

DIMENSIONALITY REDUCTION TECHNIQUES FOR ANALYSIS OF HYPERSPECTRAL DATA

A.Kiranmai¹, Y.Sai Praveen², Iyyanki. V. Murali Krishna³, V.Sowmya Devi⁴

¹Dept of CSE, GITAM University, Hyderabad

²Dept of CSE, GITAM University, Hyderabad

³Dr. Raja Ramanna Distinguished fellow, Research center Imarat, DRDO

⁴Assistant Professor, Dept of CSE, GITAM University, Hyderabad

Abstract

The Hyperspectral images provide the images with hundreds of narrow contiguous spectral channels. The spectral information provided by the hyperspectral images is high when compared to the other class of remote sensing images such as panchromatic, multispectral images. Even though the hyperspectral images contain sufficient spectral information but processing and exploiting the hyperspectral data is considered as a challenging task because of its high dimensionality. Data redundancy is one of the problems when processing hyperspectral data. And some bands in hyperspectral images are noisy and some bands are set to zero. So when we use all the bands for processing it increases the computational complexity in terms of storage and processing time. Because of this problem Dimensionality reduction techniques must be applied before we use it for processing. With that we can map high dimensionality of hyperspectral image data to lower dimensions without losing much of spectral features provided by the original hyperspectral data cube. Many techniques are developed for this dimensionality reduction; three of those techniques are analyzed in this paper.

Keywords: Dimensionality Reduction, Hyperspectral, Band Selection, Principal Component Analysis, Minimum Noise Fraction, Independent Component Analysis.

1. INTRODUCTION

In remote sensing the hyperspectral imaging and exploiting the hyperspectral data became the one of the most active research topic. Based on the number of bands, the imaging sensors in remote sensing are mainly categorized into Panchromatic and multispectral and hyperspectral. The panchromatic image contains only a single band which is a grey scale image with high spatial resolution; the multispectral image contains less than ten bands; a Hyperspectral image contains hundreds of spectral channels. The hyperspectral sensor captures the image in very narrow wavelength along the electromagnetic spectrum. It is represented as a three dimensional data cube. The x, y axis represents the spatial property of the captured image whereas the third dimension that is z axis represents the spectral property of the image in other words it represents the number of spectral channels.

Each object on the earth has a unique spectral signature. In general the regions covered by hyperspectral sensors are Visible, near infrared and SWIR regions of the electromagnetic spectrum containing hundreds of spectral channels. So every pixel contains contiguous reflectance spectrum. So the information provided by these hyperspectral images is enough to uniquely identify different materials. In hyperspectral remote sensing the processing and analysis is a challenging task. Because out of all the bands that hyperspectral image contains only some specific number of bands is useful and contains valid information and most of

the bands in these images contain redundant information [4]. Some bands may get affected by the atmospheric effects such as atmospheric water absorption which provide no valid information but increase the dimensionality of data. So if we consider all the bands for processing in remote sensing applications it not only increases the computational complexity but also affects the accuracy of the results (if noisy bands are used in processing). When it comes to processing of Hyperspectral data cost of the storage and volume is another problem. When we use all spectral channels of hyperspectral data for processing the resultant data is very huge and it also takes lots of time to process. So in order to reduce these undesired effects we must consider reducing this high dimensionality problem of the hyperspectral data. By using the dimensionality reduction technique we can eliminate these noisy, redundant data.

2. LITERATURE

Before we use the hyperspectral data for any remote sensing application, the dimensionality reduction techniques must be considered as a preprocessing step. By applying dimensionality reduction techniques we can eliminate the highly redundant data without losing much significant spectral information [5]. In general, the approaches used for dimensionality reduction can be two types. One is based on transformation technique another is band selection technique [1], [2]. Each method is having its own advantages and disadvantages.

First, in the transformation based techniques, it transforms the higher dimensions of the hyperspectral data to lower dimensions by using some mathematical function. It also takes some criterion function based on that only the transformation technique produces a particular subset as a result of transformation. Various transformation techniques are developed for dimensionality reduction each method is having different criterion function [7]. Principal Component analysis (PCA) is one of the popular dimensionality reduction technique used to analyze the remotely sensed data, which come under this category. It is linear feature transformation of dimensionality reduction technique. The main advantage of this type of techniques is their efficiency and speed.

In band selection based dimensionality reduction technique we select only a subset of the total number of bands which is having more features, which can cover maximum of the information provided by the original hyperspectral data cube. The band selection based technique can be supervised or unsupervised [6]. Like transformation techniques band selection methods also have some criterion function. By considering this criterion function these methods select a subset of bands which have significant information and select the subset of bands which less correlation or redundancy compared to the other bands in the selected subset.

3. STUDY AREA AND DATASETS

Study area is part of Hyderabad city. It falls in the longitude and latitude of 17.3700° N and 78.4800° E. The input dataset is Hyperion which is a hyperspectral image. The Hyperion image is having 30 meter spatial resolution and having 242 bands and 216X539 pixels. The dataset is atmospherically and radiometrically corrected. The hyperspectral data is geometrically corrected such that the data sets have standard geometrical projections.



Fig. 1: Study Area

4. PREPROCESSING OF DATA

The images collected hyperspectral sensor cannot be applied remote sensing applications directly. The hyperspectral data that we get can get affected by the atmospheric affects and some of the bands may contain only noise. So we need to remove those bands form processing. So preprocessing of the hyperspectral images is necessary. We have applied following preprocessing steps to our data set.

4.1 Bad Band Removal

The hyperspectral data collected by the sensor contains some zero and noisy bands which do not contain any valid information and it may decrease the accuracy of the feature extraction and classification results. By identifying and removing these bad bands from the data necessary.

4.2 Vertical Stripe Removal

Some bands in the hyperspectral images contain a vertical stripe which is known as “striping” effect. These are black lines in the images which do not contain any useful information. It can be corrected by weighted averaging technique. In this we take the average of neighbour pixel DN values and we replace this this value with affected column pixel values.

4.3 Atmospheric and Radiometric Correction

Before the reflectance from the earth surface reached by the sensor this signal get affected by the various atmospheric effects such as water vapour absorption, aerosols which attenuate the signal. Atmospheric Correction of these effects is required in order to get surface correct reflectance from imaging radiance data [11]. Radiometric correction is applied to the remotely sensed data to improve the fidelity of the brightness value.

5. DIMENSIONALITY REDUCTION TECHNIQUES

Hyperspectral images provide significant amount of spectral information which is enough for discriminating the two different materials. But these hyperspectral data contains more redundant information with respect to other bands, which increases the computational complexity in terms of Time and Size. For this reason, it is necessary to reduce the data dimensionality. Various techniques have been developed for reducing this dimensionality problem. The techniques we considered here are Principal Component Analysis and Minimum Noise Fraction and Independent Component Analysis.

5.1 Principal Component Analysis (PCA)

PCA is a dimensionality reduction technique. It is based on an orthogonal transformation. In this the components produced by Principal Component Analysis are having highest variance. It removes the correlation among the bands[8]. The PCA technique reduces the dimensionality by producing the components with high variation and less correlation. The noisy bands are segregated that they are having very less Eigen value.

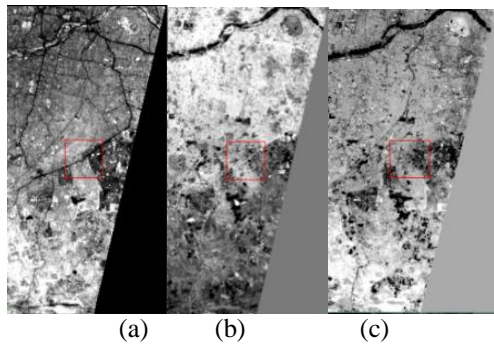


Fig. 2: PCA bands (a) PCA 1 (b) PCA 2 (c) PCA 3

The first PC band contains the largest percentage of data variance and the second PC band contains the second largest data variance, and so on.

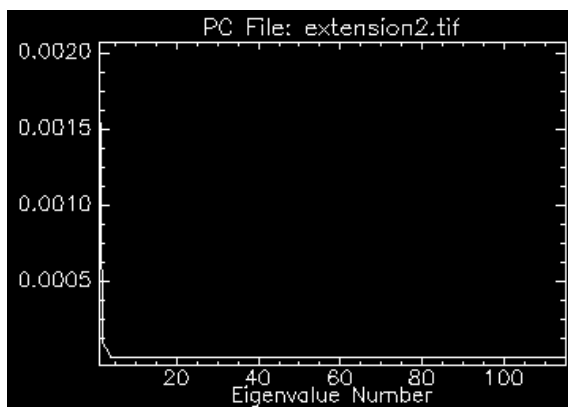


Fig. 3: Spectral Profile of PCA

Table 1 show that the first band contains the highest Eigen value and the second band contains the second highest Eigen value. The last PC bands are contains more noise and very little variance compared to other bands. By comparing the Eigen values of the PCA bands we select only those bands which are having highest Eigen values compared to the all others. By increasing the number of bands we select for processing we can increase the accuracy of the results. Figure 2 represents first three bands of PCA. From the table 1 we can observe that the Eigen value of PCA 1 is having highest value.

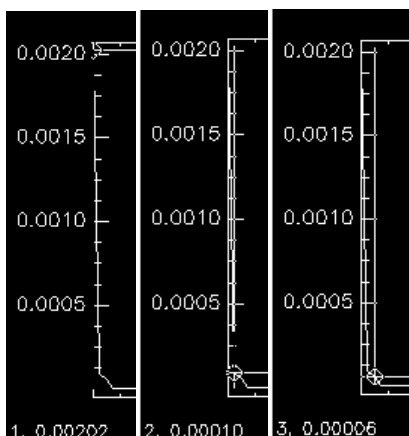


Fig.4: Eigen value of PCA bands

Table 1: First Five Eigen Values Of Pca Transformed Image

Basic Stats	Eigen Value
Band 1	0.002022
Band 2	0.000095
Band 3	0.000059
Band 4	0.000008
Band 5	0.000003

5.2 Minimum Noise Fraction

The Minimum noise fraction is another popular technique for dimensionality reduction of hyperspectral images. It is a linear transformation technique. It is having two PCA rotations. In the first rotation it uses covariance matrix to de correlate and segregate the nose in the data in other words it first computes the noise statistics of the hyperspectral data. In the second rotation it uses the noise statistics which are computed in the first rotation and the components and then produces the principal components.MNF is a technique to address the noiseand tries to maximize the Signal to Noise Ratio[10].

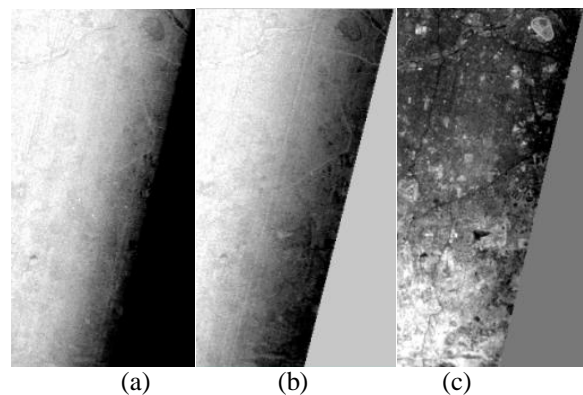


Fig.5: MNF bands (a) MNF 1 (b) MNF 2 (c) MNF 3

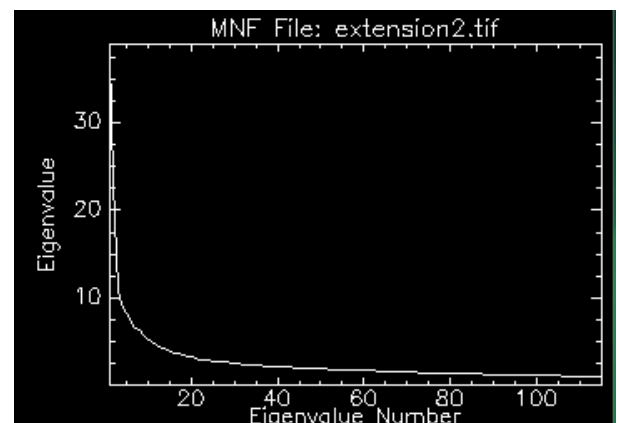


Fig. 6: Spectral Profile of MNF transformed image

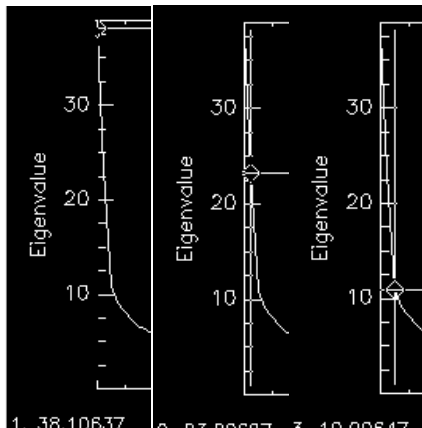


Fig. 7: Eigen value of MNF bands

Table 2: First Five Eigen Values Of Mnf Transformed Image

Basic Stats	Eigen Value
Band 1	38.106373
Band 2	23.206971
Band 3	10.906469
Band 4	9.153154
Band 5	8.201513

5.3 Independent Component Analysis

Independent Component Analysis (ICA) is a technique which is based on the assumption that each band in the input data that we consider for processing is an independent component. In this transform it extracts the interesting features using linear unmixing technique and it is based on the non-Gaussian assumption [9].

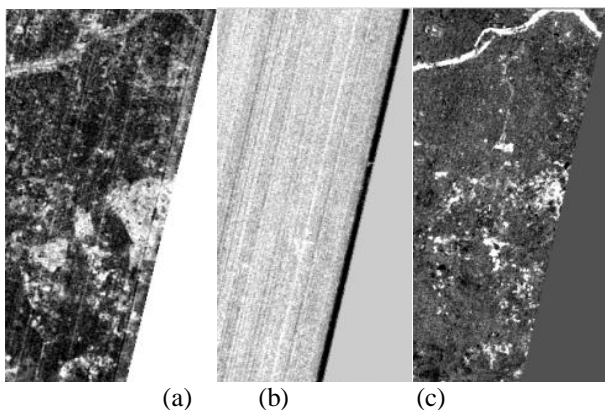


Fig. 8: ICA bands (a) ICA 1 (b) ICA2(c) ICA 3

ICA is a technique which is used to extract the end members, independent components from the input hyperspectral dataset. In figure 8 the first three bands of ICA are shown. The Eigen values of the ICA transformed image are not like the PCA or MNF. The resulted components are independent components and extract the independent features in the dataset.

7. CONCLUSION

In this paper we have studied three popular dimensionality reduction techniques and the results are evaluated. We have taken Eigen value as a measure to evaluate the feature information in the transform bands. It is observed that the accuracy of the application (for example classification, target detection etc.) result increases linearly with the increase of the number of bands. And it is also observed that the computational cost of Independent Component Analysis is more than the Principal Component analysis Technique as it treats each band as independent component. PCA has shown better performance than ICA. It is observed that PCA has covered hundred per cent of the original band information in 81 bands itself. While transformation the PCA technique only considered variance it does not involve in reducing the noise in the bands. Minimum noise fraction has also shown good results unlike PCA, it mainly can be used in truncating noise contained in hyperspectral data.

REFERENCES

- [1] Yang, He, et al. "An efficient method for supervised hyperspectral band selection." *Geoscience and Remote Sensing Letters*, IEEE 8.1 (2011): 138-142.
- [2] Pal, Mahesh, and Giles M. Foody. "Feature selection for classification of hyperspectral data by SVM." *Geoscience and Remote Sensing*, IEEE Transactions on 48.5 (2010): 2297-2307.
- [3] Koonsanit, Kittit, ChuleeratJaruskulchai, and ApisitEiumnoh. "Band selection for dimension reduction in hyper spectral image using integrated information gain and principal components analysis technique." *International Journal of Machine Learning and Computing* 2.3 (2012): 248.
- [4] Burgers, Kate, et al. "A comparative analysis of dimension reduction algorithms on hyperspectral data." *LAMDA Research Group* (2009): 1-23.
- [5] Chang, Chein-I., and HalehSafavi. "Progressive dimensionality reduction by transform for hyperspectral imagery." *Pattern Recognition* 44.10 (2011): 2760-2773.
- [6] Du, Qian, and He Yang. "Similarity-based unsupervised band selection for hyperspectral image analysis." *Geoscience and Remote Sensing Letters*, IEEE 5.4 (2008): 564-568.
- [7] Bruce, Lori Mann, et al. "Spectral reduction image processing techniques." *Geoscience and Remote Sensing Symposium*, 2003. IGARSS'03. Proceedings. 2003 IEEE International. Vol. 1. IEEE, 2003.
- [8] Rodarmel, Craig, and Jie Shan. "Principal component analysis for hyperspectral image classification." *Surveying and Land Information Science* 62.2 (2002): 115.
- [9] Wang, Jing, and Chein-I. Chang. "Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis." *Geoscience and Remote Sensing*, IEEE Transactions on 44.6 (2006): 1586-1600.
- [10] A. A. Green, M. Berman, P. Switzer and M. D. Craig, "A transformation for ordering multispectral data in

terms of image quality with implications for noise removal," in IEEE Transactions on Geoscience and Remote Sensing, vol. 26, no. 1, pp. 65-74, Jan 1988.

- [11] Gao, Bo-Cai, et al. "Atmospheric correction algorithms for hyperspectral remote sensing data of land and ocean." Remote Sensing of Environment 113 (2009): S17-S24.
- [12] Farrell Jr, Michael D., and Russell M. Mersereau. "On the impact of PCA dimension reduction for hyperspectral detection of difficult targets." Geoscience and Remote Sensing Letters, IEEE 2.2 (2005): 192-195.