AIK Method for Band Clustering Using Statistics of Correlation and Dispersion Matrix

Muhammad Ahmad [1,+], Dr. Ihsan ul Haq [1] and Qaisar Mushtaq[1]

[1]Department of Electronics Engineering, International Islamic University (IIU), Islamabad

Abstract. The proposed method is based on different characteristics of band images of hyperspectral data. Band clustering approach is based on Dispersion Matrix and Correlation Coefficient. Hyperspectral data is calculated by these approaches. Materials do not have maximum discriminatory probability within the same band. Therefore a single band clustering and selection method is not effective for multiple targets. To overcome this problem multi-techniques are used for band clustering and for each technique same method is used for end member detection. An important issue in band selection is how many bands needed to be selected in order to preserve the maximum information. Vertex component analysis (VCA) is used for unmixing of hyperspectral data. The entire work is done by using MATLAB.

Keywords: Hyperspectral; Dimensional Reduction; Dispersion; Correlation;

1. Introduction

Hyperspectral Imaging is also known as Imaging Spectroscopy [1]. Measurement, analysis, and interpretation of hyperspectral spectra acquired from a scene at a short medium or long distance by an airborne or satellite sensors. Image spectroscopy concept instigated in 1980's by A.F.H. Goetz at NASA's [1]. Airborne Visible infra red imaging spectrometer (AVIRIS) is originated by [2]. This system was able to cover the wave length region from 0.4-2.5 μm, and spectral resolution of 10nm. The sensors collect data simultaneously in hundreds of narrow and adjacent spectral bands over the wave length range from near ultraviolet through the thermal infra red as resolution of 5 nm [2-3].

In hyperspectral image data cube, each pixel is a vector whose dimensions are equal to the number of bands. The contiguousness of bands makes them highly correlated locally. High correlation between the spectral bands causes data redundancy. It is difficult to unmix, detect, and classify the targets in the presence of high correlation among the bands. It is necessary to reduce the dimensionality for high accuracy in unmixing and detection. The dimensional reduction strongly affects the performance of many supervised classification methods [4]. The high dimensionality of hyperspectral data can be reduced by taking the advantage of correlation of bands. This can be done by mapping higher dimensional space to lower uncorrelated dimensional space while preserving maximum information. In literature, many supervised and unsupervised techniques are designed according to the application, situation, and available resources. Especially regarding to hyperspectral imaging, since the data is transformed, some crucial and critical information may have been compromised and distorted, for instance, when dealing with physical measures represented in the data domain.

Much research has been done in the area of dimensional reduction of high dimensional data. Principal component analysis (PCA), using the Karhunen-Loeve (KL) transformation is a common method for dimensional reduction. While PCA gives optimal dimensional reduction, while maintaining fidelity of the signal in a mean square error sense, it is not optimal with respect to any particular signal analysis task, such

---

[+] Corresponding author. Tel.: +92-321-6617922.
 E-mail address: mahmad00@yahoo.com.

as target detection or classification, other methods including Isomap, Multidimensional scaling (MDS), Clustering and feature selection.

2. Unmixing and Target Detection

Unmixing of hyperspectral image pixel is the estimation of the pixel spectrum into a collection of constituent spectra called end member spectral signatures, and their corresponding abundance fractions [5]. Spectrum of the target and spectra of the background may also be used for sub-pixel detection to observed mixed pixel spectrum [6]. End members derivation can be grouped into two main classes. First: assume that end member exist in the image either in pure pixel or in mixed pixel [7]. Second: these methods derive the spectra of end members analytically [8]. Linear and non-linear mixture models are widely used for mixed pixels modeling. Mostly linear mixture model (LMM) has been used for source separation. In LMM, it is assumed that the observed pixel spectrum is the linear combination of a small number of unique and distinct constituent deterministic spectral signatures.

Vertex component analysis (VCA) is an unsupervised end member extraction algorithm [8]. VCA also exploits that the end members are the vertices of a simplex as N-FINDR dose. VCA also shows that the affine transformation of a simplex is also a simplex. VCA works on the assumption that pure pixels of end-members are present in data. VCA iteratively projects data onto a direction orthogonal to the subspace spanned by the end members already determined until all end-members depleted. The new end member signature corresponds to the extreme of the projection.

3. Proposed Method for Band Clustering

Clustering of band images keep the inter cluster variance minimum and the intra cluster variance maximum. Hyperspectral data spread in some directions, this data can be measured by using different statistical methods, and we used Dispersion Matrix (DISM) and Correlation Coefficient (CC) to estimates the hyperspectral data.

Suppose that we have $[\![X_i]\!]_{i=1}^L$ band images data, where L is the total number of bands. $M \times N$ Is the size of each band image and X_i is the mean of the ith band image. The statistical characteristics we used for band clustering are given below.

3.1 Dispersion Matrix (DISM)

The Dispersion matrix can be obtained by taking the difference between each value of the matrix by its mean value and it is also known as a covariance of a matrix, as

$$DISM\ (x_i) = \begin{bmatrix} (x_1 - E[x_1])\ (x_1 - E[x_1])^T & \cdots & (x_i - E[x_i])\ (x_i - E[x_i])^T \\ \vdots & \ddots & \vdots \\ (x_i - E[x_i])\ (x_i - E[x_i])^T & \cdots & (x_i - E[x_i])\ (x_i - E[x_i])^T \end{bmatrix}$$

Similarly the DISM (x_j) can be obtain, so we have estimate the data matrix (each band image) by the following way,

$$E_i = DISM(x_p, x_q) = \frac{1}{MN} \sum_{p,q=1}^{MN} \left(E\left[(x_p - E[x_p])\ (x_q - E[x_q])^T \right] \right) \quad , p \neq q \tag{1}$$

3.2 Correlation Coefficient (CC)

$$E_i = R\ (k, l) = \frac{1}{MN} \sum_{k,l=1}^{MN} \frac{C\ (k, l)}{\sqrt{C\ (k, k)\ C\ (l, l)}} \tag{2}$$

The results from above methods for L bands are given as $X = \{E_i\}_{i=1}^L$. The K-means divides the N observations into K sets (k < n), for the observations $X = [x_1, x_2, x_3 \ldots x_n]$, where $K = [S_1, S_2, S_3 \ldots S_k]$, K-Means method computes the centroid clusters by different distance metrics. We used the following distance metrics in K-Means Clustering.

3.3 Correlation (COR)

For an M × N data matrix $X = [x_1, x_2, x_3 \ldots x_n]$, the distance between the vectors x_r and x_s is,

$$X_{rs} = 1 - \frac{(x_r - \bar{x}_r)(x_s - \bar{x}_s)^T}{[(x_r - \bar{x}_r)(x_r - \bar{x}_r)^T]^{1/2}[(x_s - \bar{x}_s)(x_s - \bar{x}_s)^T]^{1/2}} \qquad (3)$$

The selected L band images for unmixing and target detection by above methods are shown in Table 1.

Table 1: Selected Bands

| CC-COR | 78 | 17 | 1 | 173 | 94 | 107 | 75 | 109 | 22 | 106 |
	13	53	163	86	66	58	71	110	34	137
	176	181								
DISM-COR	91	123	149	181	77	132	141	173	3	133
	153	130	187	85	69	15	58	72	46	23
	109	125								

4. Results and Conclusion

In the presented research work, a sub image of size 350×350 with 224 bands of a data set taken on the AVIRIS flight. Prior to the analysis of AVIRIS Cuprite image data, low SNR bands 1 – 3, 105 – 115 and 150 – 170 have been removed and the remaining 189 bands are used for experiments. Preserving the maximum information, the number of bands required are 22.

Spectral Angle Mapper (SAM) computes the spectral similarity between the found end-members pixels (a) Alunite, (b) Buddingtonite, (c) Calcite, and (k) Kaolinite and the actual end-members pixels (A) Alunite, (B) Buddingtonite, (C) Calcite, and (K) Kaolinite. The values of SAM among the same target/ end member's minerals are highlighted in Table 1 which shows a good similarity.

Table 2: Spectral Angle Mapper

CC-COR				
Found/Actual	A(161,62)	B(234,209)	C(347,30)	K(298,22)
a(160,61)	0.01649	0.16536	0.21246	0.09328
b(305,23)	0.14135	0.08415	0.10685	0.12527
c(186,152)	0.20081	0.08154	0.05938	0.21445
k(145,37)	0.08887	0.18344	0.22834	0.03407
DISM-COR				
a(161,61)	0.01718	0.16448	0.21151	0.09615
b(22,86)	0.20332	0.08714	0.08073	0.19249
c(298,23)	0.21181	0.09465	0.03932	0.22131
k(223,138)	0.10978	0.15234	0.19179	0.06133

The pulling out end member signatures and the laboratory end-members signatures are predominantly of (a) Alunite, (b) Buddingtonite, (c) Calcite, and (k) Kaolinite for 22 bands are shown in Figure 1 ((A)-(B)).
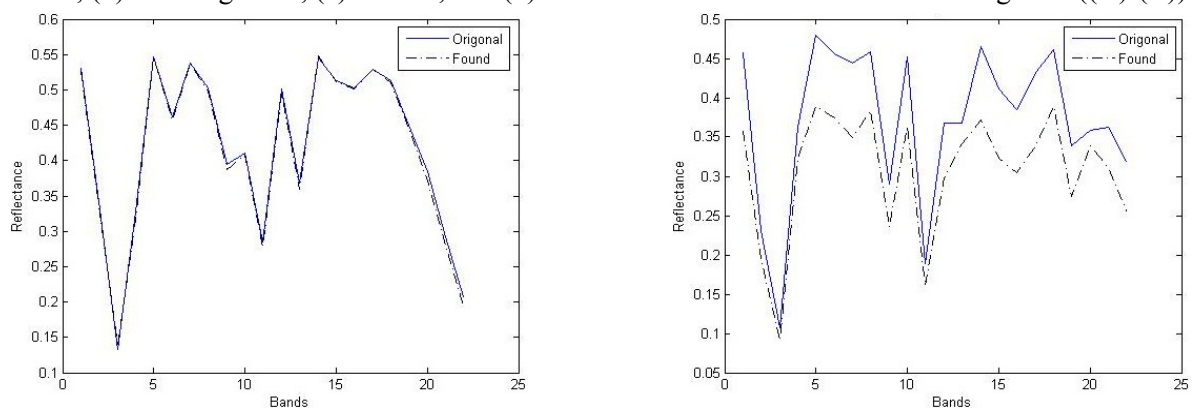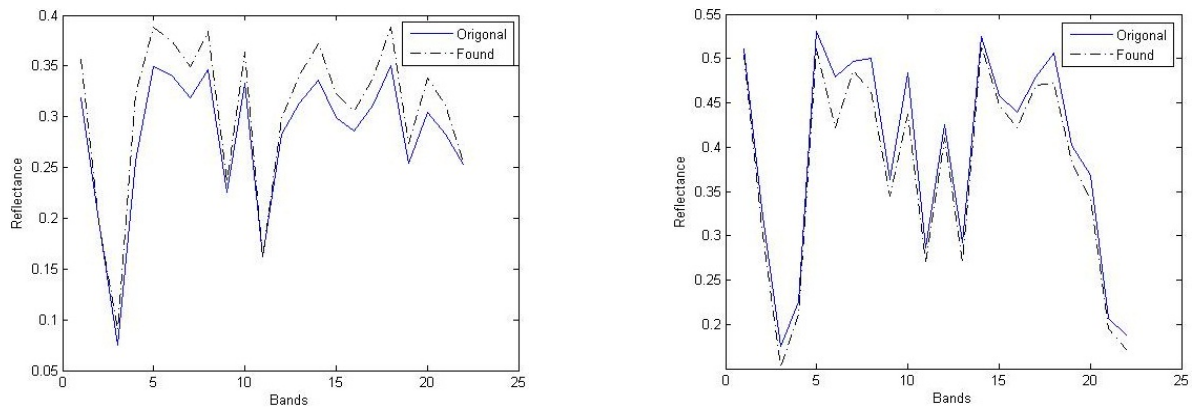
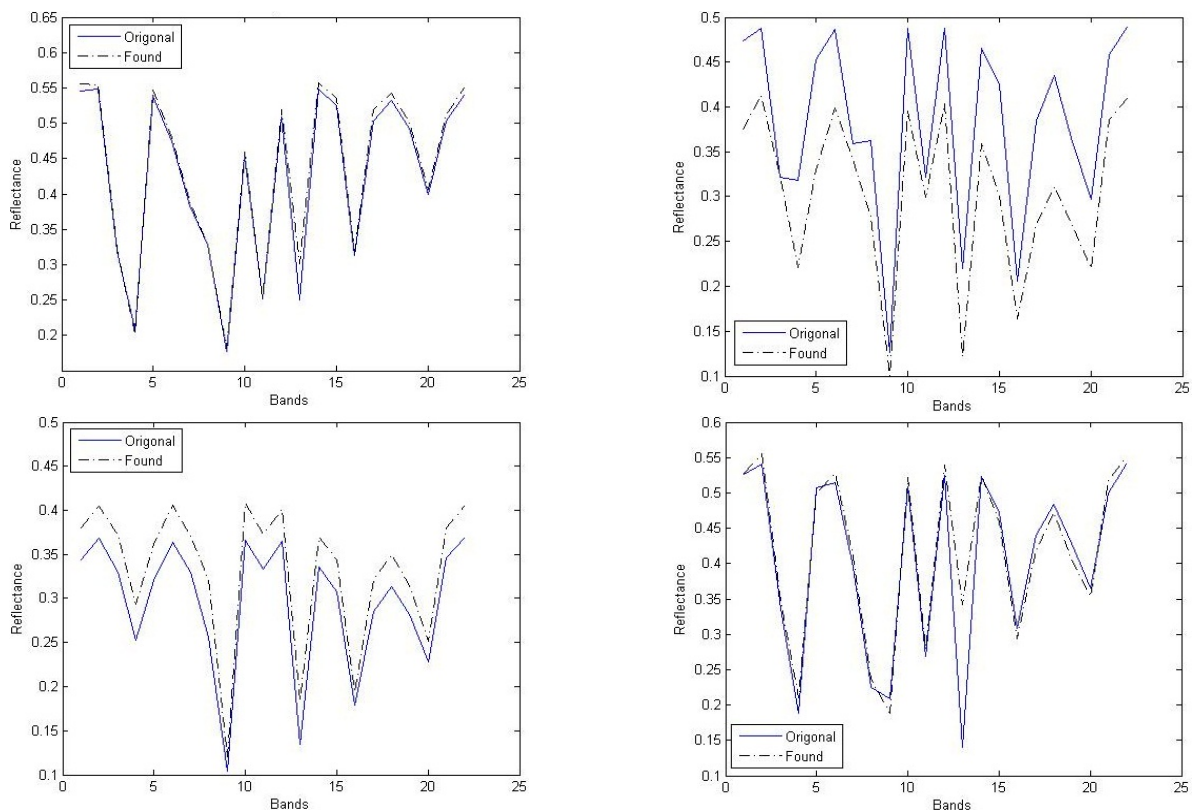Figure 1(A): F Found (doted Lines) and Laboratory Spectral Signatures (Solid Lines) by CC-COR

Figure 1(B): Found (doted Lines) and Laboratory Spectral Signatures (Solid Lines) by DISM-COR

5. Acknowledgements

6. References

[1] Goetz. F. H, Vane. G, Solomon. J, and Rock. B, *Imaging spectrometry for earth remote sensing*, in Airborne Imaging Spectrometer Data Analysis Workshop, JPL Publication, vol. 228, pp. 22–29, 1985.

[2] Green. R. O, Pavri. B, Faust. J, and Williams. O, *AVIRIS Radiometric Laboratory Calibration, In flight Validation, and a Focused Sensitivity Analysis*, NASA Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 1998.

[3] Haq. U. I, and Xu. X, *A New Approach to Band Clustering and Selection for Hyperspectral Imagery*, IEEE Proceedings on Signal Processing, pp.1198-1202, ICSP, 2008.

[4] Serpico. S, and Bruzzone. L, *A new search algorithm for feature selection in hyperspectral remote sensing images*, IEEE Transactions on Geo science and Remote Sensing, vol.39, pp. 1360–1367, 2001.

[5] Keshava. N, *Best bands selection for detection in hyperspectral processing*, IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 5, pp. 3149–3152, 2001.

[6] Manolakis. D, *Taxonomy of detection algorithms for hyperspectral imaging applications*, Optical Engineering, vol. 44, no. 6, p. 066403, 2005.

[7] Winter, *Fast autonomous spectral end-member determination in hyperspectral data*, in Proceedings of the Thirteenth International Conference on Applied Geologic Remote Sensing, vol. 11, pp. 337–344, 1999.

[8] Nascimento. J. M. P. and Dias. J. M. B, *Vertex component analysis: a fast algorithm to unmix hyperspectral data*, IEEE Transactions on Geo science and Remote Sensing, vol. 43, pp. 898–910, 2005.