Dimensionality Reduction and Classification of Hyperspectral Data

S.N. Omkar[1], Sivaranjani V[1], Senthilnath J[1], Suman Mukherjee[2]
[1]Department of Aerospace Engineering, IISc, Bangalore
[2]Verizon Data Services India Pvt. Ltd, Chennai

Abstract
Remote sensing of hyperspectral satellite images has shown to have wide applications. One problem that researchers face is the high volume and dimensionality of the data, or the curse of dimensionality. In this paper, we have used the crop stage classification problem to assess the performance of different dimensionality reduction techniques. We have made use of three techniques, Principal Component Analysis (PCA), Independent Component Analysis (ICA) and Vertex Component Analysis (VCA) for the purpose of dimensionality reduction. We then use a Multi–Layer Perceptron Neural Network for the purpose of classification. Results of each scheme are evaluated interms of the classification efficiency and computational complexity afforded. Results obtained indicate that the use of PCA is least computationally intensive, while classification results are better for the data reduced using ICA and VCA.

1. INTRODUCTION

With the advent of high resolution sensors and high speed data processing devices, the use of hyperspectral images for land resource estimation has gained considerable attention [1].

Hyperspectral image data has been used for various applications such as target detection [2], material identification and mapping [3], identifying surface properties [4], crop classification [5-7], crop stage identification [8] etc,. One of the most significant issues involved with the use of hyperspectral images is the computational burden due to the high dimensionality of the data.

Various dimensionality reduction techniques have been used in the past to overcome this problem. Principal Component Analysis (PCA) [9] is one of the most widely used dimensionality reduction technique. PCA is a method in which the directions of maximum variance are retained by the use of Eigen values of the covariance data matrix. But the directions of maximum variance, are not always the most interesting, as in the case of non-Gaussian sources [10] in which case higher order techniques such as Independent Component Analysis (ICA) [10] provide better results [11].

ICA is a spectral unmixing technique where the constituent components of the data vector that are statistically as independent as possible are estimated. There are a few problems associated with the use of ICA for dimensionality reduction [12].

First, the number of endmembers in a hyperspectral data set is not a priori known and hence endmember estimation becomes necessary. Several methods have been proposed for the estimation of endmembers. Virtual Dimensionality (VD) is defined in [13,14] as a lower limit on the number of signal sources necessary to characterize hyperspectral data from the perspective of target detection and classification. A method called Harsanyi-Farrand-Chang (HFC) for VD estimation has been proposed in [13]. It first computes the sample covariance and the sample correlation matrix and then models the difference between their corresponding eigen values as a binary hypothesis problem that can be solved using a Neyman-Pearson detector obtained by maximizing the detection power for a fixed false alarm probability [13]. A minor variation of the HFC is proposed in the noise whitened HFC (NWHFC) [13] which includes an initial noise whitening step which results in improved accuracy of VD estimates. More recently, a method called hyperspectral signal identification by minimum error (HySime) has been proposed in [15] which performs an eigen decomposition based, adaptive estimation of the signal subspace in hyperspectral data. It computes an estimate of the noise and signal correlation matrices and then selects a subset of the eigen vectors that optimally represent the signal subspace in the least square error sense (LSE). The performance of HySime, HFC and NWHFC have been in [15] where

Dimensionality Reduction and Classification of Hyperspectral Data HySime is found to show results comparable to or better than HFC and NWHFC.

Another problem with ICA is that the order in which independent components are estimated cannot be pre-determined [10]. This implies that the components determined may not be the most significant. In order to overcome some of the limitations associated with the use of ICA for hyperspectral data [16], another spectral unmixing technique termed Vertex Component Analysis (VCA) has been proposed in [17].

VCA is a geometric endmember extraction method which is specifically designed to be used with hyperspectral data [17]. The algorithm makes use of the fact that pure pixel spectra occupy vertices of a simplex and iteratively projects spectral vectors onto a direction orthogonal to the endmembers already determined.

One of the applications of hyperspectral images is crop stage classification [8]. It is seen that a change in growth stage of a crop is characterized by changes in leaf pigments such as chlorophyll and carotenoids. These changes in turn affect the reflectance spectrum of the crop for intensities of different wavelengths which is used to extract stage information.

Crop reflectance spectrum is found to be characteristic in the red edge wavelength, ie the wavelength in the range 680-780 nm, which indicates that dimensionality reduction is inherently possible in the case of crop data. This fact has been utilized to attempt crop stage classification using the intensity information of the bands in the red edge region in [8]. In this paper we have made use of the crop stage classification problem to assess the affect of dimensionality reduction in terms of classification results and the computational burden.

Various image classification techniques have been tried and tested [18]. When the data is highly nonlinear, neural network based classifiers are found to be effective. Hence for the purpose of classification, we have made use of the Multi–Layer Perceptron Neural Network (MLPNN). The Multi–Layer Perceptron (MLP) is a standard classifier used in various classification problem [5-7].

This paper is organized in the following manner. In Section 2 we have described the data used, the dimensionality reduction and classification method employed. In Section 3 we have tabulated the classification results obtained in each case along with an assessment of the computational complexity of each scheme. Finally in Section 4 we present our conclusion based on the experimental results.

2. MATERIALS AND METHODS

2.1 Spectral Data Used

In this paper, we work with images acquired using EO-1 Hyperion Satellite. It has 225 bands, with a bandwidth of the order of 10 nm ranging from 300 to 2400 nm. The image resolution is 30 by 30m. The images were first corrected using Geomatica-10.0 software, Macrovision Corporation. To read the images, an open source C language library called Geospatial Data Abstraction Library (GDAL) was used. The data consists of three stages of growth of the wheat crop. From the picture, we have manually selected pixels that represent each of the three growth stages. The description of the growth stage and the number of pixels picked are indicated in Table 1.

Table 1. Data Discription

Stage Level	Description	Number of Pixels
Stage 1	Emergence	10
Stage 2	Mature	9
Stage 3	Milking	14

2.2 Dimensionality reduction methods

2.2.1 Principal Component Analysis

Principal component analysis (PCA) [9] is a method used to identify patterns in the data and express data in such a way as to highlight similarities and differences. It can be used to reduce data dimensionality without too much loss of information. The data is first centered by subtracting the mean across each dimension. The Eigen values of the data covariance matrix are then estimated and the values that are too small are discarded. That is the number of dimensions is determined by selecting a percentage of the positive Eigen values to be retained. Thus a feature vector matrix is formed by selecting the Eigen vectors to be retained while discarding those with small Eigen values. This forms the principal components (PC's) or the directions of maximum variance of the data matrix. The new data set is computed as the product of the feature vector matrix and the data matrix.

2.2.2 Independent Component Analysis

Data representation plays an important role in simplifying data analysis. Independent component analysis (ICA) [10] is one such tool employed to represent data in a format that removes redundancy to the best possible extent. Here, the input signal is decomposed into components that are statistically as independent as possible. It should be noted that statistical independence goes beyond second order uncorrelatedness in that independence implies uncorrelatedness while the converse is not necessarily true. The primary focus in the estimation of the ICA model is non-gaussianity [11]. Some measures of non-gaussianity are Kurtosis, Negentropy and Mutual Information. A detailed description on the use of each of these measures and the advantages and disadvantages of each is dealt with in [10] along with the FastICA algorithm for the estimation of ICs.

2.2.3 Vertex Component Analysis

Vertex component analysis (VCA) [17] is another spectral unmixing technique. This algorithm makes use of the fact that the endmembers occupy the vertices of a simplex and the affine transformation of a simplex is a simplex. Owing to scale factor variations, the data resides in a convex cone. The VCA algorithm starts with projecting the convex cone onto a chosen simplex hyperplane, with vertices corresponding to the vertices of the original data. Having identified the simplex hyperplane, the algorithm iteratively projects data onto a direction orthogonal to those of the endmembers already determined. The new endmember signatures correspond to the extremes of the projection. In each iteration, a new endmember signature is determined. Hence the algorithm iterates until all the endmembers are found. One disadvantage of this scheme is that the presence of pure pixels is assumed. This could be a very stringent requirement depending on the resolution of the sensors used.

2.3 Classification Using Multi-layer Perceptron

Figure 1 shows the general architecture of the proposed algorithm. The output of the first stage i.e. data of reduced dimension is given as the input to an MLP network. The MLP network is a layered feed-forward neural network. It can be trained using the generalized back propagation algorithm [19].

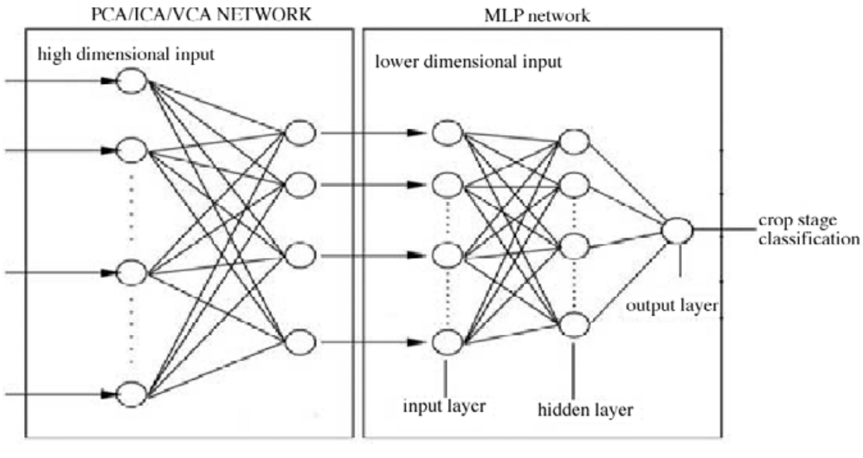

Figure 1. PCA/ICA/VCA Network for dimensionality reduction along with MLP classifier

We employ a three layered MLP neural network. The number of neurons in the first layer is the same as the number of input dimensions plus one bias. We use 20 neurons in the hidden layer and a single neuron in the output layer. The parameters used in the MLP network are shown in Table 2.

Table 2. MLP Properties

Parameter	Value: MLP Without Dimension Reduction	Value : MLP After Dimension Reduction
Architecture	226-20-1	12-20-1
Training Algorithm	Generalized back propagation	Generalized back propagation
Hidden layer activation	Sigmoidal function	Sigmoidal function
Output layer activation	Tanh function	Tanh function
Training stopping criteria	MSE<0.0001	MSE<0.001
Learning rate	0.1	0.1

3. RESULTS

3.1 Classification Results

The above PCA/ICA/VCA-MLP (Figure. 1) scheme described is used on hyperspectral image data acquired from the EO1-Hyperion.

For PCA, the highest 10 principal components are retained in order to preserve 100% of the power. The number of dimensions to be retained in the case of ICA and VCA is estimated using HySime [15], and is computed to be 11.

The MLP network is trained with four patterns for each crop stage, selected randomly from the available data. After training, the complete data set is used to test the network. The results obtained are shown in the classification matrix (Table 3 & Table 4). Table 3 shows the results obtained in the case of the MLP without dimension reduction, ICA-MLP, and VCA-MLP scheme. Table 4 shows the results obtained in the case of PCA-MLP scheme.

As can be observed in Table 3 and Table 4, classification is accurate for all three schemes in the case of stage 1 and stage 2 but in the case of stage 3, the PCA-MLP network shows some misclassification of stage 3 crops as stage 2, while the ICA-MLP and the VCA-MLP schemes show accurate classification.

Table 3. Classification Matrix for ICA-MLP &VCA MLP network & MLP without dimension reduction

Crop Stage	Stage 1	Stage 2	Stage 3	Individual efficiency
Stage 1	10	0	0	100%
Stage 2	0	9	0	100%
Stage 3	0	0	14	100%
			Overall Efficiency	100%

Table 4. Classification matrix for PCA-MLP network

Crop Stage	Stage 1	Stage 2	Stage 3	Individual efficiency
Stage 1	10	0	0	100%
Stage 2	0	9	0	100%
Stage 3	0	4	10	71.4%
			Overall Efficiency	87.8%

3.2 Complexity Assessment

The computational complexity of the MLP neural network trained with back propagation is of $O(K^2)$ where K denotes the data matrix dimension. For all three schemes used, i.e. PCA, ICA and VCA, the dimensionality of the data reduces to M<<K. In which case the complexity of the MLPNN reduces to $O(M^2)$. The computation of the number of floating point operations for the MLP network is done based on the evaluation in [20] which gives the total flop count *(fc(MLP))* to be

$$fc(MLP) = 2dn + d\sum_{S=2}^{N} T_s P_s + dP_N + 2dP_N + dP_N + d\sum_{S=1}^{N-1} P_{s+1}(V_{s+1}+1) + 2d\sum_{S=2}^{N-1} P_s P_{s+1} + dn + 2dn \quad (1)$$

The list of symbols used and their corresponding definition is given in Table 5.

Table 5. List of symbols used

Variable	Description
K	Number of dimensions of input data
M	Number of dimensions after reduction
D	Number of samples of training data
N	Number of layers of MLP
n	Total number of weight parameters for MLP network
P_S	Number of neurons in layer S including bias
T_S	Operations for evaluating [f(.)], node activation
V_S	Operations for evaluating node activation function derivative [f'(.)]
T	Total number of samples in data set
B	Number of flops for spatial whitening
J	Number of iterations

The number of floating point operations required for PCA has been computed in [21]. The flop count for PCA is hence

$$fc(PCA) = 9\mathrm{T}^3 + 2\mathrm{KT}^2 + 2\mathrm{KTM} \tag{2}$$

The HySIME algorithm used to infer the signal subspace prior to the use of ICA and VCA requires the estimation of the noise and the signal correlation matrices. The number of flops required for the noise estimation is computed in [15]. The flop count is hence approximately

$$fc(HySime) = 4\mathrm{TK}^2 + 6\mathrm{K}^3 + 2\mathrm{TK}^2 \tag{3}$$

The number of floating point operations for the fastICA algorithm used is computed the evaluation in [22] which shows the number of multiply and add operations to be

$$fc(ICA) = \mathrm{B} + \mathrm{J}\left[2\mathrm{M}(\mathrm{M}+\mathrm{T}) + 5\mathrm{TM}^2/2\right] \tag{4}$$

The number of iterations is data specific. For the data set used the average number of iterations after multiple runs was found to be 60. The number of floating point operations is taken to be twice the number of multiply and add operations.

The complexity of the VCA algorithm is evaluated as in [17] which gives the flop count to be

$$fc(VCA) = 2\mathrm{M}^2\mathrm{T} \tag{5}$$

These results have been tabulated in Table 6.

Table 6. Complexity Assessment

Scheme	Complexity Pre-Processing	Complexity MLPNN (Approximate)	Total count (Approximate)
MLP		260720000	260720000
PCA-MLP	961983	91120000	92081983
ICA-MLP	76340310	91520000	167860310
VCA-MLP	78375486	91520000	169895486

As can be observed from Tables 3, 4 and 6, the PCA-MLP scheme turns out to be least computationally expensive but its classification accuracy remains 87.8% while the ICA and VCA–MLP networks although more computationally intensive, show 100% accuracy.

4. CONCLUSION

In this paper we have made use of a MLP network to solve the problem of crop stage classification using hyperspectral images. Various pre-processing schemes have been explored to reduce data dimensionality. The computational complexity of each scheme is compared. The data acquired from the EO1 Hyperion with 225 spectral bands is used to test our scheme. The dimensionality of the data was reduced to 10 components in the case of PCA and 11 in the case of ICA and VCA. And the MLP network after training was able to classify the data with 87.8% accuracy in the case of PCA-MLP and 100% in the case of ICA-MLP and VCA-MLP.

All the schemes proposed yield a significant reduction in computational requirements of the MLP classifier. This scheme proposed is a supervised classification technique, the challenge of unsupervised classification remains.

ACKNOWLEDGMENT

This work is supported by the Space Technology Cell, Indian Institute of Science and Indian Space Research Organisation grant.

REFERENCES

[1] David Landgrebe, Hyperspectral Image Data Analysis as a High Dimensional Signal Processing Problem, *IEEE Signal Processing Magazine*, Vol. 19, No. 1, pp. 17-28, January 2002.

[2] Mohammad S. Alam, Mohammed Nazrul Islam, Abdullah Bal & Mohammad A. Karim, Hyperspectral target detection using Gaussian filter and post processing, *Optics and Lasers in Engineering*, Vol. 46, No. 11, pp. 817-822, Nov 2008

[3] Amar Kachenoura, Laurent Albera, Lotfi Senhadji & Pierre Comon, ICA : a potential tool for BCI systems, *IEEE Signal Processing Magazine* 2008, Vol. 25, No . 1, pp. 57-68, 2008

[4] E Ben-Dor, K Patkin, A Banin & A Karnieli, Mapping of several soil properties using DAIS-7915 Hyperspectral scanner data – a case study over clayey soils in Israel, *International Journal of Remote Sensing*, Vol. 23, No. 6, pp 1043-1062, 2002

[5] S.N.Omkar, J. Senthilnath, Dheevatsa Mudigere & M. Manoj Kumar, Crop Classification using Biologically inspired Techniques with High Resolution Satellite Image, *J. Indian Soc. Remote Sens*. Vol. 36, pp 175-182, June 2008

[6] Paul C. Doraiswamy, Alan J. Stern & Bakhyt Akhmedov, Crop Classification in the U.S. Corn Belt Using MODIS imagery, in the proceeding of *International Geoscience and Remote Sensing Symposium*, July 19 – 27, 2007, Barcelona, Spain.

[7] Abdulhamit Subasi & Ergun Ercelebi, Classification of EEG signals using Neural Networks and logistic regression, *Computer methods and Programs in Bio-medicine*, Vol. 78, Issue 2, pp. 87-99, May 2005

[8] Suman Mukherjee, Crop Stage Classification of Hyperspectral Data Using Red Edge Wavelength Pursuit , in *the proceeding of XXVIII INCA International Congress on Collaborative Mapping and Space Technology*, November 2008, Ahmedabad.

[9] Sunghyun Lim, Kwang Hoon Sohn & Chulhee Lee, Principal Component Analysis for compression of Hyperspectral Images, *Geoscience & Remote Sensing symposium, 2001, IGARSS01*, Vol.1, pp. 97-99, July 2001.

[10] Aapo Hyvarinen & Erkki Oja, Independent Component Analysis: Algorithms and Applications, *Neural Networks*, Vol. 13, Issue 4-5, pp 411-430, 2000.

[11] R. Mutihac, Marc M. Van Hulle, Comparison of Principal Component Analysis and Independent Component Analysis for Blind Source Separation, *Romanian Reports in Physics*, Vol. 56, No I, pp. 20-32, 2004.

[12] Jing Wang & Chein-I Chang, Independent Component Analysis-Based Dimensionality Reduction With Applications in Hyperspectral Image Analysis, *IEEE Trans. Geosci. Remote Sens.*, Vol. 44, No. 6, pp. 1586-1600, June 2006.

[13] Chein-I chang, Estimation of the Number of Spectrally Distinct Signal Sources in Hyperspectral Imagery, *IEEE Trans. Geosci. Remote Sens.*, Vol. 42, No 3, March 2004.

[14] C-I Chang, Hyperspectral Imaging: Techniques for Spectral Detection and Classification. Norwell, MA : Kluwer, 2003.

[15] Jose M. P. Nascimento & Jose M. Bioucas-Dias, Unsupervised Hyperspectral Signal Subspace Identification , *IEEE Trans. Geosci. Remote Sens.*, Vol. 46, No 8, pp 2435-2445, July 2008.

[16] Jose M.P. Nascimento, Jose m. Bioucas Dias, Does Independent Component Analysis play a role in unmixing hyperspectral data?, *IEEE Trans. Geosci. Remote Sens.*, Vol. 43, No. 1, pp. 175-187, January 2005.

[17] Jose M.P. Nascimento, Jose m. Bioucas Dias, Vertex Component Analysis : a fast algorithm to unmix hyperspectral data, *IEEE Trans. On Geosc. and Remote Sens.* Vol. 43, No. 4, pp 898-910, April 2005.

[18] D. Lu & Q. Weng, A survey of image classification methods and techniques for improving classification performance, *International Journal of Remote Sensing*, Vol. 28, No. 5, pp. 823-827, 2007

[19] D. E. Rumelhart, G. E. Hinton & R. J. Williams, Learning internal representations by error propagation, *Parallel Distributed Processing*, Vol 1, D. E. Rumelhart and J. L. Mclelland, Eds. Cambridge, MA: M.I.T. Press, pp 318-362, 1986

[20] Eiji Mizutani & Stuart E.Dreyfus, On complexity analysis of supervised MLP- learning for algorithmic comparisons, *International Joint Conference on Neural Networks*, 2001, Proceedings. IJCNN,01, Vol. 1, pp. 347-352, 2001

[21] Jun Liu, Songcan Chen, Discriminant common vectors versus neighbourhood components analysis and Laplacianfaces: A comparative study in small sample size problem, *Image and Vision Computing*, Vol. 24, No. 3, pp. 249-262, March 2006

[22] John D.Gorman, Nicola S.Subotik, Brian J.Thelen, Robust material identification in hyperspectral data via multiresolution wavelet techniques, *International conference on acoustics, speech and signal processing, ICASSP-1995*, Vol. 5, pp. 2805-2808, May 1995