Volume 2, No. 7, July 2011
Journal of Global Research in Computer Science
TECHNICAL NOTE
Available Online at www.jgrcs.info
ISSN-2229-371X

EVALUATION OF FACE RECOGNITION METHODS

Sushma Jaiswal*[1], Dr. Sarita Singh Bhadauria[2], Dr.Rakesh Singh Jadon[3]

[1]Lecturer, S.O.S. in Computer Science, Pt.Ravishankar Shukla University,Raipur(C.G.)
jaiswal1302@gmail.com
[2]Professor & Head,Department of Electornics Engineering ,Madhav Institute of Technology & Science,Gwalior(M.P.)
[3]Professor & Head,Department of Computer Applications ,Madhav Institute of Technology & Science,Gwalior(M.P.)

Abstract: Face recognition is an example of advanced object recognition. The process is influenced by several factors such as shape, reflectance, pose, occlusion and illumination which make it even more difficult. Today there exist many well known techniques to try to recognize a face. We present to the reader an investigation into individual strengths and weaknesses of the most common techniques including feature based methods, PCA based eigenfaces, LDA based fisherfaces, ICA, Gabor wavelet based methods, neural networks and hidden Markov models. Hybrid systems try to combine the strengths and suppress the weaknesses of the different techniques either in a parallel or serial manner. Today there exist many well known techniques to try to recognize a face. Experiments done with implementations of different methods have shown that they have individual strengths and weaknesses. Hybrid systems try to combine the strengths and suppress the weaknesses of the different techniques either in a parallel or serial manner. The paper is to evaluate the different techniques and consider different combinations of these.
Here we compare or evaluate templates based and geometry based face recognition, also give the comprehensive survey based face recognition methods.

Keywords-- PCA based eigenfaces, LDA based fisherfaces, ICA, and Gabor wavelet based methods, neural networks, hidden Markov models

INTRODUCTION

Face recognition is an example of advanced object recognition. The process is influenced by several factors such as shape, reflectance, pose, occlusion and illumination. A human face is an extremely complex object with features that can vary over time, sometimes very rapidly. It is covered with skin, a non-uniformly textured material that is difficult to model. Skin can change color quickly when one is embarrassed or becomes warm or cold and the reflectance properties of the skin change as the perspiration level changes.

LITERATURE REVIEW OF FACE RECOGNITION AND EVALUATION

Recent Approaches to Face Recognition

Face recognition has been an active research area over last 35 years. This research spans several disciplines such as image processing, pattern recognition, computer vision, and neural networks. It has been studied by scientists from different areas of psychophysical sciences and those from different areas of computer sciences. Psychologists and neuroscientists mainly deal with the human perception part of the topic, whereas engineers studying on machine recognition of human faces deal with the computational aspects of face recognition. Face recognition has applications mainly in the fields of biometrics, access control, law enforcement, and security and surveillance systems.

The problem of face recognition can be stated as follows: Given still images or video of a scene, identifying one or more persons in the scene by using a stored database of faces [8]. The problem is mainly a classification problem. Training the face recognition system with images from the known individuals and classifying the newly coming test images into one of the classes is the main aspect of the face recognition systems.

This problem seems to be easily solved by humans where limited memory can be the main problem; whereas the problems for a machine face recognition system are:

Facial expression change, Illumination change, Aging, Pose change, Scaling factor (i.e. size of the image), Frontal vs. profile

7. Presence and absence of spectacles, beard, mustache etc.
8. Occlusion due to scarf, mask or obstacles in front.

The problem of automatic face recognition (AFR) is a composite task that involves detection of faces from a cluttered background, facial feature extraction, and face identification. A complete face recognition system has to solve all subproblems, where each one is a separate research problem. This research work concentrates on the problem of facial feature extraction and face identification.

Most of the current face recognition algorithms can be categorized into two classes, image template based and geometry feature-based. The template based methods [1] compute the correlation between a face and one or more model templates to estimate the face identity. [4] suggest that the optimal strategy for face recognition is holistic and corresponds to template matching. In their study, they compared a geometric feature based technique with a template matching based system and reported an accuracy of 90% for the first one and 100% for the second one on a database of 97 persons. Statistical tools such as Support Vector Machines (SVM) [27,38], Principal Component Analysis (PCA) [32,36], Linear Discriminant Analysis (LDA) kernel methods [30], and neural networks [14, 21 ,29] have been used to construct a suitable set of face

templates. Other than statistical analysis and neural network approach there are other approaches known as hybrid approaches which use both statistical pattern recognition techniques and neural network systems.
Examples for hybrid approaches include the combination of PCA and Radial Basis Function (RBF) neural network [35]. Among other methods, people have used range [8], infrared scanned [43] and profile [24] images for face recognition. While templates can be viewed as features, they mostly capture global features of the face image. Facial occlusion is often difficult to handle in these approaches. The geometry feature based methods analyze explicit local facial features, and their geometric relationships. Cootes [19] have presented an active shape model in extending the approach [44]. [40] developed an elastic bunch graph matching algorithm for face recognition. Penev [28,60] developed PCA into Local Feature Analysis (LFA). This technique is the basis for one of the most successful commercial face recognition systems, FaceIt. The summary of approaches to face recognition is shown in Fig. 1.

Template based Methods

Template matching is conceptually related to holistic approach which attempts to identify faces using global representations [15]. These types of methods approach the face image as a whole and try to extract features from the whole face region and then classify the image by applying a pattern classifier. One of the methods used to extract features in a holistic system, is based on statistical approaches which are discussed in the following section.
The other algorithm proposed by Brunelli and Poggio is based on template matching. In the simplest version of template matching the image, represented by an array of intensity values, is compared using a suitable metric (typically Euclidean distance) to a single template representing the whole face. More sophisticated methods can use several templates per face to take into account the recognition from different viewpoints.
First the image is normalized using the same technique described in the previous section. Each person is stored in the database associated with template masks representing digital images of eyes, nose, mouth etc. Recognition of an unclassified image is done by comparing parts of it with all the templates stored in the database returning a matching score for each individual. The unknown individual is then classified as the one giving the highest cumulative comparison score.

Statistical Approaches

Images of faces, represented as high-dimensional pixel arrays, often belong to a manifold of intrinsically low dimension. Face recognition research has witnessed a growing interest in techniques that capitalize on this observation, and apply algebraic and statistical tools for extraction and analysis of the underlying manifold. The techniques that identify, parameterize and analyze linear subspaces are described below. Other than linear subspaces there are some statistical face recognition techniques which are based on nonlinear subspaces (like kernel-PCA and kernel-LDA), transformation (like DCT, DCT & HMM and Fourier Transform) and Support Vector Machine (SVM).
Appearance-based approaches for face recognition like PCA, LDA, and probabilistic subspace view a 2D face image as a vector in image space. A set of face images $\{x_i\}$ can be represented as a $M \times N$ matrix $X = [x_1, x_2, x_3, \ldots \ldots x_N]$, where M is total number of pixels in the images and N is the total number of samples. Each of the face images x_i belongs to one of the C classes $\{1,2,, \ldots .C\}$.

I. Template Based methods

1. **Statistical**
   - **Linear Subspaces**
     - ❖ Eigenfaces(PCA)
     - ❖ Probabilistic Eigenfaces(PPCA)
     - ❖ Fisher faces(LDA)
     - ❖ Bayesian Methods (MAP and ML)
     - ❖ ICA and source separation
     - ❖ Tensorfaces (Multi-linear SVD)
     - ❖ Two Dimensional PCA(2D-PCA)
     - ❖ Two Dimensional LDA(2D-LDA)
     - ❖ Discriminative Common Vectors(DCV)
     - ❖ HMM
     - ❖ ICP
   - **Nonlinear Subspaces**
     - ❖ Principal Curves and Nonlinear PCA
     - ❖ Kernel-PCA
     - ❖ Kernel-LDA
   - **Transformed based**
     - ❖ DCT
     - ❖ DCT and HMM
     - ❖ Fourier Transform(FT)
   - **Others**
     - ❖ SVM
2. **Neural Network**
   - ❖ Feature based back propagation NN
   - ❖ Dynamic Link Architecture (DLA)
   - ❖ Single layer Adaptive NN
   - ❖ Multilayer perceptron (MLP)
   - ❖ Probabilistic Decision based Neural Network (PDBNN)
   - ❖ Self-organizing Map (SOM)
   - ❖ Hopfield memory
3. **Hybrid**
   - ❖ PCA and RBF
4. **Others**
   - ❖ Range Data
   - ❖ Infrared Scanning
   - ❖ Profile Images

II. Geometry Feature Based Methods

- ❖ Active Shape Model
- ❖ Wavelets
- ❖ Elastic Bunch Graph Matching
- ❖ Local feature Analysis(LFA)

Figure 2.1: Summary of approaches to face recognition.

PCA (Principal Component Analysis):

The key idea behind PCA [36,47] is to find the best set of projection directions in the sample space that maximizes total scatter across all images. This is accomplished by computing a set of eigenfaces from the eigenvectors of total scatter matrix S_t, defined as:

$$S_t = \sum_{i=1}^{N}(x_i - m)(x_i - m)^T, \quad (2.1)$$

Where m is the mean face of the sample set X. The geometric interpretation of PCA is shown in Fig. 2.2. For dimensionality reduction, K (where$K < M$) eigenvectors $U = [u_1, u_2, \ldots, u_K]$ Corresponding to first K largest eigenvalues of St are selected as eigenfaces. Reduced dimension training samples, $Y = [y_1, y_2, \ldots, y_N]$ can be obtained by the transformation $Y = U^T X$.

Now, when a probe image x_t is presented for identifica tion/verification, it is projected on U to obtain a reduced vector$y_y = U^T x_t$. A response vector of length C, $R(x_t) = [r_1, r_2, \ldots, r_C]$ is calculated by measuring distances from the probe to the nearest training samples from each class. The distance function between two vectors can be expressed in the following way:

$$d = (y_i, y_j) = \|y_i - y_j\|^2 \quad (2.2)$$

The desired class label for the probe image can be obtained by minimum membership rule.

$$\mathcal{L}(x_t) = arg\,\min_c r_c. \quad (2.3)$$

PCA also known as Karhunen-Loeve (KL) transformation or eigenspace projection, a frequently used statistical technique for optimal lossy compression of data under least square sense, provides an orthogonal basis vector-space to represent original data. The first introduction of a low-dimensional characterization of faces was developed at Brown University [47], [48]. This was later extended to eigenspace projection for face recognition [49,36,50,51,52]. More recently Nayar, Nene and Murase used eigenspace projection to identify objects using a turntable to view objects at different angles [53,54]. and Drew extended greyscale eigenfaces to colour images [55].

PCA-Evaluation

Results show that eigenfaces methods are robust over a wide range of parameters and produce good recognition rates on various databases [56]. However outside this parameter range the algorithm can breakdown sharply. Results show that Eigenfaces are very robust to low resolution images as long as the preprocessing step can extract sufficient features for normalization. They also handle high resolution images very efficiently. It seems to be vital that the preprocessing step is working well. A normalization process can solve rotation issues by aligning both eyes horizontally, scale by adjusting the distance between the eyes and translation by cropping the image. Horizontal and vertical misalignment of only 5% because of difficulties in detecting face features like the eyes have severe effects on the recognition rates.
Significant variation in scale, orientation, translation and lightning will also cause it to fail. One has to remember the fact that the reflectance is different for a translated image because of a slightly different viewing angle. An image from the database of face taken from one meter away will therefore not perfectly match a face taken five meters away and zoomed in an appropriate amount.
The choice of distance measure has also proven to affect the performance of face recognition. Euclidean distance (L2 norm) is the most commonly used measure and is computational easy. Yambor, Draper and Beveridge claim that the Mahalanobis distance, which uses the eigenvalues as weighting for the contribution of each axis in the eigenspace, outperforms all the other measures when having a subspace spanned by more than 20 eigenvectors [57,58]. Beveridge conclude that PCA with Mahalanobis distance is the best combination [59,76].
Selection of k, the number of eigenfaces to keep, is also an important choice because using a low number will fail to capture all the differences in the dataset, while using a high number will be computational demanding. For large databases like the FERET database at least 200 eigenfaces are needed to sufficiently capture global variations like lighting, small scale and pose variations,[60]. The results may improve by dropping some of the eigenvectors either from the front (lighting) or the back (noise).
A solution to the fundamental problem of handling pose variations seems to be using the new eigen light-field approach, but the normalization process can become time consuming when the orientation between the face and the camera is unknown and has to be estimated. Another solution handling pose variations is having several sets of eigenvectors representing different views. The recognition results are better, but the computational cost is higher.
The low computational cost recognizing faces with the traditional eigenface method comes as a result of a high computational cost training the faces [61]. In the construction of the training set, one can imagine a new face that is not well represented by the eigenfaces calculated from this training set. In this case it becomes necessary to update the training set, which implies an update of the eigenfaces. It is always possible to do a full recalculation of the eigenfaces, but this is a time consuming process. Chandrasekaran, Manjunath, Wang, Winkeler and Zhang of University of California have proposed a method of incremental updating of the eigenspace for images being significantly outside the current object eigenspace [95,62,63].

Table: I

Advantages	Drawbacks
Robust against noise and occlusion	Removes neighborhood relationships between pixels
Robust against illumination, scaling, orientation and translation when face is correctly normalized	Sensitive to faulty normalization
Robust against facial expressions, glasses, facial hair, makeup etc.	Sensitive to perspective, viewing angle and head rotation (can be improved using eigen light-fields or other view-based methods)

Can handle high resolution images efficiently	Sensitive to large variation in illumination and strong facial expression
Can handle small training sets	Slow training/High computational cost (with large databases)
Can handle very low resolution images	----
Fast recognition/Low computational cost	---

Chain – I [3] proposed four variants of PCA namely Simultaneous PCA (SMPCA), ProGressive PCA (PGPCA), Successive PCA (SCPCA) and Prioritized PCA (PRPCA).

SMPCA solves all the roots of the characteristic polynomial equation derived from the sample covariance matrix, referred to as eigenvalues and find their corresponding eigenvectors which will be used to generate all principal components simultaneously.

PGPCA does not solve all the roots of the characteristic equations as in SMPCA. It only locates the maximum Eigenvalue for each of the sample covariance matrices formed by a sequence of reduced subspaces so that principal components can be generated one at a time progressively.

SCPCA does not solve characteristic polynomial equation. This technique uses a learning algorithm to generate the PCs. A random initial vector is used to produce a projection vector and the same process is repeated to generate successive PCs.

PRPCA uses a custom designed initialization algorithm to appropriate set of initial projection vectors for the PCA. The PCs are prioritized by the projection vectors. An example of such an PCA. The new version divides the image into equal size sub images on which PCA is applied resulting in the formation of individual eigenfaces for each sub image. Even though PCA produces good results, it is computationally very complex with increase in database size. A new PCA based algorithm using geometry and symmetry of the faces which extracts features using fast Fuzzy.

LDA (Linear Discriminant Analysis):

The objective of LDA is to find the subspace that best discriminates different face classes by maximizing between class scatter, while minimizing the within-class scatter. The eigenvectors chosen by LDA provide the best separation among the class distributions, while PCA selects eigenvectors which provide best representation of the overall sample distribution. To illustrate the difference, Fig. 2.3 shows the first projection vector chosen by PCA and LDA for a two class problem. The eigenvectors for LDA can be obtained by computing the eigenvectors of $S_w^{-1}S_b$.

Here, S_b and S_w are the between-class and with in-class scatter matrices of training samples and are defined as:

$$S_w = \sum_{i=1}^{c} \sum_{x_k \in C_i} (x_k - m_i)(x_k - m_i)^T, \qquad (2.4)$$

$$S_b = \sum_{i=1}^{c} n_i (m_i - m)(m_i - m)^T, \qquad (2.5)$$

where m_i is the mean face for i^{th} class and n_i is the number of training samples in i^{th} class. LDA subspace is spanned by a set of vectors W, which maximizes the criterion, J, defined as:

$$J = \frac{tr(S_b)}{tr(S_w)} \qquad (2.6)$$

W can be constructed by the eigenvectors of $S_w^{-1}S_b$. In most of the image processing applications, the number of training samples is usually less than the dimension of the sample space. This leads to the so-called small-sample-size (SSS) problem due to the singularity of the within-class scatter matrix. To overcome SSS problem, the following approaches are attempted: a two stage PCA+LDA approach [34], Fisherface method [2] and discriminant component analysis [46]. In all cases the higher dimension face data is projected

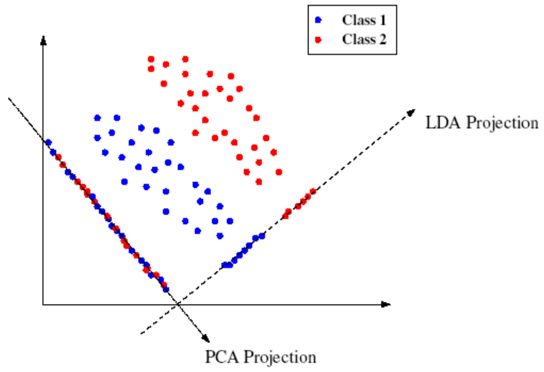

Figure 3: An example of PCA and LDA projection for a two class problem. to a lower dimension space using PCA and then LDA is applied to this PCA subspace.

Fisherfaces (LDA)-

R. A. Fisher developed Fisher's Linear Discriminant (FLD) [R. A. Fisher (1936)] in the 1930's but not until recently have Fisher discriminants been utilized for object recognition. Swets and Weng used FLD to cluster images for the purpose of identification in 1996 [34,65,67]. Also in 1997, Belhumeur, Hespanha and Kriegman of Yale University used FLD to identify faces, by training and testing with several faces under different lighting [66]. Fisher Linear Discriminant (FLD) analysis, also called Linear Discriminant Analysis (LDA) finds the line that best separates the points. For example, consider two sets of points, coloured green and blue, in two-dimensional space being projected onto a single line. Depending on the direction of the line, the points can either be mixed together (Figure 9a) or be separated (Figure 9b). In terms of face recognition this means grouping images of the same class and separate images of different classes. Images are projected from a N-dimensional space, where N is the number of pixels in the image, to a M-1 dimensional space, where M is the number of classes of images [57, 58, 67, 68]. The approach is similar to the eigenface method, which makes use of projection of training images into a subspace. The test images are projected into the same subspace and identified using a similarity measure. What differs is how subspace is calculated. The eigenface method uses PCA for dimensionality reduction, which yields directions that maximize the total scatter across all classes of images. This projection is the best for reconstruction of images from a low-dimensional basis. However, the method does not make use of between-class scatter between classes of face images belonging to the same individual. A PCA projection may not create an optimal discrimination for different classes.

The LDA method, which creates an optimal projection of the dataset, maximizes the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples. The within-class scatter matrix, also called intra-personal, represents variations in appearance of the same individual due to different lighting and face expression, while the between-class scatter matrix, also called the extra-personal, represents variations in appearance due to a difference in identity. In this way fisherfaces can project away some variation in lighting and facial expression while maintaining discriminability. [69]

LDA- Evaluation

The fisherface method is very similar to the eigenface method, but with improvement in better classification of face images by using interclass and intraclass relationships to separate them. With LDA it is possible to classify the training set to deal with different people and different facial expressions. The accuracy for handling facial expressions has shown to be better than the eigenfaces method.

The fisherfaces method is quite insensitive to large variations in lighting direction and facial expression. Compared to the eigenface method this algorithm is more complex, something which increases the computational requirements, but show lower error rates. Besides, due to the need of better classification, the dimension of the projection in face space is not as compact as in the eigenfaces approach. This results in larger storage of the faces and more processing time in recognition.

Another drawback comes from the fact that the fisherface method uses particular class information and therefore is recommended to have many images per class in the training process. On the other hand, having many images belonging to the same class can make the recognition system suffer from a lack of generalization resulting in a lower recognition rate. In general the algorithm is performing very well, but cannot always work. In general it fails when the between class scatter is inherently greater than the within class scatter. [70]

Table: II

Advantages **Advantages**	**Drawbacks** **Disadvantages**
Robust against noise and occlusion	Removes neighbourhood relationships between pixels
Robust against illumination, scaling, orientation and translation when face is correctly normalized	Sensitive to faulty normalization
Robust against facial expressions, glasses, facial hair, makeup etc.	Sensitive to perspective, viewing angle and head rotation (can be improved using fisher light-fields)
Can handle high resolution images efficiently	Does not handle small training sets well
Can handle very low resolution images	Slow training/High computational cost (with large databases)
Fast recognition/Low computational cost	

- DCV (Discriminative Common Vectors Approach): DCV [7] solves "small sample size problem" of LDA by optimizing a variant of Fisher's criterion. It searches for the optimal projection vectors in the null space of within class scatter S_w (see equation 2.4), satisfying the criterion,

$$J(W_{opt}) = arg \max_{|W^T S_w W|=0} |W^T S_b W| = \arg \max_{|W^T S_w W|=0} |W^T S_t W| \quad (2.7)$$

So, to find the optimal projection vectors in the null space of S_w, it projects the face samples onto the null space of S_w to generate common vectors for each class and then obtain the projection vectors by performing PCA on common vectors. A new set of vectors, called as discriminative common vectors, are obtained by projecting face samples on the projection vectors. Thus each class is represented by a single discriminative common vector. Among two algorithms to extract the discriminant common vectors for representing each person in the training set of face database, one algorithm uses within-class scatter matrix of the samples in the training set while the other uses the subspace methods and the Gram-Schmidt orthogonalization procedures to obtain the discriminative common vectors. These discriminative common vectors are used for classification of new faces.

- Probabilistic Eigenspace Method: Probabilistic subspace method models intra-personal and extra-personal variations to classify the face intensity difference Δ as intra-personal variation (Ω_I) for the same class and extra-personal variation (Ω_E) for different classes. The MAP similarity between two images is defined as the intra-personal a posterior probability:

$$S(I_1,I_2) = P(\Omega_I|\Delta) = \frac{P(\Delta|\Omega_I)P(\Omega_I)}{P(\Delta|\Omega_I)P(\Omega_I) + P(\Delta|\Omega_E)P(\Omega_E)} \quad (2.8)$$

To estimate $P(\Delta|\Omega_I)$ and $P(\Delta|\Omega_E)$ the eigenvectors of intra-personal and extra-personal subspaces are computed from the difference set

$$\{(x_i - x_j)|\mathcal{L}x_i) = \mathcal{L}(x_j)\} \; and \; \{(x_i - x_j)|\mathcal{L}x_i) \neq \mathcal{L}(x_j)\}$$

Respectively. The covariance matrices for intra-personal and extra-personal difference sets are defined as:

$$S_I = \sum_{\mathcal{L}(x_i)=\mathcal{L}(x_j)} (x_i - x_j)(x_i - x_j)^T, \quad (2.9)$$

$$S_E = \sum_{\mathcal{L}(x_i)\neq\mathcal{L}(x_j)} (x_i - x_j)(x_i - x_j)^T, \quad (2.10)$$

To estimate $P(\Delta|\Omega_I)$ the eigenspace of S_I is decomposed into intra-personal principal subspace F, spanned by the L largest eigenvectors, and its orthogonal complementary subspace $\bar{F}$ with dimension $M - L$. Then $P(\Delta|\Omega_I)$ can be obtained as the product of two independent marginal Gaussian densities in F and $\bar{F}$,

$$P(\Delta|\Omega_1) = \left[\frac{\exp\left(-\frac{1}{2}d_F(\Delta)\right)}{(2\pi)^{L/2}\prod_{i=1}^{L}\lambda_i^{1/2}}\right]\left[\frac{\exp\left(-\varepsilon^2(\Delta)/2\rho\right)}{(2\pi\rho)(N-L)/2}\right]$$

$$= \frac{exp\left[e - \frac{1}{2}d_F(\Delta) + \varepsilon^2(\Delta)/\rho\right]}{\left[(2\pi)^{L/2}\prod_{i=1}^{L}\lambda_i^{1/2}\right]\left[(2\pi\rho)(N-L)/2\right]} \quad (2.11)$$

Here, $d_F(\Delta) = \sum_{i=1}^{L} \frac{y_i^2}{\lambda_i}$ is a Mahalanobis distance in F and referred as "distance in- feature-space" (DIFS). y_i is the principal component of Δ projecting to the i^{th} intra-personal eigenvector, and λ_i is the corresponding eigenvalue. "$\varepsilon^2(\Delta)$, defined as "distance-from-feature-space" (DFFS), is the PCA residual (reconstruction error) in $\bar{F}$. ρ is the average eigenvalue in $\bar{F}$. $P(\Delta|\Omega_E)$ Can be Estimated in a similar way in extra-personal subspace computed from S_E. An alternative maximum likelihood (ML) measure, using the intra-personal likelihood $S'(\Delta) = P(\Delta|\Omega_I)$ is as effective as MAP measure. In face recognition, all parameters in Equation 2.11 are same except $d_F(\Delta)$ and $\varepsilon^2(\Delta)$. So, it is equivalent to evaluate the distance,

$$D_I = d_F(\Delta) + \frac{\varepsilon^2(\Delta)}{\rho}. \quad (2.12)$$

- 2D-PCA (Two Dimensional Principal Component Analysis): The 2DPCA technique [45] is based on 2D image matrices instead of 1D vectors. The image covariance matrix is constructed directly from the original image. Let A denote the image with m rows and n columns. The image matrix is projected on to n-dimension column vector x as,

$$y = ax \quad (2.13)$$

where y is an m dimensional column vector called the projected feature vector of the image A. Let G_t is the image covariance matrix and is represented as,

$$G_t = E[(A - \bar{A})^T (A - \bar{A})] \quad (2.14)$$

where $E[.]$ is the expectation operator and $\bar{A}$ is the mean image of the entire training set. The column vector x is chosen such that the generalized total scatter criterion is maximized. This means that total scatter of the projected vectors is maximized. [45] that the scatter criterion is satisfied when x is chosen as the orthonormal eigenvectors of the image covariance matrix G_t. For each eigenvectorx_s, there exists a projected feature vectory_s. Therefore, the principal component of 2D-PCA is a vector unlike that of PCA which is a scalar. If S principal components are used (corresponding to the largest S eigenvalues of G_t), the principal component vectors obtained can be represented as, $B = [y_1, y_2, \dots, y_S]$.

Thus, B is the feature matrix of the image sample A. For recognition using the nearest neighbor strategy, the distance between any two feature matrices $B_i = [y_1^{(i)}, y_2^{(i)}, \dots\dots, y_S^{(i)}]$ and $B_j = [y_1^{(j)}, y_2^{(j)}, \dots\dots, y_S^{(j)}]$ is given as,

$$d(B_i, B_j) = \sum_{i=1}^{n} \left\| y_s^{(i)} - y_s^{(j)} \right\|_2 \quad (2.15)$$

Where $\left\| y_s^{(i)} - y_S^{(j)} \right\|_2$ is the Euclidean distance between the two principal component Vectors $y_s^{(i)}$ and $y_S^{(j)}$. A test image feature B_t is assigned to class C_k if, $d(B_t, B_j) = \min_i d(B_t, B_j), and\ B_l \epsilon C_k$. Unlike the conventional PCA, the 2D-PCA does not involve computation of a large correlation matrix and therefore is relatively less computation intensive. But on the other hand, it requires more memory for storing feature matrices.

[45] tested 2D-PCA method on ORL, AR and Yale face databases. For the ORL database, the authors used two strategies for experiments: (a) 5 image samples per class for training and (b) leave one out strategy for observing the average performance. In case (a) the recognition rate is 96% and in case of (b) the same is reported to be 98.3% for ORL. Leave one out strategy was adopted for Yale database and maximum accuracy is reported to be 84.24%.

2D-LDA (Two dimensional Linear Discriminant Analysis): In the recently proposed 2D-LDA [22], the image is not reordered as a column vector. The process of projection of an image using 2D-LDA is given as follows. An image A of m rows and n columns is projected as $y = Ax$, where y is the projected vector and x is the projection vector. Optimal projection vectors are chosen when Fisher's criterion is maximized. The criterion is expressed as,

$$J(x) = \frac{x^T G_b x}{x^T G_w x} \quad (2.16)$$

where G_b and G_w are the image between-class scatter matrix and the image within-class scatter matrix as given below,

$$G_b = \sum_{k=1}^{C} n_k (\bar{A}^k - A)^T (\bar{A}^k - A) \quad (2.17)$$

$$G_w = \sum_{k=1}^{C} \sum_{A \epsilon C_k} (A - \bar{A}^k)^T (A - \bar{A}^k) \quad (2.18)$$

Where $\bar{A}^K$ is the mean image of class C_k and $\bar{A}$ is the global mean. The projection vector x $G_w^{-1} G_b$ is taken as the eigenvector of If the first S eigenvectors are used (corresponding to the largest S eigenvalues of $G_w^{-1} G_b$, the feature obtained can be represented as $B = [y_1, y_2, \dots, y_S]$. The classification using nearest neighbor strategy is similar to the classi_cation using 2D-PCA.

Among those earliest to report the work on 2D-LDA [25]. [17] address the SSS (small sample size or undersampled) problem in LDA utilizing a 2D-FDA algorithm. The recognition performance was obtained by varying the number of training samples in the range: 2-9 in case of ORL with maximum accuracy 98%, 2-12 in case of Yale-B with maximum accuracy 92%. The latest of the works on 2D-LDA [22,41].

Independent Component Analysis (ICA):

Independent Component Analysis (ICA) is a technique for extracting statistically independent variables from a mixture of them [71]. The technique is quite new and has originated from the world of signal processing. A classical example demonstrating the original problem is the cocktail-party problem where two people being in the same room speak simultaneously. Two microphones are placed at different locations recording the mixed conversations. It would be very useful if one could estimate the two original speech signals from the two mixed recordings. Surprisingly it turns out that it is enough to assume that the two speech signals are statistically independent. This is not an unrealistic assumption, but it does not need to be exactly true in practice. ICA can be used to estimate the contribution coefficients from the two signals, which allows us to separate the two original signals from each other. Hyvärinen and Oja have written a good tutorial about ICA which contains more details about the algorithms involved [72].

In a task such as face recognition, much of the important information may be contained in the high-order relationships among the image pixels. Some success has been attained using data-driven face representations based on PCA, such as eigenfaces. PCA is based on the second-order statistics of the image set, and does not address high-order statistical dependencies such as the relationships among three or more pixels. Independent component analysis (ICA) however separates the high-order moments of the input in addition to the second-order moments. ICA thus in some ways provide a more powerful data representation than PCA, as its goal is to provide an independent rather than an uncorrelated image decomposition and representation.

For finding a set of independent component images, the face images X are considered to be a linear combination of statistically independent basis images S, where A is an unknown mixing matrix. The basis images are recovered by a matrix of learned filters W, which produces statistically independent outputs U . Bartlett and Sejnowski at University of California have used ICA for face recognition [7.73.74,75]. Two approaches for recognizing faces across changes in pose were explored using ICA. The first architecture provided a set of statistically independent basis images for the faces that can be viewed as a set of independent facial features. This corresponds very much to the classical cocktail-party problem performing a blind separation of a mixture of auditory signals. These ICA basis images were spatially local, unlike the PCA basis vectors. The representation consisted of the coefficients for the linear combination of basis images that comprised each face image. The second architecture produced independent coefficient. This provided a factorial face code, in which the probability of any combination of features can be obtained from the product of their individual probabilities.

Classification was performed using nearest neighbour, with similarity measured as the cosine of the angle between representation vectors. Both ICA representations showed better recognition scores than PCA when recognizing faces across sessions, changes in expression, and changes in pose.

Independent Component Analysis based method-Evaluation-

In 1999 Liu and Wechsler also claimed that ICA produced better results or matched [59]. They showed that PCA outperformed ICA when the distance method is selected to maximize performance. Both experiments were conducted using the FERET database. The most recent contradicting results from 2001 however showed that the differences in recognition rates between PCA and ICA are only minor, and very much depend on how the algorithms in detail are implemented.

Global properties like coloring, width and length are more easily captured by PCA than ICA, since ICA basis vectors are more spatially localized than their PCA counterparts. Recognizing more localized features, like face expressions, may produce significantly different results.

Table: III

Advantages	**Drawbacks**
Considers higher-order relationships	Removes neighborhood relationships between pixels
Robust against noise and occlusion	Sensitive to faulty normalization
Robust against illumination, scaling, orientation and translation when face is correctly normalized	Sensitive to perspective, viewing angle and head rotation
Robust against facial expressions, glasses, facial hair, makeup etc.	Slow training/High computational cost (with large databases)
Fast recognition/Low computational cost	

Hidden Markov Models (HMM):

The use of hidden Markov models is a powerful statistical technique that has been applied to many subject areas, from predicting political crises to the reconstruction of DNA and the recognition of speech. The September 1964 issue of Scientific American illustrated a Markov chain by showing two containers with numbered balls in them. Numbered slips of paper associated with the balls were drawn repeatedly, with replacement, from a hat. The ball associated with the number drawn was transferred to the other container than the one it was in. Initially all the balls were in the first container, and gradually this declined exponentially until it contained only half of the balls. This modelled the physical process of allowing two separate chambers, containing a gas at different levels of pressure to be connected. One basic feature with the Markov process is that it involves probability. In addition to a random event the final result also depends on some kind of system memory, described by the number of balls in the first container.

A hidden Markov model consists of two interrelated processes. First an underlying, unobservable Markov chain with a finite number of states (N), a state transition probability matrix (A) and an initial state probability distribution (?). Transition probability is the probability that the system will change its state from one turn to the next. Second a set of probability density functions (B) associated with each state.

Using shorthand notation a discrete hidden Markov model can be defined as ? = (N, A, B, ?). In practice the state sequence is unknown (hidden) and cannot be evaluated. However, the likelihood can be evaluated by summing over all the possible state sequences. The key attraction of HMM is that there is a simple procedure for finding the parameters? Called Baum-Welch re-estimation.

In order to use HMM for recognition, an observation sequence is obtained from the test signal and then the likelihood of each HMM generating this signal is computed. The HMM which has the highest likelihood then identifies the test signal. Finding the state sequence which maximizes the probability of an observation is done using the Viterbi algorithm, which is a simple dynamic programming optimization procedure. More details describing all the technical details concerning the algorithms used can be found in a great tutorial written by Rabiner [77].

One pioneer of using hidden Markov models for face recognition was Samaria at Trinity College starting in 1994 [78,79]. Nefian and Hayes at Georgia Institute of Technology have written several papers on pseudo 2D HMM [80-86] and some on embedded HMM [92]. Eickeler, Müller and Rigoll have written about how to get high performance using pseudo 2D HMM [87,88,89]. Some attempts have also been made by Othman and Aboulnasr on 2D HMM [90,91].

HMM based methods- Evaluation

HMM-based methods have shown better performances compared to the traditional eigenfaces method. Error rates of about 5% were reported when pseudo 2D HMM was used compared to about 10% with eigenfaces on the same dataset [78]. The 1D HMM had an error rate of 13% in the same experiment. The original pseudo 2D HMM uses pixel intensities as input feature vectors. Pixels however do not represent robust features, being very sensitive to image noise as well as image rotation, shift and changes in illumination. Using them is also computational expensive both for training and recognition because of the large dimensions on the feature vectors. This can be critical for a face recognition system that operates on a large database or in real-time systems.

Investigations have been made towards using feature vectors containing coefficients from low frequencies using 2D Discrete Cosine Transform (DCT) [80] applied to each observation block. Significant improvements have been attained using DCT coefficients instead of pixel values. One useful property is that it allows recognition in the JPEG and MPEG domain because these standards use image compression based on DCT. The extraction of DCT coefficients are simply recovered with entropy decoding. This is why pseudo 2D HMM recently was suggested for MPEG-7 v2 standard [83]. Another approach is using Karhunen Loeve Transform (KLT) [74], which also has the necessary feature compression properties.

The Baum-Welch algorithm, which is used for the training of the HMM for each person, provides the HMM parameters corresponding to a local maximum of the likelihood function depending on the initial model parameters. It is therefore very important to use a good initial model for the training. Also as much training data as possible is needed in the estimation of hidden Markov model parameters, to estimate good models for recognition.

Block overlap helps in providing higher statistical resolution. However large overlap results in increasing the computational load and memory requirements for all parts of the system. Varying overlap and block size can improve recognition performance.

In order to make the system more tolerant to orientation changes, individual models will have to be trained for views of the same subject at different orientations to the camera. Test images will be five models corresponding to different face views are needed for a good face representation under a large range of orientations [52].

The time required by the recognition system is critical. It is a function of the size of the database. Recognition time must be less than the time between two consecutive occurrences of people in a scene. Depending on the parameterization used the Viterbi algorithm can require a large number of calculations. This implies that sometimes the algorithm runs slowly.

This means that new faces can be added to the database without recomputing the representations of all other learned faces.

Table: IV

Advantages	**Drawbacks**
Robust against scaling, orientation and translation when face is correctly normalized	Sensitive to faulty normalization
Robust against illumination if training data has different lighting conditions	Sensitive to occlusion
Robust against facial expressions, glasses, facial hair, makeup etc.	Sensitive to perspective, viewing angle and head rotation (can be improved training models for different views)
Easy to update	Slow training and recognition/High computational cost (can be improved using DCT or KLT feature vectors)

Neural Network based Approaches

Artificial Neural Network (ANN) [B.Yegnanarayana (1999), S.Simon Haykin(1999), C.M.Bishop(1995), R.J.Mammone(1993)] is a powerful tool for pattern recognition problems. The use of neural networks (NN) in faces has addressed several problems: gender classification, face recognition and classification of facial expressions. One of the earliest demonstrations of NN for face recall applications is reported in Kohonen's associative map [16,100]. Using a small set of face images, accurate recall was reported even when input image is very noisy or when portions of the images are missing. A few NN based face recognition techniques are discussed in the following.

Single Layer adaptive NN:

A single layer adaptive NN (one for each person) for face recognition, expression analysis and face verification was reported in [33]. A system named Wilke, Aleksander and Stonham's recognition devise (WISARD) was devised. It needs typically 200-400 presentations for training each classifier where the training patterns included translation and variation in facial expressions. One classifier was constructed corresponding to one subject in the database. Classification was achieved by determining the classifier that was giving the highest response for the given input image.

Multilayer Perceptron (MLP):

Much of the present literature on face recognition with neural networks present results with only a small number of classes (often below 20). In [11] the first 50 principal components of the images were extracted and reduced to five dimensions using autoassociative neural network. The resulting representation was classified using a standard multilayer perceptron (MLP).

Self-Organizing Map (SOM):

In [1997] Lawrence et al. presented a hybrid neural network solution which combines local image sampling, a self-organizing map (SOM) and a convolutional neural network. The SOM provides a quantization of the image samples into a topological space are also nearby in the output space, thereby providing dimensionality reduction and invariance to minor changes in the image sample. The convolutional neural network provides partial invariance to translation, rotation, scale and deformation. The recognizer provides a measure of confidence in its output. The classification error approaches zero when rejecting as few as 10% of the examples on a database of 400 images which contains a high degree of variability in expression, pose and facial details.

Hopfield Memory Model:

In [10], a Hopfield memory model for the facial images is organized and the optimal procedure of learning is determined. A method for face recognition using Hopfield memory model combined with the pattern matching is proposed. It shows better performance of database having faces of 40 subjects.

Others:

A hierarchical neural network which is grown automatically and not trained with gradient descent was used for face recognition by Weng [39]. They reported good results for discrimination of ten subjects. The ability of the compression networks was demonstrated by Cottrell and Fleming in [9,11,37]. In [37] linear autoassociative networks, nonlinear auto associative (or compression) and/or hetero-associative backpropagation networks are explored for face processing. In [23] Lin et al. proposed a face recognition technique based on Probabilistic Decision based Neural network (PDBNN). It adopts a hierarchical network structures with nonlinear basis functions and competitive credit assignment scheme. It demonstrated a successful application of PDBNN on FERET and ORL databases.
Recently, [13] described the application of mixtures of experts on gender and ethnic classification of human faces and pose classification and showed their feasibility on the FERET database. The mixture consists of ensembles of radial basis functions (RBFs). Inductive Decision Trees (IDTs) and SVMs implement the"gating network" components for deciding which of the experts should be used to determine the classification output and to restrict the support of the input space. Experimental results yield good results on gender, ethnic and pose classification, which can be effectively used in face recognition.
Recognition of visual objects is performed effortlessly in our everyday life by humans. A previously seen face is easily recognized regardless of various transformations like change in size and position. It is known that humans process a natural image in under 150 ms [85]. The brain thus performs these tasks at very high speed. Neural networks are attempts to create face recognition systems that are based on the way humans detect and recognize faces.

Multi-Layered Feed-Forward Networks-

The multi-layer perceptron (MLP) neural network is a good tool for classification purposes. It can approximate almost any regularity between its input and its output. The weights are adjusted by supervised training procedure called back-propagation (BP). Back-propagation is a kind of gradient descent method, which searches for an acceptable local minimum in order to achieve minimal error. Error is defined as the root mean square of differences between real and desired outputs from the neural network.
A typical architecture for a feed-forward network has a number of layers following each other one by one (Figure 24).

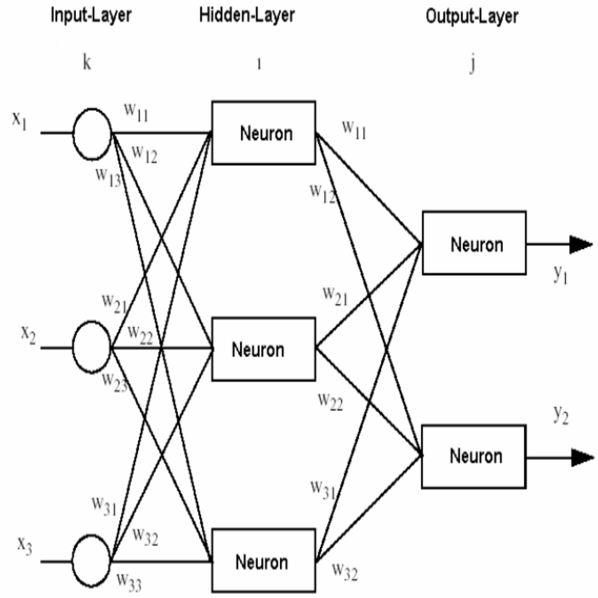

Figure 4: Feed-forward neural network

An input layer (k) consisting of input nodes and an output layer (j) consisting of output nodes. The input node a connected to the output nodes via one or more hidden layers (i) (multilayered). The nodes in the network are connected together, and each of the links has a weight associated with itself. The output value from a node is a weighted sum of all the input values to the node. By changing the different weights of the input values we can adjust the influence from different input nodes. For face recognition the input nodes will typically correspond to image pixel values from the test image to be recognized. The output layer will correspond to classes or individuals in the database. Each unit in the output layer can be trained to respond with +1 for a matching class and -1 for all others. In practice real outputs are not exactly +1 or -1, but vary in the range between these values. The closer the values of the neural network get towards the ideal, the more confidence there is towards the decision being right. Recognition is done by finding the output neuron with the maximal value. Then a threshold algorithm can be applied to reject or confirm the decision.
Experiments have also been made with ensambles of networks where each class in the database has its own neural network [93,94]. The output layer is then trained to give +1 for own person and -1 for other persons. An aggregate output consisting of outputs from all the MLP networks are then considered in the same manner as when having only one MLP and threshold rules can be applied as normal. Huang, Zhou, Zhang and Chen describe a method of pose invariant face recognition using ensambles of networks [13].

They show that the accuracy of ensambles of networks can be higher than single neural networks.
Often even a simple network can be very complex and difficult to train [96]. A typical image recognition network requires as many input nodes as there are pixels in the image. Cottrell and Flemming used two MLP networks working together [97]. The first one operates in an auto-association mode and extracts features for the second network, which operates in the more common classification mode. In this way the hidden layer output constitutes a compressed version of the input image and can be used as input to the classification network. Cottrell and Flemming also showed that a neural network using this design was not any better than an eigenface approach.
Maybe one of the more successful face recognition with neural networks is a result of the recent work of Lawrence, Giles, Tsoi and Back at NEC Research Institute. It combines local image sampling, a self organizing map (SOM) neural network and a convolutional neural network [98,99]. SOM was introduced by Kohonen [100] and is an unsupervised learning process which learns the distribution of a set of patterns without having any class information. A pattern is projected from an input space to a position in the map and information is thereby coded as the location of an activated node. Unlike most other classification or clustering techniques SOM preserves the topological ordering of classes. This feature makes it useful in classification of data which includes a large number of classes. Experiments were also made concerning using KLT instead of SOM for dimensionality reduction. A convolutional neural network was trained and compared to a standard MLP network.
A major disadvantage is that SOM as well as the convolutional network needs a considerable time to be trained.

Radial Basis Function (RBF) Networks-

RBF neural networks have recently attracted extensive research interests in the community of neural networks. Their learning speed is fast because of local-tuned neurons and they have a more compact topology than other neural networks. The RBF network is a two-layer feed-forward network, with a supervised layer from the hidden to the output nodes, and an unsupervised layer from the input to the hidden. Gaussian functions for each of the hidden units simulate the effect of overlapping and locally tuned receptive fields.
Howell and Buxton at University of Sussex have written several articles about using RBF networks for face recognition tasks [115,102,103,104,105].They experimented with using either difference of Gaussion (DoG) or Gabor wavelets as input to the network. Using Gabor wavelets as input gave the best recognition results allowing different scales and orientations to be tailored to the task at hand.
Some approaches have been made towards reducing the input size to the RBF network. Er, Wu and Lu at Nanyang Technological University [106,107] and Feitosa at University of Rio de Janeiro [Raul Queiroz Feitosa (1999)] have proposed using PCA and LDA eigenvectors as input to the RBF network to reduce dimensionality. Huang, Law and Cheung at Zhong Shan University have written an article about using ICA together with RBF networks [95]. Results show that these approaches converge faster than the conventional RBF during training, and also outperform its generalization abilities.
Gutta and Wechsler have demonstrated the capability of RBF networks to handle large databases, like FERET [110].

Dynamic Link Architecture (DLA) –

In dynamic link matching the image and all the models are represented by layers of neurons labelled by jets as local features. Jets are vectors of Gabor wavelet components. In each layer neural activity dynamics generates one small moving blob of activity. If a model is similar in feature distribution to an image, its initial connectivity matrix will connect corresponding points having high feature similiarity.
Blobs in the image and the model tend to align and synchronize by simultaneously activating and generating correlations between corresponding regions. These correlations are used to restructure and improve the connectivity matrix. This provides translational invariance as well as robustness against distortions. The main concerns with DLA is processing time and its inabilities to handle large size and orientation changes. Wiskott and Malsburg have written a good article which describes DLA and its algorithms in detail [111].

NN Based methods- Evaluation:

Neural networks have been used in many recognition tasks and have achieved high recognition rates for limited datasets. The representation of the given input to the network and the training phase is crucial for the results of the face recognition. The representation of the given input, the hidden layer network, the coupling between the network components and the transfer function are vital elements deciding the functionality and the performance of the neural network face recognition system. Achieved recognition results are dependent on the database size and the number of pictures per person. The training time is growing with the number of pictures in the training database, but once the training is done, the recognition task is performed relatively fast. The recognition process only depends on the neural network structure and not on the number of trained faces.
Much of the present literature on face recognition with neural networks presents results with only a small number of classes. Good results are reported, but the database is often quite simple, the pictures are manually aligned and there is no lighting variation, rotation or tilting. Hjelmås and Wroldsen describe a face recognition system using PCA for dimensionality reduction and feature extraction, and using a MLP neural network for classification [77]. They report of a correct classification of about 90% when using a test set containing 200 face image.

Table: V

Advantages	**Drawbacks**
Stores neighbourhood relationships	Sensitive to faulty normalization
Robust against noise and occlusion	Sensitive to illumination and face expressions
Robust against scaling, orientation and translation when face is correctly normalized	Sensitive to perspective, viewing angle and head rotation (can be improved using ensambles of networks)

Fast recognition/Low computational cost (depending only on the network and not the number of images)	Can be slow and difficult to train (especially for large databases)

Hybrid Approaches

These types of approaches use both statistical pattern recognition techniques and neural networks.

PCA and RBF:

The method by Er et al. [1999] suggests the use of RBF on the data extracted by discriminant eigenfeatures. They used a hybrid learning algorithm to decrease the dimension of the search space in the gradient method, which is crucial on optimization of high dimension problem. First, they tried to extract the face features by both PCA and LDA methods. Next, they presented a hybrid learning algorithm to train the RBF Neural Networks, so the dimension of the search space is significantly decreased in the gradient method.
[35] also studied on combining PCA and RBF neural network. Their system for face recognition consists of a PCA stage which inputs the projections of a face image over the principal components into a RBF network acting as a classifier.

Other Approaches

Range Data:

One of the different methods used in face recognition task is using the range images. In this method data is obtained by scanning the individual with a laser scanner system. This system also has the depth information so the system processes 3-dimensional data to classify face images [8].

Infrared Scanning:

Another method used for face recognition is scanning the face image by an infrared light source. Yoshitomi [4] used thermal sensors to detect temperature distribution of a face. In this method, the frontview face in input image is normalized in terms of location and size, followed by measuring the temperature distribution, the locally averaged temperature and the shape factors of face. The measured temperature distribution and the locally averaged temperature are separately used as input data to feed a neural network and supervised classification is used to identify the face. The disadvantage of visible ray image analysis is that the performance is strongly influenced by lighting condition including variation of shadow, reflection and darkness. These can be overcome by the method using infrared rays.

Profile Images:

[4] worked on profile images instead of frontal images. Their method is based on the representation of the original and morphological derived profile images. Their aim was to use the profile outline that bounds the face and the hair. They take a gray-level profile image and threshold it to produce a binary image representing the face region. They normalize the area and orientation of this shape using dilation and erosion. Then, they simulate hair growth and haircut and produce two new profile silhouettes. From these three profile shapes they obtain the feature vectors. After nor-malizing the vector components, they use the Euclidean distance measure for measuring the similarity of the feature vectors derived from different profiles.

Geometry Feature based Methods

Geometry feature based methods uses the facial feature measures such as distance between eyes, ratio of distance between eyes and nose etc., but it is significantly different from the feature-based techniques that it constructs the topological graph using the facial features of each subject.
The earliest approaches to face recognition were focused on detecting individual features such as eyes, ears, head outline and mouth, and measuring different properties such as size, distance and angles between features. This data was used to build models of faces and made it possible to distinguish between different identities. This kind of system was proposed [5] and was one of the first approaches to automated face recognition. Later work by Yuille, Cohen and Hallinan in 1989 describes a method for feature extraction using deformable templates.

Wavelets:

Wavelets represent an approach to decomposing complex signals into sums of basis functions. In this respect they are similar to Fourier decomposition approaches, but they have an important difference. Fourier functions are localized in frequency but not in space, in the sense that they isolate frequencies, but not isolated occurrences of those frequencies. This means that small changes in a Fourier transform will produce changes everywhere in time domain. Wavelets are local in both time by translations and frequency by dilations. Because of this they are able to analyze data at differentscales or resolutions much better than simple sine and cosines can. To understand this note that modelling a spike in a function, a noise dot for example, with a sum of infinite functions will be hard because of its strict locality, while functions that are already local will be naturally suited to the task. Sharp spikes and discontinuities normally take fewer wavelet bases to represent than if sine-cosine basis functions are used.

Gabor Wavelets:

Physiological studies have found simple cells in human visual cortex which are selectively tuned to orientation as well as to spatial frequency. The response of these simple cells can be approximated by 2D Gabor filters [112]. Gabor functions were first proposed by Dennis Gabor as a tool for 1D signal detection in noise [113]. Rediscovered and generalized to 2D Gabor wavelet representation for computervision was pioneered by Daugman in 1980 [114]. [63,115] have developed a face recognition system based on this representation [63]. This work has continued with elastic bunch graph matching of coefficients from Gabor filter responses[116] and the dynamic link architecture [117]. Gabor filters are now being used extensively in various computer vision applications.

Elastic Bunch Graph Matching

Different human faces have the same geometrical structure and can therefore be defined as labelled graphs. Since we want to recognize faces from different views, the nodes of the graphs consistently refer to particular fiducial points, such as eyes, mouth, the tip of the nose and other contour points. Graphs for different head pose differ in geometry and local features. Although the fiducial points refer to corresponding object locations, some may be occluded, and jets as well as distances vary due to rotation in depth. To be able to compare graphs from different poses pointers have to be established to associate corresponding nodes in the different graphs.

Kepenekci has recently proposed a method of selecting high-energy peaks of the Gabor wavelet response instead of using predefined graph nodes as in elastic graph matching [119]. This reduces computational complexity and also improves the performance in the presence of occlusions. Hjelmås reports of 85% recognition on the ORL database [120,121]

Gabor Fisher Classifier (GFC):

Liu at University of Missouri and Wechsler at Goerge Mason University have applied an enhanced Fisher Discrimination Model (EFM) to the Gabor feature vector [122,123]. The dimensionality of the vector space is reduced in order to derive a lowdimensional feature representation with enhanced discrimination power. The GFC method is robust to illumination and facial expression variability and they report about excellent performance on the FERET database compared against other methods using Gabor wavelets, eigenfaces, fisherfaces, and a combination of Gabor and eigenfaces.

Wavelets Based Methods Evaluation-

Gabor wavelets are chosen for their robustness as a data format and for their biological relevance. One of the main motivations for using such feature based methods is that representation of face images in this way becomes very compact and this lowers the computational cost. This fact especially gains importance when there is a huge database.

Since Gabor responses are DC-free they provide robustness against varying brightness in the image. Robustness against varying contrast can be obtained by normalizing the jets. The limited localization in space and frequency yields certain amount of robustness against translation, distortion, rotation and scaling. Face Bunch Graphs represent a good data structure for storing the extracted features. Simple graphs consisting of only nine nodes and six jets can theoretically represent 69 or about as many as ten million different faces.

Finding the locations and corresponding values of the fiducial points in a face image is extremely critical for the performance of the recognition system. However, some of the most successful face recognition methods are based on graph matching of Gabor filter responses. Disadvantages are the graph matching complexity manual location of training graphs and overall execution time.

Table: VI

Advantages	**Drawbacks**
Saves neighborhood relationships between pixels	Sensitive to faulty normalization
Robust against illumination, scaling, orientation and translation when face is correctly normalized	Sensitive to facial expressions, glasses, facial hair, makeup etc. (can be improved using elastic bunch graph matching)
Robust against noise	Sensitive to occlusion (can be improved using high energy feature points as graph nodes)
Robust against translation, rotation and scaling	Sensitive to perspective, viewing angle and head rotation (can be improved using elastic bunch graph matching)
Easy to update	Graph matching complexity
Fast recognition/Low computational cost	Slow training/High computational cost (with large databases)

Graph Matching based Methods:

In [18] presented dynamic link architecture for distortion invariant object recognition which employs elastic graph matching to find the closed stored graph. Objects were represented with sparse graphs whose vertices were labeled with geometrical distances. In this system, individual faces were represented by a rectangular graph, each node labeled with a set of complex Gabor wavelet coefficients, called a jet. Only the magnitudes of the coefficients were used for matching and recognition. When recognizing a face of a new image, each graph in the model gallery was matched to the image separately and the best match indicated the recognized person. They presented good results with a database of 87 subjects and test images composed of different expressions and faces turned 15 degree. The matching process was computationally expensive, taking roughly 25 seconds to compare an image with 87 stored objects when using a parallel machine with 23 transputers.

[40] extended this system to handle larger galleries and larger variations in pose and to increase the matching accuracy. Firstly, they use the phase of the complex Gabor wavelet coefficients to achieve an accurate location of the nodes and to disambiguate patterns which would be similar in the magnitudes of the coefficient. Secondly, they employ object adapted graphs, so that nodes refer to specific facial landmarks, called fiducially points. The correspondences between two faces can be found across large viewpoint changes. Thirdly, a new data structure called the bunch graph was introduced which serves as generalized representation of faces by combining jets of a small set of individual faces. This allows the system to find the fiducial points in one matching process, which eliminates the need for matching each model graph individually. This also reduces computational effort significantly. It offers good performance of about 98% for FERET database. But the drawback in this feature matching approach is that it requires manual intervention to select the fiducially points in the facial image and it requires precise location of those points.

Brunelli and Poggio developed two simple algorithms for face recognition [4]. The first one is based on the computation of a set of geometrical features, such as nose width and length, mouth position and chin shape. One motivation for using geometric methods is that in an image with sufficiently low resolution it is impossible to distinguish the fine details of a face, but often possible for a human to recognize the person. The remaining information in the low resolution image is almost pure geometrical and implies that these properties of face features are sufficient enough for face recognition. The configuration of the features can be described by a vector of numerical data representing the position and size of the main facial features, eyes and eyebrows, nose and mouth. This information can be supplemented by the shape of the face outline.

One of the most critical issues in using a vector of geometrical features is proper normalization. The extracted features have to be independent of position, scale and rotation of the face in the image plane. Translation dependency can be eliminated once the origin of coordinates is set to a point that can be detected with good accuracy in each image. Rotation invariance can be achieved by horizontally aligning the eye to eye axis and scale invariance by using the distance between the two eyes. Locating the eyes is usually performed using templates for each of the eyes.

Because almost every face has two eyes, one nose and one mouth with very similar layout face classification can be difficult, while feature extraction is easier. A very useful technique for extraction of facial features is vertical and horizontal integral projection (Figure 1). Projections can be extremely effective in determining the position of features, provided that the window on which they act is suitably located to avoid misleading interferences.

Feature based PCA*:*

Cagnoni and Poggi [6] suggested a feature based approach instead of a holistic approach to face recognition. They applied the eigenface method to sub-images (eye, nose and mouth). They also applied a rotation correction to the faces in order to obtain better results.

Template Based and Geometry Based Methods -Evaluation

The use of feature vectors seems very unstable and limited because the variation of the data from different pictures of the same face was in the same order of magnitude as the variation between different faces. The method is sensitive to inaccurate detection of features and to all sorts of disturbance such as facial expressions or varying pose.

Template-based approaches outperform geometrical methods. Templates seem to offer satisfactory results for recognition from frontal views. A more difficult problem is how to deal with non-frontal views. It should be possible to use almost the same scheme for different viewpoints at the expense of considerably greater computational complexity. Or maybe it is possible to extrapolate or guess correctly other views of the face. Humans are certainly able to recognize faces turned 20-30 degrees from the front from just one frontal view.

The recognition rate achieved with a single template (eyes, nose or mouth) is remarkable and consistent with the human ability of recognizing familiar people from a single facial characteristic. Using a eyes, nose or mouth template is most discriminating and using the whole face gives least discrimination. Integration of more features in a recognition system has a beneficial effect on robust classification. If more templates are used in parallel the score from the most similar feature can be used, scores can be added together or each feature template can be assigned a different weight.

Table: VII

Advantages	**Disadvantages**
Robust against scaling, orientation and translation when face is correctly normalized	Sensitive to faulty normalization
Can handle high resolution images efficiently	Sensitive to noise and occlusion
Saves neighborhood relationships between pixels	Templates are sensitive to illumination
Can handle very low resolution image	Sensitive to perspective, viewing angle and head rotation (can be improved using more templates)
Geometric relations are stable under varying illumination conditions	Sensitive to facial expressions, glasses, facial hair, makeup etc.
Good recognition performance	Slow training and recognition/High computational complexity

CONCLUSION

The drawbacks are compensated with regular PCA. One major advantage with this hybrid combination is that it provides methods for incremental learning of new classes. This means that new faces can be added to the database without re-computing the representations of all other learned faces.

REFERENCES

[1] Robert J. Baron. Mechanisms of Human Facial Recognition. International Journal of Man-Machine Studies, 15(2):137-178, 1981.

[2] Peter N. Belhumeur, Joao P. Hespanha, and David J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Tran. on Pattern Analysis and Machine Intelligence, 19(7):711-720, Jul. 1997.

[3] C. M. Bishop. Neural Networks for Pattern Recognition. Oxford University Press Inc., New York, 1995.

[4] R. Brunelli and T. Poggio. Face Recognition: Features versus Templates. IEEE Tran. on Pattern Analysis and Machine Intelligence, 15(10):1042-1052, October 1993.

[5] B.Yegnanarayana. Arti_cial Neural Networks. Prentice-Hall of India, New Delhi, 1999.

[6] S. Cagnoni and A. Poggi. A Modified Modular Eigenspace Approach to Face Recognition. In 10th International Conference on Image Analysis and Processing (ICIAP99), pages 490-495, Venice, Italy, September 1999.

[7] Hakan Cevikalp, Marian Neamtu, Mitch Wilkes, and Atalay Barkana. Discriminative Common Vectors for Face Recognition. IEEE Tran. on Pattern

Analysis and Machine Intelligence, 27(1):4-13, Jan. 2005.

[8] R. Chellappa, C. L. Wilson, and S. Sirohey. Human and Machine Recognition of Faces: A Survey. Proc. of the IEEE, 83(5):705-740, May 1995.

[9] G. W. Cottrell and M. K. Fleming. Face Recognition using Unsupervised Feature Extraction. In Int. J. Conf. on Neural Networks, pages 322-325, Paris, 1990.

10] Y. Dai and Y. Nakano. Recognition of Facial Images with Low Resolution using a Hopfield Memory Model. Pattern Recognition, 31(2):159-167, 1998.

[11] D. Demers and G. W. Cottrell. Nonlinear Dimensionality Reduction. In Advances in Neural Information Processing Systems, pages 580-587, 1993.

[12] M. J. Er, S. Wu, and J. Lu. Face Recognition using Radial Basis Function (RBF) Neural Networks. In 38th Conference on Decision & Control, pages 2162{2167, Phoenix, Arizona USA, 1999.

[13] S. Gutta, J. R. J. Huang, P. Jonathan, and H. Wechsler. Mixture of Experts for Classification of Gender, Ethnic Origin, and Pose of Human Faces. IEEE Trans. on Neural Networks, 11:948-959, July 2000.

[14] Jonathan Howell and Hilary Buxton. Invariance in Radial Basis Function Neural Networks in Human Face Classification. Neural Processing Letters, 2(3):26-30, 1995.

[15] J. Huang. Detection Strategies for Face Recognition Using Learning and Evolution. PhD thesis, George Mason University, May 1998.

[16] T. Kohonen. Self-Organization and Associative Memory. Springer-Verlag, New York, 1988.

[17] Hui Kong, Eam Khwang Teoh, Jian Gang Wang, and R Venkateswarlu. Two Dimensional Fisher Discriminant Analysis: Forget About Small Sample Size Problem. In International Conference on Acoustics Speech and Signal Processing, volume 2, pages 761-764, March 2005.

[18] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. Von ser Malburg, R. P. Wurtz, and W. Konen. Distortion Invariant Object Recognition in the Dynamic Link Architecture. IEEE Transactions on Computers, 42:300-310, March 1997.

[19] Andreas Lanitis, Christopher J. Taylor, and Timothy Francis Cootes. Automatic Interpretation and Coding of Face Images using Flexible Models. IEEE Tran. on Pattern Analysis and Machine Intelligence, 19(7):743-756, 1997.

[20] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back. Face Recognition: A Convolutional Neural Network Approach. IEEE Trans. on Neural Networks, 8(1):98{112, January 1997.

[21] Steve Lawrence, C. Lee Giles, Ah Chung Tsoi, and Andrew D. Back. Face Recognition: A Convectional Neural Network Approach. IEEE Trans. on Neural Networks, 8(1):98-113, 1998.

[22] Ming Li and Baozong Yuan. 2D-LDA: A Statistical Linear Discriminant Analysis for Image Matrix. Pattern Recognition Letters, 26:527-532, 2005.

[23] Shang-Hung Lin, Sun-Yuan Kung, and Long ji Lin. Face Recognition/Detection by Probabilistic Decision-based Neural Network. IEEE Trans. on Neural Networks, 8(1):114-131, January 1997.

[24] Z. Liposcak and S. Loncaric. Face Recognition from Profiles using Morphological Operations. In International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, pages 47{52, 1999.

[25] K. Liu, Y.Q. Cheng, and J.Y. Yang. Algebraic Feature Extraction for Image Recognition Based on an Optimal Discriminant Criterion. Pattern Recognition, 26(6):903-911, June 1993.

[26] R. J. Mammone. Arti_cial Neural Networks for Speech and Vision. Chapman and Hall, Cambridge, 1993.

[27] E. Osuna, R. Freund, and F. Girosi. Training Support Vector Machines: An Application to Face Detection. In IEEE Conference on Computer Vision and Pattern Recognition, pages 193-199, 1997.

[28] P. Penev and J. Atick. Local Feature Analysis: A Genera Statistical Theory for Object Representation. Network: Computation in Neural Systems, 7:477-500, 1996.

[29] T. Poggio and K. K. Sung. Example-based Learning for View-based Human Face Detection. ARPA Image Understanding Workshop, 2, November 1994.

[30] Bernhard Scholkopf, Alex J. Smola, and Andre Bernhardt. Nonlinear Component Analysis as a Kernel Eigenvalue Problem. Neural Computation, 10(5):1299-1319, 1998.

[31] Simon Haykin. Neural networks: A Comprehensive Foundation. Prentice-Hall International, New Jersey, 1999.

[32] L. Sirovich and M. Kirby. Low-dimensional Procedure for the Characterization of Human Faces. Journal of Optical Society of America, 4(3):519-524, March 1987.

[33] T. J. Stonham. Practical Face Recognition and Verification with WISARD. In Aspects of Face Processing, pages 426{441, 1984.

[34] Daniel L. Swets and John (Juyang) Weng. Using Discriminant Eigenfeatures for Image Retrieval. IEEE Tran. on Pattern Analysis and Machine Intelligence, 18(8):831-836, Aug. 1996.

[35] E. Thomaz, R. Q. Feitosa, and A. Veiga. Design of Radial Basis Function Network as Classi_er in face Recognition using Eigenfaces. In Vth Brazilian Symposium on Neural Networks, pages 118-123, 1998.

[36] Matthew Turk and Alex Paul Pentland. Eigenfaces for Recognition. Journal of Cognitive Neuroscience, 3(1):71-86, 1991.

[37] Valentin, H. Abdi, A. J. O'Toole, and G. W. Cottrell. Connectionist Models of Face Processing: A survey. Pattern Recognition, 27(9):1209-1230, 1994.

[38] Vladimir N. Vapnik. The Nature of Statistical Learning Theory. Springer Verlog, Heidelberg, DE, 1995.

[39] J. Weng, N. Ahuja, and T. S. Huang. Learning Recognition Segmentation of 3-D Objects from 2-D

Images. In Int. Workshop Face Gesture Recognition, Zurich, Switzerland, 1995.
[40] Laurenz Wiskott, Jean-Marc Fellous, Norbert Kruger, and Christoph von der Malsburg. Face Recognition by Elastic Bunch Graph Matching. IEEE Tran.on Pattern Analysis and Machine Intelligence, 19(7):775-779, July 1997.
[41] Huilin Xiong, M. N. S. Swamy, and M. O. Ahmad. Two-Dimensional FLD for Face Recognition. Pattern Recognition, 38:1121-1124, 2005.
[42] M. H. Yang. Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition using Kernel Methods. In IEEE International Conference on Face and Gesture Recognition, pages 215{220, Washington, May 2002.
[43] Y. Yoshitomi, T. Miyaura, S. Tomito, and S. Kimura. Face Identification using Thermal Image Processing. In IEEE International Workshop on Robot and Human Communication, pages 374-379, 1997.
[44] Alan L. Yuille. Deformable Templates for Face Recognition. Journal of Cognitive Neuroscience, 3(1):59-70, 1991.
[45] Jian Yang Zhang, Alejandro F. Frangi, and Jingyu Yang. Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition. IEEE Tran. on Pattern Analysis and Machine Intelligence, 26(1):131-137,Jan. 2004.
[46] W. Zhao. Discriminant Component Analysis for Face Recognition. In Proc. Int. Conf. on Pattern Recognition, volume 2, pages 818-821, Sep. 2000.
[47] ”Low-Dimensional Procedure for the Characterization of Human Faces”, L.Sirovich and M. Kirby (1987)
[48] “Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces”, M. Kirby and L. Sirovich (1990)
[49] “Eigenfaces for Recognition”, Matthew Turk, Alex Pentland (1991a)
[50] “Face Recognition Using Eigenfaces”, Matthew A. Turk and Alex P. Pentland (1991b)
[51] “View-Based and Modular Eigenspaces for Face Recognition”, Alex Pentland,Baback Moghaddam and Thad Starner (1994)
[52] “A Random Walk Trough Eigenspace”, Matthew Turk (2001)
[53] “Subspace Methods for Robot Vision”, Shree K. Nayar, Sameer A. Nene and Hiroshi Murase (1995)
[54] “Visual Learning and Recognition of 3D Objects From Appearance”, H.Murase and S. Nayar (1995)
[55] “Colour Eigenfaces”, Graham D. Finlayson, Janet Dueck, Brian V. Funt, MarkS. Drew (1996)
[56] “Experiments on Eigenfaces Robustness”, Alexandre Lemieux and Marc Parizeau (2002)
[57] “Analysis of PCA-Based and Fisher Discriminant-Based Image Recognition Algorithms”, Wendy S. Yambor (2000)
[58] “Analyzing PCA-Based Face Recognition Algorithms: Eigenvector Selection and Distance Measures”, Wendy S. Yambor, Bruce A. Draper and J. Ross Beveridge (2000)
[59] “A Nonparametric Statistical Comparison of Principal Component and Linear Discriminant Subspaces for Face Recognition”, J. Ross Beveridge, Kai She and Geof H. Givens (2001)
[60] “The Global Dimensionality of Face Space”, P. S. Penev and L. Sirovich (2000)
[61] “Numerical Analysis of Eigenfaces”, Neil Muller and B. M. Herbst (2000)
[62] “Learning in Eigenspace: Theory and Application”, B.S. Manjunath, S.Chandrasekaran, Y.F. Wang (1994)
[63] “An Eigenspace Update Algorithm for Image Analysis”, S. Chandrasekaran,B.S. Manjunath, Y.F. Wang, J. Winkeler and H. Zhang (1996)
[64] “The Use of Multiple Measurements in Taxonomic Problems”, R. A. Fisher (1936)
[65] “Using Discriminant Eigenfeatures for Image Retrieval”, D. L. Swets and J. Weng (1996)
[66] “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection”, Peter N. Belhumeur, Joao P. Hespana and David J. Kriegman (1996)
[67] “Discriminant Analysis of Principal Components for Face Recognition”, W. Zhao, A. Krishnaswamy, R. Chellappa, D. L. Swets and J. Weng (1998)
[68] “Discriminant Analysis for Recognition of Human Face Images”, K. Etemad and R. Chellappa (1997)
[69] “Face Recognition Algorithms Review”, Sunny Tang Ho Man (2001)
[70] “Analysis and Comparison of Eigenspace-Based Face Recognition Approaches”, Pablo Navarrete and Javier Ruiz-del-Solar (2002)
[71] “Face Recognition Using Independent Component Analysis and Support Vector Machines”, O. Déniz, M. Castrillón and M. Hernández (2001)
[72] “Independent Component Analysis: A Tutorial”, Aapo Hyvärinen and Erkii Oja (1999)
[73] “Viewpoint Invariant Face Recognition Using Independent Component Analysis and Attractor Networks”, Marian Stewart Bartlett and Terrence J. Sejnowski (1997)
[74] “Independent Components of Face Images: A Representation for Face Recognition”, Marian Stewart Bartlett and Terrence J. Sejnowski (1997)
[75] “Independent Component Representations for Face Recognition”, Marian Stewart Bartlett, H. Martin Lades and Terrence J. Sejnowski (1998)
[76] “PCA vs. ISA: A Comparison on the FERET Data Set”, Kyungim Baek, Bruce A. Draper, J. Ross Beveridge and Kai She (2001)
[77] “A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, Lawrence R. Rabiner (1989)
[78] “Parameterization of a Stochastic Model for Human Face Identification”, F. S. Samaria and A. C. Harter (1994)
[79] “Face Recognition Using Hidden Markov Models, Ferdinando Silvestro Samaria (1994)
[80] “Statistical Approaches to Face Recognition”, Ara V. Nefian (1996)
[81] “Hidden Markov Models for Face Recognition”, Ara V. Nefian and Monson H. Hayes III (1996)
[82] “Face Detection and Recognition Using Hidden Markov Models”, Ara V. Nefian and Monson H. Hayes III (1998)

[83] "A Hidden Markov Model-Based Approach for Face Detection and Recognition", Ara V. Nefian (1998)

[84] "Face Recognition Using an Embedded HMM", Ara V. Nefian and Monson H. Hayes III (1998)

[85] "An Embedded HMM-Based Approach for Face Detection and Recognition",Ara V. Nefian and Monson H. Hayes III (1998)

[86] "Maximum Likelihood Training of the Embedded HMM for Face Detection and Recognition", Ara V. Nefian and Monson H. Hayes III (2000)

[87] "High Performance Face Recognition Using Pseudo 2-D Hidden Markov Models", Stefan Eickeler, Stefan Müller and Gerhard Rigoll (1998)

[88] "Improved Face Recognition Using Pseudo 2-D Hidden Markov Models",Stefan Eickeler, Stefan Müller and Gerhard Rigoll (1998)

[89] "Face Database Retrieval Using Pseudo 2D Hidden Markov Models", Stefan Eickeler (2002)

[90] "Low Complexity 2-D Hidden Markov Model for Face Recognition", H.Othman and T. Aboulnasr (2000)

[91] "A Simplified Second-Order HMM with Application to Face Recognition", H. Othman and T. Aboulnasr (2001)

[92] "Standard Support for Automatic Face Recognition", Ara Nefian and Bob Davies (2001)

[93] "Access Control by Face Recognition Using Neural Networks", Dmitry Bryliuk and Valery Starovoitov (2000)

[94] "Face Recognition Using Ensambles of Networks", S. Gutta, J. Huang, B. Takacs and H. Wechsler (1996)

[95] "Pose Invariant Face Recognition", Fu Jie Huang, Zhihua Zhou, Hong-Jiang Zhang and Tsuhan Chen (2000)

[96] "Multilayer Perceptrons versus Hidden Markov Models: Comparisons and Applications to Image Analysis and Visual Pattern Recognition", Lúcio F.C. Pessoa (1995)

[97] "Face Recognition Using Unsupervised Feature Extraction", G. W. Cottrell and M. Flemming (1990)

[98] "Face Recognition: A Hybrid Neural Network Approach", Steve Lawrence, C.Lee Giles, Ah Chung Tsoi and Andrew D. Baek (1996)

[99] "Face Recognition: A Convolutional Neural Network Approach", Steve Lawrence, C. Lee Giles, Ah Chung Tsoi and Andrew D. Baek (1997)

[100] "Self-Organizing Map", T. Kohonen (1990)

[101] "Receptive Field Functions for Face Recognition", A. Jonathan Howell and Hillary Buxton (1995)

[102] "Face Recognition Using Radial Basis Function Neural Networks", A. Jonathan Howell and Hillary Buxton (1996)

[103] "Towards Unconstrained Face Recognition from Image Sequences", A. Jonathan Howell and Hillary Buxton (1996)

[104] "Automatic Face Recognition Using Radial Basis Functions", Andrew Jonathan Howell (1997)

[105] "Face Unit Radial Basis Function Networks", A. Jonathan Howell (1999)

[106] "Face Recognition Using Radial Basis Function (RBF) Neural Networks", Meng Joo Er, Shiqian Wu and Juwei Lu (1999)

[107] "Face Recognition With Radial Basis Function (RBF) Neural Networks, Meng Joo Er, Shiqian Wu, Juwei Lu and Hock Lye Toh (2002)

[108] "Comparing the Performance of the Discriminant Analysis and RBF Neural Network for Face Recognition", Raul Queiroz Feitosa (1999)

[109] "An Experimental Study: On Reducing RBF Input Dimension by ICA and PCA", Rong-Bo Huang, Lap-Tak Law and Yiu-Ming Cheung (2002)

[110] "Face Recognition Using Hybrid Classifier Systems", Srinivas Gutta and Harry Wechsler (1999)

[111] "Face Recognition by Dynamic Link Matching", Laurenz Wiskott and Christoph von der Malsburg (1999)

[112] "Mathematical Description of the Responses of Simple Cortical Cells", S. Marcelja (1980)

[113] "Theory of Communication", Dennis Gabor (1946)

[114] "Two Dimensional Spectral Analysis of Cortical Receptive Field Profile", J.G. Daugman (1980)

[115] "A Feature Based Approach to Face Recognition", B.S. Manjunath, R. Chellappa and Christoph von der Malsburg (1992)

[116] "Face Recognition by Elastic Bunch Graph Matching", Laurenz Wiskott, Jean-Marc Fellous, Norbert Krüger and Christoph von der Malsburg (1996)

[117] "Distortion Invariant Object Recognition in the Dynamic Link Architecture", Martin Lades, Jan C. Vorbrüggen, Joachim Buhmann, Jörg Lange, Christoph von der Malsburg, Rolf P. Würtz and Wolfgang Konen (1993)

[118] "Face Recognition by Eigenface and Elastic Bunch Graph Matching", Jang Kim Fung (1999)

[119] "Face Recognition Using Gabor Wavelet Transform", Burcu Kepenekci (2001)

[120] "Biometric Systems: A Face Recognition Approach", Erik Hjelmås (2000)

[121] "Feature-Based Face Recognition", Erik Hjelmås (2000)

[122] "A Gabor Feature Classifier for Face Recognition", Chenjun Liu and Harry Wechsler (2001)

[123] "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discrimi

[124] "Face Recognition: Graph Matching Versus Neural Techniques", J. Vergés- Llahí, A. Sanfeliu, F. Serratosa, R. Alquezar (1999)

[125] Chein – I Chang, Wei-min Liu, "Variants of Principal Component Analysis" , Geoscience and Remote Sensing Symposium, 2007. IGARSS 2007. IEEE International, July 2007,1083-1086 ISBN: 978-1-4244-1211-2.

[126] Neeraja, Ektha Walia, " Face Recognition using Improved fast PCA Algorithm" , Image and Signal Processing, 2008. CISP '08. Congress on Volume 1, 27-30 May 2008 Page(s):554 – 558.

