using ERS with I_f ;
 for $j=1$ to N_i **do**
 Perform LPP in each subregion \hat{R}_j , where the spectral-spatial covariance features are used to search for the k nearest neighbors.
 end for
 Combine the low-dimensional features of all the subregions on the same scale to form the low-dimensional data on this scale. Perform classification on the scales i to get preliminary output T_i .
end for
 Aggregate the classification results $T_i (i = 1, 2, \dots, S)$ using (10) and (11).
Output: Final classification result T .

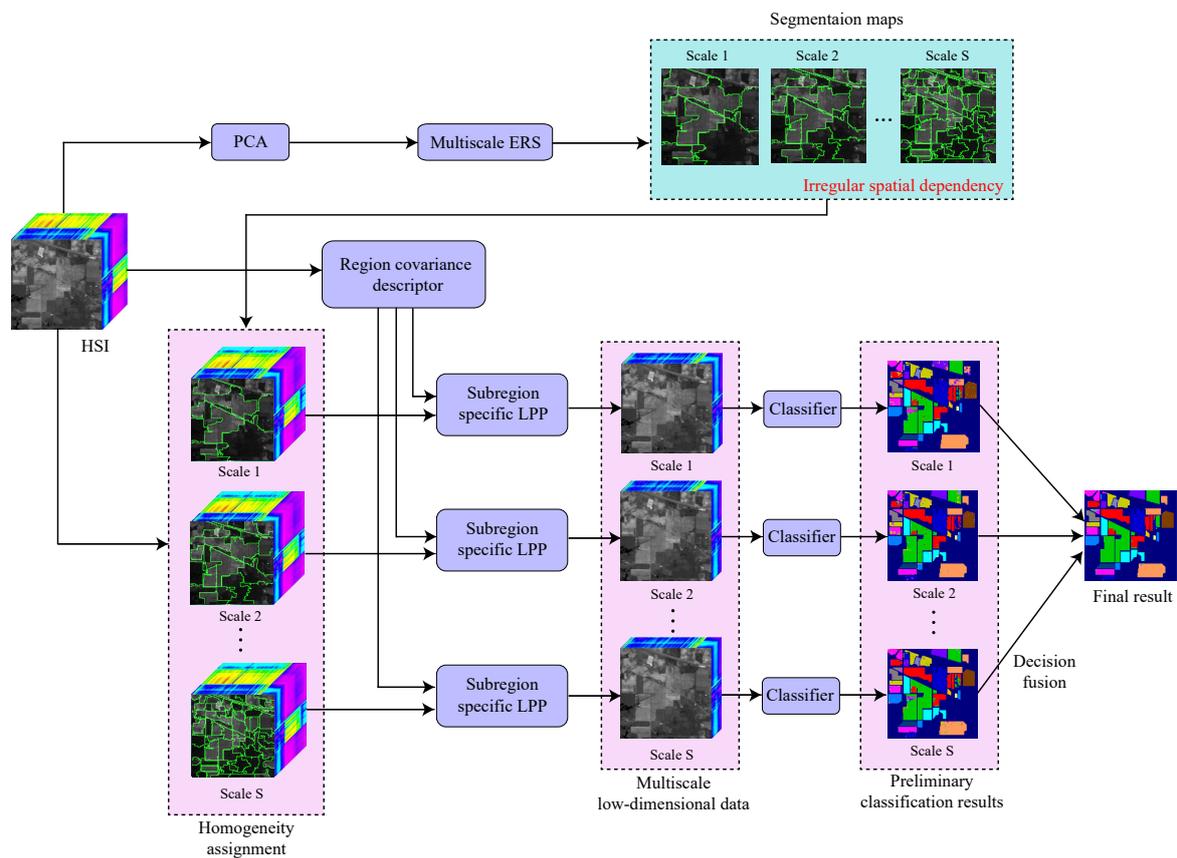


Figure 1. A flowchart of the proposed MSuperLPP for hyperspectral image (HSI) classification.

4. Experimental Results

In our experiments, three real HSI data sets are used to evaluate the proposed MSuperLPP. The first one is the Indian Pines data set, which was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over northwestern Indiana, including 220 spectral bands as well as 145×145 pixels, and its available ground-truth data contains 10249 pixels with 16 classes. 40 noisy bands are removed, and the remaining 180 bands are used for experiments. The second one is the Zaoyuan data

set, which was collected by the Operational Modular Imaging Spectrometer (OMIS) over the Zaoyuan region, China, in 2001. The data set comprises 137×202 pixels and 128 bands covering the region. After removing the noisy bands, 80 bands remain. Moreover, 23821 pixels with 8 classes are used for classification. The last one is the Salinas data set, which was collected by AVIRIS over Salinas Valley, California. The scene used for the experiments comprises 250×110 pixels and 224 bands and has 14 testing classes. After removing the noisy bands, 174 bands are Left for the experiments.

For the Indian Pines and Zaoyuan data sets, we randomly choose 5, 10, 30, and 50 pixels per class for training while the other pixels are used for test. Since some classes have a few samples, we select at most half of the total samples. For the Salinas data set, 1, 3, 5, and 10 pixels per class are selected as training samples. After a dimensionality reduction, both the nearest neighbor (NN) classifier and the support vector machine (SVM) with the radial basis function (RBF) kernel are subsequently applied to evaluate the proposed method. The free parameters of the two classifiers are set with cross-validation [38], including the number of nearest neighbors for the NN classifier and the RBF scale and the regularization coefficient for the SVM classifier. We repeat all the classification experiments ten times to avoid random bias, and the average accuracies are presented.

We compare the proposed MSuperLPP with GlobalLPP-SF (global LPP on spectral feature), GlobalLPP-SSCF (global LPP on region covariance descriptor based spectral-spatial covariance feature), SuperLPP-SF (superpixelwise LPP on spectral feature), and SuperLPP-SSCF (superpixelwise LPP on spectral-spatial covariance feature). In our experimental settings, some other parameters are empirically determined as follows. The window's size of region covariance descriptor is 3×3 for GlobalLPP-SSCF, SuperLPP-SSCF, and MSuperLPP. In MSuperLPP, the fundamental number of superpixel N_{sf} is set to 100 for Indian Pines, 40 for Zaoyuan, and 15 for Salinas, while P is set to 4 for all three data sets.

Tables 1–3 demonstrate the overall classification accuracies on the Indian Pines, Zaoyuan, and Salinas data sets, respectively. Figure 2 shows the influence of the number of training samples on the classification performance. As observed, GlobalLPP-SSCF performs better than GlobalLPP-SF while SuperLPP-SSCF is better than SuperLPP-SF, which suggests that a spectral-spatial covariance feature is beneficial to improving classification accuracy. Meanwhile, SuperLPP-SSCF and SuperLPP-SF achieve higher accuracies than GlobalLPP-SSCF and GlobalLPP-SF, respectively, which indicates that a superpixel segmentation can enhance classification accuracy. Among all the five considered methods, our proposed MSuperLPP yields the highest accuracies.

For visual inspection purposes, the classification maps obtained with five compared methods are given in Figures 3–5. Here, we only show the results obtained with the nearest neighbor classifier when the size of the training set is set to 50 for Indian Pines and Zaoyuan and 10 for Salinas, respectively. It can be seen that our MSuperLPP yields the best regional consistency and agrees more with the ground truth.

Table 1. The classification accuracy (%) for the Indian Pines data set with the nearest neighbor and SVM classifiers.

Training Size	5		10		30		50	
	NN	SVM	NN	SVM	NN	SVM	NN	SVM
GlobalLPP-SF	44.40	49.11	49.47	56.26	58.52	67.90	62.67	72.96
GlobalLPP-SSCF	62.81	63.65	68.67	68.87	77.66	77.99	80.39	80.98
SuperLPP-SF	63.74	64.10	70.73	71.65	80.92	80.33	84.03	83.14
SuperLPP-SSCF	73.50	74.37	82.23	81.35	90.43	87.91	92.28	89.38
MSuperLPP	76.49	75.91	85.53	83.60	94.55	93.10	97.02	95.57

Table 2. The classification accuracy (%) for the Zaoyuan data set with the nearest neighbor and SVM classifiers.

Training Size	5		10		30		50	
Classifier	NN	SVM	NN	SVM	NN	SVM	NN	SVM
GlobalLPP-SF	64.26	75.81	71.29	81.64	77.18	84.84	79.34	86.21
GlobalLPP-SSCF	75.12	80.18	81.06	84.92	85.77	88.50	86.63	89.19
SuperLPP-SF	70.06	80.65	75.56	86.88	81.30	89.48	83.44	90.55
SuperLPP-SSCF	80.81	84.34	84.79	87.28	88.00	90.83	89.22	92.26
MSuperLPP	82.27	86.64	90.10	90.68	93.83	93.75	94.42	94.39

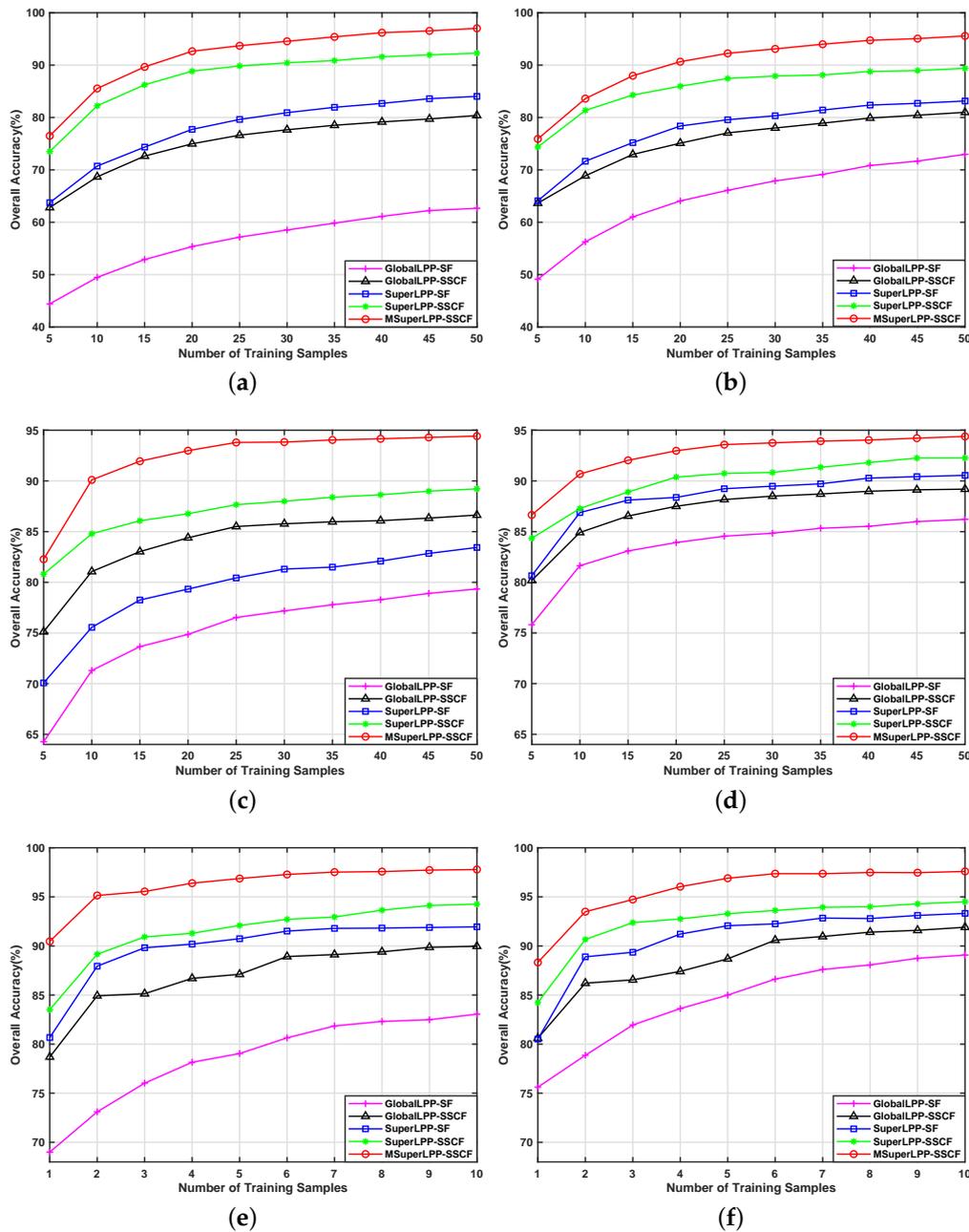


Figure 2. Classification accuracy versus training size. (a) Indian Pines with the nearest neighbor classifier. (b) Indian Pines with support vector machine (SVM) classifier. (c) Zaoyuan with nearest neighbor classifier. (d) Zaoyuan with SVM classifier. (e) Salinas with nearest neighbor classifier. (f) Salinas with SVM classifier.

Table 3. Classification accuracy (%) for the Salinas data set with nearest neighbor and SVM classifiers.

Training Size	1		3		5		10	
Classifier	NN	SVM	NN	SVM	NN	SVM	NN	SVM
GlobalLPP-SF	69.00	75.61	76.02	81.95	79.04	85.00	83.05	89.07
GlobalLPP-SSCF	78.69	80.60	85.14	86.55	87.11	88.69	89.97	91.91
SuperLPP-SF	80.67	80.51	89.82	89.36	90.74	92.06	91.95	93.32
SuperLPP-SSCF	81.77	84.23	90.92	92.37	91.97	93.28	93.82	94.50
MSuperLPP	90.46	88.32	95.55	94.73	96.87	96.89	97.78	97.59

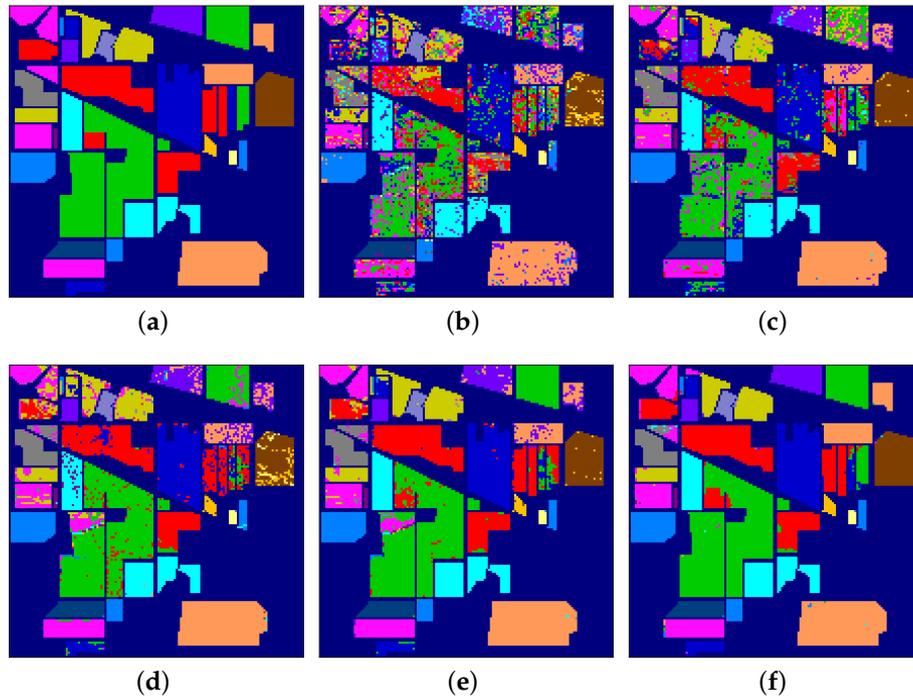


Figure 3. Classification maps obtained with the Indian Pines data set. (a) Ground truth. (b) GlobalLPP-SF. (c) GlobalLPP-SSCF. (d) SuperLPP-SF. (e) SuperLPP-SSCF. (f) MSuperLPP.

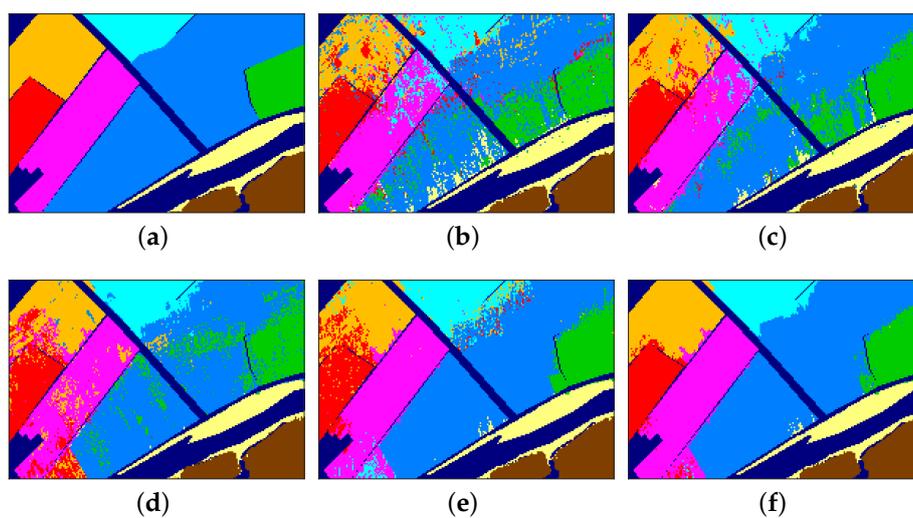


Figure 4. Classification maps obtained with the Zaoyuan data set. (a) Ground truth. (b) GlobalLPP-SF. (c) GlobalLPP-SSCF. (d) SuperLPP-SF. (e) SuperLPP-SSCF. (f) MSuperLPP.

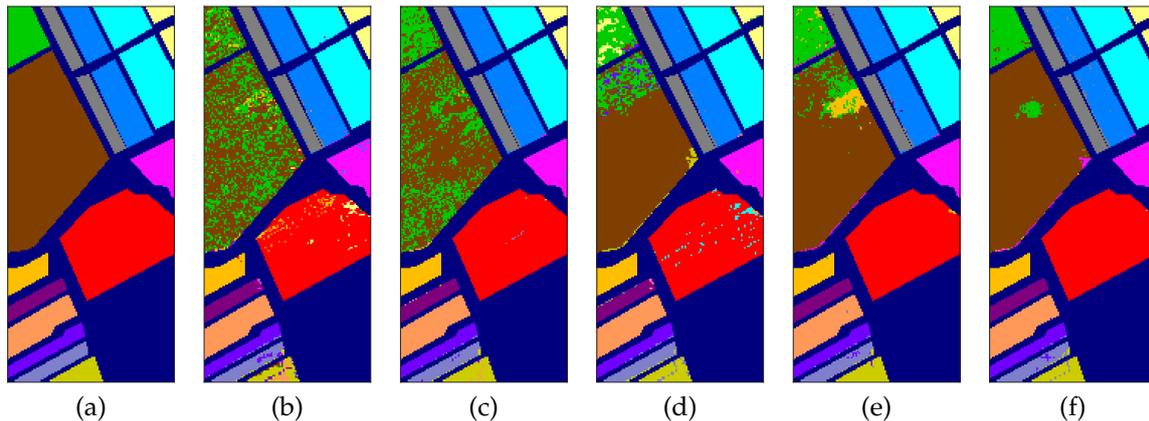


Figure 5. Classification maps obtained with the Salinas data set. (a) Ground truth. (b) GlobalLPP-SF. (c) GlobalLPP-SSCF. (d) SuperLPP-SF. (e) SuperLPP-SSCF. (f) MSuperLPP.

Therefore, it is verified by the experimental results that our MSuperLPP, involving the multiscale superpixel segmentation strategy, is able to achieve an excellent classification performance.

5. Conclusions

Taking into consideration local homogeneities on the HSI manifold connected to the irregular spatial dependency characteristic of an HSI, which is usually ignored by existing manifold learning-based dimensionality reduction methods, we propose a MSuperLPP method for HSI classification. In MSuperLPP, we adopt a divide-and-conquer strategy, first dividing the HSI into many homogeneous subregions on various scales to reveal those local homogeneities, then performing a LPP in each subregion and a preliminary classification on each scale, and finally fusing all the preliminary classification results to yield a final result. The experimental results on real HSI data sets verify the excellent performance of our method.

Author Contributions: All authors made significant contributions to the manuscript. L.H., X.C., and J.L. designed the research framework, conducted the experiments, and wrote the manuscript. L.H. and J.L. were the supervisors of the work and provided the fundings. X.X. provided many constructive suggestions on the motivation analysis and the methodology design.

Funding: This work was supported in part by the National Natural Science Foundation of China under Grant 61571195 and Grant 61771496; in part by the Guangdong Provincial Natural Science Foundation under Grant 2016A030313254, Grant 2016A030313516, and Grant 2017A030313382; and in part by the National Key Research and Development Program of China under Grant 2017YFB0502900.

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

References

1. 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](#)]
2. Chang, C.I. *Hyperspectral Data Exploitation: Theory And Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
3. Qiu, Z.; Chen, J.; Zhao, Y.; Zhu, S.; He, Y.; Zhang, C. Variety identification of single rice seed using hyperspectral imaging combined with convolutional neural network. *Appl. Sci.* **2018**, *8*, 212. [[CrossRef](#)]
4. Hughes, G. On the mean accuracy of statistical pattern recognizers. *IEEE Trans. Inf. Theory* **1968**, *14*, 55–63. [[CrossRef](#)]
5. Thenkabail, P.S.; Gumma, M.K.; Teluguntla, P.; Mohammed, I.A. Hyperspectral remote sensing of vegetation and agricultural crops. *Photogramm. Eng. Remote Sens.* **2014**, *80*, 697–723.

6. Jiménez, L.O.; Rivera-Medina, J.L.; Rodríguez-Díaz, E.; Arzuaga-Cruz, E.; Ramírez-Vélez, M. Integration of spatial and spectral information by means of unsupervised extraction and classification for homogenous objects applied to multispectral and hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 844–851. [[CrossRef](#)]
7. Ni, D.; Ma, H.B. Classification of hyperspectral image based on sparse representation in tangent space. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 786–790.
8. Shao, Z.F.; Zhang, L. Sparse dimensionality reduction of hyperspectral image based on semi-supervised local Fisher discriminant analysis. *Int. J. Appl. Earth Obs. Geoinf.* **2014**, *31*, 122–129. [[CrossRef](#)]
9. 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](#)]
10. Mojaradi, B.; Abrishami-Moghaddam, H.; Zoj, M.J.V.; Duin, R.P. Dimensionality reduction of hyperspectral data via spectral feature extraction. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 2091–2105. [[CrossRef](#)]
11. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*; John Wiley & Sons: Hoboken, NJ, USA, 2001.
12. Fukunaga, K. *Introduction to Statistical Pattern Recognition*; Elsevier: Amsterdam, The Netherlands, 1990.
13. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [[CrossRef](#)]
14. Bandos, T.V.; Bruzzone, L.; Camps-Valls, G. Classification of hyperspectral images with regularized linear discriminant analysis. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 862–873. [[CrossRef](#)]
15. He, X.F.; Niyogi, P. Locality preserving projections. In Proceedings of the Annual Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 8–13 December 2003; pp. 153–160.
16. Wang, Z.Y.; He, B.B. Locality perserving projections algorithm for hyperspectral image dimensionality reduction. In Proceedings of the International Conference on Geoinformatics, Shanghai, China, 24–26 June 2011; pp. 1–4.
17. Deng, Y.J.; Li, H.C.; Pan, L.; Emery, W.J. Tensor locality preserving projection for hyperspectral image classification. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, Fort Worth, TX, USA, 23–28 July 2017; pp. 771–774.
18. Zhai, Y.; Zhang, L.; Wang, N.; Guo, Y.; Cen, Y.; Wu, T.; Tong, Q. A modified locality-preserving projection approach for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1059–1063. [[CrossRef](#)]
19. Wang, R.; Nie, F.; Hong, R.; Chang, X.; Yang, X.; Yu, W. Fast and orthogonal locality preserving projections for dimensionality reduction. *IEEE Trans. Image Process.* **2017**, *26*, 5019–5030. [[CrossRef](#)]
20. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [[CrossRef](#)]
21. Tenenbaum, J.B.; De, S.V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323. [[CrossRef](#)]
22. Zhang, Z.Y.; Zha, H.Y. Nonlinear dimension reduction via local tangent space alignment. In Proceedings of the International Conference on Intelligent Data Engineering and Automated Learning, Hong Kong, China, 21–23 March 2003; pp. 477–481.
23. Sun, H.; Ren, J.; Zhao, H.; Yan, Y.; Zabalza, J.; Marshall, S. Superpixel based feature specific sparse representation for spectral-spatial classification of hyperspectral images. *Remote Sens.* **2019**, *11*, 536. [[CrossRef](#)]
24. Duan, W.; Li, S.; Fang, L. Spectral-spatial hyperspectral image classification using superpixel and extreme learning machines. In Proceedings of the Chinese Conference on Pattern Recognition, Changsha, China, 17–19 November 2014; pp. 159–167.
25. Zhan, T.; Sun, L.; Xu, Y.; Yang, G.; Zhang, Y.; Wu, Z. Hyperspectral classification via superpixel kernel learning-based low rank representation. *Remote Sens.* **2018**, *10*, 1639. [[CrossRef](#)]
26. 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](#)]
27. Fang, L.; He, N.; Li, S.; Plaza, A.J.; Plaza, J. A new spatial-spectral feature extraction method for hyperspectral images using local covariance matrix representation. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 3534–3546. [[CrossRef](#)]

28. He, N.; Paoletti, M.E.; Fang, L.; Li, S.; Plaza, A.; Plaza, J. Feature extraction with multiscale covariance maps for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 1–15. [[CrossRef](#)]
29. Qiao, L.; Chen, S.; Tan, X. Sparsity preserving projections with applications to face recognition. *Pattern Recogn.* **2010**, *43*, 331–341. [[CrossRef](#)]
30. Tuzel, O.; Porikli, F.; Meer, P. Pedestrian detection via classification on riemannian manifolds. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1713–1727. [[CrossRef](#)]
31. Xie, X.F.; Yu, Z.L.; Gu, Z.H.; Zhang, J.; Cen, L.; Li, Y.Q. Bilinear regularized locality preserving learning on Riemannian graph for motor imagery BCI. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2018**, *26*, 698–708. [[CrossRef](#)]
32. Wang, R.P.; Guo, H.M.; Davis, L.S.; Dai, Q.H. Covariance discriminative learning: A natural and efficient approach to image set classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012; pp. 2496–2503.
33. Guo, B.; Gunn, S.R.; Damper, R.I.; Nelson, J.D.B. Customizing kernel functions for SVM-based hyperspectral image classification. *IEEE Trans. Image Process.* **2008**, *17*, 622–629. [[CrossRef](#)]
34. Melgani, F.; Bruzzone, L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1778–1790. [[CrossRef](#)]
35. Ma, L.; Crawford, M. M.; Tian, J. Local Manifold Learning-Based k-Nearest-Neighbor for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4099–4109. [[CrossRef](#)]
36. Liu, M.Y.; Tuzel, O.; Ramalingam, S.; Chellappa, R. Entropy rate superpixel segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 2097–2104.
37. 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](#)]
38. Bishop, C.M. In *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.



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