# Quo vadis Face Recognition?

Ralph Gross, Jianbo Shi
Robotics Institute
Carnegie Mellon University
Pittsburgh, PA 15213
{rgross,jshi}@cs.cmu.edu

Jeffrey F. Cohn
Department of Psychology
University of Pittsburgh
Pittsburgh, PA 15260
jeffcohn@pitt.edu

## Abstract

*Within the past decade, major advances have occurred in face recognition. With few exceptions, however, most research has been limited to training and testing on frontal views. Little is known about the extent to which face pose, illumination, expression, occlusion, and individual differences, such as those associated with gender, influence recognition accuracy. We systematically varied these factors to test the performance of two leading algorithms, one template based and the other feature based. Image data consisted of over 21000 images from 3 publicly available databases: CMU PIE, Cohn-Kanade, and AR databases. In general, both algorithms were robust to variation in illumination and expression. Recognition accuracy was highly sensitive to variation in pose. For frontal training images, performance was attenuated beginning at about 15 degrees. Beyond about 30 degrees, performance became unacceptable. For non-frontal training images, fall off was more severe. Small but consistent differences were found for individual differences in subjects. These findings suggest direction for future research, including design of experiments and data collection.*

## 1. Introduction

Is face recognition a solved problem? Over the last 30 years face recognition has become one of the best studied pattern recognition problems with a nearly intractable number of publications. Many of the algorithms have demonstrated excellent recognition results, often with error rates of less than 10 percent. These successes have led to the development of a number of commercial face recognition systems. 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, 21], Linear Discriminant Analysis (LDA) [2], Principal Component Analysis (PCA) [27, 29, 11], Kernel Methods [25, 17], and Neural Networks [24, 7, 12, 16] 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 [15] extending the approach by Yuille [34].Wiskott et al. developed an elastic Bunch graph matching algorithm for face recognition in [33]. Penev et. al [22] developed PCA into Local Feature Analysis (LFA). This technique is the basis for one of the most successful commercial face recognition systems, FaceIt.

Most face recognition algorithms focus on frontal facial views. However, pose changes can often lead to large non-linear variation in facial appearance due to self-occlusion and self-shading. To address this issue, Moghaddam and Pentland [20] presented a Bayesian approach using PCA as a probability density estimation tool. Li et al. [17] have developed a view-based piece-wise SVM model for face recognition. In the feature based approach, Cootes et al. [5] proposed a 3D active appearance model to explicitly compute the face pose variation. Vetter et at. [32, 31] learn a 3D geometry-appearance model for face registration and matching. However, today the exact trade-offs and limitation of these algorithms are relatively unknown.

To evaluate the performance of these algorithms, Phillips et. al. have conducted the FERET face algorithm tests [23], based on the FERET database which now contains 14,126 images from 1,199 individuals. More recently the Facial Recognition Vendor Test [3] evaluated commercial systems using the FERET and HumanID databases. The test results have revealed that important progress has been made in face recognition, and many aspects of the face recognition problems are now well understood. However, there still remains a gap between these testing results and practical user experiences of commercial systems. While this gap can, and will, be narrowed through the improvements of practical details such as sensor resolution and view selection, we would like to understand clearly the fundamental capabilities and limitations of current face recognition systems.

In this paper, we will conduct a series of tests using two state of art face recognition systems on three newly constructed face databases to evaluate the effect of face pose, illumination, facial expression, occlusion and subject gender on face recognition performance.

The paper is organized as follows. We describe the three database used in our evaluation in Section 2. In Section 3 we introduce the two algorithms we used for our evaluations. The experimental procedures and results are presented in Section 4, and we conclude in Section 5.

## 2. Description of Databases

### 2.1. Overview

Table 1 gives an overview of the databases used in our evaluation.

| | 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 |

**Table 1.** Overview over databases.

### 2.2. CMU Pose Illumination Expression (PIE) database

The CMU PIE database contains a total of 41,368 images taken from 68 individuals [26]. The subjects were imaged in the CMU 3D Room [14] 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 cameras and flashes are distributed in a hemisphere in front of the subject as shown in Figure 1.

A series of images of a subject across the different 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 are 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 illumination conditions 1 and 2.

### 2.3. Cohn-Kanade AU-Coded Facial Expression Database

This is a publicly available database from Carnegie Mellon University [13]. 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 [8] and also assigned emotion-specified labels. For the current study, we selected 714 image image sequences from 105 subjects. Emotion expressions included happy, surprise, anger, disgust, fear, and sadness. Examples for the different expressions are shown in Figure 4.

### 2.4. AR Face Database

The publicly available AR database was collected at the Computer Vision Center in Barcelona [19]. 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.

(a)

(b)

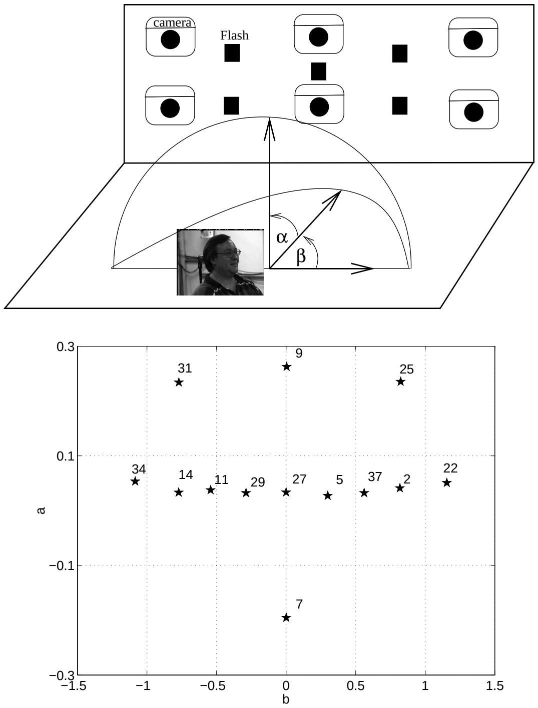


**Figure 1.** PIE database camera positions. (a) 13 synchronized video cameras capture face images from multiple angles, 21 controlled flash units are evenly distributed around the cameras. (b) A plot of the azimuth ($\beta$) and altitude ($\alpha$) angles of the cameras, along with the camera ID number. 9 of the 13 cameras sample a half circle at roughly head height ranging from a full left to a full right profile view(+/- 60 degrees); 2 cameras were placed above and below the central camera; and 2 cameras were positioned in the corners of the room.


**Figure 2.** Pose variation in the PIE database. 8 of 13 camera views are shown here. The remaining 5 camera poses are symmetrical to the right side of camera c27.

**Figure 3.** Illumination variation in the PIE database. The images in the first row show faces recorded with room lights on, the images in the second row show faces captured with only flash illumination.

During each session 13 conditions with varying facial expressions, illumination and occlusion were captured. Figure 5 shows an example for each condition.

## 3. Face Recognition Algorithms

### 3.1. MIT, Bayesian Eigenface

Moghaddam et. al. generalize the Principal Component Analysis (PCA) approach of Sirovich and Kirby [28] and Turk and Pentland [29] by examining the probability distribution of *intra-personal* variations in appearance of the same individual and *extra-personal* variations in appearance due to difference in identity. This algorithm performed consistently near the top in the 1996 FERRET test [23].

Given two face images, $I_1, I_2$, let $\Delta = I_1 - I_2$ be the image intensity difference between them, we would like to estimate the posterior probability of $P(\Omega_i|\Delta)$, where $\Omega_i$ is the intra-personal variation of subject $i$. According to Bayes rule, we can rewrite it as:

$$P(\Omega_i|\Delta) = \frac{P(\Delta|\Omega_i)P(\Omega_i)}{P(\Delta|\Omega_i)P(\Omega_i) + P(\Delta|\Omega_E)P(\Omega_E)}, \quad (1)$$

where $\Omega_E$ is the extra-personal variation of all the subjects. To estimate the probability density distributions $P(\Delta|\Omega_i)$ and $P(\Delta|\Omega_E)$, PCA is used to derived a low (M) dimension approximation of the measured feature space $\Delta \in R^N (N = O(10^4))$:

$$P(\Delta|\Omega) = \frac{exp(-\frac{1}{2}\sum_{i=1}^{M}\frac{y_i^2}{\lambda_i})}{(2\pi)^{M/2}\prod_{i=1}^{M}\lambda_i^{1/2}} \frac{exp(-\frac{\epsilon^2(\Delta)}{2\rho})}{(2\pi\rho)^{(N-M)/2}}, \quad (2)$$

where $y_i, \lambda_i$ are the eigenvectors and eigenvalues in the M dimensional principal component space, and $\epsilon(\Delta)$ is the residual error.

The algorithm finds the subject class $i$ which maximizes the posterior $P(\Omega_i|\Delta)$. Unlike FaceIt's algorithm,

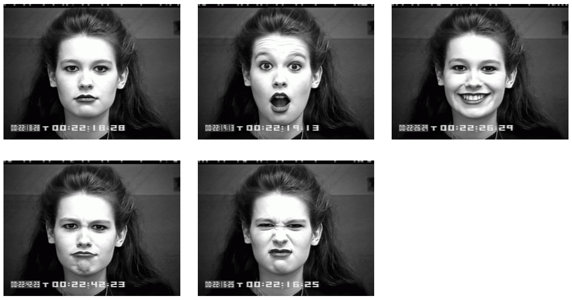

**Figure 4.** Cohn-Kanade AU-Coded Facial Expression database. Examples of emotion-specified expressions from image sequences.

this is a mostly template based classification algorithm, although some local features are implicitly encoded through the "Eigen" intra-personal and extra-personal images.

### 3.2. Visionics, FaceIt

FaceIt's recognition module is based on Local Feature Analysis (LFA) [22]. 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. [23] we distinguish between *gallery* and *probe* images. The gallery contains the images used during training of the algorithm. 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. The algorithms were not given any further information, so we only evaluate the face recognition, not the face

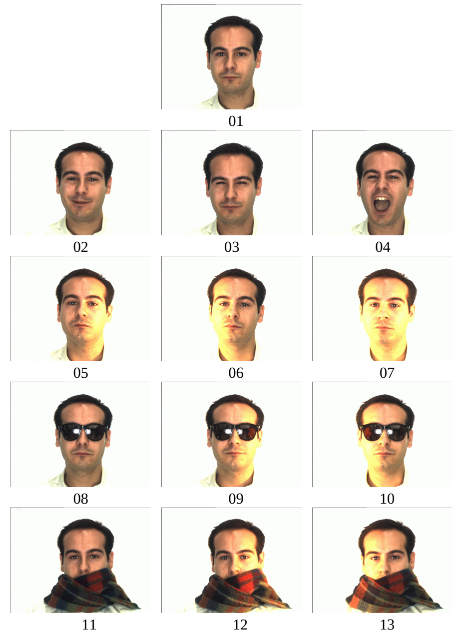


**Figure 5.** AR database. The conditions are: (1) neutral, (2) smile, (3) anger, (4) scream, (5) left light on, (6) right light on, (7) both lights on, (8) sun glasses, (9) sun glasses/left light (10) sun glasses/right light, (11) scarf, (12) scarf/left light, (13) scarf/right light

verification performance.

### 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. In order to ensure the validity of our findings in terms of face recognition accuracy, we provided both algorithms with correct locations of the left and right eyes. This is done by applying FaceIt's face finding module with a subsequent manual verification of the results. If the initial face position was incorrect, the location of the left and right eye was marked manually and the face finding module is rerun on the image. The face detection module became more likely to fail as departure from the frontal view increased.

### 4.2. Pose

Using the CMU PIE database we are in the unique position to evaluate the performance of face recognition algorithms with respect to pose variations in great detail. We exhaustively sampled the pose space by using each view in turn as gallery with the remaining views as probes. As there is only a single image per subject and camera view in the database, the gallery images are included in the probe set. Table 2 shows the complete pose confusion matrix for FaceIt. Of particular interest is the question how far the algorithm can generalize from given gallery views.

Two things are worth noting. First, FaceIt has a reasonable generalizability for frontal gallery images: the recognition rate drops to the 70%-80% range for 45 degree of head rotation (corresponds to camera positions 11 and 37 in Figure 1 ). Figure 6 shows the recognition accuracies of the different camera views for a mugshot gallery view.

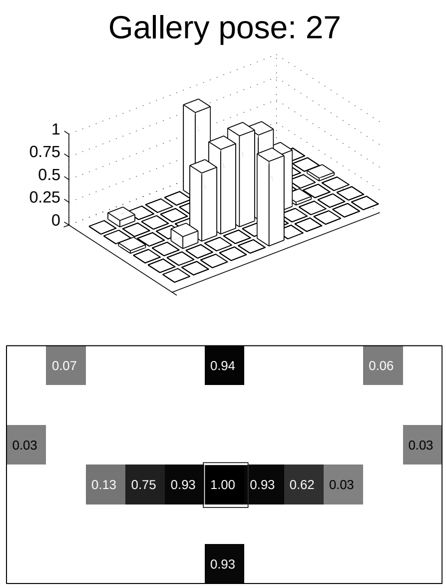


**Figure 6.** Recognition accuracies of all cameras for the mugshot gallery image. The recognition rates are plotted on the pose positions shown in figure 1(b). The darker color in the lower portion of the graph indicates higher recognition rate. The square box marks the gallery view.

Second, for most non-frontal views (outside of the 40 degree range), face generalizability goes down drastically, even for very close by views. This can be seen in Figure 7. Here the recognition rates are shown for the two profile views as gallery images. The full set of performance graphs for all 13 gallery views is shown in appendix A.

We then asked the question, if we can gain more by including 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


**Figure 7.** Recognition accuracies of all cameras for the two profile poses as gallery images (cameras 34 and 22 in 1b).

| $\beta$ | -66 | -47 | -46 | -32 | -17 | 0 | 0 | 0 | 16 | 31 | 44 | 44 | 62 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | 3 | 13 | 2 | 2 | 2 | 15 | 2 | 1.9 | 2 | 2 | 2 | 13 | 3 |
| Probe Pose | c34 | c31 | c14 | c11 | c29 | c09 | c27 | c07 | c05 | c37 | c25 | c02 | 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. The camera pose, indicated by its azimuth($\beta$) and altitude($\alpha$) angle was shown in Figure 1.

novel face poses. This test is carried out by taking two sets of face poses, $\{11, 27, 37\}$, and $\{05, 27, 29\}$ as gallery and test on all other poses. The results are presented in Table 3.

| Probe Pose | 02 | 05 | 07 | 09 | 11 | 14 | 22 | 25 | 27 | 29 | 31 | 34 | 37 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Gallery Pose | | | | | | | | | | | | | |
| 11-27-37 | 0.01 | 0.99 | 0.91 | 0.93 | 1.0 | 0.35 | 0.01 | 0.1 | 1.0 | 0.91 | 0.19 | 0.0 | 1.0 |
| 05-27-29 | 0.01 | 1.0 | 0.90 | 0.91 | 0.88 | 0.24 | 0.01 | 0.1 | 1.0 | 1.0 | 0.12 | 0.01 | 0.66 |

**Table 3.** Pose variation. Recognition rates for FaceIt with multiple poses in the gallery set.

An analysis of the results, shown in Figure 8, indicates that with this algorithm, no additional gain is achieved 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.

We conducted the same set of experiments using the MIT algorithm. The results are much worse than FaceIt's even with manual identification of the eyes. We suspect that this might be due to the extreme sensitivity of the MIT algorithm to face registration errors.

### 4.3. Illumination

For this test, the PIE and AR databases are used. We found that both algorithms performed significantly better on the illumination images than under the various pose conditions. Table 4 shows the recognition accuracies of FaceIt and the MIT algorithm on both databases. As described in section 2.2 the PIE database contains two illumination sets. The images in set *illumination 1* were taken with the room lights on. The mugshot gallery images did not have flash illumination. For the *illumination 2* set of images the room light was switched off. The illumination for the gallery images was provided by a flash directly opposite of the subject's face. In each case the probe set was made up by the remaining flash images (21 and 20 images respectively). As can be expected, the algorithms perform better in the first test.

| | Illumination | | | | |
|---|---|---|---|---|---|
| | PIE 1 | PIE 2 | AR 05 | AR 06 | AR 07 |
| FaceIt | 0.97 | 0.91 | 0.95 | 0.93 | 0.86 |
| MIT | 0.94 | 0.72 | 0.77 | 0.74 | 0.72 |

**Table 4.** 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

The results on the PIE database are consistent with the outcome of the experiments on the AR database. Here the images 5 through 7 deal with different illumination conditions, varying the lighting from the left- and right sides to both lights on.

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

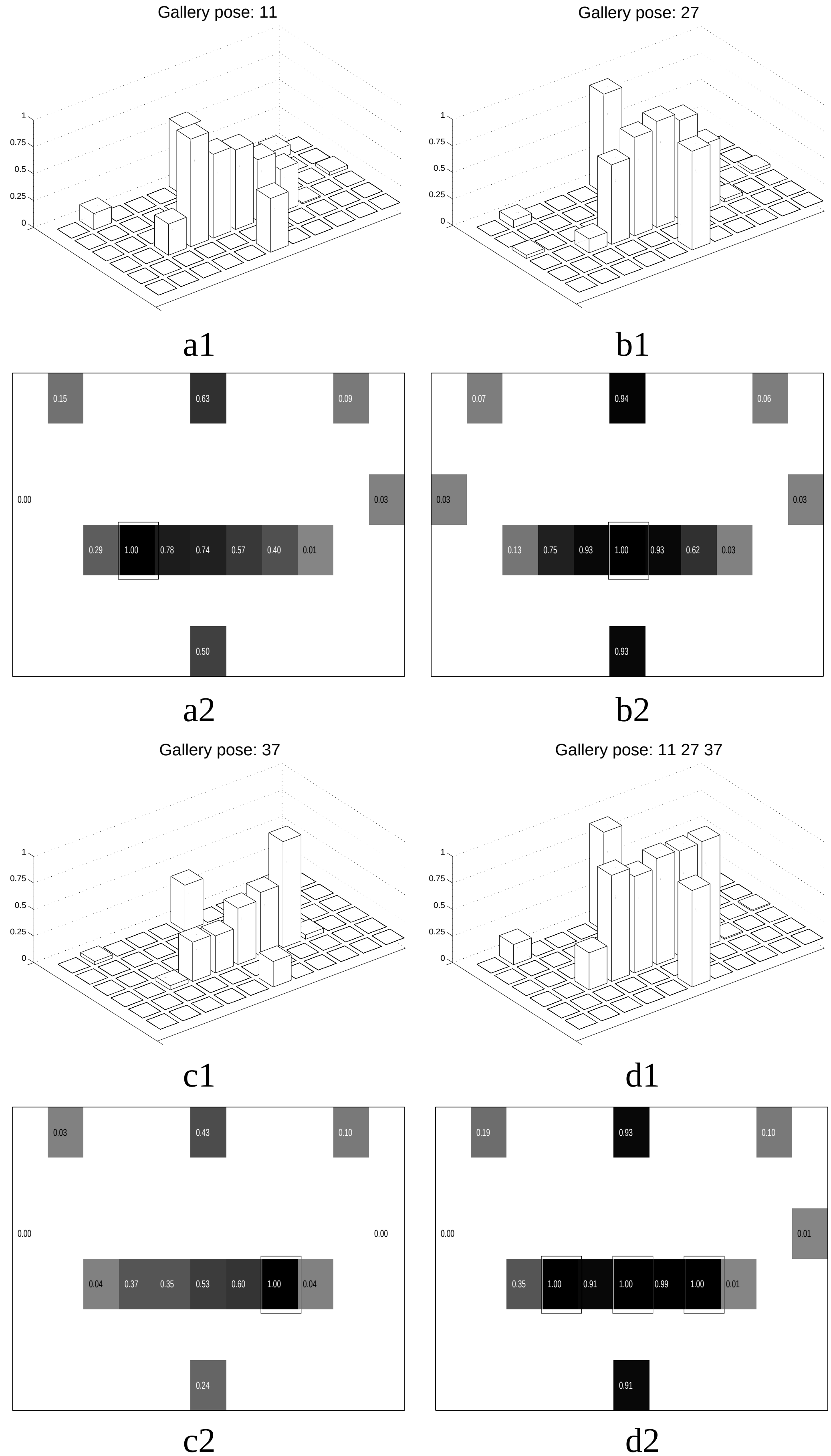


**Figure 8.** Multiple poses gallery vs. multiple single pose gallery. Subplots a1 and a2 show the recognition rate with gallery pose 11. With a single gallery image, the algorithm is able to generalize to nearby poses. Subplots b{1,2} and c{1,2} show the recognition rates for poses 27 and 37. Subplots d{1,2} show the recognition rates with gallery pose of {11,27,37}. One can see d{1,2} is the same as taking the maximum values in a-c{1,2}. In this case no additional gain is achieved by using the joint set {11,27,37} as the gallery poses.

### 4.4. Expression

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. To our surprise FaceIt and MIT performed very well on the Cohn-Kanade and the AR database. In each test we used the neutral expression as gallery image and probed the algorithm with the peak expression.

| | Expression | | | |
|---|---|---|---|---|
| | Cohn-Kanade | AR 02 | AR 03 | AR 04 |
| FaceIt | 0.97 | 0.96 | 0.93 | 0.78 |
| MIT | 0.94 | 0.72 | 0.67 | 0.41 |

**Table 5.** Expression results. AR 02, AR 03 and AR 04 refer to the expression changes in the AR database as shown in figure 5. Both algorithms perform reasonably well under facial expression, however the “scream” expression, AR 04, produces large recognition errors.

Table 5 shows the results of both algorithms on the two databases. The notable exception is the *scream* (AR04) set of the AR database.

For 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.

### 4.5. Occlusion

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 the sun glass images are according to expectations. As Table 6 shows, FaceIt is unable to handle this variation (AR08). The result further deteriorates when the left or right light is switched on (AR09 and AR10). This result is readily replicated on the images of the second session.

This test reveals that FaceIt is more vulnerable to upper face occlusion than the MIT algorithm. Facial occlusion, particularly upper face occlusion, remains a difficult problem yet to be solved. Interesting open questions are 1) what

| | Occlusion | | | | | |
|---|---|---|---|---|---|---|
| | AR 08 | AR 09 | AR 10 | AR 11 | AR 12 | AR13 |
| FaceIt | 0.10 | 0.08 | 0.06 | 0.81 | 0.73 | 0.71 |
| MIT | 0.34 | 0.35 | 0.28 | 0.46 | 0.43 | 0.40 |

**Table 6.** 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.

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.

### 4.6. Gender

Male and female faces differ in both local features and in shape [4]. 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 [4]. People readily distinguish male from female faces using these and other differences (e.g., hair style), and connectionist modeling has yielded similar results [6, 18]. 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 is seldom reported, and identification results typically fail to report 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 [4, 9], have similarly been ignored in past research.

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. Figure 9 shows the recognition rate achieved by FaceIt across the 13 variations including illumination, occlusion, and expression.

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) the recognition rate for male subjects is 83.4%, while the recognition rate for female subjects is 91.66%. It is not clear what has caused this effect. To further validate this result, a much larger database is needed.

If this result is further substantiated, it 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.

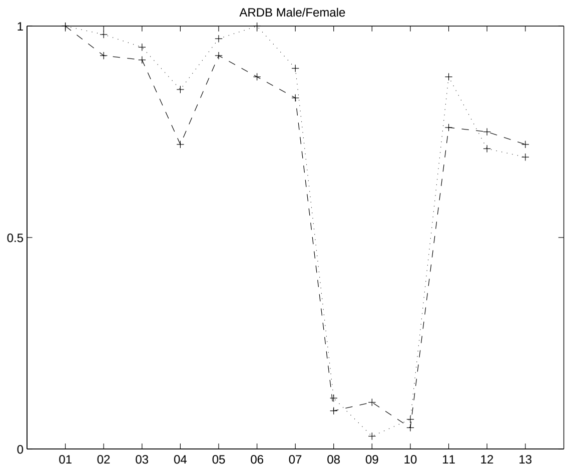


**Figure 9.** ARDB results male vs. female. The dashed line indicates the recognition rate for male subjects in the AR database shown in figure 5.

## 5. Discussion

In natural environments, pose, illumination, expression, occlusion and individual difference among people represent critical challenges to face recognition algorithms. The FERET tests [23] and the Facial Recognition Vendor Test 2000 [3] provided initial results on limited variations of these factors.

FaceIt and the MIT algorithm were overall the best performer in these tests. We evaluated both algorithms on multiple independent databases that *systematically* vary pose, illumination, expression, occlusion, and gender. We found:

1. *Pose:* Pose variation still presents a challenge for face recognition. Frontal training images have better generalizability to novel views than do non-frontal training images. For a frontal training view, we can achieve reasonable recognition rates of 70-80% for up to 45 degree head rotation. In addition, using multiple training views does not necessarily improve the recognition rate.

2. *Illumination:* Pure illumination changes on the face are handled well by current face recognition algorithms.

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 to the algorithms.

4. *Occlusion:* The performance of the face recognition algorithms under occlusion varies. FaceIt is more sensitive to upper face occlusion than MIT. FaceIt is more robust to lower face occlusion.

5. *Gender:* We found surprisingly consistent differences of face recognition rates across the gender. This result is based on testing on the AR database which has 70 male and 60 female subjects. On average the recognition rate for females is consistently about 5% higher than for males, across a range of perturbation. While the database used in these tests is too small to draw general conclusions it points into an interesting direction for future research and database collections.

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.

## 6. Acknowledgement


This research is supported by ONR N00014-00-1-0915 (DARPA HumanID) and NSF IRI-9817496.


## A. Appendix

The following figures show the recognition rates for all camera views in turn as gallery images. They are ordered according to the camera numbers roughly going from left to right in Figure 1.

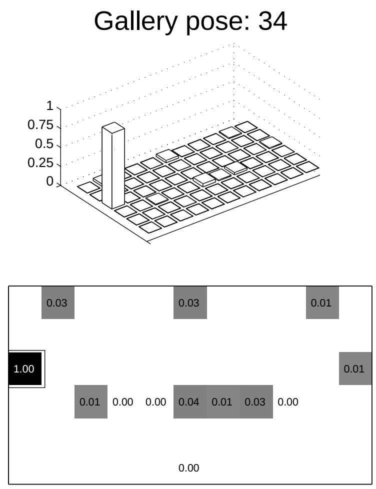


**Figure 10.** Recognition accuracies of all cameras for camera view 34 as gallery image.

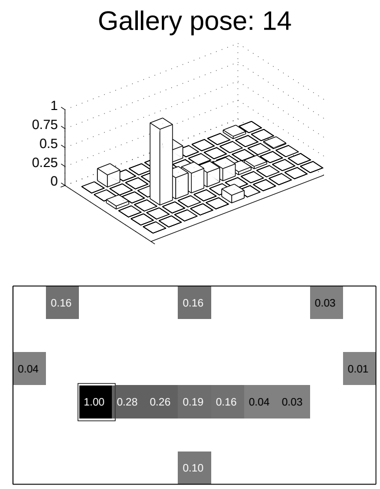


**Figure 12.** Recognition accuracies of all cameras for camera view 14 as gallery image.

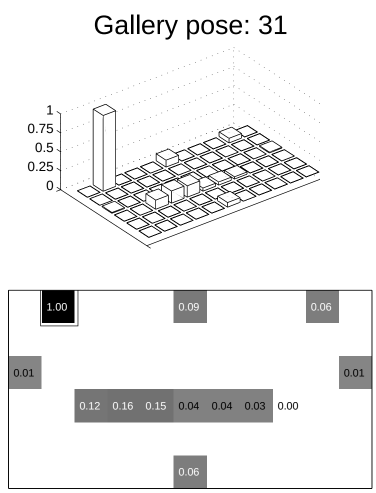


**Figure 11.** Recognition accuracies of all cameras for camera view 31 as gallery image.

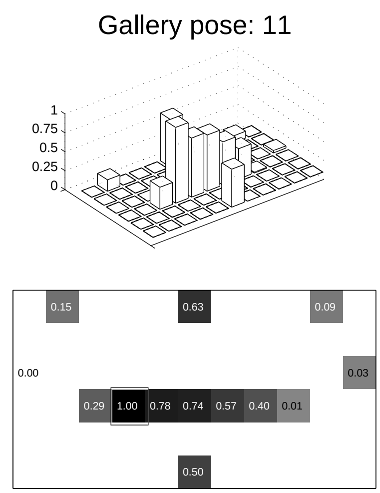


**Figure 13.** Recognition accuracies of all cameras for camera view 11 as gallery image.

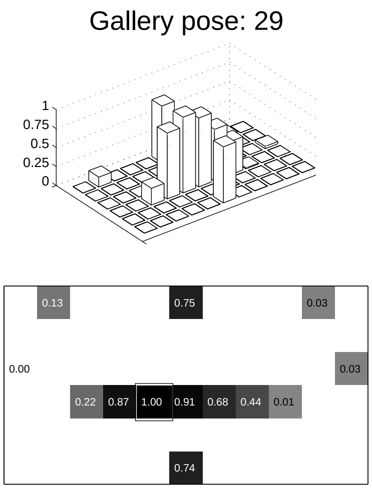


**Figure 14.** Recognition accuracies of all cameras for camera view 29 as gallery image.

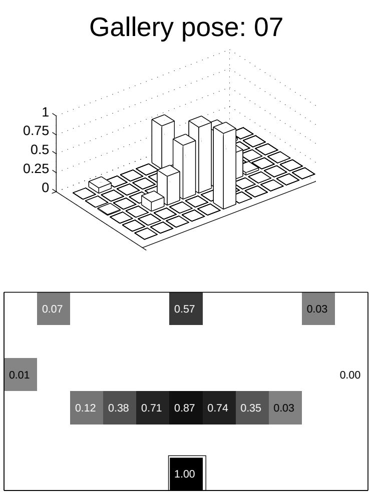


**Figure 15.** Recognition accuracies of all cameras for camera view 07 as gallery image.

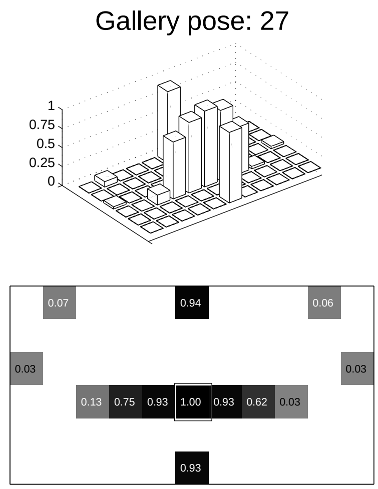


**Figure 16.** Recognition accuracies of all cameras for camera view 27 as gallery image.

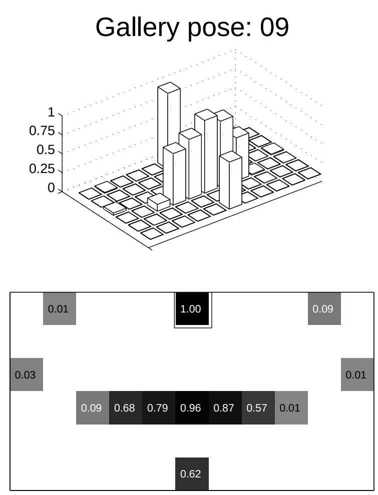


**Figure 17.** Recognition accuracies of all cameras for camera view 09 as gallery image.

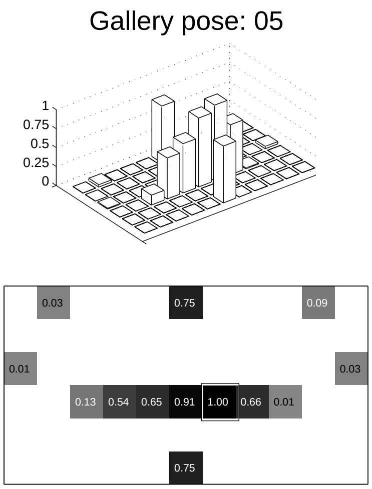


**Figure 18.** Recognition accuracies of all cameras for camera view 05 as gallery image.

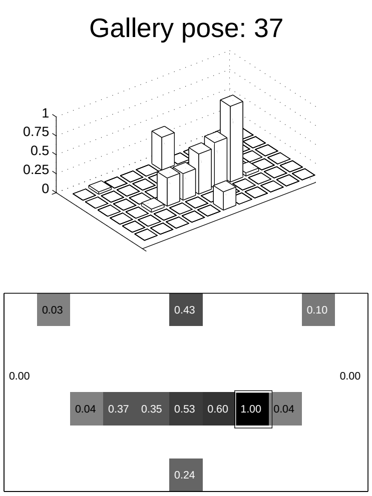


**Figure 19.** Recognition accuracies of all cameras for camera view 37 as gallery image.

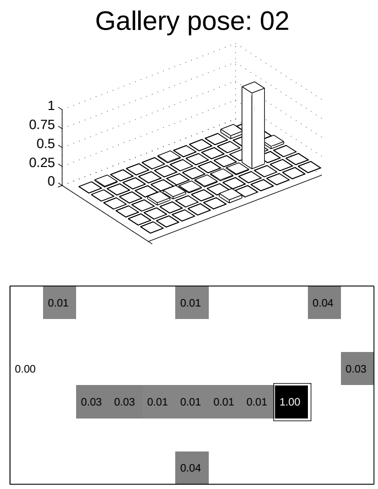


**Figure 20.** Recognition accuracies of all cameras for camera view 02 as gallery image.

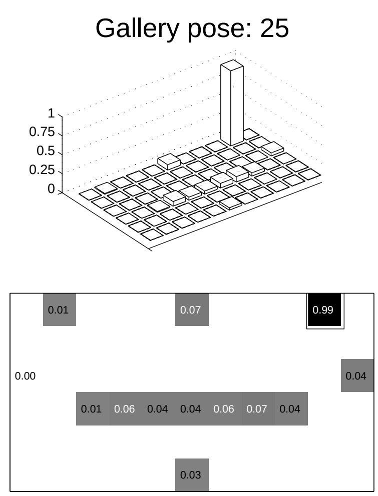


**Figure 21.** Recognition accuracies of all cameras for camera view 25 as gallery image.

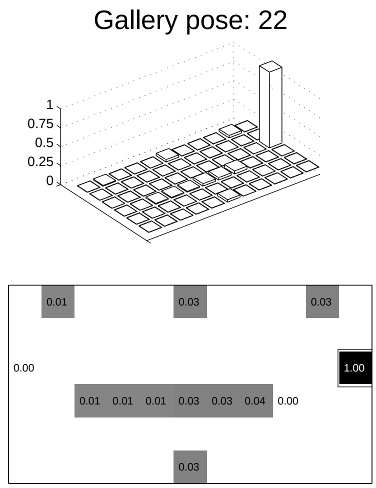


**Figure 22.** Recognition accuracies of all cameras for camera view 22 as gallery image.

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

[5] T. Cootes, K. Walker, and C. Taylor. View-based active appearance models. In *IEEE International Conference on Automatic Face and Gesture Recognition*, March 2000.

[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] M. N. Dailey and Garrison W. Cottrell. Organization of face and object recognition in modular neural network models. *Neural Networks*, 12(7/8), 1999.

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

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

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

[11] Peter J. B. Hancock, A. Mike Burton, and Vicki Bruce. Face processing: Human perception and principal components analysis. *Memory and Cognition*, 24(1):26–40, 1996.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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

[20] 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.

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

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

[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] 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.

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

[26] T. Sim, S. Baker, and M. Bsat. The CMU Pose, Illumination, and Expression (PIE) database of human faces. Technical Report CMU-RI-TR-01-02, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, January 2001.

[27] 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.

[28] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces, 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] Thomas Vetter. Synthesis of novel views from a single face image. *International Journal of Computer Vision*, 28(2):103–116, 1998.

[32] Thomas Vetter, Anya Hurlbert, and Tomaso Poggio. View-based models of 3D object recognition: Invariance to imaging transformations. *Cerebral Cortex*, 5(3):261–269, 1995.

[33] 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.

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