International Journal of Information Technology and Knowledge Management
January June 2009, Volume 2, No. 1, pp. 69-71

PRINICPAL COMPONENT ANALYSIS FOR IMAGE COMPRESSION

Sonal*, Vikas Beniwal & Sandeep Kharb*****

This paper deals with Principal Component Analysis and how it is used for getting a compressed image. In this we have shown our results that are obtained by performing image compression using Principal Component Analysis on some standard image taken from internet. We have used MATLAB for implementation.

1. Introduction

PCA has find out its most important application in the field of linear algebra. PCA is a method of extracting information from confusing data sets so used in various fields like neuroscience, computer graphics, etc [19]. PCA is a vector space transform often used to reduce multidimensional data sets to lower dimensions for analysis. Depending on the field of application, it is also named the discrete Karhunen-Loeve transform, the Hotelling transform or proper orthogonal decomposition (POD). PCA was invented in 1901 by "Karl Pearson" [20]. The structure of a self organizing system may take on a variety of different forms. It may consist of an input layer and an output layer with feed forward connections from input to output layer. Principal Component Analysis is an unsupervised approach of neural networks. It is a standard technique commonly used for data reduction in statistical pattern recognition and signal processing.

Definition: Principle Components

A set of variables that define a projection that encapsulates the maximum amount of variation in a dataset and is orthogonal (and therefore uncorrelated) to the previous principle component of the same dataset.

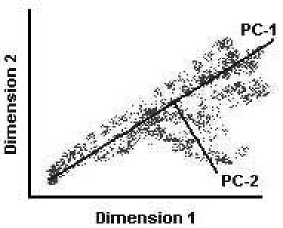


Figure 1: The Lines Represent 2 Consecutive Principle Components. Note That They are Orthogonal (At Right Angles) to Each Other

Technically, PCA is a linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. PCA can be used for dimensionality reduction in a dataset while retaining those characteristics of the dataset that contribute most to its variance, by keeping lower-order principal components and ignoring higher-order ones. Such low-order components often contain the "most important" aspects of the data. But this is not necessarily the case, depending on the application.

2. Procedure for PCA

2.1 Data Matrix

The analysis is performed on a data set of p variables (X_1, X_2 … X_p) for n individuals. From this data set, a corresponding squared covariance or correlation matrix can be calculated. For the covariance matrix the following equation can be used:

$$Cov(X_j, X_k) = \frac{\sum (X_{ij} - \overline{X_j})(X_{ik} - \overline{X_k})}{(n-1)}$$

where $\overline{X_j} = \frac{\sum_{i=1}^{n} X_{ij}}{n}$

And, $j, k = 1, 2\ldots P$.

2.2 Covariance Matrix

Covariance is defined as a measure that can be taken between any two dimensions. The covariance matrix is used when the variables are measured on comparable scales.

2.3 Computing Eigenvectors and Eigenvalues

2.4 Selecting Eigenvectors

Eigenvectors obtained from the covariance matrix are needed to be sorted in accordance with their eigenvalues.

* Lecturer (CSE Deptt. SES, BPSMV, Khanpur Kalan), E-mail: *sonalkharb@rediffmail.com*

** H.O.D (CSE & IT Dept., RIMT, Chidana), E-mail: *beniwal_vikas23@rediffmail.com*

*** Lecturer (CSE Deptt., Govt. Polytechnic, Uttawar (working as officiating incharge), E-mail: *kharbs@rediffmail.com*

Now, some of the eigenvectors has been discarded. This can be done on the basis of comparison like here we have discarded all eigenvectors whose value is less than 0.001. All other selected values are used to obtain the result.

3. IMPLEMENTATION

As implementation is concerned, MATLAB 6.5 software has been used for performing the image compression. The principal component analysis approach is applied on those images for performing the compression. The steps that have been followed are as follows:

Step 1: Get Some Data

For getting the data twenty standard images provided by AT & T [34] from the internet are used. All the images selected are of 112 × 92 dimensions. Now before performing the image compression on these images they are all resized to 60 × 60 dimensions.

Step 2: Compute & Subtract the Mean

After making a dataset in the form of matrix next step that has been performed according to the Principal Component Analysis Approach is to compute the mean of the matrix obtained

Step 3: Compute Covariance Matrix

Step 4: Compute Eigen Values and Eigen Vectors

Figure 2: Original Image from Database

Reconstructed Images

Reconstructed images are shown on a scale of 0.1 to 1.0. All the images obtained at these different values are:

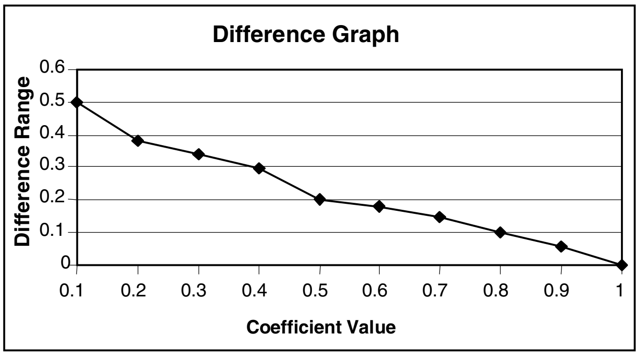
As shown in the above figures at 0.1 we have a very disturbed picture and as value of coefficient is increased the picture obtained is clearer and resembles to the original image.

Table
Difference in Values at Different Coefficients

Coefficient value	*Eigen Vectors*	*Difference range*
0.1	2	0 – 0.5
0.2	4	0 – 0.38
0.3	6	0 – 0.34
0.4	8	0 – 0.3
0.5	10	0 – 0.2
0.6	12	0 -0.18
0.7	14	0 – 0.15
0.8	16	0 – 0.1
0.9	18	0 – 0.06
1.0	20	0

As compared in the above table the difference range at 1 is about negligible, whereas highest difference range is obtained at 0.1. It has been obtained that the best value can be at 0.5. So the images compressed at 0.5 are lossy in nature but not up to that range.

The concluded table which was shown previously can also be displayed with the help of a graph as shown below:



As shown in the graph above the X- axis represents the values of coefficients. These coefficients shows the percentage of eigen vectors. The Y-axis shows the maximum difference range at different coefficients. As shown the maximum difference between the actual image and the compressed image is obtained at 0.1 (2 eigen vectors) and the minimum can be obtained at 1(20 eigen vectors) which is near to zero.

So we can conclude from above discussion that the image obtained at a 0.5 (10 eigen vectors) can be best giving a lossy image as compared to the original image.

References

[1] Ronny Luss, Alexandr'd Asprement, " DSPCA- A Toolbox for Sparse PCA", (Nov. 29, 2006).

[2] Subramanya A., "Image Compression Technique," *Potentials IEEE*, **20**, (1), (Feb-March 2001), 19–23.

[3] David Jeff Jackson & Sidney Joel Hannah, " Comparative Analysis of Image Compression Techniques," *System Theory 1993, Proceedings SSST '93, 25th Southeastern Symposium*, (7–9 March 1993), 513-517.

[4] Hong Zhang, Xiaofei Zhang & Shun Cao, "Analysis & Evaluation of Some Image Compression Techniques," *High Performance Computing in Asia-Pacific Region, 2000 Proceedings, 4th Int. Conference*, **2**, (4-17 May, 2000), 799–803.

[5] Ming Yang & Nikolaos Bourbakis ,"An Overview of Lossless Digital Image Compression Techniques," *Circuits & Systems, 2005 48th Midwest Symposium*, *IEEE* , **2**, 7– 10 Aug, 2005, 1099–1102.

[6] Milos Klima, Karel Fliegel, "Image Compression Techniques in the Field of Security Technology: Examples and Discussion,"Security Technology, 2004, *38th Annual 2004 Intn. Carnahan Conference*, (11-14 Oct., 2004), 278–284

[7] Ismail Avcibas, Nasir Memon, Bulent Sankur, Khalid Sayood, "A Progressive Lossless / Near Lossless Image Compression Algorithm," *IEEE Signal Processing Letters*, **9**, (10), (October 2002), 312–314.

[8] Dr. Charles F. Hall, "A Hybrid Image Compression Technique," Acoustics Speech & Signal Processing, *IEEE International Conference on ICASSP' 85*, **10**, (Apr, 1985), 149–152

[9] Wen Shiung Chen, En-HuiYang & Zhen Zhang, " A New Efficient Image Compression Technique with Index-Matching Vector Quantization," *Consumer Electronics, IEEE Transactions*, **43**, (2), (May 1997), 173–182.

[10] David H. Kil and Fances Bongjoo Shin, "Reduced Dimension Image Compression And its Applications," *Image Processing, 1995, Proceedings, International Conference*, **3**, (23-26 Oct.,1995), 500–503.

[11] C.K. Li and H.Yuen, "A High Performance Image Compression Technique For Multimedia Applications," *IEEE Transactions on Consumer Electronics*, **42**, (2), (2 May 1996), 239–243.

[12] Shi-Fei Ding, Zhong–Zhi Shi,Yong Liang, Feng-Xiang Jin, "Information Feature Analysis and Improved Algorithm of PCA," *Proceedings of the 4th International Conference on Machine Learning and Cybernetics*, Guangzhou, (18-21 August, 2005), 1756–1761.

[13] Vo Dinh Minh Nhat, Sung Young Lee, "Two- Dimensional Weighted PCA algorithm for Face Recognition," *Proceedings 2005 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, Espoo, Finland, (June 27-30, 2005), 219–223.

[14] Robert D. Dony, Simon Haykin, "Neural Network Approaches to Image Compression", *Proceedings IEEE*, **83**, (2), (February 1995), 288–303.

[15] Kou-Yuan Huang, "Neural Networks for Seismic Principal Components Analysis", *IEEE Transactions on Geoscience & Remote Sensing*, **37**, (1), (January 1999), 297–311.

[16] J. Hiang, "Neural Network Technology for Image Compression", *IEEE International Broadcasting Convetion*, **413**, (14-18 September, 1995), 250–257.

[17] J. S. Taur, C. W. Tao, "Medical Image Compression Using Principal Component Analysis", *Proceedings, International Conference on Image Processing*, **1**, (16-19 September, 1996), 903–906.

[18] Matthew A. Turk, Alex Pentland, "Face Recognition using Eigenfaces", *IEEE Proceedings*, (1991), 586–591.

[19] H. Chen, R. Liu, "An Alternative Proof of Convergence for Kung-Diamantras APEX Algorithm", *Potentials IEEE*, 40–49.

[20] Simone Fiori, Saverio Costa, Pietro Burrascano, "Improved – APEX for Digital Image Compression".

[21] Daoquiang Zhang, Songcan Chen, "Fast Image Compression using Matrix K-L Transform".

[22] S. Y. Kung, K. I. Diamantaras, "A Neural Network Learning Algorithm for APEX", *IEEE Potentials*, (1990), 861–864.

[23] Mathieu Delichere, Daviel Memmi, "Neural Dimensionality Reduction for Document Processing", *ESANN Proceedings*, (24-26th April, 2002), 211–216.

[24] Simone Fiori, Francesco Piazza, "A Compression of Three PCA Neural Networks", *ESANN Proceedings*, (21-23rd April,1999), 275–280.

[25] Data Compression – Encoder & PCA.

[26] Stefan Hausler- *Tutorial for Computational Intelligence on Neural Network Toolbox.*

[27] Jialong He, "Matlab Toolbox Quick Reference".

[28] Rafael C. Gonzalez, Richard E. Woods, Steven L. Eddins, "Digital Image Processing Using MATLAB", Second Indian Reprint, Pearson Education (2005).

[29] Simon Haykin, "Neural Networks," Second Indian Reprint, Pearson Education (2001).

[30] Satish Kumar, "Neural Networks- A Classroom Approach", Tata Mc-Graw Hill, (2005).

[31] *www.wikipedia.com*

[32] *www.face-reg.org/databases/*

[33] *www.answers.com*