Statistical Approach to Noisy Band Removal for Enhancement of HIRIS Image Classification

Nguyen van Huan, Hakil Kim

Image Information Research Center
School of Information and Communication Engineering, INHA University
{huan, hikim}@vision.inha.ac.kr

Abstract

The accuracy of classifying pixels in HIRIS images is usually degraded by noisy bands since noisy bands may deform the typical shape of spectral reflectance. Proposed in this paper is a statistical method for noisy band removal which mainly makes use of the correlation coefficients between bands. Considering each band as a random variable, the correlation coefficient measures the strength and direction of a linear relationship between two random variables. While the correlation between two signal bands is high, existence of a noisy band will produce a low correlation due to ill-correlativeness and undirectedness. The application of the correlation coefficient as a measure for detecting noisy bands is under a two-pass screening scheme. This method is independent of the prior knowledge of the sensor or the cause resulted in the noise. The classification in this experiment uses the unsupervised *k*-nearest neighbor algorithm in accordance with the well-accepted Euclidean distance measure and the spectral angle mapper measure. This paper also proposes a hierarchical combination of these measures for spectral matching. Finally, a separability assessment based on the between-class and within-class scatter matrices is followed to evaluate the performance.

Key words: Noisy band removal, HIRIS image classification, separability assessment

I. Introduction

Applications of hyperspectral image mainly derive from the ability of using the spectral reflectance (spectrum) of the earth's surface at different wavelengths to determine the surface's type. By taking advantage of hundreds of narrow and spectrally continuous bands, hyperspectral image provides a tool for detecting materials that can not be solved from multispectral images (Goetz, 1991, van der Meer and de Jong, 2001; Ben-Dor *et al*., 2001). Unfortunately, owing to the performance of sensors, the atmospheric conditions and several others factors, producing a noise-free image is sometimes impossible, especially when the number of bands amounts to hundreds such as those sensed by the "Airborne Visible and Infrared Imaging Spectrometer" (AVIRIS) or the EO-1 Hyperion. While, in general, spectral matching methods measure the agreeability between the shapes of an image pixel's spectrum over that from a spectral library to assess the type of surface, noisy bands deform the typical shape of the spectrum.

A few works have been dedicated to solve this problem in the remotely sensed imaging field. The Maximum Noise Fraction (MNF) transformation (Green at al, 1988), which is a modified version of Principal Components (PC) transformation, is widely accepted in practice. The MNF transforms hyperspectral data into components with increasing signal-to-noise ratio based on the covariance matrices of the data and the noise components assumed to be known a prior. The noise removal is performed on the transformed data before transformation back to the original space. However, the transformation is computationally expensive when applying to hyperspectral data of hundreds of bands, since it needs to compute the left-hand eigenvectors of the $\Sigma_{\mathbf{N}}\Sigma^{-1}$ matrix, where $\Sigma_{\mathbf{N}}$ is the covariance matrix of noise components and Σ is that of the data cube. Faulconbridge *et. al* (2006) proposed a method of detecting noisy band by first, unsupervisedly classifying the original data into clusters, then looking for spectral locations where all of the clusters are very close to one another. To measure the distance between clusters, a modified Bhattacharyya distance is used. Finally, the bands where the maximum Bhattacharyya distance is less than a given threshold are marked as noise and removed.

In this paper, a process of removing noisy bands of hyperspectral data under a screening scheme is proposed that uses the correlation coefficient of bands as an indicator. The correlation coefficient is a statistically mathematic quantity that measures the strength and direction of a linear relationship between two random variables. Normally, two successive signal bands in hyperspectral image provide very similar band images thanks to the slight difference of the reflectance when the wavelength interval is small enough to produce a high correlation. However, noise can make the bands undirected so that the value in the case is relatively small. For the classification after removing noise, the un-supervised *k*-nearest neighborhood algorithm is utilized using the distance measures Euclidean and spectral angle mapper. In addition, in order to enhance the discrimination ability, a hierarchical combination of above measures is proposed in the intention of taking strong points of each individual measure. The performance of classification is assessed by a separability measure that bases on the post-classified between-class and within-class scatter matrices.

The data used in this work is an EO-1 Hyperion, acquired on June 3, 2001 over an area about 9,200 ha near Han river of the western Seoul where covers many surface types of urban, road, forest, grass, and water body. The data has 228 bands of size 400 rows × 256 columns, ranging from 406 to 2,496 nm wavelengths. More specifications can be referred to Kim *et al.*, 2007.

II. Noisy band removal

In our data, two types of noisy bands exist: zero bands and reflectively noisy bands as given in Fig. 1 below.

Step 1: Removing zero bands

The zero bands are easily removed by considering the mean of the bands. If the mean of band is near zero, the band is marked as noise and removed; otherwise the band is treated as a potential signal band.

Step 2: Finding true signal bands from potential signal bands

The correlation coefficients between pairs of most successive potential signal bands are computed that is defined as follows:

$$\rho_{ij} = \frac{E\left(\left(X-\mu_i\right)\left(Y-\mu_j\right)\right)}{\sigma_i \sigma_j} \quad (1)$$

ρ_{ij} : correlation coefficient between bands

$\mu_{\bullet}$: mean of a band

$\sigma_{\bullet}$: variance of a band

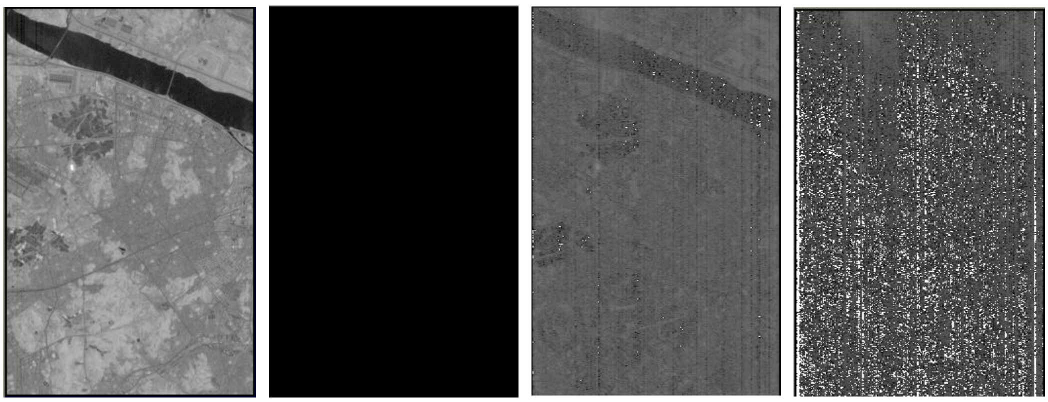


Fig 1. Examples of signal and noisy bands: a) signal band; b) zero band; c) and d) reflectively noisy bands

The correlation coefficients are then normalized to the range [0, 100] and construct a 100-bin histogram. Since adjacent signal bands normally have a high correlation distributed around the maximum, and the correlation between noisy bands or between a signal and noisy band are small and randomly distributed; the histogram will peaks at a point in the region of high correlation as depicted in Fig. 2. From the histogram, a threshold in the left of the peak is chosen for detecting true signal bands by the following rule:

- If $\rho_{i-1,i} > thres$ AND $\rho_{i,i+1} > thres$

  *i*th band is a signal band

  Else

  *i*th band is a potential noisy band

High correlations with the preceding band and the succeeding band indicate that the band under consideration is a true signal band.

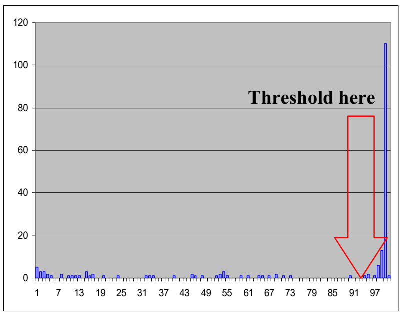


Fig. 2. 100-bin Histogram of correlation coefficient

Step 3: Refining the potential noisy bands

In the remaining bands here, sometimes a signal band is kept between two noisy bands, making the relationships to the preceding and the succeeding ill-correlative. This step aims at refining such bands as the following:

- For each potential noisy band *k*
  - Find the preceding nearest signal band (*k-m*)
  - Find the succeeding nearest signal band (*k+n*)
  - If $\rho_{k-m,k} > thres$ OR $\rho_{k,k+n} > thres$

    *k*th band is reset to a signal band

    Else

    *k*th band is a noisy band.

III. Classification

As for the classification, the unsupervised *k*-nearest neighbor algorithm (Anderberg, 1973) is implemented in accordance with two distance measures: the Euclidean Distance ED (Eq. 2) and the Spectral Angle Measure, SAM (Yuhas *et. al*, 1992) (Eq. 3). The Euclidean distance takes into consideration the brightness difference between two spectra while SAM is invariant to the brightness so that it can eliminate the influence of the shading effect.

$$ED(\mathbf{x},\mathbf{y}) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \tag{2}$$

$$SAM(\mathbf{x},\mathbf{y}) = \arccos\left(\frac{\langle \mathbf{x},\mathbf{y}\rangle}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}\right) \tag{3}$$

In an attempt of combining good points of each individual measure, a hierarchical scheme of using the above measures is proposed that operates as the diagram in Fig. 3: firstly using one measure to cluster the original image, and then use the other to further cluster each of the classes obtained.

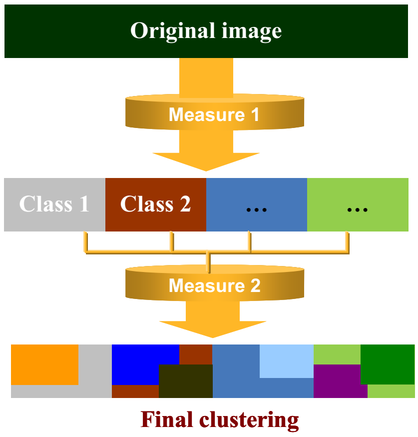


Fig. 3. Diagram of the proposed hierarchical scheme

IV. Assessment of Separability

Results of classification are assessed via the separability criteria (Fukunaga, 1990) based on a family of functions of scatter matrices, namely the within-class and between-class scatter matrices. The within-class scatter matrix indicates the scatter of samples around their respective class expected vector, defined as:

$$\begin{aligned} S_w &= \sum_{i=1}^{L} P_i E\left\{(\mathbf{X}-M_i)(\mathbf{X}-M_i)^T \mid \omega_i\right\} \\ &= \sum_{i=1}^{L} P_i \Sigma_i \end{aligned} \tag{4}$$

Σ_i : Covariance matrix of class i

P_i : A priori probability of ω_i

whereas the between-class scatter matrix indicates the scatter of the expected vectors of classes around the mixture mean:

$$S_b = \sum_{i=1}^{L} P_i \left(M_i - M_0\right)\left(M_i - M_0\right)^T$$
$$M_0 = E\{\mathbf{X}\} = \sum_{i=1}^{L} P_i M_i \qquad (5)$$

In order to give out a number for assessing the separability, the above matrices are associated as:

$$s = \frac{trS_b}{trS_w} \qquad (6)$$

trA : trace of an n × n matrix A

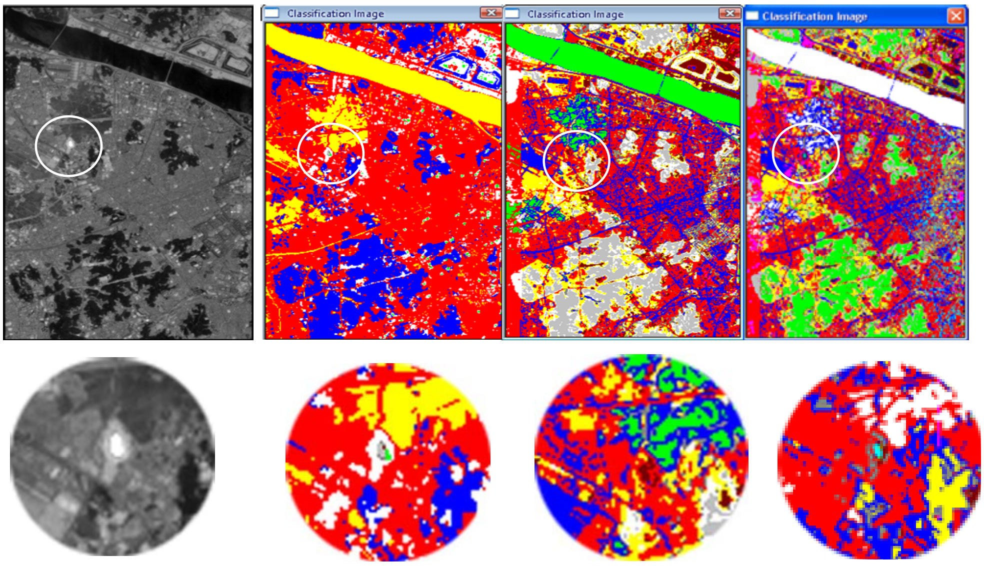
$$trA = a_{11} + a_{22} + \ldots + a_{nn} = \sum_i a_{ii}$$

so that the number would be larger (more separable or better classification) when the between-class scatter is larger or the within-class scatter is smaller.

V. Experiments

The unsupervised *k*-mean clustering is implemented so as to validate the effectiveness of the noisy band removal and the proposed hierarchical classification scheme. Due to the characteristics of our test data, the classification without removing noisy bands brings out a meaningless result. After removing the noisy bands, 135 out of 224 bands are remaining for subsequent use of classification. By controlling the threshold that is given in the section II, the level of noise to be removed can be controlled also. By observations, the removed bands are all noisy. However, in the mathematic view, the signal-to-noise ratios (SNR) of bands, defined as the ratio of the mean pixel value to the standard deviation of the pixel values, are not all of smallest due to the contrast of each band. After clustering, classes that contain less than 20 pixels are discounted from further processes.

For a thorough evaluation, we show the results of classification in two views: the observation view and the mathematic view. As shown in Fig 4 below, if the result of using ED measure can distinguish the bright spot caused by a light, it can not recognize the bridges over the river or the road maze in the scene. This happens because as mentioned above, the ED measure takes into account the brightness, therefore it is sensitive to the change of illumination. An opposite result can be found using the SAM measure which is invariant to the illumination. The proposed scheme can give a better result of successfully combining the strength of measure.



a) b) c) d)

Fig. 4. Results of classification using different measure. a) a signal band and results of classification b) using ED measure; c) using SAM measure; and d) using the hierarchical scheme

Separability scores are shown in Fig. 5. The experiment settings accord to the measures used and the order of using measures in the hierarchical scheme. In the first run, the ED measure is utilized; for the second, only use of SAM. The next three ones are performed in the order of using ED as the measure 1 in Fig. 3 then SAM as the measure 2. We in turn fixed the threshold of SAM and controlled the number of clusters by changing the threshold of ED; or fixed the ED or in the third case, controlled the number of clusters by changing both measures simultaneously. For the last case, the order of using measures is reversed: using SAM as the measure 1 in Fig. 3 then ED as the measure 2. Experiments show that the results obtained from the proposed hierarchical scheme give better classification in term of separability.

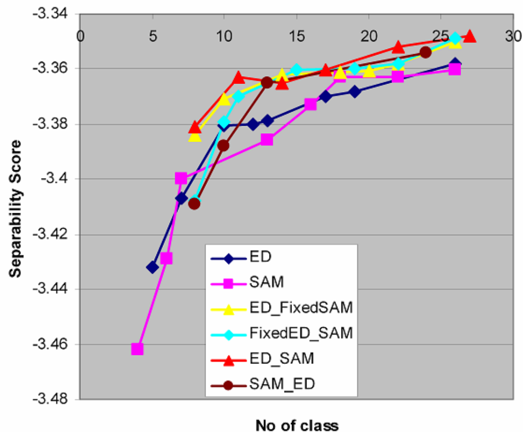


Fig. 5 Plot of separability scores at different settings

VI. Conclusions and Future Works

Noise in hyperspectral image is sometimes unavoidable that can divert the result of surface classification. The method proposed makes use of the correlation coefficients in a screening scheme to determine noisy bands. Unlike other methods that are based on the SNR as a lodestar, the proposed method is invariant to the contrast of bands. Due to the lack of a ground truth map, the separability criterion is used to assess the performance of the unsupervised clustering. The experiments show that the ED measure gave the worst classification and created the most virtual classes (classes that contain just a few pixels).

As to the future works, we intend to expand the current method in the frequency domain by considering the Fourier spectrum of bands. In addition, the performance of classification at different levels of noise removed also need to be evaluated.

Acknowledgement

This research was supported by the Agency for Defense Development, Korea, through the Image Information Research Center at Korea Advanced Institute of Science & Technology

References

AVIRIS – General Overview, NASA http://aviris.jpl.nasa.gov/. Accessed 27 Jul 2005.

Ben-Dor, E., N. Levin, and H. Saaroni, 2001. A spectral based recognition of the urban environment using the visible and near infrared spectral region. A case study over Tel Aviv, Israel, *International Journal of Remote Sensing*, 22(11): 2193-2218.

Fukunaga, K. *Introduction to Statistical Pattern Recognition*, Academic Press, 1990, 445-448.

Goetz, A. F. H., 1991. Imaging spectrometry for studying Earth, Air, Fire and Water, *EARSeL Advances in Remote Sensing*, 1: 3-15.

Green, A.A., Berman, M., Switzer, P., & Craig, M.D. 1988. A transformation for ordering multispectral data in terms of image quality with implications for noise removal: *IEEE Transactions on Geoscience and Remote Sensing*, Vol. *26*(1), 65-74.

Kim, S. H., S. J. Kang, J. H. Chi, and K. S. Lee, 2007. Absolute atmospheric correction procedure for EO-1 Hyperion data using MODTRAN code, *Korean Journal of Remote Sensing*, 23(1): 7-14

R. Ian Faulconbridge, R.I., Mark R. Pickering, M. R., and Michael J. Ryan M. J. 2006. Unsupervised band removal leading to improved classification accuracy of hyperspectral images: *Proceedings of the 29th Australasian Computer Science Conference - Volume 48, Hobart, Australia, 43 – 48*

Yuhas, R.H., Goetz, A. F. H., and Boardman, J. W., 1992, Discrimination among semiarid landscape endmembers using the spectral angle mapper (SAM) algorithm. In Summaries of the Third Annual JPL Airborne Geoscience Workshop, JPL Publication 92-14, vol. 1, pp. 147-149.

van der Meer, F. and S. de Jong, 2001. *Imaging Spectrometry: Basic Principles and Prospective Applications, Kluwer Academic Publishers*.