Kernel Fused Representation-Based Classifier for Hyperspectral Imagery

Le Gan, Peijun Du, *Senior Member, IEEE*, Junshi Xia, *Member, IEEE*, and Yaping Meng

***Abstract*— In this letter, we propose a kernel fused representation-based classifier (KFRC) for hyperspectral images (HSIs), which combines sparse representation (SR) and collaborative representation (CR) into a unified kernel representation-based classification framework. First, we present two individual kernel methods, i.e., kernel SR (KSR) and kernel CR (KCR), which kernelize the representation methods by projecting the samples into a high-dimensional kernel space to improve the samples separability between different classes. Once obtaining the two kernel representation coefficients, KFRC attempts to achieve a balance between KSR and KCR via an adjusting parameter θ in the kernel residual domain. Subsequently, the class label of each test sample is determined by the minimum residual for each class. Experimental results on two HSIs demonstrate the proposed kernel fused method performs better than the other state-of-the-art representation-based classifiers.**

***Index Terms*— Classifier fusion, collaborative representation (CR), hyperspectral image (HSI) classification, kernel trick, sparse representation (SR).**

I. Introduction

CLASSIFICATION is a consistent topic for hyperspectral image (HSI) analysis [1]–[5]. Numerous classifiers, such as support vector machines [6] and multinomial logistic regression [7], have been introduced to obtain a satisfactory accuracy for HSI classification over the past years. Recently, without considering any prior sample distribution, representation-based classifiers [8] have drawn much attention in pattern classification. As a nonparametric learning algorithm, representation-based methods can directly assign a class label to a test sample based on a structured dictionary [9], [10]. The primary thought of such types of classifiers is that the training dictionary can linearly approximate a test sample. The weight coefficients of those classifiers convey critical information that reflects the importance of different dictionary atoms. Based on the type of constraints, representation-based methods can be classified into two broad categories: sparse representation (SR), which can be solved by an ℓ_1-minimization problem, and collaborative representation (CR), corresponding to an ℓ_2-minimization problem.

SR-based classifier (SRC), originally proposed in [11] for face recognition, has been broadly applied to various kinds of HSI applications. SRC is chiefly based on the assumption that a given sample can be compactly represented by a few dictionary atoms that carry the essential information [12]. Considering the expensive computational cost of ℓ_1-norm regularization methods for high-dimensional classification tasks, Zhang *et al.* [13], [14] argued that it is not necessary to regularize the sparse coefficient with ℓ_1-norm and proposed a CR-based classifier (CRC), which is similar to SR-based method. CRC uses a linear combination of training dictionary from all the classes to reconstruct a given sample, rather than only a few atoms. In [13], the CRC is solved by an ℓ_2-norm regularized linear regression problem and a very competitive accuracy with a significantly lower complexity was achieved. Inspired by the high computational efficiency and desirable performance, the CRCs have been successfully applied to HSI classification [15], [16].

Although the individual representation-based methods have exhibited good performance for hyperspectral classification, one single kind of method can only depict the discrimination of HSI from one aspect. In the SR-based method, all atoms competitively participate in the presentation process of a given sample, whereas all atoms have an equal chance to represent a given sample in the CR-based method. Actually, information fusion technique has been broadly discussed to combine multiple pieces of information for improved performance. Zhang *et al.* [17] concatenated multiple features and proposed a multifeature SRC in the feature-level. Li *et al.* [8] combined SR- and CR-based methods via a proper weight in the residual domain and proposed a fusion-based classifier. Likewise, an elastic net representation-based classifier (ENRC) was discussed in [8] to overcome the indigenous disadvantages of SR- and CR-based methods.

As we all know, kernel methods, which project samples into a high-dimensional feature space by a nonlinear mapping, yield a significant performance improvement, because the kernel-based methods implicitly exploit the complex structure of the given sample that may not be captured by the linear model [18]. Gao *et al.* [19] proposed a kernel SRC (KSRC) to capture the nonlinear similarity of features by a kernel trick. In this letter, we propose a kernel fused

Manuscript received October 24, 2016; revised January 6, 2017; accepted January 30, 2017. Date of publication March 14, 2017; date of current version April 20, 2017. This work was supported in part by the Natural Science Foundation of China under Grant 41471275 and in part by the Key Scientific Instrument Development Foundation of China under Grant 2012YQ05025004. *(Corresponding author: Peijun Du.)*

L. Gan, P. Du, and Y. Meng are with the Key Laboratory for Satellite Mapping Technology and Applications of National Administration of Surveying, Mapping and Geoinformation of China, Nanjing University, Nanjing 210023, China, and also with the Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China, and also with the Collaborative Innovation Center of South China Sea Studies, Nanjing 210093, China (e-mail: ganleatlas@gmail.com; dupjrs@gmail.com; ypmengnju@126.com).

J. Xia is with the Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo 113-8654, Japan (e-mail: xiajunshi@gmail.com).

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

Digital Object Identifier 10.1109/LGRS.2017.2671852

representation-based classifier (KFRC) for HSI based on two individual representation methods by assuming that a test sample can be linearly approximated by the given dictionary atoms in the kernel feature space. For KFRC, samples are mapped into a high-dimensional feature space first and then two different recovered coefficients of the test sample are obtained by solving ℓ_1- and ℓ_2-minimization problems based on a structured dictionary consisting of different atoms from all of the classes in this new feature space. In order to make the representation more suitable, a new fused residual is applied by combining two single residuals between the test sample and the corresponding recovery for each class with the two different representation methods. Subsequently, the class label for the test sample is determined by minimum kernel fused residual for each class. The validity of the proposed KFRC is verified by experiments on two classical HSIs.

The remainder of this letter is organized as follows. The proposed kernel fused-based classifier is described in Section II. The effectiveness of the kernel fused classifier is demonstrated on two HSIs in Section III. Finally, Section IV summarizes this letter and makes some concluding remarks.

II. Kernel Fused Representation Classifier

A. Representation-Based Classifiers

First, we review two classical representation-based classifiers (i.e., SRC and CRC). Given a training dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_N] \in \mathbb{R}^{B\times N}$ (B is the number of bands) consisting of N atoms and class labels $\mathbf{z}_i \in \{1, 2, \ldots, C\}$, where C denotes the number of classes. SRCs [11] assume the signals belonging to the same class span the same low-dimensional subspace. A given signal $\mathbf{y} \in \mathbb{R}^B$ is approximately reconstructed by a weighted linear combination of a small number of support atoms over the training dictionary, which formulated as the following ℓ_1-minimization problem:

$$\hat{\boldsymbol{\alpha}}^{SRC} = \arg\min_{\alpha^{\text{SRC}}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}^{SRC}\|_2 + \lambda_1\|\boldsymbol{\alpha}^{SRC}\|_1 \tag{1}$$

where $\lambda_1 > 0$ is a sparsity regularization parameter. Once obtaining the sparse coefficient $\boldsymbol{\alpha}^{SRC}$, the label $\mathbf{y}$ is determined by the minimal residual between $\mathbf{y}$ and its approximation from each subdictionary

$$\text{class}(\mathbf{y}) = \arg\min_{i=1,\ldots,C} \|\mathbf{y} - \mathbf{D}\boldsymbol{\delta}_i(\boldsymbol{\alpha}^{SRC})\|_2 \tag{2}$$

where $\boldsymbol{\delta}$ is the characteristic function [11] that chooses coefficients related with the i-class and makes the rest to zero.

Similar to the SRC, in CRC [13], all the atoms are adopted to reconstruct the test signal $\mathbf{y}$, which can be formulated by the following ℓ_2-minimization problem:

$$\hat{\boldsymbol{\alpha}}^{CRC} = \arg\min_{\boldsymbol{\alpha}^{CRC}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}^{CRC}\|_2 + \lambda_2\|\boldsymbol{\alpha}^{CRC}\|_2 \tag{3}$$

where λ_2 is a positive parameter to balance the representation error and regularization term, and the corresponding closed-form solution of (3) is given by

$$\hat{\boldsymbol{\alpha}}^{CRC} = (\mathbf{D}^{\mathbf{T}}\mathbf{D} + \lambda_2\mathbf{I})^{-1}\mathbf{D}^{\mathbf{T}}\mathbf{y}. \tag{4}$$

Once obtaining the CR coefficient $\boldsymbol{\alpha}^{CRC}$, the residual $r_i^{CRC} = \|\mathbf{y} - \mathbf{D}\boldsymbol{\delta}_i(\boldsymbol{\alpha}^{CRC})\|_2$ is computed similar to SRC for class label assignment.

B. Kernel Tricks

As we all know, through suitable kernel functions that reflect the similarity among samples, a linearly inseparable sample in original space can become linearly separable in a high-dimensional kernel space [20], [21]. Let Φ define a nonlinear map corresponding to a kernel function $k(\cdot,\cdot) : \mathbb{R}^B \times \mathbb{R}^B \mapsto \mathbb{R}$. Here, we adopt the Gaussian radial basis function (RBF) kernel $k(\mathbf{x}, \mathbf{y}) = \mathbf{exp}(-\gamma\|\mathbf{x} - \mathbf{y}\|_2^2)$ ($\gamma > 0$ is the parameter of the RBF kernel), and the SR-based problem of (1) in the kernel feature space can be rewritten as

$$\hat{\boldsymbol{\alpha}}^{KSRC} = \arg\min_{\boldsymbol{\alpha}^{KSRC}} \|\Phi(\mathbf{y}) - \Phi(\mathbf{D})\boldsymbol{\alpha}^{KSRC}\|_2 + \lambda_1\|\boldsymbol{\alpha}^{KSRC}\|_1 \tag{5}$$

where $\Phi(\mathbf{y}) = [k(\mathbf{d}_1, \mathbf{y}), k(\mathbf{d}_2, \mathbf{y}), \ldots, k(\mathbf{d}_N, \mathbf{y})]^T \in \mathbb{R}^{N\times 1}$, and $\Phi(\mathbf{D}) \in \mathbb{R}^{N\times N}$ is the training kernel matrix with $\Phi(\mathbf{D})_{i,j} = k(\mathbf{d}_i, \mathbf{d}_j)$. Similarly, the class label of $\mathbf{y}$ is given as

$$\text{class}(\mathbf{y}) = \arg\min_{i=1,\ldots,C} \|\Phi(\mathbf{y}) - \Phi(\mathbf{D})\boldsymbol{\delta}_i(\boldsymbol{\alpha}^{KSRC})\|_2. \tag{6}$$

Likewise, kernel CR-based representation can be formulated as

$$\hat{\boldsymbol{\alpha}}^{KCRC} = \arg\min_{\boldsymbol{\alpha}^{KCRC}} \|\Phi(\mathbf{y}) - \Phi(\mathbf{D})\boldsymbol{\alpha}^{KCRC}\|_2 + \lambda_2\|\boldsymbol{\alpha}^{KCRC}\|_2 \tag{7}$$

and the corresponding closed-form solution is given as

$$\hat{\boldsymbol{\alpha}}^{KCRC} = (\Phi(\mathbf{D}) + \lambda_2\mathbf{I})^{-1}\Phi(\mathbf{y}). \tag{8}$$

After obtaining $\boldsymbol{\alpha}^{KCRC}$, the class label of a given sample $\mathbf{y}$ is determined according to the minimum residual $\gamma_i^{KCRC}(\mathbf{y}) = \|\Phi(\mathbf{y}) - \Phi(\mathbf{D})\boldsymbol{\delta}_i(\boldsymbol{\alpha}^{KCRC})\|_2$.

Algorithm 1 KFRC Method

Input: Available training dictionary $\{\mathbf{d}_i, \mathbf{z}_i\}_{i=1}^N$ for C class, $\mathbf{d}_i \in \mathbb{R}^B$ denotes the dictionary atom with class label $\mathbf{z}_i \in \{1, 2, \ldots, C\}$, test samples $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_M\} \in \mathbb{R}^{B\times M}$, regularization parameters λ_1, λ_2 and balancing parameter θ.

Output: An 2-D map which records the labels of the test samples $\mathbf{Y}$

Select a mercer kernel function $k(\cdot,\cdot)$ and its parameters and compute the kernel gram matrix $\Phi(\mathbf{D})$, where $\Phi(\mathbf{D})_{ij} = k(\mathbf{d}_i, \mathbf{d}_j)$.

for $i = 1, \ldots, M$ **do**

1) Calculate $\mathbf{k}(\cdot, \mathbf{y}_i) = [k(\mathbf{d}_1, \mathbf{y}_i), \ldots, k(\mathbf{d}_N, \mathbf{y}_i)]$

2) Solve the ℓ_1-minimization problem (5) to get the kernel sparse coefficient vector $\boldsymbol{\alpha}^{KSRC}$.

3) Solve the ℓ_2-minimization problem (7) according to (8) to obtain kernel collaborative weight vector $\boldsymbol{\alpha}^{KCRC}$.

4) Calculate two individual residuals $r_j^{KSRC}(\mathbf{y}_i)$ and $r_j^{KCRC}(\mathbf{y}_i)$, and the fused residual $r_j^{KFRC}(\mathbf{y}_i)$ is obtained using (9).

5) Decide the final label according to $class(\mathbf{y}_i) = \arg\min_{j=1,\ldots,C} r_j^{KFRC}(\mathbf{y}_i)$.

end for

return The estimated label $class(\mathbf{y_i})$.

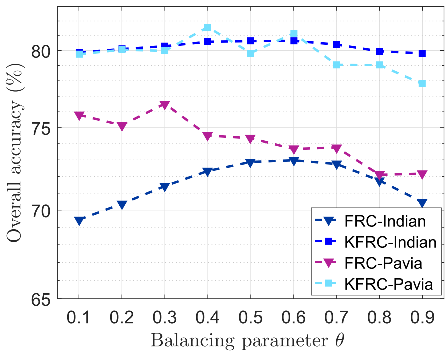

Fig. 1. OA(%) results versus different balancing parameters θ for the proposed FRCs.

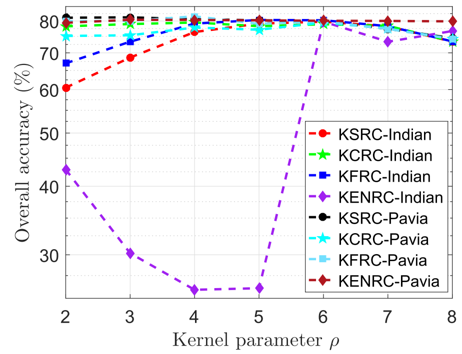

Fig. 2. OA(%) results versus different kernel parameters ρ for all kernel representation-based classifiers.

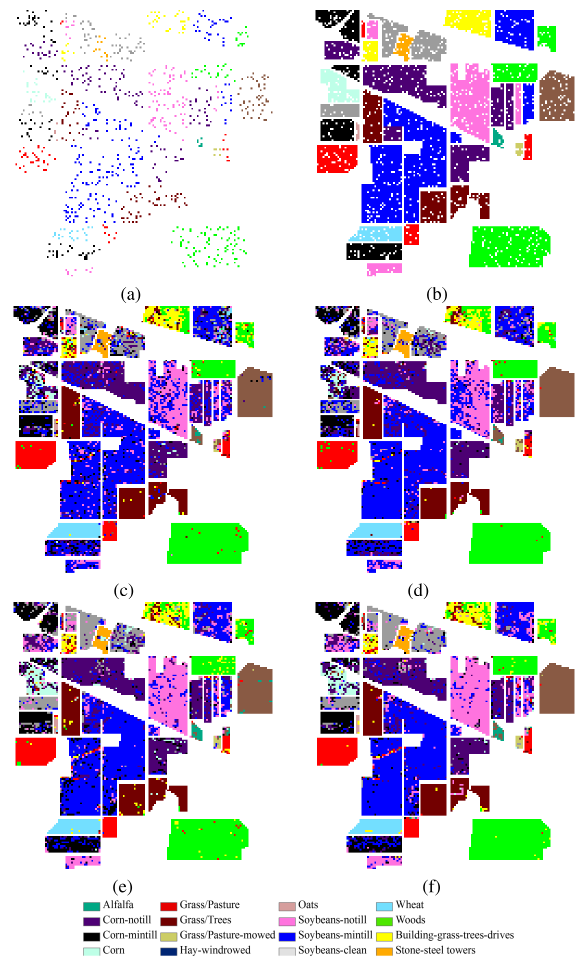

Fig. 3. Classification maps obtained by several different representation-based classifiers for the AVIRIS Indian Pines image. (a) Training samples. (b) Testing samples. (c) FRC. (d) ENRC. (e) KENRC. (f) KFRC.

C. Proposed KFRC

Li *et al.* [8] have confirmed that the classification performance was improved by SRC, while CRC brought the gain in other cases. All the dictionary atoms collaborate on the approximation of a given signal in the CR-based model with an equal chance, whereas only a few dictionary atoms have the opportunity to participate in the signal reconstruction in the SR-based model. However, if the HSI is not linearly separable, the performance of FRC may decrease. As a kernel extension of FRC [8], KFRC adopts the kernel trick to map spectral feature into a high dimensional kernel space, then two individual representation coefficients in the new feature space are obtained by solving the corresponding ℓ_1- (5) and ℓ_2-minimization problems (7), respectively. After obtaining the two coefficients $\boldsymbol{\alpha}^{KSRC}$ and $\boldsymbol{\alpha}^{KCRC}$, the individual residuals r_i^{KSRC} and r_i^{KCRC} are computed similar to SRC. For a more suitable representation in the kernel space, the fused residual $\gamma_i^{\mathrm{KFRC}}(\mathbf{y})$ can be formulated as

$$\gamma_i^{KFRC}(\mathbf{y}) = (1-\theta)\gamma_i^{KSRC}(\mathbf{y}) + \theta\gamma_i^{KCRC}(\mathbf{y}) \tag{9}$$

where θ ($0 \leq \theta \leq 1$) is a control parameter that makes a tradeoff between kernel SR- and kernel CR-based methods. Then the class label of $\mathbf{y}$ is determined according to the class with the smallest residual. Obviously, if $\theta = 0$, the fused method reduces to KSRC, and if $\theta = 1$, the fused method reduces to KCRC . The pseudocode description of the proposed KFRC is summarized in Algorithm 1.

III. Experimental Results and Discussion

We examine the proposed KFRC algorithm on two classical hyperspectral data sets, e.g., Indian Pines and the University of Pavia images. For the Indian Pines image, the scene covers a mixed agricultural/forest area, and the size is $145 \times 145 \times 200$ with a spatial resolution of 20 m. It contains 16 different classes, 10% of the labeled samples is randomly chosen as the training samples, and the rest 90% for test samples [the training and test samples are visually given in Fig. 3(a) and (b), respectively]. For the University of Pavia image, the scene covering an urban area was obtained by the ROSIS-03 sensor, and the size is $610 \times 340 \times 103$ with a spatial resolution of 1.3 m. It contains nine class, 1% samples for each class is selected as training samples, and the remaining 99% for test samples. It should be noted that RBF kernel is adopted in all kernel-based classifiers. FRC combines the SR- and CR-based representations in the residual domain via a balance parameter in spectral space. ENRC solves the ℓ_1-ℓ_2 minimization problem by elastic net model in spectral space. KENRC further explores elastic net model in the high dimensional feature space.

To obtain the optimal classification performance of various representation-based classifiers, we learn a set of the optimal parameters by cross-validation. In all the methods, the λ_1 regularization parameters for SR- (SRC and KSRC) and

TABLE I
OPTIMAL COMBINATION OF PARAMETERS FOR VARIOUS REPRESENTATION-BASED CLASSIFIERS ($L = 200$)

	Indian Pines Image				University of Pavia Image			
Parameters	ℓ_1	ℓ_2	θ	γ	ℓ_1	ℓ_2	θ	γ
SRC	$1e-3$	–	–	–	$1e-3$	–	–	–
CRC	–	$1e-5$	–	–	–	$1e-5$	–	–
FRC	$1e-4$	$1e-7$	0.6	–	$1e-3$	$1e-5$	0.3	–
ENRC	$1e-4$	$1e-7$	–	–	$1e-2$	$1e-5$	–	–
KSRC	$1e-5$	–	–	e^6/L	$1e-3$	–	–	e^3/L
KCRC	–	$1e-7$	–	e^4/L	–	$1e-5$	–	e^2/L
KFRC	$1e-5$	$1e-7$	0.6	e^5/L	$1e-3$	$1e-5$	0.4	e^4/L
KENRC	$1e-6$	$1e-7$	–	e^6/L	$1e-2$	$1e-5$	–	e^6/L

TABLE II
CLASSIFICATION ACCURACY OBTAINED BY DIFFERENT CLASSIFIERS FOR THE AVIRIS INDIAN PINES IMAGE WITH 10% TRAINING SAMPLES

Class	KNN	SRC	KSRC	CRC	KCRC	FRC	ENRC	KENRC	KFRC
1	0	35.00	42.50	2.50	47.50	12.50	30.00	**72.50**	42.50
2	49.65	55.81	69.59	65.59	72.75	69.82	65.67	70.82	**73.06**
3	31.59	40.43	74.70	42.70	78.85	44.85	44.98	65.86	**78.98**
4	1.41	35.68	46.95	15.49	47.42	25.35	32.39	**52.11**	46.95
5	54.48	81.38	91.72	88.05	**92.18**	86.67	85.29	88.97	91.72
6	93.46	95.13	**96.80**	90.87	96.04	95.89	94.52	94.06	96.35
7	26.09	**65.22**	60.87	0	60.87	47.83	52.17	60.87	**65.22**
8	99.77	95.81	98.84	99.53	99.07	**100**	98.37	96.74	99.07
9	0	50.00	**75.00**	0	**75.00**	25.00	37.50	62.50	**75.00**
10	63.89	59.54	73.83	51.66	73.60	51.31	55.31	**77.03**	75.54
11	74.72	77.87	78.13	72.07	76.14	**82.43**	72.73	75.92	79.42
12	21.16	41.95	71.16	52.43	**74.34**	56.18	49.44	69.48	73.60
13	85.87	94.02	**98.37**	96.74	**98.37**	98.37	95.65	96.74	**98.37**
14	**96.16**	94.87	94.01	95.57	92.70	95.66	93.57	93.48	94.70
15	0.58	35.16	53.03	42.07	**59.37**	40.92	49.86	50.14	57.64
16	74.07	87.65	91.36	83.95	90.12	90.12	**93.83**	81.48	91.36
OA	63.04	69.88	79.40	69.58	79.93	73.85	70.99	78.04	**81.09**
κ	57.05	65.27	76.48	64.83	77.13	69.69	66.58	75.00	**78.43**

fused-based (FRC, ENRC, KFRC, and KENRC) methods range from $1e-7$ to $1e-1$. Analogically, the λ_2 parameters for CR- (CRC, KCRC) and fused-based methods range from $1e-8$ to $1e-2$.

First, we evaluate the OA results of the presented fused-based classifiers with different balancing parameters θ. For fairness, λ_1 and λ_2 were fixed at the optimal values. Fig. 1 shows the influence of the balancing parameter. For the Indian Pines image, compared to FRC, KFRC achieves a relative stable OA result as θ vary along with the candidate set. When θ reach a small value, the "competitive" nature of SR term dominate in the fused method, as θ get larger, the “collaborative” nature makes a large role. It can be clearly observed that when $\theta = 0.6$, both the fused representation-based methods achieve the best OA results in the candidate set, which indicates the fused method achieve an optimal balance between the SR term and the CR term in the residual domain. For the University of Pavia Image, the optimal balancing parameter θ are 0.3 and 0.4 for FRC and KFRC, respectively.

Second, we examine how the kernel parameter ρ (where $\gamma = e^{\rho}/L$, $L = 200$) affects all the kernel representation-based methods. For the kernel-based methods, parameter ρ is varied in the range $\{2, 3, 4, 5, 6, 7, 8\}$, corresponding to γ in the range $\{e^2/L, \ldots, e^8/L\}$. Fig. 2 gives the sensitivity analysis of the kernel parameter. For the Indian Pines image, as we can clearly observed, the optimal γ parameters are e^6/L, e^4/L, e^5/L, and e^6/L for KSRC, KCRC, KFRC, and KENRC, respectively. Compared with the relatively stable performance to γ for KCRC in the domain, KSRC has much different classification results with various values of γ . Likewise, KFRC has an obvious OA fluctuation with different kernel parameters, and KENRC exhibits an irregular pattern in this domain. For the University of Pavia Image, the optimal γ parameters are e^3/L, e^2/L, e^4/L, and e^6/L for KSRC, KCRC, KFRC, and KENRC, respectively. Compared with Indian image, all the kernel classifiers have a relatively stable performance to γ for Pavia image in the domain. The learned optimal parameters of all the presented representation-based classifiers are listed in Table I.

Third, we evaluate the classification results achieved by the proposed kernel fused-based classifier with that obtained by the state-of-the-art methods. Following the parameter settings in Table I, we conduct each experiment 10 times, and the average classification results are presented in Tables II and III for the Indian Pines and the University of Pavia images, respectively. The corresponding classification maps for the Indian Pines image using FRC, ENRC, and the kernel methods, i.e., KENRC and KFRC are visually displayed in Fig. 3(c)–(f). For the Indian Pines image, the representation-based classifiers exceed the nearest neighbor method (KNN), the SR- (SRC and KSRC), CR- (CRC and KCRC), and the kernel-based classifiers (KSRC, KCRC, KFRC, and KENRC) outperform the nonkernel methods (SRC, CRC, FRC, and ENRC) about 10%. The fused classifiers (FRC and KFRC) usually have much better performance than individual representation-based classifiers. Moreover, the fused classifiers are superior to the classifiers solved by elastic net model, i.e., ENRC and KENRC. For the University of Pavia Image, CRC has a similar OA results with KNN, SRC has a better OA results than CRC, and the fused methods (FRC and ENRC)

TABLE III

CLASSIFICATION ACCURACY OBTAINED BY DIFFERENT CLASSIFIERS FOR THE UNIVERSITY OF PAVIA IMAGE WITH 1% TRAINING SAMPLES

Class	KNN	SRC	KSRC	CRC	KCRC	FRC	ENRC	KENRC	KFRC
1	**83.17**	72.51	42.50	68.01	47.50	81.36	77.12	78.96	74.62
2	96.04	96.55	69.59	93.76	72.75	**98.34**	97.57	95.24	95.88
3	24.34	63.06	74.70	28.62	**78.85**	62.05	52.81	47.38	57.96
4	45.19	85.43	46.95	86.03	47.42	**88.93**	85.00	75.15	88.27
5	75.00	99.92	91.72	99.85	92.18	**99.97**	99.85	98.87	99.40
6	2.45	28.24	**96.80**	34.48	96.04	28.30	30.15	43.84	53.91
7	44.12	18.60	60.87	3.04	60.87	14.43	7.52	**64.54**	46.62
8	76.82	47.52	98.84	45.60	**99.07**	43.05	52.59	77.50	70.12
9	74.49	17.50	75.00	8.32	75.00	34.69	18.78	**99.68**	94.66
OA	71.47	74.06	79.40	70.40	79.93	76.29	75.03	80.62	**81.58**
κ	59.94	64.35	76.48	59.23	**77.13**	67.07	65.39	73.62	75.11

have a slightly better OA results than SRC. Extending to kernel space, KCRC has a similar performance with KSRC, and the kernel fused methods (KFRC and KENRC) have better classification results than nonfused kernel methods (KSRC and KCRC).

IV. CONCLUSION

This letter has elaborated several pixel-wise representation-based classifiers and proposed a KFRC, which combines SR- and CR-based representations in the kernel residual domain for HSI classification. Based on a fixed dictionary, the proposed KFRC attempts to attain a balance between the SR- and CR-based methods in the decision-level via a suitable balancing parameter θ. Taking full of the "competitive" peculiarity of SRC and the "collaborative" characteristic of CRC, the fused classifier has shown stronger discriminative ability than the single methods. Compared to the FRC, the proposed KFRC method further strengthens the discrimination ability in the high dimensional feature space, making the decision boundary more separable, and achieves a high classification performance.

REFERENCES

[1] W. Li, S. Prasad, J. E. Fowler, and L. M. Bruce, "Locality-preserving dimensionality reduction and classification for hyperspectral image analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 4, pp. 1185–1198, Apr. 2012.

[2] Y. Yuan, J. Lin, and Q. Wang, "Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 2966–2977, Dec. 2016.

[3] Y. Tarabalka, J. A. Benediktsson, and 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, Aug. 2009.

[4] X. Sun, N. M. Nasrabadi, and T. D. Tran, "Task-driven dictionary learning for hyperspectral image classification with structured sparsity constraints," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4457–4471, Aug. 2015.

[5] Q. Wang, J. Lin, and Y. Yuan, "Salient band selection for hyperspectral image classification via manifold ranking," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 6, pp. 1279–1289, Jun. 2016.

[6] G. Camps-Valls and L. Bruzzone, "Kernel-based methods for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 6, pp. 1351–1362, Jun. 2005.

[7] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised hyperspectral image segmentation using multinomial logistic regression with active learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 48, no. 11, pp. 4085–4098, Nov. 2010.

[8] W. Li, Q. Du, F. Zhang, and W. Hu, "Hyperspectral image classification by fusing collaborative and sparse representations," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 9, pp. 4178–4187, Sep. 2016.

[9] L. Zhang *et al.*, "Kernel sparse representation-based classifier," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1684–1695, Apr. 2012.

[10] X. Sun, Q. Qu, N. M. Nasrabadi, and T. D. Tran, "Structured priors for sparse-representation-based hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 7, pp. 1235–1239, Jul. 2014.

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

[12] H. Zhang, J. Li, Y. Huang, and L. Zhang, "A nonlocal weighted joint sparse representation classification method for hyperspectral imagery," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2056–2065, Jun. 2014.

[13] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. Int. Conf. Comput. Vis. (ICCV)*, Washington, DC, USA, Nov. 2011, pp. 471–478.

[14] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang. (Apr. 2012). "Collaborative representation based classification for face recognition." [Online]. Available: https://arxiv.org/abs/1204.2358

[15] J. Li, H. Zhang, Y. Huang, and L. Zhang, "Hyperspectral image classification by nonlocal joint collaborative representation with a locally adaptive dictionary," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 6, pp. 3707–3719, Jun. 2014.

[16] J. Li, H. Zhang, and L. Zhang, "Column-generation kernel nonlocal joint collaborative representation for hyperspectral image classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 94, pp. 25–36, Aug. 2014.

[17] E. Zhang, X. Zhang, H. Liu, and L. Jiao, "Fast multifeature joint sparse representation for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 7, pp. 1397–1401, Jul. 2015.

[18] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.

[19] S. Gao, I. W.-H. Tsang, and L.-T. Chia, "Kernel sparse representation for image classification and face recognition," in *Proc. Eur. Conf. Comput. Vis., Heraklion (ECCV)*, Crete, Greece, Sep. 2010, pp. 1–14.

[20] J. Yin, Z. Liu, Z. Jin, and W. Yang, "Kernel sparse representation based classification," *Neurocomputing*, vol. 77, pp. 120–128, Feb. 2012.

[21] W. Li, Q. Du, and M. Xiong, "Kernel collaborative representation with tikhonov regularization for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 1, pp. 48–52, Jan. 2015.