Spectral–Spatial Sparse Subspace Clustering for Hyperspectral Remote Sensing Images

Hongyan Zhang, *Member, IEEE*, Han Zhai, *Student Member, IEEE*,
Liangpei Zhang, *Senior Member, IEEE*, and Pingxiang Li, *Member, IEEE*

***Abstract*—Clustering for hyperspectral images (HSIs) is a very challenging task due to its inherent complexity. In this paper, we propose a novel spectral–spatial sparse subspace clustering (S^4C) algorithm for hyperspectral remote sensing images. First, by treating each kind of land-cover class as a subspace, we introduce the sparse subspace clustering (SSC) algorithm to HSIs. Then, considering the spectral and spatial properties of HSIs, the high spectral correlation and rich spatial information of the HSIs are taken into consideration in the SSC model to obtain a more accurate coefficient matrix, which is used to build the adjacent matrix. Finally, spectral clustering is applied to the adjacent matrix to obtain the final clustering result. Several experiments were conducted to illustrate the performance of the proposed S^4C algorithm.**

***Index Terms*—Hyperspectral image (HSI), sparse representation, spectral clustering, subspace clustering.**

I. INTRODUCTION

BY COMBINING imaging and spectroscopy technology, hyperspectral remote sensing can simultaneously acquire spatially and spectrally continuous image data. With the diagnostic spectral information, hyperspectral data are becoming a valuable tool for monitoring the Earth's surface [1]–[5] and have been widely used in various application fields, including urban planning, surveillance, agriculture, and so on. A commonly used technique in these applications is "clustering." Hyperspectral image (HSI) clustering can be defined as the process of segmenting pixels into corresponding sets which satisfy the requirement that the differences between sets are much greater than the differences within sets. As is well known to us all, HSIs are typical high-dimensional data with large spectral variability, high dimensionality, and complex structures [6]–[8], which makes HSI clustering a very challenging task.

To date, many different clustering methods for HSIs have been proposed. The existing clustering algorithms for HSIs can be coarsely divided into the following four categories: 1) centroid-based clustering methods; 2) density-based methods; 3) biological methods; and 4) spectral-based methods.

The centroid-based clustering methods, such as k-means [9], the iterative self-organizing data analysis technique algorithm [10], and fuzzy c-means (FCM) [11], are based on the fact that similar data points generate clusters in the feature space. Such approaches optimize the clusters by giving initial clustering centers and continuously updating the location of them until they minimize the sum of squared errors. Therefore, they belong, in essence, to mountain-climbing methods, which can get easily stuck in a local optimum [12], and they are sensitive to initialization and noise. In addition, they can only work well in the case that each cluster satisfies a convex "ball-like" distribution, which is usually not the case for HSIs.

The density-based methods attempt to find clusters by calculating the local densities of the feature space, with the assumption that clusters are always dense areas separated by sparse areas. The density-based methods include the clustering by fast search and find of density peaks (CFSFDP) algorithm [13], the density-based spatial clustering of applications with noise (DBSCAN) method [14], and the clustering-in-quest method [15]. Unfortunately, these methods usually fail for HSIs because it is very difficult to find the density peaks in the sparse feature space.

The biological clustering methods, such as unsupervised remote sensing image classification using an artificial immune network [16] and automatic fuzzy clustering based on adaptive multiobjective differential evolution [17], utilize biological models to cluster the HSIs. However, the clustering results are not always satisfactory because the biological models do not always exactly fit the characteristics of the HSIs.

The spectral-based methods mainly contain two steps: 1) the construction of an adjacent matrix which describes the memberships of the data points and 2) the application of a centroid clustering algorithm to the Laplacian matrix induced by it to obtain the segmentation. Specifically, the local spectral clustering-based approaches, such as local subspace affinity [18], locally linear manifold clustering [19], and spectral local best-fit flats [20], [21], use local information around each data point to build the similarity between pairs of points, but they cannot deal with the points near the intersection of two subspaces. The reason for this is that these algorithms fail to justify which subspace these points should belong to. In this case, a large error may be introduced into the clustering model. Fortunately, the global spectral clustering-based approaches, such as spectral curvature clustering [22], try to resolve this problem by building better similarities between data points using global information, but

---

Manuscript received June 22, 2015; revised October 20, 2015, December 26, 2015, and January 21, 2016; accepted January 27, 2016. Date of publication March 3, 2016; date of current version April 27, 2016. This work was supported in part by the National Natural Science Foundation of China under Grants 41571362, 61201342, and 41431175 and in part by the Fundamental Research Funds for Central Universities under Grant 2015904020202.

The authors are with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan 430079, China (e-mail: zhanghongyan@whu.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2016.2524557

they need to have some *a priori* knowledge of the subspaces, such as the number and dimensions of the subspaces, and they assume that all the subspaces have the same dimensions [23].

Unfortunately, most of the methods described earlier suffer from significant misclassification because of the uniform feature point distribution caused by the large spectral variability of HSIs [24]. In recent years, the sparse subspace clustering (SSC) algorithm has been proposed to group data points into different subspaces by finding the sparsest representation for each data point while only selecting the data points from its own subspace to represent itself [23], [25]. The SSC algorithm has been widely applied in various computer vision fields, such as face clustering, motion segmentation, and so on. In this paper, subspace clustering is realized by solving an ℓ_1-minimization problem [26], [27], whose solution is used to define the adjacent matrix, and the segmentation result is achieved by applying spectral clustering to it. For HSIs, although pixels of the same land-cover class may have different spectra because of the varying illumination, topography, and other imaging conditions, there is a high probability that they will lie in the same subspace [24]. That is to say, each kind of land-cover material can be treated as a subspace. Therefore, subspace theory can be used to model this problem. Based on this fact, it is natural to introduce the SSC algorithm to perform the HSI clustering task. However, directly applying the SSC algorithm to HSIs usually fails to take advantage of the high spectral correlation and rich spatial information of the HSIs, and the full potential of the SSC algorithm for HSI clustering is not exploited [24].

In view of this, we propose a novel spectral–spatial sparse subspace clustering ($\mathrm{S}^4\mathrm{C}$) algorithm for HSIs. The contributions of this paper are summarized as follows. First, to the best of our knowledge, by considering the pixels of the same land-cover class as lying in one independent subspace, we are the first to segment HSI pixels into different clusters with the SSC algorithm. Second, considering the working mechanism of sparse representation, the spectral-weighted sparse subspace clustering (SWSSC) model is built to ensure that each pixel can be represented by the corresponding hyperspectral signals with the highest correlation. Third, with the spectral similarity of a local neighborhood of the HSI, the rich spatial information is also incorporated into the SSC model to further improve the performance. The experimental results demonstrate that the proposed $\mathrm{S}^4\mathrm{C}$ algorithm significantly improves the clustering performance, in both the visual and quantitative evaluations.

The remainder of this paper is organized as follows. Section II briefly introduces the SSC model for HSIs. In Section III, we propose the $\mathrm{S}^4\mathrm{C}$ algorithm for HSIs, focusing on image clustering with spectral features and detailing a new way of incorporating the spatial information. The experimental results are given in Section IV. Finally, Section V concludes this paper.

II. HSI CLUSTERING VIA THE SSC MODEL

A. General Sparse Representation Model

In the general scheme of sparse representation, we consider a set of data vectors $\mathbf{Y} = [y_1, y_2, \ldots, y_K]$ in $\mathbb{R}^D$, each of which can be approximated by a linear or an affine combination of a few atoms from a given structured "dictionary" $\mathbf{A} = [\mathbf{A}_1, \ldots \mathbf{A}_i, \ldots, \mathbf{A}_l]$, $i = 1, 2, \ldots, l$ [28], [29], where D denotes the dimension of the data vectors. The dictionary $\mathbf{A}$ is made up of l subdictionaries, with $\mathbf{A}_i \in \mathbb{R}^{D\times L_i}$ representing the ith subdictionary consisting of L_i atoms, and $L = \sum_i L_i$, where L represents the total number of atoms. Therefore, each data point can be represented with the following model:

$$y_j = \mathbf{A}c_j + g_j \quad j = 1, \ldots, K \tag{1}$$

where c_j is the representation coefficient vector of signal y_j with representation error g_j. It can also be transformed into the matrix form

$$\mathbf{Y} = \mathbf{AC} + \mathbf{G} \tag{2}$$

where $\mathbf{C} = [c_1, c_2, \ldots, c_K]$ is the representation coefficient matrix and $\mathbf{G}$ is the representation error matrix. In fact, the dictionary is usually overcomplete, and hence, it leads to an ill-posed problem with many feasible solutions [30], [31]. To resolve this issue, the principle of sparsity is invoked [32], and the sparse representation coefficient matrix can be obtained by solving the following optimization problem:

$$\min \|\mathbf{C}\|_0 \text{ s.t.} \quad \mathbf{Y} = \mathbf{AC} + \mathbf{G} \tag{3}$$

where the ℓ_0-norm represents the number of nonzero components which can promote the sparsity of the solution. This problem can be easily solved with a greedy algorithm or a relaxation algorithm [32].

B. Subspace Model

"Subspace" refers to a subset of the full space, and it has a lower dimension. In recent years, subspace models have been widely used in various fields because of their excellent properties [27], [33]. A linear subspace, as the name suggests, satisfies a linear relationship and can be seen as a space spanned by some implicit orthogonal bases. In general, the full space can be treated as a union of many linear subspaces and can be modeled as follows:

$$S = \cup_{i=1}^{l} S_i \tag{4}$$

where $S \in \mathbb{R}^D$ represents the full space, $S_i \in \mathbb{R}^{d_i}$ refers to its corresponding linear subspaces, with $d_i \ll D$, $d_1 + d_2 + \cdots + d_l = D$, and l is the number of subspaces.

The affine subspace can then be seen as a linear subspace with a constraint that the sum of all the coordinates is equal to one, which comes from the fact that a d_i-dimensional affine subspace S_i can be considered as a subset of a $(d_i + 1)$-dimensional linear subspace that includes S_i and the origin [23], [34], [35]. Therefore, it can be modeled as follows:

$$S = \cup_{i=1}^{l} S_i, \mathbf{C}^T 1 = 1 \tag{5}$$

where $\mathbf{C} \in \mathbb{R}^{L\times K}$ denotes the coordinate matrix, 1 represents a vector whose elements are all one with suitable size.

For HSIs, although pixels may vary in the spectral feature space to some extent because of the changes in the external imaging conditions, such as undulating terrain, different illumination on sunny slopes and shady slopes, and so on, we can treat those pixels of the same class as lying in a single subspace. Based on this fact, it is natural to introduce the subspace model to HSIs. Considering the hyperspectral pixels with similar energy, it is more reasonable to add a coordinate sum-to-one constraint than to release their arbitrary magnitude in the union of the subspaces. It is therefore more practical to choose an affine subspace model to model HSIs. For narrative convenience, all of the following cases adopt an affine subspace model.

C. HSI Clustering via the SSC Model

This section introduces the HSI clustering scheme with the SSC model. For HSIs, every hyperspectral pixel can be denoted as a p-dimensional vector after lexicographically reordering the 3-D data cube into a 2-D matrix, where p refers to the number of bands [36]. In this way, the HSI can be denoted by a 2-D matrix $\mathbf{Y} = [\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_{MN}]$, $\mathbf{Y} \in \mathbb{R}^{p\times MN}$, where M represents the width of the HSI data and N stands for the height of the data. Then, with the hyperspectral matrix itself being used as the dictionary, the SSC model utilizes the self-expressiveness property of the data to build the sparse representation model as follows:

$$\begin{aligned}&\min\nolimits_{\mathbf{C},\mathbf{G}} \|\mathbf{C}\|_0 + \lambda\|\mathbf{G}\|_F^2\\ &\text{s.t.}\quad \mathbf{Y} = \mathbf{YC} + \mathbf{G},\ \mathrm{diag}(\mathbf{C}) = 0,\ \mathbf{C}^T 1 = 1\end{aligned} \tag{6}$$

where $\mathbf{C} \in \mathbb{R}^{MN\times MN}$ refers to the representation coefficient matrix, $\mathbf{G}$ stands for the representation error, and the parameter λ is utilized to tradeoff the relative contribution between the sparsity of the coefficient and the magnitude of the noise. The constraint $\mathrm{diag}(\mathbf{C}) = 0$ is used to eliminate the trivial solution of writing a point as an affine combination of itself [23], [25]. In addition, the constraint $\mathbf{C}^T 1 = 1$ ensures that it is a case of an affine subspace.

Unfortunately, (6) is a nonconvex optimization problem, so there is no unique and stable solution. We can obtain a tractable convex optimization problem by relaxing (6) and replacing the ℓ_0-norm with the ℓ_1-norm, which yields the following convex surrogate:

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

where the ℓ_1-norm regularization in this formulation suggests that the signal can be faithfully represented with atoms from its own subspace [37]–[39]. The optimization problem in (7) can be solved by the alternating direction method of multipliers (ADMM), which is introduced in detail in the Appendix.

Next, the obtained sparse coefficient matrix $\mathbf{C}$ can be adopted to construct the adjacent matrix $\tilde{\mathbf{W}}$, which defines the weight on the edge between the data nodes in the following way:

$$\tilde{\mathbf{W}} = |\mathbf{C}| + |\mathbf{C}|^T. \tag{8}$$

That is, each node i connects itself to a node j by an edge whose weight is equal to $|\mathbf{C}_{i,j}| + |\mathbf{C}_{j,i}|$. The reason for the symmetrization is that, in general, a hyperspectral pixel $y_i \in S_i$ can be represented as an affine combination of certain points, including $y_j \in S_i$ [23]. However, y_j may not necessarily choose y_i in the sparse representation process. By this particular choice of weight, we make sure that nodes i and j become connected to each other if either y_i or y_j is used in the sparse representation of the other [19]. Finally, the clustering result is obtained by applying spectral clustering to the Laplacian matrix induced by the adjacent matrix, which is also called the "similarity graph" in [40]–[44]. The SSC algorithm for HSIs can be summarized as Algorithm 1.

Algorithm 1 SSC algorithm

```
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: the cluster number $l$ and the regularization
    parameter $\lambda$.
Main algorithm:
(1) Construct the sparse representation model (7) and re-
    solve it to obtain the sparse coefficient matrix $\mathbf{C}$;
(2) Normalize the columns of $\mathbf{C}$ with $\mathbf{C}_i \leftarrow (\mathbf{C}_i/\|\mathbf{C}_i\|_\infty)$;
(3) Construct the similarity graph with (8);
(4) Apply spectral clustering to the similarity graph.
Output:
  A 2-D matrix which records the labels of the clustering
  result of the HSI.
```

III. S^{4}C FOR HSIs

Directly applying the SSC algorithm to HSIs usually fails to achieve a satisfactory performance because it does not fully exploit the spectral and spatial information of the HSIs with the great potential of the SSC model. Based on this fact, in this section, we improve the performance of SSC from the spectral and spatial perspectives, respectively. First, we are committed to obtaining a more accurate representation coefficient matrix by adding a spectral constraint to ensure that the highly correlated pixels of the HSI participate in the representation process. Second, we incorporate spatial neighborhood information into the SSC scheme to utilize the spatial similarity of the spectral signatures. Finally, we explore the spectral–spatial structural information and propose the S^{4}C algorithm for HSIs.

A. SWSSC

In general, because of the high correlation between hyperspectral pixels, it is usually difficult for SSC to choose the corresponding pixels to represent the target signal in the process of image data self-expressiveness, which degrades the performance of the sparse representation. Therefore, we aim to add some necessary constraints from the spectral perspective in a reasonable way to make sure that the signals with a higher correlation are preferred in the sparse representation process. In this way, a more accurate coefficient matrix can be achieved, which is critical for yielding higher clustering accuracies [24].

Based on the fact that similar data points have a high probability of sharing the same subspace, the SWSSC model is built by imposing greater weights on similar data points and smaller weights on dissimilar ones. Thus, the representation coefficient matrix $\mathbf{C}$ for HSIs can be achieved by solving the following weighted sparse optimization problem:

$$\begin{aligned}&\min_{\mathbf{C},\mathbf{G}} \|\mathbf{W}\mathbf{C}\|_1 + \frac{\lambda}{2}\|\mathbf{G}\|_F^2\\ &\text{s.t.} \quad \mathbf{Y} = \mathbf{Y}\mathbf{C} + \mathbf{G},\ \operatorname{diag}(\mathbf{C}) = 0,\ \mathbf{C}^T 1 = 1\end{aligned} \tag{9}$$

where $\mathbf{W}$ is a weighted matrix added to the spectra.

Considering that the function of the weighted matrix is to choose the more highly correlated hyperspectral pixels to represent the target signals, the more highly correlated pixels should be given larger weights. Inspired by the centroid clustering-based methods, hyperspectral pixels within a shorter distance have a higher correlation in the feature space [45]. Therefore, it is natural to define the weighted matrix $\mathbf{W}$ in the following way:

$$\mathbf{W}_{i,j} = \frac{1}{\|y_i - y_j\|_2^2 + \gamma} i,\ j = 1, 2, \ldots, MN \text{ and } i \neq j \tag{10}$$

where y_i and y_j are vectors corresponding to the ith pixel and the jth pixel of a given HSI. γ is a very small constant to avoid the extreme case of the weight of two very similar neighboring pixels being superlarge and is set as 0.001 in practice. For convenience, the distance between data point i and data point j is calculated in the form of the Euclidean distance. It should be noted that the diagonal elements of the matrix $\mathbf{W}$ are all zero to eliminate the trivial solution of writing a point as an affine combination of itself.

Then, in order to guarantee that all the weights will fall into the range of zero to one, a normalization process for the weighed matrix $\mathbf{W}$ is conducted as follows. Through this normalization process, the influence caused by magnitude can be removed

$$\mathbf{W}_{i,j} = \frac{\mathbf{W}_{i,j}}{\mathbf{W}_{i,\cdot}},\ i = 1, 2, \ldots, MN \tag{11}$$

where $\mathbf{W}_{i,\cdot}$ represents the sum of the ith row of the weighted matrix $\mathbf{W}$.

B. SSC-S

As is well known to us all, neighboring pixels in an HSI usually consist of similar materials that have a very high probability of belonging to the same land-cover class [46]. According to the scheme of SSC, in the sparse representation procedure, adjacent pixels within a local window tend to select the same sparse basis to represent them, and the sparse coefficients will be very close to each other. Considering that accidental error is inevitable in the sparse representation procedure, it is necessary to incorporate the spatial neighborhood information into the SSC procedure to support the analysis performance [24]. Therefore, we propose to construct an SSC model incorporating spatial information (SSC-S) for HSIs, which introduces the mean constraint of the representation coefficients to correct the representation bias and obtain a more accurate representation coefficient matrix.

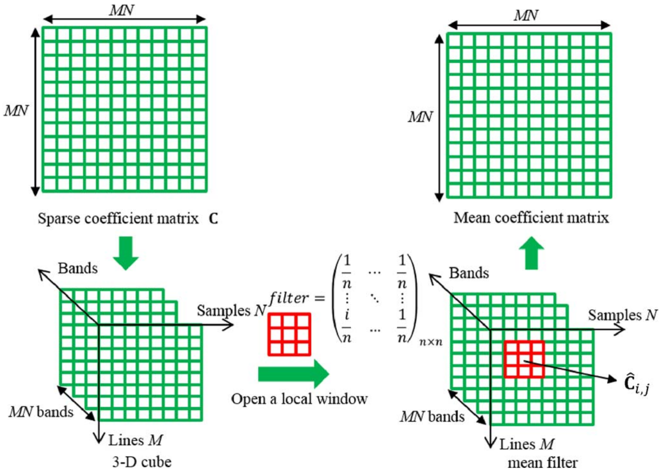


Fig. 1. Calculation progress of $\bar{\mathbf{C}}$. First, the sparse coefficient matrix $\mathbf{C}$ is reshaped to be a 3-D cube with M lines, N samples, and MN bands. Second, a local window with the size of $n \times n$ is opened and is used to calculate the mean of the window for each point. Finally, the mean 3-D cube is reshaped to a 2-D matrix with the same size as the original sparse coefficient matrix, to obtain the mean coefficient matrix $\bar{\mathbf{C}}$.

By opening a local window, the spatial information can be effectively incorporated into the SSC scheme by assuming that the representation coefficient of each hyperspectral pixel can be approximated by the mean of the coefficients of its neighboring pixels. Specifically, we first reshape the 2-D sparse coefficient matrix $\mathbf{C} \in \mathbb{R}^{MN\times MN}$ to a 3-D cube $\tilde{\mathbf{C}} \in \mathbb{R}^{M\times N\times MN}$ with the same form as the original HSI cube, to treat each coefficient vector as a "pixel" in the 3-D cube, where M represents the width of the image, N stands for the length of the image, and MN is the number of bands. It is convenient to calculate the mean representation coefficient of the spatial neighborhood within this notation framework. We then open a local window with the size of $n \times n$ at each coefficient vector and limit the difference between it and the mean of the neighboring pixels by $\|\mathbf{C} - \bar{\mathbf{C}}\|_F^2 < \varepsilon$, where ε is the restriction and $\bar{\mathbf{C}} \in \mathbb{R}^{MN\times MN}$ is the mean coefficient matrix achieved by reshaping the mean 3-D cube $\hat{\mathbf{C}} \in \mathbb{R}^{M\times N\times MN}$ to a 2-D matrix, which can be calculated in the following formulation:

$$\hat{\mathbf{C}}_{i,j} = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\tilde{\mathbf{C}}_{i,j}). \tag{12}$$

The calculation process of $\bar{\mathbf{C}}$ is shown in Fig. 1.

In this way, we can utilize $\bar{\mathbf{C}}$ to regularize $\mathbf{C}$, and the representation coefficient matrix can be obtained by solving the following optimization problem:

$$\begin{aligned}&\min_{\mathbf{C},\mathbf{G},\bar{\mathbf{C}}} \|\mathbf{C}\|_1 + \frac{\lambda}{2}\|\mathbf{G}\|_F^2 + \frac{\alpha}{2}\|\mathbf{C} - \bar{\mathbf{C}}\|_F^2\\ &\text{s.t.} \quad \mathbf{Y} = \mathbf{Y}\mathbf{C} + \mathbf{G},\ \operatorname{diag}(\mathbf{C}) = 0,\ \mathbf{C}^T 1 = 1\end{aligned} \tag{13}$$

where λ and α are both regularization parameters and α denotes the relative contribution of the spatial constraint term.

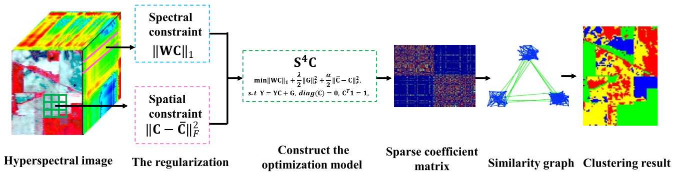


Fig. 2. Flowchart of the S^{4}C algorithm.

C. *S^{4}C Model and Flowchart*

In order to take full advantage of the different properties of the HSIs from the spectral and spatial perspectives, we propose a novel S^{4}C algorithm for HSIs by integrating the spectral and spatial constraint terms into a unified framework. The S^{4}C algorithm model can be formulated as

$$\min_{\mathbf{C},\mathbf{G},\bar{\mathbf{C}}} \|\mathbf{WC}\|_1 + \frac{\lambda}{2}\|\mathbf{G}\|_F^2 + \frac{\alpha}{2}\|\mathbf{C}-\bar{\mathbf{C}}\|_F^2$$
$$\text{s.t.}\quad \mathbf{Y}=\mathbf{YC}+\mathbf{G},\ \operatorname{diag}(\mathbf{C})=0,\ \mathbf{C}^T\mathbf{1}=1. \tag{14}$$

The sparse optimization problem in (14) can be solved with the ADMM [47]–[51], more details of which are given in the Appendix. The adjacent matrix can then be constructed with the resolved coefficient matrix $\mathbf{C}$ in the same way as SSC. After applying the spectral clustering algorithm to the Laplacian matrix induced by the adjacent matrix, the final clustering results can be obtained.

The proposed S^{4}C algorithm can be summarized as Algorithm 2.

Algorithm 2 S^{4}C algorithm for HSIs

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: the cluster number l, the size of the local window n, regularization parameters λ and α.

Main algorithm:

(1) Construct the sparse representation model (14) and resolve it to obtain the sparse coefficient matrix $\mathbf{C}$ using ADMM;

(2) Normalize the columns of $\mathbf{C}$ as $\mathbf{C}_i \leftarrow (\mathbf{C}_i/\|\mathbf{C}_i\|_\infty)$;

(3) Construct the similarity graph with (8);

(4) Apply spectral clustering to the similarity graph.

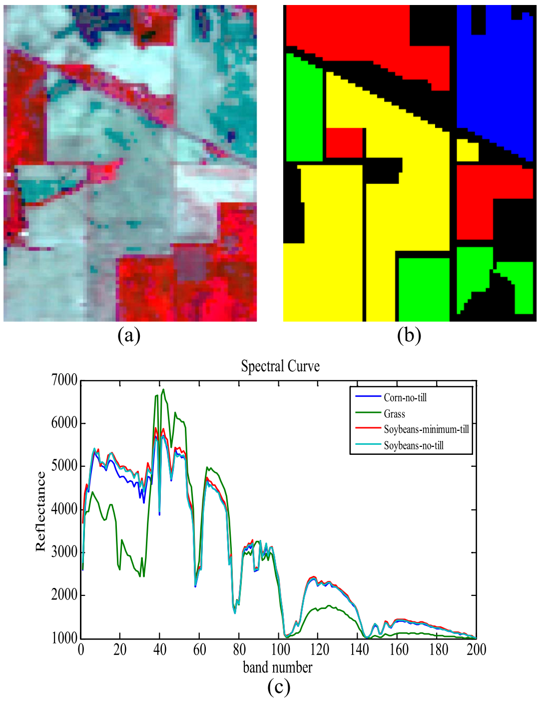
Output:

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

The flowchart of the proposed S^{4}C algorithm is illustrated in Fig. 2.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

A. *Experimental Setting*


(a) (b)

(c)

Fig. 3. AVIRIS Indian Pines image. (a) False-color image (RGB 40, 30, 20). (b) Ground truth. (c) Spectral curves of the four land-cover classes.

TABLE I
PARAMETERS OF EACH CLUSTERING METHOD FOR THE INDIAN PINES IMAGE

Methods	Parameters
FCM	$l=4,\ e=4$
FCM_S1	$l=4,\ e=4,\ \alpha=0.05$
CFSFDP	---
SSC	$l=4,\ \lambda=7.76\times10^{-7}$
SWSSC	$l=4,\ \lambda=7.76\times10^{-7}$
SSC-S	$l=4,\ \lambda=7.76\times10^{-7},\ \alpha=0.7\times10^{3}$
S^{4}C	$l=4,\ \lambda=7.76\times10^{-7},\ \alpha=0.13\times10^{3}$

l : the number of clusters; e : fuzzy exponential; λ : (sparsity/noise tradeoff) parameter; α : (spectral/spatial tradeoff) parameter.

To thoroughly evaluate the performance of the proposed algorithm, FCM [11], FCM with spatial information (FCM_S1) [52], the original SSC, and CFSFDP [13] were used as benchmarks. In addition, two simplified versions of the proposed S^{4}C method, SWSSC and SSC-S, were also included in the comparisons. CFSFDP is a powerful clustering algorithm recently published in Science, and the code is available online [53].

Three hyperspectral data sets with different imaging environment settings were used to validate the performance of the proposed method: 1) the Indian Pines data set; 2) the University of Pavia data set; and 3) the Salinas data set. The number of clusters was set as a manual input, and the parameters of each algorithm were manually adjusted to the optimum. Both the visual clustering results and quantitative evaluations (including producer's accuracy, user's accuracy, overall accuracy (OA), and kappa coefficient) are given for each experiment.

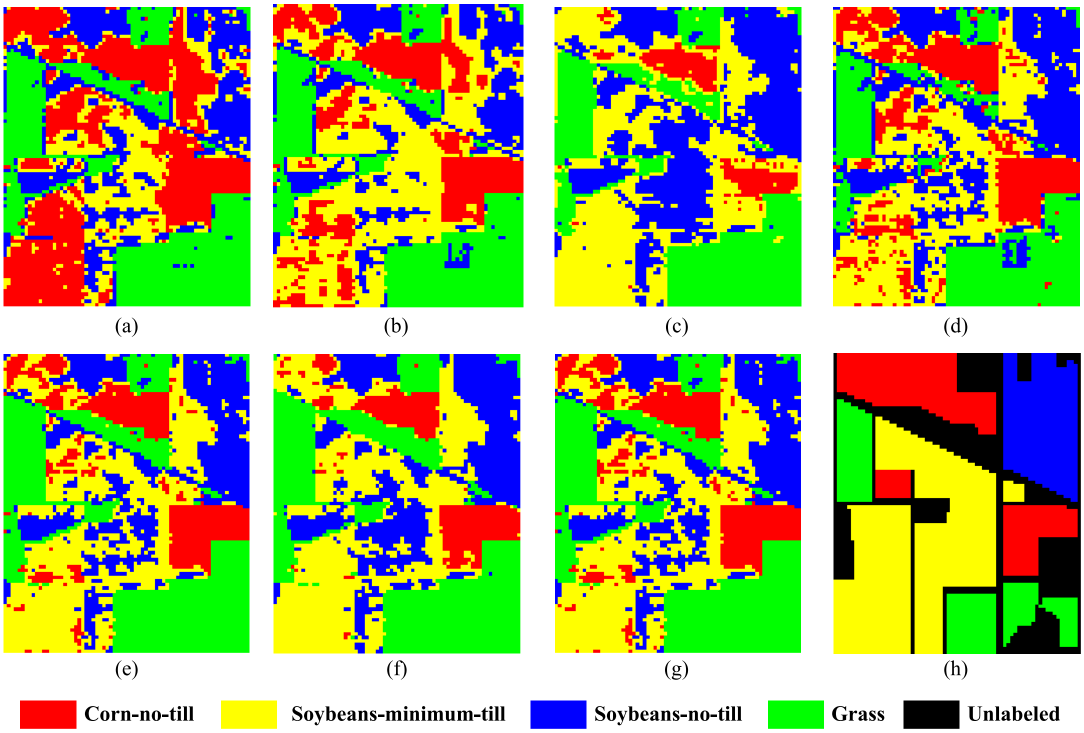


Fig. 4. Cluster maps of the different methods with the Indian Pines image: (a) FCM, (b) FCM_S1, (c) CFSFDP, (d) SSC, (e) SWSSC, (f) SSC-S, (g) S^4C, and (h) the ground truth.

TABLE II
QUANTITATIVE EVALUATION OF THE DIFFERENT CLUSTERING ALGORITHMS FOR THE INDIAN PINES IMAGE

Method	Class	FCM	FCM_S1	CFSFDP	SSC	SWSSC	SSC-S	S^4C
Producer's accuracy (%)	Corn-no-till	**71.14**	63.93	31.84	67.16	61.00	58.05	61.00
	grass	99.04	95.07	**100**	92.19	**100**	**100**	**100**
	Soybeans-no-till	44.95	44.95	**76.09**	66.12	65.30	68.85	65.30
	Soybeans-minimum-till	33.94	62.09	55.25	58.26	64.92	64.76	**65.28**
User's accuracy (%)	Corn-no-till	40.17	70.45	**99.38**	68.46	82.50	97.52	83.06
	grass	97.05	97.06	97.33	**98.68**	95.42	95.80	95.42
	Soybeans-no-till	35.26	38.34	34.17	40.03	40.75	38.15	**40.85**
	Soybeans-minimum-till	69.99	58.21	62.94	**74.04**	73.04	69.84	73.11
	OA	0.5511	0.6552	0.6081	0.6725	0.6992	0.6812	**0.7008**
	Kappa	0.4019	0.5145	0.4589	0.5488	0.5805	0.5545	**0.5825**

B. AVIRIS Data Set: Indian Pines Image

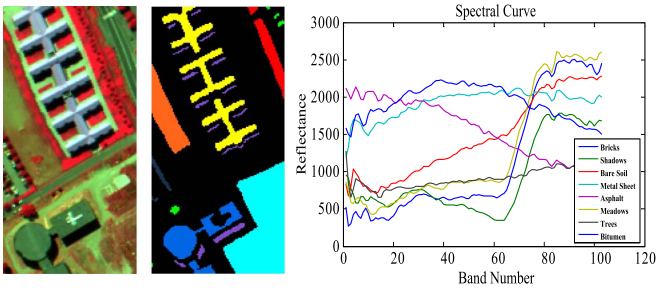
Experiment 1 was conducted using part of the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data set from the Northwestern Indiana Indian Pines test site in June 1992 [16], [17]. The size of this image is 145 × 145. A total of 20 water absorption and noisy bands (104–108, 150–163, and 220) were removed from the original 220 bands, leaving 200 spectral features for the experiment [16], [24]. Considering the computational efficiency, we cut a subimage with the size of 85 × 70, which included four main land-cover classes: corn-no-till, grass, soybeans-no-till, and soybeans-minimum-till. The clustering was a challenging task because the spectral signatures of the land-cover classes in this area are very similar and some of the spectral curves are seriously mixed, as shown in Fig. 3(c). The false-color image and the ground truth are provided in Fig. 3(a) and (b).

The parameters of each clustering method were set as shown in Table I. The cluster maps of the various clustering methods are shown in Fig. 4(a)–(g), and the corresponding quantitative evaluation of the clustering results is provided in Table II. In the table, the optimal value of each row is shown in bold, and the second best results are underlined. From Fig. 4 and Table II, it can be clearly observed that the clustering result of FCM is very poor and contains significant amounts of misclassifications, particularly for the soybeans-minimum-till class. In the lower left of the cluster map, the majority of the soybeans-minimum-till class has been misclassified into corn-no-till, with a poor precision of 33.94%. Compared with FCM, FCM_S1 improves the clustering accuracy to a large degree by making use of the spatial neighborhood information. For the soybeans-minimum-till class, the misclassification is significantly reduced, and FCM_S1 achieves a higher precision of 62.09%. However, the OA of FCM_S1 is still low due to the misclassification of the other classes. For the CFSFDP method, the soybeans-no-till and soybeans-minimum-till classes are not well separated and are frequently misclassified, and only a few parts of the corn-no-till class are successfully recognized; however, it does return a good performance for the grass class.

We now turn to the clustering results of the four subspace-based methods. Compared with CFSFDP, SSC performs much better for the corn-no-till class. Meanwhile, it effectively separates the soybeans-no-till and soybeans-minimum-till classes. However, it performs poorly for the grass class. It can be clearly seen that SWSSC obtains a better visual cluster map than SSC by effectively discriminating the three classes of corn-no-till, soybeans-no-till, and soybeans-minimum-till. The quantitative evaluation also confirms this. For example, the precision of the grass and soybeans-minimum-till classes is improved from 92.19% and 58.26% to 100% and 64.92%, respectively. The main reason for this improvement is that, by adding a weighted matrix to the image pixels, SWSSC ensures that the signals with high correlation are preferentially selected in the sparse representation process, thereby achieving a more accurate coefficient matrix, which is critical for yielding higher clustering accuracies. SSC-S can also effectively improve the performance of SSC because of the incorporation of the spatial information to correct the sparse representation bias. It can be clearly seen that SSC-S can significantly smooth the noise existing in the grass and soybeans-minimum-till classes in the left bottom of the image. For the grass, soybeans-no-till, and soybeans-minimum-till classes, the precision is greatly improved. However, SSC-S does produce some misclassifications, with a part of the corn-no-till class being misclassified as soybeans-minimum-till in the left-top corner of the image. Finally, the proposed S^{4}C method obtains the highest precision, with the best OA of 70.08% and kappa coefficient of 0.5825. It can effectively distinguish all four classes and can smooth the noise within each class. The S^{4}C method further improves the performance of SWSSC and SSC-S by integrating the constraints from the spectral and spatial perspectives of the HSI, and it fully exploits the potential of the SSC algorithm. Overall, the proposed S^{4}C method achieves the best performance for the HSI clustering task, both visually and quantitatively.

C. *ROSIS Urban Data: University of Pavia, Italy*

This scene was acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor during a flight campaign over Pavia, Northern Italy. The size of the image is 610 × 340, with 103 bands used in the experiment. The geometric resolution is 1.3 m. As with the former experiment, we took a typical area for the test data with a size of 200 × 100, containing eight main land-cover classes: metal sheet, asphalt, meadows, trees, bare soil, bitumen, bricks, and shadows [24]. This data set contains more complex land-cover classes, and the spectral signatures of some of the classes are very similar, which results in the clustering being a more challenging task. The spectral curves of the eight land-cover classes are shown in Fig. 5. The false-color image and the ground truth are also provided.



(a) (b) (c)

Fig. 5. ROSIS University of Pavia image. (a) False-color image (RGB 102, 56, 31). (b) Ground truth. (c) Spectral curves of the eight land-cover classes.

TABLE III
PARAMETERS OF EACH CLUSTERING METHOD FOR THE UNIVERSITY OF PAVIA IMAGE

Methods	Parameters
FCM	$l=8,\ e=2$
FCM_S1	$l=8,\ e=2,\ \alpha=0.5$
CFSFDP	---
SSC	$l=8,\ \lambda=1.35\times10^{-5}$
SWSSC	$l=8,\ \lambda=1.35\times10^{-5}$
SSC-S	$l=8,\ \lambda=1.35\times10^{-5},\ \alpha=1.40\times10^{3}$
S^{4}C	$l=8,\ \lambda=1.35\times10^{-5},\ \alpha=2.80\times10^{3}$

l : the number of clusters ; e : fuzzy exponential; λ : (sparsity/noise tradeoff) parameter; α : (spectral/spatial tradeoff) parameter.

The parameters of each clustering method were set as shown in Table III. From the cluster maps shown in Fig. 6 and the quantitative evaluation in Table IV, it can be clearly observed that FCM and SSC obtain poor clustering results containing a lot of salt-and-pepper noise and significant amounts of misclassifications, with low OAs of 51.59% and 43.97%, respectively. FCM_S1 improves the clustering accuracy of FCM to a certain extent. CFSFDP generates a smooth clustering result and performs better than SSC in this scene, with an OA of 52.12%. However, there are still some misclassifications in the CFSFDP result, such as the meadow and bitumen classes not being effectively recognized, and the majority of the bare soil class is misclassified as asphalt. Compared with SSC, SWSSC performs better by significantly decreasing the misclassifications to achieve an improvement of almost 15% in OA. It can also be seen that SSC-S improves the performance of SSC by effectively smoothing the noise within each class to obtain the 8.07% improvement in OA. Therefore, both SWSSC and SSC-S perform significantly better than SSC. The proposed S^{4}C method obtains the best visual result and the highest accuracy by effectively distinguishing most of the land-cover classes, with little salt-and-pepper noise within each class. It should be noted that the meadow class is effectively distinguished to a certain extent, while the recognition level is zero for the other methods. In addition, the bricks class is also distinguished well, with an accuracy of nearly 100%, which is beyond the ability of the other methods.

D. *ROSIS Urban Data: University of Pavia, Italy*

The third hyperspectral data set was acquired by the 224-band AVIRIS sensor over the Salinas Valley, CA, USA. The size of the image is 512 × 217. As with the first experiment, a total of 20 water absorption bands (108–112, 154–167, and 224) were

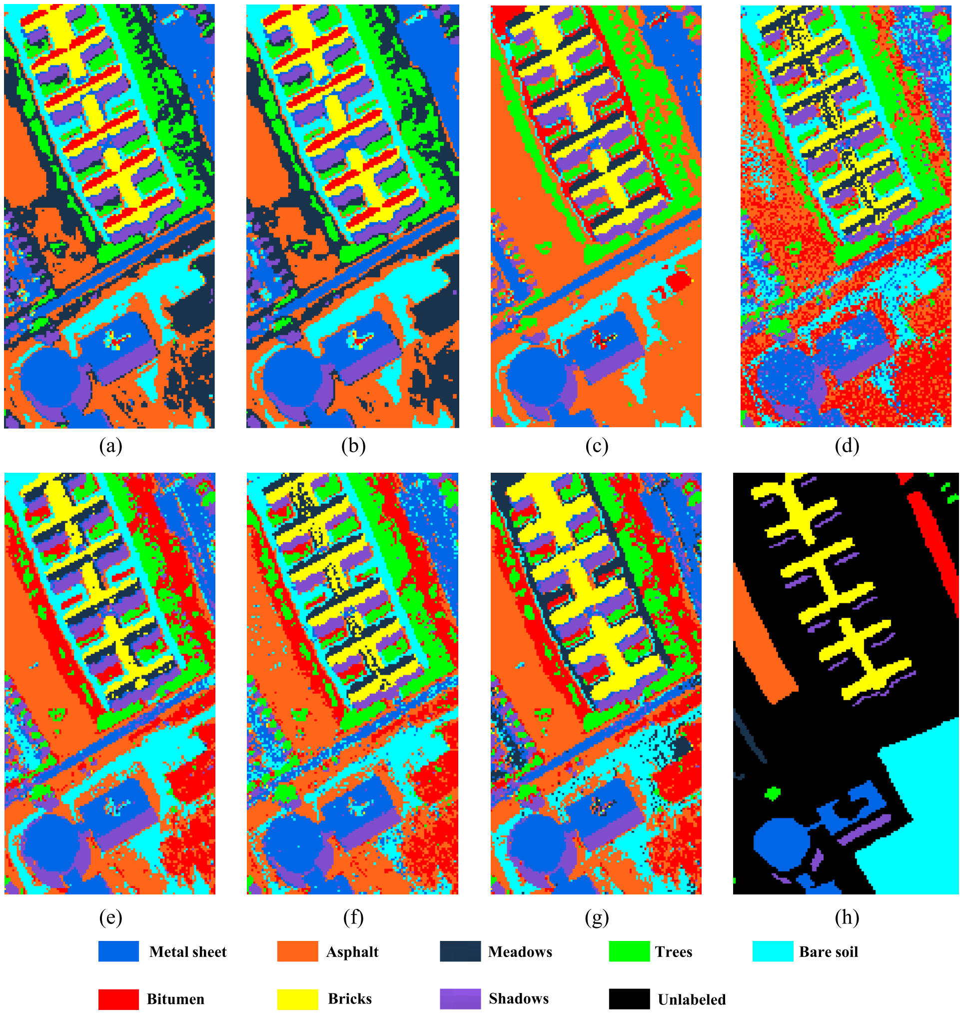
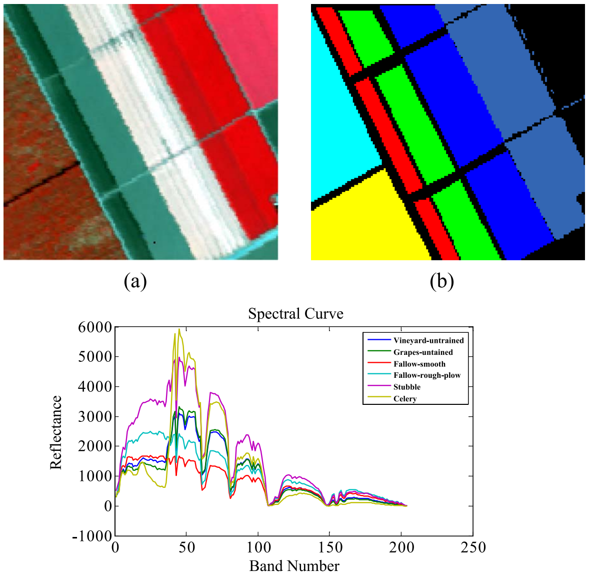

(a) (b) (c) (d)
(e) (f) (g) (h)

Fig. 6. Cluster maps of the different methods with the University of Pavia image: (a) FCM, (b) FCM_S1, (c) CFSFDP, (d) SSC, (e) SWSSC, (f) SSC-S, (g) S^{4}C, and (h) the ground truth.

TABLE IV
QUANTITATIVE EVALUATION OF THE DIFFERENT CLUSTERING ALGORITHMS FOR THE UNIVERSITY OF PAVIA IMAGE

Method	Class	FCM	FCM_S1	CFSFDP	SSC	SWSSC	SSC-S	S^{4}C
Producer's accuracy (100%)	Metal sheet	**100**	99.72	**100**	85.58	98.61	98.60	99.09
	Asphalt	60.32	58.73	74.60	52.60	87.30	**95.83**	87.30
	Meadows	0	0	0	0	0	0	**60.64**
	Trees	86.33	87.63	**100**	**100**	98.14	**100**	98.61
	Bare soil	23.29	24.97	20.59	22.04	**34.90**	25.79	31.93
	Bitumen	0	0	0	**0.71**	0	0	0
	Bricks	61.37	60.99	60.61	53.84	63.88	52.32	**98.37**
	Shadows	**100**	**100**	99.87	98.61	99.35	99.45	99.09
User's accuracy (100%)	Metal sheet	98.63	**98.90**	63.28	48.46	77.39	63.19	52.41
	Asphalt	90.48	**100**	73.44	42.84	60.44	45.10	57.29
	Meadows	0	0	0	0	0	0	**22.44**
	Trees	41.26	41.93	**100**	65.13	88.59	57.80	80.73
	Bare soil	98.03	96.53	**99.26**	64.75	72.14	87.19	96.12
	Bitumen	0	0	0	**0.21**	0	0	0
	Bricks	99.63	99.75	99.62	**99.86**	88.59	99.57	71.57
	Shadows	63.42	63.00	28.23	58.65	51.62	90.89	**99.77**
OA		0.5159	0.5230	0.5212	0.4397	0.5820	0.5204	**0.6509**
Kappa		0.4425	0.4486	0.4421	0.3495	0.5053	0.4419	**0.5852**

removed. We took a typical part of the image as the test area, with the size of 140 × 150, containing six main land-cover classes: vineyard untrained, grapes untrained, fallow smooth, fallow rough plow, stubble, and celery. This image is also difficult for clustering because of the high similarity between the spectral signatures of some land-cover classes, which can be seen in Fig. 7(c). The false-color image and the ground truth are provided in Fig. 7(a) and (b), respectively.



(a) (b)

(c)

Fig. 7. AVIRIS Salinas image. (a) Salinas false-color image (RGB 70, 27, 17). (b) Ground truth. (c) Spectral curves of the six land-cover classes.

TABLE V
PARAMETERS OF EACH CLUSTERING METHOD FOR THE SALINAS IMAGE

Methods	Parameters
FCM	$l=6,\ e=2$
FCM_S1	$l=6,\ e=2,\ \alpha=0.3$
CFSFDP	---
SSC	$l=6,\ \lambda=4.57\times10^{-6}$
SWSSC	$l=6,\ \lambda=4.57\times10^{-6}$
SSC-S	$l=6,\ \lambda=4.57\times10^{-6},\ \alpha=5.04\times10^{3}$
S^{4}C	$l=6,\ \lambda=4.57\times10^{-6},\ \alpha=4.20\times10^{3}$

l: the number of clusters; e: fuzzy exponential; λ: (sparsity/noise tradeoff) parameter; α: (spectral/spatial tradeoff) parameter.

The parameters were set as shown in Table V. From Fig. 8 and Table VI, it can be seen that FCM and SSC achieve inferior clustering results with a low degree of accuracy and CFSFDP achieves a better clustering result by successfully distinguishing most of the land-cover classes. FCM_S1 performs better than FCM. It can be clearly seen that SWSSC performs significantly better than SSC and obtains a higher accuracy. Meanwhile, the improvement in the performance of SSC-S over that of SSC is limited because, in this scene, all the land-cover classes are very smooth with concentrated distributions and less noise, which limits the function of the spatial information. Again, the proposed S^{4}C further improves the performances of SWSSC and SSC-S to achieve the best clustering result, both visually and quantitatively, with the best OA of 86.31% and kappa coefficient of 0.8312.

E. Parameter Analysis and Convergence Speed Comparison

1) Regularization Parameter λ: The regularization parameter λ acts as the tradeoff between the sparsity of the coefficient and the magnitude of the noise. It is utilized to balance the sparsity of the coefficient and the data fidelity. When λ is set too high, the data fidelity can be effectively guaranteed, but the sparsity of the coefficient is weakened. When λ is set too low, the sparsity of the coefficient can be effectively promoted, but the data fidelity cannot be guaranteed. Parameter λ is decided by the following formulation [19]:

$$\lambda = \frac{\beta}{\mu} \tag{15.a}$$

$$\mu \triangleq \min_{i} \max_{j\neq i} \left| y_i^T y_j \right| \tag{15.b}$$

where β is the adjustment coefficient and μ is a parameter related to the data set, which can be explicitly determined.

From (15.a), it can be easily concluded that the sensitivity of λ is decided by β and μ. In order to analyze the sensitivity of β, three groups of experiments were conducted on the Indian Pines image, the University of Pavia image, and the Salinas image, respectively. Fig. 9 shows the change in the OA of S^{4}C corresponding to different β values, with the other parameters fixed. The horizontal axis denotes the variation range of β, and the curves illustrate the change trend of the OA. From the results of the experiments, it can be seen that β is a relatively stable parameter with respect to different HSI data sets. That is to say, it is quite robust to the data set, and the optimal value always falls in a very narrow range of $[1, 2] \times 10^3$. Meanwhile, μ can be explicitly computed with (15.b), and it largely depends on the characteristics of the image data set and not the size of the image. Therefore, λ is adaptive to the data sets and can be easily fine-tuned. In other words, the change magnitude of the optimal values of both the parameters is very small, and they always fall into a similar narrow range.

2) Regularization Parameter α: The regularization parameter α denotes the weight of the spatial information in S^{4}C, which is a tradeoff between the spectral constraint term and the spatial constraint term. In order to analyze the sensitivity of α, a similar strategy was adopted, i.e., three groups of experiments were conducted. The change in the OA of the proposed S^{4}C algorithm corresponding to different α values, with the other parameters fixed, is shown in Fig. 10. As can be seen from Fig. 10, the precision changes significantly with different values of α, which suggests that the spatial information plays a very important role in the clustering process. However, it can also be seen that this parameter is also quite robust with respect to different data sets and is relatively stable. The optimal value of α changes in a very narrow range of $(0, 10] \times 10^3$, which means that it has a relatively strong generalization capability.

3) Analysis of the Window Size: Fig. 11 shows the change in the OA of the proposed S^{4}C algorithm corresponding to different window sizes for the three data sets. It can be clearly seen that the accuracy decreases with the increase of parameter n, and the best performance is achieved at $n = 3$. The reason for this is that, as the size of the window becomes larger, the pixels in the window will no longer be pure. This leads to a larger bias between the mean of the window and the value of the center pixel of the window. Therefore, when the spatial information of the mean constraint is incorporated in the sparse representation procedure, the bias is enlarged instead of being reduced.

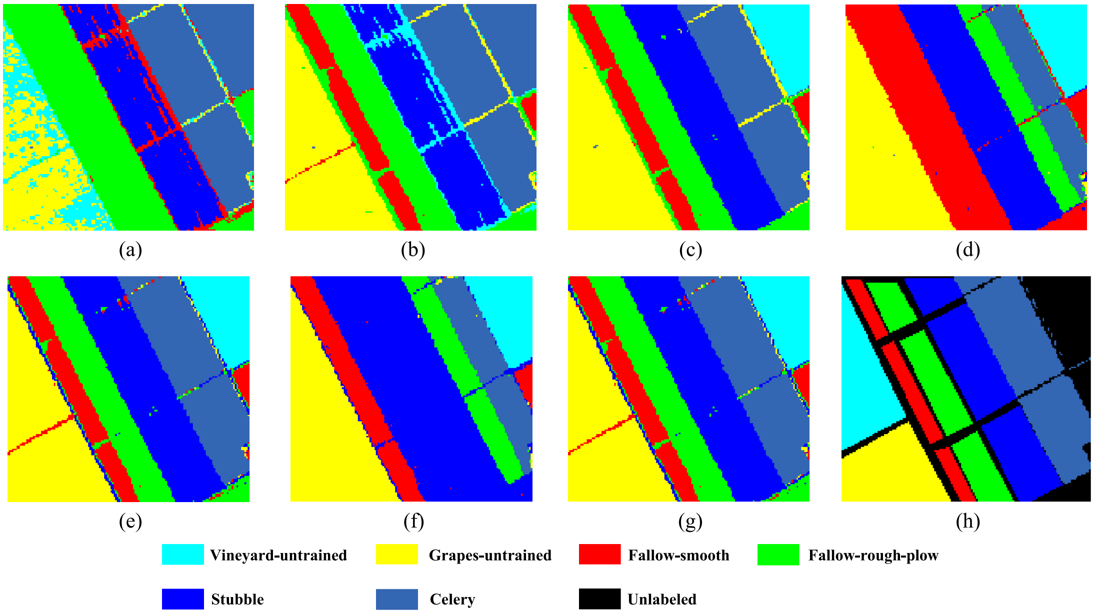
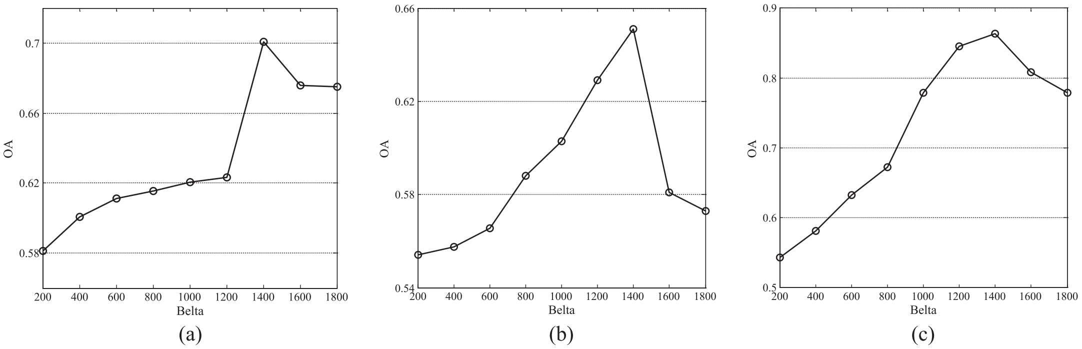


Fig. 8. Cluster maps of the different methods with the Salinas image: (a) FCM, (b) FCM_S1, (c) CFSFDP, (d) SSC, (e) SWSSC, (f) SSC-S, (g) S^{4}C, and (h) the ground truth.

TABLE VI
QUANTITATIVE EVALUATION OF THE DIFFERENT CLUSTERING ALGORITHMS FOR THE SALINAS IMAGE

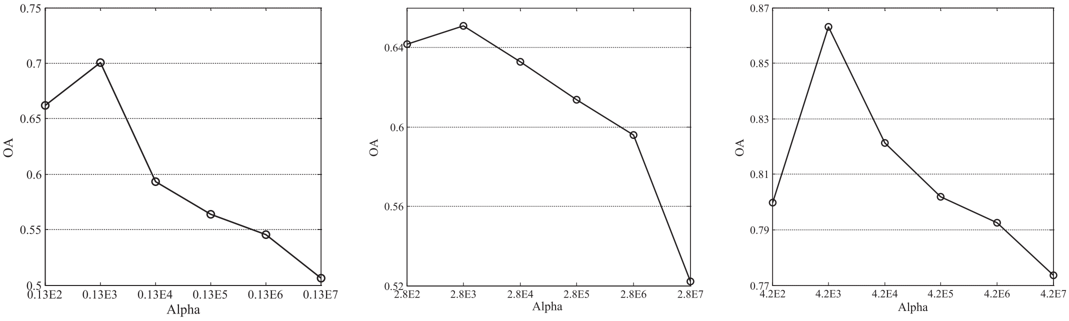
Method	Class	FCM	FCM_S1	CFSFDP	SSC	SWSSC	SSC-S	S^{4}C
Producer's accuracy (100%)	Vineyard-untrained	**43.33**	0	0	0	0	0	0
	Grapes-untrained	74.70	99.82	**100**	**100**	**100**	**100**	**100**
	Fallow-smooth	0	99.26	99.22	**99.95**	99.26	99.92	99.26
	Fallow-rough-flow	**99.96**	99.27	99.84	0	99.69	0	99.59
	Stubble	87.67	87.82	99.37	**100**	99.57	99.92	99.77
	Celery	99.24	99.27	99.38	54.13	**99.84**	55.07	**99.84**
User's accuracy (100%)	Vineyard-untrained	**61.58**	0	0	0	0	0	0
	Grapes-untrained	**58.83**	51.59	51.90	33.37	52.05	51.62	52.01
	Fallow-smooth	0	97.04	97.78	**99.55**	99.51	94.03	99.51
	Fallow-rough-flow	65.66	65.72	98.55	0	**99.83**	0	99.22
	Stubble	**100**	**100**	99.97	57.13	99.17	62.32	99.86
	Celery	99.89	**100**	99.29	**100**	96.84	51.62	97.15
	OA	0.7750	0.8311	0.8616	0.6023	0.8628	0.6043	**0.8631**
	Kappa	0.7226	0.7927	0.8293	0.5213	0.8309	0.5076	**0.8312**



(a) (b) (c)

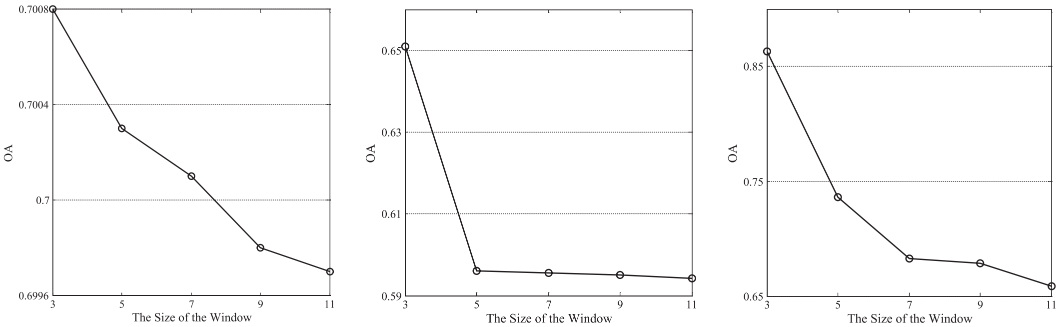
Fig. 9. Analysis of parameter β: Change in the OA with various values of β. (a) Indian Pines image. (b) University of Pavia image. (c) Salinas image.

4) Convergence Speed Comparison: Table VII illustrates the time consumption of the four subspace-based methods (SSC, SWSSC, SSC-S, and S^{4}C), FCM_S1, and CFSFDP for all three HSIs (Indian Pines image, Salinas image, and the University of Pavia image). It can be seen that, by adding the spectral constraint and incorporating the spatial information into the SSC model, SWSSC and SSC-S not only perform significantly better than SSC, but the convergence speed is also accelerated. The proposed S^{4}C method, which exploits the spectral–spatial structure information of the HSIs, also has a faster convergence speed than SSC. Although the proposed algorithms are a little slower than FCM_S1 and CFSFDP, the clustering accuracy is greatly improved. In addition, with the ongoing development of concurrent computing, the time problem will not be a major problem for much longer.



(a) (b) (c)

Fig. 10. Analysis of parameter α: Change in the OA with various values of α. (a) Indian Pines image. (b) University of Pavia image. (c) Salinas image.



(a) (b) (c)

Fig. 11. Analysis of the window size: Change in the OA with various sizes of window. (a) Indian Pines image. (b) University of Pavia image. (c) Salinas image.

TABLE VII

TIME CONSUMPTION OF FCM_S1, CFSFDP, SSC, SWSSC, SSC-S, AND S^4C FOR THE THREE EXPERIMENTAL DATA SETS: INDIAN PINES IMAGE, UNIVERSITY OF PAVIA IMAGE, AND SALINAS IMAGE

	Method	FCM_S1	CFSFDP	SSC	SWSSC	SSC-S	S^4C
Time (s)	Indian Pines	2.4778E2	3.6764E2	3.2581E3	1.7169E3	1.0327E3	1.5679E3
	University of Pavia	9.7617E2	1.8567E3	1.2934E4	9.6141E3	6.1975E3	7.3983E3
	Salinas	9.8543E2	2.1735E3	2.4034E4	1.2479E4	9.0598E3	9.3635E3

V. CONCLUSION

In this paper, we have introduced the classical SSC algorithm to HSIs by treating each kind of land-cover class as a subspace. Faced with the shortcomings of directly using SSC, we have proposed a novel S^4C algorithm for HSIs. The high spectral correlation and rich spatial information of the HSIs are simultaneously taken into consideration in the SSC model to promote the performance of the algorithm. The extensive experimental results clearly show that the proposed S^4C algorithm achieves a superior clustering performance and is a competitive algorithm.

However, the proposed algorithm still has room for improvement. For instance, the method could be improved by automatically determining the cluster number by subspace detection, and adaptively determining the regularization parameters, which will be addressed in our future work.

APPENDIX

In this section, we briefly introduce the solving process of the sparse optimization problem in (14) with the well-known ADMM algorithm [47]–[51]. First, we introduce an auxiliary matrix $\mathbf{A} \in \mathbb{R}^{MN \times MN}$ with the same size as the sparse coefficient matrix $\mathbf{C}$ to separate the variables. In this way, we only need to solve the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{C},\mathbf{A},\bar{\mathbf{C}}} \|\mathbf{W}\mathbf{C}\|_1 + \frac{\lambda}{2}\|\mathbf{Y}-\mathbf{Y}\mathbf{A}\|_F^2 + \frac{\alpha}{2}\|\bar{\mathbf{C}}-\mathbf{A}\|_F^2 \\ &\text{s.t.} \quad \mathbf{A}^T 1 = 1,\ \mathbf{A} = \mathbf{C} - \operatorname{diag}(\mathbf{C}). \end{aligned} \tag{16}$$

Two penalty terms corresponding to $\mathbf{A}^T 1 = 1$ and $\mathbf{A} = \mathbf{C} - \operatorname{diag}(\mathbf{C})$ are then added to the penalty function of (16) to obtain the following new optimization problem:

$$\begin{aligned} &\min_{\mathbf{C},\mathbf{A},\bar{\mathbf{C}}} \|\mathbf{W}\mathbf{C}\|_1 + \frac{\lambda}{2}\|\mathbf{Y}-\mathbf{Y}\mathbf{A}\|_F^2 + \frac{\alpha}{2}\|\bar{\mathbf{C}}-\mathbf{A}\|_F^2 \\ &+ \frac{\rho}{2}\|\mathbf{A}^T 1 - 1\|_2^2 + \frac{\rho}{2}\|\mathbf{A} - (\mathbf{C} - \operatorname{diag}(\mathbf{C}))\|_F^2 \\ &\text{s.t.} \quad \mathbf{A}^T 1 = 1,\ \mathbf{A} = \mathbf{C} - \operatorname{diag}(\mathbf{C}). \end{aligned} \tag{17}$$

It can be easily proved that the solutions to (14) and (16) coincide with that of (17). Next, we introduce a vector $\delta \in \mathbb{R}^{MN}$ and a matrix $\boldsymbol{\Delta} \in \mathbb{R}^{MN \times MN}$ as Lagrange multipliers for the two equality constraints in (17) to obtain the Lagrange function as

$$\begin{aligned} &\min_{\mathbf{C},\mathbf{A},\bar{\mathbf{C}}} \|\mathbf{W}\mathbf{C}\|_1 + \frac{\lambda}{2}\|\mathbf{Y}-\mathbf{Y}\mathbf{A}\|_F^2 + \frac{\alpha}{2}\|\bar{\mathbf{C}}-\mathbf{A}\|_F^2 \\ &+ \frac{\rho}{2}\|\mathbf{A}^T 1 - 1\|_2^2 + \frac{\rho}{2}\|\mathbf{A} - (\mathbf{C} - \operatorname{diag}(\mathbf{C}))\|_F^2 \\ &+ \delta^T(\mathbf{A}^T 1 - 1) + \operatorname{tr}\left(\boldsymbol{\Delta}^T(\mathbf{A} - \mathbf{C} + \operatorname{diag}(\mathbf{C}))\right) \end{aligned} \tag{18}$$

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

The aforementioned optimization problem can then be divided into three subproblems: 1) updating $\mathbf{A}$ with the other four variables fixed; 2) updating $\mathbf{C}$ by fixing the other variables and then updating $\bar{\mathbf{C}}$ with $\mathbf{C}$; and 3) updating δ and $\boldsymbol{\Delta}$ using $\mathbf{A}$ and $\mathbf{C}$. Specifically, the ADMM utilizes an iterative procedure as follows.

1) For subproblem 1), $\mathbf{A}^{(k+1)}$ is obtained by minimizing L with respect to $\mathbf{A}$, while $(\mathbf{C}^{(k)}, \bar{\mathbf{C}}^{(k)}, \delta^{(k)}, \boldsymbol{\Delta}^{(k)})$ are fixed.

We calculate the derivative of L with respect to $\mathbf{A}$ and set it to zero to obtain the calculation formula of $\mathbf{A}$ as follows:

$$(\lambda \mathbf{Y}^T\mathbf{Y} + \alpha\mathbf{I} + \rho 1 1^T + \rho\mathbf{I})\mathbf{A}^{(k+1)} = \lambda\mathbf{Y}\mathbf{Y}^T + \alpha\bar{\mathbf{C}}^{(k)} + \rho\left(1 1^T + \mathbf{C}^{(k)}\right). \quad (19)$$

2) For subproblem 2), $\mathbf{C}^{(k+1)}$ can be obtained by minimizing L with respect to $\mathbf{C}$, while $(\mathbf{A}^{(k+1)}, \bar{\mathbf{C}}^{(k)}, \delta^{(k)}, \mathbf{\Delta}^{(k)})$ are fixed

$$\mathbf{C}^{(k+1)} = \mathbf{W}\left(\mathbf{J} - \mathrm{diag}(\mathbf{J})\right), \mathbf{J} \triangleq \Gamma_{\frac{1}{\rho}}\left(\mathbf{A}^{(k+1)} + \frac{\mathbf{\Delta}^{(k)}}{\rho}\right) \quad (20)$$

where $\Gamma_{1/\rho}(\bullet)$ is a shrinkage-thresholding operator, $\Gamma_{1/\rho}(v) = (|v| - (1/\rho))_+ \mathrm{sgn}(v)$, and the operator $(\bullet)_+$ returns its arguments if it is nonnegative and returns zero otherwise. The weighted matrix $\mathbf{W}$ can be calculated with (10) and (11). We then update $\bar{\mathbf{C}}^{(k+1)}$ utilizing $\mathbf{C}^{(k+1)}$ with (12).

3) For subproblem 3), the Lagrange multipliers $\delta^{(k+1)}$ and $\mathbf{\Delta}^{(k+1)}$ are obtained through a gradient ascent update with the step size $\rho = 1000$

$$\delta^{(k+1)} = \delta^{(k)} + \rho\left(\mathbf{A}^{(k+1)}1 - 1\right) \quad (21)$$

$$\mathbf{\Delta}^{(k+1)} = \mathbf{\Delta}^{(k)} + \rho\left(\mathbf{A}^{(k+1)} - \mathbf{C}^{(k+1)}\right). \quad (22)$$

These three steps are repeated until convergence is achieved or the number of iterations exceeds the maximum iteration number. Specifically, the iteration is terminated when we have $\|\mathbf{A}^{(k)T}1 - 1\|_\infty \le \varepsilon,\ \|\mathbf{A}^{(k)} - \mathbf{C}^{(k)}\|_\infty \le \varepsilon,\ \|\mathbf{A}^{(k)} - \mathbf{A}^{(k-1)}\|_\infty \le \varepsilon$, where ε denotes the error tolerance for the primal and dual residuals.

ACKNOWLEDGMENT

The authors would like to thank Prof. D. Landgrebe at Purdue University for providing the free downloads of the Indian Pines and Salinas Airborne Visible/Infrared Imaging Spectrometer images, Prof. Gamba from the University of Pavia for providing the Reflective Optics System Imaging Spectrometer data set, and Prof. Elhamifar and Prof. Vidar from The Johns Hopkins University for sharing the sparse-subspace-clustering algorithm code.

REFERENCES

[1] D. Landgrebe, "Hyperspectral image data analysis," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 17–28, Jan. 2002.

[2] J. A. Richards and J. A. Richards, *Remote Sensing Digital Image Analysis*, vol. 3. Berlin, Germany: Springer-Verlag, 1999.

[3] W. He, H. Zhang, L. Zhang, and H. Shen, "Hyperspectral image denoising via noise-adjusted iterative low-rank matrix approximation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 6, pp. 3050–3061, Jun. 2015.

[4] W. He, H. Zhang, L. Zhang, and H. Shen, "Total-variation-regularized low-rank matrix factorization for hyperspectral image restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 178–188, Jan. 2016.

[5] W. He, H. Zhang, and L. Zhang, "Sparsity-regularized robust non-negative matrix factorization for hyperspectral unmixing," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, DOI: 10.1109/JSTARS.2016.2519498, to be published.

[6] H. Zhang, L. Zhang, and H. Shen, "A super-resolution reconstruction algorithm for hyperspectral images," *Signal Process.*, vol. 92, no. 9, pp. 2082–2096, Sep. 2012.

[7] H. Zhang, "Hyperspectral image denoising with cubic total variation model," in *Proc. ISPRS Ann. Photogramm., Remote Sens. Spatial Inf. Sci.*, 2012, vol. 7, pp. 95–98.

[8] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.

[9] S. P. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 129–137, Mar. 1982.

[10] G. H. Ball and D. J. Hall, "ISODATA, A novel method of data analysis and pattern classification," Stanford Research Inst Menlo Park, CA, USA, 1965.

[11] J. C. Bezdek, *Pattern Recognition With Fuzzy Objective Function Algorithms*. New York, NY, USA: Springer-Verlag, 2013.

[12] Y. Zhong, A. Ma, and L. Zhang, "An adaptive memetic fuzzy clustering algorithm with spatial information for remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 4, pp. 1235–1248, Apr. 2014.

[13] A. Rodriguez and A. Laio, "Clustering by fast search-and-find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, Jun. 2014.

[14] J. J. Buckley, "Fuzzy hierarchical analysis," *Fuzzy Sets Syst.*, vol. 17, no. 3, pp. 233–247, Dec. 1985.

[15] S. Vijendra, "Efficient clustering for high dimensional data: Subspace based clustering and density based clustering," *Inf. Technol. J.*, vol. 10, no. 6, pp. 1092–1105, 2011.

[16] Y. Zhong, L. Zhang, and W. Gong, "Unsupervised remote sensing image classification using an artificial network," *Int. J. Remote Sens.*, vol. 32, no. 19, pp. 5461–5483, Oct. 2011.

[17] Y. Zhong, S. Zhang, and L. Zhang, "Automatic fuzzy clustering based on adaptive multi-objective differential evolution for remote sensing imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 5, pp. 2290–2301, Oct. 2013.

[18] J. Yan and M. Pollefeys, "A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate," in *Proc. ECCV*, 2006, pp. 94–106.

[19] A. Goh and R. Vidal, "Segmenting motions of different types by unsupervised manifold clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2007, pp. 1–6.

[20] T. Zhang, A. Szlam, Y. Wang, and G. Lerman, "Hybrid linear modeling via local best-fit flats," *Int. J. Comput. Vis.*, vol. 100, no. 3, pp. 217–240, Dec. 2012.

[21] L. Zelnik-Manor and M. Irani, "Degeneracies, dependencies and their implications in multi-body and multi-sequence factorizations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2003, vol. 2, pp. II-287–II-293.

[22] G. Chen and G. Lerman, "Spectral Curvature Clustering (SCC)," *Int. J. Comput. Vis.*, vol. 81, no.3, pp. 317–330, Mar. 2009.

[23] E. Elhamifar and R. Vidar, "Sparse subspace clustering: Algorithm, theory, and application," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

[24] H. Zhai, H. Zhang, L. Zhang, P. Li, and X. Xu, "Spectral-spatial clustering of hyperspectral remote sensing image with sparse subspace clustering model," in *Proc. IEEE GRSS WHISPERS*, Tokyo, Japan, Jun. 2015.

[25] E. Elhamifar and R. Vidar, "Sparse subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2009, pp. 2790–2797.

[26] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," IEEE Trans. Inf. Theory, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[27] J. Wright, Y. A. Yang, and A. Ganesh, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[28] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representation in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.

[29] X. Y. Wang and H. Xu, "Noisy sparse subspace clustering," unpublished paper, 2013. [Online]. Available: http://arxiv.org/abs/1309.1233

[30] X. Zeng and M. A. Figueiredo, "A novel sparsity and clustering regularization," unpublished paper, 2013. [Online]. Available: http://arxiv.org/abs/1310.4945

[31] V. M. Patel, H. V. Nguyen, and R. Vidal, "Latent sparse subspace clustering," in *Proc. IEEE ICCV*, 2013, pp. 225–232.

[32] L. Lu and R. Vidal, "Combined central and subspace clustering for computer vision application," in *Proc. 23rd Int. Conf. Mach. Learn.*, 2006, pp. 593–600.

[33] J. C. Harsanyi and C.-I Chang, "Hyperspectral image classification and dimensionality reduction: An orthogonal subspace projection approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 4, pp. 779–785, Jul. 1994.

[34] E. Elhamifar and R. Vidal, "Clustering disjoint subspaces via sparse representation," in *Proc. IEEE ICASSP*, 2010, pp. 1926–1929.

[35] B. McWilliams and G. Montana, "Subspace clustering of high-dimensional data: A predictive approach," *Data Mining Knowl. Discov.*, vol. 28, no. 3, pp. 736–772, May 2014.

[36] H. Zhang, W. He, L. Zhang, H. Shen, and Q. Yuan, "Hyperspectral image restoration using low-rank matrix recovery," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4729–4743, Aug. 2014.

[37] E. Elhamifar, G. Sapiro, and R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," in *Proc. IEEE CVPR*, 2012, pp. 1600–1607.

[38] H. Shen and J. Z. Huang, "Sparse principal component analysis via regularized low rank matrix approximation," *J. Multivariate Anal.*, vol. 99, no. 6, pp. 1015–1034, Jul. 2008.

[39] R. Rubinstein, R. Zibulevsky, and M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1553–1564, Mar. 2010.

[40] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, 2002, vol. 2, pp. 849–856.

[41] V. Luxburg and Ulrike, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, Dec. 2007.

[42] Y. Boykov and G. Funka-Lea, "Graph cuts and efficient ND image segmentation," *Int. J. Comput. Vis.*, vol. 70, no. 2, pp. 109–131, Nov. 2006.

[43] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[44] Z. Wu and R. Leahy, "An optimal graph theoretic approach to data clustering: Theory and its application to image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 11, pp. 1101–1113, Nov. 1993.

[45] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010.

[46] M. Fauvel, J. A. Benediktsson, J. Chanussot, and J. R. Sveinsson, "Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 11, pp. 3804–3814, Nov. 2008.

[47] J. Eckstein and D. P. Bertsekas, "On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Programm.*, vol. 55, no. 1–3, pp. 293–318, Apr. 1992.

[48] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comput. Math. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.

[49] M. Benzi and M. A. Olshanskii, "An augmented Lagrangian-based approach to the Oseen problem," *SIAM J. Sci. Comput.*, vol. 28, no. 6, pp. 2095–2113, Dec. 2006.

[50] M. V. Afonso, J. M. Bioucas-Dias, and M. A. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems," *IEEE Trans. Image Process.*, vol. 20, no. 3, pp. 681–695, Mar. 2011

[51] J. F. Mota, J. W. F. Xavier, P. M. Q. Aguiar, and M. Puschel, "D-ADMM: A communication-efficient distributed algorithm for separable optimization," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2718–2723, May 2013.

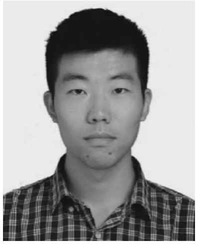
[52] S. Chen and D. Zhang, "Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 34, no. 4, pp. 1907–1916, Aug. 2004.

[53] [Online]. Available: http://people.sissa.it/~laio/Research/Res_clustering.php

Hongyan Zhang (M'13) received the B.S. degree in geographic information system and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 2005 and 2010, respectively.

He has been an Associate Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, since 2013. He has authored/coauthored more than 50 research papers. His research interests include image reconstruction for quality improvement, hyperspectral image processing, sparse representation, and low rank methods for sensing image imagery.

Dr. Zhang is a Reviewer of more than ten international academic journals, including the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING, IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and so on.



Han Zhai (S'14) received the B.S. degree from the School of Surveying and Mapping, Shandong University of Science and Technology, Wuhan, China, in 2014. He is currently working toward the M.S. degree in the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan.

His research interests include remote sensing image processing, remote sensing image information extraction, and sparse representation.

Liangpei Zhang (M'06–SM'08) received the B.S. degree in physics from Hunan Normal University, Changsha, China, in 1982, the M.S. degree in optics from the Xi'an Institute of Optics and Precision Mechanics of Chinese Academy of Sciences, Xi'an, China, in 1988, and the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1998.

He is currently the Head of the Remote Sensing Division, State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. He is also a Chang-Jiang Scholar Chair Professor appointed by the Ministry of Education of China. He is currently a Principal Scientist for the China State Key Basic Research Project (2011–2016) appointed by the Ministry of National Science and Technology of China to lead the remote sensing program in China. He has more than 410 research papers. He is the holder of 15 patents. His research interests include hyperspectral remote sensing, high-resolution remote sensing, image processing, and artificial intelligence.

Dr. Zhang is the Founding Chair of the IEEE Geoscience and Remote Sensing Society (GRSS) Wuhan Chapter. He is a Fellow of the Institution of Engineering and Technology, an Executive Member (Board of Governors) of the China National Committee of the International Geosphere–Biosphere Programme, an Executive Member of the China Society of Image and Graphics, etc. He regularly serves as a Cochair of the series International Society for Optics and Photonics (SPIE) Conferences on Multispectral Image Processing and Pattern Recognition, the Conference on Asia Remote Sensing, and many other conferences. He edits several conference proceedings, issues, and geoinformatics symposiums. He also serves as an Associate Editor of the *International Journal of Ambient Computing and Intelligence*, the *International Journal of Image and Graphics*, the *International Journal of Digital Multimedia Broadcasting*, the *Journal of Geo-spatial Information Science*, the *Journal of Remote Sensing*, and the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING; he is a Guest Editor of the *Journal of Applied Remote Sensing* and the *Journal of Sensors*. He was the General Chair for the Fourth IEEE GRSS Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing and a Guest Editor of the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING (JSTARS). He was the recipient of the Best Reviewer Awards from the IEEE GRSS for his service to the IEEE JSTARS in 2012 and the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2014. He was also the recipient of the 2010 Best Paper Boeing Award and the 2013 Best Paper ERDAS Award from the American Society of Photogrammetry and Remote Sensing. His research teams were the recipient of the top three prizes of the IEEE GRSS 2014 Data Fusion Contest, and his students have been selected as the winners or finalists of the IEEE International Geoscience and Remote Sensing Symposium Student Paper Contest in recent years.

Pingxiang Li (M'06) received the B.S., M.S., and Ph.D. degrees in photogrammetry and remote sensing from Wuhan University, Wuhan, China, in 1986, 1994, and 2003, respectively.

Since 2002, he has been a Professor with the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University. His research interests include photogrammetry and SAR image processing.