Hyperspectral Image Classification Based on Active Learning and Spectral-Spatial Feature Fusion Using Spatial Coordinates

CAIHONG MU, (Associate Member, IEEE), JIAN LIU, YI LIU, AND YIJIN LIU

1Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education, Joint International Research Laboratory of Intelligent Perception and Computation, International Research Center for Intelligent Perception and Computation, School of Artificial Intelligence, Xidian University, Xi’an 710071, China
2School of Electronic Engineering, Xidian University, Xi’an 710071, China
Corresponding author: Caihong Mu (mucaihongxd@foxmail.com)
This work was supported by the National Natural Science Foundation of China under Grant 61672405, Grant 61876141, Grant U1701267, Grant 61772399, Grant 61773304, and Grant 61773300.

ABSTRACT In Hyperspectral image (HSI) classification, combining spectral information with spatial information has become an efficient measure to obtain good classification results, where spatial information is generally introduced in an unsupervised way or some complicated way. We introduce spatial coordinates as the spatial information in a simple supervised way and propose two HSI classification algorithms, where spatial coordinates of samples are regarded as the spatial features of samples. A spectral-spatial classification algorithm is proposed, named as HSI Classification Based on Spectral-Spatial Feature Fusion using Spatial Coordinates (SSFFSC). The HSI is divided into multiple small images in spatial dimension, and samples in each small image are randomly selected as training samples. Support vector machine (SVM) is used to classify the samples to obtain the probability of samples belonging to each class according to the spatial coordinate features and spectral features respectively. The probability features are further classified by SVM to achieve the final classification result. Considering that the performance of SSFFSC relies on the partition of HSI, SSFFSC is further combined with active learning (AL) as a new method named as HSI Classification Based on Active Learning and SSFFSC (SSFFSC-AL). Partition of HSI is omitted and the training samples are selected adaptively by AL’s sampling scheme. We find spatial coordinates are useful spatial information. SSFFSC and SSFFSC-AL run fast and improve the classification accuracy effectively by using the spatial coordinates as the spatial features. Experiments demonstrate that comparing with other algorithms, SSFFSC and SSFFSC-AL can obtain higher classification accuracy in less time.

INDEX TERMS Hyperspectral image classification, spatial coordinates, active learning, spectral-spatial classification.

I. INTRODUCTION Hyperspectral image (HSI) classification is important in hyperspectral remote sensing processing [1], [2]. Unlike traditional color images or multi-spectral images, hyperspectral images have high resolution and rich spectral information [3], which brings convenience to classification. But, it also causes “Curse of Dimensionality” and “Hughes Phenomenon” [4]–[6]. That is to say, when the number of training samples is limited, the characteristic of high-dimensional data of hyperspectral images will reduce the efficiency of the classifier; as the data dimension increases, the classification accuracy will increase first and then decrease. Therefore, to obtain good classification results, one has to increase the number of training samples or improve the performance of the algorithm. However, it is costly to acquire enough training samples with true labels in reality. So designing algorithms with high efficiency and high classification accuracy becomes the focus of hyperspectral classification.

In the research of hyperspectral image classification, it is found that the phenomenon of “salt and pepper” [7] occurs in the classification only based on spectral information.
The introduction of spatial information into the classification process, which is used to be combined with spectral information, can overcome this phenomenon and get a better classification result [8]. Therefore, how to introduce spatial information becomes an important issue that should be considered. The methods of using spatial information generally include morphological profiles [9]–[11], Markov random fields [12], [13], image segmentation [14], [15] and so on [16]. In [9], Marpu et al. proposed a method using extended attribute profiles (EAP) which was developed from morphology profiles. This method firstly reduced the dimension of data. Then the EAP with standard deviation attribute was obtained by parameters which were computed by training samples. Finally, EAP were classified to get the classification result. In [14], an approach for hyperspectral image classification using image segmentation was presented. The segmentation map of hyperspectral image was first obtained by Fractional-Order Darwinian Particle Swarm Optimization (FODPSO), and then the initial result of classification based on spectral information was got by support vector machine (SVM). Finally, the majority voting method was executed in each segmented area to assign the class with the largest number of votes to all samples in the area. In [16], Hadoux et al. proposed another method to use spatial information. Partial Least Square (PLS) was first used to reduce the dimension of hyperspectral image. Then anisotropic regularization (AR) was performed on each dimensional data after dimensionality reduction. The information in the neighborhood of the sample was used to correct the sample’s feature and finally the data obtained after processing were classified to get the final result. These methods can effectively reduce the effect of “salt and pepper” and obtain good classification results with high regional consistency. However, the way of introducing spatial information in these methods may be complicated, which thus increases the cost of computation to some extent.

For the supervised classification method, the establishment of the classifier model relies on the labeled samples. In order to obtain a classifier model with relatively high classification accuracy, it is necessary to construct a training sample set containing lots of labeled samples. However, manually labeling a large amount of samples in hyperspectral image is very time-consuming and costly in reality. In view of this problem, the active learning (AL) [17] was proposed. Among the unlabeled samples, the ability of the samples to improve the performance of the classifier is different, and some samples tend to have more abundant information. For the classifier model, such samples can improve the classification performance of the classifier more effectively. AL trains the classifier model by selecting samples with rich information through the sampling scheme, labeling them manually and adding them into the training set for the following training procedure, so that the performance of the classifier is rapidly improved, and a good classification ability can be obtained when the training sample set is small in scale.

Many hyperspectral image classification methods have been proposed based on active learning, which have achieved good classification accuracy [18]–[24]. In [24], hierarchical segmentation was combined with AL, which obtained a good result by combining the advantages of spatial information and active learning. Li et al. [25] proposed a method using loopy belief propagation (LBP) and AL for hyperspectral classification. This method established a discriminant field model, which was used to comprehensively consider the spectral information and spatial information, and then LBP was used to solve the problem of maximizing a posteriori marginal (MPM) to find the label of the sample. Finally, good classification result could be obtained by AL with fewer training samples.

In recent years, some methods [26]–[28] based on Deep neural networks have been designed for hyperspectral image classification. Cheng et al. [26] proposed a simple and effective deep spatial feature extraction method and designed a spectral-spatial feature extraction method by imposing metric learning on SVM training. Zhou et al. [27] proposed an effective framework named compact and discriminative stacked autoencoder (CDSAE), which could learn discriminative low-dimensional feature mappings and train an effective classifier progressively. Cheng et al. [28] proposed a novel and effective approach to learn a rotation-invariant CNN (RICNN) model for improving the performance of object detection, which was achieved by introducing and learning a new rotation-invariant layer on the basis of the existing CNN architectures.

Spatial coordinates are also a kind of spatial information. In [29], Goovaerts presented a method which used spatial coordinates in classification of hyperspectral image. Maximum likelihood method (ML) was used at first to classify samples by their spectral features, and then indicator kriging (IK) was used to estimate the probability of samples by using the probabilistic information in the neighborhood. These probability values were combined with the probability values based on spectral information and the final result was obtained. In this method, spatial coordinates were used to determine neighbor samples of the central sample. However, in our work we find that spatial coordinates, as features of the spatial distribution of the samples, can be utilized directly for classification.

In this paper, a novel spectral-spatial classification using spatial coordinate information is proposed, which is named as Hyperspectral Image Classification Based on Spectral-Spatial Feature Fusion using Spatial Coordinates (SSFFSC). In SSFFSC, spatial coordinates of samples are regarded as the spatial features of samples and utilized for classification. The samples show a certain local consistency in spatial distribution, making it possible to classify samples by spatial coordinates. Considering spatial coordinate features are related to the spatial distribution of samples, a strategy of constructing training set is designed that samples are selected from all spatial areas as the training samples to making full use of spatial coordinates as follows. The hyperspectral image is divided into multiple small images according to the spatial dimension, and samples in each small sample
image are randomly selected, which ensure that samples are selected as the training samples from all spatial areas. Then, the spatial coordinate features and spectral features will be combined to achieve better classification result. However, the range of values of spatial coordinate features is quite different from that of spectral features. So these features should not be directly connected end-to-end, which is not good for classification. In fact, SVM is used to classify the samples to get the probability of samples belonging to each class according to the spatial coordinate features and spectral features respectively, and thus the spectral features and spatial coordinate features could be converted into corresponding probability features. Finally, the probability features are further classified by SVM to achieve the final classification result.

In SSFFSC, it can be found that the classification performance of the algorithm relies on the partition of HSI. Considering this, SSFFSC is further combined with AL, which is named as Hyperspectral Image Classification Based on Active Learning and Spectral-Spatial Feature Fusion using Spatial Coordinates (SSFFSC-AL). In SSFFSC-AL, the training samples are selected by AL’s sampling scheme which selects samples contributing most to classifier. SSFFSC-AL could achieve good classification result when the size of training sample set is small.

The main contributions of this paper can be summarized as follows.

1) Spatial coordinates of samples are directly used to classify samples, the results of which are then combined with those based on spectral features. It is a simple way to implement spectral-spatial information fusion. Compared with other methods using spectral-spatial information, this algorithm can use training sample set with the same size to obtain higher classification accuracy with less running time.

2) Considering the phenomenon that the performance of the algorithm based on spatial coordinates is related to the division of images, we introduce AL into the framework, and the proposed SSFFSC-AL can get better classification result when the size of training sample set is small. Compared with other methods using AL, SSFFSC-AL is high in accuracy and fast in running speed.

The remainder of this paper is organized as follows. Section II describes the proposed SSFFSC and SSFFSC-AL. Section III provides the comparison experiments conducted and the analysis of the parameters. Finally, Section IV summarizes the work and provides the further research.

II. PROPOSED APPROACH

A. HYPERSONTAL IMAGE CLASSIFICATION BASED ON SPECTRAL-SPATIAL FEATURE FUSION USING SPATIAL COORDINATES

There are $n$ samples in the Hyperspectral remote sensing data set $X = \{x_1, x_2, \ldots, x_i, \ldots, x_n\}$. The dimension of spectral information of each sample is $d$ and the data set contains $L$ categories. Each sample has a pair of spatial coordinates corresponding to it. In the spatial dimension of hyperspectral image, the distribution of samples in local region shows the consistency of the spatial region, which is separable. Therefore, it is feasible to conduct supervised classification using the spatial coordinates of samples.

To implement spectral-spatial information fusion, directly combining the spectral information and the spatial coordinates to form the new features is not a feasible way. Because the range of the samples’ spatial coordinates is related to the size of hyperspectral image and its value is usually from tens to hundreds. The spectral information value of the samples should be much less than the value of the spatial coordinates. Therefore another feasible approach of feature fusion should be considered. After the support vector machine (SVM) obtains the classification result, the Sigmoid fitting can get the probability distribution of the samples belonging to different categories [30], which makes it possible to design an approach of feature fusion of spatial coordinates and spectral information as follows.

To achieve correct classification of samples in hyperspectral image, the supervised classification method needs to select a part of the samples from sample set of each category. In SSFFSC, the method of constructing the training sample set is proposed as shown in Fig. 1 to make full use of spatial coordinate information. The hyperspectral image is uniformly divided into 100 small hyperspectral images.
Algorithm 1 SSFFSC Algorithm

**Input:** Hyperspectral image data set \( X = \{ x_1, x_2, \ldots, x_i, \ldots, x_n \} \), which contains \( n \) samples of \( L \) categories. The spectral dimension of each sample is \( d \).

**Output:** Classification result of Hyperspectral image \( \hat{Y} = \{ \hat{y}_1, \hat{y}_2, \ldots, \hat{y}_i, \ldots, \hat{y}_n \} \).

1. Construct the training sample set and the testing sample set as shown in Fig. 1.
2. Get the spatial probability feature \( F_a (x_i) = \{ p_{a1}, p_{a2}, \ldots, p_{ak}, \ldots, p_{aL} \} \) of the sample \( x_i \).
3. Get the spectral probability feature \( F_e (x_i) = \{ p_{el1}, p_{el2}, \ldots, p_{eik}, \ldots, p_{eIL} \} \) of the sample \( x_i \).
4. Combine the spatial probability feature with the spectral probability feature to get the spectral-spatial probability feature \( F_{ea} (x_i) = \{ p_{el1}, p_{el2}, \ldots, p_{eik}, p_{ai1}, p_{ai2}, \ldots, p_{aIL} \} \).
5. Classify the sample \( x_i \) based on the spectral-spatial probability feature by SVM and obtain the classification result \( \hat{y}_i = \text{SVM} (F_{ea} (x_i)) \).

According to the spatial relationship between pixels in the image, and the 100 small hyperspectral images can be used to obtain 100 small data sets. In each small data set, a certain percentage of samples are randomly selected as training samples and all the training samples selected from the 100 small data sets make up the final training sample set. All the other samples in hyperspectral image data set are assembled into a testing sample set.

After the construction of training sample set and testing sample set, spatial coordinates and spectral information are utilized respectively in the following classification. First, spatial coordinates are used as features to be classified by SVM. Then the probability values of each sample belonging to each category are obtained, which can be regarded as a new feature of the sample. For the sample \( x_i, i = 1, 2, \ldots, n \), its probability values can be written as a vector as follows:

\[
F_a (x_i) = \{ p_{a1}, p_{a2}, \ldots, p_{ak}, \ldots, p_{aL} \} \quad (1)
\]

where \( p_{aik} \) is obtained by the SVM classifier based on the spatial coordinates and indicates the probability that the sample \( x_i \) belongs to the class \( k (k = 1, 2, \ldots, L) \). Thus, \( F_a (x_i) \) can be treated as the spatial probability feature of \( x_i \).

Meanwhile, spectral information is also used as features to be classified by SVM. Principal component analysis (PCA) is first performed on the spectral information and partial principal components are extracted as the spectral features. Then the spectral features are classified by SVM and the probability values of each sample belonging to each category are obtained and recorded as the spectral probability feature vector as follows:

\[
F_e (x_i) = \{ p_{el1}, p_{el2}, \ldots, p_{eik}, \ldots, p_{eIL} \} \quad (2)
\]

where \( p_{eik} \) is obtained by the SVM classifier based on the spectral features and indicates the probability that the sample \( x_i \) belongs to the class \( k (k = 1, 2, \ldots, L) \).

In this way, both spectral information and spatial coordinates are converted into probability feature vectors by SVM classifiers, making the fusion possible. The spectral probability feature vector is combined with the spatial probability feature vector to get the spectral-spatial probability feature as shown in formula (3).

\[
F_{ea} (x_i) = \{ p_{el1}, p_{el2}, \ldots, p_{eik}, p_{ai1}, p_{ai2}, \ldots, p_{aIL} \} \quad (3)
\]

The final classification result is then obtained through the SVM classification based on the spectral-spatial probability features.

Algorithm 1 shows the steps of hyperspectral image classification based on spectral-spatial feature fusion using spatial coordinates.

**B. COMBINED WITH ACTIVE LEARNING**

The classification result of SVM based on spatial coordinates is dependent on the division of images. In this section, another version of classification using the active learning method is proposed to select the samples with high uncertainties corresponding to the classifiers, where the division of the image is no longer needed. All these selected samples will be labeled manually to retrain the classifiers, which can improve the classifiers’ performance rapidly. Based on this idea, we propose a sampling scheme which selects samples with high uncertainties corresponding to spectral classifier and spatial classifier, where spectral classifier means SVM classifier using spectral information as features and spatial classifier means SVM classifier using spatial coordinates as features. Finally, the image is classified based on the spectral-spatial probability.

After the SVM classification is performed, in addition to prediction labels of all the samples, the probability value of each sample belonging to the class corresponding to the prediction label can also be obtained. We define the probability value of sample \( x_i \) corresponding to spectral classifier’s prediction label as the \( P_e (x_i) \) and the probability value corresponding to spatial classifier’s prediction label as the \( P_a (x_i) \), where \( x_i \in X, i = 1, 2, \ldots, n \).

Calculate sampling measurement value \( W \) of all samples according to (4), where \( \alpha \) is sampling weight parameter, \( 0 \leq \alpha \leq 1 \).

\[
W (x_i) = \alpha \ast P_e (x_i) + (1 - \alpha) \ast P_a (x_i) \quad (4)
\]

\( P_e (x_i) \) and \( P_a (x_i) \) are the probability values that the sample \( x_i \) is judged as the corresponding class by spectral classifier and spatial classifier, respectively. They also indicate
Algorithm 2 SSFFSC-AL Algorithm

**Input:** Hyperspectral image data set \( X = \{ x_1, x_2, x_3, \cdots, x_n \} \), which contains \( n \) samples for \( L \) categories. The spectral dimension of each sample is \( d \).

**Output:** Classification result of Hyperspectral image \( \hat{Y} = \{ \hat{y}_1, \hat{y}_2, \hat{y}_3, \cdots, \hat{y}_n \} \).

1: Select a part of samples randomly in each class to constitute the initial training sample set. All the rest of samples constitute the testing sample set.

2: Perform PCA on spectral information and extract the first \( m \) principal components as spectral features.

3: Classify the samples based on spatial coordinates by SVM to obtain the classification result and the probability value \( P_a \) corresponding to the result.

4: Classify the samples based on spectral information by SVM to obtain the classification result and the probability value \( P_e \) corresponding to the result.

5: Calculate the sampling measurement value \( W \) according to (4) using \( P_a \) and \( P_e \).

6: Select some samples with the smallest \( W \) in the testing sample set and manually label these samples. These samples are removed from the testing sample set and added to the training sample set.

7: Go to step 3 if the termination condition (the preset maximal number of training samples is reached) of AL is not met.

8: Obtain the spatial probability feature \( F_a (x_i) \) of \( x_i \).

9: Obtain the spectral probability feature \( F_e (x_i) \) of \( x_i \).

10: Combine the spatial probability feature with the spectral probability feature to get the spectral-spatial probability feature \( F_{ea} (x_i) = (p_{e1}, p_{e2}, \cdots, p_{el}, p_{a1}, p_{a2}, \cdots, p_{al}) \).

11: Classify the sample \( x_i \) based on the spectral-spatial probability feature by SVM and obtain the classification result \( \hat{y}_i = \text{SVM} (F_{ea} (x_i)) \).

III. EXPERIMENTAL RESULTS AND DISCUSSION

To verify the effectiveness of the proposed method which uses spatial information, we compare the proposed SSFFSC with PSVM (Spectral Classification using PCA and SVM), SCSVM (Spatial Classification using Spatial Coordinates and SVM), FODPSO-SVM (SVM classification combined with segmentation maps obtained by fractional-order Darwin particle swarm optimization algorithms) [14], and PLS-AR-SVM (Spatial information classification using information near the sample to correct the characteristics of the sample, and using partial least squares to reduce the size of the hyperspectral image.) [16]. PSVM only uses spectral information for classification and SCSVM only uses spatial coordinates for classification. FODPSO-SVM and PLS-AR-SVM use both spectral and spatial information to classify the samples.

Furthermore, to verify the effectiveness of the proposed SSFFSC-AL, we compare it with some methods which also use active learning. These methods are Random-SVM-AL (Hyperspectral image classification using active learning based on random sampling), MS-SVM-AL (Hyperspectral image classification using active learning based on margin sampling), MCLU-SVM-AL (Hyperspectral image classification using active learning based on margin sampling), and MCLU-SVM-AL (Hyperspectral image classification using active learning based on margin sampling).
classification using active learning based on multiclass-level-uncertainty sampling) and MPM-LBP-AL (A method for hyperspectral classification using cyclic belief propagation and AL, and using LBP to solve the problem of maximizing posterior edges to find sample labels) [25]. Among them, Random-SVM-AL, MS-SVM-AL and MCLU-SVM-AL only use spectral information during the classification, while MPM-LBP-AL and SSFFSC-AL use both spectral information and spatial information. Random-SVM-AL, MS-SVM-AL and MCLU-SVM-AL were performed by the AL Toolbox [31].

All experiments are performed on a personal computer with i5-3210M central processing unit, 8-G memory, and the 64-bit Windows 7 using Matlab R2008a and Python 3.6.0. Libsvm [32] is adopted to implement SVM using OAOSVMs [33]. The overall accuracy (OA), average accuracy (AA), kappa coefficient (Kappa) [34] and classification accuracy of each class are employed to evaluate the classification performance. OA is the overall accuracy for all classes and is defined as follows:

$$\text{OA} = \frac{\sum x_{\text{test,correct}}^{(i)}}{N}$$  \hspace{1cm}(5)$$

where $x_{\text{test,correct}}^{(i)}$ is the $i$-th correctly classified test sample, $N$ is the total number of the test samples. AA represents the averaged accuracy of each class and is defined as:

$$\text{AA} = \frac{1}{M} \sum_{i=1}^{M} \frac{\sum x_{\text{correct}}^{(j)}}{N_i}$$  \hspace{1cm}(6)$$

where $M$ is the class number of a data set, $N_i$ is the total test sample number of the $i$-th class, and $x_{\text{correct}}^{(j)}$ is the $j$-th correctly classified test sample of class $i$. Kappa coefficient [34] is a statistical measure of agreement degree. The higher these measurement metrics are, the better the classification performance is.

The remainder of this section is organized as follows. Section III-A introduces the data sets, which comprise the AVIRIS Indian Pines and ROSIS Pavia University data set. They are two widely used benchmarks for hyperspectral image classification. Section III-B provides the comparison experiments of SSFFSC and four other algorithms. Section III-C shows the comparison experiments of SSFFSC-AL and four other algorithms using AL. Finally, Section III-D conducts experiments to analyze the parameters.

### A. HYPERSONTICAL DATA SETS

In this paper, two frequently-used hyperspectral data sets, Indian Pines data set and Pavia University data set, are tested to validate the effectiveness of the proposed algorithms.

1) Indian Pines data set. This hyperspectral image data set was collected by the AVIRIS sensor over the Indian Pines region in Northwestern Indiana in 1992. This scene has the size of $145 \times 145$ and the spatial resolution of 20 meters per pixel. The scene comprises 220 spectral channels in the wavelength ranging from 0.4 to $2.5 \mu m$. After several spectral bands are removed due to noise and water absorption phenomena, the remaining 200 radiance channels are used in the experiments. The Indian Pines data set contains 10366 samples with 16 mutually exclusive ground-truth classes. Fig. 3 (a) shows the grayscale map of the 170th channel in the 200 channels, while Fig. 3 (b) shows the ground-truth map available for the scene. Table 1 shows the meaning of each class and the number of samples in each class in Indian Pines data set.

2) Pavia University data set. This hyperspectral data set was collected by the ROSIS optical sensor over the urban area of the University of Pavia, Italy. The image size is $610 \times 340$, with very high spatial resolution of 1.3 meters per pixel. The number of data channels in this image is 115 with...
spectral ranging from 0.43 to 0.86μm. After 12 channels affected by noise are removed, the remaining 103 channels are used in the experiments. Pavia University data set contains 42776 samples with 9 classes. Fig. 3 (c) shows the grayscale map of the 60th channel in the remaining 103 channels, while Fig. 3 (d) shows the ground-truth map available for the scene. Table 2 shows the meaning of each class and the number of samples in each class in Pavia University data set.

### TABLE 1. The information of each class of samples in Indian Pines.

| #  | Class               | Samples |
|----|---------------------|---------|
| 1  | Alfalfa             | 54      |
| 2  | Corn-no till        | 1434    |
| 3  | Corn-min till       | 834     |
| 4  | Corn                | 234     |
| 5  | Grass-pasture       | 497     |
| 6  | Grass-trees         | 747     |
| 7  | Grass-pasture-mowed | 26      |
| 8  | Hay-windrowed       | 489     |
| 9  | Oats                | 20      |
| 10 | Soybean-no till     | 968     |
| 11 | Soybean-min till    | 2468    |
| 12 | Soybean-clean       | 614     |
| 13 | Wheat               | 212     |
| 14 | Woods               | 1294    |
| 15 | Buildings-Grass-Trees-Drives | 380 |
| 16 | Stone-Steel-Towers  | 95      |

### TABLE 2. The information of each class of samples in Pavia University.

| #  | Class            | Samples |
|----|------------------|---------|
| 1  | Asphalt          | 6631    |
| 2  | Meadows          | 18649   |
| 3  | Gravel           | 2099    |
| 4  | Trees            | 3064    |
| 5  | Painted metal sheets | 1345     |
| 6  | Bare Soil        | 5029    |
| 7  | Bitumen          | 1330    |
| 8  | Self-Blocking Bricks | 3682     |
| 9  | Shadows          | 947     |

**B. COMPARISON OF SSFFSC WITH FOUR OTHER ALGORITHMS**

In the experiment of comparing SSFFSC with other algorithms, the training samples and the testing samples are selected according to the method described in Fig. 1. Since the Indian Pines data set has only 26 and 20 samples in categories 7 and 9, while the Pavia University data set does not have a category with a particularly small sample size. Taking into account the different characteristics of the two data sets, we use the following method to select training samples in accordance with conventions. For Indian Pines data set, 10% of samples of each class are selected from each segmented small data set as training samples. The total number of the training samples and the testing samples are shown in TABLE 3. For Pavia University data set, in each segmented small data set, training samples are selected according to the corresponding proportion which satisfies that about 50 samples are finally selected as training samples from each class. The final number of training samples and testing samples are shown in TABLE 4. After using PCA in PSVM and SSFFSC, only the main first 15 principal components are extracted. The kernel function of SVM in all algorithms is chosen as radial basis kernel function (RBF) and involves two parameters: one is the penalty coefficient $C$, and the other is the parameter $\gamma$. The two coefficients in this paper are obtained by 5-fold cross validation method. To make the experimental results objective and fair, each algorithm has been simulated for 10 times with different training samples and testing samples, and then the average results are obtained and compared. All methods, containing SSFFSC and other four algorithms, are simulated under the same training sample set and testing sample set every time. TABLE 3 and TABLE 4 show the results of the experiments of Indian Pines and Pavia University. Fig. 4 and Fig. 5 provide the classification maps of two data set respectively.

From TABLE 3 and TABLE 4, it can be found that PSVM only using spectral information is inferior to other four algorithms in terms of OA, AA and Kappa. Unlike PSVM, FODPSO-SVM, PLS-AR-SVM and SSFFSC all introduce
spatial information in classification. These three algorithms achieve better classification result than PSVM, proving that the introduction of spatial information could effectively improve the classification accuracy. It could also be observed from Fig. 4 and Fig. 5 that the results of these three algorithms have better consistency in local regions, less isolated points and better classification results. Comparing SCSVM with others, it can be found that SCSVM, which only makes use of spatial coordinates, can obtain good classification result. SCSVM performs much better than PSVM on both data sets, and it even performs better than FODPSO-SVM and PLS-AR-SVM on data set of Indian Pines. However, as for data set of Pavia University, the performance of SCSVM degrades, and its accuracy is lower than that of FODPSO-SVM and PLS-AR-SVM. Why does SCSVM perform differently on two data sets? The reason can be found by comparing the ground-truth maps of Indian Pines and Pavia University. In the map of Indian Pines, the samples of the same class are generally concentrated in several blocks, which are best suited for classification based on spatial coordinates, while in the map of Pavia University, the samples of the same class may be distributed dispersedly in different regions, which are difficult to be classified by spatial coordinates. From TABLE 4, it can be seen that samples of Trees (#4) and shadows (#9) are classified by SCSVM with low accuracy, which are distributed dispersedly. In contrast, PSVM obtains

| Class | Training-Test Samples | PSVM | SCSVM | FODPSO-SVM | PLS-AR-SVM | SSFFSC |
|-------|-----------------------|------|-------|------------|------------|--------|
| 1     | -                     | -    | -     | -          | -          | -      |
| 2     | -                     | -    | -     | -          | -          | -      |
| 3     | -                     | -    | -     | -          | -          | -      |
| 4     | -                     | -    | -     | -          | -          | -      |
| 5     | -                     | -    | -     | -          | -          | -      |
| 6     | -                     | -    | -     | -          | -          | -      |
| 7     | -                     | -    | -     | -          | -          | -      |
| 8     | -                     | -    | -     | -          | -          | -      |
| 9     | -                     | -    | -     | -          | -          | -      |

OA (%) - 80.75 98.56 93.63 97.92 98.96
AA (%) - 66.11 91.23 78.27 91.18 96.27
Kappa - 0.779 0.984 0.927 0.976 0.988
Time(s) - 3.2 0.9 14.2 12.7 4.8

TABLE 4. The classification result and computation time of each method in Pavia University.
good results on these two classes. But PSVM performs poorly on most other classes (#1, 3, 6, 7, 8), while SCSVM can obtain reasonable results on these classes. As for SSFFSC that combines spectral information with spatial coordinates, it achieves relatively good classification results for each class on both data sets, which validates the effectiveness of the fusion of the spectral information and spatial coordinates.

As for three algorithms using both spectral and spatial information, the OA, AA and Kappa of SSFFSC are larger than those of FODPSO-SVM and PLS-AR-SVM on two data sets. It proves that the proposed method using spatial information in classification can effectively improve the classification accuracy. From Fig. 4 and Fig. 5, it can be seen that compared with FODPSO-SVM and PLS-AR-SVM, SSFFSC obtains a better classification result in the area of blocks in the image and there are fewer misclassified samples in the local regions. This phenomenon indicates that compared with FODPSO-SVM using the segmentation and PLS-AR-SVM using the anisotropic diffusion regularization method, SSFFSC making use of spatial coordinates performs better and can effectively reduce misclassified samples and obtain higher classification accuracy with better regional consistency. From TABLE 3 and TABLE 4, it can be seen that SSFFSC obtains relatively good classification results for each class. For OA, AA and Kappa, the proposed method obtains the best result and it has obvious advantages over others especially on AA and Kappa.

Comparing the running time of FODPSO-SVM and PLS-AR-SVM with that of SSFFSC, it can be found from TABLE 3 and TABLE 4 that SSFFSC takes the least time, which is reasonable considering that SSFFSC combines PSVM and SCSVM and both of them are simple and take less time.

On the whole, it can be found that SSFFSC is superior to four other algorithms on two data sets, which demonstrates that spatial coordinates are very useful features for classification. If combined with spectral information, spatial coordinates can improve the classification result further.

Considering the performance of SSFFSC is dependent on the division of images, active learning is a reasonable measure to make SSFFSC perform better, the effect of which will be discussed in the following section C.

C. COMPARISON OF SSFFSC-AL WITH FOUR OTHER ALGORITHMS USING ACTIVE LEARNING

In the experiment of comparing SSFFSC-AL with four other algorithms using active learning, the training samples and testing samples are selected as follows. In the experiments on Indian Pines data set, 160 samples are selected at the initial time, and each class has 10 samples. In the process of active learning, 50 samples unlabeled are chosen adaptively by AL's sampling scheme to be manually labeled in each iteration. The 50 samples selected each time may come from various categories, and a total of 900 samples are selected as the training samples during the iteration. Therefore, in total, 1060 samples are treated as the training samples. In the experiments on Pavia University data set, 90 samples are selected at first, and each class contains 10 samples. Then during the process of active learning, 10 samples unlabeled are chosen to be manually labeled in each iteration. The number of samples labeled manually is 250 and the total number of training samples is 340. The SVM used in algorithms uses RBF as the kernel function and the parameters are set according to experience. The sampling weight parameter \( \alpha \) is set as 0.3. The first 15 principal components are extracted by PCA as spectral features.

TABLE 5 shows the classification result and time of these five methods on Indian Pines data set. Fig. 6 shows the classification overall accuracy curves and Fig. 7 shows the classification maps on Indian Pines data set. TABLE 6 shows the classification result and time of these five methods on Pavia University data set. Fig. 8 shows the classification overall accuracy curves and Fig. 9 shows the classification maps on Pavia University data set.

From Fig. 6 and Fig. 8, it can be seen that SSFFSC-AL is superior to Random-SVM-AL, MS-SVM-AL and MCLU-SVM-AL. The spatial information is not introduced in these three methods and only spectral information is used for classification. So the results of these three methods are not so good. The proposed method SSFFSC-AL uses spatial coordinates as spatial features, and combines spatial features with spectral features for classification, which makes the classification result has obvious advantages.
TABLE 5. The final classification results and computation time of all the methods with active learning on Indian Pines.

| #  | Random-SVM-AL (Training-Test Samples) | MS-SVM-AL (Training-Test Samples) | MCLU-SVM-AL (Training-Test Samples) | MPM-LBP-AL (Training-Test Samples) | SSFFSC-AL (Training-Test Samples) |
|----|---------------------------------------|-----------------------------------|-------------------------------------|------------------------------------|-----------------------------------|
| 1  | 77.50 (14-40)                         | 79.49 (15-39)                     | 71.43 (12-42)                      | - (54-0)                           | 100 (28-26)                       |
| 2  | 74.58 (128-1306)                      | 82.14 (202-1232)                  | 84.52 (200-1234)                   | 99.85 (94-1340)                    | 100 (120-1314)                    |
| 3  | 75.88 (96-738)                        | 82.31 (116-718)                   | 80.64 (111-723)                    | 100 (102-732)                      | 100 (80-754)                      |
| 4  | 68.00 (34-200)                        | 66.50 (37-197)                    | 85.86 (36-198)                     | 100 (58-176)                       | 100 (51-183)                      |
| 5  | 91.96 (49-448)                        | 94.17 (34-463)                    | 95.51 (29-468)                     | 100 (75-422)                       | 100 (66-431)                      |
| 6  | 95.39 (74-673)                        | 96.90 (38-709)                    | 97.75 (35-712)                     | 100 (69-678)                       | 100 (66-681)                      |
| 7  | 93.33 (11-15)                         | 87.50 (10-16)                     | 93.75 (10-16)                      | 100 (26-0)                         | 100 (24-2)                        |
| 8  | 96.59 (49-440)                        | 97.64 (22-467)                    | 97.45 (18-471)                     | 100 (81-408)                       | 100 (33-456)                      |
| 9  | 100 (11-9)                            | 100 (10-10)                       | 100 (10-10)                        | 100 (20-0)                         | 100 (19-1)                        |
| 10 | 82.72 (94-874)                        | 84.55 (107-861)                   | 90.01 (117-851)                    | 88.69 (75-893)                     | 100 (100-868)                     |
| 11 | 83.16 (218-2250)                      | 88.08 (253-2215)                  | 90.04 (259-2209)                   | 98.76 (121-2347)                   | 100 (192-2276)                    |
| 12 | 87.41 (58-556)                        | 87.07 (58-526)                    | 91.39 (80-534)                     | 100 (76-538)                       | 100 (91-523)                      |
| 13 | 98.94 (24-188)                        | 98.49 (13-199)                    | 100 (14-198)                       | 100 (78-134)                       | 100 (39-173)                      |
| 14 | 91.37 (124-1170)                      | 96.30 (51-1243)                   | 95.50 (49-1245)                    | 100 (87-1207)                      | 100 (52-1242)                     |
| 15 | 70.90 (57-323)                        | 65.24 (52-328)                    | 70.29 (67-313)                     | 100 (87-293)                       | 100 (69-311)                      |
| 16 | 97.37 (19-76)                         | 93.98 (12-83)                     | 95.12 (13-82)                      | 100 (78-17)                        | 100 (30-65)                       |

OA (%) 84.26 88.01 90.06 98.56 100  
AA (%) 86.57 87.52 89.95 - 100  
Kappa 0.820 0.863 0.887 0.983 1.00  
Time(s) 196.3 397.3 412.3 2538.9 93.6

FIGURE 6. The curves of overall accuracy on Indian Pines.

Random-SVM-AL, MS-SVM-A and MCLU-SVM-AL are all performed by the toolbox of AL Toolbox. This toolbox uses a multi-class SVM based on OAASVMS strategy. In the

TABLE 5 and TABLE 6, the time of these three methods are longer than that of SSFFSC-AL. However, the complexity of Random-SVM-AL should be lower than that of SSFFSC-AL. Longer running time is due to AL Toolbox. So we only consider the comparison of time between SSFFSC-AL and MPM-LBP-AL here.

Comparing SSFFSC-AL with MPM-LBP-AL on Indian Pines data set, it can be observed from Fig. 6 that the performance of SSFFSC-AL is better than that of MPM-LBP-AL. When the number of training samples is small, SSFFSC-AL has particularly obvious advantages. When the number of training samples is 610, the accuracy of SSFFSC-AL reaches 99.58%. At the same time, the accuracy of MPM-LBP-AL is only 94.10%, which is much lower than that of SSFFSC-AL. After a large number of training samples are
TABLE 6. The final classification results and computation time of all the methods with active learning on Pavia University.

| #  | Random-SVM-AL (Training-Test Samples) | MS-SVM-AL (Training-Test Samples) | MCLU-SVM-AL (Training-Test Samples) | MPM-LBP-AL (Training-Test Samples) | SSFFSC-AL (Training-Test Samples) |
|----|---------------------------------------|----------------------------------|-------------------------------------|-----------------------------------|----------------------------------|
| 1  | 93.41 (51-6580)                       | 6.67 (15-6616)                  | 88.54 (58-6573)                     | 99.21 (68-6563)                  | 96.60 (67-6564)                  |
| 2  | 92.02 (109-18540)                     | 94.29 (117-18532)              | 86.72 (52-18597)                    | 99.97 (81-18568)                 | 98.38 (44-18605)                 |
| 3  | 44.08 (21-2078)                       | 34.68 (20-2079)                 | 50.27 (34-2065)                     | 92.98 (61-2038)                  | 99.37 (30-2069)                  |
| 4  | 88.94 (33-3031)                       | 85.88 (25-3039)                 | 79.17 (15-3049)                     | 99.83 (49-3015)                  | 97.62 (37-3027)                  |
| 5  | 93.96 (22-1323)                       | 95.42 (12-1333)                 | 93.93 (10-1335)                     | 100 (55-1290)                    | 99.92 (13-1332)                  |
| 6  | 40.56 (39-4990)                       | 47.18 (71-4958)                 | 60.64 (47-4982)                     | 100 (52-4977)                    | 99.98 (23-5006)                  |
| 7  | 27.00 (19-1311)                       | 29.98 (16-1314)                 | 50.58 (37-1293)                     | 99.77 (50-1280)                  | 97.38 (31-1299)                  |
| 8  | 49.26 (26-3656)                       | 84.02 (52-3630)                 | 76.26 (51-3631)                     | 98.78 (67-3615)                  | 100 (69-3613)                    |
| 9  | 22.04 (20-927)                        | 36.04 (12-935)                  | 58.18 (36-911)                      | 100 (36-911)                     | 97.50 (57-890)                   |

OA (%) | 76.45 | 67.48 | 79.24 | 99.40 | 98.43 | 98.43
AA (%) | 61.25 | 57.13 | 71.59 | 98.95 | 98.53 | 98.53
Kappa  | 0.680 | 0.567 | 0.724 | 0.992 | 0.979 | 0.979
Time(s) | 698.1 | 1302.7 | 1451.9 | 30020 | 96.1 |

FIGURE 8. The curves of overall accuracy on Pavia University.

selected, both methods achieve good results at last. According to TABLE 5 and Fig. 7, the overall accuracy of MPM-LBP-AL is 98.56 and the overall accuracy of SSFFSC-AL is 100%. At the same time, SSFFSC-AL requires less time. SSFFSC-AL needs 93.6 seconds, while MPM-LBP-AL needs 2538.9 seconds, which is much longer.

Comparing SSFFSC-AL with MPM-LBP-AL on Pavia University, the SSFFSC-AL’s performance is worse than the MPM-LBP-AL’s performance. When the number of training samples is small, the accuracy of SSFFSC-AL is obviously worse than that of MPM-LBP-AL. When the number of training samples is large, the final results of these two algorithms reach a high level. According to TABLE 6 and Fig. 9, the final accuracy of MPM-LBP-AL is 99.40% and the overall accuracy of SSFFSC-AL is 98.43%. The accuracy of the proposed method is slightly lower than that of MPM-LBP-AL. But compared with other three methods, SSFFSC-AL has an obvious advantage in overall accuracy.

Comparing the overall accuracy curves of SSFFSC-AL and MPM-LBP-AL on the two data sets, it can be seen that...
SSFFSC-AL has an obvious advantage on Indian Pines data set. But the performance of this algorithm is worse than that of MPM-LBP-AL on Pavia University data set. SSFFSC-AL mainly classifies samples by using spectral-spatial feature fusion. The spectral features are used to obtain the final classification with the spatial features (spatial coordinates). The improvement of result based on spatial coordinates can effectively improve the overall accuracy. However, the spatial coordinates of samples are related to the spatial distribution of samples. In Indian Pines image, the samples of the same class are generally concentrated in several blocks, which is advantageous for the classification based on the coordinate features. But in Pavia University, there are a lot of small and disorderly areas and spots. In this case, classification based on coordinate features is difficult to achieve its greatest advantage. Nevertheless, SSFFSC-AL still can get a result which is close to that of MPM-LBP-AL. In addition, SSFFSC-AL has a great advantage in time. According to TABLE 6, SSFFSC-AL requires 96.1 seconds for classification on Pavia University data set, while MPM-LBP-AL needs 30020 seconds.

D. PARAMETER SENSITIVITY ANALYSIS

1) NUMBER OF DIVISIONS USED IN SSFFSC

As mentioned above, good performance of SSFFSC mainly benefits from the using of coordinates, which are related to the spatial distribution of samples. Therefore, the number of divisions of HSI is closely related to the performance of SSFFSC. To verify the impact of the number of divisions of HSI on the SSFFSC, SSFFSC was performed on both data sets by using the number of divisions $4 \times 4, 5 \times 5, \cdots, 16 \times 16$, respectively. Fig. 10 shows the results of overall accuracy versus the number of divisions of HSI. From the experimental curve in Fig. 10, it can be seen that when the number of divisions is smaller than 100, OA gradually rises, and when the number of divisions is greater than 100, the curve of OA tends to be flat for Indian Pines or even descending for Pavia University. Therefore, 100 is a reasonable choice and is adopted in this paper.

2) WEIGHT PARAMETER OF SSFFSC-AL

To verify the influence of sampling weight parameter $\alpha$, SSFFSC-AL is performed in the case of 310, 460, 610, 760, 910 and 1060 training samples with different sampling weight parameter on Indian Pines data set. Fig. 11 shows the results of this experiment. Similarly, on Pavia University data set, SSFFSC-AL is performed in the case of 140, 190, 240 and 340 training samples with different sampling weight parameter and Fig. 12 shows the results.

On the Indian Pines data set, the influence of the sampling weight parameter $\alpha$ on the classification accuracy is approximately the same in these six cases. The sampling weight parameter has a very important influence on the final overall accuracy. When $\alpha = 0$, the algorithm only considers the uncertainty of samples belonging to the classifier using
spatial coordinates. Although 100% overall accuracy is got when $\alpha = 0$ and the number of training samples is 1060, it can be seen that for the case that $\alpha = 0$, the overall accuracy decreases more greatly than other cases as the number of training samples decreases. When $\alpha = 1$, the final overall accuracy is lower than most other cases. When parameter $\alpha = 1$, it means that the method only considers the classifier based on spectral features when selecting samples. The samples selected could not enhance the performance of the classifier based on spatial coordinates. So the final overall accuracy is not high in this case. According to Fig. 11, it shows that when the number of training samples is 1060, SSFFSC-AL can obtain a good result when the value of the parameter $\alpha$ is between 0 and 0.5.

On the Pavia University data set, the sampling weight parameter $\alpha$ has a great influence on the final overall accuracy. When the parameter $\alpha$ is 0 or 1, the accuracy of the proposed method is low. Because it means that when selecting samples, the method only considers the classifier based on spatial features or the classifier based on spectral features. The samples selected by only considering one classifier could not enhance the performance of the other classifier. Therefore, the final overall accuracy of SSFFSC in these cases is low. According to Fig. 12, SSFFSC-AL can obtain a good result when the number of training samples is 340 in the case that the parameter $\alpha$ is 0.3, 0.5 and 0.6.

IV. CONCLUSION AND FUTURE RESEARCH LINES

In this paper, a hyperspectral image classification method based on spatial coordinate information named as SSFFSC is proposed. Considering that the commonly methods introducing spatial information is complicated and time consuming, SSFFSC algorithm implements a simple way to classify the samples based on spatial coordinates using supervised classification method. In addition, active learning is introduced to SSFFSC to improve the performance further, which is called SSFFSC-AL. Compared with other methods using spectral-spatial information, SSFFSC can more effectively solve the phenomenon of “salt and pepper” by combining spatial information with spectral information. It obtains better classification results with less time. Comparing SSFFSC-AL with other methods using active learning, it can obtain higher classification accuracy in images where the samples of the same class are generally concentrated in several blocks. For the images where the samples of the same class are distributed disperedly in different regions, its performance may degrade slightly, but it also can obtain the satisfactory classification results. Moreover, the running time of SSFFSC-AL is significantly less than that of other methods.

In the sampling scheme used in SSFFSC-AL, the sampling weight parameter $\alpha$ is static and invariable. It could not change when the number of training samples increases. In the future, we will design the sampling weight parameter as an adaptive parameter which can change adaptively with the change of training samples set, thus to improve the performance of classifiers more effectively.

REFERENCES

[1] W. Yang, K. Hou, B. Liu, F. Yu, and L. Lin, “Two-stage clustering technique based on the neighboring union histogram for hyperspectral remote sensing images,” IEEE Access, vol. 5, no. 99, pp. 5640–5647, 2017.
[2] Y. Dong, B. Du, L. Zhang, and L. Zhang, “Exploring locally adaptive dimensionality reduction for hyperspectral image classification: A maximum margin metric learning aspect,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 10, no. 3, pp. 1136–1150, Mar. 2017.
[3] P. Liu, P.-J. Du, and K. Tan, “A novel remotely sensed image classification based on ensemble learning and feature integration,” J. Infr. Millim. Waves., vol. 33, no. 3, pp. 311–317, Jun. 2014.
[4] G. Hughes, “On the mean accuracy of statistical pattern recognizers,” IEEE Trans. Inf. Theory, vol. IT-14, no. 1, pp. 55–63, Jan. 1968.
[5] F. Melgani and L. Bruzzone, “Classification of hyperspectral remote sensing images with support vector machines,” IEEE Trans. Geosci. Remote Sens., vol. 42, no. 8, pp. 1778–1790, Aug. 2004.
[6] L. Bruce, C. Koger, and J. Li, “Dimensionality reduction of hyperspectral data using discrete wavelet transform feature extraction,” IEEE Trans. Geosci. Remote Sens., vol. 40, no. 10, pp. 2331–2338, Jun. 2002.
[7] L. Hu, C. Qi, S. Chen, and Q. Wang, “An improved heuristic optimization algorithm for feature learning based on morphological filtering and its application,” IEEE Access, vol. 6, pp. 22754–22763, 2018.
[8] M. Fauvel, Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, “Advances in spectral–spatial classification of hyperspectral images,” Proc. IEEE, vol. 101, no. 3, pp. 652–675, Mar. 2013.
[9] P. R. Marpu, M. Pedergnana, M. D. Mura, J. A. Benediktsson, and L. Bruzzone, “Automatic generation of standard deviation attribute profiles for spectral–spatial classification of remote sensing data,” IEEE Geosci. Remote Sens. Lett., vol. 10, no. 2, pp. 293–297, Mar. 2013.
[10] R. Bao, J. Xia, M. D. Mura, P. Du, J. Chanussot, and J. Ren, “Combining morphological attribute profiles via an ensemble method for hyperspectral image classification,” IEEE Geosci. Remote Sens. Lett., vol. 13, no. 3, pp. 359–363, Mar. 2016.
[11] M. D. Mura, A. Villa, J. A. Benediktsson, J. Chanussot, and L. Bruzzone, “Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis,” IEEE Geosci. Remote Sens. Lett., vol. 8, no. 3, pp. 542–546, May 2011.
[12] N. Li, “An improved classification approach based on spatial and spectral features for hyperspectral data,” Spectrosc. Spectr. Anal., vol. 34, no. 2, pp. 526–531, 2014.
[13] Q. Jackson and D. Landgrebe, “Adaptive Bayesian contextual classification based on Markov random fields,” IEEE Trans. Geosci. Remote Sens., vol. 40, no. 11, pp. 2454–2463, Nov. 2002.
[14] P. Ghamisi, M. S. Coupeiro, F. M. L. Martins, and J. A. Benediktsson, “Multilevel image segmentation based on fractional-order darwinian particle swarm optimization,” IEEE Trans. Geosci. Remote Sens., vol. 52, no. 5, pp. 2382–2394, May 2014.
[15] Y. Tarabalka, J. A. Benediktsson, J. Chanussot, and J. C. Tilton, “Multiple spectral-spatial classification approach for hyperspectral data,” IEEE Trans. Geosci. Remote Sens., vol. 48, no. 11, pp. 4122–4132, Sep. 2010.
[16] X. Hadoux, S. Jay, G. Rabatel, and N. Gorretta, “A spectral-spatial approach for hyperspectral image classification using spatial regularization on supervised score image,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 8, no. 6, pp. 2361–2369, Aug. 2015.

[17] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, “Active learning with statistical models,” Jair, vol. 4, no. 1, pp. 129–145, Jul. 2018.

[18] S. Rajan, J. Ghosh, and M. Crawford, “An active learning approach to hyperspectral data classification,” IEEE Trans. Geosci. Remote Sens., vol. 46, no. 4, pp. 1231–1242, Apr. 2008.

[19] D. Tuia, F. Ratle, F. Pacifici, M. Kanevski, and W. Emery, “Active learning methods for remote sensing image classification,” IEEE Trans. Geosci. Remote Sens., vol. 47, no. 7, pp. 2218–2232, Jul. 2009.

[20] G. Jun and J. Ghosh, “An efficient active learning algorithm with knowledge transfer for hyperspectral data analysis,” in Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS), Boston, MA, USA, Jul. 2009, pp. 52–55.

[21] B. Demir, C. Persello, and L. Bruzzone, “Batch-mode active-learning methods for the interactive classification of remote sensing images,” IEEE Trans. Geosci. Remote Sens., vol. 49, no. 3, pp. 1014–1031, Mar. 2011.

[22] S. Patra and L. Bruzzone, “A fast cluster-assumption based active-learning technique for classification of remote sensing images,” IEEE Trans. Geosci. Remote Sens., vol. 49, no. 5, pp. 1617–1626, May 2011.

[23] S. Patra and L. Bruzzone, “A batch-mode active learning technique based on multiple uncertainty for SVM classifier,” IEEE Geosci. Remote Sens. Lett., vol. 9, no. 3, pp. 497–501, May 2012.

[24] Z. Zhang, E. Pasolli, M. M. Crawford, and J. C. Tilton, “An active learning framework for hyperspectral image classification using hierarchical segmentation,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 9, no. 2, pp. 640–654, Feb. 2016.

[25] J. Li, J. M. Bioucas-Dias, and A. Plaza, “Spectral–spatial classification of hyperspectral data using loopy belief propagation and active learning,” IEEE Trans. Geosci. Remote Sens., vol. 51, no. 2, pp. 844–856, Feb. 2013.

[26] G. Cheng, Z. Li, J. Han, X. Yao, and L. Guo, “Exploring hierarchical convolutional features for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 56, no. 11, pp. 6712–6722, Nov. 2018.

[27] P. Zhou, J. Han, G. Cheng, and B. Zhang, “Learning compact and discriminative stacked autoencoder for hyperspectral image classification,” IEEE Trans. Geosci. Remote Sens., vol. 57, no. 7, pp. 4823–4833, Jul. 2019.

[28] G. Cheng, P. Zhou, and J. Han, “Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images,” IEEE Trans. Geosci. Remote Sens., vol. 54, no. 12, pp. 7405–7415, Dec. 2016.

[29] P. Goovaerts, “Geostatistical incorporation of spatial coordinates into supervised classification of hyperspectral data,” J. Geograph. Syst., vol. 4, no. 1, pp. 99–111, Mar. 2002.

[30] J. C. Platt, “Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods,” Adv. Large Margin Classifiers, vol. 10, no. 3, pp. 61–74, 1999.

[31] D. Tuia, M. Volpi, L. Copa, M. Kanevski, and J. Munoz-Mari, “A survey of active learning algorithms for supervised remote sensing image classification,” IEEE J. Sel. Top. Signal Process., vol. 5, no. 3, pp. 606–617, Jun. 2011.

[32] C.-C. Chang and C.-I. Lin, “LIBSVM: A library for support vector machines,” ACM Trans. Intell. Syst. Technol., vol. 2, no. 3, pp. 2157–6904, 2011.

[33] J. Weston and C. Watkins, “Multiclass support vector machines,” in Support Vector Machines for Pattern Classification. London, U.K.: Springer, 2005, pp. 83–128.

[34] W. Thompson and S. D. Walter, “A reappraisal of the kappa coefficient,” J. Clin. Epidemiol., vol. 41, no. 10, pp. 949–958, Jan. 1988.