Tied Factor Analysis for Face Recognition across Large Pose Differences

Simon J.D. Prince, James H. Elder, Jonathan Warrell, Fatima M. Felisberti

IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, JUNE 2008

Presented by Lan Du

July 25th, 2008

Outline

- Motivation
- Overview of Some Existing Methods for Face Recognition across Pose and the Proposed Method
- Detailed Model and Its Application
  - Observation and Identity Spaces
  - Tied Factor Analysis
  - Learning System Parameters
  - Learning Results
  - Recognition
- Experiments
- Conclusions

Motivation

One of the greatest remaining research challenges in face recognition is to recognize faces across different poses, expressions, and illuminations. Current face recognition systems require the implicit cooperation of the user.

- Face recognition from security footage.
- Face recognition in archive footage.
- Face recognition for HCI and ambient intelligence.

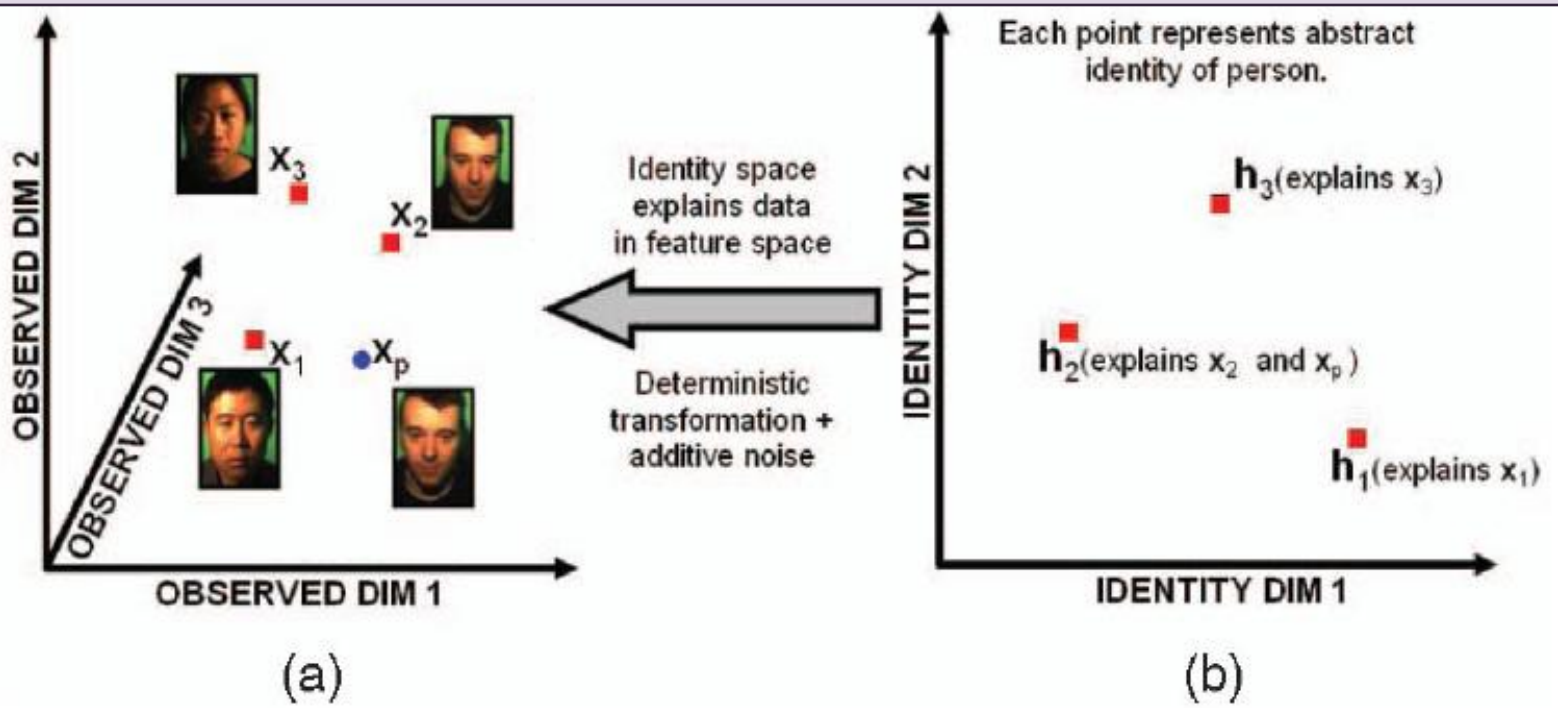
In this paper, the authors try to examine the worst case, in which there is only a single instance of each individual in a large database, and the *probe* image is taken from a very different pose than the matching *gallery* images.

Algorithms for Face Recognition across Pose

- Record each subject at each possible angle, then use a statistical model for each or create a 3D model of the head. -- require the cooperation of the user
- 3D Geometric Approaches: take a single probe image at one pose and create a full 3D head. -- complex to implement and are computationally expensive
- Statistical Approaches: the relationship between frontal and nonfrontal images is treated as a statistical learning problem. -- simpler and computationally cheaper but produce relatively poor results
  - *Global statistical models*
  - *Local statistical models:* build several models relating different parts of the face

Overview of the Proposed Method

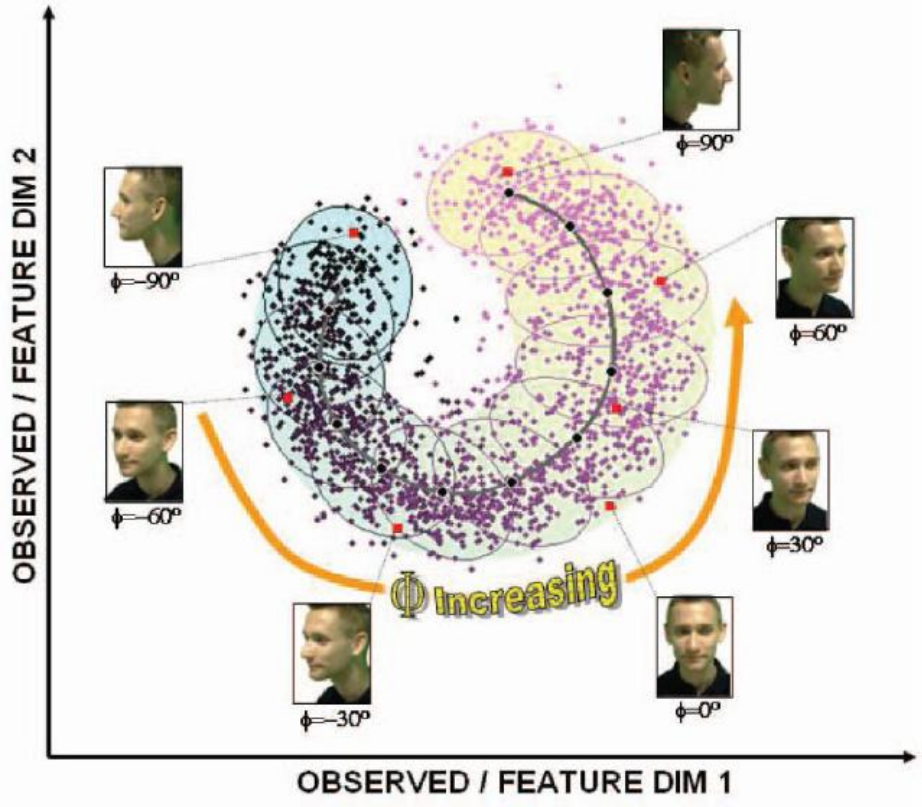
The algorithm is based on a generative model that describes how an underlying pose-invariant representation created the (pose-varying) observed data.



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).

Observation and Identity Spaces

- Observed Data: the raw gray values of the image or some simple deterministic transformation of these values, which does not attempt to compensate for pose variations. -- Observation Space
- Latent Identity Variable: a multidimensional variable that represents the identity of the individual, regardless of the pose. -- Identity Space



The effect of pose variation in the observation space. First, the mean position in the manifold changes systematically with the pose of the face. Second, for a given individual at a given pose, the position of the observation vector, relative to this mean, also varies.

Observation and Identity Spaces (*contd.*)

- Generation Process:
  - Choose the point in the identity space that corresponds to an individual.
  - Choose a pose.
  - Transform this identity variable to the observation space by using a deterministic function, which depends on the pose.
  - Add noise to the resulting observation vector.
- Latent Identity Variable -- describing the shape and structure of the face

  Deterministic Function -- representing the perspective projection process, which is parameterized by pose.

  Noise Term -- representing the measurement noise in the camera, plus all unmodeled aspects of the situation such as expression and lighting variation.

Tied Factor Analysis

Standard Factor Analysis

$$\boldsymbol{x}_{ij} = \boldsymbol{F}_i \boldsymbol{h}_{ij} + \boldsymbol{m}_i + \boldsymbol{\epsilon}_{ij}$$

$$\boldsymbol{h}_{ij} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \quad \boldsymbol{\epsilon}_{ij} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$$

$$\boldsymbol{x}_{ij} \,|\, \boldsymbol{h}_{ij} \sim \mathcal{N}(\boldsymbol{F}_i \boldsymbol{h}_{ij} + \boldsymbol{m}_i, \boldsymbol{\Sigma})$$

$$\begin{aligned} &p(\boldsymbol{x}_{ij} \,|\, \boldsymbol{F}_i, \boldsymbol{m}_i, \boldsymbol{\Sigma}) \\ &\quad = \int p(\boldsymbol{h}_{ij}) p(\boldsymbol{x}_{ij} \,|\, \boldsymbol{h}_{ij}, \boldsymbol{F}_i, \boldsymbol{m}_i, \boldsymbol{\Sigma}) \, \mathrm{d}\boldsymbol{h}_{ij} \\ &\quad = \mathcal{N}(\boldsymbol{x}_{ij} \,;\, \boldsymbol{m}_i, \boldsymbol{F}_i \boldsymbol{F}_i^{\mathrm{T}} + \boldsymbol{\Sigma}) \end{aligned}$$

$\boldsymbol{x}_{ij}$: the jth observed data vector of the ith individual

$\boldsymbol{h}_{ij}$: factor corresponding to $\boldsymbol{x}_{ij}$

$\boldsymbol{F}_i$: factor loading for the ith individual

$\boldsymbol{m}_i$: offset of the ith individual

$\boldsymbol{\epsilon}_{ij}$: noise term corresponding to $\boldsymbol{x}_{ij}$

Tied Factor Analysis

$$\boldsymbol{x}_{ijk} = \boldsymbol{F}_k \boldsymbol{h}_i + \boldsymbol{m}_k + \boldsymbol{\epsilon}_{ijk}$$

$$\boldsymbol{h}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}), \quad \boldsymbol{\epsilon}_{ijk} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_k)$$

$$\boldsymbol{x}_{ijk} \,|\, \boldsymbol{h}_i \sim \mathcal{N}(\boldsymbol{F}_k \boldsymbol{h}_i + \boldsymbol{m}_k, \boldsymbol{\Sigma}_k)$$

$$\begin{aligned} &p(\boldsymbol{x}_{ijk} \,|\, \boldsymbol{F}_k, \boldsymbol{m}_k, \boldsymbol{\Sigma}_k) \\ &\quad = \int p(\boldsymbol{h}_i) p(\boldsymbol{x}_{ijk} \,|\, \boldsymbol{h}_i, \boldsymbol{F}_k, \boldsymbol{m}_k, \boldsymbol{\Sigma}_k) \, \mathrm{d}\boldsymbol{h}_i \\ &\quad = \mathcal{N}(\boldsymbol{x}_{ijk} \,;\, \boldsymbol{m}_k, \boldsymbol{F}_k \boldsymbol{F}_k^{\mathrm{T}} + \boldsymbol{\Sigma}_k) \end{aligned}$$

$\boldsymbol{x}_{ijk}$: the jth observed data vector of the ith individual in the kth pose

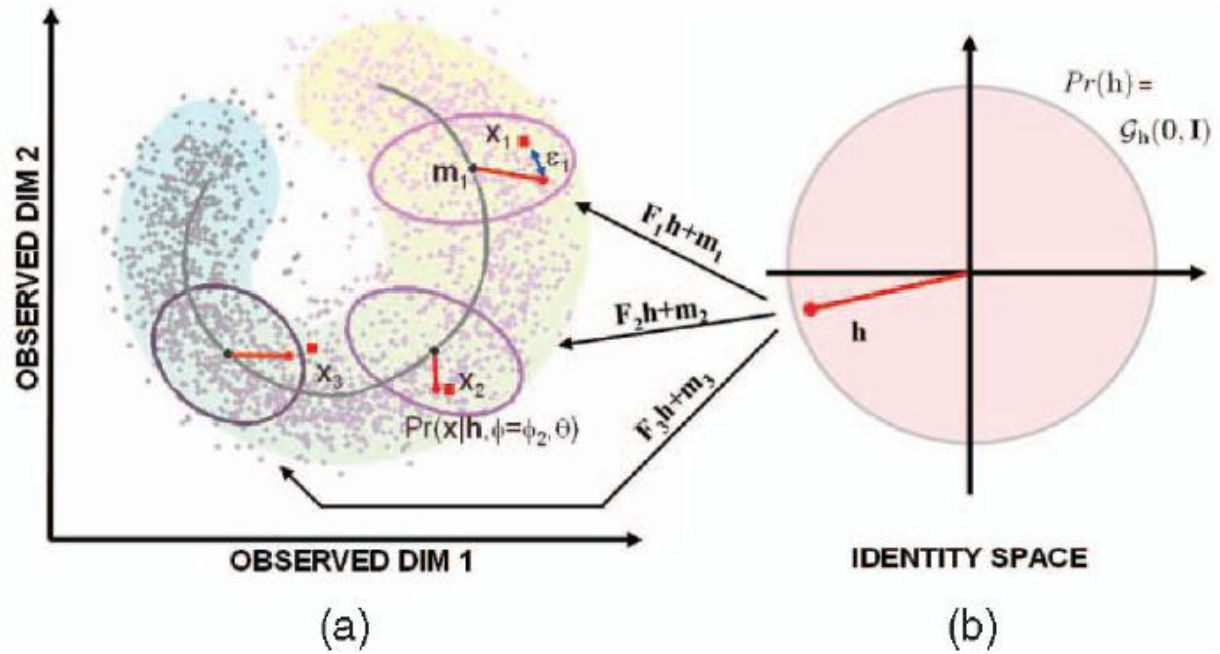
$\boldsymbol{h}_i$: factor for the ith individual, which is same at each pose (tied)

$\boldsymbol{F}_k$: factor loading of the kth pose

$\boldsymbol{m}_k$: offset of the kth pose

$\boldsymbol{\epsilon}_{ijk}$: noise term corresponding to $\boldsymbol{x}_{ijk}$

TIED FACTOR ANALYSIS (*CONTD.*)



Tied factor analysis model. (a) Observed measurement space. (b) “Identity” space. The three square symbols in (a) represent observed data for one person viewed at three poses. The circle symbol in (b) represents the latent identity variable for this person. Data in the observation space are explained by transforming latent identity variable by a pose-dependent transform and by adding noise.

LEARNING SYSTEM PARAMETERS

E-Step:

$$\mathrm{E}[\boldsymbol{h}_i \,|\, \boldsymbol{x}_{i..}] = \left(\boldsymbol{I} + \sum_{j=1}^{J}\sum_{k=1}^{K} \boldsymbol{F}_k^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{F}_k\right)^{-1} \sum_{j=1}^{J}\sum_{k=1}^{K} \boldsymbol{F}_k^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_{ijk} - \boldsymbol{m}_k)$$

$$\mathrm{E}[\boldsymbol{h}_i \boldsymbol{h}_i^{\mathrm{T}} \,|\, \boldsymbol{x}_{i..}] = \left(\boldsymbol{I} + \sum_{i=1}^{J}\sum_{k=1}^{K} \boldsymbol{F}_k^{\mathrm{T}} \boldsymbol{\Sigma}_k^{-1} \boldsymbol{F}_k\right)^{-1} + \mathrm{E}[\boldsymbol{h}_i \,|\, \boldsymbol{x}_{i..}]\mathrm{E}[\boldsymbol{h}_i \,|\, \boldsymbol{x}_{i..}]^{\mathrm{T}}$$

M-Step:

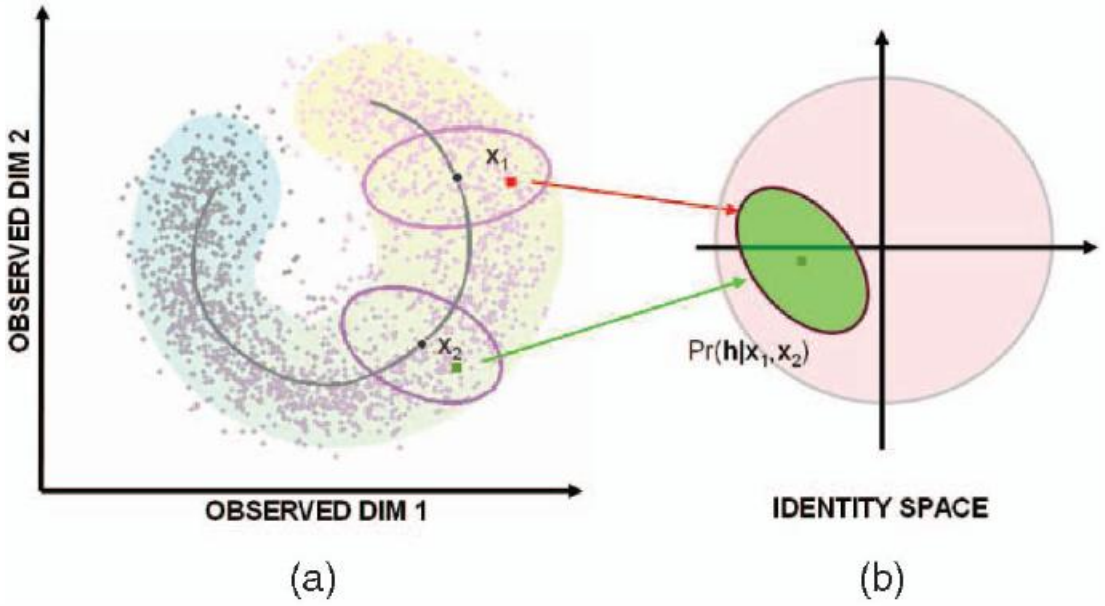
$$\tilde{\boldsymbol{F}}_k = \left(\sum_{i=1}^{I}\sum_{j=1}^{J} \boldsymbol{x}_{ijk} \mathrm{E}[\tilde{\boldsymbol{h}}_i \,|\, \boldsymbol{x}_{i..}]^{\mathrm{T}}\right) \left(\sum_{i=1}^{I}\sum_{j=1}^{J} \mathrm{E}[\tilde{\boldsymbol{h}}_i \tilde{\boldsymbol{h}}_i^{\mathrm{T}} \,|\, \boldsymbol{x}_{i..}]\right)$$

$$\boldsymbol{\Sigma}_k = \frac{1}{IJ} \sum_{i=1}^{I}\sum_{j=1}^{J} \mathrm{diag}[\boldsymbol{x}_{ijk} \boldsymbol{x}_{ijk}^{\mathrm{T}} - \tilde{\boldsymbol{F}}_k \mathrm{E}[\tilde{\boldsymbol{h}}_i \,|\, \boldsymbol{x}_{i..}] \boldsymbol{x}_{ijk}^{\mathrm{T}}]$$

$$\tilde{\boldsymbol{F}}_k = [\boldsymbol{F}_k \;\; \boldsymbol{m}_k], \quad \tilde{\boldsymbol{h}}_i = [\boldsymbol{h}_i^{\mathrm{T}} \;\; 1]^{\mathrm{T}}$$

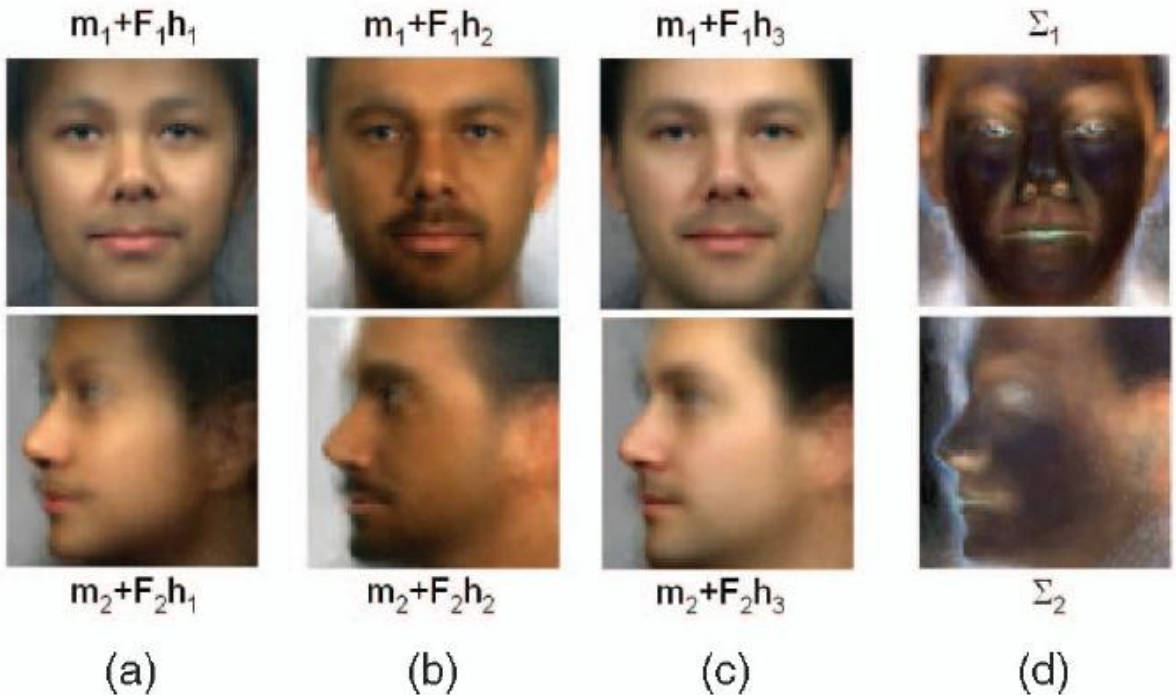
✓(b) The E-Step calculates the posterior probability distribution over the latent identity variables. (a) This is inferred from the observed data for that individual across all poses.

✓The M-Step optimizes the values of the transformation parameters for each pose by using data for that pose across all individuals.



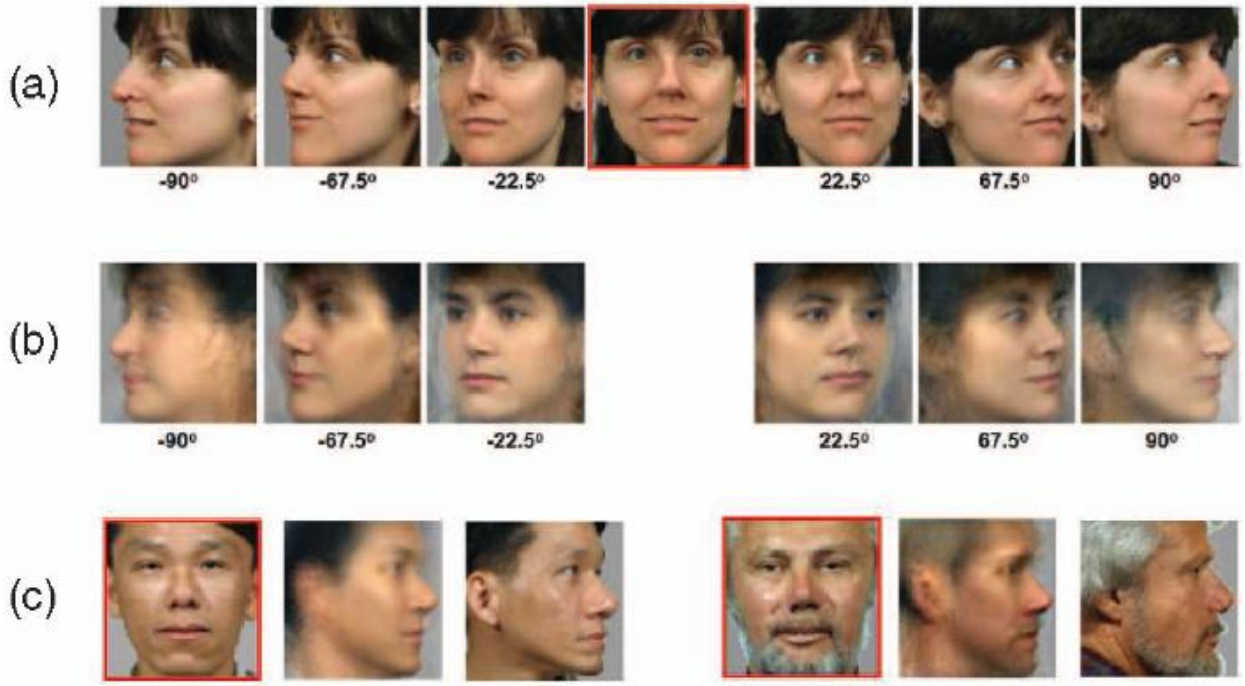
Learning Results

- FERET Dataset: 320 individuals at 7 poses -90, -67.5, -22.5, 0, 22.5, 67.5 and 90˚; 220 individuals for training and 100 individuals for testing at each pose; identifying 21 keypoints on each face by hand and extracting the corresponding image features.



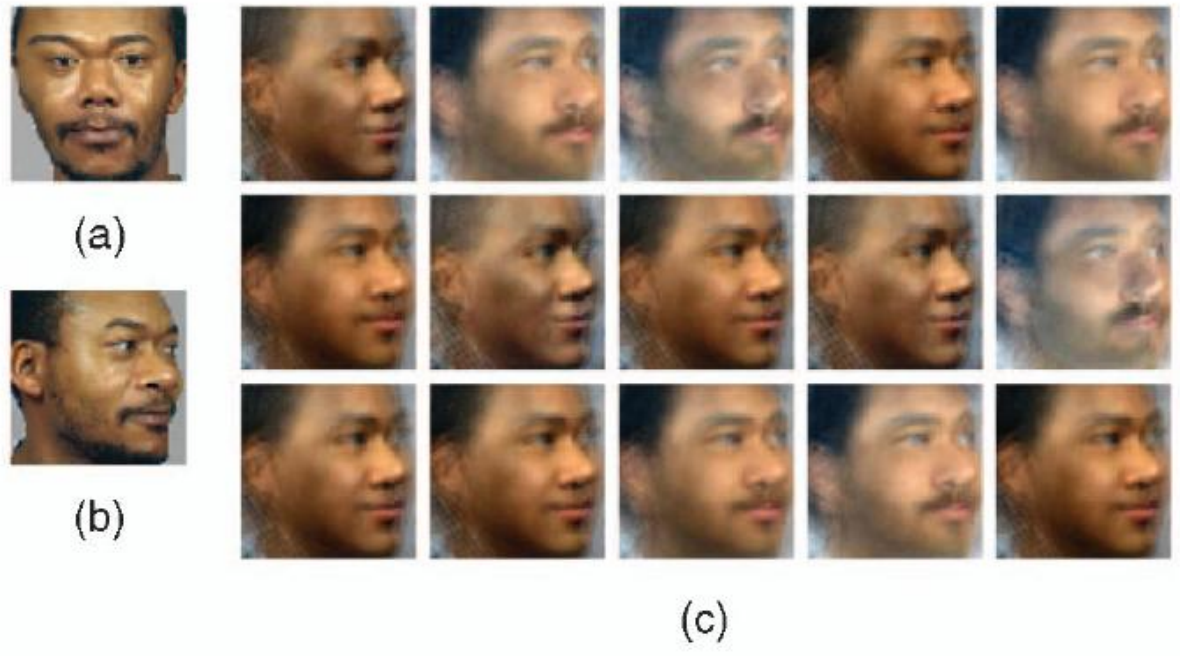
Generated face images with 16 factors. (a), (b), and (c) Three points in the identity space projected back into the observation space through frontal and profile models. (d) Per-pixel noise terms for frontal and profile models. Brighter points represent pixels with more noise.

LEARNING RESULTS (*CONTD.*)



Prediction of nonfrontal faces from frontal faces (project the mean of the latent identity variable back to the image space by using a nonfrontal transformation) 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.

LEARNING RESULTS (*CONTD.*)



Prediction of nonfrontal faces from frontal faces (project the samples of the latent identity variable back to the image space by using a nonfrontal transformation) with 16 factors. (a) Frontal image of subject. (b) Actual nonfrontal image of subject. (c) Fifteen projected samples.

RECOGNITION

$$
\begin{aligned}
p(\boldsymbol{x}_{1,\cdots,N}, \boldsymbol{x}_p \,|\, \mathcal{M}_n, \boldsymbol{\theta}) &= \int p(\boldsymbol{x}_{1,\cdots,N}, \boldsymbol{x}_p, \boldsymbol{h}_{1,\cdots,N}, \boldsymbol{h}_p \,|\, \boldsymbol{h}_p = \boldsymbol{h}_n, \boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{h}_{1,\cdots,N} \\
&= \int p(\boldsymbol{x}_1, \boldsymbol{h}_1 \,|\, \boldsymbol{\theta}_{q_1}) \, \mathrm{d}\boldsymbol{h}_1 \cdots \int p(\boldsymbol{x}_n, \boldsymbol{x}_p, \boldsymbol{h}_n \,|\, \boldsymbol{\theta}_{q_n}, \boldsymbol{\theta}_{q_p}) \, \mathrm{d}\boldsymbol{h}_n \cdots \int p(\boldsymbol{x}_N, \boldsymbol{h}_N \,|\, \boldsymbol{\theta}_{q_N}) \, \mathrm{d}\boldsymbol{h}_N \\
&= \int p(\boldsymbol{h}_1) p(\boldsymbol{x}_1 \,|\, \boldsymbol{h}_1, \boldsymbol{\theta}_{q_1}) \, \mathrm{d}\boldsymbol{h}_1 \cdots \int p(\boldsymbol{h}_n) p(\boldsymbol{x}_n, \boldsymbol{x}_p \,|\, \boldsymbol{h}_n, \boldsymbol{\theta}_{q_n}, \boldsymbol{\theta}_{q_p}) \, \mathrm{d}\boldsymbol{h}_n \cdots \int p(\boldsymbol{h}_N) p(\boldsymbol{x}_N \,|\, \boldsymbol{h}_N, \boldsymbol{\theta}_{q_N}) \, \mathrm{d}\boldsymbol{h}_N \\
&\propto \int p(\boldsymbol{x}_n, \boldsymbol{x}_p, \boldsymbol{h}_n \,|\, \boldsymbol{\theta}_{q_n}, \boldsymbol{\theta}_{q_p}) \, \mathrm{d}\boldsymbol{h}_n \Big/ \int p(\boldsymbol{x}_n, \boldsymbol{h}_n \,|\, \boldsymbol{\theta}_{q_n}) \, \mathrm{d}\boldsymbol{h}_n
\end{aligned}
$$

$$
\int p(\boldsymbol{x}_{n'}, \boldsymbol{h}_{n'} \,|\, \boldsymbol{\theta}_{q_{n'}}) \, \mathrm{d}\boldsymbol{h}_{n'} = \mathcal{N}(\boldsymbol{x}_{n'} \,;\, \boldsymbol{m}_{q_{n'}}, \boldsymbol{F}_{q_{n'}} \boldsymbol{F}_{q_{n'}}^{\mathrm{T}} + \boldsymbol{\Sigma}_{q_{n'}}), \quad n' \neq n
$$

$$
\int p(\boldsymbol{x}_n, \boldsymbol{x}_p, \boldsymbol{h}_n \,|\, \boldsymbol{\theta}_{q_n}, \boldsymbol{\theta}_{q_p}) \, \mathrm{d}\boldsymbol{h}_n = \mathcal{N}\left(\begin{bmatrix} \boldsymbol{x}_n \\ \boldsymbol{x}_p \end{bmatrix} ; \begin{bmatrix} \boldsymbol{m}_{q_n} \\ \boldsymbol{m}_{q_p} \end{bmatrix}, \begin{bmatrix} \boldsymbol{F}_{q_n} \\ \boldsymbol{F}_{q_p} \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_{q_n} \\ \boldsymbol{F}_{q_p} \end{bmatrix}^{\mathrm{T}} + \begin{bmatrix} \boldsymbol{\Sigma}_{q_n} & \\ & \boldsymbol{\Sigma}_{q_p} \end{bmatrix} \right)
$$

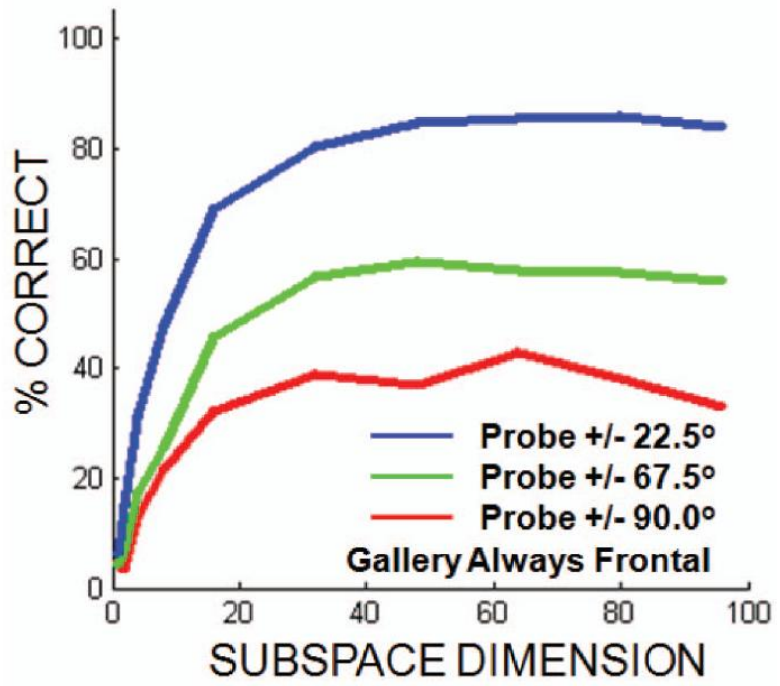
$$
\neq \int p(\boldsymbol{x}_n, \boldsymbol{h}_n \,|\, \boldsymbol{\theta}_{q_n}) \, \mathrm{d}\boldsymbol{h}_n \cdot \int p(\boldsymbol{x}_p, \boldsymbol{h}_n \,|\, \boldsymbol{\theta}_{q_p}) \, \mathrm{d}\boldsymbol{h}_n
$$

$$
p(\mathcal{M}_n \,|\, \boldsymbol{x}_{1,\cdots,N}, \boldsymbol{x}_p, \ \boldsymbol{\theta}) = \frac{p(\boldsymbol{x}_{1,\cdots,N}, \boldsymbol{x}_p \,|\, \mathcal{M}_n, \boldsymbol{\theta}) p(\mathcal{M}_n)}{\sum_{m=1}^{N} p(\boldsymbol{x}_{1,\cdots,N}, \boldsymbol{x}_p \,|\, \mathcal{M}_m, \boldsymbol{\theta}) p(\mathcal{M}_m)}
$$

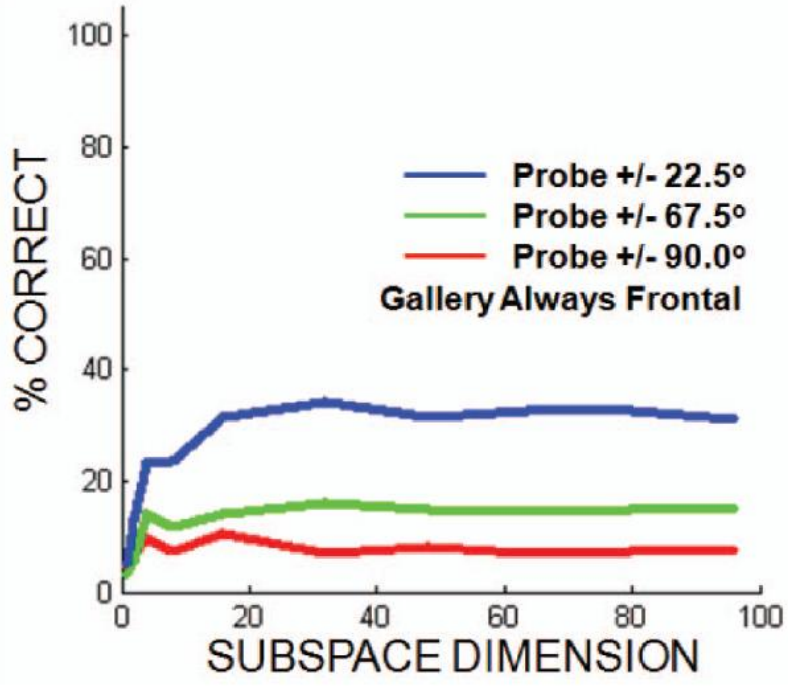
$\boldsymbol{x}_{1,\cdots,N}$	:	gallery dataset of faces, each of which belongs to a different individual
$\boldsymbol{x}_p$	:	a single probe face
$\boldsymbol{\theta}$	:	$\{\boldsymbol{F}_{1,\cdots,K}, \boldsymbol{m}_{1,\cdots,K}, \boldsymbol{\Sigma}_{1,\cdots,K}\}$
$\boldsymbol{\theta}_{q_n}$	:	$\{\boldsymbol{F}_{q_n}, \boldsymbol{m}_{q_n}, \boldsymbol{\Sigma}_{q_n}\}$
$\mathcal{M}_n$	:	the probe matches the nth gallery face

Experiment 1: Face Identification Using Raw Pixel Data

- 100 frontal testing faces as the gallery faces and a single nonfrontal face as the probe face
- "factor analysis model": only a single set of generation parameters

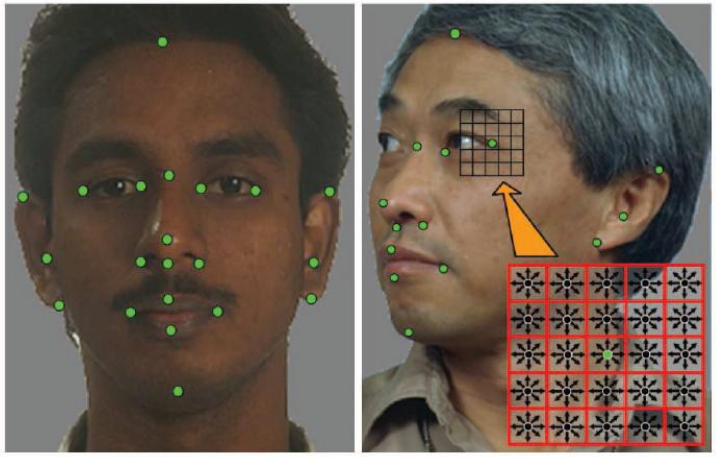


Percentage of first-match correct performance with the tied factor analysis model.



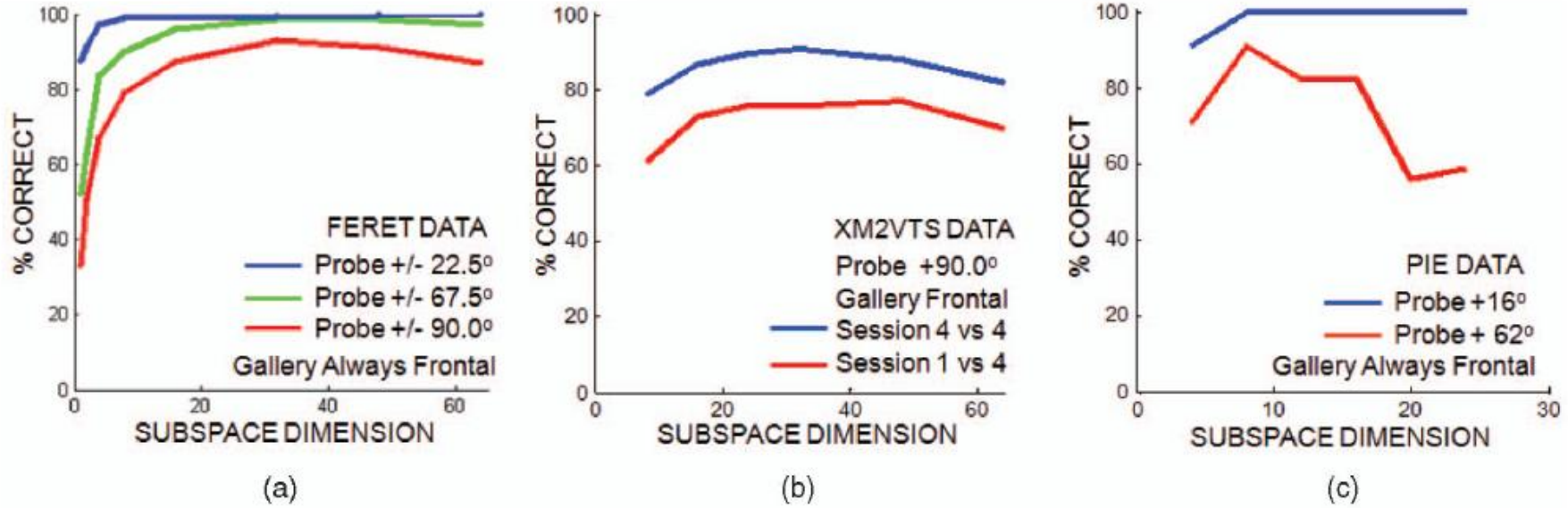
Percentage of first-match correct performance with the "factor analysis model".

Experiment 2: Face Identification with Local Gabor Data



Local measurements. (a) 21 keypoints on each face were identified by hand. (b) features were extracted at 25 spatial positions around each keypoint.

Build 21 local models to describe how these local facial features (*nose, eye, etc.*) change with pose.

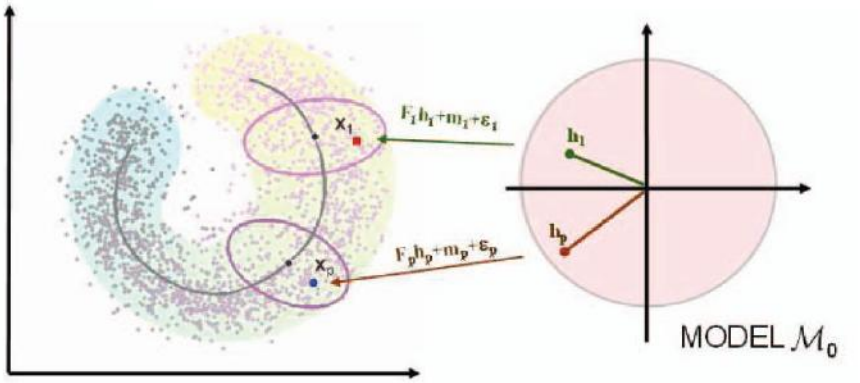


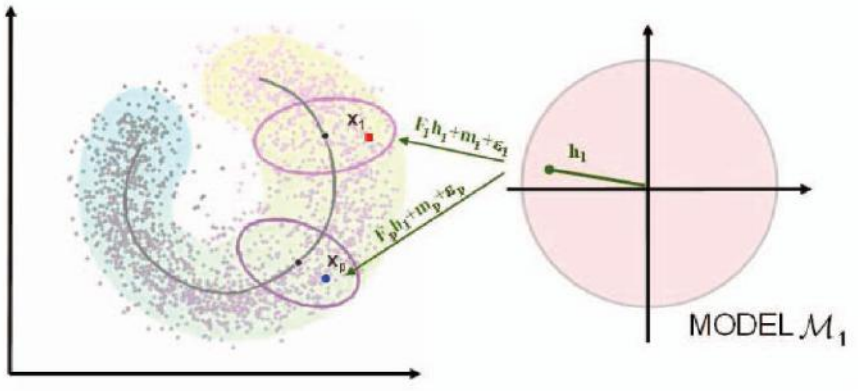
Percentage of first-match correct performance with the tied factor analysis model, combining 21 local Gabor models. (a) FERET dataset; (b) XM2VTS dataset; (c) PIE dataset.

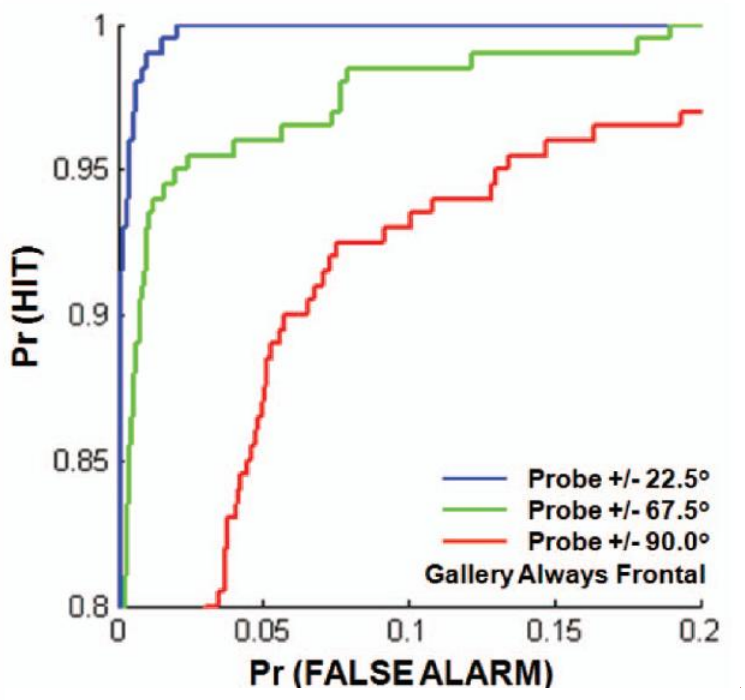
EXPERIMENT 3: FACE VERIFICATION

$$p(\boldsymbol{x}_1, \boldsymbol{x}_p \,|\, \mathcal{M}_0, \boldsymbol{\theta}) = \int p(\boldsymbol{h}_1) p(\boldsymbol{x}_1 \,|\, \boldsymbol{h}_1, \boldsymbol{\theta}_{q_1}) \,\mathrm{d}\boldsymbol{h}_1 \cdot \int p(\boldsymbol{h}_p) p(\boldsymbol{x}_p \,|\, \boldsymbol{h}_p, \boldsymbol{\theta}_{q_p}) \,\mathrm{d}\boldsymbol{h}_p$$
$$= \mathcal{N}(\boldsymbol{x}_1 \,;\, \boldsymbol{m}_{q_1}, \boldsymbol{F}_{q_1}\boldsymbol{F}_{q_1}^{\mathrm{T}} + \boldsymbol{\Sigma}_{q_1}) \cdot \mathcal{N}(\boldsymbol{x}_p \,;\, \boldsymbol{m}_{q_p}, \boldsymbol{F}_{q_p}\boldsymbol{F}_{q_p}^{\mathrm{T}} + \boldsymbol{\Sigma}_{q_p})$$

$$p(\boldsymbol{x}_1, \boldsymbol{x}_p \,|\, \mathcal{M}_1, \boldsymbol{\theta}) = \int p(\boldsymbol{h}_1) p(\boldsymbol{x}_1, \boldsymbol{x}_p \,|\, \boldsymbol{h}_1, \boldsymbol{\theta}_{q_1}, \boldsymbol{\theta}_p) \,\mathrm{d}\boldsymbol{h}_1$$
$$= \mathcal{N}\left(\begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_p \end{bmatrix} ; \begin{bmatrix} \boldsymbol{m}_{q_1} \\ \boldsymbol{m}_{q_p} \end{bmatrix}, \begin{bmatrix} \boldsymbol{F}_{q_1} \\ \boldsymbol{F}_{q_p} \end{bmatrix} \begin{bmatrix} \boldsymbol{F}_{q_1} \\ \boldsymbol{F}_{q_p} \end{bmatrix}^{\mathrm{T}} + \begin{bmatrix} \boldsymbol{\Sigma}_{q_1} & \\ & \boldsymbol{\Sigma}_{q_p} \end{bmatrix} \right)$$





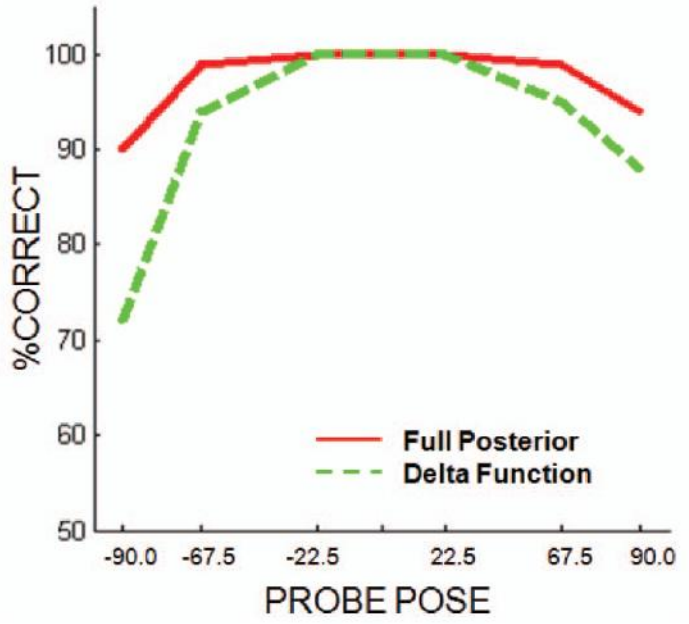


ROC curves of face verification using 21 local models.

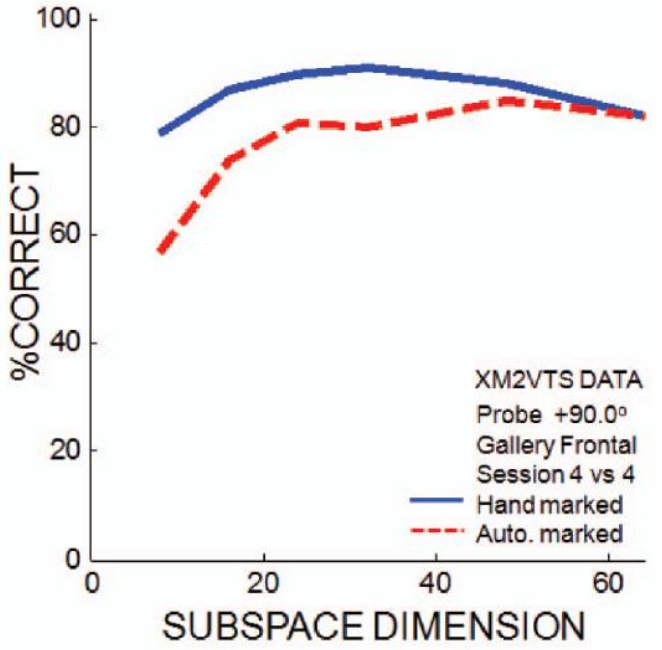
Experiments 4 and 5: Approximation of Evidence Term and Automated versus Manual Keypoint Detection

$$p(\boldsymbol{x}_{1,\cdots,N}, \boldsymbol{x}_p \,|\, \mathcal{M}_n, \boldsymbol{\theta}) \approx p(\hat{\boldsymbol{h}}_1)p(\boldsymbol{x}_1 \,|\, \hat{\boldsymbol{h}}_1, \boldsymbol{\theta}_{q_1}) \cdots p(\hat{\boldsymbol{h}}_n)p(\boldsymbol{x}_n, \boldsymbol{x}_p \,|\, \hat{\boldsymbol{h}}_n, \boldsymbol{\theta}_{q_n}, \boldsymbol{\theta}_{q_p}) \cdots p(\hat{\boldsymbol{h}}_N)p(\boldsymbol{x}_N \,|\, \hat{\boldsymbol{h}}_N, \boldsymbol{\theta}_{q_N})$$
$$\propto p(\hat{\boldsymbol{h}}_n)p(\boldsymbol{x}_p, \hat{\boldsymbol{h}}_n \,|\, \boldsymbol{\theta}_{q_p})$$

$$p(\boldsymbol{x}_n, \boldsymbol{x}_p, \hat{\boldsymbol{h}}_n \,|\, \boldsymbol{\theta}_{q_n}, \boldsymbol{\theta}_{q_p}) = p(\boldsymbol{x}_n, \hat{\boldsymbol{h}}_n \,|\, \boldsymbol{\theta}_{q_n}) \cdot p(\boldsymbol{x}_p, \hat{\boldsymbol{h}}_n \,|\, \boldsymbol{\theta}_{q_p})$$



Plot of the percentage of first-match correct performance for both full and approximate (delta function) models.



Plot of the percentage of first-match correct performance for both automated and manual keypoint registration.

Experiment 6: Comparison to Other Studies

Comparison of Face Identification Studies across Poses

STUDY	DATABASE	POSE DIFF (^{o})	% 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

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

Conclusion

- Fast
- Provides a posterior over the possible matches.
- Considers the case that the probe face is not in the database, without the need for choosing a threshold for the verification procedure.
- Only a single parameter: the dimension of the latent identity variables.
- Provides a clear way of incorporating multiple gallery or probe images.