

Dissimilarity-Weighted Sparse Representation for Hyperspectral Image Classification

Le Gan, Junshi Xia, *Member, IEEE*, Peijun Du, *Senior Member, IEEE*, and Zhigang Xu

Abstract—To improve the capability of a traditional sparse representation-based classifier (SRC), we propose a novel dissimilarity-weighted SRC (DWSRC) for hyperspectral image (HSI) classification. In particular, DWSRC computes the weights for each atom according to the distance or dissimilarity information between the test pixel and the atoms. First, a locality constraint dictionary set is constructed by the Gaussian kernel distance with a suitable distance metric (e.g., Euclidean distance). Second, the test pixel is sparsely coded over the new weighted dictionary set based on the ℓ_1 -norm minimization problem. Finally, the test pixel is classified by using the obtained sparse coefficients with the minimal residual rule. Experimental results on two widely used public HSIs demonstrate that the proposed DWSRC is more efficient and accurate than other state-of-the-art SRCs.

Index Terms—Dissimilarity-weighted, distance metrics, hyperspectral image (HSI) classification, locality constraint dictionary, sparse representation.

I. INTRODUCTION

CASSIFICATION of hyperspectral images (HSIs) plays significant roles in several applications, including environmental monitoring [1], geological explorations [2], precision farming [3], and national defenses [4]. In the HSIs, each pixel is a high-dimensional vector and its entries represent the spectral response of hundreds of spectral bands, spanning from the visible to the infrared wavelengths [5]. Wang *et al.* [6] proposed a manifold ranking-based band selection method for HSI classification. Recently, a sparse modeling technique has been widely used in [7]–[10]. Peng *et al.* [7] proposed a general framework to solve the large-scale and out-of-sample clustering problems for representation-based subspace clustering. Peng *et al.* [8] proposed a novel unsupervised subspace learning method, which automatically determines the optimal dimension of

Manuscript received May 16, 2017; revised July 31, 2017; accepted August 20, 2017. Date of publication September 14, 2017; date of current version October 25, 2017. This work was supported in part by the National 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 Z. Xu 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, 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; xzg_xzg1982@163.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.2017.2743742

feature space and obtains the low-dimensional representation of a given data set. Under the framework of graph embedding, Peng *et al.* [9] developed two algorithms by embedding the L2-graph into a low-dimensional space for robust subspace clustering and subspace learning. The basic idea is that a test pixel can be sparsely approximated as a linear combination of a few atoms from the entire training dictionary via the ℓ_0 -norm or ℓ_1 -norm regularization. The final label of the test pixel is assigned to the class whose subdictionary provides the smallest representation error. In order to capture the nonlinear similarity of samples, a kernel version of sparse representation-based classifier (KSRC) is proposed in [11]. KSRC maps the samples into a high-dimensional kernel-induced feature space first and then sparse representation-based classifier (SRC) is performed in this new feature space by utilizing kernel trick. To alleviate the high computational complexity of SRC via ℓ_1 -norm regularization, Zhang *et al.* [12] proposed a collaborative representation-based classifier (CRC) via ℓ_2 -norm regularization.

In the hyperspectral remote sensing community, Chen *et al.* [13] introduced a joint sparse representation classification (JSRC) framework via simultaneous orthogonal matching pursuit (OMP). Furthermore, the kernel version of JSRC is exploited in [14]. Liu *et al.* [15] proposed a spatial-spectral KSRC via ℓ_1 -norm minimization to improve the performance of SRC. Li and Du [16] proposed a joint within-class CRC, investigating the relationship of the hyperspectral neighbor. Furthermore, they proposed a fused representation-based classifier (FRC) by combining the CRC and the SRC. The objective of the FRC is to integrate the advantages of the CRC and the SRC to improve the classification ability. Yuan *et al.* [17] proposed the spectral-spatial classification scheme, which focuses on multitask joint sparse representation (MJSR) and a stepwise Markov random field framework.

Data locality has been a critical issue in many machine learning applications. Wang *et al.* [18] proposed a locality constraint distance metric learning for traffic congestion detection. In [19], a multistage clustering strategy was aimed to discover collective motions with both local and global consistency along time. It should be emphasized that the typical SRC or CRC model does not consider the distance or dissimilarity relationship between the test pixel and each individual dictionary atoms. Specifically, dictionary atoms, which are closer or more similar to the test pixel, play a greater role in representing the test pixel. Once such a dissimilarity between a dictionary atom and the test pixel is determined, the significance of these atoms in representing the test pixel should

be known. To enhance the discriminative ability of SRC, Lu *et al.* [20] proposed weighted SRC (WSRC) by using the weighted ℓ_1 -minimization problem to consider the dissimilarity information. A distance-weighted CRC, namely nearest regularized subspace (NRS) [21], is also proposed to exploit the spectral similarity [i.e., Euclidean distance (ED)] to control the regularization term.

Extended by the idea of WSRC, we propose a dissimilarity-WSRC (DWSRC) for HSIs, which employs a locality constraint dictionary to represent the test pixel by multiple dissimilarity measures. The dissimilarity can be measured using the distances, such as ED, Mahalanobis distance (MD), spectral angle distance (SAD), and χ^2 distance (χ^2). The classification framework contains the following main stages:

- 1) to compute the weight by the dissimilarity between the dictionary atoms and the test pixel;
- 2) to construct a locality-constrained dictionary set;
- 3) to represent the test pixel over the new weighted dictionary atoms via the standard ℓ_1 -norm minimization problem;
- 4) to determine the label of the test pixel for classification by using the minimal residual rule.

The rest of this letter is structured as follows. The proposed DWSRC is described in Section II. The effectiveness of the DWSRC is validated by conducting several experiments on two HSIs in Section III. Finally, Section IV summarizes this letter and makes some concluding remarks.

II. DISSIMILARITY-WEIGHTED SPARSE REPRESENTATION CLASSIFIER

A. General SRC

Let $\mathcal{D} = [\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_C] \in \mathcal{R}^{B \times N}$ denote a class-specific dictionary set that contains C -class training samples, where $N = \sum_{c=1}^C N_c$ (N_c denotes the number of samples in the c th class) and B denotes the number of spectral bands. For a test pixel $\mathbf{y} \in \mathcal{R}^B$, the representation coefficients of $\mathbf{y} \in \mathcal{R}^B$ over dictionary \mathcal{D} via SRC can be obtained by solving the following ℓ_1 -minimization problem:

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathcal{D}\boldsymbol{\alpha}\|_2 + \lambda \|\boldsymbol{\alpha}\|_1 \quad (1)$$

where λ denotes the ℓ_1 -regularization parameter. After obtaining the sparse coefficients $\boldsymbol{\alpha}$, the label of pixel \mathbf{y} is determined by the minimal residual between \mathbf{y} and its C approximation from each subdictionary \mathcal{D}_c

$$\text{class}(\mathbf{y}) = \arg \min_{c=1, \dots, C} \|\mathbf{y} - \mathcal{D}_c \boldsymbol{\delta}_c(\boldsymbol{\alpha})\|_2^2 \quad (2)$$

where $\boldsymbol{\delta}_c$ is the indicator function [22] that extracts coefficients related with the c th class.

B. Pixel Dissimilarity and Multiple Distance Metrics

In HSIs, neighboring pixels usually consist of a similar material, and their spectral signatures are highly correlated. Hence, the spectral dissimilarity between the pairwise spectral vector reflects the approximation of each other, and a smaller value of dissimilarity denotes a higher probability that the two samples belong to the same class. In [20], they proposed a distance-WSRC method, which can adaptively exploit the

dissimilarity information between the test sample \mathbf{y} and each dictionary atoms \mathbf{d}_i . A larger $\text{dist}(\mathbf{y}, \mathbf{d}_i)$ characterizes a larger dissimilarity between the test sample and the training dictionary. Thus, WSRC can generate more discriminative sparse coefficients that can be used to reconstruct the test sample more robustly. Multiple well-performing distances are available to quantify the dissimilarity between the test sample and each dictionary atoms, including ED, MD, SAD, and χ^2 .

C. Proposed DWSRC

To find a robust representation of the test pixel \mathbf{y} of HSIs, we propose the DWSRC classifier considering pixel locality information based on the dissimilarity between \mathbf{y} and each atom \mathbf{d}_i .

First, we adopt the Gaussian kernel to capture the nonlinear information within the given HSI. Let \mathbf{w}_i define the locality weight matrix

$$\mathbf{w}_i = \exp \left(-\frac{\text{dist}(\mathbf{y}, \mathbf{d}_i)}{\sigma} \right) \quad (3)$$

where $\text{dist}(\mathbf{y}, \mathbf{d}_i)$ denotes the distance function between \mathbf{y} and \mathbf{d}_i , and σ denotes the parameter. We further normalize $\{\mathbf{w}_i\}_{i=1, \dots, N}$ in the range $(0, 1]$ by subtracting $\max_{i=1, \dots, N} \{\mathbf{w}_i\}_{i=1, \dots, N}$ from $\{\mathbf{w}_i\}_{i=1, \dots, N}$.

Second, after obtaining the weights $\mathbf{w} = \{\mathbf{w}_i\}_{i=1, \dots, N}$ between \mathbf{y} and each atoms \mathbf{d}_i , we construct a locality-constrained dictionary set $\mathcal{D}' = [\mathcal{D}'_1, \dots, \mathcal{D}'_c, \dots, \mathcal{D}'_C] \in \mathcal{R}^{B \times N}$, and $\mathcal{D}'_c = [w_{c1}\mathbf{d}_1, w_{c2}\mathbf{d}_2, \dots, w_{cN_c}\mathbf{d}_{cN_c}]$ denotes the weighted class-specific dictionary associated with the c th class.

Third, we code \mathbf{y} over the new weighted dictionary set \mathcal{D}' by solving the following ℓ_1 -minimization problem:

$$\hat{\boldsymbol{\alpha}} = \arg \min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathcal{D}'\boldsymbol{\alpha}\|_2 + \lambda \|\boldsymbol{\alpha}\|_1. \quad (4)$$

Fourth, after obtaining the sparse coefficient vector $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_N]^T$, the label of \mathbf{y} can generally be determined by the minimal residual rule

$$\text{class}(\mathbf{y}) = \arg \min_{c=1, \dots, C} \|\mathbf{y} - \mathcal{D}'_c \boldsymbol{\delta}_c(\boldsymbol{\alpha})\|_2^2. \quad (5)$$

The proposed DWSRC is summarized in Algorithm 1.

III. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, we evaluate the performance of the proposed DWSRC on two standard HSIs, including Indian Pines and PaviaU image. The Indian image is acquired by the AVIRIS sensor from JPL, NASA, on June 12, 1992. It generates 220 bands and 20 noisy bands are removed before classification. The spatial dimension is 145×145 with a spatial resolution of 20 m, and it contains 16 ground truth classes. The PaviaU image covering an urban area was obtained from the ROSIS-03 sensor with a spatial resolution of 1.3 m. It contains 610×340 pixels with 115 spectral bands, in which 12 noisy bands are removed. There are nine ground truth classes of interests. For the two HSIs, overall accuracy (OA), average accuracy (AA), and κ are used to evaluate the classification performance. All the tests are repeated ten times to get the averaged results. All the tests are carried out using MATLAB

TABLE I
CLASSIFICATION RESULTS (MEAN \pm STD-DEV PERCENT) OBTAINED BY DIFFERENT METHODS ON THE INDIAN PINES IMAGE VERSUS A DIFFERENT SIZE OF DICTIONARY ATOMS FOR PER CLASS

Methods		OMP [24]	SRC [22]	CRC [12]	FRC [25]	KSRC [11]	NRS [21]	WSRC [20]	DWSRC- χ^2
5%	OA	65.05 \pm 0.38	58.67 \pm 0.12	63.48 \pm 1.15	68.74 \pm 0.57	78.93 \pm 0.70	76.83 \pm 0.53	79.31 \pm 0.59	79.65 \pm 0.62
	AA	60.00 \pm 1.43	47.63 \pm 0.95	50.32 \pm 0.57	54.58 \pm 0.54	72.84 \pm 1.70	67.83 \pm 1.24	70.95 \pm 1.78	71.44 \pm 1.31
	κ	59.79 \pm 0.51	52.34 \pm 0.15	57.71 \pm 1.35	63.70 \pm 0.68	75.89 \pm 0.77	73.08 \pm 0.71	76.26 \pm 0.66	76.66 \pm 0.70
10%	OA	68.75 \pm 0.94	65.76 \pm 1.04	68.35 \pm 0.37	72.47 \pm 0.69	82.34 \pm 0.44	78.98 \pm 0.72	83.01 \pm 0.64	84.34 \pm 0.54
	AA	64.78 \pm 1.67	54.38 \pm 1.03	53.33 \pm 0.43	57.29 \pm 0.84	78.00 \pm 1.51	72.92 \pm 1.93	77.39 \pm 1.60	78.03 \pm 1.32
	κ	64.06 \pm 1.07	60.46 \pm 1.16	63.20 \pm 0.45	68.07 \pm 0.85	79.83 \pm 0.50	75.53 \pm 0.88	80.52 \pm 0.77	80.91 \pm 0.65
15%	OA	70.00 \pm 0.44	68.04 \pm 0.50	71.09 \pm 0.26	74.59 \pm 0.86	84.41 \pm 0.41	79.14 \pm 0.92	85.03 \pm 0.37	85.64 \pm 0.26
	AA	65.73 \pm 1.23	56.55 \pm 0.76	54.97 \pm 0.20	59.20 \pm 0.35	81.32 \pm 1.52	73.97 \pm 0.96	79.49 \pm 1.69	80.02 \pm 1.81
	κ	65.51 \pm 0.49	63.11 \pm 0.56	66.39 \pm 0.32	70.54 \pm 0.61	82.20 \pm 0.46	75.66 \pm 1.12	82.85 \pm 0.43	83.09 \pm 0.30
20%	OA	70.94 \pm 0.55	71.30 \pm 0.14	71.74 \pm 0.49	75.72 \pm 0.44	85.65 \pm 0.46	79.42 \pm 0.80	86.49 \pm 0.13	87.08 \pm 0.17
	AA	67.85 \pm 1.36	60.15 \pm 0.68	55.29 \pm 0.32	60.23 \pm 0.60	84.76 \pm 1.14	73.23 \pm 1.78	82.08 \pm 0.60	82.89 \pm 0.69
	κ	66.57 \pm 0.71	66.89 \pm 0.18	67.12 \pm 0.57	71.85 \pm 0.51	83.62 \pm 0.28	75.94 \pm 0.98	84.53 \pm 0.14	84.87 \pm 0.20

Algorithm 1 Proposed DWSRC for HSIs

```

1: Input: A initial dictionary of training samples  $\mathcal{D} = [\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_C] \in \mathbb{R}^{B \times N}$  for  $C$  class; a test pixel set  $\mathbf{Y} = \{\mathbf{y}^m\}_{m=1,\dots,M}$ ; sparsity regularization  $\lambda$ 
2: for each test pixel  $\mathbf{y}^m$  do
3:   1) Calculate the weight matrix  $\mathbf{w}_i$  between  $\mathbf{y}^m$  and each dictionary atom  $\mathbf{d}_i$  via Gaussian kernel distance with a suitable distance metric;
4:   2) Generate the locality-constrained dictionary set  $\mathcal{D}' = [\mathcal{D}'_1, \dots, \mathcal{D}'_c, \dots, \mathcal{D}'_C] \in \mathbb{R}^{B \times N}$ ,
5:   3) Sparsely code  $\mathbf{y}^m$  over the new weighted dictionary  $\mathcal{D}'$  via  $\ell_1$ -norm minimization according to Eq. (4).
6:   4) Identify the final class of  $\mathbf{y}^m$  with the minimal residual rule according to Eq. (5).
7: end for
8: Output: An 2-D map which records the labels of the test sample set  $\mathbf{Y}$ 

```

R2016a on Intel Core i7-4790 CPU PC machine with 16 GB of RAM, and we use the SPArse Modeling Software package [23] to solve the ℓ_1 -minimization problem.

For comparison, we have implemented the OMP (i.e., ℓ_0 -norm) [24], SRC [22], CRC [12], FRC [25], KSRC [11], and WSRC [20]. The Gaussian kernel is adopted for the KSRC, and the width is set to be the averaged distance of the dictionary atoms [i.e., $\sigma = (1/N^2)\|\mathbf{d}_i - \mathbf{d}_j\|$, where N denotes the number of dictionary atoms]. The sparsity level of the OMP is set to be 10. The regularization parameters for the aforementioned classifiers are learned by cross validation. Fig. 1 plots the results of DWSRC with different regularization parameters (λ) on the Indian Pines (with the χ^2 distance) and PaviaU images (with the SAD distance), respectively. From Fig. 1, the optimized values of λ are $1e-2$ and $4e-2$ for the Indian Pines and PaviaU images, respectively.

First, we investigate the classification performance (OA, AA, and κ) of the proposed DWSRC as well as the compared methods. For the Indian Pines image, 5%, 10%, 15%, and 20% labeled pixel from the ground truth are chosen as the initial dictionary atoms, and the remaining pixels are used for the test. For the PaviaU image, 0.25%, 0.5%, 1%, and 2% labeled pixels are selected as the

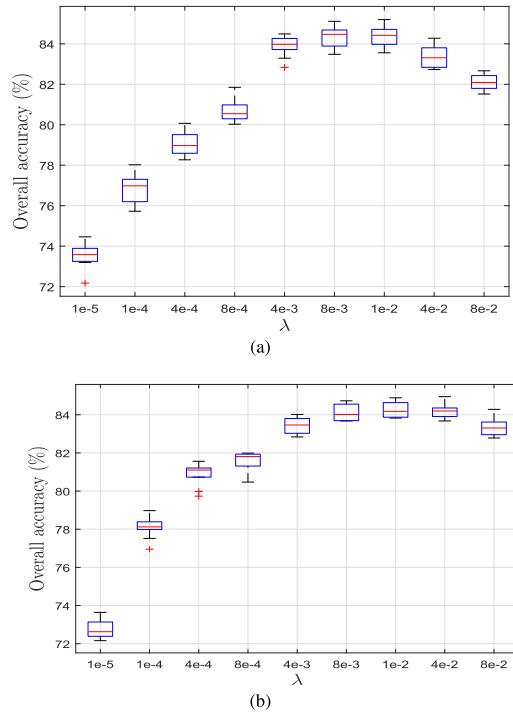


Fig. 1. Sensitivity to the change of λ on (a) Indian Pines image and (b) PaviaU image.

initial dictionary atoms and the remaining pixels are used for the test purpose. The results are presented in Tables I and II. Please note that we only present the best results in four distance metrics. As shown in Tables I and II, DWSRC- χ^2 and DWSRC-SAD achieve the best results for Indian Pines and PaviaU image in all the cases. For both the images, CRC is superior to SRC, and FRC outperforms two individual classifiers (i.e., SRC and CRC). For the Indian Pines image, FRC achieves better results than OMP. However, the situation is reversed in the PaviaU image. DWSRC performs better classification results than KSRC, because the scheme by using weighted-dissimilarity information is more efficient than the one with the kernel mapping method. Thanks to the construction of a locality constraint dictionary set by using weighted-dissimilarity information, the proposed DWSRC achieves better performance than NRS and WSRC. Fig. 3(c)–(i) shows the classification maps on the PaviaU image using ℓ_1 dictionary atoms obtained by OMP, SRC, FRC, KSRC, NRS, WSRC, and DWSRC-SAD.

TABLE II
CLASSIFICATION RESULTS (MEAN \pm STD-DEV PERCENT) OBTAINED BY DIFFERENT METHODS
ON THE PAVIAU IMAGE VERSUS A DIFFERENT SIZE OF DICTIONARY ATOMS FOR PER CLASS

Methods		OMP [24]	SRC [22]	CRC [12]	FRC [25]	KSRC [11]	NRS [21]	WSRC [20]	DWSRC-SAD
0.25%	OA	69.60 \pm 1.87	30.64 \pm 1.51	53.99 \pm 1.82	62.03 \pm 1.04	74.57 \pm 0.77	78.68 \pm 2.09	77.89 \pm 1.64	78.86 \pm 1.61
	AA	60.50 \pm 2.62	38.15 \pm 1.28	45.16 \pm 1.87	50.46 \pm 1.22	66.67 \pm 1.87	65.38 \pm 4.65	71.67 \pm 1.25	71.55 \pm 1.24
	κ	59.00 \pm 2.28	18.73 \pm 1.22	40.08 \pm 2.32	49.41 \pm 1.58	65.57 \pm 1.23	70.36 \pm 3.30	70.33 \pm 2.06	70.68 \pm 2.02
0.5%	OA	73.20 \pm 1.89	60.18 \pm 2.24	64.85 \pm 1.43	67.79 \pm 1.04	80.10 \pm 0.70	81.55 \pm 0.43	80.32 \pm 1.22	82.33 \pm 1.22
	AA	66.00 \pm 1.44	47.72 \pm 0.96	49.93 \pm 1.16	54.58 \pm 0.86	74.58 \pm 2.03	70.17 \pm 0.65	73.21 \pm 1.66	73.26 \pm 1.71
	κ	63.59 \pm 2.37	46.54 \pm 2.52	52.35 \pm 1.64	56.05 \pm 1.16	73.24 \pm 0.99	74.35 \pm 0.66	73.28 \pm 1.64	75.30 \pm 1.64
1%	OA	77.56 \pm 0.44	66.52 \pm 0.35	70.30 \pm 0.68	71.93 \pm 0.90	83.27 \pm 0.48	83.71 \pm 0.61	83.51 \pm 0.25	85.53 \pm 0.25
	AA	71.84 \pm 0.64	51.36 \pm 0.88	51.75 \pm 0.95	58.57 \pm 1.75	74.26 \pm 1.27	73.95 \pm 1.56	76.42 \pm 0.47	77.56 \pm 0.45
	κ	69.63 \pm 0.55	54.44 \pm 0.44	58.96 \pm 0.88	61.48 \pm 1.24	77.59 \pm 0.70	77.43 \pm 0.89	77.68 \pm 0.31	77.70 \pm 0.32
2%	OA	79.84 \pm 0.58	68.55 \pm 0.86	73.16 \pm 0.53	73.69 \pm 0.76	84.94 \pm 0.67	83.23 \pm 0.35	85.25 \pm 0.42	86.56 \pm 0.41
	AA	76.22 \pm 0.94	52.17 \pm 1.04	53.88 \pm 0.84	61.43 \pm 1.01	80.53 \pm 1.31	73.64 \pm 0.30	79.85 \pm 1.13	80.89 \pm 1.11
	κ	72.80 \pm 0.82	56.99 \pm 1.12	62.74 \pm 0.75	63.81 \pm 1.02	79.85 \pm 0.94	76.63 \pm 0.54	80.20 \pm 0.54	81.20 \pm 0.54

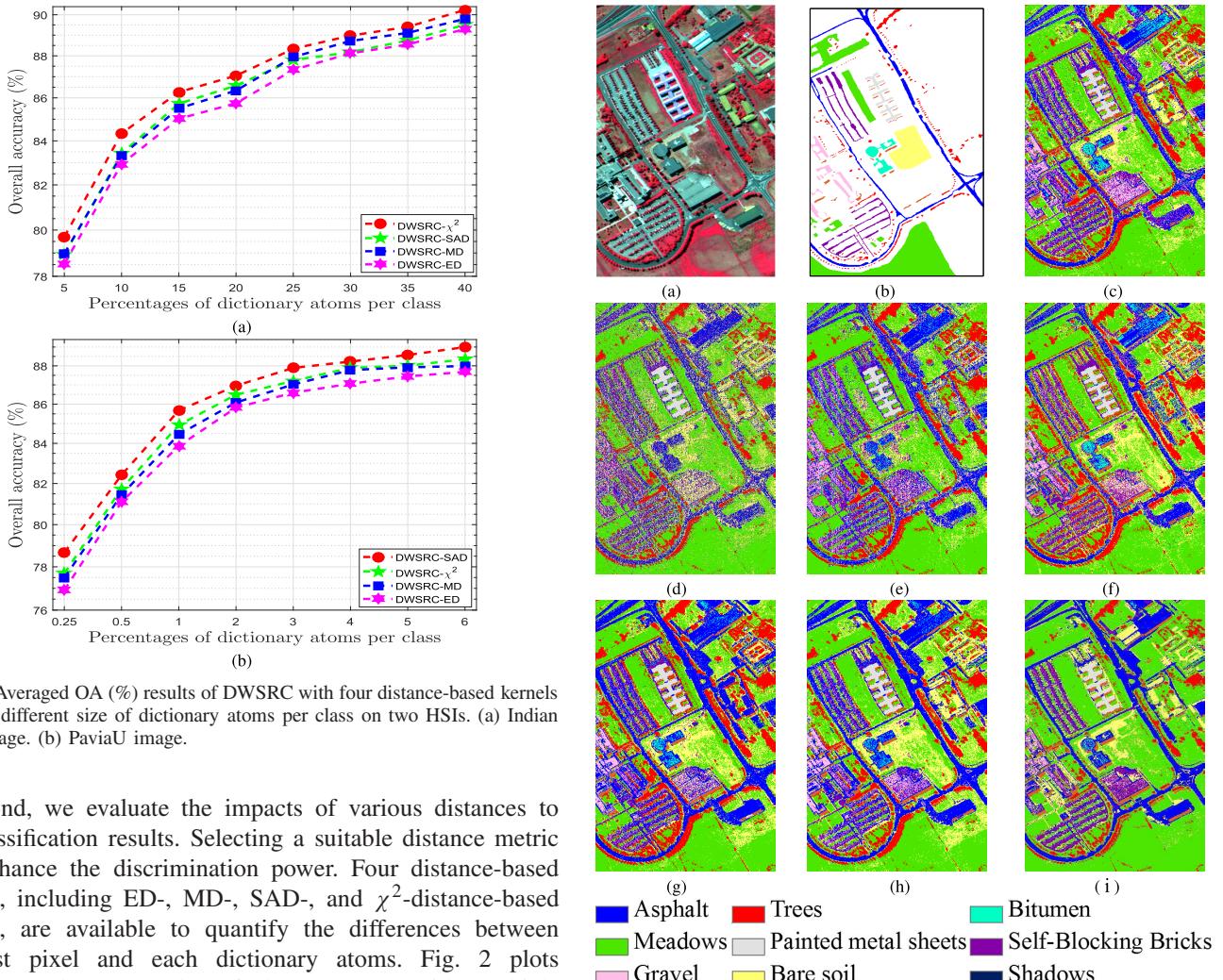


Fig. 2. Averaged OA (%) results of DWSRC with four distance-based kernels versus a different size of dictionary atoms per class on two HSIs. (a) Indian Pines image. (b) PaviaU image.

Second, we evaluate the impacts of various distances to the classification results. Selecting a suitable distance metric can enhance the discrimination power. Four distance-based kernels, including ED-, MD-, SAD-, and χ^2 -distance-based kernels, are available to quantify the differences between the test pixel and each dictionary atoms. Fig. 2 plots the averaged OA results of the DWSRC with multiple distance-based kernels by a different size of dictionary atoms per class. The range of proportions of dictionary atoms in the total labeled samples is set to be $\{5\%, 10\%, 15\%, 20\%, 25\%, 30\%, 35\%, 40\%\}$ for the Indian Pines image, and as $\{0.25\%, 0.5\%, 1\%, 2\%, 3\%, 4\%, 5\%, 6\%\}$ for the PaviaU image. From Fig. 2, a larger size of dictionary atoms tends to have better performance of DWSRC. The class separability of the two HSIs is different using various distance metrics. Thus, the optimized results are acquired by various distance metrics on two HSIs. For the Indian Pines image, χ^2 kernel achieves

Fig. 3. (a) False color image. (b) Groundtruth map. Classification maps on the PaviaU image using %1 dictionary atoms obtained by (c) OMP (OA = 78.14%), (d) SRC (OA = 66.79%), (e) FRC (OA = 72.26%), (f) KSRC (OA = 83.53%), (g) NRS (OA = 83.52%), (h) WSRC (OA = 83.67%), and (i) DWSRC-SAD (OA = 84.08%).

the highest OA in all cases. The SAD kernel method has similar results to MD kernel, and ED kernel method performs the worst classification performance. For the PaviaU image, the SAD kernel method achieves the best OA results in all cases.

TABLE III
MCNEMAR'S TEST ON THE INDIAN PINES IMAGE

Dictionary percent (%)		5	10	15	20	25
Distance	χ^2 vs. SAD	1.82	2.38	2.04	2.83	1.67
	χ^2 vs. MD	2.74	2.31	2.86	2.92	3.16
	χ^2 vs. ED	3.35	2.41	2.44	3.52	2.45
Method	DWSRC- χ^2 vs. WSRC	1.18	3.32	1.99	2.36	2.57
	DWSRC- χ^2 vs. KSRC	7.63	9.86	8.26	10.02	9.42

TABLE IV
MCNEMAR'S TEST ON THE PAVIAU IMAGE

Dictionary percent (%)		0.25	0.5	1	2	3
Distance	SAD vs. χ^2	5.14	2.87	5.28	2.47	3.01
	SAD vs. MD	7.78	6.66	7.37	6.10	8.41
	SAD vs. ED	9.28	4.46	8.38	7.18	8.07
Method	DWSRC-SAD vs. WSRC	7.82	14.89	15.58	9.79	10.69
	DWSRC-SAD vs. KSRC	18.65	16.59	17.62	18.52	20.11

Finally, we investigate the statistical differences between DWSRC with four distance-based kernels by the standardized McNemar's test. If the test statistic $|Z| > 1.96$, the two classifiers are regarded as statistically significant at the 5% level of significance [26]. The results under the different size of dictionary atoms for two HSIs are listed in Tables III and IV, respectively.

The statistical difference results indicate that DWSRC- χ^2 is significantly better than DWSRC-SAD, DWSRC-MD, and DWSRC-ED in most cases for the Indian Pines image, except for χ^2 versus SAD with 5% and 25% size of dictionary atoms. For the PaviaU image, DWSRC-SAD statistically outperforms DWSRC- χ^2 , DWSRC-MD, and DWSRC-ED. Compared with WSRC and KSRC methods, DWSRC- χ^2 is statistically significantly better for the Indian Pines image (except for the 5% dictionary percent), and DWSRC-SAD is statistically significantly better for the PaviaU image.

IV. CONCLUSION

In this letter, we have developed a novel DWSRC based on the Gaussian kernel distance for HSI, where multiple distance metrics are adopted (i.e., χ^2 distance, Manhattan distance, SAD, and ED) to construct the locality constraint dictionary set. Compared with the traditional SRC, the proposed DWSRC aims to utilize the dissimilarity information based on multiple distance kernels to quantify the significance of each dictionary atoms in representing a test pixel. Experimental results demonstrated that the proposed DWSRC achieves a better classification performance on two standard HSIs.

REFERENCES

- [1] G. Camps-Valls, D. Tuia, L. Bruzzone, and J. A. Benediktsson, "Advances in hyperspectral image classification: Earth monitoring with statistical learning methods," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 45–54, Jan. 2014.
- [2] J. Cui *et al.*, "Temperature and emissivity separation and mineral mapping based on airborne TASI hyperspectral thermal infrared data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 40, pp. 19–28, Aug. 2015.
- [3] J. Cierniewski, C. Kaźmierowski, S. Królewicz, J. Piekarczyk, M. Wróbel, and B. Zagajewski, "Effects of different illumination and observation techniques of cultivated soils on their hyperspectral bidirectional measurements under field and laboratory conditions," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2525–2530, Jun. 2014.
- [4] L. Luft, C. Neumann, M. Freude, N. Blaum, and F. Jeltsch, "Hyperspectral modeling of ecological indicators—A new approach for monitoring former military training areas," *Ecol. Indicators*, vol. 46, pp. 264–285, Nov. 2014.
- [5] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral Remote Sensing: Principles and Applications*. Boca Raton, FL, USA: CRC Press, 2008.
- [6] 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.
- [7] X. Peng, H. Tang, L. Zhang, Z. Yi, and S. Xiao, "A unified framework for representation-based subspace clustering of out-of-sample and large-scale data," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 12, pp. 2499–2512, Dec. 2016.
- [8] X. Peng, J. Lu, Z. Yi, and R. Yan, "Automatic subspace learning via principal coefficients embedding," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3583–3596, Nov. 2017.
- [9] X. Peng, Z. Yu, Z. Yi, and H. Tang, "Constructing the L2-graph for robust subspace learning and subspace clustering," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1053–1066, Apr. 2017.
- [10] L. Gan, P. Du, J. Xia, and Y. Meng, "Kernel fused representation-based classifier for hyperspectral imagery," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 5, pp. 684–688, May 2017.
- [11] S. Gao, I. W.-H. Tsang, and L.-T. Chia, "Kernel sparse representation for image classification and face recognition," in *Proc. ECCV*, 2010, pp. 1–14.
- [12] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang, "Collaborative representation based classification for face recognition," unpublished paper, 2012. [Online]. Available: <https://arxiv.org/abs/1204.2358>
- [13] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.
- [14] 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.
- [15] J. Liu, Z. Wu, Z. Wei, L. Xiao, and L. Sun, "Spatial-spectral kernel sparse representation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 6, no. 6, pp. 2462–2471, Dec. 2013.
- [16] W. Li and Q. Du, "Joint within-class collaborative representation for hyperspectral image classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2200–2208, Jun. 2014.
- [17] 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.
- [18] Q. Wang, J. Wan, and Y. Yuan, "Locality constraint distance metric learning for traffic congestion detection," *Pattern Recognit.*, to be published.
- [19] Q. Wang, M. Chen, and X. Li, "Quantifying and detecting collective motion by manifold learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4292–4298.
- [20] C.-Y. Lu, H. Min, J. Gui, L. Zhu, and Y.-K. Lei, "Face recognition via weighted sparse representation," *J. Vis. Commun. Image Represent.*, vol. 24, no. 2, pp. 111–116, Feb. 2013.
- [21] W. Li, E. W. Tramel, S. Prasad, and J. E. Fowler, "Nearest regularized subspace for hyperspectral classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 1, pp. 477–489, Jan. 2014.
- [22] 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.
- [23] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, Mar. 2010.
- [24] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.
- [25] 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.
- [26] A. Villa, J. A. Benediktsson, J. Chanussot, and C. Jutten, "Hyperspectral image classification with independent component discriminant analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 12, pp. 4865–4876, Dec. 2011.