Multimed Tools Appl (2016) 75:12137–12154
DOI 10.1007/s11042-016-3320-7

A remote sensing image classification method based on sparse representation

Shulei Wu[1,2] · **Huandong Chen**[2] · **Yong Bai**[1] · **Guokang Zhu**[3]

Received: 21 July 2015 / Revised: 19 January 2016 / Accepted: 27 January 2016 /
Published online: 20 February 2016

Abstract With the development of remote sensing image applications, sparse-based representation classification approaches have been investigated for better classification accuracy. This paper introduces an improved classification method based on sparse representation by representing the test samples through a dictionary. The key components of our proposed method rely on the feature dictionary construction, sparse representation and image reconstruction. The dictionary is obtained by training samples according to their class for a sparse linear combination. The sparse representation for the image is expressed as sparse coefficients by solving an optimization problem. We describe the method of constructing a dictionary by computing a best matrix to represent all data vectors. We also describe the algorithm used to solve for the sparse representation. Finally, we discuss the way of using the sparse vector to reconstruct the image for classification. In the experiments, the proposed method is applied to two real high spatial resolution images for the classification in comparison to Backpropagation Neural Network, Support Vector Machine, Classification and Regression Trees and K-means. The experimental results show that the proposed method performs better than the benchmark methods in terms of classification accuracy.

Keywords Image classification · Sparse representation · Image reconstruction · Remote sensing

✉ Yong Bai
baiyonghnu@163.com

[1] College of Information Science and Technology, Hainan University, No. 58 Renming Road, Haikou 570228, China

[2] College of Information Science and Technology, Hainan Normal University, No. 99 South Longkun Road, Haikou 571158, China

[3] College of Computer Science and Technology, Shanghai University of Electric Power, No. 2588 Changyang Road, Yangpu District, Shanghai 200090, China

1 Introduction

The remote sensing imagery of high spatial resolution (HSR) provides useful geometric and detailed information which can precisely represent the Earth's surface. Due to the increasing applications of HSR remote sensing imagery, a major issue of land-cover classification is how to improve the accuracy of the image processing. The common HSR remote sensing imagery is obtained from satellites, such as IKONOS, QuickBird, WorldView-2 and Pleiades. The availability of HSR increases the possibility of accurate Earth observations [27] and makes it possible to be widely used. However, urban landscapes become more complicated and have many different objects with similar spectral features. The increasing resolution does not facilitate improvement of the classification accuracy in the same level. Therefore, it is necessary to explore more effective approaches and incorporate the spatial features to deal with the HSR images.

This paper focuses on the problem of classification of a given high-resolution image according to different objects. Our approach is motivated by those research works [9, 36, 39, 40] on sparse signal representation, which suggest that the linear relationship among high-resolution signal elements. We propose an improved strategy to train a dictionary [36] by utilizing the sparsity of the input samples and construct a sparse model to classify the pixels in remote sensing image through adopting the error residual for sparse representation [9, 39]. The sparse vector representing the atoms for the test spectral pixels can be recovered by solving an optimization problem [40]. The classes of the test pixels can then be determined by the characteristics of the recovered sparse vector.

The remainder of this study is organized as follows. In Section 3, the sparse-based representation classification method is introduced. In Section 4, the results of our experiments and the analyses are described, and the effectiveness of our proposed method is demonstrated. Section 5 summarizes this work and draws the future work.

2 Related Work

Toward the classification, various classification approaches have been developed in order to improve the accuracy of the classification, including Independent Component Analysis [30], Artificial Neural Networks [15], Back Propagation Neural Network (BPNN) [4, 16, 45], Hierarchical Hybrid Fuzzy-Neural Network [37], K-Nearest Neighbor [43], likelihood classifier [31], Support Vector Machine (SVM) [3, 6, 25], Classification and Regression Trees (CART) [8], K-means [23, 34] and decision tree classification [20]. Giacinto et al. [14] proposed an approach to the automatic design of effective neural network ensembles, to select the subset formed by the most error-independent nets. Conventional cluster technique such as K-means [23, 34] has been used for image segmentation over years. Luo et al. [23] proposed a spatial constrained K-means approach to solve the image segmentation problem. Back-propagation neural network [16] algorithm, which is a gradient-based method, was explored for classification of multispectral image data. A variation of the SVM-based algorithms [41] put forward a set of tools for structured classification, and generalized the traditional non-structured classification approaches.

However, the above traditional classifiers are inadequate for HSR imagery [17]. In this context, the features [1, 11, 18, 19, 29, 32, 35] were used to enhance the spectral information and raise the classification accuracy. Ouma and Tateishi [29] presented a pre-classification

filtering method based on unsupervised multiresolution non-linear image filtering that combines spectral and textural image characteristics. The local texture characteristics were extracted via wavelet decomposition. Huang et al. [18] proposed some statistical measures to extract some structural features and used different classifiers including maximum likelihood classifier, BPNN, probability neural network based on expectation–maximization training, and SVM to process the hybrid spectral-structural features after the steps of spatial feature extraction and dimension reduction. Pingel et al. [32] developed the Morphological Filter algorithm to be competitive with other ground filtering algorithms for LIDAR and established a baseline performance for a progressive morphological filter implemented in its simplest form.

Researchers proposed to exploit spatial information for complementing the spectral feature space and enhancing separability of the spectrally similar classes [5, 11, 42]. Dópido et al. [11] developed a semisupervised self-learning framework in which the machine learning algorithm itself selects the most useful and informative unlabeled samples for hyperspectral image classification. However, this method was dependent on the assumption that the pixels with similar spectral signature belong to the same class. This might be possible for hyperspectral images, but not for multispectral images, since they contain many spectral ambiguities (e.g., roofs and roads, water and shadow). Bruzzone et al. [5] proposed a pixel-based system, which was aimed at obtaining accurate and reliable maps both by preserving the geometrical details in the images and by properly considering the spatial-context information, for the supervised classification of high spatial resolution images. Tuia et al. [42] presented a classification method for very high resolution images by exploiting efficient multisource information, both spectral and spatial through the combination of SVMs and composite kernels. Fauvel et al. [13] used kernel methods which deal with the joint use of the spatial and the spectral information through a support vector machine formulation.

Moreover, this is meaningful for classification of land cover, but not sufficient for applications of urban mapping, since the impervious surfaces need to be elaborated into more detailed objects (e.g., tree, residential area, and water). Therefore, it is desired to explore more effective algorithms, such as sparse representation and compressive sensing. Sparse representation has been an extremely powerful tool in many classical signal processing applications.

For sparse representation, Chen and Donoho proposed the so-called Basis Pursuit (BP) algorithm [7]. BP is a principle for decomposing a signal into an optimal super position of dictionary elements, and optimal means to have the smallest l^1 norm of coefficients among all such decompositions. Mallat and Zhang [24] used an over-complete redundant dictionary for signal representation. They gave rise to the Matching Pursuit (MP) algorithm for the sparse reconstruction, and pointed out that the stronger a sparse signal is, the more accurate the reconstruction will be. The MP algorithm is a greedy algorithm, but is different from the BP algorithm. MP is a local optimization algorithm, of which the final result may not converge and may not necessarily to find the global optimal solution. Differently, Tropp and Gilbert presented the Orthogonal Matching Pursuit (OMP) algorithm [38] to deal with the convergence problem to obtain the most matching signal. In optimization, OMP selects an atom set to conduct the orthogonal optimization Gram-Schmidt in each iteration. In OMP, fewer samples are required and less iterative times are needed to achieve the optimal result compared with MP. Olshausen [28] pointed out that each image has a sparse nature.

Whereafter, the theory of the sparse theory developed rapidly. It has been adopted and employed effectively in the field of image processing [2, 12]. Furthermore, a better performance of sparse optimization algorithm [33] was proposed. Donoho and Candes

presented the concept of compressive sensing [10] based on the sparse theory to further develop the sparse signal representation theory.

In very recent years, sparse representation has been further studied in literature [21, 22, 26, 42, 44, 45]. A nonlocal weighted joint sparse representation classification method [46] was proposed to improve the remote sensing image classification result, with different weights, for different neighboring pixels around the central test pixel, and the simultaneous orthogonal matching pursuit technique. Moody et al. [26] presented a technical method of land-cover unsupervised classification in multispectral satellite imagery, using sparse representations in learned dictionaries: clustering on sparse approximations and applying a Hebbian learning rule to build multispectral, multi-resolution dictionaries. In [22], Zhang et al. proposed a hyperspectral image anomaly detection approach using background joint sparse representation, which adaptively selects the most representative background bases for the local region. Zhang et al. [21] put forward a superpixel-level sparse representation classification resolution with multitask learning for hyperspectral imagery. Their proposed algorithm exploited the class-level sparsity prior for multiple-feature fusion, and the correlation and distinctiveness of pixels in a spatial local region. Yu et al. [44] proposed a remote sensing image classification method based on sparse component analysis, whose classification result is more reliable and more accurate.

3 Image Classification Model

In this paper, we focus on classification of a given high-resolution image using DigitalGlobe's WorldView-2 satellite imagery. The main contribution of this paper is to develop an efficient solution for image classification by using nearest neighbor joint sparse linear combination to build the feature dictionary and applying pursuit algorithm joint sparse representation for image reconstruction. In this section, we mainly introduce the key components of our proposed method: the feature dictionary construction, sparse representation and image reconstruction. The idea of constructing the dictionary is to find a best matrix to represent all data vectors through extracting features directly from the data itself by nearest neighbor. We select randomly the training data set to construct the feature dictionary according to their classes by a sparse linear combination. So we describe the algorithm used to solve for the sparse representation. In our method, the sparse coefficients of test samples are divided into several groups, corresponding to the dictionary components representing specific classes. The test samples of image are represented by the sparse representation. We then discuss how to determine the class of the test pixel. The proposed classification model is shown in Fig. 1. The proposed classification model mainly consists of three steps: (1) feature dictionary construction, (2) sparse representation, (3) classification decision.

3.1 Sparse representation model

Let f be a pixel observation from an input signal with $l-$dimension for classification. In the sparse representation model, test spectral pixels, which lie approximately in several subspaces, are approximately represented by a few training examples. Suppose we have T distinct classes, and any one training sample for each class have n training data. This training sample can be trained to k dictionary elements. And test samples can be modeled to the T subspaces according to the T classes from the dictionary D. If the pixel f belongs to the ith class, we can represent f

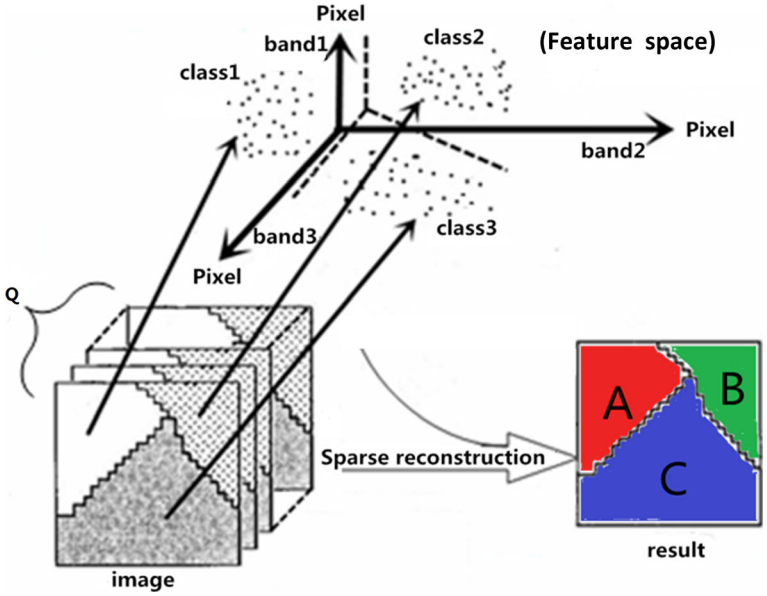


Fig. 1 Proposed classification model

through these training data as a linear combination for the *i*th class. Thus, the test pixel *f* can be expressed as

$$f = D\alpha = \begin{bmatrix} d_1^i & \cdots d_j^i \cdots & d_n^i \end{bmatrix} \begin{bmatrix} \alpha_1^i \\ \vdots \\ \alpha_n^i \end{bmatrix} = d_1^i\alpha_1^i + d_2^i\alpha_2^i + \cdots + d_n^i\alpha_n^i,\ 1 \le i \le T,\ 1 \le j \le n, \quad (1)$$

where $D = \{d_j^i\}_{j=1,i=1}^{n,T}$ is a feature dictionary which totally has *n* training data from the input sample of *i*th class and α^i is a sparse vector. The coefficients of the sparse representations α can be decomposed to *T* pieces, each α^i is a sparse vector which has only a few nonzero entries. Therefore, the sparse representation of the test pixel *f* can also be expressed as a linear combination of only the *K* dictionary atoms α_k (k = 1, . . . , *K*) which is a vector with *K* ($K = \|\alpha\|_0$) nonzero entries. Thus, *f* can be written as

$$f = D\alpha = \begin{bmatrix} d_1^i & \cdots & d_K^i \end{bmatrix} \begin{bmatrix} \alpha_1^i \\ \vdots \\ \alpha_K^i \end{bmatrix} = d_1^i\alpha_1^i + d_2^i\alpha_2^i + \cdots + d_K^i\alpha_K^i,\ 1 \le i \le T \quad (2)$$

where *K* denotes the number of nonzero elements in the vector α. Next, we will train a dictionary from a set of input samples. And we also will introduce how to obtain the sparse vector α and how to classify test samples from the sparse vector α.

3.2 Feature space construction

We consider a method for constructing the dictionary that produces sparse representations for the training examples. For sparse representation, it is a procedure of computing the representation coefficients based on the given examples and dictionary. Here, we will construct the feature dictionary from the input examples. In the proposed model, assume

that the pixels of spectral features belonging to the same class approximately lie in the same subspace. The construction strategy of the feature dictionary is to model the best centers based on the training examples to express the most distinct characteristics of the presented objects.

Given a remote sensing image with Q channels and $N \times M$ pixels as an input signal be such a set $F = \{f_{i,j}^l\}_{l=1,i=1,j=1}^{Q,N,M}$ ($1 \le l \le Q$, $1 \le i \le N$, $1 \le j \le M$), where $l = 1, 2, \cdots, Q$, and N, M is the number of rows and columns respectively. Suppose there be T distinct classes contained in the image in accordance with different plants or objects, and any one class has n training data. We select T types of representative samples from the training dataset, and input them into a sample set, $S = (s_1, \cdots, s_i, \cdots, s_T)$ ($1 \le i \le T$), where s_i is a subset corresponding to the ith class. It contains n data points $[x_1^i, x_2^i, \cdots, x_n^i]$ with l bands, where x^i is a data point in the subset s_i. Then, we construct the feature vector $D = [d_1, d_2, \cdots, d_K]$, which can be viewed as a dictionary including a total of K ($K << NM$) elements, where $D \in \mathbb{R}^{Q \times K}$. In addition, associated with this feature matrix, we have a class index table $W = [I_1, \cdots, I_i, \cdots I_K]$, where $1 \le I_i \le T$, and I_i records the class label of the feature pixel i, $i = 1, 2, \cdots, K$, that is, I_i indicates the class which the dictionary element d_i belongs to. For the given training set of image patches, each is reshaped as a two-dimensional vector. For better description, this image is rewritten as $F = \{f_{fj}^l\}_{l=1,j=1}^{Q,NM}$, ($1 \le l \le Q$, $1 \le j \le NM$), which can be represented as a sparse linear combination of these feature vectors. The representation of F may be approximate, that is $F \approx D\alpha$, which satisfies the constrain $\|F - D\alpha\|_2 \le \varepsilon$. The vector α involves the representation coefficients of the image F. We can write $f_j = D\alpha_j$, where $\alpha_j = e_j$ is a vector from the trivial basis, with all zero elements except the one in the pth position. The index p is selected such that

$$\forall p \neq q \left\| f_i - D\alpha_p \right\|_2^2 \le \left\| f_i - D\alpha_q \right\|_2^2. \tag{3}$$

For the sparse representation of the data set F, the minimization of error is computed in order to search the best possible dictionary D with K items. And it could alternatively be met by considering

$$< D, W, \alpha >= \arg\min_{\alpha} \| F - D\alpha \|_2^2 \; s.t \; \forall j, \; \alpha_j = e_j. \tag{4}$$

Algorithm 1 (**Training a dictionary**).

Task: Find a best matrix to represent all data vectors for constructing a dictionary by nearest neighbor.

Input: A remote sensing image with Q channels and $N \times M$ pixels $F = \{f_{i,j}^l\}_{l=1,i=1,j=1}^{Q,N,M}$, $1 \le l \le Q$, $1 \le i \le N$, $1 \le j \le M$.

Initialization: Randomly select k ($k = K/T$) data points from a sample subset s_i as the initial representatives $\varphi^i = [\varphi_1^i, \varphi_2^i, \cdots, \varphi_k^i]$, set i = 1 and repeat it until i reaches T.

1: Compute the k best centers from these n data points for each training sample s_i, and w_j^i records the index of the best possible point for each data sample, $w_j^i = \{p \mid \forall p \neq q, \|x_j^i - \varphi_p^i\|_2 < \|x_j^i - \varphi_q^i\|_2\}$, $1 \le p, q \le k$, $1 \le j \le n$, $1 \le i \le T$.

2: The representatives $[\psi_1^i, \psi_2^i, \cdots, \psi_k^i]$ is obtained by the following formula: $\psi^i = \{x_p^i \mid \forall p \neq q, \|x_p^i - x_m^i\|_2 < \|x_q^i - x_m^i\|_2\}$, $x_p^i, x_q^i, x_m^i \in \varphi^i$.

3: Update φ^i by $\varphi^i = \psi^i$.

4: Go back to step 2. Repeat it until ψ^i is equal to φ^i.

Output: A dictionary and a class vector.

3.3 Reconstruction and classification

We describe the way we use the sparse vector α for a test sample f_j ($1 \le j \le NM$) when reconstruct and classify it. At the moment, the dictionary D is obtained and known. Every image patch f_j could be represented sparsely over this dictionary. And the representation α_j satisfying $D\alpha_j = f_j$ is obtained by solving the following optimization problem:

$$\hat{\alpha}_j = \arg\min \left\|\alpha_j\right\|_0 \; s.t \; D\alpha_j = f_j. \tag{5}$$

In order to solve the problem of searching the sparsest representation of f_j, the equality constraint in (5) can be formulated to an inequality one as

$$\hat{\alpha}_j = \arg\min \left\|\alpha_j\right\|_0 \; s.t \; \left\|D\alpha_j - f_j\right\|_2 \le \varepsilon, \tag{6}$$

where ε is the error. The above problem can also be considered as minimizing the approximation error within a certain sparsity level. We can compute the error residual as $r_j = f_j - D\alpha_j$. Notice that the above optimization problem can be replaced by

$$\hat{\alpha}_j = \arg\min \left\|D\alpha_j - f_j\right\|_2 \; s.t \; \left\|\alpha_j\right\|_0 \le L, \tag{7}$$

where L express the sparsity level for the approximation error. Compute the residual for the ith class, that is, the error between the test sample f_j and the reconstruction from training samples in the ith class. The class of f_j can be determined by the recovered sparse vector $\hat{\alpha}_j$ as

$$c\left(f_j\right) = \arg \max_i |D^i \hat{\alpha}_j^i|, \; s.t \min \|f_j - D^i \hat{\alpha}_j^i\|_2, \; \forall i, \; 1 \le i \le T, \tag{8}$$

where $\hat{\alpha}_j^i$ denotes the portion of the recovered sparse coefficients corresponding to the training samples in the ith class.

Eventually, we can obtain the final classification result $\hat{F}$ for the image F as (9).

$$\hat{F} = \left\{ f_j \middle| f_j = color_{W(j_0)}, \forall c\left(f_j\right) \in W(j_0), \; 1 \le W(j_0) \le T, \; 1 \le j \le NM \right\} \tag{9}$$

Algorithm 2 (Reconstruction and Classification).

Task: Construct the image and determine the class of the test pixels.

Input: A normalized feature dictionary D, class vector W and sparsity level L.

Initialization: Set $j = 1$ and repeat it until j reaches NM.

1: Choose the index j_0, $1 \le j_0 \le K$, such that $\left|\varphi_{j_0}{}^T r_{j_0}\right|$ is maximized. We say that j_0 is the index of the maximum value of the product of the residual r_j and the atom φ_j, $j = 1, 2, \cdots, NM$, from the class index table W, i.e. $j_0 = \arg\max_{j=1 \ldots NM} |\langle r_j, \varphi_j \rangle|$.

2: Update the index set by $I_j = I_{j-1} \cup \{j_0\}$ and the size incremental matrix by $A_j = A_{j-1} \cup \left\{\varphi_{j_0}\right\}$, when the product of the residual r_j and the atom φ_j is the maximum value. Then, remove the current column vector φ_{j_0} from the dictionary D, denoted by $D = D \backslash \left\{\varphi_{j_0}\right\}$.

3: In this step, first, decompose A_j by $A_j = UZV^T$. Then we can obtain the orthogonal vectors U and V^T, and the singular value vector Z whose diagonal elements are called singular value. In addition to the diagonal elements, the value of its

elements is zero. Therefore, we calculate the sparse coefficient by $\alpha_j = V \times \frac{1}{Z} \times U^T$ and recompute α_j by $\alpha_j = \alpha_j \times f_j$.

4: The residual is updated by the formula as $r_j = f_j - D_{j_0}\alpha_j$, $s.t\ \|\alpha_j\|_0 \le L$.

5: The class of f_j can be determined by the recovered sparse vector $\hat{\alpha}_j$ as $c(f_j) = \arg\max_i \left|D^i\hat{\alpha}_j^i\right|$, $s.t\ \min\left\|f_j - D^i\hat{\alpha}_j^i\right\|_2$, $\forall\, i,\ 1 \le i \le T$.

6: The final classification result $\hat{F}$ for the image F is obtained as $\hat{F} = \left\{f_j | f_j = color_{W(j_0)}, \forall c(f_j) \in W(j_0), 1 \le W(j_0) \le T, 1 \le j \le NM\right\}$.

7: $j = j + 1$

Output: The coloured classification image .

4 Experimental results and analysis

In this paper, we focus on classification using DigitalGlobe's WorldView-2 satellite imagery. The sensor provides the highest resolution commercially available multispectral data and has eight multispectral bands: four standard bands (red, green, blue, and near-infrared 1) and four new bands. Ordered from shorter to longer wavelength, the list of bands is coastal blue, blue, green, yellow, red, red edge, near-infrared 1 (NIR1), and near-infrared 2 (NIR2).

In this section, two data sets are applied for the experiment. We adopt just three bands (Red, Green, and Blue) shown in Fig. 2. We illustrate the effectiveness of the proposed classification method by comparing it with other traditional classifiers, which can be divided into two categories: supervised methods and unsupervised methods according to the previous works of researchers in this field. The first category is supervised method which focuses on learning feature representation and whose training samples with identity labels are required, for example, BPNN, SVM and CART. The second category is unsupervised method, which mainly focuses on feature extraction, such as K-means. The experiments aim to compare the performance of the proposed method with the other four classifiers. Thus, the average accuracy (AA), overall accuracy (OA), Kappa (Ka), producer's accuracy (PA), and user's accuracy (UA)

(a) (b)

Fig. 2 The employed remote sensing images: (**a**) image 1 and (**b**) image 2

are used as the accuracy statistical parameters. For each image, we quantitatively and visually compare and evaluate the classification results of these methods.

4.1 Experiment I for the image 1

The first dataset in our experiments was obtained from DigitalGlobe, which was acquired on 17 May 2010, as shown in Fig. 2a. It contains eight typical classes, including the bare land, residential area, grass, tree, and four different crops, which are labeled as: 1-bare land, 2-residential area, 3-grass, 4-tree, 5-crop1, 6-crop2, 7-crop3, and 8-crop4, respectively. Please refer to Fig. 3a. We randomly select around 11 % samples with ground truth class labels to train the classifiers, and use the rest as testing samples for evaluation. The number of training and testing samples for each class is shown in Table 1.

In order to verify the superiority of our proposed method, we make classification to this image by employing the proposed method, BPNN, SVM, CART and K-means. Figure 3 shows the classification results of the five classifiers. Thereafter, we analyze and compare their experimental results. The classification maps are shown in Fig. 3b-f. It is clear that in the K-means map, shown in Fig. 3f, all kinds of objects are grievously

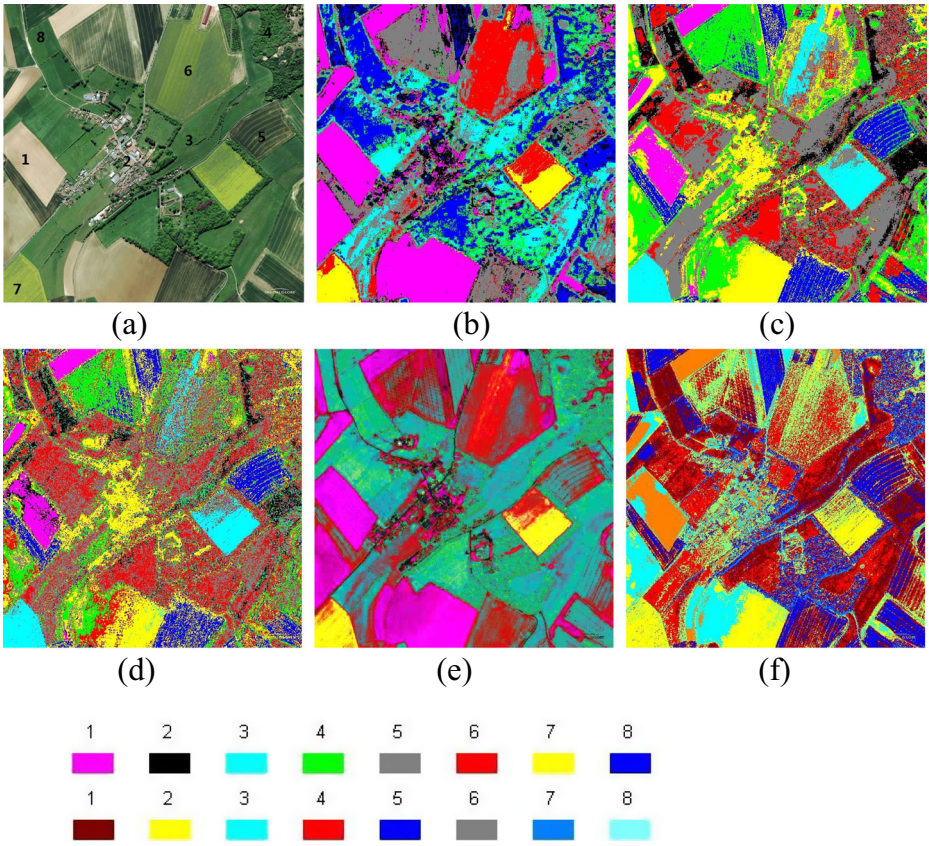


Fig. 3 Classification map for the image 1: (**a**) ground truth, (**b**) the proposed method, (**c**) SVM, (**d**) CART, (**e**) BPNN and (**f**) K-means. (Objects are labeled as: *1*-bare land, *2*-residential area, *3*-grass, *4*-tree, *5*-crop1, *6*-crop2, *7*-crop3, and *8*-crop4. And the upper legend is for Fig. 3b-e, the lower for Fig. 3f.)

Table 1 The training and testing sets for each class (labeled as 1-bare land, 2-residential area, 3-grass, 4-tree, 5-crop1, 6-crop2, 7-crop3, and 8-crop4 in Fig. 3a)

Class		Samples (pixels)	
No	Name	Train	Test
1	Bare land	4,612	39,487
2	Residential area	2,399	21,846
3	Grass	2,986	26,519
4	Tree	2,612	23,353
5	Crop1	5,978	54,604
6	Crop2	2,254	21,088
7	Crop3	905	8,264
8	Crop4	4,436	40,801
Total		26,182	235,962

misclassified; as for the BPNN method, shown in Fig. 3e, the objects illustrated with highlight colors, such as the bare lands, residential areas, and crops2, are very easy to recognize, whereas the green-colored objects, such as the crops1, crops2, crops3, crops4, and especially the grasses and trees, are seriously misclassified; also, many crop4 pixels are wrongly labeled as grasses in the classification map; for the SVM and CART classification result in Fig. 3c, d, it can be clearly seen that there is severe misclassification among these classes. Relatively, through the comparison of our proposed method with the other four methods, our proposed method has achieved a great improvement, namely, the better distinction of objects, particularly the bare lands, residential areas, grasses, trees, crops3, and crops4. Only crops1 and crops2 have been classified with confusions. The result of the proposed method is shown in Fig. 3b.

The classification accuracies for each class using different classifiers are provided in Table 2. In this Table, AA, OA, Ka, PA, and UA are the statistics of the confusion

Table 2 The classification accuracies for different methods in Fig. 2a

Class	Proposed method		BPNN		SVM		CART		K-means	
	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA
1	0.9875	0.9899	0.8775	0.9564	0.4725	1	0.4300	0.8431	0.4900	0.8750
2	0.9150	0.9734	0.8350	1	0.2400	0.2400	0.0600	0.0937	0.4500	0.2571
3	0.8750	0.8413	0.6600	0.3419	0	0	0.0900	0.0643	0.5200	0.5977
4	0.9300	0.9538	0.6900	0.4395	0	0	0	0	0.4500	0.3600
5	0.8750	0.8454	0	0	0	0	0.1100	0.0833	0.4900	0.4712
6	0.8350	0.7840	0.6400	0.4076	0.1300	0.1340	0.0600	0.0380	0	0
7	0.8950	0.9521	0.6800	0.7556	0	0	0.0500	0.0394	0.6800	0.4503
8	0.9100	0.9010	0.2100	0.6364	0	0	0	0	0.7500	0.3521
AA	0.9028		0.5741		0.1053		0.1000		0.4788	
OA	0.9122		0.6717		0.2055		0.1000		0.5127	
Kappa	0.8984		0.6038		0.0785		0.0286		0.3929	

The abbreviations in the table are the average accuracy (AA), overall accuracy (OA), producer's accuracy (PA), user's accuracy (UA)

Table 3 The confusion matrix for the proposed method in Fig. 2a

Confusion Matrix										
Class	1	2	3	4	5	6	7	8	NUM	UA
1	395	4	0	0	0	0	0	0	399	0.9900
2	5	183	0	0	0	0	0	0	188	0.9700
3	0	0	175	6	10	6	0	11	208	0.8400
4	0	0	9	186	0	0	0	0	195	0.9500
5	0	7	0	0	175	18	0	7	207	0.8500
6	0	6	4	0	15	167	21	0	213	0.7800
7	0	0	0	0	0	9	179	0	188	0.9500
8	0	0	12	8	0	0	0	182	202	0.9000
NUM	400	200	200	200	200	200	200	200	1800	
PA	0.9875	0.9150	0.8750	0.9300	0.8750	0.8350	0.8950	0.9100		
OA	0.9122									
Kappa	0.8984									

matrix. Table 3 lists the confusion matrix of the proposed method. AA is the mean of the eight class accuracies. OA is computed as the ratio between the correctly classified testing samples and all the testing samples. Ka coefficient is a quantitative analysis for the classification precision and degree of agreement between the classification map and the ground truth based on the confusion matrix.

Combining the classification maps in Fig. 3b-f with the accuracy statistics in Tables 2 and 3, we can see that according to the ground truth, some pixels of crop2 are wrongly labeled as crop1, while some pixels of crop1 are misclassified as crop2 and residential area in the classification map of the proposed method; crop1 and crop4 cannot be identified in the BPNN map; there are several colors for each object in the SVM, CART and K-means map, for example, some pixels of the bare land are misclassified as tree and crop3 in the SVM map, and some pixels of tree are wrongly labeled as residential area, crop1 and crop2 in the CART map.

From Table 2, we can observe that the proposed method has achieved the highest PA and UA, and performed better in AA, OA, and Kappa coefficient. The best classification results for different objects are achieved by the proposed method. Moreover, we can see that the proposed method performs well in many aspects. The highest accuracies are achieved for the proposed method. The OA values of the proposed method, BPNN, SVM, CART and K-means are 91.22 %, 67.17 %, 20.55 %, 10 % and 51.27 %. The Ka values are 0.8984, 0.6038, 0.0785, 0.0286 and 0.3929. The AA values are 90.28 %, 57.41 %, 10.53 %, 10 % and 47.88 %. Compared with the BPNN, SVM, CART and K-means classifier, the PA values of the proposed method for each class are increased by at least 8 %, 51.5 %, 55.75 %, and 16 % respectively, and the UA values for each class are increased by at least 3.35 %, 65 %, 14.689 %, and 11.49 % respectively. Besides, the PA values of the proposed method for all classes are increased averagely by 32.88 %, 75.75 %, 80.28 %, and 42.41 % respectively, and the UA values for all classes are increased averagely by 33.79 %, 73.34 %, 75.99 %, and 48.47 %. The best classification results of different objects are achieved for the proposed method. Table 3 shows the confusion matrix of Fig. 2a.

4.2 Experiment II for the image 2

In order to verify the stability of the proposed classification method, we select another HSR image shown of WorldView-2 in Fig. 2b. We consider the WorldView-2 true-color image with 1.8-m spatial resolution of a suburban area, which has eight bands. We adopt just three bands (Red, Green, and Blue). This image contains six main kinds of objects: 1-lake, 2-tree, 3-short bush, 4-road, 5-grass, and 6-residential area, as shown in Fig. 4a. The training and testing samples are chosen from the reference data.

We also apply the proposed method, BPNN, SVM, CART and K-means classifier to classify the image 2. The classification maps are shown in Fig. 4b-f. By comparing the classification map in Fig. 4b-f with the original image in Fig. 4a, we can see that some pixels of residential area and grass are misclassified as road in Fig. 4b; some pixels of lake, short bush and residential area are labeled as tree, some pixels of road are labeled as residential area, and some pixels of residential area are labeled as road in Fig. 4c; some pixels of lake and grass are labeled as tree, some pixels of lake and road are labeled as residential area, and some pixels of

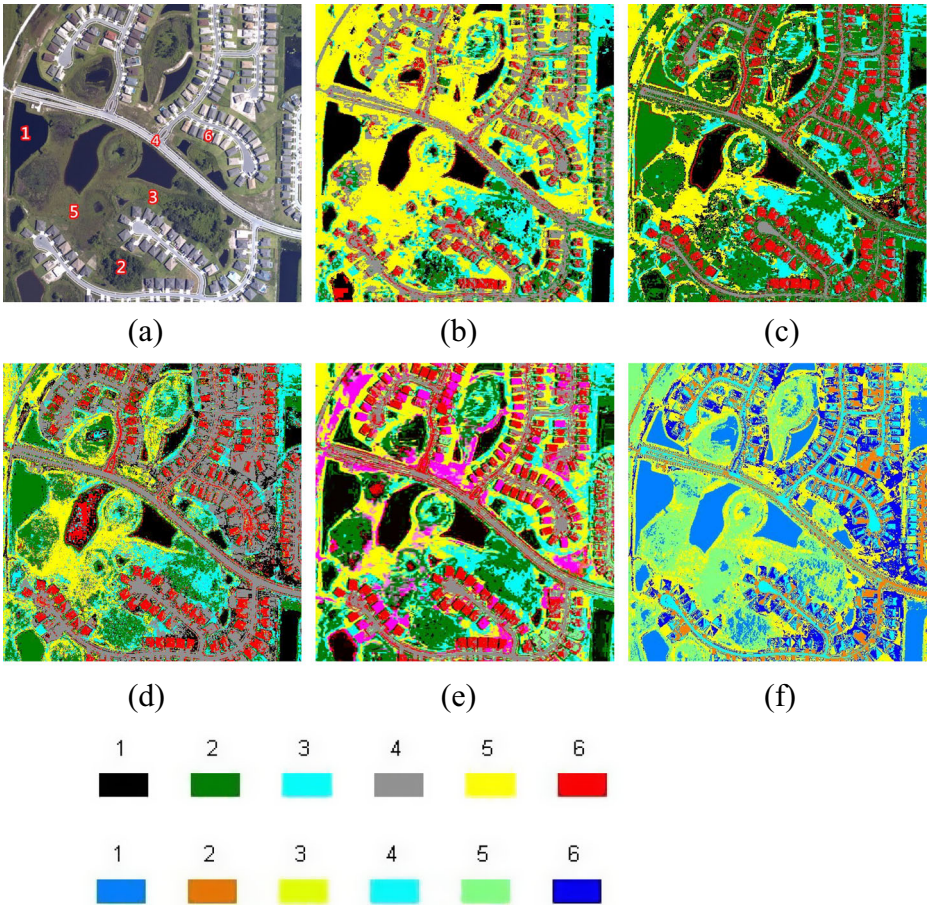


Fig. 4 Classification map for the image 2: (**a**) ground truth, (**b**) proposed method, (**c**) SVM, (**d**) CART, (**e**) BPNN and (**f**) K-means. (Objects are labeled as: *1*-lake, *2*-tree, *3*-short bush, *4*-road, *5*-grass, and *6*-residential area. And the upper legend is for Fig. 4b-e, the lower for Fig. 4f.)

Table 4 The classification accuracies for different methods in Fig. 2b

Class	Proposed method		BPNN		SVM		CART		K-means	
	PA	UA	PA	UA	PA	UA	PA	UA	PA	UA
1	0.9625	0.9722	0.8950	0.9702	0.5100	0.9623	0.4500	1	0.7750	0.9422
2	0.9200	0.9634	0.8150	0.8446	0.8300	0.4049	0.8100	0.4682	0	0
3	0.9300	0.9163	0.8100	0.7397	0.8200	0.5777	0.8150	0.5850	0.815	0.5470
4	0.8550	0.9048	0.7582	0.6571	0.8200	0.8039	0.7250	0.6842	0.725	0.6776
5	0.9450	0.8630	0.9300	0.775	0.4600	1	0.7900	1	0.79	0.6723
6	0.9000	0.8911	0.7650	0.9272	0.6500	0.6566	0.7400	0.5422	0.74	0.6884
AA	0.9188		0.8292		0.6417		0.7217		0.6408	
OA	0.9250		0.8386		0.6571		0.6521		0.6600	
Kappa	0.9082		0.8030		0.5896		0.5853		0.5872	

residential area are labeled as road in Fig. 4d; the obvious error is the misclassification of grass as residential area, some pixels of road are misclassified as residential area, and the lake and bush are both misclassified as tree in Fig. 4e. In Fig. 4f, the tree and residential area are obviously muddled, some pixels of bush are misclassified as the grass, some pixels of grass are misclassified as the lake, and the tree is seriously misclassified as the lake.

From the results in Tables 4 and 5, we can see that the highest accuracies are also achieved for the proposed method. The OA values of the proposed method, BPNN, SVM, CART and K-means are 92.5 %, 83.86 %, 65.71 %, 65.21 % and 66 % respectively. The Ka values are 0.9082, 0.8030, 0.5896, 0.5853 and 0.5872 respectively. The AA values are 91.88 %, 82.89 %, 64.17 %, 72.17 % and 64.08 % respectively. Compared with the BPNN, SVM, CART and K-means classifiers, the PA values of the proposed method for each class are increased by at least 1.5 %, 3.5 %, 11 % and 11.5 %, respectively. Moreover, compared with the BPNN, SVM, CART and K-means, the PA values of the proposed method for all objects are increased averagely by 8.99 %, 23.71 %, 19.71 % and 27.79 % respectively, and the UA values for all classes are increased averagely by 9.95 %, 18.42 %, 20.52 % and 33.06 % respectively. The

Table 5 The confusion matrix for the proposed method in Fig. 2b

Confusion Matrix								
Class	1	2	3	4	5	6	NUM	UA
1	385	2	2	2	0	5	396	0.9722
2	3	184	4	0	0	0	191	0.9634
3	4	2	186	0	11	0	203	0.9163
4	0	5	0	171	0	13	189	0.9048
5	7	3	8	10	189	2	219	0.8630
6	1	4	0	17	0	180	202	0.8911
NUM	400	200	200	200	200	200	1400	
PA	0.9625	0.9200	0.9300	0.8550	0.9450	0.9000		
OA	0.9250							
Kappa	0.9082							

Table 6 Computational cost (CPU time) of the different methods in Fig. 5

	Proposed	CART	SVM	BPNN	K-means
Image1	111.5279	0.2700	106.2532	83.1489	362.3897
Image2	48.0666	0.2237	49.0674	23.5600	298.5017

best classification results of different objects are also obtained for the proposed method. Table 5 shows the confusion matrix of Fig. 2b.

Finally, we compare the different methods in terms of computational cost, which is the CPU time computed by Matlab function and used to evaluate those methods. As can be seen from Table 6 and Fig. 5, the proposed method takes about 112 s in the first dataset and about 48 s in the second dataset to train the dictionary and make a decision; the computational cost of the proposed method is almost the same as SVM, more than the method of BPNN and CART, but far less than K-means.

5 Conclusion and Future Work

In this paper, we tackle the problem of the classification of the HSR remote sensing image using the proposed method based on sparse representation by representing the test samples through a dictionary. The dictionary is obtained by training samples according to their classes for a sparse linear combination. We discuss the specific idea of constructing the dictionary. That is how to compute a best matrix to represent all data vectors by nearest neighbor. We also describe the algorithm used to solve for the sparse representation. We then discuss how to construct the image and how to determine the classes of the test pixels. The experimental results indicate that our method has performed better and achieved higher accuracies in the verified four real remote sensing images.

In the future work, we plan to explore more properties as feature and work toward combining the proposed approach with spectral and spatial features, both at the feature and decision levels, to improve classification accuracy. And we also need to speed up the proposed method. The end goal of this work is to detect yearly and seasonal changes in vegetation cover. Additionally, we also explore how to construct dictionaries to expand to images from the same area in different seasons, and make use of them for change detection.

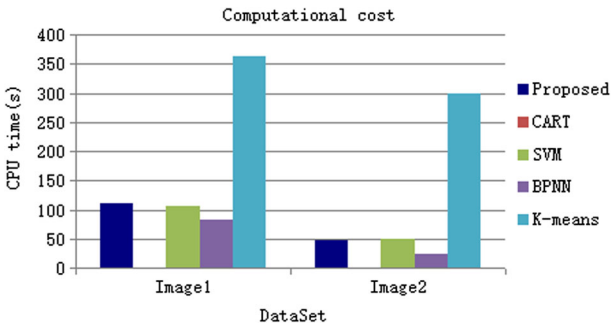


Fig. 5 Computational cost (CPU time) of the different methods

Acknowledgments This work was supported by the National Natural Science Foundation of China (No. 61163042 61503235 and 41272359), and funded by Key Discipline (Cartography and Geographic Information System of Hainan Normal University).

References

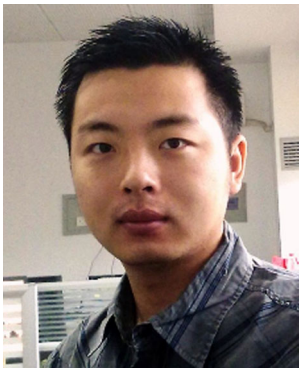
1. Aguera F, Aguilar JF, Aguilar AM (2008) Using texture analysis to improve perpixel classification of very high resolution images for mapping plastic greenhouses. ISPRS J Photogramm Remote Sens 63:635–646
2. Aharon M, Elad M, Bruckstein A (2006) The K-SVD: an algorithm for designing of overcomplete dictionaries for sparse representation. IEEE Trans Sign Process 54:4311–4322
3. Benediktsson JA, Palmason JA, Sveinsson JR (2005) Classification of hyperspectral data from urban areas based on extended morphological profiles. IEEE Trans Geosci Remote Sens 43:480–491
4. Bischof H, Schneider W, Pinz AJ (1992) Multispectral classification of Landsat-images using neural networks. IEEE Trans Geosci Remote Sens 30:482–490
5. Bruzzone L, Carlin L (2006) A multilevel context-based system for classification of very high spatial resolution images. IEEE Trans Geosci Remote Sens 44:2587–2600
6. Bruzzone L, Chi M, Marconcini M (2006) A novel transductive SVM for the semisupervised classification of remote sensing images. IEEE Trans Geosci Remote Sens 44:3363–3373
7. Chen S, Donoho D (1994) Basis pursuit. In: IEEE Conference Record of the Twenty-Eighth Asilomar Conference on Signals, Systems and Computers 1, 41–44
8. Chou PA (1991) Optimal partitioning for classification and regression trees. IEEE Trans Pattern Anal Mach Intell 4:340–354
9. Cotter SF, Rao BD, Engan K, Kreutz-Delgado K (2005) Sparse solutions to linear inverse problems with multiple measurement vectors. IEEE Trans Signal Process 53:2477–2488
10. Donoho DL (2006) Compressed sensing. IEEE Trans Inf Theory 52:1289–1306
11. Dópido I, Li J, Marpu PR, Plaza A (2013) Semisupervised self-learning for hyperspectral image classification. IEEE Trans Geosci Remote Sens 51:4032–4044
12. Elad M, Aharon M (2006) Image denoising via sparse and redundant representations over learned dictionaries. IEEE Trans Image Process 15:3736–3745
13. Fauvel M, Chanussot J, Benediktsson JA (2012) A spatial-spectral kernel-based approach for the classification of remote-sensing images. Pattern Recogn 45:381–392
14. Giacinto G, Roli F (2001) Design of effective neural network ensembles for image classification processes. Image Vis Comput 19:699–707
15. Goel PK, Prasher SO, Patel RM, Landry JA, Bonnell RB, Viau AA (2003) Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn. Comput Electron Agric 39:67–93
16. Heermann PD, Khazenie N (1992) Classification of multispectral remote sensing data using a back-propagation neural network. IEEE Trans Geosci Remote Sens 30:81–88
17. Huang X, Zhang L (2013) An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. Geosci Remote Sens IEEE Trans 51(1): 257–272
18. Huang X, Zhang L, Li P (2007) Classification and extraction of spatial features in urban areas using high-resolution multispectral imagery. IEEE Geosci Remote Sens Lett 4:260–264
19. Inglada J (2007) Automatic recognition of man-made objects in high resolution optical remote sensing images by SVM classification of geometric image features. ISPRS J Photogramm Remote Sens 62:236–248
20. Jiang LH, Wang WS, Yang XR, Xie NF, Cheng YP (2011) Classification methods of remote sensing image based on decision tree technologies. Comput Comput Technol Agric 344:353–358
21. Li J, Zhang H, Zhang L (2015) Efficient superpixel-oriented multi-task joint sparse representation classification for hyperspectral imagery. IEEE Trans Geosci Remote Sens 53(10): 5338–5351
22. Li J, Zhang H, Zhang L, Ma L (2015) Hyperspectral anomaly detection by the use of background joint sparse representation. IEEE J Sel Top Appl Earth Obs Remote Sens 8(6):2523–2533
23. Luo M, Ma YF, Zhang HJ (1992) A spatial constrained k-means approach to image segmentation. In: Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on, 2, 738–742
24. Mallat SG, Zhang Z (1993) Matching pursuits with time-frequency dictionaries. IEEE Trans Signal Process 41:3397–3415

25. Melgani F, Bruzzone L (2004) Classification of hyperspectral remote sensing images with support vector machines. IEEE Trans Geosci Remote Sens 42:1778–1790
26. Moody DI, Brumby SP, Rowland JC, Altmann GL (2014) Land cover classification in multispectral imagery using clustering of sparse approximations over learned feature dictionaries. J Appl Remote Sens 8(1): 084793–084793
27. Moser G, Serpico SB, Benediktsson JA (2013) Land-cover mapping by Markov modeling of spatial-contextual information in veryhigh-resolution remote sensing images. Proc IEEE 101(3):631–651
28. Olshausen BA (1996) Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature 381:607–609
29. Ouma OY, Tateishi R (2008) Urban-trees extraction from QuickBird imagery using multiscale spectex-filtering and non-parametric classification. ISPRS J Photogramm Remote Sens 63:333–351
30. Palmason JA, Benediktsson JA, Sveinsson JR, Chanussot J (2005) Classification of hyperspectral data from urban areas using morphological preprocessing and independent component analysis. In: Proc. 2005 Int. Conf. Geoscience and Remote Sensing Symp. (IGARSS), pp 25–29
31. Paola JD, Schowengerdt RA (1995) A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. IEEE Trans Geosci Remote Sens 33(4): 981–996
32. Pingel JT, Clarke CK, MaBride AW (2013) An improved simple morphological filter for the terrain classification of airborne LIDAR data. ISPRS J Photogramm Remote Sens 77:21–30
33. Rakotomamonjy A (2011) Surveying and comparing simultaneous sparse approximation (or group-lasso) algorithms. Signal Process 91:1505–1526
34. Ray S, Turi RH (1999) Determination of number of clusters in k-means clustering and application in colour image segmentation. In: Proceedings of the 4th international conference on advances in pattern recognition and digital techniques, pp 137–143
35. Reis S, Tasdemir K (2011) Identification of hazelnut fields using spectral and Gabor textural features. ISPRS J Photogramm Remote Sens 66:652–661
36. Rubinstein R, Bruckstein AM, Elad M (2010) Dictionaries for sparse representation modeling. Proc IEEE 98:1045–1057
37. Stoeva S, Nikov A (2000) A fuzzy backpropagation algorithm. Fuzzy Sets Syst 112(1):27–39
38. Tropp J, Gilbert AC (2007) Signal recovery from random measurements via orthogonal matching pursuit. Inf Theory IEEE Trans 53(12):4655–4666
39. Tropp JA, Gilbert AC, Strauss MJ (2006) Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit. Signal Process Spec Issue Sparse Approximations Signal Image Process 86:572–588
40. Tropp JA, Wright SJ (2010) Computational methods for sparse solution of linear inverse problems. Proc IEEE 98:948–958
41. Tuia D, Jordi MM, Kanevski M, Camps G (2011) Structured output SVM for remote sensing image classification. J Sign Process Syst 65:301–310
42. Tuia D, Ratle F, Pozdnoukhov A, Camps-Valls G (2010) Multisource composite kernels for urban-image classification. IEEE Geosci Remote Sens Lett 7:88–92
43. Yang J, Su M, Yu P (2010) A novel K-nearest neighbor classifier based on adaptive metric formed by features extracted by nonparametric feature extraction mode. Int J Adv Inf Technol 4(2):89–103
44. Yu X, Cao T, Yang C, Chen H, Wu S (2009) Remote sensing image classification based on sparse component analysis. Prog Geophys 24:2274–2279
45. Yu XC, Dai S, Hu D, Jiang QY (2011) HHFNN based on lasso function and its application remote sensing image classification. Chin J Geophys 54(6):1672–1678
46. Zhang H, Li J (2014) A nonlocal weighted joint sparse representation classification method for hyperspectral imagery. IEEE J Sel Top Appl Earth Obs Remote Sens 7:2056–2065

Shulei Wu was born in 1974. She is a professor at the College of Information Science and Technology, Hainan Normal University. She received her Bachelor Degree of Radiophysics from Sichuan University. She received her Master Degree of application of computer technology from Chongqing University. Her research interests include image processing, remote sensing image analysis, video retrieval and geo-information system.

Huandong Chen was born in 1968. He is a professor at the College of Information Science and Technology, Hainan Normal University. He received his Bachelor Degree of Mathematics from Hainan Normal University. He received his Master Degree of Communication Engineering from Hainan University. His research interests include image processing, remote sensing image analysis, and education technology.

Yong Bai received his BS degree from Xidian University, China, in 1992, and MS degree from Beijing University of Posts and Telecommunications (BUPT), China, in 1995, and Ph.D. degree from Rutgers-The State University of New Jersey in 2001. He was with PacketVideo Corporation from 2000 to 2002. He was with Motorola from 2002 to 2004. He was with CEC Wireless from 2004 to 2005. He was a senior researcher at DOCOMO Beijing Communication Labs from 2005 to 2009. He is a professor at College of Information Science & Technology, Hainan University since 2010. He acted as the Lead Guest Editor for EURASIP Journal on Wireless Communications and Networking, Special Issue on Topology Control in Wireless Ad Hoc and Sensor Networks. His current research interests include mobile communications, and maritime communications. He is a member of the IEEE.



Guokang Zhu received the joint Ph.D. degree in information and communication engineering from the Chinese Academy of Sciences, Xi'an, China, and Xi'an Jiaotong University, Xi'an, China, in 2014. He is currently a Lecturer with the College of Computer Science and Technology, Shanghai University of Electronic Power, Shanghai, China. His major research interests include computer vision, hyperspectral imagery, and machine learning.