

Image Analysis for Face Recognition

Xiaoguang Lu
Dept. of Computer Science & Engineering
Michigan State University, East Lansing, MI, 48824
Email: lvxiaogu@cse.msu.edu

Abstract

In recent years face recognition has received substantial attention from both research communities and the market, but still remained very challenging in real applications. A lot of face recognition algorithms, along with their modifications, have been developed during the past decades. A number of typical algorithms are presented, being categorized into appearance-based and model-based schemes. For appearance-based methods, three linear subspace analysis schemes are presented, and several non-linear manifold analysis approaches for face recognition are briefly described. The model-based approaches are introduced, including Elastic Bunch Graph matching, Active Appearance Model and 3D Morphable Model methods. A number of face databases available in the public domain and several published performance evaluation results are digested. Future research directions based on the current recognition results are pointed out.

1 Introduction

In recent years face recognition has received substantial attention from researchers in biometrics, pattern recognition, and computer vision communities [1][2][3][4]. The machine learning and computer graphics communities are also increasingly involved in face recognition. This common interest among researchers working in diverse fields is motivated by our remarkable ability to recognize people and the fact that human activity is a primary concern both in everyday life and in cyberspace. Besides, there are a large number of commercial, security, and forensic applications requiring the use of face recognition technologies. These applications include automated crowd surveillance, access control, mugshot identification (e.g., for issuing driver licenses), face reconstruction, design of human computer interface (HCI), multimedia communication (e.g., generation of synthetic faces),

and content-based image database management. A number of commercial face recognition systems have been deployed, such as Cognitec [5], Eyematic [6], Viisage [7], and Identix [8].

Facial scan is an effective biometric attribute/indicator. Different biometric indicators are suited for different kinds of identification applications due to their variations in intrusiveness, accuracy, cost, and ease of sensing [9] (see Fig. 1(a)). Among the six biometric indicators considered in [10], facial features scored the highest compatibility, shown in Fig. 1(b), in a machine readable travel documents (MRTD) system based on a number of evaluation factors [10].

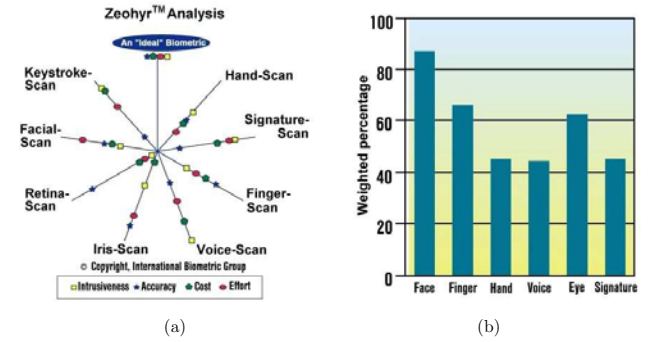


Figure 1: Comparison of various biometric features: (a) based on zephyr analysis [9]; (b) based on MRTD compatibility [10].

Global 2002 industry revenues of \$601million are expected to reach \$4.04billion by 2007 [9], driven by large-scale public sector biometric deployments, the emergence of transactional revenue models, and the adoption of standardized biometric infrastructures and data formats. Among emerging biometric technologies, facial recognition and middleware are projected to reach \$200million and \$215million, respectively, in annual revenues in 2005.

Face recognition scenarios can be classified into two types, (i) face verification (or authentication) and (ii) face identification (or recognition). In the Face Recognition Vendor Test (FRVT) 2002 [11], which was conducted by the National Institute of Standards and Technology (NIST), another scenario is added, called the 'watch list'.

- **Face verification** ("Am I who I say I am?") is a one-to-one match that compares a query

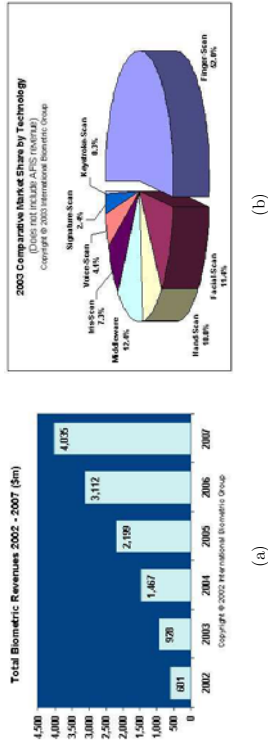


Figure 2: Face recognition market [9]. (a) Total biometric revenues 2002 - 2007. (b) Comparative market share by technology.

face image against a template face image whose identity is being claimed. To evaluate the verification performance, the verification rate (the rate at which legitimate users are granted access) vs. false accept rate (the rate at which imposters are granted access) is plotted, called ROC curve. A good verification system should balance these two rates based on operational needs.

- **Face identification** ("Who am I?") is a one-to-many matching process that compares a query face image against all the template images in a face database to determine the identity of the query face (see Fig. 3). The identification of the test image is done by locating the image in the database who has the highest similarity with the test image. The identification process is a "closed" test, which means the sensor takes an observation of an individual that is known to be in the database. The test subject's (normalized) features are compared to the other features in the system's database and a similarity score is found for each comparison. These similarity scores are then numerically ranked in a descending order. The percentage of times that the highest similarity score is the correct match for all individuals is referred to as the "top match score." If any of the top r similarity scores corresponds to the test subject, it is considered as a correct match in terms of the cumulative match. The percentage of times one of those r similarity scores is the correct match for all individuals is referred to as the "Cumulative Match Score". The "Cumulative Match Score" curve is the rank n versus percentage of correct identification, where rank n is the number of top similarity scores reported.

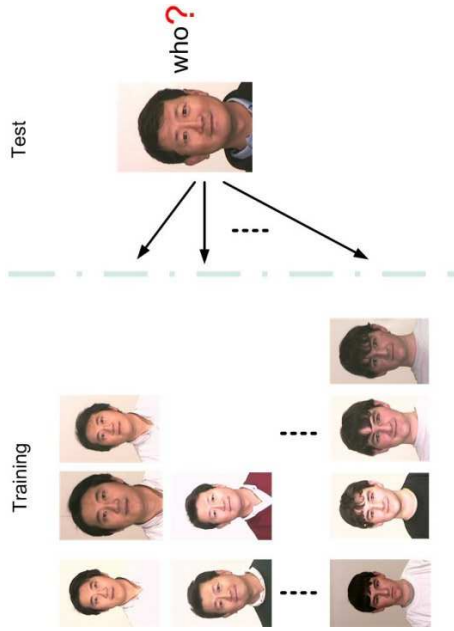


Figure 3: Face identification scenario.

- The **watch list** ("Are you looking for me?") method is an open-universe test. The test individual may or may not be in the system database. That person is compared to the others in the system's database and a similarity score is reported for each comparison. These similarity scores are then numerically ranked so that the highest similarity score is first. If a similarity score is higher than a preset threshold, an alarm is raised. If an alarm is raised, the system thinks that the individual is located in the system's database. There are two main items of interest for watch list applications. The first is the percentage of times the system raises the alarm and it correctly identifies a person on the watchlist. This is called the "Detection and Identification Rate." The second item of interest is the percentage of times the system raises the alarm for an individual that is not on the watchlist (database). This is called the "False Alarm Rate."

In this report, all the experiments are conducted in the identification scenario.

Human face image appearance has potentially very large intra-subject variations due to

- 3D head pose

- Illumination (including indoor / outdoor)

- Facial expression

- Occlusion due to other objects or accessories (e.g., sunglasses, scarf, etc.)

- Facial hair

- Aging [12].

On the other hand, the inter-subject variations are small due to the similarity of individual appearances. Fig. 4 gives examples of appearance variations of one subject. And Fig. 5 illustrates examples of appearance variations of different subjects. Currently, image-based face recognition techniques can be mainly categorized into two groups based on the face representation which they use: (i) appearance-based which uses holistic texture features; (ii) model-based which employ shape and texture of the face, along with 3D depth information.



Figure 4: Appearance variations of the same subject under different lighting conditions and different facial expressions [13].

A number of face recognition algorithms, along with their modifications, have been developed during the past several decades (see Fig. 6). In section 2, three leading linear subspace analysis schemes are presented, and several non-linear manifold analysis approaches for face recognition are briefly described. The model-based approaches are introduced in section 3, including Elastic Bunch Graph matching, Active Appearance Model and 3D Morphable Model methods. A number of face databases available in the public domain and several published performance evaluation results are provided in section 4. Concluding remarks and future research directions are summarized in section 5.

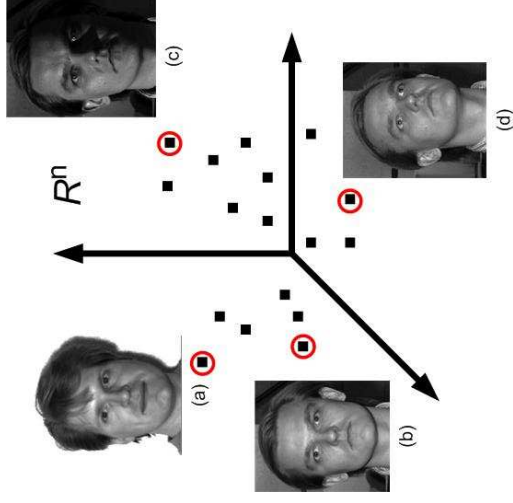


Figure 5: Inter-subject variations versus intra-subject variations. (a) and (b) are images from different subjects, but their appearance variations represented in the input space can be smaller than images from the same subject, b, c and d. These images are taken from Yale database B.

2 Appearance-based (View-based) face recognition

Many approaches to object recognition and to computer graphics are based directly on images without the use of intermediate 3D models. Most of these techniques depend on a representation of images that induces a vector space structure and, in principle, requires dense correspondence.

Appearance-based approaches represent an object in terms of several object views (raw intensity images). An image is considered as a high-dimensional vector, i.e., a point in a high-dimensional vector space. Many view-based approaches use statistical techniques to analyze the distribution of the object image vectors in the vector space, and derive an efficient and effective representation (feature space) according to different applications. Given a test image, the similarity between the stored prototypes and the test view is then carried out in the feature space.

2.2 Linear (subspace) Analysis

Three classical linear appearance-based classifiers, PCA [14], ICA [15] and LDA [16][17] are introduced in the following. Each classifier has its own representation (basis vectors) of a high dimensional face vector space based on different statistical viewpoints. By projecting the face vector to the basis vectors, the projection coefficients are used as the feature representation of each face image. The matching score between the test face image and the training prototype is calculated (e.g., as the cosine value of the angle) between their coefficients vectors. The larger the matching score, the better the match.

All the three representations can be considered as a linear transformation from the original image vector to a projection feature vector, i.e.

$$Y = W^T X, \quad (1)$$

where Y is the $d \times N$ feature vector matrix, d is the dimension of the feature vector, and W is the transformation matrix. Note that $d \leq n$.

2.2.1 PCA

The Eigenface algorithm uses the Principal Component Analysis (PCA) for dimensionality reduction to find the vectors which best account for the distribution of face images within the entire image space [14]. These vectors define the subspace of face images and the subspace is called face space. All faces in the training set are projected onto the face space to find a set of weights that describes the contribution of each vector in the face space. To identify a test image, it requires the projection of the test image onto the face space to obtain the corresponding set of weights. By comparing the weights of the test image with the set of weights of the faces in the training set, the face in the test image can be identified.

The key procedure in PCA is based on Karhunen-Loeve transformation [18]. If the image elements are considered to be random variables, the image may be seen as a sample of a stochastic process. The Principal Component Analysis basis vectors are defined as the eigenvectors of the scatter matrix S_T ,

$$S_T = \sum_{i=1}^N (x_i - \mu)(x_i - \mu)^T. \quad (2)$$

The transformation matrix W_{PCA} is composed of the eigenvectors corresponding to the d largest eigenvalues. A 2D example of PCA is demonstrated in Fig. 7. The eigenvectors (a.k.a. eigenface)

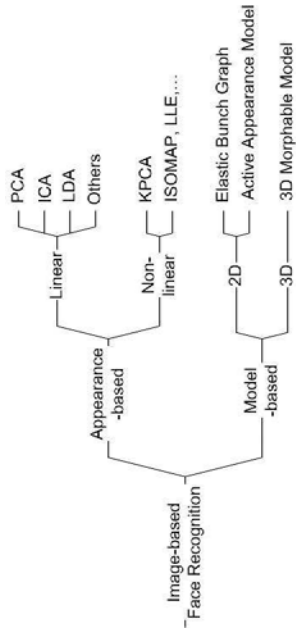


Figure 6: Face recognition methods covered in this report.

This image vector representation allows the use of learning techniques for the analysis and for the synthesis of images. Face recognition can be treated as a space-searching problem combined with a machine-learning problem.

2.1 Vector representation of images

Image data can be represented as vectors, i.e., as points in a high dimensional vector space. For example, a $p \times q$ 2D image can be mapped to a vector $x \in R^{pq}$, by lexicographic ordering of the pixel elements (such as by concatenating each row or column of the image). Despite this high-dimensional embedding, the natural constraints of the physical world (and the imaging process) dictate that the data will, in fact, lie in a lower-dimensional (though possibly disjoint) manifold. The primary goal of the subspace analysis is to identify, represent, and parameterize this manifold in accordance with some optimality criteria.

Let $X = (x_1, x_2, \dots, x_i, \dots, x_N)$ represent the $n \times N$ data matrix, where each x_i is a face vector of dimension n , concatenated from a $p \times q$ face image, where $n = p \times q$. Here n represents the total number of pixels in the face image and N is the number of different face images in the training set. The mean vector of the training images $\mu = \sum_{i=1}^N x_i$ is subtracted from each image vector.

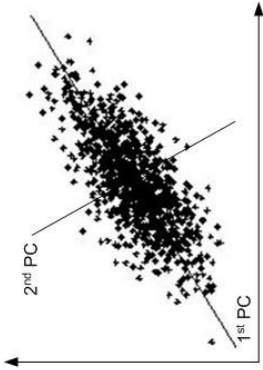


Figure 7: Principal components (PC) of a two-dimensional set of points. The first principal component provides an optimal linear dimension reduction from 2D to 1D, in the sense of the mean square error.

corresponding to the 7 largest eigenvalues, derived from ORL face database [19], are shown in Fig. 9. The corresponding average face is given in Fig. 8. ORL face samples are provided in Fig. 26. After applying the projection, the input vector (face) in an n -dimensional space is reduced to a feature vector in a d -dimensional subspace. Also the eigenvectors corresponding to the 7 smallest eigenvalues are provided in Fig. 10. For most applications, these eigenvectors corresponding to very small eigenvalues are considered as noise, and not taken into account during identification. Several extensions of PCA are developed, such as modular eigenspaces [20] and probabilistic subspaces [21].

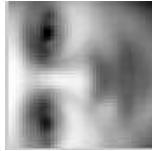


Figure 8: The average face (derived from the ORL face database [19]).

2.2.2 ICA

Independent Component Analysis (ICA) [22] is similar to PCA except that the distribution of the components are designed to be non-Gaussian. Maximizing non-Gaussianity promotes statistical

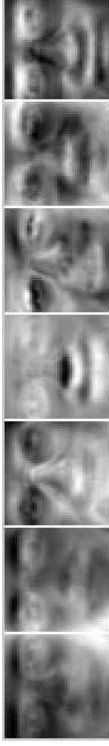


Figure 9: Eigenvectors corresponding to the 7 largest eigenvalues, shown as $p \times p$ images, where $p \times p = n$ (derived from the ORL face database [19]).



Figure 10: Eigenvectors corresponding to the 7 smallest eigenvalues, shown as $p \times p$ images, where $p \times p = n$ (derived from the ORL face database [19]).

independence. Figure 11 presents the different feature extraction properties between PCA and ICA. Bartlett et al. [15] provided two architectures based on Independent Component Analysis, statistically independent basis images and a factorial code representation, for the face recognition task. The ICA separates the high-order moments of the input in addition to the second-order moments utilized in PCA. Both the architectures lead to a similar performance. The obtained basis vectors based on fast fixed-point algorithm [24] for the ICA factorial code representation are illustrated in Fig. 12. There is no special order imposed on the ICA basis vectors.

2.2.3 LDA

Both PCA and ICA construct the face space without using the face class (category) information. The whole face training data is taken as a whole. In LDA the goal is to find an efficient or interesting way to represent the face vector space. But exploiting the class information can be helpful to the identification tasks, see Fig. 13 for an example.

The Fisherface algorithm [16] is derived from the Fisher Linear Discriminant (FLD), which uses class specific information. By defining different classes with different statistics, the images in the learning set are divided into the corresponding classes. Then, techniques similar to those used in Eigenface algorithm are applied. The Fisherface algorithm results in a higher accuracy rate in

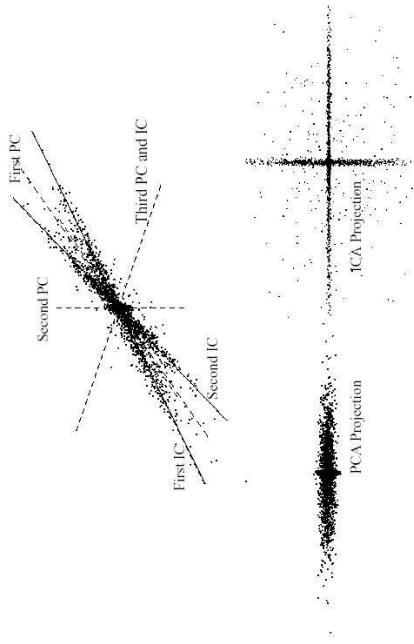


Figure 11: Top: Example 3D data distribution and the corresponding principal component and independent component axes. Each axis is a direction found by PCA or ICA. Note the PC axes are orthogonal while the IC axes are not. If only 2 components are allowed, ICA chooses a different subspace than PCA. Bottom left: Distribution of the first PCA coordinate of the data. Bottom right: distribution of the first ICA coordinate of the data [23]. For this example, ICA tends to extract more intrinsic structure of the original data clusters.



Figure 12: ICA basis vectors shown as $p \times p$ images; there is no special order for ICA basis vectors (derived from the ORL face database [19], based on the second architecture [23]). The ICA code package to compute these ICA is downloaded from <http://www.cis.hut.fi/projects/ica/fastica/>.

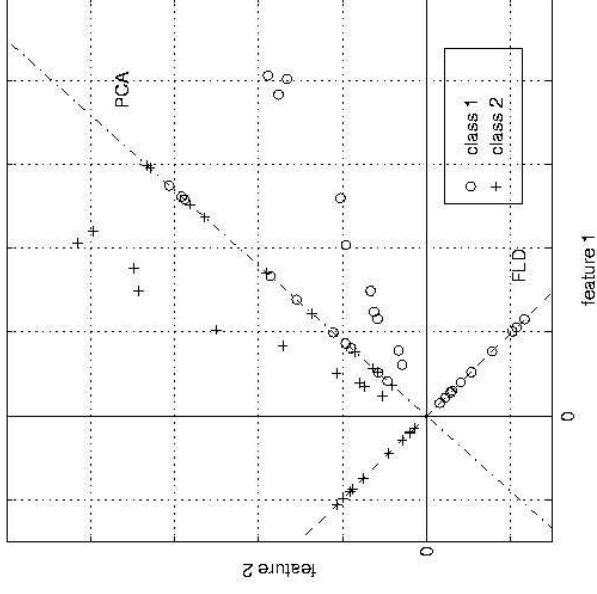


Figure 13: A comparison of principal component analysis (PCA) and Fisher's linear discriminant (FLD) for a two class problem where data for each class lies near a linear subspace [16]. It shows that FLD is better than PCA in the sense of discriminating the two classes.

recognizing faces when compared with Eigenface algorithm.

The Linear Discriminant Analysis finds a transform WLDA, such that

$$W_{LDA} = \arg \max_W \frac{W^T S_B W}{W^T S_W W}, \quad (3)$$

where S_B is the between-class scatter matrix and S_W is the within-class scatter matrix, defined as

$$S_B = \sum_{i=1}^c N_i (\mu_i - \mu)(\mu_i - \mu)^T, \quad (4)$$

$$S_W = \sum_{i=1}^c \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T. \quad (5)$$

In the above expression, N_i is the number of training samples in class i , c is the number of distinct classes, μ_i is the mean vector of samples belonging to class i and X_i represents the set of samples belonging to class i . The LDA basis vectors are demonstrated in Fig. 14.



Figure 14: First seven LDA basis vectors shown as $p \times p$ images (derived from the ORL face database [19]).

2.3 Non-linear (manifold) Analysis

The face manifold is more complicated than linear models. Linear subspace analysis is an approximation of this non-linear manifold. Direct non-linear manifold modeling schemes are explored to learn this non-linear manifold. In the following subsection, the kernel principal component analysis (KPCA) is introduced and several other manifold learning algorithms are also listed.

2.4 Kernel PCA

The kernel PCA [25] is to apply a nonlinear mapping from the input space R^M to the feature space R^L , denoted by $\Psi(x)$, where L is larger than M . This mapping is made implicit by the use of kernel functions satisfying the Mercer's theorem

$$k(x_i, x_j) = \Psi(x_i) \cdot \Psi(x_j). \quad (6)$$

where kernel functions $k(x_i, x_j)$ in the input space correspond to inner-product in the higher dimensional feature space. Because computing covariance is based on inner-products, performing a PCA in the feature space can be formulated with kernels in the input space without the explicit computation of $\Psi(x)$. Suppose the covariance in the feature space is calculated as

$$\Sigma_K = \langle \Psi(x_i) \Psi(x_i)^T \rangle. \quad (7)$$

The corresponding eigen-problem is $\lambda V = \Sigma_K V$. It can be proved [25] that V can be expressed as $V = \sum_{i=1}^N w_i \Psi(x_i)$, where N is the total number of training samples. The equivalent eigenvalue problem can be formulated in terms of kernels in the input space

$$N \lambda w = K w, \quad (8)$$

where w is a N -dimensional vector, K is a $N \times N$ matrix with $K_{ij} = k(x_i, x_j)$.

The projection of a sample x onto the n^{th} eigenvector V^n can be calculated by

$$p_n = (V^n \cdot \Psi(x)) = \sum_{i=1}^N w_i^n k(x_i, x_j). \quad (9)$$

Figure 15 gives an example of KPCA.

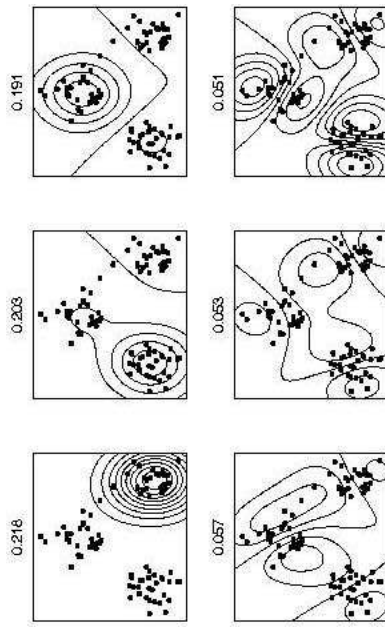


Figure 15: Contour plots of the first six principal component projections. Each contour contains the same projection values onto the corresponding eigenvectors. Data is generated by 3 Gaussian clusters. A RBF kernel is used. The corresponding eigenvalues are given above each subplot. Notice that the first three components have the potential to extract the individual clusters [25].

Similar to traditional PCA, the projection coefficients are used as features for classification. Yang [26] explored the use of KPCA for the face recognition problem. Unlike traditional PCA, KPCA can use more eigenvector projections than the input dimensionality. But a suitable kernel and correspondent parameters can only be decided empirically.

Manifold learning has attracted a lot of effort in the machine learning community recently. ISOMAP [27] and LLE [28] have been proposed to learn the non-linear manifold, where the learned manifold have been shown for some simple face images. Yang [29] applied LDA to the face recognition with the feature of the geodesic distance, which is the basis of the ISOMAP. These manifold learning algorithms are interesting, but further exploration is needed to demonstrate their performance in the face recognition in real applications.

2.5 Small Sample Size

In real applications, current appearance-based face recognition systems encounter difficulties due to the small number of available training face images and complex facial variations during the test. Human face appearances have a lot of variations resulting from varying lighting conditions, different head poses and facial expressions. In real-world situations, only a small number of samples for each subject are available for training. If a sufficient amount of enough representative data is not available, Martinez and Kak [30] have shown that the switch from nondiscriminant techniques (e.g., PCA) to discriminant approaches (e.g., LDA) is not always warranted and may sometimes lead to poor system design when small and nonrepresentative training data sets are used. Figure 16 gives an example.

Therefore, face synthesis, where additional training samples can be generated, is helpful to enhance the face recognition systems [31][32].

3 Model-based face recognition

The model-based face recognition scheme is aimed at constructing a model of the human face, which is able to capture the facial variations. The prior knowledge of human face is highly utilized to design the model. For example, feature-based matching derives distance and relative position features from the placement of internal facial elements (e.g., eyes, etc.). Kanade [33] developed one of the earliest face recognition algorithms based on automatic feature detection. By localizing the corners of the eyes, nostrils, etc. in frontal views, his system computed parameters for each face, which were compared (using a Euclidean metric) against the parameters of known faces. A more recent feature-based system, based on elastic bunch graph matching, was developed by Wiskott et al. [34] as an extension to their original graph matching system [35]. By integrating both shape

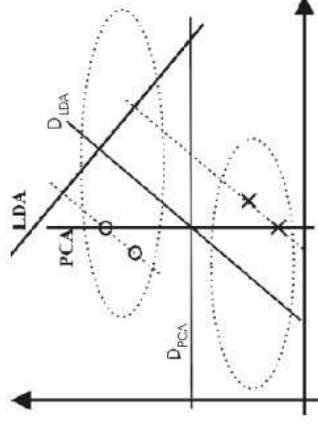


Figure 16: Suppose there are two different classes embedded in two different "Gaussian-like" distributions. However, only two sample per class are supplied to the learning procedure (PCA or LDA). The classification result of the PCA procedure (using only the first eigenvector) is more desirable than the result of the LDA. D_{PCA} and D_{LDA} represent the decision thresholds obtained by using the nearest-neighbor classification [30].

and texture, Cootes et al. [36][37] developed a 2D morphable face model, through which the face variations are learned. A more advanced 3D morphable face model is explored to capture the true 3D structure of human face surface. Both morphable model methods come under the framework of "interpretation through synthesis".

The model-based scheme usually contains three steps: 1) Constructing the model; 2) Fitting the model to the given face image; 3) Using the parameters of the fitted model as the feature vector to calculate the similarity between the query face and prototype faces in the database to perform the recognition.

3.1 Feature-based Elastic Bunch Graph Matching

3.1.1 Bunch Graph

All human faces share a similar topological structure. Wiskott et al. present a general in-class recognition method for classifying members of a known class of objects. Faces are represented as

graphs, with nodes positioned at fiducial points (such as the eyes, the tip of the nose, some contour points, etc.; see Fig. 17), and edges labeled with 2-D distance vectors.

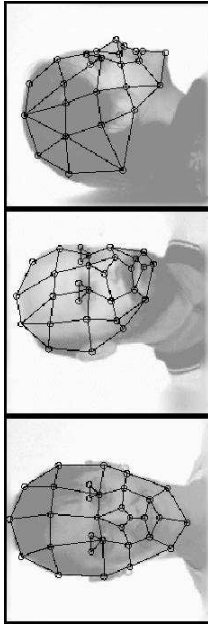


Figure 17: Multiview faces overlaid with labeled graphs [34].

Each node contains a set of 40 complex Gabor wavelet coefficients, including both phase and magnitude, known as a jet (shown in Fig. 18). Wavelet coefficients are extracted using a family of Gabor kernels with 5 different spatial frequencies and 8 orientations; all kernels are normalized to be of zero mean.

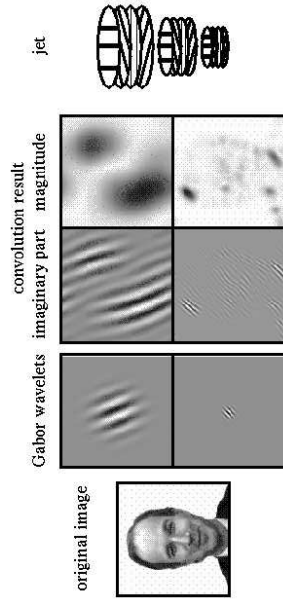


Figure 18: Jet [35].

Face recognition is based on labeled graphs. A labeled graph is a set of nodes connected by edges; nodes are labeled with jets; edges are labeled with distances. Thus, the geometry of an object is encoded by the edges while the gray value distribution is patch-wise encoded by the nodes (jets). An example is shown in Fig. 19.

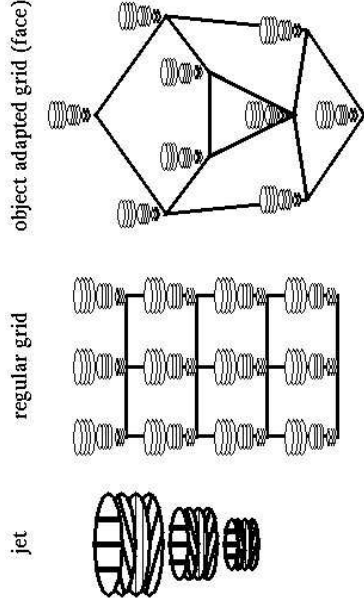


Figure 19: Labeled graph [35].

While individual faces can be represented by simple labeled graphs, a face class requires a more comprehensive representation in order to account for all kinds of variations within the class. The Face Bunch Graph has a stack-like structure that combines graphs of individual sample faces, as demonstrated in Fig. 20. It is crucial that the individual graphs all have the same structure and that the nodes refer to the same fiducial points. All jets referring to the same fiducial point, e.g., all left-eye jets, are bundled together in a bunch, from which one can select any jet as an alternative description. The left-eye bunch might contain a male eye, a female eye, both closed or open, etc. Each fiducial point is represented by such a set of alternatives and from each bunch any jet can be selected independently of the jets selected from the other bunches. This provides full combinatorial power of this representation even if it is constituted only from a few graphs.

3.1.2 Elastic Graph Matching

To identify a new face, the face graph is positioned on the face image using elastic bunch graph matching. The goal of Elastic graph matching is to find the fiducial points on a query image and thus to extract from the image a graph which maximizes the graph similarity function. This is performed automatically if the face bunch graph (FBG) is appropriately initialized. A face bunch graph (FBG) consists of a collection of individual face model graphs combined into a stack-like

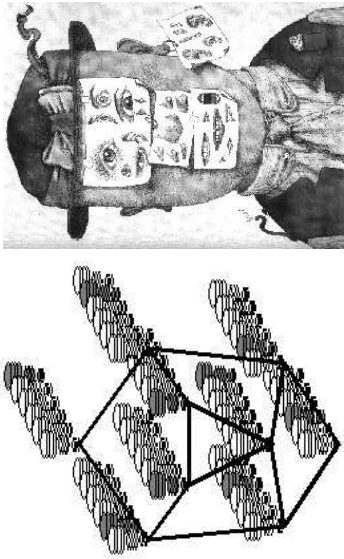


Figure 20: The left figure shows a sketch of a face bunch graph [34]. Each of the nine nodes is labeled with a bunch of six jets. From each bunch, one particular jet has been selected, indicated as gray. The actual selection depends on the situation, e.g., the face onto which the face bunch graph is matched. Though constructed from six sample faces only, this bunch graph can potentially represent $6^9 = 10,077,696$ different faces. The right figure shows the same concept interpreted slightly differently by Tullio Pericoli ("Unfinished Portrait" 1985) [http://www.cml.salk.edu/simwiskott/Projects/BunchGraph.html].

structure, in which each node contains the jets of all previously initialized faces from the database. To position the grid on a new face, the graph similarity between the image graph and the existing FBG is maximized. Graph similarity is defined as the average of the best possible match between the new image and any face stored within the FBG minus a topographical term (see Eq. 11), which accounts for distortion between the image grid and the FBG. Let S_ϕ be the similarity between two jets, defined as

$$S_\phi(I, J') = \frac{\sum_j a_j a'_j \cos(\phi_j - \phi'_j - d\vec{k}_j)}{\sqrt{\sum_j a_j^2 \sum_j a'^2_j}}, \quad (10)$$

where a_j and ϕ_j are magnitude and phase of the Gabor coefficients in the j^{th} jet, respectively; \vec{d} is the displacement between locations of the two jets; \vec{k}_j determines the wavelength and orientation of the Gabor wavelet kernels [35]. For an image graph G' with nodes $n = 1, \dots, N$ and edges

$$e = 1, \dots, E \text{ and an FBG } B \text{ with model graphs } m = 1, \dots, M, \text{ the graph similarity is defined as} \quad (11)$$

$$S_B(G', B) = \frac{1}{N} \sum_n \max_m S_\phi(I_n, J_n^{Bm}) - \frac{\lambda}{E} \sum_e \frac{(\Delta \vec{x}_e^I - \Delta \vec{x}_e^B)^2}{(\Delta \vec{x}_e^B)^2},$$

where λ determines the relative importance of jets and metric structure, J_n is the jets at nodes n , and $\Delta \vec{x}_e$ is the distance vector used as labels at edges e . After the grid has been positioned on the new face, the face is identified by comparing the similarity between that face and every face stored in the FBG. Graphs can be easily translated, rotated, scaled, and elastically deformed, thus compensating for the variance in face images which is commonly encountered in a recognition process.

3.2 AAM - A 2D Morphable Model

An Active Appearance Model (AAM) is an integrated statistical model which combines a model of shape variation with a model of the appearance variations in a shape-normalized frame. An AAM contains a statistical model of the shape and gray-level appearance of the object of interest which can generalize to almost any valid example. Matching to an image involves finding model parameters which minimize the difference between the image and a synthesized model example, projected onto the image. The potentially large number of parameters makes this a difficult problem.

3.2.1 AAM Construction

The AAM is constructed based on a training set of labeled images, where landmark points are marked on each example face at key positions to outline the main features (shown in Fig. 21).



Figure 21: The training image is split into shape and shape-normalized texture [38].

The shape of a face is represented by a vector consisting of the positions of the landmarks, $s = (x_1, y_1, \dots, x_n, y_n)^T$, where (x_j, y_j) denotes the 2D image coordinate of the j^{th} landmark

point. All shape vectors of faces are normalized into a common coordinate system. The principal component analysis is applied to this set of shape vectors to construct the face shape model, denoted as: $s = \bar{s} + P_s b_s$, where \bar{s} is a shape vector, \bar{s} is the mean shape, P_s is a set of orthogonal modes of shape variation and b_s is a set of shape parameters.

In order to construct the appearance model, the example image is warped to make the control points match the mean shape. Then the warped image region covered by the mean shape is sampled to extract the gray level intensity (texture) information. Similar to the shape model construction, a vector is generated as the representation, $g = (I_1, \dots, I_m)^T$, where I_j denotes the intensity of the sampled pixel in the warped image. PCA is also applied to construct a linear model $g = \bar{g} + P_g b_g$, where \bar{g} is the mean appearance vector, P_g is a set of orthogonal modes of gray-level variation and b_g is a set of gray-level model parameters.

Thus, all shape and texture of any example face can be summarized by the vectors b_s and b_g . The combined model is the concatenated version of b_s and b_g , denoted as follows:

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s P_s^T (s - \bar{s}) \\ P_g^T (g - \bar{g}) \end{pmatrix}, \quad (12)$$

where W_s is a diagonal matrix of weights for each shape parameter, allowing for the difference in units between the shape and gray scale models. PCA is applied to b also, $b = Qc$, where c is the vector of parameters for the combined model.

The model was built based on 400 face images, each with 122 landmark points [37]. A shape model with 23 parameters, a shape-normalized texture model with 113 parameters and a combined appearance model with 80 parameters (containing 98% variations of the observation) are generated. The model used about 10,000 pixel values to make up the face.

3.2.2 AAM Fitting

Given a new image and constructed model, the metric used to measure the match quality between the model and image is $\Delta = |\delta I|^2$, where δI is the vector of intensity differences between the given image and the image generated by the model tuned by the model parameters, called residuals. The AAM fitting seeks the optimal set of model parameters that best describes the given image. Cootes [36] observed that displacing each model parameter from the correct value induces a particular pattern in the residuals. In the training phase, the AAM learned a linear model that captured the relationship between parameter displacements and the induced residuals. During the model fitting,

it measures the residuals and uses this model to correct the values of current parameters, leading to a better fit. Figure 22 shows examples of the iterative AAM fitting process.



Figure 22: Examples of the AAM fitting iterations [38].

3.2.3 Face Recognition by AAM

For all the training images, the corresponding model parameter vectors are used as the feature vectors. The linear discrimination analysis (LDA) is utilized to construct the discriminant subspace for face identity recognition. Given a query image, the AAM fitting is applied to extract the corresponding feature vector. The recognition is achieved by finding the best match between the query feature vector and stored prototype feature vectors, both of which are projected onto the discriminant subspace.

3.3 3D Morphable Model

Human face is a surface lying in the 3D space intrinsically. Therefore, in principle, the 3D model is better for representing faces, especially to handle facial variations, such as pose, illumination. Blanz et al. [39][40] proposed a method based on a 3D morphable face model that encodes shape and texture in terms of model parameters, and an algorithm that recovers these parameters from a

single image of a face. For face identification, we use the shape and texture parameters of the model that are separated from imaging parameters, such as pose and illumination. Fig. 23 illustrates the scheme. To handle the extreme image variations induced by these parameters, one common approach taken by various groups is to use generative image models. For image analysis, the general strategy of all these techniques is to fit the generative model to a novel image, thereby parameterizing it in terms of the model. In order to make identification independent of imaging conditions, the goal is to separate intrinsic model parameters of the face from extrinsic imaging parameters. The separation of intrinsic and extrinsic parameters is achieved explicitly by simulating the process of image formation using 3D computer graphics technology.

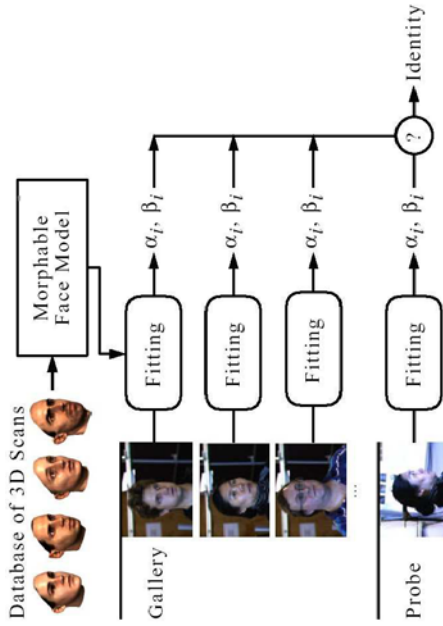


Figure 23: The three-dimensional morphable face model derived from a database of laser scans is used to encode gallery and probe images. For identification, the model coefficients of the probe image are compared with the coefficients of all gallery images [40].

3.3.1 Model Construction

Generalizing the well-known morphing process between pairs of three-dimensional objects, the morphable face model is based on a vector space representation of faces [31]. The database of laser scans used in this study contains scans of 100 males and 100 females recorded with a *Cyberware™* 3030PS scanner. Scans are stored in cylindrical coordinates relative to a vertical axis. The coordinates and texture values of all the n vertices of the reference face ($n = 75,972$) are concatenated to form shape and texture vectors

$$S_0 = (x_1, y_1, z_1, \dots, x_n, y_n, z_n)^T, \quad (13)$$

$$T_0 = (R_1, G_1, B_1, \dots, R_n, G_n, B_n)^T. \quad (14)$$

Vectors S_i and T_i of the examples $i = 1 \dots m$ in the database are formed in a common coordinate system. Convex combinations of the examples produce novel shape and texture vectors S and T . Previous results [39] indicate that the shape and texture can be combined independently:

$$S = \sum_{i=1}^m a_i S_i, T = \sum_{i=1}^m b_i T_i. \quad (15)$$

S and T can also be represented as:

$$S = \bar{s} + \sum_{i=1}^{m-1} \alpha_i s_i, T = \bar{t} + \sum_{i=1}^m \beta_i T_i, \bar{s} = \frac{1}{m} \sum_{i=1}^m S_i, \bar{t} = \frac{1}{m} \sum_{i=1}^m T_i, \quad (16)$$

where \bar{s} is the mean shape and \bar{t} is the mean texture.

3.3.2 Model Fitting

The image synthesis is to render the new projected positions of vertices of the 3D model, along with illumination and color. During the process of fitting the model to a novel image, not only the shape and texture coefficients α_i and β_i are optimized, but also the following rendering parameters, which are concatenated into a vector p : the head orientation angles ϕ , θ and γ , the head position (P_x, P_y) in the image plane, size s , color and intensity of the light sources L , as well as color constant, and gain and offset of colors, shown in Fig. 24.

The primary goal in analyzing a face is to minimize the sum of square differences over all color channels and all pixels in the input image and the symmetric reconstruction,

$$E_I = \sum_{x,y} \|I_{input}(x,y) - I_{model}(x,y)\|^2. \quad (17)$$

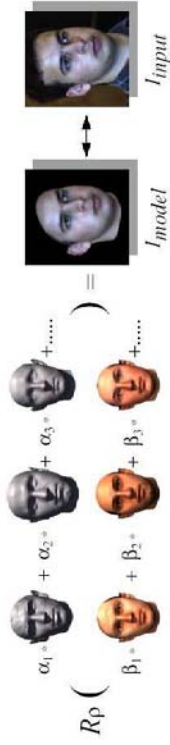


Figure 24: The goal of the fitting process is to find shape and texture coefficients α and β such that rendering R_p produces an image I_{model} that is as similar as possible to I_{input} [40].

Under a probabilistic framework, the overall cost function to be minimized is derived as [40]:

$$E = \frac{1}{\sigma_N^2} E_I + \sum_i \frac{\alpha_i^2}{\sigma_{S,i}^2} + \sum_i \frac{\beta_i^2}{\sigma_{T,i}^2} + \sum_i \frac{(\rho_i - \bar{\rho}_i)^2}{\sigma_{R,i}^2}, \quad (18)$$

A modification of stochastic gradient descent algorithm is used to optimize the cost function. The optimization is achieved globally, resulting in a set of global parameters α_{global} and β_{global} . The face model is divided into four regions – eyes, nose, mouth and the surrounding face segment. The optimization is also applied separately for each region to obtain the local parameters for each segment, i.e., $\alpha_{r1}, \beta_{r1}, \dots, \alpha_{r4}$ and β_{r4} . The fitting process is demonstrated in Fig. 25.

3.3.3 Recognition

The similarity between two faces is defined as:

$$S = \sum_{global, r1, r2, r3, r4} \left(\frac{\langle \alpha, \alpha' \rangle_M}{\|\alpha\|_M \cdot \|\alpha'\|_M} + \frac{\langle \alpha, \alpha' \rangle_M}{\|\alpha\|_M \cdot \|\alpha'\|_M} \right), \quad (19)$$

where $\langle \alpha, \alpha' \rangle = \sum_i \frac{\alpha_i \alpha'_i}{\sigma_{S,i}^2}$, $\langle \beta, \beta' \rangle = \sum_i \frac{\beta_i \beta'_i}{\sigma_{T,i}^2}$, $\|\alpha\|_M^2 = \langle \alpha, \alpha \rangle_M$. The query image will be assigned the identity with which the similarity between the query and the corresponding prototype is maximized.

Besides the above-mentioned techniques, a number of interesting approaches have been explored from different perspectives, such as local feature analysis [41] and statistical model based face recognition methods. Examples of the statistical model based scheme are 1D Hidden Markov Model (HMM) [42] and pseudo-2D HMM [43].

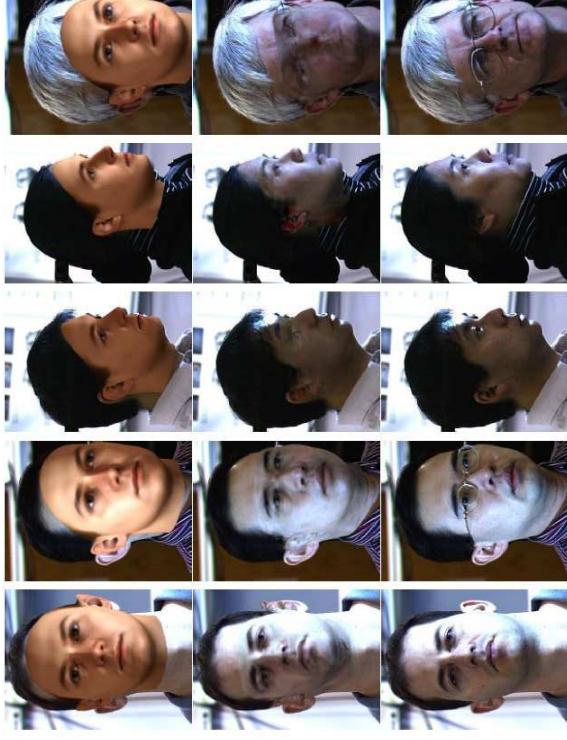


Figure 25: Examples of model fitting [40]. Top row: initial parameters, Middle row: Results of fitting, rendered on top of the input images. Bottom row: Input images. The fifth column is an example of a poor fit.

4 Databases and Performance Evaluation

A number of face databases have been collected for different face recognition tasks. Table 1 lists a selection of those available in the public domain. The AR database contains occlusions due to eye glasses and scarf. The CMU PIE database is collected with well-constrained pose, illumination and expression. FERET [44] and XM2VTS databases are the two most comprehensive databases, which can be used as a benchmark for detailed testing or comparison. The XM2VTS is especially designed for multi-modal biometrics, including audio and video cues. To continue the facial recognition technology evaluation with the state-of-the-art advances, the Face Recognition Vendor Test

(FRVT) [45] followed the original FERET, and was conducted in the year 2000 and 2002 (namely FRVT2000 and FRVT2002). The database used in FRVT was significantly extended between 2000 and 2002, including more than 120,000 face images from more than 30,000 subjects. More facial appearance variations were also considered in FRVT, such as indoor/outdoor difference.

Table 1: Selected face databases available in the public domain. ¹XM2VTS database is not available free of charge. ²FERET database is complicated, for details, see http://www.itl.nist.gov/iad/humanid/feret/feret_master.html. *For each subject, it collects video+audio+3D model. for details, see <http://www.ec.surrey.ac.uk/Research/VSSP/xm2vtsdb/>; **e: expression; i: illumination; o: occlusion; p: pose; s: scale; t: time interval, images for the same subject are taken between a short period (e.g., a couple of days), or a long period (e.g., years).

Face database	No. of subjects	No. of images	Variations included **
ORL [19]	40	400	p, e
Yale [13]	15	165	i, e
AR [46]	>120	>3,000	i, e, o, t
MIT [47]	16	432	i, p, s
UMIST [48]	20	564	p
CMU PIE [49]	68	41,308	p, i, e
XM2VTS ¹ [50]	295	*	*
FERET ² [51]	>1,000	>10,000	p, i, e, t

Some examples from the ORL database are shown in Fig. 26.

Based on the published experimental results, it is very difficult to put all the face recognition algorithms together for comparison in a rather fair protocol. There is no common benchmark database on which all the algorithms have been tested, although FERET is an excellent attempt in this direction. Researchers have their own choices on databases when doing the performance evaluation for publications. Also, for the same algorithm, the recognition accuracy may vary due to different evaluation protocols (leave-one-out, cross-validation etc.), different normalization schemes (e.g., facial area cropping styles), different image resolutions, different parameter settings (e.g., dimensionality of the subspace), etc. As a result, the reported performance in this section is a selection from several major publications, showing the results which are obtained by a subset of algorithms being applied on the same database.



Figure 26: Face samples from the ORL face database.

1) Table 2 presents the comparison results between PCA and LDA based on Yale Database. It shows that LDA is able to use a smaller size of the subspace to achieve a higher recognition accuracy.

Table 2: Leave-one-out evaluation of PCA and LDA on the Yale face database [16]. "Close crop" means the face area is cropped to contain only internal structures such as the eyebrows, eyes, nose and mouth, while "full face" cropping contains the entire face area.

Approach	Dim. of the subspace	Error rate (close crop)	Error rate (full face)
Eigenface (PCA)	30	24.4	19.4
Fisherface (LDA)	15	7.3	0.6

2) Figure 27 and Table 3 give the experimental results provided in [21], which compared PCA, ICA, KPCA and Probabilistic Subspace methods [21]. The experimental data consisted of a training set of 706 individual FERET faces and 1,123 "probe" images containing one or more views of every person in the gallery. All images are aligned and normalized [52]. Face images are downsampled to 21-by-12 pixels, thereby reducing the dimensionality of the input space to $21 \times 12 = 252$.

3) Performance of Elastic Bunch Graph Matching. The experiments are conducted on the FERET database. Results are given in table 4. The recognition results of this system are good on identical poses, e.g., frontal views against frontal views. However, across different poses, e.g., frontal views

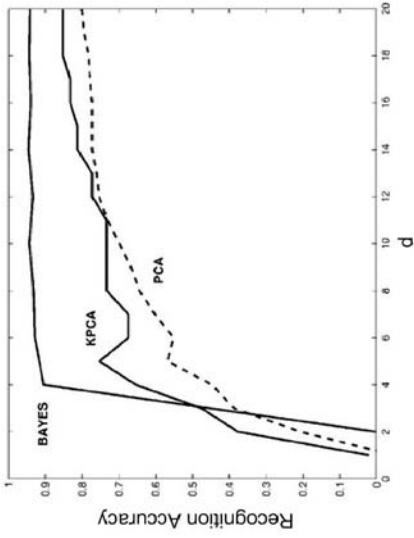


Figure 27: Recognition accuracy w.r.t. the dimensionality (d) of the subspace by 5-fold cross-validation. Data used is part of the FERET database, 1,829 images for 706 subjects [21].

Table 3: Comparison of subspace algorithms ($d = 20$) by 5-fold cross-validation [21]., where d is the dimensionality of the subspace. Data is part of FERET database, 1,829 images for 706 subjects.

	PCA	ICA	KPCA	Bayes [21]
Accuracy	77%	77%	87%	95%
Computation (floating-point operations)	10^8	10^9	10^9	10^8
Uniqueness	Yes	No	Yes	Yes
Projections	Linear	Linear	Nonlinear	Linear

against half profiles, the system performs rather poorly.

4) Performance of AAM-based face recognition. Totally 400 faces of 20 individuals were collected, with 200 images used for training and 200 for testing. The AAM is used to fit both training and test images, given the initial eye positions. The recognition accuracy is 88% [37]. No comparison with other face recognition algorithms is provided.

5) Performance of 3D Morphable Model. The database contains 68 subjects, each with 3 poses

Table 4: Recognition results of Elastic Bunch Graph Matching [34]. It shows two types of accuracies, (i) how often the correct model is identified as rank one, and (ii) how often it was among the top 10 (4%). fa: neutral frontal view; fb: frontal view with expression; hr: half-profile right (rotated by around $40-70^\circ$); hl: half-profile left; pr: profile right (rotated by around 90°); pl: profile left.

Model gallery	Probe images	First rank # (%)	Top 10 # (%)
250 fa	250 fb	245 (98)	248 (99)
250 hr	181 hl	103 (57)	147 (81)
250 pr	250 pl	210 (84)	236 (94)
249 fa + 1fb	171 hl + 79 hr	44 (18)	111 (44)
171 hl + 79 hr	249 fa + 1 fb	42 (17)	95 (38)
170 hl + 80 hr	217 pl + 33 pr	22 (9)	67 (27)
217 pl + 33 pr	170 hl + 80 hr	31 (12)	80 (32)

and 22 different illumination directions, for a total of $68 \times 3 \times 22 = 4,488$ images. The training gallery contains a single image for each of 68 subjects. All images have the same illumination direction. The remaining images ($4488 - 68 = 4,420$ images) are used as the test (probe) set. Recognition results are shown with different training pose and probe pose combinations in table 5. The performance of this method across pose is better than Elastic Graph Matching method.

Table 5: Recognition results of 3D morphable model [40].

training view	frontal	mean	testview	
			frontal (%)	side (%)
		std	94	85
		std	6.3	20.7
side	side	mean	89	90
		std	6.4	9.2
		mean	71	71
		std	9.2	12.2
profile	profile	mean	71	71
		std	9.2	12.2
		mean	71	71
		std	9.2	12.2

6) FRVT2002. FRVT is an independently administered technology evaluation of mature face

recognition systems by NIST. In 2002, ten commercial products participated in FRVT2002. The task designed for FRVT is very close to the real application scenarios. On March 2003, NIST issued the evaluation report for FRVT2002, which indicates the current state-of-the-art of the face recognition techniques [53].

Figure 28 plots identification performance of the top three commercial face recognition products, namely Cognitec, Eyematic and Identix. The database consists of 37,437 individuals. Figure 29 demonstrates that 3D morphable model [40] significantly improves the identification performance on non-frontal face recognition tasks.

FRVT2002 also shows that identification performance is dependent on the size of the database. For every doubling of the database size, performance decreases by 2% to 3% points.

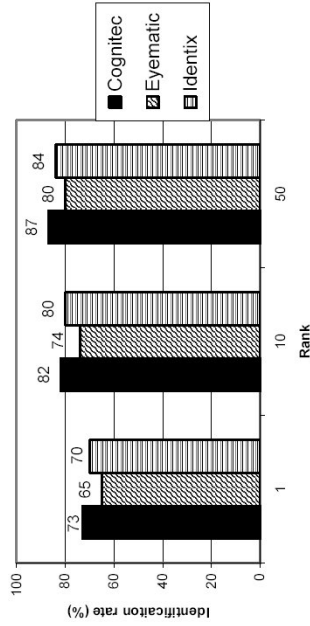


Figure 28: Identification results for the three best face recognition systems [11].

5 Summary

Image-based face recognition is still a very challenging topic after decades of exploration. A number of typical algorithms are presented, being categorized into appearance-based and model-based schemes. Table 6 provides the pros and cons of these two types of face recognition methods.

Sensitivity to variations in pose and different lighting conditions is still a challenging problem. Georgiades et al. [54] extensively explored the illumination change and synthesis for facial analysis using appearance-based approaches to achieve an illumination-invariant face recognition system.

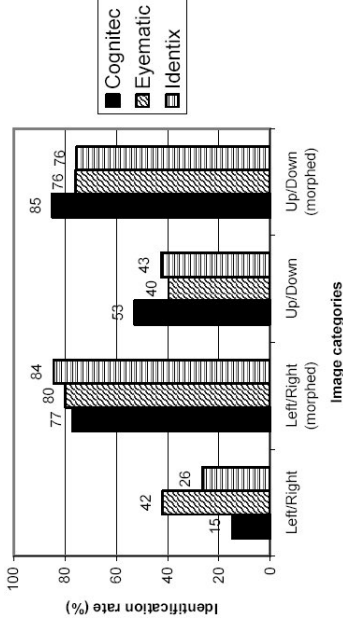


Figure 29: Evaluation of effectiveness of morphable model for non-frontal face identification tasks [11]. Performance is on a database consisting of 87 subjects.

Basri and Jacobs [55] proved that the set of all reflectance functions (the mapping from surface normals to intensities) produced by Lambertian objects under distant, isotropic lighting lies close to a 9D linear subspace. Their analysis was based on using spherical harmonics to represent lighting functions. The proposed algorithm was utilized for face recognition across illumination changes. Although a number of efforts have been made on pose-invariant face recognition, the performance of current face recognition system are still not satisfactory [11][56].

References

- [1] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, pp. 705-740, 1995.
- [2] H. Wechsler, P. Phillips, V. Bruce, F. Soulie, and T. Huang, *Face Recognition: From Theory to Applications*, Springer-Verlag, 1996.
- [3] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips, "Face recognition: A literature survey," *CVL Technical Report, University of Maryland*, 2000, <ftp://ftp.cfar.umd.edu/TRs/CVL-Reports-2000/TR4167-zhao.ps.gz>.

Table 6: Pros and cons of two major types of face recognition methods.

	Appearance-based	Model-based
Pros	<ol style="list-style-type: none"> 1. Face recognition problem is transformed to a face space analysis problem, where a number of well known statistical methods can be tried out. 2. Applicable to low resolution or poor quality images. 	<ol style="list-style-type: none"> 1. The model has intrinsic physical relationship with real faces. 2. An explicit modeling of face variations, such as pose, illumination and expression, gives the possibility to handle these variabilities in practice. 3. Integrating prior human knowledge.
Cons	<ol style="list-style-type: none"> 1. Sufficient representative data is needed to sample the underlying distribution successfully. 2. Does not utilize the prior (expert) knowledge of the human faces. 3. Subject to the limitations in facial variations, such as 3D pose, illumination and expression. 4. Correspondence needs to be established in advance, although the tangent distance may be used to tolerate small correspondence displacements. 	<ol style="list-style-type: none"> 1. Model construction is complicated and laborious. 2. Facial feature points are difficult to extract automatically with robustness. 3. Model fitting is a searching process, prone to be trapped into local minimum; recognition results highly depend on the fitting results. 4. Fitting process takes time. 5. Relatively high resolution and good quality face images are needed. 6. Appropriate initialization is needed.

- [4] S. Gong, S.J. McKenna, and A. Psarrou, *Dynamic Vision: from Images to Face Recognition*, Imperial College Press and World Scientific Publishing, 2000.
- [5] *Cognitec Systems GmbH*, <<http://www.cognitec-systems.de/Contact/contact.html>>.
- [6] *Eyematic Interfaces Inc.*, <<http://www.eyematic.com/>>.
- [7] *Viisage*, Littleton, MA, <<http://www.viisage.com/>>.
- [8] *Identix*, Minnetonka, MN, <<http://www.identix.com/>>.
- [9] *International Biometric Group*, <<http://www.biometricgroup.com/>>.
- [10] R. Hietmeyer, "Biometric identification promises fast and secure processing of airline passengers," *The Int'l Civil Aviation Organization Journal*, vol. 55, no. 9, pp. 10–11, 2000.
- [11] P.J. Phillips, P. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi, and J.M. Bone, "FRVT 2002: Overview and summary," March 2003, <<http://www.frvt.org/FRVT2002/documents.htm>>.
- [12] A. Lanitis, C. J. Taylor, and T. F. Cootes, "Towards automatic simulation of ageing effects on face images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 442–455, 2002.
- [13] *Yale University face database*, <<http://cvc.yale.edu/projects/yalefaces/yalefaces.html>>.
- [14] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, Mar. 1991.
- [15] M.S. Bartlett, H.M. Lades, and T.J. Sejnowski, "Independent component representations for face recognition," in *Proceedings of SPIE*, 1998, vol. 3299, pp. 528–539.
- [16] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
- [17] D. L. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 8, pp. 831–836, 1996.

- [18] M. Kirby and L. Sirovich, "Application of the Karhunen-Loève procedure for the characterization of human faces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, no. 1, pp. 103–108, Jan. 1990.
- [19] *ORL face database*, <<http://www.uk.research.att.com/facedatabase.html>>.
- [20] A. Penland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 1994, pp. 84–91.
- [21] B. Moghaddam, "Principal manifolds and probabilistic subspaces for visual recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 780–788, Feb. 2002.
- [22] A. Hyvarinen, "Survey on independent component analysis," *Neural Computing Surveys*, vol. 2, pp. 94–128, 1999.
- [23] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Networks*, vol. 13, no. 6, pp. 1450–1464, 2002.
- [24] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.
- [25] B. Scholkopf, A. Smola, and K. Muller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, no. 5, pp. 1299–1319, 1998.
- [26] Ming-Hsuan Yang, "Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, Washington D. C., May 2002, pp. 215–220.
- [27] J.B. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.
- [28] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.
- [29] Ming-Hsuan Yang, "Face recognition using extended isomap," in *Proc. IEEE International Conference on Image Processing*, Rochester, New York, Sep. 2002, vol. 2, pp. 117–120.
- [30] A.M. Martinez and A.C. Kak, "PCA versus LDA," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, Feb. 2001.
- [31] T. Vetter and T. Poggio, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 733–742, 1997.
- [32] W. Zhao and R. Chellappa, "SFS based view synthesis for robust face recognition," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 285–292.
- [33] T. Kamade, *Picture Processing by Computer Complex and Recognition of Human Faces*, PhD thesis, Kyoto University, 1973.
- [34] L. Wiskott, J.M. Fellous, N. Kruger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 775–779, 1997.
- [35] M. Lades, J. C. Vorbruggen, J. Buhmann, Jorg Lange, C. Malsburg, R. P. Wurtz, and W. Koenen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Computers*, vol. 42, no. 3, pp. 300–310, Jan. 1993.
- [36] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models," in *Proc. European Conference on Computer Vision*, 1998, vol. 2, pp. 484–498.
- [37] G.J. Edwards, T.F. Cootes, and C.J. Taylor, "Face recognition using active appearance models," in *Proc. European Conference on Computer Vision*, 1998, vol. 2, pp. 581–695.
- [38] T.F. Cootes, G.J. Edwards, and C.J. Taylor, "Active appearance models," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, Jun. 2001.
- [39] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proc. ACM SIGGRAPH*, Mar. 1999, pp. 187–194.
- [40] Volker Blanz, Sami Romdhani, and Thomas Vetter, "Face identification across different poses and illuminations with a 3D morphable model," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 202–207.
- [41] P. Penev and J. Atick, "Local feature analysis: a general statistical theory for object representation," *Network: Computation in Neural Systems*, vol. 7, pp. 477–500, 1996.

- [42] Ferdinando Samaria and Andy Harter, "Parameterisation of a stochastic model for human face identification," in *Proc. 2nd IEEE Workshop on Applications of Computer Vision*, Sarasota FL, Dec. 1994.
- [43] Ara V. Nefian and Monson H. Hayes, "Face recognition using an embedded HMM," in *Proc. International Conference On Audio- And Video-Based Biometric Person Authentication*, 1999, pp. 19–24.
- [44] P. Jonathon Phillips, Hyeonjoon Moon, Syed A. Rizvi, and Patrick J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.
- [45] *Face Recognition Vendor Test (FRVT)*, <<http://www.frvt.org/>>.
- [46] *AR face database*, <http://rvl.lcen.purdue.edu/aleix/aleix_face_DB.html>.
- [47] *MIT face database*, <<ftp://whitechapel.media.mit.edu/pub/images/>>.
- [48] *UMIST face database*, <<http://images.ec.umist.ac.uk/danny/database.html>>.
- [49] *CMU PIE face database*, <http://www.ri.cmu.edu/projects/project_418.html>.
- [50] *XM2VTS face database*, <<http://www.ec.surrey.ac.uk/Research/VSSP/xm2vtsdb/>>.
- [51] *FERET face database*, <<http://www.itl.nist.gov/iad/humanid/feret/>>.
- [52] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696–710, Jul. 1997.
- [53] P.J. Phillips, P. Grother, R.J. Micheals, D.M. Blackburn, E. Tabassi, and J.M. Bone, "FRVT 2002: Evaluation report," March 2003, <<http://www.frvt.org/FRVT2002/documents.htm>>.
- [54] A. S. Georgiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, Jun. 2001.
- [55] R. Basri and D. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, Feb. 2003.
- [56] Ralph Gross, Jianbo Shi, and Jeffrey Cohn, "Quo vadis face recognition?," in *Proc. Third Workshop on Empirical Evaluation Methods in Computer Vision*, Dec. 2001.