

Article

Kernel Sparse Subspace Clustering with a Spatial Max Pooling Operation for Hyperspectral Remote Sensing Data Interpretation

Han Zhai ^{1,*}, Hongyan Zhang ^{1,*}, Xiong Xu ^{2,*}, Liangpei Zhang ¹ and Pingxiang Li ¹

¹ The State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, and the Collaborative Innovation Center for Geospatial Technology, Wuhan University, Wuhan 430079, China; zhaihan@whu.edu.cn (H.Z.); zlp62@whu.edu.cn (L.Z.); pxLi@whu.edu.cn (P.L.)

² College of Surveying and Geoinformatics, Tongji University, Shanghai 200000, China

* Correspondence: zhanghongyan@whu.edu.cn (H.Z.); xvxiang@tongji.edu.cn (X.X.); Tel.: +86-27-6877-8395 (H.Z.); +86-21-6598-8851 (X.X.)

Academic Editors: Gonzalo Pajares Martinsanz, Magaly Koch and Prasad S. Thenkabail
Received: 24 January 2017; Accepted: 28 March 2017; Published: 1 April 2017

Abstract: Hyperspectral image (HSI) clustering is generally a challenging task because of the complex spectral-spatial structure. Based on the assumption that all the pixels are sampled from the union of subspaces, recent works have introduced a robust technique—the sparse subspace clustering (SSC) algorithm and its enhanced versions (SSC models incorporating spatial information)—to cluster HSIs, achieving excellent performances. However, these methods are all based on the linear representation model, which conflicts with the well-known nonlinear structure of HSIs and limits their performance to a large degree. In this paper, to overcome these obstacles, we present a new kernel sparse subspace clustering algorithm with a spatial max pooling operation (KSSC-SMP) for hyperspectral remote sensing data interpretation. The proposed approach maps the feature points into a much higher dimensional kernel space to extend the linear sparse subspace clustering model to nonlinear manifolds, which can better fit the complex nonlinear structure of HSIs. With the help of the kernel sparse representation, a more accurate representation coefficient matrix can be obtained. A spatial max pooling operation is utilized for the representation coefficients to generate more discriminant features by integrating the spatial-contextual information, which is essential for the accurate modeling of HSIs. This paper is an extension of our previous conference paper, and a number of enhancements are put forward. The proposed algorithm was evaluated on two well-known hyperspectral data sets—the Salinas image and the University of Pavia image—and the experimental results clearly indicate that the newly developed KSSC-SMP algorithm can obtain very competitive clustering results for HSIs, outperforming the current state-of-the-art clustering methods.

Keywords: hyperspectral images; subspace clustering; nonlinear techniques; kernels; spatial max pooling

1. Introduction

Hyperspectral sensors acquire nearly continuous spectral bands with hundreds of spectral channels to capture the diagnostic information of land-cover materials, opening new possibilities for remote sensing applications such as mineral exploration, precision agriculture, and disaster monitoring [1–3]. As an unsupervised information extraction technique, clustering is a very useful tool for hyperspectral image (HSI) interpretation. When labeled samples are unavailable or difficult to acquire, clustering can be an effective alternative [4]. Clustering can be defined as a partition process, which involves grouping similar pixels while separating dissimilar ones. However, due to the complex nonlinear structure and large spectral variability, clustering HSIs is a very challenging task.

The traditional clustering methods that are commonly applied to HSIs, such as k -means [5] and fuzzy c -means (FCM) [6], attempt to segment pixels using only the spectral information. However, because of their limited discriminative capability and weak adaptation to the nonlinear structure of HSIs, such methods often fail to achieve satisfactory results. In recent years, to address this issue, a number of spectral-spatial clustering methods have been developed, such as S - k -means [7] and FCM-S1 [8], which have been shown to be able to improve the performance of unsupervised clustering. In addition, supervised spectral-spatial segmentation methods, such as Markov random field (MRF)-based methods [9] and graph-based methods [10], have also been widely used in remote sensing data interpretation.

Subspace-based and sparse representation based approaches have recently drawn wide attention in the context of high-dimensional data analysis, due to their excellent performance. By modeling the data via low-dimensional manifolds, the inherent structure of the data can be better explored [11,12]. As a perfect combination of these two theories, the sparse subspace clustering (SSC) algorithm [13,14] has achieved great success in the computer vision field. By treating pixels from the same land-cover class that have different spectra as belonging to the same subspace, the SSC algorithm has also shown great potential in HSI clustering [15–17], which is mainly due to its capacity to address the spectral variability problem. However, SSC interprets the HSIs only with the spectral information, and it ignores the rich spatial information contained in the images. As a result, the obtained performance is limited. In order to account for the spatial information contained in HSIs, SSC models incorporating spatial information have become very popular. In a recent study [15], the spectral-spatial sparse subspace clustering (S^4C) algorithm was proposed to improve clustering performance by the use of a spectral weighting strategy to relieve the high correlation problem between hyperspectral pixels and a local averaging constraint to incorporate the spatial information contained in the eight-connected neighborhood. As an enhanced approach incorporating spatial information, the ℓ_2 -norm regularized sparse subspace clustering (L2-SSC) algorithm [16] was developed to exploit the inherent spectral-spatial properties of HSIs in a more elaborate way, and can achieve a good effect.

One important consideration is that the aforementioned linear methods often have difficulty in coping with the inherently nonlinear structure of HSIs. In recent years, a number of methods have been proposed to deal with the linear inseparability problem in HSI data interpretation. Among these methods, kernel-based algorithms are one of the most commonly used methods. These approaches map the HSI from the original feature space to a much higher dimensional kernel feature space, in which the classes are assumed to be linearly separable. For example, Mercier and Lennon [18] developed support vector machines for hyperspectral image classification with spectral-based kernels. Chen et al. [19] explored kernel sparse representation for HSI classification. However, to the best of our knowledge, there are very few clustering methods that can deal with the nonlinearity problem. Very recently, Morsier et al. [20] proposed a kernel-based low-rank and sparse representation algorithm to explore the nonlinearity of HSIs. However, this approach only utilizes the spectral information, without considering any spatial-contextual information.

In this paper, to address these issues, we introduce a novel kernel sparse subspace clustering algorithm with a spatial max pooling operation (KSSC-SMP) for hyperspectral remote sensing data interpretation, which is an extension of our previous conference work [21]. The proposed method simultaneously explores the nonlinear structure and the inherent spectral-spatial attributes of HSIs. On the one hand, with the kernel strategy, we map the feature points from the original feature space to a much higher dimensional space to ensure that the linearly inseparable points are more separable. On the other hand, in order to fully exploit the discriminative spectral-spatial information of HSIs and the potential of the SSC model, the spatial max pooling operation is introduced to merge the obtained representation coefficients into new features which incorporate the spatial-contextual information, thereby improving the clustering performance and guaranteeing the spatial homogeneity of the final clustering result.

The rest of this paper is organized as follows. Section 2 reviews the classical SSC model. Section 3 introduces the proposed KSSC-SMP algorithm in detail. Section 4 describes the experimental results. Section 5 analyzes the experimental results. Section 6 draws the conclusions and points out the future research lines.

2. Sparse Subspace Clustering (SSC)

In this section, we briefly review the SSC algorithm. By using the HSI dataset itself as the representation dictionary, the attributes of each data point can be comprehensively exploited within the SSC framework [13]. For an HSI with a size of $M \times N \times p$, all the pixels can be considered as being selected from a union of l affine subspaces $S_1 \cup S_2 \cup \dots \cup S_l$ of dimensions $\{d_i\}_{i=1}^l$ in the full space \mathbb{R}^p with $d_1 + d_2 + \dots + d_l = p$, where M denotes the height of the image, N represents the width of the image, and p stands for the number of spectral channels. By treating each pixel as a column vector, the HSI cube can be transformed into a 2-D matrix $\mathbf{Y} = [y_1, y_2, \dots, y_{MN}] \in \mathbb{R}^{p \times MN}$. Then, with this 2-D matrix being utilized as the representation dictionary, the sparse optimization problem can be modeled as follows:

$$\min_{\mathbf{C}} \|\mathbf{C}\|_0 \text{ s.t. } \mathbf{Y} = \mathbf{Y}\mathbf{C} + \mathbf{N}, \text{diag}(\mathbf{C}) = \mathbf{0}, \mathbf{C}^T \mathbf{1} = \mathbf{1} \quad (1)$$

where $\mathbf{C} \in \mathbb{R}^{MN \times MN}$ denotes the representation coefficient matrix, $\mathbf{N} \in \mathbb{R}^{p \times MN}$ stands for the representation error matrix, and $\mathbf{1} \in \mathbb{R}^{MN}$ is a vector whose elements are all ones. The constraint $\text{diag}(\mathbf{C}) = \mathbf{0}$ is used to eliminate the trivial solution of each pixel being represented as a linear combination of itself [13]. The condition $\mathbf{C}^T \mathbf{1} = \mathbf{1}$ means that it adopts the affine subspace model, which is a special linear subspace model [13,15]. As the ℓ_0 -norm optimization problem is non-deterministic polynomial hard (NP-hard), the relaxed ℓ_1 -norm is usually adopted [22]:

$$\min_{\mathbf{C}} \|\mathbf{C}\|_1 \text{ s.t. } \mathbf{Y} = \mathbf{Y}\mathbf{C} + \mathbf{N}, \text{diag}(\mathbf{C}) = \mathbf{0}, \mathbf{C}^T \mathbf{1} = \mathbf{1} \quad (2)$$

The optimization problem in Equation (2) can be effectively solved by the alternating direction method of multipliers (ADMM) algorithm [23–25]. We then construct the similarity graph using the sparse coefficient matrix [15–17]. Meanwhile, the symmetric form is adopted to enhance the connectivity of the graph:

$$\mathbf{W} = |\mathbf{C}| + |\mathbf{C}|^T \quad (3)$$

where $\mathbf{W} \in \mathbb{R}^{MN \times MN}$ denotes the similarity graph whose element $W_{i,j}$ represents the similarity between pixel i and pixel j . The spectral clustering algorithm is then applied to the similarity graph to obtain the final clustering result [26–28].

3. Kernel Sparse Subspace Clustering Algorithm with a Spatial Max Pooling Operation (KSSC-SMP)

In this section, we introduce the newly developed KSSC-SMP algorithm. The proposed approach attempts to fully exploit both the nonlinear structure of HSIs and the potential of the SSC model to achieve more accurate clustering results. Considering the complex nonlinear structure of HSIs, we map the feature points to a much higher dimensional kernel space with the kernel strategy to make them linearly separable. In this way, a more accurate representation coefficient matrix can be obtained. In addition, the spatial max pooling operation is introduced to yield new features with these coefficients by incorporating the spatial-contextual information.

3.1. Kernel Sparse Subspace Clustering Algorithm (KSSC)

The kernel sparse subspace clustering (KSSC) algorithm extends the SSC algorithm to nonlinear manifolds by the use of a kernel strategy to map the feature points from the original space to a much higher dimensional space, in order to make them linearly separable [29]. The kernel mapping directly takes the signal in the kernel space as a feature [30] and represents each signal with the others from its

own subspace, as in SSC. The sparse representation coefficient matrix can then be obtained by solving the following optimization problem:

$$\min_{\mathbf{C}} \|\mathbf{C}\|_1 + \lambda \|K(\mathbf{Y}) - K(\mathbf{Y})\mathbf{C}\|_F^2 \text{ s.t. } \text{diag}(\mathbf{C}) = \mathbf{0}, \mathbf{C}^T \mathbf{1} = \mathbf{1} \quad (4)$$

where λ represents a tradeoff between the data fidelity term and the sparsity term, and $K: \mathbb{R}^p \rightarrow \mathbb{R}^p$ is a mapping function from the input space to the reproducing kernel space [30]. We let $K_{\mathbf{Y}\mathbf{Y}} \in \mathbb{R}^{MN \times MN}$ denote a positive semi-definite kernel Gram matrix whose elements are computed as follows:

$$[K_{\mathbf{Y}\mathbf{Y}}]_{i,j} = [\langle K(\mathbf{Y}), K(\mathbf{Y}) \rangle]_{i,j} = K(\mathbf{y}_i)^T K(\mathbf{y}_j) = \kappa(\mathbf{y}_i, \mathbf{y}_j) \quad (5)$$

where \mathbf{y}_i is the “spectral pixel” at location i in HSI \mathbf{Y} and $\kappa: \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ represents the kernel function, which measures the similarity of two arguments denoted as a pair of pixels. Commonly used kernels include the radial basis function (RBF) kernel $\kappa(\mathbf{y}_i, \mathbf{y}_j) = \exp(-\delta \|\mathbf{y}_i - \mathbf{y}_j\|^2)$ and the polynomial kernel $\kappa(\mathbf{y}_i, \mathbf{y}_j) = \langle \langle \mathbf{y}_i, \mathbf{y}_j \rangle + a \rangle^b$, where δ , a , and b are the parameters of the kernel functions. In this paper, the RBF kernel is adopted. As the feature space of the RBF kernel has an infinite number of dimensions, and its value decreases with distance between (0, 1), it can be readily interpreted as a similarity measure [30].

3.2. Incorporating Spatial Information with the Spatial Max Pooling Operation

Through sparse representation, the coefficients can reveal the underlying cluster structure and can be directly utilized as features. However, when only using spectral information, a single representation coefficient vector of the target pixel can only provide limited discriminative information [31,32]. Based on the fact that pixels within a local patch (where the center pixel is the target pixel to be processed) have a high probability of being associated with the same thematic class, the sparse representation coefficients of these pixels are also expected to be similar. Therefore, the spatial pooling operation can be considered a reasonable way to merge these similar sparse representation coefficient vectors and yield a new feature vector with better discriminative ability, which is an approach that has been widely used in the literature [33,34].

At this point, considering the mechanism adopted by SSC (which pays more attention to larger coefficients and tends to highlight their role), the spatial max pooling operation is adopted to merge these sparse coefficient vectors into a new pooling vector, which suppresses the small elements and preserves the large ones. This is because the larger the representation coefficients, the more similar to the target pixel the corresponding representation atoms are, and the greater the contribution they make to the sparse representation of the target pixel. In this process, since all the representation coefficients within the local patch are simultaneously taken into consideration to generate a new representation coefficient vector, the spatial-contextual information can be naturally incorporated. Moreover, it should be noted that the spatial max pooling operation is performed in the representation dimension, not directly on the pixel values.

Firstly, the 2-D sparse coefficient matrix $\mathbf{C} \in \mathbb{R}^{MN \times MN}$ is reshaped into a 3-D cube $\hat{\mathbf{C}} \in \mathbb{R}^{M \times N \times MN}$ along rows, which can be seen as a “coefficient hyperspectral cube”. We then consider a local neighborhood window of size S for each representation coefficient vector, with the currently processed one as the center. Following this, the max pooling operation is performed on the local neighborhood window to select the local largest coefficient value of each band to act as one element of the newly generated coefficient vector for the target pixel. In this way, the more meaningful and discriminative information among the local neighborhood is extracted by the newly generated sparse representation coefficient vector, and the spatial-contextual information is naturally incorporated and fused in the max pooling process to generate the new spectral-spatial features. Finally, the obtained coefficient features are utilized to cluster the HSI. The process of spatial max pooling is illustrated in Figure 1.

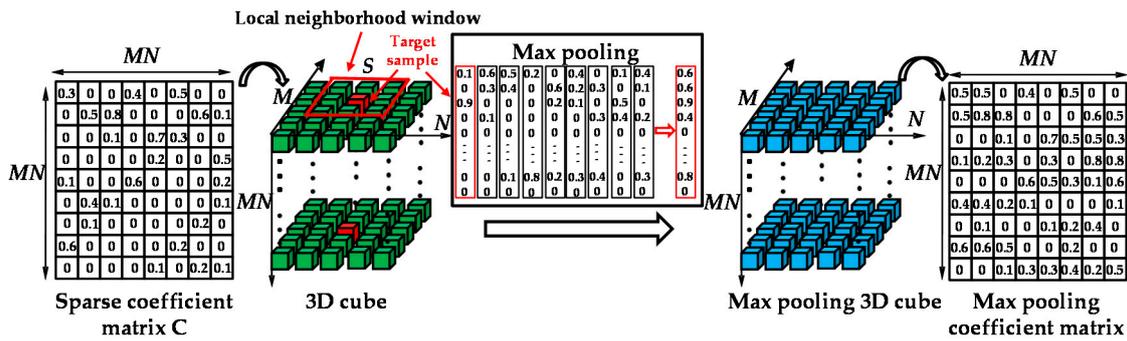


Figure 1. Graphical illustration of the process of spatial max pooling. Firstly, the 2-D sparse coefficient matrix is reshaped into a 3-D cube along rows. Then, a local neighborhood window of size S is considered for each representation coefficient vector, with the currently processed one as the center. Following this, the max pooling operation is performed on the local neighborhood window to select the local, largest coefficient value of each band to act as one element of the newly generated coefficient vector for the target pixel. Lastly, with the obtained max pooling 3-D cube reshaped into a 2-D matrix, the target max pooling coefficient matrix can be obtained.

3.3. The KSSC-SMP Algorithm

In order to further improve the clustering performance, we combine the two aforementioned schemes into a unified framework to obtain the KSSC-SMP algorithm, which can simultaneously deal with the complex nonlinear structure and utilize the spectral-spatial attributes of HSIs. The KSSC-SMP algorithm can be summarized as shown in Algorithm 1.

Algorithm 1. The kernel sparse subspace clustering algorithm with a spatial max pooling operation (KSSC-SMP) for hyperspectral remote sensing imagery

Input:

- 1) A 2-D matrix of the HSI containing a set of points $\{y_i\}_{i=1}^{MN}$, in a union of l affine subspaces $\{S_i\}_{i=1}^l$;
- 2) Parameters, including the cluster number l , the regulation parameter λ , the kernel parameter δ , and the window size S of the spatial max pooling operation.

Main algorithm:

- 1) Construct the kernel sparse representation optimization model (Equation (4)) and solve it to obtain the kernel sparse representation coefficient matrix C using ADMM;
- 2) Conduct the spatial max pooling operation on C to obtain the max pooling coefficient matrix, as shown in Figure 1;
- 3) Construct the similarity graph with the max pooling coefficient matrix;
- 4) Apply spectral clustering to the similarity graph to obtain the final clustering results;

Output:

A 2-D matrix which records the labels of the clustering result of the HSI.

The flowchart of the proposed KSSC-SMP algorithm is given in Figure 2.

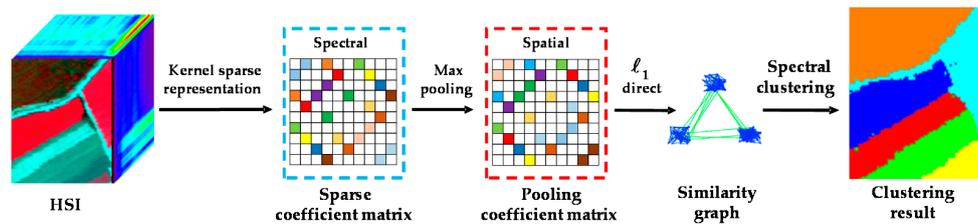


Figure 2. The flowchart of the proposed kernel sparse subspace clustering algorithm with a spatial max pooling operation (KSSC-SMP) algorithm. Firstly, the kernel sparse representation model is adopted to more precisely represent each target pixel by exploiting the inherent nonlinearity of hyperspectral images (HSIs). The spatial max pooling operation is then introduced to yield new features with these coefficients by incorporating the spatial-contextual information. Next, the similarity graph is constructed in the ℓ_1 -direct way. Lastly, by applying the spectral clustering algorithm to the similarity graph, the final clustering result can be obtained.

4. Experimental Results

In order to evaluate the effectiveness of the proposed KSSC-SMP algorithm, the following state-of-the-art clustering approaches were implemented for comparative purposes: FCM-S1 [8], clustering by fast search and find of density peaks (CFSEDP) [35], SSC [13], S⁴C [15], and L2-SSC [16]. The number of clusters was determined by carefully observing the original image, and the parameters of each clustering method were carefully optimized through the well-known grid search strategy. The thematic information was manually determined to give a label for each cluster group through cross-referencing between the clustering result and the original image. To thoroughly evaluate the performance of each clustering method, both the visual cluster map and quantitative evaluation measures are provided, including the accuracy for each class (producer’s accuracy), the overall accuracy (OA), the kappa coefficient, the z-value, and the running time.

The confusion matrix is defined as shown in Table 1, which is a specific table layout that allows visualization of the performance of an algorithm. In the clustering field, it is commonly called a “matching matrix”. Based on the confusion matrix, the producer’s accuracy, OA, and kappa are calculated as follows.

Table 1. Confusion matrix of an algorithm.

Predicted Class \ Actual Class		Predicted Class							Row Total
		1	...	<i>i</i>	...	<i>j</i>	...	<i>l</i>	
1		$p_{1,1}$...	$p_{1,i}$...	$p_{1,j}$...	$p_{1,l}$	$p_{1,+}$
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<i>i</i>		$p_{i,1}$...	$p_{i,i}$...	$p_{i,j}$...	$p_{i,l}$	$p_{i,+}$
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<i>j</i>		$p_{j,1}$...	$p_{j,i}$...	$p_{j,j}$...	$p_{j,l}$	$p_{j,+}$
⋮		⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
<i>l</i>		$p_{l,1}$...	$p_{l,i}$...	$p_{l,j}$...	$p_{l,l}$	$p_{l,+}$
Column total		$p_{+,1}$...	$p_{+,i}$...	$p_{+,j}$...	$p_{+,l}$	MN

l: the number of classes; *MN*: the total number of pixels; *i*: the *i*th class; $p_{i,j}$: the number of pixels that are actually class *i* but predicted as class *j*, where $1 \leq i, j \leq l$; $p_{i,+}$: the number of the pixels predicted as class *i*; $p_{+,i}$: the number of pixels of actual class *i*.

$$\text{Producer's accuracy} : p_i = \frac{p_{i,i}}{p_{i,+}} \tag{6}$$

$$\text{OA} = \frac{\sum_{i=1}^l p_{i,i}}{MN} \tag{7}$$

$$Kappa = \frac{MN \sum_{i=1}^l p_{i,i} - \sum_{i=1}^l (p_{i,+} \cdot p_{+,i})}{(MN)^2 - \sum_{i=1}^l (p_{i,+} \cdot p_{+,i})} \quad (8)$$

where p_i refers to the producer's accuracy of the i th class. The z-value measure is based on the contingency table, as shown in Table 2, and it can be calculated as follows [36].

Table 2. Contingency table of two different clustering methods.

Algorithm Test	Algorithm 2 Positive	Algorithm 2 Negative	Row Total
Algorithm 1 positive	$f_{1,1}$	$f_{1,2}$	$f_{1,1} + f_{1,2}$
Algorithm 1 negative	$f_{2,1}$	$f_{2,2}$	$f_{2,1} + f_{2,2}$
Column total	$f_{1,1} + f_{2,1}$	$f_{1,2} + f_{2,2}$	$f_{1,1} + f_{1,2} + f_{2,1} + f_{2,2}$

$f_{1,1}$: the number of pixels that both algorithms classified correctly; $f_{1,2}$: the number of pixels that Algorithm 1 classified correctly but Algorithm 2 misclassified; $f_{2,1}$: the number of pixels that Algorithm 2 classified correctly but Algorithm 1 misclassified; $f_{2,2}$: the number of pixels that both algorithms misclassified.

$$z = \frac{f_{1,2} - f_{2,1}}{\sqrt{f_{1,2} + f_{2,1}}} \quad (9)$$

where z represents the significance of the difference between the two clustering algorithms.

4.1. Experimental Results Obtained with the Salinas Image

The first experiment was conducted on the well-known Salinas image. This scene was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor over the Salinas Valley, CA, USA, and consists of 512×217 pixels and 224 spectral reflectance bands in the wavelength range of 0.4–2.5 μm . Twenty bands (108–112, 154–167, and 224) corresponding to water absorption and noisy bands were removed, leaving 204 bands for the experiment. In this scene, there are 16 different classes, which are mainly different kinds of vegetation. A typical area at a size of 100×80 was selected as the test data, containing six main land-cover classes: Broccoli-green-weeds-1 (Broccoli-1), Broccoli-green-weeds-2 (Broccoli-2), Soil-vinyard-develop (Soil), Lettuce-romaine-4wk (Lettuce-4wk), Lettuce-romaine-5wk (Lettuce-5wk), and Lettuce-romaine-6wk (Lettuce-6wk). To be consistent with all the unsupervised methods, all the samples in the ground truth were utilized as the test samples. The false-color composite image and the ground truth are shown in Figure 3a,b, respectively. The parameters of each clustering algorithm in this experiment were set as shown in Table 3.

Table 3. Parameters of each clustering method for the Salinas image: fuzzy c -means with spatial information (FCM-S1), clustering by fast search and find of density peaks (CFSFDP), sparse subspace clustering (SSC), spectral-spatial sparse subspace clustering (S^4C), ℓ_2 -norm regularized sparse subspace clustering (L2-SSC), and kernel sparse subspace clustering with a spatial max pooling operation (KSSC-SMP).

Method	Salinas
FCM-S1	$l = 6, e = 2, \alpha = 0.9$
CFSFDP	—
SSC	$l = 6, \lambda = 2.63 \times 10^{-6}$
S^4C	$l = 6, \lambda = 8.78 \times 10^{-7}, \alpha = 5 \times 10^3, S = 3$
L2-SSC	$l = 6, \lambda = 3.07 \times 10^{-6}, \alpha = 3 \times 10^3$
KSSC-SMP	$l = 6, \lambda = 3.07 \times 10^{-6}, \delta = 10^{-7}, S = 3$

l : cluster number; e : fuzzy exponential; λ : tradeoff parameter; α : tradeoff parameter; δ : kernel parameter; S : window size parameter.

The thematic maps obtained by each clustering algorithm with the Salinas image and the corresponding quantitative evaluation of the clustering precision are provided in Figure 3 and Table 4,

respectively. In Table 4, the best result in each row is shown in bold, with the second-best result underlined. From the figure and the table, it can be clearly observed that the proposed KSSC-SMP algorithm significantly improves the clustering performance and is superior to the other state-of-the-art clustering methods. KSSC-SMP achieves an optimal or sub-optimal precision for all the classes. As a result, it achieves the best OA of 99.89% and kappa of 0.9986.

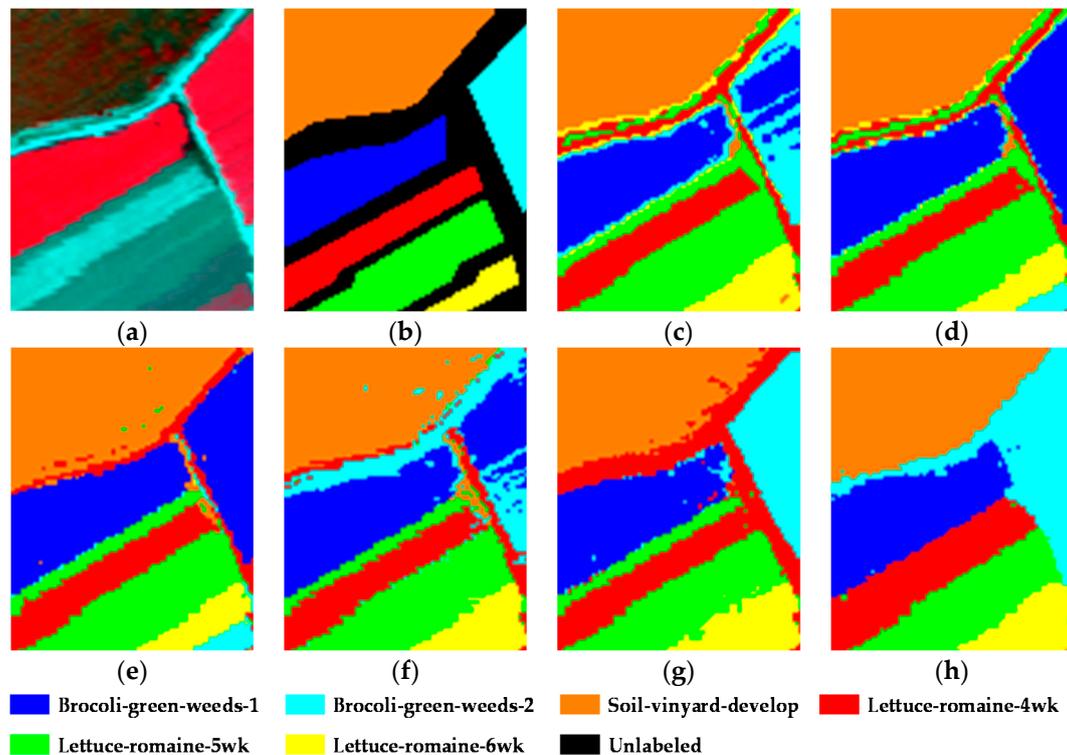


Figure 3. The cluster maps of the different methods for the Salinas image: (a) false-color image (R: 70, G: 27, B: 17); (b) ground truth; (c) FCM-S1; (d) CFSFDP; (e) SSC; (f) S⁴C; (g) L2-SSC; and (h) KSSC-SMP.

Table 4. Quantitative evaluation of the different clustering algorithms for the Salinas image: the best result in each row is shown in bold, with the second-best result underlined.

Method	Class	FCM-S1	CFSFDP	SSC	S ⁴ C	L2-SSC	KSSC-SMP
Producer's accuracy (%)	Broccoli-1	99.01	100	<u>99.90</u>	99.51	98.72	100
	Broccoli-2	<u>46.93</u>	0	0	31.75	100	100
	Soil	97.10	97.96	<u>99.49</u>	97.30	95.32	99.95
	Lettuce-4wk	93.39	93.95	94.23	93.67	<u>95.08</u>	100
	Lettuce-5wk	100	100	100	100	94.84	<u>99.46</u>
	Lettuce-6wk	100	<u>95.20</u>	93.45	100	100	100
Overall accuracy (OA) (%)		91.64	86.44	86.93	90.04	<u>96.58</u>	99.89
Kappa		0.8922	0.8241	0.8301	0.8731	<u>0.9563</u>	0.9986
z-value		13.2343	−4.4388	0	10.3494	<u>18.6001</u>	26.5168
Time (s)		59.11	<u>75.52</u>	9.24×10^2	9.59×10^2	8.95×10^2	3.81×10^2

4.2. Experimental Results Obtained with the University of Pavia Image

The second experimental dataset was the University of Pavia image scene, which was collected by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight campaign over Pavia, Northern Italy. This image has 610×340 pixels, with 103 spectral reflectance bands utilized in the experiment after removing the water absorption and noisy bands. The geometric resolution of this image is 1.3 m, with nine classes contained in the scene. Considering the computational complexity, a typical sub-set at a size of 200×100 was utilized as the test data [15–17], with eight main land-cover

classes: Metal sheet, Asphalt, Meadows, Trees, Bare soil, Bitumen, Bricks, and Shadows. All the samples in the ground truth were utilized as the test samples. The composite false-color image and the corresponding ground truth are provided in Figure 4a,b. Differing from the first experimental dataset, this scene is of a typical city area that has a more complex distribution of land-cover classes. The parameters of each clustering algorithm in this experiment were set as shown in Table 5.

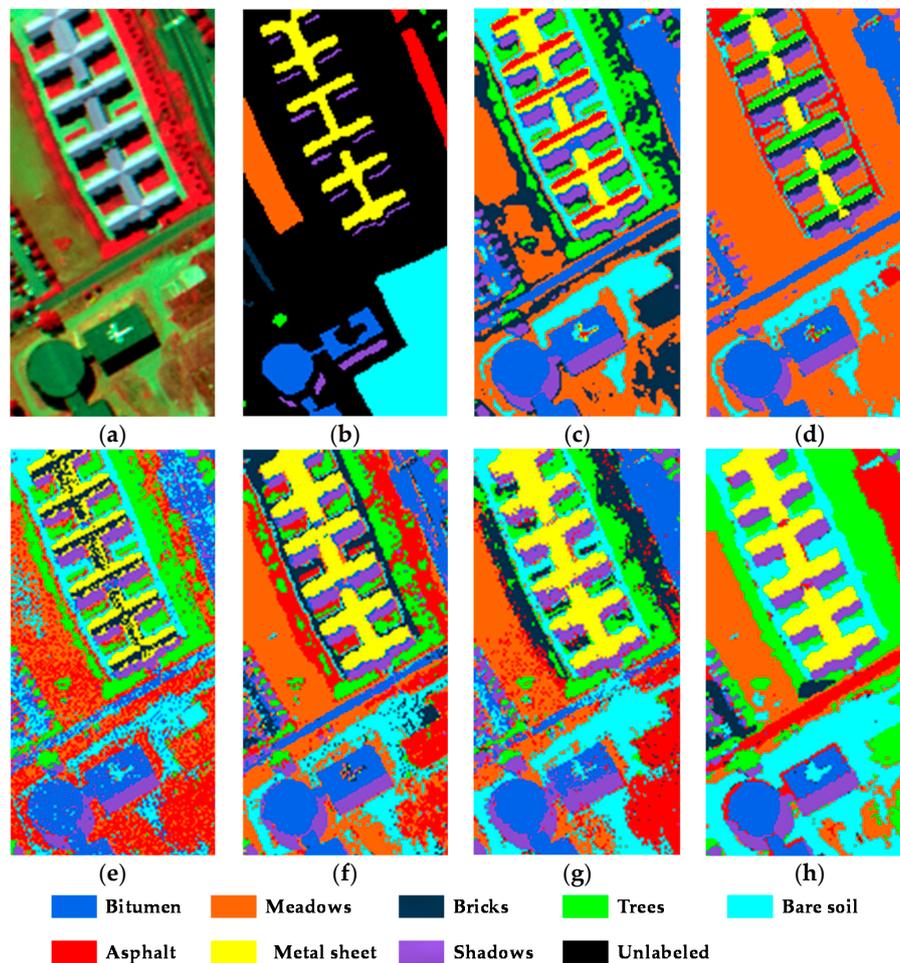


Figure 4. The cluster maps of the different methods for the University of Pavia image: (a) false-color image (R: 102, G: 56, B: 31); (b) ground truth; (c) FCM-S1; (d) CFSFDP; (e) SSC; (f) S^4C ; (g) L2-SSC; and (h) KSSC-SMP.

Table 5. Parameters of each clustering method for the University of Pavia image: fuzzy c -means with spatial information (FCM-S1), clustering by fast search and find of density peaks (CFSFDP), sparse subspace clustering (SSC), spectral-spatial sparse subspace clustering (S^4C), ℓ_2 -norm regularized sparse subspace clustering (L2-SSC), and kernel sparse subspace clustering with a spatial max pooling operation (KSSC-SMP).

Method	University of Pavia
FCM-S1	$l = 8, e = 2, \alpha = 0.9$
CFSFDP	—
SSC	$l = 8, \lambda = 1.58 \times 10^{-5}$
S^4C	$l = 8, \lambda = 1.58 \times 10^{-5}, \alpha = 3.0 \times 10^3, S = 3$
L2-SSC	$l = 8, \lambda = 1.58 \times 10^{-5}, \alpha = 9 \times 10^3$
KSSC-SMP	$l = 8, \lambda = 1.58 \times 10^{-5}, \delta = 10^{-8}, S = 3$

l : cluster number; e : fuzzy exponential; λ : tradeoff parameter; α : tradeoff parameter; δ : kernel parameter; S : window size parameter.

The cluster maps of the University of Pavia image obtained by each algorithm are shown in Figure 4, with the corresponding quantitative evaluation provided in Table 6. For consistency, the best result in each row is shown in bold and the second-best result is underlined. In this experiment, it can be clearly seen that the clustering results are consistent with those of the first experiment, and similar conclusions can be drawn. In this experiment, the proposed KSSC-SMP, again, achieves the best clustering performance in both the visual and quantitative evaluations.

Table 6. Quantitative evaluation of the different clustering algorithms for the University of Pavia image. The best result in each row is shown in bold, with the second-best result underlined.

Method	Class	FCM-S1	CFSFDP	SSC	S ⁴ C	L2-SSC	KSSC-SMP
Producer's accuracy (%)	Metal sheet	60.99	37.26	53.84	<u>99.09</u>	97.34	99.85
	Asphalt	0	0	<u>0.71</u>	0	<u>0.71</u>	100
	Meadows	87.37	100	52.60	<u>99.09</u>	92.97	<u>99.09</u>
	Trees	58.73	0	100	<u>87.30</u>	60.32	100
	Bare soil	24.97	21.41	22.04	31.93	<u>37.20</u>	51.58
	Bitumen	100	100	85.58	<u>98.37</u>	87.67	95.47
	Bricks	0	0	0	<u>60.64</u>	0	98.94
	Shadows	<u>99.72</u>	100	98.61	98.61	93.91	100
Overall accuracy (OA) (%)		52.27	46.97	43.97	<u>65.09</u>	63.30	80.02
Kappa		0.4484	0.3874	0.3495	<u>0.5852</u>	0.5618	0.7583
z-value		21.9698	6.8536	0	<u>36.5971</u>	34.1658	48.1975
Time (s)		3.69 × 10²	<u>4.81 × 10²</u>	1.29 × 10 ⁴	7.40 × 10 ³	1.33 × 10 ⁴	1.98 × 10 ³

5. Discussion

5.1. Experimental Results Analysis

From Figures 3 and 4 and Tables 4 and 6, it can be clearly observed that the clustering performance of FCM-S1 is far from satisfactory, as this clustering model assumes that each cluster satisfies a convex “ball-like” distribution, and it cannot handle the large spectral variability of HSIs [15]. As a result, misclassification and within-class noise appear in the cluster maps. Compared with FCM-S1, CFSFDP performs even more poorly with a higher level of misclassification, with only a few classes effectively recognized. This suggests that the density-based segmentation strategy does not fit well with the complex structure of HSIs. Specifically, both these methods completely fail to recognize the Asphalt and Bricks classes in the second experiment.

We now compare the clustering results of the SSC algorithm and its extensions. Due to the fact that it ignores the spatial information, the clustering results of SSC are far from satisfactory, with a large amount of salt-and-pepper noise and numbers of misclassifications. Although SSC successfully distinguishes most of the classes in the first experiment, showing great potential for hyperspectral clustering, it performs very badly in the second experiment and obtains the lowest clustering precision. Compared with SSC, S⁴C effectively improves the clustering performance, by utilizing the spectral weighting strategy to guarantee that highly correlated pixels occupy the dominant place in the representation process, and incorporates the spatial neighborhood information with the local averaging constraint. For example, in the first experiment, the recognition precision of Brocoli-green-weeds-2 and Lettuce-romaine-6wk is improved from 0% and 93.45% to 31.75% and 100%, respectively. On the other hand, L2-SSC also obtains a much better clustering precision than that of the original SSC method, which again suggests the importance of the spatial information from another perspective. However, due to the limitations of the linear representation, there are still some classes that are not effectively recognized by these two methods, such as the Asphalt and Bricks classes in the second experiment. Compared with the three linear methods, the proposed KSSC-SMP algorithm further improves the clustering performance by exploiting the nonlinear structure of the HSI and generating more discriminative features by incorporating the spatial-contextual information with the spatial max pooling operation. For example, in the second experiment, for the Asphalt and Bricks

classes, which are beyond the recognition capability of most of the methods, the recognition precision of KSSC-SMP is 100% and 98.94%, respectively, which suggests the superiority of the nonlinear method. As a result, KSSC-SMP achieves much smoother cluster maps with a much higher clustering precision in both experiments, which confirms the superiority of the nonlinear method. Increments in OA of approximately 13%, 9%, and 3% are achieved in the first experiment, compared with that of SSC, S⁴C, and L2-SSC, respectively, and increments in OA of approximately 26%, 15%, and 17% are achieved in the second experiment. Overall, KSSC-SMP outperforms the other state-of-the-art clustering methods and achieves the best clustering performance in both the visual and quantitative evaluations.

In order to reflect the significant difference between the clustering methods, we took the clustering result of SSC as the baseline and calculated the z-values [36] of the other clustering results. For the widely used 5% level of significance, we consider that there is a significant difference between the two clustering methods if $|z| > 3.84$. From Tables 4 and 6, it can be clearly observed that S⁴C and L2-SSC perform significantly better than the original SSC algorithm. In addition, the proposed KSSC-SMP algorithm also performs significantly better than the linear methods. On the other hand, from the last rows of Tables 4 and 6, it can be seen that the four SSC-based algorithms require more computation time than CFSFDP and FCM-S1, but they do provide better clustering accuracies. Fortunately, with the development of computer hardware and the parallel computing technique, this will not be a major problem for much longer.

5.2. Sensitivity Analysis of the Parameters

There are three parameters in the proposed KSSC-SMP algorithm: the regularization parameter λ , the kernel parameter δ , and the window size parameter S . When analyzing one parameter, the other two parameters were fixed at the optimal values. Both the AVIRIS Salinas image scene and the University of Pavia image scene were utilized to test the sensitivity of the parameters of KSSC-SMP.

As a tradeoff between the data fidelity term and the sparsity term, the sensitivity of the regularization parameter λ was first analyzed. For this parameter, we adopted the same strategy that was used with the other SSC-based methods in [13,15]. In practice, parameter λ is decided by the following formulation [15]:

$$\lambda = \frac{\beta}{\mu} \quad (10)$$

$$\mu \triangleq \min_i \max_{j \neq i} |y_i^T y_j| \quad (11)$$

where β is a coefficient, and μ can be exactly computed for each dataset.

From Equations (10) and (11), it can be easily found that the sensitivity of λ is actually decided by β , as μ is fixed for a certain dataset. Therefore, in practice, we only need to fine tune β . Figure 5 shows the compromise between OA and β for both experimental images. In Figure 5, the optimal clustering values of SSC, S⁴C, and L2-SSC are also provided. From the figure, it can be observed that β is independent of the dataset, to some degree, as the optimal value always falls into a relatively narrow range of [1000, 1400]. For a certain dataset, β can be easily fine-tuned. In addition, it should be noted that KSSC-SMP always achieves a much better precision than the optimal values of the other three linear SSC-based methods, which further demonstrates its effectiveness.

The kernel parameter δ controls the quality of the projection of the feature points from the original space to the kernel space, and is quite important for the proposed KSSC-SMP algorithm. Similarly, the sensitivity of the kernel parameter was analyzed through a series of experiments, as shown in Figure 6. From the figure, it can be seen that the KSSC-SMP algorithm is relatively robust for this parameter, and its performance is relatively stable. The optimal value of δ varies for the different datasets. For the Salinas image, the optimal value is 1×10^{-7} , and for the University of Pavia image, it is 1×10^{-8} . However, it can be seen that over a very large range of δ , KSSC-SMP can achieve a better performance than the three linear methods. Therefore, the KSSC-SMP algorithm is suitable for use in real applications.

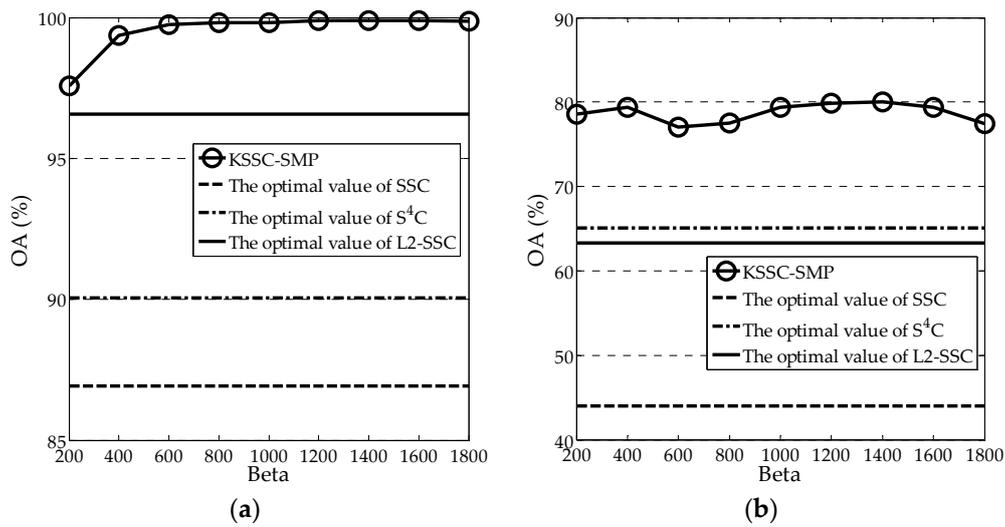


Figure 5. Analysis of parameter β in the clustering performance for both image scenes. (a) Change trend of OA versus β for the Salinas image; (b) Change trend of OA versus β for the University of Pavia image.

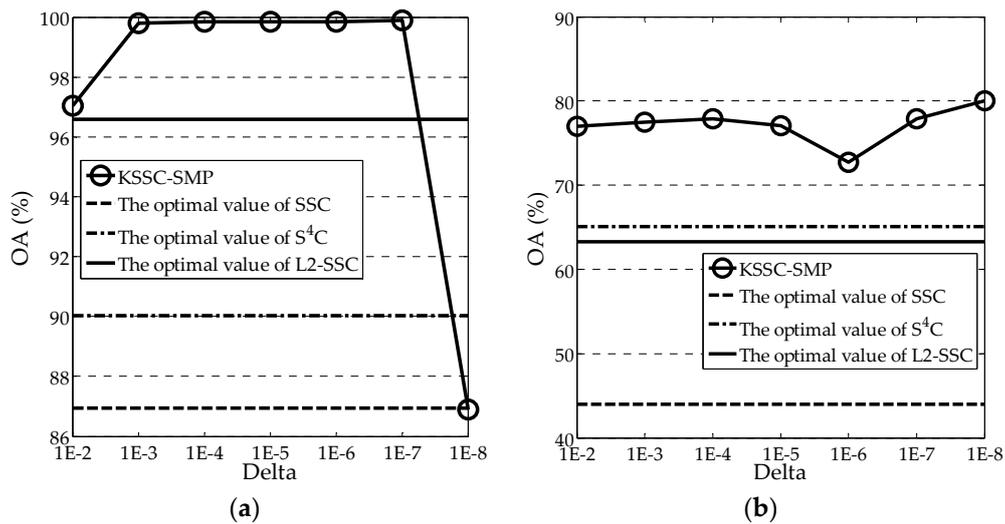


Figure 6. Analysis of parameter δ in the clustering performance for both image scenes. (a) Change trend of OA versus δ for the Salinas image; (b) Change trend of OA versus δ for the University of Pavia image.

The window size parameter S controls the spatial max pooling operation. The influence of this parameter on the clustering performance is shown in Figure 7, using a similar strategy to that adopted for the other two parameters. From the figure, it can be seen that the spatial information plays a very important role in the clustering process. For both experimental scenes, the optimal value of parameter S is 3. Both these scenes have a relatively complex distribution of land-cover classes. Generally, parameter S is closely related to the resolution of the image. Again, over a large range of parameter S , KSSC-SMP can achieve a better clustering performance than the three linear methods, which further confirms its effectiveness.

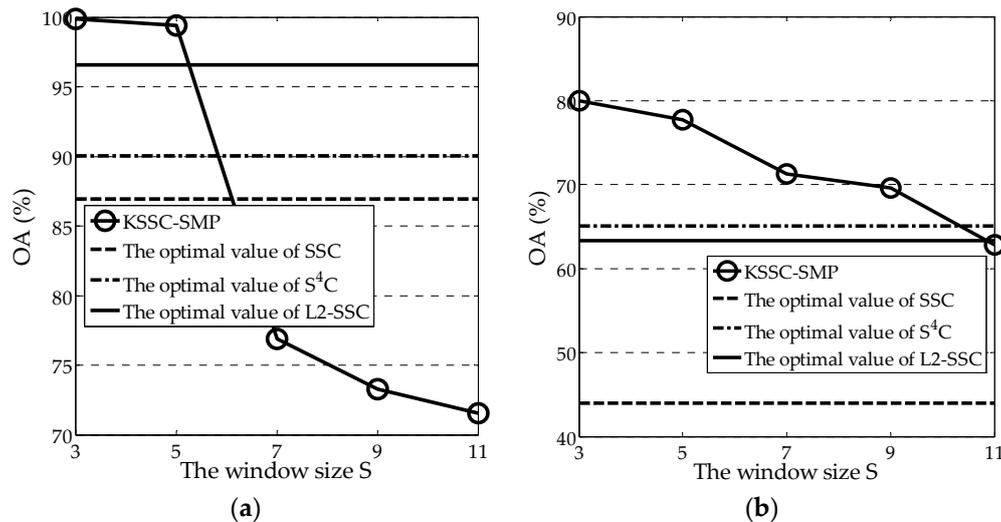


Figure 7. Analysis of the window size parameter S in the clustering performance for both image scenes. (a) Change trend of OA versus S for the Salinas image. (b) Change trend of OA versus S for the University of Pavia image.

6. Conclusions and Future Lines of Research

In this paper, we have proposed a new kernel sparse subspace clustering algorithm with a spatial max pooling operation (KSSC-SMP) for HSI data interpretation. The proposed approach focuses on exploiting the complex nonlinear structure of HSIs and obtaining a more precise representation coefficient matrix with the kernel sparse representation. In addition, the spatial max pooling operation is introduced to these coefficients to yield new features to fully exploit the discriminative spectral-spatial information of HSIs and the potential of the SSC model. Thereby, KSSC-SMP can effectively improve the clustering performance and guarantee the spatial homogeneity of the final clustering result.

Although the results show that the KSSC-SMP algorithm is very competitive, there are several aspects that could be improved. For instance, the method could be further improved by adaptively determining the regularization parameters and allowing for the extraction of more discriminative spatial features. All of these issues will be addressed in our future work.

Acknowledgments: This work was supported in part by the National Natural Science Foundation of China under grants 41571362, 61261130587, and 41431175, and by the Fundamental Research Funds for Central Universities under grant No. 2042017kf0218 and 2015904020202.

Author Contributions: Han Zhai and Hongyan Zhang conceived and designed the experiments; Han Zhai conducted the experiments; Han Zhai, Xiong Xu, and Hongyan Zhang analyzed the data; Han Zhai, Hongyan Zhang, Xiong Xu, Liangpei Zhang, and Pingxiang Li wrote the paper.

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

References

- Joyce, K.E.; Phinn, S.R.; Roelfsema, C.M. Live coral cover index testing and application with hyperspectral airborne image data. *Remote Sens.* **2013**, *5*, 6116–6137. [[CrossRef](#)]
- Priem, F.; Canters, F. Synergistic use of LiDAR and APEX hyperspectral data for high-resolution urban land cover mapping. *Remote Sens.* **2016**, *8*, 787. [[CrossRef](#)]
- Wang, K.; Yong, B. Application of the frequency spectrum to spectral similarity measures. *Remote Sens.* **2016**, *8*, 344. [[CrossRef](#)]
- Baldeck, C.A.; Asner, G.P. Estimating vegetation beta diversity from airborne imaging spectroscopy and unsupervised clustering. *Remote Sens.* **2013**, *5*, 2057–2071. [[CrossRef](#)]
- Lloyd, S.P. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [[CrossRef](#)]

6. Bezdek, J.C. *Pattern Recognition with Fuzzy Objective Function Algorithms*; Springer Science & Business Media: New York, NY, USA, 2013.
7. Luo, M.; Ma, Y.F.; Zhang, H.J. A spatial constrained k -means approach to image segmentation. In Proceedings of the International Conference on Information, Communications and Signal Processing/Pacific Rim Conference on Multimedia, Singapore, Singapore, 15–18 December 2003.
8. Chen, S.; Zhang, D. Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. *IEEE Trans. Syst. Man Cybern. Part B Cybern.* **2004**, *34*, 1907–1916. [[CrossRef](#)]
9. Tang, H.; Shen, L.; Yang, X.; Qi, Y.; Jiang, W.; Gong, A. A MRF-based clustering algorithm for remote sensing images by using the latent Dirichlet allocation model. *Procedia Earth Planet. Sci.* **2011**, *2*, 358–363. [[CrossRef](#)]
10. Cariou, C.; Chehdi, K. Unsupervised nearest neighbors clustering with application to hyperspectral images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2015**, *9*, 1105–1116. [[CrossRef](#)]
11. Vijendra, S. Efficient clustering for high dimensional data: Subspace based clustering and density based clustering. *Inf. Technol. J.* **2011**, *10*, 1092–1105. [[CrossRef](#)]
12. 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](#)]
13. Elhamifar, E.; Vidal, R. Sparse subspace clustering: Algorithm, theory, and application. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 2765–2781. [[CrossRef](#)] [[PubMed](#)]
14. Elhamifar, E.; Vidal, R. Sparse subspace clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–26 June 2009.
15. 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](#)]
16. 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**. [[CrossRef](#)]
17. Zhai, H.; Zhang, H.; Zhang, L.; Li, P. Reweighted mass center based object-oriented sparse subspace clustering for hyperspectral images. *J. Appl. Remote Sens.* **2016**, *10*, 046014. [[CrossRef](#)]
18. Mercier, G.; Lennon, M. Support vector machines for hyperspectral image classification with spectral-based kernels. In Proceedings of the IEEE Geoscience and Remote Sensing Symposium, Toulouse, France, 21–25 July 2003.
19. Chen, Y.; Nasrabadi, N.M.; Tran, T.D. Hyperspectral image classification via kernel sparse representation. *IEEE Trans. Geosci. Remote Sens.* **2013**, *51*, 217–231. [[CrossRef](#)]
20. De Morsier, F.; Borgeaud, M.; Gass, V.; Thiran, J.P.; Tuia, D. Kernel low-rank and sparse graph for unsupervised and semi-supervised classification of hyperspectral images. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 3410–3420. [[CrossRef](#)]
21. Zhang, H.; Zhai, H.; Liao, W.; Cao, L.; Zhang, L.; Pizurica, A. Hyperspectral image kernel sparse subspace clustering with spatial max pooling operation. In Proceedings of the XXIII Congress of the International Society For Photogrammetry and Remote Sensing, Prague, Czech Republic, 12–19 July 2016.
22. 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](#)]
23. Mota, J.F.; Xavier, J.W.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](#)]
24. Benzi, M.; Olshanskii, M.A. An augmented Lagrangian-based approach to the Oseen problem. *SIAM J. Sci. Comput.* **2006**, *28*, 2095–2113. [[CrossRef](#)]
25. Afonso, M.V.; Bioucas-Dias, J.M.; Figueiredo, M.A. An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. *IEEE Trans. Image Process.* **2011**, *20*, 681–695. [[CrossRef](#)] [[PubMed](#)]
26. Ng, A.Y.; Jordan, M.I.; Weiss, Y. On spectral clustering: Analysis and an algorithm. In Proceedings of Neural Information Processing Systems: Natural and Synthetic Monday, Vancouver, BC, Canada, 3 December–8 December 2001.
27. Shi, J.; Malik, J. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 888–905.
28. Wu, Z.; Leahy, R. An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1993**, *15*, 1101–1113. [[CrossRef](#)]

29. Patel, V.M.; Vidal, R. Kernel sparse subspace clustering. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Paris, France, 2014.
30. Li, J.; Zhang, H.; Zhang, L. Column-generation kernel nonlocal joint collaborative representation for hyperspectral image classification. *ISPRS J. Photogramm. Remote Sens.* **2014**, *94*, 25–36. [[CrossRef](#)]
31. He, W.; Zhang, H.; Zhang, L.; Shen, H. Total-variation regularized low-rank matrix factorization for hyperspectral image restoration. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 178–188. [[CrossRef](#)]
32. Yuan, H.; Tang, Y.Y. A novel sparsity-based framework using max pooling operation for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.* **2014**, *7*, 3570–3576. [[CrossRef](#)]
33. Gao, S.; Tsang, I.; Chia, L. Laplacian sparse coding, hypergraph Laplacian sparse coding and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2013**, *35*, 92–104. [[CrossRef](#)] [[PubMed](#)]
34. Huang, Y.; Huang, K.; Yu, Y.; Tan, T. Salient coding for image classification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Colorado Springs, CO, USA, 20–25 June 2011.
35. Rodriguez, A.; Laio, A. Clustering by fast search-and-find of density peaks. *Science* **2014**, *344*, 1492–1496. [[CrossRef](#)] [[PubMed](#)]
36. Foody, G.M. Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy. *Photogramm. Eng. Remote Sens.* **2004**, *70*, 627–633. [[CrossRef](#)]



© 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/>).