Appearance based Recognition Methodology for Recognising Fingerspelling Alphabets

M.G.Suraj, D.S.Guru
Department of Studies in Computer Science
University of Mysore, Mysore, India
Email: mgsuraj@yahoo.com, dsg@compsci.uni-mysore.ac.in

Abstract

In this paper, a study on the suitability of an appearance based model, specifically PCA based model, for the purpose of recognising fingerspelling (sign language) alphabets is made. Its recognition performance on a large and varied real time dataset is analysed. In order to enhance the performance of a PCA based model, we suggest to incorporate a sort of pre-processing operation both during training and recognition. An exhaustive experiment conducted on a large number of fingerspelling alphabet images taken from 20 different individuals in real environment has revealed that the suggested pre-processing has a drastic impact in improving the performance of a conventional PCA based model.

Keywords: Appearance based recognition, PCA model, Histogram equalisation, fingerspelling alphabet, sign language recognition.

1 Introduction

Automatic sign language recognition has enormous social significance. It would lead to integration of people with voice and hearing disability to communicate and interact with the society. Sign language recognition is being actively pursued by the research community around the world. A survey of techniques adopted for recognition [LaViola, 1999; Pavlovic *et al*, 1997] give a review of techniques used in the broader area of hand gesture recognition. Pioneering work in this area was by [Starner and Pentland, 1995] using Hidden Markov Models (HMM) which recognises 40 signs from American Sign Language by extracting a feature vector describing the hand shape, orientation and trajectory. Hand modeling was avoided by tracking the coloured hand gloves worn by the user and in addition the usage of HMM gave the time context. [Gao *et al*., 2000] use cyber gloves and positional trackers to get features for recognising the user signs employing HMM, Artificial Neural Network and Dynamic Programming techniques. To overcome the expensive cost of cyber gloves, [Hernández *et al*., 2002] use inexpensive micro accelerometers to extract postures of hand as 3-D patterns. Classification into sub classes was done by projecting the vectors on to different planes, and hierarchical decision tree was used to classify within subclasses. They report an excellent success rate for majority of signs of fingerspelling alphabets. [Fang *et al*., 2004] report the use of cyber gloves to capture features and classify large vocabulary of signs using Fuzzy Decision trees. Recently [Chalechale et al., 2005] reported the use of geometric and moment based properties of fingerspelling hand postures to classify them using Bayesian rule assuming Gaussian distribution of the properties. They report a very good success rate, but it needs to be mentioned that they use gestures of single user with black background and wearing black cloth up to wrists. All the above mentioned approaches put restrictions on the user and the environment for the systems to work.

[Kang *et al*., 2005] declare that although several aspects of directing computers using human gestures/postures have been studied in the literature, gesture/posture recognition is still an open problem. This is due to significant challenges in response time, reliability, economical constrains, and natural intuitive gesticulation restrictions.

Apart from the above mentioned image processing approaches for feature extraction, appearance based methodologies have been reported primarily in the area of face recognition. Appearance based methods started with the work of [Turk and Pentland, 1991] on face recognition using a well known statistical technique called Principal Component Analysis (PCA). Since then, various extensions and modifications have been proposed to the PCA method for face recognition, resulting in 2 Dimensional PCA: 2D PCA [Yang *et al*, 2004], 2 Directional 2 Dimensional PCA: $(2D)^2$ PCA [Zhang and Zhou, 2005] and Diagonal PCA [Zhang *et al*., 2006] methods improving the recognition rates and memory requirements over PCA. Not much research has been reported regarding the application of appearance based approach for sign language recognition, [Birk and Moeslund, 1996] mention PCA based approach for sign language recognition and [Birk *et al*., 1997] gave an introductory work to fingerspelling recognition using PCA, although the system gave excellent results of 98.4% recognition rate (defined as number of correct recognitions / total number of test samples), the system was trained using sign images of single user with a black background. They reported the exploration of the system using two users and mention good results, but it should be taken that the hand structure of the second user is similar to the first user in light of the experiments using more users in this work.

In this paper, overcoming the complexity and variability of signs in a large dataset thereby improving the performance of the PCA algorithm using standard image processing techniques is explored. Image processing methods like smoothing and contrast enhancement of the images are applied before using them in training and recognition phases. A novel approach to enhance the performance of the appearance based methodology of using PCA in general conditions is presented. No restrictions on the user and the environment are imposed except that of a plain background. In fact, the proposed method overcomes the reduced performance of PCA due to its usage in general environment. An extensive experimentation is done using PCA for recognition of fingerspelling alphabets on a large dataset created by capturing 24 fingerspelling alphabets made by 20 users, with 6 samples of each alphabet, resulting in 2880 signs. Experiments show that the proposed image processing methodology gives better results compared to conventional PCA.

The paper is organised as follows. Section 2 gives an overview of PCA. The proposed method is described in Section 3. Section 4 contains the experimental details and results. The paper concludes with discussions and future avenues in Section 5.

2 Appearance based recognition using PCA: An Overview

The concept of Eigen Faces, an appearance based recognition methodology was given by [Turk and Pentland, 1991]. The PCA procedure consists of an offline training phase and online recognition phase.

In the training phase, the samples of the data on which the system needs to recognise are used to create an Eigen matrix which transforms the samples in the image space into points in the Eigen space. Let $S = \{s_1, s_2, s_3 \dots s_n\}$ be a set of n image samples of size $N \times N$ taken for training the system. The image samples are taken as gray scale images and are transformed from 2-D matrix to 1-D column vector of size $N^2 \times 1$ by placing the image matrix columns consecutively. The transformed vector denotes a point in N^2 dimensional space. These column vectors of n images are placed column wise to form the data matrix A of dimensions $N^2 \times n$.

$$A = \left[a^1, \dots a^n\right]$$

and $\overline{A}$ be the mean vector of the data vectors in the matrix A given by

$$\overline{A} = \frac{1}{n}\sum_k a^{(k)}$$

The vectors of the data matrix A are centred by subtracting the mean vector $\overline{A}$ from all the columns vectors of A to get the covariance matrix C of the column vectors, and is given by

$$C = [A - \overline{A}][A - \overline{A}]^T$$

having the dimensions $N^2 \times N^2$. Eigen vectors E of the covariance matrix C are computed. E is a matrix of $N^2 \times N^2$ dimensions, and eigen vectors are sorted on the basis of corresponding eigen values

$$V = \{v_1, v_2, v_3 \dots v_{N^2}\}$$

of the eigen vectors, to result in the transformation matrix X where

$$X = [X^1, \dots, X^{N^2}]$$

Here the corresponding eigen values of the vectors in X are ordered such that

$$v_{X^1} > v_{X^2} > \dots > v_{X^{N^2}}$$

The data matrix A is projected on to the eigen space to get P consisting n columns, where

$$P = X^T A$$

and each column in P signifies a transformed point in the eigen space of $N^2 \times N^2$ dimensions.

The advantage of transforming the image into eigen space is that the dimensions required to represent the image can be reduced to $m << N^2$, that is, the top m co-ordinate points of a transformed column vector in P matrix would represent the image with very little loss of information.

In the recognition phase, the image I to be recognised is converted to 1-D vector form J as mentioned above and is projected on to the eigen space to get Z where

$$Z = X^T J$$

and the Euclidian distance measure d between Z and all the projected samples in P is computed,

$$d = dist\left(Z\, P^k\right) = \left\|Z - P^k\right\| = \sqrt{\sum_{j=1}^{m}\left(Z - P_j^k\right)^2}$$

the sample label k of the sample in P with minimum d is taken as the recognised label of Z.

3 Proposed scheme

PCA has been used for recognition on constrained dataset. Using PCA on real data with good recognition rate is a challenge. Towards that goal, the proposed scheme incorporates a pre-processing step in the training and recognition phase of the PCA model. The images are processed and then are used for training and recognition. Smoothing and contrast enhancement of images are the processing steps that are proposed to be incorporated.

A well known adaptive process known as wiener filter is used for smoothing, where the process adapts itself to local image variance. When the variance is large, the wiener filter performs little smoothing, when the variance is small, it performs more smoothing, thereby preserving edges in the images. The results are often better than linear filtering, however, it does require more computation time than linear filtering.

The other proposed process extends the use of sobel filter. The sobel filter has the property of detecting the edges in the horizontal or vertical direction as specified. In this work, it is first used to get the horizontal edge mapped image which is added back to the image to get an enhanced image. Horizontal edge detecting sobel mask S is given as

$$S = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Block wise application of the above mentioned mask on an image results in horizontal edge image. A *filter* can be defined over block wise application of the mask on the image I as to get an edge image E,

$$E = filter(S, I)$$

which is in turn is added to the original image I to get an enhanced image K, where

$$K = I + E$$

The resulting image K is used in training and recognition phases.

Another image processing technique is histogram equalisation of images. This process results in a contrast adjusted image by increasing the local contrast. This process employs a monotonic, non-linear mapping which re-assigns the gray values of pixels in the input image such that the output image contains a uniform distribution of gray values (i.e. a flat histogram). This technique is useful in comparison processes because of its effective detail enhancement.

A gray levels reduction technique is also proposed, it increases the contrast in the image by reducing the number of gray levels in the image from 256 to 3 levels. This reduction in image gray levels is done by partitioning the gray level range into 3 sub ranges.

Let $I = \{g_1, g_2, g_3 \ldots g_r\}$ be an image containing r gray levels. If the gray levels are in predefined ranges, they are re-assigned to arbitrarily unique 3 gray levels L_1, L_2, and L_3, i.e.,

$$I = \begin{cases} g_i = L_1 \forall i = 1, \ldots p \\ g_i = L_2 \forall i = p+1, \ldots q \\ g_i = L_3 \forall i = q+1, \ldots r \end{cases}$$

The resulting image I contains only 3 gray levels and high contrast. After extensive experimentation, the gray level sub range markers p and q are taken as 99 and 174 respectively and substituting gray levels L_1, L_2, and L_3 are taken as 50, 150 and 200 respectively. All the images used in the experiment have 256 gray levels.

After application of the image processing technique, the resulting images are used for the training and recognition phases of the PCA model.

Thus, the system training algorithm is as follows.

Algorithm: Training
Input: Fingerspelling training images
Output: Fingerspelling image features, eigen vector matrix, feature matrix
Method:

Step 1: Apply any of the discussed pre-processing techniques to the training images.
Step 2: Transform the training images into column vector by appending the columns in the image consecutively.
Step 3: Build the data matrix A of image column vectors with a label vector L having the corresponding alphabet names of the image columns in A.
Step 4: Get mean column vector M of the data matrix A.
Step 5: Subtract mean M from each of the columns of A to result in mean centered matrix A.
Step 6: Compute the covariance matrix C of A as $C = AA^T$.
Step 7: Obtain eigen vectors matrix E and eigen values vector V of C.
Step 8: Rearrange the eigen vector columns in E as the corresponding eigen values in V are sorted in descending order.
Step 9: Project the centered matrix A onto E to get feature matrix $P = E^T A$.

Training ends.

Following is the algorithm designed for recognition.

Algorithm: Recognition
Input: Fingerspelling image I to be recognised, number of dimensions to be considered m, feature matrix P, eigen vectors matrix E, mean vector M, labels vector L
Output: Classification label of input image
Method:

Step 1: Apply the respective pre-processing technique on I
Step 2: Transform the processed image I into a column vector J by placing the columns in the image consecutively.
Step 3: Subtract the mean vector M from the image vector J, $J = J - M$.
Step 4: Project the image vector J onto the eigen matrix E to get the feature vector $Z = E^T J$.
Step 5: Compute the Euclidian distance d between the feature vector Z and all the column vectors in the feature matrix P considering only m elements in the vectors and identify the column having the minimum distance d.
Step 6: Obtain the label from vector L corresponding to the column identified in P having the minimum distance to Z.

Recognition ends.

4 Experimental Results

This section presents the results of the experiments conducted to substantiate the performance enhancement of PCA using the proposed methodology.

To robustly test the performance of the proposed work, large dataset of fingerspelling signs is built up.

4.1 Dataset of fingerspelling signs

User images were captured by a web cam with plain background. The alphabets ‘j’ and ‘z’ are dynamic signs consisting of hand movement and thus, only static signs are considered in this work. Hence those two alphabet signs are not considered. Signs of 24 fingerspelling alphabets made by 20 users were captured, with 6 samples of each alphabet, resulting in 2880 signs. A few sample images are given in Figure 1.

The sign images were cropped to get hand area images, Figure 2 shows the result of cropping the sample images in Figure 1. Since PCA requires all the vectors to be of same size, all the images are resized to 50×50 dimensions, Figure 3 presents the resized sample images. Figure 4 shows the result of application of smoothing using wiener filter to the samples images. Result of extended sobel filter application is in Figure 5. Histogram equalisation result is in Figure 6 and Figure 7 shows the result of gray levels reduction.

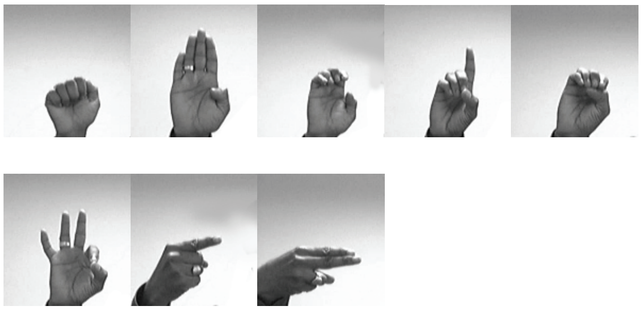

Figure 1: Captured images of alphabets a,b,c,d,e,f,g and h

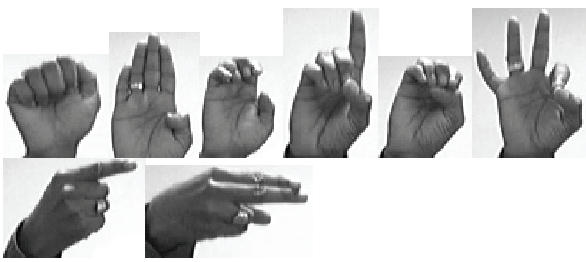

Figure 2: Segmented images of the alphabets

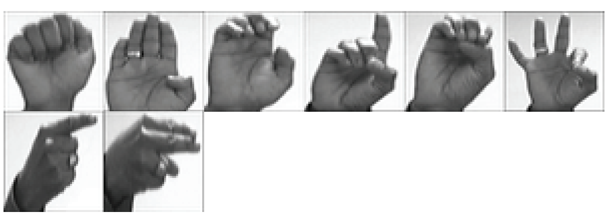

Figure 3: Resized images of the segmented images

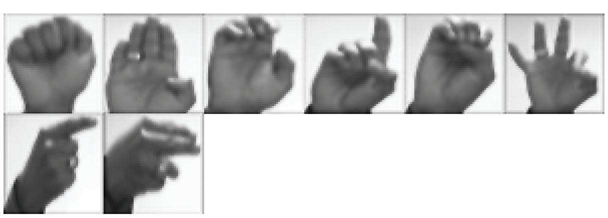

Figure 4: Result of smoothing operation using wiener filter.

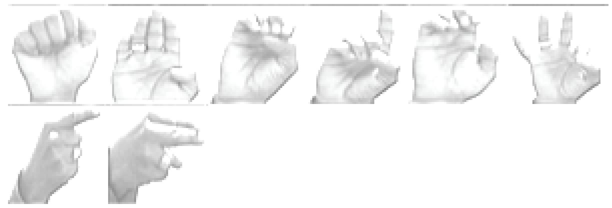

Figure 5: Result of extended sobel filter operation

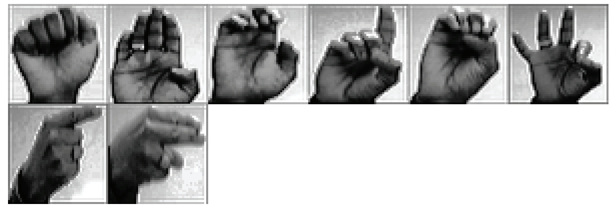

Figure 6: Result of Histogram equalisation

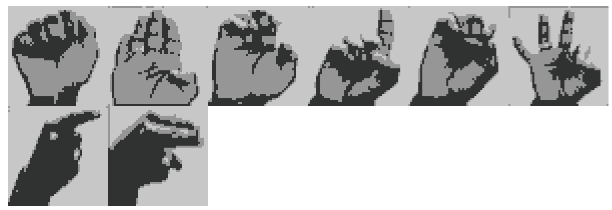

Figure 7: Application of gray levels reduction operation.

Out of the 6 samples for each alphabet from each user in the above mentioned dataset, 4 samples are considered for training the model (totalling 1920 images). Remaining 2 samples for each alphabet from each user is considered for testing in recognition phase (totalling 960 images).

During recognition, various dimensions (*m*) are considered while computing the Euclidian distance measure. Recognition rate is defined as the ratio of successful recognition of sample image to the number of samples used for testing in recognition phase. Results of the experiments are tabulated in Table 8. The best performances of different methods are highlighted.

Image Processing Methods	**Dimensions used for Recognition phase**							
	5	**10**	**15**	**20**	**25**	**30**	**35**	**40**
Original PCA	55.6	74.8	78.8	81.1	82.7	83.6	83.4	**84.3**
PCA with Smoothing	55.9	74.7	79.1	81.4	83.1	83.9	**84.5**	84.2
PCA with extended sobel filter	60.4	77.6	82.4	85.5	85.8	85.3	85.9	**86.4**
PCA with Histogram equalisation	55.1	75.1	80.0	81.8	83.8	85.4	**85.6**	85.6
PCA with gray levels reduction	57.8	77.1	81.0	84.5	84.4	**86.7**	86.5	86.2

Table 8: Recognition rates in percentages on application of image-processing techniques.

Performance results of various techniques are graphically represented in Figure 9. From Table 8 and Figure 9 it is clear that the proposed method achieves better result than standard processes that too at lower dimensions.

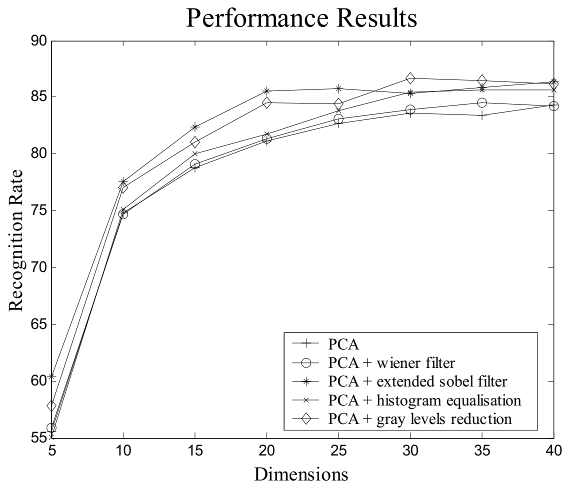

Figure 9: Performance in percentages on application of image processing techniques

[Salzberg, 1997] outlines an approach for experimental data setup to make a robust comparative analysis of classifying methods. Similar to it, the dataset of signs is divided into different combinations of subsets for training and testing phases of the methods. The dataset has 6 samples from each of the 20 users for each alphabet, out of which 4 samples are taken for training and 2 samples are taken for testing. Totally we get 15 (=$_6C_4$) combinations for dividing the dataset into a training set of 4 samples and remaining 2 samples for testing in recognition phase. After 15 runs of each algorithm on different sub sets, the average of the recognition rates is taken as the overall performance of the method.

This experimental setup is used to test the effectiveness of the proposed method, it is used along with the standard techniques mentioned above for comparing the performance of other appearance based procedures like 2D PCA, $(2D)^2$ PCA and Diagonal PCA and its effectiveness is apparent from Table 10 and its chart in Figure 11.

Image Processing Methods	**Appearance Based Methods**			
	PCA	**2D PCA**	**$(2D)^2$ PCA**	**Diagonal PCA**
Original method	86.17	88.66	84.52	88.98
Wiener filter	86.45	88.74	84.38	88.91
Extended sobel Filter	**88.75**	89.68	86.68	90.12
Histogram equalisation	87.03	89.52	86.1	**90.17**
Gray levels reduction	88.16	**89.74**	**86.86**	90.08

Table 10: Recognition rates in percentages on application of image-processing techniques

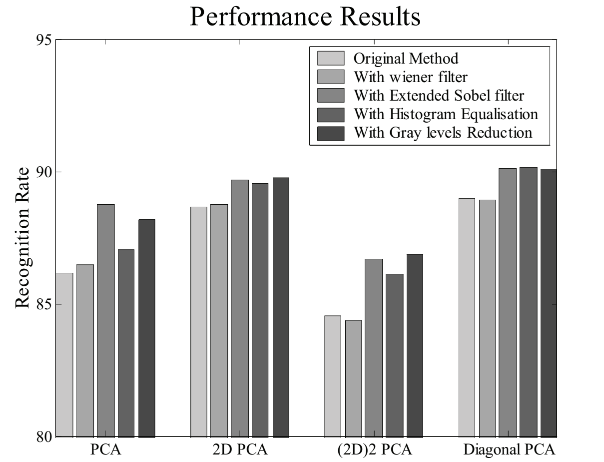

Figure 11: Performance rates in percentages on application of image-processing techniques

From Table 8, Table 10, Figure 9 and Figure 11, it is apparent that the proposed filter procedure increases the performance of the appearance based methods.

5 Discussion and Conclusion:

Application of image processing techniques definitely helps in increasing the performance of appearance based methodologies for fingerspelling recognition. Techniques like smoothing and contrast enhancement using wiener filter, extended sobel filter, histogram equalisation techniques and a novel gray levels reduction technique are applied resulting in much improvement in performance of PCA. To further show its effectiveness, other methods like 2D PCA, $(2D)^2$ PCA and Diagonal PCA are considered and their performances are compared, further reinforcing the effectiveness of the application of pre-processing techniques.

Although the gray sub range bounds in gray level reduction technique are chosen experimentally, further work in this direction would definitely give a basis for choosing the bounds and thereby making the technique adaptive.

References

[Birk and Moeslund, 1996] Henrik Birk and Thomas B. Moeslund. Recognizing Gestures From the Hand Alphabet Using Principal Component Analysis, *Master Thesis*, 1996.

[Birk *et al*., 1997] Henrik Birk, Thomas B. Moeslund and Claus B. Madsen. Real-Time Recognition of Hand Alphabet Gestures Using Principal Component Analysis, In *Proceedings of the 10th Scandinavian Conference on Image Analysis* (SCIA'97), Lappeenranta, Finland, 1997.

[Chalechale et al., 2005] Abdolah Chalechale, Farzad Safaei, Golshah Nagdy and Prashan Premaratne. Hand posture analysis for visual-based human-machine interface, In *Proceedings of the Workshop on Digital Image Computing* WDIC2005, pages 91-96. 2005.

[Fang *et al*., 2004] Gaolin Fang, Wen Gao, and Debin Zhao. Large Vocabulary Sign Language Recognition Based on Fuzzy Decision Trees, *IEEE Transactions on Systems, Man and Cybernetics – Part A: Systems and Humans*, 34(3):305-314, 2004.

[Gao *et al*., 2000] Wen Gao, Jiyong Ma, Jiangqin Wu, Chunli Wang. Sign Language Recognition based on HMM / ANN / DP*, International Journal of Pattern Recognition and Artificial Intelligence*, 14(5):587-602, 2000.

[Hernández *et al*., 2002] Jose L. Hernández-Rebollar, Robert W. Lindeman. and Nicholas Kyriakopoulos. A Multi-Class Pattern Recognition System for Practical Finger Spelling Translation, In *Proceedings of Fourth IEEE Int'l Conference on Multimodal Interfaces* (ICMI'02), pages 185-190, Pittsburgh, USA, October 14-16, 2002.

[Kang *et al*., 2005] Hyun Kang, Chang Woo Lee and Keechul Jung. Recognition-based gesture spotting in video games. *Pattern Recognition Letters*, 25(15):1701–1714. 2004.

[LaViola, 1999] Joseph J. LaViola Jr. A Survey of Hand Posture and Gesture Recognition Techniques and Technology, *Technical Report: CS-99-11*, 1999.

[Pavlovic *et al*, 1997] Vladimir I. Pavlovic, Rajeev Sharma and Thomas S. Huang. Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review, *IEEE Transactions on PAMI*, 19(7):667, 1997.

[Salzberg, 1997] Steven L. Salzberg, On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach, *Data Mining and Knowledge Discovery*, 1:317-328, 1997.

[Starner and Pentland, 1995] Thad Starner and Alex Pentland. Real-Time American Sign Language Recognition from video using Hidden Markov Model", In *Proceedings of International Workshop on Automatic Face and Gesture Recognition*, pages 189-194, Zurich, Switzerland, 1995.

[Turk and Pentland, 1991] Matthew Turk and Alex Pentland. Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, 3:72-86. 1991.

[Yang *et al*, 2004] Jian Yang, David Zhang, Alejandro F. Frangi and Jing-yu Yang. Two-dimensional PCA: A new approach to appearance based face representation and recognition, *IEEE Transactions on PAMI*, 26 (1):131–137, 2004.

[Zhang and Zhou, 2005] Daoqiang Zhang and Zhi-Hua Zhou. $(2D)^2$ PCA: 2-directional 2-dimensional PCA for efficient face representation and recognition, *Journal of Neurocomputing*, 69(1-3):224-231, 2005.

[Zhang *et al*., 2006] Daoqiang Zhang, Zhi-Hua Zhoua and Songcan Chen, Diagonal principal component analysis for face recognition, *Pattern Recognition* 39:140 – 142, 2006.