Quo vadis Face Recognition?

Ralph Gross Jianbo Shi Jeff Cohn

CMU-RI-TR-01-17

June 2001

Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

Abstract

Within the past decade, major advances have occurred in face recognition. Many systems have emerged that are capable of achieving recognition rates in excess of 90% accuracy under controlled conditions. In field settings, face images are subject to a wide range of variation that includes viewing, illumination, occlusion, facial expression, time delay between acquisition of gallery and probe images, and individual differences. The scalability of face recognition systems to such factors is not well understood. We quantified the influence of these factors, individually and in combination, on face recognition algorithms that included Eigenfaces, Fisherfaces, and FaceIt. Image data consisted of over 37,000 images from 3 publicly available databases that systematically vary in multiple factors individually and in combination: CMU PIE, Cohn-Kanade, and AR databases. Our main findings are: 1) pose variations beyond 30° head rotation substantially depressed recognition rate, 2) time delay: pictures taken on different days but under the same pose and lighting condition produced a consistent reduction in recognition rate, 3) with some notable exceptions, algorithms were robust to variation in facial expression, but not to occlusion. We also found small but significant differences related to gender, which suggests that greater attention be paid to individual differences in future research. Algorithm performance across a range of conditions was higher for women than for men.

Keywords: Face Recognition, Evaluation, Face database, Eigenface, Fisherface, FaceIt

Contents

1 Introduction

Within the past decade, major advances have occurred in face recognition. A large number of systems has emerged that are capable of achieving recognition rates of greater than 90% under controlled conditions. Successful application under real world conditions remains a challenge though. In field settings, face images are subject to a wide range of variations. These include pose or view angle, illumination, occlusion, facial expression, time delay between image acquisition, and individual differences. The scalability of face recognition systems to such factors is not well understood. Most research has been limited to frontal views obtained under standardized illumination on the same day with absence of occlusion and with neutral facial expression or slight smile. Relatively few studies, e.g., [22] have tested face recognition in the context of multiple views or explored related problems, e.g., [4]. Individual differences in subjects, such as whether accuracy is higher for one or another ethnic group, to our knowledge have not been studied.

Two notable exceptions to the homogeneity of testing conditions are the FERET competition and related studies [23] and the Facial Recognition Vendor Test [3]. In the period between August 1993 and July 1996 the FERET program collected 14,126 images from 1,199 individuals. For each subject two frontal views were recorded (sets **fa** and **fb**), where a different facial expression was requested for the fb image. For a subset of the subjects a third frontal image was taken using a different camera and under different illumination (set **fc**). A number of subjects were brought back at later dates to record "duplicate" images. For the **duplicate I** set the images were taken between 0 and 1,031 days after the initial recording (mean = 251 days). A subset of this set, the **duplicate II** set, contains images of subjects who returned between 540 and 1,031 days after the initial recording (mean = 627 days). In the final evaluation in 1996/1997 ten different algorithms, developed mostly by university research labs were evaluated. The test identified three algorithms as top performers: PCA-difference space from MIT [19], Fisher linear discriminant from the University of Maryland [33] and the Dynamic Link Architecture from the University of Southern California [31]. Furthermore the test provided a ranking of the difficulty of the different datasets in the FERET database. It was found that the **fb** set was the easiest and the duplicate II set the hardest, with the performance on the **fc** and duplicate I sets ranging in between these two.

One of the main goals of the Facial Recognition Vendor Test was the assessment of the capabilities of commercially available facial recognition systems. In the end three vendors, Visionics Corp., Lau Technologies and C-Vis completed the required tests in the given time. The imagery used in the evaluation spans a wide range of conditions: compression, distance, expression, illumination, media, pose, resolution and time. Pose was measured by asking subjects to rotate their head which was inexact. Subjects varied in their compliance and changes in expression often coocurred with head rotation. The pose variation was limited to a maximum of about $60°$. The most difficult conditions were temporal (11 to 13 months difference between recordings), pose and especially distance (change from 1.5m up to 5m). The top performing algorithm had few problems with the categories expression (regular vs. alternate), media (digital images vs. 35mm film), resolution (decreasing face sizes) and compression (up

to a factor of $30:1$). The illumination condition proved to be more difficult, especially when comparing subjects under indoor mug shot lighting with subjects recorded outside. In the majority of the experiments Visionics' FaceIt outperformed the other two vendors.

For faces to be a useful biometric, facial features used for face recognition should remain invariant to factors unrelated to person identity that modify face image appearance. While theory and some data suggest that many of these factors are difficult to handle, it is not clear where exactly the difficulties lie and what their causes may be. In this paper, we quantify the exact difficulties in face recognition as a function of variation in factors that influence face image acquisition and individual differences in subjects. We focus on six factors:

1. *Viewing angle*. The face has a 3D shape. As the camera pose changes, the appearance of the face can change due to a) projective deformation, which leads to stretching and foreshortening of different part of face, and b) self occlusion and dis-occlusion of parts of the face. If we have seen faces only from one viewing angle, in general it is difficult to recognize them from disparate angles. We investigate the functional relation between viewing angle and recognition and whether some viewing angles afford better or worse generalization to other viewing angles. To investigate these issues we use the CMU PIE database which densely samples viewing angles over an arc of 180° in the horizontal plane (from full profile left through frontal face to full profile right).

2. *Illumination*. Just as with pose variation, illumination variation is inevitable. Ambient lighting changes greatly within and between days and among indoor and outdoor environments. Due to the 3D shape of the face, direct lighting source can caste strong shadows and shading that accentuate or diminish certain facial features. Previous findings in the Facial Recognition Vendor Test suggest that illumination changes degrade recognition. However this finding is difficult to interpret. The effect of the illumination change in images can be due to either of two factors, 1) the inherent amount of light reflected off of the skin and 2) the non-linear adjustment in internal camera control, such as gamma correction, contrast, and exposure settings. Both can have major effects on facial appearance. While the latter is less noticeable for humans, it can cause major problems for computer vision. These factors were confounded in the Facial Recognition Vendor Test. In our study, we will focus on reflectance from the skin, which we refer to as illumination, using the well sampled illumination portion of the PIE database. We evaluate main effects and interactions between illumination and viewing angle and other factors.

3. *Expression*. The face is a non-rigid object. Facial expression of emotion and paralinguistic communication along with speech acts can and do produce large variation in facial appearance. The number of possible changes in facial expression is reportedly in the thousands. The influence of facial expression on recognition is not well understood. Previous research has been limited primarily to neutral expressions and slight smiles. Because facial expression affects the apparent geometrical shape and position of the facial features, the influence on

recognition may be greater for geometry based algorithms than for holistic algorithms. We use the Cohn-Kanade facial expression database to investigate these issues. This database samples well characterized emotions and expressions. We will ask the questions: 1) does facial expression pose a problem for most facial recognition system and 2) if so, what are the challenging expressions? We investigate the conditions under which facial expression may either impair or improve face recognition.

4. *Occlusion*. The face may be occluded by other objects in the scene or by sunglasses or other paraphernalia. Occlusion may be unintentional or intentional. Under some conditions subjects may be motivated to thwart recognition efforts by covering portions of their face. Since in many situations, our goal is to recognize non- or even un-cooperating subjects, we would like to know how difficult it is to recognize people given certain quantitative and qualitative changes in occlusion. We examine under which conditions such efforts may or may not be successful. To investigate occlusion we use the AR database, which has two different types of facial occlusion, one for the eyes, and one for the lower face.

5. *Time delay*. Faces change over time. There are changes in hair style, makeup, muscle tension and appearance of the skin, presence or absence of facial hair, glasses, or facial jewelry, and over longer periods effects related to aging. We use the AR database to investigate the effects of time delay and interactions between time delay and expression, illumination, and occlusion.

6. *Individual factors*. Algorithms may be more or less sensitive for men or women or members of different ethnic groups. We focus on the differences between men and women with respect to algorithm performance. Intuitively, females might be harder to recognize because of greater use and day-to-day variation in makeup or in structural facial features. Male and female faces differ in both local features and in shape [5]. Men's faces on average have thicker eyebrows and greater texture in the beard region. In women's faces, the distance between the eyes and brows is greater, the protuberance of the nose smaller, and the chin narrower than in men [5]. People readily distinguish male from female faces using these and other differences (e.g., hair style), and connectionist modeling has yielded similar results [6, 17]. Little is known, however, about the sensitivity of face identification algorithms to differences between men's and women's faces. The relative proportions of men and women in training samples are seldom reported, and identification results typically fail to mention whether algorithms are more or less accurate for one sex or the other. Other factors that may influence identification, such as differences in face shape between individuals of European, Asian, and African ancestry [5, 8], have similarly been ignored in past research. To address this issue, we will use both the AR database, which has well balanced proportions of men and women in the database, and FERET, which has a much large number of subjects.

This paper is organized as follows. In Section 2, we describe the three databases, containing 37,954 images, that form the basis of our experiments. In Section 3 we

describe face recognition systems used in this report: (1) Eigenfaces, similar to Turk and Pentland [29], which provides an essential benchmark, (2) Fisherfaces using Fisher linear discriminants similar to Belhumeur et al. [2], and (3) FaceIt, a leading commercially available face recognition system from Visionics. Eigen- and Fisherfaces are widely known and present common benchmarks for evaluating performance of other face recognition algorithms. FaceIt was the system with the best overall performance in the Facial Recognition Vendor Test and serves as an example of state-of-the-art performance in face recognition. In Section 4 we present results of each experiment. Conclusions and discussion are presented in Section 5.

2 Description of Databases

2.1 Overview

We use images from three publicly available databases in our evaluation. Table 1 gives an overview of the CMU PIE, Cohn-Kanade and the AR database.

	CMU PIE	Cohn-Kanade	AR DB
Subjects	68	105	116
Poses	13	1	1
Illuminations	43	3	3
Expressions	3	6	3
Occlusion	0	0	2
Sessions	1	1	2
Number of images	41,368	1424	3288

Table 1: Overview of the databases used in the evaluation.

2.2 CMU Pose Illumination Expression (PIE) database

The CMU PIE database contains a total of 41,368 images taken from 68 individuals [27]. The subjects were imaged in the CMU 3D Room [13] using a set of 13 synchronized high-quality color cameras and 21 flashes. The resulting images are 640x480 in size, with 24-bit color resolution. The setup of the room with the camera and flash locations is shown in Figure 1. The images of a subject across all 13 poses is shown in Figure 2.

Each subject was recorded under 4 conditions:

1. *Expression*: the subjects were asked to display a neutral face, to smile, and to close their eyes in order to simulate a blink. The images of all 13 cameras are available in the database.

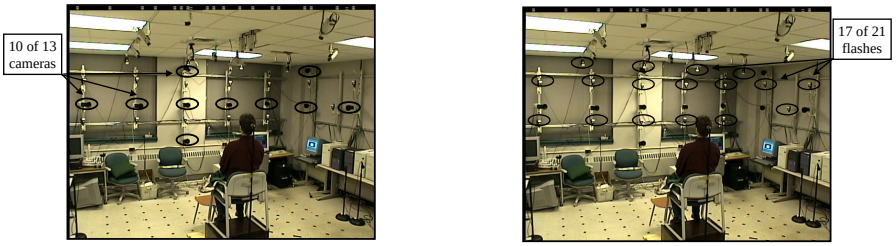
2. *Illumination 1*: 21 flashes were individually turned on in a rapid sequence. In the first setting the images were captured with the room lights on. Each camera

recorded 24 images, 2 with no flashes, 21 with one flash firing and then a final image with no flashes. Only the output of three cameras (frontal, three-quarter and profile view) was kept.

3. *Illumination 2*: the procedure for the *illumination 1* was repeated with the room lights off. The output of all 13 cameras was retained in the database. Combining the two illumination settings, a total of 43 different illumination conditions were recorded.

4. *Talking*: subjects counted starting at 1. 2 seconds (60 frames) of them talking were recorded using 3 cameras as above (again frontal, three-quarter and profile view).

Figure 3 shows examples for 12 different illumination conditions across three poses.



(a) Camera positions (b) Flash positions

Figure 1: Pictures of the CMU 3D room setup. 10 of the 13 cameras are indicated in (a). (b) shows 17 of the 21 flash locations.

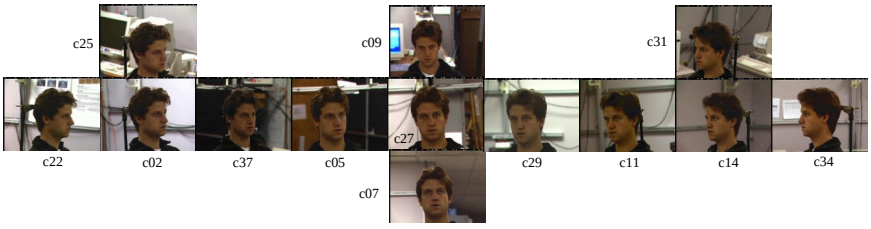


Figure 2: Pose variation in the PIE database [27]. The pose varies from full left profile (c34) to full frontal (c27) and on to full right profile (c22). The 9 cameras in the horizontal sweep are each separated by about 22.5°. The 4 other cameras include 1 above (c09) and 1 below (c07) the central camera, and 2 in the corners of the room (c25 and c31), typical locations for surveillance cameras.

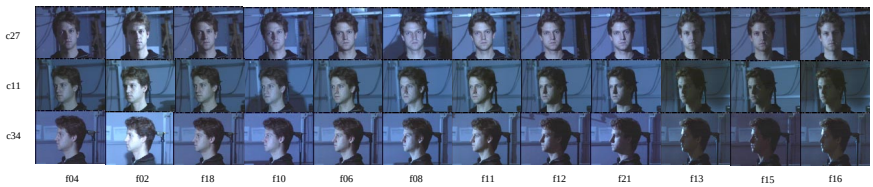


Figure 3: Illumination variation in the PIE database. The figure shows twelve flash conditions across three head poses.

2.3 Cohn-Kanade AU-Coded Facial Expression Database

This is a publicly available database from Carnegie Mellon University [12]. It contains image sequences of facial expression from men and women of varying ethnic backgrounds. The camera orientation is frontal. Small head motion is present. Image size is 640 by 480 pixels with 8-bit gray scale resolution. There are three variations in lighting: ambient lighting, single-high-intensity lamp, and dual high-intensity lamps with reflective umbrellas. Facial expressions are coded using the Facial Action Coding System [7] and also assigned emotion-specified labels. For the current study, we selected a total of 1424 images from 105 subjects. Emotion expressions included happy, surprise, anger, disgust, fear, and sadness. Examples for the different expressions are shown in Figure 4.

Figure 4: Cohn-Kanade AU-Coded Facial Expression database. Examples of emotion-specified expressions from image sequences.

2.4 AR Face Database

The publicly available AR database was collected at the Computer Vision Center in Barcelona [18]. It contains images of 116 individuals (63 males and 53 females). The images are 768x576 pixels in size with 24-bit color resolution. The subjects were recorded twice at a 2-week interval. During each session 13 conditions with varying facial expressions, illumination, and occlusion were captured. Figure 5 shows an example for each condition.

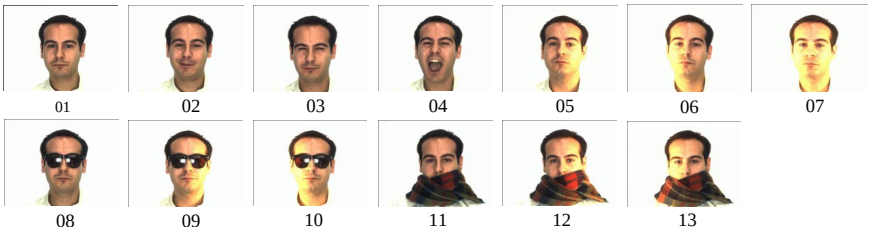


Figure 5: AR database. The conditions are: (1) neutral, (2) smile, (3) anger, (4) scream, (5) left light on, (6) right light on, (7) both lights on, (8) sun glasses, (9) sun glasses/left light (10) sun glasses/right light, (11) scarf, (12) scarf/left light, (13) scarf/right light

3 Face Recognition Algorithms

Most of the current face recognition algorithms can be categorized into two classes, image template based or 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. Statistical tools such as Support Vector Machines (SVM) [30, 20], Linear Discriminant Analysis (LDA) [2], Principal Component Analysis (PCA) [28, 29], Kernel Methods [26, 16], and Neural Networks [25, 11, 15] have been used to construct a suitable set of face templates. While these templates can be viewed as features, they mostly capture global features of the face images. 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 et al. have presented an active shape model in [14] extending the approach by Yuille [32].Wiskott et al. developed an elastic Bunch graph matching algorithm for face recognition in [31]. Penev et. al [21] developed PCA into Local Feature Analysis (LFA). This technique is the basis for one of the most successful commercial face recognition systems, FaceIt. The following sections describe the algorithms that are used in our experiments in more detail.

3.1 Principal Component Analysis

Principal Component Analysis (PCA) is a method for the unsupervised reduction of dimensionality. Assume that a set of N sample images $\{x_1, x_2, \ldots, x_N\} \in \Re^n$ is given. Each image belongs to one of m classes $\{C_1, C_2, \ldots, C_m\}$. We define the *total scatter* matrix S_T as

$$S_T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T$$

where μ is the mean of the data. PCA determines the orthogonal projection Φ in

$$y_k = \Phi^T x_k, k = 1, \ldots, N$$

that maximizes the determinant of the total scatter matrix of the projected samples $y_1, \ldots, y_N$:

$$\Phi_{opt} = \arg\max_{\Phi} \mid \Phi^T S_T \Phi \mid$$

This scatter stems from *inter-class* variations between the objects, as well as from *intra-class* variation within the object classes. Most of the differences between faces are due to external factors such as viewing direction and illumination. As PCA does not differentiate between inter-class and intra-class variation it fails to discriminate well between object classes.

3.2 Linear Discriminant Analysis

A alternative approach is Fisher's Linear Discriminant (FLD) [9], also known as Linear Discriminant Analysis (LDA) [33], which uses the available class information to compute a projection better suited for discrimination tasks. We define the *within-class* scatter matrix S_W as

$$S_W = \sum_{i=1}^{m} \sum_{x_k \in C_i} (x_k - \mu_i)(x_k - \mu_i)^T$$

where μ_i is the mean of class i. Furthermore we define the *between-class* scatter matrix S_B as

$$S_B = \sum_{i=1}^{N} N_i(\mu_i - \mu)(\mu_i - \mu)^T$$

where N_i refers to the number of samples in class i. LDA computes the projection Ψ that maximizes the ratio

$$\Psi_{opt} = \arg\max_{\Psi} \frac{\mid \Psi^T S_B \Psi \mid}{\mid \Psi^T S_W \Psi \mid}$$

Ψ_{opt} is found by solving the generalized eigenvalue problem

$$S_B \Psi = \lambda S_W \Psi$$

Due to the structure of the data the within-class scatter matrix S_W is always singular. We can overcome this problem by first using PCA to reduce the dimensionality and then applying LDA [2]. The overall projection is therefore given by $W_{opt}^T = \Psi_{opt}^T \Phi_{opt}^T$.

3.3 The Classification Algorithm

In order to determine the closest gallery vector for each probe vector we perform nearest neighbor classification using the Mahalanobis distance metric in the PCA and LDA subspaces. For input vectors μ and λ the Mahalanobis distance is defined as

$$d_M(\mu, \lambda) = (\mu - \lambda)^T \Sigma^{-1} (\mu - \lambda)$$

where Σ^{-1} is the inverse of the data covariance matrix.

3.4 FaceIt

FaceIt's recognition module is based on Local Feature Analysis (LFA) [21]. This technique addresses two major problems of Principal Component Analysis. The application of PCA to a set of images yields a global representation of the image features that is not robust to variability due to localized changes in the input [10]. Furthermore the PCA representation is non topographic, so nearby values in the feature representation do not necessarily correspond to nearby values in the input. LFA overcomes these problems by using localized image features in form of multi-scale filters. The feature images are then encoded using PCA to obtain a compact description. According to Visionics, FaceIt is robust against variations in lighting, skin tone, eye glasses, facial expression and hair style. They furthermore claim to be able to handle pose variations of up to 35 degrees in all directions. We systematically evaluated these claims.

4 Evaluation

Following Phillips et. al. [24] we distinguish between *gallery* and *probe* images. The gallery contains the images of known individuals against which unknown images are matched. The algorithms are tested with the images in the probe sets. All results reported here are based on non-overlapping gallery and probe sets (with the exception of the PIE pose test). We use the *closed universe* model for evaluating the performance, meaning that every individual in the probe set is also present in the gallery.

4.1 Face Localization and Registration

Face recognition is a two step process consisting of face detection and recognition. First, the face has to be located in the image and registered against an internal model. The result of this stage is a normalized representation of the face, which the recognition algorithm can be applied to. While FaceIt has its own face finding module we have to provide normalized images to PCA and LDA. We manually labeled the x-y positions of both eyes (pupils) and the tip of the nose in all images used in the experiments. Within each condition separately the face images are normalized for rotation, translation, and scale. The face region is then tightly cropped using the normalized feature point distances. Figure 6 shows the result of face region extraction for two cameras (c27 and c37) of the PIE database.

4.2 Generic Training Data

For the construction of the PCA and LDA representations we randomly select half of the subjects in each evaluation condition as generic training data. During this stage both algorithms are presented with images from all gallery and probe conditions. The testing is then done on the set of remaining subjects with non-overlapping gallery and probe sets. As FaceIt is already fully trained we report results over the full dataset with all subjects for all evaluation conditions.

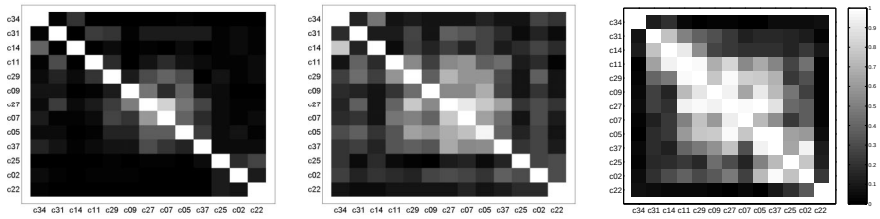
Figure 6: Face normalization. The original images from camera views c27 and c37 are shown together with the normalized and cropped face region.

4.3 Pose with Constant Illumination

Using the CMU PIE database we evaluate the performance of face recognition algorithms with respect to pose variations in great detail. We exhaustively sampled the pose space by using each pose in turn as gallery with the remaining poses as probes.

	Pose
Gallery	Each of 13 pose images in PIE, with room lighting.
Probe	All 13 pose images in PIE, with room lighting.

Figure 7 visualizes the confusion matrix for PCA, LDA and FaceIt. The numerical results for FaceIt are listed in Table 2.



(a) PCA (b) LDA (c) FaceIt

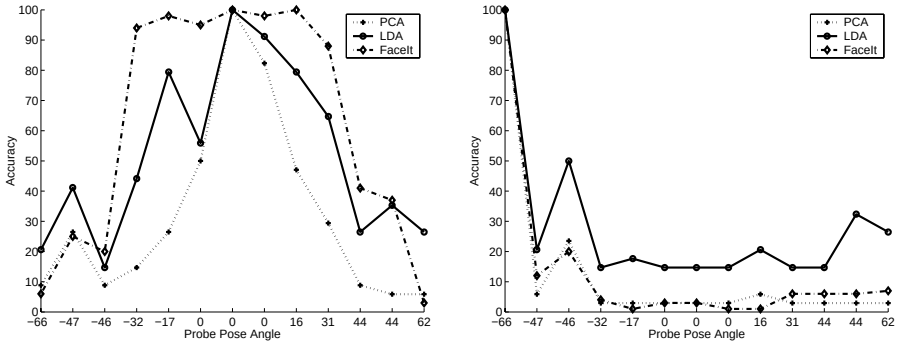
Figure 7: Comparison of the pose confusion matrix for PCA, LDA and FaceIt. The gallery poses (see Figure 2) are shown along the x-axis, the probe poses along the y-axis.

Of particular interest is the question how far the algorithm can generalize from given gallery poses. For a frontal gallery pose, the recognition rate of FaceIt drops rapidly below 90% for head rotation beyond 32° (corresponds to camera positions 11 and 37 in Figure 2), and the recognition rate of LDA drops below 80% for head rotation

β α Probe Pose	-66 3 c34	-47 13 c31	-46 2 c14	-32 2 c11	-17 2 c29	0 15 c09	0 2 c27	0 1.9 c07	16 2 c05	31 2 c37	44 2 c25	44 13 c02	62 3 c22
Gallery Pose													
c34	**1.00**	0.03	0.01	0.00	0.00	0.03	0.04	0.00	0.01	0.03	0.01	0.00	0.01
c31	0.01	**1.00**	0.12	0.16	0.15	0.09	0.04	0.06	0.04	0.03	0.06	0.00	0.01
c14	0.04	0.16	**1.00**	0.28	0.26	0.16	0.19	0.10	0.16	0.04	0.03	0.03	0.01
c11	0.00	0.15	0.29	**1.00**	**0.78**	0.63	**0.73**	0.50	0.57	0.40	0.09	0.01	0.03
c29	0.00	0.13	0.22	**0.87**	**1.00**	**0.75**	**0.91**	**0.73**	0.68	0.44	0.03	0.01	0.03
c09	0.03	0.01	0.09	0.68	**0.79**	**1.00**	**0.95**	0.62	**0.87**	0.57	0.09	0.01	0.01
c27	0.03	0.07	0.13	**0.75**	**0.93**	**0.94**	**1.00**	**0.93**	**0.93**	0.62	0.06	0.03	0.03
c07	0.01	0.07	0.12	0.38	**0.70**	0.57	**0.87**	**1.00**	**0.73**	0.35	0.03	0.03	0.00
c05	0.01	0.03	0.13	0.54	0.65	**0.75**	**0.91**	**0.75**	**1.00**	0.66	0.09	0.01	0.03
c37	0.00	0.03	0.04	0.37	0.35	0.43	0.53	0.23	0.60	**1.00**	0.10	0.04	0.00
c25	0.00	0.01	0.01	0.06	0.04	0.07	0.04	0.03	0.06	0.07	**0.98**	0.04	0.04
c02	0.00	0.01	0.03	0.03	0.01	0.01	0.01	0.04	0.01	0.01	0.04	**1.00**	0.03
c22	0.00	0.01	0.01	0.01	0.01	0.03	0.03	0.03	0.03	0.04	0.03	0.00	**1.00**

Table 2: Confusion table for pose variation. Each row of the confusion table shows the recognition rate on each of the probe poses given a particular gallery pose.

beyond 17°. Furthermore, for most non-frontal poses, face generalizability goes down drastically, even for close-by poses. This can be seen in more detail in Figure 8.



(a) Frontal pose as gallery (b) Profile pose as gallery

Figure 8: Generalizability varies with gallery pose. The frontal pose has good generalizability up to 32° for FaceIt and up to 17° for LDA. For the profile view the performance is low everywhere outside the gallery pose.

We then asked the question, if we can improve the performance by providing multiple face poses in the gallery set? Intuitively, given multiple face poses, with correspondence between the facial features, one can have a better chance of predicting novel face poses. In our experiments with FaceIt we did not find any evidence of an additional gain through multiple face gallery poses. This suggests that 3D face recognition approaches could have an advantage over naive integration of multiple face poses, such as in the proposed 2D statistical SVM or related non-linear Kernel methods.

4.4 Illumination with Frontal Pose

For this test, the PIE and AR databases are used. As described in Section 2.2 the PIE database contains two illumination sets. In *Illumination 1*, images were taken with the room lights on, whereas for the *Illumination 2* set the images were captured with the room lights turned off.

PIE Illumination 1	
Gallery	Frontal pose, room illumination without flash
Probe	Frontal pose, room illumination with 21 flash conditions
PIE Illumination 2	
Gallery:	Frontal pose, frontal flash illumination, no room light
Probe:	Frontal pose, all 21 flash conditions, no room light
AR database	
Gallery:	Frontal pose, room illumination
Probe:	Frontal pose, illumination from left, right and from both directions

Table 3 shows the recognition accuracies of the algorithms in each of the experiments. The results on the PIE database are consistent with the outcome of the experiments on the AR database. Overall, the performance of FaceIt and Fisherfaces are acceptable in most of the illumination conditions. The overall trend is that the PIE Illumination 1 experiment is the easiest, the AR experiments are slightly more difficult, and PIE Illumination 2 is the most difficult. The result is understandable as in a large number of Illumination 2 images, significant portions of the faces are invisible, see Figure 3.

	 PIE 1	 PIE 2	 ARDB 05	 ARDB 06	 ARDB 07
PCA	0.89	0.61	0.81	0.79	0.82
LDA	0.96	0.69	0.87	0.82	0.86
FaceIt	1.0	0.91	0.96	0.93	0.86

Table 3: Illumination results. PIE 1 and 2 refer to the two illumination conditions described in Section 2.2. AR05, AR06, AR07 are the left, right, both light on conditions in the AR database as shown in Figure 5.

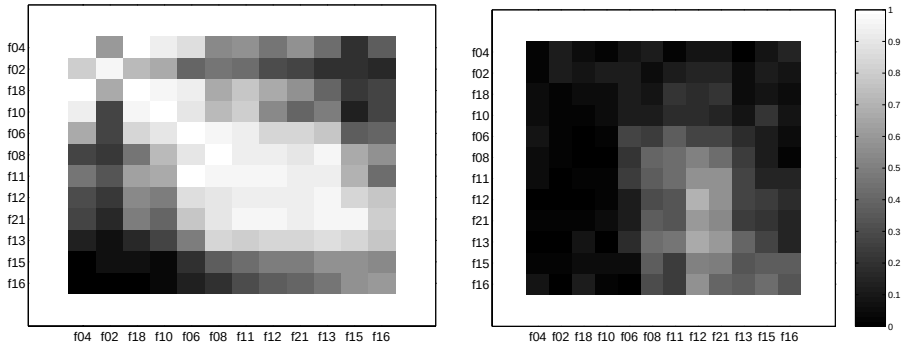
While these results may lead one to conclude that face recognition under illumination is a solved problem, we would like to caution that the illumination change may still cause a major problem when it is coupled with other changes (expression, pose, etc.).

4.5 Pose and Illumination Combined

To test this we evaluated the combined effect of pose and illumination changes on the performance of FaceIt.

Pose and Illumination	
Gallery:	Three PIE poses 05,27,29 (frontal), 12 flash conditions from the Illumination 2 set.
Probe:	PIE poses 02 (right profile) and 07 (lower frontal), 12 flash conditions from the Illumination 2 set.

Figure 9 shows the illumination confusion matrices for FaceIt. We see in Figure 9(b) that, for the right profile pose, lighting from the left produces recognition failures, since most of the face will be invisible. In Figure 9(a) we see that, while for frontal pose the lighting conditions have better generalizibility, far apart lighting angles cause difficulties for FaceIt. In separation, frontal-to-frontal recognition and recognition across illumination are well handled. However, when coupled they can cause significant degradation in face recognition accuracy.



(a) Probe 07 Pose
lower frontal

(b) Probe 02 Pose
right profile

Figure 9: FaceIt illumination-to-illumination confusion matrix for two camera probe poses. The gallery illumination conditions are shown along the x-axis, the probe illumination conditions across the y-axis. The flashes are sorted from far left to far right (as seen from the subject).

4.6 Expression with Frontal Pose and Constant Illumination

Faces undergo large deformations under facial expressions. Humans can easily handle this variation, but we expected the algorithms to have problems with the expression databases. Table 4 shows the results of the 3 algorithms in this experiments. To our surprise FaceIt and LDA performed very well on the Cohn-Kanade and the AR database, with the notable exception of the *scream* (AR04) set of the AR database. For

Expression	
Gallery:	Cohn-Kanade/AR, frontal pose, room illumination, neutral expression
Probe:	Cohn-Kanade/AR, frontal pose, room illumination, expressions

most facial expressions, the facial deformation is centered around the lower part of the face. This might leave sufficient invariant information in the upper face for recognition, which results in a high recognition rate. The expression "scream" has effects on both the upper and the lower face appearance, which leads to a significant fall off in the recognition rate. This indicates that 1) face recognition under extreme facial expression still remains an unsolved problem, and 2) temporal information can provide significant additional information in face recognition under expression.

	 Cohn-Kanade	 AR 02	 AR 03	 AR 04
PCA	0.78	0.87	0.86	0.39
LDA	0.97	0.96	0.89	0.60
FaceIt	0.97	0.96	0.92	0.76

Table 4: Expression results. AR 02, AR 03 and AR 04 refer to the expression changes in the AR database as shown in Figure 5. All three algorithms perform reasonably well under facial expression, however the "scream" expression, AR 04, produces large recognition errors.

4.7 Occlusion with Frontal Pose and Three Illumination Conditions

For the occlusion tests we look at images where parts of the face are invisible for the camera. The AR database provides two scenarios: subjects wearing sun glasses and subjects wearing a scarf around the lower portion of the face. The recognition rates for

Occlusion	
Gallery:	Frontal pose, room illumination, no occlusion
Probe:	Frontal pose, one of three illumination conditions, sunglasses or scarf

the sun glass images, as shown in Table 5, are according to expectations: it is difficult for face recognition system. The result further deteriorates when the left or right light is switched on (AR09 and AR10). Furthermore, the test reveals that FaceIt is more vulnerable to upper face occlusion than either PCA or LDA. Facial occlusion, particularly upper face occlusion, remains a difficult problem yet to be solved. Interesting open questions are 1) what are the fundamental limits of any recognition system under

various occlusions, and 2) to what extend can other additional facial information, such as motion, provide the necessary help for face recognition under occlusion.

	 AR 08	 AR 09	 AR 10	 AR 11	 AR 12	 AR 13
PCA	0.48	0.26	0.21	0.27	0.21	0.11
LDA	0.45	0.31	0.27	0.44	0.33	0.31
FaceIt	0.10	0.09	0.06	0.81	0.72	0.72

Table 5: Occlusion results. AR08, AR09, AR10 refer to the upper facial occlusions, and AR11, AR12, AR13 refer to the lower facial occlusions as shown in Figure 5. Upper facial occlusion causes a major drop in recognition rates.

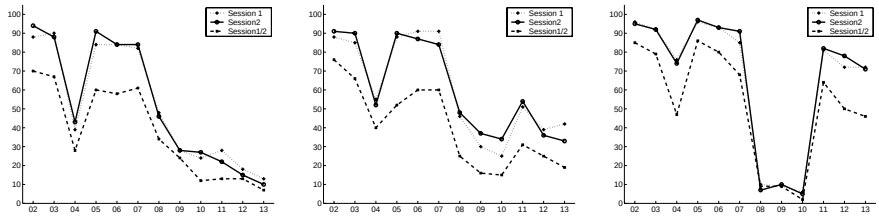
4.8 Time Delay

Figure 10 shows the performance of all three algorithms across AR database conditions for three different gallery/probe configurations. For the *session 1* and *session 2* curves the gallery and probe images were taken from the same recording session. In the majority of conditions these two curves are identical. The third curve labeled *session 1/2* shows the performance for running FaceIt with the neutral image of session 1 as gallery and the images of session 2 as probe. Even though the images for the two sessions were recorded only two weeks apart, the recognition performance degrades visibly across all conditions. This drop in performance is observable for all three algorithms.

Time Delay: Session 1	
Gallery:	Condition 01 first recording session
Probe:	Conditions 02-13 first recording session
Time Delay: Session 2	
Gallery:	Condition 01 second recording session
Probe:	Conditions 02-13 second recording session
Time Delay: Session 1/2	
Gallery:	Condition 01 first recording session
Probe:	Conditions 02-13 second recording session

4.9 Gender

We evaluated the influence of gender on face recognition algorithms on the AR database due to its balanced ratio between the female and male subjects. The results reveal a surprising trend: better recognition rates are consistently achieved for female subjects. Averaged across the conditions (excluding the tests AR08-10 where FaceIt breaks down)



(a) PCA (b) LDA (c) FaceIt

Figure 10: Results on the ARDB database for multiple sessions for PCA, LDA and FaceIt. Excluding the three worst conditions the performances drop by 18.1% for PCA, 21.9% for LDA and 18.2% for FaceIt.

the recognition rate for male subjects is 83.4%, while the recognition rate for female subjects is 91.66%. We replicated this test over the much larger FERET database (1,199 subjects). For the **fb** set FaceIt achieves a recognition rate of 93.7% for female subjects and 87.6% for male subjects. The difference in performance for male and female subjects is statistically significant (chisquare, $p = 0.0006$). This opens up many interesting questions on face recognition. In particular it raises the questions: 1) what makes one face easier to recognize than another, and 2) are there face classes with similar recognizability.

5 Discussion

To summarize the results in previous experiments, we see that:

1. *Pose:* Pose variation still presents a challenge for face recognition. Frontal training images have better generalizability to novel poses than do non-frontal training images. For a frontal training pose, we can achieve reasonable recognition rates of above 90% for 32° head rotation. In field applications, however, even this range of viewing angles may prove insufficient. Security cameras often are positioned near ceilings and corners, thus creating viewing angles that are outside of the effective limits we observed.

2. *Illumination:* Pure illumination changes on the face are handled well by current face recognition algorithms. However, face recognition systems have difficulties in extreme illumination conditions in which significant parts of the face are invisible. Furthermore, it can become particularly difficult when illumination is coupled with pose variation. Our findings for illumination are seemingly at variance with those of the Facial Recognition Vendor Test. In the latter, illumination was a significant problem for the algorithms. Because illumination and non-linear variation in camera characteristics were confounded in the FRVT, our results suggest that it was non-linear camera characteristics that were primarily responsible for the effects they interpreted as due to illumination.

3. *Expression:* With the exception of extreme expressions such as scream, the algorithms are relatively robust to facial expression. Deformation of the mouth and occlusion of the eyes by eye narrowing and closing present a problem for the algorithms.

4. *Occlusion:* The performance of the face recognition algorithms under occlusion is in general poor. There are however important differences among algorithms in this regard. FaceIt proves robust with respect to lower face occlusion but fails with upper-face occlusion. PCA and LDA show the opposite pattern. These findings suggest that optimal results might be achieved by combining features of different approaches.

5. *Time delay between gallery and probe images:* Time delay between acquisition of gallery and probe images can cause degradation in face recognition performance. In the AR database, with recording sessions just 2 weeks apart, we see a significant difference of about 20% in recognition rate. The effects of time are likely to be non-linear over longer periods of change with development. The accuracy of recognition algorithms in children and across developmental periods (e.g., childhood to adolescence) to our knowledge remains unexplored.

6. *Gender:* We found surprisingly consistent differences of face recognition rates related to gender. In two databases (AR and FERET) the recognition rate for female subjects is higher than for males across a range of perturbations. One hypothesis is that women invest more effort into modifying their facial appearance, by use of cosmetics, for instance, which leads to greater differentiation among women than men. Alternatively, algorithms may simply be more sensitive to structural differences between the faces of women and men. The finding that algorithms are more sensitive to women's faces suggests that there may be other individual differences related to algorithm performance. Algorithms may, for instance, prove more accurate for some ethnic groups or ages than others.

   These experiments in total show that challenging problems remain in face recognition. Pose, occlusion, and time delay variation in particular present the most difficulties.

   While our study has revealed many challenges for current face recognition research, the current study has several limitations. One, we did not examine the effect of face image size on algorithm performance in the various conditions. Minimum size thresholds may well differ for various permutations, which would be important to determine. Two, the influence of racial or ethnic differences on algorithm performance could not be examined due to the homogeneity of racial and ethnic backgrounds in the databases. While large databases with ethnic variation are available, they lack the parametric variation in lighting, shape, pose and other factors that were the focus of this investigation. Three, faces change dramatically with development, but the influence of change with development on algorithm performance could not be examined. Fourth, while we were able to examine the combined effects of some factors, databases are needed that support examination of all ecologically valid combinations, which may be non-additive.

The results of the current study suggest that greater attention be paid to the multiple sources of variation that are likely to affect face recognition in natural environments.

Acknowledgements

We wish to thank Takeo Kanade, Simon Baker, Iain Matthews and Henry Schneiderman for useful feedback, Peter Metes for labeling the data and Aleksandra Slavkovic for the statistical evaluations. The research described in this paper was supported by U.S. Office of Naval Research contract N00014-00-1-0915. Portions of the research in this paper use the FERET database of facial images collected under the FERET program.

References

[1] Robert J. Baron. Mechanisms of human facial recognition. *International Journal of Man-Machine Studies*, 15(2):137–178, 1981.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.

[3] D.M. Blackburn, M. Bone, and P.J. Philips. Facial recognition vendor test 2000: evaluation report, 2000.

[4] V. Blanz, S. Romdhani, and T. Vetter. Face identification across different poses and illumination with a 3d morphable model. In *5th International Conference on Automatic Face and Gesture Recognition*, 2002.

[5] V. Bruce and A. Young. *In the eye of the beholder: The science of face perception.* Oxford University Press, 1998.

[6] G.W. Cottrell and J. Metcalfe. Empath: Face, emotion, and gender recognition using holons. In R.P. Lippmann, J.E. Moody, and D.S. Touretzky, editors, *Neural information processing systems*, volume 3, pages 53–60, San Mateo, CA, 1991. Morgan Kaufmann.

[7] P. Ekman and W.V. Friesen. *Facial Action Coding System.* Consulting Psychologist Press, 1978.

[8] L.G. Farkas and I.R. Munro. *Anthropometric facial proportions in medicine.* C.C. Thomas, Springfield, IL, 1987.

[9] K. Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, 1990.

[10] R. Gross, J. Yang, and A. Waibel. Face recognition in a meeting room. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, 2000.

[11] A. Jonathan Howell and Hilary Buxton. Invariance in radial basis function neural networks in human face classification. *Neural Processing Letters*, 2(3):26–30, 1995.

[12] T. Kanade, J.F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 46–53, Grenoble, France, 2000.

[13] T. Kanade, H. Saito, and S. Vedula. The 3d room: Digitizing time-varying 3d events by synchronized multiple video streams. Technical Report CMU-RI-TR-98-34, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, December 1998.

[14] Andreas Lanitis, Christopher J. Taylor, and Timothy Francis Cootes. Automatic interpretation and coding of face images using flexible models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):743–756, 1997.

[15] Steve Lawrence, C. Lee Giles, Ah Chung Tsoi, and Andrew D. Back. Face recognition: A convolutional neural network approach. *IEEE Transactions on Neural Networks*, 8(1):98–113, 1998.

[16] Y. Li, S. Gong, and H. Liddell. Support vector regression and classification based multi-view face detection and recognition. In *IEEE International Conference on Automatic Face and Gesture Recognition*, March 2000.

[17] A.J. Luckman, N.M Allison, A. Ellis, and B.M. Flude. Familiar face recognition: A comparative study of a connectionist model and human performance. *Neurocomputing*, 7:3–27, 1995.

[18] A. R. Martinez and R. Benavente. The AR face database. Technical Report 24, Computer Vision Center(CVC) Technical Report, Barcelona, Spain, June 1998.

[19] Baback Moghaddam and Alex Paul Pentland. Probabilistic visual learning for object representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):696–710, 1997.

[20] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: An application to face detection, 1997.

[21] P. Penev and J. Atick. Local feature analysis: A general statistical theory for object representati on, 1996.

[22] A.P. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. In *Proceedings of the 2001 Conference on Computer Vision and Pattern Recognition*, 1994.

[23] P. Jonathon Phillips, Harry Wechsler, Jeffrey S. Huang, and Patrick J. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.

[24] P.J. Phillips, H. Moon, S. Rizvi, and P.J Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transations on PAMI*, 22(10):1090–1104, 2000.

[25] T. Poggio and K.-K. Sung. Example-based Learning for View-based Human Face Detection. In *1994 ARPA Image Understanding Workshop*, volume II, November 1994.

[26] B. Schoelkopf, A. Smola, and K.-R. Muller. Kernal principal component analysis. In *Artificial Neural Networks ICANN97*, 1997.

[27] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression (PIE) database. In *Proc. of the 5th IEEE International Conference on Automatic Face and Gesture Recognition*, 2002.

[28] 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.

[29] Matthew Turk and Alex Paul Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[30] Vladimir N. Vapnik. *The nature of statistical learning theory*. Springer Verlag, Heidelberg, DE, 1995.

[31] Laurenz Wiskott, Jean-Marc Fellous, Norbert Krüger, and Christoph von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775–779, July 1997.

[32] Alan L. Yuille. Deformable templates for face recognition. *Journal of Cognitive Neuroscience*, 3(1):59–70, 1991.

[33] W. Zhao, R. Chellappa, and A. Krishnaswamy. Discriminant analysis of principal components for face recognition. In *Proceedings of the 3rd International Conference on Automatic Face and Gesture Recognition*, pages 336–341, 1998.