UNSUPERVISED HYPERSPECTRAL IMAGE CLASSIFICATION USING BLIND SOURCE SEPARATION

Qian Du, Sumit Chakrarvarty

Department of Electrical Engineering and Computer Science
Texas A&M University-Kingsville
Kingsville, Texas 78363, USA

ABSTRACT

This paper presents an unsupervised classification algorithm for hyperspectral remotely sensed imagery based on blind source separation. Since the area covered by a single pixel in such an image is very large, the reflectance of a pixel is the mixture from all the materials resident in this area. A contrast function consisting of the mutual information minimization and orthogonality among the outputs, is defined to separate the assumed linear mixture so as to achieve soft classification. In order to reduce the computational complexity, a Neyman-Pearson detection theory based eigen-thresholding method is used to estimate the number of classes, followed by a band selection technique to select smaller number of bands used in the learning algorithm. The preliminary result using an AVIRIS experiment demonstrates the feasibility of the proposed algorithm.

I. INTRODUCTION

Hyperspectral remote sensing has been received lots of interest in recent years due to the fact that its very high spectral resolution provides the potential in accurate material identification. Because the area covered by a single pixel is very large (typically several square meters for the data acquired by an airborne sensor and several hundred square meters for the data acquired by a spaceborne sensor), the pixel reflectance is a mixture from all the materials resident in this area. So in Hyperspectral image processing we deal with mixed pixels instead of pure pixels as in ordinary digital image processing. One of the major difficulties of hyperspectral image processing is the spectral signature variability. Materials can be theoretically identified by their spectral characteristics (i.e., spectral signature). However, the variability of spectral signature of the same material is profound in remote sensing applications due to the variations in atmospheric conditions, sensor noise, material composition, surrounding materials, etc. So hyperspectral image processing is very challenging in that the spectral signature of a mixed pixel does not correspond to any single well-defined material. Such complexity requires us to develop efficient unsupervised algorithms that do not depend on any prior information about material spectral signatures.

Quite a few techniques have been proposed in hyperspectral image classification. But most of them are supervised methods, i.e., the number of classes and their spectral signatures are known *a priori*. In this paper an unsupervised technique is to be proposed without assuming any prior information of the image scene. Here the classification problem is a soft classification because the abundance fractions of all the materials in a pixel area are to be estimated in the classified images, which described the distributions of these materials in the image scene. If we assume the mixing process is linear, it actually is a linear unmixing problem, which can be solved by a technique of blind source separation. Several researchers have proposed such methods in [1-4]. We will introduce a different contrast function and simplify the computational complexity by taking advantage of our previous research results.

II. CONTRAST FUNCTION

In the linear mixture model [5], the mixing process is assumed linear. If these are p materials present, a pixel vector $\mathbf{r}$ can be represented as

$$\mathbf{r} = \mathbf{M}\boldsymbol{\alpha} + \mathbf{n} \tag{1}$$

where $\mathbf{M} = \lfloor \mathbf{m}_1, \mathbf{m}_2, \cdots, \mathbf{m}_p \rfloor$ is the material signature matrix and $\mathbf{m}_i$ is the spectral signature of the i-th material, $\boldsymbol{\alpha} = (\alpha_1 \alpha_2 \cdots \alpha_p)^T$ is the abundance vector and $\boldsymbol{\alpha}_i$ is the abundance of the i-th material contained in pixel **r**, **n** is the noise term.

We assume that the p abundance fractions in a pixel are mutually statistically independent and at most one of them is Gaussian distributed. Then the classification based on linear unmixing can be viewed as a blind source separation problem. A solution to this kind of problem can be obtained via independent component analysis (ICA). The goal is to find a separating matrix **W** such that the elements in $\mathbf{y} = \hat{\boldsymbol{\alpha}} = \mathbf{W}\mathbf{r}$ are as independent as possible. The mutual information is chosen as the measure of independence,

$$I(\mathbf{y}) = \sum_{i=1}^{n} H(y_i) - H(\mathbf{y}) \tag{2}$$

where $H(\mathbf{y}) = -E\lfloor \log(p_{\mathbf{y}}(\mathbf{y})) \rfloor$ is the joint entropy of random vector **y**, $H(y_i) = -E\lfloor \log(p_{y_i}(y_i)) \rfloor$ is the entropy of y_i. The learning algorithm for **W** searching is to solve a constrained problem,

$$\text{minimize } I(\mathbf{y}) \text{ subject to } E(\mathbf{y}\mathbf{y}^T) = \mathbf{I} \tag{3}$$

where **I** is the identity matrix. So the objective function is formulated as

$$C(\mathbf{W}) = I(\mathbf{y}) + \sum_{i,j=1}^{n} \beta\left(E(y_i y_j) - \delta_{ij}\right)^2 \tag{4}$$

where β is a penalty term. An optimal **W** is such that Eq. (4) can be minimized.

When **W** is a square matrix, i.e., the number of materials n in the image scene equals the number of bands L, $p_{\mathbf{y}}(\mathbf{y})$ and $p_{\mathbf{r}}(\mathbf{r})$ can be related as

$$p_{\mathbf{y}}(\mathbf{y}) = \frac{p_{\mathbf{r}}(\mathbf{r})}{\left|\frac{\partial \mathbf{y}}{\partial \mathbf{r}}\right|} \tag{5}$$

where $\left|\frac{\partial \mathbf{y}}{\partial \mathbf{r}}\right|$ is the determinant of the Jacobian matrix of **y** with respect to **r**, i.e.,

$$\left|\frac{\partial \mathbf{y}}{\partial \mathbf{r}}\right| = \begin{vmatrix} \frac{\partial y_1}{\partial r_1} & \cdots & \frac{\partial y_1}{\partial r_L} \\ \vdots & & \vdots \\ \frac{\partial y_L}{\partial r_1} & \cdots & \frac{\partial y_L}{\partial r_L} \end{vmatrix} = \begin{vmatrix} w_{11} & \cdots & w_{1L} \\ \vdots & & \vdots \\ w_{L1} & \cdots & w_{LL} \end{vmatrix} = |\mathbf{W}| \cdot$$

Therefore, $H(\mathbf{y})$ can be related with $H(\mathbf{r})$ as [6]

$$H(\mathbf{y}) = E\{\log|\mathbf{W}|\} + H(\mathbf{r}). \tag{6}$$

Taking derivative with respect to **W** yields

$$\frac{\partial H(\mathbf{y})}{\partial \mathbf{W}} = \frac{1}{|\mathbf{W}|}. \tag{7}$$

The computation of $H(y_i)$ needs to use the pdf of y_i which is unknown. One way to tackle this problem is to approximate it using Gram-Charlier expansion as [7]

$$H(y_i) \approx \frac{1}{2}\log(2\pi e) - \frac{1}{12}\left(k_3^i\right)^2 - \frac{1}{48}\left(k_4^i\right)^2 + \frac{3}{8}\left(k_3^i\right)^2 k_4^i + \frac{1}{16}\left(k_4^i\right)^3 \tag{8}$$

where $k_3^i = m_3^i$, $k_4^i = m_4^i - 3$ and $m_k^i = E\lfloor (y_i)^k \rfloor$ is the k-th moment of y_i. We know

$$\frac{\partial m_k^i}{\partial w_{st}} = E\left[(y_s)^{k-1} r_t\right]\delta_{is} \tag{9}$$

so

$$\frac{\partial \sum_{i=1}^{n} H(y_i)}{\partial w_{st}} = -\frac{1}{6}k_3^s E\left[(y_s)^2 r_t\right] - \frac{1}{24}k_4^s E\left[(y_s)^3 r_t\right] + \frac{3}{4}k_3^s k_4^s E\left[(y_s)^2 r_t\right] + \frac{3}{8}\left(k_3^s\right)^2 E\left[(y_s)^3 r_t\right] + \frac{3}{16}\left(k_4^s\right)^2 E\left[(y_s)^3 r_t\right] \tag{10}$$

As for the second term in Eq. (4),

$$\frac{\partial \sum_{i,j=1}^{n} \beta\left(E(y_i y_j) - \delta_{ij}\right)^2}{\partial w_{st}} = 2\beta \sum_{i=1}^{n} \left(E(y_i y_s) - \delta_{is}\right) E(y_i r_t) \tag{11}$$

Based on Eqs. (7), (10) and (11), we can easily calculate the learning update for **W** as

$$\delta \mathbf{W} = -\eta \frac{\partial C(\mathbf{W})}{\partial \mathbf{W}} \qquad (12)$$

where η is a positive learning rate.

III. ESTIMATION OF THE NUMBER OF CLASSES

The algorithm proposed in Section (2) is very time consuming because the size of **W** is very large. For instance, for a Hyperspectral image with 210 bans, a 210×210 separating matrix must be estimated. A possible way to reduce the size of **W** is to estimate the number of classes, i.e, the number of outputs, which in general is much smaller than the band number. We apply an eigen-thresholding technique in [8] to do the estimation.

Let $\{\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_L\}$ and $\{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L\}$ be two sets of eigenvalues generated by sample correlation matrix $\mathbf{R}_{L\times L}$ and sample covariance matrix $\mathbf{K}_{L\times L}$, respectively. By assuming that signal sources are nonrandom unknown positive constants and noise is white with zero mean, we can expect that $\hat{\lambda}_l > \lambda_l$ for $l = 1, \cdots, p$ and $\hat{\lambda}_l = \lambda_l$ for $l = p+1, \cdots, L$, where $\hat{\lambda}_l = \lambda_l + \sigma_l^2$ for $l = 1, \cdots, p$ and $\hat{\lambda}_l = \lambda_l = \sigma_l^2$ for $l = p+1, \cdots, L$.

In order to determine the *p*, a binary hypothesis problem can be formulated as follows.

$$\begin{array}{ll} H_0 : z_l = \hat{\lambda}_l - \lambda_l = 0 & \\ \textit{versus} & \text{for } l = 1, \cdots, L \\ H_1 : z_l = \hat{\lambda}_l - \lambda_l > 0 & \end{array} \qquad (13)$$

where the null hypothesis H_0 and the alternative hypothesis H_1 represent the case that the correlation-eigenvalue is equal to its corresponding covariance eigenvalue and the case that the correlation-eigenvalue is greater than its corresponding covariance eigenvalue respectively. In other words, when H_1 is true (i.e., H_0 fails), it implies that there is a signal contributing the correlation-eigenvalue in addition to noise since noise eigenvalues in $\mathbf{R}_{L\times L}$ are equal to the corresponding eigenvalues in $\mathbf{K}_{L\times L}$.

We can model the difference z_l between each pair of eigenvalues, $\hat{\lambda}_l$ and λ_l under hypotheses H_0 and H_1 as Gaussian random variables by the asymptotic conditional probability densities given as $p_0(z_l) \cong N(0, \sigma_{\lambda_l}^2)$ and $p_1(z_l) \cong N(\mu_l, \sigma_{\lambda_l}^2)$, where the mean μ_l is an unknown constant and the variance $\sigma_{\lambda_l}^2 \cong \frac{2\hat{\lambda}_l^2}{N} + \frac{2\lambda_l^2}{N}$. Then decision rule becomes

$$\begin{array}{l} \text{Choose } H_1 \text{ if } z_l \geq \tau_l \\ \text{Choose } H_0 \text{ if } z_l < \tau_l \end{array} \qquad (14)$$

where the eigen-threshold τ_l can be determined by using the Neyman-Pearson criterion so that the probability of correctly choosing H_1 is maximized subject to a designated upper bound on the false alarm probability p_f. τ_l is not fixed, but a function of eigenvalue index *i*. An estimated number of classes *p* can be obtained by counting the number of times when H_1 is true.

IV. BAND SELECTION

After the number of classes *p* is estimated, we have to select *p* band images from original data cube so that a square matrix **W** of size *p*×*p* can be constructed. We apply a joint band prioritization and band decorrelation approach in [9] to select band images to be used in the learning algorithm.

Let $\mathbf{\Sigma}_n$ be estimated noise covariance matrix and **F** is a noise-adjusted operator such that

$$\mathbf{F}^T \mathbf{\Sigma}_n \mathbf{F} = \mathbf{I}. \qquad (15)$$

Each pixel vector **r** is adjusted by $\mathbf{F}^T\mathbf{r}$. Then the variance of each adjusted band image is equivalent to the SNR in this band. The resulting band images are reordered based on their SNRs. The resulting prioritized image set is denoted as $\Omega = \{B_j\}_{j=1}^{l}$, where $B_1 \geq B_2 \geq \cdots \geq B_l$ and $\geq$ is the notation of priority order. The first *p* most distinct images are selected by comparing the divergence of each pair of band images in Ω.

V. EXPERIMENT

The data used in the experiment is the 224-band AVIRIS image shown in Fig. 1. It is a scene of 200×200 over the Lunar Crater Volcanic Field in Northern Nye County, Nevada. First, the method in Section III was applied to estimate the number of different materials, which was five. Second, five band images were selected using the approach described in Section IV. Then the algorithm in Section II was operated to find the optimal separating matrix **W**. The final classified images are shown in Fig. 2. Compared to the prior information, they correspond to "playa lake", "shade", "cinder", "vegetation" and "rhyolite", respectively.

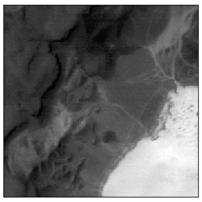

Figure 1: An AVIRIS image scene

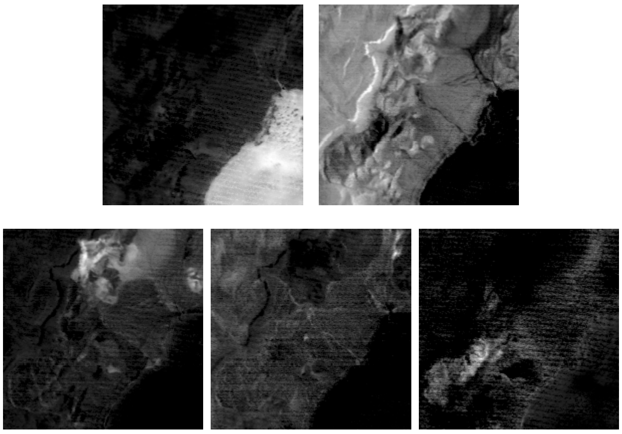

Figure 2: Unsupervised classification result

VI. CONCLUSION

We introduced an unsupervised classification algorithm based on minimizing mutual information criterion. In order to reduce the computational complexity, a Neyman-Pearson detection theory based eigen-thresholding method is used to estimate the number of classes, followed by a joint band prioritization and band decorrelation approach to select the same number of bands to be used in the proposed algorithm. The preliminary result using AVIRIS data shows its feasibility.

REFERENCE

[1] H. Szu and J. Buss, "ICA neural net to refine remote sensing with multiple labels," *Proc. SPIE* 4056, pp. 32-49, 2000.

[2] T. Tu, "Unsupervised signature extraction and separation in hyperspectral images: A noise-adjusted fast independent component analysis approach," *Optical Engineering*, vol. 39, no. 4, pp. 897-906, 2000.

[3] X. Zhang, C. H. Chen, "New independent component analysis method using higher order statistics with application to remote sensing images," *Optical Engineering*, vol. 41, no. 7, pp. 1717-1728, 2002.

[4] C.-I Chang, S.-S Chiang; J. A. Smith, I. W. Ginsberg, "Linear spectral random mixture analysis for hyperspectral imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 2 , pp. 375 –392, 2002.

[5] J. B. Adams, M. O. Smith and A. R. Gillespie, "Remote geochemical analysis: elemental and mineralogical composition," in *Image Spectroscopy: Interpretation Based on Spectral Mixture Analysis*, C. M. Pieters and P. A. J. Englert, Eds. Cambridge Univ. Press, pp. 145-166, 1993.

[6] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129-1159, 1995.

[7] S. Amari, A. Cichocki and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*, vol. 8, D. S. Touretzky, et.al, Eds. Cambridge, MA, 1996, pp. 757-763.

[8] C.-I Chang and Q. Du, "A noise subspace projection approach to determination of intrinsic dimensionality for hyperspectral imagery," *EOS/SPIE Symposium on Remote Sensing, Conference on Image and Signal Processing for Remote Sensing V*, SPIE vol. 3871, Florence, Italy, pp. 34-44, September 20-24, 1999.

[9] C.-I Chang, Q. Du, T.-L. Sun and M. L. G. Althouse, "A joint band prioritization and band decorrelation approach to band selection for hyperspectral image classification," *IEEE Trans. On Geoscience and Remote Sensing*, vol.37, pp.2631-2641, Nov.1999.