

Article

Class Probability Propagation of Supervised Information Based on Sparse Subspace Clustering for Hyperspectral Images

Qing Yan ^{1,†}, Yun Ding ^{2,†}, Yi Xia ², Yanwen Chong ³ and Chunhou Zheng ^{1,*}

¹ College of Computer Science and Technology, Anhui University, Hefei 230601, China; rubby_yan5996@sina.com

² College of Electrical Engineering and Automation, Anhui University, Hefei 230601, China; yunding92@163.com (Y.D.); xiayi@ahu.edu.cn (Y.X.)

³ State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China; ywchong@whu.edu.cn

* Correspondence: zhengch99@126.com; Tel.: +86-182-5515-2909

† These authors contributed equally to the paper as first authors.

Received: 28 August 2017; Accepted: 28 September 2017; Published: 30 September 2017

Abstract: Hyperspectral image (HSI) clustering has drawn increasing attention due to its challenging work with respect to the curse of dimensionality. In this paper, we propose a novel class probability propagation of supervised information based on sparse subspace clustering (CPPSSC) algorithm for HSI clustering. Firstly, we estimate the class probability of unlabeled samples by way of partial known supervised information, which can be addressed by sparse representation-based classification (SRC). Then, we incorporate the class probability into the traditional sparse subspace clustering (SSC) model to obtain a more accurate sparse representation coefficient matrix accompanied by obvious block diagonalization, which will be used to build the similarity matrix. Finally, the cluster results can be obtained by applying the spectral clustering on similarity matrix. Extensive experiments on a variety of challenging data sets illustrate that our proposed method is effective.

Keywords: hyperspectral images; class probability; supervised information; sparse subspace clustering

1. Introduction

Hyperspectral images (HSIs) can provide more detailed information for land-over classification and clustering with hundreds of spectral bands for each pixel [1–4]. To a certain extent, it is difficult to process the HSI data, because many hundreds of spectral bands can cause the curse of dimensionality [5,6]. In general, the processing methods proposed by most scholars can be roughly divided into two categories. The first one is supervised learning for HSIs, which is generally called classification [7–9]. HSI classification is usually limited to the number of labeled samples, since it is time-consuming to collect large numbers of training samples [10–12]. The second category is unsupervised learning named clustering, which does not need to label a huge volume of training samples.

To our knowledge, subspace clustering is an important type of technology in signal processing and pattern recognition. It has been successfully applied to face recognition [13,14] and object segmentation [15–17], etc. Subspace clustering can extract intrinsic features from the high-dimensional data embedded in low-dimensional structures. Until now, many subspace clustering methods have been published. Generalized PCA (GPCA) [18] is a typical subspace clustering method which transforms the subspace clustering into the problem of how to fit the data with polynomials. The low-rank representation (LRR) proposed in [19] seeks the lowest-rank representation among all the

data points reconstructed by a linear combination of other points in the dataset. Then, segment data points are drawn from a union of multiple subspaces. Moreover, robust latent low-rank representation for subspace clustering (RobustLatLRR) [20] seamlessly integrates subspace clustering and feature selection into a unified framework. Peng et al. [21] have proposed construction of the l_2 -graph for robust subspace learning and subspace clustering based on a mathematically trackable property of the projection space, intrasubspace projection dominance (IPD), which can be used to eliminate the effects of the errors from the projection space rather than from the input space. Then, they proposed a novel subspace clustering method called a unified framework for representation-based subspace clustering of out-of-sample and large-scale data, which address the two limitations of some subspace clustering methods, i.e., time complexities and that they cannot tackle the out-of-sample data used to construct the similarity graph. Yuan et al. [22] proposed a novel technique named dual-clustering-based HSI classification by context analysis (DCCA), which selects the most discriminative bands to represent the original HSI and reduces the redundant information of HSI to achieve the high classification accuracy. Sparse subspace clustering (SSC) [23] has been presented to cluster data points that lie in a union of low-dimensional subspaces. It is mainly divided into two steps. The method firstly computes the sparse representation coefficient matrix from the self-expressiveness model, and secondly applies spectral clustering on similarity matrix to find the cluster results of the data. Many scholars have made some improvements on SSC to raise clustering accuracy. Zhang et al. [24] have put forward the spatial information SSC (SSC-S) and spatial-spectral SSC (S^4C) algorithms, which consider the wealthy spatial information and great spectral correlation of HSIs, and achieve better clustering results. Although the fact is that the majority of subspace clustering methods perform better in some applications, they exploit the so-called self-expressive property of the data [23].

Actually, the aforementioned clustering methods have a common obvious disadvantage. They only use unlabeled samples which have no prior information. Specifically, these methods only concentrate on unlabeled information and evidently ignore supervised information propagation, which limits the clustering precision to a large degree. In particular, this is critical for those clustering methods of exploiting the self-expressive property of the data, such as LRR and SSC. They can obtain discriminant self-expressive coefficients via limited supervised information, which play a significant role in exploiting the subspace structure. Moreover, in the process of HSI clustering, with the increase of spectral bands the clustering accuracy may decrease due to the curse of dimensionality [25,26]. Consequently, it is quite necessary to add labeled information to improve the overall accuracy of traditional clustering algorithms. Further, recently, semi-supervised learning (SSL) [27] has attracted great attention over the past decade because of its ability to make use of rich unlabeled samples via a small amount of labeled samples for effective clustering [28]. For example, Fang et al. [29] have proposed a robust semi-supervised subspace clustering method based on non-negative low-rank representation to obtain discriminant LRR coefficients, which address the overall optimum problem by combining the LRR framework and the gaussian fields and harmonic functions method. Ahn et al. [15] have proposed an multiple segmentation technique based on constrained spectral clustering via supervised information, which combines with supervised prior knowledge to build a face and hair region labeler. Convincingly, Jain [30] has published a book about semi-supervised clustering analysis, which describes in detail the theoretical knowledge. Benefiting from the development of compressed sensing [31], semi-supervised sparse representation (S^3R) has been proposed in [32], which is based on an l_1 graph to utilize both labeled and unlabeled data for inference on a graph. Yang et al. [33] have proposed a new semi-supervised low-rank representation (SSLRR) graph, which uses the calculated LRR coefficients of both labeled and unlabeled samples as the graph weights. It can capture the structure of data and implement more robust subspace clustering.

As discussed above, essentially, most of those semi-supervised clustering methods can improve the precision of clustering. However, they fail to take the class structure of data samples into account. To solve this problem, Shao et al. [28] have presented a probabilistic class structure-regularized sparse representation graph for semi-supervised hyperspectral image classification, which implies the

probabilistic relationship between each sample and each class. The supervised information of labeled instances can be efficiently propagated to the unlabeled samples through class probability [28] and further facilitates cluster correctness.

In this paper, we are motivated by probabilistic class structure insight [28], and consider the intrinsic geometric structure between labeled and unlabeled data. We thus propose a novel algorithm named class probability propagation of supervised information based on sparse subspace clustering (CPPSSC) algorithm, which combines a little supervised information with the unlabeled data to acquire the class probability. The proposed method incorporates supervised information into the SSC framework by exploring class relationship among the data samples, which can obtain the more accurate sparse coefficient matrix. Such class structure information can help the SSC model to yield a discriminative block diagonalization. To a certain extent, integrating the class probability into the sparse representation process can better assign the similar HSI pixels into the same class and concretely demonstrate a better clustering effect. Benefiting from the breakthroughs in [34,35], the optimization problem of CPPSSC can be solved by the alternating direction method of multipliers (ADMM) [36], which can reduce the computation cost. Summarily, the main contribution of this paper is as follows.

Firstly, the label information is explicitly incorporated to guide sparse representation coefficients in SSC model via estimation of the probabilistic class structure, which implies the probabilistic relationship between data points with corresponding class. Moreover, this model can be better encouraged to assign more similar elements into corresponding class. Secondly, such prior information can better capture the subspace structure of data, which can improve the self-expressiveness property of the samples and preserve the subspace-sparse representation. In other words, the block diagonalization via sparse representation tends to be more apparent.

The remainder of this paper is organized as follows. In Section 2, a brief view of the general SSC algorithm in the HSI field is given. The related work of our algorithm is presented in Section 3. Experimental results and analysis will be discussed in Section 4. Section 5 concludes this paper and outlines the future work.

2. The Brief View of General SSC Algorithm in the HSI Field

Sparse subspace clustering (SSC) is a novel framework for data clustering based on spectral clustering. Generally, high-dimensional data usually lies in a union of low-dimensional subspaces, which allows sparse representation of high-dimensional data with an appropriate dictionary [37]. The underlying idea of SSC is the self-expressing property of the data, i.e., each data point in a union of subspaces can be efficiently represented as a linear combination of other points from the same subspace [23]. Firstly, let us review the content of SSC algorithm. Let $\{S_r\}_{r=1}^n$ be an array of n linear subspaces of IR^D of dimensions $\{d_r\}_{r=1}^n$. A collection of NN data points $Y \triangleq [y_1, \dots, y_{NN}] = [Y_1, \dots, Y_n]\Gamma$ lies in the union of the n subspaces, where $Y_r \in R^{D \times NN_r}$ is a matrix that lies in S_r and $\Gamma \in R^{NN \times NN}$ is an unknown permutation matrix. It is worth noting that each data point in a union of subspaces can be efficiently reconstructed by a combination of other points in the dataset. Therefore, the SSC model utilizes the self-expressiveness property of the data to build the sparse representation model as follows:

$$\begin{aligned} \min & \left\| C \right\|_1 + \frac{\lambda}{2} \left\| E \right\|_F^2 \\ \text{s.t.} & Y = YC + E, C^T \mathbf{1} = \mathbf{1}, \text{diag}(C) = 0 \end{aligned} \quad (1)$$

where $C \triangleq [c_1 \ c_2 \ \dots \ c_{NN}] \in R^{NN \times NN}$ required to be solved is an estimated matrix whose i -th column corresponds to the sparse representation of y_i . $\text{diag}(C) = 0 \in R^{NN}$ is the vector of the diagonal elements of C to eliminate the trivial solution of self-expression. $E \in R^{d \times NN}$ is the error matrix, and λ is the tradeoff parameter between the sparse coefficient and noise matrix. After the sparse solution C is obtained, normalize the columns of C as

$$c_i = \frac{c_i}{\|c_i\|_\infty} \quad (2)$$

Now we can build the similarity matrix W as Equation (3). $W \in R^{NN \times NN}$ is a symmetric nonnegative similarity matrix.

$$W_{ij} = |C_{ij}| + |C_{ji}| \quad (3)$$

Finally, we apply spectral clustering to the similarity matrix W and get the clustering results of the data: Y_1, Y_2, \dots, Y_n .

To our knowledge, each item of hyperspectral imagery data is in 3D. Before performing the SSC algorithm, each pixel can be treated as a d -dimensional vector where d is the number of spectral bands [38] and the 3D HSI data $Y \in R^{M \times N \times d}$ must be translated into 2D matrix. In this way, the HSI data can be denoted by a 2D matrix $Y = [y_1, y_2, \dots, y_{MN}] \in R^{d \times MN}$, where M represents the width of the HSI data and N is on behalf of height of the HSI data. The sparse representation coefficient of HSIs can be obtained by utilizing Equation (1) of the SSC model. $C \in R^{MN \times MN}$ is on behalf of sparse representation coefficient matrix of HSIs data. The SSC algorithm for HSIs data can be generalized in Algorithm 1.

Algorithm 1. Sparse Subspace Clustering for HSI data

Input: MN pixel points $\{y_i\}_{i=1}^{MN}$ of d dimension from n subspaces

Step 1. Calculate sparse coefficient matrix by performing SSC model (1) on points $\{y_i\}_{i=1}^{MN}$.

Step 2. Normalize the columns of C as $c_i \leftarrow \frac{c_i}{\|c_i\|_\infty}$.

Step 3. Build the similarity matrix W according to Equation (3).

Step 4. Apply the spectral clustering to the similarity W to obtain the clustering results.

Output: clusters Y_1, Y_2, \dots, Y_n .

The process of spectral clustering can be summarized as follows. Firstly, we can obtain the Laplacian matrix L formed by $L = D - W$ where $D_{ii} = \sum_j W_{ij}$ is a diagonal matrix. Then, we obtain the clustering results by applying the K-means algorithm to the normalized rows of a matrix whose columns are the bottom eigenvectors of the symmetric normalized Laplacian matrix.

3. Class Probability Propagation of Supervised Information Based on Sparse Subspace Clustering (CPPSSC)

In this section, we firstly describe the procedure of the uniform class probability structure between the supervised information and unlabeled samples. Naturally, the basic theory for our proposed model named the class probability propagation of supervised information based on sparse subspace clustering (CPPSSC) is induced.

3.1. The Procedure of the Uniform Class Probability

For labeled samples of HSIs, the associated samples are distributed in a certain class. Nevertheless, those unlabeled samples will not possess a specific class. Fortunately, we can estimate the class probability between each unlabeled sample and each specific class by partial known supervised information [28], which can be addressed by sparse representation-based classification (SRC) [39]. According to the classic sparse representation theory, ideally, a test sample in the unlabeled samples can be written as a linear combination of the training samples from the same subspace. Hence, two samples that have nonzero coefficients in the representation will be in the same class and the coefficients denote the similarity of the two samples. Inheriting its merits, SRC can be successfully applied to estimate the class probability.

Firstly, we have total HSI data samples $\{y_i\}_{i=1}^{MN}$, and the initial l training samples of d dimensionality $Y_l = [y_1, y_2, \dots, y_l] \in R^{d \times l}$ attached to n classes $\{S_r\}_{r=1}^n$, where l training samples are from n classes. Let supervised information matrix $Q^l = [q_1, \dots, q_n]$ be an $l \times n$ binary matrix

indicating the membership of each data point to each class. That is, $q_{ij} = 1$ if the i -th label information of sample belongs to class S_j and $q_{ij} = 0$ otherwise. This is assuming that each data sample lies in only one class. Hence, we have $Q^T \mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ is the vector of all ones of appropriate dimension. Let $Y_u = [y_{l+1}, y_{l+2}, \dots, y_{l+u}]$ be test samples, where total data samples are $MN = l + u$. The similarity between the test samples $y_i \in Y_u$ and training samples Y_l can be solved by the following l_1 minimization

$$\begin{aligned} \min \|a\|_1 \\ \text{s.t. } Y_l a = y_i \end{aligned} \quad (4)$$

where $a \in R^{l \times 1}$ represents the sparse coefficient. Then, the class probability vector of y_i can be calculated by

$$p_i = a^T Q^l \quad (5)$$

where $p_i = (p_{i1}, p_{i2}, \dots, p_{in}) \in R^{1 \times n}$; the entry p_{in} represents the class probability of data y_i belonging to class S_n . For unlabeled samples, we can acquire the class probability matrix $p_U \in R^{u \times n}$ via label propagation of the given samples. For labeled samples, we denote the class probability matrix $p_L \in R^{l \times n}$ of training samples. Therefore, the probability of the objective y_i and y_j being assigned to the same class can be given by

$$P_{ij} = \begin{cases} 1 & i = j \\ p_i p_j^T & i \neq j \end{cases} \quad (6)$$

Finally, we must have normalize class probability P to guarantee $P \mathbf{1} = \mathbf{1}$.

3.2. Class Probability Propagation of Supervised Information Based on Sparse Subspace Clustering Algorithm

According to Equation (1), by applying traditional SSC theory to the HSI data clustering, we can obtain the vital character representation known as sparse representation coefficient C , which has more wealthy information. Moreover, the sparse solution with nonzero entries corresponds to data points from the same subspace, whose theoretical knowledge is similar to class probability. As the name suggests, if the data points have a higher class probability, their possibility of belonging to the same thematic class is larger. In other words, the prior knowledge can be effectively transmitted to unknown test samples via class probability. All the similar samples are pleasurable assigned into the same label by larger probabilities among samples, which are the same as nonzero entries of sparse representation coefficients. Naturally, the combination of sparse representation coefficients and class probabilities has a theoretical guarantee and can improve the global similarity structure among samples. Herein, we design a new framework in which the semi-supervised class probability information is incorporated into the objective function. This combination makes full use of the abundant correlation of the sparse representation coefficient, which will promote the cluster performance. The concrete joint formula named the class probability propagation of supervised information based on sparse subspace clustering (CPPSSC) is written as

$$\begin{aligned} \min \|PC\|_1 + \frac{\lambda}{2} \|E\|_F^2 \\ \text{s.t. } Y = YC + E, C^T \mathbf{1} = \mathbf{1}, \text{diag}(C) = 0 \end{aligned} \quad (7)$$

where class probability P can be obtained by Equation (6) and sparse representation C can be estimated by the alternating direction method of multipliers (ADMM). The element-wise product of similarity between class probability P and sparse representation coefficient C can boost the globally block subspace structure, which closely enhances the compactibilities with respect to similar samples.

This new model can be solved using the alternating direction method of multipliers (ADMM) [36]. Then, we can gain the new sparse representation coefficient via the ADMM algorithm. The details of this technique are given in the next section. The class probability propagation of supervised information based on sparse subspace clustering (CPPSSC) algorithm is summarized in Algorithm 2.

Algorithm 2. CPPSSC algorithm for HSIs

Input: The HSIs containing MN pixel points $\{y_i\}_{i=1}^{MN}$ of d dimension from n subspaces.

Main algorithm:

- (1) Calculate sparse coefficient matrix by performing CPPSSC model (7) on points $\{y_i\}_{i=1}^{MN}$ using ADMM.
- (2) Normalize the columns of C as $c_i \leftarrow \frac{c_i}{\|c_i\|_\infty}$.
- (3) Build the similarity matrix W according to Equation (3).
- (4) Apply the spectral clustering to the similarity matrix W to obtain the clustering results.

Output: Clusters Y_1, Y_2, \dots, Y_n .

3.3. The CPPSSC Algorithm Solved by ADMM

In this subsection, we briefly introduce how to solve the sparse representation coefficient of the CPPSSC algorithm model in (7) via the ADMM algorithm [36,40]. The CPPSSC model (7) can be rewritten as

$$\begin{aligned} \min \|PC\|_1 + \frac{\lambda}{2} \|Y - YC\|_F^2 \\ \text{s.t. } C^T \mathbf{1} = \mathbf{1}, \text{diag}(C) = 0 \end{aligned} \tag{8}$$

Then, an auxiliary matrix $Z \in R^{MN \times MN}$ with the same size as sparse representation C can be introduced to the model (8), and reshape this formula as

$$\begin{aligned} \min \|PC\|_1 + \frac{\lambda}{2} \|Y - YZ\|_F^2 \\ \text{s.t. } Z^T \mathbf{1} = \mathbf{1}, Z = C - \text{diag}(C) \end{aligned} \tag{9}$$

Two penalty terms with $Z^T \mathbf{1} = \mathbf{1}$ and $Z = C - \text{diag}(C)$ can be incorporated into the model (9), which is equivalent to the following optimization program:

$$\begin{aligned} \min \|PC\|_1 + \frac{\lambda}{2} \|Y - YZ\|_F^2 + \frac{\rho}{2} \|Z^T \mathbf{1} - \mathbf{1}\|_2^2 + \frac{\rho}{2} \|Z - (C - \text{diag}(C))\|_F^2 \\ \text{s.t. } Z^T \mathbf{1} = \mathbf{1}, Z = C - \text{diag}(C) \end{aligned} \tag{10}$$

Next, a vector $\alpha \in R^{MN}$ and a matrix $\beta \in R^{MN \times MN}$, which are the Lagrange multipliers for the two equality constraints with $Z^T \mathbf{1} = \mathbf{1}$ and $Z = C - \text{diag}(C)$, are added into Lagrange function as

$$\begin{aligned} L = \min \|PC\|_1 + \frac{\lambda}{2} \|Y - YZ\|_F^2 + \frac{\rho}{2} \|Z^T \mathbf{1} - \mathbf{1}\|_2^2 \\ + \frac{\rho}{2} \|Z - (C - \text{diag}(C))\|_F^2 + \alpha(Z^T \mathbf{1} - \mathbf{1}) + \text{tr}(\beta^T(Z - C + \text{diag}(C))) \end{aligned} \tag{11}$$

where $\text{tr}()$ denotes the trace operator of the given matrix.

According to the ADMM optimization program, we update each of Z, C, α and β alternatively while keeping the other variables fixed.

- (1) Update for Z . We update Z by solving the following problem.

$$Z^{k+1} = \underset{Z}{\text{argmin}} L(Z, C^k, \alpha^k, \beta^k) \tag{12}$$

Firstly, we calculate the derivation of L with respect to Z and set it to zero to obtain the results.

$$(\lambda_z \times Y^T Y + \rho \times \mathbf{1}\mathbf{1}^T + \rho \times I)Z = \lambda_z \times Y^T Y + \rho \times C - \beta + \rho \times \mathbf{1}\mathbf{1}^T - \alpha \tag{13}$$

- (2) Update for C by the following problem.

$$C^{k+1} = \underset{C}{\text{argmin}} L(Z^{k+1}, C, \alpha^k, \beta^k) \tag{14}$$

We calculate the derivation of L with respect to C and set it to zero to obtain the following formula.

$$C^{k+1} = P(J - \text{diag}(J))$$

$$J \triangleq \Gamma_{\frac{1}{\rho}}(v) = (|v| - \frac{1}{\rho})_+ \text{sign}(v), v = Z + \frac{\beta}{\rho} \tag{15}$$

The solution of C with one norm can be obtained by the iterative shrinkage algorithm. The operator $(\cdot)_+$ will return its arguments if it is nonnegative and return zero otherwise.

(3) Update for α and β , named Lagrange multipliers, which can be solved by a simple gradient ascent step.

$$\alpha^{k+1} = \alpha^k + \rho((Z^{k+1})^T \mathbf{1} - \mathbf{1}) \tag{16}$$

$$\beta^{k+1} = \beta^k + \rho(Z^{k+1} - C^{k+1}) \tag{17}$$

These three steps are alternate with iterative operation until it can finish the final convergence or the largest number of iterations exceeds the predefined values. Generally, the iteration is terminated when we have $\|(Z^k)^T \mathbf{1} - \mathbf{1}\|_{\infty} \leq \varepsilon, \|Z^k - C^k\|_{\infty} \leq \varepsilon$.

The specifically algorithmic procedure to solving CPPSSC algorithm is shown in Algorithm 3, in which the more details of iteration stops about ADMM can be referenced to [34–36].

Algorithm 3. ADMM for solving problem (8)

Input: Data samples $\{y_i\}_{i=1}^{MN}$, class probability matrix P_0 , parameter λ . Initialization: $\alpha = \mathbf{0}$, $\beta = 0, P = P_0, \lambda > 0, \rho > 0, \varepsilon = 2 \times 10^{-3}$.

While not converged do

 Update Z^k by (12).

 Update C^k and P^k by (15).

 Update α^k and β^k by (16) and (17) respectively.

 Check the convergence condition $\|Z^k \mathbf{1} - \mathbf{1}\|_{\infty} \leq \varepsilon, \|Z^k - C^k\|_{\infty} \leq \varepsilon$; if not, set

$k \leftarrow k + 1$.

End while

Output: Z^{k+1} and C^{k+1} .

The concrete CPPSSC algorithm for HSI data clustering can be interpreted graphically using Figure 1.

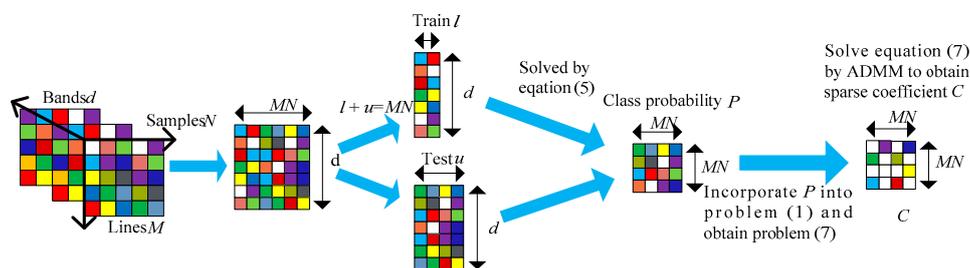


Figure 1. Calculate the sparse representation coefficient C with class probability P of our class probability propagation of supervised information based on sparse subspace clustering (CPPSSC) algorithm. ADMM: alternating direction method of multipliers.

4. Experiment and Analysis

In this section, we conduct a series of experiments to further assess the cluster effectiveness of the proposed algorithm for HSIs. To illustrate the better performance of our method, we compared our method with unsupervised clustering and semi-supervised clustering methods, respectively.

As initially mentioned, unsupervised clustering methods included SSC [23], SSC-S [24], S⁴C [24], and semi-supervised clustering methods such as S³R [32], SSLRR [33], and semi-supervised RobustLatLRR (SSRLRR) are used as benchmarks. Furthermore, S³R, SSLRR and SSRLRR make full use of 30% labeled information to obtain sparse and low rank representation coefficients in the experiment. Finally, they use related sparse representation (SR) and LRR coefficients as the weight of graph and acquire clustering results by typically normalized cuts [41]. The evaluation indicators used in this paper are user's accuracy (UA) [24], overall accuracy (OA) [42], kappa coefficient (kappa) [28], accuracy (AC) and normalized information metric (NMI) [43], which are very popular clustering indicators.

4.1. Experimental Datasets

Our proposed algorithm is evaluated using two widely used hyperspectral data sets, which are the Pavia University scene and Indian Pines. The Pavia University scene is acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor, which has the size of $610 \times 340 \times 103$ with a 1.3 m geometric resolution and has nine main classes. A typical subset of $170 \times 160 \times 103$ is selected as our objective data, with nine classes. The Indian Pines data are gathered by an Airborne Visible InfraRed Imaging Spectrometer (AVIRIS) sensor with a size of $145 \times 145 \times 220$ including sixteen classes, with a subset of $75 \times 82 \times 220$ including six classes selected as our objective data. The false color composites and the color maps of ground truth with two scenes are shown in Figures 2 and 3.

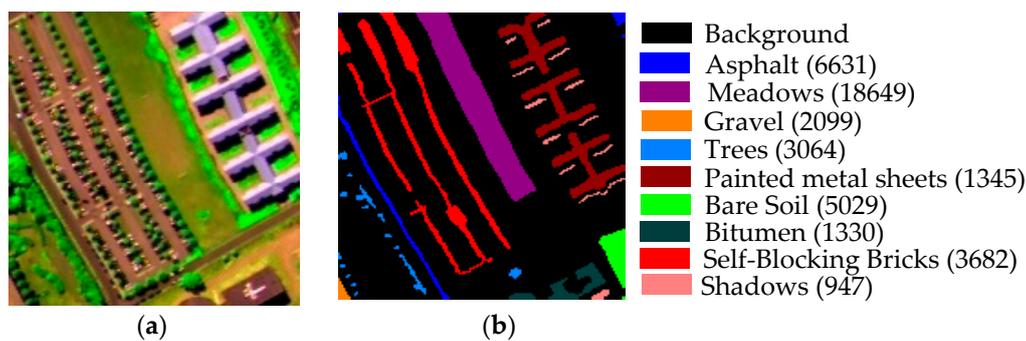


Figure 2. (a) The Pavia University scene with false color (R: 70, G: 102, B: 12); (b) Ground truth.

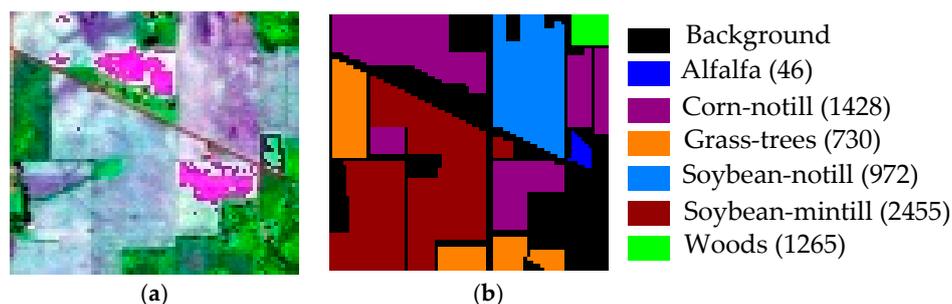


Figure 3. (a) The Indian Pines with false color (R: 10, G: 96, B: 193); (b) Ground truth.

4.2. Experimental Procedure and Analysis

4.2.1. The Quantitative Experimental Results on the Pavia University Scene

First, we conduct our CPPSSC algorithm with 30% supervised information on the Pavia University scene data set, and the experimental cluster maps compared with these benchmarks are shown in Figure 4.

From the visual effect of Figure 4, we can clearly see that our algorithm, especially for such classes as the meadows, gravel and trees, demonstrates the better clustering effect and is closer to the true ground. To a certain extent, these categories obviously achieved fewer misclassifications using our

algorithm. We can verify our observation from the Tables 1 and 2 with quantitative experimental analysis. The bold numbers are the best clustering results.

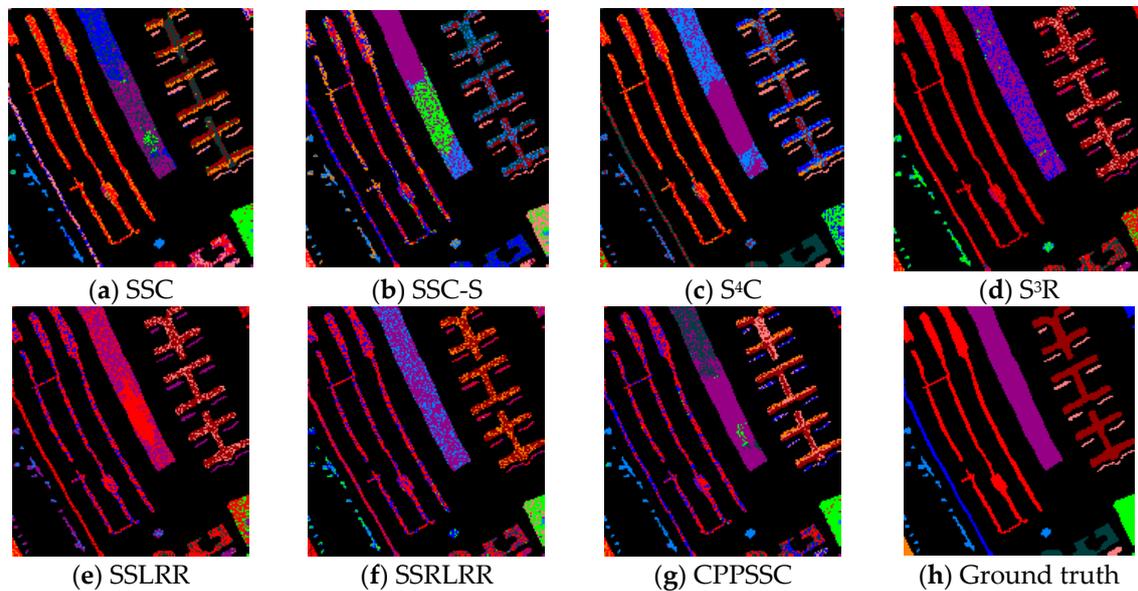


Figure 4. Cluster maps of the experimental results on Pavia University scene. SSC: sparse subspace clustering; SSC-S: spatial information SSC; S^4C :spatial-spectral SSC; S^3R :semi-supervised sparse representation; SSLRR: semi-supervised low-rank representation; SSRLRR: semi-supervised robust latent low-rank representation for subspace clustering; CPPSSC: sparse subspace clustering.

It can be seen from Table 1 that, according to the evaluation indicator with the UAs, the meadows, gravel and trees in our algorithm can possess the better performance compared with these benchmarks. The UAs are up to 69.77%, 30.36% and 91.15%, respectively. On the other hand, the quantitative analysis from Table 1 confirms our visual performance. Moreover, from the UA point of view, the SSC, S^4C and SSRLRR algorithms absolutely misclassify the pixels of gravel, while the SSC-S obtains a poor accuracy of 1.79%. Apparently, their recognition effect is not satisfied. Fortunately, our CPPSSC superiorly reaches the best UA with 30.36%, which effectively exceeds the other methods and possesses the more correct pixels. The main reason is that our algorithm can deliver the supervised information to the sparse representation process, whose theoretical knowledge is similar to S^3R and SSLRR. In addition, relatively better clustering effects for asphalt and bare soil can also be acquired by our algorithm, although they are not the best ones. From Table 2, the OA and kappa of our CPPSSC algorithm are the best compared with the other benchmarks, achieving 62.91% in OA and 0.5330 in kappa. The SSC-S and S^4C combined HSI spectral information with the wealthy spatial correlation, obtaining OAs of 48.35% and 54.87%, respectively, and also obtained kappas of 0.4037 and 0.4625 separately. It can be seen that they have a limited clustering effect for lacking known supervised information. The SSC and CPPSSC obtained OAs of 51.37% and 62.91%, and can also obtain kappa coefficients of 0.4353 and 0.5530, respectively. In other words, an CPPSSC algorithm evidently generates sharp growth, with 11.54% of OA, and achieves an improvement of 0.1177 in kappa. Compared with S^4C , our CPPSSC presents an apparent rise in OA with 8.04%, which also obtains Kappa coefficients with the distinct growth of 0.0905. This comes from the fact that our algorithm successfully utilizes the supervised information to propagate the probability about whether two samples belong to the same class and eventually deduce to intrinsic sparse representation coefficients. Although S^3R and SSRLRR can obtain better OAs at 58.14% and 58.19%, and kappas of 0.4707 and 0.4826, respectively, the CPPSSC can obtain better clustering performance via class probability structure between data samples and classes, which assists sparse representation coefficients by yielding a more discriminative diagonalization.

Table 1. The quantitative analysis on the Pavia University scene. UA: user accuracy.

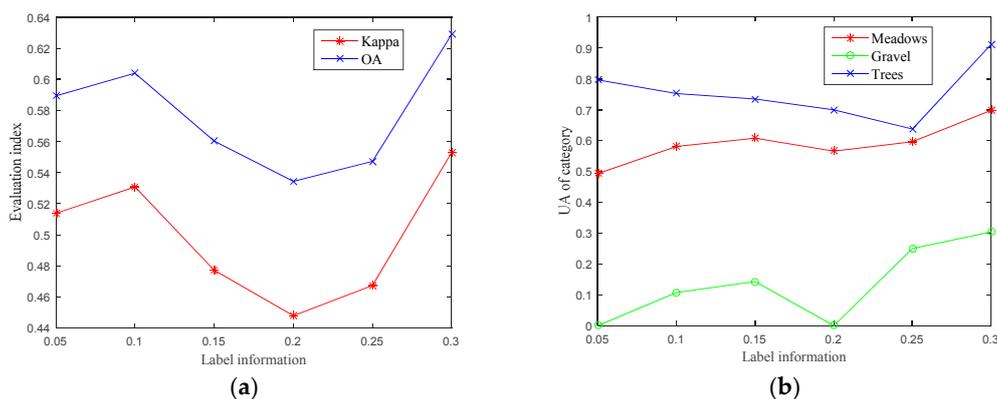
Evaluation		UAs (%)						
Class	Method	SSC	SSC-S	S ⁴ C	S ³ R	SSLRR	SSRLRR	CPPSSC
	Asphalt	30.09	64.76	0	0	0	0	57.59
	Meadows	40.02	62.91	61.74	68.97	53.81	69.56	69.77
	Gravel	0	1.79	0	30.36	30.36	0	30.36
	Trees	89.82	67.70	68.14	30.09	30.09	36.73	91.15
	Painted metal sheets	48.02	36.20	27.90	69.86	69.56	69.86	58.71
	Bare soil	93.18	31.82	63.64	34.47	29.92	64.77	92.05
	Bitumen	0	32.06	98.95	48.78	29.97	29.97	29.97
	Self-blocking bricks	64.14	39.79	61.24	67.05	69.72	65.29	62.75
	Shadows	92.95	67.22	99.17	1.66	0.41	0.41	29.88

Table 2. The overall accuracy analysis on the Pavia University scene. OA: overall accuracy.

Algorithm	SSC	SSC-S	S ⁴ C	S ³ R	SSLRR	SSRLRR	CPPSSC
Oas(%)	51.37	48.35	54.87	58.14	53.15	58.19	62.91
Kappas	0.4353	0.4037	0.4625	0.4707	0.3972	0.4826	0.5530

We also conduct experiment on the Pavia University scene by our CPPSSC algorithm by changing supervised information by 5%, 10%, 15%, 15%, 20% and 25%, compared with default 30% supervised information. The variation tendencies of OAs and kappas are shown in Figure 5.

To our best knowledge, the clustering effect will be better when the supervised information accounts for a large proportion. It can be seen clearly from Figure 5 that OA and kappa with 30% supervised information are the best compared with the other different levels of supervised information. However, the clustering results still keep fairly stable when the supervised information accounts for a proportion of 10%, which shows the low dependence of the supervised information. They also depends on other parameters such as λ . The case demonstrates that our algorithm does not rely too much on supervised information. Besides, for the meadows, gravel and trees, the classification precision also can reach better clustering validity with UAs when the known class probability information can be integrated into the traditional SSC algorithm.

**Figure 5.** (a) The variation tendency of the OAs, kappa; (b) The aforementioned three classes included the meadows, gravel and trees. They are shown with the gradual growth of supervised information.

4.2.2. The Block Diagonal Structure of Sparse Coefficients

We also conduct the experiments to confirm the block diagonal structure of sparse representation coefficient with our algorithm, and the results are listed in Figure 6.

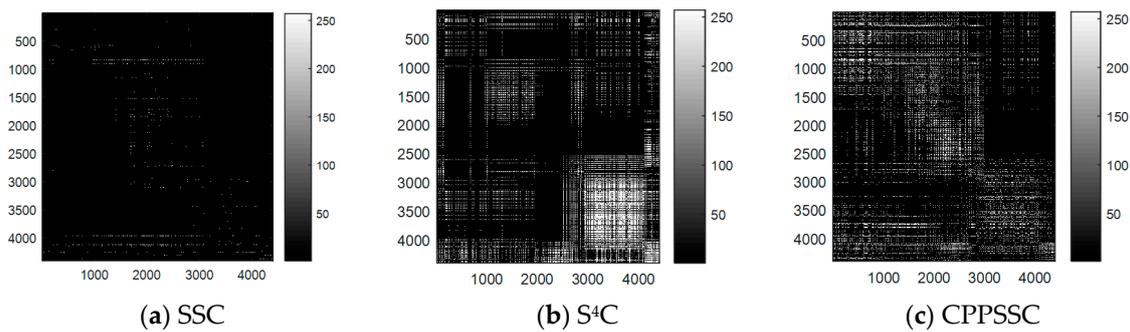


Figure 6. Visualization of the block diagonal structure of sparse coefficient with three algorithms on the Pavia University scene. For clarity, we consider the nonzero entries of sparse coefficients as number 255. Consequently, the sparse coefficient matrix is a binary matrix.

From Figure 6, we can see that the block diagonal structure of sparse coefficient with our CPPSSC algorithm is obviously better than with SSC and S^4C , which is in favor of self-expressiveness to boost the final clustering results. As illustrated in Figure 6, the white spaces indicating nonzero coefficients are the block sparse coefficients among data samples. In Figure 6a, it is difficult to form block diagonalization facing HSI data with the samples with nonzero coefficients. Although Figure 6b can show block diagonalization to a certain extent, an imperfection is that the nonzero sparse coefficients occupy a large proportion in the overall sparse coefficient matrix, which is contrary to sparse representation theory. In terms of Figure 6c, the block diagonalizations via our CPPSSC method are quite obvious compared with the other two methods; the reason is that the probabilistic class structure estimated the similarity between each sample and each class is incorporated into the sparse representation coefficients. Moreover, the global nonzero coefficients structure can be enhanced, which facilitate block diagonalization of sparse coefficients.

4.2.3. The Quantitative Experimental Results on the Indian Pines

Then, we conduct our CPPSSC algorithm with supervised information of 30% on the Indian Pines data set, and the experimental effect is shown in Figure 7.

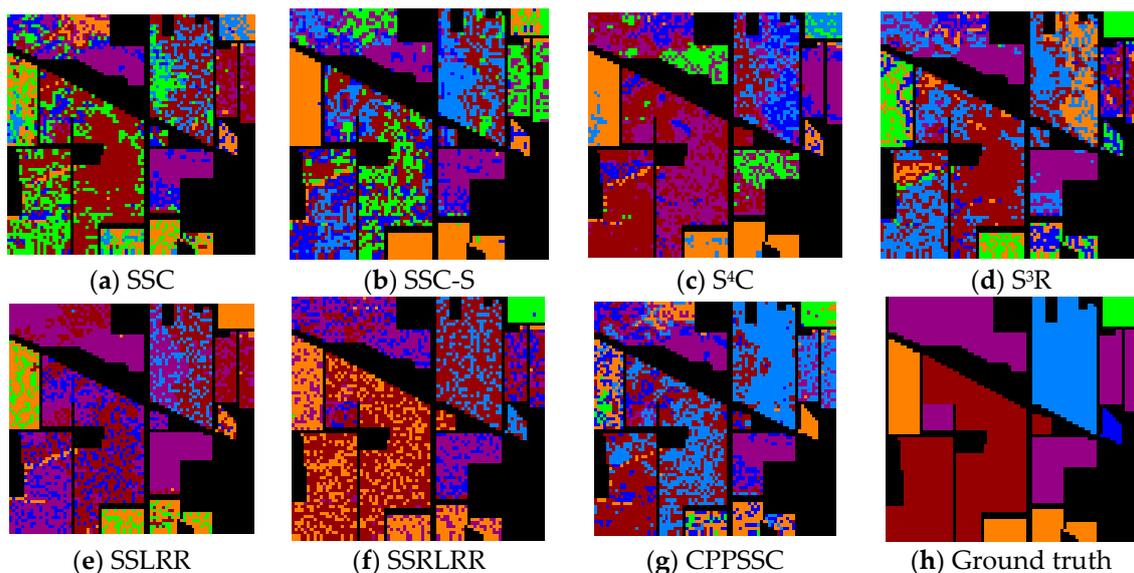


Figure 7. Cluster maps of the experimental results on Indian Pines.

Figure 7 shows the visual cluster maps with all kinds of clustering technologies. We can see that the visual cluster effect of the soybean-notill and woods with our CPPSSC is closer to that of the original cluster map. The quantitative data analysis is given in Tables 3 and 4.

Table 3. The quantitative analysis on Indian Pines.

Evaluation		UAs (%)						
Class	Method	SSC	SSC-S	S ⁴ C	S ³ R	SSLRR	SSRLRR	CPPSSC
	Alfalfa	30.43	30.43	30.43	45.65	10.87	0	0
	Corn-notill	48.96	51.35	47.13	54.06	70.62	47.37	49.20
	Grass-trees	45.45	98.84	91.68	29.98	65.57	69.63	47.58
	Soybean-notill	30.05	49.86	31.01	43.31	30.05	30.05	81.01
	Soybean-mintill	62.50	39.96	59.79	60.53	35.59	67.24	57.76
	Woods	0	44.44	31.31	100	0	98.99	90.91

Table 4. The overall accuracy analysis on Indian Pines.

Algorithm	SSC	SSC-S	S ⁴ C	S ³ R	SSLRR	SSRLRR	CPPSSC
OAs(%)	49.11	52.13	54.02	52.74	47.47	55.33	58.14
Kappas	0.3464	0.4154	0.3962	39.31	0.2996	0.3979	0.4643

Table 3 shows the UA of every land over class, which can distinguish the clustering performance of every method. Evidently, the clustering expression of soybean-notill and woods is able to present better clustering performance via our CPPSSC, with higher UAs of 81.01% and 90.91%, respectively, reducing the misclassification. Moreover, in terms of soybean-notill, the other benchmark techniques achieve a poor UA of 30%, which is less than half that of our method. In the cluster map, the majority of soybean-notill class has been misclassified into grass-trees in SSC. At the same time, the UA of soybean-notill with the SSC-S method achieves better performance than that of SSC algorithm, with a growth of 19.81%, because of adding the spatial information into HSI clustering. However, the speed of the growth is limited. For our CPPSSC algorithm, the sparse representation can take full advantage of the known information to extract the internal essence on HSI data, which achieves the qualitative upgrade of the soybean-notill clustering. The cluster precision of the woods via our algorithm is higher, up to 90.91%. Compared with S⁴C, it perfectly achieves the best clustering results for the woods class, with an improvement of almost 60% in UA. Actually, the reason for improvements in our algorithm is that signals with high correlation are preferentially selected in the sparse representation process via the spread of supervised information. The UAs of the woods in the SSC-S and S⁴C, are 44.44% and 31.31%, respectively, which are far lower results than ours. The UA of woods in CPPSSC is lower than in S³R and SSRLRR, which indirectly proves significance of supervised information for clustering.

In Table 4, with overall precision analysis on Indian Pines, the OA and kappa values by our algorithm are the best results (presented with bold), with an OA of 58.14% and a kappa of 0.4643. The SSC-S and S⁴C can obtain better cluster performance with OA precision, which is 52.13% and 54.02%. It is a fact that the two algorithms can acquire smooth growth compared with the SSC algorithm because of adding the wealthy spatial correlation of HSIs, obtaining improvements of 3.02% and 4.91% in OA, respectively. However, the promotion of the two methods has some limitations. This is because they do not utilize the known supervised information but only utilize the unknown samples to fetch information. Fortunately, our CPPSSC algorithm can rationally make full use of supervised information to spread to unknown data. Hence, our method can achieve preferable clustering precision and is more effective compared with traditional SSC algorithm, achieving an improvement of 9.03% in OA and a growth of 0.1179 in kappa. In terms of S⁴C, our CPPSSC algorithm can obtain evident OA growth with 4.12%, and also obtain the kappa growth of 0.0741 by the effective class probability model. With

respect to S^3R , SSLRR and SSRLRR algorithms, our CPPSSC performs better than these three methods. The reason is that our algorithm utilizes a small amount of labeled samples to generate class probability among samples, and exploits class structure information via sparse representation classification.

For our CPPSSC algorithm, we also carry out a series of experiments on Indian Pines with added supervised information of 5%, 10%, 15%, 20%, 25% and 30%. The experimental results are shown in Figure 8.

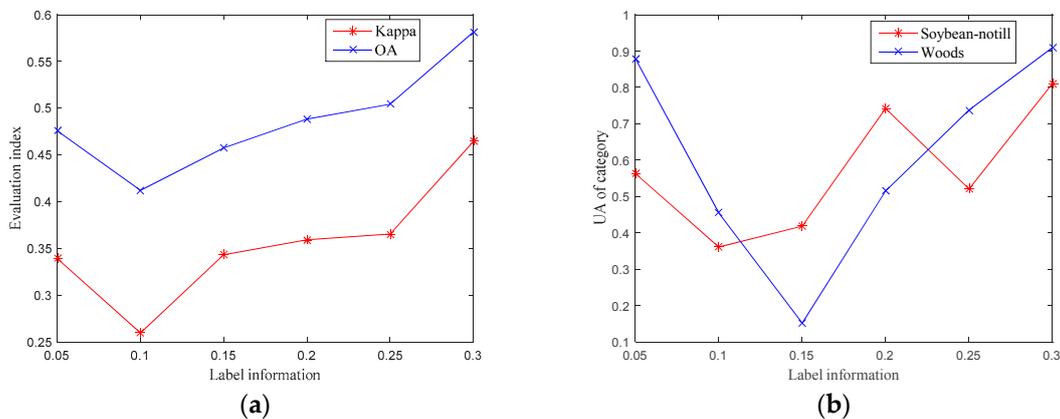


Figure 8. (a) The variation tendency of the OA, kappa; (b) The aforementioned precision of soybean-notill and woods with the gradual growth of supervised information.

In Figure 8, the clustering precision of OA and kappa firstly shows a descending trend since the supervised information is not utilized by sparse representation process at the beginning, and then ascends because of gradually added supervised information, and it propagates to the unknown samples. In other words, we only take full advantage of the testing samples via the spare process and the reduction in the available data. Hence, the variation tendency is shown to be descending in the beginning. With the increase in supervised information, quality information can be propagated to unknown samples via class probability, and then the overall clustering accuracy will be improved. The clustering precision of Indian Pines can be best reached using 30% supervised information, with 58.14% in OA and 0.4643 in kappa. Likewise, the clustering accuracy with UA of the soybean-notill and woods can be closer to ground truth when we add the supervised information with 30% into unknown HSI sample clustering.

4.2.4. The Clustering Performance Evaluated by AC and NMI on Two Data Sets

In general, the accuracy (AC) and normalized information metric (NMI) [43] are used to evaluate the performance of clustering method. Consequently, we have conducted experiments on the Pavia University scene and Indian Pines to verify the effectiveness of CPPSSC. The performance of the benchmarks and CPPSSC methods are listed in Table 5.

Table 5. The accuracy (AC) and normalized information metric (NMI) of clustering performance on two datasets.

Data Set	Algorithm							
	Evaluation	SSC	SSC-S	S ⁴ C	S ³ R	SSLRR	SSRLRR	CPPSSC
Pavia University scene	AC (%)	88.89%	86.25%	87.28%	86.12%	89.53%	87.62%	91.89%
	NMI (%)	77.74%	74.41%	76.94%	75.37%	79.09%	75.16%	84.32%
Indian Pines	AC (%)	58.46%	56.99%	63.76%	59.19%	58.76%	58.89%	68.23%
	NMI (%)	50.76%	51.02%	53.23%	52.09%	38.74%	38.82%	55.68%

To our knowledge, both the AC and NMI range from 0 to 1 and a higher value indicates a better result. From Table 5, we can see that CPPSSC achieves an 3% AC gain and 6.58% NMI gain on Pavia University scene over the SSC. Besides, it is better than the other semi-supervised clustering methods. Moreover, the CPPSSC also achieves the best clustering performance over the other benchmarks on Indian Pines. The reason might be the effectiveness of the proposed class probability, which explores the relationships between the samples and class and further adds valid similarity information to SSC.

4.2.5. The Parameters Analysis in the CPPSSC Algorithm on Two Data Sets

There are two main parameters in the CPPSSC algorithm, they are λ and γ . λ is the tradeoff parameter between the sparsity of the coefficient and the magnitude of noise. It can be decided by the following formula.

$$\lambda = \gamma / \mu \quad (18)$$

$$\mu \triangleq \min_i \max_{j \neq i} |y_i^T y_j| \quad (19)$$

The λ is actually decided by γ since μ is fixed for a certain data set. Indeed, we only need to fine-tune γ , and find the optimum values with OA and Kappa, which can be shown with a curve in Figure 9.

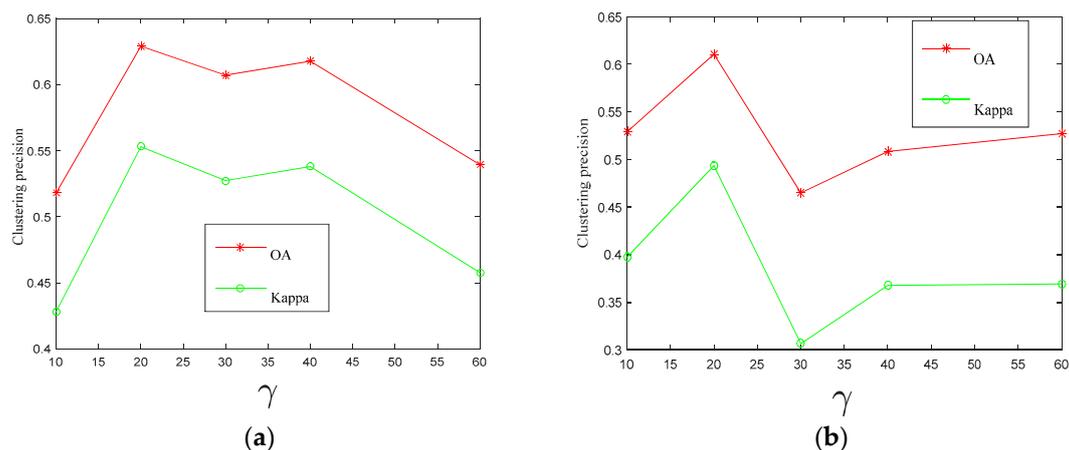


Figure 9. The variation curves of OA and kappa via fine-tuned γ with (a) Pavia University scene; (b) Indian Pines data set.

The clustering precision change curves of the OA and kappa with various values of γ are shown in Figure 9. From this figure, it can be seen that the γ is independent on the dataset to some extent. The optimizational values of the two data sets are located on $\gamma = 20$, which are perfectly shown on the two datasets. The OA and kappa values are respectively 0.6291 and 0.5530 on the Pavia University scene when $\gamma = 20$. In addition, the optimizational values of the OA and kappa are also achieved as 0.6107 and 0.4936 when $\gamma = 20$ on the Indian Pines data set. It can be seen that our CPPSSC can achieve a better clustering accuracy with $\gamma = 20$, and makes sense for HSI clustering.

To show the computation complexity of these clustering methods, we also perform the experiments on two data sets. The computational time of different clustering methods (seconds) is shown in Table 6.

From Table 6, we can see that SSC is fast in the Pavia University scene and Indian Pines, and CPPSSC spends the more time than SSC and SSC-S but less than the other clustering methods. The main reason is that CPPSSC has to spend a little time to estimate the class probability distribution of unlabeled samples.

Table 6. The computational time of different clustering methods (seconds).

	SSC	SSC-S	S ⁴ C	S ³ R	SSLRR	SSRLRR	CPPSSC
Pavia University scene	436.80	577.85	2493.73	268.32	3677.74	2140.79	751.93
Indian Pines	161.87	431.74	687.20	387.16	1469.02	2103.37	562.09

4.3. Discussion

From the experimental results, we can see that, compared with the unsupervised and semi-supervised methods, our algorithm is informative and discriminative. The reason is that firstly, SRC can obtain a good estimation of the underlying class structure of test samples by utilizing a small amount of labeled samples. This is named probabilistic class structure. Then, the class probability is incorporated into the sparse representation coefficient to strength the global similarity structure among all the samples and preserve the subspace-sparse representation by facilitating the block diagonalization of sparse coefficients.

The computational complexity of the CPPSSC algorithm depends on updating C^k and P^k in Algorithm 3. Specifically, the computation complexity of it is about $O(n^2)$, where n is the number of data samples. The computation complexity of updating Z^k , α^k , and β^k is $O(n)$. In summary, the computational complexity of CPPSSC algorithm is $O(\tau n^2)$, where τ is the number of iterations.

To obtain the uniform class probability between each unlabeled sample and each specific class addressed by SRC, we prefer to use the l_1 norm instead of the l_2 norm to deal with $\|P\|$, although the l_1 norm has been proven unimportant to classification in theory [44] and practice [45]. The reasons can be summarized as follows. First, in SRC, Wright et al. verified that the SRC coefficients solved by l_2 minimization are much less sparse than by l_1 minimization. We hope to obtain the much sparser representation of self-expressiveness of data. Second, since both the SRC and SSC are based on the l_1 norm, this leads to a combined norm that also has the structure of the l_1 norm. It will facilitate the block diagonal structure of sparse coefficients. The third reason to use the l_1 norm is the greatly theoretical guarantees for correctness of SSC, which can be applicable to detect subspace even when subspaces are overlapping [46].

A number of supervised classification methods have been suggested. Compared with supervised classification methods, the advantage of semi-supervised methods for HSIs can be summarized as follows. First, the supervised classification needs a great deal of labeled samples to improve the classifier performance [47]. However HSI classification often faces the issue of limited number of labeled data, which are often costly, effortful, and time-consuming [28]. On the other hand, we can obtain a large number of unlabeled data effortlessly. Semi-supervised learning (SSL), which can utilize both small amount of labeled instances and abundant as well as unlabeled samples, has been proposed to deal with this issue [48]. Second, in essence, semi-supervised clustering such as using the CPPSSC method adds the constraints of a small amount of labeled information to the objective function for assigning similar samples into corresponding class. These constraints are used to estimate the similarity between data points and thereby enhance the clustering performance [49]. Consequently, the semi-supervised clustering algorithms are becoming more popular because of the abundance of unlabeled data and the high cost of obtaining labeled data.

What should be denoted is that, to be honest, we cannot guarantee that all the UAs of corresponding classes via the CPPSSC algorithm are the best because of the different data structures of each class. The main reason is that CPPSSC algorithm has difficulty in estimate class probability of every data point because of the redundant information of complex HSI bands.

5. Conclusions

In this paper, we have briefly reviewed the classical unsupervised SSC algorithm to HSI clusters by considering every land over class as a subspace. Owing to the hard cluster effect of directly using unsupervised SSC, we proposed a novel class probability propagation via a supervised information

algorithm called CPPSSC for HSI clustering, which mainly takes full advantage of rationally known supervised information. In terms of the SSC-S, S^4C algorithms, they only concentrate on the spectral and spatial correlation of HSIs and neglect to add the known information. It is hard to extract the valid sparse representation and guarantee the sparse coefficient block diagonalization structure properly. Our CPPSSC algorithm can make progress on this point by mixing with class probability. Compared with S^3R , SSLRR and SSRLRR algorithms, the CPPSSC can capture global similarity structure and work well over the state-of-the-art methods.

The presented CPPSSC algorithm still has room for improvement. For instance, because of the redundant information of complex HSI bands, we will develop a structured class probability model, which may better present the correlation between each sample and each class with redundant information of complex HSI bands and fully explore the structure relationships among samples in the future work.

Acknowledgments: This work is supported by the National Science Foundation for China (Nos. 61602002 & 61572372), the An Hui University Youth Skeleton Teacher Project (E12333010289), Anhui University Doctoral Scientific Research Start-up Funding (J10113190084), LIESMARS Special Research Funding, the Anhui Provincial Natural Science Foundation (grant number 1608085MF136) and the China Postdoctoral Science Foundation (2015M582826).

Author Contributions: Qing Yan, Yun Ding and Yi Xia conceived the research and conducted the simulations; Yanwen Chong designed and implemented the algorithm; Chunhou Zheng analyzed the data, results and verified the theory; all authors participated in the writing of the manuscript.

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

References

- Landgrebe, D. Hyperspectral image data analysis. *IEEE Signal Process. Mag.* **2002**, *19*, 17–28. [[CrossRef](#)]
- Silva-Filho, A.G.; Frery, A.C.; de Araujo, C.C.; Alice, H.; Cerqueira, J.; Loureiro, J.A.; de Lima, M.E.; Oliveira, M.G.S.; Horta, M.M. Hyperspectral images clustering on reconfigurable hardware using the k-means algorithm. In Proceedings of the 16th Symposium on Integrated Circuits and Systems Design, 2003 (SBCCI 2003), Sao Paulo, Brazil, 8–11 September 2003; pp. 99–104.
- Ghasrodashti, E.K.; Karami, A.; Heylen, R.; Scheunders, P. Spatial resolution enhancement of hyperspectral images using spectral unmixing and bayesian sparse representation. *Remote Sens.* **2017**, *9*, 541. [[CrossRef](#)]
- Sun, Y.; Wang, S.; Liu, Q.; Hang, R.; Liu, G. Hypergraph embedding for spatial-spectral joint feature extraction in hyperspectral images. *Remote Sens.* **2017**, *9*, 506.
- Huang, H.; Luo, F.; Liu, J.; Yang, Y. Dimensionality reduction of hyperspectral images based on sparse discriminant manifold embedding. *ISPRS J. Photogramm. Remote Sens.* **2015**, *106*, 42–54. [[CrossRef](#)]
- He, Z.; Liu, L.; Zhou, S.; Shen, Y. Learning group-based sparse and low-rank representation for hyperspectral image classification. *Pattern Recognit.* **2016**, *60*, 1041–1056. [[CrossRef](#)]
- Xue, Z.; Du, P.; Li, J.; Su, H. Sparse graph regularization for hyperspectral remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 2351–2366. [[CrossRef](#)]
- He, L.; Li, J.; Plaza, A.; Li, Y. Discriminative low-rank gabor filtering for spectral-spatial hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2016**, *55*, 1381–1395. [[CrossRef](#)]
- Zhang, H.; Li, J.; Huang, Y.; Zhang, L. A nonlocal weighted joint sparse representation classification method for hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 2056–2065. [[CrossRef](#)]
- Lunga, D.; Prasad, S.; Crawford, M.M.; Ersoy, O. Manifold-learning-based feature extraction for classification of hyperspectral data: A review of advances in manifold learning. *IEEE Signal Process. Mag.* **2014**, *31*, 55–66. [[CrossRef](#)]
- Gao, Y.; Ji, R.; Cui, P.; Dai, Q.; Hua, G. Hyperspectral image classification through bilayer graph-based learning. *IEEE Trans. Image Process.* **2014**, *23*, 2769–2778. [[CrossRef](#)] [[PubMed](#)]
- Zhang, E.; Zhang, X.; Jiao, L.; Li, L.; Hou, B. Spectral-spatial hyperspectral image ensemble classification via joint sparse representation. *Pattern Recognit.* **2016**, *59*, 42–54. [[CrossRef](#)]
- Li, C.; Tan, Y.; Wang, D.; Ma, P. Research on 3D face recognition method in cloud environment based on semi supervised clustering algorithm. *Multimed. Tools Appl.* **2017**, *76*, 17055–17073. [[CrossRef](#)]

14. Chen, B.; Yang, J.; Jeon, B.; Zhang, X. Kernel quaternion principal component analysis and its application in RGB-D object recognition. *Neurocomputing* **2017**, *226*, 293–303. [[CrossRef](#)]
15. Ahn, I.; Kim, C. Face and hair region labeling using semi-supervised spectral clustering-based multiple segmentations. *IEEE Trans. Multimed.* **2016**, *18*, 1414–1421. [[CrossRef](#)]
16. Ma, T.; Wang, Y.; Tang, M.; Cao, J.; Tian, Y.; Al-Dhelaan, A.; Al-Rodhaan, M. LED: A fast overlapping communities detection algorithm based on structural clustering. *Neurocomputing* **2016**, *207*, 488–500. [[CrossRef](#)]
17. Hong, R.; Wang, M.; Gao, Y.; Tao, D.; Li, X.; Wu, X. Image annotation by multiple-instance learning with discriminative feature mapping and selection. *IEEE Trans. Cybern.* **2014**, *44*, 669–680. [[CrossRef](#)] [[PubMed](#)]
18. Vidal, R.; Ma, Y.; Sastry, S.S. Generalized principal component analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *27*, 1945. [[CrossRef](#)] [[PubMed](#)]
19. Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; Ma, Y. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 171–184. [[CrossRef](#)] [[PubMed](#)]
20. Zhang, H.; Lin, Z.; Zhang, C.; Gao, J. Robust latent low rank representation for subspace clustering. *Neurocomputing* **2014**, *145*, 369–373. [[CrossRef](#)]
21. Peng, X.; Yu, Z.; Yi, Z.; Tang, H. Constructing the L2-graph for robust subspace learning and subspace clustering. *IEEE Trans. Cybern.* **2017**, *47*, 1053–1066. [[CrossRef](#)] [[PubMed](#)]
22. Yuan, Y.; Lin, J.; Wang, Q. Dual-clustering-based hyperspectral band selection by contextual analysis. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 1431–1445. [[CrossRef](#)]
23. Elhamifar, E.; Vidal, R. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2765–2781. [[CrossRef](#)] [[PubMed](#)]
24. Zhang, H.; Zhai, H.; Zhang, L.; Li, P. Spectral-spatial sparse subspace clustering for hyperspectral remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3672–3684. [[CrossRef](#)]
25. Cheng, G.; Zhu, F.; Xiang, S.; Wang, Y.; Pan, C. Semisupervised hyperspectral image classification via discriminant analysis and robust regression. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 595–608. [[CrossRef](#)]
26. Gu, B.; Sheng, V.S.; Sheng, S. A robust regularization path algorithm for v-support vector classification. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, *28*, 1241. [[CrossRef](#)] [[PubMed](#)]
27. Zhuang, L.; Zhou, Z.; Gao, S.; Yin, J.; Lin, Z.; Ma, Y. Label information guided graph construction for semi-supervised learning. *IEEE Trans. Image Process. Publ. IEEE Signal Process. Soc.* **2017**, *26*, 4182–4192. [[CrossRef](#)] [[PubMed](#)]
28. Shao, Y.; Sang, N.; Gao, C.; Ma, L. Probabilistic class structure regularized sparse representation graph for semi-supervised hyperspectral image classification. *Pattern Recognit.* **2017**, *63*, 102–114. [[CrossRef](#)]
29. Fang, X.; Xu, Y.; Li, X.; Lai, Z.; Wong, W.K. Robust semi-supervised subspace clustering via non-negative low-rank representation. *IEEE Trans. Cybern.* **2016**, *46*, 1828–1838. [[CrossRef](#)] [[PubMed](#)]
30. Jain, A.; Jin, R.; Chitta, R. Semi-supervised clustering. In *Handbook of Cluster Analysis*; CSC Press: Boca Raton, FL, USA, 2015; pp. 1–35.
31. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306. [[CrossRef](#)]
32. Yan, S.; Wang, H. Semi-supervised learning by sparse representation. In Proceedings of the 2009 SIAM International Conference on Data Mining, Sparks, NV, USA, 30 April–2 May 2009; pp. 792–801.
33. Yang, S.; Wang, X.; Wang, M.; Han, Y.; Jiao, L. Semi-supervised low-rank representation graph for pattern recognition. *IET Image Process.* **2013**, *7*, 131–136. [[CrossRef](#)]
34. Zhu, F.; Fan, B.; Zhu, X.; Wang, Y.; Xiang, S.; Pan, C. 10,000+ times accelerated robust subset selection (ARSS). In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, Austin, TX, USA, 25–29 January 2015; pp. 3217–3223.
35. Nie, F.; Wang, H.; Huang, H.; Ding, C.H. Early active learning via robust representation and structured sparsity. In Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI), Beijing, China, 3–9 August 2013; pp. 1572–1578.
36. Boyd, S.; Parikh, N.; Chu, E.; Peleato, B.; Eckstein, J. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.* **2011**, *3*, 1–122. [[CrossRef](#)]
37. Wang, W.W.; Li, X.; Feng, X.C.; Wang, S.Q. A survey on sparse subspace clustering. *Acta Autom. Sin.* **2015**, *41*, 1373–1384.

38. Zhang, H.; He, W.; Zhang, L.; Shen, H.; Yuan, Q. Hyperspectral image restoration using low-rank matrix recovery. *IEEE Trans. Geosci. Remote Sens.* **2014**, *52*, 4729–4743. [[CrossRef](#)]
39. Wright, J.; Yang, A.Y.; Ganesh, A.; Sastry, S.S.; Ma, Y. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *31*, 210–227. [[CrossRef](#)] [[PubMed](#)]
40. Mota, J.F.C.; Xavier, J.M.F.; Aguiar, P.M.Q.; Puschel, M. D-ADMM: A communication-efficient distributed algorithm for separable optimization. *IEEE Trans. Signal Process.* **2013**, *61*, 2718–2723. [[CrossRef](#)]
41. Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888–905.
42. Zhai, H.; Zhang, H.; Zhang, L.; Li, P.; Plaza, A. A new sparse subspace clustering algorithm for hyperspectral remote sensing imagery. *IEEE Geosci. Remote Sens. Lett.* **2016**, *14*, 43–47. [[CrossRef](#)]
43. Cai, D.; Chen, X. Large scale spectral clustering via landmark-based sparse representation. *IEEE Trans. Cybern.* **2014**, *45*, 1669–1680. [[PubMed](#)]
44. Peng, X.; Lu, C.; Yi, Z.; Tang, H. Connections between nuclear-norm and frobenius-norm-based representations. *IEEE Trans. Neural Netw. Learn. Syst.* **2016**, 1–7. [[CrossRef](#)] [[PubMed](#)]
45. Peng, X.; Lu, J.; Yi, Z.; Yan, R. Automatic subspace learning via principal coefficients embedding. *IEEE Trans. Cybern.* **2014**, *PP*, 1–14. [[CrossRef](#)] [[PubMed](#)]
46. Soltanolkotabi, M.; Candés, E.J. A geometric analysis of subspace clustering with outliers. *Ann. Stat.* **2011**, *40*, 2012. [[CrossRef](#)]
47. Tan, K.; Zhu, J.; Du, Q.; Wu, L.; Du, P. A novel tri-training technique for semi-supervised classification of hyperspectral images based on diversity measurement. *Remote Sens.* **2016**, *8*, 749. [[CrossRef](#)]
48. Zhu, X.; Ghahramani, Z.; Lafferty, J. Semi-supervised learning using gaussian fields and harmonic functions. In Proceedings of the Twentieth International Conference on International Conference on Machine Learning, Washington, DC, USA, 21–24 August 2003; pp. 912–919.
49. Wolf, M. Adaboost on low-rank psd matrices for metric learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20–25 June 2011; pp. 2617–2624.



© 2017 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/>).