Statistical Approaches To Face Recognition

A Qualifying Examination Report

By

Ara V. Nefian

Presented to the Qualifying Examination Committee

In Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy in Electrical Engineering

Dr. Albin J. Gasiewski

Dr. Jeff Geronimo

Dr. Monson H. Hayes III

Dr. Russell M. Mersereau

Dr. Ronald W. Schafer

Georgia Institute of Technology

School of Electrical Engineering

December, 1996

Contents

List of Figures

List of Tables

1 Introduction

Face recognition from still images and video images is emerging as an active research area with numerous commercial and law enforcement applications. These applications require robust algorithms for human face recognition under different lighting conditions, facial expressions and orientations. An excellent survey on face recognition is given in [1]. A general scheme used for face recognition is illustrated in Figure 1.

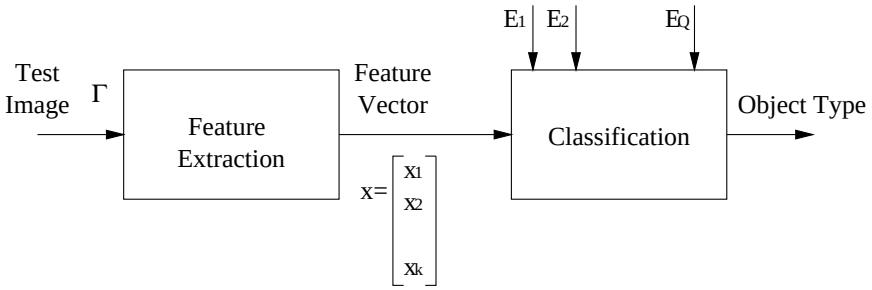

Figure 1: General face recognition scheme

The feature vector $x = [x_1, x_2, \ldots x_K]$ extracted from a test face image Γ is compared in turn to the feature vectors extracted from all example face images $E_1, E_2, \ldots, E_Q$ and a measure of similarity in the feature space is used to classify the input image as one of the example images. The ratio of correct classified face images over the total number of faces classified by the recognition system defines the recognition performance (recognition rate) of the system. Based on the feature extraction and classification techniques used, the face recognition approaches are:

- geometrical parameterization approaches
- statistical approaches
- neural networks approaches

In this paper we focus on the statistical approaches to face recognition and more specifically on:

- correlation methods
- Singular Value Decomposition methods

- Karhunen-Loeve expansion - based methods
- Fisher Linear Discriminant - based Methods
- Hidden Markov Model - based Methods

A qualitative comparison among these methods will also be discussed.

2 Correlation Methods

The most direct of the procedures used for face recognition is the matching between the test images and a set of training images based on measuring the *correlation* [2]. The matching technique, in this case is based on the computation of the normalized cross-correlation coefficient C_N [3], defined by:

$$C_N = \frac{E\{I_T T\} - E\{I_T\}E\{T\}}{\sigma(I_T)\sigma(T)}, \tag{1}$$

where I_T is the image which must be matched to the template T, $I_T T$ represents the pixel-by-pixel product, E is the average operator and σ is the standard deviation over the area being matched. This normalization rescales the template and image energy distribution so that their average and variances match. However, correlation based methods are very dependent on illumination, rotation and scale. The best results for the reduction of the illumination variations were obtained using the intensity of the gradient ($|\delta_x I_T| + |\delta_y I_T|$). Figure 2, 3 and 4 illustrate the dependency of the correlation coefficient versus image scaling, rotation around the central vertical axis lying in the image plane and rotation in the image plane for the intensity image (I) and for the same image after the illumination normalization ($D(I)$). Because the correlation method is computationally very expensive the dependency of the recognition on the resolution of the image has been investigated. In [4] it has been shown that correlation based recognition is possible at a good performance level using templates as small as 36 x 36 pixels.

2.1 Recognition Using Correlation Methods

In [3], Brunelli and Poggio describe a correlation based method for face recognition from frontal views. Their method is based on the matching of templates corresponding to facial features of

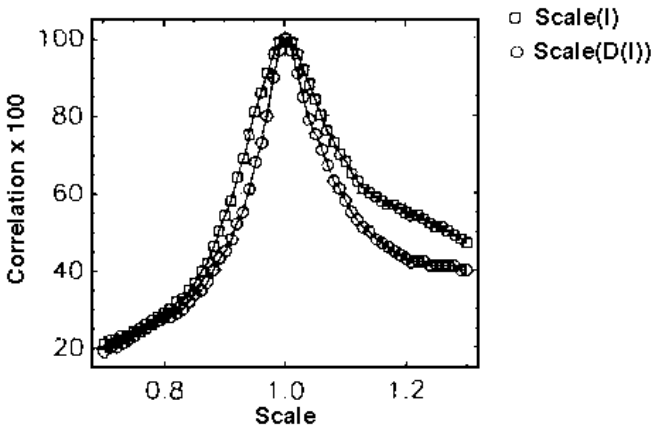

Figure 2: Correlation versus image scaling

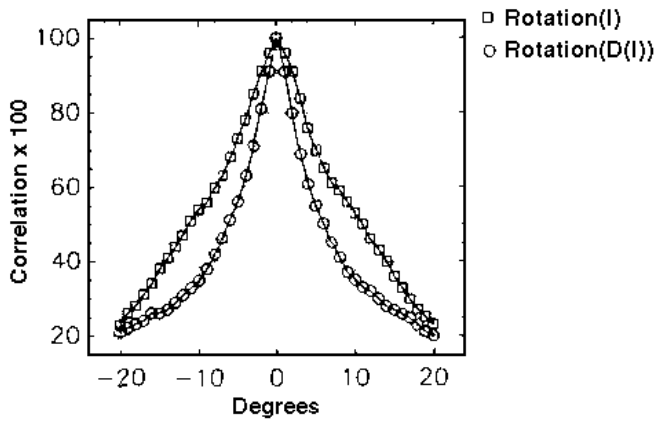

Figure 3: Correlation versus image rotation with axis perpendicular to the image plane

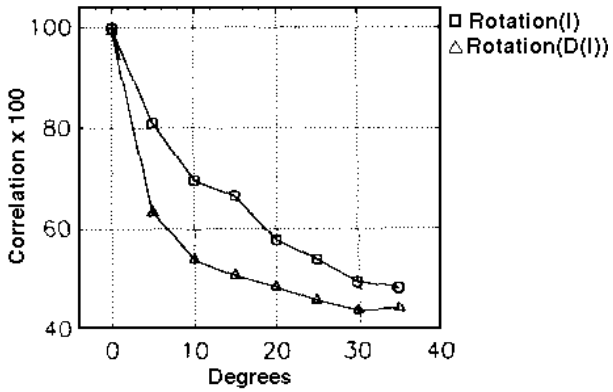

Figure 4: Correlation versus rotation in the image plane

relevant significance as the eyes, nose and mouth. In order to reduce the complexity of this approach, first the positions of these features are detected. Facial features detection has been intensively studied in the few last years [5], [6], [7], [8]. The method proposed by Brunelli and Poggio uses a set of templates to detect the eye position in a new image, by looking for the maximum absolute values of the normalized correlation coefficient of these templates at each point in the test image. To cope with scale variations, a set of five eye templates at different scales was used. However, this method is computationally expensive. Additionally, eyes of different people can be markedly different. These difficulties can be significantly reduced by using hierarchical correlation (as proposed by Burt in [9]). Once the eyes are located, the detection of other features can take advantage of their estimated position and the use of integral projections.

Feature Extraction: After the facial features are detected, a set of templates corresponding to these features in the test image is compared, in turn, with the corresponding features of all of the images in the database, returning a vector of matching scores (one per feature) computed through normalized cross correlation.

Classification: The similarity scores of different features can be integrated to obtain a global score. The cumulative score can be computed in several ways:

- choose the score of the most similar feature.

- sum the feature scores.

- sum the feature scores, using constant weights.

- sum the features scores using person-dependent weights.

After the cumulative matching scores are computed, a test face is assigned to the face class for which this score is maximized. The discriminating properties of some facial features versus interocular distance are illustrated in Figure 5. It can be seen that for large images (interocular distance bigger than 30 pixels) the whole face is the least discriminating template.

Recognition Performance: The recognition rate reported in [3] using correlation method for frontal faces is higher than 96%. The correlation method as described in this section requires a robust feature detection algorithm with respect to variations in scale, illumination and rotations in

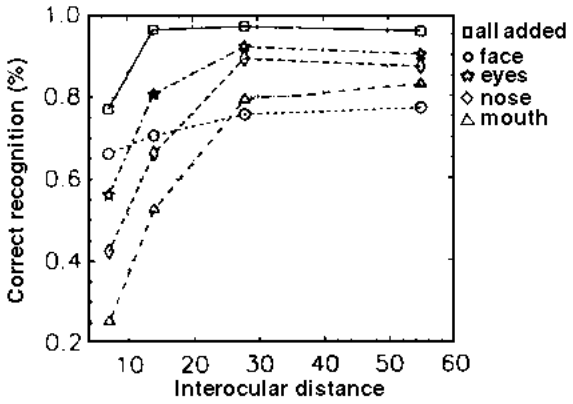

Figure 5: Average performance for recognition based on feature matching

image plane and in image depth. In addition the computational complexity of this method is very high.

2.2 Recognition Under General Viewing Conditions

In [10], Beymer extended the correlation based approach, to a view-based approach for recognizing faces under varying orientations, including rotations with axis perpendicular to the image plane (rotations in image depth). In the first stage, of this method, a pose estimation module detects the position of three facial features (the eyes and the nose lobe) to determine the face pose (orientation). These features were detected by looking for the maximum absolute values of the normalized correlation coefficients of the model templates (representing the facial features) at each point in the current image. To handle rotations out of image plane, templates from different views and different people were used. After the pose has been determined, the task of recognition is reduced to the classical correlation method discussed in the previous section in which the facial feature templates are matched to the corresponding templates of the appropriate view-based models using the cross correlation coefficient. In this case, the computation complexity increases with the number of model views for each person in the database. Furthermore, the feature detection under non frontal pose is a more difficult task.

Recognition Performances: The recognition performances of the system were evaluated on a small database of 620 test images representing 62 people. A recognition rate of 98.7% has been

reported [10]. However, the method is computationally complex requiring 10-15 min to perform the classification of one face image on a Sparc2 workstation.

3 Karhunen-Loeve Expansion - Based Methods

3.1 Recognition Using Eigenfaces

The "Eigenfaces" method proposed by Turk and Pentland [11], [12] is based on the Karhunen-Loeve expansion and is motivated by the earlier work of Sirovitch and Kirby [13], [14] for efficiently representing picture of faces. The eigenfaces method presented by Turk and Pentland finds the principal components (Karhunen-Loeve expansion) of the face image distribution or the eigenvectors of the covariance matrix of the set of face images. These eigenvectors can be thought as a set of features which together characterize the variation between face images.

Let a face image $I(x, y)$ be a two dimensional array of intensity values, or a vector of dimension n. Let the training set of images be $I_1, I_2, \ldots, I_N$. The average face image of the set is defined by $\Psi = \frac{1}{N}\sum_{i=1}^{N} I_i$. Each face differs from the average by the vector $\Phi_i = I_i - \Psi$. This set of very large vectors is subject to principal component analysis which seeks a set of K orthonormal vectors $v_k,\ k = 1, \ldots, K$ and their associated eigenvalues λ_k which best describe the distribution of data. The vectors v_k and scalars λ_k are the eigenvectors and eigenvalues of the covariance matrix:

$$C = \frac{1}{N}\sum_{i=1}^{N} \Phi_i \Phi_i^T = AA^T, \tag{2}$$

where the matrix $A = [\Phi_1, \Phi_2, \ldots, \Phi_N]$. Finding the eigenvectors of matrix $C_{n\times n}$ is computationally intensive. However, the eigenvectors of C can be determined by first finding the eigenvectors of a much smaller matrix of size $N \times N$ and taking a linear combination of the resulting vectors [12].

Feature extraction: The space spanned by the eigenvectors $v_k,\ k = 1, \ldots, K$ corresponding to the largest K eigenvalues of the covariance matrix C, is called the *face space*. The eigenvectors of matrix C, which are called *eigenfaces* form a basis set for the face images. A new face image Γ is transformed into its eigenface components (projected onto the face space) by:

$$\omega_k = < v_k, (\Gamma - \Phi) > = v_k^T (\Gamma - \Phi) \tag{3}$$

for $k = 1, \ldots, K$. The projections ω_k form the feature vector $\Omega = [\omega_1, \omega_2, \ldots, \omega_K]$ which describes the contribution of each eigenface in representing the input image.

Classification: Given a set of face classes E_q and the corresponding feature vectors Ω_q, the simplest method for determining which face class provides the best description of an input face image Γ is to find the face class j that minimizes the Euclidian distance in the feature space:

$$\xi_q = ||\Omega - \Omega_q||, \tag{4}$$

A face is classified as belonging to class E_q when the minimum ϵ_q is below some threshold θ_ϵ.

$$E_q = argmin_q\{\xi_q\} \tag{5}$$

Otherwise, the face is classified as unknown.

Recognition Performance: The Eigenfaces method was tested on a large database of 2500 face images digitized under controlled conditions. Various groups of sixteen images corresponding to sixteen different subjects were selected and used as a the training set. The recognition performances reported are 96% correct classification over lighting variations, 85% correct classification over orientation variation, and 64% correct classification over size variation. It can be seen that the approach is fairly robust to changes in lighting conditions [15], but degrades quickly as the scale changes. One can explain this by the significant correlation present between images with changes in illumination conditions; the correlation between face images at different scales is low. The recognition currently takes about 350 msec running on a SparcStation 1, using face images of size 128 x 128 pixels.

3.2 Recognition Under General Viewing Conditions

3.2.1 The Parametric Approach

In [16], Murase and Nyar extended the capabilities of the eigenface method to general 3D object recognition under different illumination and viewing conditions. Given N object images taken under P views and L different illumination conditions, an universal image set is built which contains all the available data. In this way a single "parametric space" describes the object identity as well as the viewing or illumination conditions. The eigenface decomposition of this space was used for

feature extraction and classification. However, in order to insure discrimination between different objects the number of eigenvectors used in this method was increased compared to the classical Eigenface method.

3.2.2 The View-Based Approach

Based on the eigenface decomposition, Pentland *et al* [17] developed a "view-based" eigenspace approach for human face recognition under general viewing conditions. Given N individuals under P different views, recognition is performed over P separate eigenspaces, each capturing the variation of the individuals in a common view. The "view-based" approach is essentially an extension of the eigenface technique to multiple sets of eigenvectors, one for each face orientation. To deal with multiple views, in the first stage of this approach, the orientation of the test face is determined and the eigenspace which best describes the input image is selected. This is accomplished by calculating the residual description error (distance from feature space: DFFS) for each view space. Once the proper view is determined, the image is projected onto the appropriate view space and then recognized. The view-based approach is computationally more intensive than the parametric approach because P different sets of V projections are required (V is the number of eigenfaces selected to represent each eigenspace). However, this does not imply that a factor of P times more computation is necessarily required. By progressively calculating the eigenvector coefficients while pruning alternative viewspaces, the cost of using P eigenspaces can be greatly reduced. Naturally, the view-based representation can yield more accurate representation of the underlying geometry.

Recognition Performance: The recognition performance of the view-based and parametric approaches was evaluated on a database of 189 images consisting of nine views of 21 people [17]. The nine views of each person were evenly spaced from -90^0 to 90^0. The performance of the algorithms was tested by training on a subset of the available views $\pm 90^0, \pm 45^0, 0^0$ and testing on the intermediate views $\pm 68^0, \pm 23^0$ (*interpolation performances*). The average recognition rates reported were 90% for the view-based and 88% for the parametric eigenspace methods. A second series of experiments tested the extrapolation performance by training on a range of views (e.g. -90^0 to $+45^0$) and testing on novel views outside the training range (e.g. $+68^0$ and $+90^0$). For testing views separated by $\pm 23^0$ from the training range the average recognition rates were 83% for the

view based and 78% for the parametric eigenspace method. For $\pm 45^0$ testing views, the average recognition rates were 50% (view-based) and 43% parametric.

3.3 Recognition Using Eigenfeatures

In [17], Pentland et al discussed the use of facial features for face recognition. This can be viewed as either a modular or layered representation of the face, where a coarse (low resolution) description of the whole head is augmented by additional (high resolution) details in terms of salient facial features. The eigenface technique was extended to detect facial features. For each of the facial features, a feature space is built by selecting the most significant eigenfeatures (eigenvectors corresponding to the largest eigenvalues of the features correlation matrix). In the eigenfeature representation the equivalent "distance from feature space"(DFFS) can be effectively used for the detection of facial features. The DFFS feature detection was extended to the detection of features under different viewing geometries by using either a view-based eigenspace or a parametric eigenspace.

Feature extraction: After the facial features in a test image were extracted, a score of similarity between the detected features and the features corresponding to the model images is computed. The technique used to determine this score is an extension of the eigenface method presented in section 3.1.

Classification: A simple approach for recognition is to compute a cumulative score in terms of equal contribution by each of the facial feature scores. More elaborate weighting schemes, for classification were discussed in section 2.1. Once the cumulative score is determined, a new face is classified such that this score is maximized.

Recognition Performance: The utility of a layered representation using eigenfaces and eigenfeatures was tested on a set 45 individuals with two views per person corresponding to different facial expressions (neutral vs. smiling). The neutral set of images was used as a training set and the recognition was performed on the smiling set. Since the difference between these particular facial expressions is primarily articulated in the mouth, this feature was discarded for recognition purposes. The recognition results showed that the eigenfeatures alone were sufficient in achieving a recognition rate of 95%, equal to that of the eigenfaces. When a combined representation of eigenfaces and eigenfeatures was tested a recognition rate of 98% was reported.

3.4 The Karhunen-Loeve Transform of the Fourier Spectrum

In [18], Akamatsu et al, illustrated the effectiveness of Karhunen-Loeve Transform of Fourier Spectrum in the Affine Transformed Target Image (KL-FSAT) for face recognition. First, the original images were standardized with respect to position, size, orientation using an affine transform so that three reference points satisfy a specific spatial arrangement. The position of these points is related to the position of some significant facial features. The standardization of the images used in the training set allowed to significantly increase the number of individuals to be identified (from 16 classes reported for the eigenface method to 269). The Eigenface method is applied as discussed in section 3.1. to the magnitude of the Fourier Spectrum of the standardized images (KL-FSAT). Due to the shift invariance property of the magnitude of the Fourier spectrum, the KL-FSAT performed better than classical eigenfaces method under variations in head orientations and shifting.

Recognition Performance: The authors reported recognition rates as high as 91% for a training set of 269 individuals. The testing set contained 100 images (5 samples of 20 individuals) with variations in head orientations. On the same database the eigenface recognition rate was 85%. However, the computational complexity of KL-FSAT method is significantly greater than the eigenface method due to the computation of the Fourier spectrum.

4 Linear Discriminant Methods - Fisherfaces

In [19],[20], the authors proposed a new method for reducing the dimensionality of the feature space by using Fisher's Linear Discriminant(FLD) [21]. The FLD uses the class membership information and develops a set of feature vectors in which variations of different faces are emphasized while different instances of faces due to illumination conditions, facial expressions and orientations are de-emphasized.

4.1 Fisher's Linear Discriminant

Given c classes with a priori probabilities P_i, let N_i be the number of samples of class i, $i = 1, \ldots, c$. Then the following positive semidefinite scatter matrices are defined as:

$$S_B = \sum_{i=1}^{c} P_i(\mu_i - \mu)(\mu_i - \mu)^T \tag{6}$$

$$S_W = \sum_{i=1}^{c} P_i \sum_{j=1}^{N_i}(x_j^{(i)} - \mu_i)(x_j^{(i)} - \mu_i)^T \tag{7}$$

where $x_j^{(i)}$ denotes the jth n dimensional sample vector belonging to class i, μ_i is the mean of class i:

$$\mu_i = \frac{1}{N_i}\sum_{j=1}^{N_i} x_j^{(i)}, \tag{8}$$

and μ is the overall mean of sample vectors:

$$\mu = \frac{1}{\sum_{i=1}^{c} N_i}\sum_{i=1}^{c}\sum_{j=1}^{N_i} x_j^{(i)}, \tag{9}$$

S_W is the *within-class scatter matrix* and represents the average scatter of sample vector of class i; S_B is the *between-class scatter matrix* and represents the scatter of the mean μ_i of class i around the overall mean vector μ. If S_W is non singular, the *Linear Discriminant Analysis* (LDA) selects a matrix $V_{opt} \in R^{n\times k}$ with orthonormal columns which maximizes the ratio of the determinant of the between class scatter matrix of the projected vector samples to the determinant of the within-class scatter matrix of the projected samples, i.e.

$$V_{opt} = argmax_V \frac{|V^T S_B V|}{|V^T S_W V|} = [v_1, v_2, \ldots, v_k], \tag{10}$$

where $\{v_i | i = 1, 2, \ldots, k\}$ is the set of generalized eigenvectors of S_B and S_W corresponding to the set of decreasing eigenvalues $\{\lambda_i | i = 1, 2, \ldots, k\}$, i.e.

$$S_B v_i = \lambda_i S_W v_i. \tag{11}$$

From [22], the upper bound of k is $c - 1$. The matrix V_{opt} describes the Optimal Linear Discriminant Transform or the Foley-Sammon Transform. While the Karhunen-Loeve Transform

performs a rotation on a set of axes along which the projection of sample vectors differ most in the autocorrelation sense, the Linear Discriminant Transform performs a rotation on a set of axes $[v_1, v_2, \ldots, v_k]$ along which the projection of sample vectors show maximum discrimination.

4.2 Face Recognition Using Linear Discriminant Analysis

Let a training set of N face images represent c different subjects. The face images in the training set are two-dimensional arrays of intensity values, represented as vectors of dimension n. Different instances of a person's face (variations in lighting, pose or facial expressions) are defined to be in the same class and faces of different subjects are defined to be from different classes.

Feature Extraction: The scatter matrices S_B and S_W are defined in Equations 6, 7. However, the matrix V_{opt} cannot be found directly from Equation 10 because in general the matrix S_W is singular. This stems from the fact that the rank of S_W is less than $N - c$, and in general, the number of pixels in each image (n) is much larger than the number of images in the learning set (N). There have been presented many solutions in the literature in order to overcome this problem [23], [24]. In, [19], the authors proposed a method which was called the Fisherfaces method. The problem of S_W being singular is avoided by projecting the image set onto a lower dimensional space so that the resulting within class scatter is non singular. This is achieved by using Principal Component Analysis (PCA) to reduce the dimension of the feature space to $N - c$ and then, applying the standard linear discriminant defined in Equation 10 to reduce the dimension to $c - 1$. More formally V_{opt} is given by:

$$V_{opt} = V_{fld} V_{pca}, \tag{12}$$

where

$$V_{pca} = argmax_V |V^T C V|, \tag{13}$$

and,

$$V_{fld} = argmax_V \frac{|V^T V_{pca}^T S_B V_{pca} V|}{|V^T V_{pca}^T S_W V_{pca} V|}, \tag{14}$$

where C is the covariance matrix of the set of training images and is computed from Equation 2. The columns of V_{opt} are orthogonal vectors which are called Fisherfaces. Unlike the Eigenfaces,

the Fisherfaces do not correspond to face like patterns. All example face images $E_q,\ q = 1, \ldots, Q$ in the example set S are projected on the vectors corresponding to the columns of the V_{fld} and a set of features is extracted for each example face image. These feature vectors are used directly for classification.

Classification: Having extracted a compact and efficient feature set, the recognition task can be performed by using the Euclidian distance in the feature space. However, in [20] as a measure in the feature space, is proposed a weighted mean absolute/square distance with weights obtained based on the reliability of the decision axis.

$$D(\Gamma, E) = \sum_{v=1}^{K} \frac{(\Gamma_v - E_v)^2}{\sum_{E \in S} (\Gamma_v - E_v)^2} \alpha_v, \tag{15}$$

where Γ_v and E_v are the projections of the test image Γ and example image E_v on vector v. S is the set of example images. The weights α_v are related to the discrimination power along axis v and are equal to the corresponding normalized eigenvalue:

$$\alpha_i = \frac{\lambda_i}{\sum_{i=1}^{K} \lambda_i}, \tag{16}$$

Therefore, for a given face image Γ, the best match E^0 is given by

$$E^0 = argmin_{E \in S}\{D(\Gamma, E)\}, \tag{17}$$

and the confidence measure is defined as:

$$Conf(\Gamma, E^0) = 1 - \frac{D(\Gamma, E^0)}{D(\Gamma, E^1)}, \tag{18}$$

where E^1 is the second best candidate.

Recognition Performance: In [19], the authors reported a recognition rate of 99.4% under variations in lighting, facial expressions and eye wear (glasses, no-glasses). On the same database, the recognition rate reported when using the eigenface method was 80%. The database did not include images with variations in pose or orientation. The training set contained five sets of ten face images taken under strong variations in illumination facial expression, facial details, but no variations in pose.

In [18], Akamatsu et al applied LDA to the magnitude of the Fourier Spectrum of the intensity image. The database used in the experiments contained large variations in lighting conditions as

well as variations in head orientation. The results reported by the authors showed that LDA in the Fourier domain is significantly more robust to variations in lighting than the LDA applied directly to the intensity images. However, the computational complexity of this method is significantly bigger than classical Fisherface method due to the computation of the Fourier spectrum.

5 Singular Value Decomposition Methods

5.1 Singular Value Decomposition

The Singular Value Decomposition methods for face recognition use the general result stated by the following theorem:

Theorem: Let $I_{p\times q}$ be a real rectangular matrix and $Rank(I) = r$, then there exist two orthonormal matrices $U_{p\times p}$, $V_{q\times q}$ and a diagonal matrix $\Sigma_{p\times q}$ and the following formula holds:

$$I = U\Sigma V^T = \sum_{i=1}^{r} \lambda_i u_i v_i^T, \tag{19}$$

where

$$U = (u_1, u_2, \ldots, u_r, u_{r+1}, \ldots, u_p), \tag{20}$$

$$V = (v_1, v_2, .., v_r, v_{r+1}, .., v_q), \tag{21}$$

$$\Sigma = diag(\lambda_1, \lambda_2, \ldots, \lambda_r, 0, \ldots, 0), \tag{22}$$

$\lambda_1 > \lambda_2 > \ldots > \lambda_r > 0$, λ_i^2, $i = 1, 2, \ldots, r$ are the eigenvalues of II^T and I^TI, $= \lambda_i$ are the singular values of I, u_i, v_j, $i = 1, 2, \ldots, p$, $j = 1, 2, \ldots, q$ are the eigenvectors corresponding to eigenvalues of II^T and I^TI.

5.2 Face Recognition Using Singular Value Decomposition

Let a face image $I(x, y)$ be a two dimensional $(m \times n)$ array of intensity values and $[\lambda_1, \lambda_2, \ldots, \lambda_r]$ be its singular value (SV) vector. In [24], Zhong revealed the importance of using SVD for human face recognition by prooving several important properties of the SV vector as: the stability of the

SV vector to small perturbations caused by stochastic variation in the intensity image, the proportional variance of the SV vector to proportional variance of pixels in the intensity image, the invariance of the SV feature vector to rotation transform, translation and mirror transform. The above properties of the SV vector provide the theoretical basis for using singular values as image features. However, it has been shown that compressing the original SV vector into a low dimensional space, by means of various mathematic transforms leads to higher recognition performances. Among various transformations of compressing dimensionality, the Foley-Sammon transform based on Fisher criterion, i.e. optimal discriminant vectors is the most popular one. Given N face images which represent c different subjects, the SV vectors are extracted from each image. According to Equations 6 and 7, the scatter matrices S_B and S_W of the SV vectors are constructed. It has been shown that it is difficult to obtain the optimal discriminant vectors in the case of small number of samples, i.e. the number of samples is less than the dimensionality of the SV vector because the scatter matrix S_W is singular in this case. Many solutions have been proposed to overcome this problem. Hong [24] , circumvented the problem by adding a small singular value perturbation to S_W resulting in $\tilde{S}_W$ such that $\tilde{S}_W$ becomes nonsingular. However, the perturbation of S_W introduces an arbitrary parameter, and the range to which the authors restricted the perturbation is not appropriate to ensure that the inversion of $\tilde{S}_W$ is numerically stable. Cheng et al [23], solved the problem by rank decomposition of S_W. This is a generalization of Tian's method [25], who substituted S_W^{-1} by the positive pseudoinverse S_W^+.

Feature extraction: After the set of optimal discriminant vectors $\{v_1, v_2, \ldots, v_k\}$ has been extracted, the feature vectors are obtained by projecting the SV vectors onto the space spanned by $\{v_1, v_2, \ldots, v_k\}$.

Classification: When a test image is acquired, its SV vector is projected onto the space spanned by $\{v_1, v_2, \ldots, v_k\}$ and classification is performed in the feature space by measuring the Euclidian distance in the this space and assigning the test image to the class of images for which the minimum distance is achieved.

Recognition Performance: In [23] the authors reported 100% recognition rate over a database of 64 images (24 images in the training set and 40 in the testing set) of eight different subjects. In the testing set changes were made in face orientation, camera focus and eye wear(glasses, no-glasses).

Another method to reduce the feature space of the SV feature vectors was described by Cheng et al [26]. The training set used consisted of a small sample of face images of the same person. If $I_j^{(i)}$ represents the j^{th} image face image of person i, then the average image is given by $\frac{1}{N}\sum_{j=1}^{N} I_j^i$. Eigenvalues and eigenvectors are determined for this average image using SVD. The eigenvalues are thresholded to disregard the values close to zero. Average eigenvectors (called feature vectors) for all the average face images are calculated. A test image is then projected onto the space spanned by the eigenvectors. The Frobenius norm is used as a criterion to determine which person the test image belongs. The recognition performance was evaluated on a small database of 64 face images (24 images in the training set and 40 images in the testing set) of eight different persons and 100% recognition results were reported. The feature vectors were determined by averaging three images from each person.

6 Hidden Markov Model Based Methods

6.1 Hidden Markov Models

Hidden Markov Models (HMM) are a set of statistical models used to characterize the statistical properties of a signal. Rabiner [27],[28], provides an extensive and complete tutorial on HMMs. HMM are made of two interrelated processes: (1) an underlying, unobservable Markov chain with finite number of states, a state transition probability matrix and an initial state probability distribution. (2) a set of probability density functions associated to each state. The elements of a HMM are:

- N, the number of states in the model. If S is the set of states, then $S = \{S_1, S_2, ..., S_N\}$. The state of the model at time t is given by $q_t \in S$, $1 \leq t \leq T$, where T is the length of the observation sequence (number of frames).

- M, the number of different observation symbols. If V is the set of all possible observation symbols (also called the *codebook* of the model), then $V = \{v_1, v_2, ..., v_M\}$.

- A, the state transition probability matrix, i.e. $A = \{a_{ij}\}$ where

$$a_{ij} = P[q_t = S_j | q_{t-1} = S_i], \ 1 \leq i, j, \leq N, \tag{23}$$

$$0 \leq a_{i,j} \leq 1, \tag{24}$$

$$\sum_{j=1}^{N} a_{ij} = 1, \ 1 \leq i \leq N \tag{25}$$

- B, the observation symbol probability matrix, i.e. $B = b_j(k)$, where,

$$b_j(k) = P[O_t = v_k | q_t = S_j], \ 1 \leq j \leq N, \ 1 \leq k \leq M \tag{26}$$

  and O_t is the observation symbol at time t.

- π, the initial state distribution, i.e. $\pi = \pi_i$ where:

$$\pi_i = P[q_1 = S_i], \ 1 \leq i \leq N \tag{27}$$

  Using a shorthand notation, a HMM is defined as:

$$\lambda = (A, B, \pi). \tag{28}$$

The above characterization corresponds to a *discrete* HMM, where the observations are characterized as discrete symbols chosen from a finite alphabet $V = \{v_1, v_2, ..., v_M\}$. In a *continuous density* HMM, the states are characterized by continuous observation density functions. The most general representation of the model probability density function (pdf) is a finite mixture of the form:

$$b_i(O) = \sum_{k=1}^{M} c_{ik} N(O, \mu_{ik}, U_{ik}), \ 1 \leq i \leq N \tag{29}$$

where c_{ik} is the mixture coefficient for the kth mixture in state i. Without loss of generality $N(O, \mu_{ik}, U_{ik})$ is assumed to be a Gaussian pdf with mean vector μ_{ik} and covariance matrix U_{ik}.

6.2 Face Recognition Using HMM

HMM have been used extensively for speech recognition, where data is naturally one dimensional (1D) along the time axis. However, the equivalent fully connected two dimensional HMM would lead to a very high computational problem [29]. Attempts have been made to use multi-model representations that lead to pseudo 2D HMM [30]. These models are currently used in character recognition [31] [32].

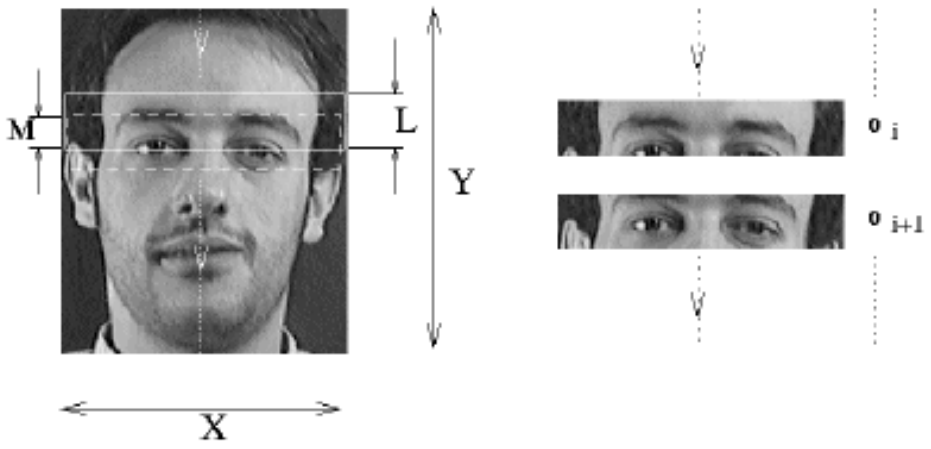

Figure 6: Image sampling technique for HMM recognition

In [33], Samaria et al proposed the use of the 1D continuous HMM for face recognition. Assuming that each face is in an upright, frontal position, features will occur in a predictable order, i.e. forehead, eyes, nose etc. This ordering suggests the use of a top-bottom model, where only transitions between adjacent states in a top to bottom manner are allowed [34]. The states of the model correspond to the facial features as forehead, eyes, nose, mouth and chin [35]. The observation sequence O is generated from an $X \times Y$ image using an $X \times L$ sampling window with $X \times M$ pixels overlap (Figure 6). Each observation vector is a block of L lines. There is an M line overlap between successive observations. The overlapping allows the features to be captured in a manner which is independent of vertical position, while a disjoint partitioning of the image could result in the truncation of features occurring across blocks boundaries. In [36], the effect of different sampling parameters has been discussed. With no overlap, if a small height of the sampling window is used, the segmented data do not correspond to significant facial features. However, as the window height increases there is a higher probability of cutting across the features.

Training: Given c face images for each subject of the training set, the goal of the training stage is to optimize the parameters $\lambda_i = (A, B, \pi)$ to "best" describe , the observations $O = \{o_1, o_2, \ldots, o_T\}$, in the sense of maximizing $P(O|\lambda)$. The general HMM training scheme is illustrated in Figure 7 and is a variant of the K-means iterative procedure for clustering data:

1. The training images are collected for each subject in the database and are sampled to generate

the observation sequence.

2. A common prototype model is constructed with the purpose of specifying the number of states in the HMM and the state transitions allowed, A (model initialization).

3. A set of initial parameter values using the training data and the prototype model are computed iteratively. The goal of this stage is to find a good estimate for the observation model probability B. In [28], it has been shown that a good initial estimates of the parameters are essential for rapid and proper convergence (to the global maximum of the likelihood function) of the reestimation formulas. On the first cycle, the data is uniformly segmented, matched with each model state and the initial model parameters are extracted. On successive cycles, the set of training observation sequences was segmented into states via the Viterbi algorithm. The result of segmenting each of the training sequences is for each of N states, a maximum likelihood estimate of the set of observations that occur within each state according to the current model.

4. Following the Viterbi segmentation, the model parameters are reestimated using the Baum-Welch reestimation procedure. This procedure adjusts the model parameters so as to maximize the probability of observing the training data, given each corresponding model.

5. The resulting model is then compared to the previous model (by computing a distance score that reflects the statistical similarity of the HMMs). If the model distance score exceeds a threshold, then the old model λ is replaced by the new model $\tilde{\lambda}$, and the overall training loop is repeated. If the model distance score falls below the threshold, then model convergence is assumed and the final parameters are saved.

Recognition: Recognition is carried out by matching the test image against each of the trained models (Figure 8). To do this, the image is converted to an observation sequence and then model likelihoods $P(O_{test}|\lambda_i)$ are computed for each λ_i $i = 1, \ldots, c$. The model with highest likelihood reveals the identity of the unknown face.

$$\nu = argmax_{1 \leq i \leq c}[P(O_{test}|\lambda_i)]. \tag{30}$$

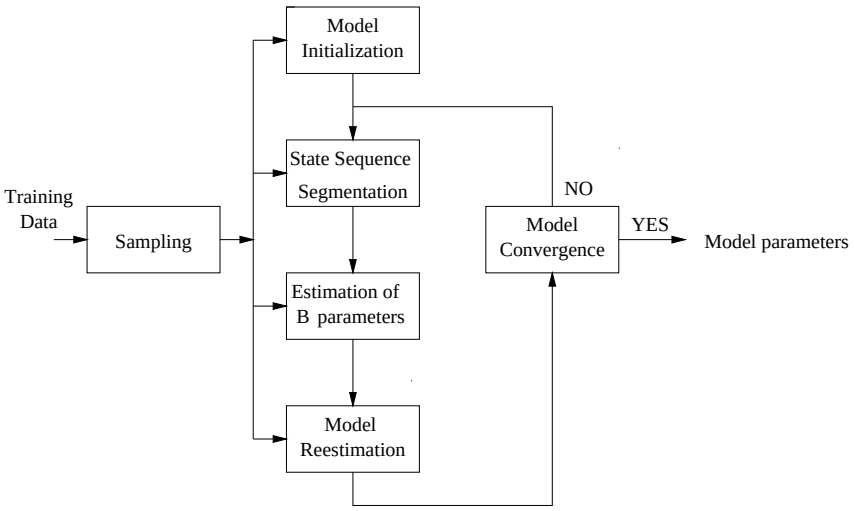

Figure 7: HMM training scheme

Recognition Performance: The recognition performances were tested on a small database of 50 images that were not part of the training dataset of 24 images [37]. The images in the test set contain faces with different facial expressions, facial details (with and without glasses) and variations in lighting. On this database the authors reported a recognition rate of 84%. On the same database the recognition rate obtained by running the eigenface method was 73%. However, the computation involved for recognition takes approximately 12 seconds to classify an image using the set of 24 training models, on a SunSparc II workstation.

7 Conclusions

In this report, most successful statistical-based approaches to face recognition were analyzed. Due to the fact that these methods were tested on different databases, a quantitative comparison cannot be presented. The recognition results of the approaches disscused are shown in Table 1. The parameters of comparison between these methods are the recognition rates reported under variations in lighting and viewing conditions facial expressions, number of classes used by the recognition system as well as computation complexity. Following conclusions were extracted from this study:

- The correlation method performs with high accuracy, if lighting and size normalization is applied, under variations in facial expression and pose. However, this method is computa-

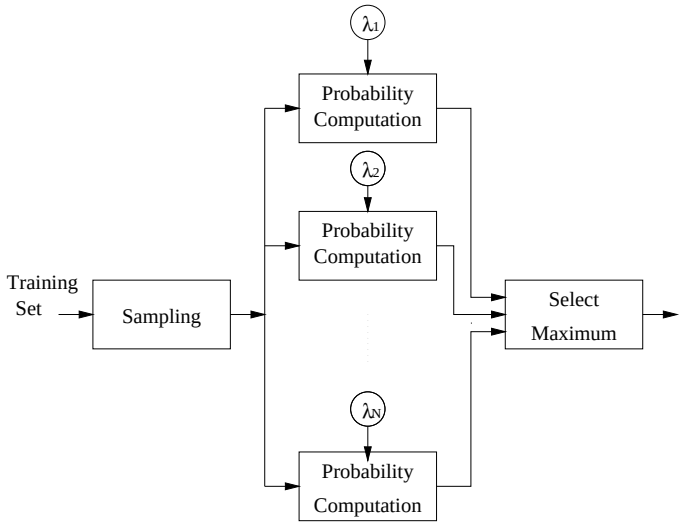

Figure 8: HMM recognition scheme

tionally very complex. The dependency of correlation value versus image rotation and scalling is disscused in section 2.

- A more efficient approach to face recognition is the eigenface method. Although the recognition performance are lower than the correlation method, the substantial reduction in computational complexity of the eigenface method makes this method very attractive. The recognition rates increase with the number of principal components used (eigenfaces) and in the limit as more principal components are used, performances approaches that of correlation. In [19] and [14], the authors reported that the performances level-off at about 45 principal components.

- In [19], it has been shown that removing first three principal components results in better recognition performances (authors reported an error rate of 20% when using the eigenface method with 30 principal components on a database strongly affected by illumination variations and only 10% error rate after removing the first three components). The recognition rates in this case were better than the recognition rates obtained using the correlation method. This was argued based on the fact that first components are more influenced by variations of lighting conditions.

- The recognition performances of the eigenface and correlation methods against large variations in illumination and pose were improved by using either a parametric approach or a view based-approach. It has been illustrated that the parametric approach has a reduced computational complexity but the view based approach is more accurate.

- KL-FSAT shows better recognition results than the classical eigenface method to small variations in head pose and face shifting. The complexity of this method is increased by the computation of the images Fourier Spectrum.

- Although the parametric and view-based approaches of the eigenface method perform very well on databases affected by large pose variations, the Fisherface method is very efficient when the testing database is affected by variations in illumination and facial expressions as well as occlusions of facial features. Its computational complexity is similar to the classical eigenfaces method. In [19] the authors reported that at the same recognition rate the Fisherface method performed faster than the Eigenface method as a smaller number of eigenvectors was required.

- The Singular Value Decomposition methods showed perfect recognition performances when tested on small database. The invariance properties of the SVD vector make this method robust to variations in illumination, orientation and facial expressions. However, the SVD method is computationally more expensive than the Eigenface method due to the computation of the SV vector for each test image.

- A layered representation of the face images (including information from facial features as well as the whole face image) leads to better recognition results. The success of the eigenfeature method empirically proved this. However, the detection of facial features under general viewing conditions is a difficult problem and needs in general significant initial guidance. In [17], the authors suggested that the use of the eigenface method to detect the face pose can be combined with the eigenfeature method to perform recognition in the selected viewspace. Furthermore, a more elaborate weighting scheme for the computation of the cumulative score of the eigenfeature method can result in better recognition performance.

- The HMM based methods which are successfully used in speech recognition, showed significantly better performances for face recognition than the Eigenface method. This is due to fact that HMM based method offers a solution to facial features detection as well as face recognition. However the 1D continuous HMM are computationally more complex than the Eigenface method. A solution in reducing the running time of this method is the use of discrete HMM. Very encouraging preliminary results (error rates below 5%) were reported in [37] when pseudo 2D HMM are used. Furthermore, the authors suggested that a Fourier representations of the images can lead to better recognition performance as frequency and frequency-space representation can lead to better data separation.

Table 1: Comparison of face recognition methods

Method	Training Set	Testing Set	Recognition Results	Type of database	Complexity	Comparaison with other methods
Correlation [3]	Not specified	Not specified	over 96%	Frontal faces, small variations in lummination, scale	high	NA
Correlation [5]	62 people 15 images per person	62 people 10 images per person	98.7%	Strong rotations in depth, small var. in scale and illumination	10-15 min on Sparc 2	NA
Eigenface	16	2500	96%	Lighting variations	350 msec on Sparc 1	NA
			85%	Orientation variations		
			64%	Variations in size		
Eigenface parametric approach	128	7562	88%	Variations in pose (interpolation)	higher than the parametric approach	90% view-based approach
			78%	Variations in pose (extrapolation)		83% view-based approach
Eigenface view based approach	128	7562	90%	Variations in pose (interpolation)	lower than the view-based approach	83% parametric approach
			83%	Variations in pose (extrapolation)		90% parametric approach
Eigen Features	45 subjects 2 images per subject	Not specified	95% combined 98%	Variations in facial expressions (neutral vs smiling)	Not specified	NA
KL-FSAT	269	5 sample of 20 individuals	91%	Variations in head orientation and shifting	higher than eigenfaces method	85% eigenface
LDA Fisherfaces	16 subjects 10 images per person	16	99.4%	Strong variations in lighting, facial expre sions and eye wear	lower than eigenfaces	eigenface(30) 80% correlation 80% eigenface (w/o 3) 90%
SVD [24]	8 subjects 3 images per person	40	100%	Small variations in illumination and face orienttaion	high due to the computation of SVD vector	NA
SVD [26]	8 subjects 3 images per person	40	100%	Small variations in illumination and face orientation	high due to the computation of SVD vector	NA
HMM	24 subjects 5 images per person	24	84%	Variations in facial expression	12 sec on Sparc 2	eigenface 73%

References

[1] R. Chellappa, C. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proceedings of IEEE*, vol. 83, May 1995.

[2] M. Bichsel and A. Pentland, "Human face recognition and face image set's topology," *CVIP:Image Understanding*, vol. 59, pp. 254–261, March 1994.

[3] R. Brunelli and T. Poggio, "Face recognition:features versus templates," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 15, October 1993.

[4] A. Yuille, "Deformable templates for face recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 59–70, 1991.

[5] L.Yuille, P. Hallinan, and D. Cohen, "Feature extraction from faces using deformable templates," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 99–111, 1992.

[6] B. Manjunath, R. Chellappa, and C. d. Malsburg, "A feature based approach to face recognition," *Proceeding of IEEE Computer Society. Conference on Computer Vision and pattern Recognition*, pp. 373–378, 1992.

[7] G. Chow and X. Li, "Towards a system for automatic facial feature detection," *Pattern Recognition*, vol. 26, no. 12, pp. 1739–1755, 1993.

[8] L. Stringa, "Eyes detection for face recognition," *Applied Artificial Intelligence*, vol. 7, pp. 365–382, Oct-Dec 1993.

[9] P. Burt, "Smart sensing within a pyramid vision machine," *Proceedings of IEEE*, vol. 76, pp. 1006–1015, August 1988.

[10] D. Beymer, "Face recognition under varying pose," in *Proceedings of 23rd Image Understanding Workshop*, vol. 2, pp. 837–842, 1994.

[11] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Proceedings of International Conference on Pattern Recognition*, pp. 586 – 591, 1991.

[12] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, March 1991.

[13] L. Sirovitch and M.Kirby, "Low-dimensional procedure for the characterization of human faces," *Journal of the Optical Society of America*, vol. 4, pp. 519–524, March 1987.

[14] M. Kirby and L. Sirovitch, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 103–108, January 1990.

[15] R. Eppstein, P. Hallinan, and A. Yuille, "5+/-2 eigenimages suffice:an empirical investigation of low dimensional lighting models," in *Proceedings of the Workshop on Physics-Based Modelling in Computer Vision*, 1995.

[16] H. Murase and S. Nayar, "Visual learning and recognition of 3-d objects from appearance," *International Journal of Computer Vision*, vol. 14, pp. 5–24, 1995.

[17] A. Pentland, B. Moghadam, T. Starner, and M. Turk, "View based and modular eigenspaces for face recognition," in *Proceedings on IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 84–91, 1994.

[18] H. F. S. Akamatsu, T. Sasaki and Y. Suenuga, "A robust face identification scheme - KL expansion of an invariant feature space," in *SPIE Proceedings::Intelligent Robots and Computer Vision X:Algorithms and Technology*, vol. 1607, pp. 71–84, 1991.

[19] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs Fisherfaces: Recognition using class specific linear projection," in *Proceedings of Fourth Europeean Conference on Computer Vision, ECCV'96*, pp. 45–56, April 1996.

[20] K. Etemad and R. Chellapa, "Face recognition using discriminant eigenvectors," in *Proceedings of ICASSP*, 1996.

[21] K. Fukunaga, *Introduction to Statistical Pattern Recognition.* Academic Press, 1990.

[22] R. Duda and P. Hart, *Pattern Classification and Scene Analysis.* New York: Wiley, 1973.

[23] Y. Cheng, K. Liu, J. Yang, Y. Zhang, and N. Gu, "Human face recognition method based on the statistical model of small sample size," in *SPIE Proceedings:Intelligent Robots and Computer Vision X:Alg. and Techn*, vol. 1607, pp. 85–95, 1991.

[24] Z. Hong, "Algebraic feature extraction of image for recognition," *Pattern Recognition*, vol. 24, pp. 211–219, 1991.

[25] Q. Tian, "Comparaision of statistical pattern-recognition algorithms for hybrid processing, ii: eigenvector-based algorithms," *Journal of the Optical Society of America*, vol. 5, pp. 1670–1672, 1988.

[26] Y. Cheng, K. Liu, J. Yang, and H. Wang, "A robust algebraic method for human face recognition," in *Proceedings of 11th International Conference on Pattern Recognition*, pp. 221–224, 1992.

[27] L. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition," *Proceedings of IEEE*, vol. 77, pp. 257–286, February 1989.

[28] L. Rabiner and B. Huang, *Fundamentals of Speech Recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[29] E.Levin and R. Pieraccini, "Dynamic planar warping for optical character recognition," in *ICAASP*, pp. 149–152, 1992.

[30] O. Agazzi, S. Kuo, E. Levin, and R. Pieraccini, "Connected and degraded text recognition using planar HMM," in *ICASSP '93*, vol. 5, pp. 113–116, 1993.

[31] S. Kuo and O. Agazzi, "Keyword spotting in poorly printed documents using pseudo 2-d Hidden Markov Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994.

[32] O. Agazzi and S. Kuo, "Hidden Markov Models based optical character recognition in presence of deterministic transformations," *Pattern Recognition*, vol. 26, no. 12, pp. 1813–1826, 1993.

[33] F. Samaria, "Face segmentation for identification using Hidden Markov Models," in *British Machine Vision Conference*, 1993.

[34] F. Samaria and F. Fallside, "Face identification and feature extraction using Hidden Markov Models," *Image Processing: Theory and Applications*, 1993.

[35] F. Samaria and F. Fallside, "Automated face identification using hidden markov models," in *Proceedings of the International Conference on Advanced Mechatronics*, 1993.

[36] F. Samaria and A. Harter, "Parametrisation of stochastic model for human face identification," in *Proceedings of the Second IEEE Workshop on Application of Computer Vision*, 1994.

[37] F. Samaria and S. Young, "HMM based arhitecture for face identification," *Image and Computer Vision*, vol. 12, October 1994.