SBRN'98 – Simpósio Brasileiro de Redes Neurais, Belo Horizonte, Minas Gerais, dezembro de 1998.

Design of Radial Basis Function Network as Classifier in Face Recognition Using Eigenfaces

Carlos Eduardo Thomaz
Raul Queiroz Feitosa
Álvaro Veiga
PUC RJ- Pontifícia Universidade Católica do Rio de Janeiro
Departamento de Engenharia Elétrica
Rua Marquês de São Vicente, 225, 22453-900 Rio de Janeiro, RJ, Brasil
{cethomaz,raul,alvf}@ele.puc-rio.br

Abstract

In this paper we investigate alternative designs of a Radial Basis Function Network acting as classifier in a face recognition system. Input to the RBF network is the projections of a face image over the principal components. A database of 250 facial images of 25 persons is used for training and evaluation. Two RBF designs are studied: the forward selection and the gaussian mixture model. Both designs are also compared to the conventional Euclidean and Mahalanobis classifiers. A set of experiments evaluates the recognition rate of each method as a function of the number of principal components used to characterize the image samples. The results of the experiments indicate that the gaussian mixture model RBF achieves the best performance while allowing less neurons in the hidden layer. The gaussian mixture model approach shows also to be less sensitive to the choice of the training set.

1. Introduction

In an increasingly computerized world, there is an overwhelming demand for automated personal identification systems. In the past few years, sophisticated methods for either verifying or recognizing the identity of an individual have been proposed, well beyond the password authentication schemes commonly employed in automated teller machines, telephone calling and credit cards. Most of those methods are based on the recognition of physiological characteristics such as hand shape, fingerprint, retinal pattern, speech and the whole face[4]. Identity verification based on face features has important advantages when compared to other approaches, particularly in applications where the subject does not wish to be explicitly identified, e.g., bank/store security, expert identification, witness face reconstruction, etc.

A central issue in pattern recognition in general, and in face recognition in particular, is the well known problem of dimensionality reduction. Face images are highly redundant, since every individual has one mouth, one nose, two eyes and so on. Instead of using N intensity values for a N pixel image, it is generally possible to characterize an image instance by a set of M features, for M<<N. The set of images samples of the face of the same individual defines a class. The features must be chosen in such a way, that it is possible to identify the right class of a face image based only on those features.

This work studies a face recognition system consisting of a PCA stage which inputs the projections of a face image over the principal components into a RBF network acting as a classifier. The main concern is to analyze how different network designs perform in a PCA+RBF face recognition system.

For performance evaluation a database with 250 images - 10 photos of 25 individuals - is used as input to the system. The results are compared with the performance of more conventional schemes, where the RBF network is replaced by the Euclidean and the Mahalanobis distances.

Two RBF designs are studied : the forward selection and the gaussian mixture model [7]. The first method optimizes the recognition performance by varying the number of neurons in the hidden layer. In the second method, the number of neurons are input to the training algorithm, which estimates the center and the width of the gaussian activation functions by maximum likelihood of the input data density.

In the experiments carried out in this work, the gaussian mixture model optimization achieves the best performance even using far less neurons than the forward selection algorithm. The results indicate also that the gaussian mixture model design is less sensitive to the choice of the training set. Both RBF designs show a better performance than the conventional distance classifiers.

The next section presents an overview of related works. PCA applied to the problem of face recognition is then briefly discussed. The RBF network designs considered in this work are then described along with the Euclidean and Mahalanobis classifiers. The remaining sections describe the experiments carried out to evaluate the performance of each approach and discuss the results.

2. Previous work on face recognition

Earlier face recognition systems were mainly based on geometric facial features and template matching [20,22]. In those works a face was characterized by a set of features such as mouth position, chin shape, nose width and length which are potentially insensitive to illumination conditions. Brunelli et al. (1993) [20] compared this approach with a traditional template matching scheme which produced higher recognition rates for the same face database (90% against 100%). Cox, Ghosn and Yianilos (1996) [11] proposed a mixture distance technique which achieved the best reported recognition rate among the geometric feature approaches using the same database. Those results were obtained in a experiment where the features were extracted manually.

Turk and Pentland (1991) [15] use the projections of the face images onto the principal components of the training images as the face features. It achieves recognition rates around 96%, 85% and 64% respectively for lighting, orientation and scale variation. Recognition rate around 95% are reported by Pentland et al. (1994) [2] for a database consisting of 3000 accurate registered and aligned faces.

Samaria & Harter (1994) [10] presented an approach based on Hidden Markov Models which achieved a recognition rate of 95% for the ORL database at the expense of a high computational overhead.

Available results on Neural Network based approaches [8] come from experiments with few individuals, what makes it difficult to compare with other reported approaches.

All those works rely on a preprocessing to detect a face in a scene and to compensate for variation of lighting, position, rotation and scale.

The work reported here studies a face recognition system consisting of a standard PCA used for dimensionality reduction, followed by a RBF network acting as a classifier. As in the most approaches mentioned before, the database used for the evaluation contains face images with moderate lighting, rotation, scale and viewing variation. A previous work [6] has indicated that a RBF network performs better than conventional distance classifiers. The present work focus on the study of alternative network optimization to maximize the recognition rate.

The RBF network for face recognition has already been studied by Howell and Buxton. Instead of using principal components, they use either the image itself , or the output of a Difference of Gaussian filter and the output of a Gabor filter [1] as the input to the RBF network. Vallentin, Abdi and Edelman [9] used PCA followed by a RBF network to model how faces are stored in human memory. Their work neither compares the performance of the RBF network with any other classifier nor analyses alternative network designs.

The main contribution of this work is a better understanding of how the parameters of the RBF network can be optimized to maximize its performance for the face recognition task.

3. Using PCA for dimensionality reduction

PCA is a well known statistical technique for dimensionality reduction. It was first suggested for the characterization of human faces by Kirby and Sirovich [14] and later extended by Turk and Pentland [15]. Many refinements to the original idea were further introduced [2,3,5,19] . Many psychologists and neurophysiologists use PCA to model the way the human brain stores, retrieves and recognizes faces [9,16,17,18].

An input image can be treated as a N-dimensional feature vector, where N is the number of pixels of the image. The intensity of each pixel is used as a single feature. Thus, an image can be considered as a sample point in an N-dimensional space, called in this context image space. Image instances can be represented by a random N-dimensional vector $\mathbf{x_i}$ obtained by concatenating the rows of the image matrix.

PCA generates a new orthonormal basis for the image space, where each component is not correlated with any other component. Each vector for the new basis is chosen so that the variance of the projection along it is maximized, subject to the orthonormality condition.

The procedure to generate such a basis can be summarized as follows. Given a set of k images $\{\mathbf{x_i}\}_{i=0,\ldots,K}$ a training set matrix $\mathbf{X}=[\mathbf{x_i}]$ can be built , where each line is a vector $\mathbf{x_i}$. Without loss of generality, $\mathbf{x_i}$ can be considered as a zero mean vector, by replacing the images $\mathbf{x_i}$ by $\mathbf{x_i} - \mathbf{E}[\mathbf{x_i}]$. The new basis vectors are obtained by solving the eingenvalue problem:

$$\Lambda = P^T \Sigma_X P$$

where $\Sigma_{\mathbf{x}}$ is the covariance matrix, defined by

$$\Sigma_X = X * X^T ,$$

P is the eigenvector matrix of $\Sigma_{\mathbf{x}}$ and Λ is the corresponding diagonal matrix of eigenvalues. The eigenvectors are called in the literature of face recognition eigenfaces. A face image is then described as a linear combination of eigenfaces.

In PCA the eigenvectors corresponding to the M largest eigenvalues (for some M) are selected to form a lower dimensional subspace, the face space. It is a proven result [21] that the residual reconstruction error generated by dismissing the N-M components is given by:

$$e^2(x_i) = \sum_{j=M+1}^{N} \lambda_j^2$$

where λ_j is the eigenvalues corresponding to the dismissed eigenvectors. For face images, the eigenvalues decay in general exponentially, so that the residual reconstruction error are low even for small M.

4. Classification schemes

A classifier is essentially a mapping of the input space onto a set of classes. The literature on pattern recognition presents a huge number of schemes to construct this mapping from data [13].

In the present work, two basic schemes were tested: RBF networks [12] and minimum distance to centroids classifiers with two different distance measures - Euclidean and Mahalanobis.

4.1 The RBF network classifier

The RBF network is an one hidden layer neural network with several forms of radial basis activation functions. The most common one is the Gaussian function defined by,

$$f_j(\mathbf{x}) = \exp\left(-\frac{\left\|\mathbf{x} - \boldsymbol{\mu}_{\mathbf{j}}\right\|^2}{2\sigma_j^2}\right)$$

where σ is the width parameter, μ is the vector determining the center of basis function f and $\mathbf{x}$ is the d-dimensional input vector .

In a RBF network, a neuron of the hidden layer is activated whenever the input vector is close enough to its center vector μ. There are several techniques and heuristics for optimizing the basis functions parameters and determining the number of hidden neurons needed to best classification [7]. This work discuss two training algorithms : forward selection (FS) and gaussian mixture model (MM). The first one allocates one neuron to each group of faces of each individual and if different faces of the same individual are not close to each other, more than one neuron will be necessary. The second training method regards the basis functions as the components of a mixture density model, whose parameters μ and σ are to be optimized by maximum likelihood [7]. In this latter, the number K of basis functions is treated as an input to the model and is typically much less than the total number of input data points $\{\mathbf{x}\}$.

The second layer of the RBF network, which is the output layer, comprises one neuron to each individual. Their output are linear functions of the outputs of the neurons in the hidden layer and is equivalent of a OR operator. The final classification is given by the output neuron with the greatest output.

With RBF networks, the regions of the input space associated to each individual can present an arbitrary form. Also, disjoint regions can be associated to the same individual to render, for example, very different angles of vision or different facial expressions.

4.2 Centroid classifier

A centroid classifier works in a far simpler and interpretable way but suffers from limitations in the form of the regions of the input space associated to each individual.

First a reduced sample is constructed by replacing each group of face images of an individual by its centroid which defined as the mean of the vectors of features for this group. Then, a new face will be classified to the closest centroid of the reduced sample.

In this work, two distance measures where compared. The first one is the usual Euclidean distance given by,

$$\mathrm{dist}(x,i) = (x - \mu_i)^T (x - \mu_i)$$

where $\boldsymbol{\mu}_{\mathbf{i}}$ is the centroid corresponding to the i^{th} individual. The second distance measure is the Mahalanobis distance, defined as,

$$\mathrm{dist}(x,i) = (x - \mu_i)^T \Sigma_x^{-1} (x - \mu_i)$$

which is the same as the Euclidean distance applied to the standard data (i.e. the covariance is the identity matrix). In case of the present application, the distance is computed over uncorrelated data and consequently $\Sigma_{\mathbf{x}}$ is a diagonal matrix with diagonal elements given by the eigenvalues $\boldsymbol{\lambda}_{\mathbf{i}}$. In this case, the Mahalanobis distance assumes the simpler form, given by

$$\mathrm{dist}(x,i) = \sum_j \lambda_j^{-1} (x_j - \mu_{ij})^2 \; .$$

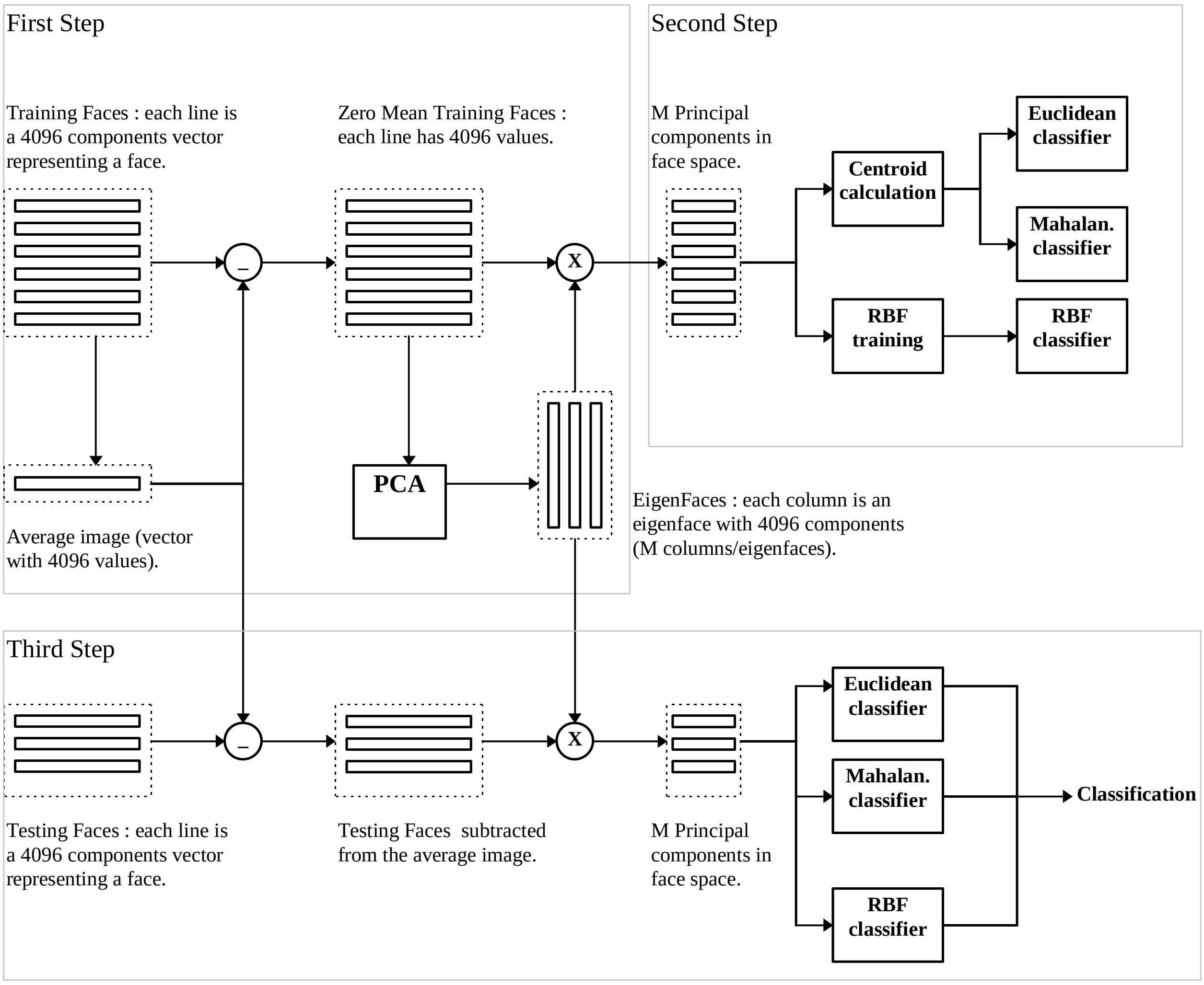

Figure 1 - Experiment Design carried out in the work.

5. Experiment design

The experiments to evaluate the methods make use of the ORL face database[1]. It contains a set of face images taken between April 1992 and April 1994 at the Olivetti Research Laboratory in Cambridge, U.K, with ten images for each of 40 individuals, a total of 400 images. In some cases the images were taken at distinct times, with varying lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All images were taken against a dark homogeneous background with the person in a upright frontal position, with tolerance for some tilting and rotation of up to about 20 degrees. Scale varies about 10%. The original size of each image is 92x112 pixels, with 256 gray levels per pixel. For implementation convenience all images were first resized to 64x64 pixels. Due to limitation of the available computational capacity the experiments took a subset containing 250 images - ten images for each of 25 individuals.

Before being used in the experiments all the images were represented as a vector, which is obtained by simply concatenating the rows of the image matrix.

Figure 1 illustrates the experiments carried out in this work. Each experiment consists of three steps: generation of the eigenfaces, training the classifier and testing the classifier.

In the first step the eigenfaces are generated. A training set is selected , by choosing randomly 7 images for each individual. The remaining 3 images are used later to test the method (step 3).Then, the average image of all training faces is calculated and subtracted from each face. Afterwards, the training matrix composed of the zero mean image vectors is used as input to compute the PCA, and a set of M different eigenfaces is generated.

In a second step the classifiers are trained. For each individual the average point of the 7 training images - the centroids - are calculated and later used by the Euclidean and Mahalanobis classifiers in the final classification step. To train the RBF classifier, the two basis function optimizations are considered. In the first one - forward

[1] The ORL database is available free of charge, see http://www.cam-orl.co.uk/facedatabase.html

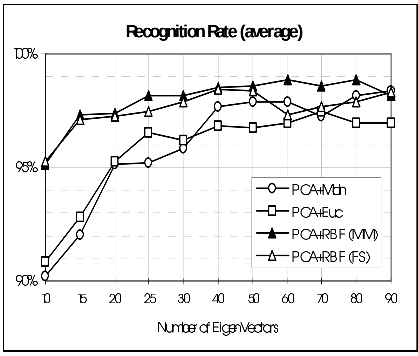

Figure 2

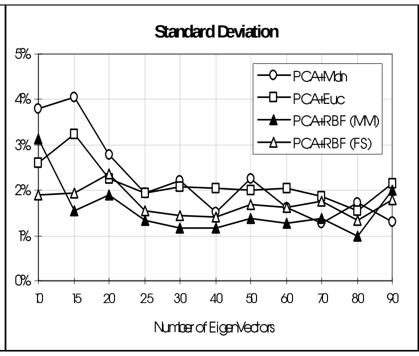

Figure 3

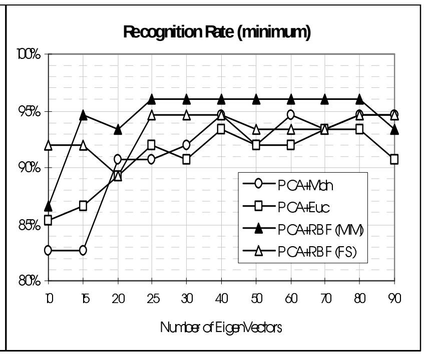

Figure 4

selection (FS) - the 7 training images per individual chosen in the first step are grouped into two subsets: 5 images to train the network and 2 images for the validation. This approach is called the “hold out method” [7]. After being projected onto the face space, the training and validation face images are used to train the RBF network. The centers of the radial basis functions are constraint to be given by the input data vectors and each radial basis function has a common variance, equals to 1. The training process creates interatively a RBF network one hidden neuron at a time. These neurons are added to the network until the sum-squared error corresponding to the validation set reaches the minimum. In the other RBF design - gaussian mixture model (MM) - the 7 training images per individual chosen and projected onto the face space are used to form a representation of the probability density of the input data. The number K of basis functions are used as another parameter to the model and then the unsupervised procedure for optimization the gaussian parameters depends only on the input data set. The basis function centers are determined by fitting the mixture model with circular covariances using the EM (expectation-maximization) algorithm and then their respective widths are set to the maximum inter-center square distance. The hidden to output weights that give rise to the least squares solution are determined using the pseudo-inverse [7]. Both RBF networks are trained to produce a 1 in the output unit corresponding to the face presented at the input layer and a 0 in every other unit.

In the third step, the performance of the classifiers is evaluated. Each test image is projected onto the eigenfaces obtained in the first step and input to each classifier (figure 1). The true/false recognitions are then stored for the computation of the recognition rate.

This three-steps procedure was repeated 25 times using different training and testing sets. The number of principal components used to represent the images (M) were varied and took eleven different values: 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, and 90 - a total of 11x25 = 275 runs were executed. Another parameter varied in the experiments was the number of hidden neurons K used as input to the RBF gaussian mixture model. For each number of eigenfaces this parameter has taken seven different values: 25, 30, 40, 50, 60, 70, 80.

6. Experiment results

The results of the experiments are summarized in figures 2 to 4. Figure 2 shows the average recognition rate for each classifier as a function of the number of eigenfaces. The curves representing the RBF gaussian mixture model (MM) correspond to a network with K=60 hidden neurons, where the classifier has reached its best performance.

For less than 20-30 eigenfaces, the both RBF classifiers have clearly the best recognition rate. It is fair to say on the basis of the results of figure 2, that the two RBFs reach the peak performance for a small number of eigenfaces than the distance classifiers.

For more than 25 eigenfaces all four classifiers have near performances - between 96% and 99%, although the RBF gaussian mixture model classifier shows the best results.

Another important aspect revealed by the experiments is shown in figure 3. It presents the standard deviation of the recognition rate computed for the 25 runs for each number of eigenfaces considered. This graph indicates how sensitive are the results to the choice of the training and testing sets. The RBF gaussian mixture model classifier presents again the lowest standard deviation for all but one number of eigenfaces.

As a complement to the figure 3, the minimum performance among the 25 runs is plotted in figure 4 as a function of the number of eigenfaces.The RBF gaussian mixture model classifier presents again the best performance in the experiments.

Another advantage of the MM approach is shown by the figure 5. The bars on the graph represent the range of the number of added hidden neurons by the FS training - the mark indicates the average value - in the 25 runs, as a function of the number of eigenfaces. It can be observed that the FS training built a network with more hidden neurons (around 100 neurons) than the MM training (60 neurons).

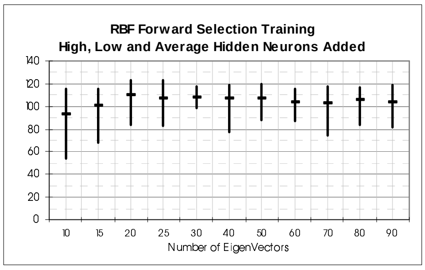

Figure 5 - RBF forward selection training

As a final evaluation, the RBF classifiers have a better performance than the conventional classifiers. Among the RBF designs analyzed in this work, the mixture model approach produces the best recognition rates with much less hidden neurons.

7. Conclusions

The performance of a face recognition method using PCA for dimensionality reduction and RBF networks are evaluated. Experiments using a database with 250 face images were carried out to compare the performance of the RBF classifiers with the performance of an Euclidean and a Mahalanobis classifiers.

The results indicated that both RBF classifiers reach its peak performance - around 98% of recognition rate - for a lower number of eigenfaces. The RBF classifiers presented also a better performance regarding the sensitiveness to the choice of the training and testing sets.

Comparing the two RBF designs, the RBF gaussian mixture model overperformed the RBF forward selection in all the result analyses taken. An important aspect also revealed by the experiments was the number of hidden neurons used in both designs. The RBF gaussian mixture model presented the best results with much less neurons.

Future work will evaluate the robustness of the PCA+RBF method separately against variations of lighting, view, scale, rotation and facial expression, and further test the method on larger face databases. This will give a more definitive conclusion about the performance of the classifier and its sensitiveness to the choice of the training and testing sets.

References

[1] A. J. Howell and H. Buxton, "A scaleable approach to face identification", Proceedings of the International Conference on Artificial Neural Networks, pp. 257-262, 1995.

[2] A. Pentland, B. Moghaddam, and T. Strainer, "View-based and modular eigenspaces for face recognition", Proc. of IEEE Conference on Computer Vision and Pattern Recognition, June, 1994.

[3] A. Pentland, et all, "Experiments with Eigenfaces", International Joint Conference on Artificial Intelligence, Chamberry, France, August ,1993.

[4] B. Miller, "Vital signs of identity", IEEE Spectrum, Feb. , pp. 22-30, 1994.

[5] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Detection", 5th Int. Conference on Computer Vision, June, 1995.

[6] C.E. Thomaz, R.Q. Feitosa and A.Veiga, "Face Recognition Using Eigenfaces and a Radial Basis Function Network", submitted to SIBGRAPI, 1998.

[7] C. M. Bishop, "Neural Networks for Pattern Recognition", Oxford, England: Oxford Press, 1996.

[8] D. Valentin, H. Abdi, A. O'Toole and G. Cottrell, "Connectionist Models of Face Processing : A Survey", Pattern Recognition, Vol. 27, pp. 1209-1230, 1994.

[9] D. Valentin, H. Abdi and B. Edelman, "What Represents a Face: A Computational Approach for the Integration of Physiological and Psychological Data", Perception, Vol. 26, 1997.

[10] F. Samaria & A. Harter, "Parametrisation of a stochastic model for human face identification", Proc. 2nd IEEE workshop on Applications of Computer Vision, 1994.

[11] I.J. Cox, J. Ghosn and P.N Yianilos, "Feature-Based Face Recognition Using Mixture Distance", IEEE Conference on Computer Vision and Pattern Recognition, June 1996.

[12] J. Moody and C. Darken, "Fast learning in networks of locally-tuned processing units", Neural Computation, 1, pp. 281-294, 1988.

[13] J. Schurman, "Pattern Classification: A Unified View of Statistical and Neural Approaches", John Willey & Sons, 1996.

[14] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces", IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 12, No. 1, Jan. 1990.

[15] M. Turk, A. Pentland, "Eigenfaces for Recognition", Journal of Cognitive Neuroscience, Vol. 3, No. 1, 1991.

[16] P. J. B. Hancock, A. M. Burton and V. Bruce, "Face processing: human perception and principal components analysis", Memory and Cognition, 1996.

[17] P. J. B. Hancock, A. M. Burton and V. Bruce, "Preprocessing Images of Faces: Correlations with Human Perceptions of Distinctiveness and Familiarity", Proc. of IEE 5th International Conference on Image Processing and its Applications, 1995.

[18] P. J. B. Hancock, V. Bruce and A. M. Burton, "Testing Principal Component Representations for Faces", Proc. of 4th Neural Computation and Psychology Workshop, 1997.

[19] R. P. N. Rao, D. H. Ballard, "Natural basis functions and topographic memory for the face recognition", in International Joint Conference on Artificial Intelligence, Montreal, Canada, pp. 10-17, 1995.

[20] R.Brunelli and T.Poggio, "Face Recognition : Features versus Templates", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, pp. 1042-1052, October 1993.

[21] S. Romdhami, "Face Recognition using Principal Components Analysis", MSc Thesis, http//eeapp.elec.gla.ac.uk/~romdhami /pca, 1996.

[22] Samal and P.A. Iyengar, "Automatic recognition and analysis of human faces and facial expressions : A survey", Pattern Recognition, vol. 25, pp. 65-77, 1992.