

## ORIGINAL PAPER

# Spectral-spatial feature extraction and supervised classification by MF-KELM classifier on hyperspectral imagery

WENTING SHANG,<sup>1</sup>  ZEBIN WU,<sup>1,2,3</sup> YANG XU,<sup>1</sup> YAN ZHANG<sup>3</sup> AND ZHIHUI WEI<sup>1</sup>

*The kernel extreme learning machine (KELM) is more robust and has a faster learning speed when compared with the traditional neural networks, and thus it is increasingly gaining attention in hyperspectral image (HSI) classification. Although the Gaussian radial basis function kernel widely used in KELM has achieved promising classification performance in supervised HSI classification, it does not consider the underlying data structure of HSIs. In this paper, we propose a novel spectral-spatial KELM method (termed as MF-KELM) by incorporating the mean filtering kernel into the KELM model, which can properly compute the mean value of the spatial neighboring pixels in the kernel space. Considering that in the situation of limited training samples the classification result is very noisy, the spatial bilateral filtering information on spectral band-subsets is introduced to improve the accuracy. Experiment results show that our method outperforms other kernel functions based on KELM in terms of classification accuracy and visual comparison.*

**Keywords:** Kernel extreme learning machine (KELM), Mean filtering (MF) kernel, Spatial bilateral filtering, Spectral band-subsets, Hyperspectral image (HSI), Supervised classification

Received 20 April 2019; Revised 22 August 2019

## 1. INTRODUCTION

Hyperspectral imaging sensors are able to capture affluent spectral information across continuous narrow bands. Where the detailed spectral information makes it possible to accurately discriminate materials of interest. Moreover, [1–4] indicate the fine spatial resolution of the sensors enables the analysis of small spatial structures in the image. Therefore, the rich spatial and spectral characteristics bring wide applications of hyperspectral image (HSI), one of the most popular research directions is the supervised HSI classification.

Extreme learning machine (ELM) has been widely used in supervised HSI classification [5] due to its high efficiency and fast learning speed. ELM [6–8] is a single hidden layer feed-forward neural network with the input weights and biases randomly generated without tedious and time-consuming parameter iterative tuning. Therefore, it maintains the advantage of remarkable training efficiency.

In order to improve the classification accuracy of ELM, the kernel function is used into the ELM model by replacing the activation function, which is termed as kernel ELM (KELM) [9]. The advantages of the KELM function in this algorithm are that the kernel functions used in this algorithm do not need to satisfy Mercer's theorem, and it can also overcome the shortage of random initialization about the input weights and biases, specifically, KELM is robust to the parameters learning [10]. It is well known that both of the Gaussian radial basis function (RBF) kernel and composite kernel (CK) [5] applied on KELM have achieved promising classification performance in HSI classification [5, 11]. However, the RBF [12] only describes the spectral similarity of the pixels without considering the spatial similarity between the neighbor pixels, and the CK by combining spatial and spectral information that also has its limitations. The [13] suggested a suitable kernel should learn all high-order similarities between neighboring samples directly, and then reflect the data lie in complex manifolds. Keeping these points in view, in this paper, an efficient KELM algorithm is proposed, which integrates spectral-spatial information by mean filtering (MF) kernel for supervised HSI classification. MF kernel not only computes the spectral similarity but also considers the spatial similarity by averaging the spatial neighboring pixels of the central pixel. Furthermore, we observe that the KELM with MF kernel shows superior performance, and significantly reduces the

<sup>1</sup>School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

<sup>2</sup>Nanjing Robot Research Institute Co. Ltd, Nanjing 211135, China

<sup>3</sup>Lianyungang E-Port Information Development Co. Ltd, Lianyungang 222042, China

**Corresponding author:**

Zebin Wu

Email: [zebin.wu@gmail.com](mailto:zebin.wu@gmail.com)

computational complexity compared with other methods [14, 15] simultaneously.

Nevertheless, considering that in the situation of limited training samples, the classification result of MF based on KLEM is very noisy, where the existence of noise in HSI not only degrades the image quality but also reduces the performance. Therefore, it is very critical to propose an effective preprocessing algorithm for HSI noise reduction. The given neighborhood spectral bands always have highly similar spatial and structural characteristics in the context of the HSI. To date, a number of theories have proposed to improve HSI classification performance by taking into spatial and spectral information. Literature [16] introduced 3D Gabor filters to extract spectral-spatial information from HSI data, and these filters can be used to generate a set of features that captures specific orientation, scale, and wavelength-dependent properties of an HSI region. In [17], Lu and Li presented a gradient-guided sparse representation method by jointly sparse-coding the similar image patches in neighborhood spectral bands of the HSI. In order to improve the classification performance, a novel multiple kernel learning (MKL) algorithm is introduced by integrating multiple types (i.e. multistructure morphological profiles (MPs), multiattribute MPs) of spectral-spatial features with the guidance of an ideal kernel [18]. What is more, in [19], Gu *et al.* proposed a multistructure-element non-linear MKL, in which the multiple structure elements are employed to generate extended MPs to present spatial-spectral information. Li *et al.* integrated EMAPs with spectral information to formulate the CK for HSI classification [20]. In [21], the bilateral filtering exploited on each band-subset to incorporate the spectral signature in the spatial neighborhood. The above methods demonstrate the efficiency to incorporate the spectral and spatial features instead of one single feature. Therefore, in this paper, we utilize the bilateral filtering on the spectral-adaptive band-subsets to exploit their spectral-spatial information. Due to the fact that the high structural similarity spectral bands have continuous spectral characteristics in each band subset, then, the spectral-adaptive band-subsets partition [17] is introduced to gather the highly similar spectral and discard the weak similar ones. The traditional filtering enforces closeness by weighting the pixel values with coefficients that decrease with distance named domain filter. However, in [22], noise values that corrupt these nearby pixels are mutually less correlated than the signal values, so noise is averaged away while signal is preserved. It also defined a non-linear range filtering that depends on image intensity or color to average image values with weights that decay with dissimilarity, then, it combines domain filtering with range filtering and denotes it as bilateral filtering, the bilateral filtering smooths each band-subset while preserves the edge information of it. Therewith, we merge the band-subsets together and form a new 3D data cube, where the newly formed cube serves as the input of the MF-KELM classifier.

To tackle the above issues, this paper makes two contributions, summarized as follows:

- (1) The proposed MF-KELM method applies the MF kernel to replace the ordinary kernel with the consideration of the underlying data spatial structure as well as the spectral structure.
- (2) The proposed Bilateral MF-KELM method, which combines the bilateral filtering and MF-KELM, utilizes the bilateral filtering preprocessing on each spectral-adaptive band-subset in terms of smoothing the image and with the edge information preserved.

The rest of the paper is organized as follows. In Section II, we introduce the MF-KELM classifier for the supervised HSI classification. Part A is a detailed description of the KELM model. Part B is the introduction of the MF kernel. Part C introduces a preprocessing method named bilateral filtering on spectral-adaptive band-subsets. Section III is the experiments and analysis, where we use our statics to prove the efficiency of our proposed method, and compare with some state-of-the-art HSI classification methods based on KELM. Section IV gives a summary of our work.

## II. HYPERSPPECTRAL IMAGE CLASSIFICATION BASED ON MEAN FILTERING KERNEL EXTREME LEARNING MACHINE

### A) Kernel extreme learning machine

ELM [9] is a random generation single-hidden layer feed-forward neural network. KELM is the evolution of ELM, which utilizes the mapping kernel function to replace the hidden layer of ELM. It shows a better performance compared with other methods.

Given  $N$  different training samples  $\{\mathbf{x}_i, \mathbf{z}_i\}_{i=1}^N$ , where  $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{iD}] \in \mathbb{R}^D$ ,  $D$  stands for the number of the spectral bands or dimensionality of the HSI,  $L$  is the number of the classes,  $N$  represents the number of training samples. The row vector  $\mathbf{z}_i = [z_{i1}, \dots, z_{ik}, \dots, z_{iL}] \in \mathbb{R}^L$  determines which class the sample belongs to, we define  $z_{ik} \in \{0, 1\}$  and  $1 \leq k \leq L$ . If  $z_{ik} = 1$  and any other elements of  $\mathbf{z}_i$  are zero, the sample belongs to the  $k$ th class. The output function of the ELM having  $P$  hidden neurons defined as follows:

$$f(\mathbf{x}_i) = \sum_{j=1}^P \beta_j \cdot G(\omega_j \cdot \mathbf{x}_i + b_j) = \mathbf{z}_i, \quad i = 1, 2, \dots, N \quad (1)$$

where  $G(\cdot)$  represents a non-linear activation function (e.g. RBF),  $\omega_j \in \mathbb{R}^D$  is the input weight vector connecting the  $j$ th hidden neuron and input neurons,  $\beta_j \in \mathbb{R}^L$  is the output weight vector connecting the  $j$ th hidden neuron and the output neurons, and  $b_j$  is the bias of the  $j$ th hidden neuron.  $\omega_j \cdot \mathbf{x}_i$  denotes the inner product of  $\omega_j$  and  $\mathbf{x}_i$ . With  $N$

samples, equation (1) can be briefly written as

$$\begin{aligned} \mathbf{H} \cdot \boldsymbol{\beta} &= \mathbf{Z} \\ \mathbf{H}(\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_p, b_1, \dots, b_p, \dots, \mathbf{x}_1, \dots, \mathbf{x}_N) & \\ &= \begin{bmatrix} G(\mathbf{x}_1) \\ \vdots \\ G(\mathbf{x}_N) \end{bmatrix} \\ &= \begin{bmatrix} G(\boldsymbol{\omega}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & G(\boldsymbol{\omega}_p \cdot \mathbf{x}_1 + b_p) \\ \vdots & \ddots & \vdots \\ G(\boldsymbol{\omega}_1 \cdot \mathbf{x}_N + b_1) & \cdots & G(\boldsymbol{\omega}_p \cdot \mathbf{x}_N + b_p) \end{bmatrix}_{N \times P} \end{aligned} \tag{2}$$

where

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}_{P \times L}, \mathbf{Z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_L \end{bmatrix}_{N \times L}.$$

Equation (2) can be written as

$$\boldsymbol{\beta} = \mathbf{H}^\dagger \mathbf{Z}, \tag{3}$$

where  $\mathbf{H}^\dagger$  is the Moore–Penrose generalized inverse of  $\mathbf{H}$ . It can be transformed into the form like  $\mathbf{H}^\dagger = \mathbf{H}^T (\mathbf{H}\mathbf{H}^T)^{-1}$ . In order to achieve more stability and generalized inverse classification result, a positive value  $\rho^{-1}$  is added to the diagonal elements of  $\mathbf{H}\mathbf{H}^T$ . Therefore, the output function of ELM classifier is expressed as follows

$$f(\mathbf{x}_i) = G(\mathbf{x}_i)\boldsymbol{\beta} = G(\mathbf{x}_i)\mathbf{H}^T \left( \frac{\mathbf{I}}{\rho} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{Z}, \tag{4}$$

where  $\mathbf{Z} \in \mathbb{R}^{N \times L}$  is the training samples label set, similar to SVM, ELM can be generalized to kernel ELM using a kernel trick. The activation function can be replaced by a kernel function; the output function of KELM is expressed as below

$$\begin{aligned} f(\mathbf{x}_i) &= G(\mathbf{x}_i)\mathbf{H}^T \left( \frac{\mathbf{I}}{\rho} + \mathbf{H}\mathbf{H}^T \right)^{-1} \mathbf{Z} \\ &= \begin{bmatrix} K_{MF}(\mathbf{x}_i, \mathbf{x}_1) \\ \vdots \\ K_{MF}(\mathbf{x}_i, \mathbf{x}_N) \end{bmatrix}^T \left( \frac{\mathbf{I}}{\rho} + \mathbf{K}_{MF} \right)^{-1} \mathbf{Z} \\ &= \mathbf{K}_{(\mathbf{x}_i)MF} \left( \frac{\mathbf{I}}{\rho} + \mathbf{K}_{MF} \right)^{-1} \mathbf{Z} \end{aligned} \tag{5}$$

where  $\mathbf{K}_{MF} = [K_{MF}(\mathbf{x}_q, \mathbf{x}_t)]_{q,t=1}^N$ ,  $\mathbf{K}_{(\mathbf{x}_i)MF} = [K_{MF}(\mathbf{x}_i, \mathbf{x}_1), \dots, K_{MF}(\mathbf{x}_i, \mathbf{x}_N)]$ .

Note that the label of the test sample is determined by the index of the output node with the largest value. The traditional RBF kernel function is replaced by MF kernel in KELM, and we observe an excellent generation result by doing this.

## B) Hyperspectral image classification based on mean filtering kernel extreme learning machine

Although the RBF kernel used in KELM has achieved promising classification performance, it does not consider the underlying data structure of HSI. In order to reflect data relations in a kernel, we adopt the MF kernel [9] and incorporate it into the KELM, and the MF kernel computes the mean value of the spatial neighboring pixels in the kernel space to estimate the central pixel.

Given  $\mathbf{x}^m \in \Omega^*$ ,  $m = 1, 2, \dots, \omega^2$ , and  $\Omega^*$  represents the spatial window,  $\omega$  is the size of the window, and  $\mathbf{x}$  is the center pixel. Let us denote  $\phi(\mathbf{x}^m)$  as the image of  $\mathbf{x}^m$  under the map  $\phi$ . Then, the MF can be represented as follows

$$MF(\phi(\mathbf{x})) = \frac{1}{\omega^2} \sum_{m=1}^{\omega^2} \phi(\mathbf{x}^m). \tag{6}$$

In order to describe the similarity between different pixels, we define the function like

$$\begin{aligned} K_{MF}(\mathbf{x}_i, \mathbf{x}_j) &= \langle MF(\phi(\mathbf{x}_i)), MF(\phi(\mathbf{x}_j)) \rangle \\ &= \left\langle \frac{1}{\omega^2} \sum_{m=1}^{\omega^2} \phi(\mathbf{x}_i^m), \frac{1}{\omega^2} \sum_{n=1}^{\omega^2} \phi(\mathbf{x}_j^n) \right\rangle \\ &= \frac{1}{\omega^4} \sum_{m=1}^{\omega^2} \sum_{n=1}^{\omega^2} K(\mathbf{x}_i^m, \mathbf{x}_j^n). \end{aligned} \tag{7}$$

### Algorithm 1. MF-KELM

**Input:** the training samples set  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ , the test samples set  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M]$ , window size  $\omega$ , a positive value  $\rho$ , number of classes  $L$ .

1. Compute the kernel  $K_{pq}$  pixel by pixel in training samples based on equation (7). Then the kernel matrix of training samples  $\mathbf{K}_{MF} \in \mathbb{R}^{N \times N}$  consists of all  $K_{pq}$ ,  $p = 1, \dots, N$ ,  $q = 1, \dots, N$ .
2. Compute the kernel  $K_{sr}$  pixel by pixel among all samples based on equation (7). Then the kernel matrix  $\mathbf{K}_{MF}^* \in \mathbb{R}^{M \times N}$  of test samples consists of all  $\mathbf{K}_{sr}$ ,  $r = 1, \dots, N$ ,  $s = 1, \dots, M$ .
3. Calculate the output of test samples  $\mathbf{Y}$  based on equation (5).
4. According to the output of step 3 to achieve the classification prediction:  $label(\mathbf{y}) = \arg \max_{t=1, \dots, L} f_t$ .

**Output:** the estimated label of  $\mathbf{Y}$ .

Where  $\mathbf{x}_i$  and  $\mathbf{x}_j$  represent the two different pixels, respectively.

Algorithm 1 describes the combination of the MF and KELM methods, and we named it MF-KELM.

Image denoising is an image preprocessing work, which has been widely used in the field of HSI processing and has shown a good performance. The following part C is the introduction of bilateral filtering for HSI denoising based on MF-KELM.

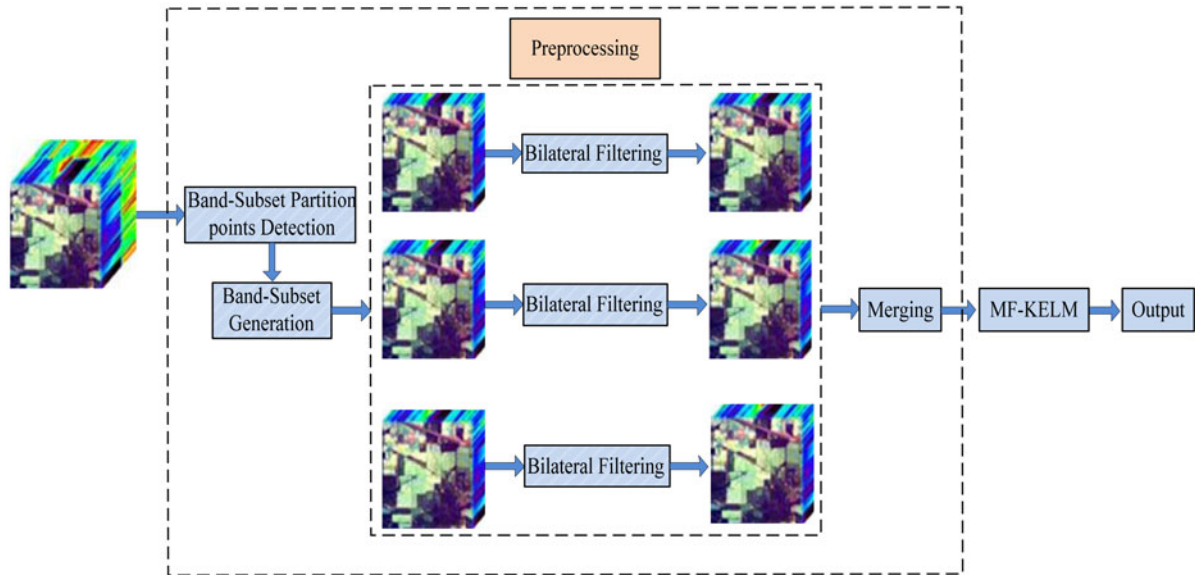


Fig. 1. The flowchart of the proposed Bilateral MF-KELM method.

### C) Spectral band-subsets-wise Bilateral MF-KELM method

Considering the highly similar structural characteristics of neighborhood spectral bands in HSI, which is urgent to exploit the redundant structural information in terms of a better structure characteristics preservation. In this case, based on the structural similarity index metric [23], the spectral-adaptive band-subsets partition [17] method is introduced as follow.

Let  $\mathbf{I} = [\mathbf{I}_1, \mathbf{I}_2, \dots, \mathbf{I}_D]^T$  represents an HSI cube, where  $\mathbf{I}_i \in \mathbb{R}^{P \times Q}$  is  $i$  th band of HSI. Then, we measure the structural similarity between spectral bands  $\mathbf{I}_i$  and  $\mathbf{I}_{i+1}$  as shown in the following:

$$SSIM(\mathbf{I}_i, \mathbf{I}_{i+1}) = \frac{(2\mu_{\mathbf{I}_i}\mu_{\mathbf{I}_{i+1}} + C_1)(2\sigma_{\mathbf{I}_i, \mathbf{I}_{i+1}} + C_2)}{(\mu_{\mathbf{I}_i}^2 + \mu_{\mathbf{I}_{i+1}}^2 + C_1)(\sigma_{\mathbf{I}_i}^2 + \sigma_{\mathbf{I}_{i+1}}^2 + C_2)}, \quad (8)$$

where  $\mu_{\mathbf{I}_i}$  ( $\mu_{\mathbf{I}_{i+1}}$ ) and  $\sigma_{\mathbf{I}_i}$  ( $\sigma_{\mathbf{I}_{i+1}}$ ) represent the mean value and the standard deviation of  $\mathbf{I}_i$  ( $\mathbf{I}_{i+1}$ ), respectively, here,  $C_1$  and  $C_2$  are constants. When applying equation (8) on adjacent spectral bands, a correlation curve  $C$  would generate, where  $C(i) = SSIM(\mathbf{I}_i, \mathbf{I}_{i+1})$ . Shen *et al.* [15] suggest the continuous spectral bands with high structural similarity correspond to relatively stable trend, while obvious drops existing in curve  $C$  demonstrate the adjacent spectral bands have much lower structural similarity. Thus, we according to the detection of sharp drops in  $C$  to achieve the band-subset partition. Let us define the band-subset as  $S_q$ , where  $S_q$  contains  $B_q$  spectral bands with the size of  $P \times Q$ .

The bilateral filtering [22] defined as the combination of the range and domain filtering, which combines the nearby image values in a non-linear way, and preserves image features with a smooth image while preserving edges. Since the homogeneous regions are commonly contained by HSI, [15] argues that for each pixel in a band-subset  $S_q$ , its neighboring pixels will likely share similar spectral characteristics or

have the same class membership. Furthermore, the bilateral filtering, which has the ability to influence both the spatial distance and spectral dissimilarity relative to the center pixel, need to extend the general form to the vector form [21]. We denote  $S_q(a, b)$  as a pixel of  $S_q$ , where  $a$  and  $b$  are the spatial positions of  $S_q$ , we adopt the same bilateral filtering definition as provided in [22]:

$$\tilde{S}_q(a, b) = \frac{\sum_{(x,y) \in \Omega(a,b)} S_q(x, y) \cdot w(x, y; a, b)}{\sum_{(x,y) \in \Omega(a,b)} w(x, y; a, b)}, \quad (9)$$

where  $S_q(x, y)$  stands for the neighboring pixel of  $S_q(a, b)$  within a spatial search window. A  $a \times b$  matrix neighborhood  $\Omega(a, b)$  is centered on the target pixel. The weight  $w(x, y; a, b)$  can be calculated as

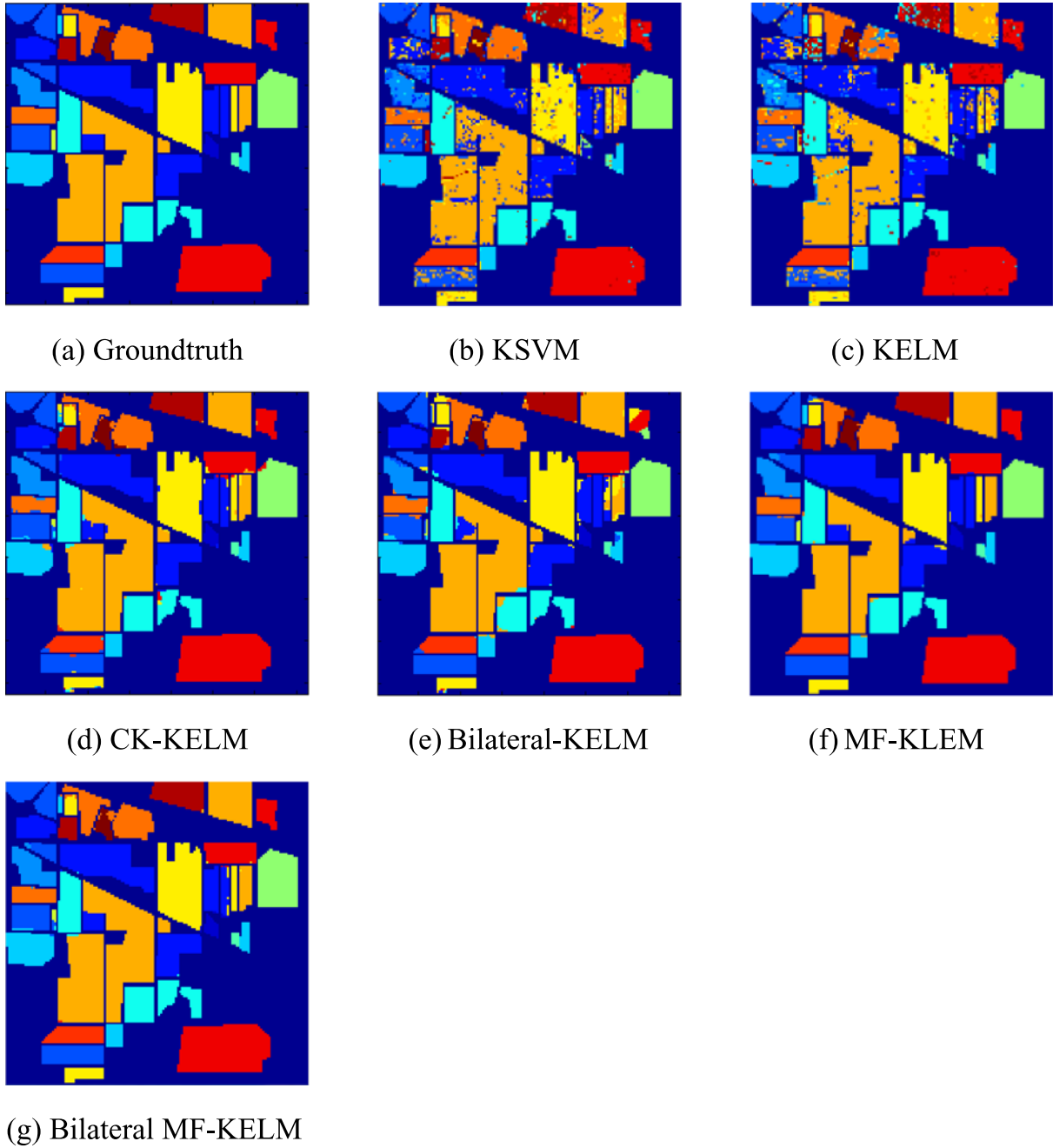
$$w(x, y; a, b) = \exp\left(\frac{\|S_q(x, y) - S_q(a, b)\|_2^2}{2\sigma_r^2}\right) \times \exp\left(\frac{\|x - a\|_2^2 + \|y - b\|_2^2}{2\sigma_d^2}\right), \quad (10)$$

$\sigma_r$  and  $\sigma_d$  are the filter parameters and adaptive corresponding to different  $S_q$ .

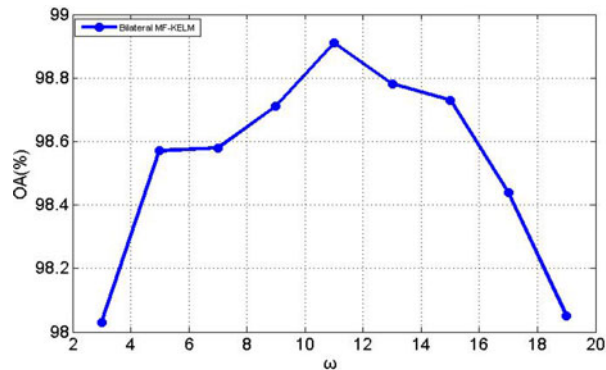
After exploiting the spectral-spatial average information of each band-subset  $S_q$  via bilateral filtering, we merge all the band-subsets into a new three-dimensional data cube, and the cube will act as robust spectral signatures for the input of MF-KELM classifier to predict the final classification results. Figure 1 shows the flowchart of the proposed Bilateral MF-KELM method. Algorithm 2 is the complete description of Bilateral MF-KELM.

### III. EXPERIMENTS AND ANALYSIS

In this section, we use the Indian Pines dataset to evaluate the performance of the proposed method. Indian pines



**Fig. 2.** Indian Pines image. (a) Ground truth; (b) KSVM (OA = 74.21%); (c) KELM (OA = 86.92%); (d) CK-KELM (OA = 94.96%); (e) Bilateral-KELM (OA = 97.29%); (f) MF-KELM (OA = 98.52%); (g) Bilateral MF-KELM (OA = 98.91%).



**Fig. 3.** The classification accuracy for different window sizes  $\omega$  of Indian Pines in Bilateral MF-KELM method.



**Table 1.** Classification accuracy (%) for Indian Pines image on the test set.

Class	Train	Test	KSVM	KELM	CK-KELM	Bilateral-KELM	MF-KELM	Bilateral MF-KELM
Alfalfa	6	48	43.13	72.22	91.67	93.75	100	100
Corn-no till	144	1290	83.43	82.50	95.12	95.89	98.84	98.22
Corn-min till	84	750	75.79	79.02	93.73	98.27	97.07	97.60
Corn	24	210	67.52	72.65	91.43	93.33	98.57	100
Grass/pasture	50	447	93.24	92.35	91.50	97.99	95.75	99.11
Grass/tree	75	672	95.68	96.65	96.13	97.17	98.96	99.85
Grass/pasture-mowed	3	23	36.96	92.31	78.26	91.30	95.65	100
Hay-windrowed	49	440	99.00	98.98	100	99.77	100	100
Oats	2	18	2.78	70.00	38.89	77.78	94.44	50.00
Soybeans-no till	97	871	76.76	85.95	94.26	96.67	97.01	98.85
Soybeans-min till	247	2221	87.87	88.90	96.40	98.69	99.14	99.50
Soybeans-clean till	62	552	85.51	80.60	94.75	91.30	95.83	99.46
Wheat	22	190	97.79	99.06	94.74	90.53	98.42	100
Woods	130	1164	96.93	94.44	95.88	99.48	100	99.83
Big-grass-tree-drives	38	342	57.37	56.84	91.23	99.71	99.71	99.42
Stone-stell towers	10	85	87.53	91.58	87.06	95.19	77.65	77.65
Overall accuracy (%)			74.21	86.92	94.96	97.29	98.52	98.91
Average accuracy (%)			85.68	84.63	84.64	94.81	98.39	98.75
$\kappa$			0.758	0.859	0.943	0.974	0.980	0.985
Time (s) (Feature Extraction)			–	–	–	47.22	–	47.22
Time (s) (Classification)			0.823	0.435	0.656	0.423	0.843	0.725

dataset was collected by the AVIRIS sensor over North western Indiana, and it is an agriculture/forestry landscape with spectrally similar classes and high in-class spectral variability. Figure 2(a) shows the ground truth of Indian Pines dataset. The size of this image is  $145 \times 145$ . A total of 20 water absorption and noisy bands (104–108, 150–163, and 220) were removed from the original 220 bands. It contains 16 classes, and 10% of pixels per class randomly selected as the training dataset. In order to evaluate the extensibility and effectiveness of our proposed Bilateral MF-KELM method, we compare it with KSVM [19], KELM [9], CK-KELM, Bilateral-KELM [15], and MF-KELM [10] on the Indian Pines

#### Algorithm 2. Bilateral MF-KELM

**Input:** the HSI data cube  $\mathbf{I}$ , the training samples set  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ , the test samples set  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_M]$ ,  $C_1$ ,  $C_2$ , window size  $\omega$ , a positive value  $\rho$ , number of classes  $L$ .

1. Partition the spectral-adaptive band-subsets of HSI, and apply the bilateral filtering to each band-subset, merge them into a new 3D cube data. The above steps depend on equations (8), (9), and equation (10).
2. According to step 1, the new training set  $\mathbf{X}^*$  and new test set  $\mathbf{Y}^*$  are obtained.
3. Take  $\mathbf{X}^*$  and  $\mathbf{Y}^*$  as the input of algorithm 1, then, get the final classification result of the test dataset  $\mathbf{Y}^*$ .

**Output:** the estimated label of  $\mathbf{Y}$ .

We adopt the overall accuracy (OA), average accuracy (AA), and the  $\kappa$  coefficient of agreement ( $k$ ) to measure the classification accuracies, and we also use the Running Time(s) to measure the computational cost, including the Feature Extraction Time(s) and Classification Time(s). All experiments were carried out using MATLAB on an Intel Xeon 2.4.00-GHz machine with 32 GB of RAM.

MF kernel is sensitive to the window size  $\omega$ . Figure 3 shows the sensitivity of the MF kernel-based method Bilateral MF-KELM to the window size. With the increase of  $\omega$ , the classification accuracy also increases gradually, but the OA saturates for  $\omega = 11$ . As  $\omega$  continues to grow, the decreasing performance suggests that the neighboring relations would be oversmooth. For the preprocessing method, the filter parameters value of  $\sigma_r$  and  $\sigma_d$  are determined by cross-validation. We set the window size of the bilateral filter that depends on the structure of different scenes to be  $9 \times 9$ . In Table 1, the OA of Bilateral MF-KELM method is 98.91%, which is 24.7, 11.99, 3.95, 1.62, and 0.39% higher than the KSVM, KELM, CK-KELM, Bilateral-KELM, and MF-KELM, respectively. Figures 2(a)–2(g) illustrate the classification maps of the methods mentioned above, it is obvious that the proposed Bilateral MF-KELM method outperforms other compared methods. In addition, the Bilateral MF-KELM mainly contains two steps: the Feature Extraction and the MF-KELM, here, the time complexity of them are  $O(PQD|\Omega(a, b)|^2)$  and  $O(\omega^4 DN^2 + N^3)$ , respectively, where  $|\Omega(a, b)|$  represents the number of pixels in the neighborhood, then, the whole cost is  $O(PQD|\Omega(a, b)|^2 + \omega^4 DN^2 + N^3)$ . The traditional KELM time complexity is  $O(DN^2 + N^3)$ . It is worth to note that a higher value of  $\omega$  indicates a higher computation in Bilateral MF-KELM. Therefore, it is suitable to select a much smaller  $\omega$  value while achieving a good performance.

#### IV. CONCLUSION

In this paper, we proposed a novel spectral-spatial Bilateral MF-KELM method of HSI classification, where we incorporate the MF kernel into the KELM model by replacing the original RBF kernel. MF-KELM captures not only the spectral similarity of pixels, but also the spatial similarity

of a central pixel by averaging its spatially neighboring pixels in the kernel space. Moreover, in order to overcome the drawback that MF-KELM is unable to eliminate noise effectively, we apply bilateral filtering on each band-subset to reduce the image noise for the HSI preprocessing. By comparing the Bilateral KELM with other different HSI supervised classification methods, obvious visual and numerical merit can be demonstrated in real data experiments.

## ACKNOWLEDGEMENTS

This work was supported in part by the National Natural Science Foundation of China under Grant 61471199, Grant 61772274, Grant 61701238, Grant 91538108, Grant 11431015, and Grant 61671243, in part by the Jiangsu Provincial Natural Science Foundation of China under Grant BK20170858, in part by the Fundamental Research Funds for the Central Universities under Grant 30917015104, in part by the China Postdoctoral Science Foundation under Grant 2017M611814. The authors would like to thank Professor Landgrebe from Purdue University for providing the Indian Pines dataset, Professor Gamba from the University of Pavia for providing the University of Pavia dataset, Professor Lin from the National Taiwan University for providing LIBSVM toolbox, and Professor Huang from Nanyang Technological University for providing ELM toolbox.

## REFERENCES

- [1] Fauvel M.; Tarabalka Y.; Benediktsson J. A.; Chanussot J.; Tilton J. C.: Advances in spectral-spatial classification of hyperspectral images. *Proc. IEEE*, **101** (3) (2013), 652–675.
- [2] Guo A. J. X.; Zhu F.: Spectral-spatial feature extraction and classification by ANN supervised with center loss in hyperspectral imagery. *IEEE Trans. Geosci. Remote Sens.*, **57** (3) (2019), 1755–1767.
- [3] Gao L.; Du Q.; Zhang B.; Yang W.; Wu Y.: A comparative study on linear regression-based noise estimation for hyperspectral imagery. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **6** (2) (2013), 488–498.
- [4] Zhang B.; Li S.; Jia X.; Gao L.; Peng M.: Adaptive Markov random field approach for classification of hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.*, **8** (5) (2011), 973–977.
- [5] Mahesh P.; Maxwell A. E.; Warner T. A.: Kernel-based extreme learning machine for remote-sensing image classification. *Remote Sens. Lett.*, **4** (9) (2013), 853–862. doi: 10.1080/2150704X.2013.805279.
- [6] Huang G. B.; Zhou H. M.; Ding X. J.: Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man Cybernet. Part B: Cybernet.*, **4** (2) (2012), 513–529.
- [7] Huang G. B.; Bai Z.; Chi M. V.: Local receptive fields based extreme learning machine[J]. *IEEE Comput. Intell. Mag.*, **10** (2) (2015), 18–29.
- [8] Huang G. B.; Zhu Q. Y.; Siew C. K.: Extreme learning machine: a new learning scheme of feed-forward neural networks. *Inst. Electric. Electron. Eng. Inc.*, **2** (2) (2004), 985–990.
- [9] Huang G. B.; Ding X.; Zhou H.: Optimization method based extreme learning machine for classification[J]. *Neurocomputing*, **74** (1–3) (2010), 155–163.
- [10] Liu J.; Wu Z.; Wei Z.; Xiao L.; Sun L.: Spatial-spectral kernel sparse representation for hyperspectral image classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, **6** (6) (2013), 2462–2471.

- [11] Huang G. B.; Zhou H.; Ding X.: Extreme learning machine for regression and multiclass classification. *Syst. Man Cybernet. Part B: IEEE Trans. Cybernet.*, **42** (2) (2012), 513–529.
- [12] Ring M.; Eskofier B. M.: An Approximation of the Gaussian RBF Kernel for Efficient Classification with SVMs [M], New York, NY, USA, 2016.
- [13] Liu J.; Wu Z.; Xiao Z.; Yang J.: Region-Based Relaxed Multiple Kernel Collaborative Representation for Hyperspectral Image Classification. *IEEE Access*, **5** (1) (2017), 20921–20933.
- [14] Camps-Valls G.; Gomez-Chova L.; Muñoz Maré J.; Vila-Francés J.; Calpe-Maravilla J.: Composite kernels for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.*, **3** (1) (2006), 93–97.
- [15] Shen Y.; Xu J.; Li H.: ELM-based spectral-spatial classification of hyperspectral images using bilateral filtering information on spectral band-subsets[C]. *Geosci. Remote Sens. Symp. IEEE*, (2016), 497–450.
- [16] Bau T. C.; Sarkar S.; Healey G.: Hyperspectral region classification using a three-dimensional Gabor filter bank. *IEEE Trans. Geosci. Remote Sens.*, **48** (9) (2010), 3457–3464.
- [17] Lu T.; Li S.: Gradient-guided sparse representation for hyperspectral image denoising, in *2015 IEEE Int. Geoscience and Remote Sensing Symp. (IGARSS)*, Milan, 2015, 1128–1131.
- [18] Gao W.; Peng Y.: Ideal kernel-based multiple kernel learning for spectral-spatial classification of hyperspectral image [J]. *IEEE Geosci. Remote Sens. Lett.*, **14** (7) (2017), 1051–1055.
- [19] Gu Y.; Liu T.; Jia X.: Nonlinear Multiple Kernel Learning With Multiple-Structure-Element Extended Morphological Profiles for Hyperspectral Image Classification[J]. *IEEE Transactions on Geoscience and Remote Sensing*, (2016), 1–13.
- [20] Li J.; Reddy Marpu P.; Plaza A.; Bioucas-Dias J. M.; Benediktsson J. A.: Generalized composite kernel framework for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, **51** (9) (2013), 4816–4829.
- [21] Peng H.; Rao R.: Hyperspectral image enhancement with vector bilateral filtering. *16th IEEE International Conference on Image Processing (ICIP)* (2009), 3713–3716.
- [22] Tomasi C.; Manduchi R.: Bilateral filtering for gray and color images, in *Sixth Int. Conf. on Computer Vision (IEEE Cat. No.98CH36271)*, Bombay, India, 1998, 839–846.
- [23] Wang Z.; Bovik A. C.; Sheikh H. R.; Simoncelli E. P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, **13** (4) (2004), 600–612.

**Wenting Shang** received the B.S. degree from Zhengzhou University, Zhengzhou, China, in 2014. She is currently pursuing the Ph.D. degree with the Nanjing University of Science and Technology. Her current research interests include hyperspectral target detection and hyperspectral image classification.

**Zebin Wu** (M'13–SM'17) was born in Zhejiang, China, in 1981. He received the B.Sc. and Ph.D. degrees in computer science and technology from the Nanjing University of Science and Technology, Nanjing, China, in 2003 and 2007, respectively. In 2016, he joined the Department of Mathematics, University of California at Los Angeles, Los Angeles, CA, USA, as a Visiting Scholar. He was a Visiting Scholar with the Hyperspectral Computing Laboratory, Department of Technology of Computers and Communications, Escuela Politécnica, University of Extremadura, Cáceres, Spain, from 2014 to 2015. He is currently a Professor with

the School of Computer Science and Engineering, Nanjing University of Science and Technology. His research interests include hyperspectral image processing, high-performance computing, and computer simulation.

**Yang Xu** (S'14–M'16) received the B.Sc. degree in applied mathematics and the Ph.D. degree in pattern recognition and intelligence systems from the Nanjing University of Science and Technology (NUST), Nanjing, China, in 2011 and 2016, respectively. He is currently a Lecturer with the School of Computer Science and Engineering, NUST. His research interests include hyperspectral image classification, hyperspectral detection, image processing, and machine learning.

**Yan Zhang** was born in Jiangsu, China, in 1981. He received the B.Sc. degree from the Suzhou University in 2004.

**Zhihui Wei** was born in Huaian, China, in 1963. He received the B.Sc. and M.Sc. degrees in applied mathematics and the Ph.D. degree in communication and information system from South East University, Nanjing, China, in 1983, 1986, and 2003, respectively. He is currently a Professor and a Doctoral Supervisor with the Nanjing University of Science and Technology, Nanjing. His research interests include partial differential equations, mathematical image processing, multiscale analysis, sparse representation, and compressive sensing.