ISAS'99 – International Conference on Information Systems Analysis and Synthesis, Orlando, EUA, agosto de 1999.

Comparing the Performance of the Discriminant Analysis and RBF Neural Network for Face Recognition

Raul Queiroz Feitosa[1,2]
Carlos Eduardo Thomaz[2]
Álvaro Veiga[2]

UERJ[1]
University of Rio de Janeiro
Department of Computer Science
r. São Francisco Xavier, 524,
20559-900, Rio de Janeiro, Brazil

PUC RJ[2]
Catholic University of Rio de Janeiro
Department of Electrical Engineering
r. Marquês de São Vicente 225,
22453-900, Rio de Janeiro, Brazil

{raul, cethomaz, alvf}@ele.puc-rio.br

ABSTRACT

Among the many methods proposed in the literature for face recognition, those relying on the so called eigenfaces have been explored with great interest in the last few years. In those methods the face images are initially subjected to a PCA stage (Principal Component Analysis) for dimensionality reduction and then applied to a classifier. This work evaluates and compares two eigenface based face recognition systems, using two different classifiers: a) the LDA (Linear Discriminant Analysis) classifier, and b) a Gaussian Mixture Model RBF (Radial Basis Function) neural network. Extensive experiments using the ORL Face Database indicate that the more general model underlying the RBF classifier does not bring any significant performance improvement compared with the simpler and less computation intensive LDA approach.

Keywords: Face Recognition, Eigenfaces, PCA, LDA and RBF neural network..

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 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 p features, for $p << n$. The set of face images 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 two face recognition systems. Both systems have a PCA stage for dimensionality reduction. It computes the projections of a face image over the principal components (the so called eigenfaces). The face images represented in this lower dimensional space (the face space) are the input to a classifier. In the first system a linear classifier based on the Linear Discriminant Analysis (LDA) is used. The second uses a nonlinear classifier based on a RBF neural network.

The LDA approach assumes that the population of each group, corresponding to the different images of the same person, is normally distributed around its centroid in the discriminant space. It further assumes that all groups have the same covariance matrix. Independent experiments [8,13,22] have indicated that the LDA approach reaches the best performance among the proposed linear methods for face recognition using eigenfaces.

The RBF network is an one hidden layer neural network with radial basis activation functions. The most commonly used activation function is the Gaussian function. There are several techniques and heuristics for optimizing the basis functions parameters and determining the number of hidden neurons for best classification rates. This work implements the Gaussian Mixture Model algorithm to train the network [6]. The model describes samples of each individual as mixture of Gaussian distributions, allowing the RBF classifier to describe regions with arbitrary form, including the case of disjoint regions. Since the LDA approach uses a single Gaussian to represent the population of a group in the input space, the RBF network has the potential for a better performance.

This work analyzes both face recognition systems to determine how much the superior generality of the model underlying the RBF Gaussian Mixture Model network impacts the performance when compared to the LDA approach.

The Olivetti Face Database (ORL) containing 10 photos of 40 individuals was used in a set of experiments carried out to compare the performance of both face recognition systems. In each experiment 200 photos (5 of each individual) were chosen at random to calculate the eigenfaces and to train both classifiers. The other 200 photos were used for performance evaluation.

The results of the experiments carried out in this work showed no clear superiority of RBF or LDA methods. The ability of the RBF network to use more than one Gaussian to describe the population of each group brought no significant performance improvement, when compared to the less computation intensive LDA classifier.

2. PREVIOUS WORK ON FACE RECOGNITION

Earlier face recognition systems were mainly based on geometric facial features and template matching [20,21]. 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. [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 [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 an experiment where the features were extracted manually.

The Principal Component Analysis technique 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]. Several psychologists and neurophysiologists use PCA to model the way the human brain stores, retrieves and recognizes faces [9,16,17,18]. The experiments of Turk and Pentland [15] achieved recognition rates around 96%, 85% and 64% respectively for lighting, orientation and scale variation. Recognition rate around 95% are reported by Pentland and Moghaddam (1994) [2] for a database consisting of 3000 accurate registered and aligned faces.

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

All those works, as well as this one, 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 two face recognition systems consisting of a standard PCA used for dimensionality reduction, followed respectively by a LDA classifier and by a RBF network.

The LDA approach was originally proposed by Swets and Weng [8]. In the first step a n-pixel face image is projected onto the face subspace, whose basis is given by the p ($p < n$) eigenvectors of the sampled covariance matrix corresponding to the highest eigenvalues (the eigenfaces). In the second step the LDA maps the projection of the input image on the face space onto a discriminant space. A set of discriminant functions, based on general quadratic distance measures, is then used as classifier.

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.

The main contribution of this work is a better understanding of the models underlying the LDA and RBF classifiers for the face recognition application. The experiments carried on with this purpose have shown that the simpler and less computational expensive LDA classifier has almost the same performance as the RBF Gaussian Mixture neural network.

3. USING PCA FOR DIMENSIONALITY REDUCTION

An input image with n pixels can be treated as a point in a n-dimensional axes space, called, in this context, the image space. The coordinates of this point represent the values of each pixel of the image i and form a feature vector $\mathbf{x}_i = [x_{i1},..., x_{in}]^T$ obtained by concatenating the columns of the image matrix. For this representation to make sense for classification, it is necessary that two images that look alike correspond to two close points in the image space.

The intensities of adjacent pixels of a face image are usually highly correlated, so that a face image contains much redundant information. As a consequence, the image-points will not occupy evenly the image space and can be projected onto a lower dimensional subspace without significant loss of information.

Lets consider a set of K images and define the data matrix $\mathbf{X}$ with rows given by the feature vectors $\mathbf{x}_i^T$, $i = 1,..,K$,

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1^T \\ \vdots \\ \mathbf{x}_K^T \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{K1} & \cdots & x_{KN} \end{bmatrix}. \tag{1}$$

A column of matrix $\mathbf{X}$ represents the values of a particular pixel observed across the K images. Without loss of generality the mean of each column can be assumed to be zero. This can be achieved by subtracting the column mean from each column. With zero mean images the sampled covariance matrix of the images is given by the n x n matrix,

$$\mathbf{S}_{\mathrm{X}} = \frac{1}{K-1}\mathbf{X}^T\mathbf{X}. \tag{2}$$

Consider now the representation of those images on another basis (matrix $\mathbf{Y}$). This is achieved by applying a linear transformation:

$$\mathbf{y}_i = \mathbf{A}^T\mathbf{x}_i , \tag{3}$$

where the columns of the transformation matrix $\mathbf{A}$ are the axes of the new basis expressed in the original basis. This corresponds geometrically to project the points $\mathbf{x}_i$'s over the new axes. The set of images in matrix $\mathbf{X}$ can be projected on the new basis by the transformation

$$\mathbf{Y} = \mathbf{XA}. \tag{4}$$

Since the mean of $\mathbf{x}_i$ is zero, the mean of $\mathbf{y}_i$ will also be zero. Moreover, the sample covariance matrix expressed in the new basis will be obtained by,

$$\mathbf{S}_{\mathrm{Y}} = \frac{1}{K-1}\mathbf{Y}^T\mathbf{Y} = \mathbf{A}^T\mathbf{S}_{\mathrm{X}}\mathbf{A}. \tag{5}$$

The new basis, defined by transformation matrix $\mathbf{A}$ columns, can be determined so as to satisfy several criterion, motivating different approaches for dimensionality reduction. In this work we concentrate on the maximum variance of projections given by the Principal Components Analysis (PCA) technique.

PCA gives the set of axes over which the projections of the whole sample has the maximum dispersion, subjected to the orthonormality condition. Figure 1 illustrates the axes found by PCA for a 2 dimensional example. Note that the first axis **e’**$_1$ is in the direction of the maximum variance. Axis **e’**$_2$ is in the next direction of maximum variance and is orthogonal to **e’**1.

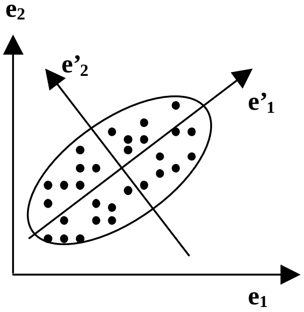

Figure 1: The geometric interpretation of principal components for a two dimensional feature space.

The axes of the new orthonormal basis having those properties are the principal components. They are given by the eigenvectors of $\mathbf{S_X}$. In the context of face recognition those axes are frequently named eigenfaces. Therefore, the new features associated with the columns of **Y** will be uncorrelated with variance given by the eigenvalues of $\mathbf{S_X}$ or, equivalently, $\mathbf{S_Y}$ will be a diagonal matrix.

Important for the dimensionality reduction is the property that the proportion of the total variance explained by one eigenface is given by the ratio of the corresponding eigenvalue to the trace of $\mathbf{S_Y}$. Suppose that the eigenfaces of $\mathbf{S_X}$ are ranked in eigenvalues decreasing order. To reduce dimensionality, only projections over the p eigenvectors ($p < n$) corresponding to the p greatest eigenvalues are considered.

By dismissing the projections over the remaining n-p eigenvectors, with lower eigenvalues, will result in a reconstruction error, whose mean square value is given by

$$error(p) = \sum_{j=p+1}^{n} \lambda_j \,, \tag{6}$$

where λ_j is the jth eigenvalue of $\mathbf{S_X}$.

The number p of principal components retained is chosen so that the error is less than some given percentage m of the sum of all eigenvalues. On this way, p is the minimum integer number for which the condition below holds:

$$\frac{\sum_{j=p+1}^{n} \lambda_j}{\sum_{j=1}^{n} \lambda_j} < m \,. \tag{7}$$

Therefore, the columns of the transformation matrix **P** generated by PCA are the first p eigenvectors of $\mathbf{S_X}$. Swets and Weng have called the so selected features the *Most Expressive Features* (MEFs) [8], since they give the minimum mean square reconstruction error.

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 [12].

This work evaluates and compares two eigenface based recognition systems, using two different classifiers: a) the LDA (Linear Discriminant Analysis) classifier, and b) a RBF (Radial Basis Function) neural network.

The LDA approach

The classifier based on the Linear Discriminant Analysis can be seen as involving two stages. In the first stage a new (discriminant) lower dimensional basis is chosen. In the second stage the projections of the sample on the new discriminant basis are fed to a distance classifier.

The goal of PCA on choosing a lower dimensional basis is to minimize the reconstruction error. This is not the major concern in pattern recognition applications, whose goal is to maximize the recognition rate. Figure 2 shows two dimensional example, where the projections on the principal components (**e’**$_1$, **e’**$_2$) can not properly separate the two groups. The axis over which the projections of the both populations are clearly separated is the **e”**$_1$ axis, quite different from the basis provided by PCA. The set of axes which satisfies the condition of maximal separation is called the Most Discriminant Features (MDFs) [8].

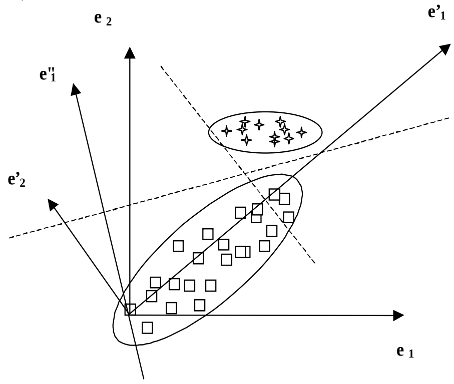

Figure 2: The geometric interpretation of Most Discriminant Features for a two dimensional feature space.

The Linear Discriminant Analysis (LDA) provides a procedure to determine a set of axes whose projections of different groups have the maximum separation. This procedure can be described as follows.

Suppose that the sample consists of K face images from where k_j images are of individual j, for $j = 1,...,g$, so that $K = k_1 +...+ k_g$. Let $\overline{\mathbf{x}}_j$ be the mean feature vector of images from individual j, defined by

$$\overline{\mathbf{x}}_j = \frac{1}{k_j} \sum_{x_k \in individual\ j} \mathbf{x}_k \,. \tag{8}$$

The *sampled between individuals covariance* matrix is defined as

$$\mathbf{B} = \sum_{j=1}^{g} \left(\overline{\mathbf{x}}_j - \overline{\mathbf{x}}\right)\left(\overline{\mathbf{x}}_j - \overline{\mathbf{x}}\right)^T \,, \tag{9}$$

where $\overline{\mathbf{x}}$ is the grand mean vector of all observations of all groups. Let the *sampled within individuals covariance* matrix be defined as

$$\mathbf{W} = \sum_{j=1}^{g} \sum_{x_k \in \, indvidual \; j} \left(\mathbf{x}_k - \bar{\mathbf{x}}_j\right)\left(\mathbf{x}_k - \bar{\mathbf{x}}_j\right)^T \quad . \qquad (10)$$

The maximum separation problem can be stated as to find the projection matrix **Q** such that

$$\max_{\mathbf{Q}}[\det(\mathbf{B})\det(\mathbf{W}^{-1})] \quad , \qquad (11)$$

with solution given by the eigenvectors of $\mathbf{C} = \mathbf{W}^{-1}\mathbf{B}$. $\mathbf{C}$ is a n x n matrix with maximum rank equal to min(g-1,n), since **B** is obtained from a summation of only g terms. Therefore, the new basis consists of the min(g-1,n) eigenvectors with nonzero eigenvalues.

Since in most practical applications $g < n$, this procedure can significantly reduce the dimensionality. An important property of the new basis allows a further reduction. The contribution of each axis to the measure of the spread of the populations is proportional to the corresponding eigenvalue. If the eigenvectors are ranked in eigenvalues decreasing order, one can take only the first $q <$ min(g-1,n) eigenvectors, with the greatest eigenvalues, still preserving much of the group separation. The parameter q is chosen such that some given percentage r of the separation between groups are preserved. Then, the value of q is the minimum integer number for which the condition

$$\frac{\sum_{j=1}^{q} \lambda_j}{\sum_{j=1}^{n} \lambda_j} > r \ , \qquad (12)$$

holds.

Usually, in face recognition applications, the amount of face images in the training set (K) is less than the number of pixels in the images (n). In such cases the sampled within individuals covariance matrix **W** - Eq. (11) - is not invertible. To circumvent this problem, Swets and Weng [8] propose to use the $p \leq K$ projections over the eigenfaces produced by a previous PCA algorithm as input to the LDA stage. Therefore, this work does not use the pixel intensities as input to the LDA classifier, but the principal components generated by a previous PCA stage, according to Swets and Wengs proposal.

The RBF network classifier

The RBF classifier 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,

$$f_j(\mathbf{x}) = \exp \frac{\left\| \mathbf{x} - \boldsymbol{\mu}_j \right\|^2}{2\sigma_j^2} \ , \qquad (13)$$

where σ_j is the width parameter, μ_j is the vector determining the center of basis function f and $\mathbf{x}$ is the n-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_j. 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 implements the Gaussian Mixture Model algorithm to train the network. The model regards the basis functions as the components of a mixture density model, whose parameters σ_j and μ_j are to be optimized by maximum likelihood [7]. The number of hidden neurons or, equivalently, the number of basis functions is treated as an input to the model and is typically much less than the total number of input data points $\{\mathbf{x}_i\}$.

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 to an 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.

5. EXPERIMENT DESIGN

The Face Database

The experiments to evaluate both recognition systems make use of the ORL Face Database. 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 an 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. Figure 3 shows an example of the set of images for one subject.

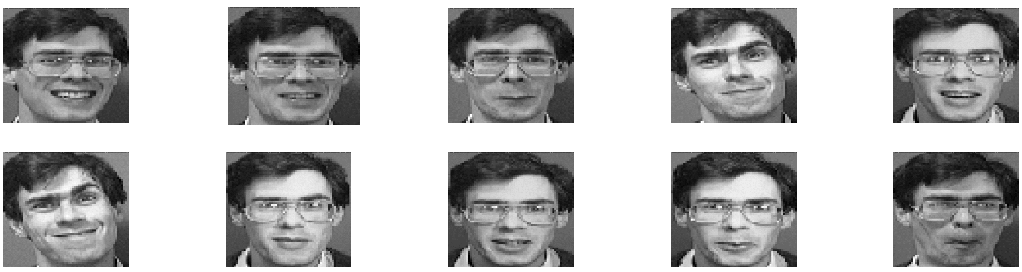

Figure 3 – A set of ten images for one subject from the ORL database.

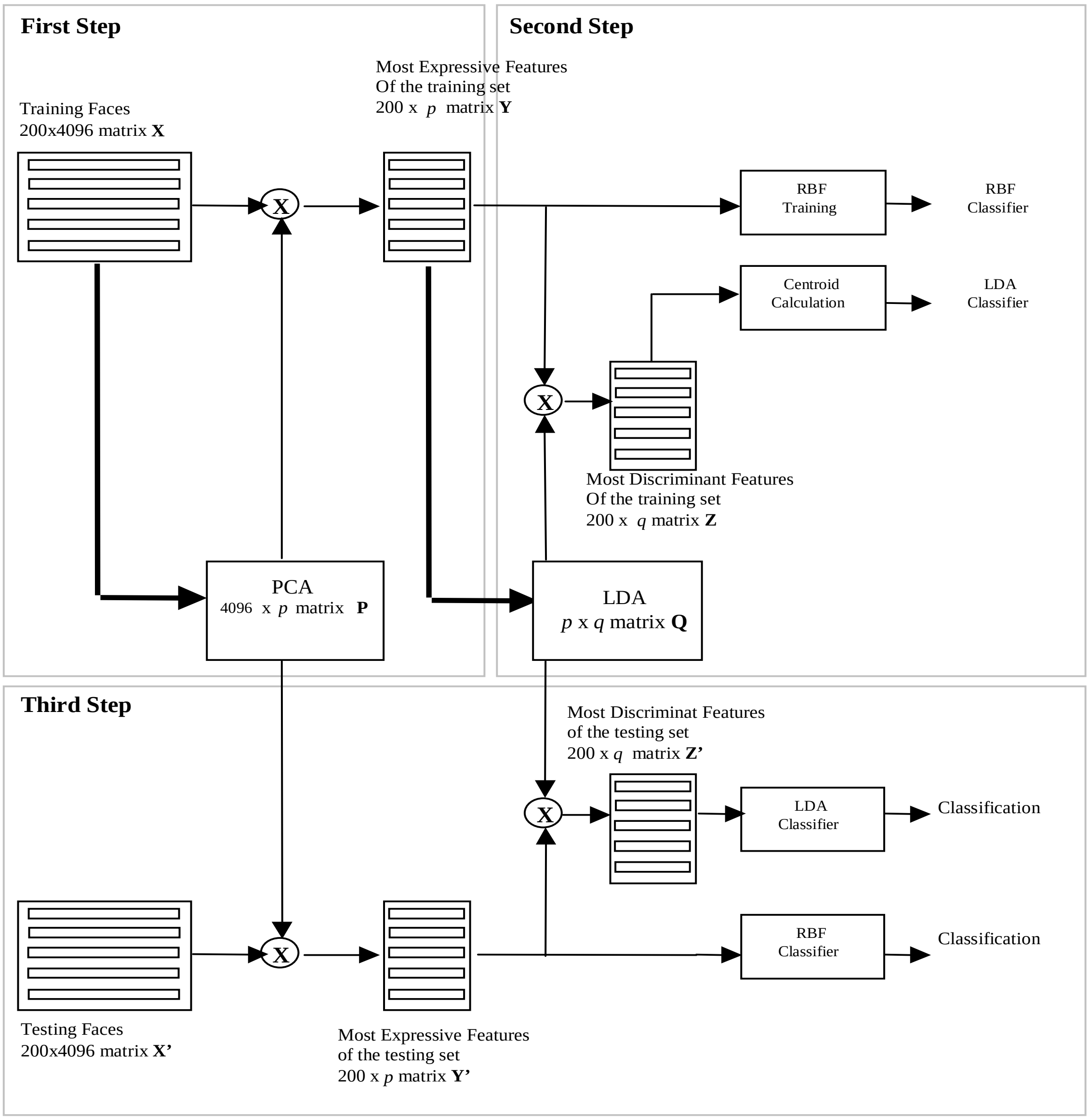

Figure 4: The experiments carried out to evaluate the recognition systems.

The Experiments

Figure 4 illustrates the experiments carried out in this work. Before being used in the experiments all the images are represented as a row vector, which is obtained by simply concatenating the columns of the image matrix together.

Each experiment consists of three steps: generation of the PCA eigenvectors, training the classifiers and testing the classifiers.

First Step – PCA eigenvectors generation: In the first step the PCA eigenvectors are generated. A training set is selected by choosing randomly 5 images for each individual, a total of 200 images, each one containing 64x64=4096 pixels. The remaining 5 images per individual are used later to test both methods (step 3). The average image of all training faces is then computed and subtracted from each face producing the matrix **X**. This zero mean training matrix is used as input to compute the PCA, and the *p* eigenvectors (eigenfaces) with the greatest eigenvalues are selected, forming the PCA transformation matrix **P**. To determine the maximum number of eigenfaces to retain, the total variance explained by each principal component was taken into account. Both systems reached the top performance around 50 eigenfaces.

Second Step – Training the classifiers: In the second step the classifiers are trained. To train the RBF classifier, the matrix **X** containing the 200 training images is projected onto the face space. The matrix **Y**, containing the training faces represented in the Most Expressive Features, is so computed. Those matrices are used to estimate the probability density of the input data. The number of basis functions is an input to the model. The basis function centers are determined by fitting the mixture model with circular covariances using the EM (expectation-maximization) algorithm [7] and their respective widths are set to the maximum inter-center square distance. The hidden to output weights that gives rise to the least squares solution are determined using the pseudo-inverse [7]. The RBF network is trained to produce the value 1 in the output unit corresponding to the face presented at the input layer and value 0 in every other unit.

The second design method, the LDA classifier, uses also the matrix **Y** as its training set. The within-group covariance matrix and the between-group covariance matrix of the face features represented in the MEF space are computed. The discriminant axes are then calculated according to Eq. (11) obtaining the transformation matrix **Q**. The projections of the training faces in terms of the Most Discriminat Features (the matrix **Z** in figure 4) are then computed. This step ends with the computation of the centroid for each individual in the new discriminant space.

Third Step – Testing the classifiers: The performance of the classifiers is evaluated in the third step. Each test image (matrix **X'**) is projected onto the MEFs using the matrix **P** obtained in the first step. The resulting matrix **Y'** is subjected to the RBF classifier.

In the system using the LDA classifier a second projection takes place onto the MDF space, whose projection matrix **Q** has been computed in the second step.

The true/false recognition of both classifiers 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 *p* of eigenfaces used during the analysis assumed ten different values: 10,20,30,40,50,60,70,80,90 and 100. In the evaluation of the RBF classifier 70,80,90,100, and 110 neurons in the hidden layer were considered.

6. EXPERIMENT RESULTS

The results of the experiments are summarized in figures 5 to 8.

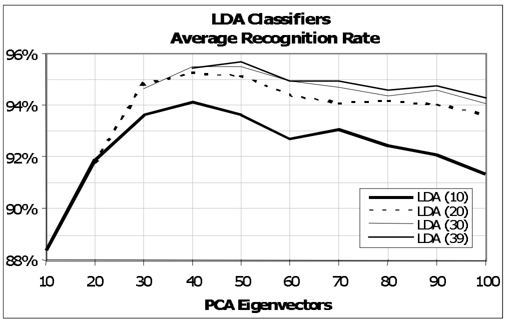

Figure 5: Average recognition rate for LDA method as a function of the number of eigenfaces for 10- 20- 30- and 39 dimensional discriminant space.

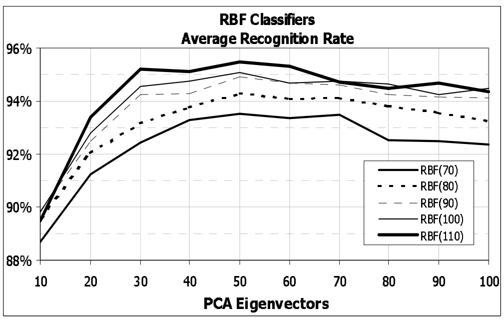

Figure 6: Average recognition rate for the RBF method as a function of the number of eigenfaces for 70, 80, 90, 100 and 110 neurons in the hidden layer.

Figure 5 shows the average recognition rate obtained in 25 runs for the LDA method as a function of the number of eigenfaces. The four curves represent the results of 10, 20, 30 and 39 dimensional discriminant spaces generated by the LDA algorithm. It can be observed that the performance grows by increasing the number of MDFs (the dimension of the discriminant space - *q*), and reaches, in these experiments, the maximum value for 39 MDFs. The maximum average recognition rate for the LDA - 95.7% - was reached for 39 MDFs computed on 50 MEFs.

Figure 6 shows the recognition rate of the RBF classifier. The five distinct curves correspond to different numbers of neurons in the RBF hidden layer. With more than 110 hidden neurons, the classifier brings no improvement of the recognition rates and the training process has become unstable. It can be seen in figure 6 that the best average recognition rate of the RBF approach - 95.5% - was obtained for 50 eigenfaces working with 110 neurons in the RBF hidden layer.

The average recognition rates for the best performance configurations for each methods – LDA classifier with 39 MDFs and the RBF classifier with 110 neurons in the hidden layer - are shown in figure 7. For more than 40 eigenfaces both classifiers had almost the same recognition rates.

Figure 8 shows how many times, during the 25 runs, each method has won, that is to say, has had a better performance than the other. Again, the results in the figure 8 indicate a similar performance of both approaches.

As a final evaluation, the results show no clear superiority of any method as far as the recognition rate is concerned. The ability of the RBF network to use more than one gaussian activation function to describe the population brings no important performance improvement when compared to the LDA approach. The more complex model underlying the RBF network implies in a higher computation cost, and its training is sometimes unstable.

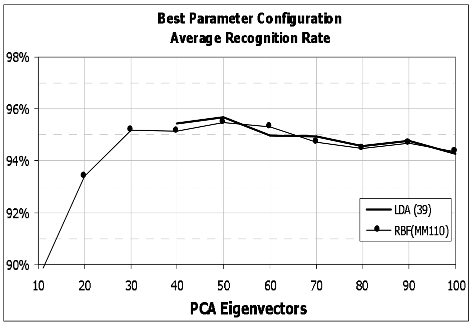

Figure 7: Average recognition rate for best parameter configuration of both methods as a function of the number of eigenfaces.

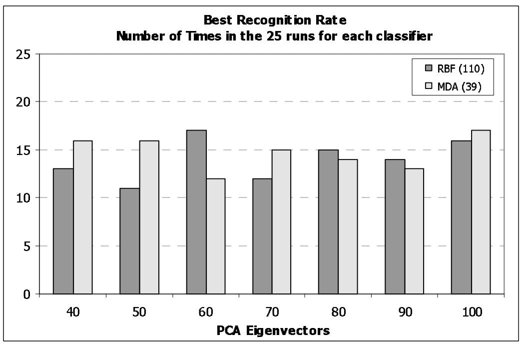

Figure 8: Number of times in the 25 runs each classifier had the best recognition rate as a function of the number of eigenfaces (in some cases both classifiers had the same performance).

7. CONCLUSIONS

Two eigenface based face recognition systems were evaluated: one based on the Linear Discriminant Analysis and the other based on a Radial Basis Function Neural Network.

The ORL Face database containing 10 photos of 40 individuals was used for the training and testing purposes. The parameters of both systems (number of eigenfaces, number of discriminant features and number of neuron in the hidden layer of the RBF network) were varied in a wide range of values. For each set of parameter values the experiments were executed 25 times for a different choice of the training and testing sets.

The performance measures on the experiments have indicated no recognition rate superiority of any method. On the other hand, the RBF approach has involved a higher computational cost than LDA method, and also sometimes has presented problems of convergence for high dimensional space. Therefore, it is fair to say that the experiments carried out in this work favor the LDA approach because of its simplicity and reliability.

8. 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 al., "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, A. Veiga, "Desing of Radial Basis Function Network as Classifier in Face Recognition Using Eigenfaces," 5th Brazilien Symposium on Neural Networks, Belo Horizonte, 1998.

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

[8] D. L. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieva," IEEE trans. on PAMI, Vol. 18, pp. 831-836, Aug. 1996.

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

[13] K. Etemad and R. Chellappa, "Discriminant analysis for Recognition of Human Face Images," Journal of Optical Society of America A, pp. 1724-1733, Aug., 1997

[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 and A. Pentland, "Eigenfaces for Recognition, "Journal of Cognitive Neuroscience, Vol. 3, pp. 72-85, 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] 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.

[22] W. Zhao, R. Chellappa and A. Krishnaswamy, "Discriminant Analysis of Principal Components for Face Recognition," Proc. 2nd International Conference on Automatic Face and Gesture Recognition, pp. 336-341, Japan, April, 1998.