Fast Multifeature Joint Sparse Representation for Hyperspectral Image Classification

Erlei Zhang, Xiangrong Zhang, *Member, IEEE*, Hongying Liu, *Member, IEEE*, and Licheng Jiao, *Senior Member, IEEE*

***Abstract*—Since hyperspectral images (HSIs) usually have complex content and chaotic background, multiple kinds of features would be helpful for the classification task. Recently, representation-based methods with multifeature combination learning have been proposed. However, multifeature learning and the extended contextual information require much more computational burden, particularly for a large-scale dictionary case. In this letter, we propose a fast joint sparse representation classification method with multifeature combination learning for hyperspectral imagery. Once getting several complementary features (spectral, shape, and texture), the proposed model simultaneously acquires a representation vector for each kind of feature and imposes the joint sparsity $\ell_{\mathrm{row},0}$-norm regularization on the representation coefficients. The regularization can enforce the coefficients to share a common sparsity pattern, which preserves the cross-feature information. A new version of the simultaneous orthogonal matching pursuit is presented to solve the aforementioned problem because of its optimization with strong convergence guarantee and efficiency. Moreover, to further improve the classification performance, we incorporate contextual neighborhood information of the image into each kind of feature. Compared with state-of-the-art algorithms, it has been proved that the proposed algorithm with much less memory requirements performs tens to hundreds of times faster than those on real HSIs, while providing the same (or even better) accuracy.**

***Index Terms*—Feature extraction, hyperspectral image (HSI) classification, joint sparse representation, multifeature, orthogonal matching pursuit (OMP).**

I. Introduction

A HYPERSPECTRAL image (HSI) is obtained by an imaging spectrometer with a large spectral wavelength range, spanning the visible spectrum to the infrared one. HSI has paved the way for remote sensing applications in various fields [1]. Classification of HSI is an extremely important task for image understanding. HSI classification process aims at converting data, a high-dimensional vector, into meaningful information where pixels are classified into certain classes.

Recently, sparse representation-based models have aroused much concern. As a signal processing method, through taking advantage of the signal's sparseness in some domain and solving the underdetermined linear system, compressed sensing theory [2] can reconstruct the entire signal. In spite of the high dimensionality of HSI, the signals in the same class usually lie in a low-dimensional subspace, which is spanned by the dictionary atoms (training samples) of the same class. The sparse representation has been also introduced into HSI classification [3]–[6]. In classification, an unknown test pixel can be sparsely represented by a few atoms from one training dictionary. The recovered sparse coefficients, including the positions and the weight values of those selected atoms, determine the class label of the tested pixels by the minimal reconstruction error.

Although the preceding methods have good performance, one single kind of feature can only depict the HSI from one perspective. In addition, it is obvious that none of the common feature descriptors has the same discriminative power for all classes. In order to solve this problem, some representation-based methods have been extended to multitask learning (MTL) [7]–[11]. The basic idea of representation-based MTL is to find out a very few common classes of training samples that are most correlated to the query sample from related representation models [8], [9]. Since different tasks (e.g., each kind of feature is one task) may support different sparse representation coefficients, the constraint of joint sparsity across different tasks provides additional useful information to solve the classification problem. However, time consumption of sparsity constraint with ℓ_1-norm is still a problem, although these methods obtain an improved recognition rate [8], [9]. In [10], the authors proposed a joint collaborative representation (CR) model, which was similar to the sparse representation classifier (SRC), with the ℓ_2-norm instead of the ℓ_1-norm [12]. However, when dealing with a large-scale training sample set as the dictionary, CR-based MTL (CRC-MTL) and joint CRC-MTL (JCRC-MTL), utilizing the matrix inverse operation, require more computational time and memory cost [10]. Thus, although these methods achieve state-of-the-art performance, it is necessary to develop a fast method.

Motivated by the multitask joint representation models mentioned above, in this letter, we propose a fast multifeature joint SRC (MF-JSRC) for HSI classification. The proposed

Manuscript received October 28, 2014; revised December 29, 2014; accepted February 2, 2015. This work was supported in part by the National Basic Research Program of China (973 Program) under Grant 2013CB329402; by the National Natural Science Foundation of China under Grant 61272282, Grant 61203303, Grant 61272279, Grant 61377011, and Grant 61373111; by the Program for New Century Excellent Talents in University under Grant NCET-13-0948; by the Program for New Scientific and Technological Star of Shaanxi Province under Grant 2014KJXX-45; and by the China Postdoctoral Science Foundation under Grant 65ZY1425.

The authors are with the Key Laboratory of Intelligent Perception and Image Understanding of the Ministry of Education of China, Institute of Intelligent Information Processing, Xidian University, Xi'an 710071, China (e-mail: zhangerlei0123@163.com; xrzhang@mail.xidian.edu.cn; hyliu@xidian.edu.cn).

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

Digital Object Identifier 10.1109/LGRS.2015.2402971

method not only can utilize the $\ell_{\mathrm{row},0}$-norm penalty to enforce joint sparsity across multiple features representation tasks but also can obtain significant gains in speed with less memory requirements. Once getting several complementary features (spectral, shape, and texture), the proposed method simultaneously acquires a representation vector for each kind of feature and imposes the joint sparsity $\ell_{\mathrm{row},0}$-norm regularization on the representation coefficients. The regularization encourages the coefficients to share a common sparsity pattern, which can preserve the cross-feature information. The formulated objective consists of a squared reconstruction error term and a sparse $\ell_{\mathrm{row},0}$-norm regularization term. Multifeature joint sparse recovery problems are NP-hard problems, but they can be approximately worked out by greedy algorithms, among which we mostly refer to the orthogonal matching pursuit (OMP) [13] because of its optimization with strong convergence guarantee. Another advantage of OMP algorithm is that it can avoid the large matrix inverse operation because of parts of dictionaries selected for computation. That is why our method has less computational complexity and memory requirements. In addition, neighborhood information is integrated in the multifeature joint SR framework by constructing a joint signal set composed of neighborhood pixels around the unlabeled pixel. The validity of the proposed MF-JSRC algorithm is confirmed by experiments on real HSIs.

The remainder of this letter is structured as follows. Details of the proposed fast joint sparsity representation-based method are described in Section II. The efficiency and effectiveness of the proposed method are demonstrated in Section III by experimental results on real HSI. Finally, Section IV summarizes this letter and presents some remarks.

II. Fast Multifeature Joint Sparse Representation

A. Multifeature Pixelwise Sparse Representation Model

The sparse representation model [14] assumes that samples belonging to the same class approximately lie in a low-dimensional subspace, which is spanned by the training samples of the same class. In other words, a signal can be approximated by a sparse linear combination of a few atoms from a given dictionary. In the multiple features case, we suppose that each pixel has S different kinds of features. For the unlabeled pixel $\boldsymbol{x}$, we denote $\boldsymbol{x}^s \in \Re^{b^s}$ as the sth ($s = 1, \ldots, S$) feature vector, where b^s is the dimension of the sth feature vector. $\boldsymbol{D}^s = [\boldsymbol{D}_1^s, \ldots, \boldsymbol{D}_c^s, \ldots, \boldsymbol{D}_C^s]$ is the dictionary of the sth feature ($\boldsymbol{D}_c^s \in \Re^{b^s \times N_c}$, $c = 1, \ldots, C$, is the cth class subdictionary whose columns (atoms) are extracted from the training samples; C is the number of classes; N_c is the number of atoms in subdictionary $\boldsymbol{D}_c^s$; $N = \sum_{c=1}^{C} N_c$). $\boldsymbol{\alpha}^s \in \Re^N$ is the coding vector of $\boldsymbol{x}^s$ over $\boldsymbol{D}^s$. The representation $\boldsymbol{\alpha}^s$ satisfying $\boldsymbol{x}^s = \boldsymbol{D}^s\boldsymbol{\alpha}^s$ is obtained by solving the following optimization problem:

$$\hat{\boldsymbol{\alpha}}^s = \arg\min_{s=1,\ldots,S} \|\boldsymbol{x}^s - \boldsymbol{D}^s\boldsymbol{\alpha}^s\|_2 \quad \text{s.t.} \quad \|\boldsymbol{\alpha}^s\|_0 \le K \tag{1}$$

where K is a predefined upper bound on the sparsity level, representing the maximum number of the nonzero coefficients in $\hat{\boldsymbol{\alpha}}^s$.

Since each kind of feature of the same training samples constructs corresponding subdictionaries $\{\boldsymbol{D}^s\}_{s=1,\ldots,S}$, respectively, it is reasonable that these features may share some similarities. Therefore, it can be assumed that the representation coefficients of these features $\{\boldsymbol{\alpha}^s\}_{s=1,\ldots,S}$ over their associated subdictionaries should be similar. We assume that the positions of nonzero coefficients tend to be the same. On the other hand, because different types of features differ from each other, the representation coefficients need not be identical. Thus, they can preserve additional complementary information. Under this assumption, the multifeature pixelwise sparse representation model (MF-SRC) can be formulated as

$$\hat{\boldsymbol{A}} = \arg\min \sum_{s=1}^{S} \|\boldsymbol{x}^s - \boldsymbol{D}^s\boldsymbol{\alpha}^s\|_2 \quad \text{s.t.} \quad \|\boldsymbol{A}\|_{\mathrm{row},0} \le K \tag{2}$$

where $\boldsymbol{A} = [\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^S]$.

Once $\hat{\boldsymbol{A}}$ is obtained, we can calculate the residual errors between $\boldsymbol{x}^s$ $(s = 1, \ldots, S)$ and the approximations obtained over their corresponding subdictionaries $\{\boldsymbol{D}_i^s\}_{\substack{s=1,\ldots,S \\ i=1,\ldots,C}}$, i.e.,

$$\text{label}(\boldsymbol{x}) = \arg\min_{i=1,\ldots,C} r_i(\boldsymbol{x}) = \arg\min_{i=1,\ldots,C} \sum_{s=1}^{S} \|\boldsymbol{x}^s - \boldsymbol{D}_i^s\hat{\boldsymbol{\alpha}}_i^s\|_2 \tag{3}$$

where $\hat{\boldsymbol{\alpha}}_i^s$ is the subset of the coefficient vector $\hat{\boldsymbol{\alpha}}^s$ associated with class i.

B. Multifeature Joint Sparse Representation Model

The MF-SRC model (proposed in Section II-A) can take the spatial information into consideration by these meaningful features of the unlabeled pixel. In fact, since HSI pixels in a small spatial neighborhood tend to be highly correlated and share many similarities, neighboring pixels are always assumed consisting of similar materials. Previous work in [3] suggested that, by combining these contextual neighboring pixels with the unlabeled pixel into JSRC, the classification accuracy can be significantly improved. In the case of only the spectral value feature, these pixels in a neighborhood can be simultaneously approximated by a linear combination of common training pixels, which are assigned a different set of coefficients. Similarly, in multiple-feature cases, these neighboring pixels can be also used for every kind of feature. That is, pixels $\{\boldsymbol{x}_t^s\}_{t=1,\ldots,L}$ (where L is the neighborhood size) of every type of feature (e.g., spectral value, Gray-Level Co-occurrence Matrix (GLCM), Differential Morphological Profiles (DMP) shape feature, and 3-D Wavelet Transform (3D-WT) for the HSI) in a spatial neighborhood are approximated by a linear combination of common atoms from the given dictionary $\boldsymbol{D}^s$. A visual illustration of the proposed multifeature joint sparse model is shown in Fig. 1. The model represents the test data by a sparse linear combination of training data. Sparse coefficients of different pixels in a small region from different types of features share the same sparsity pattern.

Suppose we get a matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t, \ldots, \boldsymbol{x}_L]$ via simultaneously stacking all the pixels in the neighborhood pixels centered at the hyperspectral pixel $\boldsymbol{x}_t$ of size L. For S different

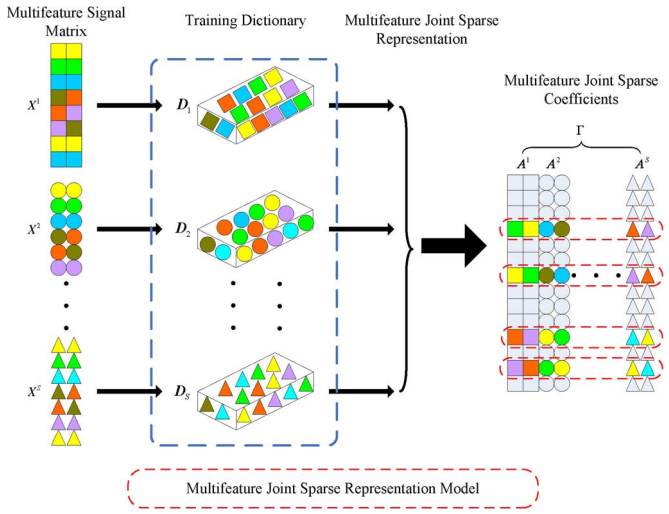

Fig. 1. Illustration of MF-JSRC [see formula (5)]. Each patch (square, circle, and triangular) in the column is a coefficient value, and each column represents a sparse coefficient vector corresponding to one pixel, in which white blocks denote zero values, whereas color blocks stand for nonzero values.

types of features, in the same way, we can construct the matrix $\{\boldsymbol{X}^s\}_{s=1,\ldots,S} = \{[\boldsymbol{x}_1^s, \ldots, \boldsymbol{x}_t^s, \ldots, \boldsymbol{x}_L^s]\}_{s=1,\ldots,S}$, which contains S matrices sized $b^s \times L$ for each neighborhood patch. In this letter, we construct dictionaries $\{\boldsymbol{D}^s\}_{s=1,\ldots,S}$ with the training pixels, including multiple features. Using the JSRC model, $\{\boldsymbol{X}^s\}_{s=1,\ldots,S}$ can be represented by

$$\boldsymbol{X}^s = [\boldsymbol{x}_1^s, \boldsymbol{x}_2^s, \ldots, \boldsymbol{x}_L^s] = \boldsymbol{D}^s \boldsymbol{A}^s \tag{4}$$

where $\{\boldsymbol{A}^s\}_{s=1,\ldots,S} \in \Re^{N\times L}$ is a set of the coding coefficient matrices associated with the corresponding feature dictionary $\{\boldsymbol{D}^s\}_{s=1,\ldots,S}$. Two constraints are taken into consideration in this scheme. The first constraint assumes that the signals in the same class usually lie in a low-dimensional subspace, which is spanned by the dictionary atoms (training samples) of the same class. The second one supposes that the hyperspectral pixels in a small neighboring window should share the same low-dimensional dominant subspace.

Let $\boldsymbol{\Gamma} = [\boldsymbol{A}^1, \boldsymbol{A}^2, \ldots, \boldsymbol{A}^S]$ be the matrix formed by concatenating the coefficient matrices. Then, following the joint model with multifeature learning, we can determine the row-sparse matrix $\boldsymbol{\Gamma}$, i.e.,

$$\hat{\boldsymbol{\Gamma}} = \arg\min_{\boldsymbol{\Gamma}} \sum_{s=1}^{S} \|\boldsymbol{X}^s - \boldsymbol{D}^s \boldsymbol{A}^s\|_F \quad \text{s.t.} \quad \|\boldsymbol{\Gamma}\|_{\text{row},0} \le K. \tag{5}$$

The first penalty aims at minimizing the representation error for all kinds of features. The second penalty directs the minimization task toward the sparsest possible representation. The aforementioned problems are too complex, but we commonly use approximation algorithm such as OMP [13] to solve them. The OMP algorithm augments the support set by choosing one atom at each time until the approximation error decreases to the preset threshold or K atoms are selected.

After obtaining $\hat{\boldsymbol{\Gamma}}$, we can calculate the residual errors between $\boldsymbol{X}^s (s = 1, \ldots, S)$ and the approximations obtained over their corresponding subdictionaries $\{\boldsymbol{D}_i^s\}_{\substack{s=1,\ldots,S \\ i=1,\ldots,C}}$. Then, the label of the center pixel $\boldsymbol{x}_t$ is determined to be the class that yields the minimal total residual as

$$\text{label}(\boldsymbol{x}_t) = \arg\min_{i=1,\ldots,C} r_i(\boldsymbol{X}^s) = \arg\min_{i=1,\ldots,C} \sum_{s=1}^{S} \|\boldsymbol{X}^s - \boldsymbol{D}_i^s \boldsymbol{A}_i^s\|_F \tag{6}$$

where $\boldsymbol{A}_i^s$ is the subset of the coefficient matrix $\boldsymbol{A}^s$ associated with class i.

Algorithm 1: Multifeature Joint Sparse Representation Model

Input: multifeature test matrix $\{\boldsymbol{X}^s\}_{s=1,\ldots,S}$, structural dictionary $\{\boldsymbol{D}^s\}_{s=1,\ldots,S}$, number of classes C, sparsity level K.
Output: $\hat{\boldsymbol{\Gamma}}$, joint sparse representation coefficients matrix.
Initialization: set iteration counter $iter = 1$; index set $\Lambda_0 = \phi$; residual matrix $\{\boldsymbol{R}_0^s\}_{s=1,\ldots,S} = \{\boldsymbol{X}^s\}_{s=1,\ldots,S}$.
While stopping criterion has not been met **do**
Step 1: Computer residual correlation matrix for each feature: $\boldsymbol{Z}(s,i) = \|(\boldsymbol{R}_{iter-1}^s)^T \boldsymbol{d}_i^s\|_p$, $s = 1, \ldots, S$, $i = 1, \ldots, N$, $p \ge 1$, $\boldsymbol{d}_i^s$ is ith atom of $\boldsymbol{D}^s$;
Step 2: Find index of the atom that best approximates all residual: $\lambda_{iter} = \arg\max_{i=1,2,\ldots,N} \|\boldsymbol{Z}(:,i)\|_q$, $q \ge 1$;
Step 3: Update the index set: $\boldsymbol{\Lambda}_{iter} = \boldsymbol{\Lambda}_{iter-1} \cup \lambda_{iter}$;
Step 4: Estimate the sparse representation coefficients $\boldsymbol{A}_{iter}^s$:

$$\boldsymbol{A}_{iter}^s = \left(\left(\boldsymbol{D}_{\boldsymbol{\Lambda}_{iter}}^s\right)^T \boldsymbol{D}_{\boldsymbol{\Lambda}_{iter}}^s\right)^{-1} \left(\boldsymbol{D}_{\boldsymbol{\Lambda}_{iter}}^s\right)^T \boldsymbol{X}^s, \; s = 1, \ldots, S;$$

Step 5: Update the residual matrix: $\boldsymbol{R}_{iter}^s = \boldsymbol{X}^s - \boldsymbol{D}_{\boldsymbol{\Lambda}_{iter}}^s \boldsymbol{A}_{iter}^s$, $s = 1, \ldots, S$;
Step 6: Update joint sparse representation coefficients:

$$\hat{\boldsymbol{\Gamma}} = \left[\boldsymbol{A}_{iter}^1, \boldsymbol{A}_{iter}^2, \ldots, \boldsymbol{A}_{iter}^S\right], s = 1, \ldots, S;$$

Step 7: $iter \leftarrow iter + 1$.
End while

III. Experimental Results and Discussion

We evaluate the proposed MF-SRC and MF-JSRC on a real HSI. In addition, we make a comparison between existing methods (including CR- and SR-based methods) and the proposed methods.

The real HSI in our experiments is the commonly used Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Indian Pines image. The AVIRIS sensor generates 220 bands across the spectral range from 0.2 to 2.4 μm. In the experiments, the number of bands is reduced to 200 by removing 20 water absorption bands. This image has a spatial resolution of 20 m per pixel and spatial dimension of 145×145. It contains 16 ground truth classes, most of which are different types of crops (e.g., corns, soybeans, and wheat). A total size of $n = 1036$ (10% of each class) samples are used for training and the remaining 9330 samples for test.

TABLE I
PARAMETERS FOR FEATURES

Feature	Parameters	Dimension
GLCM	Base image: PC1, PC2, PC3, PC4; Measure: angular second moment, contrast, entropy, variance, correlation; Window: 3,7,11; Direction: averaging the extracted features over four directions.	60
DMP	Base image: PC1, PC2 ; Size of structuring elements: 3,5,7,9 ; Morphological operators: opening and closing.	16
3D-WT	Base image :PC1, PC2, PC3, PC4; Level : 2.	60

TABLE II
CLASSIFICATION ACCURACY (%) FOR THE INDIAN PINES IMAGE ON THE TEST SET

	Methods	OA	AA	Kappa
Single feature methods	CRC [12]	70.67	54.06	0.658
	JCRC [12]	78.59	59.95	0.751
	SRC [14]	76.58	72.29	0.733
	JSRC [3]	91.30	88.46	0.901
Multifeature methods	CRC-MTL [10]	87.93	75.10	0.862
	JCRC-MTL [10]	91.03	78.58	0.897
	SRC-MTL [8, 9]	86.01	83.85	0.839
	JSRC-MTL [8, 9]	95.02	86.41	0.943
	MF-SRC	91.75	93.57	0.906
	MF-JSRC	**97.78**	**97.82**	**0.975**

In our experiments, four different types of features (original spectral value feature, GLCM feature [15], DMP feature [16], and 3D-WT feature [6]) are used to describe each pixel. Due to space limitation, the detailed feature extraction procedure can be found in [6], [15], and [16]; and an overview of feature extraction methods can be found in [17]. The parameter values for different kinds of features are set to be the same as the corresponding references and listed in Table I.

Here, the proposed MF-JSRC classifier is compared with the widely used classification methods, including CR-based methods (CRC, JCRC [12], CRC-MTL, and JCRC-MTL [10]) and SR-based methods (SRC-Pixel-wise [14], JSRC [3], SRC-MTL, and JSRC-MTL [8], [9]). The regularization parameters for the CRC, JCRC, CRC-MTL, and JCRC-MTL algorithms range from $1e^{-6}$ to $1e^{-1}$. SRC-Pixel-wise utilizes the spectral information only, whereas JSRC uses the local spatial information within one fixed single scale for classification. The regularization parameters for the SRC-MTL and JSRC-MTL algorithms range from $1e^{-6}$ to $1e^{-1}$. The neighborhood size for JCRC, JCRC-MTL, JSRC, JSRC-MTL, and MF-JSRC is set 3×3.

First, we analyze the classification results of the proposed MF-JSRC algorithm and the other representation-based methods. We do the experiment ten times to get the average results shown in Table II. Comparing results of multifeature-based methods with the ones of single-feature-based methods, it can be seen that all feature combination methods dramatically improve the classification performance. The proposed MF-JSRC method outperforms other compared approaches in terms of the overall accuracy (OA), the average accuracy (AA), and the κ coefficient measure (Kappa). Compared with JCRC-MTL and JSRC-MTL, the gains (in OA, AA, and Kappa coefficients) of MF-JSRC are more than 6% and 2% respectively, which demonstrate the effectiveness of the proposed multifeature strategy. In addition, it is also shown that, with the help of the neighboring pixels, the classification accuracy (JCRC, JSRC, JCRC-MTL, JSRC-MTL, and MF-JSRC) can be further improved than those methods without considering neighboring pixels (CRC, SRC, CRC-MTL, SRC-MTL, and MF-SRC). Overall, it demonstrates that the proposed MF-JSRC algorithm achieves the best performance.

TABLE III
RUNNING TIME (SECONDS) FOR THE CLASSIFICATION OF THE INDIAN PINES IMAGE

	Percentages (%) of training samples							
Methods	5	10	15	20	25	30	35	40
CRC	0.12	0.28	0.38	0.55	1.02	1.39	--	--
JCRC	0.45	0.82	1.22	1.61	2.37	2.58	--	--
SRC	0.15	0.30	0.44	0.60	1.07	1.53	1.81	2.02
JSRC	0.49	0.81	1.22	1.76	2.55	2.93	3.24	3.85
CRC-MTL	29.30	210.52	690.46	--	--	--	--	--
JCRC-MTL	30.79	219.01	696.01	--	--	--	--	--
SRC-MTL	28.24	76.47	143.09	230.61	336.10	460.41	--	--
JSRC-MTL	53.25	167.31	323.55	550.75	754.07	--	--	--
MF-SRC	0.37	0.64	1.34	2.06	2.50	3.08	3.50	3.89
MF-JSRC	1.02	1.83	2.84	4.13	5.32	5.99	7.83	7.88

Next, we compare the running time of the various classification algorithms. All the programs are executed using MATLAB in the environment of an Intel Core i3-550 CPU 3.19 GHz and 4 GB of RAM. For the Indian Pines image, different percentages (from 5% to 40% per class) of samples are randomly chosen to construct a training sample set, and 100 samples are randomly chosen to be test samples. Table III shows the average running time for each case of each classifier in detail. The "- -" symbol means the lack of results because of insufficient memory during calculation.

As shown in Tables II and III, CRC is the fastest but serves the worst classification performance; JCRC and JSRC have better classification performance than CRC and SRC, respectively, while they need more time; JSRC costs comparable time and achieves better classification results than JCRC. SRC-MTL and JSRC-MTL utilize accelerated proximal gradient method [9], whereas CRC-MTL and JCRC-MTL, including the matrix inverse operation, are slower than SRC-MTL and JSRC-MTL, respectively. It should be also noted that CR-based methods, i.e., the SRC-MTL and JSRC-MTL algorithms, lack results when dealing with a large-scale training sample set as the dictionary. There are some reasons for explaining this phenomenon. MTL and the extended contextual information increase the computational load [10]. When dealing with a large-scale training sample set as the dictionary, CRC-MTL and JCRC-MTL, utilizing the matrix inverse operation, will require more computing time and computer hardware. For the SRC-MTL and JSRC-MTL methods, satisfying recognition accuracy can be obtained within a few hundred times of iteration [9].

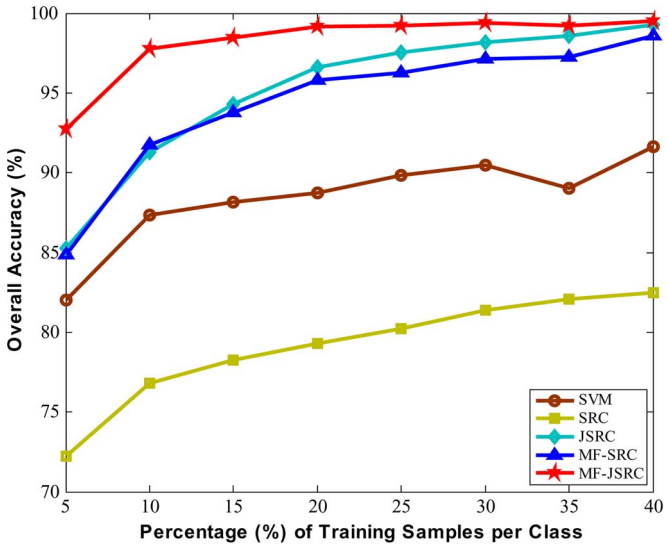

Fig. 2. Effect of the number of training samples for the Indian Pines image.

Thus, the SRC-MTL and JSRC-MTL methods require much computing time. The optimization problems in the proposed MF-SRC and MF-JSRC methods are NP-hard, but they can be approximately solved by greedy pursuit algorithms [13]. Our methods, using OMP and subdictionary selected, have strong convergence guarantee and can avoid large matrix inverse operation. Thus, the proposed MF-SRC and MF-JSRC can achieve comparable classification accuracy and perform tens to hundreds of times faster than the other multifeature methods.

Finally, the effects of different numbers of training samples to MF-SRC and MF-JSRC are examined on the Indian Pines image and shown in Fig. 2. Support vector machines [18], SRC-Pixel-wise, and JSRC are used as baseline methods. Different percentages (from 5% to 40% per class) of samples are randomly chosen to construct a training sample set, and the remaining samples are used as test samples. The results of accuracy value are averaged over five runs. As shown in Fig. 2, with the training samples increase, the performances of SVM, SRC-Pixel-wise, single-feature JSRC, and the proposed MF-SRC, MF-JSRC classifiers generally improve. Overall, the proposed MF-JSRC method can consistently outperform the other approaches on all the training samples.

IV. Conclusion

This letter has presented a fast joint sparse representation classification method with multifeature combination for HSI. The proposed model simultaneously represents pixels of multifeature with an $\ell_{\mathrm{row},0}$-norm regularization. Moreover, an efficient algorithm based on Simultaneous Orthogonal Matching Pursuit (SOMP) is proposed to solve the optimization problem. In addition, the proposed MF-JSRC utilizes neighborhood information to further improve the performance for classification. On real HSIs, the extensive experimental results confirm that the MF-JSRC algorithm can provide a satisfying classification performance with fast speed.

It should be pointed out that the proposed MF-JSRC algorithm could still be further improved in certain aspects, for instance, adaptively selecting the more-correlated neighboring pixels to the test sample, acquiring adaptive weights for each kind of feature, and dealing with multisource feature space (spectral, optical, and radar features).

References

[1] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 3, pp. 480–491, Mar. 2005.

[2] D. L. Donoho , "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] 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.

[4] U. Srinivas, Y. Chen, V. Monga, N. M. Nasrabadi, and T. D. Tran, "Exploiting sparsity in hyperspectral image classification via graphical models," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no. 3, pp. 505–509, May 2013.

[5] 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.

[6] Y. Qian, M. Ye, and J. Zhou, "Hyperspectral image classification based on structured sparse logistic regression and three-dimensional wavelet texture features," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2276–2291, Apr. 2013.

[7] S. Shekhar, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Joint sparse representation for robust multimodal biometrics recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 113–126, Jan. 2014.

[8] X. Zheng, X. Sun, K. Fu, and H. Wang, "Automatic annotation of satellite images via multifeature joint sparse coding with spatial relation constraint," *IEEE Geosci. Remote Sens. Lett.*, vol. 10, no.4 pp. 652–656, Jul. 2013.

[9] X.-T. Yuan, X. Liu, and S. Yan, "Visual classification with multitask joint sparse representation," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4349–4360, Oct. 2012.

[10] 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 Sens.*, vol. 52, no. 9, pp. 5923–5936, Sep. 2014.

[11] Y. Meng, L. Zhang, D. Zhang, and S. Wang, "Relaxed collaborative representation for pattern classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 2012, pp. 2224–2231.

[12] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang, "Collaborative representation based classification for face recognition," Unpublished Tech. Rep., 2012, [Online]. Available: http://arxiv.org/abs/1204.2358

[13] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

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

[15] F. Pacifici, M. Chini, and W. J. Emery, "A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification," *Remote Sens. Environ.*, vol. 113, no. 6, pp. 1276–1292, Jun. 2009.

[16] D. Tuia, F. Pacifici, M. Kanevski, and W. J. Emery, "Classification of very high spatial resolution imagery using mathematical morphology and support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 47, no. 11, pp. 3866–3879, Nov. 2009.

[17] X. Jia, B.-C. Kuo, and M. M. Crawford, "Feature mining for hyperspectral image classification," *Proc. IEEE*, vol. 101, no. 3, pp. 676–697, Mar. 2013.

[18] F. Melgani and L. Bruzzone, "Classification of hyperspectral remote sensing images with support vector machines," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 8, pp. 1778–1790, Aug. 2004.