# A morphable model for the synthesis of 3D faces

**Feng Wang**

**AIRD, Coretronic Co.**

June 24, 2019

**The slides and a list of references can be found from**
https://github.com/fwcore/object-detection

# Outlines

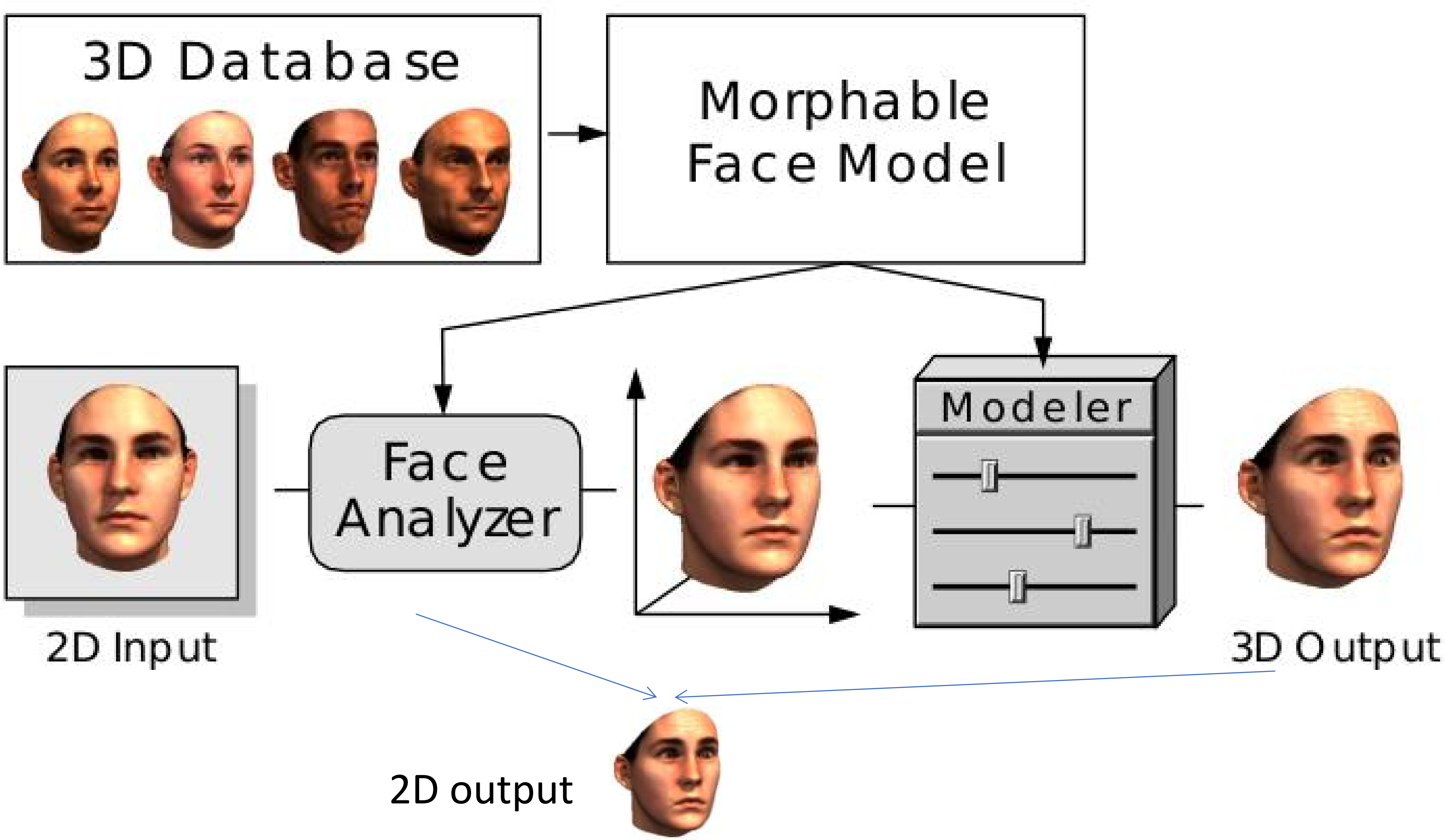
- **2D-to-3D and 3D-to-2D workflow**

- **3D morphable face model**
  - build model from 3D face database
  - modeling facial attributes
  - 3D correspondence using optic flow

- **Matching a morphable model to images**
  - matching 3D scan
  - matching 2D image

Blanz & Vetter (1999) A morphable model for the synthesis of 3D faces.

# 2D-to-3D and 3D-to-2D workflow



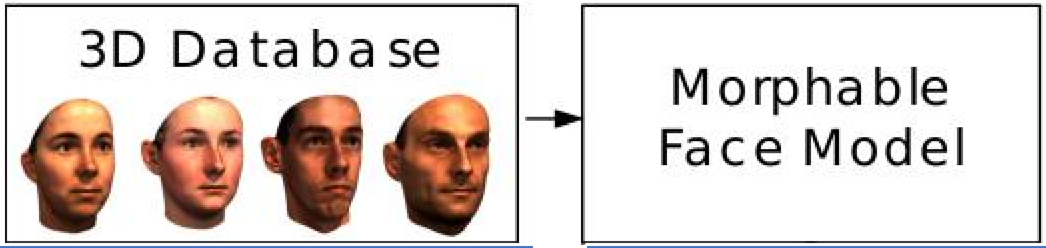

Blanz & Vetter (1999) A morphable model for the synthesis of 3D faces.

# Outlines

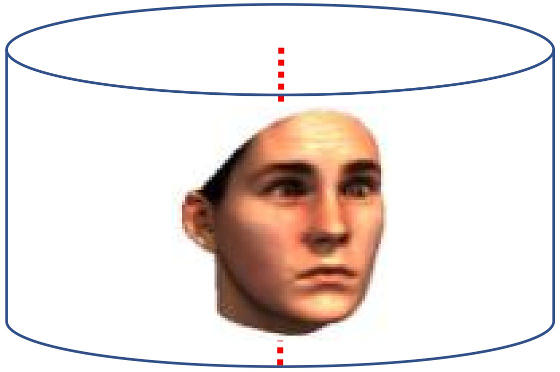
- 2D-to-3D and 3D-to-2D workflow
- **3D morphable face model**
  - build model from 3D face database
  - modeling facial attributes
  - 3D correspondence using optic flow
- **Matching a morphable model to images**
  - matching 3D scan
  - matching 2D image

Blanz & Vetter (1999) A morphable model for the synthesis of 3D faces.

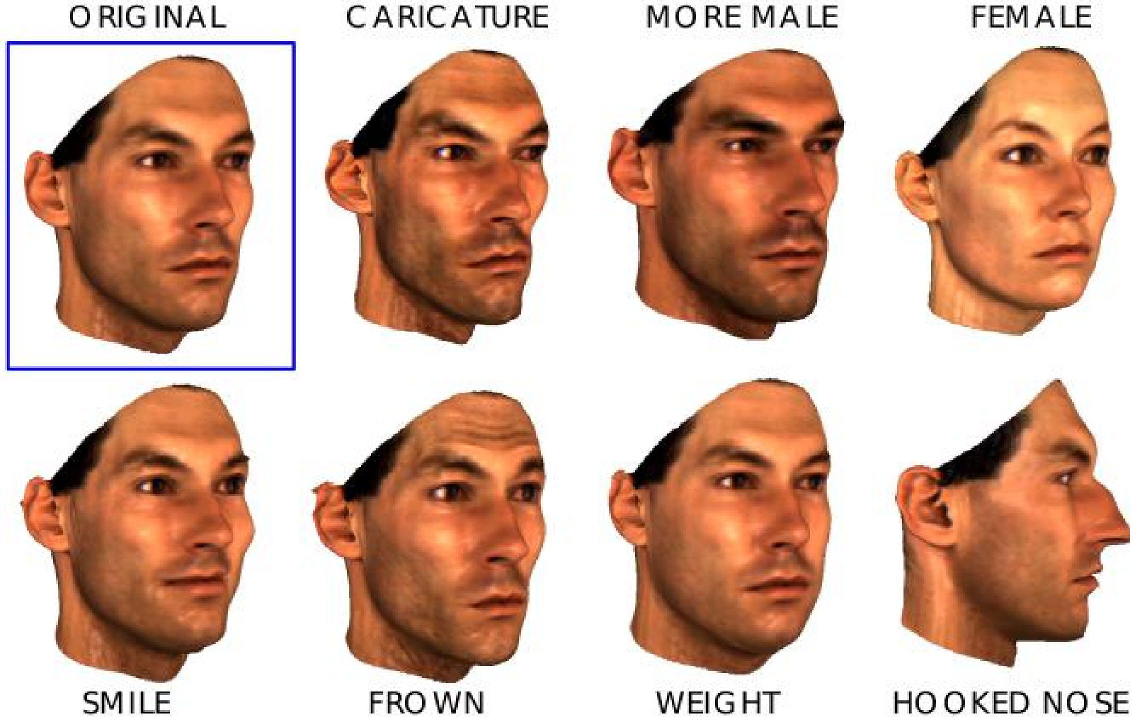
# 3D morphable face model



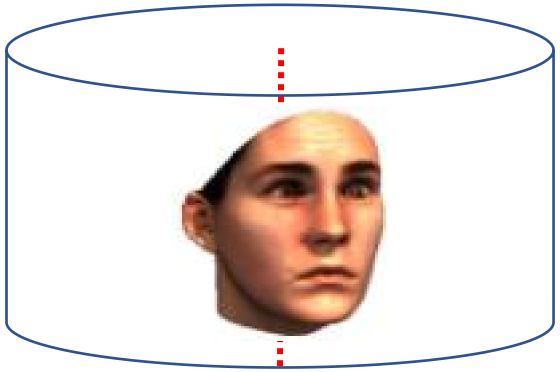

| 3D database (1999) |
|---|
| Cyberware™ laser scans |
| 100 young male and 100 young female |
| $r(h, \phi)$, $R(h, \phi)$, $G(h, \phi)$, $B(h, \phi)$ |
| 512 equally-spaced in $\phi$, ~70,000 vertices |
| no makeup, accessories, facial hair. Wearing bathing cap, and removed digitally. |



| Spec. of morphable face model (1999) |
|---|
| being real face-like |
| easy to change facial attributes |



# build model from 3D face database



**3D face**

- shape vector (x1, y1, z1, x2, y2, z2, ...)
- texture vector (R1, G1, B1, R2, G2, B2, ...)

**3n+3n dimensional vector**

n: number of vertices

m: number of faces in database

m << n

**Assumptions:**

- Gaussian distribution in 6n dim. space
- shape and texture are independent (to generate more faces)
- *face subregions are independent (to further increase the possible faces). Require blending the borders later on.*

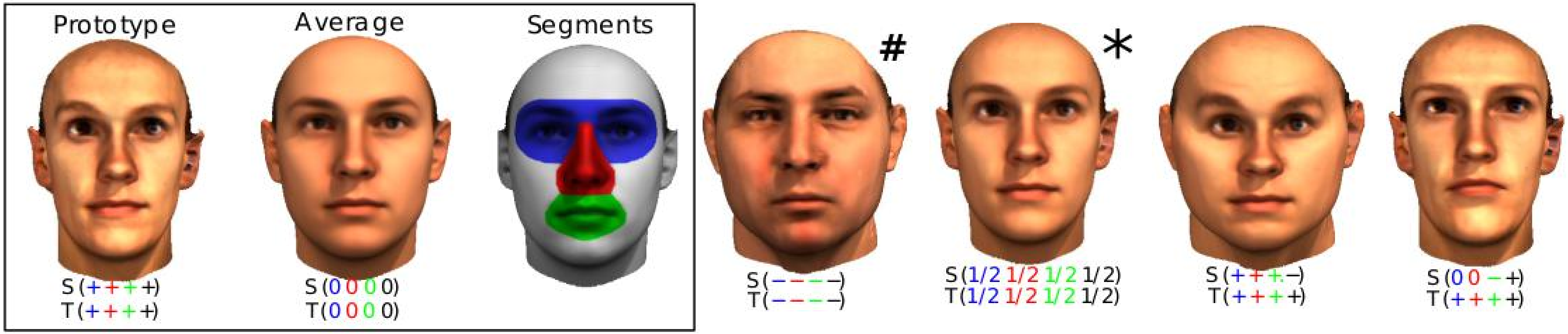
**Gaussian = PCA**:

PCs are decorrelated and Gaussian distributed.

$$S_{model} = \overline{S} + \sum_{i=1}^{m-1} \alpha_i s_i \, , \quad T_{model} = \overline{T} + \sum_{i=1}^{m-1} \beta_i t_i$$

$$p(\vec{\alpha}) \sim exp[-\frac{1}{2} \sum_{i=1}^{m-1} (\alpha_i / \sigma_i)^2],$$

# modeling facial attributes: linear model




$$\Delta S = \sum_{i=1}^{m} \mu_i (S_i - \overline{S}), \quad \Delta T = \sum_{i=1}^{m} \mu_i (T_i - \overline{T})$$

$\mu_i$: facial attribute (obtained by regression)

gender, fullness of faces, darkness of eyebrows,
double chins, hooked vs. concave noses.

# Preparing database: 3D correspondence using optic flow

***Key problem*: build a point-to-point correspondence between vertices of faces.**
***Mathematically*: consistently label vertices by 1 to n**

**3D scans**
- $r(h, \phi)$
- $R(h, \phi)$, $G(h, \phi)$, $B(h, \phi)$

?

**3D face**
- shape vector (x1, y1, z1, x2, y2, z2, ...)
- texture vector (R1, G1, B1, R2, G2, B2, ...)

unregistered face

reference face

***minimize***

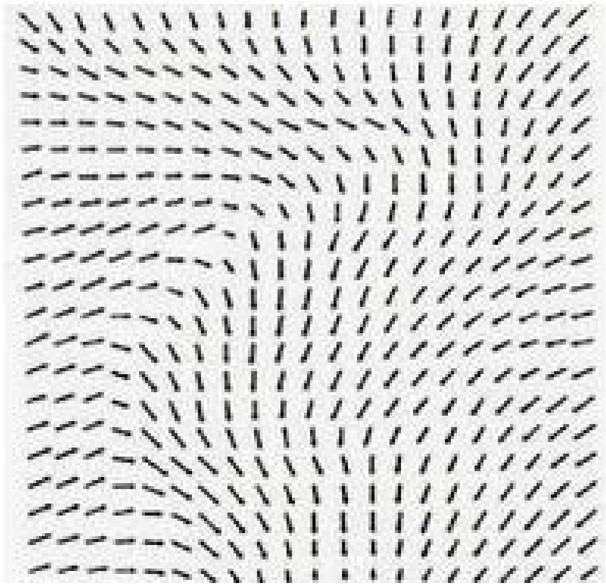
$$||\mathbf{I}_1(h,\phi) - \mathbf{I}_2(h+\delta h, \phi+\delta\phi)||$$

+

quadratic coupling between neighboring flow

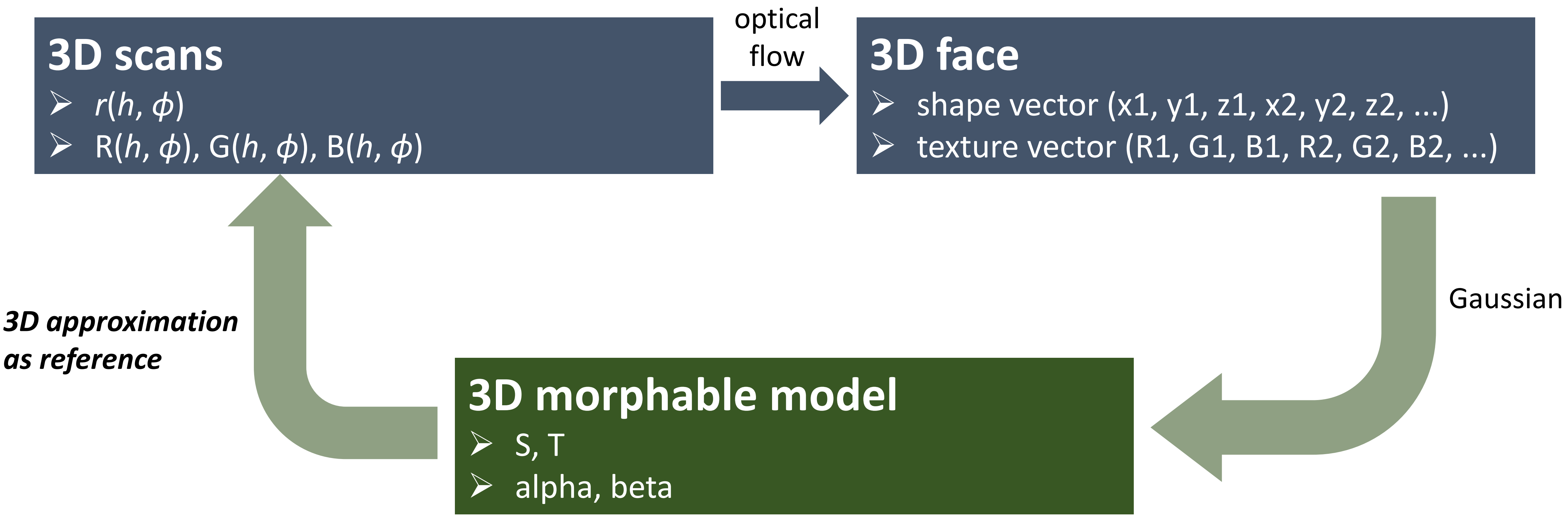
$(\delta h(h,\phi), \delta\phi(h,\phi))$

flow-field/correspondence



$\mathbf{I}(h,\phi) = (R(h,\phi), G(h,\phi), B(h,\phi), r(h,\phi))^T$ + mean curvature (optional)

# Preparing database: bootstrapping




*<u>We need tech to approximate 3D scans by 3D morphable model.</u>*

# Outlines

- **2D-to-3D and 3D-to-2D workflow**
- **3D morphable face model**
  - build model from 3D face database
  - modeling facial attributes
  - 3D correspondence using optic flow
- **Matching a morphable model to images**
  - matching 3D scan
  - matching 2D image

Blanz & Vetter (1999) A morphable model for the synthesis of 3D faces.

# matching 3D scan

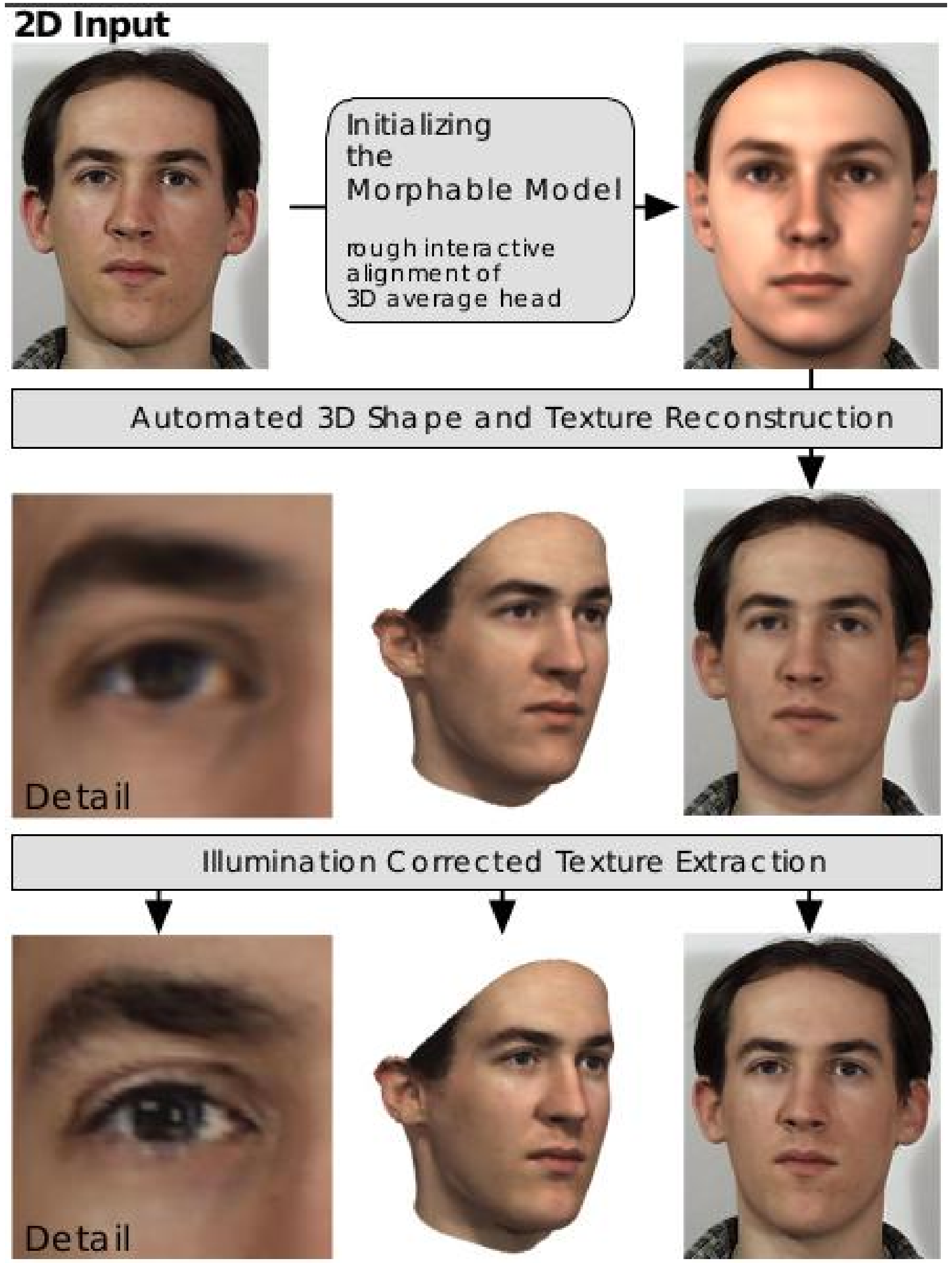
mapping $C : \mathcal{R}^3 \rightarrow \mathcal{R}^2, \quad (x, y, z) \mapsto (h, \phi)$

$$\mathbf{I}(h, \phi) = (R(h, \phi), G(h, \phi), B(h, \phi), r(h, \phi))^T$$

***minimize by varying α, β***

$$E = \sum_{h,\phi} \|\mathbf{I}_{input}(h, \phi) - \mathbf{I}_{model}(h, \phi)\|^2$$

# matching 2D image




## *3D Model parameters*

- ***α, β***
- ***rendering ρ:***
  - ***camera position,***
  - ***object scale,***
  - ***image plane rotation & translation,***
  - ***intensity of ambient light,***
  - ***intensity of directed light,***
  - ***color contrast, offset, and gain in RGB channel***
  - ***(fixed & estimated by user) camera distance***
  - ***(fixed & estimated by user) light direction***
  - ***(fixed & estimated by user) surface shininess***

# matching 2D image

***3D Model***
- ***α, β***
- ***ρ***

perspective

projection

2D image

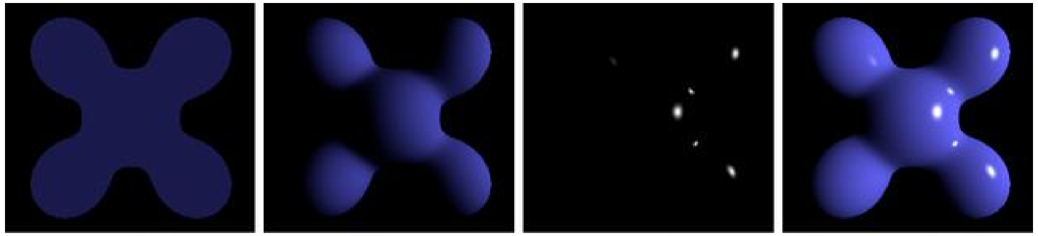
$$\mathbf{I}_{model}(x,y) = (I_{r,mod}(x,y), I_{g,mod}(x,y), I_{b,mod}(x,y))^T$$

$$I_{r,model,k} = (i_{r,amb} + i_{r,dir} \cdot (\mathbf{n}_k \mathbf{l}))\bar{R}_k + i_{r,dir}\, s \cdot (\mathbf{r}_k \mathbf{v}_k)^{\nu}$$

ambient light

directed light

specular refection




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

***minimize*** $$E = \frac{1}{\sigma_N^2} E_I + \sum_{j=1}^{m-1} \frac{\alpha_j^2}{\sigma_{S,j}^2} + \sum_{j=1}^{m-1} \frac{\beta_j^2}{\sigma_{T,j}^2} + \sum_j \frac{(\rho_j - \bar{\rho}_j)^2}{\sigma_{\rho,j}^2}$$

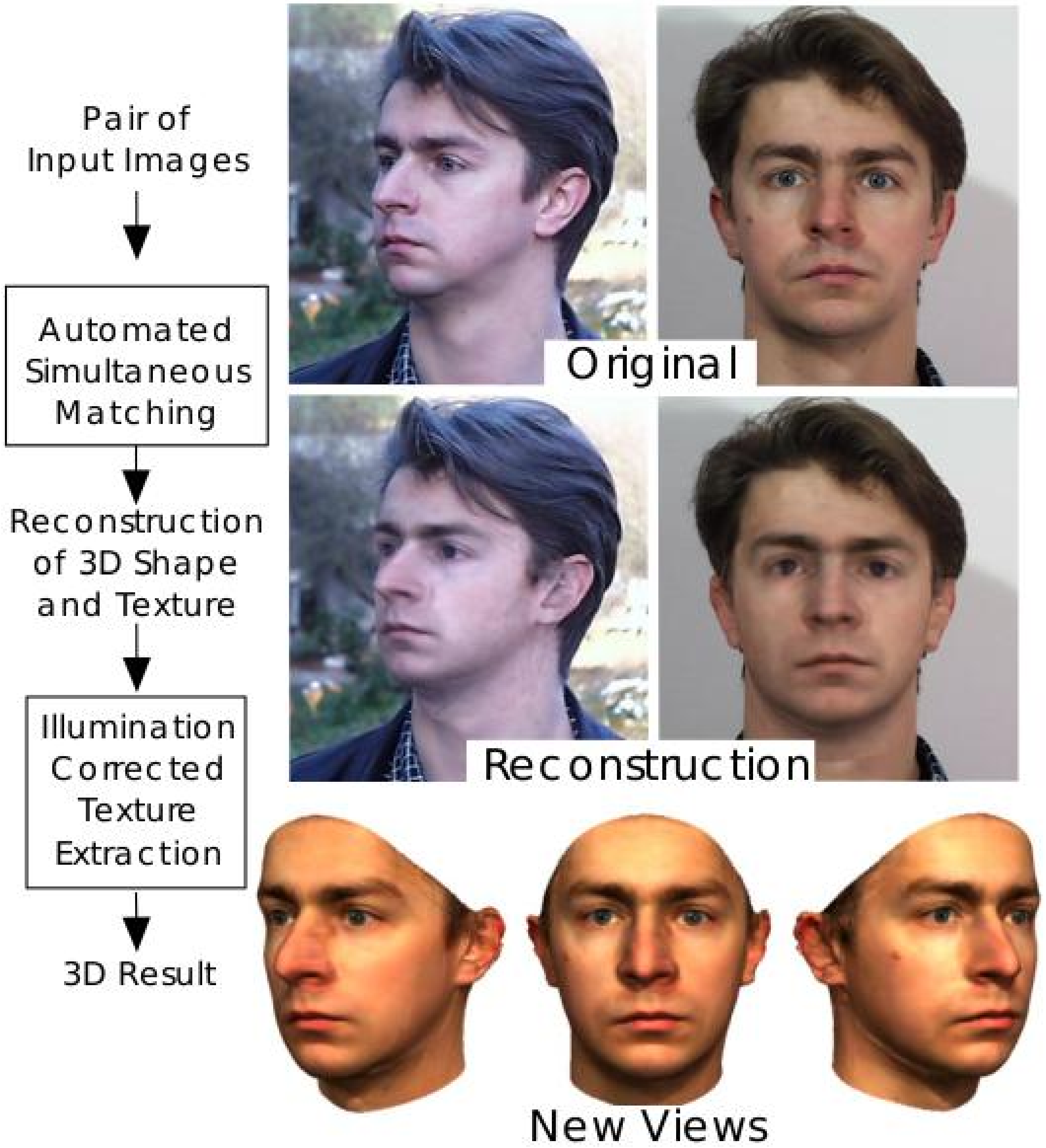
L2 regularization

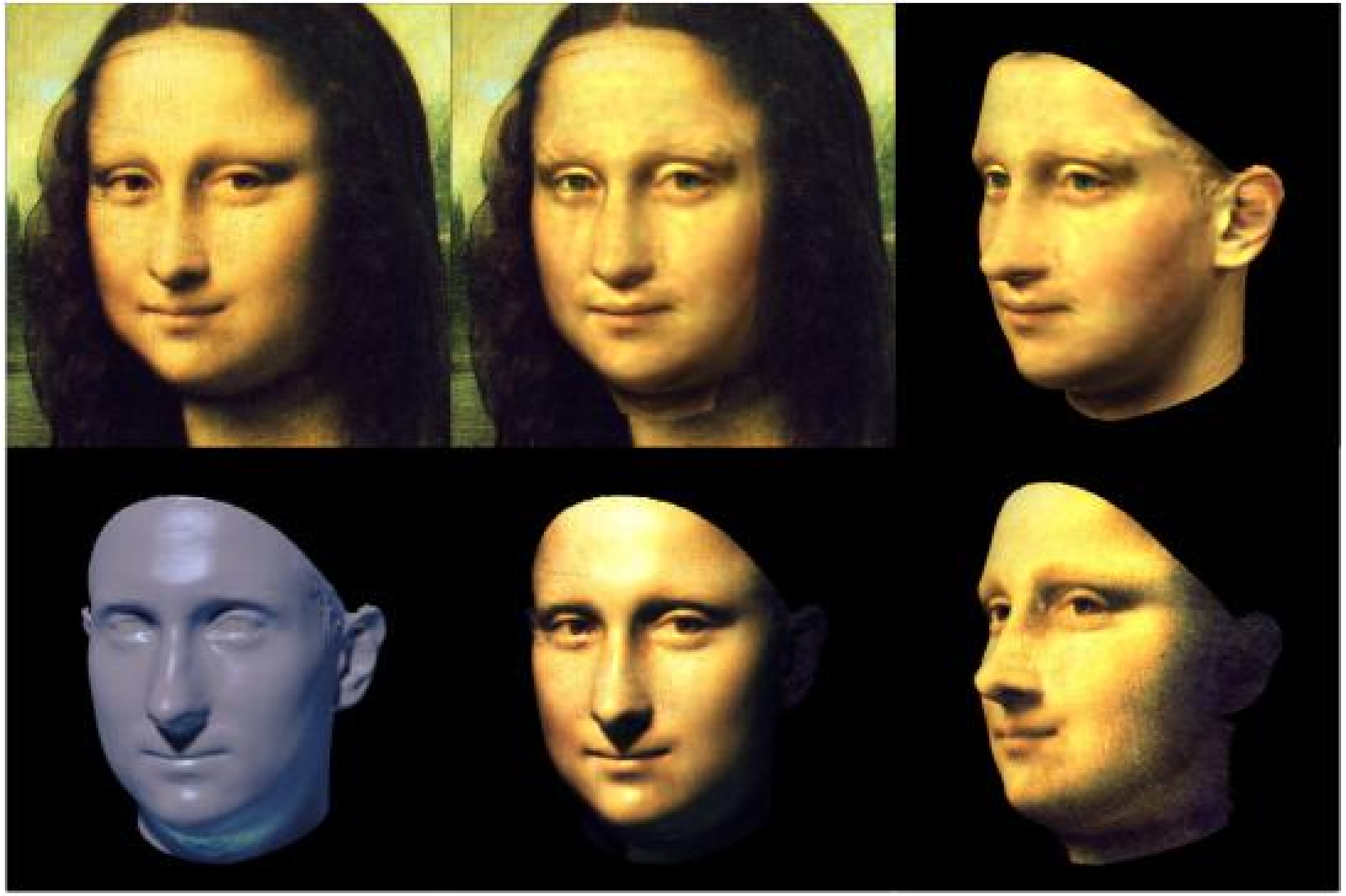
# Coarse-to-fine

1. start iteration with down-sampled input image and low-resolution 3D model.
2. slowly add more ***$\alpha$, $\beta$*** and principle components during optimization
3. start with strong regularization and then slowly reduce it
4. finally, fixed ***$\rho$*** and change to segmented 3D model to improve details
5. illumination-corrected texture extraction: extract raw-image details and overlay.

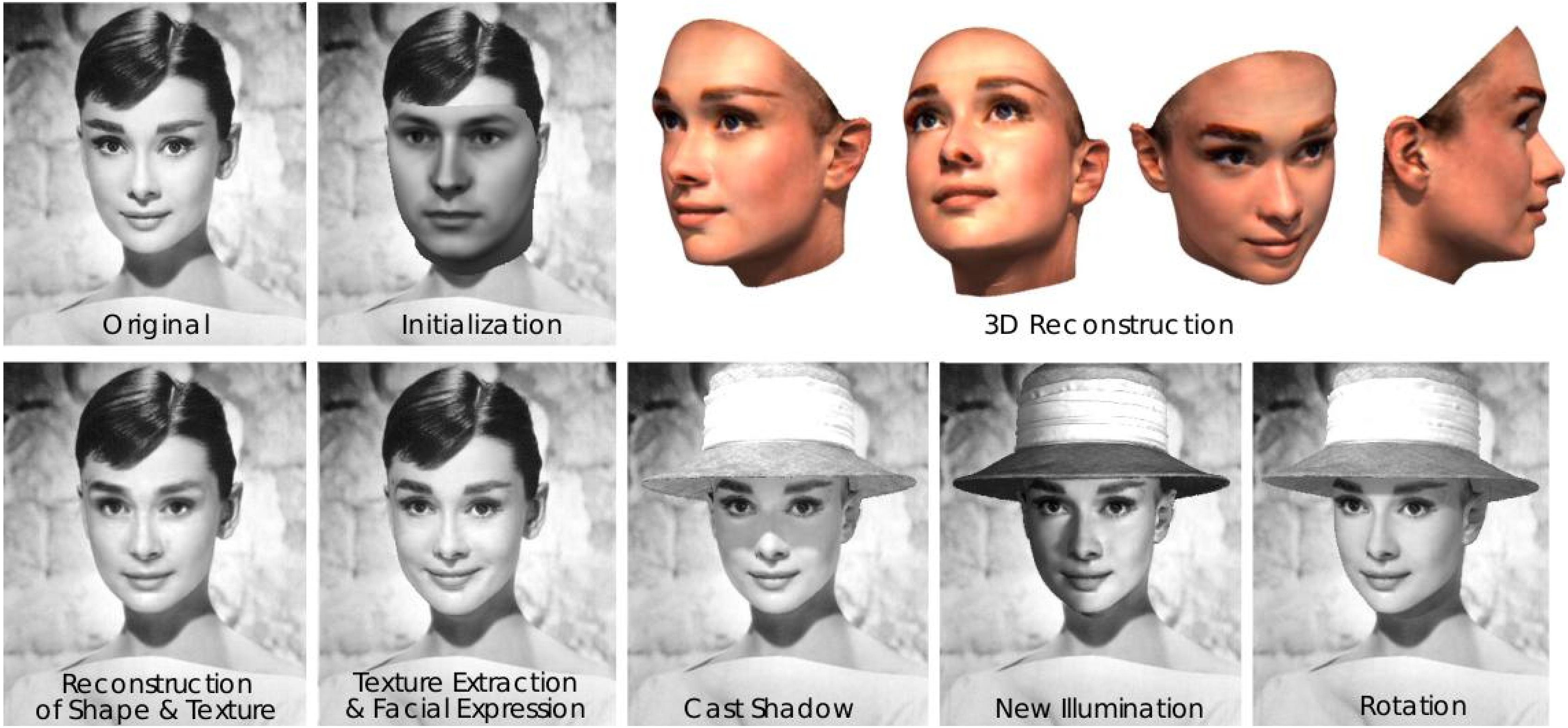
# Examples

two -> one



Performance: *50 min on SGI R1000 processor (1999)*


Original
Initialization
3D Reconstruction
Reconstruction of Shape & Texture
Texture Extraction & Facial Expression
Cast Shadow
New Illumination
Rotation

# Open questions

1. Extending the database:
   - mainly Caucasian of middle age
   - facial expressions, visemes, face variation during speech (limited by the slow laser scan ~ 10 sec/scan)

2. Extending the model:
   - include hair