A Band Selection Method based on Synthesized Information Content for Hyper-spectral Pixel Un-mixing

Yuan Bo[1] , Wang Ning[1], Li Chuanrong[2]

[1] Key Laboratory of Quantitative Remote Sensing Information Technology, Academy of Opto-Electronics, Chinese Academy of Sciences, Beijing, China.100094
[2] Academy of Opto-Electronics, Chinese Academy of Sciences, Beijing, China.100094

E-mail: byuan@aoe.ac.cn

Abstract. Hyper-spectral remote sensing has the defects of huge data size and massive information redundancy, especially when processing pixel un-mixing, which always has high computation complexity. This paper proposes a band selection method for hyper-spectral pixel un-mixing, based on synthesized parameters of information content. It uses Kullback-Leibler divergence and mutual information with respective weights to construct a new comprehensive information matrix. The matrix can indicate the overall distribution of data's spectral information. According to the comprehensive information matrix, the method can select a small-number band combination from the massive bands of initial data in an iterative way. The experimental results show that, the method is effective in decreasing data volume and retaining effective spectral information，its result is better than those of similar algorithms. This method can be chosen as an effective preprocessing step for hyper-spectral pixel un-mixing.

1. Introduction

The "mixed pixel" phenomenon, which prevails in hyper-spectral images, makes pixel un-mixing a key step in quantitative analysis of hyper-spectral data. Un-mixing methods need to extract pattern from spectral information contained in initial data, as well as decreasing data volume and information redundancy. So initial spectral information and structure need to be retained as much as possible.

The solution of problems mentioned above can be categoried of dimension reduction and band selection. The latter is a better option, since the mathematical manipulation in dimension reduction would destruct data's initial spectral structure, while band selection would not. There are many band selection methods in present, including Maximum Variance Principle Component Analysis (MVPCA), Adaptive Band Selection (ABS)[1] and so on. As these methods are not designed for hyper-spectral data or un-mixing process specifically, their performances of retaining initial spectral information are not always satisfying[2].

This paper proposes a new band selection method aiming at hyper-spectral pixel un-mixing. Experimental result shows that, it can decrease data size significantly and retain valid spectral information as much as possible, which will be very helpful in processing pixel un-mixing methods.

2. Method

The redundant information among adjacent bands is the key point and inner motivation of hyper-spectral data's band selection. This paper selects K-L divergence and mutual information as the basic parameters for measuring inherent valid spectral information.

2.1. Number of bands to be selected

Pixel un-mixing methods generally call for that the number of mixed signals is no less than the number of end-members[3]. The number of end-members can be estimated by Principal Component Analysis[4] (PCA), Minimum Noise Fraction[5] (MNF) or Singular Value Decomposition (SVD). This paper adopts PCA to estimate the band number.

2.2. K-L divergence matrix

In probability theory, the Kullback-Leibler divergence (K-L divergence) is a non-symmetric measure of the difference between two probability distributions *P* and *Q*, it's defined as

$$D_{KL}(P \| Q) = \sum_i P(i) \ln \frac{P(i)}{Q(i)} \tag{2-1}$$

The K-L divergence is only defined if P and Q both sum to 1 and if $Q(i) > 0$ for any i such that $P(i) > 0$.As for hyper-spectral data, a big K-L divergence value between two bands means distinct information difference and big amount of effective information. So K-L divergence can be chosen as an indicator of the information content difference among spectral bands.

$D_{KL}(P \| Q)$ can only present the information difference between two bands, while the band selection process needs an overall distribution presentation of information difference among all bands. A K-L divergence matrix M_{KL} is defined in this paper. As for hyper-spectral data $X \in R^{N \times M}$ (N is the number of all pixels, M is the number of all bands), M_{KL} is a $M \times M$ square matrix, as shown in (2-2).

$$M_{KL} = \begin{bmatrix} 0 & D_{KL}(X_1 \| X_2) & \cdots & D_{KL}(X_1 \| X_M) \\ D_{KL}(X_2 \| X_1) & 0 & & \cdots \\ \cdots & & \cdots & D_{KL}(X_{M-1} \| X_M) \\ D_{KL}(X_M \| X_1) & \cdots & D_{KL}(X_M \| X_{M-1}) & 0 \end{bmatrix} \tag{2-2}$$

$D_{KL}(X_i \| X_j)$ is the K-L divergence of all pixels from band i spectral reflectance to band j, it can be calculated by formula (2-1). When calculating $D_{KL}(X_i \| X_j)$, $X(t)$ should be normalized beforehand to fulfill the constraint of "sum of $X(t)$ equal to 1". Reflectivity value of hyper-spectral data is all nonnegative, and all of the zero value should be set as a tiny positive value (such as 0.001) to fulfill the constraint condition of "$X_i(t) > 0$， and $X_j(t) > 0$".

2.3. Average mutual information matrix

Besides K-L divergence, this paper adopts average mutual information (AMI) to weigh the information content contained in hyper-spectral data more thoroughly. The mutual information (MI) of two random variables is a quantity that measures their mutual dependence, AMI is the statistical average value in the joint probability space.

The MI of two discrete random variables X and Y can be defined as (2-3).

$$I(X;Y) = \sum_{y \in Y} \sum_{x \in X} p(x,y) \log \left(\frac{p(x,y)}{p(x)p(y)} \right) \tag{2-3}$$

This paper uses AMI to overcome the randomness of MI, and obtains a certain quantity, so it defines an AMI matrix I. I is a $M \times M$ square matrix， as shown in equation (2-4).

$$I=\begin{bmatrix} I(X_1;X_1) & I(X_1;X_2) & \cdots & I(X_1;X_M) \\ I(X_2;X_1) & \cdots & & \cdots \\ \cdots & & \cdots & I(X_{M-1};X_M) \\ I(X_M;X_1) & \cdots & I(X_M;X_{M-1}) & I(X_M;X_M) \end{bmatrix} \tag{2-4}$$

$I(X_i;X_j)$ is the average mutual information of all pixels' reflectivity between band *i* and band *j*, it can be calculated according to equation (2-3).

AMI matrix I and K-L divergence matrix M_{KL} represent the inter-band distribution of information from different aspects, they are not correspond with each other strictly. So the adoption of average mutual information isn't needless, it can present the overall information distribution of hyper-spectral data more comprehensively together with K-L divergence, reduce the negative influence of utilizing single parameter (such as abnormal value), and improve the algorithm's stability.

2.4. Comprehensive information matrix

To indicate the information distribution among all bands more comprehensively and suppress the one-sidedness of single parameter, this paper combines the K-L divergence matrix M_{KL} and AMI matrix I with respective weights, and produces a new comprehensive information matrix *S*. *S* is treated as the direct reference of band selection. A bigger element in *S* means more information difference (valid information) and less information dependence (information redundancy) between the two corresponding bands. As a $M \times M$ square matrix，*S* is shown in equation (2-5).

$$S=\begin{bmatrix} S_{(1,1)} & S_{(1,2)} & \cdots & S_{(1,M)} \\ S_{(2,1)} & \cdots & & \cdots \\ \cdots & & \cdots & S_{(M-1,M)} \\ S_{(M,1)} & \cdots & S_{(M,M-1)} & S_{(M,M)} \end{bmatrix} \tag{2-5}$$

$S_{(i,j)}$ indicates the comprehensive spectral information difference (valid spectral information) of all pixels between band *i* and band *j*. $S_{(i,j)}$ can be calculated as (2-6).

$$S_{(i,j)}=c_1\times D_{KL}(X_i \square X_j)-c_2\times I(X_i;X_j) \tag{2-6}$$

c_1 and c_2 are the weights of M_{KL} and I respectively, and can be determined by the mean values of matrix M_{KL} and I's elements. Since the ratio between c_1 and c_2 is what we are really interested in, for simplification, the computational formula of *S* can be set as (2-7).

$$S=\frac{c_1}{c_2}\times M_{KL}-I \tag{2-7}$$

2.5. Band selection

It's unpractical to select all bands by utilizing comprehensive information matrix at one time. The method of this paper selects one band from remaining bands at a time, which contains the most comprehensive spectral information content together with selected bands.

Considering a random column of *S*: $s_i = \{S(1,i),S(2,i),...,S(M,i)\}$, it indicates the amount of effective information between band *i* and all of the other bands. Supposing the index combination of selected bands is $K = \{K_1,K_2,\ldots,K_U\}$, calculate $S_K = S(K_1,i)+S(K_2,i)+\ldots+S(K_U,i)$, select index *i* from remaining unselected band indices, which can make S_K get it maximum value (It means that, the new band combination composed by band *i* and previous band combination, contains the

maximum valid information content in some certain number of bands). Then remove band *i* from unselected bands, add it into the selected combination.

Matrix *S* is the synthesis of K-L divergence matrix and average mutual information matrix, the position with bigger value of *S* means bigger K-L divergence and smaller average mutual information at the same time, which indicates that, the information difference is big and the information redundancy is small, so it's the ideal place for band selection.

The process of band selection is listed as follows:

① Estimate the number of end-members through PCA transformation;

② Calculate the K-L divergence Matrix M_{KL} and the average mutual information matrix *I*, then get the comprehensive information matrix *S*;

③ Normalize matrix *S*, Select the first band according to the maximum value of *S*;

④ Select the second band according to the first selected band and *S*, and select all of the other bands in an iterative way. Specifically, find the second band which has the maximum comprehensive effective information (together with the first selected band) firstly. Then find the band which has the maximum comprehensive effective information, together with the updated selected band combination. By parity of reasoning, select all of the other bands.

3. Experimental process and result

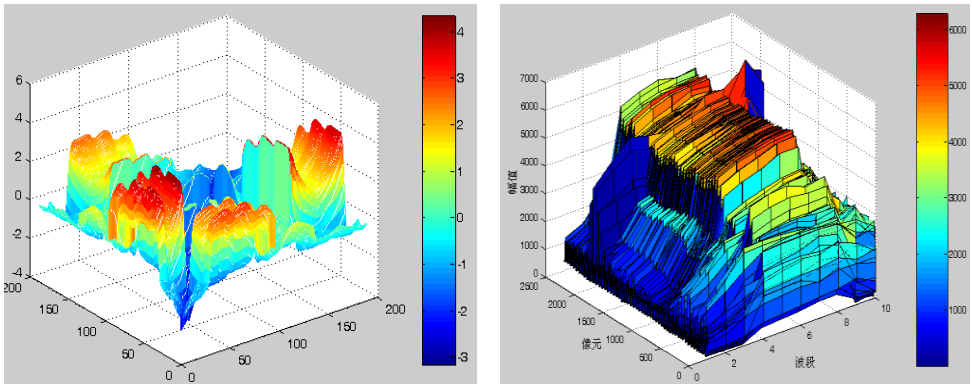
3.1. Experimental data

Experimental data is shown in figure 1. It's an AVIRIS (Airborne Visible Infrared Imaging Spectrometer) data of San Diego (an American city) district. The AVIRIS payload's band number is 224, band range is 0.4~2.5μm, spectral resolution is 10nm, spatial resolution is 3.5m, and the valid band number of experimental data is 189, the image's line number and column number are both 50.

Figure 1. Experimental data of AVIRIS in San Diego, USA.

3.2. Experimental process

The final results of band selection are 10 bands, their indices are listed as follows: {2，7，27，37，43，49，63，68，136，189}.



(a) 3D Visualization of the comprehensive information matrix.
(b) 3D visualization of the 10 selected bands' spectrum.
Figure 2. Experimental Results

3.3. Analysis and comparison

As shown in figure 2, in the combination of 10 selected bands, the inter-band difference, and the overall spectral difference among all bands is very distinct, which means that the band combination has little information redundancy and enough effective information, so it has good expected effect in the pixel un-mixing process of hyper-spectral data.

If we use the mean value of selected bands' K-L divergence matrix as the approximate representation of retained effective information, then the comparison of effective information retained by 3 different band selection methods, including Maximum Variance Principle Component Analysis (MVPCA), Adaptive Band Selection (ABS) and the proposed method, is listed in table 1.

Table 1. The comparison of retained effective information among different band selection methods

Band selection method	Average K-L divergence of band selection result	Ratio between average K-L divergence of different methods (Result of the method proposed in this paper is assumed to be 1)
MVPCA	0.2237	0.8221
ABS	0.1965	0.7222
The method of this paper	0.2721	1

As shown in table 1, comparing with the other two methods, the proposed method can retain more effective information for pixel un-mixing of hyper-spectral data.

Finally, compute the similarity of the two abundance matrices. (Similarity of two matrices is a non-dimensional value, with range of [-1, 1], 0 represents complete irrelevant, 1 or -1 represents perfect positive correlation or perfect negative correlation.) The first is the abundance matrix of the 189-bands initial data, the second is the abundance matrix of the 10-selected-bands data. They are both estimated by pixel un-mixing algorithm of basic NMF. The result of similarity is 0.8922, which is very close to 1. The results indicate that the proposed method has ideal effect in pixel un-mixing.

4. Conclusions

This paper proposes a new band selection method specifically for the pixel un-mixing of hyper-spectral data. This method has definite physical meanings, it adopts multiple information parameters to demote the effective spectral information of hyper-spectral data comprehensively, and selects spectral bands accordingly. The experimental results prove that it's effective in decreasing data quantity and retaining effective spectral information. Since the method can improve the feasibility and

efficiency of pixel un-mixing process, and maintain the un-mixing precision, it can be chosen as an effective preprocessing step of hyper-spectral pixel un-mixing.

In the follow-up study，the method's performance of local convergence and stability needs further research; more similar combination of information parameter, or different weight coefficients need to be tested and compared based on this paper's method; and more complex un-mixing algorithms need to be adopted, to testify the performance of this method.

References

[1] A Mohan, G Sapiro and E Bosch 2007 Spatially Coherent Nonlinear Dimensionality Reduction and Segmentation of Hyper-spectral Images *IEEE Geosic. Remote Sens.* **4** 206

[2] Peter Bajcsy and Peter Groves 2004 Methodology for Hyper-spectral Band Selection *Photogrammetric Engineering and Remote Sensing journal* **70** 793

[3] Daniel D Lee and H Sebastian Seung 2000 Algorithms for Non-negative Matrix Factorization *Advances in Neural Information Processing Systems 13: Proceedings of the 2000 Conference* **2** 556

[4] Abdi H, and Williams L J 2010 Principal component analysis *Wiley Interdisciplinary Reviews: Computational Statistics* **1** 433

[5] J R Harris, D Rogge, R Hitchcock, O Ijewliw and D Wright 2005 Mapping lithology in Canada's Arctic: application of hyper-spectral data using the minimum noise fraction transformation and matched filtering *Canadian Journal of Earth Sciences* **42** 2173