Hierarchical Guidance Filtering Based Ensemble Classification for Hyperspectral Images

Bin Pan, Zhenwei Shi and Xia Xu

Abstract

Joint spectral and spatial information should be fully exploited in order to achieve accurate classification results for hyperspectral images. In this paper, we propose an ensemble framework which combines spectral and spatial information in different scales. The motivation of the proposed method derives from the basic idea: By integrating many individual learners, ensemble learning can achieve better generalization ability than a single learner. In the proposed work, the individual learners are obtained by joint spectral-spatial features generated from different scales. Specially, we develop two techniques to construct the ensemble model, namely, hierarchical guidance filtering (HGF) and matrix of spectral angle distance (mSAD). HGF and mSAD are combined via a weighted ensemble strategy. HGF is a hierarchical edge-preserving filtering operation which could produce diverse sample sets. Meanwhile, in each hierarchy, different spatial contextual information is extracted. With the increase of hierarchy, the pixels spectra tend smooth, while the spatial features are enhanced. Based on the outputs of HGF, a series of classifiers can be obtained. Subsequently, we define a low-rank matrix, mSAD, to measure the diversity among training samples in each hierarchy. Finally, an ensemble strategy is proposed using the obtained individual classifiers and mSAD. We term the proposed method as HiFi-We. Experiments are conducted on two popular data sets, Indian Pines and Pavia University, as well as a challenging hyperspectral data set used in 2014 Data Fusion Contest (GRSS_DFC_2014). An effectiveness analysis about the ensemble strategy is also displayed.

Index Terms

Hyperspectral image classification, hierarchical guidance filtering, ensemble learning.

The work was supported by the National Natural Science Foundation of China under the Grants 61671037, the Beijing Natural Science Foundation under the Grant 4152031, the funding project of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University under the Grant BUAA-VR-16ZZ-03. *(Corresponding author: Zhenwei Shi.)*

Bin Pan, Zhenwei Shi (Corresponding Author) and Xia Xu are with Image Processing Center, School of Astronautics, Beihang University, Beijing 100191, China, and with Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, China, and also with State Key Laboratory of Virtual Reality Technology and Systems, School of Astronautics, Beihang University, Beijing 100191, China, (e-mail: panbin@buaa.edu.cn; shizhenwei@buaa.edu.cn; xuxia@buaa.edu.cn).

I. INTRODUCTION

Hyperspectral sensors can provide images with hundreds of continuous spectral bands as well as high spatial resolution. During the past two decades, hyperspectral images (HSIs) processing techniques have been widely used in many fields such as spectral unmixing [1], mineral identification [2] and environmental monitoring [3]. To better utilize the HSI data, many HSI processing techniques are developed. One of the most important techniques is per-pixel classification, i.e., assign a unique class label to each pixel. However, HSI classification is still challenging, due to many reasons such as the Hughes phenomenon [4].

A popular strategy to improve the classification accuracy is designing multi-feature systems. Gu *et al.* proposed a multiple-kernel learning method by extracting the variation from the different features space [5]. In [6], Gu *et al.* improved the multi-kernel models using low-rank nonnegative matrix factorization. In [7] and [8], spatial information was utilized to enhance the performance of multiple-kernel models. Vector stacking is also a typical approach to address multi-feature problem, which refers to concatenating the multiple features and putting them into a single classifier. In [9], Chen *et al.* combined the magnitude and shape feature spaces via a stacked generalization. In [10], Huang *et al.* compared the performance of vector stacking with other multi-feature methods. However, vector stacking approach does not necessarily lead to better results, because studies have shown that the classification accuracy may vary as a function of the number of selected features [11].

Recently, ensemble learning based methods were developed for HSI classification. By integrating many individual learners, ensemble learning can achieve better generalization performance [12]. In [13], Dietterich *et al.* considered that the incorporation of individual learners could outperform single learner mainly because the following three aspects: First, a single learner may fall into local minima; Second, the ensemble strategy could slightly expand the hypothesis space; At last, because there may be several hypothesises that achieve the same performance in training sets, combining many individual learners could reduce the risk of false hypothesis. For the task of HSI classification, researchers have proposed many ensemble learning based methods. Random forest methods are typical ensemble approaches, and the use of random forest was investigated in [14]–[17], etc. In [18], Xia *et al.* utilized rotation forest [19] for HSI classification and achieved better results than random forest. Support vector machine (SVM) is also sdudied in some ensemble based HSI classification methods [20]. In [21], Pal proposed two ensemble approaches based on SVM using boosting and bagging. In [22], Huang *et al.* combined the spectral, structural and semantic features to construct an SVM ensemble approach. Santos *et al.* performed a combination of six different classification models based on SVM and multilayer perceptron neural network [23]. Xia *et al.* proposed a rotation-based SVM ensemble strategy with limited training samples [24]. Diverse ensemble based HSI classification methods were reported in [25]–[28], etc.

Many studies have also demonstrated that the use of spatial information could significantly improve the classification accuracy [29]–[34]. A common strategy to express the spatial information is using a neighborhood system [31], for example, Markov random field [35], [36] and attribute profile (AP) [37]–[39]. Wavelet-based and Gabor-based methods are also reported in many works. Based on the properties of HSI data, three-dimensional wavelet

are used to extract the texture feature [40]–[42]. In [43], Li *et al.* proposed a Gabor-filtering-based method using nearest regularized subspace (NRS). More recently, edge-preserving filtering (EPF) has become an active research topic in natural scene image processing [44]–[49]. The basic idea of EPF is to remove small details and noise from the image while preserving large-scale edges automatically. In [50], EPF is used to address the task of HSI classification for the first time. In [28], Xia *et al.* combined independent component analysis (ICA) and EPF via an ensemble strategy. However, since EPF is still a kind of smoothing filtering method, it is difficult to determine what level of filtering is the most appropriate. Stronger smoothing could result in better spatial representation, but at the same time lead to more loss of spectral information.

In this paper, we present a novel ensemble learning based HSI classification method, which is composed of joint spectral-spatial features of different scales. Firstly, in order to exploit the joint spectral-spatial information, we propose a hierarchical feature extraction strategy, hierarchical guidance filtering (HGF). HGF is an extension of guided filtering (GF) [44] and rolling guidance filtering (RGF) [46], which is able to generate a series of joint spectral-spatial features. Spatial contextual information of different scales are obtained by the filtering in different hierarchies. Secondly, instead of using complicated optimization techniques, we define a metric matrix, matrix of spectral angle distance (mSAD), to evaluate the feature quality in each hierarchy. Based on the obtained hierarchical features and the evaluation results, a popular ensemble strategy, the weighted voting, is employed to determine the final classification results. We term the proposed method as HiFi-We.

The initial motivations of our study include two aspects: First, we want to combine the joint spectral-spatial information in different scales. The classification model should be determined from a more representative feature space. Second, spectral-spatial features extracted from different scales should have different contributions. More reliable and qualified features should get higher confidence. The solution comes as no surprise: We propose the HGF to obtain a series of spectral-spatial features from different scales; then, we design an ensemble model to simultaneously utilize these features; a new weighting method, mSAD, is also developed.

The major contributions of HiFi-We can be summarized as follows:

- A new ensemble-based HSI classification method is proposed, where joint spectral-spatial information of different scales are combined.
- We develop the hierarchical guidance filtering (HGF) to extract more various spectral-spatial features.
- The matrix of spectral angle distance (mSAD) is designed and used to generate the weight coefficients in the ensemble model.

The remainder of this paper is organized as follows. In section II, we give detailed description of the proposed HSI classification method. Experimental results and analysis are displayed in section III. We conclude this paper in section IV.

II. The proposed method

Ensemble learning can achieve better performance than the best single learner via combining many individual learners [12]. Based on this idea, we propose an HSI classification method using ensemble learning. The proposed

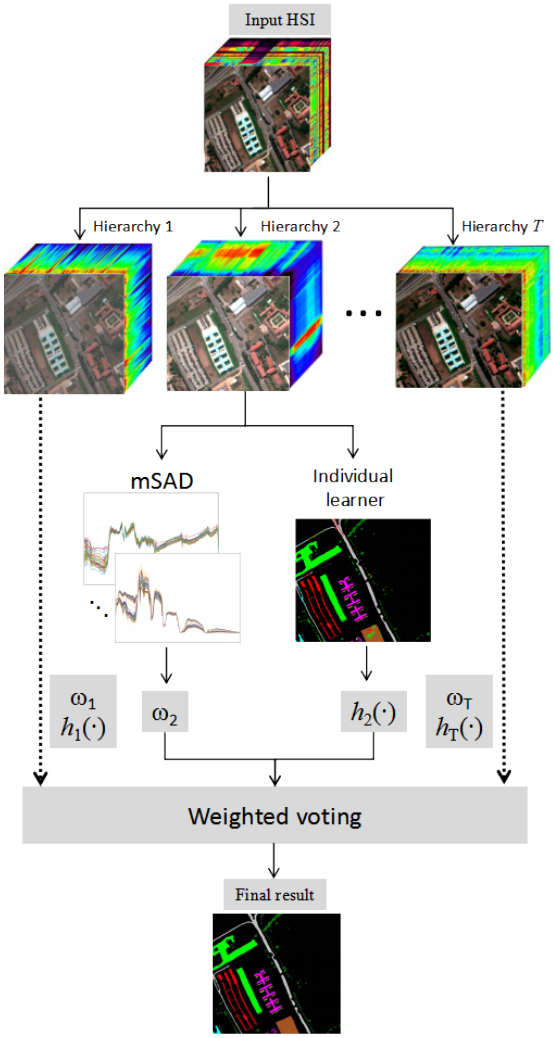


Fig. 1: The flowchart of HiFi-We.

ensemble method contains three component: HGF, mSAD and weighted voting based classification. To generate various joint spectral-spatial features, hierarchical guidance filtering is developed. Based on HGF, an individual learner can be obtained in each hierarchy. Then, the mSAD is designed to evaluate the contribution of each individual learner. At last, weighted voting is conducted to get the final classification results. The flowchart of the proposed method is exhibited in Fig. 1.

A. HGF

As a kind of EPF, HGF is able to remove noise and small details while preserve the overall structure of the image. Therefore, it can be utilized as an implementation to extract the spatial contextual information for HSI classification.

(a) (b) (c) (d) (e)

Fig. 2: Examples of HGF and RGF for Indian Pines data set. HGF and RGF are both conducted on the band 2 of the data set. (a) Guidance image, using the first principal component. (b) Original image in band 2. (c) Results of RGF, rolling 5 times. (d) Results of HGF, 5 hierarchies. (e) Results of HGF, 50 hierarchies.

In the first hierarchy, we conduct a guided filtering (GF) [44], [45] for each the bands of the HSI. Let $\mathbf{Q}^p$ denotes the pth bands of the filtering output, then the output of GF can be expressed by

$$\mathbf{Q}_i^p = a_k^p \mathbf{G}_i + b_k^p, \forall i \in \omega_k, \tag{1}$$

where ω_k is a window around pixel k with size (2r+1)×(2r+1), r is the window radius, i is one of a pixel in ω_k, $\mathbf{G}$ is a guidance image, and a_k^p and b_k^p are coefficients to be estimated. Eq. (1) indicates that the output of the filtering is a linear transform of the guidance image. Conduct the gradient operation for Eq. (1), we can find that

$$\nabla \mathbf{Q}_i^p = a^p \nabla \mathbf{G}_i. \tag{2}$$

According to Eq. (2), the filtering output $\mathbf{Q}^p$ has an edge only if G_i also has an edge, and this is just the reason for edge preserving. Then, we need to determine the linear coefficients a_k^p and b_k^p based on the input HSI data $\mathbf{I}$ and the guidance image $\mathbf{G}$. The following cost function is minimized in the window ω_k:

$$E(a_k^p, b_k^p) = \sum_{i \in \omega_k} ((a_k^p \mathbf{G}_i + b_k^p - \mathbf{I}_i^p)^2 + \epsilon {a_k^p}^2), \tag{3}$$

where ϵ is a parameter controlling the smooth degree. Larger ϵ corresponds to stronger penalization for a_k^p, which leads to smoother output. Eq. (3) guarantees the similarity between input and output of the filtering, meanwhile, noise and small details are removed. Eq. (3) is a linear ridge regression [51], thus it can be solved by

$$\begin{aligned} a_k^p &= \frac{\frac{1}{|w|}\sum_{i \in \omega_k} \mathbf{I}_i^p \mathbf{G}_i - \mu_k \bar{\mathbf{I}}_k^p}{{\sigma_k}^2 + \epsilon}, \\ b_k^p &= \mathbf{I}_k^p - a_k^p \mu_k^p, \end{aligned} \tag{4}$$

where μ_k and σ_k are the mean value and standard variance of $\mathbf{G}$ in ω_k, $\bar{\mathbf{I}}_k^p$ is the mean value of $\mathbf{I}$ in ω_k, and $|\omega|$ is the number of pixels in ω_k.

After obtaining (a_k^p, b_k^p), the output $\mathbf{Q}_i^p$ can be determined by Eq. (1). Then, we use $\mathbf{Q}_i^p$ as the input of the next hierarchy. In other words, the outputs of current hierarchy are considered as the inputs of the next hierarchy. From Fig. 2(d)(e) we can find that the outputs in the 5th and the 50th have certain difference. These difference are

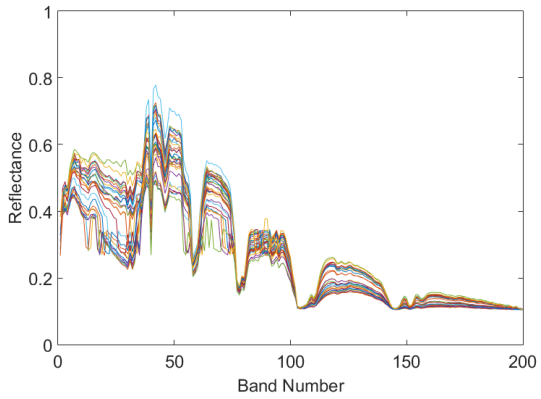
generated because HGF could introduce some variations on both spectral and spatial characteristics for HSI data. In this work, we consider these variations as the joint spectral-spatial information in different scales. To ensure that the spectral information does not suffer severe loss after many hierarchies, here we give parameters r and ϵ as small values as possible, for example, $r = 1$ and $\epsilon = 0.01$.

Generally, the better spatial smoothness, the greater loss of spectral characteristics. It is quite difficult to determine what degree of smooth is the best. In the proposed work, we try to address this problem via a hierarchy strategy, i.e., HGF. HGF is developed to enhance the diversity of samples, where we can get an individual learner based on the output data in every hierarchy. That is to say, for each individual learner, both the number and the dimensionality of the training samples keep the same. To some extent, HGF can be regarded as a linear transform of GF. Compared with some traditional subset selection methods such as bootstrapping and bands selection, using HGF can not only avoid the information loss in each individual learner, but also provide more abundant feature expression.

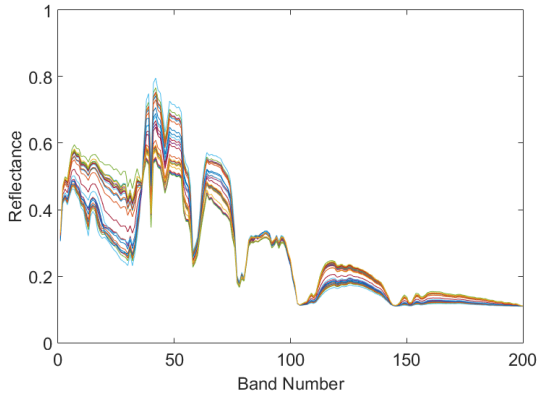
The idea of HGF is similar to that of RGF [46]. However, there are at least two characteristics of HGF seem to challenge RGF. First, in HGF, we run a guided filtering in each hierarchy, while RGF usually adopts joint bilateral filtering [52] in each iteration. Since guided filtering belongs to a linear transform whereas joint bilateral filtering is based on nonlinear model, HGF is more efficient than RGF. Note that RGF can also use guided filtering in each rolling. However, with the increase of rolling times, the results of RGF will get more blurry, as shown in [46]. In other words, guided filtering is not suitable for RGF. More importantly, the guidance images used in HGF and RGF are different. Because RGF is originally designed for natural scene images where only one or three bands are observed, the spectral diversity and correlation are not considered. In RGF, the guidance image is the original input. In this case, with the increase of rolling times, the result image will be more similar to the original image. However, if the original image is seriously polluted by noise, the RGF can hardly lead to satisfying results, as shown in Fig. 2(b)(c). In HGF, we use the first principal component (PC) of the HSI as the guidance image considering that it could provide an ideal representation of the image. Therefore, in HGF, higher hierarchy could generate images more similar to the guidance image, namely, 1st PC (Fig. 2). In Table II and IV, experimental results also demonstrate the superiority of HGF when compared with RGF. Note that HGF is a global operation, which means that it should be conducted on both the training and testing sets.

B. mSAD

Based on HGF, we can obtain many groups of features. The number of features groups is determined by the number of hierarchies, i.e., each hierarchy's outputs correspond to a certain group of features. However, the contributions of different groups may be not equal. Generally, features with high quality have greater weights. Here, we define the term matrix of spectral angle distance (mSAD) to represent the "quality" of samples. The mSAD is based on the assumption that samples of the same class should present similar spectral characteristics. For example, samples in Fig. 3(b) are closer between each other than those in Fig. 3(a), i.e., samples in Fig. 3(b) have higher quality. In this case, the power of feature expression is enhanced, meanwhile, the number of training samples required is declined.



(a)



(b)

Fig. 3: Spectral characteristics (a) before and (b) after HGF (50 hierarchies). We take class *Soybean-notill* in Indian Pines data set for example.

Let $\mathbf{X}_c = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_i, \cdots, \mathbf{x}_n]$ denote a group of the training samples in class c, $\mathbf{x}_i \in \mathbb{R}^{L\times 1}$ is a pixel spectrum with L bands, n is the number of training samples in class c. The spectral angle distance (SAD) between two spectral vector $\mathbf{x}_i$ and $\mathbf{x}_j$ can be expressed by

$$SAD(\mathbf{x}_i, \mathbf{x}_j) = \arccos\left(\frac{\mathbf{x}_i^{\mathrm{T}}\mathbf{x}_j}{\|\mathbf{x}_i\|_2 \cdot \|\mathbf{x}_j\|_2}\right). \tag{5}$$

SAD can be used to measure the difference degree between two pixels, where lower value corresponds to smaller difference. Based on SAD, we first obtain a square matrix:

$$\hat{\mathbf{S}}_c = \begin{bmatrix} SAD(\mathbf{x}_1, \mathbf{x}_1) & \cdots & SAD(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & & \vdots \\ SAD(\mathbf{x}_n, \mathbf{x}_1) & \cdots & SAD(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix} \in \mathbb{R}^{n\times n}. \tag{6}$$

Then, we define the mSAD for $\mathbf{X}_c$ by removing the diagonal elements from $\hat{\mathbf{S}}_c$

$$\begin{aligned}\mathbf{S}_c =&[s_{ij}]\\ =&\begin{bmatrix} SAD(\mathbf{x}_1,\mathbf{x}_2) & \cdots & SAD(\mathbf{x}_1,\mathbf{x}_n)\\ \vdots & & \vdots \\ SAD(\mathbf{x}_n,\mathbf{x}_1) & \cdots & SAD(\mathbf{x}_n,\mathbf{x}_{n-1})\end{bmatrix} \in \mathbb{R}^{n\times(n-1)}.\end{aligned} \tag{7}$$

$\mathbf{S}_c$ is the mSAD for class c. Ideally, $\mathbf{S}_c$ should be $\mathbf{O}_{n\times(n-1)}$, i.e., all the samples in the training set are the same. According to the hypothesis that the testing set shares the consistent distribution as the training set, samples in testing set are also the same as those in the training set, or at least very similar. In this case, only limited samples are necessary for training a powerful model. In real HSI data, this ideal situation is impossible. However, since samples in the same class usually present close spectral characteristics, the $\mathbf{S}_c$ should be low rank. Fewer outliers correspond to lower rank of $\mathbf{S}_c$. Therefore, we use the rank of $\mathbf{S}_c$ to measure the quality of training samples. Usually, the rank of $\mathbf{S}_c$ is relaxed by nuclear norm [53], i.e.,

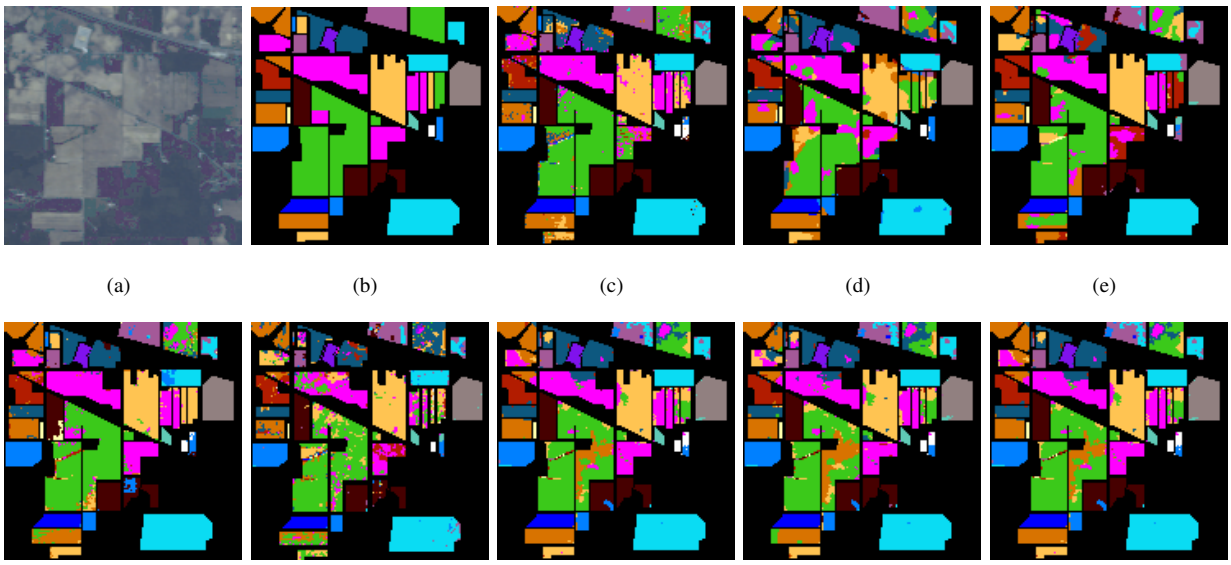
$$R_c = \mathrm{rank}(\mathbf{S}_c) \doteq \|\mathbf{S}_c\|_* = \sum_i \sigma_i(\mathbf{S}_c), \tag{8}$$

where $\sigma_i(\mathbf{S}_c)$ denotes the singular values of $\mathbf{S}_c$, and R_c is the nuclear norm of $\mathbf{S}_c$. Higher R_c indicates that samples in class c are more discrepant. In this case, we can consider that these samples have low "quality", and vice versa.

Because there are many classes in HSI data, we use the mean value of all the R_c to calculate the weight. Based on Eq. (5)-(8), the weight of the tth hierarchy can be obtained by

$$\omega_t = (\frac{1}{C}\sum_{c=1}^{C} R_c)^{-1}, \tag{9}$$

where C is the number of classes in an HSI. Note that the reciprocal is adopted in Eq. (9), because the weight values are negatively related to the within-class diversity.



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

Fig. 4: Indian Pines data set. (a) False color composite image. R-G-B=bands 36-17-11. (b) Ground truth. Each color corresponds to a specific class. Results by (c) GCK (d) NRS (e) EPF-G (f) IIDF (g) NSSNet (h) RGF-W (i) HGF-V and (j) HiFi-We.

Algorithm 1 The proposed HiFi-We method

Input:$\mathbf{I}$, T, r, ϵ, $h(\cdot)$

Initialize: training set, testing set

1.**HGF**

Obtain $\mathbf{G}$ based on PCA for $\mathbf{I}$

For t=1:T

Generate $\mathbf{Q}_t$ by Eq. (1)-(4)

End for

2.**mSAD**

For t=1:T

Determine ω_t by Eq. (5)-(9)

End for

3.**Weighted voting**

Classify by $h_t(\cdot)$

Combine ω_t and the results of $h_t(\cdot)$ by Eq. (10)

Output: Ensemble-based classification results for $\mathbf{I}$

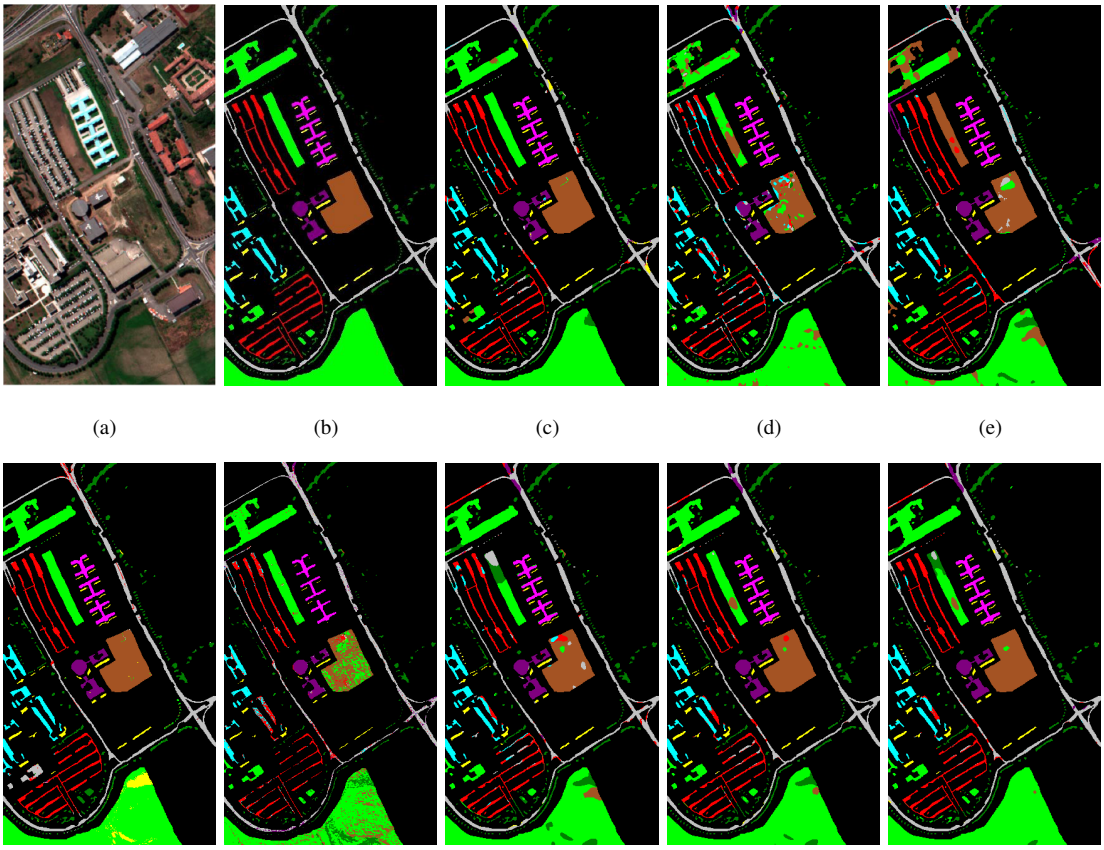


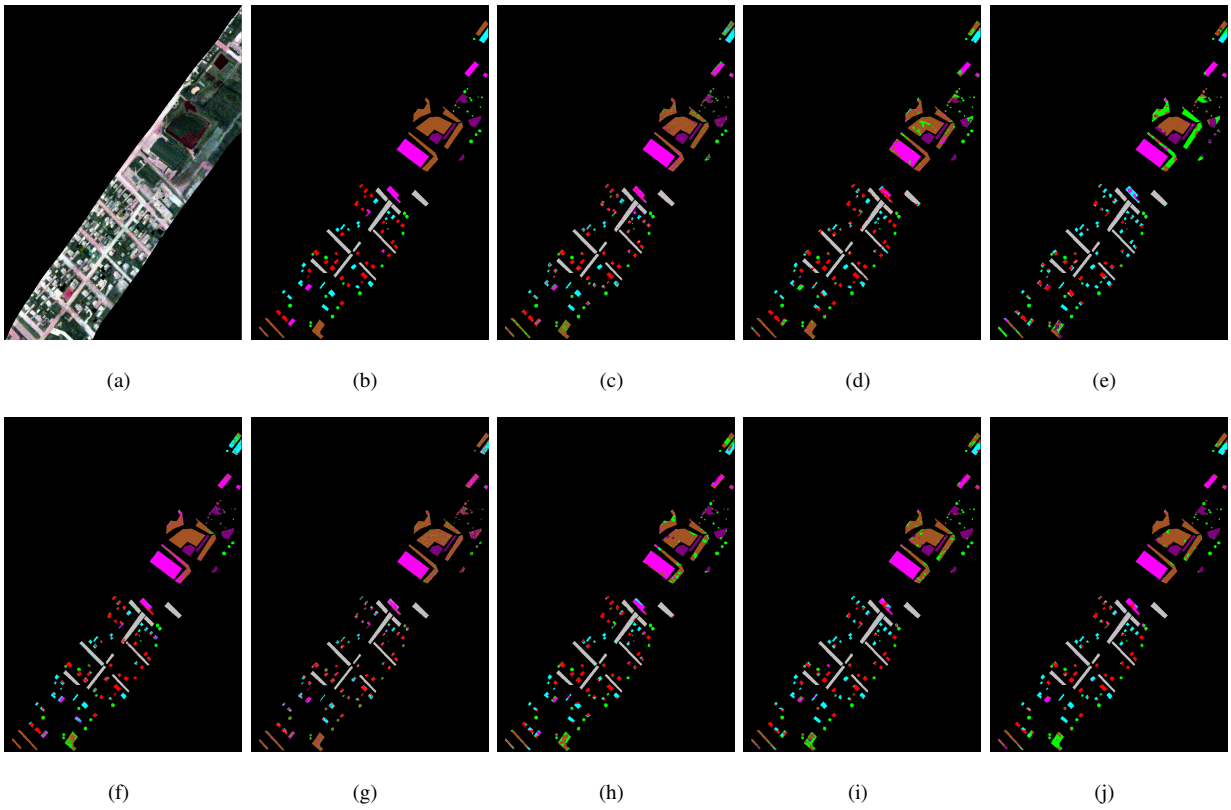
Fig. 5: Pavia University data set. (a) False color composite image. R-G-B=bands 10-27-46. (b) Ground truth. Each color corresponds to a specific class. Results by (c) GCK (d) NRS (e) EPF-G (f) IIDF (g) NSSNet (h) RGF-W (i) HGF-V and (j) HiFi-We.

C. Weighted Voting for Classification

Research has shown that although the class posterior probabilities estimated by individual classifiers are often not very accurate, soft voting usually presents better performance than hard voting [12]. Therefore, in this paper, we adopt a soft voting strategy to determine the labels of test samples. In each hierarchy, we use logistic regression (softmax) classifier to obtain the class posterior probabilities for a test sample. Let h_t denote the classifier in the tth hierarchy, $h_t^c(\mathbf{x}) \in [0,1]$ denote the probability of classifying sample $\mathbf{x}$ to class c. Then the final classification result for $\mathbf{x}$ is determined by

$$H(\mathbf{x}) = \arg\max_c \sum_{t=1}^{T} \omega_t h_t^c(\mathbf{x}), \tag{10}$$

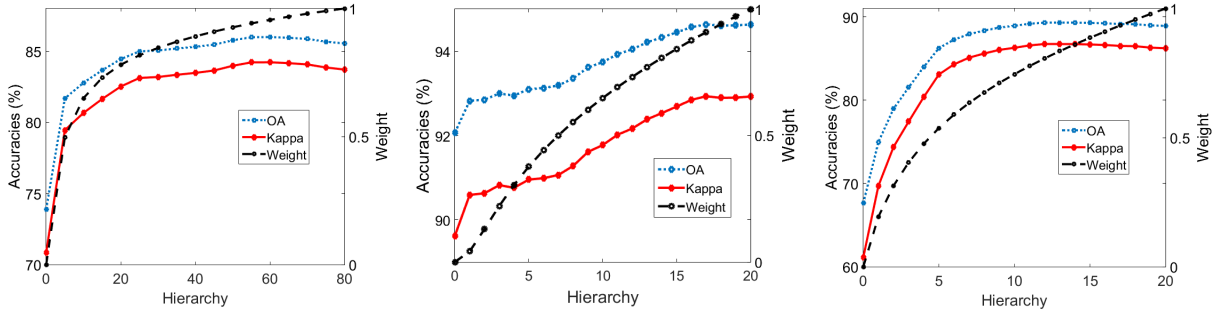
where T is the number of hierarchies, and $H(\mathbf{x})$ is the predicted label. In different hierarchies, the features "$\mathbf{x}$" are not the same. $\mathbf{x}$ is determined by the outputs of HGF. Note that logistic regression is an available classifier, but not the only one. Algorithm 1 depicts the overall process of the proposed method.



(a) (b) (c) (d) (e)

(f) (g) (h) (i) (j)

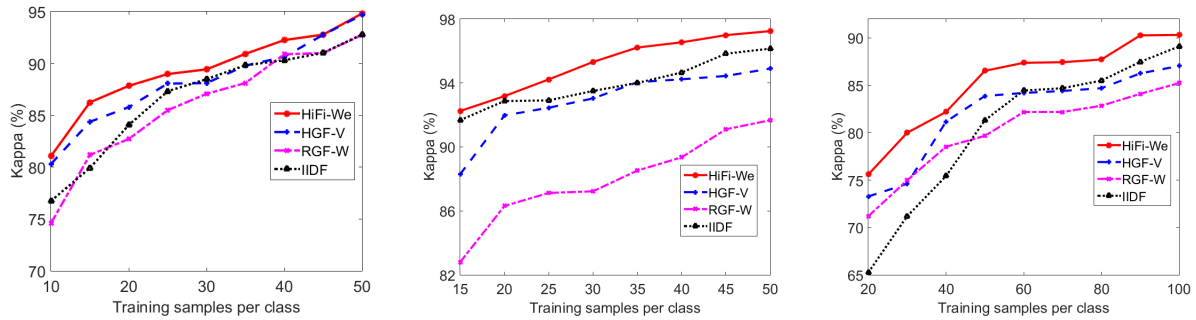
Fig. 6: GRSS_DFC_2014 data set. (a) False color composite image. R-G-B=bands 30-45-66. (b) Ground truth. Each color corresponds to a specific class. Results by (c) GCK (d) NRS (e) EPF-G (f) IIDF (g) NSSNet (h) RGF-W (i) HGF-V and (j) HiFi-We.



(a) (b) (c)

Fig. 7: Influence of T and ω. Results on (a) Indian Pines, (b) Pavia University and (c) GRSS_DFC_2014 data sets. OA and κ correspond to the left coordinate axis, and ω corresponds to the right one.

Readers may doubt the effectiveness of the above ensemble strategy. According to [12], the generalization error



(a) (b) (c)

Fig. 8: Influence of training samples on (a) Indian Pines, (b) Pavia University and (c) GRSS_DFC_2014 data sets.

of an ensemble model is determined by

$$E = \bar{E} - \bar{A}, \tag{11}$$

where $\bar{E}$ is the average error of individual learners, $\bar{A}$ denotes the average ensemble "ambiguity", and E is the generalization error. For a given test sample $\mathbf{x}$, $A(\mathbf{x})$ is obtained by

$$A(\mathbf{x}) = \sum_{t=1}^{T} \omega_t (h_t(\mathbf{x}) - H(\mathbf{x}))^2. \tag{12}$$

Eq. (11) is called error-ambiguity decomposition, which demonstrates that an ensemble model is effective as long as $\bar{E}$ is reduced and $\bar{A}$ is enhanced. Unfortunately, Eq. (11) cannot be optimized directly because $\bar{A}$ is obtained only after the ensemble model is determined [12]. Moreover, it is hard to expand Eq. (11)(12) from regression to classification task. In this paper, the statistical significance analysis and experimental discussion are conducted to verify the effectiveness of the proposed ensemble approach. Details are shown in section III(D)

III. Experiments and Discussion

A. Experimental Setup

In this section, we conduct three types of experiments. We have published the Matlab demo in our homepage[1].

First, we compare the proposed HiFi-We with some state-of-the-art HSI classification algorithms, including GCK [7], NRS [43], EPF-G [50], IIDF [54] and NSSNet [34]. GCK is a multiply kernel learning based method. NRS is developed by Gabor filtering. EPF-G and IIDF are based on edge-preserving filtering (we use GF in this experiment) and intrinsic image decomposition, respectively. NSSNet is a recently proposed deep learning based HSI classification approach. Overall accuracy (OA), average accuracy (AA) and Kappa coefficient (κ) are used to evaluate the performances of all the methods. In order to verify the effectiveness of the "HGF" and "mSAD", we replace HGF by RGF (RGF-W), as well as use simple majority voting (HGF-V). HGF-V and RGF-W can be regarded as extensions of the HiFi-We. The performance of RGF-W and HGF-V are also reported.

[1]Available online: http://levir.buaa.edu.cn/Code.htm

Second, the influence of important parameters is discussed. The parameters used in HiFi-We are listed in Table I.

TABLE I: PARAMETERS USED IN HIFI-WE.

Symbol	Value	Name
T	80/20	Number of hierarchies
r	1	Local window radius in HGF
ϵ	1	Regularization parameter in HGF

At last, we give statistical evaluation about the effectiveness of the proposed ensemble strategy. Furthermore, we also verify that the improvement achieved by the proposed method is significant.

Three data sets are used in our experiments, namely Indian Pines, Pavia University[2] and GRSS_DFC_2014 [55].

- Indian Pines is a widely used data set for HSI classification. It was acquired by airborne visible/infrared imaging spectrometer (AVIRIS) in Northwestern Indiana, with the wavelengths ranges from 0.4 to 2.5 μm. The spatial resolution is 20m, and the size is 145×145 pixels. After removing the water absorption bands, there are 200 spectral bands remain. Totally 10249 pixels are labeled, and they are classified into 16 classes. Figure 4(a)(b) are false color composite image and the corresponding ground truth for this data.
- Pavia University data set was collected over the city of Pavia, Italy, by reflective optics system imaging spectrometer (ROSIS-3) sensor. This data set contains 42776 labeled pixels which are composed of 9 different classes. It has 1.3m spatial resolution and 610$\times$340 pixels size. After removing the noise bands totally 103 channels are preserved. A false color composite image and the corresponding ground truth image are shown in Fig. 5(a)(b).
- GRSS_DFC_2014 is the long-wave infrared (LWIR, thermal infrared) hyperspectral data set used in the 2014 IEEE GRSS Data Fusion Contest. It was acquired by an 84-channel airborne LWIR hyperspectral imager covering an urban area near Thetford Mines in Québec, Canada, with the wavelengths between 7.8 to 11.5 μm and approximately 1m spatial resolution. The size of this data set is 795×564 pixels. 22532 labeled pixels and a ground truth with 7 land cover classes are provided. Since this data set is collected from LWIR bands, its quality is much lower than that of Indian Pines. Therefore, this data set is more challenging. A false color composite image and the corresponding ground truth are shown in Fig. 6(a)(b).

B. Classification Results

We validate the superiority of the proposed HiFi-We method in the three data sets. All the methods are conducted 50 times and the average results are reported. The standard deviations of HiFi-We are also revealed. The number of hierarchies in Indian Pines, Pavia University and GRSS_DFC_2014 data sets are set as 80, 20 and 20, respectively.

[2]Available online: http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes

TABLE II: CLASSIFICATION ACCURACIES OF DIFFERENT METHODS ON INDIAN PINES DATA SET (%).

Class	Samples		Methods							
	Train	Test	GCK	NRS	EPF-G	IIDF	NSSNet	RGF-W	HGF-V	HiFi-We
Alfalfa	20	26	96.94	99.42	62.07	87.22	95.92	98.38	99.31	**99.46±1.35**
Corn-notill	20	1408	75.67	70.45	74.83	80.45	56.58	77.66	80.59	**81.91±5.58**
Corn-mintill	20	810	81.59	74.25	74.68	75.89	66.56	86.10	90.03	**91.49±4.52**
Corn	20	217	93.17	95.10	42.07	66.03	91.32	95.51	96.36	**96.78±3.84**
Grass-pasture	20	463	89.70	87.51	**94.83**	93.49	89.13	90.38	89.75	90.06±3.88
Grass-trees	20	710	97.57	92.35	94.39	97.67	95.25	96.97	96.59	**97.92±1.80**
Grass-pasture-mowed	14	14	97.54	**100.0**	91.01	54.09	99.50	**100.0**	97.00	96.75±5.54
Hay-windrowed	20	458	99.42	98.56	99.85	**99.92**	96.45	98.66	99.10	99.39±0.92
Oats	10	10	**100.0**	99.59	78.18	44.83	99.60	99.60	**100.0**	**100.0±0.00**
Soybean-notill	20	952	80.98	73.53	69.21	73.58	76.64	81.57	87.94	**88.16±6.63**
Soybean-mintill	20	2435	79.87	69.93	84.44	**92.37**	56.37	74.29	76.64	79.82±5.86
Soybean-clean	20	573	84.07	81.23	59.43	79.13	78.20	80.56	91.94	**93.31±3.18**
Wheat	20	185	**99.56**	98.76	99.50	99.54	99.45	98.41	99.45	99.41±0.29
Woods	20	1245	93.74	87.16	97.27	**99.07**	91.57	94.26	95.62	96.96±2.79
Buildings-Grass-Trees-Drives	20	366	93.16	90.80	71.08	84.73	72.87	**96.36**	94.01	95.23±2.72
Stone-Steel-Towers	20	73	95.71	98.65	81.16	94.63	**99.12**	97.26	99.01	99.07±0.65
OA			85.54	79.38	78.45	85.90	73.72	84.79	87.46	**89.06±1.70**
AA			91.17	88.58	79.63	82.67	85.28	91.62	93.33	**94.11±0.77**
κ			83.61	76.60	75.64	84.03	70.42	82.78	85.79	**87.51±1.90**

TABLE III: CLASSIFICATION ACCURACIES OF DIFFERENT METHODS ON PAVIA UNIVERSITY DATA SET (%).

Class	Samples		Methods							
	Train	Test	GCK	NRS	EPF-G	IIDF	NSSNet	RGF-W	HGF-V	HiFi-We
Asphalt	20	6611	80.91	84.16	**96.54**	94.66	82.17	83.96	88.44	89.48±2.26
Meadows	20	18629	97.78	84.08	94.66	**99.32**	79.43	90.12	95.24	95.94±1.42
Gravel	20	2079	74.59	83.69	80.74	**96.07**	81.25	87.18	91.15	91.61±1.66
Trees	20	3044	80.00	91.68	73.77	86.11	91.94	89.55	92.03	**92.79±1.12**
Painted metal sheets	20	1325	99.73	**99.99**	94.61	99.09	99.87	96.54	98.68	99.14±0.48
Bare	20	5009	87.68	86.45	60.07	94.69	75.13	93.78	98.06	**98.62±1.53**
Bitumen	20	1310	79.03	86.29	76.99	92.59	92.89	91.41	98.53	**99.15±0.36**
Self-Blocking Bricks	20	3662	70.02	77.74	84.95	84.95	78.43	84.85	91.27	**91.72±1.68**
Shadows	20	927	62.35	92.91	**98.39**	91.61	98.15	97.58	93.28	94.14±1.53
OA(%)			86.35	85.11	83.54	94.62	81.66	89.49	93.91	**94.93±0.63**
AA(%)			81.34	87.44	84.53	93.23	86.58	91.00	94.07	**95.26±0.34**
$\kappa\times100$			82.29	80.85	79.11	92.97	76.57	86.31	91.97	**93.29±0.82**

TABLE IV: CLASSIFICATION ACCURACIES OF DIFFERENT METHODS ON GRSS_DFC_2014 DATA SET (%).

Class	Samples		Methods							
	Train	Test	GCK	NRS	EPF-G	IIDF	NSSNet	RGF-W	HGF-V	HiFi-We
Road	50	4393	95.75	94.64	97.21	98.49	98.37	98.24	98.79	**98.95±0.82**
Trees	50	1043	63.67	76.76	17.09	48.31	27.89	76.83	80.62	**81.09±6.25**
Red roof	50	1804	64.15	73.90	52.93	67.14	60.90	77.21	80.87	**83.33±3.71**
Grey roof	50	2076	68.78	**85.40**	56.89	74.54	66.24	75.29	80.83	83.89±4.30
Concrete roof	50	3838	78.70	72.19	**92.89**	82.30	79.07	83.98	86.74	87.88±3.91
Vegetation	50	7307	66.97	83.93	87.65	**96.17**	84.47	76.23	82.43	85.57±3.75
Bare soil	50	1721	81.62	89.49	77.03	89.66	68.45	92.43	93.42	**93.98±3.19**
OA			75.65	83.37	68.53	84.83	78.62	83.21	86.91	**88.72±1.51**
AA			74.24	82.30	68.81	79.51	69.81	82.89	86.24	**87.81±1.49**
κ			70.33	79.46	62.50	81.32	73.10	79.45	83.88	**86.06±1.82**

TABLE V: MCNEMAR'S TEST FOR INDIAN PINES DATA SET.

	h_1	h_2	h_3	h_4	h_5	h_6	h_7	h_8	h_9	h_{10}	h_{11}	h_{12}	h_{13}	h_{14}	h_{15}	h_{16}	h_{17}	h_{18}	h_{19}	h_{20}
h_1	0.00	14.84	20.48	22.40	22.97	24.30	24.92	25.35	25.82	26.56	27.44	27.97	28.74	28.95	29.48	29.77	30.13	30.21	30.39	30.69
h_2	14.84	0.00	13.12	15.39	15.79	17.02	17.63	18.04	18.63	19.42	20.45	20.99	21.80	21.98	22.51	22.76	23.17	23.29	23.42	23.82
h_3	20.48	13.12	0.00	7.91	8.89	10.77	11.62	12.25	12.82	13.68	14.90	15.47	16.38	16.55	17.14	17.44	17.91	18.06	18.21	18.61
h_4	22.40	15.39	7.91	0.00	4.26	7.21	8.37	9.17	9.83	10.85	12.24	12.85	13.86	14.06	14.66	14.94	15.45	15.61	15.74	16.12
h_5	22.97	15.79	8.89	4.26	0.00	6.04	7.36	8.23	8.85	9.95	11.43	12.03	13.11	13.29	13.91	14.19	14.75	14.89	15.02	15.40
h_6	24.30	17.02	10.77	7.21	6.04	0.00	4.16	5.57	6.47	7.88	9.66	10.38	11.54	11.73	12.43	12.70	13.32	13.46	13.60	13.98
h_7	24.92	17.63	11.62	8.37	7.36	4.16	0.00	3.71	4.88	6.62	8.68	9.42	10.68	10.87	11.64	11.88	12.52	12.64	12.79	13.20
h_8	25.35	18.04	12.25	9.17	8.23	5.57	3.71	0.00	3.15	5.35	7.76	8.59	9.94	10.14	10.99	11.21	11.90	12.03	12.16	12.54
h_9	25.82	18.63	12.82	9.83	8.85	6.47	4.88	3.15	0.00	4.48	7.30	8.14	9.61	9.77	10.58	10.80	11.46	11.58	11.69	12.09
h_{10}	26.56	19.42	13.68	10.85	9.95	7.88	6.62	5.35	4.48	0.00	5.74	6.76	8.48	8.65	9.56	9.80	10.52	10.65	10.77	11.21
h_{11}	27.44	20.45	14.90	12.24	11.43	9.66	8.68	7.76	7.30	5.74	0.00	3.68	6.19	6.48	7.64	7.96	8.77	8.94	9.09	9.60
h_{12}	27.97	20.99	15.47	12.85	12.03	10.38	9.42	8.59	8.14	6.76	3.68	0.00	5.08	5.35	6.71	7.02	7.94	8.13	8.29	8.84
h_{13}	28.74	21.80	16.38	13.86	13.11	11.54	10.68	9.94	9.61	8.48	6.19	5.08	0.00	2.29	4.50	5.03	6.19	6.40	6.58	7.25
h_{14}	28.95	21.98	16.55	14.06	13.29	11.73	10.87	10.14	9.77	8.65	6.48	5.35	2.29	0.00	4.01	4.45	5.77	5.98	6.16	6.80
h_{15}	29.48	22.51	17.14	14.66	13.91	12.43	11.64	10.99	10.58	9.56	7.64	6.71	4.50	4.01	0.00	2.29	4.19	4.51	4.77	5.55
h_{16}	29.77	22.76	17.44	14.94	14.19	12.70	11.88	11.21	10.80	9.80	7.96	7.02	5.03	4.45	2.29	0.00	3.96	4.04	4.28	5.16
h_{17}	30.13	23.17	17.91	15.45	14.75	13.32	12.52	11.90	11.46	10.52	8.77	7.94	6.19	5.77	4.19	3.96	0.00	1.80	2.41	3.68
h_{18}	30.21	23.29	18.06	15.61	14.89	13.46	12.64	12.03	11.58	10.65	8.94	8.13	6.40	5.98	4.51	4.04	1.80	0.00	1.61	3.24
h_{19}	30.39	23.42	18.21	15.74	15.02	13.60	12.79	12.16	11.69	10.77	9.09	8.29	6.58	6.16	4.77	4.28	2.41	1.61	0.00	2.89
h_{20}	30.69	23.82	18.61	16.12	15.40	13.98	13.20	12.54	12.09	11.21	9.60	8.84	7.25	6.80	5.55	5.16	3.68	3.24	2.89	0.00

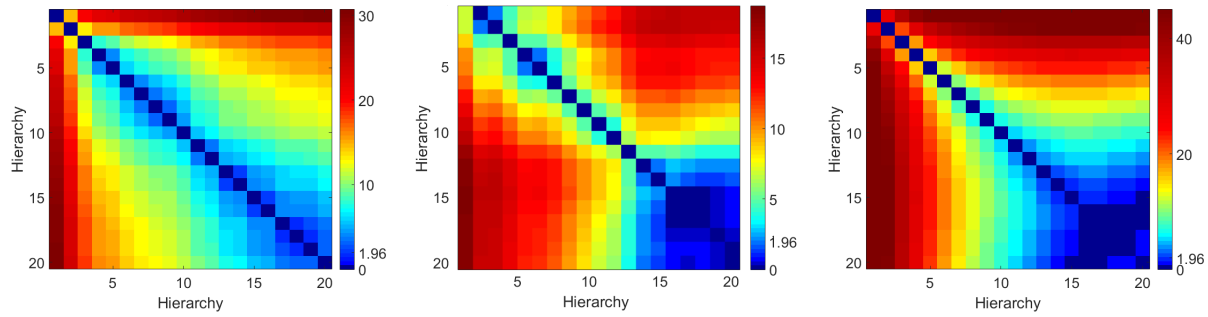
1) Results on Indian Pines data set: Indian Pines data set is widely used in many works. Here, we only use 20 samples in each class for training, and the rests for testing. Fig. 4(c)-(j) display the overall classification maps of all the compared methods. We can see that strong spatial correlation is observed. In Table II, the quantitative results of different methods are reported. Compared with GCK, EPF, IIDF and NSSNet, the proposed method achieves about 4% advantage in OA, AA and κ. Among all the 16 classes, HiFi-We performs best in 8 classes. Specially, 13 of all

TABLE VI: MCNEMAR'S TEST FOR PAVIA UNIVERSITY DATA SET.

	h_1	h_2	h_3	h_4	h_5	h_6	h_7	h_8	h_9	h_{10}	h_{11}	h_{12}	h_{13}	h_{14}	h_{15}	h_{16}	h_{17}	h_{18}	h_{19}	h_{20}
h_1	0.00	6.82	6.75	8.55	10.06	10.32	11.26	12.59	13.82	14.93	15.78	17.25	18.15	18.40	18.70	18.50	18.36	18.37	18.37	18.27
h_2	6.82	0.00	1.89	4.67	6.60	6.90	7.94	9.43	10.79	11.97	12.85	14.41	15.34	15.57	15.85	15.66	15.51	15.49	15.49	15.36
h_3	6.75	1.89	0.00	4.88	7.00	7.04	8.13	9.68	11.09	12.29	13.14	14.73	15.65	15.80	16.03	15.75	15.52	15.46	15.40	15.21
h_4	8.55	4.67	4.88	0.00	4.89	4.92	6.26	7.98	9.53	10.82	11.72	13.39	14.32	14.44	14.65	14.34	14.10	14.03	13.98	13.77
h_5	10.06	6.60	7.00	4.89	0.00	1.83	3.87	6.09	7.87	9.32	10.29	12.06	13.06	13.17	13.38	13.02	12.73	12.65	12.58	12.35
h_6	10.32	6.90	7.04	4.92	1.83	0.00	3.70	6.19	8.02	9.49	10.40	12.18	13.20	13.26	13.47	13.04	12.67	12.54	12.46	12.16
h_7	11.26	7.94	8.13	6.26	3.87	3.70	0.00	5.01	7.11	8.64	9.63	11.52	12.60	12.62	12.82	12.33	11.91	11.73	11.61	11.30
h_8	12.59	9.43	9.68	7.98	6.09	6.19	5.01	0.00	4.91	6.84	7.97	10.12	11.28	11.28	11.47	10.93	10.45	10.24	10.09	9.78
h_9	13.82	10.79	11.09	9.53	7.87	8.02	7.11	4.91	0.00	4.70	6.20	8.76	10.06	10.02	10.18	9.57	9.08	8.86	8.68	8.36
h_{10}	14.93	11.97	12.29	10.82	9.32	9.49	8.64	6.84	4.70	0.00	4.02	7.38	8.86	8.78	8.92	8.24	7.70	7.47	7.28	6.98
h_{11}	15.78	12.85	13.14	11.72	10.29	10.40	9.63	7.97	6.20	4.02	0.00	6.39	8.02	7.77	7.85	7.10	6.53	6.31	6.14	5.85
h_{12}	17.25	14.41	14.73	13.39	12.06	12.18	11.52	10.12	8.76	7.38	6.39	0.00	4.80	4.72	5.08	4.43	3.98	3.88	3.81	3.60
h_{13}	18.15	15.34	15.65	14.32	13.06	13.20	12.60	11.28	10.06	8.86	8.02	4.80	0.00	1.88	2.81	2.38	2.09	2.13	2.15	2.01
h_{14}	18.40	15.57	15.80	14.44	13.17	13.26	12.62	11.28	10.02	8.78	7.77	4.72	1.88	0.00	2.09	1.59	1.34	1.45	1.50	1.39
h_{15}	18.70	15.85	16.03	14.65	13.38	13.47	12.82	11.47	10.18	8.92	7.85	5.08	2.81	2.09	0.00	0.23	0.22	0.50	0.64	0.59
h_{16}	18.50	15.66	15.75	14.34	13.02	13.04	12.33	10.93	9.57	8.24	7.10	4.43	2.38	1.59	0.23	0.00	0.08	0.45	0.61	0.54
h_{17}	18.36	15.51	15.52	14.10	12.73	12.67	11.91	10.45	9.08	7.70	6.53	3.98	2.09	1.34	0.22	0.08	0.00	0.56	0.70	0.59
h_{18}	18.37	15.49	15.46	14.03	12.65	12.54	11.73	10.24	8.86	7.47	6.31	3.88	2.13	1.45	0.50	0.45	0.56	0.00	0.43	0.31
h_{19}	18.37	15.49	15.40	13.98	12.58	12.46	11.61	10.09	8.68	7.28	6.14	3.81	2.15	1.50	0.64	0.61	0.70	0.43	0.00	0.00
h_{20}	18.27	15.36	15.21	13.77	12.35	12.16	11.30	9.78	8.36	6.98	5.85	3.60	2.01	1.39	0.59	0.54	0.59	0.31	0.00	0.00

TABLE VII: MCNEMAR'S TEST FOR GRSS_DFC_2014 DATA SET.

	h_1	h_2	h_3	h_4	h_5	h_6	h_7	h_8	h_9	h_{10}	h_{11}	h_{12}	h_{13}	h_{14}	h_{15}	h_{16}	h_{17}	h_{18}	h_{19}	h_{20}
h_1	0.00	28.10	36.97	42.68	46.97	49.85	51.96	53.48	54.64	55.62	56.36	56.87	57.29	57.65	57.82	57.93	57.89	57.80	57.67	57.44
h_2	28.10	0.00	21.74	28.97	33.88	36.97	39.17	40.72	41.87	42.92	43.67	44.17	44.59	44.91	45.03	45.06	44.93	44.75	44.54	44.22
h_3	36.97	21.74	0.00	18.03	24.40	27.94	30.35	31.96	33.16	34.25	35.02	35.52	35.92	36.22	36.31	36.31	36.13	35.88	35.61	35.24
h_4	42.68	28.97	18.03	0.00	15.48	20.10	23.03	24.87	26.23	27.44	28.26	28.78	29.21	29.50	29.56	29.53	29.31	29.03	28.72	28.30
h_5	46.97	33.88	24.40	15.48	0.00	12.04	16.10	18.37	20.00	21.40	22.32	22.88	23.34	23.65	23.71	23.69	23.45	23.15	22.81	22.37
h_6	49.85	36.97	27.94	20.10	12.04	0.00	9.99	13.08	15.10	16.74	17.81	18.43	18.94	19.26	19.34	19.33	19.08	18.76	18.40	17.94
h_7	51.96	39.17	30.35	23.03	16.10	9.99	0.00	8.02	10.80	12.78	14.02	14.72	15.30	15.66	15.76	15.75	15.50	15.18	14.80	14.33
h_8	53.48	40.72	31.96	24.87	18.37	13.08	8.02	0.00	6.85	9.47	10.98	11.81	12.49	12.90	13.02	13.02	12.76	12.43	12.04	11.55
h_9	54.64	41.87	33.16	26.23	20.00	15.10	10.80	6.85	0.00	6.34	8.35	9.36	10.15	10.61	10.75	10.78	10.51	10.17	9.78	9.28
h_{10}	55.62	42.92	34.25	27.44	21.40	16.74	12.78	9.47	6.34	0.00	5.28	6.70	7.72	8.27	8.47	8.54	8.29	7.95	7.56	7.07
h_{11}	56.36	43.67	35.02	28.26	22.32	17.81	14.02	10.98	8.35	5.28	0.00	4.08	5.51	6.23	6.52	6.66	6.44	6.12	5.74	5.25
h_{12}	56.87	44.17	35.52	28.78	22.88	18.43	14.72	11.81	9.36	6.70	4.08	0.00	3.56	4.56	4.95	5.17	4.98	4.67	4.31	3.82
h_{13}	57.29	44.59	35.92	29.21	23.34	18.94	15.30	12.49	10.15	7.72	5.51	3.56	0.00	2.75	3.38	3.70	3.55	3.28	2.93	2.46
h_{14}	57.65	44.91	36.22	29.50	23.65	19.26	15.66	12.90	10.61	8.27	6.23	4.56	2.75	0.00	1.86	2.34	2.25	2.02	1.70	1.23
h_{15}	57.82	45.03	36.31	29.56	23.71	19.34	15.76	13.02	10.75	8.47	6.52	4.95	3.38	1.86	0.00	1.49	1.41	1.18	0.89	0.42
h_{16}	57.93	45.06	36.31	29.53	23.69	19.33	15.75	13.02	10.78	8.54	6.66	5.17	3.70	2.34	1.49	0.00	0.45	0.36	0.12	0.34
h_{17}	57.89	44.93	36.13	29.31	23.45	19.08	15.50	12.76	10.51	8.29	6.44	4.98	3.55	2.25	1.41	0.45	0.00	0.05	0.18	0.67
h_{18}	57.80	44.75	35.88	29.03	23.15	18.76	15.18	12.43	10.17	7.95	6.12	4.67	3.28	2.02	1.18	0.36	0.05	0.00	0.30	0.86
h_{19}	57.67	44.54	35.61	28.72	22.81	18.40	14.80	12.04	9.78	7.56	5.74	4.31	2.93	1.70	0.89	0.12	0.18	0.30	0.00	0.94
h_{20}	57.44	44.22	35.24	28.30	22.37	17.94	14.33	11.55	9.28	7.07	5.25	3.82	2.46	1.23	0.42	0.34	0.67	0.86	0.94	0.00



(a) (b) (c)

Fig. 9: McNermar's test for (a) Indian Pines data set, (b) Pavia University data set and (c) GRSS_DFC_2014 data set.

classes present over 90% accuracies. RGF-W and HGF-V also perform well, however, slightly gaps could still be observed. This results indicate that the improvements in HGF and mSAD are valid. It seems that HGF contributes more to the final results.

Fig. 8(a) presents the influence of training samples size on this data set. All the compared methods achieve similar accuracies, especially when the training samples number is large. The advantage of HiFi-We mainly reflects in the case of limited samples, such as 10-20 per class. Performing well with limited samples is meaningful, since a very simple method may also work well with abundant samples.

2) Results on Pavia University data set: Fig. 5(c)-(j) and Table III show the classification results on Pavia University data set. Compared with Indian Pines, all the methods perform better on Pavia University data set. This maybe because the latter has higher spatial resolution. HiFi-We method still outperforms others, and the superiority on AA is more obvious. When training samples are limited, the AA becomes a more important measure. Since several methods have achieved above 93% accuracies, in this case 1.5-2% advantages are also significant. Among all the 9 classes, HiFi-We performs best in 4 ones, and exceeds 90% in 8 ones. The influence of training samples size on this data set is shown in Fig. 8(b). Increasing training samples will lead to better performance. When the training samples number is up to 50 per class, HiFi-We presents nearly 98% κ.

3) Results on GRSS_DFC_2014 data set: This data set is collected from LWIR bands, and the imaging quality is relatively low. Thus, this data set is more challenging. For all the compared methods, 50 samples in each class are used for training. The visual classification maps of the all the methods are revealed in Fig. 6(c)-(j). In Table IV, we reveal the objective evaluation about the classification accuracies. About 5% advantage is observed in OA, AA and κ. HiFi-We presents the best performance in 4 classes, and above 80% accuracy in all the classes. Compared with RGF-W and HGF-V, HiFi-We outperforms them by 2%-5%. Both the HGF and mSAD work, and HGF still contributes more to the accuracies. Similar conclusion could be reached from the experiments in Indian Pines and Pavia University data sets.

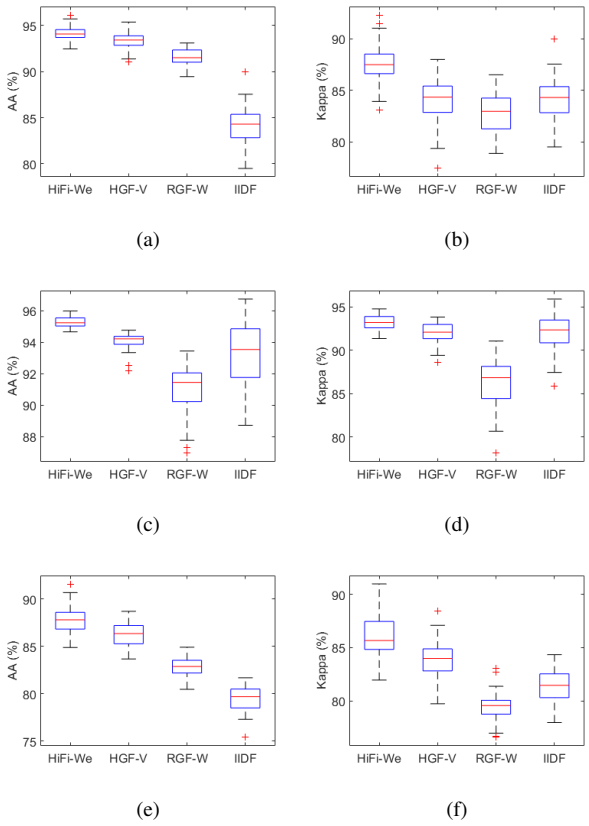


Fig. 10: Statistical evaluation. (a) AA and (b) κ for Indian Pines, (c) AA and (d) κ for Pavia University, (e) AA and (f) κ for GRSS_DFC_2014 data sets.

The influence of training samples size is shown in Fig. 8(c). In this data set, the gaps are more apparent. When the training samples number is below 50 per class, the proposed method presents 5%-10% advantage. This mainly because the quality of this data is relatively lower than the other two, while the HGF operation has actually removed some noise and improved the image quality.

C. Parameters Analysis

As we mentioned above, to avoid the spectral information loss, the parameters r and ϵ should be as small as possible. In this case, the different scales' spatial information is mainly extracted by different hierarchies. Therefore, the number of hierarchies T is the most important parameter in HiFi-We. In addition, in section II(B), we claim that the weighting strategy by Eq. (5)-(9) are meaningful. Here, we design a uniform experiment to evaluate the influence of T and the weights ω, as shown in Fig. 7. Fig. 7 is also an illustration about the influence of different scales' spatial information. We can see the classification accuracies keep increasing at the original hierarchies, and after that they tend stable. This result indicates that the different scales' spatial information is not the same, i.e., it really has some influence on the final classification results. Note that the OA and κ in this figure correspond

to the results of each individual learner. Sharp increases in OA and κ are observed at the first several hierarchies. For Indian Pines data set, the OA and κ keep stable after the 20th hierarchy. Similar phenomena appear at Pavia University and GRSS_DFC_2014 data sets after the 15th and 10th hierarchies, respectively. Further increasing the number hierarchies contribute little to the individual learners. Moreover, we find that the curves of ω present similar tendency as OA and κ. In general, the hierarchical strategy could really lead to diverse classification results, at the same time, the accuracies and the weights have consistent trend.

D. Statistical Evaluation

In this part, we give statistical evaluation about the effectiveness of the ensemble strategy. According to Eq. (11), the classification error can be reduced by ensemble, as long as $\bar{A}$ is enhanced and $\bar{E}$ is reduced, with the increase of individual learners. For $\bar{A}$, we use McNemar's test [56] to evaluate the difference of individual learners. McNemar's test has been widely used in many ensemble based works [22], [24], [28], which is defined by

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}}, \tag{13}$$

where f_{12} denotes the number of samples correctly classified by learner 1 while incorrectly by learner 2. The difference between learner 1 and 2 is statistically significant if $|Z|$ is above 1.96. The evaluation matrix for the two data sets are displayed in Table V, VI, VII and Fig. 9. Note that absolute values have been adopted, and only the first 20 learners are reported. In Fig. 9(a)-(c), we can clearly find that the values around the diagonal are relatively lower, and this is in line with our expectation. Most blocks in Fig. 9(a) are close to deep color (red), which demonstrates that the diversity of the first 20 learners in Indian Pines data set is strong. However, we notice that the diversity in Pavia University and GRSS_DFC_2014 presents significant decline after around 10th hierarchy. Thus, further increase hierarchies would not improve the performance of the ensemble model. Quantitative results are shown in Table V, VI and VII. In Table V, only two pairs of learners score lower than 1.96. However, an obvious tendency is shown that with the increase of hierarchy number, $|Z|$ tends to decrease. The results indicate that the diversity of individual learners is enhanced by HGF, at least in the first 20 hierarchies. In Table VI, about 90% of all values are above 1.96 (after removing the diagonal), which indicates that the diversity in Pavia University data set is relatively high. For GRSS_DFC_2014 data set, descent rate of $|Z|$ is much faster. The diversity of individual learners increases little after the 14 hierarchy. This results can be also implied from Fig. 7(b). Overall, we can safely infer that hierarchical strategy is generally effective, but it is not necessary to set a very high hierarchy number.

Furthermore, Fig. 7 demonstrates that the classification accuracies keep growing with the increase of hierarchy. This results guarantee that $\bar{E}$ could be reduced after adding new individual learners. However, it is not to say more individual learners will lead to better accuracies. When the accuracies curves present steady shapes, further increasing individual learners may not reduce $\bar{E}$, instead, it will harm the diversity of the ensemble system.

To verify that the improvement achieved by HiFi-We is significant, we use box plot to describe the detailed statistics, as shown in Fig. 10. We compare HiFi-We with HGF-V, RGF-W and IIDF, because they present the best

performance among all the compared methods. Moreover, paired t-test results also show that the improvements on OA, AA and κ are statistically significant (at the level of 95%) in most cases.

IV. Conclusion

The initial motivation of this work is to develop a simply but effective HSI classification model which could combine spectral and spatial information in different scales. The most immediate idea is using ensemble learning. However, to ensure that the ensemble model really work, we must design diversity enhancing as well as valid ensemble strategies. In this paper, we propose a novel ensemble based method for HSI classification. The major contributions of our work include two folds: HGF and mSAD. HGF is an edge-preserving filtering operation which is able to generate diverse sample sets. Joint spectral-spatial information in different scales are extracted and utilized by HGF. Considering that the samples generated in each hierarchy may have different quality and confidence, we propose a measurement strategy called mSAD. Finally, the HGF and mSAD are unified via weighted voting.

To evaluate the performance of the proposed method, we conduct contrast experiments with some state-of-the-art methods on two popular data sets and a challenging data set. The results indicate that the proposed method works well, and the effectiveness is verified via statistical evaluation.

There are several future works associated with the proposed method. 1) It would be interesting to extend the proposed HGF and mSAD to other application, such as hyperspectral quality evaluation. 2) In this paper, we only study the influence of samples to the ensemble system. More attention could be paid to the design of classifiers.

V. Acknowlegement

The authors would like to thank Telops Inc. (Québec, Canada) for acquiring and providing the data used in this study, the IEEE GRSS Image Analysis and Data Fusion Technical Committee and Dr. Michal Shimoni (Signal and Image Centre, Royal Military Academy, Belgium) for organizing the 2014 Data Fusion Contest, the Centre de Recherche Public Gabriel Lippmann (CRPGL, Luxembourg) and Dr. Martin Schlerf (CRPGL) for their contribution of the Hyper-Cam LWIR sensor, and Dr. Michaela De Martino (University of Genoa, Italy) for her contribution to data preparation.

References

[1] X. Xu and Z. Shi, "Multi-objective based spectral unmixing for hyperspectral images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 124, pp. 54–69, 2017.

[2] F. J. van Ruitenbeek, P. Debba, F. D. van der Meer, T. Cudahy, M. van der Meijde, and M. Hale, "Mapping white micas and their absorption wavelengths using hyperspectral band ratios," *Remote Sensing of Environment*, vol. 102, no. 3, pp. 211–222, 2006.

[3] B. Pan, Z. Shi, Z. An, Z. Jiang, and Y. Ma, "A novel spectral-unmixing-based green algae area estimation method for GOCI data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016.

[4] G. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Transactions on Information Theory*, vol. 14, no. 1, pp. 55–63, 1968.

[5] Y. Gu, C. Wang, D. You, Y. Zhang, S. Wang, and Y. Zhang, "Representative multiple kernel learning for classification in hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 7, pp. 2852–2865, 2012.

[6] Y. Gu, Q. Wang, H. Wang, and D. You, "Multiple kernel learning via low-rank nonnegative matrix factorization for classification of hyperspectral imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2739–2751, 2015.

[7] J. Li, P. Reddy Marpu, A. Plaza, J. M. Bioucas-Dias, and J. Atli Benediktsson, "Generalized composite kernel framework for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 9, pp. 4816–4829, 2013.

[8] Y. Gu, T. Liu, X. Jia, and J. A. Benediktsson, "Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3235–3247, 2016.

[9] J. Chen, C. Wang, and R. Wang, "Using stacked generalization to combine SVMs in magnitude and shape feature spaces for classification of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2193–2205, 2009.

[10] X. Huang and L. Zhang, "Comparison of vector stacking, multi-SVMs fuzzy output, and multi-SVMs voting methods for multiscale VHR urban mapping," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 2, pp. 261–265, 2010.

[11] M. Pal and G. M. Foody, "Feature selection for classification of hyperspectral data by SVM," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 5, pp. 2297–2307, 2010.

[12] Z. H. Zhou, "Ensemble methods: Foundations and algorithms," vol. 8, no. 1, pp. 77–79, 2012.

[13] T. G. Dietterich, "Ensemble methods in machine learning," in *International Workshop on Multiple Classifier Systems*, 2000, pp. 1–15.

[14] J. Ham, Y. Chen, M. M. Crawford, and J. Ghosh, "Investigation of the random forest framework for classification of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 492–501, 2005.

[15] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognition Letters*, vol. 27, no. 4, pp. 294–300, 2006.

[16] C. W. Chan and D. Paelinckx, "Evaluation of random forest and adaboost tree-based ensemble classification and spectral band selection for ecotope mapping using airborne hyperspectral imagery," *Remote Sensing of Environment*, vol. 112, no. 6, pp. 2999–3011, 2008.

[17] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 67, no. 1, pp. 93–104, 2012.

[18] J. Xia, P. Du, X. He, and J. Chanussot, "Hyperspectral remote sensing image classification based on rotation forest," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 1, pp. 239–243, 2014.

[19] J. J. Rodrłguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A new classifier ensemble method." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1619–30, 2006.

[20] G. Mountrakis, J. Im, and C. Ogole, "Support vector machines in remote sensing: A review," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 247–259, 2011.

[21] M. Pal, "Ensemble of support vector machines for land cover classification," *International Journal of Remote Sensing*, vol. 29, no. 10, pp. 3043–3049, 2008.

[22] X. Huang and L. Zhang, "An svm ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 257–272, 2013.

[23] A. B. Santos, A. De, Albuquerque Araujo, and D. Menotti, "Combining multiple classification methods for hyperspectral data interpretation," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 6, no. 3, pp. 1450–1459, 2013.

[24] J. Xia, J. Chanussot, P. Du, and X. He, "Rotation-based support vector machine ensemble in classification of hyperspectral data with limited training samples," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 3, pp. 1519–1531, 2016.

[25] M. Chi, K. Qian, J. A. Benediktsson, and R. Feng, "Ensemble classification algorithm for hyperspectral remote sensing data," *IEEE Geoscience and Remote Sensing Letters*, vol. 6, no. 4, pp. 762–766, 2009.

[26] B. Waske, V. D. L. Sebastian, J. A. Benediktsson, A. Rabe, and P. Hostert, "Sensitivity of support vector machines to random feature selection in classification of hyperspectral data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 7, pp. 2880–2889, 2010.

[27] A. Samat, P. Du, S. Liu, J. Li, and L. Cheng, "E2LMs: Ensemble extreme learning machines for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 4, pp. 1060–1069, 2014.

[28] J. Xia, L. Bombrun, T. Adalı, Y. Berthoumieu, and C. Germain, "Spectral–spatial classification of hyperspectral images using ICA and

edge-preserving filter via an ensemble strategy," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 8, pp. 4971–4982, 2016.

[29] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 45–54, 2014.

[30] P. Ghamisi, M. Dalla Mura, and J. A. Benediktsson, "A survey on spectral–spatial classification techniques based on attribute profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 5, pp. 2335–2353, 2015.

[31] W. Li and Q. Du, "A survey on representation-based classification and detection in hyperspectral remote sensing imagery," *Pattern Recognition Letters*, 2015.

[32] Y. Zhong, Y. Wu, X. Xu, and L. Zhang, "An adaptive subpixel mapping method based on map model and class determination strategy for hyperspectral remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1411–1426, 2015.

[33] J. Zhao, Y. Zhong, Y. Wu, L. Zhang, and H. Shu, "Sub-pixel mapping based on conditional random fields for hyperspectral remote sensing imagery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 6, pp. 1049–1060, 2015.

[34] B. Pan, Z. Shi, N. Zhang, and S. Xie, "Hyperspectral image classification based on nonlinear spectral–spatial network," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 12, pp. 1782–1786, 2016.

[35] Y. Tarabalka, M. Fauvel, J. Chanussot, and J. A. Benediktsson, "SVM- and MRF-based method for accurate classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 4, pp. 736–740, 2010.

[36] P. Ghamisi, J. A. Benediktsson, and M. O. Ulfarsson, "Spectralc-spatial classification of hyperspectral images based on hidden Markov Random Fields," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2565–2574, 2014.

[37] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480–491, 2005.

[38] J. A. Benediktsson, "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *International Journal of Remote Sensing*, vol. 31, no. 22, pp. 5975–5991, 2010.

[39] 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 Geoscience and Remote Sensing Letters*, vol. 10, no. 2, pp. 293–297, 2013.

[40] L. Shen and S. Jia, "Three-dimensional gabor wavelets for pixel-based hyperspectral imagery classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 12, pp. 5039–5046, 2011.

[41] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 4, pp. 2276–2291, 2013.

[42] X. Guo, X. Huang, and L. Zhang, "Three-dimensional wavelet texture feature extraction and classification for multi/hyperspectral imagery," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 12, pp. 2183–2187, 2014.

[43] W. Li and Q. Du, "Gabor-filtering-based nearest regularized subspace for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 7, no. 4, pp. 1012–1022, 2014.

[44] K. He, J. Sun, and X. Tang, "Guided image filtering," in *European Conference on Computer Vision (ECCV)*. Springer, 2010, pp. 1–14.

[45] ——, "Guided image filtering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 6, pp. 1397–1409, 2013.

[46] Q. Zhang, X. Shen, L. Xu, and J. Jia, "Rolling guidance filter," in *European Conference on Computer Vision (ECCV)*. Springer, 2014, pp. 815–830.

[47] B. Ham, M. Cho, and J. Ponce, "Robust image filtering using joint static and dynamic guidance," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2015, pp. 4823–4831.

[48] Y. Li, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep joint image filtering," in *European Conference on Computer Vision (ECCV)*. Springer, 2016, pp. 154–169.

[49] B. Pan, Z. Shi, and X. Xu, "R-VCANet: A new deep-learning-based hyperspectral image classification method," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017.

[50] X. Kang, S. Li, and J. A. Benediktsson, "Spectral–spatial hyperspectral image classification with edge-preserving filtering," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 5, pp. 2666–2677, 2014.

[51] T. Hastie, R. J. Tibshirani, and J. H. Friedman, *The Elements of Statistical Learning*. Springer, 2001.

[52] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *IEEE International Conference on Computer Vision (ICCV)*. IEEE, 1998, pp. 839–846.

[53] J. Wright, A. Ganesh, S. Rao, Y. Peng, and Y. Ma, "Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization," in *Advances in Neural Information Processing Systems (NIPS)*, 2009, pp. 2080–2088.

[54] X. Kang, S. Li, L. Fang, and J. A. Benediktsson, "Intrinsic image decomposition for feature extraction of hyperspectral images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2241–2253, 2015.

[55] 2014 IEEE GRSS Data Fusion Contest. Online: http://www.grss-ieee.org/community/technical-committees/data-fusion/.

[56] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," *Photogrammetric Engineering and Remote Sensing*, vol. 70, no. 5, pp. 627–633, 2004.