ATLANTIS PRESS

Advances in Computer Science Research, volume 62

2nd Joint International Information Technology, Mechanical and Electronic Engineering Conference (JIMEC 2017)

Hyperspectral Image Classification Using Compressive Sampling Measurements

Chen Xinmeng, Li Yuting, Liu Jiying, Zhu Jubo

College of Science, National University of Defence Technology

Keywords: **Compressive Sensing, Classification, Hyperspectral Image, OMP.**

Abstract

In this paper, we develop a new approach for hyperspectral image classification directly from the compressive sensing measurements without reconstructing the original hyperspectral image first. The proposed method is based on the fact that each pixel in the hyperspectral image lies in a low-dimensional subspace, and thus it can be represented as a sparse linear combination of vectors in a dictionary obtained from training samples. In compressive sensing theory, with the sparsity prior, we can reconstruct the original signal from the random sampling measurements using appropriate algorithms. And finally the recovered sparse vector is used to determine the class label of the test pixel by the nearest neighbor classifier. The proposed method can fulfil the classification task and reconstruction at the same time.

I. Introduction

Hyperspectral imaging (HSI) acquires data spanning hundreds of contiguous wavelength bands in a certain spectral range. Pixels in hyperspectral images are therefore represented by vectors whose entries correspond to the intensity at different spectral bands. Since different materials reflect electromagnetic energy differently at specific wavelengths, it can be used to identify different materials. Hyperspectral images (HSI) are widely used in many areas, such as military surveillance [1], environmental monitoring and mineral detection and exploration [2]. On the other hand, HSI suffers from huge amounts of data. In this regard, we use the compressive sensing (CS) [3], [4], [5] technique to compress the data while sampling. CS theory is a new approach to reconstruct a signal from sub-Nyquist random linear measurements with a designed algorithm during last decade. The theory of CS shows that, when the signal is sparse enough, it can be recovered from its compressive measurements and the key observation of HSI data is that despite its huge size, there exist massive correlations both spatially and spectrally, thus exploiting them properly will lead sparse signal.

Hyperspectral image classification is to categorize each spectral pixel belonging to one of the classes based on the spectral information [6]. Many methods have been developed to achieve this objective, including sparse representation based classification method [7]. Therefore, we could do compressive sampling once we have the sparse information of the signal. In this paper, we develop a classification approach that labels each pixel in the image directly from the compressive measurements, without resorting to first reconstruct the image cube. The proposed algorithm is based on a sparse model where a test pixel can be represented as a sparse linear combination of vectors in a dictionary which is obtained from the training samples. The class of the test pixel is then determined by the characteristics of the recovered sparse vector. Thus we can reconstruct and classify the hyperspectral datacube at the same time, saving the computational time and enhance the accuracy compared to classifing after recovering first.

The application of CS to hyperspectral images is an active area of research. [8], [9], [10] introduce methods with similarities to the one described in this work, combining the unmixing task with compressive sensing. The authors describe a new method to unmix the original hyperspectral datacube directly from the compressive sensing measurements, without reconstructing the datacube first. And they inspire this work.

ChenXinmeng, Li Yuting, Liu Jiying and Zhu Jubo are with the College of Science, National University of Defence Technology.

II. Problem Formulation

In this section, we first introduce a sparse representation based HSI classification algorithm by representing the test sample by a sparse linear combination of training samples from a dictionary. And then simulate the compressive sampling process by left multiplying a sampling matrix. We then discuss the standard algorithms use to reconstruct the original hyperspectral image, as well as the procedure determine the label of the test pixel.

Let x be a L-dimensional hyperspectral pixel vector. Suppose we have C different classes, and the cth class has N_c training samples $\{a_j^c\}_{j=1,\ldots,N_c}$. If x belongs to the cth class, then the spectral signature of x lies in a low-dimension subspace spanned by the training samples. Specifically, x can be a linear combination of the dictionary:

$$\begin{aligned} x &= \alpha_1^c a_1^c + \alpha_2^c a_2^c + ... + \alpha_{N_m}^c a_{N_m}^c && (1)\\ &= [a_1^c\ a_2^c\ ...\ a_{N_m}^c][\alpha_1^c\ \alpha_2^c\ ...\ \alpha_{N_m}^c]^T && (2)\\ &= \mathbf{A}^c\alpha^c && (3) \end{aligned}$$

where $\mathbf{A}^c$ is a $C \times N_m$ class subdictionary for the mth class, whose columns are the training samples of the mth class. α^m is a sparse vector whose entries are the weights of the corresponding atom in the dictionary.

For a test pixel without class information, we can model it as a vector lying in a subspace combining by the C subspaces corresponding to C classes. Therefore, x can be represented as

$$x = \mathbf{A}^1\alpha^1 + \mathbf{A}^2\alpha^2 + ... + \mathbf{A}^C\alpha^C \quad (4)$$
$$= \mathbf{A}\alpha \quad (5)$$

where $\mathbf{A} = [\mathbf{A}^1, \mathbf{A}^2, ..., \mathbf{A}^C]$ is a $L \times N$ matrix, $N = \sum_{c=1}^{C} N_c$. And α is a N-dimension long column vector cncatenated by all sparse vector α^c.

In this paper, we perform spectral compressive sensing per pixel,

$$\mathbf{y} = \mathbf{M}\mathbf{x} \quad (6)$$

where $\mathbf{M} \in \mathbb{R}^{l \times L}$ is the sensing matrix, y is the measurement of dimension l. Substituting (4) for $\mathbf{x}$ in (6), we can get

$$\mathbf{y} = \mathbf{M}\mathbf{A}\alpha \quad (7)$$

Then we would like to reconstruct and classify the original datacube at the same time. Given the measuerment matrix $\mathbf{M}$ and the dictionary $\mathbf{A}$, the sparse vector α can be achieved by solving the following optimization problem:

$$\hat{\alpha} = \arg\min \|\alpha\|_0 \quad subject\ to: x = \mathbf{M}\mathbf{A}\alpha \quad (8)$$

Considering the noise in empirical data, the equality constraint in (8) can be relaxted to an inequality one:

$$\hat{\alpha} = \arg\min \|\alpha\|_0 \quad subject\ to: \|\mathbf{M}\mathbf{A}\alpha - x\|_2 \leq \sigma \quad (9)$$

where σ is the error tolerance. The aforementioned problem can also be written as minimizing a fidelity term with a certain sparsity level:

$$\hat{\alpha} = \arg\min \|\mathbf{M}\mathbf{A}\alpha - x\|_2 \quad subject\ to: \|\alpha\|_0 \leq K_0 \quad (10)$$

K_0 is a given upper bound on the sparsity level.

Note that the dictionary A is structured in the sense that it exhibits high mutual coherence between vector columns. Here we can use principle component transformation to get the orthonormal basis $\hat{A}$, which is obtained from the set of training samples:

- $A_1, A_2, ..., A_C$ are the set of the training samples corresponding to C classes. The matrix A is formed by combining the set of training pixels as column vectors. This is, $A = [A_1, A_2, ..., A_C]$.
- The covariance matrix is computed as: $\Gamma = (A - \overline{A})(A - \overline{A})^T$; where $\overline{A}$ is the mean spectral vector.
- Seek the set of orthonormal vectors u_i that best describes the covariance matrix: $\Gamma u_i = \alpha_i u_i$. The set of eigenvectors, $\Psi_i = \overline{A}u_i$, corresponds to the vectors in the dictionary $\Psi = \{\psi_1, \psi_2, ..., \psi_L\}$.
- Every hyperspectral pixel observation x can be linearly decomposed into its eigenvectors components, i.e., $x = [\psi_1, \psi_2, ..., \psi_L]\theta$. θ is a sparse vector.

Principal component analysis (PCA) [11] is one of the most used approaches for hyperspectral iamge dimensionality reduction and compression since it preserves most information of the signal in just a few of its principal components. Thus, the original sparse vector α representing the test pixel x can be transformed into a new sparse vector θ which represents the pixel in the orthogonal basis Ψ. Specifically, the entries are ordered in decreasing order.

Therefore the optimization is converted into:

$$\hat{\theta} = \arg\min \|\mathbf{M}\Psi\theta - x\|_2 \quad subject\ to: \|\theta\|_0 \leq K_0 \quad (11)$$

where $\hat{\theta}$ is the estimated sparse vector and we can directly use it to classify the corresponding pixel. However, it's NP-hard, and it can be approximately sloved by greedy pursuit algorithms, such as Orthogonal Matching Pursuit (OMP) [12] or Subspace Pursuit (SP) [13]. Both OMP and BP are used to locate the support of the saprse vector that approximately solve the problem in (11). But the atoms are selected from the dictionary in different ways. The OMP algorithm augments the support set by one index at each iteration until K_0 atoms are selected or the approximation error is within a pre-defined threshold. While the SP algorithm maintains a set of K_0 indices. At each iteration, the index set is refined by adding K_0 new candidates to the current list and then discarding K_0 insignificant ones from the list of $2K_0$ candidates. With the backtracking mechanism, SP is able to find the K_0 most significant atoms. The computational complexity is in the order of $\mathcal{O}(BNK_0)$ for OMP and is upperbounded by $\mathcal{O}(BNK_0)$ for SP. Alsoit can be solved by the $\ell 1$-regularized least square solution by interior point method [14] or the gradient projection for sparse reconstruction (GPSR) method [15]. We choose to use a greedy pursuit method, which is generallized OMP method in this paper due to it's faster than solving a convex optimization problem with the $\ell 1$-regularized method.

III. Algorithms

In this paper, we choose to use a generallized OMP algorithm, named Simultaneous Orthogonal Matching Pursuit (SOMP) [16], summarized as follows:

Algorithm 1 SOMP

Require: $L \times N$ dictionary $A = [a_1, a_2, ...,]$, $L \times 1$ data vector x, a stopping criteria{Make sure all columns in A and x have unit norm}

Ensure: Residual $R_0 = X$, active set $\Lambda_0 = \emptyset$, iteration counter k=1.

1: **repeat**
2: Find the index of the atom that best approximates all residuals: $\lambda_k = \arg\max \|R_{k-1}^T a_i\|_p$, $p \geq 1$
3: Update the index set $\Lambda_k = \Lambda_{k-1} \bigcup \{\lambda_k\}$
4: Calculate $P_k = (A_{\Lambda_k}^T A_{\Lambda_k})^{-1} A_{\Lambda_k}^T x$
5: Compute the residual $R_k = x - A_{\Lambda_k} P_k$
6: $k \leftarrow k + 1$
7: **until** Stopping criteria is satisfied

We could get the final optimal index set $\Lambda = \Lambda_{k-1}$ through SOMP, the location of nonzero entries of the sparse representation is indexed by the final optimal index. After getting the estimated sparse vectors, we use such coefficients to deduce the class of the corresponding pixel, which is

$$Class(x) = \arg\min_c \|\hat{\theta} - \theta_c^q\|_2^2; \forall q = 1, 2, ..., N_c \quad (12)$$

This classifier is known as the nearest neighbor method [17]. Many other modified versions of the nearest neighbor method, such as local manifold learning weighted k-NN [18], can also be used to improve the performance of the classifier.

In summary, we can classify the pixel and reconstruct the original hyperspectral image at the same time with the measurements. It's an inverse problem and we can utilize greedy pursuit algorithms to approximate the solution. First we are supposed to transform the dictionary to an orthonormal basis and after we get the sparse vector, the class of the pixel can be determined using the nearest neighbor classifier. We summarize the algorithm in the following figure, named C-SOMP:

Algorithm 2 C-SOMP

Require: The measurement matrix M, the dictionary A, the measurements x and a stopping criteria{Make sure all columns in A and x have unit norm}

Ensure: Residual $R_0 = X$, active set $\Lambda_0 = \emptyset$, iteration counter k=1.

1: Compute the orthonormal dictionary with the training samples using the aforementioned method. And the outcome is denoted as $\overline{A}$.
2: **repeat**
3: Find the index of the atom that best approximates all residuals: $\lambda_k = \arg\max \|R_{k-1}^T a_i\|_p$, $p \geq 1$
4: Update the index set $\Lambda_k = \Lambda_{k-1} \bigcup \{\lambda_k\}$
5: Calculate $P_k = ((M\overline{A})_{\Lambda_k}^T (M\overline{A})_{\Lambda_k})^{-1} (M\overline{A})_{\Lambda_k}^T x$
6: Compute the residual $R_k = x - (M\overline{A})_{\Lambda_k} P_k$
7: $k \leftarrow$ k+1
8: **until** stopping criteria is satisfied
9: $Class(x) = \arg\min_c \|\hat{\theta} - \theta_c^q\|_2^2$; $\forall q = 1, 2, ..., N_c$

IV. Experiments

In this paper, the dictionary A consists of training samples is sampled directly from the image of interest. And we use random matrix following a Gaussian i.i.d. as the sampling matrix.

A. Data Sets

We use the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) Indian Pines data [19] to verify the effectiveness of the proposed method. The AVIRIS Indian Pines image was captured over the agricultural Indian Pine test site in northwestern Indiana. The image has 220 data channels across the spectral range from 0.2 to $2.4\mu m$, and each band is of size 145 145, with a spatial resolution of 20 m per pixel. In the experiments, 20 water absorption bands were discarded [20]. The reference map for Indian Pines contains 16 classes, most of which are different types of crops (e.g., corns, soybeans, and wheat). Fig. IV-A and IV-A shows the color composite of the Indian Pines image and the corresponding reference data.

B. Quantitative Metrics

An objective metric overall accuracy (OA) is adopted in the experiments to evaluate the quality of classification results. The OA measures the percentage of pixels that are correctly classified.

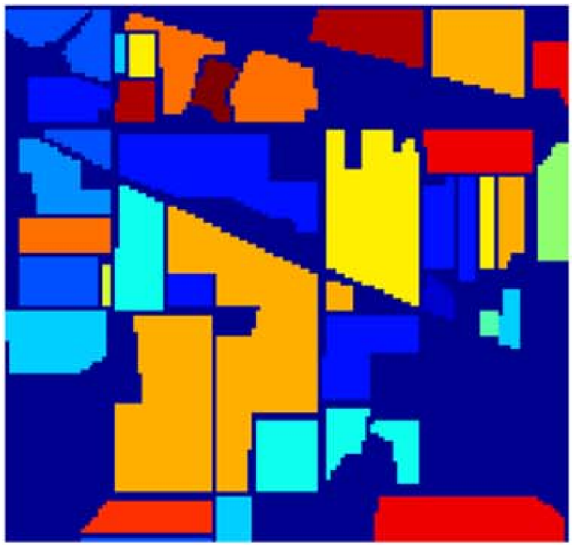

Fig. 1. Ground truth classification.

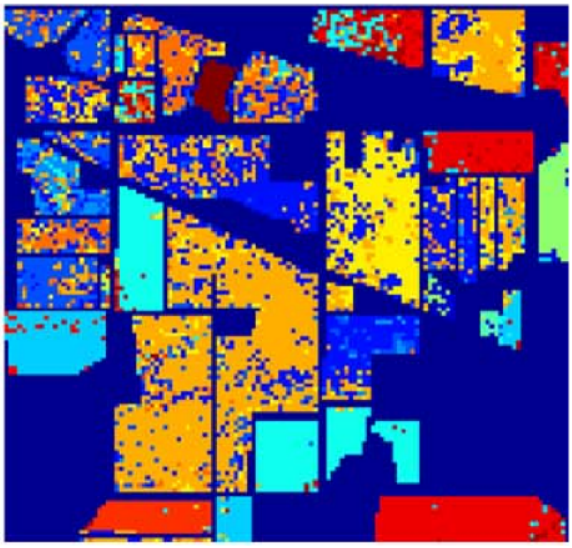

Fig. 2. Outcome of 25% sampling ratio.

C. Experimental Results

In the experiment, we use random matrix following a Gaussian i.i.d. as our sampling matrix. The sampling ratio is 0.25, 0.30, 0.35, 0.40, 0.50. 10% of the pixels is randomly chosen to be used as training samples from the image. The remaining 90% of the pixels are used for testing. Using the algorithm described above, we successfully do the reconstruction and the classification task at the same time. The outcome are summarized in the table:

Sampling ratio	OA
0.25	61.9%
0.30	65.8%
0.35	68.4%
0.40	74.4%
0.50	76.6%

Specifically, we show the outcome of a 25% sampling ratio

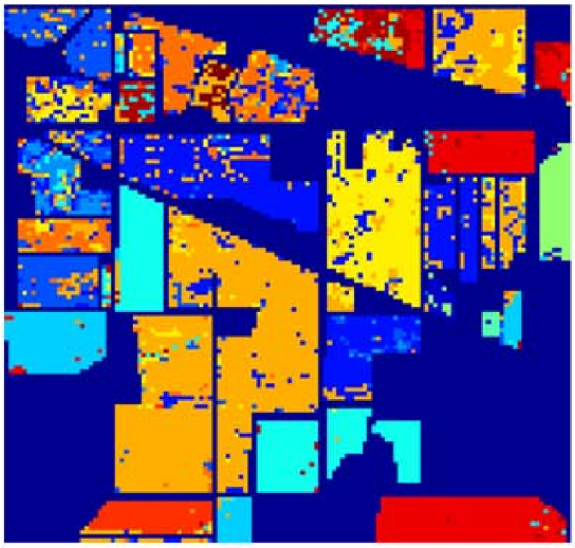

Fig. 3. Outcome of 50% sampling ratio.

in Fig.2, with the OA of 61.9% and 50% sampling ratio in Fig.3, with the OA of 76.6%. From these two images, we can see that we successfully reconstructed the images and showed that it is possible to classify the pixels directly from the compressed measurements, avoiding to reconstruct the original hyperspectral datacube first. However, the performance of the algorithm is not good enough and we can further improve the results by incorporating the spatial information.

V. Conclusion

In this paper, we have exploited a compressed sensing (CS) scheme for multidimensional signal acquisition and classification with a particular focus on hyperspectral images. In the proposed method, each HSI pixel is assumed to be sparsely represented by samples in a given dictionary obtained by the training samples. And then the sparse representation of a test pixel is recovered by solving an optimization problem with sparsity constraint and nearest neighbor method is used to label each test pixel. By such method, we successfully reconstructed the images and showed that it is possible to accurately classify the pixels directly from the compressed measurements, avoiding to reconstruct the original hyperspectral datacube first. Experimental results on an hyperspectral image show the classification accuracy for the proposed methodology. However, we only exploited the spectral information in the paper and this approach can further developed by taking advantage of the spatial information. From a computational point of view, we suffer from heavy computational burden and further improvement is under development.

References

[1] Sek M Chai, Antonio Gentile, Wilfredo E Lugo-Beauchamp, José L Cruz-Rivera, and D Scott Wills. Hyper-spectral image processing applications on the simd pixel processor for the digital battlefield. In *Computer Vision Beyond the Visible Spectrum: Methods and Applications, 1999.(CVBVS'99) Proceedings. IEEE Workshop on*, pages 130–138. IEEE, 1999.

[2] Dimitris Manolakis, Christina Siracusa, and Gary Shaw. Hyperspectral subpixel target detection using the linear mixing model. *IEEE Transactions on Geoscience and Remote Sensing*, 39(7):1392–1409, 2001.

[3] Emmanuel J Candès and Michael B Wakin. An introduction to compressive sampling. *IEEE signal processing magazine*, 25(2):21–30, 2008.

[4] Emmanuel J Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on information theory*, 52(2):489–509, 2006.

[5] Emmanuel J Candes and Terence Tao. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE transactions on information theory*, 52(12):5406–5425, 2006.

[6] Gustavo Camps-Valls, Devis Tuia, Lorenzo Bruzzone, and Jon Atli Benediktsson. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Processing Magazine*, 31(1):45–54, 2014.

[7] Yi Chen, Nasser M Nasrabadi, and Trac D Tran. Hyperspectral image classification using dictionary-based sparse representation. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10):3973–3985, 2011.

[8] Chengbo Li, Ting Sun, Kevin F Kelly, and Yin Zhang. A compressive sensing and unmixing scheme for hyperspectral data processing. *IEEE Transactions on Image Processing*, 21(3):1200–1210, 2012.

[9] Mohammad Golbabaee, Simon Arberet, and Pierre Vandergheynst. Compressive source separation: Theory and methods for hyperspectral imaging. *IEEE Transactions on Image Processing*, 22(12):5096–5110, 2013.

[10] Gabriel Martín, José M Bioucas-Dias, and Antonio Plaza. Hyca: A new technique for hyperspectral compressive sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 53(5):2819–2831, 2015.

[11] Ian Jolliffe. *Principal component analysis*. Wiley Online Library, 2002.

[12] Joel A Tropp and Anna C Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on information theory*, 53(12):4655–4666, 2007.

[13] Wei Dai and Olgica Milenkovic. Subspace pursuit for compressive sensing signal reconstruction. *IEEE Transactions on Information Theory*, 55(5):2230–2249, 2009.

[14] S Kim, Kwangmoo Koh, Michael Lustig, Stephen Boyd, and Dimitry Gorinevsky. A method for large-scale l1-regularised least squares problems with applications in signal processing and statistics. *IEEE Journal on Selected Topics in Signal Processing*, 1:606–617, 2007.

[15] Robert D Nowak, Stephen J Wright, et al. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE Journal of selected topics in signal processing*, 1(4):586–597, 2007.

[16] Joel A Tropp, Anna C Gilbert, and Martin J Strauss. Algorithms for simultaneous sparse approximation. part i: Greedy pursuit. *Signal Processing*, 86(3):572–588, 2006.

[17] James M Keller, Michael R Gray, and James A Givens. A fuzzy k-nearest neighbor algorithm. *IEEE transactions on systems, man, and cybernetics*, (4):580–585, 1985.

[18] Li Ma, Melba M Crawford, and Jinwen Tian. Local manifold learning-based-nearest-neighbor for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 48(11):4099–4109, 2010.

[19] D Landgrebe. Aviris nw indianas indian pines 1992 data set, 1992.

[20] J Anthony Gualtieri and Robert F Cromp. Support vector machines for hyperspectral remote sensing classification. In *The 27th AIPR Workshop: Advances in Computer-Assisted Recognition*, pages 221–232. International Society for Optics and Photonics, 1999.