Face Recognition using Tensor Analysis

Prahlad R. Enuganti

The University of Texas at Austin

Final Report

EE381K – 14 Multidimensional Digital Signal Processing

May 16, 2005

Submitted to

Prof. Brian Evans

ABSTRACT

Over the past fifteen years many methods have been developed to tackle the problem of recognizing human faces. Face recognition is currently one of the most researched areas in pattern recognition. Its popularity stems from the fact that its applications are used in a variety of real life situations ranging from human - computer interaction to authentication and surveillance. Although various machine learning techniques have been developed, their success is limited because of the restrictions imposed by data acquisition systems. This paper tries to develop couple of robust novel techniques, based on Tensor Analysis and Isometric Feature Mapping (ISOMAP), to recognizes human faces.

INTRODUCTION

Human recognition processes consider a broad spectrum of stimuli obtained from many, if not all, of the senses. The human brain is a complex system that probably applies contextual knowledge to recognize faces. It is futile to even attempt developing a computer system using existing technologies that can closely resemble the remarkable ability of facial recognition in humans. However, the key advantage that such a computer system would have over a human classifier is due to the limitation of the human brain to accurately remember a large database of individuals. Over the past couple of decades, face recognition has emerged as one of the primary areas of research in pattern recognition. The fact that it has numerous potential applications in biometrics, surveillance, human-computer interaction, video based communication, and the emergence of technologies that enable the implementation of these algorithms in real-time are the main reasons for this trend. All of the existing FRT systems suffer from a dip in performance whenever the data acquisition systems suffer from a change in pose, illumination and expression. FRT systems can be broadly classified into two groups depending on whether they make use of still images or video. In this paper, the focus was only to develop an FRT system that made use of static images. The aim of this project can be described as follows: Given an image of a person under varying conditions of illumination or pose, verify/identify the person in the stored database of facial images.

BACKGROUND : PREVIOUS WORK

One of the first attempts at automatic face recognition was made by Kanade [1] in 1973. He used a robust feature detector to locate feature geometric points on the facial image. This was a very simple algorithm that when tested on a database consisting of images obtained from 20 individuals performed at an accuracy of 45 ~ 75 %. Since Kanade's algorithm, several algorithms have been developed to tackle the problem of facial recognition. Some of involve feature extraction [2] while others were based on principal component analysis [3], Support Vector Machines[5], Wavelet transform [9], Gabor filters [10], etc.

	Resistance to			Computational	Classification
	Illumination	**View**	**Expression**	**Efficiency**	**Quality**
Technique					
Geometric Features [2]	good	Poor	Good	good	very poor
Eigenfaces [3]	average	Poor	Average	good	average
SVM [5,6]	average	Average	Average	good	very good
Depth and Texture Maps [7]	good	Good	Good	average	very good
Multiresolution Analysis [8]	good	Good	Very good	average	very good
Gabor Feature Classifier [9]	good	Good	Good	average	very good
Tensor Analysis [13]	very good	Very good	Very good	average	very good

Table 1

Table 1 presents a qualitative study of some of the FRT techniques that have been developed over the years. Although some of these algorithms were fast and accurate for small databases, their performance suffered when additional constraints such as varying illumination and pose were imposed on the image acquisition system [4]. From Table 1, it also appears that Tensor Analysis could solve some of the performance related problems faced by the earlier FRT systems.

MATERIALS AND METHODS

TENSOR ANALYSIS

Vasilescu *et al.* [10] tried to solve the problem of facial recognition using Tensor Analysis. They identified the analysis of an ensemble of facial images resulting from the confluence of multiple factors related to scene structure, illumination, and viewpoint as a problem in multilinear algebra in which the image ensemble is represented as a higher-dimensional tensor. A tensor D can be expressed as a multilinear model of factors as follows [11]:

$$D = Z \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \ldots \times_N \mathbf{U}_N \tag{1}$$

where Z, known as the core tensor, is analogous to the diagonal matrix in standard SVD and $\mathbf{U}_1$ to $\mathbf{U}_N$ contain the orthonormal vectors spanning the column space of $D_{(n)}$ resulting from the *mode-n flattening of D* [11]. Using the "N-mode SVD" algorithm, a multilinear extension of conventional matrix singular value decomposition (SVD), the core tensor Z is obtained. This image data tensor is decomposed to separate and parsimoniously represent the constituent factors. In case of facial image data used in our experiments, the various variables are people, views, illumination and pixels. Therefore applying the SVD algorithm results in the following expression

$$D = Z \times_1 \mathbf{U}_{views} \times_2 \mathbf{U}_{illumination} \times_3 \mathbf{U}_{pixels} \times_4 \mathbf{U}_{people} \tag{2}$$

Multilinear analysis enables us to represent each person regardless of pose and illumination with the combination of different base tensors (similar to the case of EigenFaces)

$$B = Z \times_1 \mathbf{U}_{views} \times_2 \mathbf{U}_{illumination} \times_3 \mathbf{U}_{pixels} \tag{3}$$

ISOMETRIC FEATURE MAPPING (ISOMAP)

The common problem that is faced when working with high dimensional data such as gene expressions or large image databases is to find lower dimensional structures hidden in a much higher dimensional observation space. Isometric Feature Mapping, popularly known as ISOMAP is often used to solve dimensionality reduction problems [12]. Some of the traditional methods for dimensionality reduction are Principal Component Analysis (PCA) and Multidimensional Scaling (MDS). However, these techniques assume that the data points lie on a linear subspace of the high dimensional input space and cannot be used to capture any inherent non-linearity of the data. The advantage of ISOMAP over these linear techniques and other non-linear techniques is that it is capable of efficiently calculating a globally optimal solution.

Let us briefly consider the Swiss Roll dataset shown in 1 [12]. As can be seen from the figure, it is possible for two points to be extremely close in the original data as measured by their Euclidean distances but can be extremely far apart in the lower dimensional manifold when measured by the geodesic or shortest path distances.

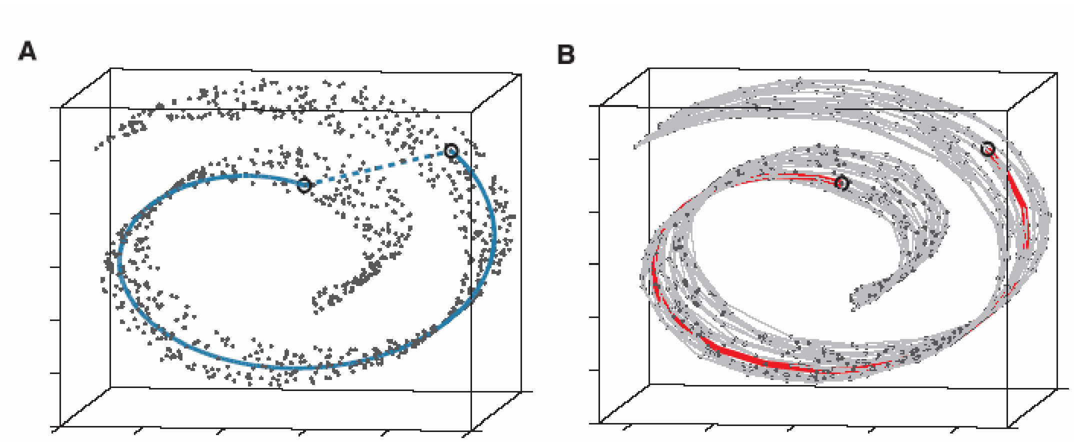

Figure 1: Swiss Roll Data Set

A) Distance when measured by Euclidean distances in higher dimensional data space

B) Distance between 2 points measured by their geodesic Distances

The advantage of ISOMAP in such cases is that it is capable of preserving local geodesic distances.

EXPERIMENTS AND RESULTS

The Images for testing the algorithm were taken from the PIE (Pose Expression Illumination) Database [13]. Images of 23 different individuals taken from 13 different cameras and 21 different illumination conditions were considered. Figure 2 shows some of these images under varying pose and illumination.

Figure 2: Image with A) pose 2 and lighting 5. B) pose 2 and lighting 20. C) Pose 5 and lighting 5

It is noticed that in each image there exists a certain smaller region of interest containing the face of the person. Therefore an elliptical mask is used to extract only that part of the image which contains the face. The pixels that fall within the confines of this mask are then filed into a single vector, thereby representing each image by a vector of length 6385. There were two different experiments performed to recognize faces.

EXPERIMENT 1: The approach of the first experiment was to perform tensor decomposition of the image tensor D of dimension 13 x 21 x 6385 x 23. From [11], the first step is to perform the SVD of the flattened mode – n matrix $D_{(n)}$. $U_{(n)}$ is obtained by performing the SVD of $D_{(n)}$ *i.e.* $D_{(n)} = U_{(n)}*S*V^T$. For n = 1, D(n) has the dimensions 13 x 3083955. Due to the extremely large size of the matrix, performing SVD is practically impossible as it would require calculation of 3 million eigen vectors in order to evaluate $U_{(n)}$. The trick used is to simplify this problem is to perform the SVD of $D_{(n)}^T$, which is comparatively much simpler as only 13 eigen values have to be computed. This technique worked well for calculating U_{view}, U_{illum} and

U_{people} but failed to calculate U_{pixels} as it was unable to calculate eigenvectors of 6385x6385 matrix.

EXPERIMENT 2: This technique focused on developing a face recognition system by using the ISOMAP technique. The database consists of 273 images for each of the 23 individuals. Each Image contains 6385 pixels and therefore can be thought of as a point in the 6385 point space. The aim of the ISOMAP algorithm is to reduce the dimensionality of this high dimensional data by finding a lower dimensional manifold that preserves the local geodesic distances. The steps involved in implementation of ISOMAP are:

1. Constructing the neighborhood graph: The first step was to create a neighborhood graph. *k* nearest neighbor rule was used to determine the number of neighbors that are considered for each of the data points. During the experiment, it was noticed that when *k* was chosen to be 10, the ISOMAP technique failed. This reason behind this could be that the points corresponding to images of the same person are close to each other and therefore when neighborhood was taken as 10, it resulted in disjoint graph leading to the failure of ISOMAP.

2. Computing the shortest paths: This was done using Floyds Algorithm [14], rather than Djikstra's Algorithm which was proposed in the original ISOMAP paper [12]. Djikstra's algorithm contains *for loops* which are inefficient to implement in MATLAB.

3. Construct *d* – dimensional embedding: The final step is to apply classical Multidimensional Scaling (MDS) to embed the data in the lower dimensional space.

The ISOMAP Algorithm was tested on the image database with 5 levels of illumination and 5 different poses for 23 different people. From Fig 3, it can be observed that the lowest dimensional manifold that best captures the data is 3. This is exactly what is expected since there are only three variables that are changing: pose, illumination and person's identity.

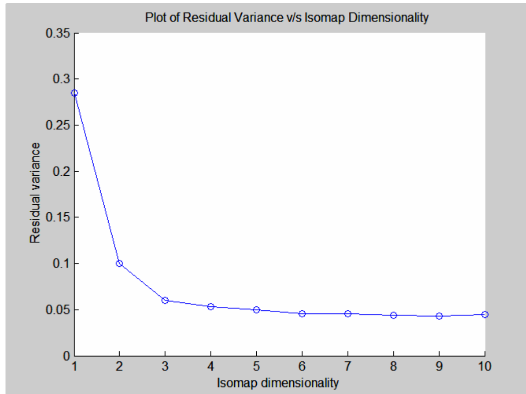

Fig 3: Plot of Residual Variance v/s Dimensionality

It was also a pleasant surprise to note that the recognition accuracy of **100%** was achieved.

CONCLUSIONS AND FUTURE WORK

From the results of the second experiment, it is evident that ISOMAP is an excellent dimensionality reduction technique for facial data. A nearest neighbor classifier when applied to the lower dimensional embedded data results in extremely high face recognition accuracy. The fault of this technique, as with any ISOMAP based technique, is the fact that it is only asymptotically guaranteed to converge to the actual structure. Hence it is likely that the obtained model can change as number of data points increase. There are some intrinsic defects present in the PIE dataset used for these experiments. The location of face is not same in all images for all the subjects. Although, the results of this study are not affected by these faults, it could create major problems as the number of individuals in the database increases. In order for Tensor Analysis to be useful, it is required that we have each and every image of all the subjects (*i.e.* with all poses and illuminations). However, it is not reasonable assumption in case of real life applications. The possibility of applying co-clustering technique in such cases is worth exploring.

ACKNOWLEDGEMENTS

I would like to thank Mr. Ralph Gross in the Robotics Institute at Carnegie Mellon University (CMU) for providing me with the CMU PIE (Pose Expression and Illumination) Database [13] without which this project would not have been possible. I would also like to thank Shalini Gupta, University of Texas at Austin for providing me with some of the preprocessing code.

REFERENCES

[1] T. Kanade, *Picture processing by computer complex and recognition of human faces*, Doctoral dissertation, Department of Information Science, Kyoto University, 1973

[2] R. Brunelli and T. Poggio, "*Face recognition through geometrical features*," in European Conference on Computer Vision (E*CCV)* 1992, pp. 792-800.

[3] M.A. Turk, A.P. Pentland, "Eigenfaces for recognition", *Journal of Cognitive Neuroscience* 3 (1) , 1991 71-86

[4] A. Jones, *CS 851 Recognizing Patterns and Digital Signatures Course Notes*, Department of Computer Science, The University of Virginia , Charlottesville, VA , Spring 2004.

[5] G. Guo, S.Z. Li and K.L. Chan, " Support Vector Machines for Face Recognition", *Image and Vision Computing* 19 (2001) 631-638

[6] Y. Lia, S. Gong, J. Sherrah and H. Liddell, " Support vector machine based multi-view face recognition and detection", *Image and Vision Computing* 22 (2004) 413-427.

[7] C. BenAbdelkader and P.A. Griffin, "Comparing and combining depth and texture cues for face recognition", *Image and Vision Computing* 23 (2005) 339–352.

[8] H. K. Ekenel and B. Sankur, "Multiresolution Face Recognition", *Image and Vision Computing* 23 (2005) 1-9.

[9] C. Li and H. Wechsler, "A Gabor feature classifier for face recognition" in *Eighth IEEE International Conference on Computer Vision, 2001.* Volume: 2, 7-14 July 2001 Pages: 270 – 275.

[10] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear Image Analysis for Facial Recognition", *International Conference on Pattern Recognition (ICPR' 02)*, Quebec City, Canada, August, 2002

[11] M. A. O. Vasilescu and D. Terzopoulos, "Tensor Textures: Multilinear Image based Rendering", *Proc. Association for computer machinery Special Interest Group in Graphics (ACM SIGGRAPH) 2004 Conference* Los Angeles, CA, August, 2004

[12] J.B. Tenenbaum, V. de Silva, and J.C. Langford. "A global geometric framework for nonlinear dimensionality reduction". *Science*, 290:2319 to 2323, 2000.

[13] T. Sim, S. Baker, M. Bsat, " The CMU Pose, Illumination and Expression (PIE) Database", *Proceedings of the 5th International Conference on Automatic Face and Gesture Recognition*, May 2002

[14] L. Allison, T. I. Dix and C. N. Yee, "Shortest Path and Closure Algorithms for Banded Matrices", *http://www.csse.monash.edu.au/~lloyd/tildeAlgDS/2D/1991-Banded/*.