

ORIGINAL PAPER

Multiple feature regularized kernel for hyperspectral imagery classification

XU YAN,¹ PENG JIANGTAO² AND DU QIAN¹ 

In this paper, a multiple feature regularized kernel is proposed for hyperspectral imagery classification. To exploit the label information, a regularized kernel is used to refine the original kernel in the Support Vector Machine classifier. Furthermore, since spatial features have been widely investigated for hyperspectral imagery classification, different types of spatial features including spectral feature, local feature (i.e. local binary pattern), global feature (i.e. Gabor feature), and shape feature (i.e. extended multiattribute profiles) are included to provide distinguish discriminative information. Finally, a majority voting-based ensemble approach, which combines different types of features, is adopted to further increase the classification performance. Combining different discriminative feature information can improve the classification performance since one type of feature may result in poor performance, especially when the number of training samples is limited. Experimental results demonstrated that the proposed approach has superior performance compared with the state-of-the-art classifiers.

Keywords: Hyperspectral image, Classification, Feature fusion

Received 17 May 2019; Revised 25 February 2020

I. INTRODUCTION

Hyperspectral images (HSIs) contain hundreds of contiguous spectral bands ranging from visible to infrared bands, which can provide rich spectral information [1]. Compared with multispectral images with only a few spectral bands, they offer an advantage in terms of classification and detection [2]. Due to the rich spectral bands, HSIs have a variety of applications including environmental pollution control, agriculture precision farming, mineral exploration, etc [3–6]. Accurate classification is very important to those applications. The key for accurate classification is to learn an accurate similarity metric between samples. The traditional pixel-wise algorithms assume the samples from the same class have similar spectral characteristics and simply compare the similarity between pixels. However, for a complicate HSI scene, different materials might have a similar spectral signature, and pixels from the same class usually have various spectral signature. Thus, it is quite difficult to distinguish them by simply comparing the spectral distance. Additionally, the limited number of training samples and the high dimensionality, also known as the Hughes Phenomenon, prohibit the performance of HSI classification [7].

To overcome the above-mentioned problems, kernel-based approaches have been applied for HSI classification and shown excellent performance [8–10]. Kernel methods project the original samples into a high-dimensional space, where the samples are more separable. The commonly used kernels include Gaussian radial basis function (RBF), Polynomial, and linear kernels [11]. In [12], kernel-based Support Vector Machine (SVM) is proposed for HSI classification. A region-kernel-based SVM [13], which measures the region-to-region similarity, is able to capture the spatial-spectral similarity and has shown improved performance. Other variants such as generalized composite kernel [14] and sample-cluster composite kernels [15] are also introduced for spectral-spatial HSI classification. In [16], a band-weighted SVM, which generates a band weight vector to regularize the original SVM, is presented for HSI classification. Kernel collaborative representation with Tikhonov regularization is presented for HSI classification in [17, 18].

Although kernel methods have achieved excellent performance for HSI classification, the kernels are constructed using the distance between the training samples only while ignoring the available label information. The label information among the training samples can be used to help refine the original kernel. In [19–21], the ideal kernel method is presented which incorporated the label information into the original kernel and has shown better performance compared to using the original kernel only. An ideal regularized composite kernel (IRCK) framework is applied for HSI classification in [22], which combines spatial information, spectral information, and label information simultaneously.

¹Department of Electrical and Computer Engineering, Mississippi State University, Starkville, USA

²Hubei Key Laboratory of Applied Mathematics, Faculty of Mathematics and Statistics, Hubei University, Wuhan, China

Corresponding author: J. Peng, Email: pengtj982@126.com

Multiple Kernel Learning (MKL) [23], which combines multiple kernels to improve the performance, has been applied for HSI. A composite kernel approach, which balances the spatial and spectral information, is presented in [24], and four types of different composite kernels are proposed including summation kernel, weighted summation kernel, stacked kernel, and cross-information kernel. In [25], a two-step discriminative MKL (DMKL), which can increase the between-class scatter while decrease the within-class scatter in the reproduce kernel Hilbert space, is presented for HSI classification.

The abovementioned IRCK method incorporates the spatial information in terms of the spatial mean pixel of the neighborhood; however, it does not consider other different types of spatial information, which may provide complementary discriminative information. Different types of spatial features have been widely investigated for HSI classification. For instance, in [26], a Gabor-filtering-based nearest regularized subspace is presented for HSI classification. In [27], local binary pattern is applied to extract local image features such as edges, corners, and spots, and then extreme learning machine is applied to achieve excellent performance for HSI classification. In [28], extended multi-attribute profiles feature (EMAP) is presented for the analysis of hyperspectral imagery, which can efficiently extract the spatial information for classification purposes.

Furthermore, for a complicated HSI scene, one type of feature may not guarantee good performance, especially when the number of training samples is small. Different types of features can be used to provide more discriminative information, and multiple feature learning can be a good solution to solve the above limitation. For instance, a multiple feature learning framework, which integrates both linear and non-linear features, is introduced for HSI classification [29]. In [30], a non-linear joint collaborative representation model with multiple feature learning is presented to solve the small sample set problem in HSI. An efficient patch alignment framework, which combines multiple features, is proposed in [31].

Inspired by the idea of multiple feature learning, in this paper, we propose to combine different types of features for the IR kernel. Four types of features including spectral feature, Gabor feature, EMAP feature, and LBP feature are investigated due to their potential for HSI classifications. Moreover, a majority voting-based ensemble approach is adopted to make more robust classifications. The rest of the paper is organized as follows. The related work on regularized kernel is introduced in Section II. Section III shows the proposed framework. In Section IV, the experimental results and analysis are provided. Finally, Section V draws the conclusion.

II. REGULARIZED KERNEL

A) Standard kernel

Suppose we have a set of HSI training samples $\mathcal{S} = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)\}$ with $\mathbf{x}_i \in \mathbb{R}^d$, where d

denotes the dimensionality of a sample. The kernel measures the similarity between two samples. The most commonly used kernels are linear kernel, polynomial kernel, and RBF kernel. We choose the RBF kernel in our experiment due to its wide application for HSI classification, and the RBF kernel is calculated as

$$K_{ij} = K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\delta^2}\right), \quad (1)$$

where δ is the band width of the RBF kernel.

B) Ideal kernel

The ideal kernel is defined as [20]

$$T_{ij} = T(\mathbf{x}_i, \mathbf{x}_j) = \begin{cases} 1, & y_i = y_j \\ 0, & y_i \neq y_j \end{cases}. \quad (2)$$

The ideal kernel is inspired by the idea that if two samples should be considered as "similar" if and only they belong to the same class. The ideal kernel has included the label information.

C) Ideal regularized kernel

The similarity measurement between samples in the standard kernel does not consider the labeled information, and a more desirable kernel can be learned by incorporating the labeled information. In [21], an ideal regularization (IR) learning framework is proposed, which can efficiently incorporate the label information into the standard kernel. The following equation learns a more desirable matrix \mathbf{K} given the initial kernel \mathbf{K}_o [21]

$$\min_{\mathbf{K} \succeq \mathbf{0}} (D(\mathbf{K}, \mathbf{K}_o) + \gamma \Omega(\mathbf{K})), \quad (3)$$

where D represents the von Neumann divergence between \mathbf{K} and \mathbf{K}_o , Ω is a regularization term, and γ is the regularization parameter. $\mathbf{K} \succeq \mathbf{0}$ indicates that \mathbf{K} is a symmetric positive semidefinite matrix. The von Neumann divergence is expressed as

$$D(\mathbf{K}, \mathbf{K}_o) = \text{tr}(\mathbf{K} \log \mathbf{K} - \mathbf{K} \log \mathbf{K}_o - \mathbf{K} + \mathbf{K}_o), \quad (4)$$

where $\text{tr}(\bullet)$ represents the trace of a matrix and $\Omega(\mathbf{K})$ is defined as $\Omega(\mathbf{K}) = -\text{tr}(\mathbf{K}T)$. The functionality of this term is to incorporate the label information into the standard kernel. Therefore, the final objective function is:

$$\min_{\mathbf{K} \succeq \mathbf{0}} \text{tr}(\mathbf{K} \log \mathbf{K} - \mathbf{K} \log \mathbf{K}_o - \mathbf{K} + \mathbf{K}_o - \gamma \text{tr}(\mathbf{K}T)). \quad (5)$$

Taking the derivative regarding \mathbf{K} and setting the equation to zero, the following equation for \mathbf{K} is derived as

$$\mathbf{K} = \exp(\log \mathbf{K}_o + \gamma T) = \mathbf{K}_o * \exp(\gamma T), \quad (6)$$

where $*$ represents the element-wise dot product between two matrices. Using the Taylor expansion of equation (6),

it can be also expressed as:

$$K = K_o + \gamma K_o * T + \frac{\gamma^2}{2!} K_o * T^2 + \dots \quad (7)$$

The above equation indicates that the IR kernel K can be considered as a linear combination of the standard kernel and the ideal kernels.

The out-of-sample extension is investigated for samples that are not encountered before. Denote $S = K_o^{-1}(K + K_o)K_o^{-1}$, then the kernel between the two new points s and t can be calculated as

$$K(s, t) = -K_o(s, t) + \sum_{i,j=1}^N S(i, j)K_o(s, x_i)K_o(x_j, t). \quad (8)$$

D) Majority voting-based ensemble approach

We propose to use a majority voting-based ensemble approach to combine the output of SVM using the IR kernel with different types of features. Suppose we are using m features and $f_i(x)$ represents the output using the i -th feature, the final output is calculated as

$$\tilde{y} = \text{mode}\{f_1(x), f_2(x), \dots, f_m(x)\}. \quad (9)$$

III. PROPOSED FRAMEWORK

In this paper, a majority voting-based multi-feature IR kernel method which can efficiently deal with the small training sample problem is proposed. Figure 1 shows the proposed framework. First, principal component analysis [32] is conducted on the original dataset to extract the principal components of the original dataset. Different types of features including EMAP features, Gabor features, and LBP features are then extracted from the principle components. Classification is conducted on each type of feature with the IR kernel-based SVM. Finally, a majority voting-based ensemble approach is adopted to combine the results of the classification output.

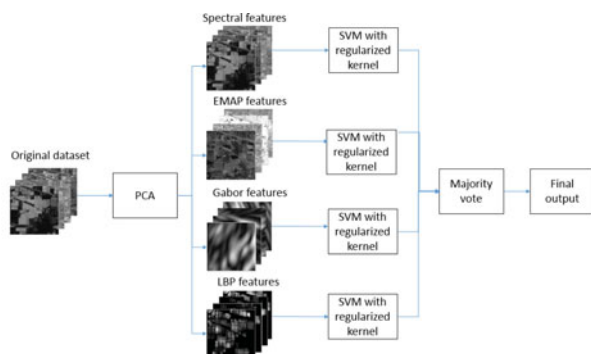


Fig. 1. Proposed framework.

IV. EXPERIMENTAL RESULTS

A) Experimental dataset

Three popular HSI datasets are used in our experiments.

- (1) *Indian Pines*: This dataset is collected by the Airborne Visible and Infrared Imaging Spectrometer (AVIRIS) sensor over the Indian Pines site in June 1992. The spatial size is 145×145 with the spatial resolution of 20 m/pixel. The 200 spectral bands are ranging from 400 to 2500 nm. After bad bands removal, 200 spectral bands remain. There are a total of 16 classes in the scene. The false color infrared of bands 50, 27, and 17 is shown in Fig. 2(a), and the groundtruth is shown in Fig. 2(b).
- (2) *University of Pavia*: The second dataset is acquired by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor, and the scene covers the University of Pavia with 115 spectral bands ranging from 0.43 to 0.86 μm . The spatial size is 610×340 with the spatial resolution of 1.3 m/pixel including nine classes. After removing the noisy band, 103 bands remain. Figures 3(a) and 3(b) show the color infrared composite of bands 60, 30, 2 and the groundtruth, respectively.
- (3) *Salinas*: The third dataset is collected by the AVIRIS sensor over Salinas Valley, California. The spatial size is 512×217 with the spatial resolution of 3.7 m/pixel. After removing the 20 water absorption bands, 200 bands remain. The color infrared composite of bands 47, 27, 13 and the groundtruth is shown in Fig. 4(a), and the groundtruth of this dataset is shown in Fig. 4(b). There are a total of 16 classes.

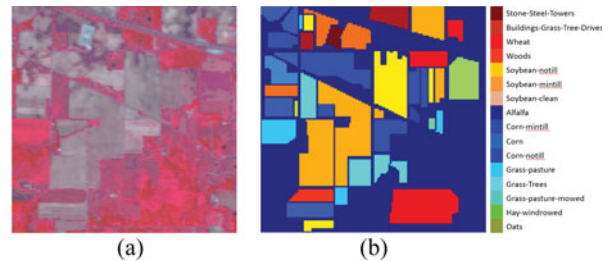


Fig. 2. Indian pines dataset. (a) Color-infrared-composite of bands 50, 27, and 17. (b) Groundtruth.

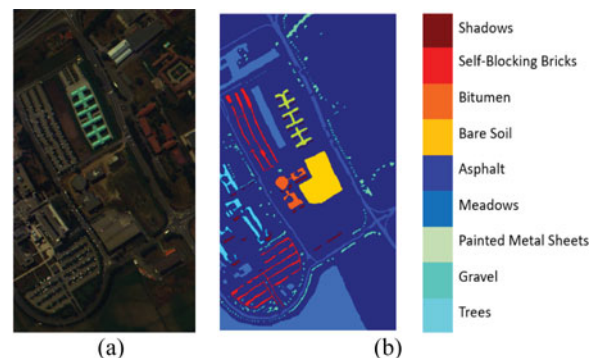


Fig. 3. University of Pavia dataset. (a) Color-infrared-composite of bands 60, 30, and 2. (b) Groundtruth.

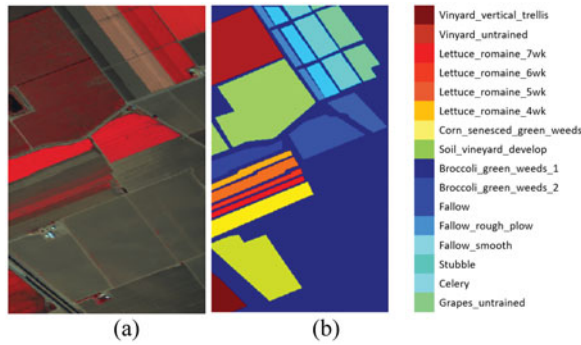


Fig. 4. Salinas dataset. (a) Color-infrared-composite of bands 47, 27, and 13. (b) Groundtruth.

B) Experimental setup

The number of training samples per class is denoted as N . In the experiment, N is chosen to be $\{3, 5, 7, \dots, 13\}$, and for those classes less than N samples, half of the total samples are chosen for training. The rest of the labeled samples are chosen as the testing set. Gaussian kernel is investigated, and the LIBSVM [33] is used to implement SVM. The number of PCs for the Gabor features, EMAP features, and LBP features are set to 10, 4, and 4, respectively [26–28]. The EMAP features are generated according to [28], and the parameters for LBP are set according to [27]. Each experiment is repeated 50 times to avoid bias.

C) Results

Tables 1–3 list the overall accuracies (OAs), average accuracies (AAs) and the κ coefficient of Indian Pines, University of Pavia, and Salinas, respectively, for different types of features using the standard and the regularized kernel; furthermore, a majority voting-based ensemble approach is presented with the standard and regularized kernel to better utilize the discriminative information.

Several observations can be drawn from Table 1. First, the regularized versions consistently outperform the standard kernels using all the four features. Specifically, for the EMAP feature, the regularized kernel can approximately improve the OAs of standard kernel around 5% when the number of training samples is small. For the spectral, Gabor, and LBP, the IR kernels can outperform the standard kernel, indicating the advantage of incorporating the label information into the kernel metric. Second, the majority voting-based ensemble approach can achieve higher OAs compared using a single feature for both the standard and IR kernels. The explanation is that by combining multiple features, complementary discrimination information brought by multiple features is used. Thus, better performance can be achieved. Additionally, the classification performance is improved with the increased number of training samples.

From the results for the University of Pavia, it can be observed that the IR kernel versions can significantly outperform the standard kernel on both the EMAP and Gabor features. For instance, the OAs increase approximately by 3 and 5% on EMAP and Gabor feature, respectively, for the

Table 1. Classification accuracies for different features and ensemble approach on Indian Pines.

N		Spectral		EMAP		Gabor		LBP		Ensemble	
		Standard	IR	Standard	IR	Standard	IR	Standard	IR	Standard	IR
3	OA	42.69 ± 3.37	42.14 ± 4.48	57.20 ± 5.24	62.51 ± 5.44	47.56 ± 3.40	47.69 ± 3.50	46.26 ± 2.35	47.49 ± 2.73	65.84 ± 2.95	68.39 ± 3.17
	AA	55.99 ± 2.81	55.72 ± 2.96	69.68 ± 3.17	73.79 ± 2.89	68.08 ± 2.49	68.59 ± 2.78	64.90 ± 2.23	65.88 ± 2.27	80.37 ± 1.75	82.38 ± 1.42
	K	36.30 ± 3.49	35.76 ± 4.55	52.15 ± 5.58	57.98 ± 5.86	42.42 ± 3.61	42.68 ± 3.73	40.85 ± 2.45	42.18 ± 2.87	61.87 ± 3.24	64.72 ± 3.43
5	OA	48.34 ± 3.59	48.11 ± 3.78	67.49 ± 3.21	72.60 ± 4.53	56.40 ± 3.74	57.31 ± 2.75	56.18 ± 2.93	57.28 ± 2.60	76.19 ± 2.06	78.75 ± 2.66
	AA	61.69 ± 2.54	61.41 ± 2.62	77.40 ± 2.31	81.57 ± 2.89	74.40 ± 2.44	75.39 ± 1.88	73.25 ± 2.32	74.17 ± 1.66	87.42 ± 1.13	89.03 ± 1.49
	K	42.57 ± 3.78	42.17 ± 4.00	63.48 ± 3.53	69.16 ± 5.01	51.74 ± 4.03	52.92 ± 2.93	51.49 ± 3.14	52.68 ± 2.69	73.29 ± 2.26	76.13 ± 2.94
7	OA	50.87 ± 3.45	52.59 ± 2.92	73.04 ± 3.23	77.81 ± 3.13	61.63 ± 2.29	62.81 ± 3.07	62.41 ± 2.44	64.40 ± 2.80	81.26 ± 1.94	84.28 ± 1.77
	AA	65.05 ± 2.30	65.53 ± 2.37	81.67 ± 1.88	85.43 ± 1.85	77.69 ± 1.47	78.85 ± 1.99	78.19 ± 1.63	79.35 ± 1.89	90.62 ± 0.80	92.31 ± 0.87
	K	45.32 ± 3.39	47.08 ± 3.10	69.67 ± 3.56	74.94 ± 3.47	57.33 ± 2.46	58.86 ± 3.29	58.20 ± 2.58	60.38 ± 3.02	78.91 ± 2.14	82.27 ± 1.96
9	OA	54.64 ± 2.89	54.24 ± 2.88	77.59 ± 2.91	81.33 ± 3.68	67.37 ± 2.45	67.61 ± 2.73	67.89 ± 2.2	69.84 ± 2.30	86.24 ± 1.61	87.71 ± 1.99
	AA	67.64 ± 2.07	67.62 ± 1.77	84.77 ± 1.73	87.82 ± 2.01	81.58 ± 1.45	81.79 ± 1.76	81.41 ± 1.35	83.10 ± 1.58	93.28 ± 0.70	94.24 ± 0.79
	K	49.37 ± 3.14	48.99 ± 3.01	74.71 ± 3.22	78.89 ± 4.08	63.66 ± 2.62	64.01 ± 2.93	64.10 ± 2.35	66.31 ± 2.45	86.11 ± 2.22	88.45 ± 1.80
11	OA	55.99 ± 2.71	56.99 ± 2.47	79.91 ± 2.70	83.28 ± 3.24	70.22 ± 2.61	71.72 ± 2.52	71.99 ± 2.28	72.86 ± 2.49	88.52 ± 1.54	90.14 ± 1.40
	AA	69.23 ± 1.25	70.04 ± 1.63	86.59 ± 1.43	89.10 ± 1.94	83.47 ± 1.46	84.26 ± 1.53	84.50 ± 1.50	85.39 ± 1.46	94.60 ± 0.62	95.46 ± 0.64
	K	50.79 ± 2.83	51.93 ± 2.64	77.30 ± 2.99	81.04 ± 3.64	66.74 ± 2.83	68.50 ± 2.71	68.64 ± 2.48	69.65 ± 2.67	87.01 ± 1.72	88.82 ± 1.57
13	OA	58.47 ± 2.14	59.64 ± 2.22	81.54 ± 2.57	84.09 ± 3.12	73.18 ± 2.39	73.68 ± 2.57	74.16 ± 2.52	76.38 ± 1.79	90.12 ± 1.32	91.61 ± 1.32
	AA	70.93 ± 1.65	72.23 ± 1.67	87.74 ± 1.47	90.14 ± 1.48	85.30 ± 1.40	85.32 ± 1.57	86.04 ± 1.22	87.38 ± 1.09	95.52 ± 0.57	96.34 ± 0.49
	K	53.53 ± 2.28	54.81 ± 2.36	79.11 ± 2.86	81.95 ± 3.49	70.01 ± 2.61	70.64 ± 2.76	71.02 ± 2.71	73.49 ± 1.98	88.81 ± 1.47	90.48 ± 1.49

Table 2. Classification accuracies for different features and ensemble approach on the University of Pavia.

N		Spectral		EMAP		Gabor		LBP		Ensemble	
		Standard	IR	Standard	IR	Standard	IR	Standard	IR	Standard	IR
3	OA	55.17 ± 4.94	56.77 ± 6.49	73.5 ± 6.67	77.70 ± 5.38	55.85 ± 5.59	57.67 ± 7.54	43.74 ± 5.58	44.61 ± 5.29	78.47 ± 3.09	81.40 ± 3.60
	AA	67.79 ± 3.04	67.84 ± 3.01	82.12 ± 2.88	85.80 ± 2.94	56.72 ± 3.45	58.33 ± 4.10	49.90 ± 3.58	51.73 ± 2.89	82.23 ± 2.22	84.80 ± 1.89
	K	45.32 ± 4.77	46.75 ± 6.49	66.88 ± 7.45	71.97 ± 6.21	44.44 ± 5.66	46.64 ± 7.47	33.30 ± 4.85	34.46 ± 4.69	72.06 ± 3.68	75.85 ± 4.23
5	OA	60.22 ± 4.41	60.92 ± 5.07	82.32 ± 4.47	86.12 ± 4.21	63.45 ± 5.71	66.59 ± 5.61	52.37 ± 5.47	51.27 ± 5.04	85.99 ± 2.87	88.49 ± 2.49
	AA	71.62 ± 2.18	71.21 ± 2.36	88.17 ± 2.6	91.79 ± 1.96	63.92 ± 2.71	66.67 ± 3.20	59.28 ± 2.15	58.60 ± 58.0	88.35 ± 1.61	90.41 ± 1.46
	K	50.85 ± 4.46	51.4 ± 5.16	77.44 ± 5.33	82.30 ± 5.11	53.44 ± 5.79	57.26 ± 6.06	42.82 ± 4.97	41.84 ± 4.88	81.63 ± 3.47	84.89 ± 3.14
7	OA	64.45 ± 5.21	65.03 ± 4.69	87.25 ± 5.00	89.89 ± 3.63	67.28 ± 4.29	72.36 ± 4.99	58.07 ± 3.76	58.63 ± 3.85	90.14 ± 2.13	92.59 ± 1.57
	AA	73.87 ± 1.89	74.34 ± 1.93	91.10 ± 2.32	93.94 ± 1.82	67.21 ± 2.09	71.99 ± 2.51	65.03 ± 2.98	65.04 ± 2.27	91.2 ± 1.47	93.59 ± 1.18
	K	55.66 ± 5.37	56.37 ± 4.79	83.62 ± 6.01	86.97 ± 4.51	57.89 ± 4.55	64.45 ± 5.53	49.16 ± 3.73	49.70 ± 3.77	86.97 ± 2.66	90.22 ± 2.04
9	OA	65.90 ± 4.10	66.63 ± 4.49	87.62 ± 3.54	92.02 ± 3.76	71.07 ± 4.00	74.97 ± 4.66	61.65 ± 3.5	63.45 ± 2.71	91.96 ± 1.34	94.34 ± 1.22
	AA	75.41 ± 1.65	75.49 ± 1.62	92.64 ± 1.81	95.59 ± 1.29	70.88 ± 2.05	74.67 ± 2.33	68.02 ± 2.00	70.13 ± 1.98	93.12 ± 1.05	95.1 ± 0.73
	K	57.4 ± 4.30	58.13 ± 4.79	84.11 ± 4.36	89.72 ± 4.67	62.62 ± 4.40	67.73 ± 5.34	53.14 ± 3.60	55.18 ± 2.90	89.38 ± 1.72	92.52 ± 1.57
11	OA	67.41 ± 3.76	67.96 ± 4.40	91.29 ± 3.54	92.76 ± 3.5	73.91 ± 4.36	78.40 ± 3.80	65.51 ± 2.74	67.16 ± 2.68	94.01 ± 1.11	95.61 ± 0.97
	AA	76.42 ± 1.67	76.77 ± 1.68	94.94 ± 1.25	96.10 ± 1.16	73.54 ± 2.19	77.11 ± 2.18	71.87 ± 2.01	73.1 ± 2.37	94.85 ± 0.81	96.09 ± 0.78
	K	59.15 ± 4.07	59.85 ± 4.68	88.76 ± 4.34	90.65 ± 4.33	66.21 ± 4.96	71.88 ± 4.49	57.47 ± 2.89	59.35 ± 2.90	92.06 ± 1.44	94.19 ± 1.27
13	OA	70.58 ± 3.44	70.54 ± 3.18	91.76 ± 2.86	94.88 ± 3.11	75.58 ± 4.09	80.32 ± 3.08	68.37 ± 3.40	69.25 ± 3.26	95.05 ± 0.87	96.58 ± 0.92
	AA	77.83 ± 1.48	78.30 ± 1.42	95.19 ± 1.29	96.92 ± 1.10	75.37 ± 2.13	79.49 ± 1.42	74.52 ± 1.79	75.49 ± 1.82	95.57 ± 0.66	96.93 ± 0.55
	K	62.76 ± 3.77	62.83 ± 3.49	89.33 ± 3.57	93.35 ± 3.89	68.38 ± 4.75	74.42 ± 3.57	60.73 ± 3.68	61.83 ± 3.57	93.44 ± 1.15	95.47 ± 1.20

Table 3. Classification accuracies for different features and ensemble approach on Salinas.

N		Spectral		EMAP		Gabor		LBP		Ensemble	
		Standard	IR	Standard	IR	Standard	IR	Standard	IR	Standard	IR
3	OA	78.95 ± 3.12	79.24 ± 3.21	82.90 ± 3.41	86.20 ± 3.24	40.93 ± 2.72	41.42 ± 2.54	56.18 ± 2.95	57.15 ± 2.72	84.01 ± 2.02	85.73 ± 2.20
	AA	86.12 ± 2.36	86.26 ± 2.17	90.62 ± 1.88	92.18 ± 1.98	48.92 ± 2.73	49.77 ± 2.42	63.91 ± 2.86	64.50 ± 2.96	91.80 ± 1.03	92.56 ± 0.92
	K	76.66 ± 3.42	77.00 ± 3.49	81.03 ± 3.73	84.70 ± 3.57	36.27 ± 2.74	36.83 ± 2.61	51.55 ± 3.16	52.51 ± 2.88	82.26 ± 2.21	84.17 ± 2.40
5	OA	81.52 ± 2.64	81.91 ± 2.79	86.45 ± 3.03	89.44 ± 2.39	48.27 ± 2.17	48.93 ± 2.57	63.80 ± 2.07	64.89 ± 2.38	87.46 ± 1.67	88.98 ± 1.76
	AA	88.93 ± 1.56	89.28 ± 1.65	92.86 ± 1.38	94.58 ± 1.22	57.01 ± 2.09	58.02 ± 2.33	72.22 ± 1.72	73.33 ± 2.09	94.20 ± 0.67	94.89 ± 0.75
	K	79.54 ± 2.87	79.95 ± 3.05	84.96 ± 3.32	88.72 ± 2.65	44.00 ± 2.25	44.70 ± 2.68	59.91 ± 2.24	61.18 ± 2.54	86.08 ± 1.84	87.76 ± 1.94
7	OA	82.89 ± 2.50	84.21 ± 2.15	87.46 ± 2.78	91.25 ± 1.74	53.82 ± 2.22	54.85 ± 1.96	68.76 ± 2.00	69.31 ± 1.77	88.99 ± 1.33	91.34 ± 1.08
	AA	90.19 ± 1.36	90.92 ± 1.13	93.67 ± 1.15	95.57 ± 0.76	63.17 ± 1.95	64.34 ± 1.88	77.13 ± 1.73	78.30 ± 1.6	95.09 ± 0.54	96.19 ± 0.48
	K	81.02 ± 2.73	82.47 ± 2.36	86.08 ± 3.06	90.2 ± 1.93	49.90 ± 2.35	51.05 ± 2.02	65.41 ± 2.15	66.09 ± 1.91	87.76 ± 1.47	90.38 ± 1.19
9	OA	84.41 ± 1.99	85.09 ± 1.70	88.49 ± 3.21	92.43 ± 2.41	58.28 ± 1.92	59.08 ± 1.91	71.53 ± 1.88	72.42 ± 2.10	90.8 ± 1.47	92.46 ± 1.25
	AA	91.18 ± 0.95	91.70 ± 0.90	94.53 ± 0.99	96.12 ± 0.98	67.38 ± 1.62	68.42 ± 1.65	79.99 ± 1.46	81.33 ± 1.63	95.99 ± 0.49	96.78 ± 0.47
	K	82.70 ± 2.17	83.46 ± 1.86	87.23 ± 3.52	91.79 ± 2.67	54.63 ± 2.02	55.55 ± 2.00	68.53 ± 2.03	69.53 ± 2.29	89.53 ± 1.62	91.8 ± 1.39
11	OA	85.12 ± 1.92	85.38 ± 1.87	90.18 ± 1.92	92.88 ± 1.83	61.74 ± 1.49	63.02 ± 2.08	74.03 ± 1.67	74.86 ± 1.48	91.99 ± 1.06	92.92 ± 0.95
	AA	91.90 ± 0.91	92.19 ± 0.84	95.16 ± 0.78	96.33 ± 0.82	70.78 ± 1.33	72.35 ± 1.84	82.68 ± 1.26	83.80 ± 1.28	96.63 ± 0.41	97.04 ± 0.39
	K	83.49 ± 2.11	83.77 ± 2.05	89.09 ± 2.12	92.09 ± 2.03	58.31 ± 1.60	59.75 ± 2.21	71.29 ± 1.82	72.22 ± 1.61	91.09 ± 1.18	92.12 ± 1.05
13	OA	85.81 ± 1.38	85.96 ± 1.65	91.09 ± 1.78	93.6 ± 1.31	64.56 ± 1.63	66.7 ± 1.65	76.09 ± 1.50	76.97 ± 1.27	92.79 ± 0.81	93.81 ± 0.81
	AA	92.40 ± 0.75	92.68 ± 0.82	95.65 ± 0.70	96.64 ± 0.60	73.59 ± 1.34	75.52 ± 1.66	84.40 ± 0.97	85.31 ± 1.08	96.99 ± 0.32	97.38 ± 0.36
	K	84.26 ± 1.52	84.41 ± 1.82	90.11 ± 1.97	92.89 ± 1.45	61.35 ± 1.72	63.69 ± 1.76	73.54 ± 1.64	74.52 ± 1.36	91.98 ± 0.89	93.11 ± 0.90

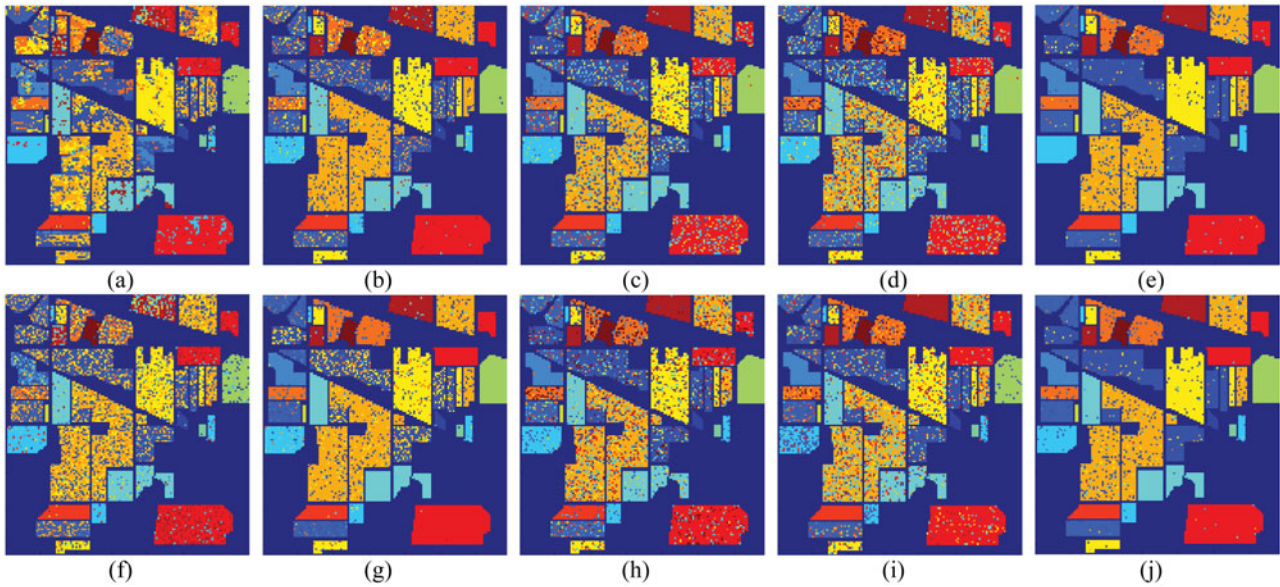


Fig. 5. Overall accuracies of IRCK, DMKL, and the proposed method for three datasets.

University of Pavia dataset. The IR kernel also has better performance for the spectral feature and LBP feature. Moreover, the ensemble approach using the IR kernel has the best performance, indicating the advantage of combining complementary discriminative information. It can be observed that the ensemble approach has a smaller variance compared with using a specific feature alone. This is expected because by combining different types of features, a more robust result can be achieved.

For the results on Salinas as shown in Table 3, several conclusions can be drawn. First, it can be concluded that IR can provide better performance compared with the standard kernel for all types of features. For this dataset, ensemble approaches have slightly better performance compared with using the EMAP feature alone. Gabor feature does not have a good performance on this dataset, including the Gabor feature for this dataset can affect the accuracy. We also notice that the regularized kernel has poor performance if the standard kernel has poor performance. In addition, with only seven and 11 training samples per class for Indian Pines, University of Pavia, and Salinas, the OA can be above 90%. The explanation maybe that when the number of training samples is limited, the similarity measurement between samples is not reliable, and thus, the label similarity between different classes can learn a desirable kernel metric, which is essential for good classification performance. Another conclusion is that the ensemble approach provides good performance with a limited number of training samples.

Figure 5 compares the OAs of IRCK, DMKL, and our proposed method using 15 samples per class. It can be concluded that the proposed method has superior performance over IRCK on three different datasets. More specifically, the proposed method has around 6.1, 4.7, and 1.4% higher OA than IRCK for Indian Pines, University of Pavia, and Salinas, respectively. Compared with DMKL, the proposed

method has around 7 and 4% higher OA than DMKL on Indian Pines and the University of Pavia, respectively. On the Salinas dataset, the proposed method slightly degrades from DMKL. In Fig. 6, the classification maps of Indian Pines are shown for four different types of features. The proposed method has around 0.6, 4, and 1.5% higher OA for Indian Pines, University of Pavia, and Salinas, respectively. The observation is that few misclassified samples exist in the classification map of the ensemble approaches. This is expected since different features can provide discriminative information, which can lead to better performance.

D) Parameters analysis

The parameters in this paper are analyzed in this section. For the SVM, we will first analyze the RBF kernel σ and the regularization term C first. For σ , it is chosen from the range $\{2^{-4}, 2^{-3}, \dots, 2^4\}$. C is chosen in the range $\{1, 10, 100, 10^5\}$. Figures 7(a)–7(c) show the OA with different σ and C for EMAP features on Indian Pines, University of

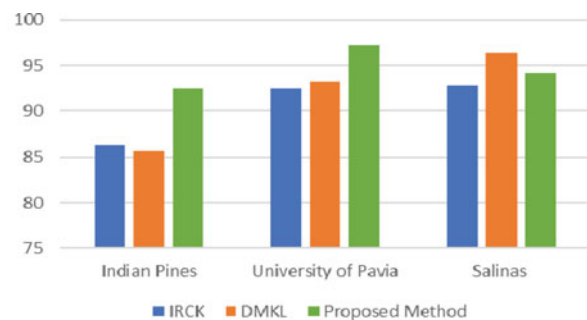


Fig. 6. Classification maps for Indian Pines. The first and second rows correspond to Standard and IR kernels. (a) Spectral-Sta(OA = 63.31%). (b) EMAP-Sta(OA = 83.9%). (c) Gabor-Sta(OA = 77.8%). (d) LBP-Sta(OA = 73.2%). (e) Ensemble-Sta(OA = 92.4%). (f) Spectral-IR(OA = 64.4%). (g) EMAP-IR(OA = 86.4%). (h) Gabor-IR(OA = 78.1%). (i) LBP-IR(OA = 76.8%). (j) Ensemble-IR(OA = 93.8%).

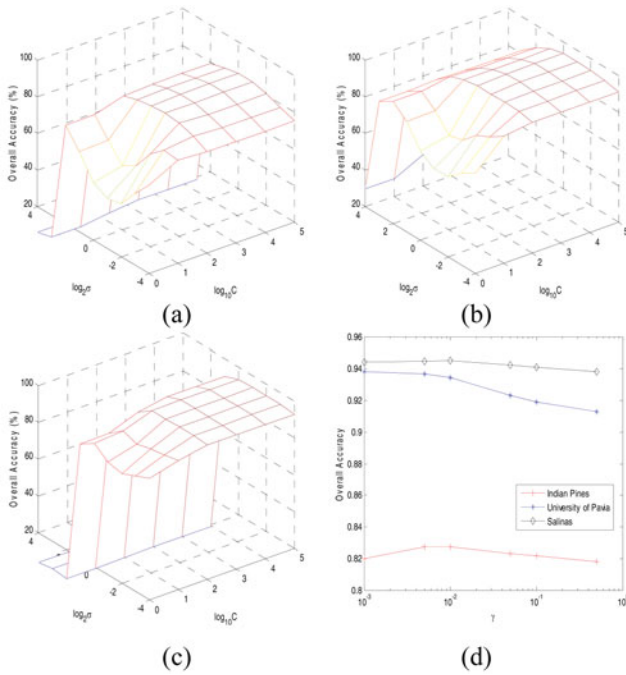


Fig. 7. Parameters tuning using EMAP feature. (a) σ and C for Indian Pines. (b) σ and C for the University of Pavia. (c) σ and C for Salinas. (d) γ for all datasets.

Pavia, and Salinas, respectively, using 13 training samples per class. A holdout validation sample set is used for testing. It can be concluded from Fig. 7 that C does not affect the OA much especially when it is larger than 100, and any number above 100 will guarantee good performance. For σ , a relatively small number will generate good classification performance. In order to analyze the effect of the γ parameter in IR, in the following experiment, C is set as 1000, and σ is chosen as 2^{-1} for the EMAP features of different datasets to ensure good performance.

The parameter γ plays an important role in the classification performance. We investigate the effect of γ on the EMAP feature for Indian Pines, University of Pavia, and Salinas, respectively. Figure 7(d) shows the OAs of the three datasets with respect to different γ in the range $\{1e^{-3}, 5e^{-3}, \dots, 5e^{-1}\}$. It can be seen that the OA stays stable when γ is very small, and when γ increases, the OAs of all three datasets will decrease. This may be due to the reason that the label information is dominant in the generated kernel while ignoring the spectral similarity between the training samples. In our experiment, we set C as 1000, σ as 2^{-1} , and γ as $5e^{-3}$ for the EMAP features for three datasets. For other types of features, a similar parameter tuning process is adopted.

V. CONCLUSIONS

In this paper, a novel multiple feature-based IR kernel is presented for HSI classification. The proposed framework incorporates the label information in conjunction with the complementary discriminative information with multiple

features. Experimental results show our proposed approach can achieve superior classification performance.

ACKNOWLEDGEMENT

The authors would like to thank Prof. D. Landgrebe for providing the Indian Pines data set, Prof. P. Gamba for providing the University of Pavia data set, and Dr. J. Anthony Gualtieri for providing Salinas data set.

FINANCIAL SUPPORT

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61871177 and 11771130.

REFERENCES

- [1] Plaza A. et al.: Recent advances in techniques for hyperspectral image processing. *Remote Sens. Environ.*, **113** (2009), S110–S122.
- [2] Xu Y.; Du Q.; Li W.; Chen C.; Younan N.H.: Nonlinear classification of multispectral imagery using representation-based classifiers. *Remote Sens.*, **9** (7) (2017), 662.
- [3] Lillesand T.M.; Kiefer R.W.: Remote Sensing and Image Interpretation, vol. 4, Wiley, Hoboken, NJ, USA, 2015.
- [4] Bioucas-dias J.M.; Plaza A.; Camps-valls G.; Scheunders P.; Nasrabadi N.M.; Chanussot J.: Hyperspectral remote sensing data analysis and future challenges. *IEEE Geosci. Remote Sens. Mag.*, **1** (2) (2013), 6–36.
- [5] Nasrabadi N.M.: Hyperspectral target detection: an overview of current and future challenges. *IEEE Signal Process. Mag.*, **31** (1) (2014), 34–44.
- [6] Mei S.; Ji J.; Geng Y.; Zhang Z.; Li X.; Du Q.: Unsupervised spatial-spectral feature learning by 3D convolutional autoencoder for hyperspectral classification. *IEEE Trans. Geosci. Remote Sens.*, **57** (9) (2019), 6808–6820.
- [7] Yang H.; Du Q.; Chen G.: Particle swarm optimization-based hyperspectral dimensionality reduction for urban land cover classification. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, **5** (2) (2012), 544–554.
- [8] Liu W.; Fowler J.E.; Zhao C.: Spatial logistic regression for support-vector classification of hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.*, **14** (3) (2017), 439–443.
- [9] Gan L.; Du P.; Xia J.; Meng Y.: Kernel fused representation-based classifier for hyperspectral imagery. *IEEE Geosci. Remote Sens. Lett.*, **14** (5) (2017), 684–688.
- [10] Kuo B.-C.; Ho H.-H.; Li C.-H.; Hung C.-C.; Taur J.-S.: A kernel-based feature selection method for SVM With RBF kernel for hyperspectral image classification. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, **7** (1) (Jan. 2014), 317–326.
- [11] Melgani F.; Bruzzone L.: Classification of hyperspectral remote sensing. *IEEE Trans. Geosci. Remote Sens.*, **42**, (2004), 1778–1790.
- [12] Camps-Valls G.; Bruzzone L.: Kernel-based methods for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, **43** (6) (2005), 1351–1362.
- [13] Peng J.; Zhou Y.; Chen C.L.P.: Region-kernel-based support vector machines for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, **53** (9) (2015), 4810–4824.

- [14] Li J.; Marpu P.R.; 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.
- [15] Gomez-Chova L.; Camps-Valls G.; Bruzzone L.; Calpe-Maravilla J.: Mean map kernel methods for semisupervised cloud classification. *IEEE Trans. Geosci. Remote Sens.*, **48** (1) (2010), 207–220.
- [16] Sun W.; Liu C.; Xu Y.; Tian L.; Li W.: A band-weighted support vector machine method for hyperspectral imagery classification. *IEEE Geosci. Remote Sens. Lett.*, **14** (10) (2017), 1710–1714.
- [17] Li W.; Du Q.; Xiong M.: Kernel collaborative representation with Tikhonov regularization for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.*, **12** (1) (2015), 48–52.
- [18] Ma M.; Mei S.; Wan S.; Hou J.; Wang Z.; Feng D.D.: Video summarization via block sparse dictionary selection. *Neurocomputing*, **378** (2020), 197–209.
- [19] Kwok J.T.; Tsang I.W.: Learning with idealized kernels, in *Proc. Int. Conf. Mach. Learn.*, 2003, 400–407.
- [20] Pan B.; Lai J.; Shen L.: Ideal regularization for learning kernels from labels. *Neural Netw.*, **56** (2014), 22–34.
- [21] Pan B.; Chen W.S.; Xu C.; Chen B.: A novel framework for learning geometry-aware kernels. *IEEE Trans. Neural Netw. Learn. Syst.*, **27** (5) (2016), 939–951.
- [22] Peng J.; Chen H.; Zhou Y.; Li L.: Ideal regularized composite kernel for hyperspectral image classification. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, **10** (4) (2017), 1563–1574.
- [23] Mehmet G.; Alpayd E.: Multiple kernel learning algorithms. *J. Mach. Learn. Research*, **12** (2011), 2211–2268.
- [24] Camps-Valls G.; Gomez-Chova L.; Muñoz-Mari J.; Vila-Francés J.; Calpe-Maravilla J.: Composite kernels for hyperspectral image classification. *IEEE Geosci. Remote Sens. Lett.*, **3** (1) (2006), 93–97.
- [25] Wang Q.; Gu Y.; Tuia D.: Discriminative multiple kernel learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, **54** (7) (2016), 3912–3927.
- [26] Li W.; Du Q.: Gabor-filtering-based nearest regularized subspace for hyperspectral image classification. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, **7** (4) (2014), 1012–1022.
- [27] Li W.; Chen C.; Su H.; Du Q.: Local binary patterns and extreme learning machine for hyperspectral imagery classification. *IEEE Trans. Geosci. Remote Sens.*, **53** (7) (2015), 3681–3693.
- [28] Dalla Mura M.; Atli Benediktsson J.; Waske B.; Bruzzone L.: Extended profiles with morphological attribute filters for the analysis of hyperspectral data. *Int. J. Remote Sens.*, **31** (22) (2010), 5975–5991.
- [29] Li J. *et al.*: Multiple feature learning for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.*, **53** (3) (2015), 1592–1606.
- [30] Li J.; Zhang H.; Zhang L.: A nonlinear multiple feature learning classifier for hyperspectral images with limited training samples. *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, **8** (6) (2015), 2728–2738.
- [31] Zhang L.; Zhang L.; Tao D.; Huang X.: On combining multiple features for hyperspectral remote sensing image classification. *IEEE Trans. Geosci. Remote Sens.*, **50** (3) (2012), 879–893.
- [32] Rodarmel C.; Shan J.: Principal component analysis for hyperspectral image classification. *Inf. Syst.*, **62** (2) (2002), 115–115.
- [33] Chang C.-C.; Lin C.-J.: LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, **2** (3) (2011), 1–27.

Yan Xu received his Ph.D. degree in 2019 from the Department of Electrical and Computer Engineering, Mississippi State University, USA.

Jiangtao Peng received the B.S. and M.S. degrees from Hubei University, Wuhan, China, in 2005 and 2008, and his Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, China, in 2011. He is currently a Professor in the Faculty of Mathematics and Statistics, Hubei University, China. His research interests include machine learning and hyperspectral image processing.

Qian Du received the Ph.D. degree in electrical engineering from the University of Maryland – Baltimore County, Baltimore, MD, USA, in 2000. She is currently a Bobby Shackouls Professor with the Department of Electrical and Computer Engineering, Mississippi State University, Starkville, MS, USA. Her research interests include hyperspectral remote-sensing image analysis and applications, pattern classification, data compression, and neural networks. Dr. Du is a fellow of SPIE and IEEE. Since 2016, she has been the Editor-in-Chief of the IEEE JSTARS.