Tied Factor Analysis for Face Recognition across Large Pose Differences

Simon J.D. Prince, *Member*, *IEEE*, James H. Elder, *Member*, *IEEE*, Jonathan Warrell, *Member*, *IEEE*, and Fatima M. Felisberti

Abstract—Face recognition algorithms perform very unreliably when the pose of the probe face is different from the gallery 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 identity space, the representation for each individual does not vary with pose. We model the measured feature vector as being generated by a pose-contingent linear transformation of the identity variable in the presence of Gaussian 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. We use the EM algorithm to estimate the linear transformations and the noise parameters from training data. We propose a probabilistic distance metric that allows a full posterior over possible matches to be established. We introduce a novel feature extraction process and investigate recognition performance by using the FERET, XM2VTS, and PIE databases. Recognition performance compares favorably with contemporary approaches.

Index Terms—Computing methodologies, pattern recognition, applications, face and gesture recognition.

1 Introduction

Face recognition systems can now achieve high performance under controlled image conditions. One of the greatest remaining research challenges in face recognition is to recognize faces across different poses, expressions, and illuminations [42]. In this paper, we address face recognition across poses, although our method is equally applicable to illuminations or expressions. In particular, we examine 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 gallery image. Under these conditions, commercial systems flounder: In the 2002 Face Recognition Vendor Test (FRVT) [25], 10 commercial systems were tested in an identification task, using 87 subjects with a 45 degree horizontal pose difference. The best achieved less than 50 percent correct rank-1 identification. In this paper, we present an algorithm that can produce significantly improved recognition performance, even when the pose variation is very significant.

Although the problem of face recognition across poses may seem esoteric, it has important real-world applications. Current face recognition systems require the implicit cooperation of the user, who is required to stand in a certain place, face the camera, and maintain a neutral expression. However, there are many situations where such a cooperation is not possible:

- *Face recognition from security footage.* People may be entirely unaware that the camera is present, and the positioning of cameras makes it unlikely that a pure frontal image will ever be captured. Indeed, in our previous work, we have developed a novel sensor that can capture high-quality human faces over a wide area [9].
- *Face recognition in archive footage.* There are many applications in which face recognition might be applied to archived photo or video footage. Examples include the semiautomatic labeling of identity in collections of photos, generating cast lists for movies and Internet searches for a given face image.
- *Face recognition for HCI and ambient intelligence.* There is a trend for computational devices to become smaller and more ubiquitous and to have more natural styles of interaction with the users. It is likely that future computational devices will have the ability to recognize their users rather than to demand an explicit logon procedure. It would be preferable for the user not to have to cooperate with this procedure by standing in a certain position.

In this paper, we present a method for face recognition across poses, which can potentially be applied to all of these goals.

1.1 Distance-Based Methods for Face Recognition

In this section, we consider why common face recognition methods fail when the pose varies and why common methods for suppressing "within-individual variance" cannot help. Most face recognition methods have the following common structure: the observed images are registered to a standard face shape. The registered image data is transformed to create a *feature vector* in a space of reduced dimensionality. The probe image and all of the gallery images are transformed this

- *S.J.D. Prince and J. Warrell are with the Department of Computer Science, University College London, London WC1E 6BT, UK. E-mail: {s.prince, j.warrell}@cs.ucl.ac.uk.*
- *J.H. Elder is with the Centre for Vision Research, York University, Room 003G, Computer Science Building, 4700 Keele Street, North York, Ontario, Canada M3J 1P3. E-mail: jelder@yorku.ca.*
- *F.M. Felisberti is with the Department of Psychology, Kingston University, Penrhyn Road, Kingston upon Thames, Surrey KT1 2EE, UK. E-mail: f.felisberti@kingston.ac.uk.*

Manuscript received 9 May 2007; revised 29 Nov. 2007; accepted 24 Jan. 2008; published online 22 Feb. 2008.
Recommended for acceptance by H. Wechsler.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2007-05-0270.
Digital Object Identifier no. 10.1109/TPAMI.2008.48.

way. The distance in feature space between the probe image and all of the gallery images is calculated. The probe image is associated with the closest gallery image. The logic for this approach is that for a suitable choice of transformation, the signal-to-noise ratio in the feature space is improved, relative to that of the original space. We will refer to this class of face recognition algorithms as "distance-based methods."

Within the class of distance-based methods, 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 [33] used principal component analysis to model image space as a multidimensional Gaussian and selected the projections onto the largest eigenvectors. Other work has variously investigated using different linear weighted pixel sums, [1], [13], analogous nonlinear techniques [39], and different distance measures [24].

Unfortunately, these distance-based methods fail when the pose of the probe and gallery faces significantly differ. The reason is that the pose change causes corresponding changes in the extracted feature vector. Indeed, variation attributable to pose may dwarf variation due to differences in identity. For example, in most common feature spaces, profile faces of different individuals are much closer to each other than the profile and frontal views of the same individual. Under these circumstances, it is inevitable that recognition performance will be poor.

An obvious approach to making recognition robust to pose is to remove all directions in the feature space that covary strongly with this variable. This technique is used to eliminate lighting variation, where one convention is to drop the largest eigenvector and retain 60 percent of the remaining eigenvectors [38]. A more elaborate version of the same idea is to measure the amount of signal (interpersonal variation) and noise (here primarily variation due to pose) along each direction and select feature directions where the signal-to-noise ratio is optimal. This approach was proposed by Belhumeur et al. [1] for frontal face recognition. A drawback of these approaches is that the discarded or suppressed dimensions still contain a significant portion of the signal and their elimination ultimately impedes recognition performance.

1.2 Algorithms for Face Recognition across Pose

The simplest method of generalizing across poses is to record each subject at each possible angle and use a statistical model for each [2], [16], [23]. A related approach is to take several images of the subject and use these to build a statistical model that can interpolate to unseen views [34]. Other methods make explicit use of geometric information and use several photos to create a 3D model of the head, which can then be rerendered at any given pose to compare with a given probe [11], [41]. The fourth approach is to actively seek to take an image from the correct pose [10]. It is also possible to use 3D measurements to perform face recognition and eliminate pose by aligning probe and gallery models. All of these methods are valid, and some produce high-quality results. However, they all require the cooperation of the user, multiple images, or special capture methods. They are consequently unsuitable for the tasks described above, where we may have only a single image of the individual, and we will not consider them further.

The second class of algorithms take a single probe image at one pose and create a full 3D head model for the subject based on just one image, including parameters representing the pose and illumination. We will term this the *geometric approach.* Face recognition can be performed in two distinct ways. The first method is to directly compare the parameters representing the shape and texture of the 3D model [6], [17], [29]. In the second approach, the 3D model can be used to rerender the face at a new pose, and 2D methods can be used [5], [40]. There does not seem to be very much empirical difference between these methods [5].

These geometric methods represent the state of the art in pose-invariant recognition. The system described in [5] achieved 86 percent first-match recognition performance on a database of 87 people, with a pose variation of $\pm 45^{\circ}$. Unfortunately, their approach is slow, since it requires iterative optimization of the model parameters. Their algorithm takes on the order of tens of minutes to create a 3D model from an image. This problem is partly mitigated if the second recognition style (rerendering) is employed, as the models are built for the gallery images offline, but the registration of a new individual to the system is still slow. These systems are not currently suitable for an application such as an Internet search for faces.

The third and most common approach to face recognition across poses is the *statistical approach.* Here, domain-specific information about the 3D world is eschewed, and the relationship between frontal and nonfrontal images is treated as a statistical learning problem. Similar to the geometric approach, there are two basic methods for face recognition. In the first approach, the statistical relationship is used to rerender frontal faces in nonfrontal views, or vice versa, and then, standard 2D face recognition methods are used. For example, Beymer and Poggio [3] used image warping to predict nonfrontal images from frontal images. Similarly, Wallhoff et al. [36] synthesized profile faces from frontal images by using a technique based on neural networks and hidden Markov models.

The second type of statistical approach aims at transforming features to a pose-invariant space. Identity is assigned based on distance in this space. For example, Maurer and von der Malsberg [20] extracted Gabor jet features at several positions on the face and then transformed these features when the face was nonfrontal to predict how they would appear for a frontal face. Similarly, Sanderson et al. [30] developed a Bayesian classifier based on mixtures of Gaussians and transformed the parameters of the model for nonfrontal views.

A further example of the statistical approach is the "eigenlightfields" work of Gross et al. [12]. In this approach, pose-invariant face recognition is treated 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 possible poses. The missing information can be estimated from the visible data based on prior knowledge of the joint probability distribution of the complete data set. This joint probability distribution is modeled as a multivariate Gaussian distribution in the complete set of images. In practice, the first few eigenvectors of this distribution are used for the recognition decision. Prince and Elder [28] presented a heuristic algorithm that extracted eigenfeatures for both frontal and nonfrontal faces. These features then undergo a pose-dependent transformation to a new feature space,

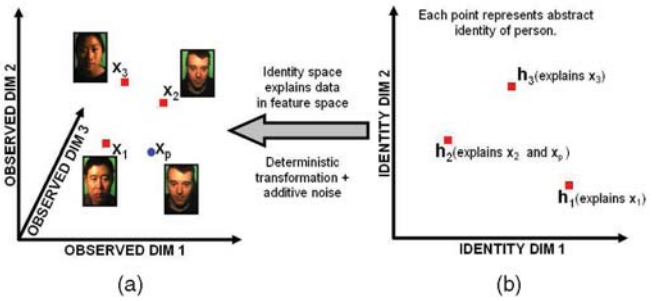

Fig. 1. The latent identity variable approach. (a) Three gallery faces (square symbols) and a probe face (circular symbol) represented in multivariate observation space. Each position in this space represents a different image. (b) The "identity space," in which each position depicts a different individual. Each image in (a) is modeled as having been generated from a particular point in the identity space in (b).

where the representation *does not vary with pose.* A closely related method based on the Linear Discriminant Analysis (LDA) was proposed by Kim and Kittler [15].

The statistical approaches discussed above have the advantage of relative speed and simplicity of implementation compared to the geometric approach. Unfortunately, the recognition performance of these methods is not, in general, as good as for 3D geometric approaches. For example, the eigenlightfields methods [12] yields 75 percent first-match correctness for a database of 100 people, with a mean absolute pose difference of 30 degrees.

One potential way of improving the results of the statistical approach is to build several models relating different parts of the face. Most of the above methods build models relating the whole face image at one pose to the whole face image at the other. We shall term these *global statistical models.* Lucey and Chen [18] introduced a method that separately models the statistical relationship between local patches of the frontal gallery image and the entire profile image. Other authors [14], [20] have developed models that relate several distinct regions of the frontal image to their counterparts in non-frontal images. We term these *local statistical models*.

1.3 Overview

There is no current method for face recognition that provides good performance at large pose differences. The current state of the art is based on constructing a geometric 3D model from a single image and iteratively estimating pose, illumination, and model parameters. Unfortunately, such methods are complex to implement and are computationally expensive. The alternative is to build a statistical model. Such methods are simpler and computationally cheaper but produce relatively poor results. In this paper, we develop a statistical model that is fast and simple to implement and produces results that are superior to the current state of the art.

Our algorithm has the following distinctive characteristics:

- It is *probabilistic* and provides a posterior probability for the matching to a gallery (identification) or for whether the two faces match or belong to different people (verification).
- The algorithm is based on a generative model that describes how an underlying pose-invariant representation created the (pose-varying) observed data (see Fig. 1). This is in contrast to most existing algorithms, where the direction of information flow is from the observed image to the pose-invariant representation.
- In matching, we ask the question: "What is the probability that two images were created from the same underlying representation?" However, we acknowledge that this underlying representation is uncertain and never form an explicit point estimate.
- We acknowledge that modeling the relationship between the entire frontal and nonfrontal faces (the global approach) is too challenging. Instead, we build several local models describing how each individual facial feature (nose, eye, etc.) changes with pose. We combine information from each model by using naive Bayes to make a final recognition decision.

In Section 2, we introduce the problem of pose variation as seen from the space of observed data. We propose a simpler underlying representation, which we term *identity space*, and a generative model that creates the complex observed data from the simpler identity space. In Section 3, we demonstrate how we can learn the parameters of the mapping between these two data spaces from training data by using the Expectation-Maximization (EM) algorithm. In Section 4, we present several ways of visualizing the results of this learning procedure. Subsequently, in Section 5, we demonstrate how our generative model can be used to perform recognition decisions. In Section 6, we present a series of experiments investigating the performance of this model. In Section 7, we compare the theoretical properties and empirical performance of our algorithm to contemporary approaches.

2 Observed and Identity Spaces

2.1 Observed Image Data

In this section, we discuss the characteristics of the observed image data. We define observed data to mean either the raw gray values of the image or some simple deterministic transformation of these values, which does not attempt to compensate for pose variations. We assume that the observed data is vectorized to form an observed data vector.

For most common choices of observed data vector, the majority of positions in the 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 nonlinear probabilistic region tracing through multidimensional observation space. The manifold has two key characteristics that must be captured by our model. First, the mean position in the manifold changes systematically with the pose of the face. Second, for a given individual, the position of the observation vector, relative to this mean, varies. These two characteristics are illustrated in Fig. 2.

These properties account for why face recognition is poor when the observed vectors are used directly, and there is significant pose variation. The first property implies that a face belonging to a particular individual can appear in very different parts of the manifold, depending on its pose. As shown in Fig. 2, there is no simple distance metric in this space that supports good recognition performance. The second property implies that, even if we were to compensate for the average shift due to the pose change, the performance would probably not improve.

2.2 Identity Space Representation

Since the observed feature space is problematic in terms of recognition, we hypothesize an underlying representation

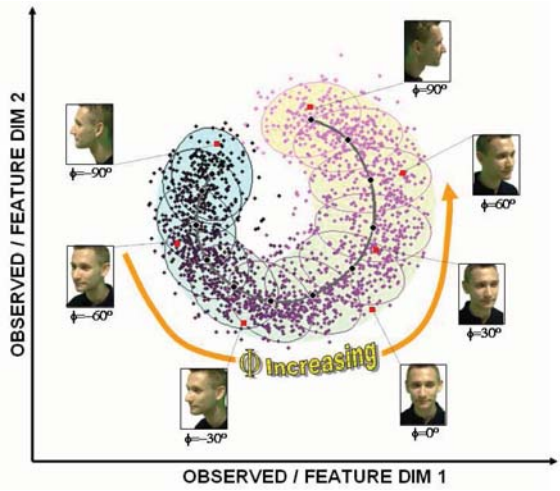

Fig. 2. The effect of pose variation in the observation (and feature) space. Face pose is coded by intensity so that faces with poses near -90° are represented by dark points and faces with poses near 90° are represented by light points. The pose variable is quantized into K bins, and each bin is represented by a Gaussian distribution (ellipses). The K Means of these Gaussians trace a path through a multidimensional 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 not a simple metric in this space, which will identify these points with one another.

with more optimal properties. At the core of our algorithm is the notion that there genuinely exists a multidimensional variable $\mathbf{h}$ that represents the identity of the individual, regardless of the pose. We term the space of possible values for this variable as *identity space*, and the variable itself is termed a *latent identity variable*.

Latent identity variables (LIVs) have this key property: If two LIVs take the same value, they represent the same person. If they take different values, they represent different people. In general, latent identity variables may be discrete or continuous and may have a variety of topological properties. In this paper, we will consider identity as a vector of real values representing a point in a multidimensional space, but we stress that this need not always be the case.

2.3 From Identity Space to Observed Space

In this section, we describe a Bayesian generative model that creates data that closely follows the face manifold from the simpler underlying identity representation. The identity variable takes the role of a *latent* or *hidden* variable in the context of this model. In particular, it is assumed that each observed data can be described as the result of the following process:

1. Choose the point in the identity space that corresponds to the individual for which we create image data from some prior distribution.
2. Choose a pose (also from a prior distribution).
3. Transform this identity variable to the observation space by using a deterministic function. This function depends on the pose.
4. Add noise to the resulting observation vector.

Step 2 in the above generation process induces the pose dependence of the observed data vector: the transformation from the identity space to the observation space is different for different poses. The addition of the noise in Step 4 has two implications. First, it provides an explanation as to why repeated images of the same person at the same pose are not exactly the same. Second, it means that for a given observed feature vector, we can never be exactly sure which identity was responsible. The best that we can do is to calculate a posterior distribution over possible values.

Note that this structure broadly describes the actual generation process. One can consider the latent identity variable as describing the shape and structure of the face. The function relating the identity variable to the observed image represents the perspective projection process, which is parameterized by pose. The noise term represents the genuine measurement noise in the camera, plus all unmodeled aspects of the situation such as expression and lighting variation.

In principle, we could describe the full 3D geometric projection process in this framework, but in practice, we use a simpler generative model. This does not have any physical validity but can still be used to make accurate inferences about identity, together with appropriate measures of uncertainty. We now provide details of this generative model.

2.4 Tied Factor Analysis

Pose is assumed to be discretized into K different bins. For notational convenience, we will assume that there are J examples of K poses for each of I different individuals. We denote the jth image of individual i in the kth pose by $\mathbf{x}_{ijk}$. We assume that this data was generated from an underlying latent identity variable, which we denote $\mathbf{h}_i$. The dimensionality of the observed and the identity spaces are, in general, different, and it is usual for the identity space to be of smaller dimensionality than the observed space. The deterministic mapping between the identity and the observed spaces is affine. It comprises a set of offsets $\mathbf{m}_{1,\ldots,K}$ and a set of linear functions (matrices) $\mathbf{F}_{1,\ldots,K}$. There is one offset and one linear function specialized for each discretized pose k. The generative process can hence be described as

$$\mathbf{x}_{ijk} = \mathbf{F}_k \mathbf{h}_i + \mathbf{m}_k + \epsilon_{\mathbf{ijk}}, \tag{1}$$

where ϵ_{ijk} is a zero-mean multivariate Gaussian noise term with an unknown *diagonal* covariance matrix Σ_k. Note that the noise depends on the particular pose chosen. More formally, we write the model in terms of conditional probabilities

$$Pr(\mathbf{x}_{ijk}|\mathbf{h}_i) = \mathcal{G}_{\mathbf{x}}[\mathbf{F}_k \mathbf{h}_i + \mathbf{m}_k, \Sigma_k], \tag{2}$$

$$Pr(\mathbf{h}_i) = \mathcal{G}_{\mathbf{h}}[\mathbf{0}, \mathbf{I}], \tag{3}$$

where $\mathcal{G}_{\mathbf{a}}[\mathbf{b}, \mathbf{C}]$ denotes a Gaussian in $\mathbf{a}$ with mean $\mathbf{b}$ and covariance $\mathbf{C}$.

This model is closely related to factor analysis. The factors $\mathbf{F}_k$ depend on pose, but the factor loadings $\mathbf{h}_i$ are the same at each pose (tied). Hence, we term this generative model *tied factor analysis.* The relationship between the observation space $\mathbf{x}$ and the identity space $\mathbf{h}$ is indicated in Fig. 3. It can be seen that vectors in widely varying parts of the original image space can be generated from the same point in the identity space as required.

To complete the definition of the generative model, we need to define a prior on the latent identity variables $\mathbf{h}$. The

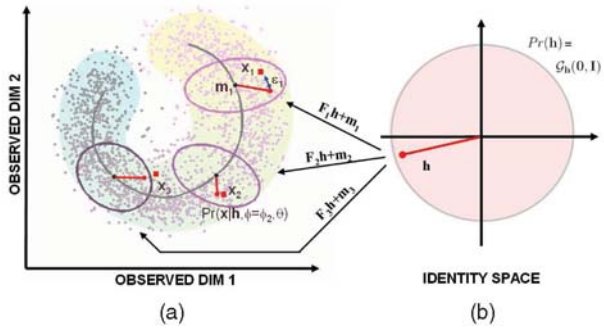

Fig. 3. Tied factor analysis model. (a) Observed measurement space. (b) "Identity" space. Latent identity variables in this space have a prior distribution $Pr(\mathbf{h}) = \mathcal{G}_{\mathbf{h}}(\mathbf{0}, \mathbf{I})$. The three square symbols in (a) represent observed data for one person viewed at three poses $k = \{1, 2, 3\}$. The circle symbol in (b) represents the latent identity variable for this person. Data in the observation space $\mathbf{x}_k$ are explained by transforming latent identity variable $\mathbf{h}$ by a pose-dependent affine transform $\mathbf{F}_k\mathbf{h} + \mathbf{m}_k$ and by adding noise ϵ_k.

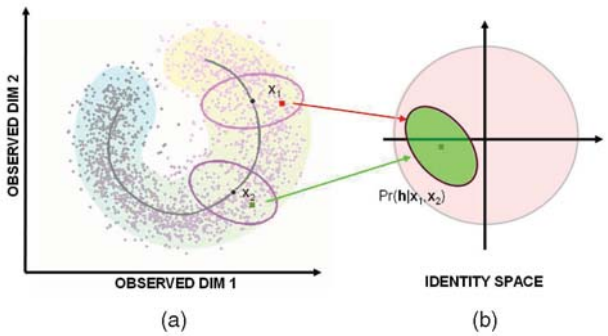

Fig. 4. (b) In the E-Step, we aim at calculating the posterior probability distribution over the latent identity variables. (a) This is inferred from the observed images. Here, two image data points $\mathbf{x}_1$ and $\mathbf{x}_2$ at different poses are used to find the posterior over $\mathbf{h}$.

prior is assumed to be a zero-mean Gaussian with identity covariance $\mathbf{I}$. This is required to ensure that the learning process (see Section 3) converges. In common with other approaches, we will consider the pose to be known a priori for all images, rather than treating it as a random variable. We do not consider this to be a serious restriction, however: The original problem was that the observed image is highly dependent on the pose, so it follows that coarse pose information is easy to recover from the images.

Note that in our generative model, information flows in the *opposite* direction to the distance-based approach. Instead of taking the observed data and transforming it forward to a feature space, we hypothesize underlying latent identity variables that explain the observed data.

3 Learning System Parameters

In this section, we aim at learning the unknown parameters of the generative model. These are the functions $\mathbf{F}_k$, the means $\mathbf{m}_k$, and the noise parameters Σ_k. We aim at adjusting the parameters $\boldsymbol{\theta} = \{\mathbf{F}_{1,\ldots,K}, \mathbf{m}_{1,\ldots,K}, \Sigma_{1,\ldots,K}\}$ to increase the joint likelihood $Pr(\mathbf{x}, \mathbf{h}|\boldsymbol{\theta})$ of the observed image data $\mathbf{x}$ and the associated identity variables $\mathbf{h}$. Unfortunately, we cannot observe the identity vectors directly: we can only infer a posterior distribution over them for some fixed set of parameters θ. This type of chicken-and-egg problem is suited to the EM algorithm [8]. We iteratively maximize

$$Q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1}) = \sum_{i=1}^{I} \int Pr(\mathbf{h}_i|\mathbf{x}_{i\bullet\bullet}, \boldsymbol{\theta}_{t-1}). \left[\sum_{j=1}^{J}\sum_{k=1}^{K} \log Pr(\mathbf{x}_{ijk}|\mathbf{h}_i, \boldsymbol{\theta}_t) + \log Pr(\mathbf{h}_i)\right] d\mathbf{h}_i, \tag{4}$$

where t represents the iteration index, and $\mathbf{x}_{i\bullet\bullet}$ denotes all the data associated with individual i (i.e., all J repetitions at each of the K poses). The first of these probability terms will be calculated in the E-Step. The second two terms were given by (2) and (3).

The EM algorithm alternately finds the expected values for the unknown identity variables $\mathbf{h}$ (the E-Step) and then maximizes a lower bound on the overall likelihood of data as a function of the parameters $\boldsymbol{\theta}$ (the M-Step). More precisely, the E-Step calculates the expected values of the identity variable $\mathbf{h}_i$ for each individual i by using the data for that individual across all poses $\mathbf{x}_{i\bullet\bullet}$. The M-Step optimizes the values of the transformation parameters $\{\mathbf{F}_k, \mathbf{m}_k, \Sigma_k\}$ for each pose k by using data for that pose across all individuals and repetitions $\mathbf{x}_{\bullet\bullet k}$. These steps are repeated until convergence.

E-Step. For each individual, we estimate the distribution of $\mathbf{h}_i$, given the parameter estimates $\boldsymbol{\theta}_{t-1}$ at the previous time $t-1$ and all the data associated with that individual (see Fig. 4). The posterior distribution for the latent identity variable can be calculated using Bayes' rule

$$Pr(\mathbf{h_i}|\mathbf{x}_{i\bullet\bullet}, \boldsymbol{\theta}_{t-1}) = \frac{Pr(\mathbf{x}_{i\bullet\bullet}|\mathbf{h}_i, \boldsymbol{\theta}_{t-1})Pr(\mathbf{h}_i)}{\int Pr(\mathbf{x}_{i\bullet\bullet}|\mathbf{h}_i, \boldsymbol{\theta}_{t-1})Pr(\mathbf{h}_i)d\mathbf{h}_i}. \tag{5}$$

We assume that the likelihood of each data point from individual i is independent so that

$$Pr(\mathbf{x}_{i\bullet\bullet}|\mathbf{h}_i, \boldsymbol{\theta}_{t-1}) = \prod_{j=1}^{J}\prod_{k=1}^{K} Pr(\mathbf{x}_{ijk}|\mathbf{h}_i, \boldsymbol{\theta}_{t-1}), \tag{6}$$

where the terms on the right-hand side are calculated from the forward model (2). Since all terms on the right-hand side of (5) are normally distributed, the left-hand side is also normally distributed and can be represented with a mean vector and a covariance matrix. The first two moments of this distribution can be shown to equal

$$\begin{aligned} \mathrm{E}[\mathbf{h}_i|\mathbf{x}_{i\bullet\bullet}] &= \left(\mathbf{I} + \sum_{j=1}^{J}\sum_{k=1}^{K} \mathbf{F}_k^T\Sigma_k^{-1}\mathbf{F}_k\right)^{-1} \cdot \\ &\quad \sum_{j=1}^{J}\sum_{k=1}^{K} \mathbf{F}_k^T\Sigma_k^{-1}(\mathbf{x}_{ijk} - \mathbf{m}_k) \\ \mathrm{E}[\mathbf{h}_i\mathbf{h}_i^T|\mathbf{x}_{i\bullet\bullet}] &= \left(\mathbf{I} + \sum_{j=1}^{J}\sum_{k=1}^{K} \mathbf{F}_k^T\Sigma_k^{-1}\mathbf{F}_k\right)^{-1} \\ &\quad + \mathrm{E}[\mathbf{h}_i|\mathbf{x}_{i\bullet\bullet}]\mathrm{E}[\mathbf{h}_i|\mathbf{x}_{i\bullet\bullet}]^T. \end{aligned} \tag{7}$$

More details on these calculations are provided in the Appendix.

M-Step. For each pose k, we maximize the objective function $Q(\boldsymbol{\theta}_t, \boldsymbol{\theta}_{t-1})$, as defined in (4), with the respect to the parameters $\boldsymbol{\theta}$. For simplicity, we estimate the mean $\mathbf{m}_k$ and linear transform $\mathbf{F}_k$ at the same time. To this end, we create

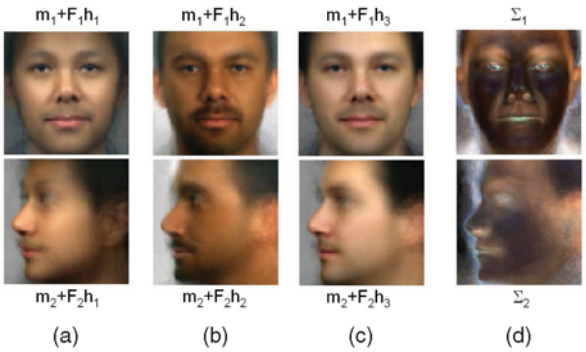

Fig. 5. Tied factor analysis model, with 16 factors learned from FERET training data. (a), (b), and (c) Three points in the identity space projected back into the observation space through frontal and profile models. In each case, the frontal and profile images look like the same person. (d) Per-pixel noise terms Σ for frontal and profile models. Brighter points represent pixels with more noise.

Fig. 6. Prediction of nonfrontal faces from frontal faces by using the tied factor analysis model with 16 factors. (a) Actual images of subject (not in the training database). The frontal image (highlighted in red) is used to predict nonfrontal faces as described in the text. (b) Predicted images for six different poses. (c) (left) One more good example of profile image prediction (left to right: frontal, predicted profile, and actual profile) and (right) one poor example.

new matrices $\tilde{\mathbf{F}}_k = [\mathbf{F}_k\ \mathbf{m}_k]$ and $\tilde{\mathbf{h}}_i = [\mathbf{h}_i^T\ 1]^T$. The first log probability term in (4) can be written as

$$\log[Pr(\mathbf{x}_{ijk}|\mathbf{h}_i, \boldsymbol{\theta}_t)] = \kappa + \frac{1}{2}\left(\log|\Sigma_k^{-1}| - (\mathbf{x}_{ijk} - \tilde{\mathbf{F}}_k\tilde{\mathbf{h}}_i)^T\Sigma_k^{-1}(\mathbf{x}_{ijk} - x\tilde{\mathbf{F}}_k\tilde{\mathbf{h}}_i)\right), \tag{8}$$

where κ is an unimportant constant. We substitute this expression into (4) and take derivatives with respect to each $\tilde{\mathbf{F}}_k$ and Σ_k. The second log term in (4) had no dependence on these parameters and disappears from the derivatives. These derivative expressions are equated to zero and are rearranged to provide the following update rules:

$$\tilde{\mathbf{F}}_k = \left(\sum_{i=1}^{I}\sum_{j=1}^{J}\mathbf{x}_{ijk}\mathrm{E}\left[\tilde{\mathbf{h}}_i|\mathbf{x}_{i\bullet\bullet}\right]^T\right). \left(\sum_{i=1}^{I}\sum_{j=1}^{J}\mathrm{E}\left[\tilde{\mathbf{h}}_i\tilde{\mathbf{h}}_i^T\middle|\mathbf{x}_{i\bullet\bullet}\right]\right)^{-1}, \tag{9}$$

$$\Sigma_k = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J}\mathbf{diag}\left[\mathbf{x}_{ijk}\mathbf{x}_{ijk}^T - \tilde{\mathbf{F}}_k\mathrm{E}\left[\tilde{\mathbf{h}}_i|\mathbf{x}_{i\bullet\bullet}\right]\mathbf{x}_{ijk}^T\right], \tag{10}$$

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

4 Learning Results

Before explaining how face recognition can be performed with this model, we describe the results of the learning process and confirm that the model has successfully learned the relationship between frontal and nonfrontal faces. We extracted 320 individuals from the FERET database [27] at seven poses pl, hl, ql, fa, qr, hr, and pr and categories -90, -67.5, -22.5, 0, 22.5, 67.5, and 90°). We divided these into a training set of 220 individuals and a test set of 100 individuals at each pose. Images were segmented from the background by using an iterative graph-cuts procedure and were placed against a mid-gray background. We identified 21 image features on each face by hand (automated feature detection is investigated in Section 6.5). These were used to register each image to a standard template by using a piecewise linear warp. Each image was resized to $70 \times 70 \times 3$. We concatenate the pixel values from the red, green, and blue (RGB) channels of the input image to make one long observation vector.

We learn the parameters $\boldsymbol{\theta} = \{\mathbf{F}_{1,\ldots,K}, \mathbf{m}_{1,\ldots,K}, \Sigma_{1,\ldots,K}\}$ from the training set. We build six models, each describing the variation between one of the six nonfrontal poses and the frontal pose. In each case, we applied 10 iterations of the EM algorithm. The only free parameter in the model is the number of dimensions of the hidden variables $\mathbf{h}$ (and, hence, the number of columns in the matrices $\mathbf{F}_{1,\ldots,K}$). This parameter is explored in the subsequent recognition experiments.

In Fig. 5, we visualize the resulting values. In Figs. 5a, 5b, and 5c, we take three different values of the latent identity variable $\mathbf{h}_1$, $\mathbf{h}_2$, and $\mathbf{h}_3$ and generate observations from them at two different poses. In each case, the generated images look like the same person: the algorithm has successfully learned the relationship between faces at different poses. In Fig. 5d, the diagonal noise terms Σ are shown for the frontal and profile cases. These indicate which parts of the image are least predictable from the deterministic part of the model. Unsurprisingly, this tends to be at high contrast features and on the edge of the face.

A second way of investigating this model is to use it to predict nonfrontal faces from frontal images. In order to do this, we calculate the posterior distribution over the latent identity variable $\mathbf{h}$, given the frontal face [see (5)]. We then project the mode of this distribution back to the observed space $\mathbf{x}$ by using one of the nonfrontal factor models $(\mathbf{F}_k, \mathbf{m}_k)$. The results of this process are demonstrated in Figs. 6a and 6b for each of the six tied factor models. These predictions resemble the actual images of the person at different poses. One can see that the pose bins in the FERET database are not very accurate: In several cases, the model predicts a face at a slightly different pose from the actual position. In Fig. 6c, we show (left) one more good example and (right) a bad one. The training data contained no one with white facial hair and, hence, the prediction is poor.

The previous investigation yields the most likely nonfrontal face, given the observed frontal face; however, our model is fundamentally Bayesian in nature and describes a probability distribution over the predicted images. In order to visualize this, we employ the following procedure: As

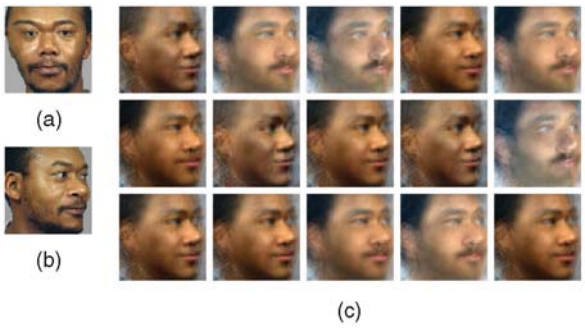

Fig. 7. Prediction of nonfrontal faces by using the tied factor analysis model with 16 factors. (a) Frontal image of subject. (b) Actual nonfrontal image of subject. (c) Fifteen samples from distribution of predicted images. More properly, we should have added independent noise at each pixel sampled from Σ, but these images are harder to interpret.

before, we calculate the posterior distribution over the latent identity variable, given the frontal face. However, we now sample from this Gaussian posterior and project each sample back down to the image space by using one of the nonfrontal models. Such image samples for one face are shown in Fig. 7. They all resemble the profile face.

5 Recognition

In the previous sections, we described how we can learn the parameters $\boldsymbol{\theta} = \{\mathbf{F}_{1,\ldots,K}, \mathbf{m}_{1,\ldots,K}, \Sigma_{1,\ldots,K}\}$. We have also demonstrated how we can use this model to predict how a face will look at a different pose. In principle, we could use this capability to convert all images in our test database to frontal and then use conventional face recognition techniques as in [5]. However, it is not clear how we can exploit knowledge about the uncertainty in the predicted image (as in Fig. 7). In this section, we present a probabilistic approach to face identification. We discuss face verification in Section 6.3. In both cases, the approach has the following characteristics:

- The criteria for a gallery and probe face matching are that the observed data vectors are explained by *exactly the same* value of the identity variable.
- Since our observations are noisy, we can never be sure which value the identity variable takes. Hence, we integrate out the hidden identity variable to give a final formulation that *does not depend on an estimate of* $\mathbf{h}$.
- The final decision is based on a calculation of the relative likelihood that the observed vectors were explained by different configurations of the underlying set of identity variables.

In face identification tasks, we are given a gallery database of faces $\mathbf{x}_{1,\ldots,N}$, each of which belongs to a different individual. We are also given a single probe face $\mathbf{x}_p$. Note that this represents a change in notation from that used in the previous sections. Our goal is to determine the posterior probability that each gallery face matches the probe face.

We frame the recognition task in terms of model comparison. We compare evidence for N models, which we denote by $\mathcal{M}_{1,\ldots,N}$. The nth model $\mathcal{M}_n$ represents the case where the probe matches the nth gallery face: we assume that there are only N underlying identity variables $\mathbf{h}_{1,\ldots,N}$, each of which generated the corresponding observed feature vector $\mathbf{x}_{1,\ldots,N}$. For the nth model, the nth identity

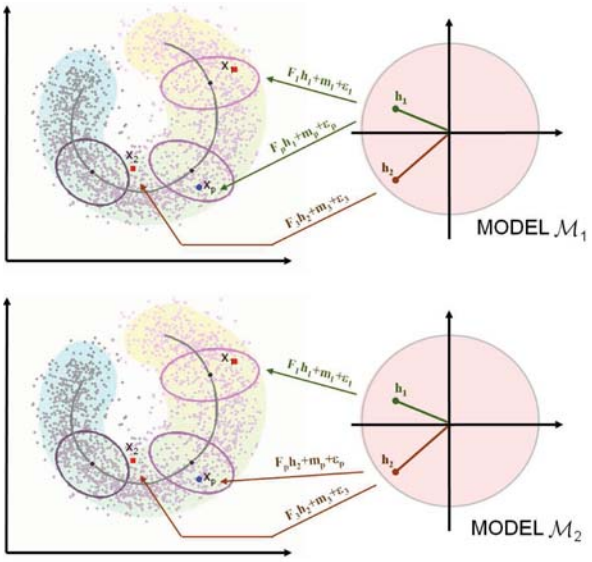

Fig. 8. Face identification. Given a probe face $\mathbf{x}_p$ and two gallery faces $\mathbf{x}_1$ and $\mathbf{x}_2$, there are two associated models, $\mathcal{M}_1$ and $\mathcal{M}_2$. In $\mathcal{M}_1$, the identity space variable $\mathbf{h}_1$ explains both the first gallery image $\mathbf{x}_1$ and the probe image $\mathbf{x}_p$. The second identity variable $\mathbf{h}_2$ explains the second observed image $\mathbf{x}_2$. This corresponds to the case where the probe image $\mathbf{x}_p$ matches gallery image $\mathbf{x}_1$. In the second model $\mathcal{M}_2$, the generation of the probe image is ascribed to the second identity variable $\mathbf{h}_2$. This corresponds to the case where the probe image $\mathbf{x}_p$ matches gallery image $\mathbf{x}_2$.

variable $\mathbf{h}_n$ is also deemed responsible for having generated the probe feature vector $\mathbf{x}_p$ (i.e., $\mathbf{h}_p = \mathbf{h}_n$). Fig. 8 shows this scheme for a gallery of two individuals.

The evidence for model $\mathcal{M}_n$ is given by

$$\begin{aligned} &Pr(\mathbf{x}_{1,\ldots,N}, \mathbf{x}_p|\mathcal{M}_n) \\ &= \int Pr(\mathbf{x}_{1,\ldots,N}, \mathbf{x}_p, \mathbf{h}_{1,\ldots,N}, \mathbf{h}_p|\mathbf{h}_p = \mathbf{h}_n) d\mathbf{h}_{1,\ldots,N} \\ &= \int Pr(\mathbf{x}_1, \mathbf{h}_1) d\mathbf{h}_1 \ldots \int Pr(\mathbf{x}_n, \mathbf{x}_p, \mathbf{h}_n) d\mathbf{h}_n \ldots \\ &\quad \int Pr(\mathbf{x}_N, \mathbf{h}_N) d\mathbf{h}_N \\ &= \int Pr(\mathbf{x}_1|\mathbf{h}_1) Pr(\mathbf{h}_1) d\mathbf{h}_1 \ldots \\ &\quad \int Pr(\mathbf{x}_n, \mathbf{x}_p|\mathbf{h}_n) Pr(\mathbf{h}_n) d\mathbf{h}_n \ldots \\ &\quad \int Pr(\mathbf{x}_N|\mathbf{h}_N) Pr(\mathbf{h}_N) d\mathbf{h}_N. \end{aligned} \tag{11}$$

Note that we marginalize over the uncertain identity variable rather than commit ourselves to one value. The terms in the last line were defined in (2) and (3) and, for our model, are Gaussian, so these integrals are tractable. Each has the following form:

$$\int Pr(\mathbf{x}_{1,\ldots,Q}, \mathbf{h}) d\mathbf{h}, \tag{12}$$

where the number of images Q takes values 1 or 2, depending on the term from (11), but might take larger values if we were assessing hypotheses that more images belonged to the same person. In order to calculate this

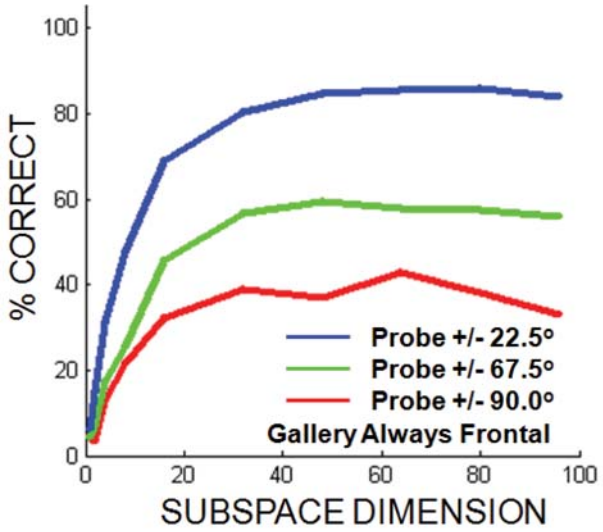

Fig. 9. Percentage of first-match correct performance with the tied factor analysis model as a function of the dimension of the latent identity variables. There were 100 frontal gallery faces and a single nonfrontal face, with an absolute pose difference that is different for each curve.

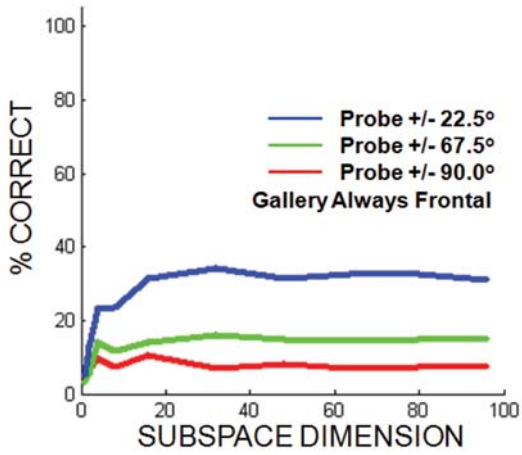

Fig. 10. Percentage of first-match correct performance with the factor analysis model as a function of the dimension of the latent identity variables. There were 100 frontal gallery faces and a single nonfrontal face, with an absolute pose difference that is different for each curve. Note that the performance is uniformly worse than in Fig. 9.

integral, we reformulate the generative equation as a standard factor analyzer, for which the solution is known:

$$\begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_Q \end{bmatrix} = \begin{bmatrix} \mathbf{F}_1 \\ \vdots \\ \mathbf{F}_Q \end{bmatrix} \mathbf{h}_i + \begin{bmatrix} \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_Q \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_Q \end{bmatrix},$$

$$\text{or} \quad \mathbf{x}' = \mathbf{F}' \mathbf{h}_i + \mathbf{m}' + \epsilon',$$

where $\mathbf{F}_q$, $\mathbf{m}_q$, and ϵ_q represent the factor matrix, mean, and noise term associated with the pose of the qth face. These equations now take the form of a standard factor analyzer, and the likelihood (i.e., the solution to the integral) is $\mathcal{G}_{\mathbf{x}'}\left[\mathbf{m}', \mathbf{F}'\mathbf{F}'^T + \Sigma'\right]$, where Σ' is $\mathbf{diag}[\Sigma_1, \ldots, \Sigma_P]$.

Having calculated the evidence for each different model, it is simple to calculate the posterior over the possible matches by using the Bayes rule

$$Pr(\mathcal{M}_n|\mathbf{x}_{1\ldots N}, \mathbf{x}_p, \boldsymbol{\theta}) = \frac{Pr(\mathbf{x}_{1\ldots N}, \mathbf{x}_p|\mathcal{M}_n, \boldsymbol{\theta})Pr(\mathcal{M}_n)}{\sum_{m=1}^{N} Pr(\mathbf{x}_{1\ldots N}, \mathbf{x}_p|\mathcal{M}_m, \boldsymbol{\theta})Pr(\mathcal{M}_m)}. \quad (13)$$

Note that the terms $Pr(\mathcal{M}_n)$ are the prior probability for each model. In our experiments, this is set to the uniform value of $1/N$ for each model. However, it is conceivable that in a real application, some users are expected to be seen more often than others and these values might vary. The final recognition decision is made by choosing the maximum a posteriori model.

6 Experiments

6.1 Experiment 1: Face Identification Using Raw Pixel Data

In order to test the tied factor analysis model, we first use the pixel values from the $70 \times 70 \times 3$ images as the observed data (as in Section 4). We use 100 individuals from the FERET database [27] who were not part of the training set. On each trial, the algorithm takes a nonfrontal probe image and aims at identifying which of the 100 frontal gallery faces is the correct match. For this and all subsequent experiments, it is assumed that the pose of each face is correctly identified. A tied factor analysis model is used, which was trained relating only the two poses that feature in the experiment. For each trial, we calculate the likelihood of the data under each of the 100 models and consider the maximum a posteriori model from (13) to be the estimated match.

There are only two parameters in the experiment. The first is the pose of the nonfrontal faces: we investigate ± 22.5, 67.5, and $90°$. The second is the dimension of the latent identity variables. In Fig. 9, we plot the percentage of first-match correct performance as a function of both of these parameters. We have pooled the data from left-facing and right-facing faces for each magnitude of pose difference, so each point on the graph was generated from a total of 200 trials. The peak performance is 83 percent for $\pm 22.5°$, 59 percent for $\pm 67.5°$, and 41 percent for $\pm 90.0°$. There is a steady increase in performance with the dimension of the subspace until approximately 64 dimensions, after which performance plateaus or exhibits a small decline.

In order to better analyze the success of our method, we compare to a case where no effort has been made to compensate for the pose variation. In Fig. 10, we present results from a "factor analysis model." This is exactly the same as the tied model, but now, there is only a single set of generation parameters $\mathbf{F}$, $\mathbf{m}$, and Σ. These are learned using data from both the frontal and nonfrontal poses. It is clear that performance here is worse: the peak performance is 36 percent for $\pm 22.5°$, 16 percent for $\pm 67.5°$, and 12 percent for $\pm 90.0°$. We would expect much the same performance from the eigenfaces algorithm [33].

From these experiments, we conclude that the tied factor analysis model significantly improves performance relative to a model where no attempt is made to compensate for pose differences. However, performance stills fall short of the state of the art: Blanz et al. [5] achieved 86 percent first-match recognition performance with the pose variation of $\pm 45°$, whereas this model only manages 83 percent performance with $\pm 22.5°$. One limitation of our method may be that it is unrealistic to expect this simple model, applied to global pixel

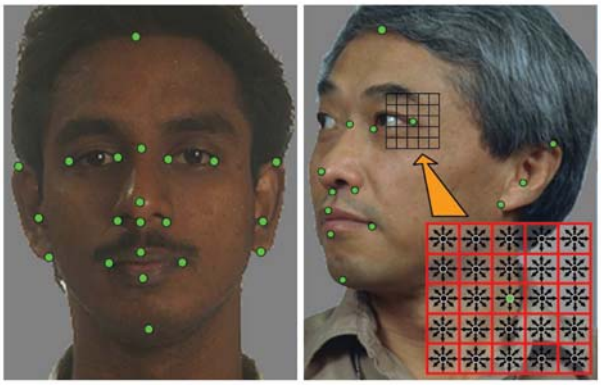

Fig. 11. Local measurements. (a) For registration, 21 positions on each face were identified by hand. (b) In recognition, the subset of features that were visible at all poses were chosen. For each feature, image gradients in eight directions and three scales were extracted at 25 spatial positions around the feature point.

data, to accurately describe the variations in the entire image. Hence, we now improve the preprocessing and combine a series of local models to make the recognition decision.

6.2 Experiment 2: Face Identification with Local Gabor Data

In order to register the images, we identified keypoints on each face. We now build a separate tied factor analysis model to describe the data around these keypoints. We only build models for the subset of 14 points that are not occluded, even when the face is in full profile. We start with a 400×400 image and calculate the image gradient in eight directions and three scales and the mean intensity at 25 spatial positions around each feature. This is done in each RGB channel to give a total of 1,875 measurements, which are normalized to length one. The 25 sampling positions are arranged in an 5×5 axis-oriented grid. This measurement scheme is shown in Fig. 11. In training, we learn 14 separate tied factor analyzers, where each is associated with one feature. In recognition, we calculate a likelihood for each possible model of the data for each features. We treat each of these likelihoods as independent and take the product to calculate the final likelihood for the Bayes rule in (13).

In Fig. 12a, we repeat the previous experiment by using these local features. Once more, we present results as a function of the difference in pose between gallery and probe images and the subspace dimension, which was always the same for each of the 14 local models. Once more, performance increases as a function of this subspace dimension but now peaks at around 32 dimensions. The peak performance is 100 percent for $\pm 22.5^\circ$, 99 percent for $\pm 67.5^\circ$, and 92 percent for $\pm 90.0^\circ$. Comparison with Fig. 9 demonstrates that the local features yield a significant improvement.

In Figs. 12b and 12c, we run the same experiment by using the XM2VTS and PIE databases, using identical preprocessing. For the XM2VTS database, we train with the first 195 individuals and test with the remaining 100. We use frontal gallery faces and left-facing profile faces. When these come from the same recording session, the results are very similar to those for the FERET database, with a peak performance of 91 percent. When the faces were taken from different recording sessions (first versus fourth), the performance drops to give a peak of 77 percent. For the PIE database, we used the first 34 individuals for training and the last 34 for testing. For frontal gallery images (pose condition C27) and probe faces at 16° (pose C05), we yield a 100 percent correct performance. With a pose difference of 62° (pose C22), we get a peak of a 91 percent performance. We conclude that our algorithm works well for several different data sets. The remaining experiments (with the exception of Experiment 5) are confined to the FERET data.

We conclude that building a number of local Gabor models vastly improves the performance in cross-pose recognition. There are several reasons for this improvement: 1) the underlying image resolution was greater, 2) Gabor features are known to support better face recognition performance than raw pixel values, and 3) the local features ensure that no erroneous correlations are learned between disparate parts of the face. In Section 7, we compare these results with those from other algorithms.

6.3 Experiment 3: Face Verification

In face verification tasks, we are given a probe face $\mathbf{x}_p$, and we have to decide if it belongs to a particular gallery face $\mathbf{x}_1$. Our goal is to determine the posterior probability that the probe face matches the gallery face.

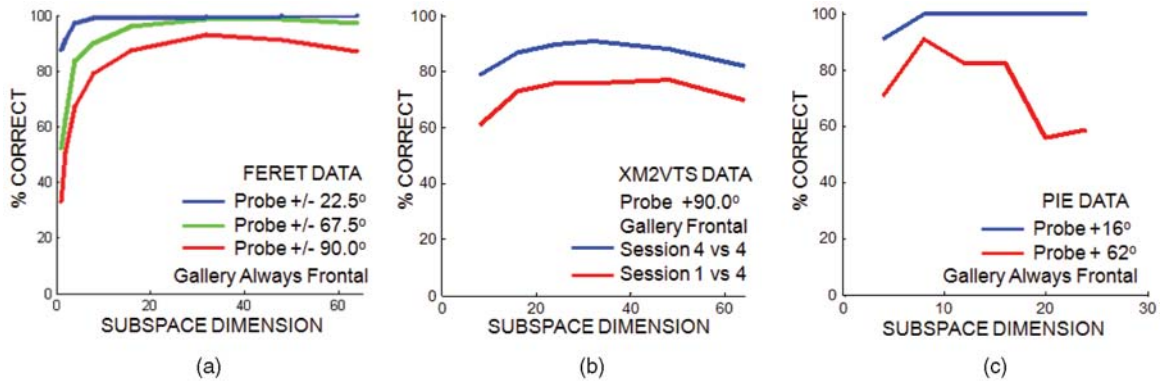

Fig. 12. (a) Percentage of first-match correct performance for FERET data with the tied factor analysis model, combining 14 local Gabor models as a function of the dimension of the latent identity variables. There were 100 frontal gallery faces and a single nonfrontal face, with an absolute pose difference that is different for each curve. (b) Performance for the XM2VTS database with a frontal gallery image and a profile probe image from the same (4/4) and different (1/4) sessions. (c) Performance for the PIE database with a frontal gallery image (pose C27) and a nonfrontal probe at 16° (pose C05) or 62° (pose C22).

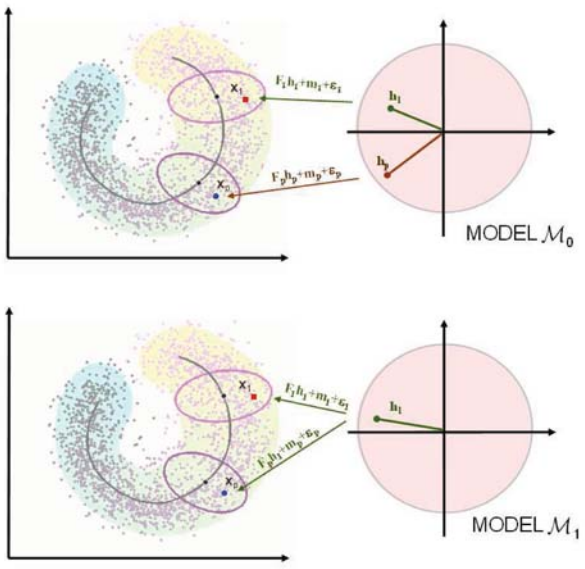

Fig. 13. Face verification. Given a probe face $\mathbf{x}_p$ and a single gallery face $\mathbf{x}_1$, our task is to decide whether they match or not. Once more we construct two models, $\mathcal{M}_0$ and $\mathcal{M}_1$. In model $\mathcal{M}_0$, the two faces are assumed to come from different individuals. In this case, two identity variables $\mathbf{h}_p$ and $\mathbf{h}_1$ explain the observed data. In model $\mathcal{M}_1$, the faces are assumed to be from the same individual, so a single identity variable explains both sets of observed data.

As in the face identification task, we associate each hypothesis with a model. Model $\mathcal{M}_0$ represents the case where the probe face does not match the gallery case. In this case, one latent identity variable $\mathbf{h}_p$ is associated with the probe face $\mathbf{x}_p$, and a different latent identity variable $\mathbf{h}_1$ is associated with the observed gallery face $\mathbf{x}_1$. Model $\mathcal{M}_1$ represents the case where the probe face matches the gallery face. In this case, a single latent identity variable $\mathbf{h}_1$ explains both observed data vectors $\mathbf{x}_p$ and $\mathbf{x}_1$. This scheme is illustrated in Fig. 13.

The evidence for models $\mathcal{M}_0$ and $\mathcal{M}_1$ are

$$Pr(\mathbf{x}_1, \mathbf{x}_p|\mathcal{M}_0) = \int Pr(\mathbf{x}_1|\mathbf{h}_1)Pr(\mathbf{h}_1)dh_1, \int Pr(\mathbf{x}_p|\mathbf{h}_p)Pr(\mathbf{h}_p)dh_p, \tag{14}$$

$$Pr(\mathbf{x}_1, \mathbf{x}_p|\mathcal{M}_1) = \int Pr(\mathbf{x}_1, \mathbf{x}_p|\mathbf{h}_1)Pr(\mathbf{h}_1)dh_1, \tag{15}$$

where these terms are calculated using (12). Once more, we find a posterior distribution over the hypotheses by using the Bayes rule. The verification decision is determined by the model with the maximum a posteriori probability.

Note that each model is explained by a different number of latent identity variables. One might naively think that the model with more parameters will always explain the data better. However, since these terms are integrated out, this is not the case, and it is valid to compare the models. This is a Bayesian model comparison procedure [19].

In this experiment, we investigate verification performance for the FERET database. We use the same test set as in the previous experiments. On each trial, the algorithm considers one of the frontal gallery images. Each of the nonfrontal images is presented in turn. Hence, there are 99 impostors for every one true match. The priors for models $\mathcal{M}_0$ and $\mathcal{M}_1$ are set to reflect this. On each trial, the posterior probability for these two estimates is compared to accept or reject the match. We vary the threshold for this posterior value between 0 and 1 to plot out a receiver operating characteristic (ROC) curve for each pose difference. The results are shown in Fig. 14 for a combination of local models, each of which used a subspace of 32 dimensions. We compare results to the previous studies in Section 7. For now, we conclude that the tied factor analysis model can be productively applied to face verification.

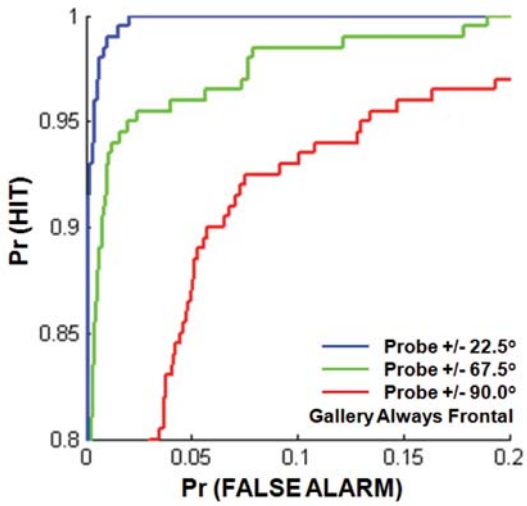

Fig. 14. Face verification using 14 local models, each with a 32-dimensional latent identity variable. ROC curve plotted for three different pose differences. Once more, left and right profile results are amalgamated.

6.4 Experiment 4: Approximation of Evidence Term

In the previous experiments, the recognition decision has been based on comparing evidence for different models. In order to calculate these evidence terms, we integrate over the uncertainty in the latent identity variables $\mathbf{h}$ (for example, see (11)). Here, it is possible to calculate this in closed form, but in cases where this integral is intractable, it is possible to approximate the uncertainty in $\mathbf{h}$ by a delta function at the maximum a posteriori value $\hat{\mathbf{h}}$. In this case, the solution for model $\mathcal{M}_n$ in face identification becomes

$$Pr(\mathbf{x}_{1\ldots N}, \mathbf{x}_p|\mathcal{M}_n) \approx Pr(\mathbf{x}_1|\hat{\mathbf{h}}_1)Pr(\hat{\mathbf{h}}_1)\ldots Pr(\mathbf{x}_n, \mathbf{x}_p|\hat{\mathbf{h}}_n)Pr(\hat{\mathbf{h}}_n)\ldots Pr(\mathbf{x}_N|\hat{\mathbf{h}}_N)Pr(\hat{\mathbf{h}}_N). \tag{16}$$

The maximum a posteriori value $\hat{\mathbf{h}}$ can be calculated using (5). The two evidence terms in (14) for face verification can be approximated in a similar way.

In Fig. 15, we reproduce results for the FERET database in Fig. 12. We now plot performance as a function of pose with a fixed subspace dimension of 32 (labeled "full posterior"). We compare these to the equivalent results from the approximated model (labeled "delta function"). The figure shows that identification performance is worse by using the approximated model: our main algorithm successfully exploits the estimated uncertainty in identity.

6.5 Experiment 5: Automated versus Manual Keypoint Detection

All of the previous experiments have used manually placed keypoints. While keypoint localization for frontal faces is quite reliable, the same is not necessarily true for profile faces.

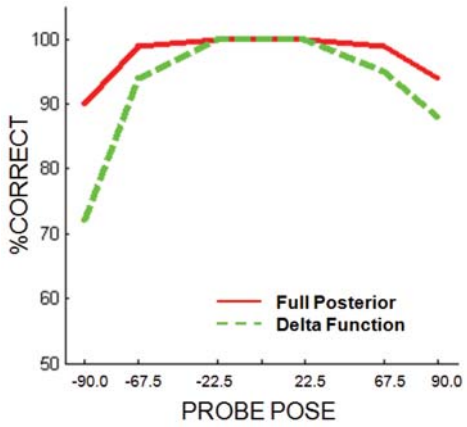

Fig. 15. Plot of the percentage of first-match correct performance as a function of probe pose (the gallery pose is always frontal) for both full and approximate (delta function) models. See Section 6.4 for more details.

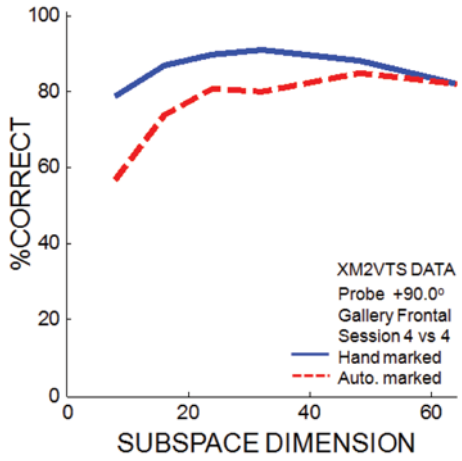

Fig. 16. Plot of the percentage of first-match correct performance as a function of subspace size for the XM2VTS database for two methods of keypoint registration. First, we plot performance where all features were manually labeled (peak performance of 91 percent). Second, we plot performance where the probe (profile) keypoints were located automatically (peak performance of 85 percent).

Consequently, in this experiment, we retain the manual placement for the gallery (frontal) images: it is reasonable to assume that they could be manually labeled for many real applications. However, we use automatic localization for probe (profile) faces. We found the keypoints for the last 100 faces from session 4 of the XM2VTS database with the following procedure. For each feature, we trained a scanning window Adaboost detector similar to that of Viola and Jones [35] but using Gabor responses rather than integral image features. These were trained using data from the first 195 individuals from the XM2VTS database and the entire FERET database. We multiplied the response of these detectors by -1 and treated the result as a cost for the feature position. This was weighted suitably and combined additively with the negative log likelihood of a factor analyzer model of the 28 feature position measurements (two for each of the 14 features). Hence, the total cost function favors sets of keypoints that agree with the local data but have a globally sensible configuration. We optimized this cost function by using coordinate ascent.

We compare manual and automatic labelings in Fig. 16. There is some decline in performance with automated labeling, but the results are still good. There is a decrease of 6 percent between the peaks of the graphs from 91 percent to 85 percent. We do not claim to have developed a particularly sophisticated keypoint detector, and this gap will probably be closed with more development. We conclude that performance is not critically contingent on manual feature labeling.

TABLE 1
Comparison of Face Identification Studies across Poses

STUDY	DATABASE	POSE DIFF (°)	% CORRECT
Gross et al. [12]	FERET (100)	30[1]	75
Gross et al. [12]	PIE (34)	6 / 62	39 / 93
Blanz et al. [5]	FRVT (87)	45	86
Zhang and Samaras [40]	CMU PIE (68)[2]	45 / 90	92 / 55
Chai et al. [7]	CMU PIE (68)	16/ 45	99.85 / 89.7
Wallhoff [36]	Mugshot(100)	90	60
Maurer [20]	US ARL (90)	45	53
Kim and Kittler [15]	XM2VTS (125)	30	53
Kanade and Yanade [14]	CMU PIE (34)	45 / 67.5 / 90	100 / 80 / 40
Our method	FERET (100)	22.5 / 67.5 / 90	100 / 99 / 92
Our method	XM2VTS (100)	90	91
Our method	PIE (100)	16 / 62	100 / 91

Note that the difficulty of the task depends on the number of individuals in the gallery. This is given in brackets after the database name. In each case, the best result is given, where there were several modifications to the basic method. Our method produces results that compare favorably to all contemporary approaches.

[1] *Gross et al. have an average pose difference of* $30°$. *The worst case absolute pose difference was* $60°$.

[2] *These results are better than they appear, as there was also considerably variation in lighting in this experiment.*

7 Discussion

7.1 Empirical Comparison to Other Studies

When comparing identification results, there are three factors that must be carefully considered: 1) one must remember that the difficulty of the task is dependent on the number of individuals in the gallery (100 for our experiments). When there are more individuals, there are more people to confuse the probe with, and the task becomes harder. 2) Moreover, the particular database may influence the difficulty. For example, in the CMU PIE database [31], images at different poses were captured at exactly the same time, which means that expression is always matched. In the FERET database [27], the images are not taken at the same time but are taken at the same session. In other data sets such as the XM2VTS database [21], some images are captured across different sessions. Even within a single database, it has been shown that different subsets may produce differing results [26]. 3) The degree of manual intervention should also be noted: Our algorithm assumed that the pose was known and used between 14 and 21 hand-labeled keypoints, depending on the pose (see Experiment 5 for results without manual labeling).

With these considerations in mind, we present a summary of identification performance from other studies in Table 1. Notably, Gross et al. [12] report 75 percent first-match results over 100 test faces from a different subset of the FERET database, with a mean difference in the absolute pose of $30°$ and a worst case difference of $60°$ by using only three manually marked feature points. Our system gives 99 percent performance, with a pose difference of $67.5°$ for every pair but uses more manual annotation. In the same study, they also report 39 percent and 93 percent performance for the PIE database conditions C22 ($62°$) and C05 ($16°$), respectively, with a large number (> 39) of manually labeled keypoints. For the same conditions, we report 91 percent and 100 percent, respectively, with less annotation.

TABLE 2
Comparison of Face Verification Studies across Poses

STUDY	DATABASE	POSE DIFF (^{o})	Pr(HIT)	Pr(FA)
Lucey and Chen [18] (EER)	FERET	60 / 90	0.9 / 0.84	0.1 / 0.16
Sanderson [30] (EER)	FERET	60	0.86	0.14
Our Method (EER)	FERET	22.5 / 67.5 / 90.0	0.99 / 0.96 / 0.925	0.01 / 0.04 / 0.075
Blanz et al. [5]	FRVT	45	0.79	0.01
Human Performance	XM2VTS	90.0	0.86	0.03
Our Method	FERET	22.5 / 67.5 / 90	0.99 / 0.93 / 0.80	0.01 / 0.01 / 0.03

In each case, the best result is given, where there were several modifications to the basic method. Our method produces results that compare favorably with these contemporary approaches.

Blanz et al. [5] report results for a test database of 87 subjects with a horizontal pose variation of $\pm 45^\circ$ from the FRVT 2002 database, using, on the average, 11 manually established feature points. They investigate both full coefficient-based 3D recognition (84.5 percent) performance and estimating the 3D model and creating a frontal image to compare to the test database (86.25 percent correct). Our system produces better performance at larger pose differences for comparable databases (indoors, with frontal and profile images taken in same session). Probably, the best previous results are those of Chai et al. [7], who used the PIE database and got an average performance of 98.5 percent with small pose differences of 16 degrees (pose C05) and an 89.7 percent performance with pose differences of 45° (pose C11) with only two manually registered points.

In Table 2, we reproduce verification results from several studies that have attempted to cope with pose differences. The best previous results are those of Blanz et al. [5], who report a 79.3 percent hit rate with a 1 percent false-alarm rate and 45° of pose differences. Our method yields a 93.5 percent hit rate at the same false-alarm rate, with a larger pose difference of $\pm 67.5^\circ$.

We have not found reliable data suggesting how well humans can perform the task of face recognition across poses. Wallhoff et al. [36] state that "some preliminary tests in their laboratory resulted in a recognition rate of 70 to 80 percent for several test persons," performing a frontal to profile identification task with a gallery containing 100 people and using the Mugshot database. Our model produces results of 92 percent for this task on the FERET database. In Table 2, we report human performance from an informal verification experiment using the XM2VTS database. Three subjects viewed 1,000 pairs of faces (one frontal and one profile) for 200 ms each and were asked to judge if they were the same person or different persons. The mean hit rate was 86 percent, with 0.03 percent false alarms. This is superior to the performance of our algorithm, which achieves only 80 percent performance with the same false alarm rate for the FERET data set. Moreover, we consider this to be a lower bound on human performance, as the subjects reported that many of their errors were due to their lack of attention or wrong keypresses. We tentatively conclude that our system cannot yet compete with human performance.

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 a highly peaked posterior. Nonetheless, the simple linear transformation has sensible properties: if we consider faces that look similar at one pose, they probably also look similar to each other at another pose. These linear transformations maintain this relationship in the observed feature space. In Section 6.4, we have demonstrated that our method exploits knowledge about uncertainty in the identity of the individual: performance decreased when we used a point estimate of identity. We conclude that our Bayesian approach, in which we are not required to fix a single estimate of identity, has some empirical advantages.

It is notable that our best performance comes from combining results from a set of local models rather than attempting to model the entire image with a single model. There are two things to note. First, our naive Bayes formulation is probably suboptimal, since the features overlap and are not independent. Second, it should be mentioned that this study is, by no means, the first to use local features. Notably, the elastic graph matching approach in [20], [37] used a series of local features that are compared across poses. However, many methods use global features (e.g., [12]) or relate local features at one pose to global features at another [18].

7.2 Relation to Previous Work

Our algorithm has a strong Bayesian flavor and aims at providing a posterior probability over possible models of the data. Several other probabilistic models for face recognition have been presented. First, Moghaddam [22] suggested taking the difference between probe and gallery images and estimating the likelihood that this came from a within or between-individual difference distribution that was learned in the training stage. This method only produces a posterior probability for face verification. However, it has been employed for both verification and identification tasks and produces good results for frontal images. Lucey and Chen [18] implemented the method in [22] for face recognition across large pose differences and found that the results were poor. This is probably because the pixelwise differences become increasingly meaningless as the pose difference increases.

Zhou and Chellappa [43] also presented a probabilistic method that addressed pose variation. The key differences with our model were 1) they did not integrate out uncertainty in identity, 2) they only address identification and do not provide a probabilistic method for verification tasks, and 3) their method is designed to take multiple images, which makes it hard to directly compare results with ours. They

produce an 82 percent identification performance, with a gallery of 34 people from the PIE database using multiple probes at one pose and multiple gallery images at a different pose to make these comparisons.

Tenenbaum and Freeman [32] investigated the use of bilinear models for separating style from content. The model presented in this paper is very similar to their asymmetric bilinear model, in which the pose is considered as style and the identity of the individual is considered as content. The main differences are that 1) we have an axis-oriented per-pixel noise term, 2) our learning algorithm is different, and 3) we use our model to make a different type of inference. Tenenbaum and Freeman are variously concerned with extrapolating styles to new content examples or with classifying new content based on a number of training examples seen at different styles. They do not address the fundamental decision required for face recognition: Were two or more unseen examples generated from the same content, regardless of their style?

7.3 Advantages of Identity Space Approach

Our system has several desirable properties. First, it is fast relative to that of [6], as it only involves linear algebra in relatively low dimensions and does not require an expensive nonlinear 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 threshold using a generalization of the verification procedure. In order to do this, we simply add an extra model to the face identification procedure that associates a separate identity variable with the probe rather than forcing it to share a variable with one of the gallery images. Fourth, there is only a single parameter: the dimension of the latent identity variables. In fact, even this could be estimated using a variational factor analysis formulation [4]. Fifth, the Bayesian approach provides a clear way of incorporating multiple gallery or probe images: if images are known to come from the same individual, they are forced to share an identity variable in every competing model for the data.

8 Conclusions

We have presented a novel generative model for describing image variation in face data across different poses. Our model was applied to both face identification and verification tasks. Our method produces results that are favorable to the previous state of the art in both tasks. It is also considerably simpler and faster to implement than many other algorithms. The system described here is a pure machine learning approach that knows very little about geometry, lighting, or the structure of faces. Although we have achieved good results, it would be more sensible to incorporate this information, and in our future work, we will investigate more complex generative models that exploit information about the real-world generative process.

Appendix

We expand on the derivation of the posterior moments given in (7). Recall that in the E-Step, we aim at finding the posterior distribution over the identity variable $\mathbf{h}_i$. This is given by

$$\begin{aligned} &Pr(\mathbf{h_i}|\mathbf{x}_{i\bullet\bullet}, \boldsymbol{\theta}_{t-1}) \\ &\quad = \frac{\prod_{j=1}^{J}\prod_{k=1}^{K} Pr(\mathbf{x}_{ijk}|\mathbf{h}_i, \boldsymbol{\theta}_{t-1}) Pr(\mathbf{h}_i)}{\int Pr(\mathbf{x}_{i\bullet\bullet}) d\mathbf{h}} \\ &\quad = \kappa_1 \prod_{j=1}^{J}\prod_{k=1}^{K} \mathcal{G}_{\mathbf{x}_{ijk}}[\mathbf{F}_k\mathbf{h}_i + \mathbf{m}_k, \Sigma_k].\mathcal{G}_{\mathbf{h}_i}[\mathbf{0}, \mathbf{I}], \end{aligned} \tag{17}$$

where κ_1 represents the constant in the denominator. In order to calculate this posterior in closed form, we re-express the first term in the numerator as a Gaussian in $\mathbf{h}$ by using the relationship

$$\begin{aligned} &G_{\mathbf{x}}[\mathbf{F}\mathbf{h} + \mathbf{m}, \Sigma] \\ &\quad \propto G_{\mathbf{h}}\Big[(\mathbf{F}^T\Sigma^{-1}\mathbf{F})^{-1}\mathbf{h}^T\Sigma^{-1}(\mathbf{x} - \mathbf{m}), (\mathbf{F}^T\Sigma^{-1}\mathbf{F})^{-1}\Big], \end{aligned} \tag{18}$$

which gives

$$\begin{aligned} &Pr(\mathbf{h_i}|\mathbf{x}_{i\bullet\bullet}, \boldsymbol{\theta}_{t-1}) = \kappa_2 \prod_{j=1}^{J}\prod_{k=1}^{K} \\ &\mathcal{G}_{\mathbf{h}_i}\Big[(\mathbf{F}_k^T\Sigma_k^{-1}\mathbf{F}_k)^{-1}\mathbf{F}_k^T\Sigma_k^{-1}(\mathbf{x}_{ijk} - \mathbf{m}_k), (\mathbf{F}_k^T\Sigma_k^{-1}\mathbf{F}_k)^{-1}\Big]. \\ &\mathcal{G}_{\mathbf{h}_i}[\mathbf{0}, \mathbf{I}]. \end{aligned} \tag{19}$$

The numerator now consists of a product of Gaussian terms in the same variable, so the posterior probability will also be a Gaussian. We use the second Gaussian relation

$$\begin{aligned} &G_{\mathbf{x}}(\mathbf{a}, \mathbf{A})G_{\mathbf{x}}(\mathbf{b}, \mathbf{B}) \propto \\ &\quad G_{\mathbf{x}}\Big((\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}(\mathbf{A}^{-1}\mathbf{a} + \mathbf{B}^{-1}\mathbf{b}), (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1}\Big), \end{aligned} \tag{20}$$

so we can show that

$$\begin{aligned} &Pr(\mathbf{h_i}|\mathbf{x}_{i\bullet\bullet}, \boldsymbol{\theta}_{t-1}) \\ &\quad = \kappa_3 \mathcal{G}_{\mathbf{h}_i}\left[\mathbf{C}^{-1}\sum_{j=1}^{J}\sum_{k=1}^{K}\mathbf{F}_k^T\Sigma_k^{-1}(\mathbf{x}_{ijk} - \mathbf{m}_k), \mathbf{C}^{-1}\right], \end{aligned} \tag{21}$$

where

$$\mathbf{C}^{-1} = \left(\mathbf{I} + \sum_{j=1}^{J}\sum_{k=1}^{K}\mathbf{F}_k^T\Sigma_k^{-1}\mathbf{F}_k\right)^{-1}. \tag{22}$$

This distribution has the moments given in (7). Notice that the final constant κ_3 takes the value 1, as the posterior distribution has to sum to one.

Acknowledgments

The authors would like to thank Francisco Estrada, Jania Aghajanian, and Alastair Moore for reading early drafts of this work. The term "tied factor analysis" was suggested by Geoff Hinton. This work was supported by the EPSRC and by OCE-Etech, GEOIDE, and PRECARN.

REFERENCES

[1] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class-Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, pp. 711-720, 1997.

[2] D. Beymer, "Face Recognition under Varying Pose," Technical Report AIM-1461, Massachussetts Inst. of Technology AI Laboratory, 1993.

[3] D. Beymer and T. Poggio, "Face Recognition from One Example View," Technical Report AIM-1536, Massachussetts Inst. of Technology AI Laboratory, Sept. 1995.

[4] C. Bishop, *Pattern Recognition and Machine Learning.* Springer, 2007.

[5] V. Blanz, P. Grother, P.J. Phillips, and T. Vetter, "Face Recognition Based on Frontal Views Generated from Non-Frontal Images," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* pp. 454-461, 2005.

[6] V. Blanz, S. Romdhani, and T. Vetter, "Face Identification across Different Poses and Illumination with a 3D Morphable Model," *Proc. Fifth IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 202-207, 2002.

[7] X. Chai, S. Shan, X. Chen, and W. Gao, "Locally Linear Regression for Pose-Invariant Face Recognition," *IEEE Trans. Image Processing,* vol. 16, pp. 1716-1725, 2007.

[8] A. Dempster, N. Laird, and D. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *J. Royal Statistical Soc. B,* vol. 39, pp. 1-38, 1977.

[9] J. Elder, S. Prince, Y. Hou, M. Sizinstev, and E. Olevskiy, "Pre-Attentive and Attentive Detection of Humans in Wide-Field Scenes," *Int'l J. Computer Vision,* vol. 72, no. 1, pp. 47-66, 2007.

[10] K. Fukui and O. Yamaguchi, "Face Recognition Using Multi-Viewpoint Patterns for Robot Vision," *Proc. 11th Int'l Symp. Robotics Research,* pp. 192-201, 2003.

[11] A. Georghiades, P. Belhumeur, and D. Kriegman, "From Few to Many: Illumination Cone Models and Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, pp. 129-139, 2001.

[12] R. Gross, I. Matthews, and S. Baker, "Appearance-Based Face Recognition and Light Fields," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, pp. 449-465, 2004.

[13] X. He, S. Yan, Y. Hu, P. Nihogi, and H. Zhang, "Face Recognition Using Laplacianfaces," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, pp. 328-340, 2005.

[14] T. Kanade and A. Yamada, "Multi-Subregion-Based Probabilistic Approach toward Pose-Invariant Face Recognition," *Proc. IEEE Int'l Symp. Computational Intelligence in Robotics and Automation,* pp. 954-959, 2003.

[15] T. Kim and J. Kittler, "Locally Linear Discriminant Analysis for Multimodally Distributed Classes for Face Recognition with a Single Model Image," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, pp. 318-327, 2005.

[16] Y. Li, S. Gong, and H. Liddell, "Constructing Facial Identity Surfaces in a Nonlinear Discriminating Space," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 258-265, 2001.

[17] X. Liu and T. Chen, "Pose-Robust Face Recognition Using Geometry Assisted Probabilistic Modelling," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 502-509, 2005.

[18] S. Lucey and T. Chen, "Learning Patch Dependencies for Improved Pose Mismatched Face Verification," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 17-22, 2006.

[19] D. MacKay, *Information Theory, Learning and Algorithms.* Cambridge Univ. Press, 2003.

[20] T. Maurer and C. von der Malsburg, "Single-View Based Recognition of Faces Rotated in Depth," *Proc. Int'l Workshop Automatic Face and Gesture Recognition,* pp. 80-85, 1995.

[21] K. Messer, J. Matas, J. Kittler, and J. Luettin, "XM2VTSDB: The Extended M2VTS Database," *Proc. Second Int'l Conf. Audio and Video-Based Biometric Person Authentication,* pp. 72-77, 1999.

[22] B. Moghaddam, "Principal Manifolds and Probabilistic Subspaces for Visual Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, pp. 780-788, 2002.

[23] A. Pentland, B. Moghaddam, and T. Starner, "View-Based and Modular Eigenspaces for Face Recognition," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* pp. 84-91, 1994.

[24] V. Perlibakas, "Distance Measures for PCA-Based Face Recognition," *Pattern Recognition Letters,* vol. 25, pp. 711-724, 2004.

[25] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and J. Bone, *FRVT Evaluation Report,* http://www.frvt.org/FRVT2002/documents.htm, 2003.

[26] P. Phillips, H. Moon, S.A. Rizvi, and P. Rauss, "The FERET Evaluation Methodology for Face Recognition Algorithms," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, pp. 1090-1104, 2000.

[27] P. Phillips, H. Wechsler, J. Huang, and P.J. Rauss, "The FERET Database and Evaluation Procedure for Face Recognition Algorithms," *Image and Vision Computing,* vol. 16, pp. 295-306, 1998.

[28] S. Prince and J. Elder, "Invariance to Nuisance Parameters in Face Recognition," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* pp. 446-453, 2005.

[29] S. Romdhani, V. Blanz, and T. Vetter, "Face Identification by Fitting a 3D Morphable Model Using Linear Shape and Texture Error Functions," *Proc. Seventh European Conf. Computer Vision,* 2002.

[30] C. Sanderson, S. Bengio, and Y. Gao, "Transforming Statistical Models for Non-Frontal Face Verification," *Pattern Recognition,* vol. 39, pp. 288-302, 2006.

[31] T. Sim, S. Baker, and M. Bsat, "The CMU Pose, Illumination and Expression Database of Human Faces," CMU Technical Report CMU-RI-TR-01-02, 2001.

[32] J. Tenenbaum and W. Freeman, "Separating Style and Content with Bilinear Models," *Neural Computation,* vol. 12, pp. 1247-1283, 2000.

[33] M. Turk and A. Pentland, "Face Recognition Using Eigenfaces," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* pp. 586-591, 1991.

[34] M. Vasilescu and D. Terzopoulos, "Multilinear Image Analysis for Facial Recognition," *Proc. 16th Int'l Conf. Pattern Recognition,* pp. 205-211, 2002.

[35] P. Viola and M. Jones, "Robust Real-Time Face Detection," *Int'l J. Computer Vision,* vol. 57, pp. 1473-1505, 2004.

[36] F. Wallhoff, S. Muller, and G. Rigoll, "Hybrid Face Recognition Systems for Profile Views Using the Mugshot Database," *Proc. Second IEEE ICCV Workshop Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems,* pp. 149-156, 2001.

[37] L. Wiskott, J. Fellous, N. Kruger, and C. der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, pp. 775-779, 1997.

[38] W. Yambor, B. Draper, and R. Beveridge, "Analyzing PCA-Based Face Recognition Algorithms: Eigenvector Selection and Distance Measures," *Proc. Second Workshop Empirical Evaluation Methods in Computer Vision,* 2000.

[39] M. Yang, "Kernel Eigenfaces versus Kernel Fisherfaces: Face Recognition Using Kernel Methods," *Proc. Fifth IEEE Int'l Conf. Face and Gesture Recognition,* 2002.

[40] L. Zhang and D. Samaras, "Pose Invariant Face Recognition under Arbitrary Unknown Lighting Using Spherical Harmonics," *Proc. ECCV Int'l Workshop Biometric Authentication Workshop,* 2004.

[41] W. Zhao and R. Chellappa, "SFS-Based View Synthesis for Robust Face Recognition," *Proc. Fifth IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 285-292, 2002.

[42] W. Zhao, R. Chellappa, A. Rosenfeld, and J. Phillips, "Face Recognition: A Literature Survey," *ACM Computing Surveys,* vol. 12, pp. 399-458, 2003.

[43] S. Zhou and R. Chellappa, "Probabilistic Identity Characterization for Face Recognition," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 805-812, 2002.

Simon J.D. Prince received the PhD degree from the University of Oxford in 1999. His PhD work was focused on the study of human stereo vision. He was a postdoctoral research scientist in Oxford, Singapore, and Toronto. He is currently a senior lecturer in the Department of Computer Science, University College London. He has a diverse background in biological and computing sciences and has published papers across the fields of biometrics, psychology, physiology, medical imaging, computer vision, computer graphics, and human computer interaction. He is a member of the IEEE and the ACM Computing Society and is the technical meetings organizer of the British Machine Vision Association.

James H. Elder received the BASc degree in electrical engineering from the University of British Columbia in 1987 and the PhD degree in electrical engineering from McGill University in 1995. From 1995 to 1996, he was with the NEC Research Institute, Princeton, New Jersey. He joined the faculty of York University in 1996, where he is currently an associate professor. He is an associate editor for the *ACM Transactions on Applied Perception* and is a cochair of the Fifth IEEE Workshop on Perceptual Organization in Computer Vision (POCV 2006). His research interests include computer and human vision. His recent work has focused on natural scene statistics, perceptual organization, contour processing, attentive vision systems, and face detection and recognition. He received the Young Investigator Award from the Canadian Image Processing and Pattern Recognition Society in 2001. He is a member of the IEEE.

Jonathan Warrell received the BA degree in music from the University of Cambridge, the MSc degree in computer science from the University College London, and the PhD degree in music theory and analysis from King's College London. He is a research fellow in the Department of Computer Science, University College London. His research interests include object recognition, generative modeling, and machine learning. He is a member of the IEEE.

Fatima M. Felisberti received the PhD degree for her work on extraocular photoreception in insects from the University of Sao Paulo, Sao Paulo, Brazil, in 1992 and the PhD degree for her work on the effect of general anesthetics on neocortical neurones from the Max Planck Institute for Biological Cybernetics in 1996. Subsequently, she was with Nottingham University, exploring the effect of long-range interactions in the visual thalamus of mammals and primates, and then with City University, working on the role of visual distracters in target detection in humans. She was also with the Royal Holloway University of London, conducting research on motion discrimination with transparent displays. She is currently a senior lecturer in the Psychology Research Unit, Kingston University. Her current research program addresses cognitive aspects of face and emotion recognition and evolutionary psychology. She is a member of the AVA, IBRO, and APS.