

Tied factor analysis for face recognition across large pose changes

Simon J.D. Prince¹ and James H. Elder²

¹Department of Computer Science, University College London, UK, s.prince@cs.ucl.ac.uk

² Centre for Vision Research, York University, Toronto, Canada, jelder@yorku.ca

Abstract

Face recognition algorithms perform very unreliably when the pose of the probe face is different from the stored face: typical feature vectors vary more with pose than with identity. We propose a generative model that creates a one-to-many mapping from an idealized “identity” space to the observed data space. In this identity space, the representation for each individual does not vary with pose. The measured feature vector is generated by a pose-contingent linear transformation of the identity vector in the presence of noise. We term this model “tied” factor analysis. The choice of linear transformation (factors) depends on the pose, but the loadings are constant (tied) for a given individual. Our algorithm estimates the linear transformations and the noise parameters using training data. We propose a probabilistic distance metric which allows a full posterior over possible matches to be established. We introduce a novel feature extraction process and investigate recognition performance using the FERET database. Recognition performance is shown to be significantly better than contemporary approaches.

1 Introduction

In face recognition, there is commonly only one example of an individual in the database. Recognition algorithms extract feature vectors from a probe image and search the database for the closest vector. Most previous work has revolved around selecting optimal feature sets. The dominant paradigm is the “appearance based” approach in which weighted sums of pixel values are used as features for the recognition decision. Turk and Pentland [11] used principal components analysis to model image space as a multidimensional Gaussian and selected the projections onto the largest eigenvectors. Other work has used more optimal linear weighted pixel sums, or analogous non-linear techniques [1, 7].

One of the greatest challenges for these methods is to recognize faces across different poses and illuminations [13]. In this paper we address the worst case scenario in which there is only a single instance of each individual in a large database and the probe image is taken from a very different pose than the matching test image. Under these circumstances, most methods fail, since the extracted feature vector varies considerably with the pose. Indeed, variation attributable to pose may dwarf the variation due to differences in identity. Our strategy is to build a generative model that explains this variation. In particular we develop a one-to-many transformation from an idealized “identity” space in which each individual has a unique vector regardless of pose, to the conventional feature space where features vary with pose.

The simplest approach to making recognition robust to pose is to remove all feature measurements that co-vary strongly with this variable. A more sophisticated approach

2

is to measure the amount of signal (inter-personal variation) and noise (variation due to pose in this case) along each dimension and select features where the signal:noise ratio is optimal [1]. A drawback of these approaches is that the discarded dimensions may contain a significant portion of the signal and their elimination ultimately impedes recognition performance. Another obvious method to generalize across pose is to record each subject in the database at each possible angle, and use an appearance based model for each [8]. Another approach is to use several photos to create a 3D model of the head which can then be re-rendered at any given pose to compare with given probe [5, 12]. Unfortunately, these methods require extensive recording and the cooperation of the subject.

Several previous studies have presented algorithms which can take a single probe image at one pose and attempt to match it to a single test image at a different pose. One approach is to create a full 3D head model for the subject based on just one image [10, 2] and compare 3D models. This approach is feasible, but the computation involved is too significant for a practical face recognition system. This problem can partially be alleviated by projecting the test models to 2D images at all possible orientations in advance [3]. However, registration of a new individual is still computationally expensive. An alternative approach is to treat this as a learning problem in which we aim to predict frontal images from non-frontal ones: the “Eigen-light fields” by Gross et al. [6] treats matching as a missing data problem - the single test and probe images are assumed to be parts of larger data vector containing the face viewed from all poses. The missing information is estimated from the visible data, based on prior knowledge of the joint covariance structure. The complete vector can then be used for the matching process.

The emphasis in these algorithms is on creating a model which can predict how a given face will appear when viewed at different poses. Prince and Elder [9], presented a heuristic algorithm to construct a single feature *which does not vary with pose*. This seems a natural formulation for a recognition task. In this paper, we develop this idea in a full Bayesian probabilistic setting. In Section 2 we introduce the problem of pose variation as seen from the observation space. We then introduce the idea of a pose-invariant vector space, and describe a pose-contingent mapping from this invariant space to explain the original measured features. We then describe how the direction of inference can be reversed, so that a pose-invariant feature vector can be estimated given the image measurements. In Section 2.3 we use this reverse inference to iteratively estimate the parameters of the mapping using the EM algorithm. We introduce a recognition method based on Bayesian model comparison. We introduce a set of observation features that are particularly suited to recognition across large pose variations. We compare our algorithm to contemporary work and show that it produces superior results.

2 Methods

For most choices of feature vector, the majority of positions in the vector space are unlikely to have been generated by faces. The subspace to which faces commonly project is termed the face manifold. In general this is a complex non-linear probabilistic region tracing through multi-dimensional space. Figure 1 shows that the mean position of this region changes systematically with pose. Moreover, for a given individual, the position of the observation vector relative to this mean also varies. This accounts for the poor recognition performance when measurement vectors are compared across different poses: There is no simple distance metric in this space that supports good recognition performance.

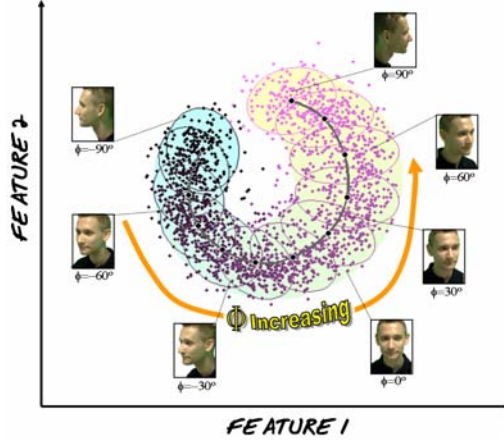


Figure 1: The effect of pose variation in the observation space. Face pose is coded by intensity, so that faces with poses near -90° are represented by dark points and faces with poses near 90° are represented by light points. The pose-variable is quantized into K bins, and each bin is represented by a Gaussian distribution (ellipses). The K means of these Gaussians trace a path through multi-dimensional space as we move through each successive pose bin (solid gray line) The shaded region represents the envelope of the K covariance ellipses. Notice that the same individual appears at very different positions in the manifold depending on the pose at which their image is taken. There is clearly no simple metric in this space which will identify these points with one another.

2.1 Modelling Feature Generation

At the core of our algorithm is the notion that there genuinely exists a multidimensional vector \mathbf{c} that represents the identity of the individual regardless of the pose. It is assumed that the image data, \mathbf{x}_k at pose ϕ_k can be generated from this pose-invariant “identity” representation, using a parameterized function \mathbf{F}_k that depends on the pose. In particular, we assume that there is a linear function which is specialized for generating the image data at pose ϕ_k . The forward model is hence:

$$\mathbf{x}_k = \mathbf{F}_k \mathbf{c} + \mathbf{m}_k + \eta_k \quad (1)$$

where \mathbf{m}_k is the mean vector for this pose bin and η_k is a zero-mean noise term distributed as, $\mathcal{G}_x(\mathbf{0}, \Sigma_k)$ with unknown diagonal covariance matrix Σ_k . We denote the unknown parameters of the system, $\{\mathbf{F}_{1..K}, \mathbf{m}_{1..K}, \Sigma_{1..K}\}$ with the symbol θ . This model is equivalent to factor analysis where the factors \mathbf{F}_k depend on pose, but the factor loadings \mathbf{c} are the same at each pose (tied).

The pose-invariant vectors, \mathbf{c} are assumed a priori to be distributed as a zero mean Gaussian with identity covariance, $\mathcal{G}_c(\mathbf{0}, \mathbf{I})$. The dimensionality of the invariant space \mathbf{c} is a parameter of the system. The relationship between the standard feature space, \mathbf{x} and the identity space \mathbf{c} is indicated in Figure 2. It can be seen that vectors in widely varying parts of the original image space can be generated from the same point in identity space.

4

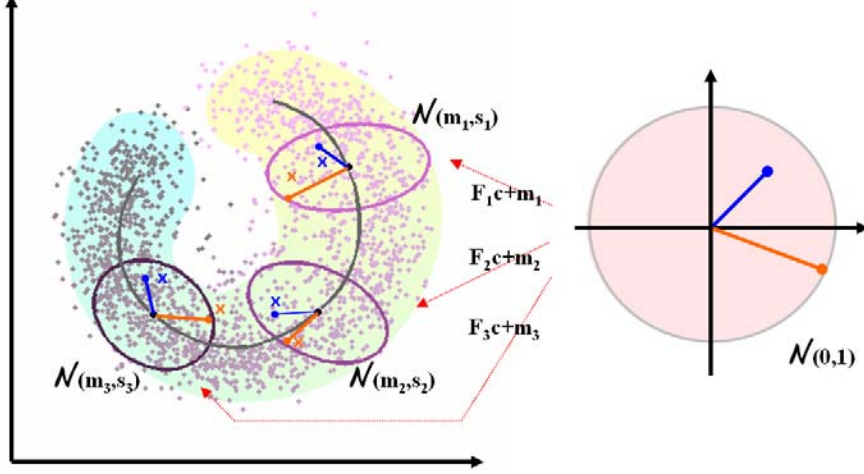


Figure 2: Forward model for feature generation. Left hand side represents the image measurement space. Right hand side represents a second pose-invariant “identity” space. The three blue crosses represent image measurements for one person viewed at three poses, $k = \{1, 2, 3\}$. Orange crosses represent measurements for a second individual viewed at a the same 3 poses. These data originate from two points (one for each individual) in a zero-mean pose-invariant feature space (solid circles on right hand side). Image measurements for pose k are created by multiplying the invariant vectors by linear transformation \mathbf{F}_k and adding, \mathbf{m}_k , (see Equation 1). The resulting data (solid circles on left hand side) are observed under noisy conditions (crosses).

2.2 Estimating Pose-Invariant Vectors

In the previous section we have described how pose-dependent vectors can be generated from an underlying pose invariant representation. It is also necessary to model the inverse process. In other words, we wish to estimate the invariant vector for a given individual, \mathbf{c} given image measurements, \mathbf{x}_k at some pose, $\phi = \phi_k$. The posterior distribution for the invariant vector can be calculated using Bayes’ rule:

$$p(\mathbf{c}|\mathbf{x}, \phi=\phi_k, \theta) = \frac{p(\mathbf{x}|\mathbf{c}, \phi=\phi_k, \theta)p(\mathbf{c})}{\int p(\mathbf{x}|\mathbf{c}, \phi=\phi_k, \theta)p(\mathbf{c})d\mathbf{c}} \quad (2)$$

where:

$$p(\mathbf{x}|\mathbf{c}, \phi=\phi_k, \theta) = \mathcal{G}_{\mathbf{x}}(\mathbf{F}_k\mathbf{c} + \mathbf{m}_k, \Sigma_k) \quad (3)$$

$$p(\mathbf{c}) = \mathcal{G}_{\mathbf{c}}(\mathbf{0}, \mathbf{I}) \quad (4)$$

Since all terms in Equation 2 are normally distributed, it follows that the posterior must also be normally distributed. After some manipulation it can be shown that:

$$p(\mathbf{c}|\mathbf{x}, \phi=\phi_k, \theta) = \mathcal{G}(\mathbf{F}_k^T(\mathbf{F}_k\mathbf{F}_k^T + \Sigma_k)^{-1}(\mathbf{x} - \mathbf{m}_k), \mathbf{I} - \mathbf{F}_k^T(\mathbf{F}_k\mathbf{F}_k^T + \Sigma_k)^{-1}\mathbf{F}_k) \quad (5)$$

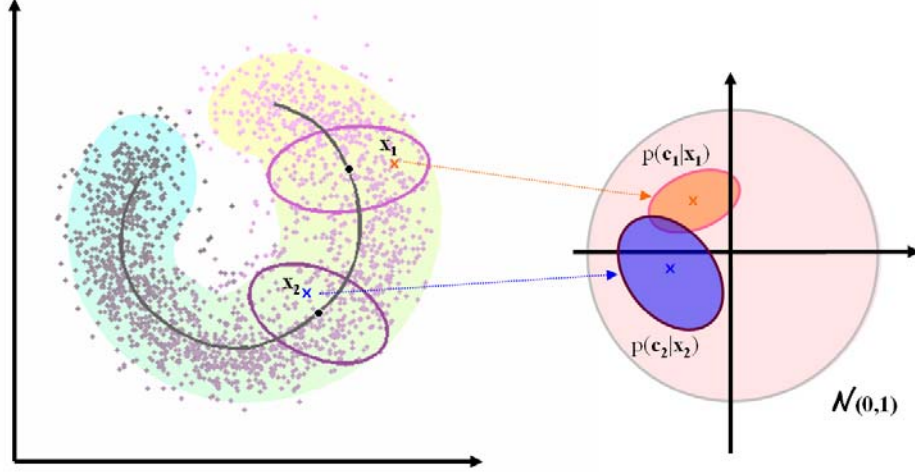


Figure 3: The probability distribution for the pose-invariant vector (right hand side) is inferred from the measured vector (left-hand side). Two image data points \mathbf{x}_1 and \mathbf{x}_2 at different poses are transformed to the invariant space. Intuitively, the probability that the two measured feature vectors belong to the same individual is determined by the degree of overlap of the two distributions in the pose-invariant space.

This formulation assumes that we know the pose, ϕ_k of the image under consideration. This is illustrated in Figure 3. Each data point in the original space is associated with one of the pose bins and transformed into the identity space, to yield a Gaussian posterior.

2.3 Learning System Parameters

We now have a prescription to generate new pose-invariant feature vectors from the initial image measurements. However, this requires knowledge of the functions, \mathbf{F}_k , the means, \mathbf{m}_k and the noise parameters, Σ_k . These must be learnt from a training data set with two important characteristics. First, the value of the pose must be known for each member. In this sense our algorithm is partially supervised. Second, each individual in the training database appears with at least two different poses. These characteristics provide sufficient information to learn the relationship between images of the same face at different poses.

We aim to adjust the parameters, $\theta = \{\mathbf{F}_{1..K}, \mathbf{m}_{1..K}, \Sigma_{1..K}\}$ to increase the joint likelihood $p(\mathbf{x}, \mathbf{c}|\theta)$ of the measured image data \mathbf{x} and the invariant vectors, \mathbf{c} . Unfortunately, we cannot observe the invariant vectors directly: we can only infer them, and this in turn requires the unknown parameters, θ . This type of chicken-and-egg problem is suited to the EM algorithm[4]. We iteratively maximize:

$$Q(\theta_t, \theta_{t-1}) = \sum_{i=1}^I \sum_{k=1}^K \int p(\mathbf{c}_i | \mathbf{x}_{i1..ik}, \theta_{t-1}) \log[p(\mathbf{x}_{ik} | \mathbf{c}_i, \theta_t) p(\mathbf{c}_i)] d\mathbf{c}_i \quad (6)$$

where t represents the iteration index and the three probability terms on the right hand side are given by Equations 5,1 and 4 respectively. The term \mathbf{x}_{ik} represents training face data for individual i at pose ϕ_k . For notational convenience we assume that we have one training face vector, \mathbf{x}_{ik} for each individual i at every pose ϕ_k . In practice this is not a

6

necessary requirement: if data is missing (all individuals are not seen at all poses) these terms are simply dropped from the summation.

The EM algorithm alternately finds the expected values for the unknown pose-invariant vectors \mathbf{c} (the Expectation- or E-Step) and then maximizes the overall likelihood of data as a function of the parameters θ (the Maximization- or M-Step). More precisely, the E-Step calculates the expected values of the invariant vector \mathbf{c}_i for each individual i , using the data for that individual across all poses, $\mathbf{x}_{i1\dots iK}$. The M-Step optimizes the values of the transformation parameters $\{\mathbf{F}_k, \mathbf{m}_k, \Sigma_k\}$ for each pose, k , using data for that pose across all individuals, $\mathbf{x}_{1k\dots Ik}$. These steps are repeated until convergence.

E-Step: For each individual, we estimate the distribution of \mathbf{c}_i given the current parameter estimates θ_{t-1} . We assume that the probability distributions for \mathbf{c}_i given each data point $\mathbf{x}_{i1\dots iK}$ are independent so that:

$$p(\mathbf{x}_{i1\dots iK} | \mathbf{c}_i, \theta) = \prod_{k=1}^K p(\mathbf{x}_{ik} | \mathbf{c}_i, \phi = \phi_k, \theta) \quad (7)$$

where the terms on the right hand side are calculated from the forward model (Equation 3). Since all terms are normally distributed, the left hand side is also normally distributed and can be represented with a mean vector and covariance matrix. We use Bayes rule to combine this new likelihood estimate with the prior over the invariant space as in Equation 2. This yields a posterior distribution similar to that in Equation 5. The first two moments of this distribution can be shown to equal:

$$\begin{aligned} E[\mathbf{c}_i | \mathbf{x}, \theta] &= \left(\mathbf{I} + \sum_{k=1}^K \mathbf{F}_k^T \Sigma_k^{-1} \mathbf{F}_k \right)^{-1} \sum_{k=1}^K \mathbf{F}_k^T \Sigma_k^{-1} (\mathbf{x}_k - \mathbf{m}_k) \\ E[\mathbf{c}_i \mathbf{c}_i^T | \mathbf{x}, \theta] &= \left(\mathbf{I} + \sum_{k=1}^K \mathbf{F}_k^T \Sigma_k^{-1} \mathbf{F}_k \right)^{-1} + E[\mathbf{c}_i | \mathbf{x}, \theta] E[\mathbf{c}_i | \mathbf{x}, \theta]^T \end{aligned} \quad (8)$$

M-Step: For each pose, ϕ_k we maximize the objective function, $Q(\theta_t, \theta_{t-1})$, defined in Equation 6 with the respect to the parameters, θ . For simplicity, we estimate the mean, \mathbf{m}_k and linear transform \mathbf{F}_k at the same time. To this end, we create new matrices $\tilde{\mathbf{F}}_k = [\mathbf{F}_k \ \mathbf{m}_k]$ and $\tilde{\mathbf{c}}_i = [\mathbf{c}_i^T \ 1]^T$. The first log probability term in Equation 6 can be written

$$\log[p(\mathbf{x}_{ik} | \mathbf{c}_i, \theta_t)] = \kappa + \frac{1}{2} (\log |\Sigma_k^{-1}| + (\mathbf{x}_{ik} - \tilde{\mathbf{F}}_k \tilde{\mathbf{c}}_i)^T \Sigma_k^{-1} (\mathbf{x}_{ik} - \tilde{\mathbf{F}}_k \tilde{\mathbf{c}}_i)) \quad (9)$$

where κ is an unimportant constant. We substitute this expression into Equation 6 and take derivatives with respect to each $\tilde{\mathbf{F}}_k$, and Σ_k . The second log term in Equation 6 had no dependence on these parameters and disappears from the derivatives. These derivative expressions are equated to zero and re-arranged to provide the following update rules:

$$\tilde{\mathbf{F}}_k = \left(\sum_{i=1}^I \mathbf{x}_{ik} E[\tilde{\mathbf{c}}_i | \mathbf{x}, \theta]^T \right) \left(\sum_{i=1}^I E[\tilde{\mathbf{c}}_i \tilde{\mathbf{c}}_i^T | \mathbf{x}, \theta] \right)^{-1} \quad (10)$$

$$\Sigma_k = \frac{1}{I} \sum_{i=1}^I \mathbf{diag} [\mathbf{x}_{ik} \mathbf{x}_{ik}^T - \tilde{\mathbf{F}}_k E[\tilde{\mathbf{c}}_i | \mathbf{x}, \theta] \mathbf{x}_{ik}] \quad (11)$$

where **diag** represents the operation of retaining only the diagonal elements from a matrix.

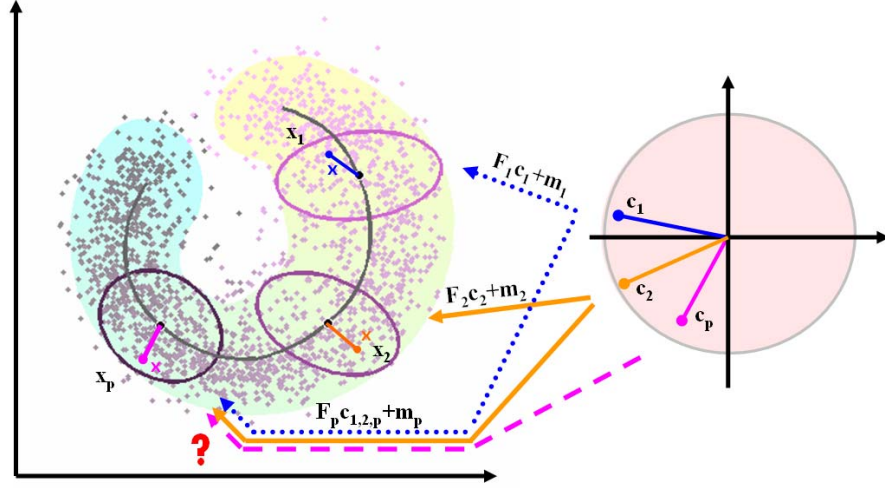


Figure 4: Recognition is posed in terms of Bayesian model comparison. Consider two test faces, \mathbf{x}_1 and \mathbf{x}_2 , and a probe face \mathbf{x}_p . The recognition algorithm compares the evidence for three models: (i) The probe face was generated from the same invariant vector as test face 1. (ii) The probe face was generated from the same invariant vector as test face 2. (iii) The probe face was generated from a third identity vector, \mathbf{c}_p .

2.4 Face Recognition

In the previous section we described how to learn the parameters, $\theta = \{\mathbf{F}_{1...K}, \mathbf{m}_{1...K}, \Sigma_{1...K}\}$. Now we use these parameters to perform face recognition. We are given a test database of faces, $\mathbf{x}_{1...N}$, each of which belongs to a different individual. We are also given a single probe face, \mathbf{x}_p . Our task is to determine the posterior probability that each test face matches the probe face. We may also wish to consider the possibility that the probe face is not present in the test set.

We pose the recognition task in terms of model comparison. We compare evidence for $N+1$ models, which we denote by $\mathcal{M}_{0...N}$. Model \mathcal{M}_0 represents the case where the probe face is not in the test database. We hypothesize that each test feature vector \mathbf{x}_n was generated by a distinct pose-invariant vector, \mathbf{c}_n , and that the probe face \mathbf{x}_p was generated by a different pose-invariant vector, \mathbf{c}_p . The n 'th model, \mathcal{M}_n represents the case where the probe matches the n 'th test face in the database: we assume that there are only N underlying pose-invariant vectors, $\mathbf{c}_{1...N}$, each of which generated the corresponding test feature vector $\mathbf{x}_{1...N}$. The n 'th pose-invariant vector \mathbf{c}_n is also deemed responsible for having generated the probe feature vector \mathbf{x}_p (i.e. $\mathbf{c}_p = \mathbf{c}_n$). Hence, models $\mathcal{M}_{1...n}$ have N parameter vectors $\mathbf{c}_{1...N}$ and model \mathcal{M}_0 has one further parameter vector, \mathbf{c}_p . The evidence for models \mathcal{M}_0 and \mathcal{M}_n are given by:

$$\begin{aligned}
 p(\mathbf{x}_{1...N}, \mathbf{x}_p | \mathcal{M}_0) &= \int p(\mathbf{x}_{1...N}, \mathbf{x}_p, \mathbf{c}_1 \dots \mathbf{c}_N, \mathbf{c}_p) d\mathbf{c}_{1...N,p} \\
 &= \int p(\mathbf{x}_1 | \mathbf{c}_1) p(\mathbf{c}_1) d\mathbf{c}_1 \dots \int p(\mathbf{x}_N | \mathbf{c}_N) p(\mathbf{c}_N) d\mathbf{c}_N \int p(\mathbf{x}_p | \mathbf{c}_p) p(\mathbf{c}_p) d\mathbf{c}_p
 \end{aligned} \tag{12}$$

8

$$\begin{aligned}
p(\mathbf{x}_{1\dots N}, \mathbf{x}_p | \mathcal{M}_n) &= \int p(\mathbf{x}_{1\dots N}, \mathbf{x}_p, \mathbf{c}_1 \dots \mathbf{c}_N, \mathbf{c}_p | \mathbf{c}_p = \mathbf{c}_n) d\mathbf{c}_{1\dots N} \\
&= \int p(\mathbf{x}_1 | \mathbf{c}_1) p(\mathbf{c}_1) d\mathbf{c}_1 \dots \int p(\mathbf{x}_n, \mathbf{x}_p | \mathbf{c}_n) p(\mathbf{c}_n) d\mathbf{c}_n \dots \int p(\mathbf{x}_N | \mathbf{c}_N) p(\mathbf{c}_N) d\mathbf{c}_N
\end{aligned} \tag{13}$$

Since all the terms in these expressions are Gaussian it is possible to find closed form expressions for the evidence, obviating the need for explicit integration over multidimensional space. For each different model type, it is possible to calculate the posterior distribution for the parameters, \mathbf{c} using Equation 2. It is also possible to approximate the posterior distributions by delta functions at the maximum a-posteriori solutions, $\hat{\mathbf{c}}_{1\dots N, p}$ in which case the solution for model \mathcal{M}_n becomes:

$$\begin{aligned}
p(\mathbf{x}_{1\dots N}, \mathbf{x}_p | \mathcal{M}_0) &\approx p(\mathbf{x}_1 | \hat{\mathbf{c}}_1) p(\hat{\mathbf{c}}_1) \dots p(\mathbf{x}_N | \hat{\mathbf{c}}_N) p(\hat{\mathbf{c}}_N) p(\mathbf{x}_p | \hat{\mathbf{c}}_p) p(\hat{\mathbf{c}}_p) \\
p(\mathbf{x}_{1\dots N}, \mathbf{x}_p | \mathcal{M}_n) &\approx p(\mathbf{x}_1 | \hat{\mathbf{c}}_1) p(\hat{\mathbf{c}}_1) \dots p(\mathbf{x}_n, \mathbf{x}_p | \hat{\mathbf{c}}_n) p(\hat{\mathbf{c}}_n) \dots p(\mathbf{x}_N | \hat{\mathbf{c}}_N) p(\hat{\mathbf{c}}_N)
\end{aligned} \tag{14}$$

The solution for model \mathcal{M}_0 is similar. We calculate a posterior over the possible models using a second level of inference:

$$p(\mathcal{M}_n | \mathbf{x}_{1\dots N}, \mathbf{x}_p, \theta) = \frac{p(\mathbf{x}_{1\dots N}, \mathbf{x}_p | \mathcal{M}_n, \theta) p(\mathcal{M}_n)}{\sum_{m=0}^N p(\mathbf{x}_{1\dots N}, \mathbf{x}_p | \mathcal{M}_m, \theta) p(\mathcal{M}_m)} \tag{15}$$

where the terms $p(\mathcal{M}_n)$ represent prior probabilities for each of the models.

3 Results

Twenty unique positions on the frontal faces were identified by hand. For non-frontal faces, the subset of these features which were visible were located. A Delaunay triangulation of each image was calculated based on these points, which were then warped into a constant position for each pose under consideration. It is unrealistic to expect a linear transformation to be able to model the change across the whole image under severe pose changes. Hence we build 10 local models of image change at 10 of the original feature points that are visible at all (leftward facing) poses. We calculate the average gradient of the image in 8 directions in 25 bins around this feature point, as well as the mean intensity, for each RGB channel to give a total of 775 measurements (see Figure 5). We perform PCA on these measurements and project them into a subspace of dimension 100. We choose 30 to be the dimension of the invariant identity space in all cases. We treat these 10 local models as independent and multiply the evidence (Equation 12) for each.

We extracted 320 individuals from the FERET test set at seven poses (pl, hl, ql, fa, qr, hr, pr categories, $-90, -67.5, -22.5, 0, 22.5, 67.5, 90^\circ$). We divided these into a training set of 220 individuals and a test set of 100 individuals at each pose. We learn the parameters $\theta = \{\mathbf{F}_{1\dots K}, \mathbf{m}_{1\dots K}, \Sigma_{1\dots K}\}$ from the training set. We build six models, each describing the variation between one of the six non-frontal poses and the frontal pose. In each case we wish to identify which of the 100 frontal test images corresponds to a probe face at the second pose under consideration. We do this for all 100 probe faces and report the percentage of times that the maximum a posteriori model is correct. In this paper we do not consider the possibility that the probe face was not in the database (i.e. $Pr(\mathcal{M}_0) = 0$). This will be investigated in a separate publication. The results are

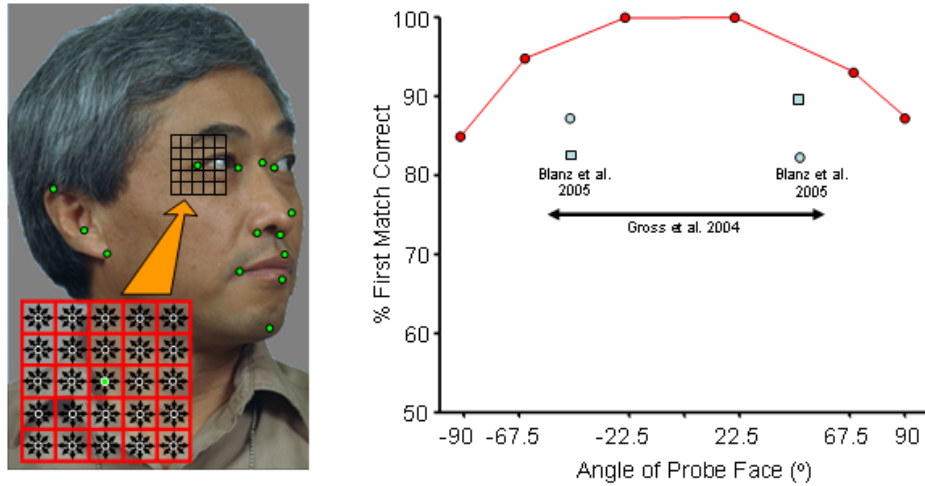


Figure 5: (Left) Features were marked by hand on each face and the smoothed intensity gradient was sampled at 25 positions and eight orientations around each. The mean intensity was also sampled at each point. (Right) Results for 100 frontal test faces in FERET database as a function of probe pose. Results from [6] and [3] (■=3D model, ●=reprojection) are indicated. Performance is significantly better with our system (see main text for a detailed comparison).

shown in Figure 5. Average performance at $\pm 22.5^\circ$ pose is 100% for this size database. Performance at $\pm 67.5^\circ$ is 95%. Even at extreme pose variations of $\pm 90^\circ$, we get 86% correct first choice matches.

4 Discussion

Our results compare favorably with previous studies. Gross et al. [6] report 75% first match results over 100 test faces from a different subset of the FERET database with a mean difference in absolute pose of 30° , and a worst case difference of 60° . Our system gives 95% performance with a pose difference of 62.5° for every pair. Blanz et al. [3] report results for a test database of 87 subjects with a horizontal pose variation of either $\pm 45^\circ$ from the Face Recognition Vendor Test 2002 database. They investigate both full co-efficient based 3D recognition (84.5%) performance and estimating the 3D model and creating a frontal image to compare to the test database (86.25% correct). Our system produces better performance over a larger pose difference in a larger database. To the best of our knowledge, there are no characteristics of our test database that make it easier or harder than those used in the above studies.

Moreover, our system has several desirable properties. First, it is fast relative to the sophisticated scheme of Blanz et al. [2] as it only involves linear algebra in relatively low dimensions and does not require an expensive non-linear optimization process. Second, it is fully probabilistic and provides a posterior over the possible matches. In a real system, this can be used to defer decision making and accumulate more data when the posterior does not have a clear spike. Third, it is possible to meaningfully consider the case that the probe face is not in the database without the need for arbitrarily choosing an acceptance

10

threshold. Fourth, there are only two parameters which need to be chosen. These are the dimensions of the observation space and the invariant identity space. Fifth, the system is relatively simple to train from two dimensional image pairs which are readily available.

It is interesting to consider why this relatively simple generative model performs so well. It should be noted that the model does not try to describe the true generative process, but merely to obtain accurate predictions together with valid estimates of uncertainty. Indeed, the performance for any given feature model (nose, eye etc.) is poor, but each provides independent information which is gradually accrued into highly peaked posterior. Nonetheless, a simple linear transformation is intuitively reasonable: if we consider faces that look similar at one pose, they probably also look similar to each other at another pose. The linear transformation maintains this type of local relationship.

In future work, we intend to investigate more complex generative models. The probabilistic recognition metric proposed in this paper is valid for all generative models, and works even when the “identity” space is discrete or admits no simple distance metric. We will also incorporate geometrical information which is not currently used in our system.

References

- [1] P.N. Belhumeur, J. Hespanha and D.J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection,” *PAMI*, Vol. 19, pp. 711-720, 1997.
- [2] V. Blanz, S. Romdhani and T. Vetter, “Face identification across different poses and illumination with a 3D morphable model,” *Int’l Conf. Face and Gesture Rec.* pp. 202-207, 2002.
- [3] V. Blanz, P. Grother, P. J. Phillips and T. Vetter, “Face Recognition Based on Frontal Views Generated from Non-Frontal Images,” *CVPR.*, pp. 454-461, 2005.
- [4] A. Dempster, N. Laird and D. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. Roy. Statist. Soc. B*, vol. 39, pp. 1-38, 1977.
- [5] A. Georghiades, P. Belhumeur and D. Kriegman, “From few to many: illumination cone models and face recognition under variable lighting and pose,” *PAMI*, vol. 23, pp. 129-139, 2001.
- [6] R. Gross, I. Matthews and S. Baker, “Appearance-Based Face Recognition and Light Fields.” *PAMI*, Vol. 26, pp. 449-465, 2004.
- [7] M.H. Yang “Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods” *Int’l Conf. Face and Gesture Recog.*, pp. 215-220, 2002.
- [8] A. Pentland, B. Moghaddam and T. Starner, “View-based and modular eigenspaces for face recognition,” *CVPR*, pp. 84-91, 1994.
- [9] S.J.D. Prince and J. Elder, “Invariance to nuisance parameters in face recognition,” *CVPR*, pp. 446-453, 2005.
- [10] S. Romdhani, V. Blanz and T.Vetter, “Face identification by fitting a 3D morphable model using linear shape and texture error functions,” *ECCV*, 2002.
- [11] M. Turk and A. Pentland, “Face Recognition using Eigenfaces,” *CVPR*, pp.586-591, 1991.
- [12] W. Zhao and R. Chellappa, “SFS based view synthesis for robust face recognition,” *Proc. International Conf. on Automatic Face and Gesture Rec.*, pp 285-292, 2002.
- [13] W. Zhao, R.Chellappa, A.Rosenfeld and J. Phillips, “Face Recognition: A literature Survey,” *ACM Computing Surveys*, Vol. 12, pp. 399-458, 2003.