DIMENSIONALITY REDUCTION OF HYPERSPECTRAL IMAGES FOR COLOR DISPLAY USING SEGMENTED INDEPENDENT COMPONENT ANALYSIS

Yingxuan Zhu, Pramod K. Varshney, and Hao Chen

Department of Electrical Engineering and Computer Science, Syracuse University, {yzhu03, varshney, hchen21}@syr.edu

ABSTRACT

The problem of dimensionality reduction for color representation of hyperspectral images has received recent attention. In this paper, several independent component analysis (ICA) based approaches are proposed to reduce the dimensionality of hyperspectral images for visualization. We also develop a simple but effective method, based on correlation coefficient and mutual information (CCMI), to select the suitable independent components for RGB color representation. Experimental results are presented to illustrate the performance of our approaches.

***Index Terms*—** Hyperspectral imaging, ICA, visualization, dimensionality reduction

1. INTRODUCTION

There has been a growing interest in the acquisition of hyperspectral remotely sensed imagery. Hyperspectral imagery is available in hundreds of bands rather than a few bands as in multispectral imagery. Consequently, it is desirable to develop computational techniques to be able to process the high-dimensional data in a computationally efficient manner that yields accurate results. There has been a great deal of research on dimensionality reduction of hyperspectral data for image analysis tasks such as classification and change detection [1-3]. In this paper, we consider the problem of dimensionality reduction of hyperspectral images for color display. This is an important problem as this will enable humans to make decisions based on image data. Naturally, the goal is to represent the high-dimensional dataset in three dimensions with as little loss of information as possible, and should be easy for the humans to understand and visualize.

There has been some recent work on this problem [4-7, 11]. A partitioned principal component transformation (PCT) method was proposed in [4] to reduce the number of bands for displaying the hyperspectral remote-sensing images. Tsagaris, et al. also used the idea of segmented PCT method where they employed three different methods for segmentation of bands and compared their performance for color representation [5]. The principal-components-based display strategy is also used in [6] for spectral imagery in the HSV color space. Jacobson and Gupta discussed the design goals for such systems and provided a display method using fixed linear spectral weighting envelopes in [7].

It is well known that PCT based methods for signal/image processing work well if the signals can be modeled accurately in terms of their second-order statistics. However, it has been observed that accurate models for hyperspectral datasets require higher-order statistics and dimensionality reduction methods based on independent component analysis (ICA) have been proposed [1,2]. The goal of these ICA based methods is to mostly improve classification. In this paper, we develop ICA based methods for dimensionality reduction suitable for the color display application. It is anticipated that the ICA based approach will result in a much better color display as it is based on a more accurate model that takes into account higher-order statistics. Three different segmentation approaches are used to partition the hyperspectral bands prior to applying ICA. All three segmentation approaches are based on the characteristics of hyperspectral images. Their performances are illustrated by means of an example.

2. ICA BASED DIMENSIONALITY REDUCTION

2.1. Independent component analysis

Independent component analysis was originally developed to separate individual signals from linear mixtures of the signals. These mixture signals can be simply written as,

$$x = As \tag{1}$$

where x is the random vector of the mixture signals, s is the vector of original signals, and A is the mixing matrix. Setting up the statistical model to find out the original signals is called independent component analysis, or ICA [8].

The general idea behind ICA is to try to estimate the inverse matrix of $A, W = A^{-1}$, and reconstruct s by $s = W$x, while having a high fidelity of the original vector s. In order to accomplish this goal, several algorithms have been developed according to different statistical criteria. These criteria include kurtosis and negentropy. Detailed discussion on ICA can be found in [8, 9]. Due to its computational advantage, the experiments in this paper are all based on

FastICA. The Matlab™ implementation of the algorithm used is available in [10].

2.2. Selection of ICA components

Unlike principal component analysis, the order of the independent components after each application of ICA is generally not the same. To the best of our knowledge, there are no specific methods to determine the rank or ordering of the components. It is done in an ad hoc manner and is application specific. In essence, here we want to reduce the dimensionality and choose the components which may contribute the best quality to our application namely the final display. In this paper, we choose those components which provide the most information with respect to the original data set. We employ a measure based on both correlation coefficient and mutual information to find the independent components which have strong relationships with the original bands.

Specifically, assume that the hyperspectral image dataset is an $M*N*L$ matrix H, where L is the number of spectral bands. The FastICA algorithm is applied to H, and the resulting output Matrix A consisting of the independent components is also an $M*N*L$ matrix. The three RGB components need to be selected from these L independent components. The correlation between the i^{th} band image in H and the j^{th} independent component in A can be written as:

$$C_{ij} = \frac{\sum_m \sum_n (H_{mni} - \overline{H_i})(A_{mnj} - \overline{A_j})}{\sqrt{(\sum_m \sum_n (H_{mni} - \overline{H_i})^2)(\sum_m \sum_n (A_{mnj} - \overline{A_j})^2)}} \tag{2}$$

with $\overline{H_i}$ and $\overline{A_j}$ indicate the means of the i^{th} and j^{th} images in H and A, respectively. We will disregard the sign of C_{ij} and will employ its absolute value.

The mutual information is a measure of statistical dependency between two images. The frequency of occurrence of a pixel intensity value corresponds to the probability of that value, $p_{H_i}(x)$ and $p_{A_j}(y)$. Similarly, $p_{H_iA_j}(x,y)$ represents the joint probability of a pair of pixel intensities. Finally, the mutual information of the i^{th} image in H and j^{th} component in A is defined as

$$I_{H_iA_j} = \sum_y \sum_x p_{H_iA_j}(x,y) \log \frac{p_{H_iA_j}(x,y)}{p_{H_i}(x)p_{A_j}(y)} = H_{H_i} + H_{A_j} - H_{H_iA_j} \tag{3}$$

where H_{H_i} and H_{A_j} are entropies of H_i and A_j, respectively. $H_{H_iA_j}$ is the joint entropy of H_i and A_j.

2.3 Dimensionality reduction approach

We first apply ICA to the given hyperspectral dataset. Once the independent components have been obtained, we need to select three components for display purposes. In our work reported here, the components which have the most information of the original hyperspectral images are selected. The proposed idea is realized by using both the correlation coefficient (CC) and the mutual information (MI) criteria for selection. We call this approach ICA-CCMI. The main idea of ICA-CCMI is summarized in the following steps.

1) Apply FastICA to the original dataset to obtain A.

2) Use Equation (2) to calculate the correlation coefficient between every pair of independent components and hyperspectral bands. The average correlation of the j^{th} component in A with hyperspectral image H is defined as:

$$\overline{C_j} = \sum_i |C_{ij}| / L \tag{4}$$

The larger the average correlation coefficient value of an independent component, the closer the relationship of this component with the original hyperspectral image dataset.

3) Similar to step 2), compute the average mutual information for the j^{th} components of A as:

$$\overline{I_{A_j}} = \sum_i I_{H_iA_j} / L \tag{5}$$

Again larger the value of $\overline{I_{A_j}}$, more the information regarding the original dataset the j^{th} component has.

4) Find the maximum value of $\overline{C_j}$, called $\max \overline{C_j}$. Normalize each $\overline{C_j}$ using $\max \overline{C_j}$ as:

$$\overline{C_j}^N = \overline{C_j} / \max \overline{C_j} \tag{6}$$

5) Normalize each $\overline{I_{A_j}}$ using $\max \overline{I_{A_j}}$ as:

$$\overline{I_{A_j}}^N = \overline{I_{A_j}} / \max \overline{I_{A_j}} \tag{7}$$

6) Find the product[1] of $\overline{C_j}^N$ and $\overline{I_{A_j}}^N$, and then the selection criterion for independent components is:

$$CCMI_j = \overline{C_j}^N \times \overline{I_{A_j}}^N \tag{8}$$

Component(s) with the largest value(s) of $CCMI_j$ are to be selected for display.

The whole procedure is implemented in Matlab on a computer with CPU 3.0 GHz, RAM 1G, and Windows XP. It takes around 55 seconds for a 512*512*21 dataset, and about 380 seconds for a 512*512*63 dataset.

3. SEGMENTATION APPROACHES

One may employ ICA-CCMI to the entire hyperspectral dataset and select the three bands with largest values of $CCMI_j$ for display purposes. However, there are two obvious drawbacks when ICA-CCMI is applied to the entire dataset. One is the huge computational burden, and the other is that

[1] In this paper, we employ the product to combine the two individual selection criteria. Other combination methods are also possible and are currently being investigated.

the use of all the bands may not reveal the information in certain bands. To alleviate this situation we consider segmentation or partitioning of the bands prior to the application of ICA. We describe partitioning approaches based on a specific dataset. The 126 bands (435-2489 nm) hyperspectral data used in this paper were acquired from HYMAP. They are divided into three subgroups by different partitioning approaches according to specific criteria.

Segmented ICA based on equal subgroups:

This segmentation method is simple and direct. It divides the total number of bands into three equal size subgroups. In our experiment, the first subgroup covers the region from band 1st to band 42nd (435 - 1047nm); the second subgroup is from band 43rd to band 84th (1062 - 1684 nm); and the third subgroup contains the remaining 42 bands (1697-2489 nm). Theoretically, the equal subgroup partitioning approach can significantly reduce the computational load, but there is no special information included in each subgroup for further image analysis. After the application of ICA-CCMI to each subgroup, the most suitable component from each subgroup is selected to construct the final RGB image.

Segmented ICA based on correlation coefficient:

The significant degree of correlation between bands results in a high degree of spectral redundancy [5]. Figure 1 shows the correlation coefficients between every pair of spectral bands for the dataset, where the white points represent high degree of correlation and dark points represent less correlation between two bands. According to the correlation values in this figure, the 126 bands can be separated into three subgroups where bands in each subgroup exhibit high correlation amongst each other. The first subgroup is from band 1st to band 21st (435 - 740 nm), the second subgroup covers the region from band 22nd to band 63rd (740 - 1405 nm), and the third subgroup is from band 64th to band 126th (1420 - 2489 nm). These three subgroups can be handled as described earlier.

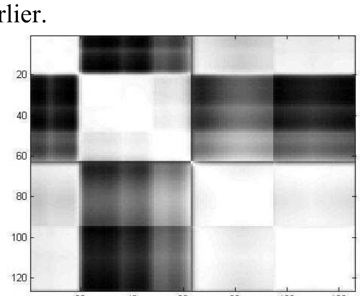


Figure 1. Correlation coefficient matrix of bands

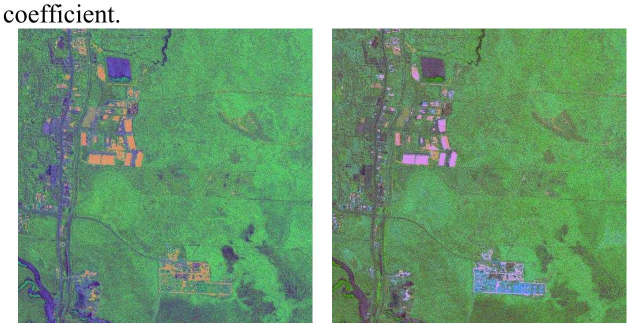
Segmented ICA based on RGB spectrum:

According to the human visual system and the region of the electromagnetic spectrum, three subgroups can be selected from the 126 bands to represent the blue, green and red regions of the visible part of the electromagnetic spectrum [7]. The first subgroup is from band 1st to band 5th (435-495 nm), the second subgroup covers the regions from band 6th to band 12th (511-603 nm) and the third subgroup is from band 13th to band 19th (619-710 nm). These three subgroups represent the regions of blue, green and red, respectively. In particular, instead of using all of the bands, this partitioning method just deals with the bands in the spectrum regions of blue, green and red, and ICA-CCMI based methodology is applied to these subgroups.

4. EXPERIMENTAL RESULTS

The partitioning approaches discussed above describe three ways in which one could divide the hyperspectral bands. In this section, the RGB visualization results are presented in Figures 2 - 4 to compare different segmented ICA approaches. As discussed above, we select one of the independent components from each subgroup to display the final RGB image. But there is no specific criterion to decide which of the three independent components will be mapped to the R component, the G component, or the B component, except for the segmented ICA approach based on RGB spectrum. So any set of selected independent components can be mapped to R, G, or B colors. Here we show results for one arbitrary mapping between independent components and R, G, and B colors.

Figure 2(a) shows the RGB display of segmented ICA based on equal subgroups. In our experiment, even though different combinations of the selected independent components have different RGB visualization results, the performance in terms of the structures and edges is relatively unaffected by different combinations. Figure 2(b) illustrates the RGB display of segmented ICA based on correlation coefficient.



(a) (b)

Figure 2. Display results of segmented ICA
(a) equal subgroups and (b) correlation coefficient

Figure 3 shows the results of segmented ICA and PCA based on RGB spectrum, where the B component is from the first subgroup, the G component is from the second subgroup, and the R component is from the third subgroup. As opposed to using ICA-CCMI algorithm in the selection of components, PCA representation is formed by the first principal components of each subgroup. The true color representation from the information provided in the HYMAP dataset is shown in Figure 4. Based on Figures 3 and 4, the performance of ICA is much closer to the true color image than that of PCA with RGB spectrum partitioning method when they are displayed in color.

(a) (b)

Figure 3. Display results of segmented ICA based on RGB spectrum, (a) ICA, (b) PCA

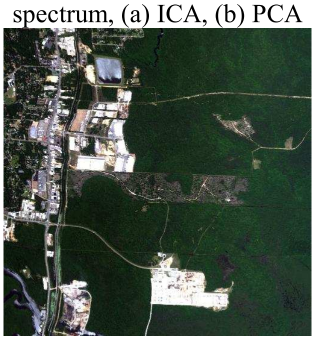


Figure 4. The true color image of the hyperspectral dataset

Entropy can be used to represent the information content of each component and, therefore, the quality of representation [5]. The entropies of each selected component of ICA and PCA visualization are calculated and compared in TABLE I. The RGB components of PCA visualization are the first principal components of each subgroup, which have most of the energy of that subgroup. Obviously, the segmented ICA based on correlation coefficient has larger entropy values than the segmented PCA method.

TABLE I ENTROPY FOR DIFFERENT PARTITIONING APPROACHES

Entropy	Subgroup 1		Subgroup 2		Subgroup 3	
Partitioning approaches	ICA	PCA	ICA	PCA	ICA	PCA
Equal Subgroup	6.840	6.528	7.011	7.163	6.792	6.728
Correlation coefficient	7.472	6.440	7.506	6.541	7.14	6.41

5. CONCLUSIONS

In this paper, an effective dimensionality reduction method for color displays, ICA-CCMI, is proposed for hyperspectral images. Our method is based on segmentation of the hyperspectral data bands and the application of ICA on each segment. Three partitioning approaches are used in our experiment, based on equal subgroups, correlation coefficients and RGB electromagnetic spectrum, respectively. Correlation coefficient and mutual information are used to select the independent component from each segment for color display. The most salient feature of the ICA-CCMI is that it does not require training data, or determination of any parameters for the independent components and the hyperspectral images. ICA-CCMI is applied to a real hyperspectral dataset, and the display results show that the ICA-CCMI algorithm is a practical and attractive method for dimensionality reduction.

There are a number of issues that are worth considering in future work. For example, additional statistical and quantitative evaluation metrics are necessary to thoroughly evaluate the display results, and extensive experimentation with other datasets is desirable.

6. REFERENCES

[1] C. I Chang, *Hyperspectral Imaging: Techniques for Spectral Detection and Classification,* Kluwer Academic/Plenum Publishers, New York, 2003.

[2] P. K. Varshney and M. K. Arora (eds.), *Advanced Image Processing Techniques for Remotely Sensed Hyperspectral Data*, Springer Verlag, 2004.

[3] Special issues on hyperspectral image analysis*, IEEE Signal Processing Magazine*, vol. 19, no. 1, 2002

[4] X. Jia and J. A. Richard, “Segmented Principal Components Transformation for Efficient Hyperspectral Remote-Sensing Image Display and Classification”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 37, No. 1, pp. 538-542, January 1999.

[5] V. Tsagaris, V. Anastassopoulos and G. A. Lampropoulos, “Fusion of Hyperspectral Data Using Segmented PCT for Color Representation and Classification”, *IEEE Transactions on Geoscience and Remote Sensing*, vol.26, no.10, pp. 2365-2375, Jan. 2005.

[6] J. S. Tyo, A. Konsolakis, D. I. Diersen, and R. C. Olsen, “Principal-Components-Based Display Strategy for Spectral Imagery”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 41, No. 3, pp. 708-718, March 2003.

[7] N. P. Jacobson and M. R. Gupta, “Design Goals and Solutions for Display of Hyperspectral Images”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 43, No. 11, pp. 2684-2692, Nov. 2005.

[8] A. Hyvärinen and E. Oja, “Independent component analysis: algorithms and applications”, *Neural Networks*, vol. 13, Issue 4, pp. 411--430, 2000.

[9] A. Hyvärinen, “Survey on Independent Component Analysis”, *Neural Computing Surveys*, 2, pp. 94-128, 1999.

[10] [Online]. Available: http://www.cis.hut.fi/projects/ica/fastica/.

[11] J. Wang and C.-I. Chang, “Applications of Independent Component Analysis in Endmember Extraction and Abundance Quantification for Hyperspectral Imagery”, *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 44, No. 9, pp. 2601-2616, Sep. 2006