

Density Peak-Based Noisy Label Detection for Hyperspectral Image Classification

Bing Tu^{ID}, Member, IEEE, Xiaofei Zhang, Student Member, IEEE, Xudong Kang^{ID}, Senior Member, IEEE, Guoyun Zhang, Member, IEEE, and Shutao Li^{ID}, Senior Member, IEEE

Abstract—Mislabeled training samples may have a negative effect on the performance of hyperspectral image classification. In order to solve this problem, a new density peak (DP) clustering-based noisy label detection method is proposed, which consists of the following steps. First, the distances among the training samples of each class are calculated using four representative distance metrics, i.e., the Euclidean distance (ED), orthogonal projection divergence (OPD), spectral information divergence (SID), and correlation coefficient (CC). Then, the local density of each training sample can be obtained using the DP clustering algorithm. Finally, a local density-based decision function is used to detect the noisy labels. The effectiveness of the proposed method is evaluated using the support vector machines on several real hyperspectral data sets. Experimental results demonstrate that the proposed noisy label detection method indeed helps in improving the classification performance.

Index Terms—Density peak (DP) clustering, hyperspectral image (HSI), noisy label detection, support vector machines (SVMs).

I. INTRODUCTION

HYPERSPECTRAL images (HSIs) contain rich spectral information that is helpful in many remote sensing applications, such as environment protection [1], [2], climatic prediction [3], geologic surveying [4]–[6], and city surveillance [7], [8]. HSI classification is an important technology that has an important role in analyzing HSIs [9], [10]. Some representative supervised classifiers, such as Bayesian estimation [11], decision tree [12], neural networks [13], support vector machines (SVMs) [14], sparse representation (SR) [15], genetic algorithm [16], and kernel-based methods [17], have been successfully applied for HSI classification. Meanwhile, some spectral–spatial classification

Manuscript received March 14, 2018; revised May 26, 2018 and August 1, 2018; accepted August 21, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 51704115, in part by the Key Laboratory Open Fund Project of Hunan Province University under Grant 17K040 and Grant 15K051, and in part by the Science and Technology Program of Hunan Province under Grant 2016TP1021. (Corresponding author: Xudong Kang.)

B. Tu, X. Zhang, and G. Zhang are with the School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang 414000, China (e-mail: tubing@hnist.edu.cn; xiaofei_zh@foxmail.com; gyzhang@hnist.edu.cn).

X. Kang and S. Li are with the College of Electrical and Information Engineering, Hunan University, Changsha 410082, China (e-mail: xudong_kang@163.com; shutao_li@hnu.edu.cn).

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

Digital Object Identifier 10.1109/TGRS.2018.2867444

methods are proposed to further improve the classification performance of HSIs [18]–[25]. One basic assumption of these methods is that the training samples used for learning are correct and highly reliable. However, this assumption is not always true in some real applications.

Performing a supervised classification on HSIs requires that some samples in the hyperspectral data are labeled artificially. However, this manual labeling process is quite hard and may be unreliable because field measurement is required by carrying a specific technical system equipped with GPS. Otherwise, labeling a HSI only according to visual interpretation may be quite unreliable. Even with a careful labeling process, it is found that incorrect labels may still exist in the training set. These incorrect labels are referred as “noisy” labels in this paper, and the problem is referred as the “noisy label” problem. Specifically, the possible reasons of introducing noisy labels are presented as follows.

- 1) Existing GPSs may produce inaccurate estimations for spatial locations. In this situation, it is hard to determine the precise location of a hyperspectral pixel, and thus, field measurements based on GPS may produce labeling errors.
- 2) The process of field measurement is quite hard and time-consuming. When labeling a scene that is very complex and contains many irregular-shaped land covers, visual interpretation-based labeling may produce noisy labels for the training set.
- 3) Some small-scale objects of different classes may locate inside of some large-scale land covers. However, these small-scale objects may be labeled as the same class as those of surrounding regions in the labeling process.
- 4) For some scenes, such as ocean and wetland, a ground investigation is not able to be performed since these scenes may be unreachable for human beings. In this situation, visual interpretation-based human labeling unavoidably produces noisy labels.

Some studies in computer vision and remote sensing fields have been focused on the “noisy label” problem. For example, in the machine learning field, Xiao *et al.* [26] propose a machine learning framework to train convolutional neural networks with a limited number of clean labels and millions of easily obtained noisy labels. Lu *et al.* [27] formulate a novel L_1 -optimization-based learning model to detect the weak and noisy labels. In the remote sensing field, Foody [28] find that the noisy labels may affect the accuracy of the SVM

on the airborne thematic mapper data set. Hou *et al.* [29] propose a semisupervised probability graphic-based classification framework for classification of Polarimetric Synthetic Aperture Radar data. Zhang *et al.* [30] design an index named perturbing influence value that is based on the SVM regression model so as to detect the mislabeled samples of microarray. Tang *et al.* [31] present a robust two-stage reranking method for remote sensing image retrieval that can still operate normally even though the labels of some selected images are incorrect. Although some studies have been performed to address the “noisy label” problem in related fields, these methods cannot be directly applied to HSI classification due to the high dimensionality and nonlinear structure of hyperspectral data sets. In order to address this problem, Kang *et al.* [32] introduce for the first time some typical reasons that may produce the noisy labels in HSI classification and propose an edge-preserving filtering (EPF) and spectral detection-based method to correct the mislabeled training samples. Tu *et al.* [33] detect the HSI noisy label and solve the effectiveness of noisy label by fusing spectral angle and the local outlier factor (SALOF), and the experimental results show that the SALOF method can detect the noisy labels effectively.

In recent years, the density peak (DP) clustering method is introduced in [34], which has been successfully applied in the field of hyperspectral remote sensing. For example, in [35]–[37], DP clustering has been applied for the band selection of HSIs. Chen *et al.* [38] present a density metric-based HSI clustering method. These related studies have confirmed that the DP could serve as an effective method for classifying the pixels of HSI effectively.

In this paper, a DP-based detection method is introduced for the first time to detect the mislabeled samples in the training set. The proposed method consists of the following major steps. First, the distances among the training samples of each class are calculated. Then, the local density of each training sample can be obtained by the DP clustering algorithm. Finally, a decision function is designed to detect the noisy labels. The major contributions of the proposed method are presented as follows.

- 1) DP clustering is introduced for the first time to clean the training set for HSI classification. It is found that mislabeled samples usually have very low local densities, which is the basic motivation behind this paper.
- 2) Four representative distance metrics are analyzed in the proposed detection framework, in which the correlation coefficient (CC) is found to be a robust metric for detecting mislabeled samples.
- 3) The effectiveness of the proposed method is demonstrated using several real hyperspectral data sets and multiple classifiers, including spectral and spectral–spatial methods. Furthermore, the proposed detection method can be implemented very efficiently, and thus, it is easy to use in real applications.

The remainder of this paper is organized as follows. Section II reviews the related work. Section III describes the proposed method in detail. In Section IV, the experimental

results are presented and analyzed. Finally, conclusions are given in Section V.

II. RELATED WORK

A. Supervised Learning With Mislabeled Samples

Designing supervised learning methods that can learn from a training set with mislabeled samples is a problem of great practical importance. Given the importance of learning from such mislabeled samples, a significant amount of practical work has been done, which can be generally divided into two categories.

1) *Robust Classifiers With Mislabeled Samples*: In order to handle the mislabeled samples in the training set, some approaches rely on training robust classifiers that are not very sensitive to the mislabeled samples in the training set [27], [29], [30]. Noise-tolerant learning using statistical queries is one representative method, which learns by using some statistical quantities computed from the examples [39]. However, this approach is limited to binary features. Risk minimization under different loss functions is a practical strategy to improve the robustness of existing classifiers. However, as found in [40], the convex loss functions used for many existing classifiers may lead to an overfitting effect. It is suggested that looking for techniques to minimize risk under 0–1 loss function may be a promising approach for design of robust classifiers.

In recent years, deep learning with mislabeled samples has also been researched. For example, Xiao *et al.* [26] proposed a deep learning-based image understanding framework, in which the original convolutional neural networks are improved to address the training set with only a limited number of clean labels and millions of easily obtained noisy labels.

2) *Cleaning of the Mislabeled Samples*: Besides designing classifiers robust to noisy labels, some data cleaning methods are proposed to remove or correct mislabeled samples from the training set. For example, Wu *et al.* [41] propose a refinement method to remove some tags by the sentiments of adjective–noun pairs and tags. By this way, the performance of the trained deep learning-based visual sentiment analysis model is improved. Pelletier *et al.* [42] present a new iterative learning framework as a solution to filter mislabeled data from the training set that is based on the well-known random forest algorithm. In this paper, a DP-based noisy label detection method is proposed for HSI classification, which also aims at cleaning the noisy labels in the training set.

B. Density Peak-Based Clustering

DP-based clustering is a representative clustering algorithm that is able to process data with arbitrary distributions. Unlike *k*-means [38] that requires the number of cluster centers as input, the parameter of the DP algorithm, i.e., a cutoff distance, can be determined more easily. Specifically, the DP clustering algorithm is based on the following assumptions.

- 1) The cluster centers are usually surrounded by those samples with lower local densities.
- 2) Those samples with lower local densities are usually far from other class centers [34], [38].

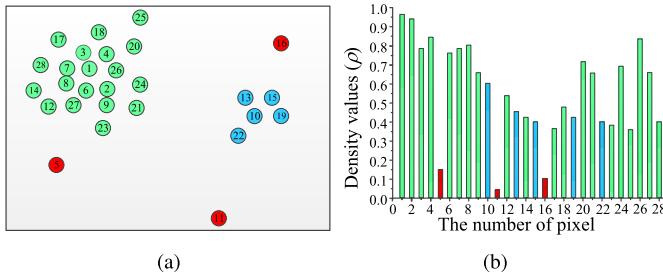


Fig. 1. Graph example illustrating the basic idea of the DP Clustering algorithm. (a) Distribution of the pixels belonging to different classes. (b) Local densities of the pixels calculated using the DP clustering algorithm.

Fig. 1 gives an example illustrating the basic idea of the DP clustering algorithm. As shown in Fig. 1, the cluster centers (1 and 10) are surrounded by the samples from the same class which have lower local densities with respect to the cluster centers. Furthermore, those samples located far from the class centers usually have very low local densities.

Here, the major steps of the DP clustering algorithm are illustrated. Given a data set $\mathbf{x} = \{\mathbf{x}_i\}_{i=1}^n$, where n represents the number of samples, the Euclidean distance (ED) \mathbf{d}_{ij} between the samples \mathbf{x}_i and \mathbf{x}_j can be obtained as follows:

$$\mathbf{d}_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|_2^2. \quad (1)$$

With the above-obtained distance information, the local density ρ_i of each sample i could be calculated as follows:

$$\rho_i = \sum_j \chi(\mathbf{d}_{ij} - d_c) \quad (2)$$

where $\chi(\mathbf{d}_{ij} - d_c) = 1$ if $(\mathbf{d}_{ij} < d_c)$. Otherwise, $\chi(\mathbf{d}_{ij} - d_c) = 0$. d_c is the cutoff distance that indicates the radius of the search region. Furthermore, it is obvious that the local density ρ_i reflects the number of samples that are closer than d_c to the pixel i .

After computing ρ_i , δ_i is defined as follows:

$$\delta_i = \begin{cases} \max_j(\mathbf{d}_{ij}), & \text{if } \rho_i = \max(\rho) \\ \min_{j: \rho_j > \rho_i} (\mathbf{d}_{ij}), & \text{Otherwise.} \end{cases} \quad (3)$$

Specifically, if the i th sample has the highest density, δ_i is defined as the maximum distance from the sample i to any other samples. Otherwise, δ_i equals the minimum distance from the sample i to any other samples with higher densities. The cluster centers are recognized as those samples with higher δ_i and ρ_i . Therefore, a score γ_i is defined as follows:

$$\gamma_i = \rho_i \times \delta_i. \quad (4)$$

The samples with higher γ_i values are more possible to be the cluster centers. Therefore, the cluster centers in the data set can be found using a ranking process.

The proposed DP clustering-based noisy label detection method does not require finding the cluster centers. The objective of the proposed method is to detect those samples that are far from the cluster centers. In (2), those samples are detected using a hard threshold, i.e., the cutoff distance d_c .

In contrast, in this paper, a soft Gaussian kernel function is used to define the local probability of each sample as follows:

$$\rho_i = \sum_j e^{-\left(\frac{\mathbf{d}_{ij}}{d_c}\right)^2}. \quad (5)$$

The advantage of the soft Gaussian function is that it can decrease the negative impact of the statistical errors caused by the fewer samples. Similar local probability functions have been successfully applied for band selection of HSIs [35], [37], [43].

III. PROPOSED APPROACH

In this paper, the DP clustering algorithm is introduced for the first time in a noisy label detection framework for HSI classification (see Fig. 2). Specifically, the proposed method consists of three steps. First, the distances of the training samples in each class are calculated. Second, the local densities of the training samples in each class are estimated using the local density function. Finally, the pixels with very low densities are detected according to the measured local densities. The details of the proposed method are presented in the following.

A. Calculating the Distances of the Training Samples

Let $\mathbf{x} = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M\}$ refers to the original training set, in which M is the number of classes, and \mathbf{x}^m refers to the training samples in the m th class. For two training samples belonging to the m th class, i.e., \mathbf{x}_a^j and \mathbf{x}_b^j , the distance d_{ab}^m between two samples can be measured. In this paper, four types of distances, i.e., the ED [44], orthogonal projection divergence (OPD) [45], spectral information divergence (SID) [46], and CC [20], are considered. The definitions of these distance metrics are presented in the following.

1) Euclidean Distance:

$$\mathbf{d}_{ab}^j = \|\mathbf{x}_a^j - \mathbf{x}_b^j\|_2^2. \quad (6)$$

2) Orthogonal Projection Divergence:

$$\mathbf{d}_{ab}^j = \mathbf{x}_a^j \mathbf{W}_a \mathbf{x}_a^j + \mathbf{x}_b^j \mathbf{W}_b \mathbf{x}_b^j \quad (7)$$

where $\mathbf{W}_a = 1 - \mathbf{x}_a^j (\mathbf{x}_a^j \mathbf{x}_a^j)^{-1} \mathbf{x}_a^j$ and $\mathbf{W}_b = 1 - \mathbf{x}_b^j (\mathbf{x}_b^j \mathbf{x}_b^j)^{-1} \mathbf{x}_b^j$.

3) Spectral Information Divergence:

$$\mathbf{d}_{ab}^j = \sum \mathbf{u} \log \left(\frac{\mathbf{u}}{\mathbf{v}} \right) + \sum \mathbf{v} \log \left(\frac{\mathbf{v}}{\mathbf{u}} \right) \quad (8)$$

where $\mathbf{u} = (\mathbf{x}_a^j / \sum \mathbf{x}_a^j)$ and $\mathbf{v} = (\mathbf{x}_b^j / \sum \mathbf{x}_b^j)$ refer to the desired probability vectors resulting from the pixel vectors \mathbf{x}_a^j and \mathbf{x}_b^j .

4) Correlation Coefficient:

$$\mathbf{d}_{ab}^j = \frac{\text{cov}(\mathbf{x}_a^j, \mathbf{x}_b^j)}{\sqrt{\text{var}(\mathbf{x}_a^j)} \cdot \sqrt{\text{var}(\mathbf{x}_b^j)}} \quad (9)$$

where $\text{var}(\mathbf{x}_a^j)$ and $\text{var}(\mathbf{x}_b^j)$ refer to variances of the pixel vectors \mathbf{x}_a^j and \mathbf{x}_b^j .

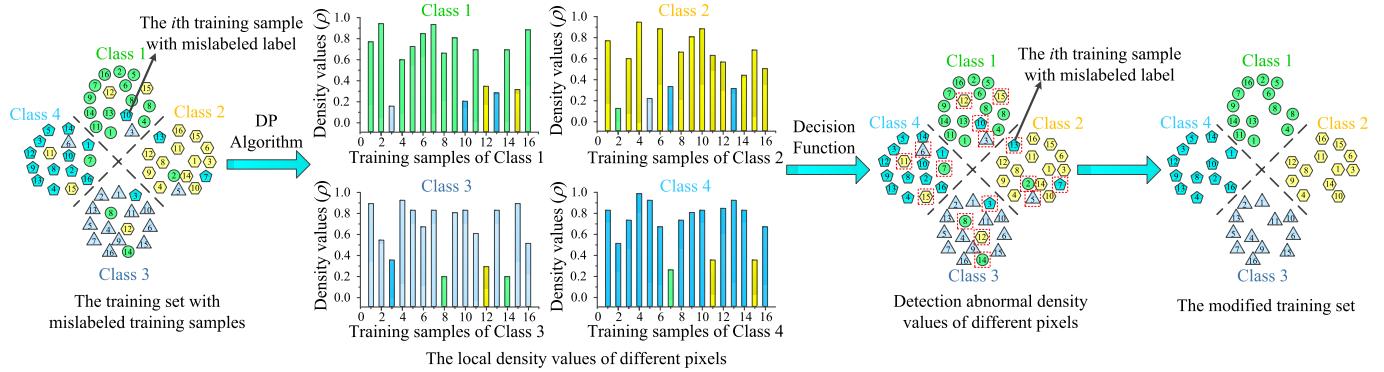


Fig. 2. Graph example illustrating the principle of the proposed method. Different colors refer to different labels of the training samples.

Through calculating the distances among the n th pixel and the pixels in the m th class, a distance array of the n th pixel can be constructed as follows:

$$\mathbf{d}_n^m = [\mathbf{d}_{n1}^m, \mathbf{d}_{n2}^m, \dots, \mathbf{d}_{nN_m}^m]^T \quad (10)$$

where N_m refers to the number of pixels in the m th class. By this way, a distance matrix \mathbf{D}^m can be constructed as $\mathbf{D}^m = \{\mathbf{d}_1^m, \mathbf{d}_2^m, \dots, \mathbf{d}_{N_m}^m\}$.

B. Calculating the Local Densities of the Training Samples

First, the cutoff distance d_c^m is calculated as follows:

$$d_c^m = \mathbf{S}^m(t) \quad \text{s.t. } t = \left\langle \frac{N_m \cdot (N_m - 1)}{100} \cdot p \right\rangle \quad (11)$$

where \mathbf{S}^m is a matrix that sorts the nonzero elements in the upper triangular matrix of \mathbf{D}^m from the smallest to the largest, p is a free parameter that will be analyzed in Section IV-B, and $\langle \cdot \rangle$ refers to the round operation.

With the above-obtained cutoff distance, the local densities $\rho^m = \{\rho_1^m, \rho_2^m, \dots, \rho_{N_m}^m\}$ of the pixels in the m th class can be calculated as follows:

$$\rho^m = \sum e^{-\left(\frac{D^m}{d_c^m}\right)^2}. \quad (12)$$

C. Detecting the Mislabeled Samples

Once the local densities of the training samples in different classes are obtained, the mislabeled samples can be easily detected and removed as follows:

$$\mathbf{Y}_i^m = \begin{cases} \mathbf{X}_i^m & \text{if } \rho_{N_m}^m \geq \lambda \cdot \bar{\rho}^m \\ \emptyset & \text{Otherwise} \end{cases} \quad (13)$$

where $\mathbf{Y} = \{\mathbf{Y}^1, \mathbf{Y}^2, \dots, \mathbf{Y}^M\}$ refers to the resulting training set, in which the noisy labels have been detected and removed. λ is a free parameter.

IV. EXPERIMENTS

In this section, experiments are performed on the Kennedy Space Center (KSC), Salinas, Indian Pines, and the University of Pavia data sets. The information of the data sets is presented in the following.

A. Data Sets

- 1) *KSC*: The image was acquired by the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) over the KSC, Merritt Island, FL, USA. The image is of size 512×614 pixels, in which 48 bands are removed as water absorption and low SNR bands. The false color composite of the KSC image, the reference classification map, and the corresponding color labels are shown in Fig. 5(a)–(c).
- 2) *Salinas*: The Salinas image data set was acquired by the AVIRIS over Salinas Valley in California, USA. The image contains 224 bands and is of size 512×217 pixels with a spatial resolution of 3.7 m. Twenty water absorption bands are removed. Fig. 6(a)–(c) shows the false color composite of the Salinas image, the reference classification map, and the color labels.
- 3) *Indian Pines*: The Indian Pines image shows the Indian Pines Test Field in the northwest of Indiana, USA, which was captured by the AVIRIS. The image is of size 145×145 pixels with a spatial resolution of 20 m/pixel and 220 bands. With 20 water absorption wave bands removed, the remaining 200 bands are used in the experiment. As the scene is captured in June, some crops, such as corn and soybean, are still in the early stage of growth. In the reference classification map obtained from site exploration, the scene is divided into 16 different classes.
- 4) *University of Pavia*: The University of Pavia data set was obtained over the University of Pavia, photographed by the Reflective Optics System Imaging Spectrometer (ROSIS-3). The image is of size 610×340 pixels with a spatial resolution of 1.3 m/pixel and 115 bands. After removing 12 bands containing serious image noise, the experiments are performed using the remaining 103 bands.

This paper adopts the SVM classifier to demonstrate the effectiveness of the proposed noisy label detection method as it is one of the most widely used pixelwise classifiers. The SVM method is implemented using the LIBSVM library [47], and the parameters of the SVM are decided using a fivefold cross validation. Furthermore, in all experiments, three widely used quality indexes, i.e., the overall accuracy (OA), the average

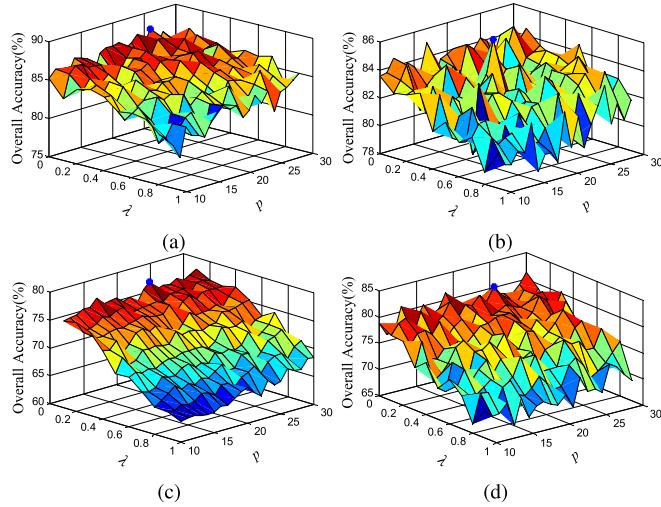


Fig. 3. Influence of the parameters p and λ to the performance of the proposed method. (a) KSC data set with 20 true and three mislabeled samples. (b) Salinas data set with 20 true and three mislabeled samples. (c) Indian Pines data set with 80 true and 10 mislabeled samples. (d) University of Pavia data set with 50 true and five mislabeled samples.

accuracy (AA), and the Kappa coefficient, are used to evaluate the performance of the proposed method. OA measures the percentage of all correctly classified pixels. AA represents the average value of the percentage of the correctly classified pixels for each class. The Kappa coefficient estimates the percentage of classified pixels corrected by the number of agreements that would be expected purely by chance. All experiments are repeated 10 times with randomly selected training samples so as to obtain the mean and standard variances of OA, AA and Kappa. The training sets are constructed using the samples in the ground truth. For each class, some pixels randomly selected from other classes will be added so as to simulate the “noisy label” problem.

B. Analysis the Influence of the Parameters

In this section, the influence of the parameters p and λ to the performance of the proposed method is analyzed. Fig. 3 shows the experimental results obtained at the KSC, Salinas, Indian Pines, and University of Pavia data sets, respectively. For the KSC data set, 20 true samples and three mislabeled samples are selected randomly for each class. For the Salinas data set, 20 true samples and three mislabeled samples are selected randomly as training samples for each class. For the Indian Pines data set, 80 true samples and 10 mislabeled samples are selected randomly for each class. For the University of Pavia data set, 50 true and five mislabeled samples are selected as the training set. The values of p and λ are selected from the intervals 10~30 and 0~1, respectively. Based on the experimental results presented in Fig. 3, it can be found that p is actually not related to the number of pixels in the training set N . Taking Fig. 3(a) and (c), for example, the number of training samples varies significantly from 23 per class to 55 per class, while a fixed $p = 20$ is always able to obtain optimal classification performance for the proposed method. Furthermore, it can be seen that, when the number of training

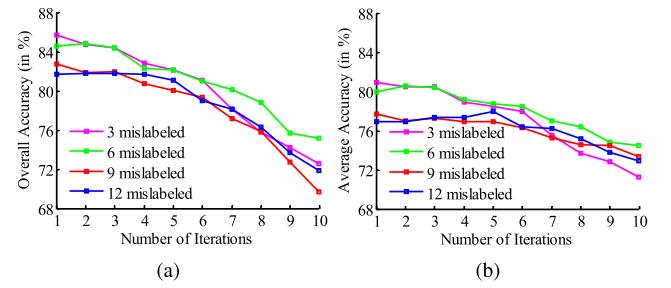


Fig. 4. Performance of the proposed method with different numbers of iterations on the KSC data set. (a) OA. (b) AA.

TABLE I
PERFORMANCE OF THE PROPOSED METHOD USING DIFFERENT DISTANCE METRICS FOR DETECTION OF MISLABELED SAMPLES ON THE KSC DATA SET

Metrics	ED [44]	OPD [45]	SID [46]	CC [20]
OA	86.26(1.84)	86.53(0.71)	86.74(0.86)	87.16(0.60)
AA	81.32(1.74)	81.03(1.78)	81.08(1.97)	81.43(1.17)
Kappa	84.70(2.05)	84.77(1.70)	84.69(1.45)	84.98(1.45)

samples is varying, the optimal p and λ actually vary in a very small range. It is found that $p = 20$ and $\lambda = 0.1$ are always able to obtain relatively optimal classification accuracies. Therefore, given a new data set, $p = 20$ and $\lambda = 0.1$ are suggested to be used as the default parameters in the proposed method.

C. Analysis of the Influence of Different Components

In this section, the first experiment is performed on the KSC data set with a training set consists of 20 true labeled training samples and three mislabeled training samples for each class. In this experiment, the performance of the proposed method with different distance metrics, i.e., the ED [44], OPD [45], SID [46], and CC [20], is analyzed in Table I. It can be seen that the CC metric obtains the best performance in terms of classification accuracies. Therefore, the CC metric is adopted for the proposed method and used in the following experiments.

The second experiment is performed to analyze the effectiveness of different density functions, i.e., cutoff function and Gaussian kernel function (see Table II). The experiment is conducted on the KSC data set with a training set that consists of 20 true labeled training samples and different numbers of mislabeled samples varying from 3 to 12 for each class. As shown in Table II, the Gaussian function leads to higher classification accuracy with respect to the cutoff function used in the original DP clustering algorithm. Therefore, in this paper, the Gaussian-based density function is adopted.

D. Analysis of the Influence of Iterations

The numbers of mislabeled samples that are detected correctly and falsely on the KSC, Salinas, Indian Pines, and University of Pavia data sets are shown in Table III. It can be seen that a small number of samples are detected falsely. Furthermore, it can be seen that a number of mislabeled

TABLE II
PERFORMANCE OF THE PROPOSED METHOD USING DIFFERENT DENSITY FUNCTIONS, I.E., CUTOFF AND GAUSSIAN

Density Function	20 (true) + 3 (mislabeled) Cut-off	20 (true) + 6 (mislabeled) Gaussian	20 (true) + 9 (mislabeled) Cut-off	20 (true) + 9 (mislabeled) Gaussian	20 (true) + 12 (mislabeled) Cut-off	20 (true) + 12 (mislabeled) Gaussian
OA	85.84(2.02)	86.14(1.80)	84.47(2.11)	84.83(1.91)	83.26(2.25)	83.62(1.49)
AA	81.00(1.55)	81.59(1.78)	79.95(2.19)	80.14(1.62)	78.23(2.29)	78.76(1.25)
Kappa	84.28(1.35)	85.28(1.61)	82.47(2.27)	83.65(1.77)	81.62(2.51)	82.45(1.12)

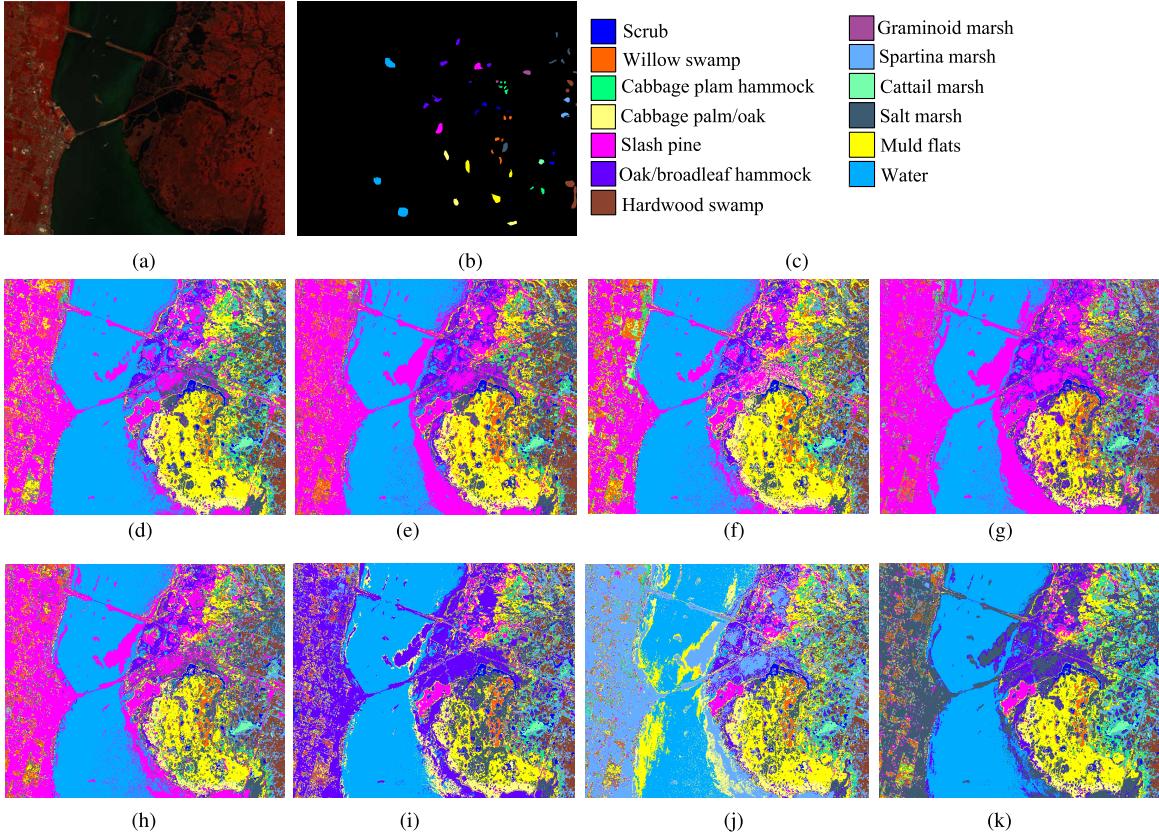


Fig. 5. KSC data set and the classification maps of the compared methods. (a) False-color composite image. (b) Reference. (c) Class names. (d)–(k) Classification maps obtained by the SVM trained using the improved training set (the second row) and the original training set (the third row) with OA = 87.12%, 85.48%, 84.53%, 80.45%, 84.16%, 83.14%, 82.27%, and 78.33%, respectively. The number of true samples is 20 and the number of mislabeled samples is 3 for (d) and (h), 6 for (e) and (i), 9 for (f) and (j), and 12 for (g) and (k).

samples are still existed in the training set when the number of mislabeled samples is very large. An intuitive solution to this problem is to perform the proposed method iteratively. However, as shown in Fig. 4, the multiple iterations of the proposed detection algorithm actually decrease the performance of the proposed method since the simple thresholding-based decision function used in the proposed method decreases the number of training samples in each iteration. Designing an optimal decision function so as to build an effective iterative detection framework could be the focus of our future research. Furthermore, the number of mislabeled samples in the training set is also an important prior knowledge for the detection algorithm that could also be researched in the future.

E. Performance Evaluation Using the SVM

In this section, the performance of proposed method is evaluated on the KSC, Salinas, Indian Pines, and University

of Pavia data sets. The proposed method adopts the parameter setting presented earlier. The experiments on the KSC and Salinas data sets are performed with 20 true samples and different numbers of mislabeled samples in the range of 2–20 for each class. For the Indian Pines image, the experiments are performed with 80 true samples and different numbers of mislabeled samples in the range of 5–50. For the University of Pavia image, the experiments are performed with 50 true samples and different numbers of mislabeled samples in the range of 4–20.

Fig. 5(d)–(k) shows the classification maps of the KSC data set, which are obtained with the SVM trained using the improved and original training sets, respectively. It can be seen that the classification maps obtained by using the original noisy training set contain misclassifications for several classes, such as the water class [see Fig. 5(d) and (h)]. In contrast, the classification maps obtained by the proposed method

TABLE III

NUMBERS OF MISLABELED SAMPLES ADDED IN THE KSC, SALINAS, INDIAN PINES, AND UNIVERSITY OF PAVIA DATA SETS, AND THE NUMBERS OF THE CORRECT AND INCORRECT DETECTIONS (THE AVERAGE VALUES OBTAINED IN 10 REPEATED EXPERIMENTS ARE REPORTED)

dataset	KSC				Salinas				Indian Pines				University of Pavia			
Total	39	78	117	156	48	96	144	192	100	200	300	400	45	90	135	180
Correct	23.0	39.8	52.2	57.7	40.7	67.0	85.9	87.4	54.0	72.3	76.2	79.0	23.5	32.8	46.1	57.7
Incorrect	6.7	7.1	5.3	4.4	11.9	8.9	6.6	3.7	10.8	8.8	8.3	7.5	22.1	21.7	21.1	20.7

TABLE IV

CLASSIFICATION ACCURACIES OF THE KSC DATA SET, WHICH ARE OBTAINED BY THE SVM AND THE PROPOSED METHOD TRAINED WITH 20 TRUE SAMPLES AND DIFFERENT NUMBERS OF MISLABELED SAMPLES. FOR THE PROPOSED METHOD, TWO TIME INFORMATION (DETECTING TIME/CLASSIFICATION TIME) IS RECORDED

Class	SVM 20 (true)	The number of true samples and mislabeled samples							
		20 (true) + 3 (mislabeled)		20 (true) + 6 (mislabeled)		20 (true) + 9 (mislabeled)		20 (true) + 12 (mislabeled)	
		SVM Proposed	SVM Proposed	SVM Proposed	SVM Proposed	SVM Proposed	SVM Proposed	SVM Proposed	SVM Proposed
1	95.57(1.82)	95.29(1.74)	96.12(2.19)	93.25(3.25)	94.77(2.01)	94.63(2.91)	95.03(2.03)	93.15(3.99)	93.71(4.16)
2	85.50(3.96)	78.37(5.89)	85.28(5.44)	73.25(6.42)	83.76(5.92)	75.11(6.73)	82.22(6.94)	71.89(5.37)	80.76(7.60)
3	85.11(6.99)	85.07(4.94)	85.50(5.76)	85.45(5.39)	86.44(3.97)	78.85(6.26)	81.66(5.64)	82.09(7.40)	86.66(5.51)
4	64.88(5.96)	60.31(7.14)	64.75(3.78)	58.02(6.92)	60.89(6.10)	58.77(4.66)	63.92(5.20)	54.77(6.65)	59.83(5.26)
5	60.82(5.93)	63.32(9.30)	64.01(10.70)	58.76(4.89)	59.96(9.22)	55.17(13.4)	61.79(9.54)	55.11(8.09)	61.16(9.84)
6	61.65(9.30)	57.07(8.00)	57.67(5.81)	53.09(8.63)	54.49(7.30)	52.24(7.65)	55.11(11.04)	49.49(6.31)	53.85(9.50)
7	69.18(3.78)	65.31(7.00)	71.71(4.53)	62.00(7.16)	65.18(6.99)	59.26(9.05)	66.73(4.58)	62.33(6.16)	63.32(7.86)
8	80.97(7.06)	78.23(9.33)	77.94(6.84)	73.83(6.75)	73.93(8.51)	69.57(11.3)	71.38(9.03)	67.70(8.86)	66.59(7.27)
9	90.09(2.89)	85.95(6.31)	87.87(4.26)	85.38(4.09)	86.70(4.71)	83.63(4.40)	82.01(6.42)	82.02(5.51)	84.42(5.65)
10	90.59(6.50)	86.44(9.38)	90.79(9.02)	86.40(7.69)	90.21(9.19)	86.76(8.15)	87.53(10.30)	86.68(9.36)	88.39(8.99)
11	92.27(5.95)	89.28(8.36)	91.07(7.58)	90.72(7.39)	89.90(9.52)	94.42(5.36)	88.06(9.63)	89.93(4.66)	86.63(7.91)
12	94.10(3.68)	92.15(5.08)	89.49(6.09)	90.18(6.34)	85.91(7.28)	88.25(6.73)	89.70(7.13)	85.53(8.82)	82.99(6.67)
13	99.98(0.04)	98.84(1.56)	99.92(0.16)	98.69(1.42)	99.02(0.87)	99.40(0.78)	97.38(2.37)	98.10(1.93)	98.31(1.73)
OA	87.70(1.44)	84.95(1.99)	86.99(1.65)	83.33(1.10)	85.33(2.21)	82.15(2.60)	84.16(1.93)	77.06(1.45)	80.51(1.51)
AA	82.36(1.61)	79.66(2.23)	81.47(1.86)	77.62(1.29)	79.01(2.26)	76.62(2.23)	78.35(2.27)	75.29(1.54)	77.41(1.48)
Kappa	86.29(1.60)	83.24(2.21)	84.85(1.83)	81.44(1.23)	82.57(2.44)	80.16(2.81)	81.18(2.13)	78.92(1.60)	80.30(1.68)
Time (s)	27.13	36.98	0.30/28.51	44.41	0.21/34.38	52.72	0.27/41.79	60.01	0.35/49.67

TABLE V

CLASSIFICATION ACCURACIES OF THE SALINAS DATA SET, WHICH ARE OBTAINED BY THE SVM AND THE PROPOSED METHOD TRAINED WITH 20 TRUE SAMPLES AND DIFFERENT NUMBERS OF MISLABELED SAMPLES. FOR THE PROPOSED METHOD, TWO TIME INFORMATION (DETECTING TIME/CLASSIFICATION TIME) IS RECORDED

Class	SVM 20 (true)	The number of true samples and mislabeled samples							
		20 (true) + 3 (mislabeled)		20 (true) + 6 (mislabeled)		20 (true) + 9 (mislabeled)		20 (true) + 12 (mislabeled)	
		SVM Proposed	SVM Proposed	SVM Proposed	SVM Proposed	SVM Proposed	SVM Proposed	SVM Proposed	SVM Proposed
1	99.32(0.92)	98.90(1.21)	99.52(0.70)	97.07(4.20)	99.14(0.79)	98.85(1.36)	98.95(1.68)	97.39(2.88)	98.83(1.73)
2	99.17(0.46)	98.32(2.22)	98.57(1.55)	98.73(0.92)	99.27(0.30)	98.81(0.69)	99.16(0.30)	97.83(3.35)	99.30(0.26)
3	90.60(2.46)	87.49(4.45)	87.95(2.48)	89.06(3.43)	90.61(1.94)	87.75(4.76)	87.21(3.85)	86.28(12.2)	90.20(3.16)
4	96.81(0.66)	93.80(7.01)	91.19(9.30)	95.05(5.23)	96.94(0.70)	96.20(1.86)	96.40(1.00)	93.90(8.74)	96.66(0.57)
5	98.21(0.97)	98.23(1.31)	98.74(1.02)	97.22(2.59)	98.47(1.17)	98.18(1.35)	98.65(1.07)	97.03(2.68)	98.46(1.89)
6	99.70(1.09)	99.65(0.71)	100.00(0.0)	99.31(1.60)	100.00(0.0)	99.14(1.93)	99.95(0.07)	96.32(6.47)	99.93(0.14)
7	98.16(1.33)	96.97(1.35)	97.92(1.42)	97.41(3.56)	94.10(9.78)	97.17(1.57)	98.27(1.24)	96.73(2.44)	97.12(1.39)
8	71.25(2.66)	70.56(4.06)	71.34(3.15)	69.71(4.13)	70.93(4.88)	71.78(2.70)	71.89(3.28)	70.57(3.38)	69.91(2.65)
9	99.12(0.39)	98.55(1.19)	99.01(0.79)	98.64(0.79)	99.05(0.47)	98.54(0.84)	98.95(0.74)	98.69(0.74)	99.21(0.51)
10	83.06(3.67)	79.21(4.71)	76.14(8.37)	81.85(5.10)	73.22(12.0)	80.44(6.50)	76.04(5.80)	81.05(5.86)	75.21(7.51)
11	88.33(3.80)	86.04(6.15)	86.32(5.43)	87.68(4.96)	88.74(4.96)	87.30(3.75)	88.66(5.41)	83.30(8.48)	88.23(5.76)
12	95.55(1.27)	94.17(1.97)	94.34(1.35)	93.88(2.25)	94.34(1.73)	92.50(3.71)	92.57(7.71)	92.51(2.91)	94.64(1.32)
13	93.94(2.43)	89.45(10.6)	91.36(5.84)	90.80(6.10)	93.22(2.48)	91.57(3.70)	91.22(4.74)	88.53(7.17)	89.96(10.7)
14	89.03(7.08)	85.90(9.90)	85.93(11.7)	89.48(7.34)	90.22(10.7)	88.03(12.2)	92.85(4.87)	84.27(12.5)	88.99(13.8)
15	52.66(4.27)	52.86(6.38)	53.26(2.24)	51.09(4.31)	51.94(5.50)	51.95(4.15)	47.80(3.98)	50.36(4.39)	53.96(4.37)
16	93.92(4.54)	95.27(2.65)	88.04(7.22)	93.48(7.63)	85.37(13.4)	95.31(4.35)	89.74(9.97)	89.63(9.82)	91.65(10.1)
OA	84.20(1.46)	82.39(1.77)	83.83(2.24)	81.92(1.13)	83.47(1.18)	80.54(1.11)	83.19(1.64)	78.55(1.33)	82.71(2.58)
AA	79.30(1.21)	89.08(1.58)	89.43(1.10)	89.40(1.06)	89.86(1.45)	89.59(0.84)	88.93(1.00)	88.31(1.53)	88.53(1.69)
Kappa	79.30(1.18)	81.58(1.93)	82.51(2.75)	81.16(1.26)	81.76(1.43)	81.63(1.20)	81.96(1.83)	80.23(1.44)	82.37(2.72)
Time (s)	20.35	31.55	0.36/24.56	37.83	0.41/27.11	45.24	0.47/37.51	53.48	0.54/42.14

reduce such misclassifications obviously. Table IV shows the classification accuracies of the compared methods. It can be seen that the proposed method can improve the classification accuracies for most of the classes. Furthermore, the computing time spent in classification can be also decreased by removing some mislabeled training samples (see the final row of Table IV). Next, the performance of the SVM trained using 20 correct training samples is also given in Table IV. In other

words, if the proposed method can detect and remove all mislabeled training samples, the performance of the proposed method should be the same as the SVM trained using the 20 correct training samples.

Table V and Fig. 6(d)–(k) show the experimental results of the Salinas data set. As shown in Table V, when the ratio of the mislabeled samples increases, the improvement in classification accuracies becomes more obvious. For example,

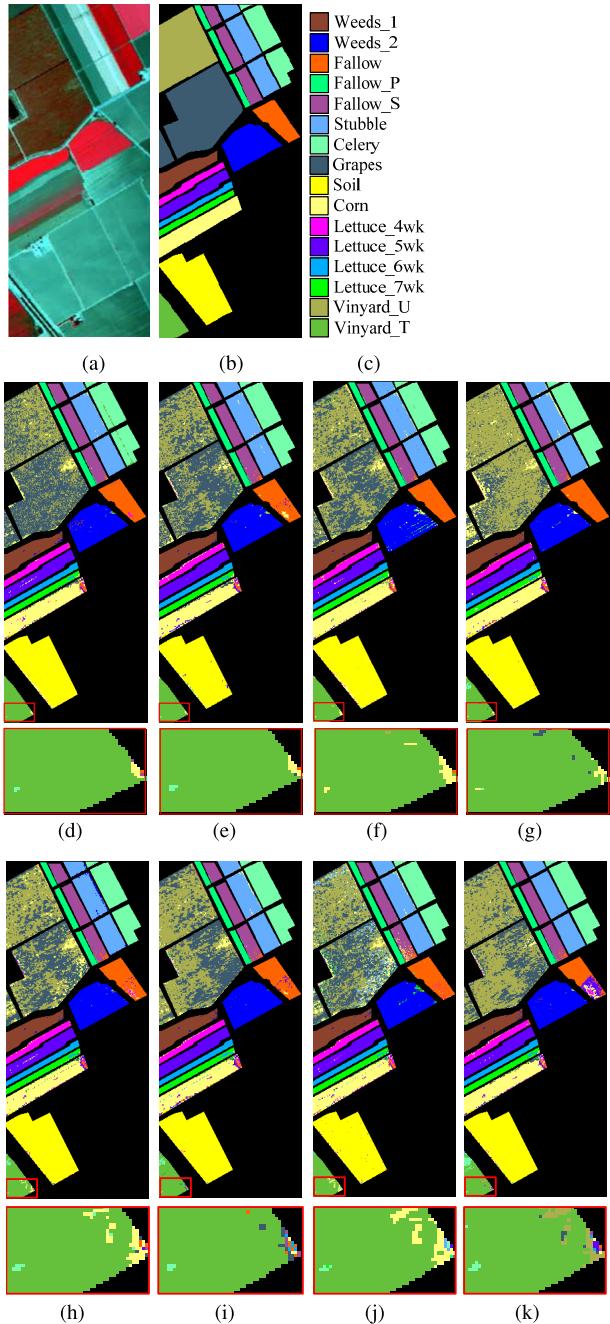


Fig. 6. Salinas data set and the classification maps of the compared methods. (a) False-color composite image. (b) Reference. (c) Class names. (d)–(k) Classification maps obtained by the SVM trained using the improved training set (the second row) and the original training set (the third row) with OA = 84.45%, 83.64%, 83.02%, 81.89%, 83.85%, 82.68%, 82.01%, and 76.61%. The number of true samples is 20 and the number of mislabeled samples is 3 for (d) and (h), 6 for (e) and (i), 9 for (f) and (j), and 12 for (g) and (k).

when the training set contains 20 true samples and three mislabeled training samples, the classification accuracy can be improved about 1% by using the proposed noisy label detection method. When the number of mislabeled training samples becomes 6, 9, and 12, the amounts of improvement in classification accuracies become 1.38%, 2.91%, and 3.24%, respectively. As shown in the close-up comparisons presented

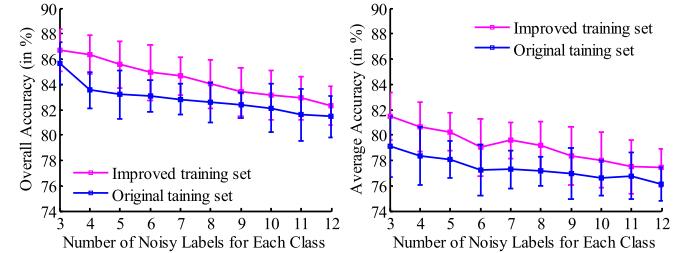


Fig. 7. OAs and AAs obtained by the SVM trained using the improved training set and the original training set. Experiment is performed on the KSC data set with 20 true samples and different numbers of mislabeled samples (vary from 3 to 12) for each class.

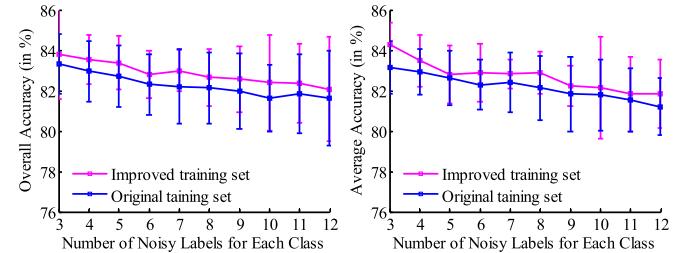


Fig. 8. OAs and AAs obtained by the SVM trained using the improved training set and the original training set. Experiment is performed on the Salinas data set with 20 true samples and different numbers of mislabeled samples (vary from 3 to 12) for each class.

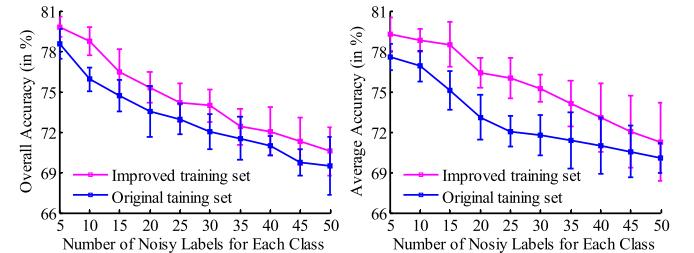


Fig. 9. OAs and AAs obtained by the SVM trained using the improved training set and the original training set. Experiment is performed on the Indian Pines data set with 80 true samples and different numbers of mislabeled samples (vary from 5 to 50) for each class.

in Fig. 6, the classification maps obtained by the proposed method are more similar as the reference classification map.

The classification results for the Indian Pines and University of Pavia data sets are reported in Tables VI and VII, respectively. Similar to the experimental results obtained on the KSC and Salinas data sets, the proposed method shows good performance for increasing the classification accuracies. Taking Table VII, for example, the proposed method can improve the accuracies by 1.5%–2%.

Figs. 7–10 also give the average values and the standard variances of the OAs and AAs obtained in 10 repeated experiments. Clearly, the SVM trained with the improved training set usually shows better classification performance with respect to the SVM trained using the original training set. This experiment verifies that the performance of SVM can be improved by detecting and removing noisy labels in the training set. However, if there are too many mislabeled training samples, the performance of the proposed method

TABLE VI

CLASSIFICATION ACCURACIES OF THE INDIAN PINES DATA SET, WHICH ARE OBTAINED BY THE SVM AND THE PROPOSED METHOD TRAINED WITH 20 TRUE SAMPLES AND DIFFERENT NUMBERS OF MISLABLED SAMPLES. FOR THE PROPOSED METHOD, TWO TIME INFORMATION (DETECTING TIME/CLASSIFICATION TIME) IS RECORDED

Class	SVM 80 (true)	The number of true samples and mislabeled samples							
		80 (true) + 10 (mislabeled)		80 (true) + 20 (mislabeled)		80 (true) + 30 (mislabeled)		80 (true) + 40 (mislabeled)	
		SVM	Proposed	SVM	Proposed	SVM	Proposed	SVM	Proposed
1	74.06(2.75)	71.54(2.97)	72.06(3.19)	69.91(3.16)	70.30(2.80)	68.89(3.71)	70.20(3.44)	67.08(4.00)	68.51(4.39)
2	66.74(3.13)	62.57(5.13)	62.68(4.14)	59.00(6.53)	59.71(4.74)	58.15(8.07)	58.97(4.16)	55.25(4.73)	56.93(6.80)
3	85.46(3.26)	84.21(2.47)	85.59(4.82)	78.80(4.43)	79.32(5.39)	83.05(5.74)	78.43(4.50)	78.86(8.76)	79.93(8.22)
4	94.31(1.10)	91.06(2.39)	93.68(1.64)	88.69(3.06)	88.59(3.95)	88.47(4.92)	89.10(2.62)	87.48(5.24)	88.70(3.04)
5	98.42(0.82)	94.16(3.08)	98.58(1.09)	88.67(4.47)	97.71(1.88)	93.10(8.94)	94.01(4.70)	91.66(8.21)	94.78(4.00)
6	69.70(2.46)	64.86(2.02)	64.64(2.84)	63.69(3.33)	64.31(3.35)	60.58(3.70)	61.17(5.13)	60.57(2.94)	61.80(3.98)
7	86.90(1.38)	84.39(2.00)	85.93(1.29)	82.97(1.52)	83.50(1.53)	81.72(3.10)	83.90(2.28)	82.54(3.12)	80.41(3.30)
8	67.08(3.81)	63.92(4.12)	54.40(3.79)	60.09(4.98)	56.01(3.96)	48.90(10.6)	52.69(7.75)	44.91(8.58)	46.09(9.55)
9	96.23(0.71)	94.24(1.24)	96.27(0.94)	93.48(1.20)	94.27(1.94)	93.73(2.90)	92.85(2.82)	93.17(2.67)	93.84(2.40)
10	59.84(2.85)	58.51(3.92)	55.46(2.36)	55.57(5.44)	49.48(4.20)	53.33(6.14)	48.54(4.91)	47.59(9.29)	52.52(6.59)
OA	80.11(0.80)	76.97(0.76)	78.88(1.04)	73.63(1.51)	75.82(1.14)	72.45(1.44)	73.96(1.19)	70.94(1.03)	73.05(1.80)
AA	79.87(0.87)	76.95(1.03)	77.83(0.83)	74.09(1.18)	75.02(1.13)	72.99(2.06)	73.59(1.01)	72.35(1.53)	72.91(2.54)
Kappa	77.07(0.91)	73.68(1.00)	74.40(1.13)	70.95(1.24)	71.53(1.24)	68.46(1.57)	69.01(1.31)	67.83(1.79)	68.87(1.93)
Time (s)	76.71	100.21	1.88/91.81	126.01	2.41/110.41	154.01	2.74/136.43	184.07	3.73/158.71

TABLE VII

CLASSIFICATION ACCURACIES OF THE UNIVERSITY OF PAVIA DATA SET, WHICH ARE OBTAINED BY THE SVM AND THE PROPOSED METHOD TRAINED WITH 20 TRUE SAMPLES AND DIFFERENT NUMBERS OF MISLABLED SAMPLES. FOR THE PROPOSED METHOD, TWO TIME INFORMATION (DETECTING TIME/CLASSIFICATION TIME) IS RECORDED

Class	SVM 50 (true)	The number of true samples and mislabeled samples							
		50 (true) + 5 (mislabeled)		50 (true) + 10 (mislabeled)		50 (true) + 15 (mislabeled)		50 (true) + 20 (mislabeled)	
		SVM	Proposed	SVM	Proposed	SVM	Proposed	SVM	Proposed
1	94.69(1.74)	90.38(6.08)	91.15(7.14)	92.09(4.43)	95.72(7.93)	86.72(7.58)	87.32(7.22)	89.46(5.52)	89.63(7.56)
2	94.79(0.73)	92.47(1.77)	94.22(1.45)	92.87(1.53)	93.80(1.31)	92.99(1.52)	93.04(1.79)	91.34(1.88)	92.64(1.89)
3	67.58(2.33)	61.95(4.09)	59.94(7.19)	59.19(3.33)	58.11(6.05)	57.45(4.69)	58.27(7.73)	56.84(5.59)	56.90(6.42)
4	79.74(9.02)	74.30(6.45)	79.38(7.63)	70.75(9.64)	75.79(10.1)	70.26(10.9)	72.40(9.18)	62.83(7.78)	70.52(12.7)
5	96.38(1.60)	90.22(6.01)	93.58(9.02)	85.95(15.0)	94.23(7.10)	84.99(11.8)	88.88(8.61)	88.27(9.75)	88.67(12.8)
6	66.47(7.68)	58.13(7.35)	60.42(6.70)	54.41(6.97)	56.00(8.13)	51.86(9.32)	52.05(7.57)	47.53(7.90)	49.53(7.49)
7	57.29(6.65)	50.09(4.54)	48.44(6.65)	46.70(5.69)	43.99(6.13)	45.20(4.90)	41.08(4.24)	43.93(3.77)	40.77(4.31)
8	79.83(4.03)	80.36(5.02)	86.10(7.22)	74.78(4.81)	74.95(7.67)	75.44(5.70)	76.12(5.65)	75.18(5.11)	73.33(6.03)
9	99.91(0.06)	84.76(14.8)	88.24(16.3)	88.39(12.0)	74.66(19.8)	82.66(13.4)	81.25(18.5)	88.27(12.7)	77.30(19.5)
OA	84.26(2.32)	79.00(2.60)	82.95(3.76)	76.42(1.96)	78.91(3.37)	74.00(2.57)	76.97(2.84)	72.75(3.26)	75.27(2.82)
AA	81.85(1.86)	75.85(2.27)	76.16(3.06)	73.90(2.59)	73.93(3.41)	71.95(1.95)	72.16(2.57)	71.74(2.10)	73.70(3.15)
Kappa	79.65(2.79)	73.21(3.10)	75.12(3.25)	70.27(2.24)	71.87(3.82)	67.53(2.92)	68.44(3.14)	66.01(3.59)	67.72(3.10)
Time (s)	19.84	27.80	0.70/20.10	32.59	0.75/25.46	40.01	0.89/31.88	45.63	0.99/37.01

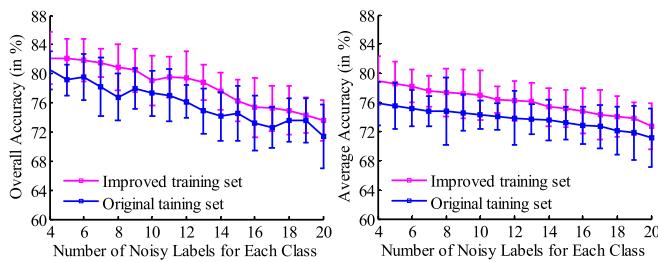


Fig. 10. OAs and AAs obtained by the SVM trained using the improved training set and the original training set. Experiment is performed on the University of Pavia data set with 50 true samples and different numbers of mislabeled samples (vary from 4 to 20) for each class.

may be unsatisfactory. Taking Fig. 7, for example, when there are 12 mislabeled training samples for the KSC data set, the improvement of OA is quite limited.

Finally, the computing time (in seconds) of different methods on the four real HSIs is presented in the bottom row of Tables IV–VII. The computing time was recorded by a

computer with a four-core processor of 2.70 GHz and an 8-GB RAM. It can be seen that the computing time of the proposed method is less than the SVM method trained with the original training set mislabeled training set. Furthermore, the time spent in the detection process is usually less than 1 s. The reason is that the major computing burden of the proposed detection algorithm is to calculate the CC between each samples of each class, for which the complexity of obtaining CC is $O(1)$, and thus, the complexity of detection is $O(\sum_{m=1}^M N_m^2)$, where M is the number of the classes and N_m is the number of training samples in the m th class.

F. Performance Evaluation Using Other Classifiers

In this section, other spectral classifiers, such as the SR classifier (SRC) [48], the basic thresholding classifier (BTC) [49], the kernel BTC (KBTC) [50], and the sparse autoencoder method [51] methods, are used to further demonstrate the effectiveness of the proposed noisy label detection method. The experimental results obtained on the University of Pavia data set with 50 true training samples and 10 mislabeled

TABLE VIII

CLASSIFICATION ACCURACIES OBTAINED BY DIFFERENT SPECTRAL CLASSIFICATION METHODS ON THE UNIVERSITY OF PAVIA DATA SET. * MARKS THE METHODS TRAINED USING THE IMPROVED TRAINING SET. THE NOISY TRAINING SET CONTAINS 50 TRUE SAMPLES AND 10 MISLABLED SAMPLES

Class	BTC	BTC*	KBTC	KBTC*	SRC	SRC*	SAE	SAE*	SVM	SVM*
1	59.65(5.84)	63.59(2.11)	62.61(4.28)	66.51(3.74)	47.43(4.33)	49.66(2.21)	52.59(5.01)	56.16(4.11)	92.09(4.43)	95.72(7.93)
2	59.60(5.09)	67.11(2.32)	52.84(5.67)	65.18(4.61)	46.18(3.75)	57.82(4.02)	47.82(4.12)	48.94(4.61)	92.87(1.53)	93.80(1.31)
3	74.91(2.24)	78.69(4.41)	70.88(3.24)	74.22(3.24)	61.24(3.02)	63.32(3.11)	50.50(4.80)	56.41(6.19)	59.19(3.33)	58.11(6.05)
4	92.41(1.94)	92.40(1.51)	80.02(4.76)	89.91(4.12)	78.45(3.72)	84.54(3.26)	69.17(3.31)	75.66(4.26)	70.75(9.64)	75.79(10.1)
5	99.58(0.31)	99.59(0.23)	93.16(4.89)	99.24(0.91)	96.49(2.34)	98.26(1.62)	93.42(3.49)	95.62(3.15)	85.95(15.0)	94.23(7.10)
6	68.73(3.40)	71.95(3.83)	66.37(4.23)	74.94(5.21)	48.59(3.51)	56.26(2.22)	57.62(2.78)	63.36(2.13)	54.41(6.97)	56.00(8.13)
7	87.78(2.29)	86.93(2.88)	82.73(4.14)	87.33(2.32)	81.92(3.51)	86.72(2.47)	68.68(5.51)	69.55(4.73)	46.70(5.69)	43.99(6.13)
8	57.29(3.92)	55.42(4.15)	65.95(3.91)	69.24(4.11)	59.66(2.01)	61.42(5.47)	54.81(8.07)	55.99(6.43)	74.78(4.81)	74.95(7.67)
9	63.47(4.53)	71.22(2.41)	96.74(2.27)	98.46(2.31)	92.51(2.89)	96.99(0.65)	94.95(3.61)	95.96(1.25)	88.39(12.0)	74.66(19.8)
OA	65.72(2.20)	74.08(1.42)	62.95(2.37)	71.06(1.42)	54.44(1.61)	62.49(2.02)	55.90(2.21)	57.37(1.86)	76.42(1.96)	78.91(3.37)
AA	68.71(0.79)	76.14(0.91)	66.59(0.98)	70.61(0.45)	58.05(0.77)	61.31(1.22)	65.95(1.51)	67.53(1.17)	73.90(2.59)	73.93(3.41)
kappa	62.67(2.32)	73.87(1.18)	61.96(2.40)	69.64(1.28)	52.12(1.55)	59.22(2.19)	48.41(2.36)	50.22(2.81)	70.27(2.24)	71.87(3.82)

TABLE IX

CLASSIFICATION ACCURACIES OBTAINED BY DIFFERENT SPECTRAL–SPATIAL CLASSIFICATION METHODS ON THE UNIVERSITY OF PAVIA DATA SET. * MARKS THE METHODS TRAINED USING THE IMPROVED TRAINING SET. THE NOISY TRAINING SET CONTAINS 50 TRUE SAMPLES AND 10 MISLABLED SAMPLES

Class	EMP	EMP*	EPF	EPF*	SC-MK	SC-MK*	IFRF	IFRF*	GFDN	GFDN*
1	92.78(2.67)	95.30(1.09)	96.81(2.79)	97.58(4.00)	89.18(5.25)	92.76(3.40)	86.14(7.63)	86.82(4.30)	72.36(9.09)	78.76(10.1)
2	77.30(5.07)	81.44(6.31)	98.05(1.03)	98.98(1.90)	73.19(5.40)	74.40(4.19)	99.01(0.57)	99.07(0.52)	51.94(14.5)	57.92(5.46)
3	95.33(2.46)	93.71(5.17)	84.90(9.94)	85.40(9.45)	98.24(2.09)	98.07(2.59)	77.91(9.58)	80.46(9.20)	78.02(15.0)	90.09(7.99)
4	96.17(1.74)	96.87(1.78)	75.21(13.6)	87.34(10.8)	87.27(5.71)	87.92(4.96)	91.21(3.80)	91.25(6.43)	80.08(13.7)	88.22(5.51)
5	97.67(1.58)	98.58(1.27)	96.82(4.78)	98.10(4.22)	97.98(1.89)	99.81(0.15)	96.13(6.15)	96.89(6.24)	95.14(3.03)	99.41(0.89)
6	92.01(4.78)	92.86(3.55)	61.81(12.4)	68.45(14.1)	94.95(2.89)	93.35(5.50)	95.42(5.69)	96.05(6.10)	71.60(16.6)	81.13(7.19)
7	99.04(0.68)	99.44(1.74)	68.56(11.9)	71.13(10.3)	99.84(0.18)	99.92(0.09)	81.09(8.84)	81.56(7.67)	91.37(4.50)	94.79(5.24)
8	97.66(0.90)	93.34(3.43)	87.17(5.76)	86.03(5.57)	95.96(2.01)	94.01(4.73)	74.74(5.70)	75.36(6.78)	73.43(26.2)	87.50(6.20)
9	98.81(1.69)	99.29(0.44)	95.94(7.23)	94.34(11.0)	97.00(2.49)	98.62(2.11)	61.99(9.68)	57.28(13.6)	99.02(0.80)	99.19(1.15)
OA	86.32(1.32)	89.03(3.12)	84.96(5.03)	87.74(4.58)	83.94(1.95)	88.02(1.10)	87.61(1.75)	90.19(1.36)	66.04(8.60)	73.41(3.15)
AA	94.09(0.79)	94.20(1.39)	85.03(3.92)	86.82(3.37)	92.62(0.84)	93.10(0.85)	84.85(3.02)	85.30(2.49)	79.22(8.24)	86.33(2.51)
kappa	83.59(2.74)	85.93(3.87)	80.99(6.05)	84.31(5.59)	80.46(2.99)	81.44(2.23)	87.18(2.91)	87.27(2.24)	59.70(10.4)	68.54(3.53)

samples are shown in Table VIII. The classification accuracies of different spectral classification methods trained using the original noisy training set and improved training set (mark with *) are compared. It shows that the spectral classifiers using the improved training set always obtain better classification accuracies than those trained with the original noisy training set. For example, for the BTC, KBTC, and SRC methods, the improvements of OA are nearly 5%–8%. For different classes, the improved training set can always lead to higher classification accuracy. In this experiment, the SVM method trained with the improved training set shows the best performance in terms of the highest OA, AA, and Kappa. This experiment further demonstrates the effectiveness of the proposed method in detecting noisy labels.

Furthermore, another experiment is performed to evaluate the effectiveness of the proposed method for spectral–spatial classification methods, i.e., the extended morphological profile (EMP) [52], EPF [53], superpixel-based classification via multiple kernels (SC-MK) [54], image fusion and recursive filtering (IFRF) [55], and the Gabor filtering-based deep learning methods (GFDN) [51], on the University of Pavia data set. Table IX shows the experimental results of different spectral–spatial classification methods trained with the original training set and improved training set (marked with *). It can be seen that the classification accuracies obtained with the improved training set are higher than those obtained with

the original training set. This experiment further demonstrates that the proposed method is applicable for both spectral and spectral–spatial classifiers.

V. CONCLUSION

In this paper, the influence of noisy labels on the performance of HSI classification is analyzed. Furthermore, a novel DP clustering-based method is proposed to detect the noisy labels for HSI classification. Experimental results on four real hyperspectral data sets show the effectiveness of the proposed method in terms of subjective and objective evaluations. However, one limitation of the proposed method is that it has not considered the spatial information in the detection process. Therefore, combining the spectral and spatial information of HSI to further improve the detection performance will be the focus of our future research.

ACKNOWLEDGMENT

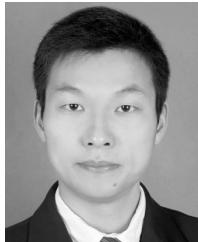
The authors would like to thank the Editor-in-Chief, the anonymous Associate Editor, and reviewers for their insightful comments and suggestions that have greatly improved this paper.

REFERENCES

- [1] Z. Wu, Y. Li, A. Plaza, J. Li, F. Xiao, and Z. Wei, "Parallel and distributed dimensionality reduction of hyperspectral data on cloud computing architectures," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 6, pp. 2270–2278, Jun. 2016.

- [2] X. Kang, S. Li, L. Fang, M. Li, and J. A. Benediktsson, "Extended random walker-based classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 144–153, Jan. 2015.
- [3] M. Garcia and S. L. Ustin, "Detection of interannual vegetation responses to climatic variability using AVIRIS data in a coastal savanna in California," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 7, pp. 1480–1490, Jul. 2001.
- [4] J. Ciermiewski and M. Gulinski, "Furrow microrelief influence on the directional hyperspectral reflectance of soil at various illumination and observation conditions," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4143–4148, Nov. 2010.
- [5] X. Kang, X. Xiang, S. Li, and J. A. Benediktsson, "PCA-based edge-preserving features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7140–7151, Dec. 2017.
- [6] X. Kang, P. Duan, S. Li, and J. A. Benediktsson, "Decolorization-based hyperspectral image visualization," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 4346–4360, Aug. 2018.
- [7] X. Kang, X. Zhang, S. Li, K. Li, J. Li, and J. A. Benediktsson, "Hyperspectral anomaly detection with attribute and edge-preserving filters," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5600–5611, Oct. 2017.
- [8] X. Kang, Y. Huang, S. Li, H. Lin, and J. A. Benediktsson, "Extended random walker for shadow detection in very high resolution remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 2, pp. 867–876, Feb. 2018.
- [9] X. Jin and Y. Gu, "Superpixel-based intrinsic image decomposition of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 8, pp. 4285–4295, Aug. 2017.
- [10] S. Li, Q. Hao, G. Gao, and X. Kang, "The effect of ground truth on performance evaluation of hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2018.2849225](https://doi.org/10.1109/TGRS.2018.2849225).
- [11] Q. Jackson and D. A. Landgrebe, "Adaptive Bayesian contextual classification based on Markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2454–2463, Nov. 2002.
- [12] H. R. Bittencourt, D. A. de Oliveira Moraes, and V. Haertel, "A binary decision tree classifier implementing logistic regression as a feature selection and classification method and its comparison with maximum likelihood," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jan. 2008, pp. 1755–1758.
- [13] F. Ratle, G. Camps-Valls, and J. Weston, "Semisupervised neural networks for efficient hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 5, pp. 2271–2282, May 2010.
- [14] 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.
- [15] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.
- [16] D. G. Stavrokoudis, G. N. Galidakis, I. Z. Gitas, and J. B. Theocaris, "A genetic fuzzy-rule-based classifier for land cover classification from hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 1, pp. 130–148, Jan. 2012.
- [17] J. Liu, Z. Wu, Z. Wei, L. Xiao, and L. Sun, "Spatial-spectral kernel sparse representation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 6, pp. 2462–2471, Dec. 2013.
- [18] X. Zhang, Z. Gao, L. Jiao, and H. Zhou, "Multifeature hyperspectral image classification with local and nonlocal spatial information via Markov random field in semantic space," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1409–1424, Mar. 2018.
- [19] X. Zhang, Y. Liang, C. Li, N. Huyan, L. Jiao, and H. Zhou, "Recursive autoencoders-based unsupervised feature learning for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 11, pp. 1928–1932, Nov. 2017.
- [20] B. Tu, X. Zhang, X. Kang, G. Zhang, J. Wang, and J. Wu, "Hyperspectral image classification via fusing correlation coefficient and joint sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 3, pp. 340–344, Mar. 2018.
- [21] L. Fang, N. He, S. Li, A. J. Plaza, and J. Plaza, "A new spatial-spectral feature extraction method for hyperspectral images using local covariance matrix representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3534–3546, Jun. 2018.
- [22] N. He, L. Fang, S. Li, P. Ghamisi, and J. A. Benediktsson, "Hyperspectral images classification by fusing extinction profiles feature," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2017, pp. 2267–2270.
- [23] N. He, L. Fang, S. Li, A. Plaza, and J. Plaza, "Remote sensing scene classification using multilayer stacked covariance pooling," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2018.2845668](https://doi.org/10.1109/TGRS.2018.2845668).
- [24] B. Tu, W. Kuang, G. Zhao, and H. Fei, "Hyperspectral image classification via superpixel spectral metrics representation," *IEEE Signal Process. Lett.*, vol. 25, no. 10, pp. 1520–1524, Oct. 2018.
- [25] G. Zhao, B. Tu, H. Fei, N. Li, and X. Yang, "Spatial-spectral classification of hyperspectral image via group tensor decomposition," *Neurocomputing*, to be published, doi: [10.1016/j.neucom.2018.07.052](https://doi.org/10.1016/j.neucom.2018.07.052).
- [26] T. Xiao, T. Xia, Y. Yang, C. Huang, and X. Wang, "Learning from massive noisy labeled data for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 2691–2699.
- [27] Z. W. Lu, Z. Fu, T. Xiang, P. Han, L. Wang, and X. Gao, "Learning from weak and noisy labels for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 486–500, Mar. 2017.
- [28] G. M. Foody, "The effect of mis-labeled training data on the accuracy of supervised image classification by SVM," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2015, pp. 4987–4990.
- [29] B. Hou, Q. Wu, Z. Wen, and L. Jiao, "Robust semisupervised classification for PolSAR image with noisy labels," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 11, pp. 6440–6455, Nov. 2017.
- [30] C. Zhang et al., "Methods for labeling error detection in microarrays based on the effect of data perturbation on the regression model," *Bioinformatics*, vol. 25, no. 20, pp. 2708–2714, Aug. 2009.
- [31] X. Tang, L. Jiao, W. J. Emery, F. Liu, and D. Zhang, "Two-stage reranking for remote sensing image retrieval," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 10, pp. 5798–5817, Oct. 2017.
- [32] X. Kang, P. Duan, X. Xiang, S. Li, and J. A. Benediktsson, "Detection and correction of mislabeled training samples for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, to be published, doi: [10.1109/TGRS.2018.2823866](https://doi.org/10.1109/TGRS.2018.2823866).
- [33] B. Tu, C. Zhou, W. Kuang, L. Guo, and X. Ou, "Hyperspectral imagery noisy label detection by spectral angle local outlier factor," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1417–1421, Sep. 2018.
- [34] A. Rodriguez and A. Laio, "Machine learning. Clustering by fast search and find of density peaks," *Science*, vol. 344, no. 6191, pp. 1492–1496, 2014.
- [35] X. Luo, R. Xue, and J. Yin, "Information-assisted density peak index for hyperspectral band selection," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 10, pp. 1870–1874, Oct. 2017.
- [36] G. Tang, S. Jia, and J. Li, "An enhanced density peak-based clustering approach for hyperspectral band selection," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Nov. 2015, pp. 1116–1119.
- [37] S. Jia, G. Tang, J. Zhu, and Q. Li, "A novel ranking-based clustering approach for hyperspectral band selection," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 1, pp. 88–102, Jan. 2016.
- [38] Y. Chen, S. Ma, X. Chen, and P. Ghamisi, "Hyperspectral data clustering based on density analysis ensemble," *Remote Sens. Lett.*, vol. 8, no. 2, pp. 194–203, 2017.
- [39] M. Kearns, "Efficient noise-tolerant learning from statistical queries," in *Proc. ACM Symp. Theory Comput.*, 1993, pp. 392–401.
- [40] N. Manwani and P. Sastry, "Noise tolerance under risk minimization," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 1146–1151, Jun. 2013.
- [41] L. Wu, S. Liu, M. Jian, J. Luo, X. Zhang, and M. Qi, "Reducing noisy labels in weakly labeled data for visual sentiment analysis," in *Proc. Int. Conf. Image Process.*, Sep. 2017, pp. 1322–1326.
- [42] C. Pelletier, S. Valero, J. Ingla, G. Dedieu, and N. Champion, "Filtering mislabeled data for improving time series classification," in *Proc. 9th Int. Workshop Anal. Multitemporal Remote Sens. Images*, Jun. 2017, pp. 1–4.
- [43] K. Sun, X. Geng, and L. Ji, "Exemplar component analysis: A fast band selection method for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 5, pp. 998–1002, May 2015.
- [44] M. Cui and S. Prasad, "Class-dependent sparse representation classifier for robust hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2683–2695, May 2015.
- [45] H. Su, Y. Sheng, H. Yang, and Q. Du, "Orthogonal projection divergence-based hyperspectral band selection," *Spectrosc. Spectral Anal.*, vol. 31, no. 5, pp. 1309–1313, May 2011.
- [46] C. Chang, "Spectral information divergence for hyperspectral image analysis," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, vol. 1, Jun. 1999, pp. 509–511.
- [47] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.

- [48] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
- [49] M. A. Toksöz and I. Ulusoy, "Hyperspectral image classification via basic thresholding classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 7, pp. 4039–4051, Jul. 2016.
- [50] M. A. Toksöz and I. Ulusoy, "Hyperspectral image classification via kernel basic thresholding classifier," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 2, pp. 715–728, Feb. 2017.
- [51] X. Kang, C. Li, S. Li, and H. Lin, "Classification of hyperspectral images by Gabor filtering based deep network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 4, pp. 1166–1178, Apr. 2018.
- [52] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.
- [53] X. Kang, S. Li, and J. A. Benediktsson, "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666–2677, May 2014.
- [54] L. Fang, S. Li, W. Duan, J. Ren, and J. A. Benediktsson, "Classification of hyperspectral images by exploiting spectral-spatial information of superpixel via multiple kernels," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 12, pp. 6663–6674, Dec. 2015.
- [55] X. Kang, S. Li, and J. A. Benediktsson, "Feature extraction of hyperspectral images with image fusion and recursive filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3742–3752, Jun. 2014.



Bing Tu (M'17) received the M.S. degree from the Guilin University of Technology, Guilin, China, in 2009, and the Ph.D. degree from the Beijing University of Technology, Beijing, China, in 2013.

From 2015 to 2016, he was a Visiting Researcher with the Department of Computer Science and Engineering, University of Nevada at Reno, Reno, NV, USA, supported by the China Scholarship Council. Since 2018, he has been an Associate Professor with the School of Information Science and Engineering, Hunan Institute of Science and Technology, Yueyang, China. His research interests include sparse representation, pattern recognition, and analysis in remote sensing.



Xiaofei Zhang (S'17) received the B.S. degree from the Hunan Institute of Science and Technology, Yueyang, China, in 2017, where he is currently pursuing the M.S. degree.

His research interests include the image processing, pattern recognition, hyperspectral image classification, and noisy label detection.



Xudong Kang (S'13–M'15–SM'17) received the B.S. degree from Northeast University, Shenyang, China, in 2007, and the Ph.D. degree from Hunan University, Changsha, China, in 2015.

In 2015, he joined the College of Electrical Engineering, Hunan University, Changsha. His research interests include hyperspectral feature extraction, image classification, image fusion, and anomaly detection.

Dr. Kang received the Second Prize in the Student Paper Competition at International Geoscience and Remote Sensing Symposium (IGARSS) 2014. In IGARSS 2017, he was selected as the Best Reviewer for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS in 2016.



Guoyun Zhang (M'17) received the B.S. degree from Xiangtan University, Xiangtan, China, in 1993, and the M.S. and Ph.D. degrees from Hunan University, Changsha, China, in 2002 and 2006, respectively.

In 1995, he joined the Hunan Institute of Science and Technology, Yueyang, China, where he is currently a Full Professor with the School of Information Science and Technology. In 2013, he was a Visiting Scholar with George Fox University, Newberg, OR, USA. His research interests include image processing, pattern recognition, and digital signal processing application technology.



Shutao Li (M'07–SM'15) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Hunan University, Changsha, China, in 1995, 1997, and 2001, respectively.

In 2001, he joined the College of Electrical and Information Engineering, Hunan University, where he is currently a Full Professor. In 2001, he was a Research Associate with the Department of Computer Science, Hong Kong University of Science and Technology, Hong Kong, where he was a Visiting Professor in 2005. From 2002 to 2003, he was a Post-Doctoral Fellow with the Royal Holloway College, University of London, Surrey, U.K. He is also a Chang-Jiang Scholar Professor appointed by the Ministry of Education of China. He has authored/co-authored over 180 refereed papers. His research interests include compressive sensing, sparse representation, image processing, and pattern recognition.

Dr. Li is a member of the Editorial Board of the Information Fusion, Sensing and Imaging. He received the National Science Fund for Distinguished Young Scholars in China in 2013. He also received two Second-Grade National Awards at the Science and Technology Progress of China in 2004 and 2006. He is an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING and the IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT.