

Joint and Implicit Registration for Face Recognition

Peng Li

University College London
Department of Computer Science
London WC1E 6BT, UK

p.li@cs.ucl.ac.uk

Simon J. D. Prince

University College London
Department of Computer Science
London WC1E 6BT, UK

s.prince@cs.ucl.ac.uk

Abstract

Contemporary face recognition algorithms rely on precise localization of keypoints (corner of eye, nose etc.). Unfortunately, finding keypoints reliably and accurately remains a hard problem. In this paper we pose two questions. First, is it possible to exploit the gallery image in order to find keypoints in the probe image? For instance, consider finding the left eye in the probe image. Rather than using a generic eye model, we use a model that is informed by the appearance of the eye in the gallery image. To this end we develop a probabilistic model which combines recognition and keypoint localization. Second, is it necessary to localize keypoints? Alternatively we can consider keypoint position as a hidden variable which we marginalize over in a Bayesian manner. We demonstrate that both of these innovations improve performance relative to conventional methods in both frontal and cross-pose face recognition.

1. Introduction

Automated face recognition systems find application in access control, image search, security and other areas. However, widespread deployment has not yet been achieved as current systems are not sufficiently reliable.

Part of the problem is that face recognition systems consist of a pipeline of sequential operations [25]: in the *face detection* stage the face is approximately localized in the image. The face pixels may then be *segmented* from the background pixels. The system then *localizes keypoints* such as the eyes, nose etc. as shown in Figure 1, with a view to registering the face more carefully. There follows a *measurement* stage in which data are extracted from the registered image. Finally *inferences* are made about identity. A problem at any part of the pipeline causes overall performance to degrade.

Most research on face recognition has concerned the last (inference) stage [20, 1, 25, 23, 7, 16, 21, 24]. Usually,

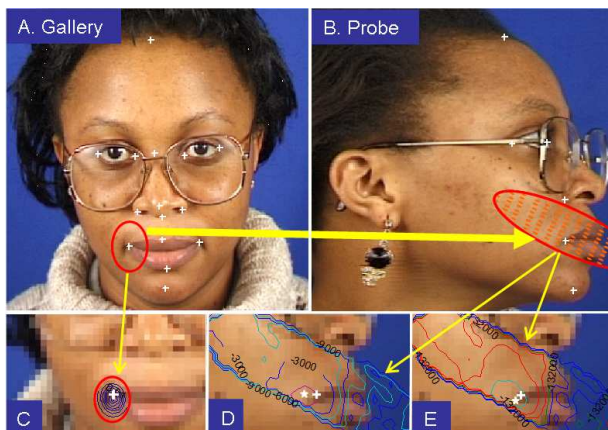


Figure 1. Localizing the left corner of mouth of a profile probe image (B) with the help of a frontal gallery image (A). Ground truth keypoints are indicated by white crosses. For each feature we have a prior distribution, shown for the frontal image in (C). The predicted distribution of the keypoint in the probe image is estimated by considering the probe image alone (D) or in combination with the gallery image (E). The MAP estimation of the keypoint position is indicated by white stars.

the extracted image measurements are mapped to a lower dimensional feature space, in which the distance between points is used to make decisions about similarity. Proposed mappings have included linear approaches such as Principal Component Analysis (PCA) [20], Linear Discriminant Analysis (LDA) [1] and Laplacianfaces [7] as well as non-linear approaches such as Kernel Linear Discriminant Analysis (KLDA) [23]. Other authors have proposed algorithms that embed similar recognition decisions in a probabilistic framework [14, 16, 9]. Many of these methods perform well for frontal faces under controlled conditions, and attempts are ongoing to extend them to cases where illumination, expression and pose may vary [11, 2, 17, 22].

Unfortunately, these achievements are diminished if the preceding pipeline is not reliable. A particular bottleneck is

keypoint localization. In *global* approaches, the keypoints are used to register the image to a common template [20, 1]. In *local* approaches, data are explicitly extracted from the region around the keypoints [3, 12, 8]. In either case recognition performance degrades if the keypoints are not accurately localized [18], especially in cross-pose recognition [13, 21, 17]. Unfortunately, automated facial keypoint detection remains a hard problem despite much investigation [3, 12, 8, 5].

In order to scientifically isolate the recognition stage, it is common to use manually labelled features [21, 24, 19, 10], while others simply do not detail how keypoints were localized. Even face recognition databases commonly provide manually labelled keypoint positions [15].

In this paper, we investigate whether it is even necessary to estimate keypoint positions. We are inspired by recent methods for inference in face recognition which do not estimate identity, but marginalize over all possible identities in a Bayesian manner [16, 9]. In this paper, we marginalize over both the identity and keypoint positions. One implication of combining keypoint localization and inference is that we can use one image to help with localization in the other: we no longer look for a generic eye, but the specific eye that matches the other image.

In Section 2 we introduce the probabilistic recognition framework. In Section 3 we describe our method for using the gallery image to help localize the keypoints in the probe image. In Section 4 we show that marginalizing over keypoint position is superior to finding the maximum a posteriori (MAP) position. In Section 5 we extend our model to cope with faces which differ in pose.

2. Probabilistic Face Recognition Framework

In this paper, we adapt the ‘‘Probabilistic Linear Discriminant Analyzer’’ (PLDA) model [16], which is a probabilistic version of [1]. In our initial description, we assume that we have already localized keypoints and extracted a feature vector from each. We model the vector \mathbf{x} associated with each keypoint separately. Let \mathbf{x}_{ij} be the feature vector extracted from the j 'th example of the i 'th individual. We describe this as a sum of signal and noise components:

$$\mathbf{x}_{ij} = \mu + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij} + \epsilon_{ij} \quad (1)$$

The first component, $\mu + \mathbf{F}\mathbf{h}_i$, represents the identity signal. It does not contain any elements that depend on the particular instance j of the given person's face. The term \mathbf{h}_i is termed a *latent identity variable* or LIV. It can be thought of as an idealized representation of human identity. The second component $\mathbf{G}\mathbf{w}_{ij} + \epsilon_{ij}$ represents within-individual noise and is different for every image. The term \mathbf{w}_{ij} is referred to as a *latent noise variable*.

The term μ represents the overall mean of the data. The matrix \mathbf{F} defines a basis for the identity (between individ-

ual) subspace. The columns of \mathbf{F} are similar to the eigenfaces of [20]. The term \mathbf{h}_i represents the position in this subspace (similar to the weighting of eigenfaces). Similarly, the matrix \mathbf{G} defines a basis for the within-individual variation, and \mathbf{w}_{ij} represents the position within this subspace. The term ϵ_{ij} is mean zero Gaussian noise, with diagonal covariance Σ . It accounts for any further image variation that is not well described by the previous components.

More formally, we can re-write the model in terms of conditional probabilities:

$$Pr(\mathbf{x}_{ij}|\mathbf{h}_i, \mathbf{w}_{ij}) = \mathcal{G}_{\mathbf{x}}[\mu + \mathbf{F}\mathbf{h}_i + \mathbf{G}\mathbf{w}_{ij}, \Sigma] \quad (2)$$

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

$$Pr(\mathbf{w}_{ij}) = \mathcal{G}_{\mathbf{w}}[\mathbf{0}, \mathbf{I}] \quad (4)$$

where $\mathcal{G}_a[\mathbf{b}, \mathbf{C}]$ denotes a Gaussian distribution in \mathbf{a} with mean \mathbf{b} and covariance \mathbf{C} . We have also specified priors over the latent variables $\mathbf{h}_i, \mathbf{w}_{ij}$ to complete the model. The unknown parameters $\theta = \{\mu, \mathbf{F}, \mathbf{G}, \Sigma\}$ can be learnt using the Expectation Maximization (EM) algorithm [4] as described in [16].

2.1. Inferences about Identity

Given face data \mathbf{x}_p and \mathbf{x}_g from the probe and gallery images respectively, we wish to know if they were generated from the same identity or whether it is better to explain the data with separate identities. We compare two generative models for the data and choose the one with higher likelihood. The model \mathcal{M}_d describes the case when the features come from *different* individuals. The model \mathcal{M}_s describes the case when the features come from the *same* individual. We treat each in turn.

When the probe and gallery image do not match (\mathcal{M}_d), we treat them as independent and model them separately. For the probe image we have:

$$\mathbf{x}_p = \mu + [\mathbf{F} \ \mathbf{G}] \begin{bmatrix} \mathbf{h} \\ \mathbf{w} \end{bmatrix} + \epsilon \quad (5)$$

or

$$\mathbf{x}_p = \mu + \mathbf{A}\mathbf{y} + \epsilon. \quad (6)$$

This has the form of a standard factor analyzer:

$$Pr(\mathbf{x}_p|\mathbf{y}, \mathcal{M}_d) = \mathcal{G}_{\mathbf{x}_p}[\mu + \mathbf{A}\mathbf{y}, \Sigma] \quad (7)$$

$$Pr(\mathbf{y}) = \mathcal{G}_{\mathbf{y}}[\mathbf{0}, \mathbf{I}] \quad (8)$$

The likelihood of observing the probe image assuming that there was no match can be calculated by marginalizing over the hidden variable \mathbf{y} . From Equation 6 it is easy to see that the first two moments of the distribution of the are given by:

$$\begin{aligned}
E[\mathbf{x}_p] &= \mu \\
E[(\mathbf{x}_p - \mu)(\mathbf{x}_p - \mu)^T] &= E[(\mathbf{A}\mathbf{y} + \epsilon)(\mathbf{A}\mathbf{y} + \epsilon)^T] \\
&= \mathbf{A}\mathbf{A}^T + \Sigma
\end{aligned}$$

In fact, it can be shown the after marginalizing over the hidden variables, the likelihood of the data has a Gaussian form so that:

$$Pr(\mathbf{x}_p|\mathcal{M}_d) = \mathcal{G}_{\mathbf{x}_p}[\mu, \mathbf{A}\mathbf{A}^T + \Sigma] \quad (9)$$

A similar equation can be developed for the gallery image. The likelihood when the images do not match is hence

$$Pr(\mathbf{x}_p, \mathbf{x}_g|\mathcal{M}_d) = Pr(\mathbf{x}_p|\mathcal{M}_d)Pr(\mathbf{x}_g|\mathcal{M}_d) \quad (10)$$

If the faces do match (\mathcal{M}_s) we use the generative equation:

$$\begin{bmatrix} \mathbf{x}_p \\ \mathbf{x}_g \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \end{bmatrix} + \begin{bmatrix} \mathbf{F} & \mathbf{G} & \mathbf{0} \\ \mathbf{F} & \mathbf{0} & \mathbf{G} \end{bmatrix} \begin{bmatrix} \mathbf{h} \\ \mathbf{w}_p \\ \mathbf{w}_g \end{bmatrix} + \begin{bmatrix} \epsilon_p \\ \epsilon_g \end{bmatrix} \quad (11)$$

or $\mathbf{x}' = \mu' + \mathbf{B}\mathbf{y}' + \epsilon'$. The likelihood of the data under this model is:

$$Pr(\mathbf{x}_p, \mathbf{x}_g|\mathcal{M}_s) = Pr(\mathbf{x}'|\mathcal{M}_s) = \mathcal{G}_{\mathbf{x}'}[\mu', \mathbf{B}\mathbf{B}' + \Sigma'] \quad (12)$$

where

$$\Sigma' = \begin{bmatrix} \Sigma & \mathbf{0} \\ \mathbf{0} & \Sigma \end{bmatrix} \quad (13)$$

3. Recognition and Keypoint Localization

Now we turn to the question of how to combine recognition and keypoint localization. Once more, we treat each keypoint independently. The (x,y) position of the keypoint in the probe image \mathcal{I}_p is denoted by \mathbf{t}_p . The (x,y) position of the same keypoint in the gallery image \mathcal{I}_g is denoted by \mathbf{t}_g . Feature extraction is denoted by the function ϕ which takes the image and keypoint positions so that:

$$\mathbf{x}_p = \phi(\mathcal{I}_p, \mathbf{t}_p) \quad (14)$$

$$\mathbf{x}_g = \phi(\mathcal{I}_g, \mathbf{t}_g) \quad (15)$$

The concatenated feature vectors \mathbf{x}' can hence be calculated as:

$$\mathbf{x}' = \begin{bmatrix} \mathbf{x}_p \\ \mathbf{x}_g \end{bmatrix} = \begin{bmatrix} \phi(\mathcal{I}_p, \mathbf{t}_p) \\ \phi(\mathcal{I}_g, \mathbf{t}_g) \end{bmatrix} \quad (16)$$

3.1. Model 1: Finding keypoints using one image

Keypoints are usually found by a separate process from the recognition model. However, since we have explicitly described our data in terms of a generative model, it is possible to use the same model to find the keypoint. We maximize the posterior probability of the extracted feature as a function of the keypoint position.

$$\begin{aligned}
\mathbf{t}_p^* &= \arg \max_{\mathbf{t}_p} Pr(\mathbf{x}_p|\mathcal{M}_d, \mathbf{t}_p)Pr(\mathbf{t}_p) \\
\mathbf{t}_g^* &= \arg \max_{\mathbf{t}_g} Pr(\mathbf{x}_g|\mathcal{M}_d, \mathbf{t}_g)Pr(\mathbf{t}_g) \quad (17)
\end{aligned}$$

where the relation between \mathbf{t}_p and \mathbf{x}_p is given by Equation 14 and the conditional relation between \mathbf{t}_g and \mathbf{x}_g is given by Equation 15. We use these optimal keypoint values \mathbf{t}_p^* and \mathbf{t}_g^* to calculate the features \mathbf{x}_p and \mathbf{x}_g and evaluate the likelihoods that the images match or do not match using Equations 10 and 12 respectively.

3.2. Model 2: Finding keypoints using both images

The previous method is similar to conventional approaches: the probe image is used to find the probe keypoint and the gallery image is used to find the gallery keypoint. However, since we also have a joint probability model of the probe and gallery images (Equation 12) a new possibility emerges: perhaps we can exploit the matching process to find both keypoints simultaneously. We now propose a different method to find keypoints depending on whether we are hypothesizing that the faces are the same (\mathcal{M}_s) or different (\mathcal{M}_d). When they are the same, we optimize over the joint likelihood:

$$\begin{aligned}
\mathcal{M}_s : \\
\mathbf{t}_p^*, \mathbf{t}_g^* &= \arg \max_{\mathbf{t}_p, \mathbf{t}_g} Pr(\mathbf{x}'|\mathcal{M}_s, \mathbf{t}_p, \mathbf{t}_g)Pr(\mathbf{t}_p, \mathbf{t}_g) \quad (18)
\end{aligned}$$

This has the natural interpretation of using not only our knowledge of how the keypoint looks in general (embodied in the model parameters), but also using the probe keypoint to find the gallery keypoint and vice-versa.

When we hypothesize that the faces do not match, we proceed as in Section 3.1 and find \mathbf{t}_p and \mathbf{t}_g separately:

$$\begin{aligned}
\mathcal{M}_d : \\
\mathbf{t}_p^* &= \arg \max_{\mathbf{t}_p} Pr(\mathbf{x}_p|\mathcal{M}_d, \mathbf{t}_p)Pr(\mathbf{t}_p) \\
\mathbf{t}_g^* &= \arg \max_{\mathbf{t}_g} Pr(\mathbf{x}_g|\mathcal{M}_d, \mathbf{t}_g)Pr(\mathbf{t}_g) \quad (19)
\end{aligned}$$

Our estimate of keypoint position differs based on our interpretation of whether the images match or not.

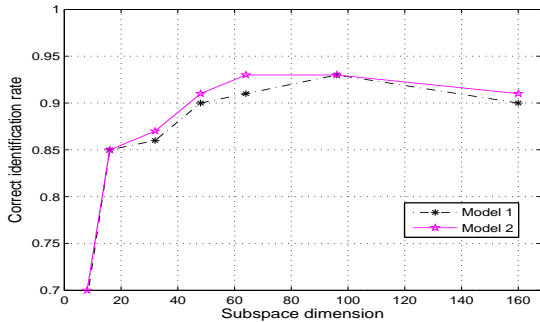


Figure 2. Frontal face identification results with keypoints estimated using a single image (Model 1) or both images (Model 2). Performance is better when both images are used.

3.3. Methods

We investigate face identification using the XM2VTS frontal face data set. We use 1200 images from the first 195 identities for training. The first image of the first session and that of the fourth session from the last 100 identities are used as gallery and probe respectively. We train the PLDA model with 6 iterations using the EM algorithm.

The original face images are RGB color images of size 400×400 . We ran a sliding window face detector over the images at several scales to identify the region of the image that is most likely to contain a face. We reshaped the resulting bounding box to 100×100 pixels using a similarity transform with bicubic interpolation. We did not use any further image warping. We applied histogram equalization to the resulting images.

Thirteen keypoints including the eye corners, nose and mouth were investigated (see Figure 1). These keypoints were manually labeled by two subjects to provide ground truth for the training and test data. In all the experiments in this paper, the keypoints in gallery images were manually labelled, but those of the probe images were unlabelled. We extract a feature vector that consists of the responses of Gabor filters at 8 orientations and 3 scales in a 6×6 grid around each keypoint. A separate PLDA model was built for each keypoint. The keypoint positions are assumed to be independent from one another.

The prior distribution of each keypoint is modeled as a two dimensional Gaussian. The mean and covariance are calculated using the coordinates of manually labeled keypoints of the training data. The locations of each keypoint are discretized with single pixel resolution over a region covering a Mahalanobis distance of ≤ 2.5 . This describes more than 99% of the probability density.

For model 1, we multiply the matching likelihood $Pr(\mathbf{x}_p | \mathcal{M}_d, t_p)$ by the prior probability over positions $Pr(t_p)$, then locate the keypoint by finding the position

with the maximum probability in a probe image as in Equation 17. For model 2, we multiply the matching likelihood $Pr(x_p, x_g | \mathcal{M}_s, t_p, t_g)$ by the prior probability over positions $Pr(t_p)$, then locate the keypoint by finding the position with the maximum probability in a probe image as in Equation 18. Note that $Pr(t_p, t_g)$ in Equation 18 becomes $Pr(t_p)$ as the gallery image t_g is fixed.

3.4. Results and Discussion

We compare Model 1 and Model 2 in two tasks. First, we consider face identification. Second, we investigate the ability to localize the keypoints. We investigate both metrics as a function of the subspace dimension (number of columns in \mathbf{F} and \mathbf{G}). These are always set to be equal although this need not necessarily be the case.

Percent first match identification rate is plotted as a function of subspace dimension in Figure 2. It is observed that finding the probe keypoints using both gallery and probe images (Model 2) produces greater or equal performance to that when we find the probe keypoints using the probe image alone (Model 1). In other words, matching performance is worse when we search for a generic keypoint, than if we use information from the hypothesized matching image to search for the expected specific keypoint.

We can also assess performance in terms of the mean localization error of the 13 keypoints. This is reported in Figure 3. The localization error is described using normalized Euclidean distance (fraction of inter-ocular distance) following [6]. This metric makes the results independent of image resolution. The results mirror the pattern for face identification: Localization is also best when we treat gallery and probe images together rather than separately. This supports our hypothesis that the gallery image can help the keypoint localization in the probe image. Furthermore, the localization errors of both algorithms are very close to that of the human labeling.

An interesting observation is that smaller localization error doesn't always mean a higher identification rate. For a given dimensionality this is generally true (see Figure 4). However, the localization is best with a small number of factors, whereas recognition is best with an intermediate number. We conjecture that low frequency components are primarily being used in the localization process and that higher frequency components which are useful for recognition, may actually impede localization.

4. Registration-Free Face Recognition

The second question we address is whether it is necessary to localize the keypoints in the probe images at all. Previously we aimed to *find* the keypoint positions. Now we treat the location of each keypoint as a hidden variable and *marginalize* over it in a Bayesian manner. Once again,

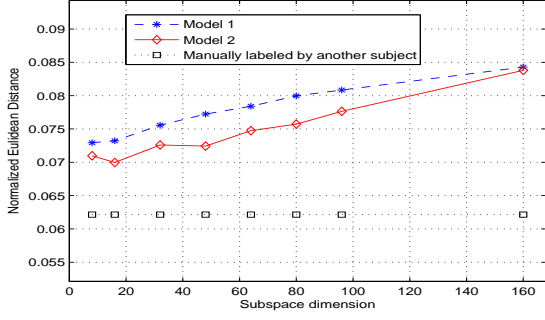


Figure 3. Normalized registration errors for keypoints as estimated using a single image (Model 1) or both images (Model 2). We also compare to manual labelling by a second subject.

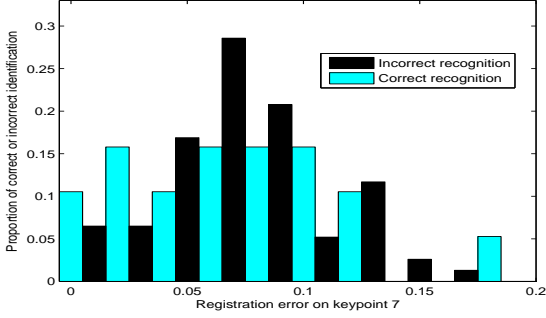


Figure 4. Localization error for keypoint 7 at the base of the septum for cases where recognition was correct and incorrect. Localization error is generally smaller (mean = 0.081 vs. 0.092) when the decision was correct, despite this being only one of many possible positions that contributed to the decision.

we have the choice of doing this in each image separately or treating both images together.

4.1. Model 3: Marginalizing over keypoint position using one image

When the faces differ we simply integrate over the feature position in the likelihood calculation:

$$Pr(\mathbf{x}_p|\mathcal{M}_d) = \int Pr(\mathbf{x}_p|\mathcal{M}_d, \mathbf{t}_p)Pr(\mathbf{t}_p)d\mathbf{t}_p \quad (20)$$

$$Pr(\mathbf{x}_g|\mathcal{M}_d) = \int Pr(\mathbf{x}_g|\mathcal{M}_d, \mathbf{t}_g)Pr(\mathbf{t}_g)d\mathbf{t}_g \quad (21)$$

When we evaluate the hypothesis that the faces are the same, we first find a probability distribution over the possible feature based on the images separately

$$\begin{aligned} Pr(\mathbf{t}_p|\mathcal{M}_d) &\propto Pr(\mathbf{x}_p|\mathcal{M}_d, \mathbf{t}_p)Pr(\mathbf{t}_p) \\ Pr(\mathbf{t}_g|\mathcal{M}_d) &\propto Pr(\mathbf{x}_g|\mathcal{M}_d, \mathbf{t}_g)Pr(\mathbf{t}_g) \end{aligned} \quad (22)$$

and then we integrate over both unknown feature positions based on the probability distributions estimated from the individual images.

$$Pr(\mathbf{x}_p, \mathbf{x}_g|\mathcal{M}_s) = \int \int Pr(\mathbf{x}_p, \mathbf{x}_g|\mathcal{M}_s, \mathbf{t}_p, \mathbf{t}_g)Pr(\mathbf{t}_p|\mathcal{M}_d)Pr(\mathbf{t}_g|\mathcal{M}_d)d\mathbf{t}_gd\mathbf{t}_p \quad (23)$$

4.2. Model 4: Marginalizing over keypoint positions using both images

Similar to estimating feature position in Section 3.2, it is possible to use both images to infer the probability distribution over the keypoint positions. In this scenario, we calculate the likelihood of the faces matching as:

$$\begin{aligned} \mathcal{M}_s : \\ Pr(\mathbf{x}_p, \mathbf{x}_g) = \int \int Pr(\mathbf{x}'|\mathcal{M}_s, \mathbf{t}_p, \mathbf{t}_g)Pr(\mathbf{t}_p, \mathbf{t}_g)d\mathbf{t}_pd\mathbf{t}_g \end{aligned} \quad (24)$$

The likelihood of the faces not matching is:

$$\begin{aligned} \mathcal{M}_d : \\ Pr(\mathbf{x}_p|\mathcal{M}_d) = \int Pr(\mathbf{x}_p|\mathcal{M}_d, \mathbf{t}_p)Pr(\mathbf{t}_p)d\mathbf{t}_p \\ Pr(\mathbf{x}_g|\mathcal{M}_d) = \int Pr(\mathbf{x}_g|\mathcal{M}_d, \mathbf{t}_g)Pr(\mathbf{t}_g)d\mathbf{t}_g \end{aligned} \quad (25)$$

The likelihood of the features not matching is calculated as in Section 4.1. Hence, in this scheme, the probability distribution over keypoint position differs depending on whether we are assessing the probability that the images match or not.

Note that the integral in the above equations is calculated over all the possible positions of a keypoint. In practice, these positions are discretized and the integral can be approximated using summation over these discretized positions. The prior probability of each position $Pr(\mathbf{t}_p)$ can be calculated using the training data, modeled as Gaussian.

4.3. Methods

We compare marginalization over possible keypoint positions treating each image separately (Model 3) to finding the maximum a posteriori keypoint position as in Model 1. We also investigate using both images and marginalizing over keypoint position (Model 4). The experimental setting is the same as in Section 3.3.

For model 1, we locate the keypoint by finding the position with the maximum probability as in Equation 17. For model 3, we multiply the matching likelihood by the prior probability of each discretized position in a probe image as

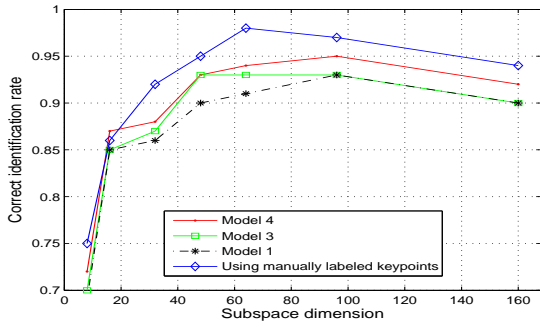


Figure 5. Identification performance for marginalizing over keypoint position (model 3) is superior to using the best estimate of keypoint position (model 1). If we marginalize over keypoint position and use information from both images in keypoint localization (model 4) performance improves further, and approaches the results with manual labelling.

shown in Equation 22, then sum them up as in Equation 20 which formulates the final likelihood of this keypoint. For model 4, we multiply the matching likelihood by the prior probability over positions $Pr(t_p)$, then sum them up as in Equation 24 which formulates the final likelihood of this keypoint.

4.4. Results

Once again we will first consider face recognition performance in an identification setting. Figure 5 shows that model 3 outperforms model 1. In other words, it is better to treat the probe keypoint position as a random variable and marginalize than it is to find maximum a posteriori keypoint position. The most likely explanation for this phenomenon is that there are some points where the likelihood surface over keypoint position is flat and possibly multimodal. When we are forced to choose a single estimate of position we occasionally make drastic mistakes. However, when we integrate over this distribution, the bulk of the probability may be close to the correct position, although the maximum is quite wrong.

The previous experiments have shown that (i) treating two images together is better than treating them separately and (ii) marginalizing over keypoint positions is better than finding the MAP keypoint. In Figure 5 we also investigate the combination of these approaches (model 4). These two manipulations have additive effects on identification performance: the best results are produced by marginalizing over position and using both images in estimation.

For completeness, we also plot the results for images with labels that were marked by hand by a second individual (Figure 5). We note that we have improved performance to a maximum of 95% but not to the level of manually marked keypoints which have maximum performance of 98%. We

note that the identification rate achieved by using manually labeled keypoints is somewhat lower than that presented in [16]. However, in their experiment, the gallery images are also included in the training whereas in our experiments training and test sets were completely disjoint. There are also various other small methodological differences including the image resolution, choice of keypoints and degree of image warping.

5. Cross-Pose Face Recognition

Finding features is more problematic in non-frontal images and hampers our ability to perform cross-pose recognition [17]. In this section, we investigate applying the same ideas to cross-pose recognition. We adapt the tied PLDA model of Prince et al. [17] to this task. The idea behind tied PLDA is to provide a generative explanation for the data that is parameterized by the viewing conditions. The j 'th feature vector from the i 'th individual in the k 'th pose is described as:

$$\mathbf{x}_{ijk} = \mu_k + \mathbf{F}_k \mathbf{h}_i + \mathbf{G} \mathbf{w}_{ijk} + \epsilon_{ijk} \quad (26)$$

In this paper, we compare profile ($k=1$) probe faces to frontal ($k=2$) gallery faces. The associated tied PLDA model hence consists of two sets of parameters $\theta_1 = \{\mu_1, \mathbf{F}_1, \mathbf{G}_1, \Sigma_1\}$ for generation of non-frontal faces and $\theta_2 = \{\mu_2, \mathbf{F}_2, \mathbf{G}_2, \Sigma_2\}$ for generation of profile faces. By analogy with section 2.1 we have two sets of equations for when the images do not match:

$$\mathbf{x}_p = \mu + [\mathbf{F}_1 \quad \mathbf{G}_1] \begin{bmatrix} \mathbf{h} \\ \mathbf{w} \end{bmatrix} + \epsilon \quad (27)$$

$$\mathbf{x}_g = \mu + [\mathbf{F}_2 \quad \mathbf{G}_2] \begin{bmatrix} \mathbf{h} \\ \mathbf{w} \end{bmatrix} + \epsilon \quad (28)$$

$$(29)$$

and when they do:

$$\begin{bmatrix} \mathbf{x}_p \\ \mathbf{x}_g \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \end{bmatrix} + \begin{bmatrix} \mathbf{F}_1 & \mathbf{G}_1 & \mathbf{0} \\ \mathbf{F}_2 & \mathbf{0} & \mathbf{G}_2 \end{bmatrix} \begin{bmatrix} \mathbf{h} \\ \mathbf{w}_p \\ \mathbf{w}_g \end{bmatrix} + \begin{bmatrix} \epsilon_p \\ \epsilon_g \end{bmatrix} \quad (30)$$

We calculate likelihoods for the two models \mathcal{M}_s and \mathcal{M}_d by integrating over the hidden models in exactly the same manner as in Section 2.1.

5.1. Methods

We investigate the same two hypotheses as for frontal face identification. The procedure was very similar to former experiments. We use 2392 images from the first 195 identities for training among which half of them are frontal

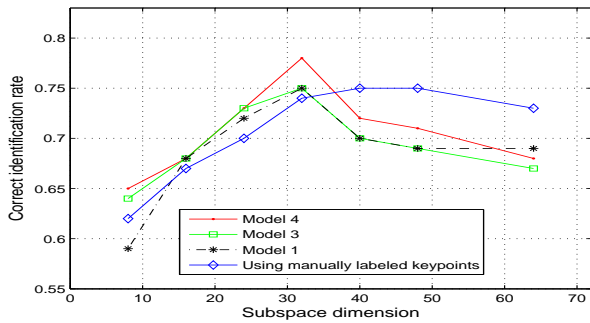


Figure 6. Identification results for cross-pose face recognition task. As for the frontal case, best performance is for the case where we exploit both images to find the probe keypoints and marginalize over the unknown position (Model 4). In this case performance exceeds that with human labelled keypoints.

faces and the other half are right profile faces (i.e. 90 degrees pose difference) as shown in Figure 1. In both training and testing, we assume that the pose of the face is known. The frontal images of the first session from the last 100 identities were used for the gallery. The profile images of the fourth session from the last 100 identities were used as the probe. We train the tied PLDA model with 6 iterations of the EM algorithm as detailed in [16]. Six keypoints from the eye corners, nose and mouth are investigated. These are illustrated in Figure 1 (B). Once again the keypoints were manually labelled by two subjects. The training keypoint positions were used to learn the model. For the test set, the manually labelled keypoints were used in the gallery images, but the probe image keypoint position was assumed to be unknown. The manually labelled keypoints for the probe images are only used to test the accuracy of registration.

5.2. Results

We report the results of cross-pose face identification in Figure 6. The results are similar to those for frontal faces. Treating gallery and probe images together improves performance as does marginalizing over keypoint positions. The best identification rate of the proposed method is 78% and that of using manually labeled keypoints is 76%. We don't claim that our proposed method outperforms using manually labeled keypoints, but the performance is comparable.

The registration errors are shown in Figure 7. Similarly to the frontal case, treating the gallery and probe images together improves performance relative to finding the keypoint using only a single image. With both methods predicted keypoint positions are closer to the manually labeled positions than the manual labels of the second subject. This is probably due to the fact that two of the keypoints (the chin and hairline) are quite ambiguous and hard to label.

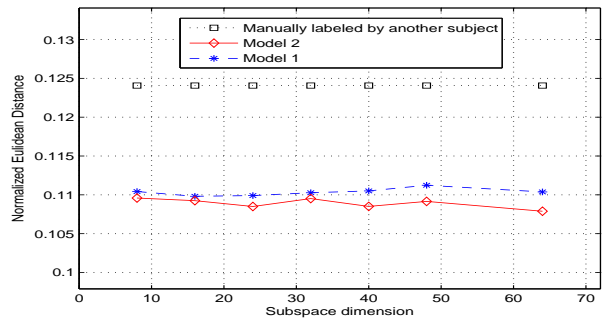


Figure 7. Normalized registration error of probe keypoints in cross-pose task decreases when we use information from the probe and gallery image (Model 2) rather than just from the probe (Model 1). Surprisingly, both methods estimate keypoint positions better than a second human subject.

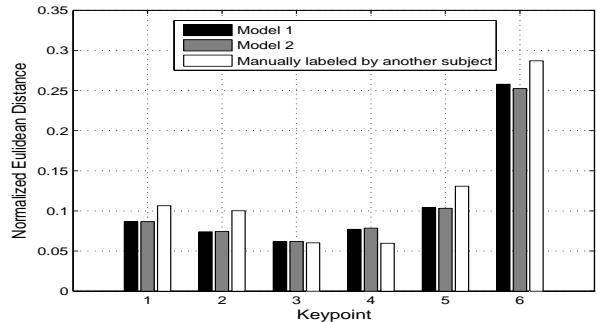


Figure 8. Normalized registration errors on 6 probe keypoints in cross-pose task. The subspace dimension is 32. The first and second largest errors are observed for keypoint 5 and 6 (the chin and hairline) because they are quite ambiguous and hard to label. Compared to these two keypoints, keypoint 3 and 4 (the corners of nose and mouth) are much easier for human labelling.

We plot the registration errors on these 6 keypoints for a subspace dimension in Figure 8. This seems to be in agree with our explanation. Since our algorithm has very good knowledge of the prior position it never make any drastic mistakes.

6. Conclusion

We have proposed to integrate face registration and recognition. Rather than using a generic keypoint model to localize them in a probe image, we use a model that is informed by the appearance of the keypoint in the gallery image. Furthermore, we consider keypoint position as a hidden variable which we marginalize over in a Bayesian manner. Experimental results on XM2VTS database demonstrate that both of these innovations improve performance in both frontal and cross-pose face recognition.

One weakness of this model is that we are forced to discretize keypoint positions in order to effectively marginalize over them as there is no closed form solution to this integral. The discretization means additional computational complexity and also means that we may have to store the original image rather than preprocessed low-dimensional projections.

A second weakness is that the model is overconfident in practice and this diminishes the effect of the priors. One possible solution would be to construct a similar model based on the multivariate Student t-distribution rather than Gaussian statistics. The long tails of the t-distribution effectively moderate the confidence and allow the prior to play a larger role.

In future work, we would also like to investigate joint models of keypoint localization: the position of one keypoint provides information about the position of the others. In practice such information could be incorporated efficiently by defining the relations between keypoints to have the structure of a tree and applying a dynamic programming technique in inference.

7. Acknowledgement

The authors would like to acknowledge the support of the EPSRC (Grant No EP/E065872/1). We would like to thank Jania Aghajanian, Jonathan Warrell, Alastair Moore, Umar Mohammed and Yun Fu for their valuable comments.

References

- [1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *PAMI*, 19(7):711–720, 1997.
- [2] A. Bronstein, M. Bronstein, and R. Kimmel. Expression-Invariant Representations of Faces. *IEEE Transactions on Image Processing*, 16(1):188–197, 2007.
- [3] T. Cootes, G. Edwards, and C. Taylor. Active Appearance Models. *PAMI*, 23(6):681–685, 2001.
- [4] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977.
- [5] L. Ding and A. Martinez. Precise detailed detection of faces and facial features. In *CVPR*, 2008.
- [6] M. Everingham and A. Zisserman. Regression and Classification Approaches to Eye Localization in Face Images. In *FGR*, pages 441–448, 2006.
- [7] X. He, S. Yan, Y. Hu, and P. Niyogi. Face Recognition Using Laplacianfaces. *PAMI*, 27(3):328–340, 2005.
- [8] J. Ilonen, J. Kamarainen, P. Paalanen, M. Hamouz, J. Kittler, and H. Kalviainen. Image Feature Localization by Multiple Hypothesis Testing of Gabor Features. *IEEE Transactions on Image Processing*, 17(3):311–325, 2008.
- [9] S. Ioffe. Probabilistic Linear Discriminant Analysis. In *ECCV*, pages 531–542, 2006.
- [10] X. Jiang, B. Mandal, and A. Kot. Eigenfeature Regularization and Extraction in Face Recognition. *PAMI*, 30(3):383–394, 2008.
- [11] D. Liu, K. Lam, and L. Shen. Illumination invariant face recognition. *Pattern Recognition*, 38(10):1705–1716, 2005.
- [12] M. Mahoor and M. Abdel Mottaleb. Facial Features Extraction in Color Images Using Enhanced Active Shape Model. In *FGR*, pages 144–148, 2006.
- [13] A. M. Martínez. Recognizing Imprecisely Localized, Partially Occluded, and Expression Variant Faces from a Single Sample per Class. *PAMI*, 24(6):748–763, 2002.
- [14] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recognition. *Pattern Recognition*, 33(11):1771–1782, 2000.
- [15] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the Face Recognition Grand Challenge. In *CVPR*, pages 947–954, 2005.
- [16] S. J. D. Prince and J. H. Elder. Probabilistic Linear Discriminant Analysis for Inferences About Identity. In *ICCV*, 2007.
- [17] S. J. D. Prince, J. H. Elder, J. Warrell, and F. M. Felisberti. Tied Factor Analysis for Face Recognition across Large Pose Differences. *PAMI*, 30(6):970–984, 2008.
- [18] E. Rentzeperis, A. Stergiou, A. Pnevmatikakis, and L. Polymenakos. *Artificial Intelligence Applications and Innovations (AIAI06)*, chapter Impact of Face Registration Errors on Recognition, pages 187–194. Springer, Berlin Heidelberg, June 2006.
- [19] S. Shan, B. Cao, Y. Su, L. Qing, X. Chen, and W. Gao. Unified Principal Component Analysis with generalized Covariance Matrix for face recognition. In *CVPR*, 2008.
- [20] M. Turk and A. Pentland. Face Recognition Using Eigenfaces. In *CVPR*, pages 586–591, 1991.
- [21] H. Wang, S. Yan, T. Huang, J. Liu, and X. Tang. Misalignment-robust face recognition. In *CVPR*, 2008.
- [22] X. Xie, W. Zheng, J. Lai, and P. Yuen. Face illumination normalization on large and small scale features. In *CVPR*, 2008.
- [23] J. Yang, A. F. Frangi, J. Yu Yang, and Z. Jin. KPCA Plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition. *PAMI*, 27(2):230–244, 2005.
- [24] T. Zhang, B. Fang, Y. Tang, G. He, and J. Wen. Topology Preserving Non-negative Matrix Factorization for Face Recognition. *IEEE Transactions on Image Processing*, 17(4):574–584, 2008.
- [25] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.