Collaborative Representation and Sparsity are Both Indispensable for Hyperspectral Imagery Classification

Ji-tai A, Shu-cai HUANG, Xiao-fei LI and Yi-dong TANG*
Changle East Road No.1, Xi’an, Shan’xi, China
*Corresponding author

Keywords: Hyperspectral imagery, Classification, Sparse representation, Collaborative representation.

Abstract. As a recently proposed technique, sparse representation based classifier (SRC) has been widely used for hyperspectral imagery classification and detection. The collaborative representation (CR) and the sparse coding are two key points in SRC scheme. More recently, the proposition that which one of them plays a dominant role in SRC scheme has attracted much attention from researchers in fields of image processing, computer vision, and pattern recognition. In this paper, we first discuss why CR or sparsity works and why one of them alone is not sufficient, and then analyze how CR and sparsity interact with each other. Although we focus on how sparsity augments CR, the necessity of CR for sparsity is also illustrated in both pixel-wise model and joint sparsity model. Inspired by the analysis, we indicate that CR is a powerful tool for solving the high-dimensional pattern recognition with small sample in SRC scheme; sparsity augments CR-based classification in stabilizing, making sure unique solution and rejecting outlying samples. In other words, CR and sparsity complement each other and are both indispensable for hyperspectral imagery classification. The experimental results on simulated data and real hyperspectral imagery confirm the conclusion.

Introduction

HYPERSPECTRAL imagery provides a wealthy of spectral information, which spans the visible to infrared spectrum. Different materials usually reflect electromagnetic energy differently at specific wavelengths. This makes it possible to uniquely identify various materials based on their spectral signatures. One of the most important applications of HSI is image classification, where pixels are labeled to one of the classes based on their spectral characteristics and training samples for each class. Various techniques have been developed for HSI classification. Comparing with previous approaches, the support vector machine (SVM) [1] [2] [3] has been proved to be a powerful tool to solve supervised classification problem in remote sensing and performs well. And variations of SVM based algorithms have also been proposed to improve its performance [4] [5].

Recently, a simple approach called sparse representation [6] [7] [8] based classification (SRC) has been proposed to solve many computer vision tasks [9]–[13]. Sparse representation has also shown quite impressive results on HSI detection and classification [14]. It was originally a common belief that sparsity of the representation is the key to success for this classification scheme. However, more recently, researchers have started arguing the role of sparsity in SRC scheme. Some researchers [15] claim that it is the collaborative representation (CR), but not the sparsity, that makes the scheme effective. While some researchers believe that the sparsity plays an explicit role in accurate CR-based classification, hence it should not be completely ignored for anything else. Some propose an analytic scoring function which depends only on the given feature vector and the size of the data set to predict whether the collaborative representation model should be sparse or non-sparse. A novel algorithm that combines sparsity and dense collaborative representation is proposed for target detection in hyperspectral imagery. Essentially, the main divergence they argue is if it is necessary to involve the sparsity constraint in CR-based classification scheme, and which one plays a key role for classification. However, should it really be the bone of contention for us? Considering they all admit the importance of CR, what we should concern about is actually how CR and sparsity promote each other in SRC scheme.
In a sense, the SRC can be treated as a CR-based classifier (CRC) with sparse constraint. For a better understanding and comparison, we define the general SRC as sparse CRC (SCRC) and the CRC without sparse constraint as dense CRC (DCRC). The main idea of this paper is the roles that CR and sparsity play in CR-based classifier but not algorithm designing. We first extend the analysis in [15], and discuss why CR or sparsity works and why one of them alone is not sufficient. Then, we analyze the interactions between CR and sparsity, and indicate that CR and sparsity complement each other in HSI classification, and neither can work effectively without each other. On one hand, sparsity improves the discriminative power of representation error when too many samples are used or subjects look similar with each other, leading to a robust classification. On the other hand, sparse representation codes a test pixel over an over-completed dictionary. However, HSI classification is a typical high dimensional problem with small sample size, leading to an under-completed dictionary. One of the tools for solving the “lack of samples” problem is CR that takes the samples from all the other classes as the possible samples of each class. In turn again, the CR coding may produce same class-specific residuals for different classes, leading to wrong classification in labeling process. In this case, sparsity can be used to make sure the uniqueness and right classification. Although we focus on how sparsity augments CR in stabilizing, making sure unique solution and rejecting outlying samples, the necessity of CR for sparsity is also illustrated in both pixel-wise model and joint sparsity model. Finally, we evaluate SCRC and DCRC as well as their simultaneous joint visions on a real HSI data set to verify our conclusion.

The rest of the paper is organized as follows. Section II briefly reviews general SRC. Section III analyzes how CR and sparsity interact and augment with each other. Experimental results and discussion on data set are provided in Section IV. Finally, we conclude the paper in Section V.

Sparse Representation Based Classifier

In general SRC scheme, it is assumed that the spectral signatures of pixels belonging to same class approximately lie in the same low-dimensional subspace. Denote sub-dictionary by the data set of training samples of class $i$ $X_i \in \mathbb{R}^{B \times N_i}$, where $B$ is the number of spectral bands. Suppose that we have $K$ classes of subjects, hence the structured dictionary $X = \{X_1, X_2, \ldots, X_K\} \in \mathbb{R}^{B \times N}$, $N = \sum_{i=1}^{K} N_i$ is formed from the sub-dictionaries. A given query pixel $y \in \mathbb{R}^B$ can be coded over $X$ with sparsity constraint.

$$\hat{\alpha} = \arg\min_{\alpha} \|y - X\alpha\|_2 \quad \text{s.t.} \|\alpha\|_0 \leq K_0$$

where $\alpha = [\alpha_1; \alpha_2; \ldots; \alpha_K] \in \mathbb{R}^{K \times n}$ and $\alpha$ is the coding vector associated with class $i$. The $\ell_0$-norm regularized minimization problem (1) can be solved by OMP algorithm. Once the coding vector is obtained, the class label of $y$ can be determined by the minimum reconstruction residual associated with each class wise sub-dictionary.

$$\text{Class}(y) = \min_{i \in \{1,2,\ldots,K\}} \|y - X_i\hat{\alpha}\|_2$$

where $i \in \{1,2,\ldots,K\}$ is the class index.

Collaborative Representation and Sparsity Constraint

As mentioned in the previous section, CR and sparsity constraint are two key points in the coding process of SRC. In this section, concerning how CR and sparsity interact with each other in SRC scheme, we focus on how sparsity augments CR, and also illustrate the necessity of CR for sparsity.

Sparsity is Necessary for Robustness

In practice, if the number of training samples of each class is relatively large, we can represent the query pixel class by class. As shown in [15] when subjects look similar to each other (a common
fact in HIS scene), sub-dictionary $X_i$ and sub-dictionary $X_j$ are not incoherent but can be highly correlated i.e. $X_j = X_i + \Delta$. Without regularization term, the coding vectors of $y$ by class $i$ and class $j$, $a_i$ and $a_j$, can be calculated by least square method. Let $e_i = y - X_i a_i$ and $e_j = y - X_j a_j$. Suppose that $X_i, X_j \in \mathbb{R}^{m \times n}$, and $\Delta$ is small such that

$$
\xi = \frac{\|e_i\|}{\|y\|} \leq \frac{\sigma_i(X_i)}{\sigma_j(X_j)}
$$

(3)

where $\sigma_i(X_i)$ and $\sigma_j(X_j)$ are respectively the largest and smallest eigenvalues of $X_i$. Then, we can get the relationship between $e_i$ and $e_j$ as follows

$$
\frac{\|e_i - e_j\|}{\|y\|} \leq \xi (1 + \kappa_2(X_i)) \min\{1, m - n\} + O(\xi^2)
$$

(4)

where $\kappa_2(X_i)$ is the $\ell_2$-norm conditional number of $X_i$. We can see from (4) that if $\Delta$ is small, the distance between $e_i$ and $e_j$ can be very small. This makes the classification so unstable that even a small disturbance can result in a wrong classification. Furthermore, with enough training samples for each class, we can use $X_j$ to represent $y$ well even when $y$ doesn’t belong to class $j$. In other words, the discriminative power of the representation error will reduce if too many samples are used or subjects look similar with each other. To illustrate the observation, we respectively use training samples from Corn-notill and Corn-min in Indian Pines image data set to represent a given Corn-notill sample, and use samples from Soybeans-min and Soybeans-clean to represent a given Soybeans-min training sample under different sparsity levels. The curves of representation error versus the sparsity level are drawn in Fig. 1. We can see From Fig. 1 that when the sparsity level is low, both of the right class and wrong class have relatively big error. And the representation error decreases with the sparsity level and tends to saturate as sparsity level reaches a certain value. However, the discriminative power of the representation error reduces if sparsity is too high.

![Figure 1](image-url)

Figure 1. The curve of representation error versus the sparsity level.

As a result, the robustness of classification can be enhanced by imposing some sparsity constraint on coding vector. If $y$ belongs to class $i$, it is more likely that only a few samples in $X_i$ can represent $y$ with a good accuracy, while more samples in $X_j$ are required to represent $y$ with nearly same accuracy. Hence, under a certain sparsity, the representation error of $y$ by $X_i$ will be obviously lower than that by $X_j$, leading to an enhanced discriminative power of representation error. This is exactly the fundamental mechanism in which SRC works.

**Sparsity Needs CR**

Although sparsity is the foundation of SRC, if we want to use fewer atoms to represent $y$, we must relax the orthogonality imposed on dictionary. In other words, we must allow more atoms (another method is dimension reduction) to be involved in dictionary so that we have more choices to represent query pixel, leading to an over-complete dictionary and a sparser representation of query.
Unfortunately, HSI classification is a typical high dimensional problem with small sample size, generally leading to an under-completed sub-dictionary $X_i$. If we use the under-complete $X_i$ to represent $y_i$, the representation error can be large, even when $y_i$ is from class $i$. In SRC, this “lack of sample” problem is solved by taking the training samples from all the other classes as the possible samples of each class, i.e. collaborative representation. The reason is that samples of different classes share similarities; hence some samples from class $j$ may be very helpful to represent the query pixel from class $i$. To understand the working mechanism of CR-based classification, we define the matrix $X = [X_1,X_2,...,X_k]$ as the dictionary of all samples, and write the subspace spanned by the columns of $X$ as $\mathcal{Y}$. In SCRC scheme, it codes the query pixel $y$ over the dictionary $X$ with $\ell_0$-norm constraint, and then identifies $y$ individually. When we ignore the $\ell_0$-norm constraint in (1), the representation becomes a least square problem $\hat{a} = \arg \min \|y - Xa\|^2_F$. Then the representation $\hat{y} = Xa$ is actually the perpendicular projection of $y$ onto $\mathcal{Y}$, and Fig. 2 shows geometrically illustration of the representation.

As shown in Fig. 2, without loss of generality, we can decompose $\hat{y}$ into two components

$$\hat{y} = \hat{z} + \hat{z}_i$$ \hspace{1cm} (5)

where $\hat{z}_i = X_i\hat{a}_i$ is the representation associate with class $i$, and $\hat{z}_i = \sum_{j \neq i} X_j\hat{a}_j$ is the sum of representation associate with all the other classes except class $i$. The representation error of each class $e_i = \|y - X_i\hat{a}_i\|^2$ is used for classification. It can be derived that

$$e_i = \|y\|^2 + \|y - X_i\hat{a}_i\|^2 - 2\hat{y}^T X_i\hat{a}_i$$ \hspace{1cm} (6)

Actually, it is the term $e_i = \|y - X_i\hat{a}_i\|^2$ that works for classification because $\|y - \hat{y}\|^2$ is a constant for all classes. Based on sine theorem, we can readily have

$$\frac{\|\hat{y}\|^2}{\sin (\gamma)} = \frac{\|\hat{y} - X_i\hat{a}_i\|^2}{\sin (\beta)}$$ \hspace{1cm} (7)

where $\beta$ is the angle between $\hat{y}$ and $X_i$, and $\gamma$ is the angle between $X_i$ and $\hat{z}_i$. Finally the representation error can be represented as

$$e_i = \|y - X_i\hat{a}_i\|^2 = \frac{\sin^2 (\beta)\|y\|^2}{\sin^2 (\gamma)}$$ \hspace{1cm} (8)

In (8), the ratio $\sin^2 (\beta)/\sin^2 (\gamma)$ not only considers if the angle $\beta$ is small (leading to small intra-class distance) but also considers if the angle $\gamma$ is big (leading to big inter-class distance). Zhang et al. [1] noted it is the “double checking” that makes the classification more effective and robust, and the $\ell_0$-norm constraint is just used for stabilize the least square solution. Not only that,
they also believe that it is not necessary to use this strong \(\ell_0\)-norm whose effects can be replaced by \(\ell_2\)-norm. The objective of dense collaborative representation is to allow all the training samples to participate in the representation of \(y\). As a result, the objective function of collaborative representation with \(\ell_2\)-norm regularization is formulated as

\[
\hat{a} = \arg\min_a \|y - Xa\|^2 + \lambda \|a\|^2
\]

where \(\lambda\) is a regularization parameter. Taking derivative with regard to \(a\) and setting the resultant equation to zero yields

\[
\hat{a} = (X^T X + \lambda I)^{-1} X^T y
\]

Once the coding vector of the representation is obtained, the label of \(y\) can be determined in the same way as SRC in (2).

Why CR Needs Sparsity in Turn

According to Zhang et al. [1], the ratio \(\sin^2(\beta)/\sin^2(\gamma)\) is powerful for classification. However, for \(\beta, \gamma \in [0, 2\pi]\), there is no unique minimum for the given squared ratio in (8). As shown in Fig. 3(a), any vector starting from \(o\) and ending at a point on broken line circle (e.g. \(p, q\)), i.e., representation associate with other class, will have the same \(e\) as \(z\). This will finally result in a wrong classification. This fact is obviously ignored by Zhang et al [1]. That is, the sparsity constraint is not only used to stabilize the least square solution, but to make sure the uniqueness of the minimum for the given squared ratio in (8). As shown in Fig. 3(b), \(z_i\) is composed of \(x_i^1, x_i^2, x_i^3\), and \(z_j\) is composed of \(x_j^1, x_j^2, x_j^3\), atoms from \(X_i\) and \(X_j\). Although the class-specific residuals are equal, we can still label the query pixel as class \(i\) since \(z_i\) requires lesser number of atoms to produce the same class-specific residual. Hence, the performance of CR-based classification is augmented by sparsity constraint.

![Figure 3. Geometric illustration of the working mechanism of CR-based classification.](image)

In the previous subsection, the number of the components, i.e. sparsity level, is used for pixel-wise classification when the class-specific residuals have equal length. How can sparsity make sure the uniqueness when class-specific residuals and sparsity level for \(X_i\) and \(X_j\) are both equal? Although this rarely happens to high dimensional problem, the joint sparsity which exploits interpixel correlation serves this purpose. In a typical HSI, the neighboring pixels usually consist of similar materials, and thus, their spectral characteristics are highly correlated. Given two neighboring pixels centered at the test pixel, \(y_i\) and \(y_j\). In joint sparsity model (JSM), they are coded by the structured dictionary simultaneously, and it is assumed that the underlying sparse vectors associated with them share a common sparsity support. Geometric illustration of the jointly sparse representation of \(y_i\) and \(y_j\) is shown in Fig. 4. The representations of \(y_i\) and \(y_j\) associate with class \(i\) and class \(j\) are shown with real line and broken line respectively. In Fig. 4, two class-specific representations for \(y_i\), i.e. vectors \(e_i\) and \(e_j\), not only have equal error lengths but also have same
number of components. Consequently, the classification will be inaccurate no matter the error or the sparsity or both of them are used for decision making. However, since the sparse vectors associated with \( y_1 \) and \( y_2 \) share a common sparsity support, the components used to represent \( y_1 \), i.e. \( x_i^1 \) paired with \( x_i^2 \) and \( x_i^1 \) paired with \( x_i^2 \), also must be used to represent \( y_2 \), just with different weight from \( y_1 \). In this case, although it is not possible to use the representation of \( y_1 \) for classification, the class-specific residuals for \( y_2 \) (\( \varepsilon_i \) and \( \varepsilon_j \)) are most likely no longer equal and the fact can be used for classification. Intuitively, \( i \) (not \( j \)) represents the correct class of the test pixel since components from class \( i \) produce a smaller class-specific error for \( y_2 \). Hence, cooperating with spatial correlation, the sparsity constraint further makes sure the uniqueness and improves the classification performance. The DCRC and SCRC in JSM are respectively denoted as simultaneous DCRC (SDCRC) and simultaneous SCRC (SSCRC).

Figure 4. Geometric illustration of the working mechanism of jointly sparse representation.

Considering the spatial correlation, the neighboring pixels are simultaneously represented in collaborative representation. Without sparsity constraint (i.e., dense CR), the solution for \( \min_{\alpha} \|Y - X\alpha\|_2 + \lambda \|\alpha\|_1 \) is essentially computed column by column. In this case, it just focuses on jointly calculating the total residuals from all the neighboring pixels, and then uses the total residuals for labeling. In other words, dense CR in JSM wastes the spatial correlation information during coding step. Different from dense CR, the sparse CR takes use of spatial correlation during both coding step and labeling step. Hence, the sparsity constraint augments joint collaborative representation on the utilization of spatial information.

In addition, it is almost impossible for us to obtain labeled samples of every class for training in hyperspectral imagery scene. In other words, we may have no training sample allied to a given query pixel at all. This makes that no matter how we classify the query pixel is wrong. Hence, before classifying a given query pixel, we must first decide if it is a valid sample from one of the classes in the HSI data set. The ability to detect and then reject invalid test pixel is important for hyperspectral classification task. Conventional classifiers such as nearest neighbor (NN) and nearest space (NS) usually use the representation error for validation. However, with an over-complete dictionary, the smallest representation error of the invalid test pixel is not so large, even an invalid pixel will has as small representation error as valid pixels, leading to inaccurate validation. Since the coefficients are computed globally in sparse representation scheme, the distribution of the coefficients contains important information about the validity of the query pixel. In detail, a valid pixel has sparse coding vector whose nonzero entries concentrate on one subject while an invalid pixel has sparse coefficients spread widely across the entire training set. However, in dense CR, since the samples from different classes share similarity and all the training samples participate in the representation, both of the coefficients of valid and invalid samples would spread widely across the entire training set. This makes the validation ability of coefficients weak. Therefore, we note that the sparsity constraint contributes to validation as well. What calls for special attention is that the coefficients computing must be interpreted as \( \ell_0 \)-norm minimization within a certain representation error since the validation is based on the distribution of the coefficients. As a result, when we need to use the coefficients for validation, the problem (1) is revised as follows
\[ \hat{\alpha} = \arg \min_{\alpha} \| \alpha \|_0 \quad \text{s.t.} \quad \| y - X \alpha \|_2 \leq \varepsilon \]  

(11)

where \( \varepsilon \) is the error tolerance. The above problem can also be solved by OMP algorithm.

To support our argument, we choose 5 classes of subjects from AVIRIS San Diego, CA, USA as valid, and randomly select 81 samples per class for training. A synthetic data set, 64×64 pixels in size, is constructed of an invalid (denote by Class 0) and the 5 classes of valid samples (denote by Class 1, 2, 3, 4, and 5). Considering the spectral mixture and noise, the spectral signatures are mixed with a background spectrum with abundance of 0.6, and Gaussian noise is then added to achieve a 30:1 signal-to-noise ratio. Then a valid and an invalid pixel are respectively represented by sparse CR (SCR) and dense CR (DCR) under same error level. Fig. 5 shows the distribution of coefficients. We see from Fig. 5 (a) and (b) that, the coefficients of valid pixel by SCR concentrate on first 81 samples (9 out of 21) while the coefficients of invalid pixel nearly spread uniformly across the entire training set (4, 6, 3, 4 and 5 nonzero entries drop into corresponding subjects respectively). Thus, we can distinguish valid pixel from invalid pixel with the distribution of coefficients. However, we see from Fig. 5 (c) and (d) that, the coefficients of valid and invalid samples by DCR both spread widely across the 405 training samples. In this case, the distribution of coefficients cannot be used for query pixel validation in DCR scheme.

![Figure 5. Distribution of coefficients of (a) valid pixel by SCR, (b) invalid pixel by SCR, (c) valid pixel by DCR, (d) invalid pixel by DCR.](image)

**Experiments and Results**

The data set in this paper is the Indian Pines image, and we download the MATLAB data files for experiments (The files can be found on the Group de Intelligence Computational at http://www.ehu.eus/ccwintco/uploads/2/22/Indian_pines.mat.). It was acquired by Airborne Visible/Infrared Imaging Spectrometer (AVIRIS), which generates 220 bands across the spectral range 0.2 to 2.4 \( \mu m \). 20 noisy bands are removed due to water absorption bands before classification. This image is 145×145 pixels in size with a spatial resolution of 20 m/pixel, and contains 16 ground-truth classes. For each class, we randomly choose around 10% of the labeled pixels for constructing the dictionary and use the remaining 90% pixels for testing, as shown in Table I and Fig. 6. In our experiments, overall accuracy (OA) and average accuracy (AA) are used to evaluate the classification performance. The OA is the ratio between correctly classified test samples and the total number of test samples, the AA is the mean of the 16 class accuracies.

In this sub-section, we removed the invalid pixels before classification as some related works do. Firstly, we examine how parameters affect the classification performance on the Indian Pines image for parameters optimization. Then, we compare the classification performance of classifiers with their own optimized parameters.
Table 1. Indian Pines Ground-Truth Classes and Train/Test Sets.

| No. | Name                      | Samples  |       |
|-----|---------------------------|----------|-------|
| 1   | Alfalfa                   | 6        | 40    |
| 2   | Corn-notill               | 137      | 1291  |
| 3   | Corn-min                  | 80       | 750   |
| 4   | Corn                      | 23       | 214   |
| 5   | Grass/Pasture             | 48       | 435   |
| 6   | Grass/Trees               | 72       | 658   |
| 7   | Grass/Pasture-mowed       | 3        | 25    |
| 8   | Hay-windrowed             | 47       | 431   |
| 9   | Oats                      | 2        | 18    |
| 10  | Soybeans-notill           | 93       | 879   |
| 11  | Soybeans-min              | 235      | 2220  |
| 12  | Soybean-clean             | 59       | 534   |
| 13  | Wheat                     | 21       | 184   |
| 14  | Woods                     | 124      | 1141  |
| 15  | Building-Grass-Trees-Drives | 37   | 349   |
| 16  | Stone-steel Towers        | 10       | 83    |
|     | Total                     | 997      | 9252  |

Table 2. Indian Pines: OA with varying parameters.

Parameter Analysis: For optimal parameter selection, we compute the OA of DCRC with varying regularization parameter $\lambda$ ranging from $10^{-7}$ to 0.1, as well as the OA of SCRC with varying sparsity level $\kappa$ ranging from 1 to 40, as listed in Table 2 and Table 3. One can see from Table 2 and Table 3 that for the pixel-wise CR model, $\lambda = 10^{-4}$ leads to the greatest OA for DCRC, and $\kappa = 2$ leads to the greatest OA for SCRC. Similarly, we also compute the OA of SDCRC with varying neighbor size $\omega$ and $\lambda$, as well as the OA of the SSCRC with varying neighbor size $\omega$ and sparsity level $\kappa$, as shown in Fig. 7. One can see from Fig. 7 (a) that $\lambda = 10^{-4}$ leads to the greatest OA at all neighbor sizes for SDCRC in the JSM. For a fixed $\lambda$, the performance of SDCRC changes a little with neighbor size. For SSCRC, as shown in Fig. 7 (b), $\omega = 2$ leads to the greatest OA at all sparsity levels. For a fixed $\omega$, the performance of SSCRC generally has two local highest points and tends to saturate as sparsity level $\kappa$ reaches 140–200. As a result, we fix $\lambda = 10^{-4}$ and $\omega = 3$ for SDCRC, and fix $\kappa = 180$ and $\omega = 3$ for SSCRC.
Table 2. Overall Accuracy of Dense CRC with Varying $\lambda$ for the Indian Pines Data.

| $\lambda$ | $10^{-7}$ | $10^{-6}$ | $10^{-5}$ | $10^{-4}$ | $10^{-3}$ | $10^{-2}$ | 0.1 |
|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----|
| OA        | 1.16      | 69.68     | 69.45     | 67.54     | 62.75     | 57.89     | 54.44 |

Table 3. Overall Accuracy of Sparse CRC with Varying $\kappa$ for the Indian Pines Data.

| $\kappa$ | 1  | 2  | 4  | 8  | 12 | 16 | 20 | 40 | 50 | 60 | 180 |
|----------|----|----|----|----|----|----|----|----|----|----|-----|
| OA       | 71.80 | 71.80 | 71.40 | 67.30 | 66.97 | 66.93 | 66.95 | 66.72 | 66.53 | 66.35 | 65.05 |

Figure 7. Parameter analysis. OA values of (a) SDCRC at different $\omega$ and $\lambda$, (b) SSCRC at different $\omega$ and $\kappa$.

Classification performance: With optimized parameters, Table 4 shows classification accuracy for each single class, OA, and AA. Fig. 8 shows the ground-truth and classification maps. Fig. 8(b) and (c) respectively show the classification maps of DCRC and SCRC in pixel-wise model, while Fig. 8(d) and (e) respectively show the classification maps of SDCRC and SSCRC in JSM. We can clearly see that sparsity constraint has significantly improved the classification performance whether or not the spatial correlation is incorporated. The SSCRC outperforms the other three classifiers, with greatest OA and AA. It is sparsity constraint that further improves the classification accuracy in cooperation with spatial correlation, as mentioned in section III.

Table 4. Classification Accuracy (%) for AVIRIS Indian Pines.

| Class | DCRC | SCRC | SDCRC | SSCRC |
|-------|------|------|-------|-------|
| 1     | 4.35 | 41.30| 0     | 78.26 |
| 2     | 67.16| 52.80| 77.87 | 75.84 |
| 3     | 44.82| 60.48| 60.84 | 58.19 |
| 4     | 12.24| 46.84| 5.91  | 64.56 |
| 5     | 66.67| 82.40| 79.92 | 91.10 |
| 6     | 93.56| 92.74| 99.73 | 99.45 |
| 7     | 0    | 92.86| 0     | 92.86 |
| 8     | 97.28| 94.56| 100   | 100   |
| 9     | 0    | 40.00| 0     | 10    |
| 10    | 43.00| 68.52| 41.15 | 57.00 |
| 11    | 79.06| 73.40| 96.99 | 89.53 |
| 12    | 48.23| 50.93| 69.81 | 78.08 |
| 13    | 97.56| 94.63| 98.54 | 97.56 |
| 14    | 98.42| 94.39| 100   | 99.76 |
| 15    | 37.05| 43.52| 47.67 | 71.76 |
| 16    | 82.80| 92.47| 100   | 100   |
| OA    | 69.68| 71.80| 79.64 | 82.68 |
| AA    | 54.51| 70.12| 61.15 | 79.00 |
Conclusion
The main idea of this paper is the intertwined roles that CR and sparsity play in CR-based classifier but not the algorithms design. And in contrast to the existing notion about which one is more important, this paper tried to reveal that the CR and sparsity are both indispensable for classification. Both of them play an explicit role in CR-based classification for hyperspectral imagery. No matter in pixel-wise model or JSM, the CR is crucial for sparse representation when lacking training sample, while the sparsity augments CR-based classification in several ways. The experimental results clearly demonstrate the conclusion.

Acknowledgement
This work was supported by the National Natural Science Foundation of China under Grant 61273275 and Aeronautical Science Foundation of China under Grant 20130196004.

References
[1] F. Melgani and L. Bruzzone, “Classification of hyperspectral remote sensing images with support vector machines”, IEEE Trans. Geosci. Remote Sens., Vol. 42, No. 8, pp. 1778–1790, 2004.
[2] B. E. Boser, I. M. Guyon, and V. N. Vapnik, “A training algorithm for optimal margin classifiers”, in Proc. 5th Annu. Workshop Comput. Learn. Theory, 1992, pp. 144–152.
[3] J. A. Gualtieri and R. F. Cromp, “Support vector machines for hyperspectral remote sensing classification”, in Proc. SPIE 27th AIPR Workshop: Adv. Comput.-Assist. Recognit., Washington, DC, Oct. 1998, Vol. 3584, pp. 221–232.
[4] G. Camps-Valls and L. Bruzzone, “Kernel-based methods for hyperspectral image classification”, IEEE Trans. Geosci. Remote Sens., Vol. 43, No. 6, pp. 1351–1362, 2005.
[5] Bruzzone, M. Chi, and M. Marconcini, “A novel transductive SVM for the semi-supervised classification of remote sensing images”, IEEE Trans. Geosci. Remote Sens., Vol. 44, No. 11, pp. 3363–3373, 2006.
[6] J. Wright, A. Y. Yang, A. Ganesh, S. Sastry, and Y. Ma, “Robust face recognition via sparse representation”, IEEE Trans. Pattern Anal. Mach. Intell, Vol. 31, No. 2, pp. 210–227, 2009.
[7] A. M. Bruckstein, D. L. Donoho, and M. Elad, “From sparse solutions of systems of equations to sparse modeling of signals and images,” SIAM Rev., Vol. 51, No. 1, pp. 34–81, 2009.
[8] M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing. New York: Springer-Verlag, 2010.
[9] M. Elad, M. A. T. Figueiredo, and Y. Ma, “On the role of sparse and redundant representations in image processing,” Proc. IEEE, Vol. 98, No. 6, pp. 972–982, Jun. 2010.
[10] J. Wright, A. Y. Yang, A. Ganesh, S. Sastry, and Y. Ma, “Robust face recognition via sparse representation,” IEEE Trans. Pattern Anal. Mach. Intell. Vol. 31, No. 2, pp. 210–227, Feb. 2009.
[11] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. Huang, and S. Yan, “Sparse representation for computer vision and pattern recognition,” Proc. IEEE, Vol. 98, No. 6, pp. 1031–1044, Jun. 2010.

[12] A. M. Bruckstein, D. L. Donoho, and M. Elad, “From sparse solutions of systems of equations to sparse modeling of signals and images,” SIAM Rev., Vol. 51, No. 1, pp. 34–81, 2009.

[13] Y. Xu, D. Zhang, J. Yang, and J.-Y. Yang, “A two-phase test sample sparse representation method for use with face recognition,” IEEE Trans. Circuits Syst. Video Technol., Vol. 21, No. 9, pp. 1255–1262, Sep. 2011.

[14] Zongze Yuan, Hao Sun, Kefeng Ji, et al, “Local sparsity divergence for hyperspectral anomaly detection,” IEEE Geosci. Remote Sens Lett, Vol. 11, No. 10, pp. 1697–1701, 2014.

[15] Y. Chen, N. M. Nasrabadi, T. D. Tran, “Hyperspectral image classification via kernel sparse representation,” IEEE Trans. Geosci. Remote Sens., Vol. 51, No. 1, pp. 217–231, 2013.