Spatial-Aware Collaborative Representation for Hyperspectral Remote Sensing Image Classification

Junjun Jiang, *Member, IEEE,* Chen Chen, *Member, IEEE,* Yi Yu, Xinwei Jiang, and Jiayi Ma *Member, IEEE*

Representation-residual based classifiers have attracted much attention in recent years in hyperspectral image (HSI) classification. How to obtain the optimal representation coefficients for the classification task is the key problem of these methods. In this letter, spatial-aware collaborative representation (SaCR) is proposed for HSI classification. In order to make full use of the spatial-spectral information, we propose a closed-form solution, in which the spatial and spectral features are both utilized to induce the distance-weighted regularization terms. Different from traditional collaborative representation based HSI classification algorithms, which model the spatial feature in a preprocessing or post-processing stage, we directly incorporate the spatial information by adding a spatial regularization term to the representation objective function. Experimental results on three HSI datasets verify that our proposed approach outperforms the state-of-the-art classifiers.

***Index Terms*—Hyperspectral image classification, collaborative representation (CR), spatial regularization, spectral spatial information.**

I. Introduction

Hyperspectral remote sensing (HRS) allows the simultaneous acquisition of full portrayal of each material's spectral reflectance and increases the possibility of discriminating ground objects more accurately [1], [2], [3], [4]. Recently, representation-based classification methods are of great interest in hyperspectral image (HSI) classification due to their excellent performances [5], [6], [7], [8], [9], [10], [11], [12]. The common idea is to represent a test sample by the training samples and assign the class that provides the lowest reconstruction residual. We term this strategy as "representation-residual" based classification in this letter.

For example, sparse representation (SR) classification was first proposed in [13], and has been successfully applied in various applications [14], [15], [16], [17], [18], [19], [20]. Chen et al. [5] introduced SR to HSI classification based on the assumption that HSI pixels that belong to the same class approximately lie in a low-dimensional subspace. In [21], Zhang et al. argued that it is the collaborative representation (CR) rather than the computationally expensive SR constraint that actually determines the classification performance. Inspired by this observation, Li et al. [8] developed a joint CR (JCR) classification method with spatial and spectral features from surrounding pixels, and they further extended JCR to the kernel version and the weighted version in [9] and [10], respectively. In addition to the raw pixel feature, Gabor feature [11] or multi-feature [7] has been proposed to improve the performance of JCR. Instead of exploring the neighbor pixels, Li et al. [6] constructed a joint matrix using the nonlocal pixels of a test pixel, and proposed a nonlocal joint collaborative representation classification method.

The essential of utilizing the spatial information of these collaborative representation methods is to construct a joint vector [8], [11], [9], or matrix [5], [6] before performing sample representation. By incorporating the spatial information of HSI, one can expect to obtain a better representation than using spectral information only [22], [23], [24], [25].

To simultaneously incorporate the spectral and spatial information to the representation framework, in this letter we propose a spatial-aware collaborative representation (SaCR). In SaCR, we design two explicit regularization terms: one is for modeling the spectral similarity constraint and the other is for the spatial similarity constraint. Based on these two regularization terms, we develop a closed-form solution to effectively fuse the spectral and spatial information. Experimental results on three HSI datasets demonstrate the proposed method outperforms the state-of-the-art spectral-spatial based CR classification techniques.

II. Related Work

Suppose there are N training HSI pixels $\mathbf{x}_i \in \mathbb{R}^d$ (d dimensional feature space), $i = 1, 2, \cdots, N$, chosen from C different classes. The training set can be also denoted by a matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N] \in \mathbb{R}^{d \times N}$. Let $w_i \in \{1, 2, \cdots, C\}$ be the label of the i-th pixel and n_l be the number of available training samples from the l-th class, thus $\sum_{l=1}^{C} n_l = N$.

Given a test pixel $\mathbf{y} \in \mathbb{R}^d$, CR based methods utilize all the training samples to represent it:

$$J(\boldsymbol{\alpha}) = ||\mathbf{y} - \mathbf{X}\alpha||_2^2 + \lambda\Omega(\boldsymbol{\alpha}). \tag{1}$$

Here, the regularization parameter λ balances the trade-off between the reconstruction residual and the prior of $\boldsymbol{\alpha}$. After solving the optimal representation coefficients $\boldsymbol{\alpha}^* = \arg\min_{\boldsymbol{\alpha}} J(\boldsymbol{\alpha})$, the class label of the test pixel $\mathbf{y}$ is determined according to the minimum residual,

$$\text{class}(y) = \arg\min_{l=1,2,\ldots,C} ||\mathbf{y} - \mathbf{X}_l\boldsymbol{\alpha}_l^*||_2^2, \tag{2}$$

The research was supported by the National Natural Science Foundation of China (61501413, 61503288, 61402424), and by the Fundamental Research Funds for the Central Universities, China University of Geosciences (Wuhan) under Grant CUGL160412.

J. Jiang and X. Jiang are with the School of Computer Science, China University of Geosciences, Wuhan 430074, China (junjun0595@163.com, ysjxw@hotmail.com).

C. Chen is with the University of Central Florida, Orlando, USA (chenchen870713@gmail.com).

Y. Yu is with the Digital Content and Media Sciences Research Division, National Institute of Informatics, Tokyo 101-8430, Japan (yiyu@nii.ac.jp).

J. Ma is with the Electronic Information School, Wuhan University, Wuhan 430072, China (e-mail: jiayima@whu.edu.cn).

where $\mathbf{X}_l$ and $\boldsymbol{\alpha}_l^*$ are the subsets of $\mathbf{X}$ and $\boldsymbol{\alpha}^*$, respectively.

Different priors $\Omega(\boldsymbol{\alpha})$ have been proposed to regularize the least squares problem. In SR classification, a test sample is sparsely represented by an ℓ_0 norm regularization. In [21], Zhang et al. argued that it is the CR nature (i.e., using all the training samples) rather than the sparsity that improves classification accuracy.

To make the representation more flexible, nearest regularized CR method [26] introduces a locality constraint to regularize the solution of (1) by giving different freedom to training samples according to their Euclidean distances from $\mathbf{y}$:

$$J(\boldsymbol{\alpha}) = ||\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}||_2^2 + \lambda||\boldsymbol{\Gamma}\alpha||_2^2 \tag{3}$$

where $\boldsymbol{\Gamma}$ is a biasing Tikhonov matrix defined by with $\boldsymbol{\Gamma}_{ii} = ||\mathbf{y} - \mathbf{x}_i||_2, i = 1, 2, ..., N$.

If the distance between a training sample $\mathbf{x}_i$ and the test sample $\mathbf{y}$ is large, $\mathbf{x}_i$ will be given a small contribution (i.e., the corresponding representation coefficient α_i is small), and vice versa.

In HSI, neighboring pixels usually consist of similar materials with high probability. Thus, their spectral characteristics are highly correlated. The spatial correlation across neighboring pixels can be indirectly incorporated through a joint model. For example, Xiong et al. [8] proposed to simultaneously represent a test pixel and its neighbors by spatially averaging the test and training pixels:

$$J(\boldsymbol{\alpha}) = ||\tilde{\mathbf{y}} - \tilde{\mathbf{X}}_l\boldsymbol{\alpha}_l||_2^2 + \lambda||\boldsymbol{\Gamma}_l\boldsymbol{\alpha}_l||_2^2 \tag{4}$$

where $\tilde{\mathbf{y}}$ denotes the average spectral features for the test pixel $\mathbf{y}$ centered in a window with m neighbors, and $\tilde{\mathbf{X}}_l$ is the averaged value for each element in matrix $\mathbf{x}_l$. In [10], they further developed an improved version to utilize more appropriate weights for surrounding pixels.

III. The Proposed Method

A. SaCR

In this section, we introduce our proposed CR model by explicitly modeling the spectral and spatial feature. Specifically, we develop a closed-form solution based on spatial-aware collaborative representation (SaCR), in which the spectral and spatial information is used together to induce regularization terms. Different from traditional CR models, which use neighbor pixels to construct a joint vector by averaging [8], [10] or matrix by combing [7], [6], [5] (they all can be seen as a predefining stage), we directly incorporate the spatial feature into the objective function (6) by adding a spatial feature induced regularization term:

$$J(\boldsymbol{\alpha}) = ||\mathbf{y} - \mathbf{X}\boldsymbol{\alpha}||_2^2 + \lambda||\boldsymbol{\Gamma}\boldsymbol{\alpha}||_2^2 + \gamma||\mathrm{diag}(\mathbf{s})\boldsymbol{\alpha}||_2^2, \tag{5}$$

where $\mathbf{s} = [s_1, s_2, \cdots, s_N]$, s_i is associated with each training samples which encourages representation coefficients that are spatially coherent with respect to the training data, and diag($\mathbf{s}$) returns a square diagonal matrix with the elements of vector $\mathbf{s}$ on the main diagonal. Here, the regularization parameters λ and γ control the contributions of the locality prior (second term) and spatial prior (third term), respectively. For more details about the effects of these two parameters, please refer to the Section IV-B. Analogous to the role of the Tikhonov matrix $\boldsymbol{\Gamma}$, $\mathrm{diag}(\mathbf{s})$ acts similar effects. If one training sample $\mathbf{x}_i$ is neighbor to the test pixel, i.e., s_i is small (the penalty on representation coefficient α_i is small), then α_i is likely to be a relatively large value and $\mathbf{x}_i$ contributes significantly to the reconstruction of the test pixel, and vice versa.

Denoting the pixel coordinate of sample $\mathbf{x}_i$ and $\mathbf{y}$ as (p_i, q_i) and (p_y, q_y) respectively, the spatial coherent between $\mathbf{x}_i$ and $\mathbf{y}$ can be measured as

$$s_i = [\mathrm{dist}((p_i, q_i), (p_y, q_y))]^c, \tag{6}$$

where $\mathrm{dist}(\cdot)$ denotes Euclidean distance and c is a smooth parameter adjusting the distance decay speed for the spatial prior. Usually we further normalize $\mathbf{s}$ to be between (0, 1] by dividing $\mathtt{max}(\mathbf{s})$ from $\mathbf{s}$. The introduction of the smooth parameter c gives the model more flexibility. It allows to emphasize more on neighbor pixels with better discriminability. When the value of c is very large, our proposed model will heavily penalize pixels that are far away from the test pixel $\mathbf{y}$ by assigning weights close to 0 to them.

From Eq. (5), we learn that the proposed SaCR model consists of three parts: the first part is the data reconstruction term to ensure the reconstructed sample ($\mathbf{x}\boldsymbol{\alpha}^*$) being similar to $\mathbf{y}$, the second part is the spectral induced penalty term that enforces similar training pixels to the test pixel to have large representation coefficients and vice versa, and the last part is the spatial induced penalty term that enforces the neighbor training pixels of the test pixel (i.e., training pixels that are spatially close to the test pixel) to have large representation coefficients and vice versa. λ and γ are two regularization parameters to balance the contributions of the three terms.

The optimization of (5) is similar to (4), which can be derived analytically as:

$$\boldsymbol{\alpha}^* = (\mathbf{x}^T\mathbf{X} + \lambda\boldsymbol{\Gamma} + \gamma\mathrm{diag}(\mathbf{s}))\mathbf{x}^T\mathbf{y}. \tag{7}$$

B. JSaCR

To take into consideration of the contextual information of center pixels, we further extend our SaCR method to a joint version (named JSaCR for short):

$$J(\boldsymbol{\alpha}) = ||\tilde{\mathbf{y}} - \tilde{\mathbf{X}}\alpha||_2^2 + \lambda||\tilde{\boldsymbol{\Gamma}}\alpha||_2^2 + \gamma||\mathrm{diag}(\tilde{\mathbf{s}})\boldsymbol{\alpha}||_2^2, \tag{8}$$

where $\tilde{\mathbf{y}}$ is defined as in JCR [1], $\tilde{\mathbf{X}}$ is the average value for each element in matrix $\mathbf{x}$, i.e., $\tilde{\mathbf{x}}_j = (1/m)\sum_{i=1}^m \mathbf{x}_i$, $j = 1, 2, \cdots, N$, $\tilde{\boldsymbol{\Gamma}}$ and $\tilde{\mathbf{s}}$ are defined by $\tilde{\mathbf{X}}$ and $\tilde{\mathbf{y}}$.

When compared with SaCR and JSaCR, we learn that JSaCR can be summarized as the following two steps: (i) average filtering and (ii) performing SaCR on the filtered image. Note that the average filtering usually can smooth the random noise in HSI image, so JSaCR can be expected to have better results than SaCR.

C. Relation to Existing Methods

It is worth mentioning that our proposed JSaCR method aims to simultaneously incorporate the spectral and spatial information to the HSI classification task, where spatial and

TABLE I
THE PARAMETER SETTINGS OF OUR PROPOSED SACR AND JSACR METHODS FOR THE THREE HSI DATASETS.

Parameters	Indian Pine		University of Pavia		Salinas	
	SaCR	JSaCR	SaCR	JSaCR	SaCR	JSaCR
c	4	4	8	8	2	2
λ	0.01	0.01	0.01	0.01	0.01	1e-06
γ	1e+04	1	1e+06	1e+04	10	0.01

spectral features are both utilized to induce the distance-weighted regularization terms, and this makes JSaCR different from the existing spectral and spatial features based HSI classification methods. Specifically,

- CR [21] can be seen as a special case of the proposed SaCR method when we set $\mathbf{\Gamma}$ to the identity matrix and γ to zero (see Eq. (1) and Eq. (5));
- When we set γ to zero, the proposed SaCR method will reduce to WCR [10] (see Eq. (3) and Eq. (5));
- Our proposed JSaCR method will reduce to JCR [8] when we set γ to zero, please refer to the objective functions (4) and (8).
- When compared with the traditional CR based methods, e.g., WCR and JCR, our proposed SaCR and JSaCR both have the spectral penalization factor and spatial penalization factor, thus the proposed algorithm will take much time to calculate the similarity.

IV. EXPERIMENTS AND ANALYSIS

In this section, we investigate the effectiveness of the proposed SaCR classification algorithm and its joint version (JSaCR) using three HSI datasets. The classifiers including support vector machine (SVM) [27], [28], SVM with composite kernel that combines the spectral and spatial information via a weighted kernel summation (denoted by SVM-CK) [28], SR classification [13], JSR classification [5], CR classification [21], weighted collaborative representation (WCR) classification [10] and JCR classification [8] are used for comparison in this letter. The classification performance is measured by overall accuracy (OA)[1] on the three HSI datasets.

A. Experimental Datasets

The first HSI dataset is the Indian Pine, acquired by National Aeronautics and Space Administrations (NASA) Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) sensor, which generates 145×145 pixels and 220 bands in the 0.4-2.45 m region, of which 20 noisy bands are removed before classification. It contains 16 ground-truth classes, and the class-specific numbers of test and training samples are shown in Table I. The second dataset is the University of Pavia, which is collected by the Reflective Optics System Imaging Spectrometer (ROSIS) sensor and contains a spatial coverage of 610×340 pixels. It generates 115 spectral bands, of which 12 noisy bands are removed. There are nine ground-truth classes of the dataset. The third dataset is the Salinas, collected by the 224-band Airborne Visible Infrared Imaging Spectrometer (AVIRIS) sensor over Salinas Valley, California, which generates 512×217 pixels

[1]OA is the number of correctly predicted pixel/total of pixel to predict.

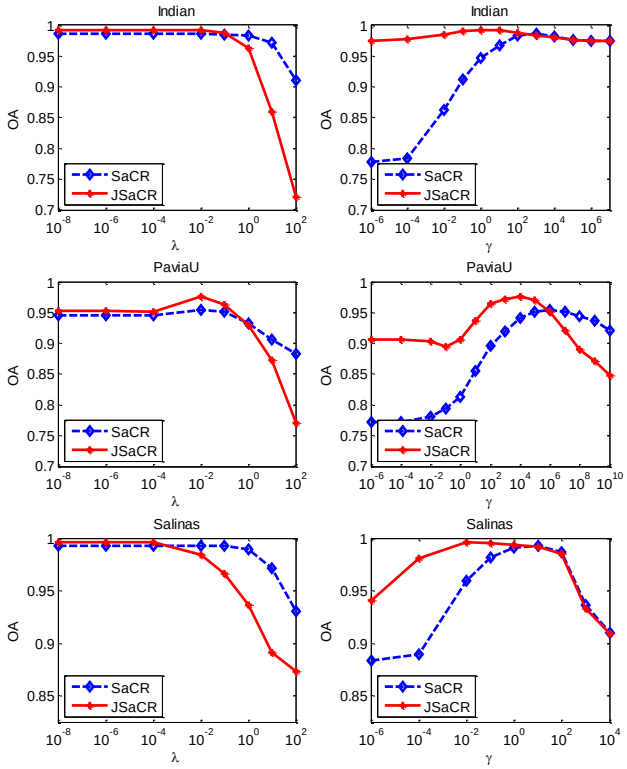

Fig. 1. Classification accuracy with varying regularization parameters λ and γ of SaCR and JSaCR on Indian Pine (top row), University of Pavia (middle row), and Salinas (bottom row), respectively.

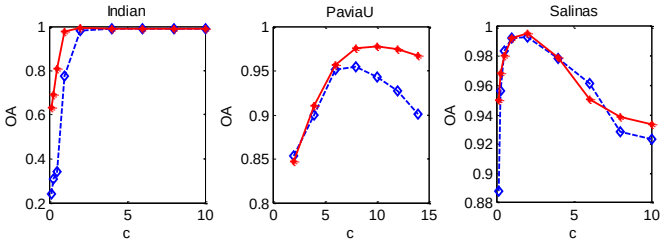

Fig. 2. Classification accuracy with varying regularization parameter c of SaCR and JSaCR on Indian Pine (left column), University of Pavia (middle column), and Salinas (right column), respectively.

and 204 bands over 0.4-2.5 m with spatial resolution of 3.7 m, of which 20 water absorption bands are removed before classification. For the above three datasets, the test and training data is randomly selected from the available ground truth maps. The class-specific numbers of test and training samples are shown in Table II, Table III, and Table IV. To avoid any bias, all the experiments are repeated 10 times, and we report the average classification accuracy.

B. Parameter Settings

To demonstrate the effectiveness of our proposed approach, we study the effect of the three regularization parameters λ, γ and c. In general, the fivefold cross-validation strategy based on training samples is considered for parameter tuning. Fig. 1 plots the curves of the OA values on the three HSI datasets as a function of the regularization λ and γ. It is worth noting that each left subfigure shows the OA values according to γ when λ is set to the optimal value, and each right subfigure presents the OA values according to λ when

TABLE II
CLASSIFICATION ACCURACY (%) FOR THE INDIAN PINE DATASET.

Class	# samples		Classification Algorithms								
	Train	Test	SVM	SVM-CK	SR	JSR	CR	WCR	JCR	SaCR	JSaCR
Alfalfa	5	41	39.02	92.68	10.00	25.00	56.10	36.59	97.56	92.68	100.00
Corn-notill	143	1285	58.29	89.56	53.38	62.96	51.05	83.50	98.21	95.10	98.05
Corn-mintill	83	747	51.27	91.47	45.95	52.86	16.20	53.15	98.93	98.93	99.33
Corn	24	213	36.15	79.62	32.50	37.50	5.16	27.70	99.06	100.00	100.00
Grass-pasture	48	435	82.76	93.95	82.50	88.75	71.49	87.59	97.93	96.55	97.93
Grass-trees	73	657	88.13	99.38	90.27	97.30	93.76	97.11	99.24	97.26	99.54
Grass-pasture-mowed	3	25	0.00	91.60	40.00	70.00	60.00	48.00	100.00	96.00	100.00
Hay-windrowed	48	430	95.35	99.74	91.25	100.00	94.42	98.84	100.00	100.00	100.00
Oats	2	18	0.00	76.11	30.00	40.00	27.78	22.22	94.44	38.89	100.00
Soybean-notill	97	875	49.49	89.57	56.73	64.69	20.69	69.03	97.83	96.91	98.40
Soybean-mintill	246	2209	80.26	95.27	70.08	80.08	89.72	82.66	98.19	99.55	99.50
Soybean-clean	59	534	37.27	86.39	43.67	53.67	6.93	59.74	97.75	99.06	99.25
Wheat	21	184	94.57	98.70	90.00	98.00	90.22	99.46	100.00	100.00	100.00
Woods	127	1138	93.41	98.59	88.41	96.35	97.10	95.34	99.82	100.00	100.00
Buildings-Grass-Trees-Drives	39	347	59.65	91.73	37.89	59.47	41.21	48.41	98.85	99.14	100.00
Stone-Steel-Towers	9	84	36.90	93.45	76.00	86.00	80.95	80.95	98.81	97.62	100.00
Overall Accuracy (%)			69.98	93.46	66.20	75.38	63.39	78.69	98.62	98.20	99.23

γ is set to the optimal value. From Fig. 1, we can see that: i) by setting proper values of λ or γ, JSaCR is better than SaCR, which implies the effectiveness of utilizing contextual information of center pixels; ii) when γ is set to the optimal value, the increase of λ brings small gains. However, when λ is set to the optimal value, the increase of γ brings relatively large gains. This indicates that the spatial induced penalty term plays a relatively more important role than the spectral induced penalty term does in the sample representation; iii) when $\gamma = 0$, the proposed SaCR method reduces to the WCRC method, the performance of SaCR is restricted. This can be explained by that the spatial induced constraint is essential for HSI classification task. Table I tabulates the parameter settings for the three HSI datasets when our proposed SaCR and JSaCR achieve the best performances.

In addition to the analysis of λ and γ, we also evaluate the performance of our proposed method with different smooth parameter c (as shown in Fig. 2). The smooth parameter c has an important influence on the classification accuracy. It cannot capture the spatial structure information if the smooth parameter is set too large or too small. To obtain the best performances, the Indian Pines and University of Pavia datasets employ larger values of c than the Salinas dataset. This is mainly because that the former two datasets have a lot of disconnected classes, while the latter exhibits more spatial homogeneity.

From the objective function (5), we learn that large values of γ and c mean that spatial constraint account for main contribution, and SaCR method tends to select a small number of neighbor pixels. This point could be learned from Table I. SaCR uses relatively small values of parameters γ and c for the Indian Pine dataset and the Salinas dataset (γ is 1e+4 or 10, c is 4 or 2), and large values for the Salinas dataset (γ is 1e+6 and c is 9).

C. Classification Performance

The performances of SaCR and JSaCR are shown in Table II, Table III and Table IV. From these results, we can conclude that the introduction of the contextual information greatly improves the performance of the original pixel based method, e.g., as for the Indian Pine dataset, SVM-CK has 23.48% gain over SVM, JSR has 9.18% gain over SR, JCR has 35.23% gain over CR, and JSaCR has 1.03% gain over SaCR. Similar results can be also observed from the other two datasets. The difference between JCR and our proposed JSaCR is that JSaCR additionally imposes a spatial feature induced regularization term. From the reported results, we can see that this spatial regularization is important in the representation residual based HSI classification method. The improvements of JSaCR over JCR are 0.51%, 5.19%, and 3.63% on the Indian Pine dataset, the University of Pavia dataset and the Salinas dataset, respectively.

V. CONCLUSIONS

In this letter, we proposed a novel CR HSI classification method based SaCR. It has a closed-form solution to incorporate the spatial and spectral information simultaneously. Meanwhile, we further developed a joint SaCR (JSaCR) modeling that takes into consideration of the contextual information of the center pixel. Extensive experimental results on three benchmark HSI datasets verified the effectiveness of incorporating the spatial feature induced regularization term. Comparison results demonstrated that our proposed JSaCR algorithm can obtain better performance than the state-of-the-art spectral-spatial HSI classification methods.

REFERENCES

[1] J. Lin, Q. Wang, and Y. Yuan, "In defense of iterated conditional mode for hyperspectral image classification," in *ICME*, July 2014, pp. 1–6.

[2] 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, June 2016.

[3] C. Chen, W. Li, H. Su, and K. Liu, "Spectral-spatial classification of hyperspectral image based on kernel extreme learning machine," *Remote Sensing*, vol. 6, no. 6, pp. 5795–5814, 2014.

[4] B. Du, M. Zhang, L. Zhang, R. Hu, and D. Tao, "Pltd: Patch-based low-rank tensor decomposition for hyperspectral images," *IEEE Transactions on Multimedia*, vol. 19, no. 1, pp. 67–79, Jan 2017.

[5] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sensing*, vol. 49, no. 10, pp. 3973–3985, 2011.

TABLE III
CLASSIFICATION ACCURACY (%) FOR THE UNIVERSITY OF PAVIA DATASET.

Class	# samples		Classification Algorithms								
	Train	Test	SVM	SVM-CK	SR	JSR	CR	WCR	JCR	SaCR	JSaCR
Asphalt	30	6601	75.22	86.24	52.71	59.64	72.58	63.26	87.74	83.56	92.11
Bare soil	30	18619	72.12	85.70	66.85	82.30	69.41	82.58	94.95	97.26	99.56
Bitumen	30	2069	67.76	85.12	60.95	51.43	66.75	67.47	85.94	98.79	98.89
Bricks	30	3034	96.31	94.50	93.46	74.51	94.03	95.81	93.31	93.31	96.47
Gravel	30	1315	93.61	99.59	100.00	97.01	99.32	99.77	100.00	99.92	100.00
Meadows	30	4999	58.43	87.59	62.95	64.94	60.25	58.91	95.58	99.84	100.00
Metal sheets	30	1300	84.38	95.80	91.04	74.63	54.77	96.08	99.08	99.38	99.23
Shadows	30	3652	70.02	84.02	57.07	66.85	68.29	81.52	86.12	98.36	98.63
Trees	30	917	98.69	99.97	95.74	78.72	86.48	98.04	96.18	94.77	98.04
Overall Accuracy (%)			73.94	87.51	67.40	73.48	71.20	78.19	92.90	95.42	98.09

TABLE IV
CLASSIFICATION ACCURACY (%) FOR THE SALINAS DATASET.

Class	# samples		Classification Algorithms								
	Train	Test	SVM	SVM-CK	SR	JSR	CR	WCR	JCR	SaCR	JSaCR
Brocoli_green_weeds_1	30	1979	93.53	99.80	99.00	99.55	97.83	99.55	99.95	100.00	99.90
Brocoli_green_weeds_2	30	3696	96.43	99.68	99.46	99.40	99.30	99.89	100.00	100.00	100.00
Fallow	30	1946	84.58	99.79	88.89	97.79	85.10	95.73	100.00	100.00	100.00
Fallow_rough_plow	30	1364	98.83	99.71	95.71	99.19	98.53	99.12	97.21	99.85	98.90
Fallow_smooth	30	2648	94.60	98.15	90.30	84.97	98.15	97.13	98.94	98.98	98.83
Stubble	30	3929	97.94	99.82	100.00	100.00	99.21	99.77	100.00	99.85	100.00
Celery	30	3549	97.72	99.58	99.44	99.75	99.38	99.66	100.00	99.92	100.00
Grapes_untrained	30	11241	55.24	89.56	49.65	61.18	69.12	75.20	91.90	99.42	99.80
Soil_vinyard_develop	30	6173	96.03	97.99	97.10	99.21	95.43	99.92	100.00	100.00	100.00
Corn_senesced_green_weeds	30	3248	82.82	95.72	87.80	89.56	92.76	92.95	97.72	98.21	98.00
Lettuce_romaine_4wk	30	1038	96.53	98.94	94.34	98.17	92.10	92.49	100.00	98.36	99.81
Lettuce_romaine_5wk	30	1897	97.73	100.00	89.58	97.36	99.89	99.84	99.84	99.95	99.68
Lettuce_romaine_6wk	30	886	96.50	99.10	82.61	97.40	97.07	97.97	100.00	98.87	99.21
Lettuce_romaine_7wk	30	1040	91.54	97.98	92.59	93.85	93.94	92.88	99.04	97.02	99.33
Vinyard_untrained	30	7238	57.94	89.10	60.88	66.34	67.02	71.50	85.04	97.72	99.28
Vinyard_vertical_trellis	30	1777	97.97	99.04	98.89	99.94	99.27	98.87	99.61	100.00	100.00
Overall Accuracy (%)			81.37	95.57	81.12	85.39	86.87	89.78	95.98	99.28	99.61

[6] 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 Sensing*, vol. 52, no. 6, pp. 3707–3719, 2014.

[7] J. Li, H. Zhang, L. Zhang, X. Huang, and L. Zhang, "Joint collaborative representation with multitask learning for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 52, no. 9, pp. 5923–5936, 2014.

[8] W. Li and Q. Du, "Joint within-class collaborative representation for hyperspectral image classification," *IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2200–2208, 2014.

[9] 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, 2015.

[10] M. Xiong, Q. Ran, W. Li, J. Zou, and Q. Du, "Hyperspectral image classification using weighted joint collaborative representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 6, pp. 1209–1213, 2015.

[11] S. Jia, L. Shen, and Q. Li, "Gabor feature-based collaborative representation for hyperspectral imagery classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 53, no. 2, pp. 1118–1129, 2015.

[12] J. Jiang, C. Chen, X. Song, and Z. Cai, "Hyperspectral image classification using set-to-set distance," in *ICASSP*, March 2016, pp. 3346–3350.

[13] 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, 2009.

[14] J. Ma, J. Zhao, J. Tian, X. Bai, and Z. Tu, "Regularized vector field learning with sparse approximation for mismatch removal," *Pattern Recognition*, vol. 46, no. 12, pp. 3519–3532, 2013.

[15] L. Wang, K. Lu, P. Liu, R. Ranjan, and L. Chen, "Ik-svd: dictionary learning for spatial big data via incremental atom update," *Computing in Science & Engineering*, vol. 16, no. 4, pp. 41–52, 2014.

[16] J. Jiang, X. Ma, Z. Cai, and R. Hu, "Sparse support regression for image super-resolution," *IEEE Photonics Journal*, vol. 7, no. 5, pp. 1–11, Oct 2015.

[17] J. Ma, C. Chen, C. Li, and J. Huang, "Infrared and visible image fusion via gradient transfer and total variation minimization," *Inf. Fusion*, vol. 31, pp. 100–109, 2016.

[18] J. Jiang, J. Ma, C. Chen, X. Jiang, and Z. Wang, "Noise robust face image super-resolution through smooth sparse representation," *IEEE Trans. on Cybern.*, vol. PP, no. 99, pp. 1–12, 2016.

[19] W. Song, P. Liu, and L. Wang, "Sparse representation-based correlation analysis of non-stationary spatiotemporal big data," *International Journal of Digital Earth*, vol. 9, no. 9, pp. 892–913, 2016.

[20] J. Wei, Y. Huang, K. Lu, and L. Wang, "Nonlocal low-rank-based compressed sensing for remote sensing image reconstruction," *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 10, pp. 1557–1561, 2016.

[21] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *ICCV*, 2011, pp. 471–478.

[22] Y. Yuan, J. Lin, and Q. Wang, "Hyperspectral image classification via multitask joint sparse representation and stepwise MRF optimization," *IEEE Trans. Cybernetics*, vol. PP, no. 99, pp. 1–12, 2015.

[23] J. Liu, Z. Wu, J. Li, A. Plaza, and Y. Yuan, "Probabilistic-kernel collaborative representation for spatial-spectral hyperspectral image classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 54, no. 4, pp. 2371–2384, April 2016.

[24] H. Pu, Z. Chen, B. Wang, and G. M. Jiang, "A novel spatial-spectral similarity measure for dimensionality reduction and classification of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sensing*, vol. 52, no. 11, pp. 7008–7022, Nov 2014.

[25] C. Li, Y. Ma, X. Mei, C. Liu, and J. Ma, "Hyperspectral image classification with robust sparse representation," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 5, pp. 641–645, May 2016.

[26] W. Li, E. W. Tramel, S. Prasad, and J. E. Fowler, "Nearest regularized subspace for hyperspectral classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 52, no. 1, pp. 477–489, 2014.

[27] R. Archibald and G. Fann, "Feature selection and classification of hyperspectral images with support vector machines," *IEEE Geosci. Remote Sens. Lett.*, vol. 4, no. 4, pp. 674–677, 2007.

[28] G. Camps-Valls, L. Gomez-Chova, J. Muñoz-Marí, J. Vila-Francés, and J. Calpe-Maravilla, "Composite kernels for hyperspectral image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 3, no. 1, pp. 93–97, 2006.