International Journal of Applied Earth Observation and Geoinformation 52 (2016) 275–283

Contents lists available at ScienceDirect

International Journal of Applied Earth Observation and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Generalization of spectral fidelity with flexible measures for the sparse representation classification of hyperspectral images

Bo Wu [a,b,*], Yong Zhu [b], Xin Huang [c], Jiayi Li [c]

[a] State key laboratory of Poyang lake wetland and watershed research, ministry of education, Jiangxi normal university, Nanchang 330022, China
[b] Spatial Information Research Center of Fujian, Fuzhou University, Fuzhou, Fujian 350002, China
[c] School of Remote Sensing and Information Engineering, Wuhan University 129, Luoyu road, Wuhan 430079, China

ARTICLE INFO

Article history:
Received 30 January 2016
Received in revised form 30 May 2016
Accepted 7 June 2016

Keywords:
Spectral similarity measures
Sparse representation classification
Unified framework
Hyperspectral image

ABSTRACT

Sparse representation classification (SRC) is becoming a promising tool for hyperspectral image (HSI) classification, where the Euclidean spectral distance (ESD) is widely used to reflect the fidelity between the original and reconstructed signals. In this paper, a generalized model is proposed to extend SRC by characterizing the spectral fidelity with flexible similarity measures. To validate the flexibility, several typical similarity measures—the spectral angle similarity (SAS), spectral information divergence (SID), the structural similarity index measure (SSIM), and the ESD—are included in the generalized model. Furthermore, a general solution based on a gradient descent technique is used to solve the nonlinear optimization problem formulated by the flexible similarity measures. To test the generalized model, two actual HSIs were used, and the experimental results confirm the ability of the proposed model to accommodate the various spectral similarity measures. Performance comparisons with the ESD, SAS, SID, and SSIM criteria were also conducted, and the results consistently show the advantages of the generalized model for HSI classification in terms of overall accuracy and kappa coefficient.

1. Introduction

Recent advances in sparse representation learning have achieved significant success in the adaptation of sparse representation for hyperspectral image (HSI) classification, and there has been an increasing tendency to explore sparsity as prior information for remote sensing applications in recent years. Chen et al. (2011) first applied sparse representation to HSI classification by utilizing the sparsity of the input sample with respect to a given overcomplete dictionary consisting of several training samples. Given a test spectral vector, it can then be sparsely represented with several atoms in the dictionary and the associated sparse coding coefficients. The class of the test pixel can then be determined by comparing the similarity between the reconstructed and test pixels. To exploit the contextual information in homogeneous regions, joint sparse representation classification (JSRC) was also proposed to enhance the ability of the classification (Chen et al., 2011).

Several variants of the typical SRC model have also been proposed and have achieved better performances in HSI classification. Realizing that not all pixels provide the same contribution to the central pixel, Zhang et al. (2014b) proposed a non-local weighted JSRC method by considering the different contributions of the neighboring pixels in the classification process. In addition, multiscale adaptive SRC was formulated by considering that regions of different scales can incorporate complementary yet correlated information (Fang et al., 2014). Manifold-based SRC has also been proposed to exploit the local structure of the test pixel to enhance smoothness across neighboring homogeneous regions (Tang et al., 2014). Since the complex arrangement of land cover on the Earth's surface inevitably results in many mixed pixels, a nonlinear extension of SRC with spatial and spectral information based on a kernel metric has also been proposed, and has performed well in HSI classification (Chen et al., 2013; Liu et al., 2013).

Although the aforementioned SRC methods can perform well, most of them adopt the Euclidean spectral distance (ESD) as the measuring criterion to reflect the spectral fidelity between the original and reconstructed pixel signals, due to its ease of use and popularity. However, the ESD measure is sensitive to the absolute magnitude of the spectra, and it fails to take into account the characteristics of the spectral signal, such as the spectral shape and spatial information. On the other hand, the magnitude of the spectra usually varies from site to site in a real image, due to external factors such as the atmospheric effect, environmental radiation, shading, etc. These physical phenomena can cause multiple signal scattering, resulting in a nonlinear mixture effect, especially for heterogeneous area. Consequently, we can infer that the linear-based similarity

* Corresponding author.
E-mail addresses: wavelet778@sohu.com (B. Wu), xhuang@whu.edu.cn (X. Huang).

http://dx.doi.org/10.1016/j.jag.2016.06.006
0303-2434/

measures such as the ESD would produce significant classification errors in such circumstances.

Some researchers have recently proposed other spectral fidelity measures to overcome the disadvantages of the ESD-based measure. In the remote sensing community, spectral information divergence (SID) is now one of the most widely accepted spectral similarity measures for material identification (Chang, 2000). SID measures the discrepancy between the probabilistic behaviors of the spectral signatures of two pixels from the aspect of information theory. A previous study confirmed the ability of SID to effectively discriminate spectral properties (Van Der Meer, 2006). More recently, a sparse representation classifier using the SID measure was proposed to reduce the influence of spectral variation, and achieved a promising classification accuracy (Zhang et al., 2014a). Inspired by the structural similarity index (Wang et al., 2004), we incorporated a new structural similarity index measurement (SSIM) into an SRC model to capture the spectral and spatial properties, and the experiments demonstrated that SSIM-based SRC outperforms ESD-based SRC in most cases (Wu et al., 2015). The superiority of the SSIM-based or SID-based SRC may be because their nonlinear spectral measures provide more powerful discriminative abilities to classify mixed pixels. On the other hand, due to the fact that the algorithm of traditional ESD-based SRC is not applicable to SID- or SSIM-based SRC methods, the solutions used to solve the respective models are different. To solve the SID-based problem, the authors developed a new orthogonal matching pursuit method to update the selected atoms via minimizing the correlation matrix between the reconstructed residual and the dictionary (Zhang et al., 2014a). Meanwhile, the solution of SSIM-based SRC adopted the same idea as the traditional ESD-based problem by adding a scaling coefficient to its optimal sparse codes (Rehman et al., 2012).

Although SRC has been extensively studied from various aspects, there are still some areas that remain untouched. Given that several spectral similarity measures are available, it is still difficult to know which to select to characterize the spectral fidelity when one needs to perform an SRC task. In addition, since the existing algorithms for ESD-, SSIM- and SID-based SRC are different, there is not yet a generalized SRC algorithm, such that different spectral fidelity items can be universally applied. We attempt to address these questions in this paper, where we make three possible contributions. The first is that we generalize the SRC model with a unified framework to adapt to various spectral similarity measures. One advantage of using this generalized model is that it provides a possible way to accommodate any spectral similarity measure that the user decides to use. Another contribution is that we present a unified solution to solve the generalized SRC model by adopting a gradient descent technique. Finally, a systematic comparison of the performance of HSI classification using various spectral similarity measures is conducted, i.e. ESD, the spectral angle measure (SAS), SSIM, and SID.

The remainder of this paper is structured as follows. Section II formulates the unified framework with flexible spectral similarity measures for SRC. In Section III, two real HSIs are used to test the effectiveness of the proposed framework. Finally, Section IV summarizes the paper.

2. The generalized SRC model

2.1. Formulation of the generalized SRC model

Suppose that we have M distinct classes, and the mth class has N_m training samples$\boldsymbol{D}_m = \{d_i^m\}_{i=1,2,\ldots N_m}$, respectively. We let $\boldsymbol{D} = [\boldsymbol{D}_1 \quad \boldsymbol{D}_2 \quad \ldots \quad \boldsymbol{D}_M] \in \mathbb{R}^{\boldsymbol{B}\times\boldsymbol{N}}(\boldsymbol{B} < N)$ be a structural dictionary, whose columns (which we refer to as atoms) are extracted from the training sample pixels, where $\boldsymbol{B}$ refers to the number of bands of the HSI, and D_i is a sub-matrix belonging to the same class of the overcomplete dictionary. The typical SRC model can be represented as:

$$\hat{\boldsymbol{\alpha}} = arg\min_{\boldsymbol{\alpha}} \|y - \boldsymbol{D}\boldsymbol{\alpha}\|_2^2 \text{s.t} \|\boldsymbol{\alpha}\|_0 \le K_0 \tag{1}$$

$$class(\boldsymbol{y}) = arg\min_i \|\boldsymbol{y} - \boldsymbol{x}_i\|_2^2, i = 1, 2, \ldots, \boldsymbol{M} \tag{2}$$

where $\hat{\boldsymbol{\alpha}}_i$ denotes the sparse coding coefficient associated with atoms D_i, i.e. the sub-matrix belonging to the ith class, and $\boldsymbol{x}_i = \boldsymbol{D}_i\hat{\boldsymbol{\alpha}}_i$ indicates the reconstructed spectral signals with the estimated sparse coding coefficient $\hat{\boldsymbol{\alpha}}_i$. It is clear from Eq. (2) that the class label is determined by the minimal residual error $r_i = \|\boldsymbol{y} - \boldsymbol{x}_i\|_2^2$ associated with each class, where the reconstructed vector is dominated by the coefficients corresponding to the indices of the training samples (atoms) in the ith class. To enhance the robustness, the JSRC model is formed to include the contextual information, which can be formulated as:

$$\hat{\boldsymbol{S}} = arg\min_S \|\boldsymbol{Y} - \boldsymbol{D}\boldsymbol{S}\|_2^2 s.t. \|\boldsymbol{S}\|_{\boldsymbol{row},0} \le K_0 \tag{3}$$

$$class(\mathrm{y}) = arg\min_i \|\boldsymbol{Y} - \boldsymbol{D}_i\hat{\boldsymbol{S}}_i\|_2^2, i = 1, 2, \ldots, M \tag{4}$$

where $\boldsymbol{Y}$ is all the pixels around the test sample $\boldsymbol{y}$ within a local window, and $\boldsymbol{S}$ is the associated group sparse coding coefficients.

As can be seen from the formulations in Eqs. (1)–(4), they all adopt the ESD measure to reflect the spectral fidelity. Intuitively, the use of different spectral fidelity items can produce different sets of atoms, and thus different classification results. Therefore, we propose a generalized sparse representation classification model (GSRC) to extend the ESD-based model to accommodate various similarity measures, such that any pixel can be formulated as a weighted sum of the selected atoms. To this aim, we formulate GSRC by characterizing the spectral fidelity as a general objective function:

$$\hat{\boldsymbol{\alpha}} = arg\min_{\boldsymbol{\alpha}} \boldsymbol{f}(\boldsymbol{y}, \boldsymbol{x}) s.t. \boldsymbol{x} = \boldsymbol{D}\boldsymbol{\alpha}, \|\boldsymbol{\alpha}\|_0 \le K_0 \tag{5}$$

$$class(\boldsymbol{y}) = arg\min_i \boldsymbol{f}(\boldsymbol{y}, \boldsymbol{x}_i), \boldsymbol{x}_i = \boldsymbol{D}_i\hat{\boldsymbol{\alpha}}_i i = 1, 2, \ldots, M \tag{6}$$

where $\boldsymbol{f}$ is a user-specified objective function, which defines the general similarity measure to reflect the spectral fidelity between the original signal $\boldsymbol{y}$ and the reconstructed signal $\boldsymbol{x}$.

It is clear that GSRC has the advantage of being able to accommodate any spectral similarity measure. For example, the ESD-based SRC model can be modeled by specifying the objective function $\boldsymbol{f}$ as ESD to measure the similarity, i.e.

$$f_{ESD} = ESD(\boldsymbol{y}, \boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{x}\|_2^2 \tag{7}$$

Other spectral fidelity items can also be specified as the objective function, such as SAS. SAS calculates the angle between two spectra and uses it as a discriminative measure, and thus has an advantage with respect to reducing or avoiding the effect of brightness deformation (Du et al., 2004). To cast the objective function as a minimization formulation, $\boldsymbol{f}$ for SAS becomes:

$$f_{SAS} = 1 - SAS(\boldsymbol{y}, \boldsymbol{x}) = 1 - \frac{\sum_{i=1}^{B} x_i y_i}{\sum_{i=1}^{B} x_i^2 \sum_{i=1}^{B} y_i^2} \tag{8}$$

Analogously, the objective function $\boldsymbol{f}$ for SSIM is given as:

$$f_{SSIM} = 1 - SSIM(\boldsymbol{y}, \boldsymbol{x}) = 1 - \frac{2\mu_x\mu_y + c1}{\mu_x^2 + \mu_y^2 + c1} \frac{2\sigma_{xy} + c2}{\sigma_x^2 + \sigma_y^2 + c2} \tag{9}$$

where $\mu_x and \mu_y$ are the means of vectors x and y, respectively; σ_{xy} is the sample covariance; andσ_x^2 and σ_y^2 are their variances. The constants $c1$ and $c2$ are stabilizing constants to account for the saturation effect of the human visual system.

The objective function $\boldsymbol{f}$ for SID is formulated as:

$$f_{SID} = SID(\boldsymbol{y}, \boldsymbol{x}) = \sum_{i=1}^{B} \boldsymbol{p}_i log(\frac{\boldsymbol{p}_i}{\boldsymbol{q}_i}) + \sum_{i=1}^{B} \boldsymbol{q}_i log(\frac{\boldsymbol{q}_i}{\boldsymbol{p}_i}) \quad (10)$$

where $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$, (where $i = 1, 2, \ldots, B$) are the two probability functions produced by the reconstructed spectral vector and the original spectral vector, respectively, which are defined as:

$$\boldsymbol{p}_i = \frac{y_i}{\sum_{i=1}^{\boldsymbol{B}} y_i}, \quad \boldsymbol{q}_i = \frac{x_i}{\sum_{i=1}^{\boldsymbol{B}} x_i} \quad (11)$$

In Eq. (10), SID(y, x) is also known as Kullback-Leibler information, which offers a new view of spectral similarity from an information theory perspective, by calculating the relative entropy to account for the spectral information provided by each pixel.

2.2. The joint generalized sparse representation classification (JGSRC) model

Since the GSRC model does not consider the contextual information, the JGSRC model is also developed by exploiting the spatial correlation across the neighboring samples in GSRC, which can be mathematically formulated as follows:

$$\hat{\boldsymbol{S}} = \arg\min_{\boldsymbol{S}} \boldsymbol{f}(\boldsymbol{Y}, \boldsymbol{X}), s.t. \boldsymbol{X} = \boldsymbol{DS}, \|\boldsymbol{S}\|_0 \leq K_0 \quad (12)$$

$$class(\boldsymbol{y}) = \arg\min_{i} \boldsymbol{f}(\boldsymbol{Y}, \boldsymbol{X}_i), X_i = \boldsymbol{D}_i \hat{\boldsymbol{S}}_i i = 1, 2, \ldots, \boldsymbol{M} \quad (13)$$

where the variables are defined in the same way as Eqs. (3) and (4). In comparison with GSRC, the main difference is that JGSRC utilizes all the sparse vectors from the local patch to compute the residual error, which has the potential to improve the classification performance. As a result, JGSRC can significantly outperform GSRC in most cases.

2.3. Solution to the JGSRC

Since the SRC model defined by the $\boldsymbol{l_0}$ norm is a nondeterministic polynomial time hard problem, it is usually resolved by the orthogonal matching pursuit (OMP) method. Specifically, for a test pixel, the OMP algorithm tends to incrementally search for a representative atom iteratively, and then the residual of the fidelity item is updated. The next atom is then selected as the one that vector. Following a similar idea, maximizes the relationship between the dictionary and the residual we propose to use a gradient descent algorithm for the general similarity measure to indicate which atom is to be selected iteratively.

$$grad = \frac{\partial(\boldsymbol{f}(\boldsymbol{R}, \boldsymbol{D\alpha}))}{\partial(\boldsymbol{\alpha})} \quad (14)$$

where $\boldsymbol{R}$ denotes the residual item between the original and reconstructed vectors, $\boldsymbol{f}$ is a general spectral similarity measure, and $\boldsymbol{\alpha}$ is the associated sparse coding coefficient. If the objective function is specified as ESD, i.e. $\boldsymbol{f} = \|\boldsymbol{y} - \boldsymbol{D\alpha}\|_2^2$, then the gradient is:

$$grad = \frac{\partial(\boldsymbol{f}(\boldsymbol{y}, \boldsymbol{D\alpha}))}{\partial(\boldsymbol{\alpha})} = \frac{\partial(\|\boldsymbol{y} - \|\boldsymbol{D\alpha}_2^2)}{\partial(\boldsymbol{\alpha})} = -\boldsymbol{2D^T}(\boldsymbol{y} - \boldsymbol{D\alpha}) = -\boldsymbol{2D^T R} \quad (15)$$

It can be seen that Eq. (15) is exactly the same formulation as the one derived from the typical ESD-based model (Chen et al., 2011). Therefore, if the maximum gradient is an indicator for the atom selection, the proposed gradient descent algorithm boils down to the traditional OMP algorithm used in the ESD-based SRC model. Given other similarity measures, such as f_{SAS}, f_{SSIM}, and f_{SID}, the set of atoms for representation can be iteratively selected with their respective gradient formulations. The respective gradients for the f_{SAS}, f_{SID}, and f_{SSIM} functions are formulated as follows:

$$\frac{\partial(f_{SAS})}{\partial\boldsymbol{\alpha}} = \frac{1}{\|\boldsymbol{y}\|_2} \frac{\|\boldsymbol{x}\|_2^2 \boldsymbol{D}'\boldsymbol{y} - (\boldsymbol{D\alpha})'\boldsymbol{y} * \boldsymbol{D}'\boldsymbol{D\alpha}}{\|\boldsymbol{x}\|_2^2 * \|\boldsymbol{x}\|_2} \quad (16)$$

$$\frac{\partial(f_{SID})}{\partial\boldsymbol{\alpha}} = \boldsymbol{D}'(log(\frac{D\alpha}{y}) + \boldsymbol{I}) - \boldsymbol{D}'\frac{y}{\boldsymbol{D\alpha}} \quad (17)$$

$$\frac{\partial(f_{SSIM})}{\partial\boldsymbol{\alpha}} = \frac{2}{N_w B_1^2 B_2^2}[A_1 B_1 (B_2 a - A_2 y + B_1 B_2 (A_2 - A_1)\mu_a - A_1 A_2 (B_2 - B_1)\mu_y)] \quad (18)$$

where $A_1 = 2\mu_a\mu_y + c1$, $B_1 = \mu_a^2 + \mu_y^2 + c1$, $A_2 = 2\sigma_{ay} + c2$, $B_2 = \sigma_a^2 + \sigma_y^2 + c2$, and N_w is the number of pixels in a local window. μ_a, σ_a^2, and σ_{ay} represent the sample mean of a, the sample variance of a, and the sample covariance of a and y.

By characterizing the spectral fidelity item with a general objective function, GSRC can be processed with various similarity measures, on the condition that the objective function is convex. Considering that GSRC is only a special case of JGSRC with the neighboring window size equal to one, we mainly focus on the JGSRC model in this paper. The implementation details are summarized as Algorithm 1.

Algorithm 1. Joint generalized sparse representation classification (JGSRC):

Input: *A dictionary of training samples* $\boldsymbol{D}$ *for M classes, sparsity level* K_0, *a pixel vector to be classified* $\boldsymbol{y}$, *and all the pixels around the central test sample* $\boldsymbol{y}$ *within a local window, denoted as* $\boldsymbol{Y}$.

Initialize: *Normalize* $\boldsymbol{D}$; Index set $\boldsymbol{\Omega^1} = [\emptyset]$; $\boldsymbol{U} = [\text{ones}]$; $n = 1$

Process: $\boldsymbol{R} = \boldsymbol{Y}$

while$(n \leq K_0)$

1 *set* $\boldsymbol{\alpha}' = [\boldsymbol{0}]$;

2 *calculate* $\boldsymbol{grad} = \frac{\partial(f(\boldsymbol{R}, \boldsymbol{D\alpha}'))}{\partial(\alpha)}$;

3 $\boldsymbol{s} = \boldsymbol{grad}^2$ *or* $|\boldsymbol{grad}|$;

4 *stretch the gradient* $\boldsymbol{s} = \boldsymbol{s./U}$;

5 update index set $\boldsymbol{\Omega^n} = \boldsymbol{\Omega^{n-1}} \cup \boldsymbol{i^n}$, $\boldsymbol{i^n} = \text{argmax}_i\, \boldsymbol{s}$;

6 *calculate the orthogonal projector* $\boldsymbol{P}_{\boldsymbol{\Omega^n}} = (\boldsymbol{I} - D_{\Omega^n}(D_{\Omega^n}^T D_{\Omega^n})^{-1} D_{\Omega^n}^T)$;

7 *calculate* $\boldsymbol{\alpha}$ *using (5);*

8 *update residual* $\boldsymbol{R^n} = \boldsymbol{P}_{\boldsymbol{\Omega^n}} \boldsymbol{Y}$;

9 *update* $\boldsymbol{U} = \boldsymbol{P}_{\boldsymbol{\Omega^n}} \boldsymbol{D}$, ***renormalize*** $\boldsymbol{u}_k = \boldsymbol{u}_k / \|\boldsymbol{u}_k\|_2$, *for* $\boldsymbol{k} \in \boldsymbol{\Omega^n}$

10 $n = n + 1$

end

classification :

calculate the class of the test sample through Eq. (6)

Output: *the label of the test sample* $\boldsymbol{y}$.

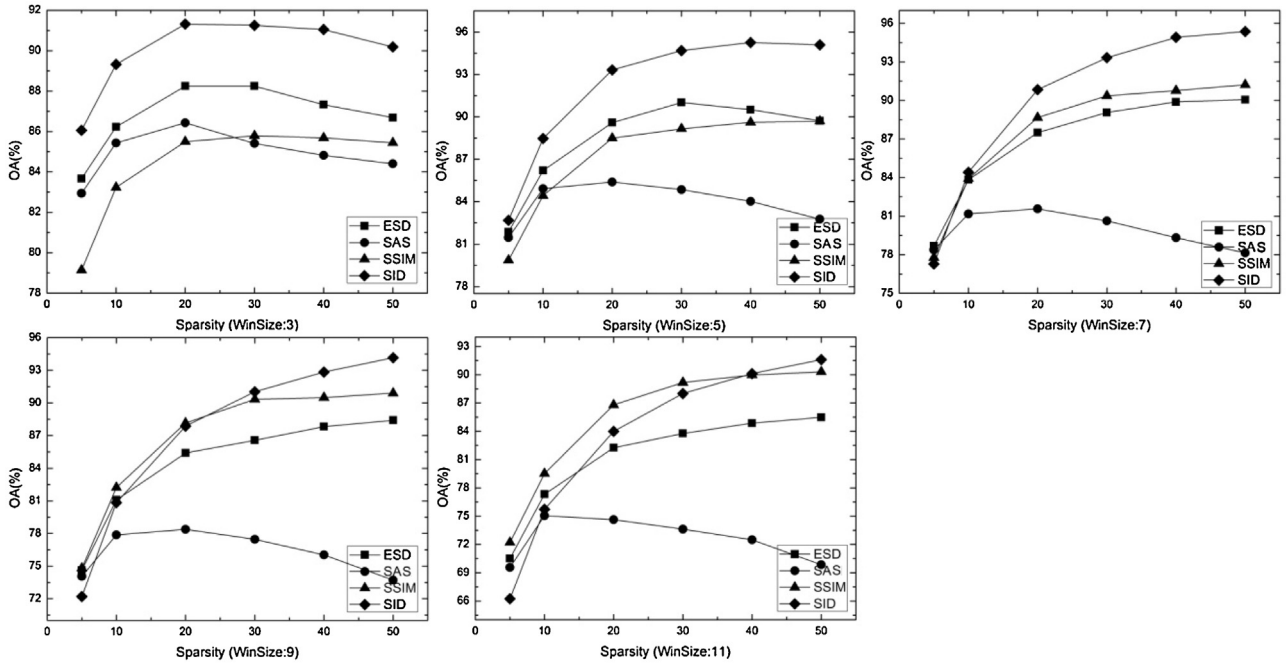

Fig. 1. Comparison of the OAs of the different methods with varying window sizes.

Although Algorithm 1 appears similar to OMP, it should be noted that the proposed method significantly differs from the traditional JSRC algorithms in two aspects. The first is that we utilize a gradient descent algorithm for the general objective function, rather than a fixed measuring function to determine the subset of atoms for the representation of the test pixel vector. The other difference is that we not only project the observed test pixel vector orthogonal to the selected atoms to calculate the residual, we also project the remaining dictionary atoms and renormalize them, such that all the atoms are unit norm again. One advantage of using renormalization of the remaining dictionary atoms is to force the sparsity of the residual to decrease along with its rank (Natarajan, 1995) in the iterative processing.

3. Experimental results and analysis

To test the effectiveness of the proposed JGSRC, two actual HSIs were selected for the experimental evaluation. The classical JSRC with ESD measure classifier was implemented as the baseline for the comparison. The primary objectives of the experiments were: 1) to evaluate the JGRSC model by comparing the performances of the specified similarity measures; and 2) to test the improvement of the proposed algorithm compared with the existing SSIM- and SID-based JSRC algorithms (Zhang et al., 2014a; Wu et al., 2015). For a fair comparison, all the algorithms were coded in MATLAB R2012b. Although many accuracy measures are available (Pontius and Millones, 2011), two widely used measures—the overall accuracy (OA) and Kappa coefficient—were chosen to measure the accuracy of the HSI classification tasks because they are the simplest yet most popular accuracy measures and are recommended for use as a primary measure (Liu, 2007).

3.1. Performance evaluation and comparison

3.1.1. Dataset 1: AVIRIS

The first dataset was the 220-band Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) image taken over the Indian Pines test site in North-western Indiana in June 1992, with 224 bands across the spectral range from 0.2 to 2.4 μm. This image is a classical benchmarking dataset used to validate algorithm accuracy (Wu et al., 2014). Since all the pixels have been labeled, the data are often used to validate different algorithms from various aspects. In our experiments, the number of bands was reduced to 200 by removing 20 water absorption bands and four noise bands. This image has a 20-m spatial resolution per pixel and a spatial dimension of 145 × 145 pixels. It contains 16 crop-type classes and a total of 10 366 labeled pixels, from which we randomly chose 10% as the training samples and used the rest for testing. The numbers of training and test samples for each class are summarized in Table 1.

Table 1
Training and test samples for the AVIRIS Indian Pines dataset. The classes are alfalfa [1], corn-notill [2], corn-min [3], corn [4], grass/pasture [5], grass/trees [6], grass/pasture-mowed [7], hay-windrowed [8], oats [9], soybeans-notill [10], soybeans-min [11], soybean-clean [12], wheat [13], woods [14], building-grass-trees [15], and stone-steel towers [16].

Class No.	1	2	3	4	5	6	7	8
Training	5	143	83	24	49	73	3	48
Testing	41	1285	747	213	343	675	25	430
Class No.	9	10	11	12	13	14	15	16
Training	2	98	246	60	21	127	39	10
Testing	18	874	2209	533	184	1138	347	83

Two important parameters, i.e. the sparsity level K_0 and window size T_0, needed to be refined for the JGSRC model, so that its best performance could be evaluated. The sparsity level K_0 controls how many atoms are selected for the representation of the test pixel. To determine K_0, we compared the OAs of the various methods by varying the sparsity level from $K_0 = 5$ to $K_0 = 50$, and the K_0 value associated with the highest accuracy was considered to be the best parameter setting. The results are plotted in Fig. 1, where the subplots from left to right and top to bottom are the window sizes equal to 3, 5, 7, 9, and 11, respectively.

It can be seen from each subplot that the increase of the sparsity level usually has a positive effect in improving the OA value. This is because the data contain a mixture of the classes' signatures, and a small number of atoms is inadequate to exactly identify the mixed pixels. We can also see that the classification accuracies acquired by the use of the SID and SSIM measures are significantly higher than those obtained by the ESD-based SRC in most cases. The

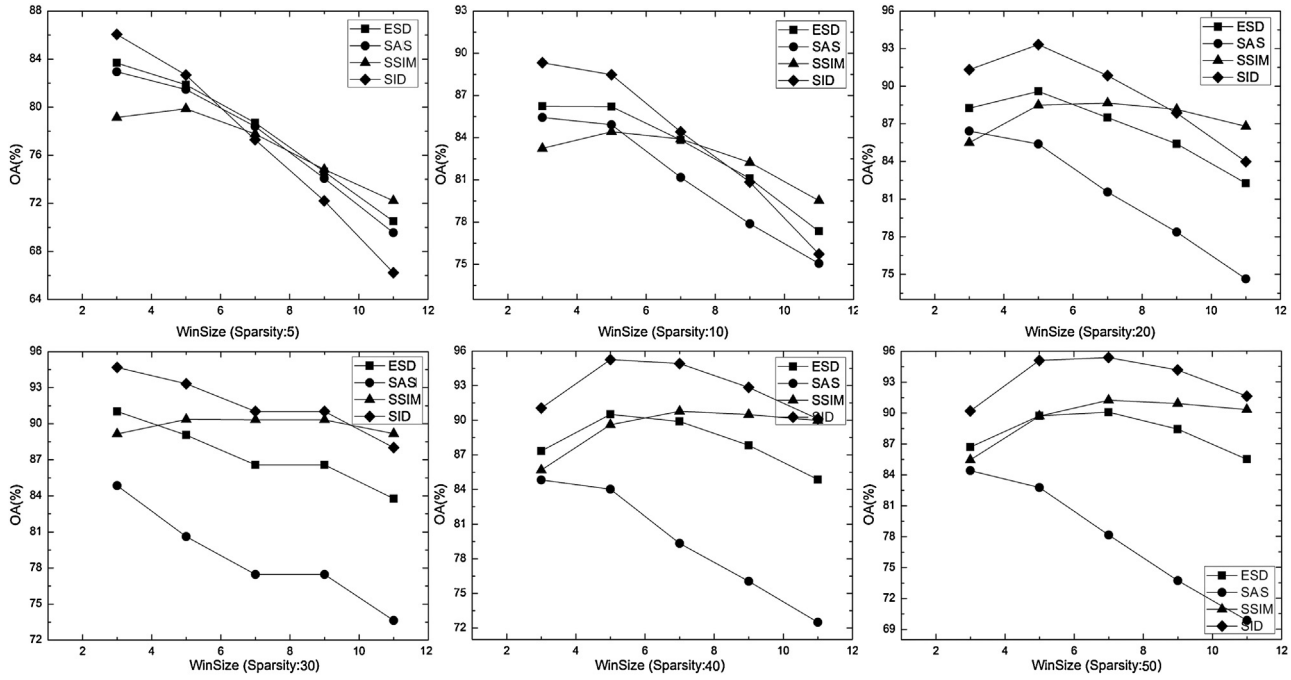

Fig. 2. Comparison of the OAs of the different methods with varying sparsity levels.

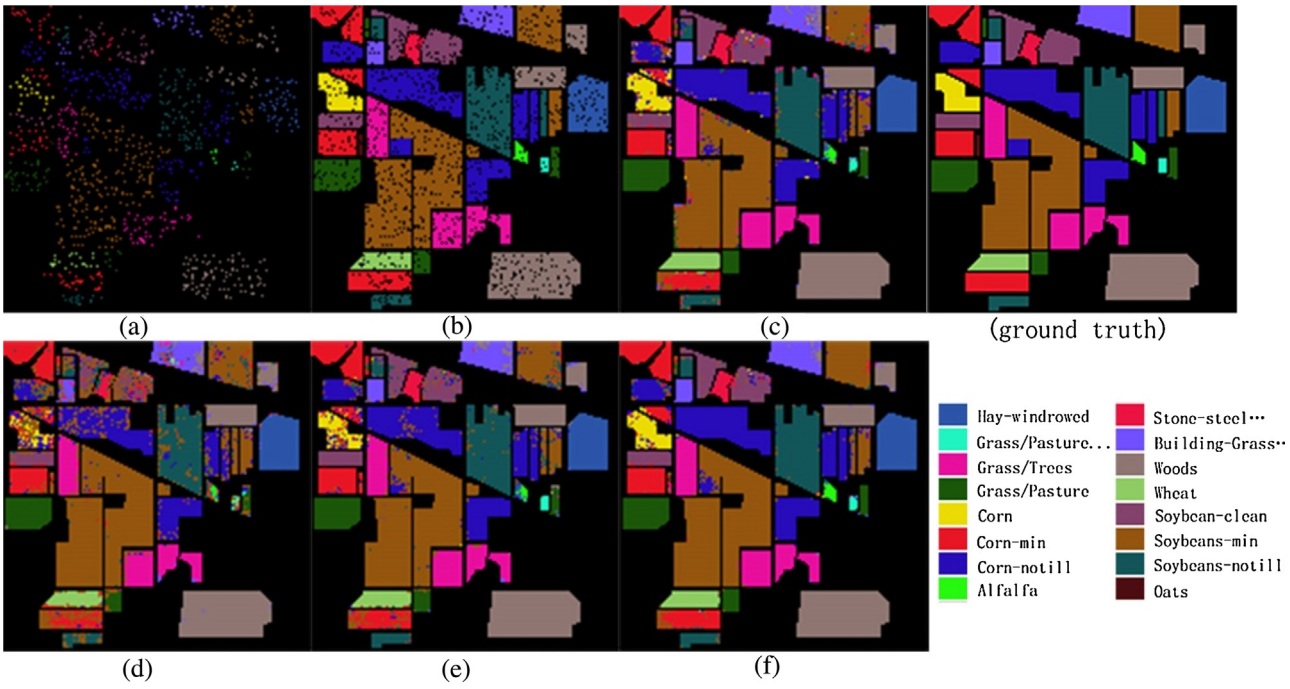

Fig. 3. Classification maps with different fidelity measures for the AVIRIS Indian Pines dataset: (a) training samples; (b) test samples; (c) ESD; (d) SAS; (e) SSIM; (f) SID.

SID measure achieves the best results among all the measures, and the highest OA is 95.37% when the sparsity and window size are equal to 50 and 7, respectively. Therefore, the results of this experiment imply that the use of different spectral similarity measures can generate totally different classification results.

Another important parameter is the window size T_0, which indicates how many neighboring pixels around the central pixel are simultaneously sparsely coded. In general, this parameter reflects how smooth the image is on the whole. The generated results are plotted in Fig. 2, where in each subplot, the vertical direction is the OA, and the horizontal direction indicates the size of the local window. It can be seen that in the case of a fixed sparsity level, the OAs for all the measures drastically decrease with the increase of the window size, especially when K_0 is small. One possible reason for this is that the neighborhood pixels around the central test sample may contain more heterogeneous materials with the increase of the window size, which invalidates the assumption that all the neighborhood pixels coming from the same class in the group sparse representation model shares a common sparsity support. Moreover, the neighboring pixels in a larger window may not be faithfully approximated by few atoms if the value of K_0 is small. In addition, it can be inferred from Fig. 1 and Fig. 2 that all four clas-

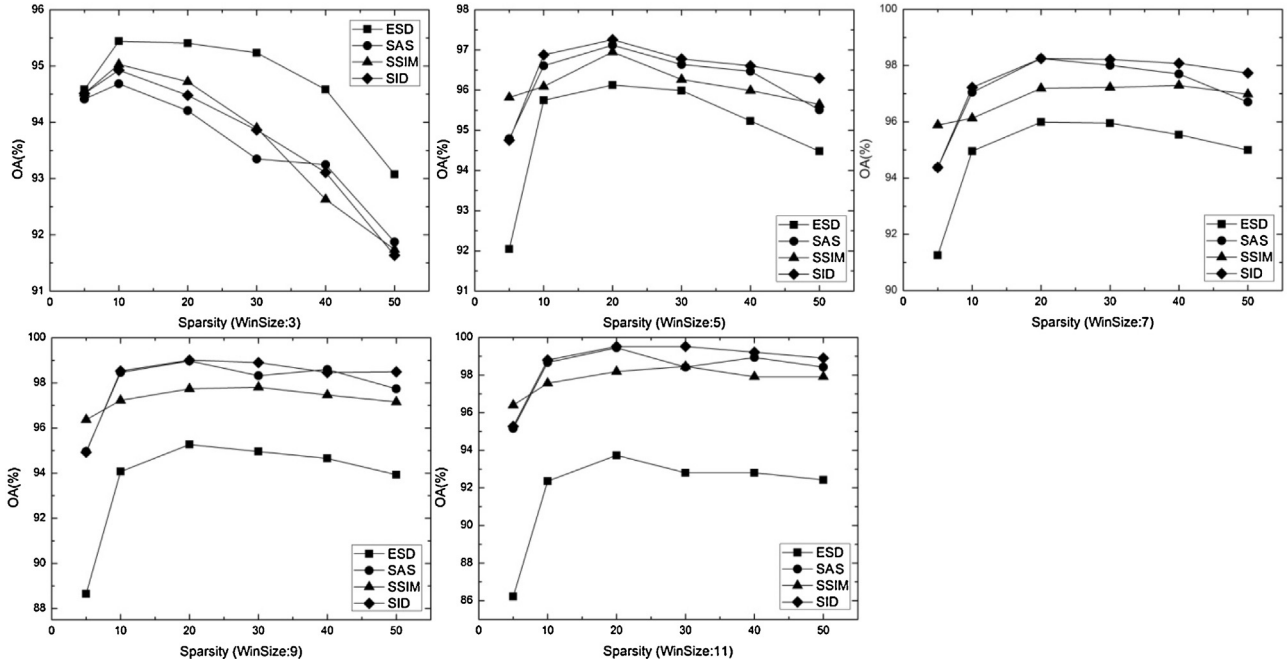

Fig. 4. Comparison of the OAs of the different methods with varying window sizes with the Hyperion Okavango Delta dataset.

Table 2
Comparison of the classification accuracies of the four different measures, i.e. ESD, SAS, SSIM, and SID.

Classes	ESD (%)	SAS (%)	SSIM(%)	SID (%)
Alfalfa	53.66	21.95	34.15	**82.93**
Corn-notill	86.07	67.86	89.34	**93.23**
Corn-min	83.00	69.08	83.53	**90.36**
Corn	84.04	22.54	61.03	**84.51**
Grass/Pasture	96.08	85.02	91.47	**96.54**
Grass/Trees	**98.93**	88.43	95.13	97.87
Grass/Pasture-mowed	68.00	24.00	**96.00**	92.00
Hay-windrowed	**100.00**	98.37	99.07	99.77
Oats	**22.22**	0.00	0.00	16.67
Soybeans-notill	83.64	72.88	87.53	**89.93**
Soybeans-min	90.04	95.20	94.79	**98.14**
Soybean-clean	83.86	44.09	83.49	**89.68**
Wheat	**98.91**	75.54	91.85	94.02
Woods	98.42	96.49	**98.77**	**98.77**
Building-Grass-Trees	83.29	67.72	87.61	**93.95**
Stone-steel Towers	**100.00**	50.60	93.98	97.59
OA	89.89	79.33	**90.77**	**94.81**
Kappa κ	0.885	0.759	**0.894**	**0.941**

sifiers generally have a positive relationship with K_0 and T_0, which suggests that a larger window size associated with a higher sparsity level can lead to a higher classification accuracy. Finally, the optimal parameter settings of the sparsity level and the size of window are $K_0 = 50$ and $T_0 = 7 \times 7$, respectively.

Using these parameter settings, we finally classified the AVIRIS image with the four specified fidelity measures. Fig. 3(a) and Fig. 3(b) show the spatial distributions of the training and test samples, respectively, and the SRC-classified maps with the ESD, SAS, SSIM, and SID measures are shown in Fig. 3(c)–(f), respectively. A visual comparison shows that SID achieves the best performance, and most of the land-cover types are correctly classified. In contrast, the classification map obtained with SAS generates the lowest accuracy, because it exhibits a lot of salt and pepper noise in several of the land-cover types, such as soybean and wheat.

These visual analyses are quantitatively illustrated by the confusion matrix in Table 2. A direct comparison of the results shows that SID achieves the best performance in terms of the OA and kappa coefficient indictors, with the highest values of 94.8% and 0.941, respectively, which are much higher than the values achieved by ESD, with 89.9% and 0.885. The improvements of OA and kappa are about 5% and 0.06, respectively. The second best results are obtained by SSIM, and SAS obtains the lowest accuracy, which is consistent with the results shown in Fig. 3. We can also see that the classification accuracy of oats is low for all the methods. This is mainly due to the small number of training samples for this class, and the high degree of mixing of the classes' signatures in this image.

3.1.2. Dataset 2: Hyperion

The second HSI used in our experiments was from the Okavango Delta, Botswana, acquired by the Hyperion sensor onboard NASA's Earth Observing-1 (EO-1) satellite on May 31, 2001, which can be freely download from website http://www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes. The Hyperion sensor acquires data at a 30-m pixel resolution over a 7.7 km strip in 242 bands covering the 400–2500 nm portion of the spectrum in 10 nm windows. This dataset from north-west Botswana comprises permanent marshlands and seasonally flooded plains, and consist of observations from 14 identified classes representing the different land-cover types (Ham et al., 2005). In our experiment, the uncalibrated and noisy bands covering the water absorption features were first removed, and the remaining 145 bands were used in the experiment. A total of 3248 labeled pixels were available, from which we randomly chose 10% as the training samples and the rest were used for the testing. Details about the training and test sets are provided in Table 3.

In order to optimize the two important parameters, i.e., sparsity level K_0 and window size T_0, we again plotted the relationships between their OAs and parameters K_0 and T_0 for the different methods, which are shown in Figs. 4 and 5, respectively.

It can be seen that the corresponding best parameters for K_0 and T_0 are 20 and 11×11, respectively, for this dataset. For the Hyperion Okavango Delta dataset, the optimal sparsity level is lower than that for the AVIRIS Indian Pines dataset. This is because most of the data are homogeneous areas, which contain fewer mixed signatures. Therefore, for most of the test pixels, selecting a small

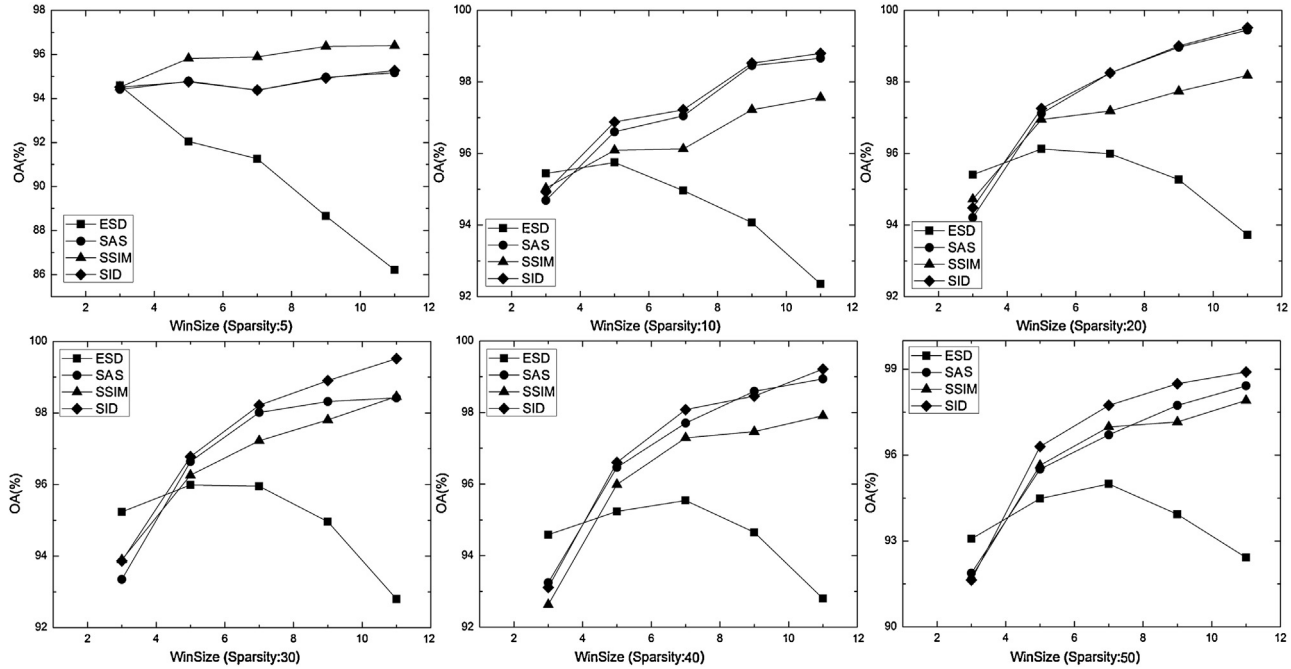

Fig. 5. Comparison of the OAs of the different methods with varying sparsity levels with the Hyperion Okavango Delta dataset.

Table 3
The 14 ground-truth classes of the Hyperion Okavango Delta dataset, showing the numbers of training and test sets for each class. The class names are water [1], hippo grass [2], floodplain grass 1 [3], floodplain grass 2 [4], reeds [5], riparian [6], fire scar [7], island interior [8], acacia woodland [9], acacia shrubland [10], acacia grassland [11], short mopane [12], mixed mopane [13], and exposed soils [14].

Class No.	1	2	3	4	5	6	7
Training	27	11	26	22	27	27	26
Testing	243	90	225	193	242	242	233
Classes No.	8	9	10	11	12	13	14
Training	21	22	25	31	19	27	10
Testing	182	282	223	274	162	241	85

number of atoms is adequate to reconstruct the test pixels. Moreover, it can be inferred from Fig. 5 that all the methods except for the ESD-based classifier exhibit a positive relationship with the spatial window size, indicating that the nonlinear spectral similarity measures have more powerful discriminative abilities, and hence a larger window size can lead to a higher classification accuracy.

The quantitative evaluation results, including the classification accuracy for every class, the OA, and the kappa coefficient, are shown in Table 4. Here, it can be seen that the SID-based technique shows the best performance, with the highest OA and kappa coefficient of 98.08% and 0.98, respectively. We can also see that the SAS-, SSIM-, and SID-based models significantly outperform the ESD-based model, implying that ESD is not the best choice of similarity measure in terms of classification accuracy.

3.2. Testing the improvements of the proposed algorithm

In this section, we aim to demonstrate whether or not the proposed JGSRC algorithm outperforms its counterparts developed in other studies, i.e., SSIM-based JSRC (Wu et al., 2015) and SID-based JSRC (Zhang et al., 2014a). Note that SAS- and ESD- based SRC models were not included in the comparison, because a SAS-based SRC method is not yet available in the literature, while ESD-based SRC is the same as the GSRC model shown in Eq. (14). Two issues are addressed. One issue is whether or not the proposed algorithm consistently outperforms the other methods. To this aim, we varied the parameters of window size and sparsity level to obtain their corresponding optimum classification accuracies for the comparison.

Table 4
Comparison of the classification accuracies of the four different measures for the Hyperion Okavango Delta dataset.

Class	ESD (%)	SAS (%)	SSIM (%)	SID(%)
1	100.0	100.0	100.0	100.0
2	97.78	100.0	100.0	100.0
3	99.11	99.56	99.11	100.0
4	100.0	100.0	100.0	100.0
5	69.83	97.11	85.95	94.63
6	88.43	87.60	84.30	90.91
7	100.0	100.0	100.0	100.0
8	100.0	100.0	100.0	100.0
9	98.23	96.81	90.43	97.52
10	99.55	99.10	99.10	98.65
11	99.27	100.0	100.0	100.0
12	98.15	95.68	98.77	96.30
13	97.51	95.85	95.85	98.34
14	90.59	98.82	97.65	98.82
OA	95.54	97.70	97.29	98.08
Kappaκ	0.952	0.975	0.971	0.980

Table 5 reports the comparison between the JGSRC model and the SSIM- and SID-based JSRC models (Wu et al., 2015; Zhang et al., 2014a) for the AVIRIS Indian Pines dataset. It can be seen that the proposed JGSRC generally outperforms its counterparts when the sparsity level is greater than 30. This is because JGSRC requires more independent atoms to represent the test pixel in the nonlinear measuring space. Closer observation shows that the JGSRC method obtains much better gains than the other methods with the increase of K_0. For example, with a fixed window size of 7×7, the gains of the JGSRC method for the SSIM and SID measures are 14.3% and 18.1%, respectively, when the sparsity level K_0 is varied from 5 to 50. In contrast, the gains of the counterpart methods are 13.2% and −2.14%, respectively. This experiment implies that the accuracies of the other algorithms generally reach saturation point with a low value of K_0, while the JGSRC method continuously improves with the increase of K_0, which is due to its ability to select independent atoms with the gradient descent algorithm. Moreover, the steps of projecting the remaining dictionary atoms and renormal-

Table 5
Comparison of the JGSRC model with the SSIM- and SID-based JSRC models, respectively, for the AVIRIS Indian Pines dataset.

3 × 3	K0											
	SSIM (%)						SID (%)					
	5	10	20	30	40	50	5	10	20	30	40	50
Others	82.37	86.09	87.77	**88.56**	87.91	85.99	**88.18**	**88.18**	88.10	87.71	87.30	86.68
Proposed	79.14	83.23	85.50	**85.79**	85.68	85.44	86.06	89.33	91.35	**91.26**	91.05	90.19
5 × 5	5	10	20	30	40	50	5	10	20	30	40	50
Others	**88.51**	88.47	88.22	87.11	86.96	85.84	74.42	82.65	85.14	86.66	89.75	**91.28**
Proposed	74.81	82.23	88.15	90.32	90.49	**90.90**	72.21	80.84	87.86	91.08	92.84	**94.16**
7 × 7	5	10	20	30	40	50	5	10	20	30	40	50
Others	77.84	85.16	87.52	89.02	**91.58**	91.22	**92.85**	92.81	92.61	91.83	91.29	90.44
Proposed	77.77	83.90	88.66	90.36	90.77	**92.09**	77.28	84.41	90.84	93.33	94.91	**95.37**
9 × 9	5	10	20	30	40	50	5	10	20	30	40	50
Others	74.42	82.65	85.14	86.66	89.75	**90.90**	**88.51**	88.47	88.22	87.11	86.96	85.84
Proposed	74.81	82.23	88.15	90.32	90.49	**91.28**	72.21	80.84	87.86	91.03	92.84	**94.16**
11 × 11	5	10	20	30	40	50	5	10	20	30	40	50
Others	**82.84**	82.78	82.48	81.19	81.15	80.28	70.85	79.25	82.50	84	87.06	**88.72**
Proposed	72.22	79.53	86.80	89.17	89.98	**90.31**	66.23	75.73	83.99	88.01	90.12	**91.60**

Table 6
The average Z-values for the proposed method and the other classifiers for both experimental datasets.

	AVIRIS dataset (%)		Hyperion dataset (%)	
	SSIM	SID	SSIM	SID
Others	91.22	93.66	96.81	96.30
Proposed	92.09	95.37	98.46	99.62
Gain	0.87	1.71	1.65	3.32
Z- value	1.32	2.16	2.27	3.58

izing them in the JGSRC algorithm play a crucial role in the process of atom selection.

Another issue is whether or not the proposed algorithm significantly outperforms its counterparts. To this end, we utilized the nonparametric McNemar's test (Foody, 2004) to assess the statistical significance in accuracy improvement. This test is based on the standardized Z-test statistic:

$$Z = \frac{c_{12} - c_{21}}{\sqrt{c_{12} + c_{21}}} \tag{16}$$

where c_{12} denotes the number of samples classified correctly and incorrectly by the proposed model and the other models, respectively. Accordingly, c_{12} and c_{21} are the counts of the classified samples on which the considered first and second models disagree. A lower prediction error (higher accuracy) is identified by the sign of Z. A positive sign indicates that the results from c_{12} are more accurate than the results from model c_{21}. At the commonly used 5% level of significance, the difference in the accuracies between the first and second models can be considered to be statistically significant if $Z > 1.96$. This experiment allowed us determine whether the differences in classification accuracy were significant.

The average Z-test values for the proposed method and the other classifiers with both experimental datasets are shown in Table 6. From Table 6, it is clear that, in most cases, the proposed method performs significantly better than the other methods in terms of classification accuracy.

4. Conclusions

This paper has presented a general SRC model for HSI classification to accommodate various spectral similarity measures. Two real HSI images were used to demonstrate the effectiveness of the proposed model, and the experiments confirmed that the generalized model can deal with various spectral similarity measures, and the classification accuracies of both the SSIM- and SID-based SRC were consistently better than that of the ESD-based SRC. However, the SAS-based SRC exhibited an unstable performance with the test datasets, because it is prone to misclassification in the case of containing a mixture of the classes' signatures. The experiments also showed that the operation of projecting the remaining dictionary atoms can have a positive effect on the HSI classification and can significantly improve the classification accuracy in most cases.

Acknowledgements

This work was supported in part by the Natural Science Foundation of China under Grant 41571330, by the Natural Science Foundation of Fujian province (grant no. 2015J01163), and the Director's Opening Fund of Key Laboratory of Poyang Lake Wetland and Watershed Research(Jiangxi Normal University), Ministry of Education (ZK2013004).

References

Chang, C., 2000. An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis. IEEE Trans. Inf. Theory 46 (5), 1927–1932.
Chen, Y., Nasrabadi, N.M., Tran, T.D., 2011. Hyperspectral image classification using dictionary-based sparse representation. IEEE Trans. Geosci. Remote Sens. 49 (10), 3973–3985.
Chen, Y., Nasrabadi, N.M., Tran, T.D., 2013. Hyperspectral image classification via kernel sparse representation. IEEE Trans. Geosci. Remote Sens. 51 (1), 217–231.
Du, Y., Chang, C.I., Ren, H., Chang, C., Jensen, J.O., 2004. New hyperspectral discrimination measure for spectral characterization. Opt. Eng. 43 (8), 1777–1786.
Fang, L., Li, S., Kang, X., Benediktsson, J., 2014. Spectral-spatial hyperspectral image classification via multiscale adaptive sparse representation. IEEE Trans. Geosci. Remote Sens. 52 (12), 7606–7618.
Foody, G.M., 2004. Thematic map comparison: evaluating the statistical significance of differences in classification accuracy. Photogramm. Eng. Remote Sens. 70 (5), 627–633.
Ham, J., Chen, Y., Crawford, M.M., Ghosh, J., 2005. Investigation of the random forest framework for classification hyperspectral data. IEEE Trans. Geosci Remote Sens. 43 (3), 492–501.
Liu, J., Wu, Z., Wei, Z., Xiao, L., Sun, L., 2013. Spatial-spectral kernel sparse representation for hyperspectral image classification. IEEE Sel. Top. Appl. Earth Obs. 6 (6), 2462–2471.
Liu, C., 2007. Comparative assessment of the measures of thematic classification accuracy. Remote Sens. Environ. 107, 606–616.
Natarajan, B.K., 1995. Sparse approximate solutions to linear systems. SIAM J. Comput. 24, 227–234.
Pontius, R., Millones, M., 2011. Death to Kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. Int. J Remote Sens. 32, 4407–4429.
Rehman, A., Rostami, M., Wang, Z., Brunet, D., Vrscay, E.R., 2012. SSIM-inspired image restoration using sparse representation. EURASIP J. Adv. Signal Process. 2012 (1), 1–12.

Tang, Y., Yan, H., Li, L., 2014. Manifold- based sparse representation for hyperspectral image classification. IEEE Trans. Geosci. Remote Sens. 52 (12), 7738–7749.
Van Der Meer, F., 2006. The effectiveness of spectral similarity measures for the analysis of hyperspectral imagery. Int. J. Appl. Earth Obs. Geoinf. 8 (1), 3–17.
Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. 13 (4), 600–612.
Wu, B., Zhang, L., Zhao, Y., 2014. Feature selection via Cramer's V-test discretization for remote sensing image classification. IEEE Trans. Geosci Remote Sens. 52 (5), 2593–2606.
Wu, B., Zhu, Y., Chen, C., Wang, X., 2015. SSIM-based sparse classifier for hyperspectral imagery classification. In: Proc. IGIT 2015 Int. Conf. Geo. Inform. Tech. Appl. Towards GEOSS, Jan., 16-17, Szekesfehervar, Hungary, pp. 38–46.
Zhang, E., Zhang, X., Yang, S., 2014a. Improving hyperspectral image classification using spectral information divergence. IEEE Geosci. Remote Sens. Lett. 11 (1), 249–253.
Zhang, H., Li, J., Huang, Y., Zhang, L., 2014b. A nonlocal weighted joint sparse representation classification method for hyperspectral imagery. IEEE Sel. Top. Appl. Earth Obs. 7 (6), 2057–2066.