The Fast Spectral Clustering Based on Spatial Information for Large Scale Hyperspectral Image

YIWEI WEI [ID], CHAO NIU, YITING WANG, HONGXIA WANG, AND DAIZHI LIU

Rocket Force University of Engineering, Xi'an 710025, China

Corresponding author: Chao Niu (chao_niu1986@163.com)

ABSTRACT Spectral clustering is one of the most popular clustering approaches and has been applied in Hyperspectral Image (HSI) clustering well. However, most of these methods are not suitable for large scale HSI. In this paper, based on anchor graph and spatial information, we propose a novel method, called fast spectral clustering based on spatial information (FSCS), which could deal with large scale HSI and have better performance in user's accuracy, average accuracy, overall accuracy and so on. Firstly, based on the physical characteristic of HSI, FSCS algorithm combines the spatial information with spectral information, and uses the spatial nearest points to reconstructs the center point and reveal the local spatial structure. As a result, the correlation of pixels is strengthened and the clustering accuracy is improved. Secondly, the new adjacency matrix is constructed based on anchor graph and thus computational complexity is reduced significantly. Finally, in order to avoid tuning the heat-kernel parameter, the parameter-free strategy is adopted in FSCS. Experiments demonstrate the efficiency and effectiveness of the proposed FSCS algorithm.

INDEX TERMS Hyperspectral images, spatial information, fast spectral clustering, anchor-based methods.

I. INTRODUCTION

Hyperspectral image (HSI) is remotely sensed image captured by Imaging spectrometer. Hyperspectral remote sensing technology collects a large number of narrow spectral image data in the range from visible light to thermal infrared band [1]–[3]. HSI combines spatial information with spectral information together. Now HSI is widely used in military detection, environmental monitoring, geological exploration, precision agriculture, atmospheric remote sensing, hydrology and other fields [4], [5].

In recent years, HSI clustering has attracted more and more attention [6], [7], which aims to partition a given image into groups so that pixels in the same group are as similar to each other as possible. So far, there are many clustering methods proposed for HSI, such as k-means [5], fuzzy c-means (FCM) [8], density based methods [9], and automatic fuzzy clustering method. In recent years, spectral clustering has become one of the most popular clustering algorithms for HSI, such as spectral curvature clustering [10], sparse subspace clustering (SSC) [11], [12], fast spectral clustering [4] and so on.

However, HSI data is enormous which will cost too much time to obtain useful information [13]. The clustering

methods are not applicable to deal with large-scale HSI [13]. Its computational complexity is $O(N^2c + N^2D)$, where N, c and D are the number of samples, classes and features, respectively. Furthermore, the clustering methods have to tune the heat-kernel parameter [11] these algorithms do not use the the spatial information of HSI, which can strengthen the correlation of pixels and increase the accuracy [14]. Because neighboring pixels in image are more likely to be contained in the same features, an improved nonnegative matrix under approximation (NMU) has been put forward [15]. This variant of NMU is particularly well suited for image analysis as it incorporates the spatial information.

To tackle these problems, inspired by latest progress on fast spectral clustering, large-scale spectral clustering [16], [17] and large-scale spectral based on the dimensionality reduction [18], [19], a new approach is proposed, called fast spectral clustering based on spatial information (FSCS) for large-scale HSI. This algorithm constructs a new model. Experiments on HSI data sets demonstrate the efficiency and effectiveness of the proposed FSCS algorithm such as the quantitative and visual evaluations, less time and so on. The main contributions of our work are as follows:

(i) Combining spatial information with spectral information of HSI, the center point is reconstructed by using the spatial nearest points. As a result, the redundant

information in the image is effectively reduced and the better quantitative and visual evaluations are obtained;

(ii) Constructing anchor-based method in order to deal with large-scale HSI, therefore the computational complexity can be greatly reduced;

(iii) Applying parameter-free effective neighbor assignment strategy to construct the similarity matrix. As a result, the new algorithm avoids tuning the heat-kernel parameter.

II. THE SPECTRAL CLUSTERING ALGORITHM

The objective function of the spectral clustering [20] is defined as:

$$\min_{F^T F = I} Tr\left(F^T L F\right) \tag{1}$$

where $F \in \mathbb{R}^{n \times c}$ is the class indicator matrix of all data. The optimal solution of F in Eq. (1) is the first c smallest eigenvectors of L.

The method of generaly graph-based spectral clustering has achieved excellent performance in discovering novel classes. However, it has limitations in the neighborhood weighted graph construction. Specially, the adjacency matrix can be designed by using all data points, thus the complexity of the algorithm is $O(n^2(d + c))$, which puts forward higher requirements on the performance of the computer. With the rapid increasement in data size, the method can not handle large data, which is faced in HSI clustering.

Furthermore, the spectral clustering methods have good performance, but they don't take advantage from the spatial information of HSI, which can strengthen the correlation of pixels and improve the clustering accuracy by using physical characteristics of HSI.

Remark 1: the complexity of the graph-based spectral clustering algorithm is $O(n^2(d + c))$, which is not suitable to handle large-scale HSI clustering problem.

III. THE FAST SPECTRAL CLUSTERING BASED ON SPATIAL INFORMATION FOR LARGE SCALE HSI

In this section, in order to deal with the large-scale HSI, we introduce the fast spectral clustering algorithm. Firstly, we use weighted joint spatial-spectral reconstruction method to smooth the HSI and reduce the interference of singular points. Secondly, we use anchor-based graph to construct adjacency matrix which can reduce time consumption quite a lot. Thirdly, in order to avoid tuning extra parameters in spectral clustering, the parameter-free strategy is adopted. Finally, the fast spectral clustering method is proposed to obtain the label of HSI.

A. WEIGHTED SPATIAL-SPECTRAL RECONSTRUCTION METHOD

Weighted spatial-spectral reconstruction method can smooth the HSI and reduce the interference of singular points. The principal is: the coordinate of sample point x_{ij} in HSI is (i, j),

while the neighbor space of x_{ij} is

$$N(x_{ij}) = \{x_{pq} | p \in [i - a, i + a], q \in [j - a, j + a]\}, \tag{2}$$

where $a = (\omega - 1)/2$, ω means the width (or scale) of the nearest neighbor window, which is usually odd.

We use the following rule to reconstruct the pixel point x_{ij}, and get reconstructed pixels.

The reconstructed pixels can be defined as:

$$\hat{x}_{ij} = \frac{\sum_{x_{pq} \in N(x_{ij})} w_{pq} x_{pq}}{\sum_{x_{pq} \in N(x_{ij})} w_{pq}} \tag{3}$$

where the weight of each pixel w_{pq} in the space of neighbor $N(x_{pq})$ to the central pixel x_{ij} is $w_{pq} = exp(-\gamma_0 ||x_{ij} - x_{pq}||_2^2)$. γ_0 is spectral factor. By assigning a small weight to the points which are quite different from the central pixel in the space of the nearest neighbor, we reduce the interference of the singular points, and then achieve smoothly HSI.

The scale of neighbor window ω adjusts the size of neighbor space, while the spectral factor γ_0 adjusts the degree of mutual influence between pixels. Therefore, the selection of parameters ω and γ_0 will directly affect the results of image processing. Neighbor space of pixels in different cases is shown in Fig. 1, from which we can see each pixel in a HSI is represented by a square grid; a light gray grid as the central pixel and a dark gray grid as the filled mode.

For the normal position, the pixels don't need to be filled in the near $N(x_{ij})$. But for the pixels located at the edge or corner of the image they are preprocessed by the method in Fig. 1. When the pixels are in normal positions as shown in Fig.1(a), its neighbor space $N(x_{ij})$ don't need to be filled. When they are in the edge area or in the corner as at Fig.1(b) and Fig.1(c), the vacant pixels in the space of neighbor need to be filled with the nearest pixels; or filled with x_{ij} directly if there are no nearest pixels in the corner positions.

In this subsection, based on the physical characteristics of HSI, we use the spatially nearest points to reconstruct the center point and to reveal the local spatial structure. By effectively reducing the redundant information in the image, HSI method increases the consistency of the same kinds of pixels as well as enhancing the difference of the different kinds of pixels. As a result, the correlation of pixels is strengthened and the clustering accuracy is improved.

The computational complexity of Eq. (3) is $O(2\omega^2 D - 2D + 2)$. Thus, for the N pixels in the HSI, the computational complexity of weighted spatial-spectral reconstruction algorithm is $O(DN\omega^2)$, where D is the number of spectral bands, N is the number of pixels and ω is the scale of neighbor window.

Remark 2: the correlation of pixels is strengthened and the clustering accuracy is improved by reconstructing the spatial-spectral information. Furthermore, the complexity of the method is $O(DN\omega^2)$, where N is the number of pixels, D is spectral bands and and ω is the scale of the neighbor window.
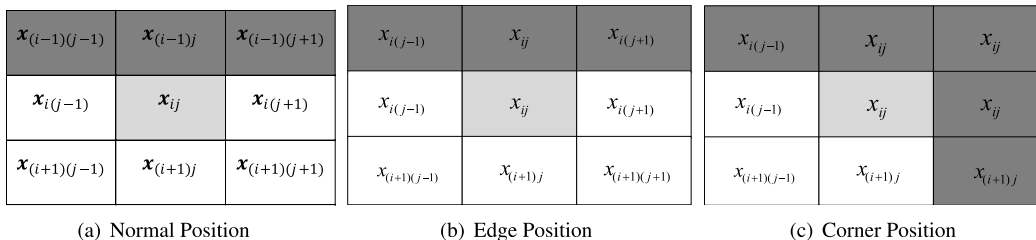
$x_{(i-1)(j-1)}$	$x_{(i-1)j}$	$x_{(i-1)(j+1)}$
$x_{i(j-1)}$	x_{ij}	$x_{i(j+1)}$
$x_{(i+1)(j-1)}$	$x_{(i+1)j}$	$x_{(i+1)(j+1)}$

(a) Normal Position

$x_{i(j-1)}$	x_{ij}	$x_{i(j+1)}$
$x_{i(j-1)}$	x_{ij}	$x_{i(j+1)}$
$x_{(i+1)(j-1)}$	$x_{(i+1)j}$	$x_{(i+1)(j+1)}$

(b) Edge Position

$x_{i(j-1)}$	x_{ij}	x_{ij}
$x_{i(j-1)}$	x_{ij}	x_{ij}
$x_{(i+1)(j-1)}$	$x_{(i+1)j}$	$x_{(i+1)j}$

(c) Corner Position

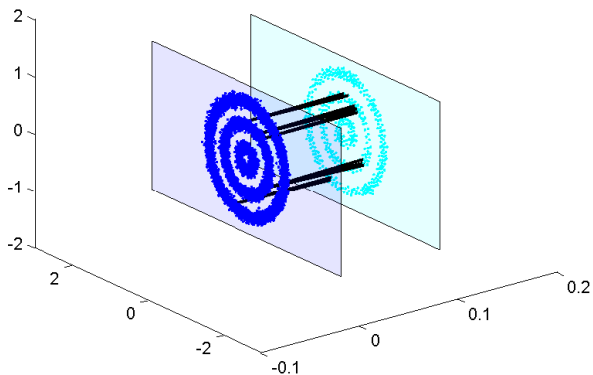**FIGURE 1.** The nearest neighbor space of central pixel in different situation.

FIGURE 2. The construction of Z (the blue points are the original data and the green points are the anchor points). For convenience, only a tiny fraction of the inter-layer edges, which present the weights between the original data and the anchors, are shown.

B. THE CONSTRUCTION OF ANCHOR GRAPH

In order to reduce the computational cost, especially for large-scale HSI, the anchor-based strategy has been applied in recent years and has achieved good results [16], [17], [21]. In general, the anchor-based approaches seek m anchor points from the original data points, where $m \ll n$. A designed matrix $\mathbf{Z} \in \mathbb{R}^{n \times m}$, which can measure the adjacency between data points and anchor points. The illustrative example of constructing Z is shown in Fig. 2. The construction is built on three-ring synthetic data, which consists of 10000 data points. We select 500 anchors from the original data randomly.

Let $\mathbf{U} = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m]^T \in \mathbb{R}^{m \times d}$ denotes the generated anchors. z_{ij} is the (i, j)-th element of $\mathbf{Z}$ representing the adjacency between the i-th data point and j-th anchor point, which can be defined as follows [16], [22]:

$$z_{ij} = \frac{K(x_i, u_j)}{\sum_{s \in \Phi_i} K(x_i, u_j)}, \quad \forall j \in \Phi_i, \tag{4}$$

where $\Phi_i \subset \{1, \ldots, m\}$ denotes the indexes of the k nearest neighbors of $\boldsymbol{x}_i$ in U. $K()$ represents a kernel function and Gaussian kernel $K(\boldsymbol{x}_i, \boldsymbol{u}_j) = exp(-\|\boldsymbol{x}_i - \boldsymbol{u}_j\|_2^2/2\sigma^2)$ is usually adopted, where σ is the heat-kernel parameter. In order to obtain excellent classification results, many experiments are needed to select a suitable heat-kernel parameter.

Remark 3: Compared with the spectral clustering method, anchor-based methods have less computational costs and the

complexity of the algorithm is $O(ndm)$; however, most methods require tuning the heat-kernel parameter.

C. THE CONSTRUCTION OF ADAPTIVE SIMILARITY MATRIX

In order to avoid tuning extra parameters in Section 2.2, by adopting the parameter-free strategy [23], we can obtain the i-th row of $\mathbf{Z}$ in Eq. (4) by solving the following problem:

$$\min_{z_i^T \mathbf{1} = 1, z_{ij} \geq 0} \sum_{j=1}^m \|\boldsymbol{x}_i - \boldsymbol{u}_j\|_2^2 z_{ij} + \gamma z_{ij}^2, \tag{5}$$

where z_i^T denotes the i-th row of $\mathbf{Z}$. Eq.(12) is parameter-free which has been proved in Reference [23].

According to Reference [23], z_i is sparse, which has exactly k nonzero values; the learned $\mathbf{Z}$ is also sparse, therefore the computational burden of subsequent processing can be largely alleviated. The square of Euclidean distance between $\boldsymbol{x}_i$ and $\boldsymbol{u}_j$ can be defined as $d(\boldsymbol{x}_i, \boldsymbol{u}_j) = \|\boldsymbol{x}_i - \boldsymbol{u}_j\|_2^2$. The parameter γ can be set as $\gamma = \frac{k}{2} d(i, k+1) - \frac{1}{2} \sum_{j=1}^k d(i, j)$. The solution to Eq. (5) is

$$z_{ij} = \frac{d_{i,k+1} - d_{i,j}}{k d_{i,k+1} - \sum_{j=1}^k d(i, j)}. \tag{6}$$

For more detailed derivation, see reference [23]. Furthermore, the affinity matrix $\boldsymbol{W}$ can be obtained [16] as follows:

$$\boldsymbol{W} = \boldsymbol{Z} \boldsymbol{\Lambda}^{-1} \boldsymbol{Z}^T, \tag{7}$$

where $\boldsymbol{\Lambda} \in \mathbb{R}^{m \times m}$ is the diagonal matrix and $\boldsymbol{\Lambda}_{jj} = \sum_{i=1}^n z_{ij}$.

D. THE FAST SPECTRAL CLUSTERING BASED ON SPATIAL INFORMATION FOR LARGE-SCALE HSI

The objective function of fast clustering for HSI can be defined as:

$$\min_{\boldsymbol{F}^T \boldsymbol{F} = \boldsymbol{I}} Tr\left(\boldsymbol{F}^T L \boldsymbol{F}\right) \tag{8}$$

and the Laplacian matrix L can therefore be written as:

$$L = D - W. \tag{9}$$

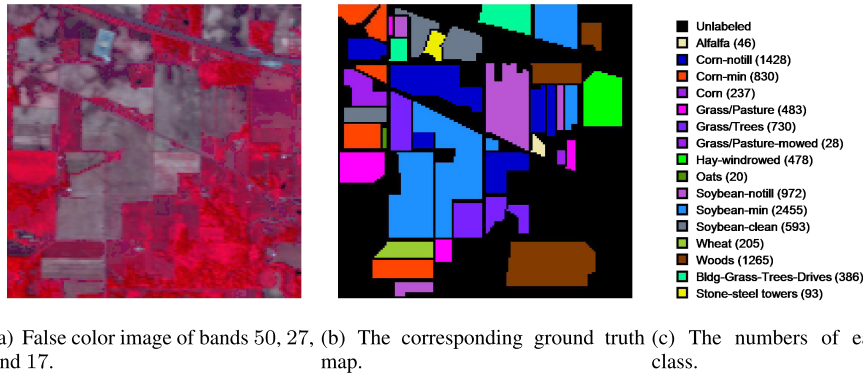where D is a diagonal matrix whose diagonal elements are row sums of W, According to the graph-based learning,

(a) False color image of bands 50, 27, (b) The corresponding ground truth (c) The numbers of each and 17.   map.   class.

FIGURE 3. **The Indian pines hyperspectral image.**

the cost function associated with F in the Eq. (1) can be written as:

$$J(F) = Tr(F^T L F) - \lambda Tr(F^T F - I) \qquad (10)$$

where λ is a regular parameter.

The optimal solution for the Eq. (10) can thus be computed as follows:

$$\frac{\partial J}{\partial F}|_{F=F^*} = 2LF^* - 2\lambda F^* = 0 \qquad (11)$$

The optimal solution can be derived from Eq. (11) as follows:

$$LF^* = \lambda F^* \qquad (12)$$

This means that the optimal solution is the eigenvalue decomposition on L whose computational complexity is $O(n^2 c)$. In order to reduce the computational complexity, we try several methods.

The similarity matrix W is automatically normalized and the degree matrix $D = I$, the problem (8) can be written as follows:

$$\min_{F^T F = I} Tr\left(F^T W F\right) \qquad (13)$$

According to Eq. (7), we can obtain $W = A^T A$, where $A = Z\Lambda^{-\frac{1}{2}}$. So,it is easy to know that the solution of Eq.(13) can be obtained by the SVD (singular value decomposition) of A. As a result, the computational complexity can be reduced to $O(m^2 c + nmc)$.

By means of the SVD on B, the solution of relaxed continuous can be obtained, which is composed of eigenvectors corresponding to the smallest c eigenvalues. The k-means clustering method can be used to calculate the discrete solution.

To summarize, we demonstrate the steps of FSCS in Algorithm 1.

E. COMPUTATIONAL COMPLEXITY ANALYSIS
The computational complexity of FSCS can be divided into the following parts:

Algorithm 1 The FSCS Algorithm

Input: data matrix $X \in \mathbb{R}^{n \times d}$, the number of classes c, the number of anchors m, the number of neighbors k.
     1) Generate m anchors by random selection.
     2) Obtain the matrix Z according to Eq. (6).
     3) Obtain the matrix W according to Eq. (7).
     4) Calculate the matrix A, where $A = Z\Lambda^{-\frac{1}{2}}$.
     5) Obtain the relaxed continuous solution of F by performing SVD on matrix A.
     6) Obtain the class indicator of data points by the k-means clustering method.

Output: the class indicators c.

---

1) $O(DN\omega^2)$ is needed in the weighted spatial-spectral reconstruction method, where N, D and ω are the number of pixels, spectral bands and the scale of the neighbor window, respectively.
2) $O(1)$ is needed to obtain m anchor points by random selection.
3) $O(NDm)$ is needed to obtain the matrix Z.
4) $O(m^2 c + Nmc)$ is needed to obtain the relaxed continuous solution of F by performing SVD on matrix B.
5) $O(Nmct)$ is needed to perform k-means on the relaxed discrete solution for final clustering results, where t is the iterative number.

In summary, the computational complexity of our method can be approximated as $O(ND\omega^2) + O(NDm)$.

IV. EXPERIMENT
In this section, we introduce the experiment data used in FSCS firstly, then, parameter setting has been studied. Overall accuracy (OA) is viewed as the evaluation standard to select the best parameters. In order to confirm the result of FSCS, we compare proposed method with classic clustering methods (named k-means, FCM, FCM_S1, SC). Finally, clustering time of different methods is analysed.
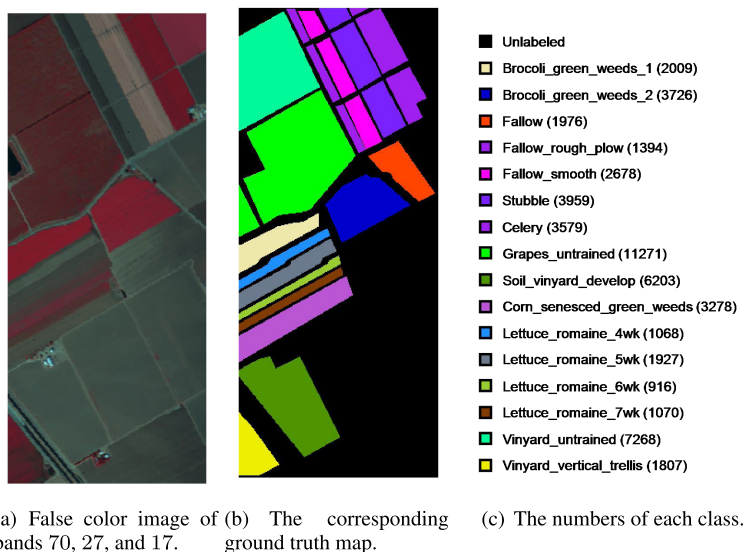
(a) False color image of bands 70, 27, and 17.
(b) The corresponding ground truth map.
(c) The numbers of each class.

FIGURE 4. The Salinas hyperspectral image.

(a) False color image of bands 40, 30, and 20.
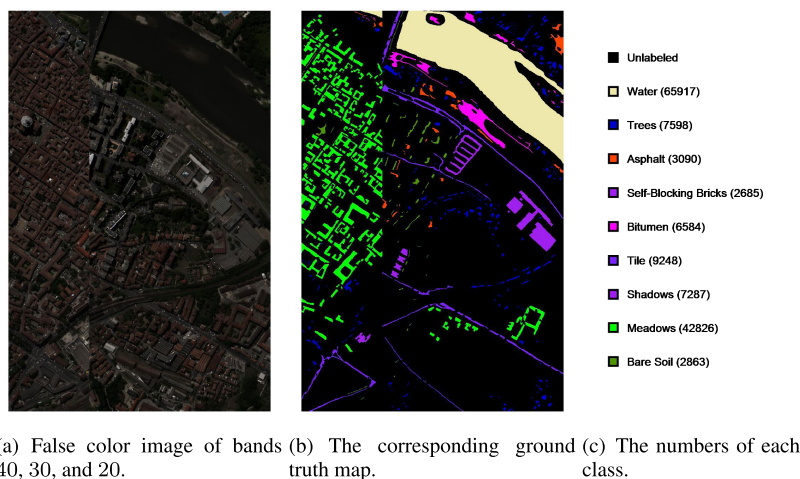(b) The corresponding ground truth map.
(c) The numbers of each class.

FIGURE 5. The Pavia centre hyperspectral image.

A. HSI DATA

Three famous HSI data (named Indian Pines, Salinas, and Pavia Centre)[1] are used in clustering experiments. Here we explain some details of these three data sets.

Indian pines data set was collected by AVIRIS sensor in a woodland. The dimension of data image is 145×145. There are 200 bands being used in experiment because some of the bands are affected by noise. In addition, the number of this data set is 21025 and the classes is 16. In figure 3, the false color image composition of bands 50, 27, and 17, the ground truth map and numbers of each class are shown in Fig3. (a), Fig3. (b), and Fig3. (c), respectively.

Salinas data set was also collected by AVIRIS. The dimension of this data is 512×217. There are 204 bands available

in our experiment. The number of this data set is 111104 and classes are 16. In figure 4, the false color image composition of bands 60, 30, and 2, the ground truth map and numbers of each class are shown in Fig4. (a), Fig4. (b), and Fig4. (c), respectively.

Pavia centre data set was collected by ROSIS sensor. In original data, dimension is 1096×1096. Because some of the pixels are image background which contains no information, we use the data of dimension reduced in our experiment. In new data, dimension is 1096×715. There are 102 bands available in the experiment. In addition, the number of this data set is 783640 and there are 9 classes. In figure 5, the false color image composition of bands 40, 30, and 20, the ground truth map and numbers of each class are shown in Fig5. (a), Fig5. (b), and Fig5. (c), respectively.

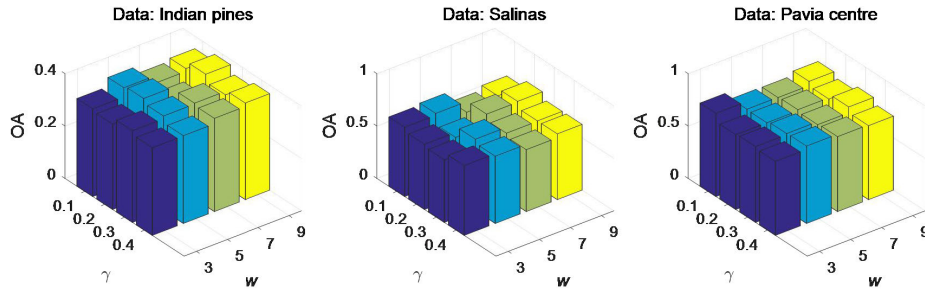[1]http://www.ehu.eus/ccwintco/index.php

FIGURE 6. Parameters setting experiment result.

(a) Ground truth   (b) k-means   (c) FCM

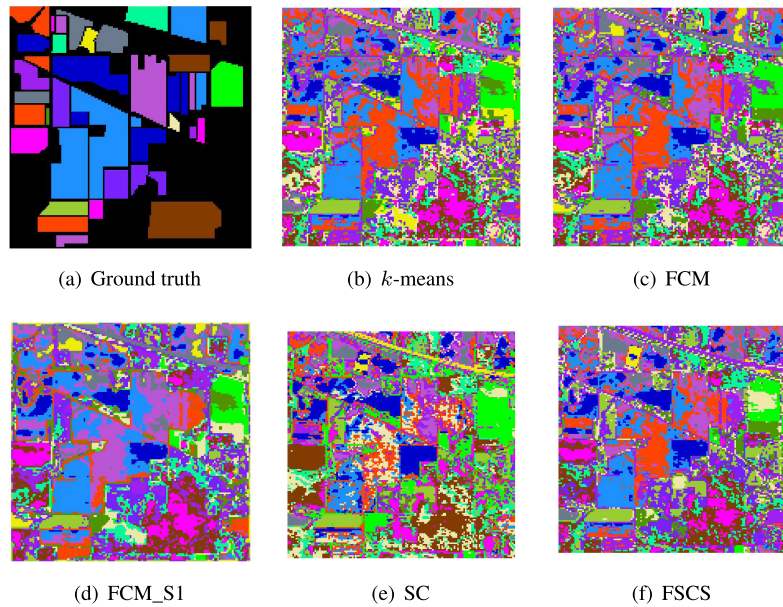(d) FCM_S1   (e) SC   (f) FSCS

FIGURE 7. Clustering maps on different method of Indian pines data.

B. EXPERIMENTAL RESULTS

The main evaluation of clustering results includes four standards: user's accuracy (UA), average accuracy (AA), overall accuracy(OA) [24]–[26], and Kappa coefficient [27]. UA represents clustering accuracy of each class (labeled data) in one image. It tells us the clustering accuracy of each labeled category. AA is the mean of class-specific accuracies which is the mean accuracy of UA. It shows us the clustering effect of labeled data in the statistical level. OA represents clustering accuracy of all data (labeled data and background) in one image. It shows us the clustering effect of all data in the statistical level. The value of UA, AA, OA ranges from 0 to 1. Here, values of higher accuracy are better than of lower one. Moreover, Kappa coefficient is a consistency test method to measure the consistency between original label and predicted label. The value of Kappa ranges from 0 to 1. Large kappa value means better consistency. In our experiments, we use the same initialization and repeat 30 times independently.

Because the total number of pixels in Indian Pines data set is 21025, we consider it as small number data set. In the experiment, we set parameter $w = 9$ and $\gamma = 0.2$, respectively.

TABLE 1. Experiment result of Indian pines data.

Class	k-means	FCM	FCM_S1	SC	FSCS
Alfalfa	0	0	0.0652	0.0217	**0.1087**
Corn-notill	0.2850	0.2773	0.2829	**0.4468**	0.3445
Corn-min	**0.4434**	0.4096	0.1157	0.2819	0.3867
Corn	0.1603	0.1392	0.2068	0.0633	**0.2236**
Grass/Pasture	**0.4969**	0.4865	0.4907	0.0104	0.3830
Grass/Trees	0.4082	0.4425	**0.4986**	0.1699	0.4137
Grass/Pasture-mowed	0.6071	**0.7857**	0	0	0
Hay-windrowed	0.7594	**0.7950**	0.6925	0.8326	0.5774
Oats	0.4500	0.4000	0.4500	0	**0.5500**
Soybean-notill	0.1862	0.2840	**0.3920**	0.1996	0.3025
Soybean-min	0.3837	0.3381	**0.3792**	0.2452	0.3605
Soybean-clean	0.1686	0.1417	0.2125	0.1636	**0.2192**
Wheat	**0.9902**	0.9610	0.8683	0.9805	0.9756
Woods	0.4024	0.4261	0.4498	**0.5557**	0.4949
Bldg-Grass-Trees-Drives	0.1736	0.1788	0.1451	**0.1969**	0.1736
Stone-steel towers	0	0	0	0.5914	**0.7957**
AA	0.3697	0.3791	0.3281	0.2975	**0.3944**
OA	0.3651	0.3641	0.3643	0.3262	**0.3826**
Kappa	0.2937	0.2944	0.2896	0.2653	**0.3140**

The experimental results (UA, AA, OA, and Kappa) are reported in Table 1, in which the best results are shown in bold. From Table 1, we can find that the AA, OA, and Kappa of FSCS performs better than other methods.
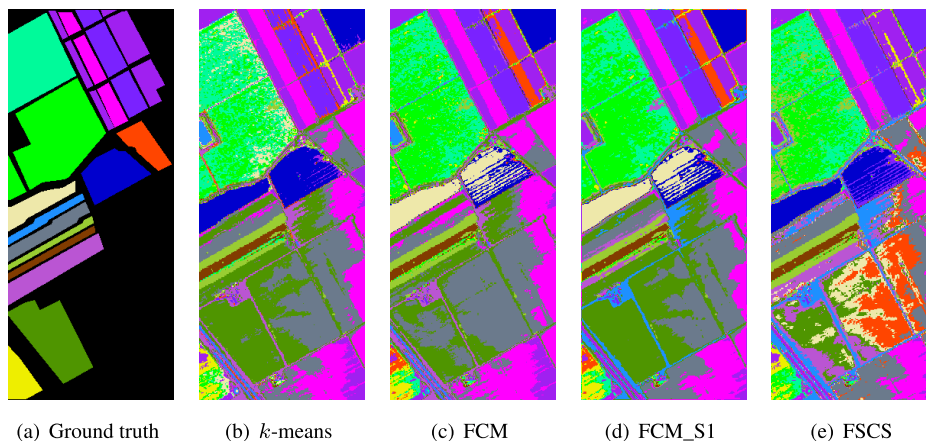
(a) Ground truth	(b) *k*-means	(c) FCM	(d) FCM_S1	(e) FSCS

FIGURE 8. Clustering maps on different method of Salinas data.

TABLE 2. Experiment result of Salinas data.

Class	*k*-means	FCM	FCM_S1	FSCS
Brocoli weeds 1	0	**0.9970**	0.9826	0.9801
Brocoli weeds 2	**0.9235**	0.3988	0.4353	0.8610
Fallow	**0.7586**	0	0	0.1776
Fallow_rough	**0.9684**	0.9656	0.9555	0.8307
Fallow_smooth	**0.9649**	0.9574	0.8215	0.9578
Stubble	0.8788	0.8287	**0.9525**	0.9518
Celery	0.9042	0.9410	0.5960	**0.9860**
Grapes_untrained	0.4245	0.4256	**0.7305**	0.4317
Soil	0.9882	**0.9924**	0.7553	0.8538
Corn	0.0012	0.0024	0.3200	**0.5946**
Lettuce_4wk	0.0356	**0.0721**	0	0
Lettuce_5wk	0.4717	0.9850	0.9523	**0.9938**
Lettuce_6wk	0.9716	**0.9869**	0.9814	0.9847
Lettuce_7wk	0.8785	0.8879	0.8888	**0.8944**
Vinyard_untrained	**0.6340**	0.5853	0.4608	0.4631
Vinyard_trellis	0.4189	0.4322	**0.4588**	0.0802
AA	0.6389	0.6537	0.6432	**0.6901**
OA	0.6401	0.6258	0.6441	**0.6640**
Kappa	0.6019	0.5876	0.6039	**0.6328**

TABLE 3. Experiment result of Pavia centre data.

Class	*k*-means	FCM	FCM_S1	FSCS
Water	**0.9936**	0.9919	0.9907	0.9814
Trees	0.6490	**0.6649**	0.5555	0.5032
Asphalt	0.1748	0.0793	**0.2942**	0.0032
Self-Blocking Bricks	0.0369	0.1121	**0.1240**	0
Bitumen	0.3171	0.5319	0.5611	**0.5878**
Tiles	**0.8914**	0.2450	0.3696	0.8391
Shadows	0.1772	0.8076	**0.8777**	0.8313
Meadows	0.5004	**0.5325**	0.5280	0.4789
Bare Soil	0.0010	0	0	**0.9990**
AA	0.4157	0.4406	0.4779	**0.5804**
OA	0.7032	0.7121	0.7218	**0.7400**
Kappa	0.5917	0.6032	0.6170	**0.6497**

However, the UA of FSCS cannot get the best result in some classes. To analyse the clustering result more clearly, we list the clustering maps in Figure 7. From figures 7, we can see that FSCAG produces more homogenous areas and better clustering maps than other algorithms, which clearly reflects the importance of incorporating spatial information. Therefore, FSCS can handle clustering work in the most block of HSI.

To verify the performance of FSCS in large-scale data set, we do the experiment on Salinas data set. In the experiment, we set parameter $w = 9$ and $\gamma = 0.2$. The experimental result (UA, AA, OA, and Kappa) is given on Table 2 and the best results are shown in bold. From Table 2, the AA, OA, and Kappa of FSCS are 0.6901, 0.6640, and 0.6328 which are higher than other methods. Although the performance of FSCS is inferior in each block clustering, the proposed method has the best comprehensive clustering effect. From Fig. 8, we see that FSCAG produces more homogeneous areas and better clustering maps than other algorithms.

Another large-scale HSI data is Pavia Centre. We do clustering experiment on it. In the experiment, we set parameter $w = 3$ and $\gamma = 0.1$. The experimental results (UA, AA, OA, and Kappa) are reported on Table 3 and and the best results are shown in bold. The AA of FSCS is almost 10 % higher than k-means, FCM, and FCM_S1, which is greatly improved than classic clustering methods. The clustering result map is reported in Fig. 9. From Fig. 9, we see that FSCAG produces more homogenous areas and better clustering maps than other algorithms. The experimental results once again confirm that the proposed method has a good clustering effect in large-scale HSI clustering.

The Hyperspectral data is complex, and there are a large difference between data as well as between classes, which leads to a large difference in clustering accuracy between different classes of the algorithm. All spectral clustering algorithms will encounter this situation. For example in Salinas data in Table. 2, K-means has poor results with classes as Brocoli weeds 1, FCM and FCM_S1 has poor results with classes as Fallow, FCM_S1 and FSCS has poor results with classes as Lettuce_4wk. In a word, in these three data sets experiments from table 1 to 3, the proposed method performs better than other methods in AA, OA and Kappa. In UA standards, although some class clustering are not the best
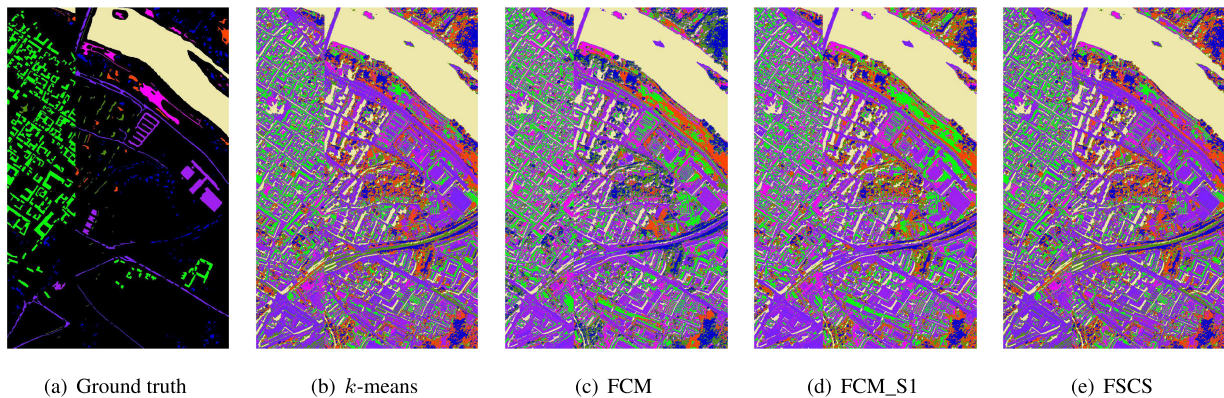
| (a) Ground truth | (b) k-means | (c) FCM | (d) FCM_S1 | (e) FSCS |

FIGURE 9. Clustering maps on different method of Pavia centre data.

TABLE 4. Clustering time (s).

Data set	k-means	FCM	FCM_S1	SC	FSCS
Idian Pines	3.8180	**2.6341**	4.2999	32.8821	2.9009
Salinas	16.0862	13.8198	23.7273	OM	**5.7286**
Pavia Centre	154.1043	45.5498	75.9006	OM	**41.2789**

one, the comprehensive clustering effect is better than others. Therefore, FSCS is a robust clustering method.

C. EXPERIMENTAL TIME ANALYSIS

Clustering time is paid close attention in this paper, for which we make a record in Table 4. From Table 4, we can see that FSCS can reach higher speed in clustering. For the two spectrally based clustering methods, In Indian Pines experiment, the proposed method only needs 2.9009(s), which is 10 times faster than the SC method and very close to the fastest method which needs 2.6341(s). In Salinas and Pavia Centre experiment, SC cannot work due to "out of memory (OM)" problem. Fortunately, FSCS can reach the fastest result than other methods which shows that it is suitable to handle large-scale data. In additional, compared to traditional clustering methods, the FSCS also has better performance. For example, in Salinas data the clustering time reduces a lot and in Pavia Centre the accuracy improves a lot (the AA of FSCS is 0.5804, the AA of FCM is 0.4406).

D. PARAMETER ANALYSIS

In the proposed method, two parameters (window scale neighbor w and spectral factor γ) have an important effect in clustering result. Parameter w acts as a denoising window scare of weighted spatial-spectral reconstruction method. Generally speaking, the size of homogeneous region determines the size of the parameters w. When the homogeneous region is large, we set a large w. Similarly, we set a small w when the homogeneous region is small. As we can see in figure (3-b) and figure (4-b), Indian Pines and Salinas dataset ground truth map show that each homogeneous region are clear and large, so we set a large w in experiment. Moreover, in figure (5-b) Pavia Centre data ground truth map shows

that same homogeneous region are patchy in left so we set a small w in experiment. This is because large homogeneous region is slowly varying among different categories of pixels while small homogeneous region is in a rapid change. Parameter γ acts as a is a regularization parameter. It can be set as $\gamma = \frac{k}{2}d(i, k+1) - \frac{1}{2}\sum_{j=1}^{k} d(i, j)$ [23]. However, the joint action of these two parameters will be a future work to study.

To research the joint action of different parameter pairs, we do experiment in three dataset. As in the reference [12], OA is regarded as the evaluation of clustering result. In figure 6, we show the experimental results of different combinations of parameters. In Indian Pines and Salinas dataset, large w can obtain best result. In Pavia center dataset, small w can obtain good result. The result validates the above analysis. Furthermore, parament γ in small value will obtain better result in the joint action.

V. DISCUSSION

Hyperspectral image clustering is playing an important role in HSI information analysis. In this paper, we proposed a novel HSI fast clustering method based on spectral clustering. Our aim is to improve the clustering effect and reduce clustering time consumption in large-scale HSI data. As shown in experimental results, better clustering results have been obtained and clustering time has been shortened.

From Table 1, 2, and 3, we can find that FSCS adopts better clustering model and achieves higher clustering accuracy. In reality, HSI data may be polluted by noise which would result in constructing wrong adjacency matrix. Therefore, we reconstruct HSI data by using weighted spatial-spectral method. The new image has taken advantage of both the spatial and spectral information at the same time. As a result, the redundant information in the image is effectively reduced and better quantitative and visual evaluations are obtained. Furthermore, most of the graph-based methods, including SSC [11], SCC [10], AGR [17] and so on, usually adopt the strategy of kernel-based neighbor assignment, therefore, a lot of experiments are needed to select the appropriate heat-kernel parameter, which will also consume a lot of time. So as to solve this problem, we adopted parameter-free

approach to construct the similarity matrix between original points and anchor points. This approach avoids tuning heat-kernel parameter.

Finally, we adopted anchor-based strategy to obtain the adjacency matrix to reduce clustering time. with this strategy, we only need to calculate bipartite graph matrix between original data set and anchor data set. The adjacency matrix can be gained in a simple way [17]. For having a quantitative analysis of clustering time, we calculated time complexity in principle. The computational complexity of our algorithm can be approximately calculated as $O(ND\omega^2) + O(NDm)$, while computational complexity of the spectral clustering (SC) is $O(N^2c + N^2D)$, where N, c, D and ω are the number of samples, classes, features and the weight of each pixel, respectively. After analysis, we found that the clustering time increases with the increasing size of data. As shows in Table 4, in Indian Pines experiments, FCSC is 10 times faster than SC method. In Salinas and Pavia Centre experiment, SC cannot work due to "out of memory (OM)" problem. Meanwhile, FCSC has better result than other unsupervised clustering methods in both of these two large-scale data sets.

However, FSCS has limitation in anchor points selection. In some dataset, the number of different categories are unequal. Since we use randomly selecting method to obtain anchor points, large number categories has a higher probability of being picked. In FSCS, adjacency matrix is obtaining by Eq.(7). It is deeply effected by the bipartite graph matrix between original data and anchor points data. Therefore, large number categories clustering result will be better than smaller one. How to analyze and select anchor points will be our next research focus.

VI. CONCLUSION
In the paper, combining the anchor graph and spatial information of HSI, we propose a novel method, called fast spectral clustering based on spatial information (FSCS). Owing to using the spatial information of HSI, FSCS has better clustering result in UA, AA, OA, and Kappa. In order to deal with lager-scale HSI, we construct the similarity matrix based on the anchor graph. As a result, the computational complexity of our FSCS algorithm can be reduced to $O(nmd)$, while computational complexity of the spectral clustering is $O(n^2d)$. Furthermore, with the parameter-free strategy, the FSCS avoids tuning the heat-kernel parameter. Experiments demonstrate the efficiency and effectiveness of the proposed FSCS algorithm for large-scale HSI.

REFERENCES
[1] Z. Xue, P. Du, J. Li, and H. Su, "Sparse graph regularization for hyperspectral remote sensing image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 4, pp. 2351–2366, Apr. 2017.

[2] 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.

[3] H. Zhang, H. Zhai, L. Zhang, and P. Li, "Spectral–spatial sparse subspace clustering for hyperspectral remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3672–3684, Jun. 2016.

[4] R. Wang, F. Nie, and W. Yu, "Fast spectral clustering with anchor graph for large hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 2003–2007, Nov. 2017.

[5] Z. Wang, B. Du, L. Zhang, L. Zhang, and X. Jia, "A novel semisupervised active-learning algorithm for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 6, pp. 3071–3083, Jun. 2017.

[6] 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.

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

[8] J. A. Hartigan and M. A. Wong, "A K-means clustering algorithm," *Appl. Stat.*, vol. 28, no. 1, pp. 100–108, 1979.

[9] J. C. Bezdek, "Pattern recognition with fuzzy objective function algorithms," *Adv. Appl. Pattern Recognit.*, vol. 22, no. 1171, pp. 203–239, 1981.

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

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

[12] H. Zhai, H. Zhang, L. Zhang, P. Li, and A. Plaza, "A new sparse subspace clustering algorithm for hyperspectral remote sensing imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 1, pp. 43–47, Jan. 2017.

[13] F. He, R. Wang, and W. Jia, "Fast semi-supervised learning with anchor graph for large hyperspectral images," *Pattern Recognit. Lett.*, to be published. doi: 10.1016/j.patrec.2018.08.008.

[14] X. Yang, W. Yu, R. Wang, G. Zhang, and F. Nie, "Fast spectral clustering learning with hierarchical bipartite graph for large-scale data," *Pattern Recognit. Lett.*, to be published. doi: 10.1016/j.patrec.2018.06.024.

[15] G. Casalino and N. Gillis, "Sequential dimensionality reduction for extracting localized features," *Pattern Recognit.*, vol. 63, pp. 15–29, Mar. 2017.

[16] Y. Zhou, J. Peng, and C. L. P. Chen, "Dimension reduction using spatial and spectral regularized local discriminant embedding for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 2, pp. 1082–1095, Feb. 2015.

[17] W. Liu, J. He, and S.-F. Chang, "Large graph construction for scalable semi-supervised learning," in *Proc. ICML*, 2010, pp. 679–686.

[18] W. Zhao and S. Du, "Spectral–spatial feature extraction for hyperspectral image classification: A dimension reduction and deep learning approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4544–4554, Aug. 2016.

[19] X. Yang, G. Liu, Q. Yu, and R. Wang, "Stable and orthogonal local discriminant embedding using trace ratio criterion for dimensionality reduction," *Multimedia Tools Appl.*, vol. 77, no. 3, pp. 3071–3081, 2018.

[20] A. N. Ng, M. J. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. 14th Int. Conf. Neural Inf. Process. Syst., Natural Synth. (NIPS)*. Cambridge, MA, USA: MIT Press, 2001, pp. 849–856.

[21] F. Nie, W. Zhu, and X. Li, "Unsupervised large graph embedding," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 2422–2428.

[22] Y. Li, F. Nie, H. Huang, and J. Huang, "Large-scale multi-view spectral clustering via bipartite graph," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 2750–2756.

[23] F. Nie, X. Wang, M. I. Jordan, and H. Huang, "The Constrained Laplacian Rank algorithm for graph-based clustering," in *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 1969–1976.

[24] Y. Yuan, J. Lin, and Q. Wang, "Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2966–2977, Dec. 2016.

[25] Y. Li, W. Xie, and H. Li, "Hyperspectral image reconstruction by deep convolutional neural network for classification," *Pattern Recognit.*, vol. 63, pp. 371–383, Mar. 2017.

[26] R. S. Majdar and H. Ghassemian, "A probabilistic SVM approach for hyperspectral image classification using spectral and texture features," *Int. J. Remote Sens.*, vol. 38, no. 15, pp. 4265–4284, 2017.

[27] J. A. Richards and X. Jia, *Remote Sensing Digital Image Analysis: An Introduction*. New York, NY, USA: Springer-Verlag, 1999.
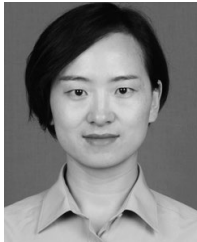
YIWEI WEI received the B.S. degree from Air Force Engineering University, in 2008, and the M.S. degree from the Rocket Force University of Engineering, Xi'an, Shaanxi, China, in 2013. She is currently pursuing the Ph.D. degree and a Lecturer. Her research interests include image processing, deep learning, and space whether.
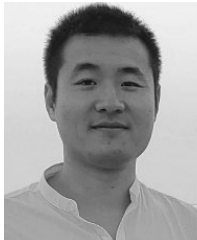
CHAO NIU received the B.S., M.S., and Ph.D. degrees from the Rocket Force University of Engineering, in 2007, 2010, and 2015, respectively, where he is currently a Lecturer. His research interests include military geophysics, intelligence analysis, and space whether.

YITING WANG received the B.S. degree in signal processing from the Chengdu University of Information Technology, in 2010, and the M.S. and Ph.D. degrees from the Rocket Force University of Engineering, in 2012 and 2017, respectively, where she is currently a Lecturer. Her research interests include hyperspectral image processing and application, deep learning, and quantitative remote sensing.

HONGXIA WANG received the B.S. degree in 1991 and the Ph.D. degree in 2009. She is currently a Professor and a Ph.D. supervisor. She has published more than 100 articles, 46 of them are indexed by SCI, EI, or ISTP. Her research interests include photoelectric information extraction, transportation, and processing.

DAIZHI LIU is currently a Leading Talented Missile Expert, also a Professor with the Rocket Force University of Engineering, and also a Ph.D. Student advisor. He has published three monographs and more than 200 articles. He enjoyed special government subsidy and received more than 50 awards of different kinds, including five second-class awards for military scientific and technological progress. He has been in charge of more than 30 national or regional important projects. He is an Editor-in-Chief of more than ten books.

● ● ●