SPARSITY-BASED CLASSIFICATION OF HYPERSPECTRAL IMAGERY

Yi Chen[1]*, Nasser M. Nasrabadi*[2] *and Trac D. Tran*[1]

[1]Department of Electrical and Computer Engineering, The Johns Hopkins University
3400 N. Charles Street, Baltimore, MD 21218
[2]US Army Research Laboratory
2800 Powder Mill Road, Adelphi, MD 20783

ABSTRACT

In this paper, a new sparsity-based classification algorithm for hyperspectral imagery is proposed. This algorithm is based on the concept that a pixel in hyperspectral imagery lies in a low-dimensional subspace and thus can be represented by a sparse linear combination of the training samples. The sparse representation can be recovered by solving a constrained optimization problem. Once the sparse vector is obtained, the class of the test sample can be directly determined by the behavior of the vector on reconstruction. In addition to the constraints on sparsity and reconstruction accuracy, we also exploit the fact that hyperspectral images are usually smooth within a neighborhood. In our proposed algorithm, a smoothness constraint is imposed by forcing the Laplacian of the reconstructed image to be minimum in the optimization process. The proposed sparsity-based algorithm is applied to several hyperspectral imagery to classify the pixels into target and background classes. Simulation results show that our algorithm outperforms the classical hyperspectral target detection algorithms.

1. INTRODUCTION

In this paper, we consider a two-class classification problem for hyperspectral imagery (HSI) where pixels are labeled as target or background based on their spectral characteristics. A number of algorithms have been proposed for this purpose based on statistical hypothesis testing techniques [1]. Among these approaches, spectral matched filter (SMF) [2], matched subspace detectors (MSD) [3], and adaptive subspace detectors (ASD) [4] have been widely used to detect targets of interests.

We propose a classification algorithm based on sparse representation. We use the same sparsity model in [5] where a test sample is approximately represented by very few training samples from both target and background dictionaries, and the sparse representation can be recovered and used directly for classification. In addition to the constraints on sparsity and reconstruction accuracy as in [5], we show that it is necessary to exploit the fact that neighboring HSI pixels usually have a similar spectral characteristics as well. To achieve this, we impose a smoothing constraint on the reconstructed image by forcing the Laplacian of the reconstructed image to be zero. The proposed approach has several advantages over the aforementioned classical techniques. First, there is no explicit assumption on the statistical distribution characteristics. Furthermore, the target dictionary can be easily augmented to account for various illumination and atmospheric conditions, making the dictionary invariant to the environmental variations [6]. Moreover, the sparsity model in our approach has the flexibility of imposing additional restrictions corresponding to the characteristics of HSI such as smoothness across neighboring hyperspectral pixels.

The paper is structured as follows. Our sparsity-driven classification algorithm is presented in Section 2. The effectiveness of the proposed method is demonstrated by simulation results presented in Section 3. Conclusions are drawn in Section 4.

2. SPARSITY-BASED CLASSIFICATION

In this section, we introduce a sparsity-based classification algorithm by sparsely representing the test sample using training samples. Firstly, we describe the details of the sparse subspace model used in the proposed algorithm.

2.1. Sparsity Model

Let $\boldsymbol{x}$ be a B-dimensional hyperspectral pixel observation, where B is the number of spectral bands. If $\boldsymbol{x}$ is a background pixel, its spectrum approximately lies in a low-dimensional subspace spanned by the N_b background training samples $\left\{\boldsymbol{a}_i^b\right\}_{i=1,2,\ldots,N_b}$. Then, $\boldsymbol{x}$ can be approximately represented by a linear combination of the training samples as follows.

$$\boldsymbol{x} \approx \alpha_1 \boldsymbol{a}_1^b + \alpha_2 \boldsymbol{a}_2^b + \cdots + \alpha_{N_b} \boldsymbol{a}_{N_b}^b = \begin{bmatrix} \boldsymbol{a}_1^b & \boldsymbol{a}_2^b & \cdots & \boldsymbol{a}_{N_b}^b \end{bmatrix} \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{N_b} \end{bmatrix}^T = \boldsymbol{A}_b \boldsymbol{\alpha}, \tag{1}$$

where $\boldsymbol{A}_b$ is the $B \times N_b$ background dictionary and $\boldsymbol{\alpha}$ is an unknown vector whose entries are the abundances of the corresponding atoms in $\boldsymbol{A}_b$. In our model, $\boldsymbol{\alpha}$ turns out to be a sparse vector (i.e., a vector with only few non-zero entries).

Similarly, a target pixel $\boldsymbol{x}$ can also be sparsely represented by the N_t target training samples $\{\boldsymbol{a}_i^t\}_{i=1,2,\ldots,N_t}$ as

$$\boldsymbol{x} \approx \beta_1 \boldsymbol{a}_1^t + \beta_2 \boldsymbol{a}_2^t + \cdots + \beta_{N_t} \boldsymbol{a}_{N_t}^t = \begin{bmatrix} \boldsymbol{a}_1^t & \boldsymbol{a}_2^t & \cdots & \boldsymbol{a}_{N_t}^t \end{bmatrix} \begin{bmatrix} \beta_1 & \beta_2 & \cdots & \beta_{N_t} \end{bmatrix}^T = \boldsymbol{A}_t \boldsymbol{\beta}, \tag{2}$$

where $\boldsymbol{A}_t$ is the target dictionary and $\boldsymbol{\beta}$ is a sparse vector whose entries contain the abundances of the target atoms in $\boldsymbol{A}_t$.

An unknown test sample $\boldsymbol{x}$ lies in the union of the background and target subspaces, and can be written as

$$\boldsymbol{x} = \boldsymbol{A}_b \boldsymbol{\alpha} + \boldsymbol{A}_t \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{A}_b & \boldsymbol{A}_t \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} = \boldsymbol{A} \boldsymbol{\gamma}, \tag{3}$$

where $\boldsymbol{A}$ consists of both background and target training samples and $\boldsymbol{\gamma}$ is a sparse $(N_b + N_t)$-dimensional vector formed by concatenating the two sparse vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Next, we show how to obtain $\boldsymbol{\gamma}$ and label the class of a test sample from $\boldsymbol{\gamma}$.

2.2. Reconstruction and Classification

Given the test sample $\boldsymbol{x}$ and dictionary $\boldsymbol{A}$, the vector $\boldsymbol{\gamma}$ can be obtained by solving the following optimization problem:

$$\hat{\boldsymbol{\gamma}} = \arg\min \|\boldsymbol{\gamma}\|_0 \quad \text{subject to} \quad \boldsymbol{A}\boldsymbol{\gamma} = \boldsymbol{x}, \tag{4}$$

where $\|\cdot\|_0$ denotes ℓ_0-norm which is defined as the number of non-zero entries in the vector. If the solution is sufficiently sparse, the problem in (4) can be relaxed to a linear programming problem which can be solved efficiently [7]. Alternatively, it can be solved by greedy pursuit algorithms such as the one in [8].

Once the sparse vector $\boldsymbol{\gamma}$ is obtained, the class of $\boldsymbol{x}$ can be determined by comparing the residuals $r_b(\boldsymbol{x}) = \|\boldsymbol{x} - \boldsymbol{A}_b\hat{\boldsymbol{\alpha}}\|_2$ and $r_t(\boldsymbol{x}) = \left\|\boldsymbol{x} - \boldsymbol{A}_t\hat{\boldsymbol{\beta}}\right\|_2$, where $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{\beta}}$ represent the recovered sparse coefficients corresponding to the background and target dictionaries, respectively. In our approach, the algorithm output is calculated by

$$D(\boldsymbol{x}) = r_b(\boldsymbol{x})/r_t(\boldsymbol{x}). \tag{5}$$

If $D(\boldsymbol{x}) > \delta$ with δ being a prescribed threshold, then $\boldsymbol{x}$ is determined as a target pixel; otherwise, $\boldsymbol{x}$ is labeled as background.

2.3. Classification with Smoothing Constraint

Hyperspectral imagery is usually smooth in the sense that neighboring pixels usually consist of similar materials and thus their spectral characteristics are highly correlated. To exploit the smoothness property of HSI, we incorporate a smoothing term in the sparsity-based algorithm. Let $\boldsymbol{I}$ represent the hyperspectral image and $\hat{\boldsymbol{I}}$ be its reconstruction. Let $\boldsymbol{x}_1$ be a pixel of interest, and $\boldsymbol{x}_i, i = 2, \ldots, 5$ be its four nearest neighbors in the spatial domain. While searching for the sparsest representation of the test sample $\boldsymbol{x}_1$, we simultaneously minimize the reconstructed image Laplacian $\triangledown^2\hat{\boldsymbol{I}}$ at the point $\boldsymbol{x}_1$, which is calculated as $4\hat{\boldsymbol{x}}_1 - \hat{\boldsymbol{x}}_2 - \hat{\boldsymbol{x}}_3 - \hat{\boldsymbol{x}}_4 - \hat{\boldsymbol{x}}_5$. In this way, the reconstructed test sample is forced to have a similar spectral characteristics as its four nearest neighbors; hence, smoothness is enforced across the spectral pixels in the reconstructed image.

Let $\boldsymbol{\gamma}_i$ be the sparse coefficient associated with $\boldsymbol{x}_i$. Then, the smoothness-constrained problem can be formulated as

$$\min \sum_{i=1}^{5} \|\boldsymbol{\gamma}_i\|_0 \quad \text{subject to:} \quad \boldsymbol{A}(4\boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2 - \boldsymbol{\gamma}_3 - \boldsymbol{\gamma}_4 - \boldsymbol{\gamma}_5) = \boldsymbol{0}, \quad \boldsymbol{x}_i = A\boldsymbol{\gamma}_i \ \text{ for } \ i = 1, \ldots, 5. \tag{6}$$

In (6), the first set of linear constraints forces the reconstructed image Laplacian to become zero such that the reconstructed neighboring pixels have similar spectral characteristics, and the second set minimizes reconstruction errors. The optimization problem in (6) can be rewritten as

$$\min \quad \|\boldsymbol{\gamma}\|_0 \quad \text{subject to:} \quad \tilde{\boldsymbol{A}}\boldsymbol{\gamma} = \tilde{\boldsymbol{x}}, \tag{7}$$

$$\text{where} \quad \tilde{\boldsymbol{A}} = \begin{bmatrix} 4\boldsymbol{A} & -\boldsymbol{A} & -\boldsymbol{A} & -\boldsymbol{A} & -\boldsymbol{A} \\ \boldsymbol{A} & & & & \\ & \boldsymbol{A} & & & \\ & & \boldsymbol{A} & & \\ & & & \boldsymbol{A} & \end{bmatrix}, \ \boldsymbol{\gamma} = \begin{bmatrix} \boldsymbol{\gamma}_1 \\ \vdots \\ \boldsymbol{\gamma}_5 \end{bmatrix}, \ \text{and} \ \tilde{\boldsymbol{x}} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{x}_1 \\ \vdots \\ \boldsymbol{x}_5 \end{bmatrix}.$$

The problem in (7) is the standard form of a linearly-constrained sparsity-minimization problem and can be solved using the various available solvers as previously mentioned.

Classification can be performed based on the behavior of the sparse coefficients as it was done in Section 2.2. The algorithm output is computed as in (5) by the ratio of residuals. If the output $D(\boldsymbol{x})$ is greater than a prescribed threshold δ, then the test sample is labeled as a target; otherwise it is labeled as background.

3. SIMULATION RESULTS AND ANALYSIS

The proposed algorithm, as well as the classical algorithms SMF, MSD, and ASD, are applied to two HSI. The results are compared both visually and quantitatively by the receiver operating characteristics (ROC) curves, which describes the probability of detection as a function of the probability of false alarms. The two test images, the desert radiance II data collection (DR-II) and forest radiance I data collection (FR-I), are from a hyperspectral digital imagery collection experiment (HYDICE) sensor. We use 150 of the 210 bands generated by the HYDICE sensor, removing the absorption and low-SNR bands. The DR-II image contains 6 military targets and the FR-I image contains 14 targets as seen in Fig. 1(a) and Fig. 2(a), respectively.

We show a comparison between the performances of the sparsity-based technique and the classical target detection algorithms for the DR-II and FR-I images. For both images, the target dictionary $\boldsymbol{A}_t$ contains $N_t = 18$ atoms from the leftmost target, and the background signatures are generated locally for each test sample to better adapt to the local statistics. The output of the proposed smoothness-constrained approach for DR-II is shown in Fig. 1(b). For visual comparison, the outputs of other algorithms are also displayed in Figs. 1(c)-(f). We see that the sparsity-based algorithm with smoothing constraint leads to the best visual quality. Similar results can be observed in Fig. 2 for the FR-I image.

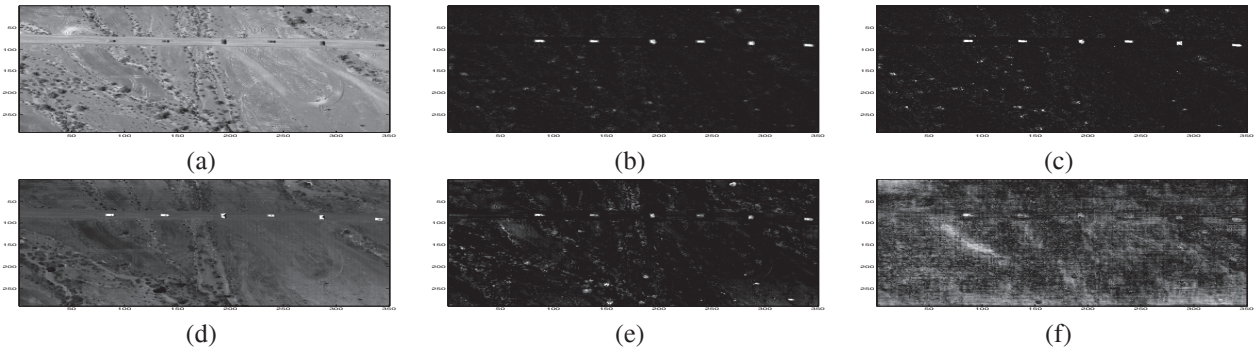


Fig. 1. (a) The mean DR-II image. Outputs for DR-II with (b) sparsity-based algorithm with smoothing constraint using (7), (c) sparsity-based algorithm without smoothing constraint using (4), (d) MSD, (e) SMF, and (f) ASD.

The ROC curves for DR-II and FR-I images are shown in Fig. 3. Under the same settings, we compare the performance of the proposed sparsity-based algorithm to the previously-developed detectors. Obviously, the proposed classification algorithm with the smoothness constraint significantly outperforms the other detectors.

4. CONCLUSIONS

In this paper, we propose a classification algorithm for hyperspectral imagery based on sparse representation of the test samples. In the proposed algorithm, the sparse representation is recovered by solving a constrained optimization problem

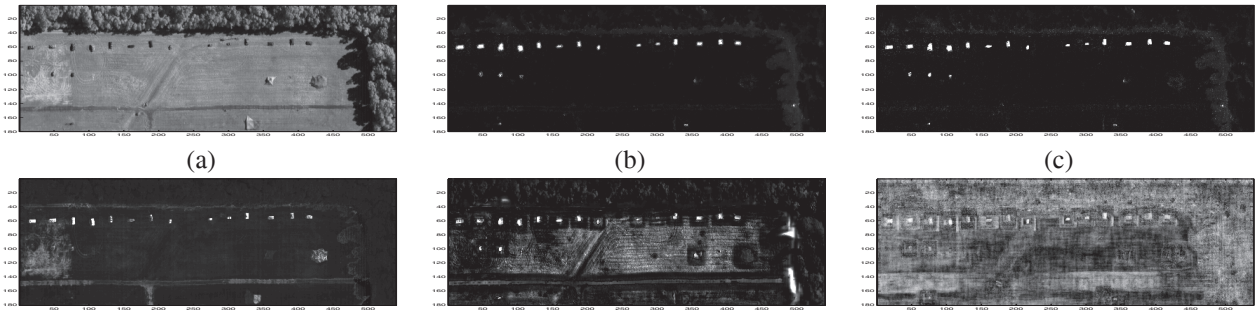
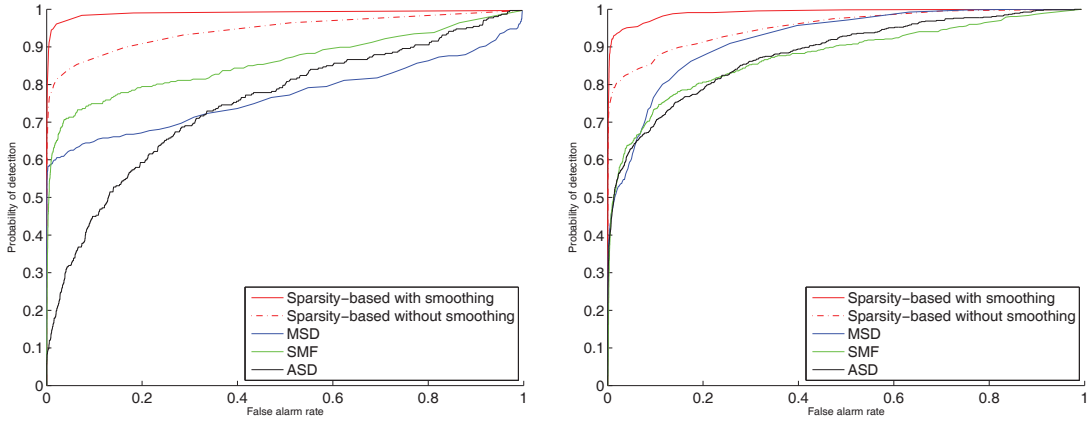


Fig. 2. (a) The mean FR-I image. Outputs for FR-I with (b) sparsity-based algorithm with smoothing constraint using (7), (c) sparsity-based algorithm without smoothing constraint using (4), (d) MSD, (e) SMF, and (f) ASD.



(a) (b)

Fig. 3. ROC curves for (a) DR-II and (b) FR-I.

that addresses the sparsity, reconstruction accuracy, and smoothness of the reconstructed image simultaneously, and then the classification decision is obtained directly from the recovered sparse vectors. The new algorithm outperforms the previously-developed detectors in terms of both qualitative and quantitative measures, as demonstrated by experimental results in several real hyperspectral imageries.

5. REFERENCES

[1] D. Manolakis and G. Shaw, "Detection algorithms for hyperspectral imaging applications," *IEEE Signal Processing Magazine*, vol. 19, no. 1, pp. 29–43, Jan. 2002.

[2] F. C. Robey, D. R. Fuhrmann, E. J. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[3] L. L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. on Signal Processing*, vol. 42, no. 8, pp. 2146–2157, Aug. 1994.

[4] S. Kraut, L. L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. on Signal Processing*, vol. 49, no. 1, pp. 1–16, Jan. 2001.

[5] J. Wright, A. Y. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[6] B. Thai and G. Healey, "Invariant subpixel material detection in hyperspectral imagery," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 40, no. 3, pp. 599–608, Mar. 2002.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM journal on scientific computing*, vol. 20, no. 1, pp. 33–61, 1998.

[8] W. Dai and O. Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," Jan. 2009, Preprint, `arXiv:0803.0811v3[cs.NA]`.