Ensemble EMD-based Spectral-Spatial Feature Extraction for Hyperspectral Image Classification

Qianming Li, *Student Member, IEEE,* Bohong Zheng, Bing Tu, *Member, IEEE,* Jinping Wang, *Student Member, IEEE,* Chengle Zhou, *Student Member, IEEE*

***Abstract*—Hyperspectral images (HSIs) have fine spectral information and rich spatial information, of which the feature quality is one of the key factors that affect the classification performance. Therefore, how to extract essential features and eliminate redundant features from hyperspectral data is the main research focus of this paper. Here, we propose a spectral-spatial feature extraction method based on ensemble empirical mode decomposition (SFEEMD) for HSI classification, which contains several steps as follows: Firstly, the dimension reduction for HSI is performed by using principal component analysis method. Secondly, in order to decrease the sensitivity to noise and extract rough outline features, the adaptive total variation filtering (ATVF) is conducted on the selected principal components. Furthermore, by using the ensemble empirical mode decomposition (EEMD) to resolve each spectral band into sequence components, the features of HSIs can be better coalesced into the transform domain. Finally, the first K principal components of the input image and the outputs of the ATVF and EEMD are integrated into a stacking system to obtain the final feature image, which is then classified by a pixel-wise classifier. The experimental results of three authentic hyperspectral data sets show that the proposed algorithm obtains superior classification performance compared with other methods.**

***Index Terms*—Hyperspectral image classification, ensemble empirical mode decomposition, feature extraction, adaptive total variation.**

I. Introduction

HYPERSPECTRAL images (HSIs) can discover the particular distinctions hidden in the confined spectrum through the broadband spectrum coverage with hundreds of bands. In other words, the wide spectral coverage and high spectral resolution of HSIs tremendously enhances peoples cognitive competence of ground objects. Due to these advantages, hyperspectral imaging technique plays a vital role in geological exploration, urban expansion, agricultural and forestry monitoring, military, and other industries [1]–[5].

This research was supported in part by the National Natural Science Foundation of China (Grant Nos. 51478470, and 61977022), the Science Foundation for Distinguished Young Scholars of Hunan Province under Grant (Grant No. 2020JJ2017), the Hunan Province Postgraduate Research and Innovation Project (Grant No. CX2018B075), the Key Project of Central South University Postgraduate Independent Innovation Project (Grant No. 502221802). (Qianming Li and Bohong Zheng contributed equally to this work.) (Corresponding author: Bing Tu and Chengle Zhou.)

Q. Li and B. Zheng are with the School of Architecture and Art, Central South University, Changsha 410083, China (e-mail: lqm188@csu.edu.cn; zhengbohong@csu.edu.cn).

B. Tu, J. Wang and C. Zhou are with the School of Information Science and Technology, Hunan Institute of Science and Technology, Yueyang 414000, China (e-mail: tubing@hnist.edu.cn; jinping_wang@foxmail.com; chengle_zhou@foxmail.com).

Although HSI provides abundant spectral information, the colossal number of spectral bands will inevitably lead to a high calculation cost and the "Hughes" phenomenon [6]. Specifically, when the quantity of training samples is fixed, with the increase of data dimension, the classification accuracy obtained by a statical model-based classifier appears to increase first and decline later. Therefore, it is obviously undesirable to directly input the high-dimensional data into a classifier. Moreover, other problems such as spectral unmixing and noisy label problem in HSI also increase the difficulty for hyperspectral image classification (HIC) [7], [8]. Recently, many different types of advanced machine learning methods have been proposed to solve these problems, such as K-nearest neighbor [9], decision tree method [10], Bayesian theory [11], [12], support vector machine (SVM) [13], [14], and neural network [15].

Feature extraction [16] is one of the highlights in the HSI processing field. The objective of existing hyperspectral feature extraction methods is usually to map high-dimensional hyperspectral data to low-dimensional feature spaces, such as independent component analysis (ICA) [17] and principal component analysis (PCA) [18]. For example, Falco *et al.* have researched the influence of different ICA methods on the classification performances [19]. Furthermore, a large number of non-linear HSI dimension reduction methods have also been proposed for HIC [20]. For instance, kernel local fisher discriminant analysis (KLFDA) [21] and the method based on chaos theory [22]. The merit of these non-linear methods is that they can seek the essence of things from the observed phenomenon, and thus, finding the inherent law of data formation.

The feature extraction and classification methods mentioned above are mainly based on the spectral characteristics only involving a single pixel, without considering the relevance between spatially adjacent pixels. However, in recent years, researchers indicate that the spatial information of HSIs also plays a considerable part in HIC. Spectral-spatial classification based on image segmentation is the most prevailingly employed method, which primarily includes watershed segmentation [23], partition clustering [24], hierarchical segmentation, superpixel segmentation and so on [25], [26]. This kind of methods first segregate an HSI into multiple non-overlapping areas in the image according to the brightness, color or texture information. Then, the label of each area is assigned with the most frequent label (obtained using a spectral classifier) of this region. Markov Random Field (MRF) as a space majorization tool has been triumphantly applied to spectral-

spatial classification. Other methods for spectral-spatial HIC are based on account of probability optimization. For instance, Tarabalka *et al.* [27] presented an approach based on SVM and MRF. This approach first used the SVM to acquire the classified probability of apiece hyperspectral pixel, and then used the MRF to optimize the classified probability map to express the HSI in the model, and finally, maximized the probability of apiece pixel to acquire classification results. Li *et al.* [28], [29] learned posterior probability distribution by multivariate logistic regression, and then united the learned posterior probability distribution with the MRF model to further enhance the classification accuracy. In addition, Wang and Du *et al.* [30] proposed a batch-processing active learning framework mode, which preserves the source distribution as much as possible to inquire the most informative samples, thus confirming the most typical and indeterminate queries. Ghamisi *et al.* [31] proposed an approach based on SVM and hidden MRF for HIC. Based on the similar idea, Kang *et al.* [32] used stochastic rovers for probability majorization, which is demonstrated to be very useful when training samples are relatively restricted. In recent years, sparse representation also becomes a hot research topic in HIC field as a novel signal expression algorithm [33]–[37].

Besides the researches in spectral-spatial classification, the extraction of spectral-spatial features of HSIs has also been carried out. Benedkitsson *et al.* proposed a morphological filtering algorithm for HSI feature extraction [38]. Mura *et al.* proposed two methods to analyze hyperspectral data, namely the extended morphological character contour and extended contour with morphological character filter [39], [40]. Ghamisi *et al.* presented an automatic framework based on spectral-spatial classification of morphological character contour and supervised feature extraction [41]. Kang *et al.* presented an edge-preserving filtering algorithm for HSI feature extraction [42]. Prasad *et al.* used redundant discrete wavelet transform (RDWT) to extract the HSI features [43]. Furthermore, by using spatial correlation features in patch, Su *et al.* proposed a joint collaborative representation classification with correlation matrix (CRC-CM) for HSI, which can maintain the local inherent structure in the band [44]. Su *et al.* also proposed two novel multifeature learning algorithms that directly or indirectly update the dictionary [45]. To provide multifeature complementarity, four different types of features are adopted. In addition, Zhang *et al.* proposed a method for cloud detection of high-resolution satellite images using various features of ground objects (such as color, texture, and shape) [46]. This method not only improves the overall accuracy rate, but also reduces the false alarm rate. Wang *et al.* proposed an end-to-end adaptive spectral-spatial multiscale network to extract multiscale context information for HSI classification [47]. The above methods make full use of spectral-spatial features and multifeatures of HSI, and the classification accuracy is effectively improved.

A decision fusion system is used to blend the decisions of diverse classifiers so as to acquire the optimal classification consequence. In comparison with the wavelet transform, the non-linear signal processing algorithm called empirical mode decomposition (EMD) [48] has higher efficiency, which decomposes the signal according to the time-scale features. However, assuming that the value does not satisfy the physical standard deviation between the two signals, the original signal characteristics will disappear in the modulated signal [49]. To conquer this difficulty, Wu *et al.* proposed the ensemble empirical mode decomposition (EEMD) [50], [51], which can get more comprehensive performance. Moreover, Motin *et al.* proposed an algorithm based on combining empirical mode decomposition and principal component analysis (EEMD-PCA) as a new method to estimate heart rate and respiratory rate simultaneously from photoplethysmographic signal [52]. Su *et al.* proposed a novel bagging-based tangent space collaborative representation classification (TCRC) and boosting-based TCRC methods [53]. The main idea of bagging-based TCRC is to use bootup sample method to generate various TCRC classification results, thus improving the accuracy and diversity of a single classifier. Liu *et al.* proposed two integrated methods based on EEMD to retrieve wind direction from rain-contaminated X-band nautical radar sea surface images [54].

In the algorithm research, the total variation (TV) model algorithm has attracted the attention of many scholars [55]–[58]. Several TV algorithms began to be applied in digital image processing, which showed outstanding performance on image denoising [59]–[61]. Based on TV model, Liu *et al.* [62] proposed a model that used an edge detection filter to adaptively select parameters, and a shock filter combined with anisotropic diffusion to process noisy images, which can quickly solve the algorithm for image denoising adaptive total variation (ATV) model. Furthermore, it proposed a fast-iterative algorithm to solve the proposed adaptive model based on Bergman iteration regularization method. Because the proposed model can maintain a balance between noise smoothing and edge retention, it has been extensively used in sparse reconstruction [64], low-dose X-ray cone-beam computed tomography (CBCT) [65], and compressed perception (CP) [66].

Considering that adaptive total variational filtering (ATVF) can smoothly reduce noise, that is, effectively and quickly reduce noise sensitivity. Therefore, this method is selected to acquire rough outline features. However, in the selection of ATVF filtered HSI features, only the high-frequency part of the details is visually reflected, while the grayscale change information of the low-frequency part is ignored. To address this problem, a hyperspectral feature extraction named spectral-spatial feature-based ensemble empirical mode decomposition (SFEEMD) is proposed in this paper. The major steps of this paper are as follows: Firstly, the output feature image of ATVF is integrated into the transform domain by using the EEMD method, which could decompose each spectral band into sequence components, and use them to enhance the identification function of spectrum. Secondly, the first K principal components of the input image and the outputs of the ATVF and EEMD are integrated into a stacking system to obtain the final feature image, which provides a better pixel-level representation feature for HIC. Furthermore, this method performs experiments on some authentic hyperspectral data sets and compares the products with various HIC methods. Our proposed method exhibits excellent classification perfor-

mances on multiple quality indicators, including Kappa coefficient, average classication accuracy, and overall classication accuracy.

The major contributions of this paper are summarized as follows:

1) A simple SFEEMD method is proposed to extract discriminative spectral-spatial features for HSI classification, which is designed by integrating the first K principal components of the input image and the outputs of the ATVF and EEMD into a stacking system.

2) SFEEMD method uses ATVF and EEMD model to perform noise reduction and feature reconstruction on HSI, which can quickly remove noise while preserving details such as the edge contour and texture of the image. The proposed SFEEMD method is more robust and generalized by comparing with other methods.

The rest of this paper is listed as follows: Section II introduces the EEMD method, ATVF model, and fast ATV algorithm. Section III describes our proposed SFEEMD method for spectral-spatial HIC. The information of data sets used in this paper and the analysis and discussion of experimental results are provided in Section III-A. Finally, Section V concludes this paper briefly.

II. Related Work

A. Ensemble Empirical Mode Decomposition

EMD is an adaptive signal processing technique with wildly applications and largely contributions. Later, the EEMD method emerged, which reduces the influence of modal mixing in the EMD process, and making this method show a better level. For EEMD, the percentage p of the additional homogeneous white noise and the quantity of semaphores in the integration are always determined in advance. The EEMD consists of diverse intrinsic mode functions (IMFs) that must meet under requirements:

- The quantity of zero points and partial extremum points in the formula should be equivalent during the entire time scope, otherwise the largest deviation is 1.
- The mean values of the lower and upper envelopes are decided by the partial minimum and maximum respectively. At any time, the value has to be 0.

Assume $X_{c,a}^{b}(m,n)$ indicates the result of the bth iteration of the ath IMF in cth band. Therefore, the screening process is summed up, and the iterative process starts from one-dimensional to realize IMFs.

1) Find all partial minimum and maximum points from every input band.

2) Insert the minimum and maximum value respectively, to acquire the lower $E_{min}(m,n)$ and upper envelope $E_{max}(m,n)$.

3) Compute the average of the lower and upper envelopes $Z_A^{(B)}(m,n)$, the formula is as below:

$$Z_A^{(B)}(m,n) = \frac{(E_{max}(m,n) + E_{min}(m,n))}{2} \tag{1}$$

4) From the input semaphore, to reduce the average value of the envelope $D_A^{(B)}(m,n) = X_{c,a}^{(b)}(m,n) - Z_a^{(b)}(m,n)$

5) Duplicate operations 1)-4) until the envelope semaphore meets the present IMF, that is, $SCD < \tau$ and converges with the IMF.

$$SCD = \frac{max(abs(Z_a^{(b)}(m,n)))}{max(abs(R_a(m,n)))} < \tau \tag{2}$$

6) Duplicate operations 1)-5) to produce a remainder $R_a(m,n) = R_a(m,n) - IMF_{c,a}(m,n)$, if the remainder does not involve an extremum value, the EMD procedure will terminate.

In summary, the original spectral band image $X_c(m,n)$ is acquired. The reconstruction of IMF performance and remainder is as below:

$$X_c(m,n) = \sum_{a=1}^{A} IMF_{c,a}(m,n) + R_a(m,n) \tag{3}$$

B. Adaptive Total Variational Filtering Model

Rudin *et al.* [57] proposed a total variational (TV) minimization model for image noise reduction, which is expressed as follows:

$$TV\ \mathrm{F}_{\mu}(u) = \arg\min_{u} TV(u) + \frac{\mu}{2}\int_{\Omega} |f(x) - u(x)|^2 dx \tag{4}$$

where $TV(u)$ is the TV of the noise data; μ is a regularization parameter, which is in connection with the filtering extent of acquired solution and the noise statistics; Ω is the image space and $x \in \Omega$; $f(x)$ is the noise image, $u(x)$ is the original image. Let $TV(u) = |\nabla_x u| + |\nabla_y u|$, the image method noise of the TV mini-mization is shown as follows [64]:

$$u(x) - TV\ \mathrm{F}_{\mu}(u)(x) = -\frac{1}{\mu} curv(TV\ \mathrm{F}_{\mu}(u)(x)) \tag{5}$$

where $curv(TV\ \mathrm{F}_{\mu}(u))$ indicates the curvature of all level sets to $TV\ \mathrm{F}_{\mu}(u)$. This paper sets the value of μ with a small regularization parameter, which effectively remove the noise from the image f. Since the details of the local structure of diverse regions and smooth regions, a small μ will cause the loss of noisy image texture information and details, and even blur the edges.

Therefore, an adaptive parameter χ_f is introduced to eliminate image noise and retain image textures and details. Namely, the TV model of image denoising that controls propagation. But, the parameter χ_f is quite sensitive to noise. To lessen the sensitivity of χ_f, this paper applies shock filtering and nonlinear anisotropic diffusion filtering to pre-process the image, thus, the effect of smoothing noise and enhancing image edges is simultaneously obtained. Then, the adaptive parameter χ_f is computed by the image edge detection operator. Alvarez *et al.* [67] united the diffusion and shock terms, and proposed the anisotropic diffusion with shock filter (ADSF) method:

$$\frac{\partial f}{\partial t} = -sign(G_\sigma \otimes f_{NN}) sign(G_\sigma \otimes f_{NN}) |\nabla f| + c f_{TT} \tag{6}$$

where G_σ represents a gaussian function with standard deviation, $\otimes$ denotes the convolution operator, c reveals a positive

constant, f_{TT} and f_{NN} are the two order directional derivatives of the tangent direction $(\bar{T})$ and the gradient direction $(\bar{N})$ of the corresponding evolution curve respectively.

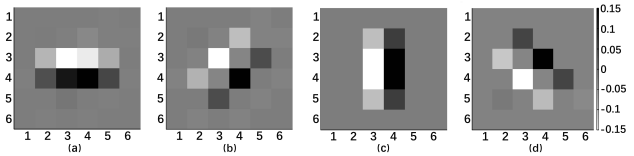


Fig. 1. The edge detection operator d_β: (**a**) β = 0°; (**b**) β = 45°; (**c**) β = 90°; (**d**) β = 135°.

Let $\hat{f}$ be the pre-processed image. It can be seen in Fig. 1, four 6×6 sized edge detection operators are presented in this paper [68]. When $\beta = 0^o$, we have $d_0 = \frac{1}{12} \bullet \begin{bmatrix} & O_1 & \\ O_2 & M & O_2 \\ & O_1 & \end{bmatrix}$, $M = \begin{bmatrix} -1 & -2 & -2 & -1 \\ 1 & 2 & 2 & 1 \end{bmatrix}$, $O_1 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$, $O_2 = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$. The relevant edge detection operator d_β can be acquired from d_0 and rotation angle $\beta \in \Phi(\Phi = \{0^o, 45^o, 90^o, 135^o\})$.

Based on the definition of the relevant edge detection operator d_β, an adaptive parameter χ_f is computed via the formula below:

$$\chi_f = \frac{1}{1 + \sqrt{\sum_{\beta \in \Phi} (d_\beta \otimes \hat{f})^2}} \in (0, 1] \tag{7}$$

Adding the χ_f into TV model of image denoising, a new approach which is named as ATV model is proposed. It is shown as follows:

$$ATV\ \mathrm{F}_{\chi,\mu}(u) = \arg\min_u \chi_f TV(u) + \frac{\mu}{2} \int_\Omega |f(x) - u(x)|^2 dx \tag{8}$$

The ATV image denoising model is as indicated below:

$$u(x) - ATV\ \mathrm{F}_{\chi,\mu}(u)(x) = -\frac{\chi_f}{\mu} curv(ATV\ \mathrm{F}_{\chi-\mu}(u)(x)) \tag{9}$$

C. Fast ATV Algorithm

Solving the ATV model using the proposed fast settling algorithm [62]. Set $u^1 = f$, $b_x^0 = b_y^0 = 0$ (z = 1,2,3,...), a fast iterative algorithm (FIA) is as below:

$$cut(h, \frac{1}{z}) := h - \frac{h}{|h|} \bullet max(|h| - \frac{1}{z}, 0) \tag{10}$$

$$b_x^z := cut(H \nabla_x u^z + b_x^{z-1}, \frac{1}{\lambda}) \tag{11}$$

$$b_y^z := cut(H \nabla_y u^z + b_y^{z-1}, \frac{1}{\lambda}) \tag{12}$$

$$u^{z+1} := f - \frac{\lambda}{u}(\nabla_x^T b_x^z + \nabla_y^T b_y^z) \tag{13}$$

As Jia attested in [63]: For z = 0,1,..., let b_x^z, b_y^z, u^{z+1} be produced by the iterative algorithm (11) - (13). On account of $\widehat{\mu} = \mu/H$, $H \in (0, 1]$, if $0 < \lambda/\widehat{\mu} < 8$, then $\lim_{z \to \infty} u^z = u^*$. Therefore, we can acquire $0 < H \bullet \lambda/\mu < 1/8$ from $0 < \lambda/\widehat{\mu} < 8$. To make random $H \in (0, 1]$ satisfy with $0 < H \bullet \lambda/\mu < 1/8$, we have to ensure that $0 < \lambda/\mu < 1/8$, so FIA can converge to the optimum solution of ATV.

III. The Proposed Method

A. Motivation

In the actual image acquisition, transmission and storage processes, HSI contains noise in most cases due to external factors such as equipment and environment. The presence of noise will affect the extraction of effective information in HSI. The traditional TV denoising algorithm needs to know the noise variance and step effect, while the ATVF algorithm realizes image denoising by calculating the angle matrix of the local direction of the image and iteratively solving the optimization minimization algorithm. In this paper, we propose to use the ATVF to denoise HSI after dimensionality reduction. In the ATVF, the image is adaptively divided into flat regions and edge regions based on the local gray average gradient value of the HSI, which can quickly remove noise while retaining the detailed information such as the edge contour and texture of the image. At the same time, the EEMD algorithm is an intuitive and a priori new adaptive signal time-frequency research method, especially suitable for the analysis and processing of nonlinear and non-stationary signals. It can further decompose and reconstruct the HSI features after denoising by ATVF. Specifically, EEMD can adaptively decompose the original features into a finite number of eigenmode functions according to the local feature scale, representing spectral information from the highest frequency to the lowest frequency. In this way, it can effectively solve the situation that the features of the same category contain pixels from other categories, thereby improving the accuracy of HSI classification.

B. The Proposed SFEEMD Method for Spectral-Spatial HIC

Fig. 2 displays the HIC framework using the SFEEMD classification method. HSI normally has hundreds of spectrum bands, which is different from three-band or single-band images. Due to the PCA can lower the spectrum dimensionality of HSI on the premise of retaining the mean square sense information, it is applied to dimensionality reduction to obtain the first K principal components. The processing is as follows:

$$\mathbf{O}^K = PCA(\mathbf{I}) \tag{14}$$

where $\mathbf{O}^K$ refer to the image with the first K principal components, after bands reduction over the input image $\mathbf{I}$.

Then, the ATVF is conducted on $\mathbf{O}^K$. The expression is defined as follows:

$$\mathbf{N}^K = ATVF_{\mu,\lambda,In}(\mathbf{O}^K) \tag{15}$$

where $\mathbf{N}^K$ represents the feature image obtained from the $\mathbf{O}^K$. μ, λ, In are the filtering parameters respectively.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2020.3018710, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

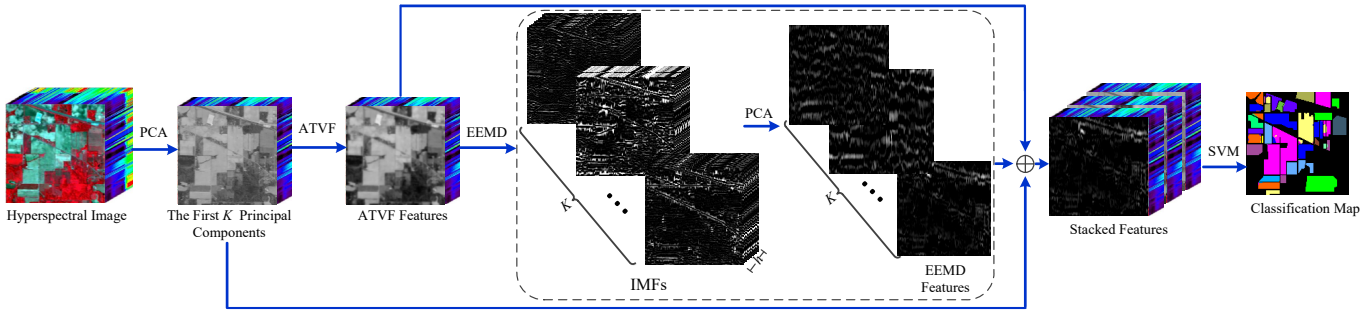


Fig. 2. Schematic diagram of the SFEEMD HIC.

Since the ATVF features image ignores the details of low-frequency grayscale change information in HSI characteristic selection, we adopt the EEMD method to extract the unique spatial structure of the intrinsic modal component image $\mathbf{N}^K$ to enhance the image details. Each spectral band of HSI can be decomposed into a sequence of IMFs by the EEMD. On this basis, the PCA can acquire the first principal component of IMFs, and the spatial-spectral features can be integrated into one component. After that, in accordance with the correlation between IMFs and residuals R, the processed HSI can be reestablished as follows:

$$\mathbf{X}^K = \sum_{m=1}^{g} IMF_{K,m} + R_m \tag{16}$$

where $\mathbf{X}^K$ is the outline feature extracted by the EEMD method. g is the quantity of the IMFs in each band.

At last, a common pixel-by-pixel SVM classifier, which has robust performance to high-dimensional data, is applied to classify all pixel points of the novel characteristic data in HSI. The EEMD feature images $\mathbf{X}^K$, the ATVF feature image $\mathbf{N}^K$, and K principal components $\mathbf{O}^K$ are merged into a stack system as $\hat{\mathbf{X}}^{3K}$. Algorithm 1 summarizes the primary steps of our proposed SFEEMD algorithm.

Algorithm 1. HIC method via SFEEMD

Inputs: HSI, feature dimension K, ATVF parameters: μ , λ, EEMD parameters p;

Step 1: According to (14), the PCA is applied to lower the dimensionality of the HSI;

Step 2: Based on (15), The K-dimensional feature data goes through the ATVF process to extract the resulting features $\mathbf{N}^K$;

Step 3: Based on (16), EEMD is performed on $\mathbf{N}^K$ image to extract the resulting features $\mathbf{X}^K$;

Step 4: Based on a stacked system, the SVM is applied to process pixel-by-pixel classification for $\hat{\mathbf{X}}^{3K}$;

Output: Classification result.

IV. Experiments

A. Data Sets

1) Indian Pines Data Set: The image has 220 bands with the spatial resolution of 20m, and the image size is 145×145. This data set is acquired in June by Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) facility and displayed the "Indian Pines Test Field" in northwest Indiana. The 200 bands are left and applied in the test, after deleting 20 water absorption bands. Due to season, climate, and other factors, some crops are in the early growth stage. Through the classification analysis, the image is divided into 16 different objectives. Fig. 3 is a pseudo-color composite picture with corresponding reference date and classification code.

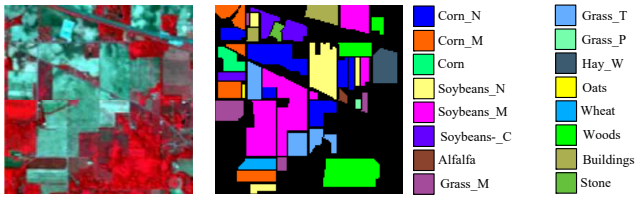


Fig. 3. **(a)** Pseudo-color composite image of the Indian Pines. **(b)** and **(c)** The Indian Pines image with corresponding reference date.

2) Salinas Data Set: The image has 224 bands with the spatial resolution of 3.7m, and the image size is 512×217. Images are gathered by AVIRIS facility over Salinas Valley, California. The absorption bands of water are removed before classification, including bands 108-112, 154-167 and 224. Fig. 4 is a pseudo-color composite picture with corresponding reference date.

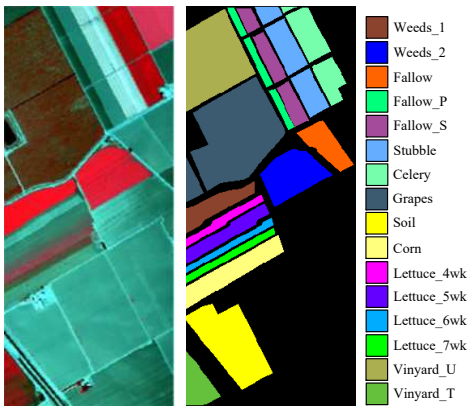


Fig. 4. **(a)** Pseudo-color composite image of the Salinas. **(b)** and **(c)** The Salinas image with corresponding reference date.

3) University of Pavia Data Set: The image has 115 bands with the spatial resolution of 1.3m, and the image size is 610×340. Images acquisition the University of Pavia, Italy, are gathered by Reflective Optics System Imaging Spectrometer

(ROSIS) facility. The 103 bands are left and applied in the test, after deleting 12 water noise bands. Fig. 5 is a pseudo-color composite picture with corresponding reference date.

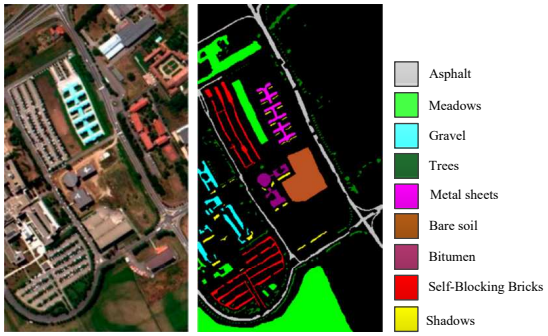


Fig. 5. **(a)** Pseudo-color composite image of the University of Pavia. **(b)** and **(c)** The University of Pavia image with corresponding reference date.

B. Parameter Settings

This part dissects the effect of parameters on the classification performance of the proposed approach. K refers to the feature dimension. μ and λ are the filtering parameters in the ATVF algorithm. p is the parameter in the EEMD. Three commonly criteria are used for evaluating the HIC accuracy: overall classification accuracy (OA), average classification accuracy (AA), and Kappa coefficient (Kappa). Parameters μ and λ are fixed when quantitatively analyzing the relationships between the feature dimension K and the used criteria (OA, AA, Kappa). It can be seen from Fig. 6 that under lower feature dimension K, the SVM and SFEEMD approaches have significantly lower accuracies. As the feature dimension K increases, the three criteria also increase. When K reaches 21, the classification accuracies turn to steady state. This is because low feature dimension may cause a large amount of meaningful information missing. Therefore, we select the feature dimension K of 21 to capture superior property. Compared with two methods, it is obvious that when the feature dimension $K > 9$, the classification accuracy of the proposed SFEEMD method is significantly superior to that of the SVM algorithm.

Then, the influence of distinct parameters λ, μ, p, and K on classification accuracy is analyzed. In Figs. 7 - 9, we randomly choose 10% and 2% samples for training. When analyzing one parameter, the others are set as constants, so as to observe the variation trend of OA. Fig. 7(a) shows, when μ = 0.032, p = 0.31, K = 21, the relationships between OA and λ. It can be observed that when λ = 0.004, OA could reach 99.2846%. Fig. 7(b) reveals, when λ = 0.004, p = 0.31, K = 21, the influence of μ on the classification accuracy. Fig. 7(c) demonstrates, when λ = 0.004, μ = 0.032, K = 21, the effect of p on the classification accuracy. Therefore, the parameters are set to λ = 0.004, μ = 0.032, p = 0.31, K = 21. With the same parameter settings, the classification accuracies of the remaining data sets are shown in Figs. 8 and 9.

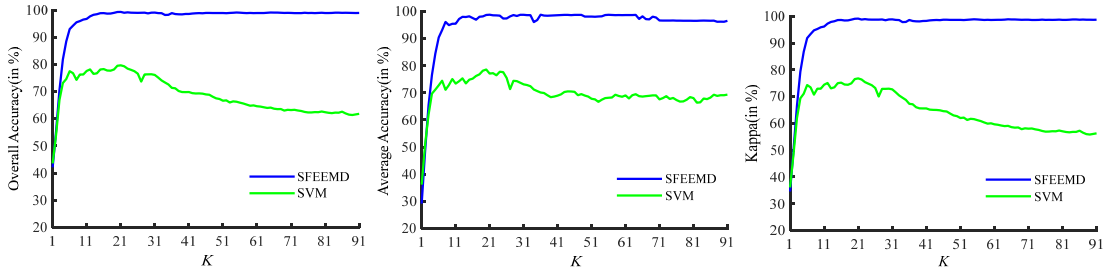
C. Experiments With Different Data Sets

To prove the effectiveness of the proposed SFEEMD method, the proposed method is compared with ATVF method, EEMD method and some combination methods. The experiment is conducted on the Indian Pines data set, and the experimental results are shown in Table I. In the experiment, 10% or 1% samples were randomly selected as training samples, and the remaining samples were used for testing. It can be found that the classification accuracy of the ATVF method, EEMD method, and their combined methods are lower than those of the proposed SFEEMD method. Moreover, the classification accuracy of the method combining the PCA method with the ATVF method or the EEMD method is also not higher than the proposed SFEEMD method. The reason is that the proposed SFEEMD method combines the features extracted by the PCA method, ATVF method and EEMD method, thereby further effectively improving the classification accuracy. The original features, edge features and deep spatial features of HSIs are extracted by PCA method, ATVF method and EEMD method, respectively.

The presented SFEEMD approach is compared with the commonly used spectral-spatial classification approaches. Common methods include SVM [69], sparse representation classification (SRC) method [70], joint sparse representation classification (JSRC) method [33], variable splitting and segmented lagrangian (LOR) [71], extended morphological profiles (EMP) [38], edge preserving filtering (EPF) [42], logistic regression and multi-level logistic (LMLL) [72]. The SVM method used the Gaussian kernel and five-layer cross validation. The feature extraction based on PCA and ICA were achieved by using the ENVI remote sensing software which widely used internationally. SVM was used to classify after 20 principal components and independent components were extracted. Based on the EMP method, this paper uses the first three principal components of HSI and morphology operators to perform four opening closing and reconstruction operations and establishes multi-scale morphological traits. Based on the JSRC algorithm, the experiment uses the software and the default parameters to establish features. The experiments used MATLAB codes and default parameters, and compared with the rest of other methods.

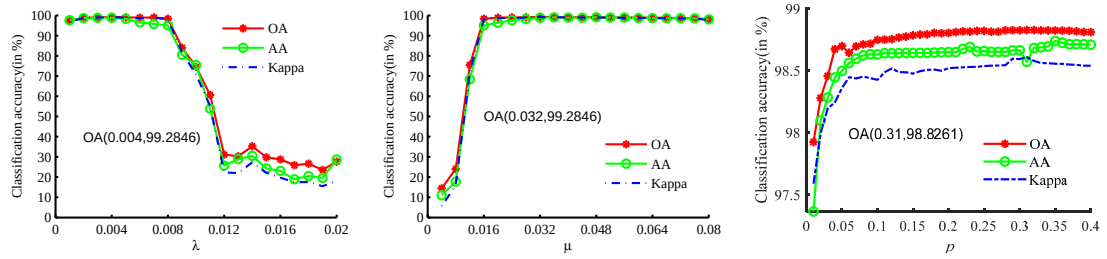
The first experiment is executed on the Indian Pines data set. Table II displays the quantity of the test samples and training samples in the examination. It is worth mentioning that to compare the classification accuracy of diverse algorithms more impersonally, all the examinations must be repeated 50 times. Then, the standard deviation of distinct approaches classification accuracy can be calculated. Next, the classification accuracy and map acquired by distinct approaches are demonstrated in Table II and Fig. 10, respectively. It can well perceived that the classification accuracies of SVM and SRC algorithms that merely consider spectral information is only 83.28% and 68.46%, while the classification algorithms (EMP, EPF, LMLL, SFEEMD) based on spectralspatial is often much higher than 93%. When the spatial information has similar spectral characteristics, and applied it in the spectral-spatial classification. Besides, in the aspects of OA, AA, and Kappa,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2020.3018710, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing



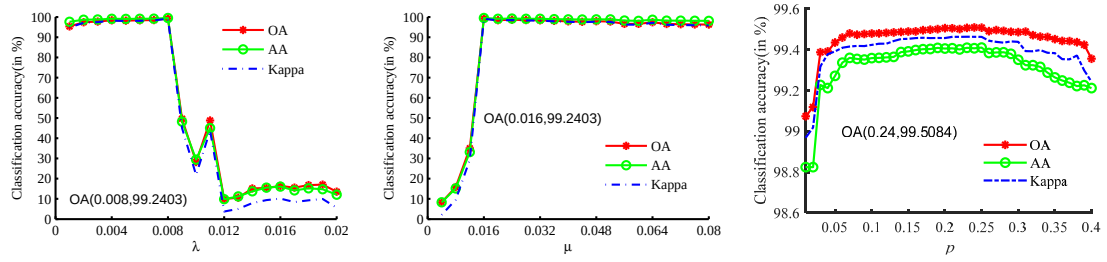
(a) (b) (c)

Fig. 6. Effect analysis of the parameter K to SFEEMD and SVM approaches on Indian Pines image: (**a**) $\lambda = 0.004, \mu = 0.032, p = 0.3$; (**b**) $\lambda = 0.004, \mu = 0.032, p = 0.3$; (**c**) $\lambda = 0.004, \mu = 0.032, p = 0.3$.



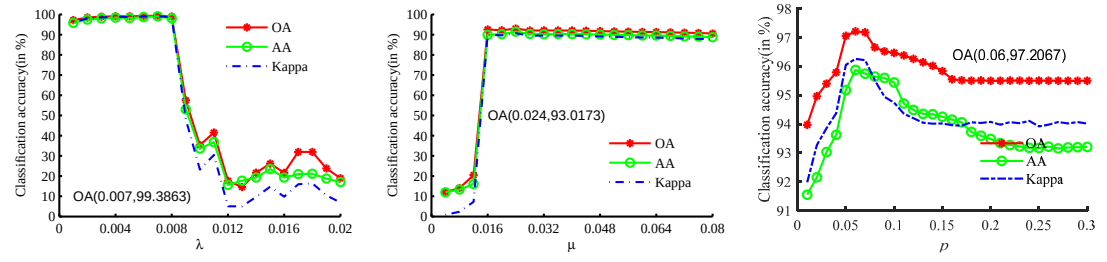
(a) (b) (c)

Fig. 7. Effect analysis of the parameters λ, μ, p on Indian Pines image: (**a**) μ = 0.032, p = 0.31, K = 21; (**b**) λ = 0.004, p = 0.31, K = 21; (**c**) λ = 0.004, μ = 0.032, K = 21.



(a) (b) (c)

Fig. 8. Effect analysis of the parameters λ, μ, p on Salinas image: (**a**) μ = 0.032, p = 0.24, K = 21; (**b**) λ = 0.004, p = 0.24, K = 21; (**c**) λ = 0.004, μ = 0.032, K = 21.



(a) (b) (c)

Fig. 9. Effect analysis of the parameters λ, μ, p on University of Pavia image: (**a**) μ = 0.032, p = 33, K = 21; (**b**) λ = 0.004, p = 33, K = 21; (**c**) λ = 0.004, μ = 0.032, K = 21.

TABLE I
CLASSIFICATION ACCURACIES OF THE ATVF, EEMD, PCA+ATVF, PCA+EEMD, ATVF+EEMD, AND SFEEMD ALGORITHMS FOR THE INDIAN PINES DATA SET WITH 10% AND 1% TRAINING SAMPLES.

The number of training samples account for 10% of the reference data						
Index	ATVF	EEMD	PCA+ATVF	PCA+EEMD	ATVF+EEMD	SFEEMD
OA	98.33	98.28	97.70	98.10	98.40	**98.84**
AA	98.27	98.05	98.44	98.15	98.68	**99.01**
Kappa	98.14	98.08	97.38	98.09	98.38	**98.68**
The number of training samples account for 1% of the reference data						
Index	ATVF	EEMD	PCA+ATVF	PCA+EEMD	ATVF+EEMD	SFEEMD
OA	85.73	84.50	82.96	83.45	85.95	**86.69**
AA	85.88	82.71	88.40	87.53	87.95	**91.66**
Kappa	84.68	82.26	80.36	80.96	85.92	**86.16**

TABLE II
CLASSIFICATION ACCURACIES OF THE SVM [69], SRC [70], JSRC [33], LOR [71], EMP [38], EPF [42], LMLL [72], AND SFEEMD ALGORITHMS FOR THE INDIAN PINES DATA SET WITH 10% AND 1% TRAINING SAMPLES.

The number of training samples account for 10% of the reference data.										
Class	**Training**	**Test**	**SVM**	**SRC**	**JSRC**	**LOR**	**EMP**	**EPF**	**LMLL**	**SFEEMD**
Alfalfa	8	38	76.94	53.26	90.88	87.56	95.53	73.08	94.26	**100.0**
Corn_N	137	1291	79.49	53.83	81.33	79.60	87.38	93.34	97.16	**98.01**
Corn_M	83	747	80.55	51.42	61.04	71.94	92.68	96.24	90.41	**97.86**
Corn	24	213	67.20	38.71	68.54	64.96	83.99	88.29	**98.42**	97.55
Grass_M	48	435	89.27	81.92	92.87	93.04	92.74	99.01	98.39	**99.53**
Grass_T	73	657	89.86	91.36	59.51	97.54	98.31	92.12	99.22	**100.0**
Grass_P	8	20	88.62	85.20	43.00	92.27	91.00	95.00	93.18	**100.0**
Hay_W	48	430	97.24	92.15	83.26	99.85	99.86	97.00	98.87	**100.0**
Oats	8	12	48.59	65.83	41.67	97.73	97.50	96.40	97.51	**100.0**
Soybean_N	97	875	77.37	65.61	67.54	79.65	86.51	98.98	84.55	**99.18**
Soybean_M	239	2216	81.12	70.56	83.26	87.05	96.25	94.13	97.86	**98.79**
Soybean_C	59	534	77.99	43.39	55.81	81.15	87.40	96.73	**98.55**	96.47
Wheat	21	184	92.68	90.41	91.85	99.46	98.21	**100.0**	99.85	**100.0**
Woods	122	1143	92.56	89.43	95.98	95.01	99.50	99.29	98.37	**100.0**
Buildings	39	347	72.96	35.63	78.39	68.19	96.08	92.40	88.66	**99.14**
Stone	47	46	**98.59**	89.42	90.36	73.17	93.01	92.50	98.25	97.62
OA			83.28	68.46	82.65	85.28	93.54	95.52	95.97	**98.84**
AA			81.94	68.45	71.96	85.51	93.5	94.76	96.12	**99.01**
Kappa			80.86	98.32	80.22	83.14	92.63	94.88	95.38	**98.68**
The number of training samples account for 1% of the reference data.										
Class	**Training**	**Test**	**SVM**	**SRC**	**JSRC**	**LOR**	**EMP**	**EPF**	**LMLL**	**SFEEMD**
Alfalfa	2	44	38.30	56.05	87.12	54.96	85.68	65.15	61.78	**100.0**
Corn_N	14	1414	52.42	41.32	60.09	75.16	52.18	**95.79**	86.90	80.34
Corn_M	8	822	66.99	30.16	59.53	64.64	55.34	64.75	**81.21**	67.64
Corn	2	235	21.00	17.17	66.52	36.67	21.40	41.43	43.02	**96.59**
Grass_M	5	478	61.26	58.21	74.97	77.69	64.79	85.80	80.49	**100.0**
Grass_T	7	723	81.35	78.96	73.08	94.90	93.54	67.94	99.30	**100.0**
Grass_P	2	26	48.57	83.92	**100.0**	68.53	87.69	42.42	93.01	70.27
Hay_W	5	473	95.65	80.86	90.35	98.40	90.51	97.40	98.78	**100.0**
Oats	3	17	17.86	70.94	64.71	93.58	90.59	89.47	96.58	**100.0**
Soybean_N	10	962	48.62	49.33	68.75	66.06	67.61	75.88	84.77	**85.77**
Soybean_M	25	2430	47.09	62.79	86.78	76.37	78.67	62.42	**90.88**	83.37
Soybean_C	6	534	49.04	22.97	62.75	72.23	41.64	30.72	90.97	**91.96**
Wheat	2	203	82.79	75.07	80.95	99.55	97.00	**100.0**	99.78	**100.0**
Woods	13	1252	87.25	80.99	95.23	93.12	83.65	90.25	**95.61**	90.65
Buildings	4	382	36.65	18.81	69.11	54.66	72.85	97.10	65.18	**100.0**
Stone	2	91	**100.0**	85.21	94.14	51.05	80.99	81.00	61.84	**100.0**
OA			58.11	55.37	76.23	76.93	70.38	72.75	85.82	**86.69**
AA			58.43	57.05	77.13	73.6	72.76	75.41	82.42	**91.66**
Kappa			50.78	48.93	71.96	72.7	66.19	68.32	83.06	**84.68**

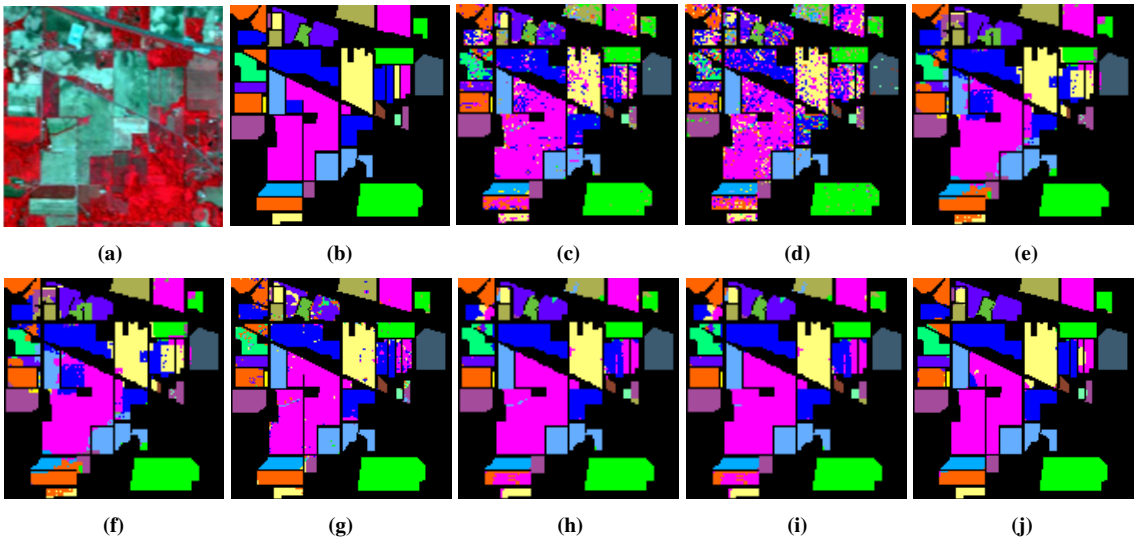


Fig. 10. Classification results (Indian Pines image) obtained by (**a**) the false-color composite of the Indian Pines image; (**b**) the original image; (**c**) the SVM method (83.28%); (**d**) the SRC method (68.46%); (**e**) the JSRC method (82.65%); (**f**) the LOR method (85.28%); (**g**) the EMP method (93.54%); (**h**) the EPF method (95.52%); (**i**) the LMLL method (95.97%); (**j**) the SFEEMD method (98.84%).

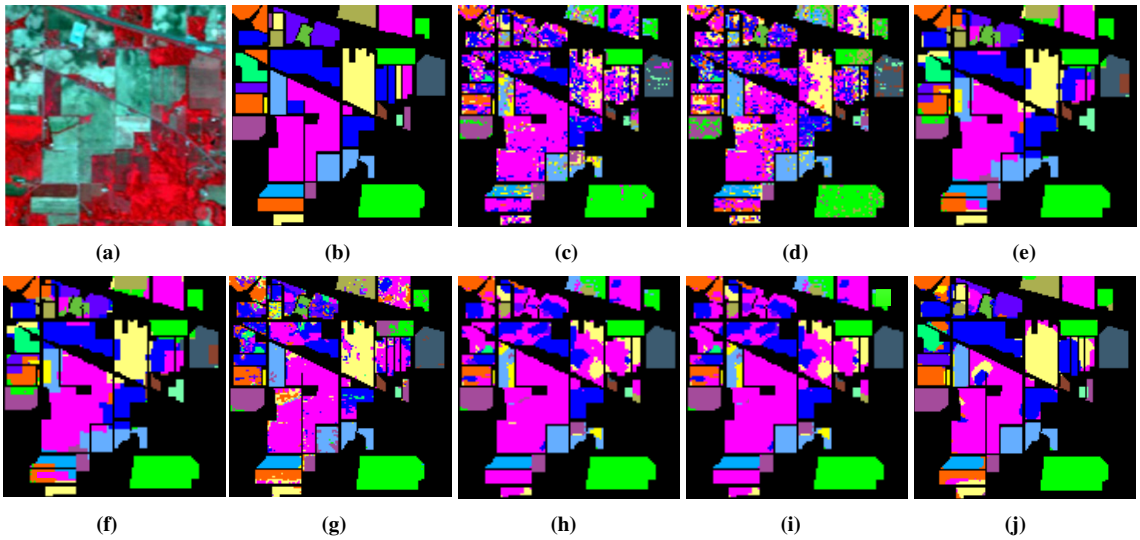


Fig. 11. Classification results (Indian Pines image) obtained by (**a**) the false-color composite of the Indian Pines image; (**b**) the original image; (**c**) the SVM method(58.11%); (**d**) the SRC method(55.37%); (**e**) the JSRC method(76.23%); (**f**) the LOR method(76.93%); (**g**) the EMP method(70.38%); (**h**) the EPF method(72.75%); (**i**) the LMLL method(85.82%); (**j**) the SFEEMD method(86.69%).

SFEEMD also outperforms than EMP, etc. approaches. Table II also illustrates the influences of diverse approaches on the OA, AA, Kappa of diverse sorts. Our presented SFEEMD method is preceded than SVM method in the classification accuracy of certain spectrum ground features. For instance, the accuracy of Building classification increased from 72.96% to 99.14%, and the accuracy of Corn classification increased from 67.20% to 97.55%. The optimal products in the Table are highlighted in bold.

To make the result credible in Indian Pines data set, the training samples by stochastically chosen have reduced to 1%, and the remaining 99% data is used for testing. Classification maps are obtained by diverse classification in Fig. 11. Obviously, compared with other spatial-spectral approaches, the classification results with SFEEMD obtains the highest accuracy. Moreover, the bottom table of Table II reveals the effects of diverse approaches on OA, AA, and Kappa of diverse sorts in this scenario. It can be observed that although the scale of training samples is tiny, it can still obtain a high classification accuracy. For some classes are compared with SVM, the accuracy of Alfalfa classification increased from 38.30% to 100.0%, the accuracy of Corn classification increased from 21.00% to 96.59%. In addition, the SFEEMD could still reach to 86.69% classification accuracy even a small sample training. Meanwhile, the proposed approach also exhibited desirable classification results when compared to other spectralspatial approaches.

The second experiment is performed on the Salinas data set, randomly chosen training data is occupied 2%, and the remaining 98% data is used for testing. Fig. 12 and Table III exhibit the classification accuracies of diverse algorithms, and compare with AA, OA, and Kappa. As the calculation

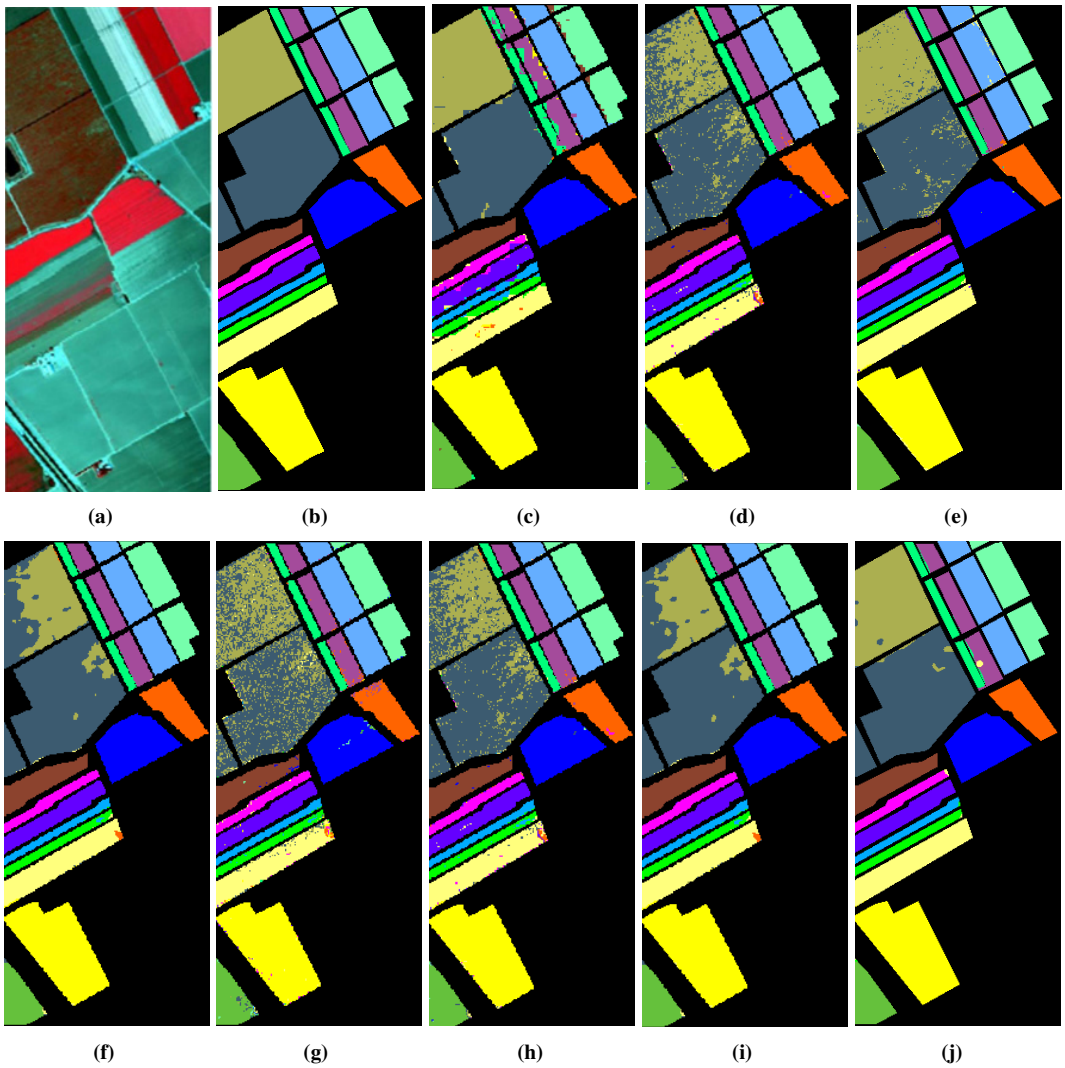


Fig. 12. Classification results (Salinas image) obtained by (**a**) the false-color composite of the Salinas image; (**b**) the original image; (**c**) the SVM method(90.68%); (**d**) the SRC method(87.16%); (**e**) the JSRC method(96.13%); (**f**) the LOR method(91.29%); (**g**) the EMP method(97.65%); (**h**) the EPF method(95.35%); (**i**) the LMLL method(93.20%); (**j**) the SFEEMD method(99.51%).

results in Table III, we can observe that the proposed approach exhibited desirable OA when compared to other approaches under the default parameter settings. For instance, Our presented SFEEMD algorithm has certain advantages over SVM, the accuracy of Vinyard_U classification increased from 77.89% to 98.53%, and the accuracy of Fallow classification increased from 94.73% to 100.0%. The SFEEMD retains edge data to make sufficient use of detailed information, and the accuracy test value is superior to other methods. Reducing the training samples to 0.2%, the products are showed in bottom of Table III. Classification products (the Salines data set) are acquired by diverse classification in Fig. 13. In the Table III, the Kappa of the SFEEMD is also improved significantly and optimal OA, AA, and Kappa could be obtained. Compared with the SVM method, it is not difficult to found that, when training with small samples, the recognition accuracy of using SVM method decreased dramatically, especially in class 15 (from 56.65% to 90.45%). As a comparison, we can still maintain relatively high recognition accuracy in all of the classes. Furthermore, after repeating with 100 times, standard deviation of the classification accuracy of SFEEMD is pretty low and stable.

To further verify the results of diverse training samples of the SFEEMD, we conducted another experiment on the Pavia University data set. In the experiment, 40, 80, 120, 160, 200, and 240 pixels for each classes are stochastically selected to establish a balance training set. Fig. 14 indicates that the OA for the presented SFEEMD with diverse quantity of training samples. And the accuracies of the SFEEMD are summarized in the Table IV. The SFEEMD can improve the misclassification phenomenon from Fig. 14 and increase the robustness from Table IV.

The calculation interval of the proposed method on two actual data sets is given. A 4 cores computer with a 2.9 GHz CPU and a 8-GB memory is used to logging data, and the quantity of training samples is set as the number listed in the Table II - III. We can observe from Table V, the computing cost of the feature extraction course (ATVF and EEMD) is irrelevant to the quantity of training samples and is just related to the size of image. The running time comparison

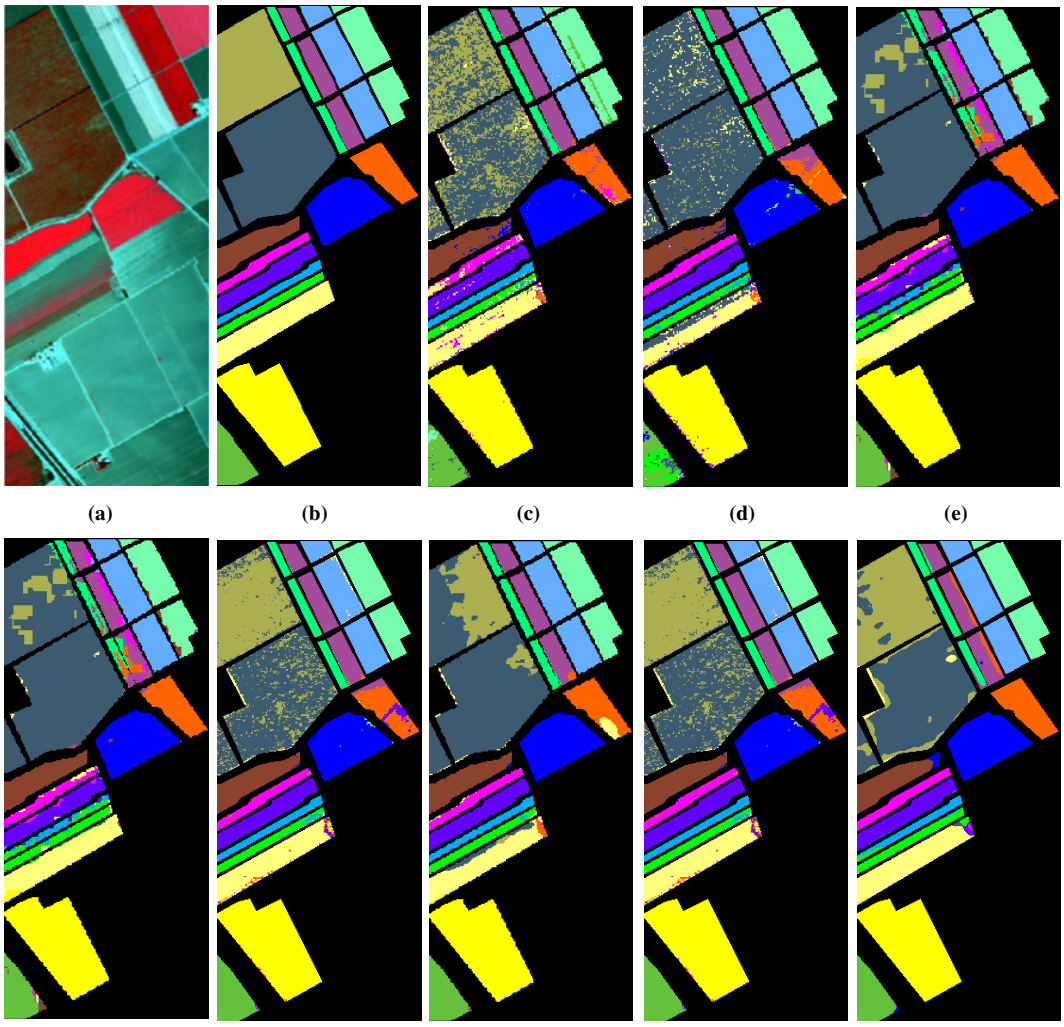


Fig. 13. Classification results (Salinas image) obtained by (**a**) the false-color composite of the Salinas image; (**b**) the original image; (**c**) the SVM method(83.01%); (**d**) the SRC method(80.32%); (**e**) the JSRC method(88.22%); (**f**) the LOR method(88.45%); (**g**) the EMP method(92.18%); (**h**) the EPF method(85.98%); (**i**) the LMLL method(91.97%); (**j**) the SFEEMD method(94.51%).

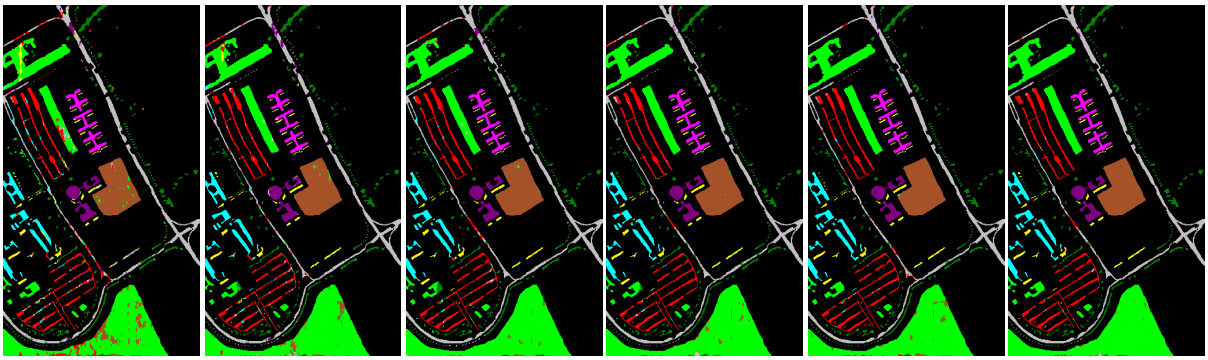


Fig. 14. Classification products (University of Pavia image) under diverse quantity of training samples: (**a**) the training samples is 40, OA=91.21%; (**b**) the training samples is 80, OA=94.27%; (**c**) the training samples is 120, OA=96.15%; (**d**) the training samples is 160, OA=96.83%; (**e**) the training samples is 200, OA=97.48%; (**f**) the training samples is 240, OA=97.83%.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JSTARS.2020.3018710, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing

TABLE III
CLASSIFICATION ACCURACIES OF THE SVM [69], SRC [70], JSRC [33], LOR [71], EMP [38], EPF [42], LMLL [72], AND SFEEMD METHODS FOR THE SALINAS DATA SET WITH 2% AND 0.2% TRAINING SAMPLES.

The number of training samples account for 2% of the reference data.										
Class	**Training**	**Test**	**SVM**	**SRC**	**JSRC**	**LOR**	**EMP**	**EPF**	**LMLL**	**SFEEMD**
Weeds_1	40	1969	99.80	98.61	97.21	97.86	99.72	98.57	90.10	**100.0**
Weeds_2	75	3651	99.45	98.74	99.76	**100.0**	99.69	99.89	**100.0**	**100.0**
Fallow	40	1936	94.73	95.56	99.31	98.55	98.89	94.76	99.79	**100.0**
Fallow_P	28	1366	96.87	99.19	86.58	98.81	99.08	97.50	98.90	**99.71**
Fallow_S	54	2624	98.87	94.28	85.67	97.71	96.89	**100.0**	99.03	99.62
Stubble	79	3880	**100.0**	99.41	97.81	99.29	99.41	99.95	99.78	**100.0**
Celery	72	3507	99.83	99.14	95.72	99.46	99.45	**100.0**	99.67	99.94
Graps	225	11046	79.31	72.40	99.04	81.88	96.12	86.34	88.58	**99.35**
Soil	124	6079	99.44	97.95	**100.0**	99.94	99.59	99.59	99.97	99.44
Corn	21	1047	89.52	90.28	96.87	94.86	96.27	95.45	96.99	**99.35**
Lettuce_4wk	21	1047	92.42	94.57	91.91	91.33	97.28	98.44	94.26	**100.0**
Lettuce_5wk	39	1888	97.40	99.81	89.18	99.81	99.80	99.32	**100.0**	99.89
Lettuce_6wk	18	898	95.55	97.16	67.97	97.52	98.83	95.60	97.58	**100.0**
Lettuce_7wk	21	1049	94.81	92.76	70.86	94.80	93.87	**99.29**	98.36	99.04
Vinyard_U	145	7123	77.89	62.08	98.03	71.41	94.61	93.01	71.29	**98.53**
Vinyard_T	36	1771	99.01	94.83	99.55	97.19	99.44	**100.0**	98.76	**100.0**
OA			90.68	87.16	96.13	91.29	97.65	95.35	93.20	**99.51**
AA			94.66	92.92	92.39	95.03	98.06	97.47	96.38	**99.68**
Kappa			89.60	85.71	95.69	90.30	97.38	94.81	92.42	**99.45**
The number of training samples account for 0.2% of the reference data.										
Class	**Training**	**Test**	**SVM**	**SRC**	**JSRC**	**LOR**	**EMP**	**EPF**	**LMLL**	**SFEEMD**
Weeds_1	4	2005	99.25	96.68	98.45	98.48	94.53	95.11	99.30	**99.85**
Weeds_2	8	3718	99.22	97.69	99.93	99.29	98.69	**99.92**	99.83	99.76
Fallow	4	1972	75.69	71.11	85.87	92.86	89.73	99.33	94.25	**100.0**
Fallow_P	3	1391	97.15	98.72	68.27	97.65	98.36	97.34	98.01	**99.24**
Fallow_S	5	2673	98.49	93.01	83.39	97.19	93.46	**99.43**	98.82	64.43
Stubble	8	3951	98.78	99.60	99.43	99.36	97.79	99.35	99.90	**100.0**
Celery	7	3572	92.83	96.46	95.12	99.38	99.47	98.70	99.65	**100.0**
Graps	22	11249	68.42	89.71	89.85	82.34	87.33	81.63	91.83	**97.44**
Soil	12	6191	99.00	96.35	**100.0**	98.89	98.50	98.41	99.46	99.41
Corn	7	3271	74.90	73.82	88.83	88.20	88.88	93.06	92.09	**95.82**
Lettuce_4wk	2	1066	92.63	90.48	75.2	73.18	93.91	93.48	70.87	**99.53**
Lettuce_5wk	4	1923	80.48	**99.43**	86.18	97.72	98.62	64.64	99.13	96.47
Lettuce_6wk	2	914	79.88	95.80	67.03	**99.02**	98.68	95.97	98.02	82.35
Lettuce_7wk	2	1068	**99.28**	89.59	81.55	86.30	90.91	99.01	89.79	94.81
Vinyard_U	14	7254	56.65	13.63	67.02	61.32	82.15	56.56	66.67	**90.45**
Vinyard_T	4	1803	97.18	56.31	95.95	91.32	89.82	84.56	97.01	**100.0**
OA			83.01	80.32	88.22	88.45	92.18	85.98	91.97	**94.51**
AA			88.19	84.9	86.48	87.11	93.80	92.31	93.48	**94.97**
Kappa			81.07	77.9	86.85	91.41	91.29	84.47	91.03	**93.89**

TABLE IV
CLASSIFICATION ACCURACIES OF DIVERSE QUANTITY OF TRAINING SAMPLES FOR THE UNIVERSITY OF PAVIA DATA SET. NUMBERS IN THE PARENTHESES DENOTE THE STANDARD DEVIATION OF THE ACCURACIES ACQUIRED IN REPETITIVE TRIALS.

Class	40	80	120	160	200	240
Asphalt	90.88(2.87)	94.44(1.57)	96.13(0.91)	96.93(0.69)	97.40(0.42)	97.79(0.47)
Meadows	98.16(0.62)	99.06(0.21)	99.46(0.18)	99.63(0.12)	99.68(0.10)	99.80(0.04)
Gravel	78.33(3.62)	86.57(5.43)	92.48(1.13)	94.88(1.31)	95.38(1.47)	96.74(0.98)
Trees	87.24(4.83)	92.14(2.12)	91.41(2.21)	91.51(3.19)	93.03(1.54)	93.98(1.33)
Metal sheets	98.34(0.90)	98.83(0.81)	99.27(0.64)	99.36(0.53)	99.30(0.59)	99.54(0.37)
Bare soil	87.26(5.70)	89.55(4.67)	94.58(3.37)	96.02(1.76)	96.71(1.43)	97.19(1.02)
Bitumen	76.06(4.84)	82.50(3.65)	88.66(3.57)	93.66(2.78)	93.30(2.43)	95.31(1.38)
Self-Blocking Bricks	79.97(3.26)	83.20(3.83)	88.49(0.86)	90.61(1.07)	91.57(1.43)	93.53(1.31)
Shadows	75.62(9.68)	86.73(5.60)	91.09(3.38)	93.48(1.78)	93.95(1.37)	94.14(1.66)
OA	91.21(0.96)	94.27(0.48)	96.15(0.27)	96.83(0.22)	97.48(0.23)	97.83(0.27)
AA	86.53(1.20)	91.25(0.79)	93.86(0.61)	95.12(0.42)	95.98(0.46)	96.45(0.47)
Kappa	88.45(1.22)	92.41(0.61)	94.87(0.37)	95.77(0.28)	96.61(0.31)	97.08(0.37)

TABLE V
THE CALCULATION INTERVAL OF THE SFEEMD ON TWO ACTUAL DATA SETS, WHEN THE QUANTITY OF TRAINING SAMPLES IS SET ACCORDING TO AS THOSE EXHIBITED IN TABLE II - III.

Data sets	Indian Pines		Salinas	
	10%	1%	2%	0.20%
ATVF	0.24	0.25	0.36	0.36
EEMD	39.06	39.48	227.18	231.8
SVM	38.32	2.15	27.4	2.05

TABLE VI
EXECUTION TIMES (IN SECONDS) FOR ALL METHODS ON INDIAN PINES DATA SETS.

Methods	Indian Pines		Methods	Indian Pines	
	10%	1%		10%	1%
SVM	119.14	3.68	EMP	28.75	2.09
SRC	7.30	3.16	EPF	116.29	4.16
JSRC	22.94	13.13	LMLL	30.97	7.17
LOR	30.97	7.17	SFEEMD	264.11	20.61

for different methods is shown in Table VI. It can be found that the SFEEMD takes the longest time than others. This is because that EEMD method requires many iterations to extract the intrinsic mode component. To solve the above-mentioned issue, a faster EEMD implementation using GPU programming will be studied.

V. CONCLUSION

This paper proposes the SFEEMD algorithm, an HSI spectral-spatial classification approach on account of feature extraction, which can enhance the classification accuracy of ground features with a low amount of calculation. Compared with the spectral-spatial classification approaches mentioned in this paper, the SFEEMD algorithm has obvious superiority, due to the capability of reducing the sensitivity to noise smoothing and denoising. By further extracting the feature filtered image into the transform domain, the edge details are retained for feature extraction. The results illustrate that an advanced approach availably enhances the classification accuracy for HSI and has superior noise suppression performance. Compared with the traditional ATVF with PCA algorithm, the SFEEMD algorithm can exploit deep in the spatial information of HSI. The experimental products on several authentic HSIs datasets indicate that our presented SFEEMD approach produces high-precision classification outcomes, particularly, with wide uniform areas in the images.

REFERENCES

[1] X. Lu, W. Zhang, X. Li. "A hybrid sparsity and distance-based discrimination detector for hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1704-1717, Mar. 2018.

[2] B. Chai, P. Li. "Annual Urban Expansion Extraction and Spatio-Temporal Analysis Using Landsat Time Series Data: A Case Study of Tianjin, China," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 8, pp. 2644-2656, Aug. 2018.

[3] C. Zhang, J.M. Kovacs. "The application of small unmanned aerial systems for precision agriculture: a review," *Precis. Agric.*, vol. 13, no. 6, pp. 693-712, Jul. 2012.

[4] J. Feng, X. Feng, J. Chen, X. Cao, T. Yu. "Generative Adversarial Networks Based on Collaborative Learning and Attention Mechanism for Hyperspectral Image Classification," *Remote Sens.*, vol. 12, no. 7, pp. 1149, Apr. 2020.

[5] B. Tu, X. Yang, C. Zhou, D. He, A. Palaza. "Hyperspectral Anomaly Detection using Dual Window Density," *IEEE Trans. Geosci. Remote Sens.*, 2020, to be published (doi: 10.1109/TGRS.2020.2988385).

[6] G. Hughes. "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inf. Theory*, vol. 14, no. 1, pp. 55-63, Jan. 1968.

[7] I. Dopido. "A Quantitative and Comparative Assessment of Unmixing-Based Feature Extraction Techniques for Hyperspectral Image Classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 2, pp. 421-435, Apr. 2012.

[8] B. Tu, C. Zhou, D. He, S. Huang, A. Palaza. "Hyperspectral Classification with Noisy Label Detection via Superpixel-to-Pixel Weighting Distance," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 4116-4131, Jun. 2020.

[9] B. Tu, X. Zhang, X. Kang, J. Wang, J.A. Benediktsson. "Spatial Density Peak Clustering for Hyperspectral Image Classification with Noisy Labels," *IEEE Trans. Geosci. Remote Sens.* vol. 57, no. 7, pp. 5085-5097, Jul. 2019.

[10] G. Zhou, R. Zhang, D. Zhang. "Manifold Learning Co-Location Decision Tree for Remotely Sensed Imagery Classification," *Remote Sens.*, vol. 8, no. 10, pp. 855, Oct. 2016.

[11] C. Chen, P. Ho. "Statistical pattern recognition in remote sensing," *Pattern Recognit.*, vol. 41, no. 9, pp. 2731-2741, Sep. 2008.

[12] N. Bali, A. Mohammad-Djafari. "Bayesian approach with hidden markov modeling and mean field approximation for hyperspectral data analysis," *IEEE Trans. Image Process.*, vol. 17, no. 2, pp. 217-225, Jan. 2008.

[13] G. Cavallaro, R. Morris, M. Richerzhagen, J. Benediktsson, A. Plaza. "On Understanding Big Data Impacts in Remotely Sensed Image Classification Using Support Vector Machine Methods," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 10, pp. 4634-4646, Aug. 2015.

[14] H. Yu, L. Gao, J. Li, S. Li, B. Zhang, J. Benediktsson. "Spectral-Spatial Hyperspectral Image Classification Using Subspace-Based Support Vector Machines and Adaptive Markov Random Fields," *Remote Sens.*, vol. 8, no. 5, pp. 355, Apr. 2016.

[15] R. Hang, Q. Liu, D. Hong and P. Ghamisi, "Cascaded Recurrent Neural Networks for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384-5394, Aug. 2019.

[16] D. Hong, X. Wu, P. Ghamisi, J. Chanussot, N. Yokoya and X. Zhu, "Invariant Attribute Profiles: A Spatial-Frequency Joint Feature Extractor for Hyperspectral Image Classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3791-3808, Jun. 2020.

[17] A. Villa, A. Benediktsson, J. Chanussot, C. Jutten. "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865-4876, Dec. 2011.

[18] J. Xia, J. Chanussot, P. Du, X. He. "(Semi-) Supervised Probabilistic Principal Component Analysis for Hyperspectral Remote Sensing Image Classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2224-2236, Sep. 2014.

[19] N. Falco, A. Benediktsson, L. Bruzzone. "A study on the effectiveness of different independent component analysis algorithms for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2183-2199, Jul. 2014.

[20] D. Hong, N. Yokoya, J. Chanussot, J. Xu, and X. Zhu, "Learning to Propagate Labels on Graphs: An Iterative Multitask Regression Framework for Semi-supervised Hyperspectral Dimensionality Reduction," *ISPRS J. Photogramm. Remote Sens.*, vol. 158, pp. 35-49, Oct. 2019.

[21] W. Li, S. Prasad, J.E. Fowler, L.M. Bruce. "Locality-preserving discriminant analysis in kernel-induced feature spaces for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 5, pp. 894-898, May 2011.

[22] A. Zhang, Y. Xie. "Chaos theory-based data-mining technique for image endmember extraction: Laypunov index and correlation dimension (l and d)," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 4, pp. 1935-1947, Jun. 2014.

[23] Y. Tarabalka, J. Chanussot, J.A. Benediktsson. "Segmentation and classification of hyperspectral images using watershed transformation," *Pattern Recognit.*, vol. 43, no. 7, pp. 2367-2379, Jul. 2010.

[24] Y. Tarabalka, J.A. Benediktsson, J. Chanussot, "Spectral-spatial classification of hyperspectral imagery based on partitional clustering techniques," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 8, pp. 2973-2987, Apr. 2009.

[25] Z. Zhang, E. Pasolli, M. Crawford, M. Melba, J.C. Tilton. "An Active Learning Framework for Hyperspectral Image Classification Using Hierarchical Segmentation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 640-654, Dec. 2015.

[26] Y. Liu, G. Cao, Q. Sun, M. Siegel. "Hyperspectral classification via deep networks and superpixel segmentation," *Int. J. Remote Sens.*, vol. 36, pp. 3459-3482, Jul. 2015.

[27] Y. Tarabalka, M. Fauvel, J. Chanussot, J.A. Benediktsson. "Svm and mrf-based method for accurate classification of hyperspectral images," *IEEE Geosci. Remote Sens. Lett.*, vol. 7, no. 4, pp. 736-740, May 2010.

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

[29] J. Li, X. Huang, P. Gamba, J.M. Bioucas-Dias, L. Zhang, J.A. Benediktsson, A. Plaza. "Multiple feature learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 3, pp. 1592-1606, Aug. 2015.

[30] Z. Wang, B. Du, L. Zhang. "A batch-mode active learning framework by querying discriminative and representative samples for hyperspectral image classification," *Neurocomput.*, vol. 179, pp. 88-100, Feb. 2016.

[31] P. Ghamisi, J.A. Benediktsson, M.O. Ulfarsson. "Spectral-spatial classification of hyperspectral images based on hidden markov random fields," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2565-2574, Jan. 2014.

[32] X. Kang, S. Li, L. Fang, M. Li, J.A. Benediktsson. "Extended random walker-based classification of hyperspectral images," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 1, pp. 144-153, May 2015.

[33] J. Liu, Z. Wu, Z. Wei, L. Xiao. "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, Apr. 2013.

[34] Z. Xue, P. Du, H. Su, S. Zhou. "Discriminative Sparse Representation for Hyperspectral Image Classification: A Semi-Supervised Perspective," *Remote Sens.*, vol. 9, no. 4, pp. 386, Apr. 2017.

[35] J.A. Benediktsson, W. Fu, S. Li, L. Fang, X. Kang. "Hyperspectral Image Classification Via Shape-Adaptive Joint Sparse Representation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, Feb. 2016.

[36] L. Fang, S. Li, X. Kang, J.A. Benediktsson. "Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 12, pp. 7738-7749, Dec. 2014.

[37] Z. Qin, W. Yang, R. Yang, X. Zhao, T. Yang. "Dictionary-Based, Clustered Sparse Representation for Hyperspectral Image Classification," *J. Spectro.*, pp. 1-6, Jan. 2015.

[38] J.A. Benediktsson, J.A. Palmason, 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.

[39] M.D. Mura, A. Villa, A.J. Benediktsson, J. Chanussot, L. Bruzzone. "Classification of hyperspectral images by using extended morphological attribute profiles and independent component analysis," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 3, pp. 542-546, May 2011.

[40] M.D. Mura, A.J. Benediktsson, B. Waske, L. Bruzzone. "Extended profiles with morphological attribute filters for the analysis of hyperspectral data," *Int. J. Remote Sens.*, vol. 31, no. 22, pp. 5975-5991, Dec. 2010.

[41] P. Ghamisi, A.J. Benediktsson, G. Cavallaro, A. Plaza. "Automatic framework for spectral–spatial classification based on supervised feature extraction and morphological attribute profiles," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2147-2160, Jun. 2014.

[42] X. Kang, S. Li, A.J. Benediktsson. "Spectral-spatial hyperspectral image classification with edge-preserving filtering," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 5, pp. 2666-2677, May 2014.

[43] S. Prasad, W. Li, J.E. Fowler, L.M. Bruce. "Information fusion in the redundant-wavelet-transform domain for noise-robust hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 9, pp. 3474-3486, Sep. 2012.

[44] H. Su, B. Zhao, Q. Du, P, Du. "Kernel collaborative representation with local correlation features for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 2, pp. 1230-1241, Feb. 2019.

[45] H. Su, B. Zhao, Q. Du, P, Du, Z. Xue. "Multi-feature dictionary learning for collaborative representation classification of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2467-2484, Apr. 2018.

[46] D. Wang, B. Du, L. Zhang, Y. Xu. "Adaptive spectral-spatial multiscale contextual feature extraction for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, pp. 1-17, 2020, (doi: 10.1109/TGRS.2020.2999957.)

[47] D. Zhang, Q. Zhou, X. Shen, Y. Li. "Cloud detection in high-resolution remote sensing images using multi-features of ground objects," *J. Geovis. Spat. Anal.*, vol. 3, no. 14, Aug. 2019.

[48] N. Huang, Z. Shen, S. Long, M. Wu, H. Shi, Q. Zheng, N. Yen, C. Tung, H. Liu. "The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. R. Soc. London A: Math. Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903-995, 1998.

[49] G. Rilling, P. Flandrin, P. Goncalves. "On empirical mode decomposition and its algorithms," *IEEE-EURASIP Workshop Nonlinear Signal Image Process*, vol. 3, pp. 8-11, Jan. 2003.

[50] Z. Wu, N. Huang, X. Chen. "The multi-dimensional ensemble empirical mode decomposition method," *Adv. Adapt Data Anal.*, vol. 1, no. 3, pp. 339-372, 2009.

[51] Z. Wu, N. Huang. "Ensemble empirical mode decomposition: a noise-assisted data analysis method," *Adv. Adapt Data Anal.*, vol. 1, no. 1, pp. 1-41, 2009.

[52] M. A. Motin, C. K. Karmakar, M. Palaniswami. "Ensemble empirical mode decomposition with principal component analysis: a novel approach for extracting respiratory rate and heart rate from photoplethysmographic signal," *IEEE J. Biomed. Health Inf.*, vol. 22, no. 3, pp. 766-774, May 2018.

[53] H. Su , Y. Yu, Q. Du, P. Du. "Ensemble learning for hyperspectral image classification using tangent collaborative representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 6, pp. 3778-3790, Jun. 2020.

[54] X. Liu, W. Huang, E. W. Gill. "Wind direction estimation from rain-contaminated marine radar data using the ensemble empirical mode Decomposition Metho," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 3, pp. 1833-1841, Mar. 2017.

[55] R. Jia, H. Zhao, W. Zhao. "Convergence analysis of the bregman method for the variational model of image denoising," *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 367-379, Nov. 2009.

[56] R. Jia, H. Zhao. "A fast algorithm for the total variation model of image denoising," *Appl. Comput. Harmon. Anal.*, vol. 33, no. 2, pp. 231-241, May 2010.

[57] L.I. Rudin, S. Osher, E. Fatemi. "Nonlinear total variation based noise removal algorithms," *Physica D.*, vol. 60, pp. 259-268, Nov. 1992.

[58] Y. Wang, X. Chen, Z. Han, S. He. "Hyperspectral Image Super-Resolution via Nonlocal Low-Rank Tensor Approximation and Total Variation Regularization," *Remote Sens.*, vol. 9, no. 12, pp. 1286, Dec. 2017.

[59] P. Duan, X. Kang, S. Li, P. Ghamisi. "Noise-Robust Hyperspectral Image Classification via Multi-Scale Total Variation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. pp, no. 99, pp. 1-15, May 2019.

[60] Z. Qin, D. Goldfarb, S. Ma. "An alternating direction method for total variation denoising," *Optim. Method. Softw.*, vol. 30, no. 3, pp. 594-615, Sep. 2015.

[61] H. Zhang, Q. Peng. "Adaptive image denoising model based on total variation," *Optoelectron. Eng.*, vol. 3, pp. 50-53, Mar. 2006.

[62] W. Liu, C. Wu, T. Xu. "Adaptive total variation model for image denoising with fast solving algorithm," *Appl. Res. Comput.*, vol. 12, pp. 109, Dec. 2011.

[63] R.-Q. Jia, H. Zhao. "A fast algorithm for the total variation model of image denoising," *Adv. Comput. Math.*, vol. 33, no. 2, pp. 231-241, May 2010.

[64] A. Buades, B. Coll, J.M. Morel. "A non-local algorithm for image denoising," *in Proc. IEEE Comput. Vision Pattern Recogn. CVPR.*, vol. 2, pp. 60-65, Jul. 2005.

[65] H. Huang, T. Hsiao. "Accelerating an ordered-subset low-dose x-ray cone beam computed tomography image reconstruction with a power factor and total variation minimization," *PLoS One*, vol. 11, no. 4, pp. e0153421, 2016.

[66] G. Chen, J. Zhang, D. Li. "Fractional-order total variation combined with sparsifying transforms for compressive sensing sparse image reconstruction," *J. Visual Commun. Image Represent.*, vol. 38, pp. 407-422, Jul. 2016.

[67] L. Alvarez, L. Mazorra. "Signal and image restoration using shock filters and anisotropic diffusion," *SIAM J. Numer. Anal.*, vol. 31, no. 2, pp. 590-605, 1994.

[68] M.S. Almeida, L.B. Almeida. "Blind and semi-blind deblurring of natural images," *IEEE Trans. Image Process*, vol. 19, no. 1, pp. 36-52, Jan. 2010.

[69] F. Melgani, L. Bruzzone. "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778-1790, Sep. 2004.

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

[71] J. Bioucas-Dias, M. Figueiredo. "Logistic regression via variable splitting and augmented lagrangian tools," *Inst. Superior Técnico, TULisbon, Tech. Rep*, 2009.

[72] J. Li, J.M. Bioucas-Dias, A. Plaza. "Hyperspectral image segmentation using a new bayesian approach with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3947-3960, Oct. 2011.