

Article

Semi-Supervised Classification for Hyperspectral Images Based on Multiple Classifiers and Relaxation Strategy

Fuding Xie ¹, Dongcui Hu ¹, Fangfei Li ¹, Jun Yang ¹  and Deshan Liu ^{2,*}

¹ College of Urban and Environment, Liaoning Normal University, Dalian 116029, China; xiefd@lnnu.edu.cn (F.X.); hudongcui@outlook.com (D.H.); lff940410@outlook.com (F.L.); yangjun@lnnu.edu.cn (J.Y.)

² College of Computer Science, Liaoning Normal University, Dalian 116081, China

* Correspondence: liudeshan@lnnu.edu.cn; Tel.: +86-411-8215-3535

Received: 11 May 2018; Accepted: 19 July 2018; Published: 23 July 2018



Abstract: Hyperspectral image (HSI) classification is a fundamental and challenging problem in remote sensing and its various applications. However, it is difficult to perfectly classify remotely sensed hyperspectral data by directly using classification techniques developed in pattern recognition. This is partially owing to a multitude of noise points and the limited training samples. Based on multinomial logistic regression (MLR), the local mean-based pseudo nearest neighbor (LMPNN) rule, and the discontinuity preserving relaxation (DPR) method, in this paper, a semi-supervised method for HSI classification is proposed. In pre-processing and post-processing, the DPR strategy is adopted to denoise the original hyperspectral data and improve the classification accuracy, respectively. The application of two classifiers, MLR and LMPNN, can automatically acquire more labeled samples in terms of a few labeled instances per class. This is termed the pre-classification procedure. The final classification result of the HSI is obtained by employing the MLRsub approach. The effectiveness of the proposal is experimentally evaluated by two real hyperspectral datasets, which are widely used to test the performance of the HSI classification algorithm. The comparison results using several competing methods confirm that the proposed method is effective, even for limited training samples.

Keywords: hyperspectral image; semi-supervised classification; multinomial logistic regression; local mean-based pseudo nearest neighbor; discontinuity preserving relaxation

1. Introduction

Owing to the special advantages of a wide spectral range, high spectral resolution, and continuous spectral curve, hyperspectral remote-sensing images have been widely applied in earth observation [1,2]. As a fundamental and challenging problem in remote sensing, hyperspectral image (HSI) classification and its various applications have attracted increasing attention in recent years. Many methods have been introduced to classify HSI, attempting to detect a perfect classification result for a specific HSI. According to whether the class labels of the samples are used in the classification process, the existing classification approaches are generally partitioned into three categories: unsupervised/cluster [3,4], supervised [5–7], and semi-supervised [8,9]. Although a satisfactory classification result can be obtained by the supervised classification method, the acquisition of labeled samples is laborious and time consuming and also depends on expert knowledge. Contrarily, sufficient unlabeled samples can be acquired. The unsupervised classification does not require prior knowledge of the hyperspectral dataset; however, the cluster purity is often not satisfied with the corresponding absence of the discriminative information. Compared with the unsupervised

and supervised classifications, semi-supervised classification applies partial labeled samples and a large number of unlabeled instances, aiming at achieving a better classification effect. The classification results provided by semi-supervised methods generally have a close relationship to the size of the labeled samples. A number of studies show that semi-supervised HSI classification is a powerful and promising technique. Nevertheless, it is still a challenging problem to propose more powerful semi-supervised methods to classify remotely sensed hyperspectral data in cases of limited training samples.

Unlike other datasets, the hyperspectral dataset is depicted by both spectral information and definite spatial information. Tobler's first law [10] shows that all attribute values on a geographic surface are related to each other, but closer values are more strongly related than those from farther away. Tobler's first law implies that the spatial neighboring pixels in remote-sensing images are likely to be the same class label. Therefore, it is natural to consider the spectral and spatial information together in HSI classification. Plenty of spatial-spectral-based classification methods have been developed in the past decades [11–15].

Hyperspectral data are known to contain plenty of noise points. These noise points can evidently affect the final classification accuracy. The spectral-spatial technique is also used to reduce noise points for the HSI pre-processing task or to improve the classification accuracy for an obtained classification result in post-processing. As a typical application of this technique, probabilistic relaxation (PR) uses the local relationship among spatially adjacent pixels to correct spectral or spatial distortion. In the probability sense, PR is applied to exploit the continuity of neighboring labels [16].

Perhaps the most sophisticated PR technology is based on the Markov random fields (MRFs). In particular, the MRFs have achieved prominent performance on refining the classification results by characterizing the spatial information [17–21]. For instance, Yu et al. proposed a novel classification method by integrating an adaptive MRF with a subspace-based support vector machine (SVM) classifier [17]. The class labels predicted in the pre-classification process were altered through an adaptive MRF approach. In [18], an adaptive MRF was applied to optimize the classified image provided by the threshold exponential spectral angle map classifier, which improved the classification performance markedly. To an extent, it proved that the incorporation of a classifier and spatial smoothing technique can yield the ideal result. A novel probabilistic label relaxation strategy was proposed in [22], which combined the evidence of adjacent label assignments to effectively eliminate label ambiguity. Nevertheless, it is often observed that the effect of using spatial information as a relaxation is not ideal. On the one hand, it significantly improves the classification accuracy of smooth image regions. On the other hand, the boundary of the class is blurred by the forced smoothness. Therefore, a more powerful PR strategy to handle this problem is urgently required. Wang et al. proposed a novel two-step MRF regularization method [23], which addressed the problem of over-smoothing at the image boundary areas based on the spatial regularizing methodology of the MRF. By detecting the discontinuities of a band image in advance, the results could be locally smoothed without crossing the boundaries by the discontinuity preserving relaxation (DPR) method [24]. More recently, a new relaxation method [16], DPR was introduced to logically smooth the original hyperspectral image or the classification results, using both spatial and spectral information while maintaining the discontinuity extracted from the hyperspectral data cube.

Currently, many machine-learning algorithms have been widely used in HSI classification, such as SVM [25], multinomial logistic regression (MLR) [26–28], k -nearest neighbor (KNN) [29,30], the local mean-based pseudo nearest neighbor (LMPNN) method [31], and artificial neural networks (ANNs) [32,33]. Using unlabeled samples actively selected from the hyperspectral dataset, Prabhakar et al. extended the MLR algorithm to a semi-supervised learning of the posterior class distribution, promoting classification results with smaller training datasets and less complexity [26]. Based on the integration of MLR with the subspace projection method, MLRsub was proposed in [27]. It assumes that the samples in each class could be located approximately into a lower dimensional subspace. The use of the subspace projection method can effectively avoid the spectral confusion caused by mixed

pixels [28]. To further enhance the KNN-based classification performance, LMPNN [31] was presented in 2014, which is based upon the concept of local mean-based k -nearest neighbor rule (LMKNN) [34] and PNN [35] rules. Additionally, a novel semi-supervised classification approach for hyperspectral images was introduced by Tan et al. [29]. To further improve the classification accuracy, the authors combined the KNN and MLR to determine the class labels of the selected unlabeled samples.

For limited training samples, it is typically difficult to provide a satisfactory classification result for an HSI. Motivated by the work of the classifier combination, a pre-classification technique has been developed, attempting to address the problem mentioned above. Based on the pre-classification, relaxation strategy and the MLRsub algorithm, a semi-supervised method for HSI classification consisting of four steps is proposed in this work.

The primary contributions of this paper are as follows:

1. In pre-processing and post-processing, the DPR strategy combined with the Roberts cross operator is adopted to denoise the original hyperspectral data and improve the classification accuracy, respectively.
2. A new classifier combination for the pre-classification process of HSIs is proposed, which addresses the problem of automatically labeling samples based on a small training set. Two classifiers, MLRsub and LMPNN, are used together to perform the pre-classification of automatically predicting more labeled samples in terms of a few labeled instances per class.
3. A novel semi-supervised classification scheme is built by four steps: pre-processing, pre-classification, classification, and post-processing.

The remainder of this paper is organized as follows. In Section 2, we first briefly introduce the related classifiers and DPR algorithm and subsequently present the proposed classification method. Section 3 evaluates the performances of the proposal. Some summarizing remarks and conclusions are presented in the last section.

2. The Proposed Semi-Supervised Classification Method

We first define the notations used in this study and subsequently introduce the DPR method, MLRsub, and LMPNN classifiers. Finally, the proposed semi-supervised classification method is depicted in detail.

Let $X = (x_1, x_2, \dots, x_n)$ be a hyperspectral dataset with n pixels, where $x_i = (x_{i1}, x_{i2}, \dots, x_{id})^T$ indicates a spectral vector associated with pixel i . $y = (y_1, y_2, \dots, y_n)$ denotes an image of class labels, $y_i \in \{1, 2, \dots, K\}$ assuming that K classes exist in X .

Let $N(x_i)$ denote a collection of spatial neighbors of pixel i . We herein adopt the Moore neighbor, which is defined as follows:

$$N(x_i) = \{x_j \mid |i_1 - j_1| \leq 1 \text{ or } |i_2 - j_2| \leq 1\} \quad (1)$$

where (i_1, i_2) and (j_1, j_2) are the spatial coordinates of pixels i and j , respectively.

2.1. Relaxation Method

Relaxation is a technique that utilizes the local spatial relationship among neighboring pixels to denoise hyperspectral data and enhance the spatial texture information in pre-processing and improve classification accuracy in post-processing. The relaxation strategy is in fact a moving window smoothing technology widely used to reduce noise in a time series. Recently, combinations of SVM and MFR [36] or MLR and MFR [28] methods have shown outstanding performance in HSI classification. Although the use of spatial information in the relaxation strategy can effectively refine the classification result in smooth image areas, it also blurs the boundaries of classes, rendering it difficult to obtain a better classification. To preserve the advantage and overcome the disadvantage of the method based

on relaxation, Li et al. [16] proposed a DPR strategy for HSI classification. This method attempted to accurately preserve the edges of class boundaries while smoothing remotely sensed hyperspectral data.

The DPR strategy can be described as follows.

For a given a hyperspectral dataset X , let $P = [p_1, \dots, p_n] \in R^{k \times n}$, $p_i = [p_i(1), \dots, p_i(k)]^T$, and $p_i(j)$ be the probability of pixel i belonging to the j -th class. Let $U = [u_1, \dots, u_n] \in R^{k \times n}$, and $u_i = [u_i(1), \dots, u_i(k)]^T$ be the final k -dimensional vector of the probabilities provided by the relaxation process.

A relaxation strategy will be obtained by solving the following optimization problem:

$$\begin{cases} \min_u (1 - \gamma) \|U - P\|^2 + \gamma \sum_i \sum_{x_j \in N(x_i)} \delta_j \|u_j - u_i\|^2 \\ \text{s.t.} : u_i \geq 0, 1^T u_i = 1 \end{cases} \quad (2)$$

where γ is a parameter balancing the relative impact of both terms in Equation (2) and δ_j is a value of the pixel j of the edge image. δ is calculated by Equation (3):

$$\delta = \exp - \left(\sum_{i=1}^d \text{sobel} (X^{(i)}) \right) \quad (3)$$

where Sobel () represents the Sobel filter that detects the discontinuities in a band image. $X^{(i)}$ denotes the i -th band of the original data cube X . More details of this method can be found in [16].

In Equation (2), the first term measures the misfit of the data, and the second term facilitates the smooth solutions weighted by the parameter δ . When no discontinuity exists between the adjacent pixels to which it is connected, δ is large. Conversely, when discontinuities exist, δ is small [16].

The Roberts cross operator, initially proposed by Roberts in 1963 [37], is one of the first edge detectors and is used in image processing for edge detection. The basic idea of the Roberts cross operator is to approximate the gradient of an image by discrete differentiation, which is realized through calculating the squared sum of the differences between diagonally adjacent pixels. When the Roberts cross operator appears on the text, it is simple, easy to compute, and more accurate to detect an edge location. Considering these merits of the Roberts cross operator, we replace the Sobel filter with the Roberts cross operator in Equation (3). The experimental results on three hyperspectral datasets show that this substitution can achieve the ideal classification results.

2.2. Multinomial Logistic Regression (MLR)

Unlike the hard classification method, MLR, a probabilistic classifier suggested by Böhning [38], is a soft classification technique to calculate the posterior class density $p(y_i|x_i)$. It indicates that each pixel of the HSI belongs to different classes with different probabilities. In fact, the classification result provided by MLR can more accurately reflect the relationship between pixels and classes.

The MLR classifier is modeled by the following:

$$p(y_i = k|x_i, \omega) = \frac{\exp(\omega^{(k)} h(x_i))}{\sum_{j=1}^k \exp(\omega^{(j)} h(x_i))} \quad (4)$$

where $h(x) = [h_1(x), \dots, h_t(x)]^T$ is a vector of t fixed functions of the input, which is often termed as features. $\omega^{(k)}$ is the set of logistic regressors for class k that can be obtained by the LORSAL (logistic regression via variable splitting and augmented Lagrangian algorithm) [39], and $\omega = (\omega^{(1)T}, \dots, \omega^{(K-1)T})$. LORSAL promotes the sparsity of the weights by a Thikonov regularization and an L_1 weight on the priors.

The function h can use linear or nonlinear functions to handle different problems. In [2,39], the Gaussian radial basis function kernel was adopted to compute the function h . The kernel technique is

often used to handle linear inseparability in low-dimensional space in machine learning and to classify remotely sensed hyperspectral data [2,39]. In this work, we prefer to use the following input function $h(x_i)$ in Equation (4) [27,40]:

$$h(x_i) = [||x_i||^2, ||x_i^T U^{(1)}||^2, \dots, ||x_i^T U^{(K)}||^2]^T \quad (5)$$

where $U^{(j)} = \{u_1^{(j)}, \dots, u_{r^{(j)}}^{(j)}\}$ is a set of $r^{(j)}$ —dimensional orthonormal basis vectors for the subspace associated with class j ($r^{(j)} \ll d$).

2.3. Local Mean-Based Pseudo Nearest Neighbor (LMPNN)

As a popular algorithm in machine learning, the KNN rule [41] is a simple and effective classification method. It yields a classification decision by computing the k -nearest neighbors of an unlabeled sample and a simple vote principle. However, the classification result of the KNN classifier is sometimes dependent on the choice of the k value. Furthermore, it does not always provide an optimal solution for an unbalanced dataset. To handle these issues, a multitude of improvements in KNN-based approaches have been suggested in recent years, such as the weighted k -nearest neighbor rule [42], pseudo nearest neighbor rule [35], local mean-based k -nearest neighbor rule [34], and local mean-based pseudo nearest neighbor [31]. The basic idea behind these methods is to assign a greater weight to the nearest neighbor or replace real neighbors with pseudo neighbors.

The LMPNN rule can be described by the following steps [31]:

(1) For the unlabeled sample x , search for the k -nearest neighbors from each class of the training set; $x_1^i, x_2^i, \dots, x_k^i$ denote the k -nearest neighbors selected from the i -th class and are arranged in the ascending order according to their distances from x .

(2) Compute the local average \bar{x}_j^i of the first j -nearest neighbors of sample x from the i -th class

$$\bar{x}_j^i = \frac{1}{j} \sum_{s=1}^j x_s^i. \quad (6)$$

(3) Assign weight $1/j$ to the local mean value \bar{x}_j^i ($j = 1, 2, \dots, k$).

(4) Calculate the distance from the sample x to the i -th class by Equation (7)

$$d(x, \bar{x}^i) = d(x, \bar{x}_1^i) + \frac{1}{2}d(x, \bar{x}_2^i) + \dots + \frac{1}{k}d(x, \bar{x}_k^i) \quad (7)$$

where $d(x, \bar{x}_j^i)$ denotes the distances from x to \bar{x}_j^i and $d(x, \bar{x}^i)$ the distance from x to the i -th class.

(5) Determine the class label of sample x . The sample x belongs to the c -th class if the following is true.

$$c = \arg \min_i d(x, \bar{x}^i). \quad (8)$$

2.4. The Proposed Method

It is known that classification results obtained by semi-supervised approaches generally depend on the volume of the training set. For limited training samples, certain semi-supervised classification methods do not always provide satisfactory classification results. Obviously, the acquisition of a number of labeled samples is difficult and sometimes even impossible.

In this subsection, we suggest a semi-supervised technique for HSI classification based on the DPR strategy. The proposal will be introduced in following three steps.

Step 1: Data Pre-Processing

In this step, the DPR strategy is used to denoise hyperspectral data. Unlike the DPR technique proposed by Li et al. [16], in this study, the Sobel filter is replaced by the Roberts cross operator in Equation (3). It can be rewritten in the following form:

$$\delta = exp - \left(\sum_{i=1}^d Roberts (X^{(i)}) \right). \quad (9)$$

For each band in the dataset X , we use Equation (10) to update it constantly as follows:

$$\tilde{x}_{is}^{(t+1)} = \frac{(1 - \gamma)x_{is} + \gamma \sum_{x_j \in N(x_i)} \delta_j \tilde{x}_{js}^{(t)}}{(1 - \gamma) + \gamma \sum_{x_j \in N(x_i)} \delta_j} \quad (s = 1, 2, \dots, d) \quad (10)$$

where $\tilde{x}_{is}^{(t)}$ denotes the value of the s -th band of the pixel x_i in the t -th iteration.

The update process terminates if Equation (11) is met.

$$\|E_s^{(t+1)} - E_s^{(t)}\| < \varepsilon, \quad E_s^{(t+1)} = \frac{\|\tilde{x}_{*s}^{(t+1)} - \tilde{x}_{*s}^{(t)}\|}{\|\tilde{x}_{*s}^{(t)}\|} \quad (11)$$

where \tilde{x}_{*s} indicates the s -th band image of the HSI and ε is a predetermined threshold.

The experimental results show that the substitution of the edge detection operator can improve the classification accuracy.

Step 2: Classification

The proposed semi-supervised classification method includes two stages: pre-classification and classification.

In the pre-classification stage, two classifiers, MLRsub and LMPNN, are used to predict the class labels of unlabeled pixels using limited training samples per class. Pre-classification can also be considered as a technique for automatically labeling samples or expanding the training set in terms of a few labeled instances.

Specifically, for any unlabeled pixel x_i , we use the MLRsub classifier and LMPNN classifier together to obtain its class label provided that pixel x_i belongs to the k -th class by the MLRsub algorithm, denoted as $y_{MLR} = k$, and the s -th class by the LMPNN method (i.e., $y_{LMPNN} = s$). If $y_{MLR} = y_{LMPNN} = k$, subsequently pixel x_i is stamped on the label of the k -th class (i.e., $y_i = k$). Otherwise, sample x_i is placed into a set in which the class labels of all samples are to be further determined. This procedure is termed pre-classification in our study.

We empirically discovered that it is effective to automatically acquire more labeled samples in terms of a few labeled instances per class. It is inevitable that some pixels are mislabeled during pre-classification. Therefore, this procedure is not executed iteratively to avoid more samples being misclassified.

For samples that have not been labeled during pre-classification, we apply the MLRsub method again to label them based on those labeled samples obtained in pre-classification along with the initial training set. Currently, we have obtained the final classification result.

Step 3: Post-Processing

To improve the classification accuracy, it is necessary to reprocess the obtained classification result. In this step, we employ the DPR strategy introduced in Section 2.1 to correct the class labels of some misclassified samples.

3. Experimental Results

We conducted some experiments to validate the performance of our proposal. Two airborne visible/infrared imaging spectrometer sensor (AVIRIS) HSI datasets (i.e., Indian Pines and Salinas)

were chosen to illustrate our method. As a benchmark, these two datasets are often used to test the performance of classification algorithms for HSIs. The performance of the proposed algorithm was also compared with four competing classification techniques.

In our experiments, the initial small-size training sets have been obtained by randomly selecting 5, 10, and 15 samples from each class, respectively. To reduce the deviation of randomly labeling for the classification result, all experiments were performed over 10 independent trials with a random choice of the training set. The final result is obtained by the mean value of the result and its standard deviation. To properly assess the classification results of the HSI, two popular indices, the overall accuracy (OA) and kappa coefficient (KC), were adopted in this study.

To facilitate the reading of the obtained experimental results, we shall introduce the following shorthand notations.

PMKM: Pre-processing + MLRsub + KNN + MLRsub.

PMLM: Pre-processing + MLRsub + LMPNN + MLRsub.

PMKMP: Pre-processing + MLRsub + KNN + MLRsub + Post-processing.

PMLMP: Pre-processing + MLRsub + LMPNN + MLRsub + Post-processing.

3.1. Datasets and Classification Results

The Indian Pines image was captured at the Indian Pines test station in northwestern Indiana, United States of America (USA), in 1992. The scene is composed of crops, forests, and other natural perennial vegetation. It has 220 successive spectral channels that range from 400 nm to 2500 nm and 145 pixels by 145 pixels with a spatial resolution of 20 m/pixel. Owing to the atmospheric water absorption, the spectral bands 104–108, 150–163, and 220 are removed to prevent them from affecting the classification performance, resulting in 200 available spectral bands. A total of 16 classes of terrestrial objects are available in the ground-truth map. The false color composition image and ground truth of the Indian Pines dataset are shown in Figure 1.

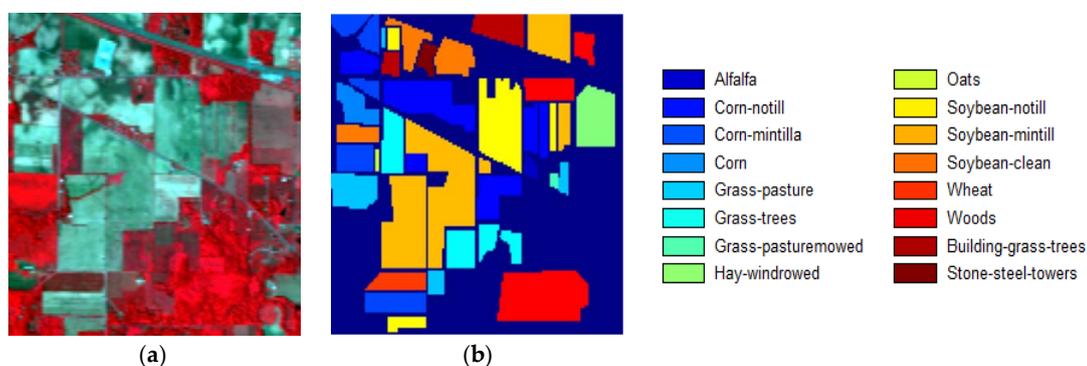


Figure 1. Indian Pines dataset. (a) False color composition; (b) Ground-truth.

The Salinas dataset was gathered by the AVIRIS sensor over the Salinas Valley in California, USA, with a geometric resolution of 3.7 m per pixel. The size of the Salinas image is $512 \times 217 \times 224$, where 20 water absorption spectral bands with insufficient information were removed; thus 204 available spectral bands remained. A total of 16 classes of ground-truth are included in the reference image. Figure 2 depicts the false-color composite image and the ground-truth image of the Salinas dataset.

These two hyperspectral datasets are available from http://www.ehu.es/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes.

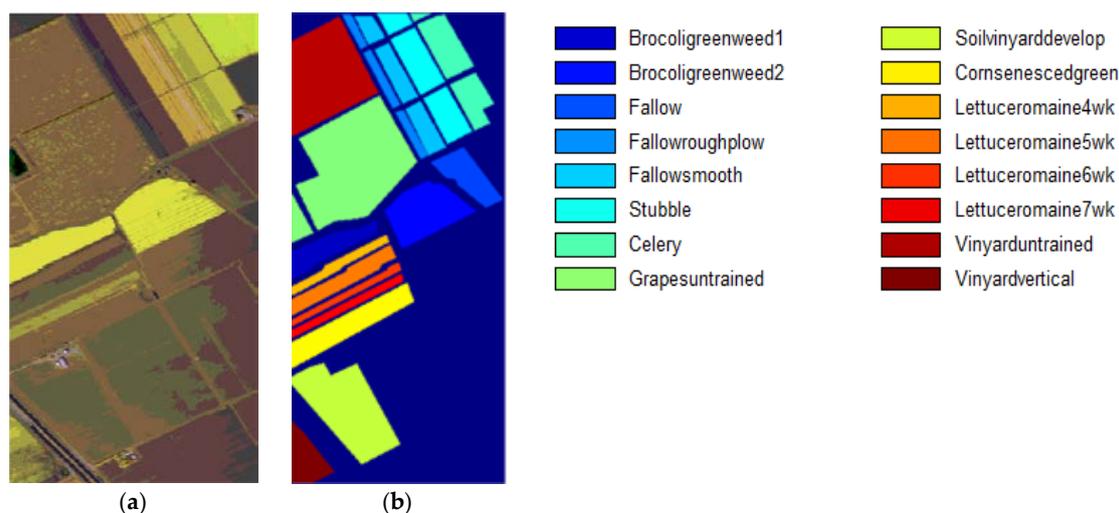


Figure 2. Salinas dataset. (a) False color composition; (b) Ground-truth.

Based on the initial training sets and the proposed method, the classification results on the Indian Pines and Salinas datasets are reported in Table 1. For Indian Pines dataset, the best classification accuracy over OA is 91.18% on average, in the case of randomly selecting 15 instances from each class. In other words, this result is obtained from approximately 2.3% of the total labeled samples. The classification accuracy of this dataset may not be as high as anticipated because of its serious data disequilibrium. The perfect classification result on the Salinas dataset has been obtained even if only 0.15% of the samples are labeled (label 5 samples per class). This fact successfully demonstrates that the proposal is valid for HSI classification with very small training sets. Table 1 also indicates that the classification result provided using the LMPNN classifier is better than that achieved by applying the KNN method of pre-classification. It is noteworthy that the application of post-processing technology has significantly improved the classification result. The classification accuracy and the standard deviation of each class in the Indian Pines and Salinas datasets are recorded in Tables 2 and 3, respectively. In Figure 3, we present the classification map obtained by PMLMP with different numbers of labeled sample.

Table 1. Classification results of the Indian Pines and Salinas datasets by different methods with different label proportions.

Dataset	Method	5		10		15	
		OA (%)	KC (%)	OA (%)	KC (%)	OA (%)	KC (%)
Indian Pines	PMKM	74.85 ± 1.10	71.55 ± 1.43	84.53 ± 1.69	82.48 ± 2.13	88.73 ± 3.99	87.22 ± 4.91
	PMLM	75.05 ± 3.57	71.87 ± 4.48	84.57 ± 2.12	82.55 ± 2.67	88.95 ± 1.25	87.48 ± 1.59
	PMKMP	76.89 ± 1.36	73.80 ± 1.56	86.45 ± 1.08	84.64 ± 1.36	90.46 ± 4.10	89.17 ± 5.14
	PMLMP	77.23 ± 2.85	74.27 ± 3.71	86.69 ± 1.39	84.92 ± 1.72	91.18 ± 2.30	89.98 ± 3.90
Salinas	PMKM	92.16 ± 1.06	91.30 ± 1.28	95.11 ± 5.76	94.58 ± 7.08	96.77 ± 1.29	96.41 ± 1.59
	PMLM	95.22 ± 3.65	94.69 ± 4.48	96.02 ± 6.84	95.58 ± 8.45	97.12 ± 2.50	96.80 ± 3.08
	PMKMP	93.05 ± 1.68	92.28 ± 2.04	95.90 ± 4.02	95.45 ± 4.93	97.42 ± 5.30	97.13 ± 6.55
	PMLMP	96.00 ± 2.98	95.56 ± 3.66	96.82 ± 4.00	96.46 ± 4.99	97.96 ± 3.83	97.74 ± 4.73

Table 2. Classification accuracy over overall accuracy (OA (%)) and standard deviations for each class in the Indian Pines dataset.

	5				10				15			
	PMKM	PMLM	PMKMP	PMLMP	PMKM	PMLM	PMKMP	PMLMP	PMKM	PMLM	PMKMP	PMLMP
Alfalfa (46)	93.70 ± 1.90	95.22 ± 2.00	93.48 ± 1.77	92.39 ± 1.85	94.35 ± 2.34	94.35 ± 2.93	94.13 ± 1.47	94.35 ± 2.06	97.61 ± 2.16	96.09 ± 1.71	95.00 ± 2.24	93.91 ± 1.05
Corn-no till (1428)	43.10 ± 10.25	48.40 ± 11.81	45.87 ± 13.45	51.23 ± 12.04	72.25 ± 4.22	68.93 ± 6.29	68.12 ± 10.32	68.93 ± 5.26	76.67 ± 5.45	80.08 ± 5.36	80.86 ± 5.90	84.46 ± 7.66
Corn-min till (830)	60.18 ± 12.57	60.80 ± 13.36	63.00 ± 16.62	68.82 ± 13.69	81.14 ± 8.86	78.88 ± 5.72	85.70 ± 8.58	78.88 ± 5.81	80.48 ± 9.82	83.94 ± 7.43	85.11 ± 8.12	85.16 ± 9.20
Corn (237)	85.15 ± 8.05	75.61 ± 17.69	84.51 ± 8.77	78.99 ± 19.81	83.88 ± 11.39	88.44 ± 9.43	84.94 ± 11.38	88.44 ± 7.78	90.89 ± 4.53	90.55 ± 4.36	90.51 ± 4.85	89.79 ± 6.37
Grass-pasture (483)	75.01 ± 8.56	79.05 ± 9.36	76.87 ± 11.00	73.31 ± 19.42	89.65 ± 3.62	87.74 ± 4.98	91.78 ± 3.85	87.74 ± 4.24	91.16 ± 4.07	89.23 ± 9.05	92.63 ± 3.79	91.93 ± 4.33
Grass-trees (730)	99.74 ± 0.56	99.49 ± 0.50	99.96 ± 0.10	98.59 ± 2.88	99.62 ± 0.64	98.84 ± 2.03	99.82 ± 0.35	98.84 ± 0.59	99.62 ± 0.61	99.71 ± 0.46	99.99 ± 0.00	99.99 ± 0.00
Grass-pasture-mowed (28)	99.64 ± 1.13	100.00 ± 0.00	98.21 ± 3.86	100.00 ± 0.00	100.00 ± 0.00	98.57 ± 4.52	87.14 ± 16.43	98.57 ± 14.60	99.29 ± 2.26	100.00 ± 0.00	92.50 ± 10.13	94.29 ± 12.42
Hay-windrowed (478)	93.39 ± 3.91	93.77 ± 3.65	95.31 ± 2.64	95.42 ± 3.52	94.85 ± 4.26	92.76 ± 12.51	95.77 ± 3.67	92.76 ± 12.85	96.51 ± 4.81	96.99 ± 3.18	94.21 ± 0.81	98.37 ± 7.93
Oats (20)	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
Soybean-no till (972)	78.77 ± 11.50	82.35 ± 8.19	80.94 ± 12.01	83.08 ± 9.77	89.87 ± 7.02	86.77 ± 5.39	91.97 ± 6.61	86.77 ± 4.64	92.12 ± 5.74	91.87 ± 5.16	92.87 ± 5.13	93.55 ± 5.62
Soybean-min till (2455)	75.82 ± 3.87	71.60 ± 5.59	77.30 ± 5.31	73.57 ± 5.33	78.37 ± 4.98	79.97 ± 4.25	82.65 ± 2.62	79.97 ± 5.07	85.62 ± 3.13	85.92 ± 2.96	88.47 ± 3.58	88.21 ± 3.29
Soybean-clean (593)	69.12 ± 8.33	59.98 ± 15.99	66.31 ± 16.11	66.58 ± 7.46	67.93 ± 14.29	80.12 ± 10.21	73.66 ± 14.83	80.12 ± 10.32	87.74 ± 7.74	82.56 ± 9.52	83.93 ± 6.89	84.65 ± 10.30
Wheat (205)	99.27 ± 0.81	99.32 ± 0.57	99.66 ± 0.33	99.61 ± 0.20	99.46 ± 0.96	99.46 ± 0.74	99.80 ± 0.24	99.46 ± 0.20	99.80 ± 0.35	99.46 ± 0.71	99.66 ± 0.20	99.61 ± 0.24
Woods (1265)	90.84 ± 10.19	92.51 ± 4.89	94.33 ± 7.55	94.96 ± 5.29	98.09 ± 3.12	97.64 ± 3.70	97.95 ± 3.28	97.64 ± 3.10	97.33 ± 2.82	96.93 ± 2.81	98.58 ± 2.17	98.63 ± 2.17
Buildings-Grass-Trees-Drives (386)	73.76 ± 13.78	79.72 ± 7.94	75.67 ± 14.74	79.27 ± 13.67	91.01 ± 6.32	90.78 ± 7.99	95.13 ± 7.83	90.78 ± 8.95	94.84 ± 4.16	97.98 ± 1.02	95.39 ± 4.17	97.93 ± 6.57
Stone-Steel-Towers (93)	96.24 ± 2.83	93.87 ± 4.97	95.16 ± 4.19	91.72 ± 4.51	96.67 ± 3.53	94.84 ± 2.57	93.01 ± 5.27	94.84 ± 4.18	95.59 ± 2.97	93.66 ± 4.10	93.44 ± 3.50	92.69 ± 3.77

Table 3. Classification accuracy over OA (%) and its standard deviations for each class in the Salinas dataset.

	5				10				15			
	PMKM	PMLM	PMKMP	PMLMP	PMKM	PMLM	PMKMP	PMLMP	PMKM	PMLM	PMKMP	PMLMP
Brocoli_green_weeds_1 (2009)	99.67 ± 0.56	99.27 ± 1.87	99.78 ± 0.44	99.40 ± 1.75	99.76 ± 0.28	99.54 ± 0.69	99.93 ± 0.22	99.69 ± 0.51	99.72 ± 0.42	99.79 ± 0.37	99.67 ± 0.48	99.92 ± 0.26
Brocoli_green_weeds_2 (3726)	99.90 ± 0.17	99.81 ± 0.30	100.00 ± 0.00	99.97 ± 0.00	99.82 ± 0.44	99.92 ± 0.14	99.97 ± 0.10	100.00 ± 0.00	99.98 ± 0.00	99.92 ± 0.17	100.00 ± 0.00	99.99 ± 0.00
Fallow (1976)	93.84 ± 9.67	97.02 ± 3.54	95.29 ± 9.86	98.07 ± 3.02	99.89 ± 0.22	99.35 ± 1.15	99.61 ± 1.20	99.72 ± 0.57	99.89 ± 0.22	99.92 ± 0.14	99.98 ± 0.00	99.98 ± 0.00
Fallow-rough-plow (1394)	98.27 ± 0.83	97.70 ± 1.71	98.21 ± 1.20	97.67 ± 2.52	98.42 ± 0.32	98.01 ± 0.63	98.68 ± 0.72	97.96 ± 1.16	97.77 ± 0.84	98.14 ± 0.54	97.71 ± 1.47	98.12 ± 1.01
Fallow-smooth (2678)	95.81 ± 4.48	96.52 ± 2.36	97.68 ± 3.66	97.45 ± 2.03	97.98 ± 1.08	98.85 ± 1.21	99.01 ± 0.82	99.19 ± 0.83	98.69 ± 0.77	97.66 ± 1.66	99.28 ± 0.47	98.41 ± 1.57
Stubble (3959)	99.74 ± 0.22	99.78 ± 0.20	99.96 ± 0.00	99.96 ± 0.00	99.78 ± 0.24	99.73 ± 0.24	99.97 ± 0.00	99.96 ± 0.00	99.88 ± 0.00	99.86 ± 0.10	99.96 ± 0.00	99.97 ± 0.00
Celery (3579)	99.79 ± 0.10	99.80 ± 0.10	99.94 ± 0.00	99.96 ± 0.00	99.84 ± 0.10	99.80 ± 0.10	99.95 ± 0.00	99.95 ± 0.00	99.82 ± 0.14	99.81 ± 0.14	99.95 ± 0.00	99.94 ± 0.00
Grapes-untrained (11271)	78.22 ± 8.09	86.79 ± 2.82	79.95 ± 8.33	88.78 ± 2.48	82.72 ± 3.00	87.91 ± 1.29	85.02 ± 2.89	90.20 ± 0.81	89.55 ± 1.62	84.54 ± 5.18	91.46 ± 1.16	86.69 ± 5.73
Soil-vineyard-develop (6203)	99.49 ± 0.74	99.65 ± 0.67	99.76 ± 0.55	99.84 ± 0.39	99.99 ± 0.00	99.95 ± 0.14	100.00 ± 0.00	99.99 ± 0.00	99.99 ± 0.00	99.98 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
Corn-senesced-green-weeds (3278)	85.61 ± 5.28	89.99 ± 2.94	87.30 ± 5.68	91.35 ± 2.74	92.30 ± 3.68	92.72 ± 2.51	94.29 ± 2.79	93.49 ± 2.25	94.47 ± 1.65	92.46 ± 6.59	94.96 ± 1.60	92.13 ± 7.74
Lettuce_roumaine_4weeks (1068)	99.30 ± 0.62	99.56 ± 0.56	99.65 ± 0.90	99.78 ± 0.44	99.12 ± 0.88	99.64 ± 0.35	99.50 ± 0.86	99.87 ± 0.35	98.90 ± 1.15	99.54 ± 0.69	99.77 ± 0.36	100 ± 0.00
Lettuce_roumaine_5weeks (1927)	97.37 ± 4.58	98.83 ± 2.14	97.25 ± 5.06	98.94 ± 1.83	98.52 ± 2.32	99.69 ± 0.33	99.24 ± 1.50	99.77 ± 0.26	99.75 ± 0.37	99.91 ± 0.14	99.68 ± 0.37	99.83 ± 0.24
Lettuce_roumaine_6weeks (916)	98.84 ± 1.36	97.67 ± 1.79	98.12 ± 1.85	96.62 ± 2.32	99.34 ± 0.47	99.04 ± 0.96	98.40 ± 0.99	98.22 ± 1.46	99.42 ± 0.49	98.89 ± 1.14	98.58 ± 1.24	98.72 ± 1.34
Lettuce_roumaine_7weeks (1070)	97.83 ± 1.27	97.10 ± 3.47	97.54 ± 1.98	96.02 ± 4.97	98.63 ± 0.63	98.30 ± 1.10	97.88 ± 1.18	98.57 ± 1.11	98.47 ± 0.89	98.76 ± 0.89	98.11 ± 1.32	97.84 ± 2.33
Vineyard-untrained (7268)	89.43 ± 7.66	94.72 ± 2.14	90.83 ± 7.50	95.72 ± 1.93	96.62 ± 2.54	94.94 ± 1.85	97.16 ± 2.07	96.11 ± 2.22	97.26 ± 1.66	95.15 ± 5.37	98.16 ± 1.76	96.12 ± 4.9
Vineyard-vertical-trellis (1807)	92.40 ± 5.64	96.44 ± 3.59	93.56 ± 5.19	97.60 ± 3.07	98.17 ± 1.09	98.57 ± 0.98	98.97 ± 0.85	99.13 ± 0.71	97.75 ± 3.12	97.20 ± 2.94	98.38 ± 2.86	98.38 ± 2.6

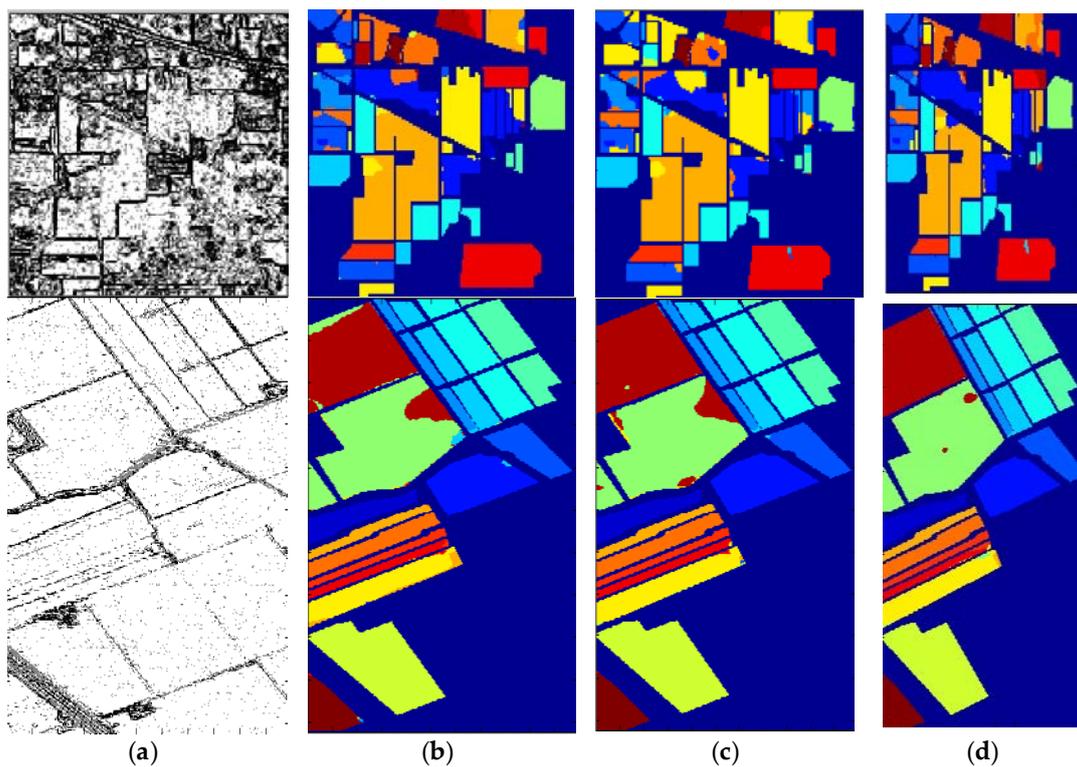


Figure 3. Discontinuity map and classification map obtained by pre-processing + MLRsub + the local mean-based pseudo nearest neighbor (LMPNN) rule + multinomial logistic regression with subspace projection (MLRsub) + post-processing (PMLMP) with different numbers of labeled samples. The above row is for the Indian Pine dataset and the below for the Salinas dataset. (a) Discontinuity map. (b–d) classification map with 5, 10, and 15 labeled samples per class, respectively.

3.2. Comparative Tests

To objectively compare our method with other competitive approaches, it is necessary to select the same dataset, labeling proportion, and methods related to the MLR technique. Therefore, we have performed comparative experiments on the Indian Pine and Salinas datasets for the case of randomly selecting 15 instances from each class. Table 4 shows the comparison results provided by MLR, MLR-MLL (multilevel logistic prior) [43], ppMLR, prMLRpr [16] approaches, and the proposed method. For the Indian Pine dataset, no significant change is shown in the classification accuracy obtained by ppMLRpr and PMLMP. A similar classification result (OA = 90.44%) has also been achieved by the MLR + KNN + SNI method [2]. However, for the Salinas dataset, the method we proposed provides a perfect classification result. The classification accuracy of this study is approximately 4% higher than that of the ppMLRpr algorithm. Tables 1 and 4 show that the results obtained by the PMLM and PMLMP techniques are still higher than those acquired by the ppMLRpr method, even for the random selection of five samples per category. This fact fully demonstrates the validity of the proposed method for remote-sensing data classification, even for the case of a small sample set.

Table 5 displays the superiority of using the Roberts cross operator over the Sobel filter in the proposed algorithm. For the random selection of 15 samples per class, the classification accuracies on both datasets have improved by at least 1%. This explains, to some extent, the rationality of the replacement scheme in pre-processing and post-processing.

Table 4. Comparison results of several methods on two hyperspectral datasets.

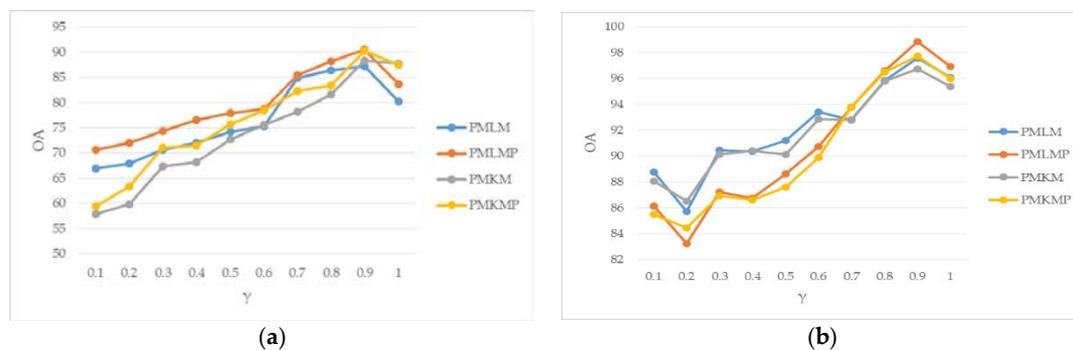
Method	Indian Pine		Salinas	
	OA (%)	KC (%)	OA (%)	KC (%)
MLR	64.30 ± 2.29	60.03 ± 2.45	85.28 ± 1.51	83.67 ± 1.66
MLR-MLL	75.09 ± 2.86	72.03 ± 3.10	89.02 ± 6.54	87.80 ± 7.28
ppMLR	88.36 ± 1.67	86.88 ± 1.86	93.30 ± 1.70	92.56 ± 1.89
ppMLRpr	91.05 ± 1.87	89.87 ± 2.09	93.79 ± 4.46	93.11 ± 4.91
PMKM	88.73 ± 3.99	87.22 ± 4.91	96.77 ± 1.29	96.41 ± 1.59
PMLM	88.95 ± 1.25	87.48 ± 1.59	97.12 ± 2.50	96.80 ± 3.08
PMKMP	90.46 ± 4.10	89.17 ± 5.14	97.42 ± 5.30	97.13 ± 6.55
PMLMP	91.18 ± 2.30	89.98 ± 3.90	97.96 ± 3.83	97.74 ± 4.73

Table 5. Comparison results of using the Roberts cross operator and Sobel filter in our proposal.

Method	Operator	Indian Pine		Salinas	
		OA (%)	KC (%)	OA (%)	KC (%)
PMKM	Sobel	85.15 ± 0.81	83.23 ± 0.96	95.70 ± 0.84	95.23 ± 1.03
	Roberts	88.73 ± 2.00	87.22 ± 2.46	96.77 ± 0.65	96.41 ± 0.80
PMKMP	Sobel	86.95 ± 0.49	85.23 ± 0.59	96.60 ± 3.55	96.22 ± 4.39
	Roberts	90.46 ± 2.05	89.17 ± 2.57	97.42 ± 2.65	97.13 ± 3.28
PMLM	Sobel	84.74 ± 0.77	82.78 ± 0.93	96.04 ± 1.25	95.60 ± 1.54
	Roberts	88.95 ± 0.63	87.48 ± 0.80	97.12 ± 1.25	96.80 ± 1.54
PMLMP	Sobel	86.33 ± 0.84	84.54 ± 1.02	96.90 ± 0.83	96.56 ± 1.00
	Roberts	91.18 ± 1.15	89.98 ± 1.95	97.96 ± 1.92	97.74 ± 2.37

3.3. Parameter Analysis

It is arduous to select the optimal parameter value in the classification algorithm. Our experience shows that the classification result generally depends on the selection of the parameter value. Figure 4 reflects the correlation between the classification result and parameter γ used in the relaxation strategy. The classification accuracy becomes increasingly better with the increase in γ . The best classification result is achieved with $\gamma = 0.9$, which is consistent with Li's selection [16]. This shows that the selection of parameter γ is independent of the selection of the operators. Parameter $\gamma = 1$ means that the boundary operator has complete control while the smoothness is ignored. Figure 3a shows that the use of the boundary operator cannot determine all the boundaries accurately; therefore, the classification accuracy is reduced.

**Figure 4.** A line chart of classification results varying with relaxation parameter on two hyperspectral images (HSIs). (a) Indian Pines dataset; (b) Salinas dataset.

Generally, the choice of the k value has a direct effect on the classification result provided by the LMPNN classifier or algorithms related to this classifier. To achieve the ideal classification result, it is clear that different k values need to be prespecified for different datasets. Figure 5 shows the relationship between the classification accuracy and k value in two datasets. Although the classification result has a slight change with the increase in k , we can still obtain the best result. Thus, in our algorithm, k is 2 for the Indian Pines dataset and 3 for the Salinas dataset.

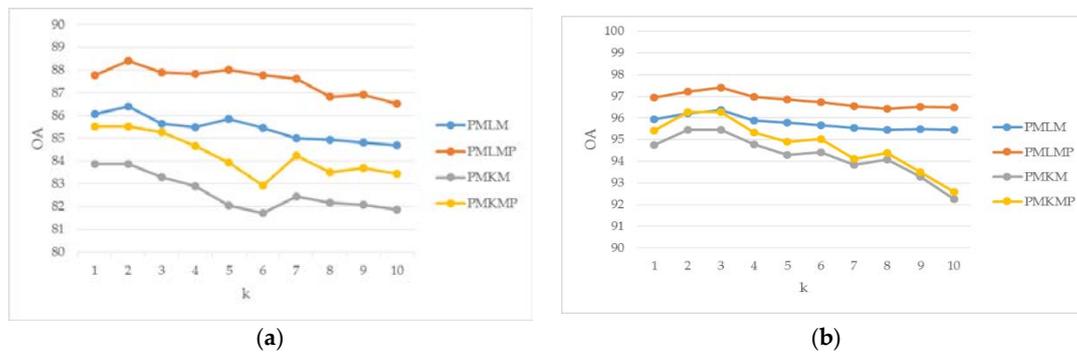


Figure 5. The relation between the classification result and parameter k . (a) Indian Pines; (b) Salinas.

4. Conclusions

In this study, we developed a novel semi-supervised method for HSI classification based on the DPR strategy and pre-classification technique. Using the DPR strategy in pre-processing and post-processing achieves the purpose of denoising and improving the classification accuracy, respectively. The pre-classification technique is aimed to address the problem of limited training samples in a semi-supervised classification. Our experimental results using the Indian Pines and Salinas datasets show that in the DPR strategy, the substitution of the Roberts cross operator with the Sobel filter can yield better classification results. The comparative test results show that the proposed method is superior to several existing methods. In particular, the proposal can provide perfect classification results for limited training samples.

Author Contributions: D.H., F.X. and D.L. conceived of and designed the experiments; D.H. and F.L. performed the experiments; J.Y., D.L., and D.H. analyzed the data and developed the graphs and tables; and F.L. and F.X. wrote the paper.

Funding: This work was supported by the National Natural Science Foundation of China (grants 41771178 and 61772252).

Acknowledgments: In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

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

References

1. Wilson, T.; Felt, R. Hyperspectral remote sensing technology (HRST) program. In Proceedings of the 1998 IEEE Aerospace Conference, Snowmass at Aspen, CO, USA, 28 March 1998; pp. 193–200.
2. Tan, K.; Hua, J.; Li, J.; Du, P. A novel semi-supervised hyperspectral image classification approach based on spatial neighborhood information and classifier combination. *ISPRS J. Photogramm. Remote Sens.* **2015**, *105*, 19–29. [[CrossRef](#)]
3. Paoli, A.; Melgani, F.; Pasoli, E. Clustering of hyperspectral images based on multi-objective particle swarm optimization. *IEEE Trans. Geosci. Remote Sens.* **2009**, *47*, 4175–4188. [[CrossRef](#)]
4. Wu, J. Unsupervised intrusion feature selection based genetic algorithm and FCM. In *Lecture Notes in Electrical Engineering*; Springer: London, UK, 2012; Volume 154, pp. 1005–1012.

5. Camps-Valls, G.; Gomez-Chova, L.; Munoz-Mari, J.; Rojo-Alvare, J.L.; Martinez-Ramon, M. Kernel-based framework for multitemporal and multisource remote sensing data classification and change detection. *IEEE Trans. Geosci. Remote Sens.* **2008**, *46*, 1822–1835. [[CrossRef](#)]
6. Bioucas-Dias, J.; Plaza, A.; Camps-Valls, G.; Scheunders, P.; Nasrabadi, N.; Chanussot, J. Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.* **2013**, *1*, 6–36. [[CrossRef](#)]
7. Ghamisi, P.; Hofle, B. LiDAR Data Classification using Extinction profiles and a composite kernel support vector machine. *IEEE Geosci. Remote Sens. Lett.* **2017**, *14*, 659–663. [[CrossRef](#)]
8. Feng, J.; Jiao, L.C.; Zhang, X.; Sun, T. Hyperspectral band selection based on trivariate mutual information and clonal selection. *IEEE Trans. Geosci. Remote Sens.* **2014**, *57*, 4092–4105. [[CrossRef](#)]
9. Feng, J.; Jiao, L.; Liu, F.; Sun, T.; Zhang, X. Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images. *Pattern Recognit.* **2016**, *51*, 295–309. [[CrossRef](#)]
10. Tobler, W. Computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* **1970**, *46*, 234–240. [[CrossRef](#)]
11. Heras, D.; Argüello, F.; Quesada-Barriuso, P. Exploring ELM-based spatial-spectral classification of hyperspectral images. *Int. J. Remote Sens.* **2014**, *35*, 401–423. [[CrossRef](#)]
12. Franchi, G.; Angulo, J. Morphological principal component analysis for hyperspectral image analysis. *ISPRS Int. J. Geo-Inf.* **2016**, *5*, 83. [[CrossRef](#)]
13. Ghamisi, P.; Dalla Mura, M.; Benediktsson, J. A Survey on Spectral-spatial classification techniques based on attribute profiles. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 2335–2353. [[CrossRef](#)]
14. Liu, J.; Xiao, Z.; Chen, Y.; Yang, J. Spatial-spectral graph regularized kernel sparse representation for hyperspectral image classification. *ISPRS Int. J. Geo-Inf.* **2017**, *6*, 258. [[CrossRef](#)]
15. Paul, S.; Nagesh Kumar, D. Spectral-spatial classification of hyperspectral data with mutual information based segmented stacked autoencoder approach. *ISPRS J. Photogramm. Remote Sens.* **2018**, *138*, 265–280. [[CrossRef](#)]
16. Li, J.; Khodadadzadeh, M.; Plaza, A.; Jia, X.; Bioucas-Dias, J.M. A discontinuity preserving relaxation scheme for spectral-spatial hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2017**, *9*, 625–639. [[CrossRef](#)]
17. Yu, H.; Gao, L.; Li, J.; Li, S.S.; Zhang, B.; Benediktsson, J.A. Spectral-spatial hyperspectral image classification using subspace-based support vector machines and adaptive Markov random fields. *Remote Sens.* **2016**, *8*, 355. [[CrossRef](#)]
18. Li, H.; Zheng, H.; Han, C.; Wang, H.; Miao, M. Onboard spectral and spatial cloud detection for hyperspectral remote sensing images. *Remote Sens.* **2018**, *10*, 152. [[CrossRef](#)]
19. Ghamisi, P.; Benediktsson, J.A.; Ulfarsson, M.O. Spectral-spatial classification of hyperspectral images based on hidden Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 2565–2574. [[CrossRef](#)]
20. Li, W.; Prasad, S.; Fowler, J.E. Hyperspectral image classification using gaussian mixture models and Markov random fields. *IEEE Geosci. Remote Sens. Lett.* **2013**, *11*, 153–157. [[CrossRef](#)]
21. Sun, L.; Wu, Z.; Liu, J.; Xiao, L.; Wei, Z. Supervised spectral-spatial hyperspectral image classification with weighted Markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2015**, *53*, 1490–1503. [[CrossRef](#)]
22. Deng, W.; Iyengar, S.S. A New Probabilistic Relaxation Scheme and Its Application to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1996**, *18*, 432–437. [[CrossRef](#)]
23. Wang, L.; Dai, Q.; Huang, X. spatial regularization of pixel-based classification maps by a two-step MRF method. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 2407–2410.
24. Gao, Q.; Lim, S.; Jia, X. Hyperspectral image classification using joint sparse model and discontinuity preserving relaxation. *IEEE Geosci. Remote Sens. Lett.* **2017**, *99*, 78–82. [[CrossRef](#)]
25. Lin, Z.; Yan, L. A support vector machine classifier based on a new kernel function model for hyperspectral data. *Mapp. Sci. Remote Sens.* **2016**, *53*, 85–101. [[CrossRef](#)]
26. Prabhakar, T.V.N.; Xavier, G.; Geetha, P.; Soman, K.P. Spatial preprocessing based multinomial logistic regression for hyperspectral image classification. *Procedia Comput. Sci.* **2015**, *46*, 1817–1826. [[CrossRef](#)]
27. Khodadadzadeh, M.; Li, J.; Plaza, A.; Bioucas-Dias, J.M. A subspace-based multinomial logistic regression for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.* **2014**, *11*, 2105–2109. [[CrossRef](#)]
28. 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](#)]

29. Denoeux, T. A k -nearest neighbor classification rule based on Dempster-Shafer theory. *IEEE Trans. Syst. Man Cybern.* **2008**, *25*, 804–813. [[CrossRef](#)]
30. Gou, J.; Zhan, Y.; Rao, Y.; Shen, X.; Wang, X.; He, W. Improved pseudo nearest neighbor classification. *Knowl.-Based Syst.* **2014**, *70*, 361–375. [[CrossRef](#)]
31. Zhang, Q.; Zhang, W.; Sun, Y.; Hu, P.; Tu, K. Detection of cold injury in peaches by hyperspectral reflectance imaging and artificial neural network. *Food Chem.* **2016**, *192*, 134–141.
32. Ahmed, A.; Duran, O.; Zweiri, Y.; Smith, M. Hybrid spectral unmixing: Using artificial neural networks for linear/non-linear switching. *Remote Sens.* **2017**, *9*, 775. [[CrossRef](#)]
33. Mitani, Y.; Hamamoto, Y. A local mean-based nonparametric classifier. *Pattern Recognit. Lett.* **2006**, *27*, 1151–1159. [[CrossRef](#)]
34. Zeng, Y.; Yang, Y.; Zhao, L. Pseudo nearest neighbor rule for pattern classification. *Expert Syst. Appl.* **2009**, *36*, 3587–3595. [[CrossRef](#)]
35. Tarabalka, Y.; Fauvel, M.; Chanussot, J.; Benediktsson, J.A. SVM and MRF-based method for accurate classification of hyperspectral images. *IEEE Geosci. Remote Sens. Lett.* **2010**, *7*, 736–740. [[CrossRef](#)]
36. Li, J.; Bioucas-Dias, J.; Plaza, A. Spectral–spatial hyperspectral image segmentation using subspace multinomial logistic regression and markov random fields. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 809–814. [[CrossRef](#)]
37. Roberts, L.G. *Machine Perception of Three-Dimensional Solids*; Outstanding Dissertations in the Computer Sciences; Garland Publishing: New York, NY, USA, 1963.
38. Böning, D. Multinomial logistic regression algorithm. *Annu. Inst. Stat. Math.* **1992**, *44*, 197–200. [[CrossRef](#)]
39. Li, J.; Bioucas-Dias, J.; Plaza, A. Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 844–856. [[CrossRef](#)]
40. Camps-Valls, G.; Bruzzone, L. Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2005**, *43*, 1351–1362. [[CrossRef](#)]
41. Cover, T.M.; Hart, P.E. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [[CrossRef](#)]
42. Dudani, S.A. The distance-weighted k -nearest neighbor rule. *IEEE Trans. Syst. Man Cybern.* **1976**, *6*, 325–327. [[CrossRef](#)]
43. Li, J.; Bioucas-Dias, J.; Plaza, A. Semi-supervised hyperspectral image segmentation using multinomial logistic regression with active learning. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 4085–4098.



© 2018 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/>).