proceedings

Proceedings

Comparison and evaluation of dimensionality reduction techniques for hyperspectral data analysis†

K. Nivedita Priyadarshini *, V. Sivashankari, Sulochana Shekhar and K. Balasubramani

[1] Department of Geography, School of Earth Sciences, Central University of Tamil Nadu, Thiruvarur – 610 005, Tamil Nadu, India

* Correspondence: nivi.darshini@yahoo.com; Tel.: +91-809-814-6588

† Presented at the 2nd International Electronic Conference on Geosciences (IECG 2019), 8-15 June 2019; Available online:

Published:

Abstract: Hyperspectral datasets provides explicit ground covers with hundreds of bands. Filtering contiguous hyperspectral datasets discriminates surface features potentially. Hence, number of spectral bands are minimized without losing original information called Dimensionality Reduction (DR). Redundant bands portray a fact that neighboring bands are highly correlated with each other sharing similar information. The benefit of utilizing dimensionality reduction is to slacken complexity of data during processing and transform original data to remove correlation among bands. In this paper, two DR methods like Principal Component Analysis (PCA) and Minimum Noise Fraction (MNF) are applied to AVIRIS NG dataset of Kalaburagi for discussion.

Keywords: Dimensionality reduction; PCA; MNF

1. Introduction

Hyperspectral sensor images serves for various applications of remote sensing containing spatial and spectral information about the surface of the earth. These imagery are acquired from both airborne and spaceborne platforms rendering data with high spectral resolution narrow contiguous bands that are redundant and complex for processing [1]. In order to reduce the complexity of the data, various transforms are applied to scale and adjust the image data thus eliminating noise from each band [2]. Transforms like Principal Component Analysis (PCA), Minimum Noise Fraction (MNF) and Independent Component Analysis (ICA) produces new components that are ordered by image quality. Such Dimensionality Reduction (DR) techniques are applied to raw datasets for effective noise removal and intense smoothening is employed for components with high noise and low signal content to each band of original data [3]. Further, techniques like Optimized MNF (OMNF) calculates Noise Covariance Matrix (NCM) via Spatial Spectral De – Correlation (SSDC) perform spectral unmixing producing components for better classification. The inherent dimensionalities of bands are resolved for visual inspection of the eigenvalues that are further utilized for endmember extraction process [4]. Determining appropriate and evident DR technique is still a tough task while handling hyperspectral imageries. Despite some disadvantages, PCA is a commonly used DR technique as it has significant effect on classification. Classification results are greatly influenced by MNF when the ground cover features exhibit homogeneity. Efficiency of the classifiers improves better with reduced components that reveal apparently informative bands [5, 6]. This paper compares and evaluates the performance of two defined DR methods namely PCA and MNF by interpreting the eigen values acquired while processing. The cumulative variances of reduced components are explored for bands showing high correlation to low order bands containing noise affecting the quality of the original hyperspectral imagery. Bands comprising definite and highly information can be further utilized for processing.

2. Material and methods

2.1. Dataset Used

AVIRIS NG (Airborne Visible/Infrared Imaging Spectrometer – Next Generation) hyperspectral dataset for a portion of Kalaburagi region, Karnataka with a total extent from 76°.04′ and 77°.42′ East longitude, and 17°.12′ and 17°.46′ North latitude is used in this study. AVIRIS NG provides spectrum of 432 bands ranging from a wavelength of 376 – 2500 nm. The dataset is atmospherically corrected to obtain reflectance spectrum of totally 425 bands which serves for detailed processing.

2.2. Methodology

The dimensionality reduction method applied for the dataset is explained in the following subsections.

2.2.1. Optimal band selection through dimensionality reduction

Determination of DR method is a challenging job for simplifying the process. For the selected hyperspectral AVIRIS NG dataset, PCA and MNF are applied to find the intrinsic dimensions that disclose urban areas clearly from the considered study region. From the correlation matrix computed and the derived eigenvalues, a scree plot is procured graphically for additional information. Through visual observation of eigenimages, the components that provide useful details with less noise that are ordered from high to low are taken into account. The appropriate band selection through PCA and MNF are briefly described below.

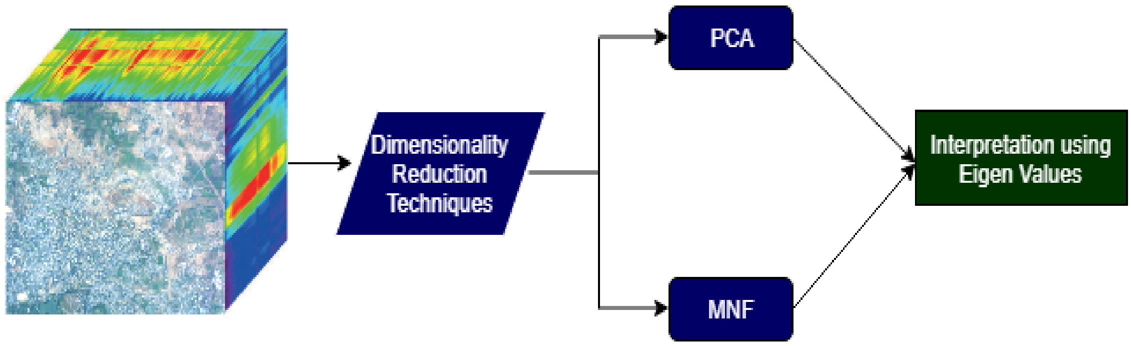


Figure 1. Workflow

2.2.2. Principal Component Analysis

PCA is a preprocessing linear transformation technique that reorganizes the variance from multiple bands to a new set of bands. Input bands processed after PCA provides uncorrelated bands that have maximized data variance. In this study, for the given hyperspectral data, statistics representing covariance and correlation matrix are computed on preference. The process of PC rotation normalizes the calculated matrices to unit variance and in order to equalize influence by bands, it reduces the bands with higher variance and vice versa [7]. In general, the PC finds the relationship between dependent and independent variable that prioritize them in which the first principal component spreads along a straight line which shows the longest dimension of the data and the second component is projected perpendicular to it thus fitting the error produced by the first component. On fitting the principal components to the variance of the data, eigenvalues and eigenvectors can be assessed [8]. The goal of the study is to reduce the n dimensional data that is projected into a k dimensional subspace in which $k < n$ is implemented through which informative bands are retained thus increasing computational accuracy. Bands ranging from λ_{50} = 621.87 nm to λ_{194} = 1343.11 nm, λ_{224} = 1493.37 nm to λ_{280} = 1773.86 nm and λ_{335} = 2049.34 nm to λ_{395} = 2349.86 nm, a total of 263 spectral bands where λ_k is k^{th} spectral band with its corresponding wavelength are chose by avoiding unessential bands. With respect to the eigenvalues, first 9 components that render information from the dataset are considered for analysis.

2.2.3. Minimum Noise Fraction

Noise characteristics changes from band to band and provides coarsely ordered bands as per image quality. MNF transform discussed in this paper, has the capability to provide optimal high to low order to the bands with respect to image quality. MNF defines how the spatial information can be used to interpret the covariance structure of the signal and noise. The transform segregates the image in the data and its inherent dimensionality is usually through two cascaded PC transformation. At first, noise covariance matrix is computed to decorrelate and rescale the noise from the data called 'noise whitening'. Later process is the eigen decomposition on the modified matrix to order the bands with respect to signal to noise. Higher eigenvalues (>1) contains more information in the bands and values near 1 render noisy datasets. Similar to the PCA, the top most components are filtered for the study and the rest are zeroed out [9]. A total number of 263 spectral bands considered for PCA were taken into account for applying MNF transform.

3. Results and Discussion

3.1. Transformation results

DR techniques retrieve information that can be further processed for endmember extraction and classification. From the results it is evident that, noise from the data is segregated although bands are ordered by decreasing signal to noise ratio. It is observed that first 9 components of each transform renders highly informative bands from the whole of eigenvalues displayed. Bands that comprise urban information are seen clearly in PCA transformed bands of 1, 3, 5 and 9 and MNF transformed bands of 1, 4, 7 and 8. Rest of the components also comprises details along with greatly enhanced noise patterns. It is noted that, there is much less noise concentration and correlation between the bands with shorter wavelengths. The remaining bands have a salt and pepper noise effect thus slackening the quality of the transformed image. Clean transformed band information are retrieved through linear transformation with respect to reflectance characteristics of each band providing noise – free components are chose for further data analysis processing. The Figures 2 and 3 shows the dimensionally reduced bands by applying PCA and MNF respectively.

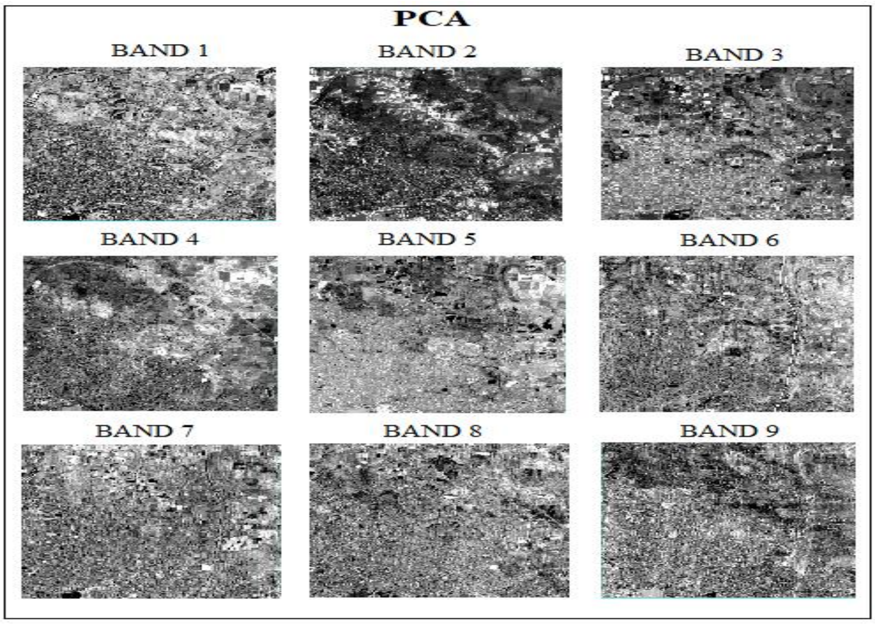


Figure 2. Principal Component Analysis

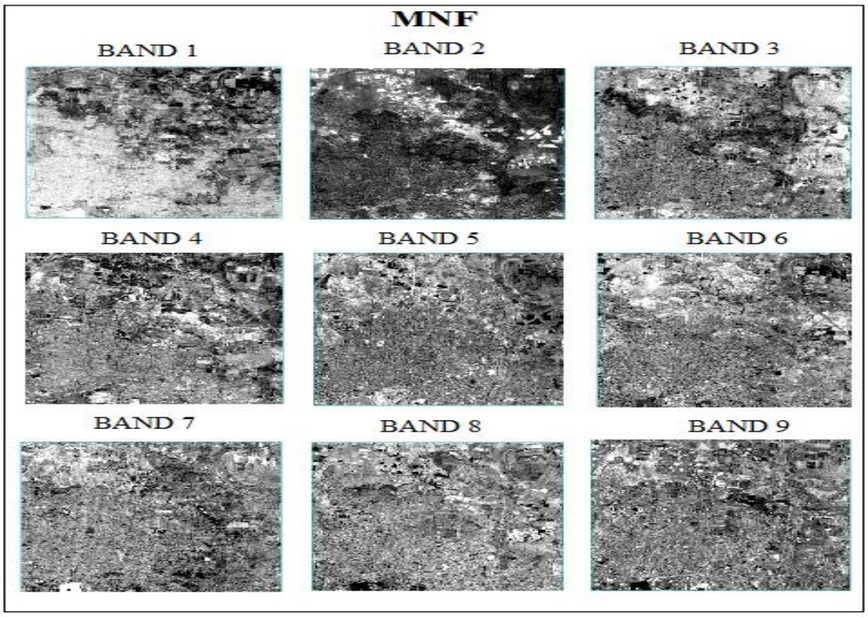


Figure 3. Minimum Noise Fraction

3.2. Interpretation of eigenvalues through scree plot

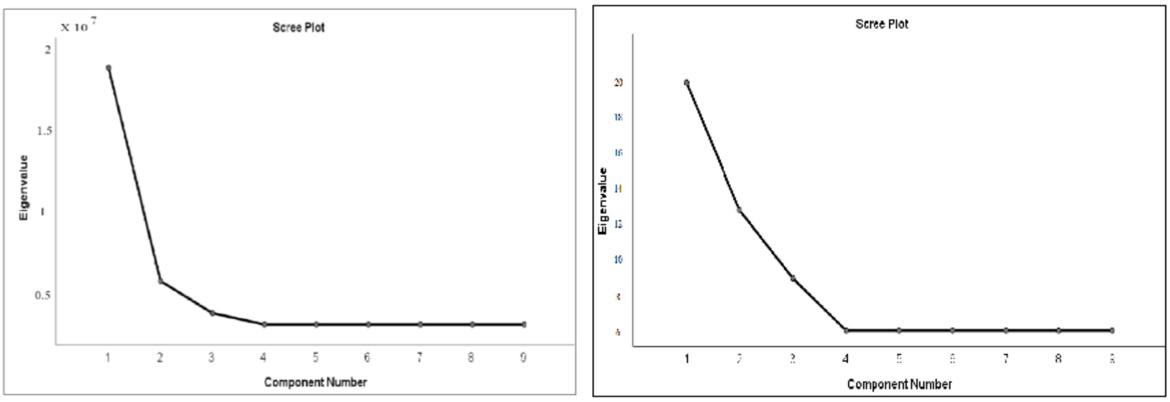
The scree test is a criterion used as an optimal method for reducing dimensionality band from the hyperspectral dataset [6]. From the scree plot, it is shown that first 9 principal components are retained and the bands corresponding to these components provide large data variance when compared with the low order PC. From the MNF results it is seen that, first 9 eigenimages contain coherent information while the others contain noise. In general, the eigenvalues of the components are decreasing denoting a lot of available information from higher to lower order. PCA eigenvalues are larger when compared to MNF indicating how well the data is spread out on the direction with much variance. However, MNF usually produce a higher signal to noise ratios when compared to that of PCA. Hence, denoising the image is more effective when using the MNF dimensionality reduction transformation. The results are analyzed by incorporating factor analysis method in which the acquired components and the total calculated are plotted against a scree graph. The scree graph represents the range of the eigenvalues plotted with respect to the principal components as it subsequently tapers towards lower order disclosing the noise in the hyperspectral imagery. Eigenvalues with the corresponding components for PCA and MNF along with the generated scree plot are shown in Table 1 and Table 2, Figure 4 (a), (b) respectively.

Table 1. PCA eigenvalue table

PCA	Eigenvalues	Percentage	Cumulative
1	18654093.4388	96.28	96.28
2	6258454.6755	1.30	97.58
3	4700620.3861	0.77	98.35
4	2242108.3396	0.44	98.79
5	911200.4619	0.41	99.20
6	754888.8681	0.33	99.53
7	339785.7655	0.19	99.72
8	136636.2315	0.15	99.87
9	70024.2476	0.06	99.93

Table 2. MNF eigenvalue table

MNF	Eigenvalues	Percentage	Cumulative
1	20.0705	16.43	16.43
2	12.9622	1.03	15.40
3	9.4114	1.05	14.35
4	7.7143	1.17	13.18
5	6.5404	1.32	11.86
6	6.3144	1.37	10.49
7	5.5829	1.62	8.87
8	5.0555	1.97	6.90
9	4.9387	2.71	4.19



(a) (b)

Figure 4. (a) Scree plot for PCA (b) MNF

4. Conclusion

It is concluded from the study that, MNF transform outperforms well than PCA dimensionality reduction technique. This is proved through visual inspection of resulting gray- scale eigenimages and scores provided by the given hyperspectral AVIRIS NG dataset. The first 9 components of the reduced dataset yielding at most information is considered for further analysis.

Author Contributions: Conceptualization & Analysis, K.N.P. and V.S.; Writing-Original Draft Preparation, K.N.P and V.S; Review & Editing, S.S. and K.B; Supervision, S.S. and K.B; Funding Acquisition, S.S.

Funding: This research was carried out under the support of DST funded Network Project on Imaging Spectroscopy & Applications (NISA).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Goetz, A.F.H.; Vane, G.; Solomon, E.J.; Rock, B.N. Imaging spectrometry for earth remote sensing, *Science* 1985, 228(4704), 1147-1153. 10.1126/science.228.4704.1147
2. Roger, R.E. Principal components transform with simple, automatic noise adjustment. Int. J. Remote Sens. 1996, 17(14), 2719-2727. 10.1080/01431169608949102
3. Green, A.A.; Berman, M.; Switzer, P.; Craig, M.D. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. IEEE Trans. Geosci. Remote Sens. 1988, 26(1), 65-74. 10.1109/36.3001
4. Gao, L.; Zhang, B.; Sun, X.; Li, S.; Du, Q.; Wu, C. Optimized maximum noise fraction for dimensionality reduction of Chinese HJ-1A hyperspectral data. EURASIP Journal on Advances in Signal Processing 2013, 65. doi:10.1186/1687-6180-2013-65

5. Liu, X.; Zhang, B.; Gao, L.; Chen, D. A maximum noise fraction transform with improved noise estimation for hyperspectral images. Science in China Series F: Information Sciences 2009, 52(9), 1578-1587. doi:10.1007/s11432-009-0156-z
6. Arslan, O.; AkyÜRek, Ö.; Kaya, Ş. A comparative analysis of classification methods for hyperspectral images generated with conventional dimension reduction methods. Turkish Journal of Electrical Engineering & Computer Sciences 2017, 25, 58-72. doi:10.3906/elk-1503-167
7. Principal Component Analysis. Available online: https://www.harrisgeospatial.com/docs/PrincipalComponentAnalysis.html (Accessed on 20 05 2019).
8. A Beginner's Guide to Eigenvectors, Eigenvalues, PCA, Covariance. Available online: https://skymind.ai/wiki/eigenvector (Accessed on 20 05 2019).
9. Minimum Noise Fraction Transform. Available online: https://www.harrisgeospatial.com/docs/MinimumNoiseFractionTransform.html (Accessed on 21 05 2019).