Face Recognition based on a 3D Morphable Model

Zhang Kun

School Of Computing
National University of Singapore

Main paper

- Volker Blanz.
  Face Recognition based on a 3D Morphable Model.
  In Proc. Int. Conf. Automatic Face and Gesture
  Recognition, 2006.

Motivation

- It’s surprisingly difficult to find features in images of faces that remain invariant with respect to changes in pose and illumination

- Changes in pose and illumination are much less complex in 3D domain than in images

- An obvious invariant feature in different images of a rigid object is the 3D surface geometry with the local reflection properties of the material

Outline

Morphable Model

Estimation

Recognition

The morphable model is derived from 3D scans of 100 males and 100 females.

For each person, the information of 3D surface geometry and texture is stored in vectors Si and Ti.

$$\mathbf{S}_{\mathrm{i}} = (x_1, y_1, z_1, x_2, \ldots, x_n, y_n, z_n)^T.$$

$$\mathbf{T}_{\mathrm{i}} = (R_1, G_1, B_1, R_2, \ldots, R_n, G_n, B_n)^T.$$

Each vector Si is the 3D shape, stored in terms of x, y, z-coordinates of vertices k$\in \{1, \cdots, n\}$, $n = 75972$.

In the same way, texture vectors are formed from red, green, and blue values of all vertices' surface color.

We perform a Principle Component Analysis on the set of shape and texture vectors Si and Ti of example faces.

We obtain shape eigenvectors si, shape variances $\sigma^2_{S,1}$ and texture eigenvectors ti, texture variances $\sigma^2_{T,i}$.

Eigenvectors form an orthogonal basis, morphable models are defined such that any linear combination of the eigenvectors.

$$\mathbf{S} = \bar{\mathbf{s}} + \sum_i \alpha_i \cdot \mathbf{s}_i, \qquad \mathbf{T} = \bar{\mathbf{t}} + \sum_i \beta_i \cdot \mathbf{t}_i.$$

And PCA provides an estimation of probability density within face space:

$$p_S(\mathbf{S}) \sim e^{-\frac{1}{2}\sum_i \frac{\alpha_i^2}{\sigma_{S,i}^2}}, \quad p_T(\mathbf{T}) \sim e^{-\frac{1}{2}\sum_i \frac{\beta_i^2}{\sigma_{T,i}^2}}.$$

- The fitting algorithm optimizes shape coefficients α = (α1, α2, …) and texture coefficients β = (β1, β2, …) along with 22 rendering parameters, concatenated into a vector ρ.

- Cost Function

  Given an input image

$$\mathbf{I}_{input}(x,y) = (I_r(x,y), I_g(x,y), I_b(x,y))^T,$$

  the cost function

$$E_I = \sum_{x,y} \|\mathbf{I}_{input}(x,y) - \mathbf{I}_{model}(x,y)\|^2.$$

  Given the manually defined feature points($q_{x,j}$, $q_{y,j}$) and the image positions ($p_{x,kj}$, $p_{y,kj}$) of the corresponding vertices kj,

$$E_F = \sum_{j} \| \begin{pmatrix} q_{x,j} \\ q_{y,i} \end{pmatrix} - \begin{pmatrix} p_{x,k_j} \\ p_{y,k_j} \end{pmatrix} \|^2$$

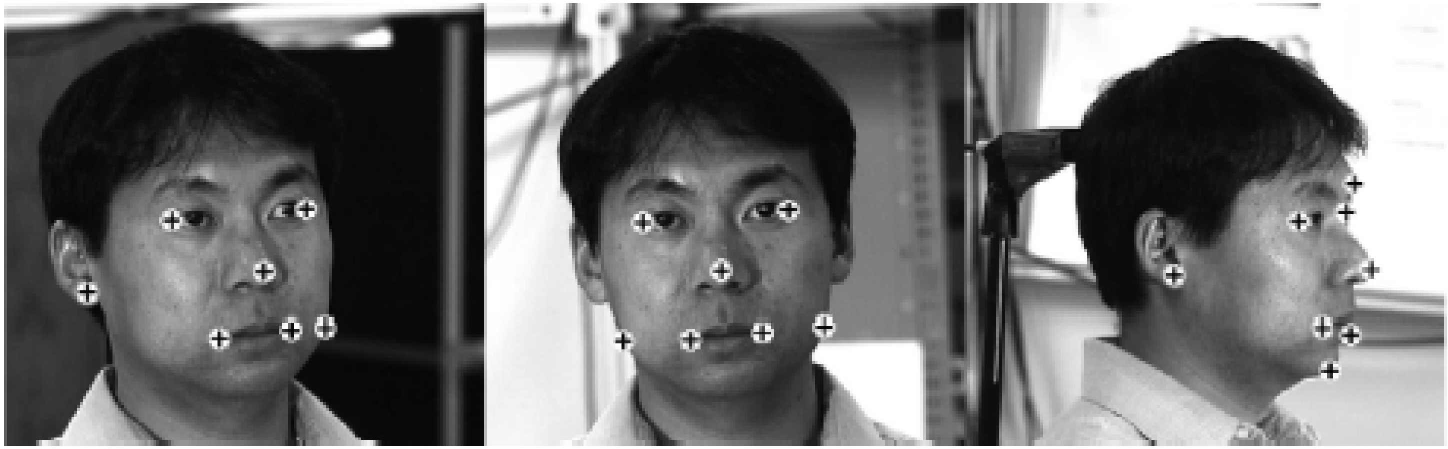


Figure 1. Up to seven feature points were manually labeled in front and side views, up to eight were labeled in profile views

- Given the input image $\mathbf{I}_{input}$ and the feature points F, the task is to find the model parameters with maximum posterior probability P(α, β, ρ| I_{input}, F). According to Bayes rules,

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\rho} \mid \mathbf{I}_{input}, F) \sim p(\mathbf{I}_{input}, F \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\rho}) \cdot P(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\rho})$$

- If we neglect correlations between these variables, the right hand side is

$$p(\mathbf{I}_{input} \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\rho}) \cdot p(F \mid \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\rho}) \cdot P(\boldsymbol{\alpha}) \cdot P(\boldsymbol{\beta}) \cdot P(\boldsymbol{\rho}).$$

- The prior probabilities P(α) and P(β) were estimated with PCA. And we assumed P(ρ) is a normal distribution and use the starting values for $\overline{\boldsymbol{\rho}}_i$

- Posterior probability is then maximized by minimizing

$$E = -2 \cdot log\, p(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\rho} \mid \mathbf{I}_{input}, F)$$

$$E = \frac{1}{\sigma_I^2} E_I + \frac{1}{\sigma_F^2} E_F + \sum_i \frac{\alpha_i^2}{\sigma_{S,i}^2} + \sum_i \frac{\beta_i^2}{\sigma_{T,i}^2} + \sum_i \frac{(\rho_i - \overline{\rho}_i)^2}{\sigma_{R,i}^2}$$

- The optimization is performed with a Stochastic Newton Algorithm: the optimum:

$$\boldsymbol{\alpha}^* = \boldsymbol{\alpha} - \mathbf{H}^{-1} \nabla E.$$

- In each iteration, we perform small steps $\alpha \mapsto \boldsymbol{\alpha} + \lambda(\boldsymbol{\alpha}^* - \boldsymbol{\alpha})$ with a factor $\lambda \ll 1$

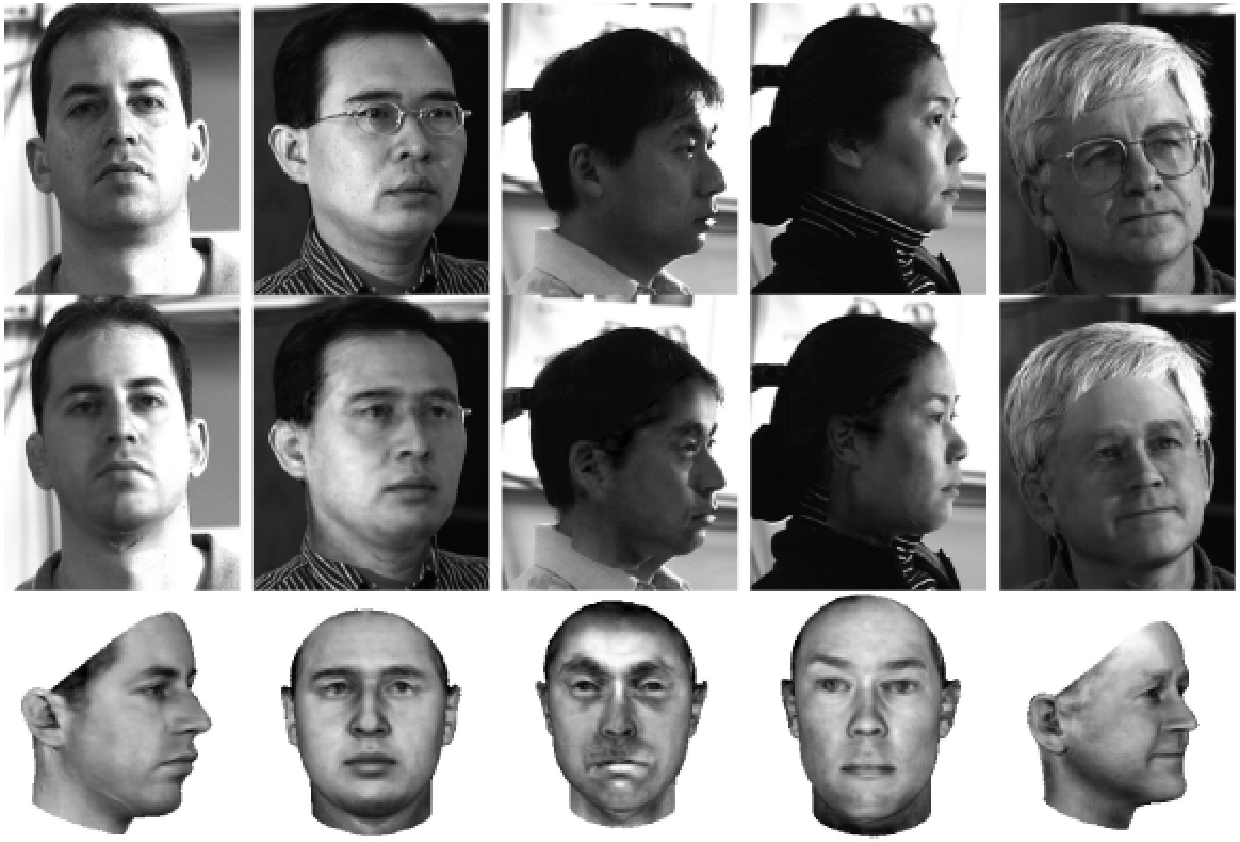
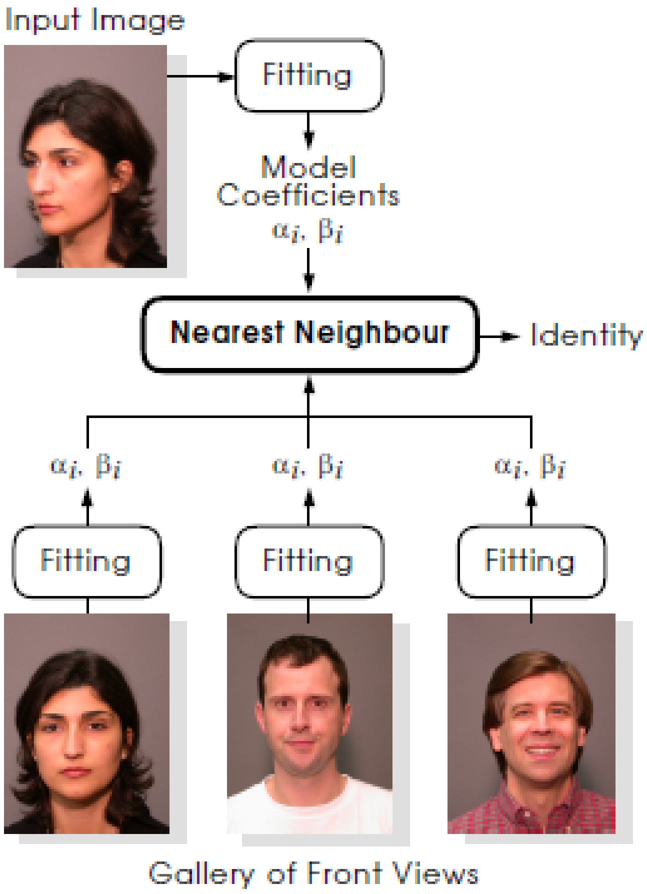


Figure 2. Three dimensional reconstructions. Top: originals; middle: reconstructions rendered into originals; bottom: novel views

Coefficient-Based Approach

- ➢The 3D shape and texture coefficients are obtained by fitting the model to the input image.



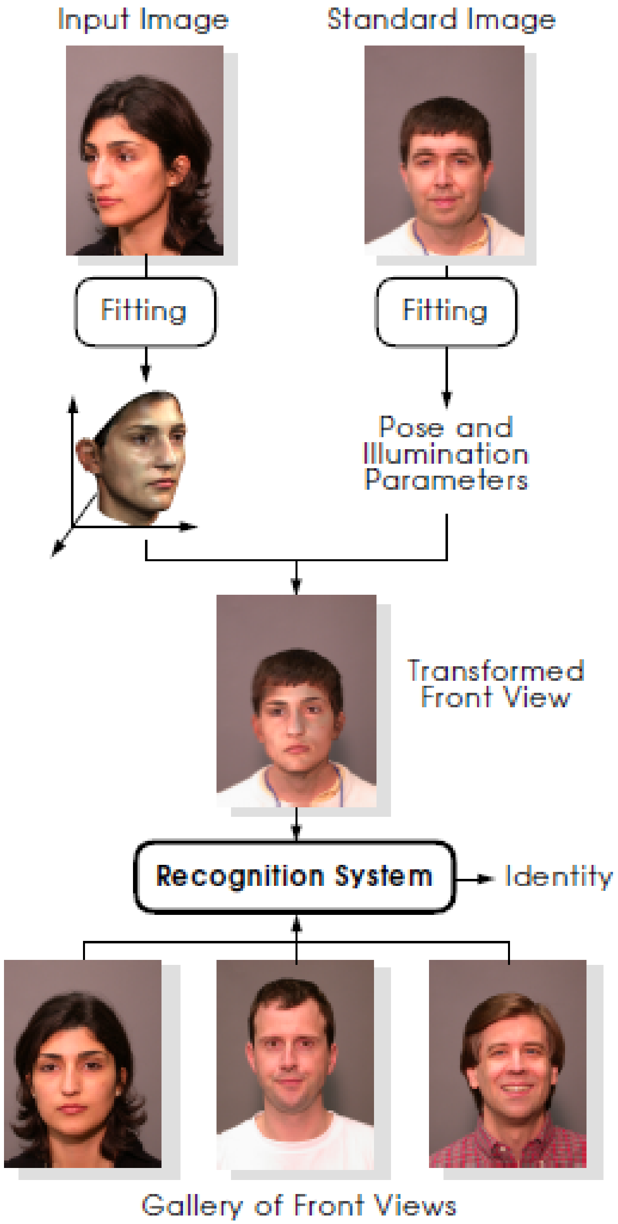
- ➢All probe and gallery images are processed by the model fitting algorithm. Identification is a nearest-neighbour search in this low-dimensional representation.

Viewpoint Normalization Approach

➢The previous approach is applicable across large changes in viewpoint, at the price of a relatively high computational cost.

➢Most face recognition algorithms that are commercially available today are restricted to images with close-to-frontal views only, but they are more computationally efficient.

➢In the viewpoint normalization approach, we use the morphable model as a preprocessing tool for generating frontal views from non-frontal images which are then input to the image-based recognition systems.



➢ The morphable model is used to estimate 3D shape and texture of the face, and this face is rendered in a frontal pose and standard size and illumination

➢ Restricted to the face, the morphable model can’t rotate the hairstyle and shoulders. So the algorithm inserts the face into an existing frontal portrait automatically.

Synthetic Training Set Approach

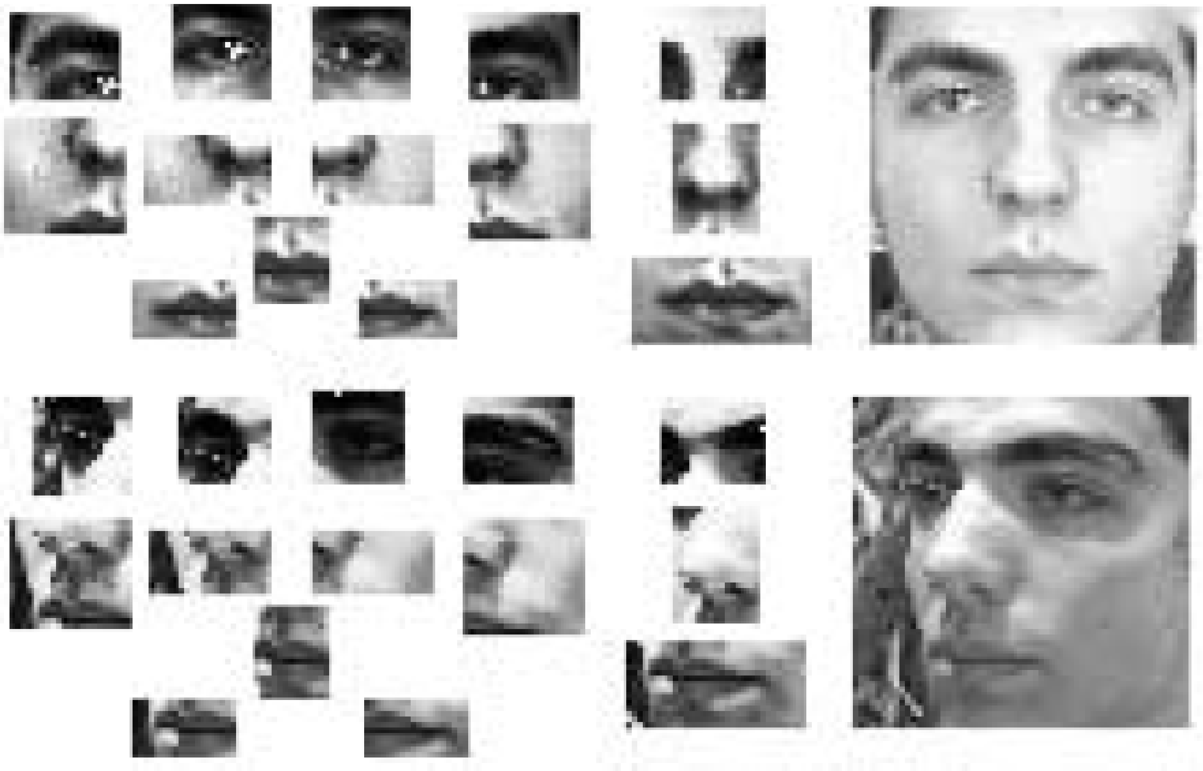
- Instead of generating a single view from each given input image, a 3D face model can also help to build a large variety of different views.

- From a small number of images of an individual, a 3D face model is reconstructed, and 7700 images per individual are rendered at different poses and illuminations.

- Along with the synthetic images, the rendering procedure also provides the 2D positions of features, and image regions around these features are cropped automatically.

- These subimages are used for training a support vector machine classifier for each individual.

Synthetic Training Set Approach



- Fourteen 2D subimages are extracted from a frontal view and a half profile view of a face

NUS
National University of Singapore

Thank you