Improving Hyperspectral Image Classification Using Spectral Information Divergence

Erlei Zhang, Xiangrong Zhang, *Member, IEEE,* Shuyuan Yang, *Member, IEEE,* and Shuang Wang, *Member, IEEE*

***Abstract*—In order to improve the classification performance for hyperspectral image (HSI), a sparse representation classifier based on spectral information divergence (SID) is proposed. SID measures the discrepancy of probabilistic behaviors between the spectral signatures of two pixels from the aspect of information theory, which can be more effective in preserving spectral properties. Thus, the new method measures the similarity between the reconstructed pixel and the true pixel by SID instead of by the L2 norm used in traditional sparse model. Moreover, the spatial coherency across neighboring pixels sharing a common sparsity pattern is taken into account during the construction of SID-based joint sparse representation model. We propose a new version of the orthogonal matching pursuit method to solve SID-based recovery problems. The proposed SID-based algorithms are applied to real HSI for classification. Experimental results show that our algorithms outperform the classical sparse representation based classification algorithms in most cases.**

***Index Terms*—Hyperspectral image (HSI) classification, joint sparse representation, orthogonal matching pursuit, sparse representation classification, spectral information divergence (SID).**

I. Introduction

HYPERSPECTRAL image (HSI) is generated from airborne and satellite sensors, containing the information of hundreds of narrow spectral bands spanning the visible to infrared spectrum. HSI has found numerous applications in various fields, such as military, agriculture, and mineralogy. A very important application of HSI is image classification where pixels are labeled to one of the classes. Recently, applications of sparse representation have been extended to the area of computer vision and pattern recognition [1] with the development of the compressed sensing framework [2], [3] as well as sparse modeling of signals and images [4]. Especially, the classifier based on sparse representation draws the attention of many people. Despite the high dimensionality of HSI, the signals in the same class usually lie in a low-dimensional subspace. Under this assumption, a number of acquisition setups based on compressed sensing idea have been proposed to acquire the HSI by a few measures [5]. In [6], the authors proposed an analysis tool using unsupervised learning of sparse hyperspectral dictionaries for large geological surveys with little or no ground truth. In [7], the authors showed that the data of each hyperspectral class approximately lie in a low-dimensional linear subspace and proposed a fast homotopy-based sparse representation approach. In [8], a new sparsity-based algorithm was proposed, which relied on the observation that a hyperspectral pixel can be sparsely represented by a linear combination of a few training samples from a structured dictionary, and it incorporated the contextual information into the sparse recovery optimization problem. In [9], the authors projected the samples into a high-dimensional feature space and proposed the kernel versions of the single-pixel and multipixel joint sparsity-based recovery problems.

However, these methods fail to take into account the characteristics of the spectral signal or spatial information. Considering the characteristics of the spectral signal from the aspect of information theory, this letter introduces a new classifier, spectral information divergence (SID)-based sparse representation classifier (SRC). Then, we propose SID-based joint sparse representation model (JSRM), which incorporates the contextual information into the SID-based sparse recovery optimization problem in order to improve the classification performance.

On one hand, an HSI pixel is actually a column vector with dimensions equal to the number of spectral bands and contains valuable spectral information that can be used to account for pixel variability, similarity, and discrimination. Due to improved spectral resolution and large scene coverage, a pixel is generally a mixture of different materials with various abundance fractions. These materials absorb or reflect within each spectral band. However, because of atmospheric effects, the spectral information of a pixel varies during data acquisition. Those variations are caused by uncertainty or randomness. In order to determine the spectral similarity between two pixels, Chang [10] has demonstrated that SID can characterize spectral variability more effectively than the commonly used Spectral Angle Mapper (SAM). Traditional SRCs are based on the reconstruction error between the reconstruction pixel and the true pixel, combining reconstruction and the sparsity of the

Manuscript received August 14, 2012; revised November 22, 2012 and March 17, 2013; accepted March 18, 2013. Date of publication June 26, 2013; date of current version November 8, 2013. This work was supported in part by the National Natural Science Foundation of China under Grant 61272282, Grant 61003198, Grant 61001206, and Grant 61173092, the Shaanxi Province Natural Science Foundation under Grant 2011JQ8020, and the Fundamental Research Funds for the Central Universities under Grant K50511020011 and Grant K50510020009.

The authors are with the Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Institute of Intelligent Information Processing, Xidian University, Xi'an 710071, China (e-mail: xrzhang@mail.xidian.edu.cn; zhangerlei@stu.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.2013.2255097

sparse representation. The reconstruction error is measured by the L2 norm, which is also known as the Euclidean distance. Under certain circumstances, the Euclidean distance is equivalent to the SAM. So in order to describe spectral variability, similarity, and discrimination of HSI more effectively, we propose a new idea of combining the reconstruction property measured by SID and sparsity of the sparse representation. On the other hand, HSI naturally has large homogeneous regions. Many previous works have shown the importance of contextual information in HSI classification. In this letter, we propose SID-based JSRM to exploit the spatial correlation across neighboring pixels, the sparsity of the sparse representation, and data spectral characteristics simultaneously. Then we solve the SID-based recovery problems by a new version of orthogonal matching pursuit (OMP). Experimental results prove their rationality.

The remainder of this paper is structured as follows. Details of the SID measure, SID-based SRC, and SID-based JSRM are described in Section II. The effectiveness of the proposed method is demonstrated in Section III by experimental results on several real HSIs. Finally, Section IV summarizes this paper and makes some remarks.

II. SID-based Sparse Representation

A. Sparse Representation Based Classifier

In the traditional sparse model, the pixels belonging to the same class are assumed to be homogeny and approximately lie in a low-dimensional subspace. That is, a signal can be approximated by a sparse linear combination of a few atoms from a given dictionary. At first, we consider the reconstruction problem of finding the sparse representation coefficient $\boldsymbol{\alpha}$ for a test sample $\boldsymbol{x}$. In the case of training samples $\boldsymbol{A}$ (known as dictionary), the representation α satisfying $\boldsymbol{A}\,\boldsymbol{\alpha} = \boldsymbol{x}$ is obtained by solving the following optimization problem:

$$\hat{\alpha} = \arg\min \ ||\alpha||_0 \qquad s.t. \qquad \boldsymbol{A}\alpha = \boldsymbol{x}. \tag{1}$$

In order to meet the requirements of the real data, we relax the equality constraint in (1) to an inequality one

$$\hat{\alpha} = \arg\min \ ||\alpha||_0 \qquad s.t. \qquad ||\boldsymbol{A}\alpha - \boldsymbol{x}\ || \quad \leq \varepsilon \tag{2}$$

where ε is the error tolerance. The above problems can also be interpreted as minimizing the approximation error within a certain sparse degrees

$$\hat{\alpha} = \arg\min \ ||\boldsymbol{A}\alpha - \boldsymbol{x}||_2 \quad s.t. \quad ||\alpha||_0 \leq T \tag{3}$$

where T is a given upper bound on the sparse degrees. The aforementioned problems are NP-hard, but we commonly use such as OMP greedy algorithm to solve these problems. The OMP algorithm augments the support set by one atom at each time until T atoms are selected or the approximation error is within a preset threshold.

B. Spectral Information Divergence

Given a hyperspectral pixel (vector) $\boldsymbol{x} = (x_1, x_2, \cdots, x_D)^T$, each element x_i is a value of band B_i acquired at a particular wavelength λ_i. In fact, all components x_i in $\boldsymbol{x}$ are nonnegative. Hence we define a probability measure P for $\boldsymbol{x}$. Due to the nature of radiance or reflectance, we can define a probability measure P for $\boldsymbol{x}$ by

$$p_i = p(x_i) = \frac{x_i}{\sum_{i=1}^{D} x_i}. \tag{4}$$

The vector $P(\boldsymbol{x}) = (p_1, p_2, \cdots, p_D)^T$ is the desired probability vector resulting from the pixel vector $\boldsymbol{x} = (x_1, x_2, \cdots, x_D)^T$. As we know, any pixel $\boldsymbol{x}$ in an HSI can be viewed as a single information source consisting of D bands, so its spectral variability can be described by statistics governed by $P(\boldsymbol{x})$.

Assume that another pixel $\boldsymbol{y}$, the probability vector is $Q(\boldsymbol{y}) = (q_1, q_2, \cdots, q_D)^T$, where $q_i = q(y_i) = y_i \Big/ \sum_{i=1}^{D} y_i$. Based on information theory, we can define the self-information of pixels $\boldsymbol{x}$ and $\boldsymbol{y}$, respectively, which is a combination of the ith band self-information as follows:

$$\begin{aligned} &I_i(x) = -\log(p_i), \\ &I(\boldsymbol{x}) = (I_1(x), I_2(x), \cdots, I_D(x))^T; \\ &I_i(y) = -\log(q_i), \\ &I(\boldsymbol{y}) = (I_1(y), I_2(y), \cdots, I_D(y))^T. \end{aligned} \tag{5}$$

The relative entropy of $\boldsymbol{y}$ with respect to $\boldsymbol{x}$ can be defined by

$$\begin{aligned} CE(\boldsymbol{x}, \boldsymbol{y}) &= \sum_{i=1}^{D} p_i CE(x_i, y_i) = \sum_{i=1}^{D} p_i (I_i(y) - I_i(x)) \\ &= \sum_{i=1}^{D} p_i \log\left(\frac{p_i}{q_i}\right) = P^T(x) * (I(\boldsymbol{y}) - I(\boldsymbol{x})). \end{aligned} \tag{6}$$

In (6), $\mathrm{CE}(\boldsymbol{x}, \boldsymbol{y})$ is also known as the Kullback–Leibler information measure, directed divergence, or cross-entropy [11]. As a result of (6), we can get a symmetric hyperspectral similarity measure, referred to as SID. SID can be used to measure the spectral similarity between two pixels $\boldsymbol{x}$ and $\boldsymbol{y}$

$$\begin{aligned} SID(\boldsymbol{x}, \boldsymbol{y}) &= CE(\boldsymbol{x}, \boldsymbol{y}) + CE(\boldsymbol{y}, \boldsymbol{x}) \\ &= P^T(\boldsymbol{x}) * (I(\boldsymbol{y}) - I(\boldsymbol{x})) + Q^T(\boldsymbol{y}) * (I(\boldsymbol{x}) - I(\boldsymbol{y})). \end{aligned} \tag{7}$$

It offers a new view of the spectral similarity from information theory perspective, which calculates relative entropy to account for the spectral information provided by each pixel.

C. SID-Based SRC

In sparse representation models (1)–(3), the essence of $||\mathbf{A}\alpha - \boldsymbol{x}||_2$ is to measure the similarity degree between the reconstruction of the pixels and real pixel by calculating the Euclidean distance. OMP algorithm is used to approximately solve the problem in (3). In each iteration, coherence parameter $\mathrm{CP} \stackrel{\mathrm{def}}{=} | < \boldsymbol{x}, \hat{\boldsymbol{x}} > |$ is calculated to measure the similarity degree of the selected atom and residual vectors. In [10], Chang has demonstrated that SID can characterize spectral variability more effectively than the commonly used Euclidean distance. So we proposed SID-based SRC for the classification of HSI.

Given the dictionary $\boldsymbol{A}$ constructed by training samples, the vector $\boldsymbol{\alpha}$ can be recovered by solving the following sparse recovery problem:

$$\hat{\alpha} = \arg\min \ ||\alpha||_0 \qquad s.t. \qquad SID(\boldsymbol{A}\alpha, \boldsymbol{x}) \ \leq \varepsilon \tag{8}$$

or

$$\hat{\alpha} = \arg\min SID \ (\boldsymbol{A}\alpha, \boldsymbol{x}) \quad s.t. \qquad ||\alpha||_0 \leq \boldsymbol{T} \tag{9}$$

where T is sparsity level. Similar to the traditional sparse recovery problems, the sparse recovery problems are NP-hard problems, which can be approximately solved by greedy algorithms [12], [13], or relaxed to convex programming which can be solved in polynomial time [14].

In this letter, a new method analogous to greedy pursuit algorithms—OMP—is used to approximately solve the problem in (9). In new version of OMP, the support of the solution is also sequentially updated like OMP. In each iteration, the atoms in the dictionary $\boldsymbol{A}$ are sequentially selected. The atom is selected that yields the best approximation to the residual vectors by minimizing SID measure.

Once the sparse coefficient vector $\hat{\alpha}$ in (9) is obtained, the class of $\boldsymbol{x}$ can be determined accordingly. Suppose we have M distinct classes. Define the mth residual $RES^m(\boldsymbol{x})$ (i.e., error between the test sample and the reconstruction by the subdictionary consisting of training samples in the mth class) as

$$RES^m(\boldsymbol{x}) = ||\boldsymbol{x} - \boldsymbol{A}^m\hat{\alpha}^m||_2, \quad m = 1, 2, \cdots, M \tag{10}$$

or

$$RES^m(\boldsymbol{x}) = SID(\boldsymbol{x}, \boldsymbol{A}^m\hat{\alpha}^m), \quad m = 1, 2, \cdots, M \tag{11}$$

where $\hat{\alpha}^m$ denotes the portion of the recovered sparse coefficients corresponding to the training samples in the mth class. The class of $\boldsymbol{x}$ is then determined as the one with the minimal residual

$$Class(\boldsymbol{x}) = \arg \min_{m=1,2,\cdots,\mathrm{M}} RES^m(\boldsymbol{x}). \tag{12}$$

D. SID-Based Joint Sparse Representation Model

For HSI, the neighboring pixels composed of similar materials, their spectral characteristics are highly correlated. We propose SID-based JSRM to exploit the spatial correlation across neighboring pixels and data spectral characteristic simultaneously. We assume that each of HSI pixels in a small spatial neighborhood is approximated by a sparse linear combination of a few atoms from a given structured dictionary with a different set of coefficients, the selected atoms being the same.

In a small neighborhood N^L, there are L pixels. Let $\boldsymbol{X} = [\boldsymbol{x}_1\boldsymbol{x}_2 \cdots \boldsymbol{x}_L]$be a $D \times L$ matrix, where the column vectors $\{\boldsymbol{x}_i\}i = 1, 2, \cdots, L \in N^L$ are pixels in a spatial neighborhood in the HSI. Now, using the JSRM, $\boldsymbol{X}$ can be represented by

$$\begin{aligned} \boldsymbol{X} &= [\boldsymbol{x}_1\boldsymbol{x}_2 \cdots \boldsymbol{x}_L] = [\boldsymbol{A}\boldsymbol{\alpha}_1\boldsymbol{A}\boldsymbol{\alpha}_2 \cdots \boldsymbol{A}\boldsymbol{\alpha}_L] \\ &= \boldsymbol{A}[\boldsymbol{\alpha}_1\boldsymbol{\alpha}_2 \cdots \boldsymbol{A}\boldsymbol{\alpha}_L] = \boldsymbol{A}\boldsymbol{Z} \end{aligned} \tag{13}$$

where the sparse vectors $\{\boldsymbol{\alpha}_i\}i = 1, 2, \cdots, L$ share the same support S and, thus, $Z \in \Re^{N\times L}$is a sparse matrix with only T nonzero rows. The row-sparse matrix $Z \in \Re^{N\times L}$ can be recovered by solving the following optimization problem:

$$\hat{\boldsymbol{Z}} = \arg\min SID(\boldsymbol{X}, \boldsymbol{A}\boldsymbol{Z}) s.t. \ \|\boldsymbol{Z}\|_{r,0} \leq \boldsymbol{T} \tag{14}$$

or

$$\begin{aligned} &\hat{\boldsymbol{Z}} = \arg\min \sum_{i=1}^{L} SID(\boldsymbol{x}_i, \boldsymbol{A}\boldsymbol{\alpha}_i) \\ &s.t. \ \|\boldsymbol{Z}\|_{\mathrm{r},0} \leq T, \\ &\boldsymbol{Z} = [\boldsymbol{\alpha}_1\boldsymbol{\alpha}_2 \cdots \boldsymbol{\alpha}_L]. \end{aligned} \tag{15}$$

where

$$\|\boldsymbol{Z}\|_{r,0} = \sum_{i=1}^{N} \Gamma(\left\|z^{\mathbf{i}}\right\|_2 > 0) \tag{16}$$

where $\Gamma(\bullet)$denotes the indicator function and $z^{\mathbf{i}}$denotes the ith row of Z, and $\|Z\|_{r,0}$denotes the number of non-zero rows of Z. The problem in (14) can be approximately solved by the simultaneous versions of OMP as

Algorithm 1 SID-Based Joint Sparse Representation Model

Input: Given the matrix $\boldsymbol{A} \in \Re^{D\times N}$, the vector $\boldsymbol{X} \in \Re^{D\times L}$, and sparsity level T.

Initialization: $k = 1$, and compute the correlation matrix $\Phi_A \in \Re^{N\times N}$and $\Phi_{A,X} \in \Re^{N\times L}$whose (i,j)th entry is $SID(\boldsymbol{a_i}, \boldsymbol{a_j})andSID(\boldsymbol{a_i}, \boldsymbol{x_j})$, respectively. Set the initial solution$\mathbf{Z}^0 = \mathbf{0}$; Let the initial solution support $S^0 = Support\{\alpha^0\} = \varphi$. The initial the correlation matrix between residual and dictionary $C^0 = \Phi_{A,X}$.

While stopping criterion has not been met **do**

Step 1: Sweep: Find index of the atom that best approximates all residual: $\lambda_k = \arg \min_{i=1,2,\cdots,N} \left\|C_{i,:}^{k-1}\right\|_p \ p = 1,2$, or $\lambda_k = \arg \min_{i=1,2,\cdots,N} \min (C_{i,:}^{k-1})$;.

Step 2: Update Support: Find a minimizer λ_k and $\lambda_k \notin S^{k-1}$.Then update $S^k = S^{k-1} \cup \lambda_k$;

Step 3: Update the correlation matrix: Compute $C^k = \Phi_{A,X} - (\Phi_A)_{:,S^k}((\Phi_A)_{S^k,S^k} + \mu\mathrm{I})^{-1}(\Phi_{A,X})_{S^k,:} \in \Re^{N\times L}$;

Step 4: $k \leftarrow k + 1$, if $k > T$, stop.

End while

Note that a regularization term μI in *Step 3* is added in order to have a stable inversion. μ is a small scalar and is chosen as $\mu = 10^{-5}$ in our implementation. This parameter, however, does not seriously affect the classification performance, because the matrix is usually invertible and regularization is not really needed.

III. Experimental Results and Discussion

We applied the proposed SID-based SRC approach and SID-based JSRM to the following real HSI.

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,

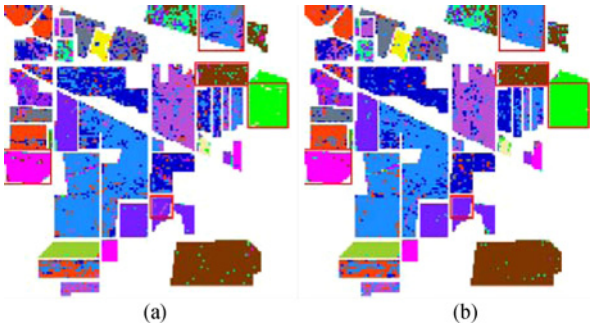

Fig. 1. Subjective classification results of Indian Pines image. (a) SRC map. (b) SID-based SRC map.

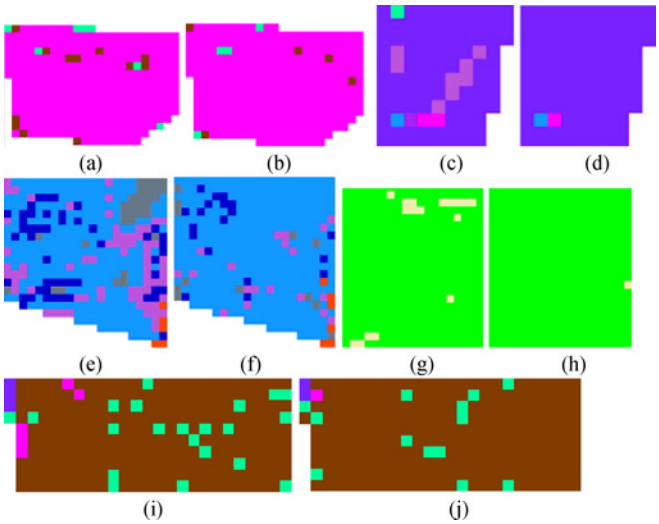

Fig. 2. Classification results of Indian Pines image. (a), (c), (e), (g), and (i) are results of SRC map. (b), (d), (f), (h), and (j) are results of SID-based SRC map.

the number of bands is reduced to 200 by removing 20 water absorption bands. This image has spatial resolution of 20 m per pixel and spatial dimension 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.

Fig. 1 shows the two corresponding classification maps of Indian Pines image got by SRC method and the new method. For a clearer comparison of experimental results of two algorithms, the amplified regions labeled by the red lined box in Fig. 1 are displayed in Fig. 2. From Fig. 2, it can be seen that our method is better than traditional SRC.

The classification accuracy for the overall accuracy (OA), average accuracy (AA), and the κ coefficient measure (kappa) are shown in Table I, respectively, using different classifiers on the test set. We can observe that the SID-based SRC accuracies of the proposed approach are better than the SRC method. It is noted that for the sake of fair comparison, we apply the classification criteria of (10) in the experiments.

The classification results of Indian Pines image using SOMP method [8] and our SID-based JSRM are shown in Table I. In contrast, SID-based JSRM provides a better performance.

TABLE I
CLASSIFICATION ACCURACY (%) BY DIFFERENT METHODS FOR INDIAN PINES CLASSIFICATION

Class	SRC	SID-based SRC	SOMP	SID-based JSM
1	**75.93**	72.22	85.42	**98.15**
2	63.95	**77.27**	94.88	**96.16**
3	67.39	**70.62**	94.93	**95.32**
4	57.69	**58.12**	**91.43**	90.60
5	91.75	**92.76**	89.49	**97.79**
6	95.58	**96.65**	**98.51**	96.79
7	**92.31**	**92.31**	**91.30**	80.77
8	95.30	**98.57**	99.55	**100**
9	**65.00**	50.00	0	**75.00**
10	75.62	**76.65**	89.44	**95.14**
11	77.27	**83.31**	**97.34**	97.24
12	61.56	**65.31**	88.22	**88.93**
13	**98.58**	**98.58**	**100**	99.53
14	94.05	**94.90**	99.14	**99.54**
15	**48.42**	45.53	**99.12**	91.84
16	**93.68**	91.58	**96.47**	87.37
OA	77.60	**81.68**	95.28	**96.19**
AA	78.38	**79.02**	88.45	**93.14**
kappa	0.7448	**0.7908**	0.946	**0.956**

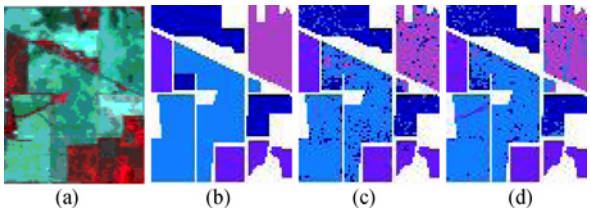

Fig. 3. Subjective classification results of Indian subset image. (a) Original HSI. (b) Ground-truth map. (c) SRC map. (d) SID-based SRC map.

And from the results of Table I, it can be clearly seen that the methods combining spatial information can greatly improve the classification performance.

Then we demonstrate the effect of the sparsity level T and the number of training samples on the performance of the SRC and SID-based SRC. In this experiment, in order to validate the method more conveniently, we still perform experiments on the subimage of the Indiana image (later referred to as Indiana subset). Indiana subset includes [27–94] $\times$ [31–116] of the total Indian Pines image. The subimage is size of 68×86 pixels, containing four types of labeled samples [Corn-notill(1008), Grass/Trees (732), Soybeans-notill(727), and Soybeans-min(1926)]. We randomly choose 10% of Indiana subset data in each class as training samples and use the remaining 90% as test samples. Corresponding two classification maps are shown in Fig. 3. The sparsity level T is chosen between $T = 1$ and $T = 40$. The overall classification accuracy of the entire test set for Indian Pines are shown in Fig. 4. As the sparsity level T increase, the classification accuracy is slightly reduced. It is shown that more atoms do not necessarily lead to the higher classification accuracy.

The number of training samples also has an important effect on the performance. We test SRC and SID-based SRC on Indiana subset with various numbers of training samples from 3 to 100. Fig. 5 shows the effect of the number of training

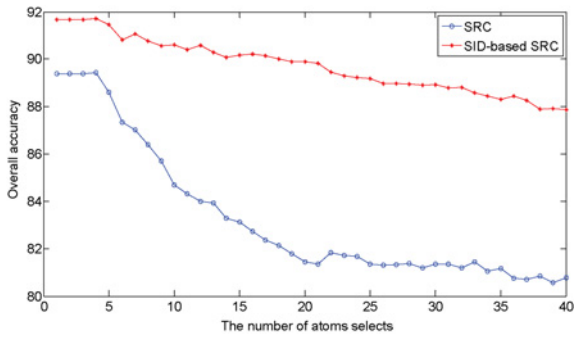

Fig. 4. Effect of the sparsity level T to classification results of Indian Pines subset.

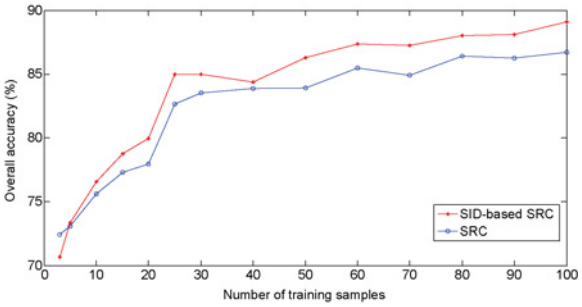

Fig. 5. Effect of the number of training samples for Indian Pines subset.

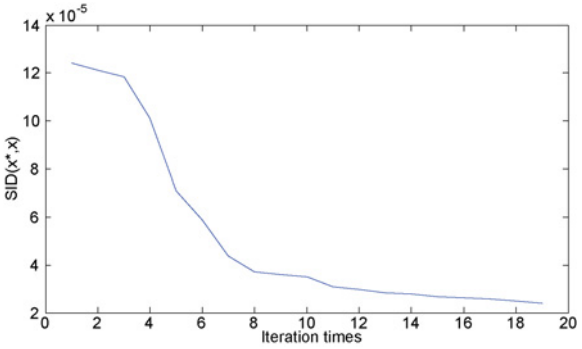

Fig. 6. Effect of the number of atoms for the SID of the data and its approximation.

samples on classification accuracy. It is obvious from Fig. 5 that in most cases the OA increases as the number of training samples increases, and our proposed method shows higher accuracy than SRC.

Finally, we prove the convergence of our method by experiments. From Fig. 6, it can be seen that the value of SID of the sample $\boldsymbol{x}$ and its approximation $\boldsymbol{x}^*$ decreases as the number of selected atoms increase. Fig. 6 intuitively proves the convergence of the optimization problem we need to solve.

IV. Conclusion

A novel SID-based sparse representation method for the classification of HSI was presented in this letter. Moreover, the SID-based joint sparse representation method exploited the spatial correlation across neighboring pixels, data spectral characteristics, and the sparsity of the sparse representation simultaneously. Experimental results demonstrated that the proposed method yielded more accurate classification maps. Therefore, it provided a new way to improve the classification methods based on sparse representation.

References

[1] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

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

[4] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009

[5] T. Sun and K. Kelly, "Compressive sensing hyperspectral imager," in *Proc. Frontiers Optics/Laser Science XXV/Fall 2009 OSA Optics Photonics Tech. Dig., OSA Tech. Dig. (CD)*, paper CTuA5 [Online]. Available: http://www.opticsinfobase.org/abstract.cfm?URI=COSI-2009-CTuA5

[6] A. Charles, B. Olshausen, and C. J. Rozell, "Sparse coding for spectral signatures in hyperspectral images," in *Proc. Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, 2010, pp. 191–195.

[7] Q. S. ul Haq, L. Tao, F. Sun, and S. Yang, "A fast and robust sparse approach for hyperspectral data classification using a few labeled samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 6, pp. 2287–2302, Jun. 2012.

[8] 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.

[9] 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.

[10] C. Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1927–1932, Aug. 2000.

[11] S. Kullback, *Information Theory and Statistics*. New York, NY, USA: Wiley, 1959.

[12] 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.

[13] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[14] E. van den Berg and M. P. Friedlander, "Theoretical and empirical results for recovery from multiple measurements," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2516–2527, May 2010.