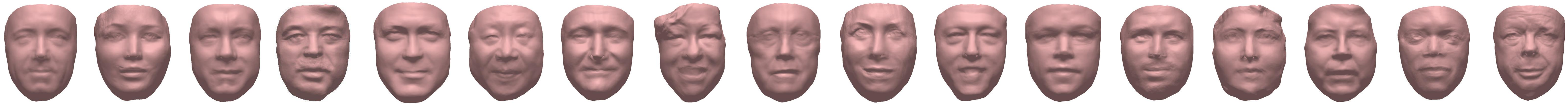
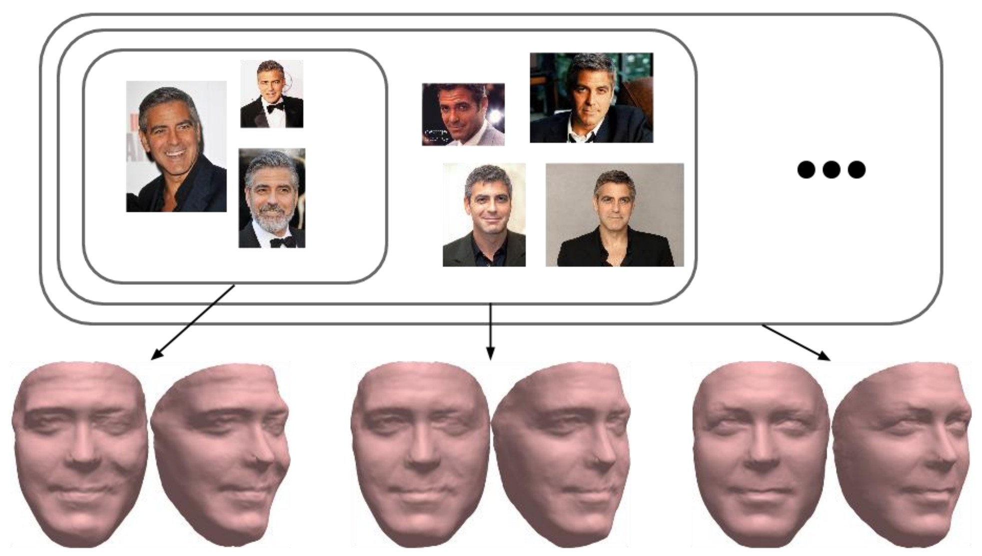
Adaptive 3D Face Reconstruction from Unconstrained Photo Collections

Joseph Roth, Yiying Tong, and Xiaoming Liu
Department of Computer Science and Engineering, Michigan State University



Problem Statement

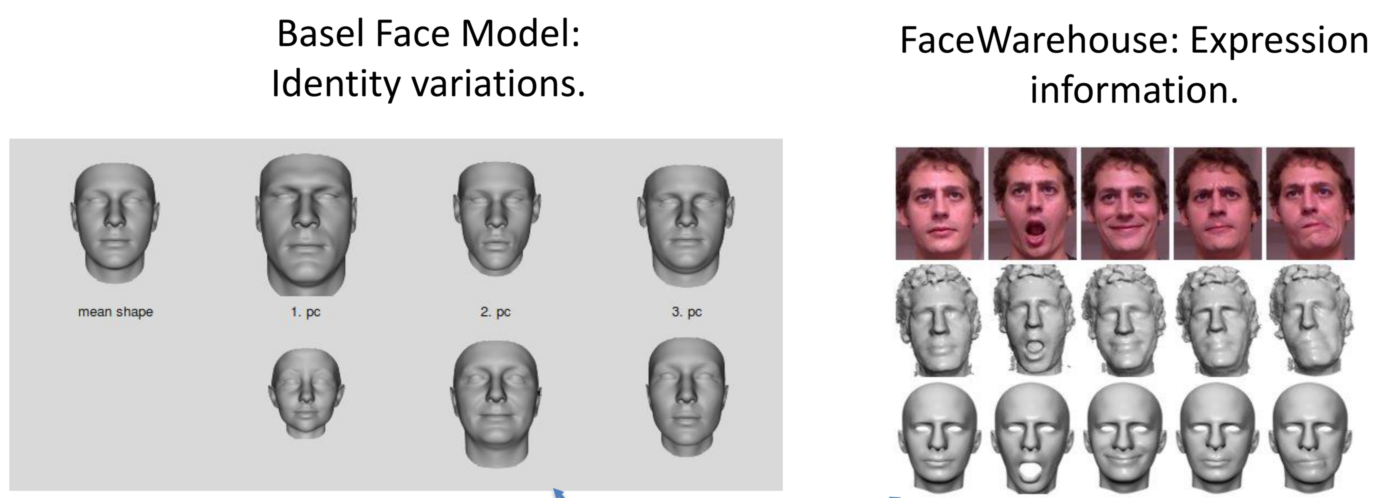


Reconstruct a detailed 3D face model from a photo collection of images with unknown pose, expression, and illumination.
The reconstruction adapts to the number and quality of images.

Applications

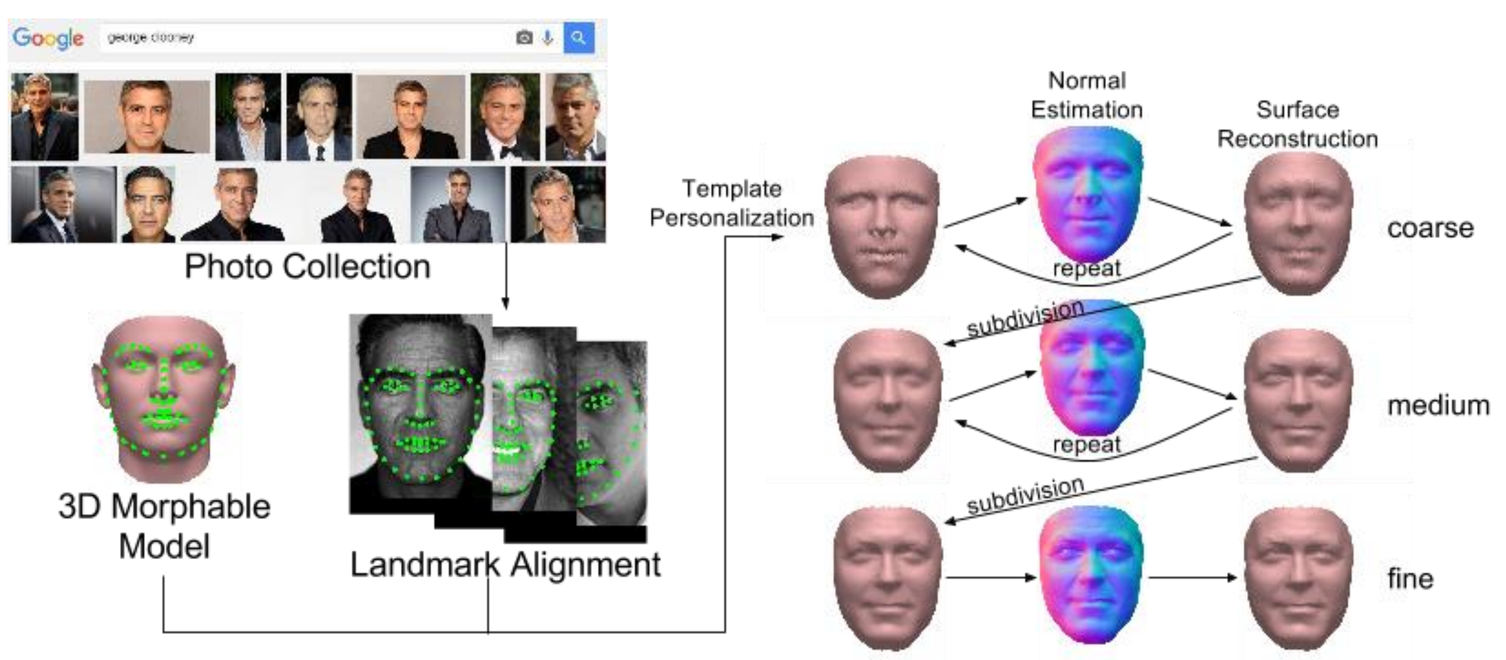
- 3D-assisted face recognition (Blanz & Vetter '03, Hu *et al.* '04).
- Facial animation (Cao *et al.* '14).
- 3D expression recognition (Wang *et al.* '06).
- Consumer entertainment, e.g., personalized bobbleheads.

3D Morphable Model



$$\bar{\mathbf{X}}+\sum_{k=1}^{199}\mathbf{X}_k^{\mathrm{id}}\alpha_k^{\mathrm{id}}+\sum_{k=1}^{29}\mathbf{X}_{ki}^{\mathrm{exp}}\alpha_k^{\mathrm{exp}}$$

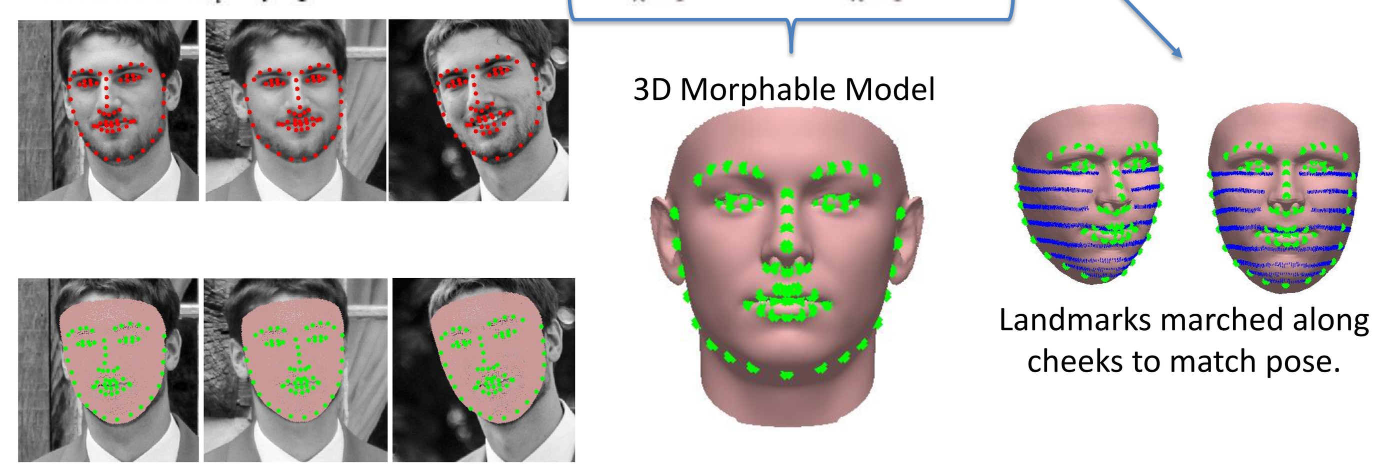
Approach Overview



Template Personalization

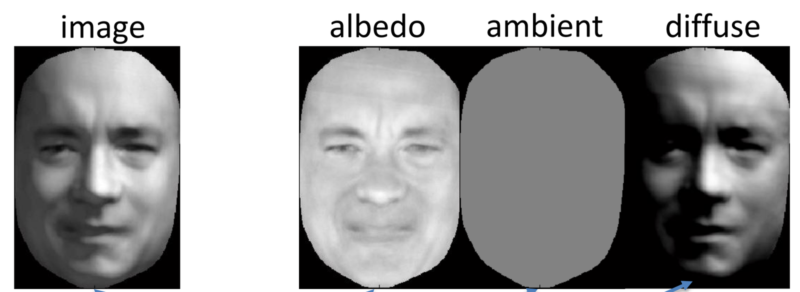
Fit 3D Morphable model jointly to images based on 2D landmark projection error.

$$\underset{s_i,\mathbf{R}_i,\mathbf{t}_i,\alpha^{\mathrm{id}},\alpha_i^{\mathrm{exp}}}{\arg\min}\sum_{i=1}^{n}\frac{1}{n}\|\mathbf{W}_i-(\underbrace{s_i}_{\text{scale}}\underbrace{\mathbf{R}_i}_{\text{rotation}}[\underbrace{\bar{\mathbf{X}}+\sum_{k=1}^{199}\mathbf{X}_k^{\mathrm{id}}\alpha_k^{\mathrm{id}}+\sum_{k=1}^{29}\mathbf{X}_{ki}^{\mathrm{exp}}\alpha_k^{\mathrm{exp}}}_{\text{3D Morphable Model}}]_{\mathrm{land}_i}+\underbrace{\mathbf{t}_i}_{\text{translation}})\|_F^2$$



Photometric Normal Estimation

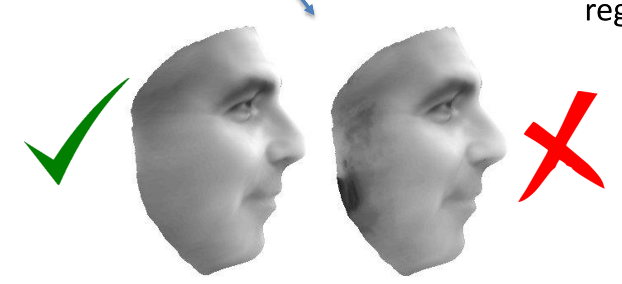
Lambertian reflectance model. Intensity in image is a linear combination of the surface normals weighted by the lighting.



$$\mathbf{I}=\rho(k_a+k_d(l^x,l^y,l^z)^{\intercal}\mathbf{n})$$

Lighting and Albedo Estimation
Project template onto each face to find vertex correspondence across all images. Some parts of the face may be obscured in a given image.

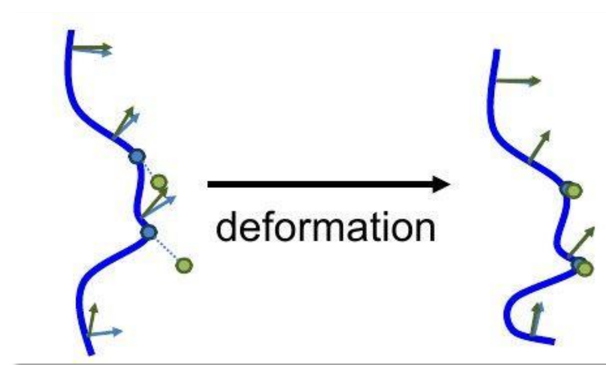
$$\underset{\rho_j,\mathbf{l}_i,\mathbf{n}_j}{\arg\min}\sum_{i=1}^{n}\sum_{j=1}^{p}\|f_{ij}\underbrace{d_{ij}}_{\text{dependability}}-\rho_j\mathbf{l}_i^{\intercal}\mathbf{n}_jd_{ij}\|^2+\underbrace{\lambda_n\|\mathbf{n}_j-\mathbf{n}_j^t\|^2}_{\text{Adaptive template regularization}}$$



Surface Reconstruction

Surface Reconstruction

- Deform the surface to better match the landmark constraints and the surface normal constraints.
- Additional boundary constraint to maintain consistency.



$$\mathbf{X}^{k+1}=\arg\min_{\mathbf{X}}\underbrace{\lambda_l\sum_{i=1}^{n}\|s_i\mathbf{R}_i[\mathbf{X}]_{\mathrm{land}_i}+t_i-\mathbf{W}\|_F^2}_{\text{Landmarks}}+\underbrace{\lambda_n\|\mathbf{X}\mathcal{L}-\mathbf{NH}\|_F^2}_{\text{Surface Normals}}+\underbrace{\lambda_b\|\mathbf{X}\mathcal{L}_b-\mathbf{X}^k\mathcal{L}_b\|_F^2}_{\text{Boundary}}$$

Results

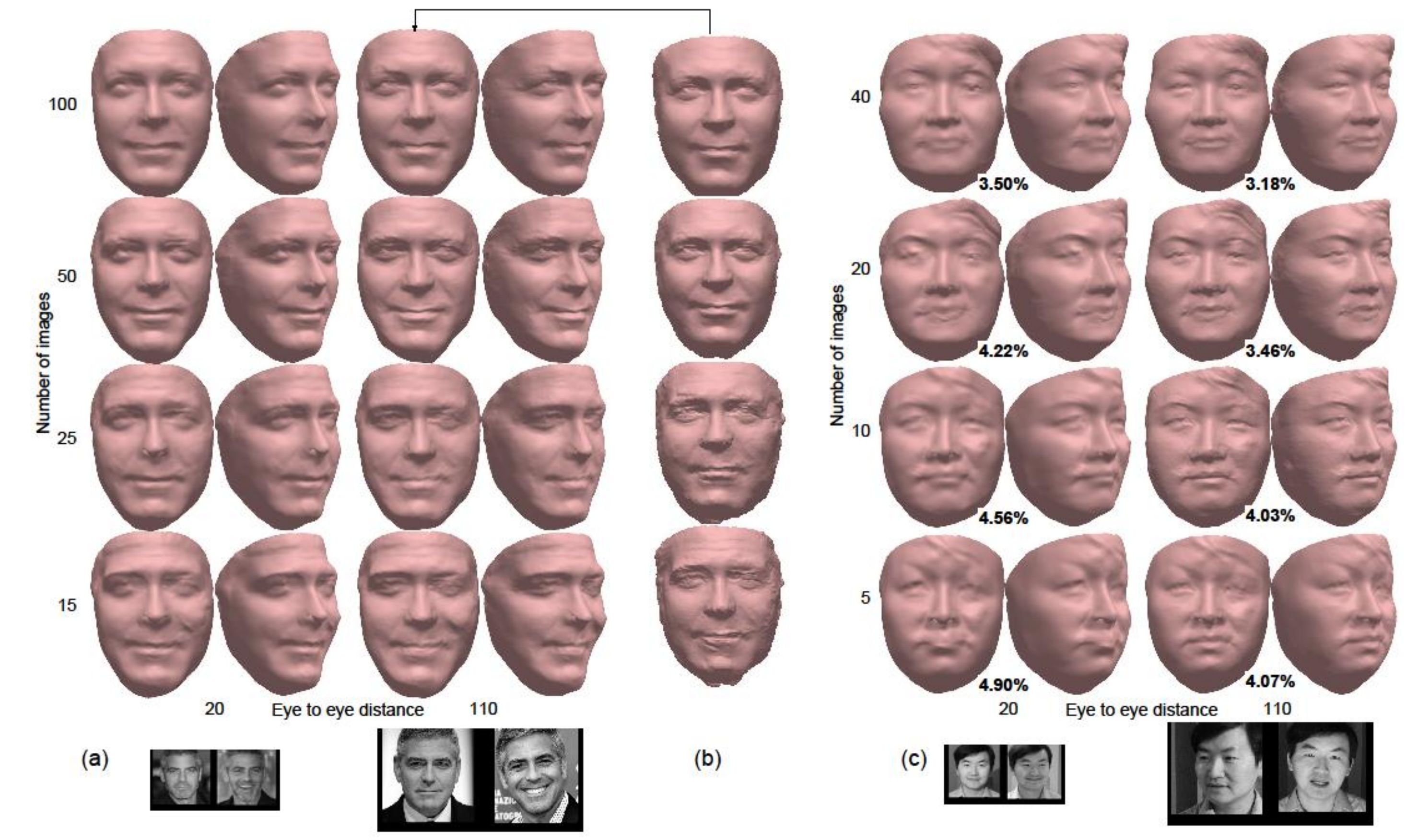


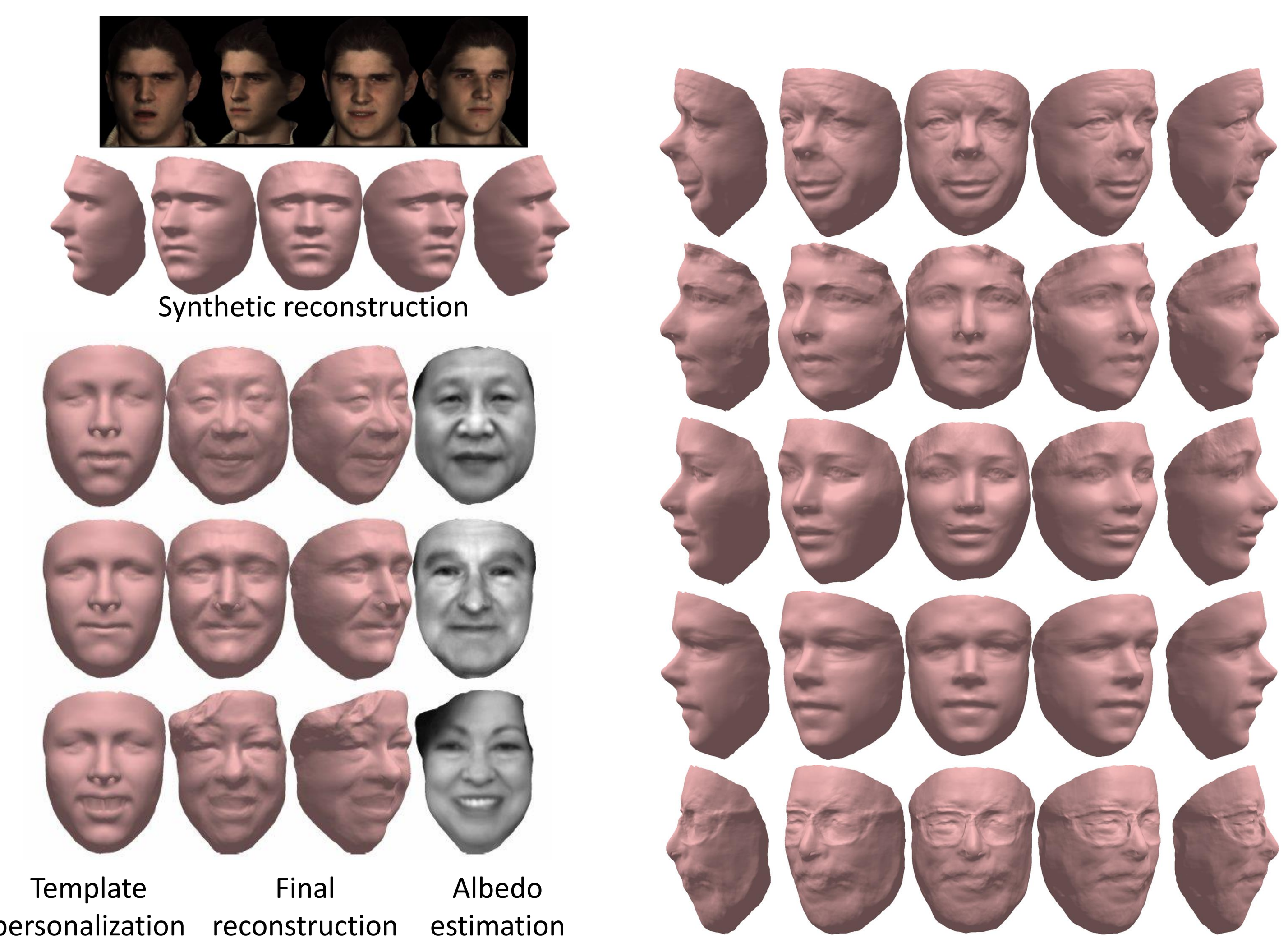
Table: Error comparison for synthetic data.

Method	Neutral	30 Yaw	Expression
Ours	3.22%	3.82%	4.40%
Roth et al. 2015	6.13%	7.48%	6.59%

Table: Error comparison of personal collection.

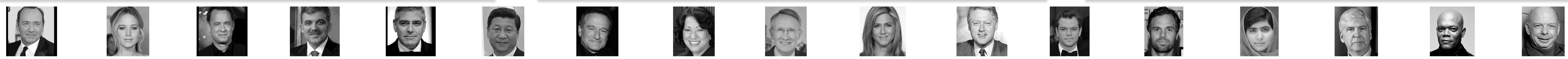
# Images	1	5	10	20	40
Ours	4.19%	4.07%	4.03%	3.46%	3.13%
Roth et al. 2015	-	8.77%	5.40%	4.73%	4.13%

Error is the surface to surface distance, defined as the mean closest distance from each vertex to the other surface. Expressed as a percentage of the interpupillary distance.



Conclusions

- 3D Morphable Model fit jointly across entire collection.
- Adaptable regularization in joint Lambertian image rendering formulation allows it to work photo collections of *any size* and diverse ethnicities and gender.
- Coarse to fine scheme improves alignment as well as efficiency.



Kevin Spacey | Jennifer Lawrence | Tom Hanks | Abdullah Gul | George Clooney | Jinping Xi | Robin Williams | Sandra Sotomayor | Harry Reid | Jennifer Aniston | Bill Clinton | Matt Damon | Mark Ruffalo | Malala Yousafzai | Rick Snyder | Samuel Jackson | Wallace Shawn