Making faces

Digital Visual Effects, Spring 2005

Yung-Yu Chuang

2005/6/8

with slides by Richard Szeliski, Steve Seitz and Alex Efros

Announcements

- Project #3 artifacts voting

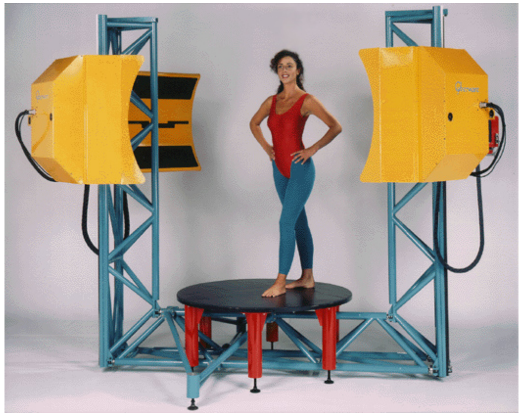
Outline

- 3D acquisition for faces
- Statistical methods
- Face models from single images
- Image-based faces
- Relighting for faces

3D acquisition for faces

Cyberware scanners

face & head scanner



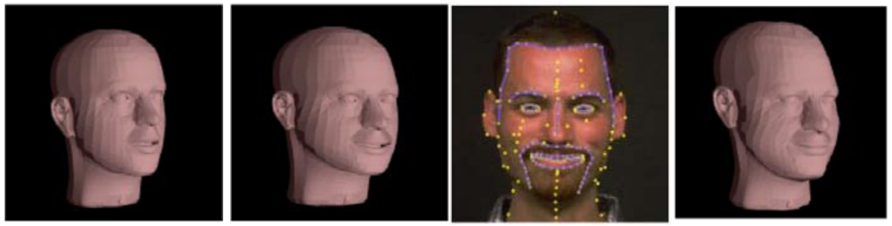
whole body scanner

Making facial expressions from photos

- Similar to Façade, use a generic face model and view-dependent texture mapping
- Procedure
  1. Take multiple photographs of a person
  2. Establish corresponding feature points
  3. Recover 3D points and camera parameters
  4. Deform generic face model to fit points
  5. Extract textures from photos

Reconstruct a 3D model

input photographs

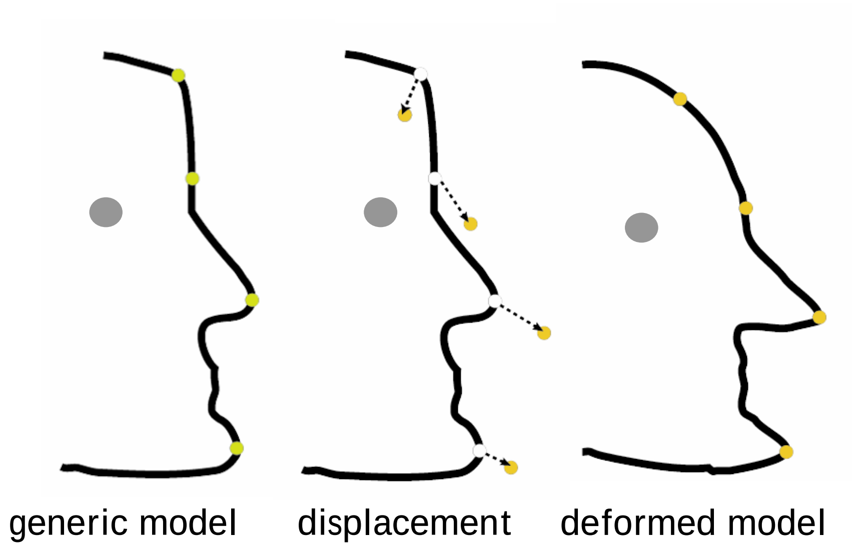


generic 3D face model | pose estimation | more features | deformed model

Mesh deformation

- Involves two steps:
  – Compute displacement of feature points
  – Apply scattered data interpolation

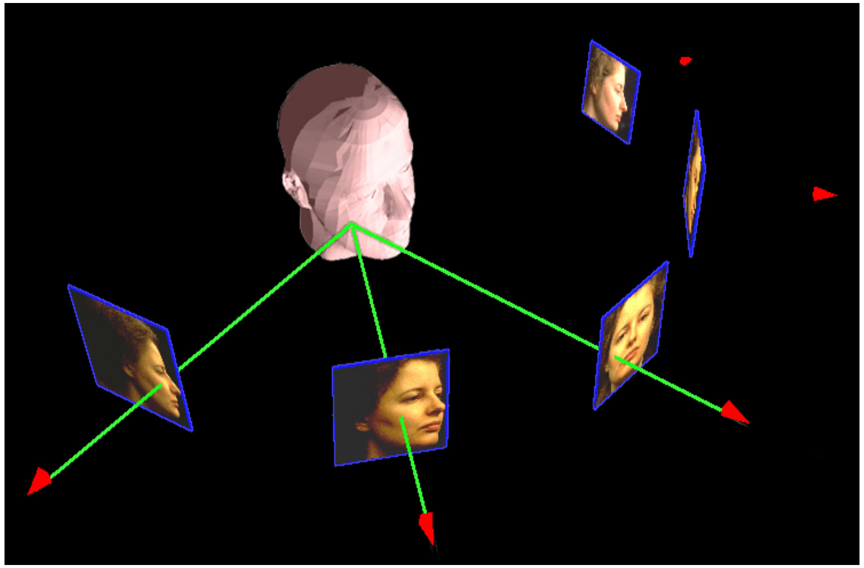
Mesh deformation



Texture extraction

- The color at each point is a weighted combination of the colors in the photos
- Texture can be:
  - View-independent
  - View-dependent
- Considerations for weighting
  - Occlusion
  - Smoothness
  - Positional certainty
  - View similarity

Texture extraction



Texture extraction

Texture extraction

view-independent view-dependent

Model reconstruction

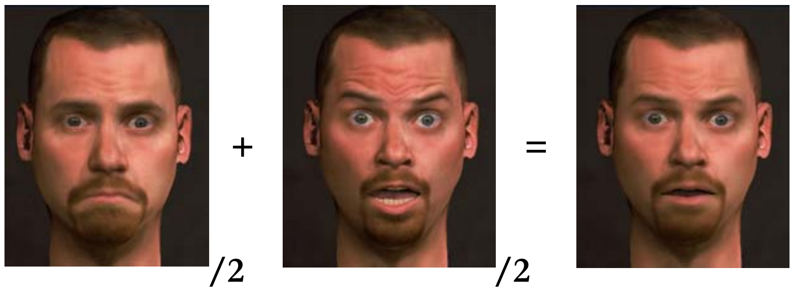
Use images to adapt a generic face model.

Creating new expressions

- In addition to global blending we can use:
  - Regional blending
  - Painterly interface

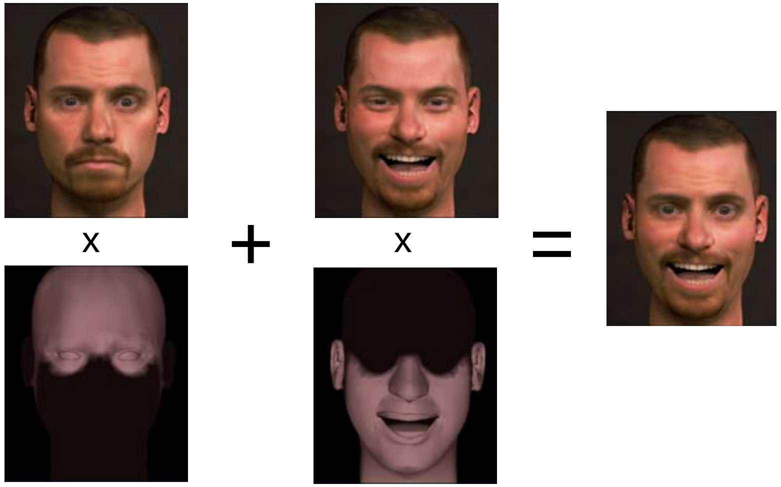
Creating new expressions

New expressions are created with 3D morphing:



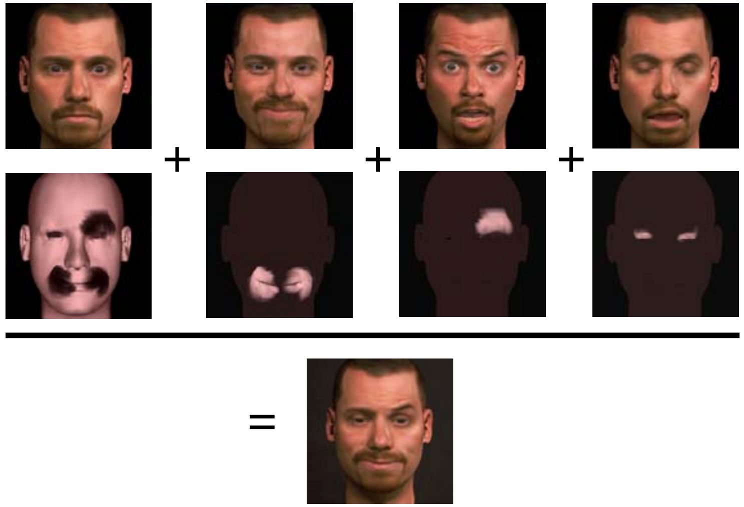
Applying a global blend

Creating new expressions



Applying a region-based blend

Creating new expressions

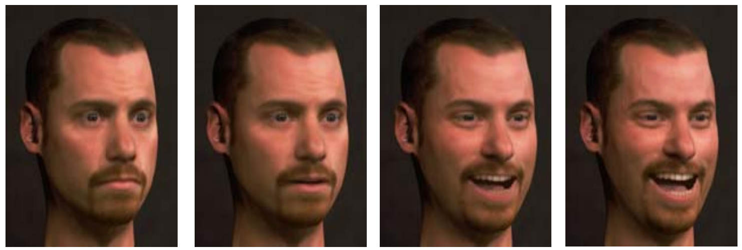


Using a painterly interface

Drunken smile

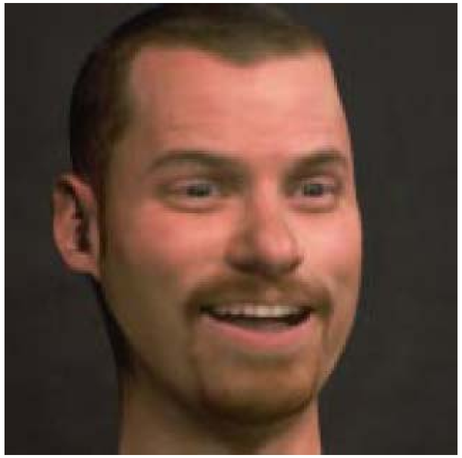
Animating between expressions

Morphing over time creates animation:

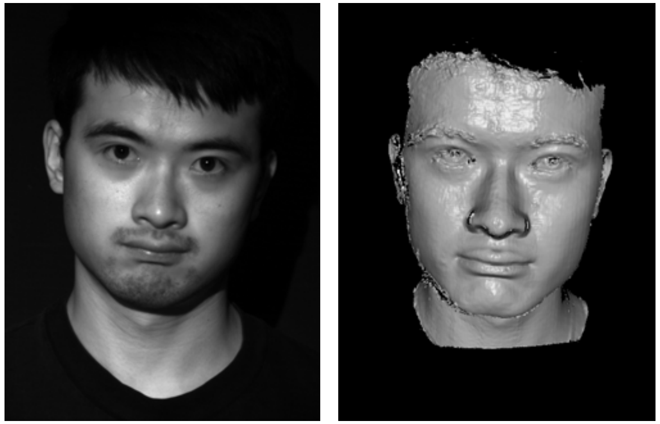


“neutral” → “joy”

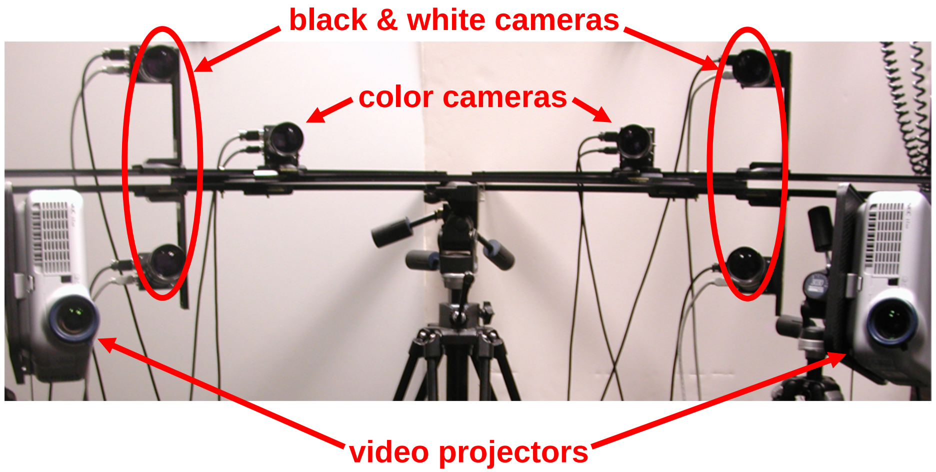
Video

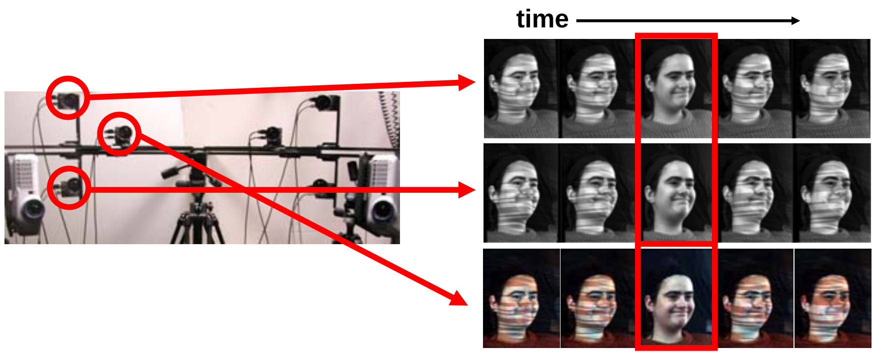


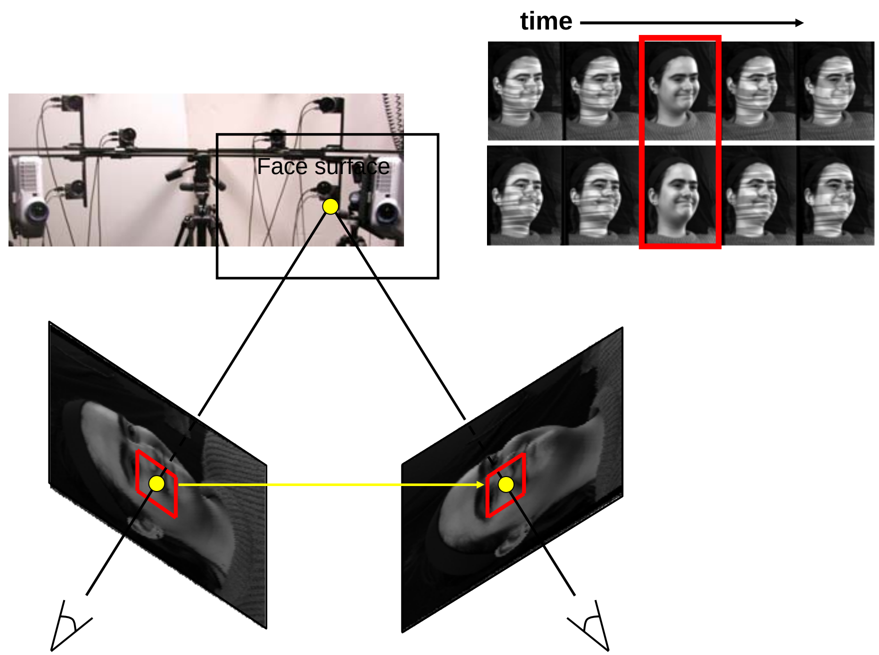
Spacetime faces

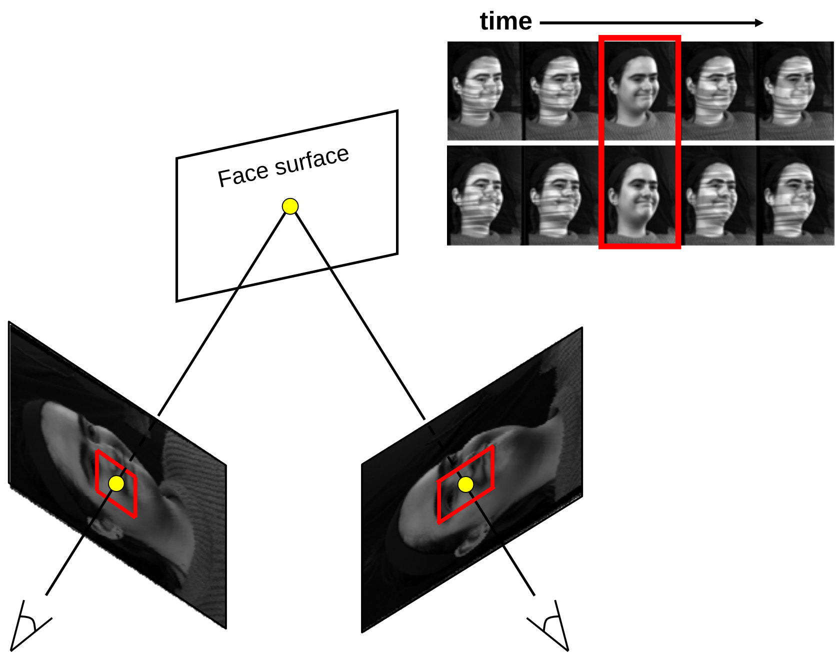


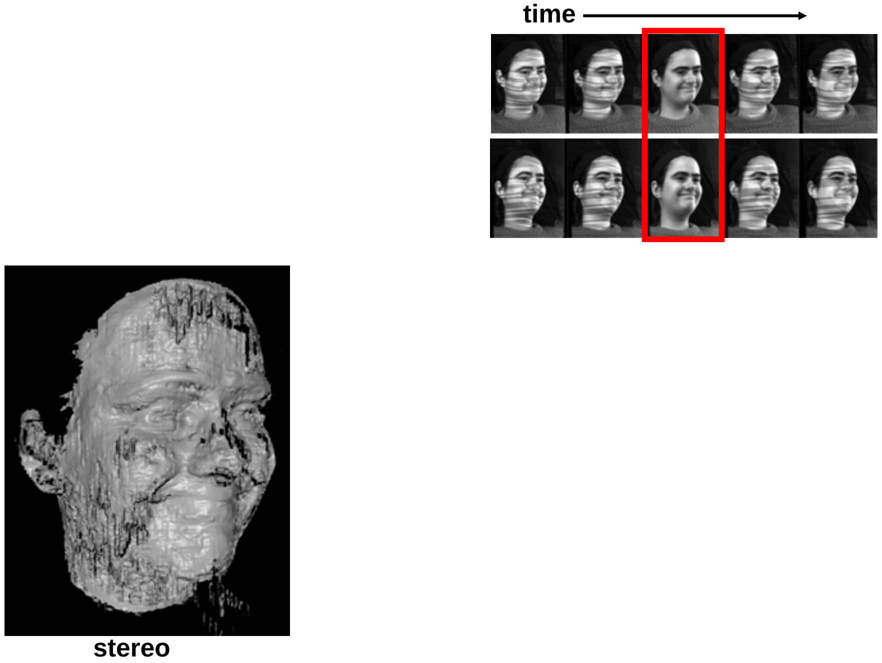
Spacetime faces

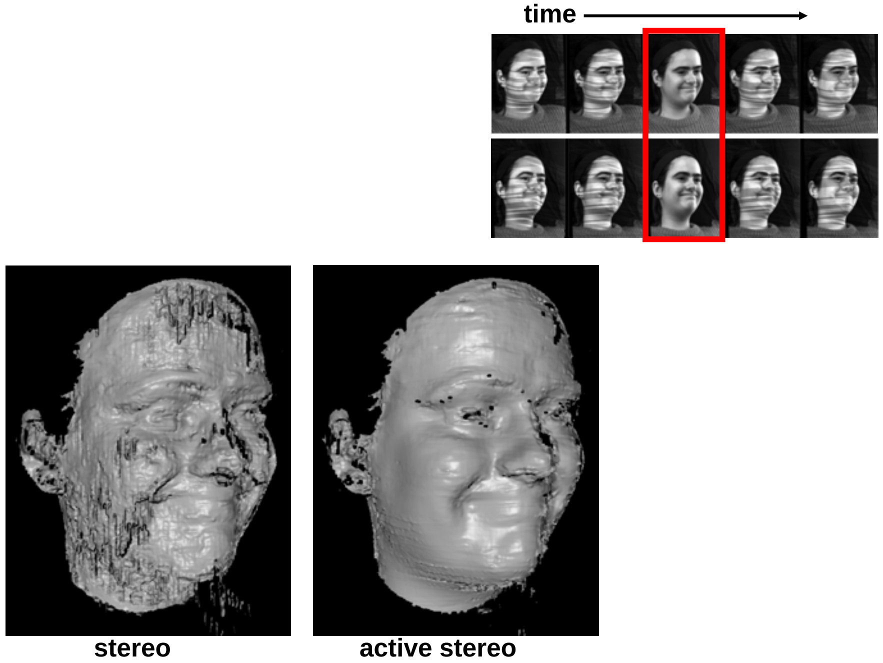


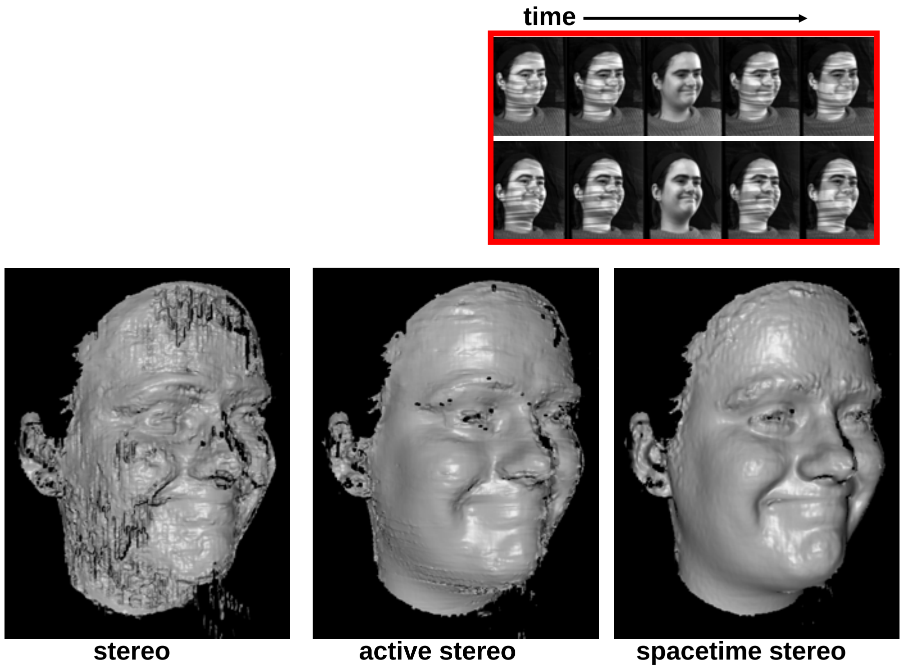



time
Face surface

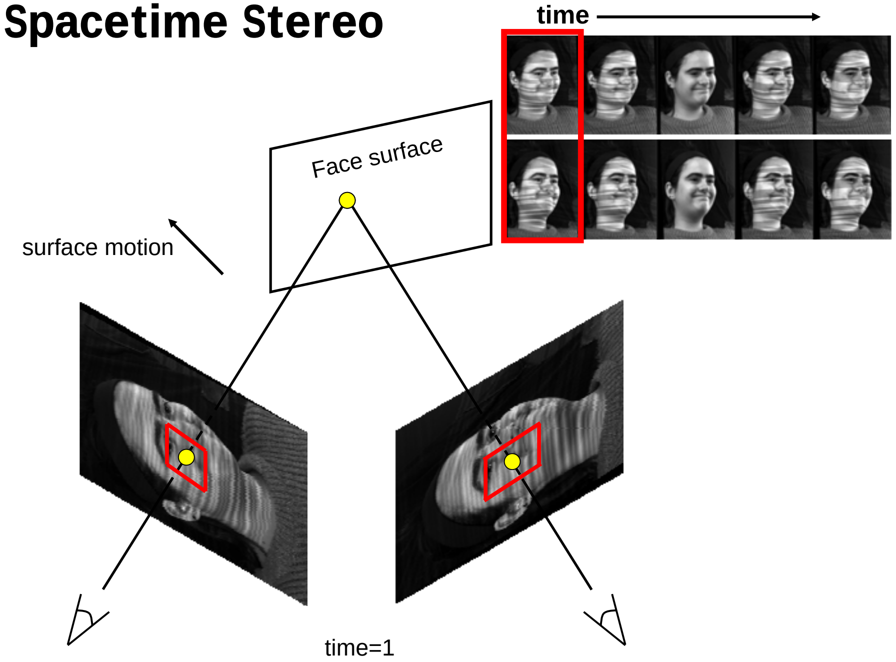

time
Face surface


time
stereo

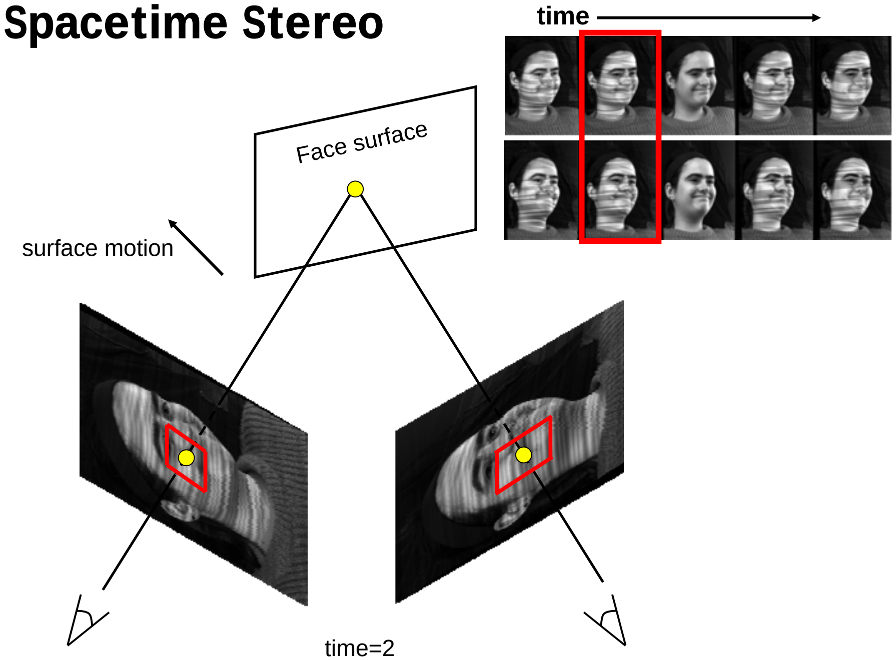

time
stereo
active stereo



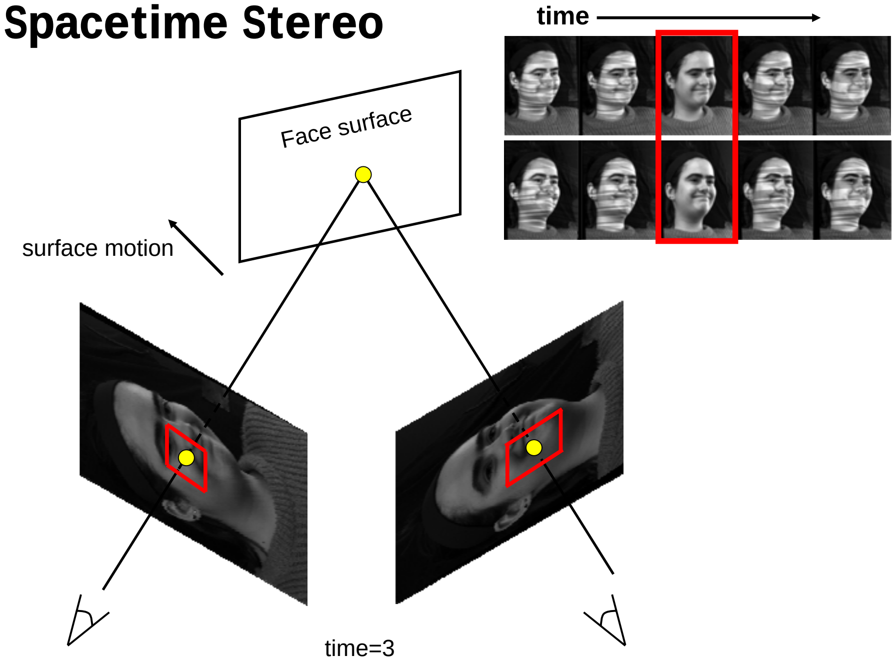
Spacetime Stereo



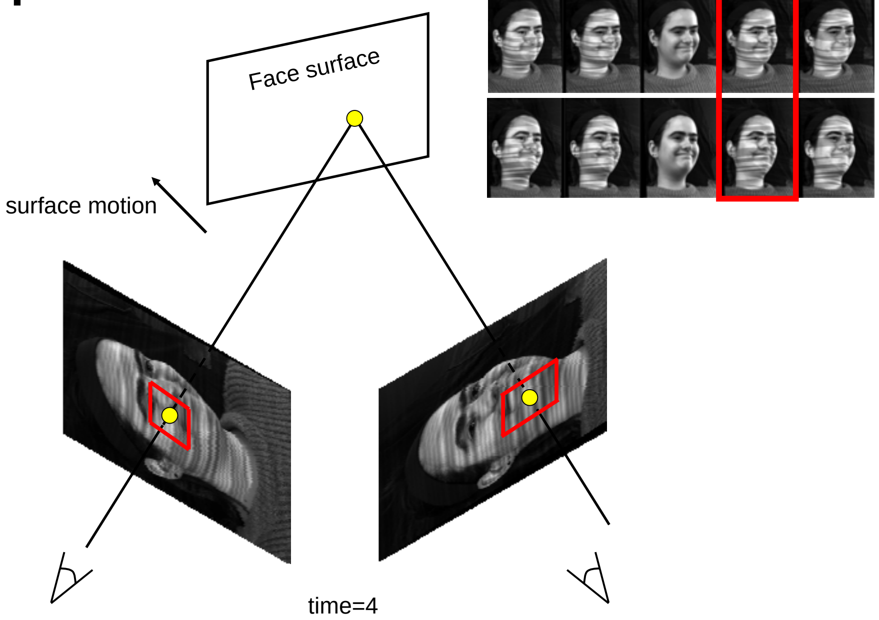
Spacetime Stereo



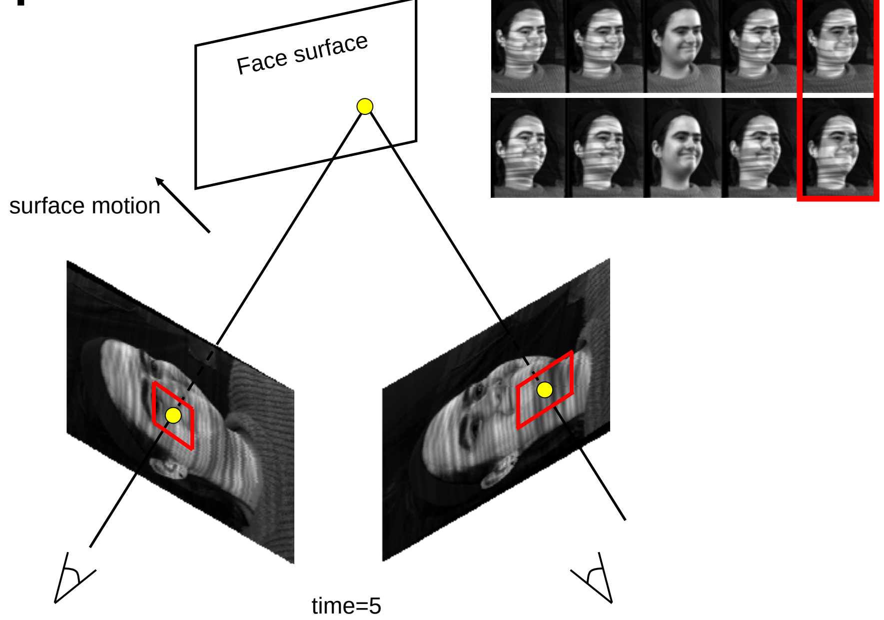
Spacetime Stereo



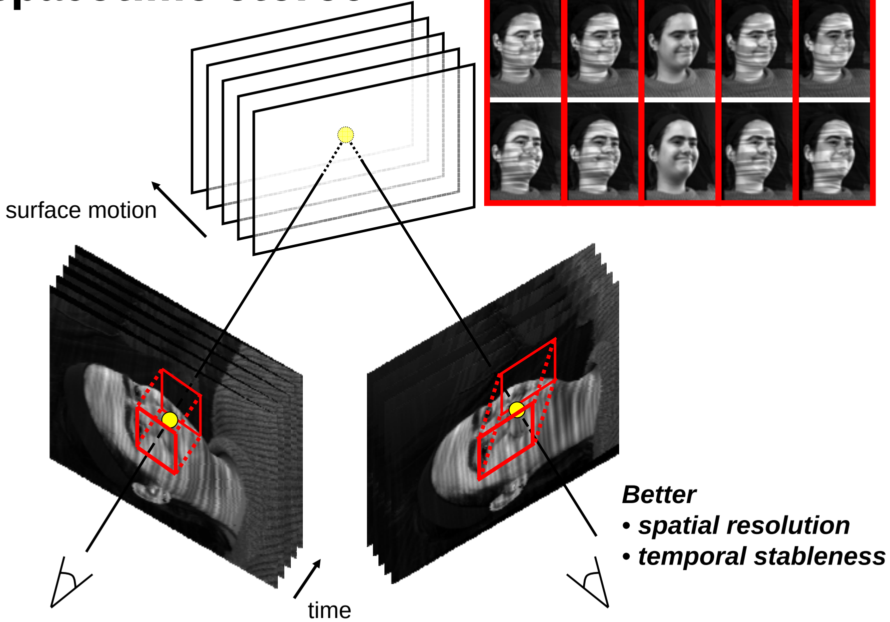
Spacetime Stereo



Spacetime Stereo

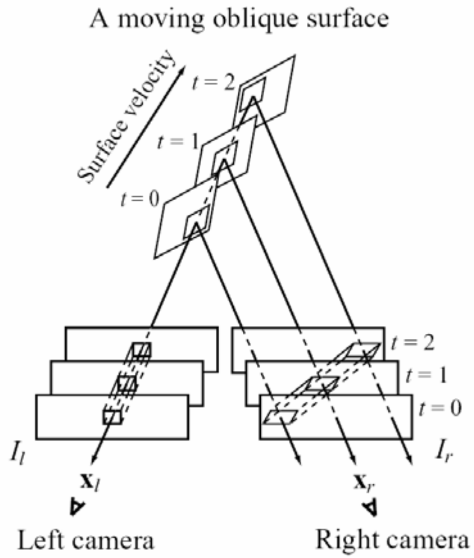


Spacetime Stereo

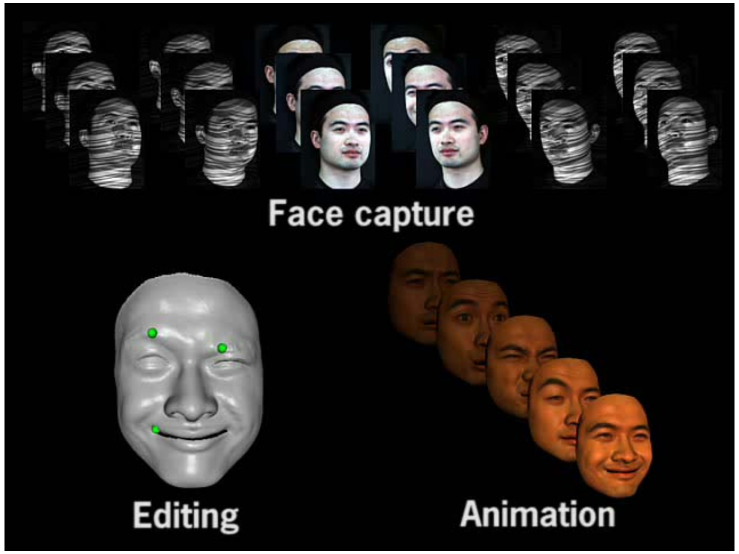


Better
- ***spatial resolution***
- ***temporal stableness***

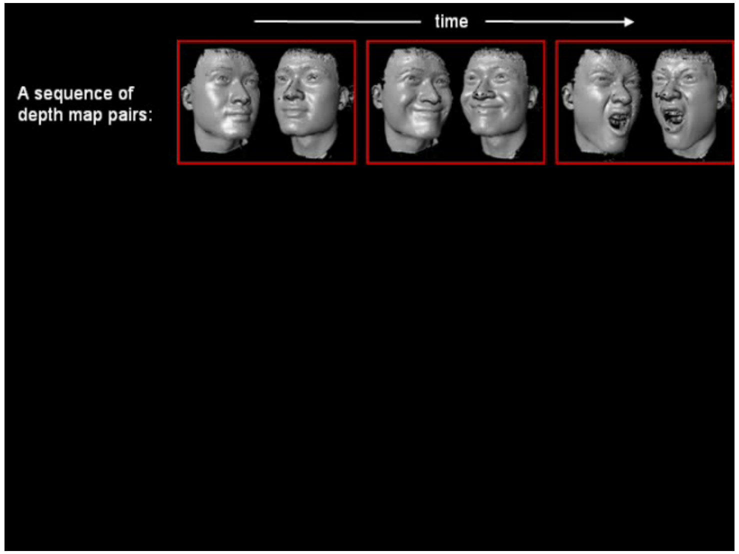
Spacetime stereo matching



Video

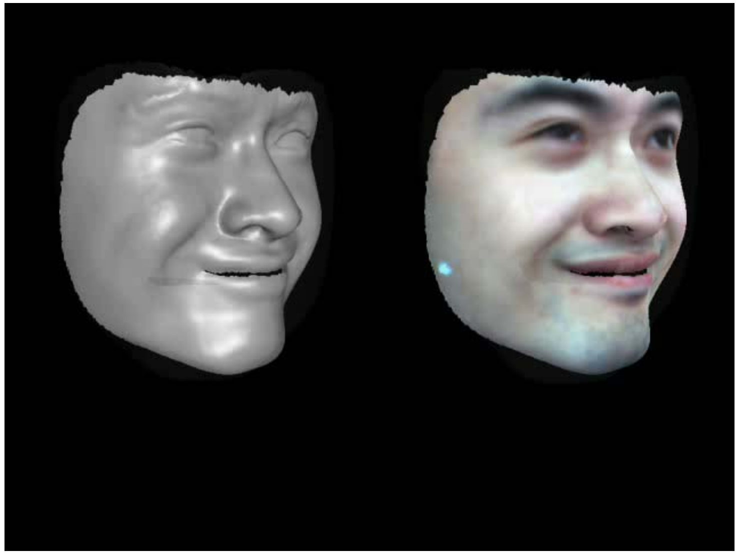


Fitting

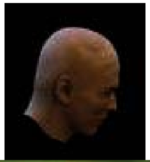


FaceIK

Animation



3D face applications: The one



3D face applications: Gladiator

extra 3M

3D face applications: Spiderman 2

Statistical methods

Statistical methods

para-meters $z \longrightarrow \boxed{f(z)+\varepsilon} \longrightarrow y$ observed signal

$$
\begin{aligned}
z^* &= \max_z P(z \mid y) \\
&= \max_z \frac{P(y \mid z)P(z)}{P(y)} \\
&= \min_z L(y \mid z) + L(z)
\end{aligned}
$$

Example:
super-resolution
de-noising
de-blocking
Inpainting
...

Statistical methods

para-meters $z \longrightarrow \boxed{f(z)+\varepsilon} \longrightarrow y$ observed signal

$$z^* = \min_z L(y \mid z) + L(z)$$

data evidence $\dfrac{\|y - f(z)\|^2}{\sigma^2}$

a-priori knowledge

Statistical methods

There are approximately 10^{240} possible 10×10 gray-level images. Even human being has not seen them all yet. There must be a strong statistical bias.

Takeo Kanade

Approximately $8\mathrm{X}10^{11}$ blocks per day per person.

Generic priors

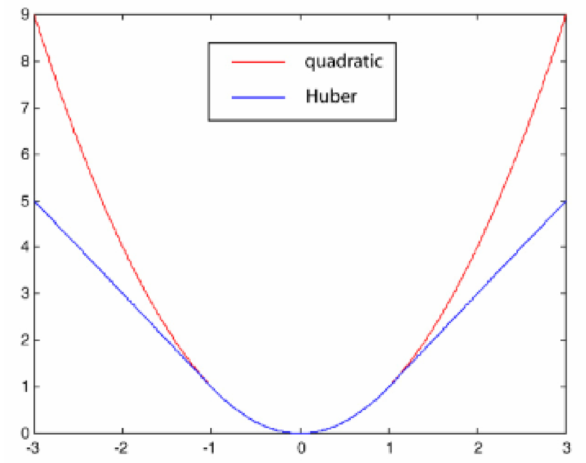
"Smooth images are good images."

$$L(z) = \sum_x \rho(V(x))$$

Gaussian MRF $\rho(d) = d^2$

Huber MRF $\rho(d) = \begin{cases} d^2 & |d| \le T \\ T^2 + 2T(|d| - T) & d > T \end{cases}$

Generic priors

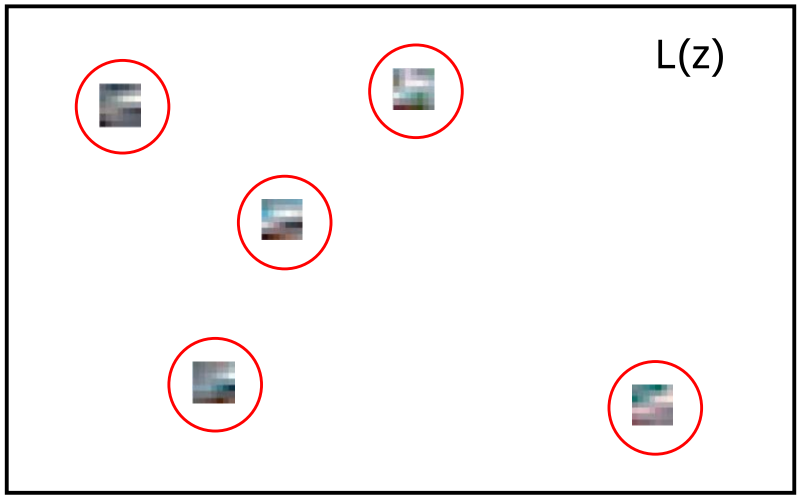


Example-based priors

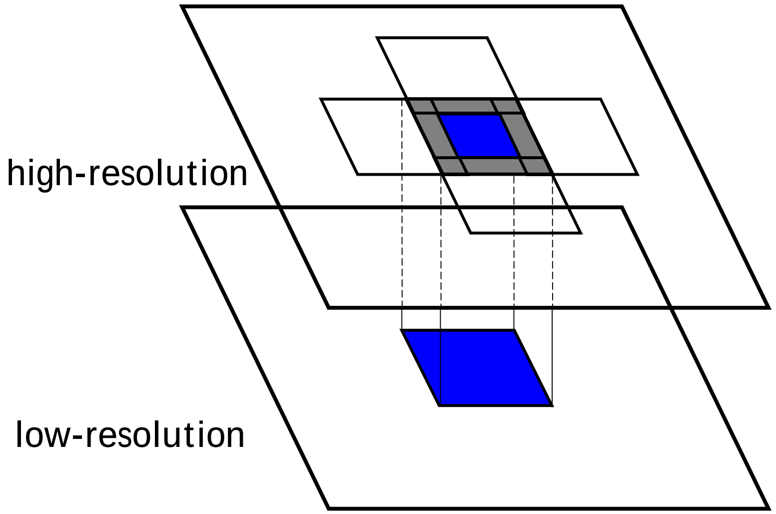
"Existing images are good images."

six 200×200 Images ⇒ 2,000,000 pairs

Example-based priors



Example-based priors



Model-based priors

"Face images are good images when working on face images …"

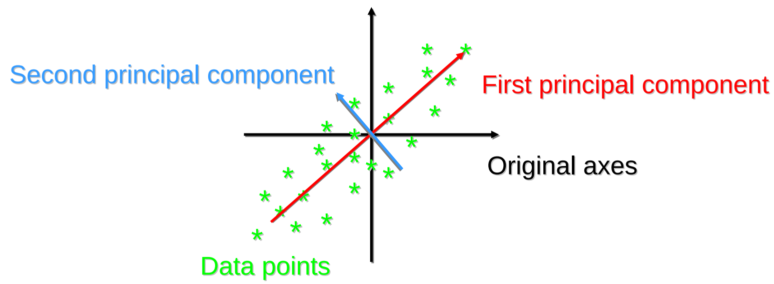
Parametric model $\quad Z=WX+\mu \quad L(X)$

$$z^* = \min_z L(y \mid z) + L(z)$$

$$\begin{cases} X^* = \min_x L(y \mid WX + \mu) + L(X) \\ z^* = WX^* + \mu \end{cases}$$

PCA

- Principal Components Analysis (PCA): approximating a high-dimensional data set with a lower-dimensional subspace



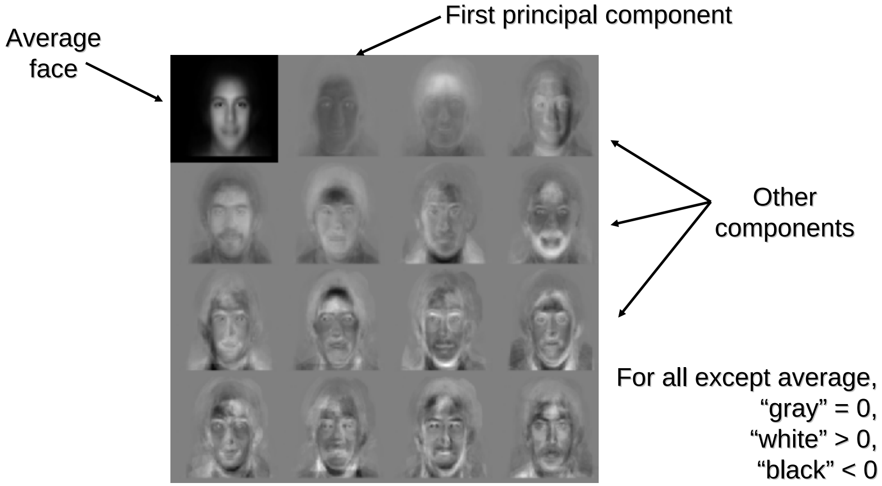
PCA

- Given n k-d points
- Calculate the mean
- Calculate the covariance matrix
- SVD (eigen-analysis) on the covariance matrix

SVD

$$\begin{pmatrix} \\ \mathbf{A} \\ \\ \end{pmatrix} = \begin{pmatrix} \\ \mathbf{U} \\ \\ \end{pmatrix} \begin{pmatrix} w_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & w_n \end{pmatrix} \begin{pmatrix} & \mathbf{V} & \end{pmatrix}^{\mathrm{T}}$$

PCA on faces: “eigenfaces”



Model-based priors

“Face images are good images when working on face images …”

Parametric model $\quad Z=WX+\mu \quad L(X)$

$$z^* = \min_z L(y \mid z) + L(z)$$

$$\begin{cases} X^* = \min_x L(y \mid WX + \mu) + L(X) \\ z^* = WX^* + \mu \end{cases}$$

Super-resolution

(a) (b) (c) (d) (e) (f)

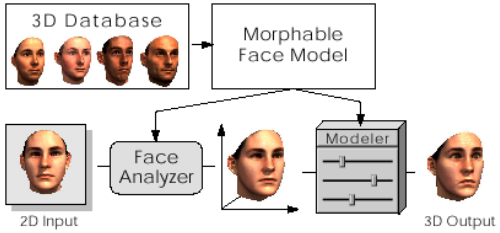
(a) Input low 24×32 (b) Our results (c) Cubic B-Spline

(d) Freeman et al. (e) Baker et al. (f) Original high 96×128

Face models from single images

Morphable model of 3D faces

- Start with a catalogue of 200 aligned 3D Cyberware scans



- Build a model of *average* shape and texture, and principal *variations* using PCA

Morphable model

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

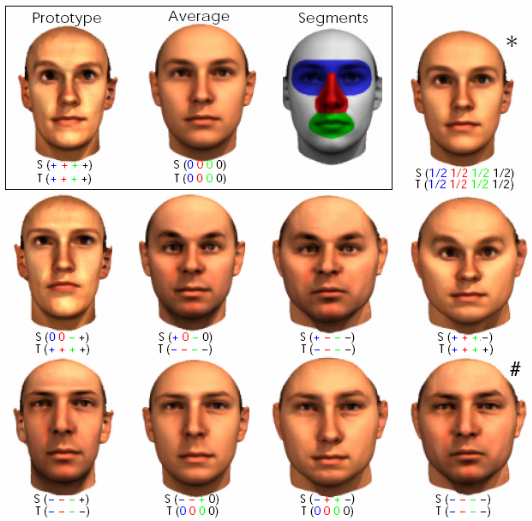
$\vec{\alpha}, \vec{\beta} \in \Re^{m-1}$. The probability for coefficients $\vec{\alpha}$ is given by

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

Morphable model of 3D faces

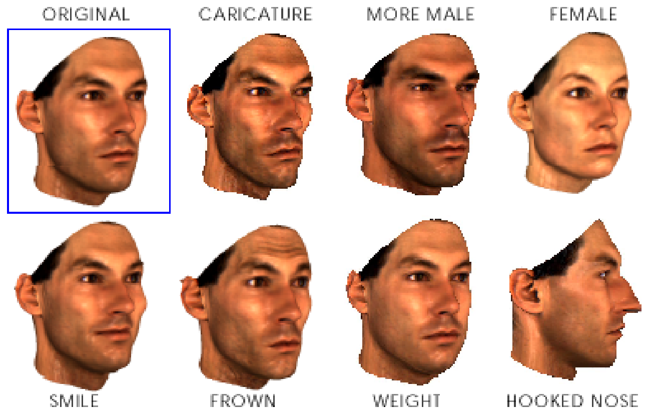
Divide face into 4 regions (eyes, nose, mouth, head)

For each new *prototype*, find amount of deviation from the reference shape and texture.

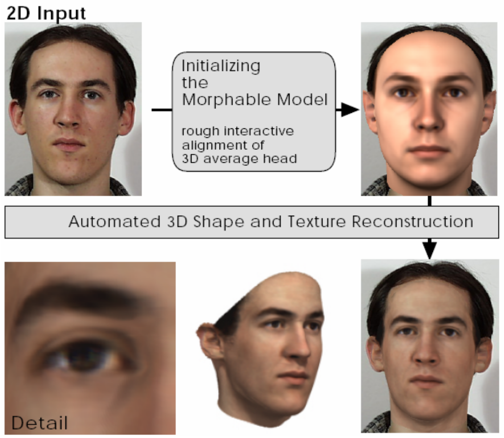


Morphable model of 3D faces

- Adding some variations

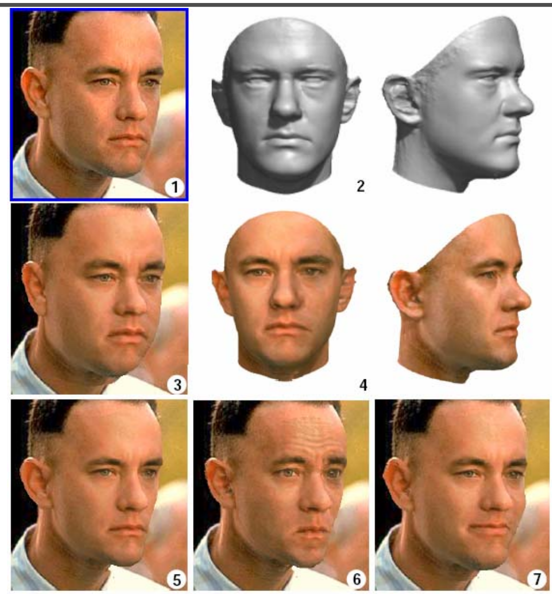


Reconstruction from single image

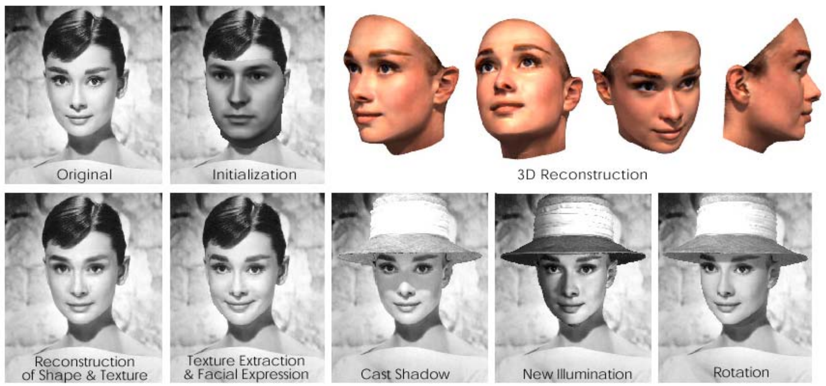


$$E = \frac{1}{\sigma_N^2} E_I + \sum_{j=1}^{m-1} \frac{\alpha_j^2}{\sigma_{S,j}^2} + \sum_{j=1}^{m-1} \frac{\beta_j^2}{\sigma_{T,j}^2} + \sum_j \frac{(\rho_j - \bar{\rho}_j)^2}{\sigma_{\rho,j}^2}$$

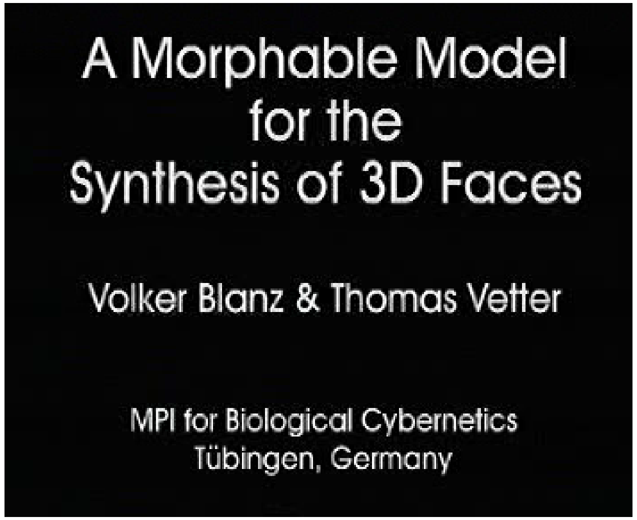
Modifying a single image



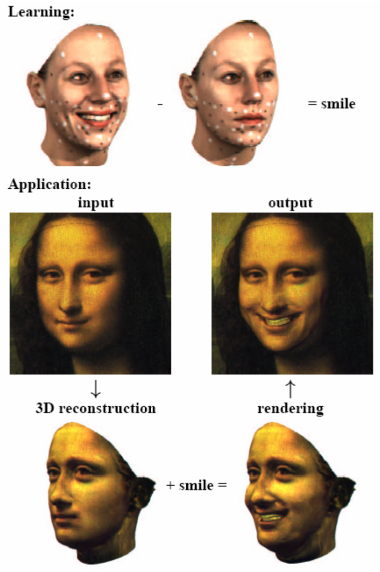
Animating from a single image

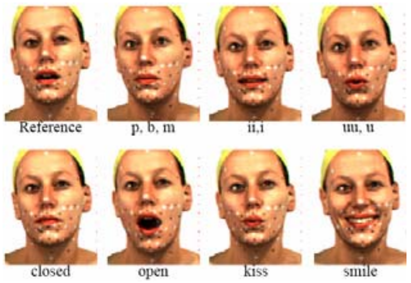


Video



Reanimating faces





35 static scans at different expressions

Videos

exercise

speech

Exchanging faces

Exchanging faces

Exchanging faces

Morphable model for human body

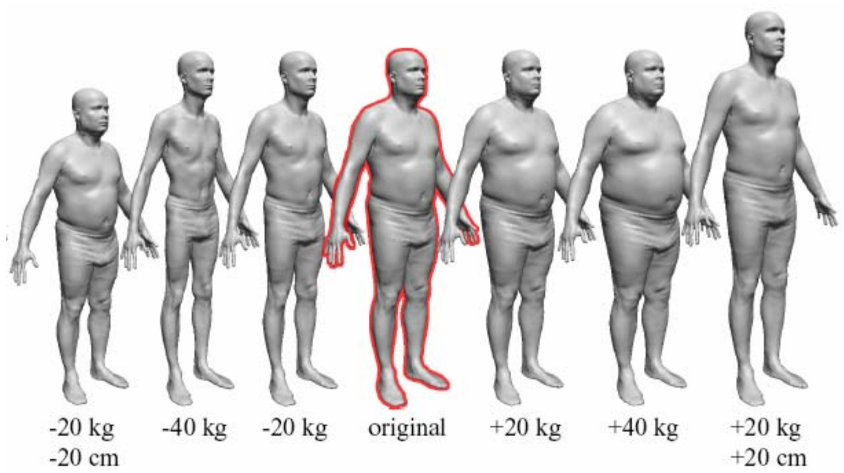
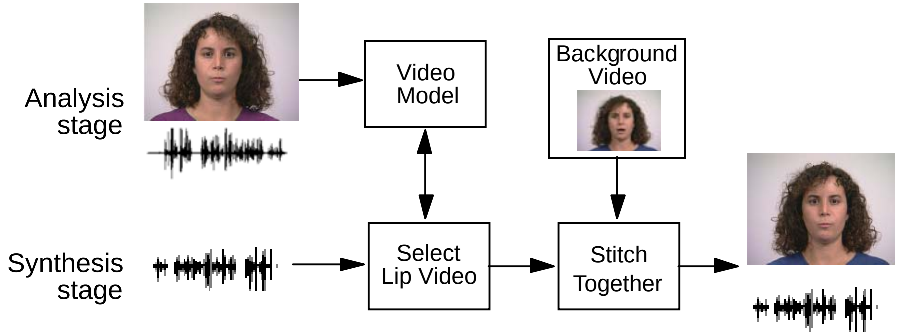


Image-based faces
(lip sync.)

Video rewrite



Results

- Video database
  - 8 minutes of Ellen
  - 2 minutes of JFK
    - Only half usable
    - Head rotation

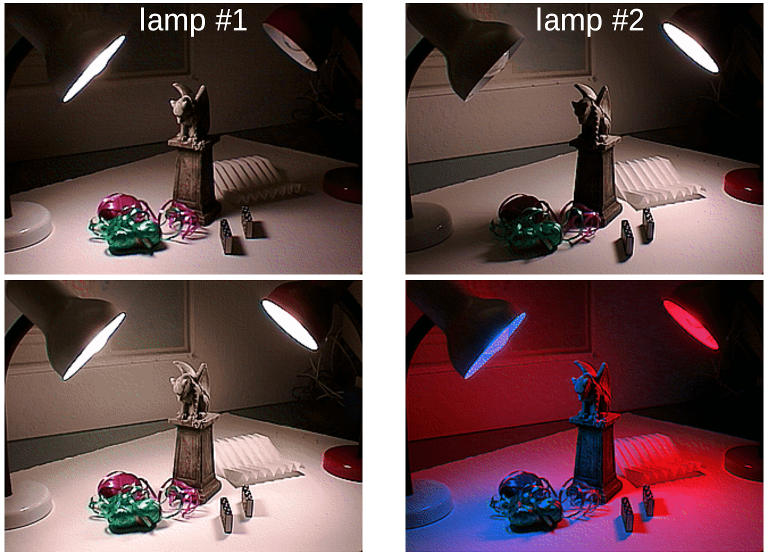
training video

Read my lips.

I never met Forest Gump.

Relighting faces

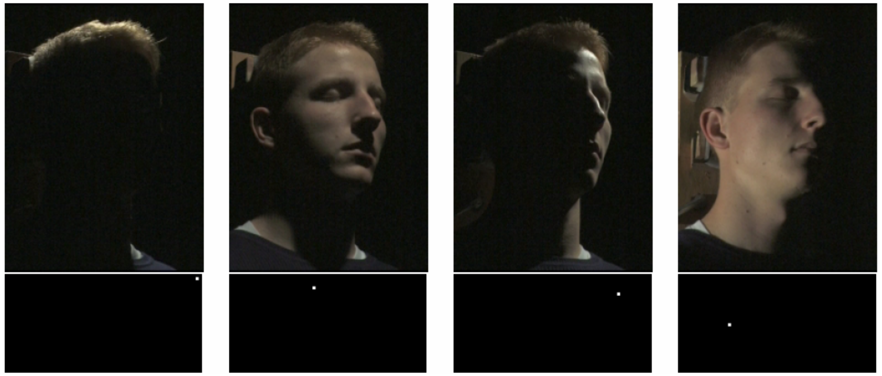
Light is additive



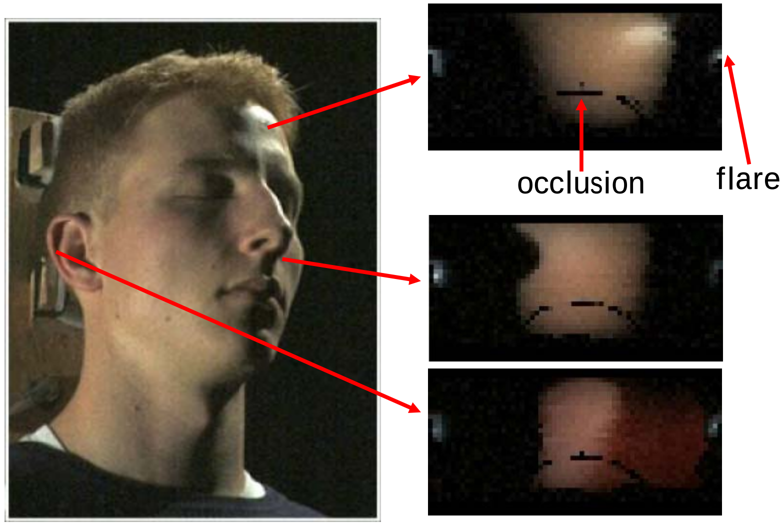
Light stage 1.0

Light stage 1.0

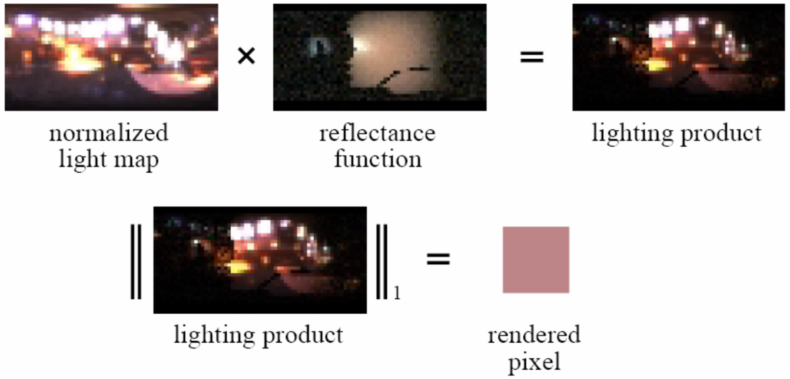
Input images



Reflectance function

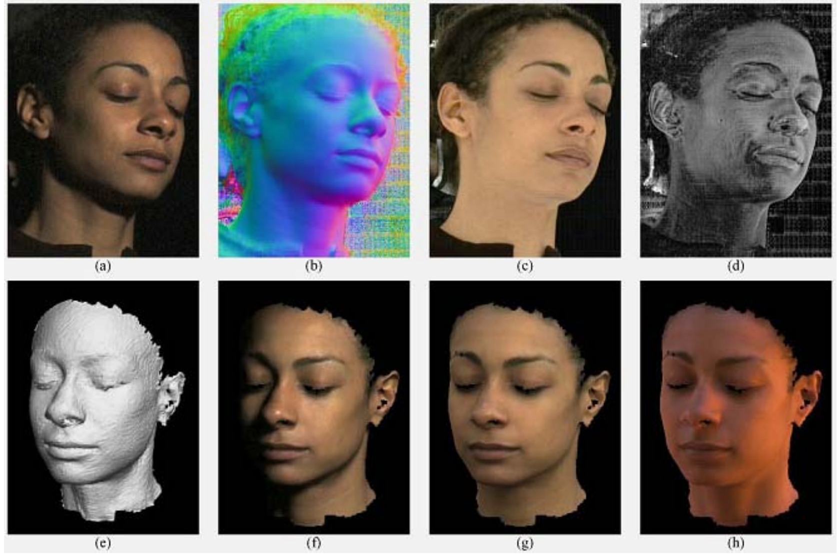


Relighting

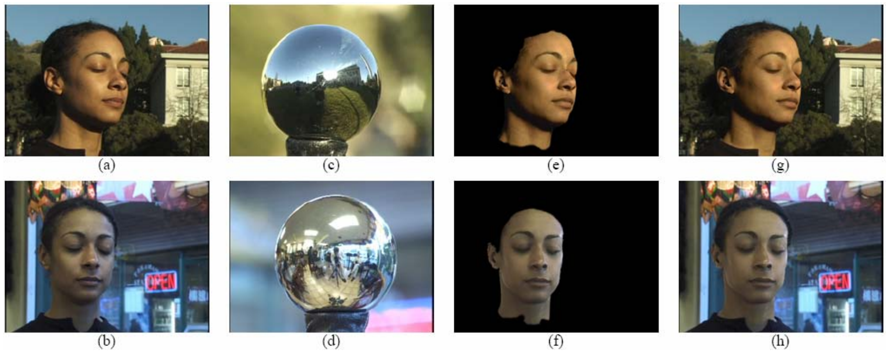


Results

Changing viewpoints



Results



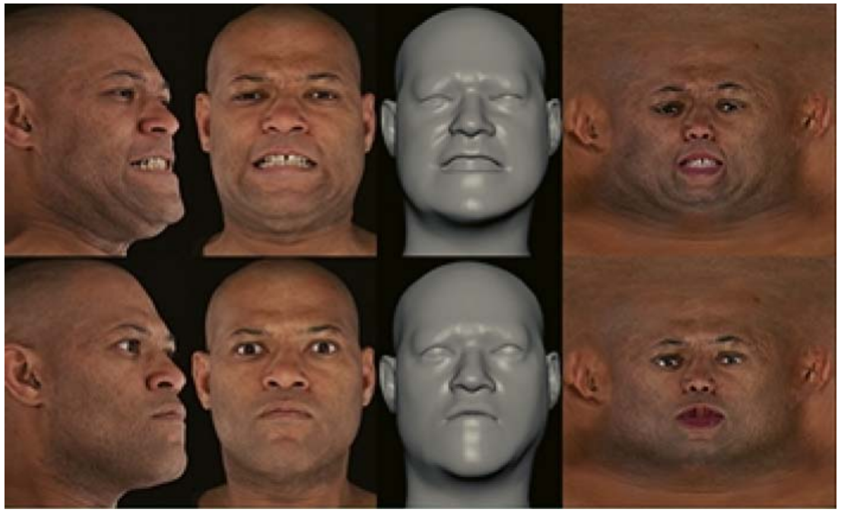
Video

Spiderman 2

real synthetic

Light stage 3

Application: The Matrix Reloaded



Application: The Matrix Reloaded

Reference

- F. Pighin, J. Hecker, D. Lischinski, D. H. Salesin, and R. Szeliski. Synthesizing realistic facial expressions from photographs. SIGGRAPH 1998, pp75-84.
- Brian Guenter, Cindy Grimm, Henrique Malvar, Daniel Wood, Making Faces, SIGGRAPH 1998.
- Li Zhang, Noah Snavely, Brian Curless, Steven M. Seitz, Spacetime Faces: High Resolution Capture for Modeling and Animation, SIGGRAPH 2004.
- Blanz, V. and Vetter, T., A Morphable Model for the Synthesis of 3D Faces, SIGGRAPH 1999, pp187-194.
- V. Blanz, C. Basso, T. Poggio and T. Vetter, Reanimating Faces in Images and Video, EUROGRAPHICS 2003.
- V. Blanz, K. Scherbaum, T. Vetter, H.P. Seidel, Exchanging Faces in Images, EUROGRAPHICS 2004.
- George Borshukov et al., Universal Capture - Image-based Facial Animation for "The Matrix Reloaded", SIGGRAPH 2003 Sketch.

Reference

- George Borshukov et al., Realistic Human Face Rendering for "The Matrix Reloaded", SIGGRAPH 2003 Sketch.
- Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, Mark Sagar, Acquiring the Reflectance Field of a Human Face, SIGGRAPH 2000.
- Paul Debevec, Chris Tchou, Andreas Wenger, Tim Hawkins, Andy Gardner, Brian Emerson, Ansul Panday, A Lighting Reproduction Approach to Live-Action Compositing, SIGGRAPH 2002.
- Mark Sagar, Reflectance Field Rendering of Human Faces for "Spider-Man 2", SIGGRAPH 2004 Sketch.
- Christoph Bregler, Malcolm Slaney, Michele Covell, Video Rewrite: Driving Visual Speeach with Audio, SIGGRAPH 1997.
- Tony Ezzat, Gadi Geiger, Tomaso Poggio, Trainable Videorealistic Speech Animation, SIGGRAPH 2002.
- Brett Allen, Brian Curless, Zoran Popovic, The Space of Human Body Shapes: Reconstruction and Parameterization From Range Scans, SIGGRAPH 2003.