Faces and Image-Based Lighting

Digital Visual Effects, Spring 2008

Yung-Yu Chuang

2007/6/3

Announcements

- Project #3 artifacts voting
- Final project:
  - Demo on 6/25 (Wednesday) 13:30pm in this room
  - Reports and videos due on 6/26 (Thursday) 11:59pm

Outline

- Image-based lighting

- 3D acquisition for faces
- Statistical methods (with application to face super-resolution)
- 3D Face models from single images
- Image-based faces
- Relighting for faces

Image-based lighting

Rendering

- Rendering is a function of geometry, reflectance, lighting and viewing.

- To synthesize CGI into real scene, we have to match the above four factors.

- Viewing can be obtained from *calibration* or *structure from motion*.

- Geometry can be captured using *3D photography* or made by hands.

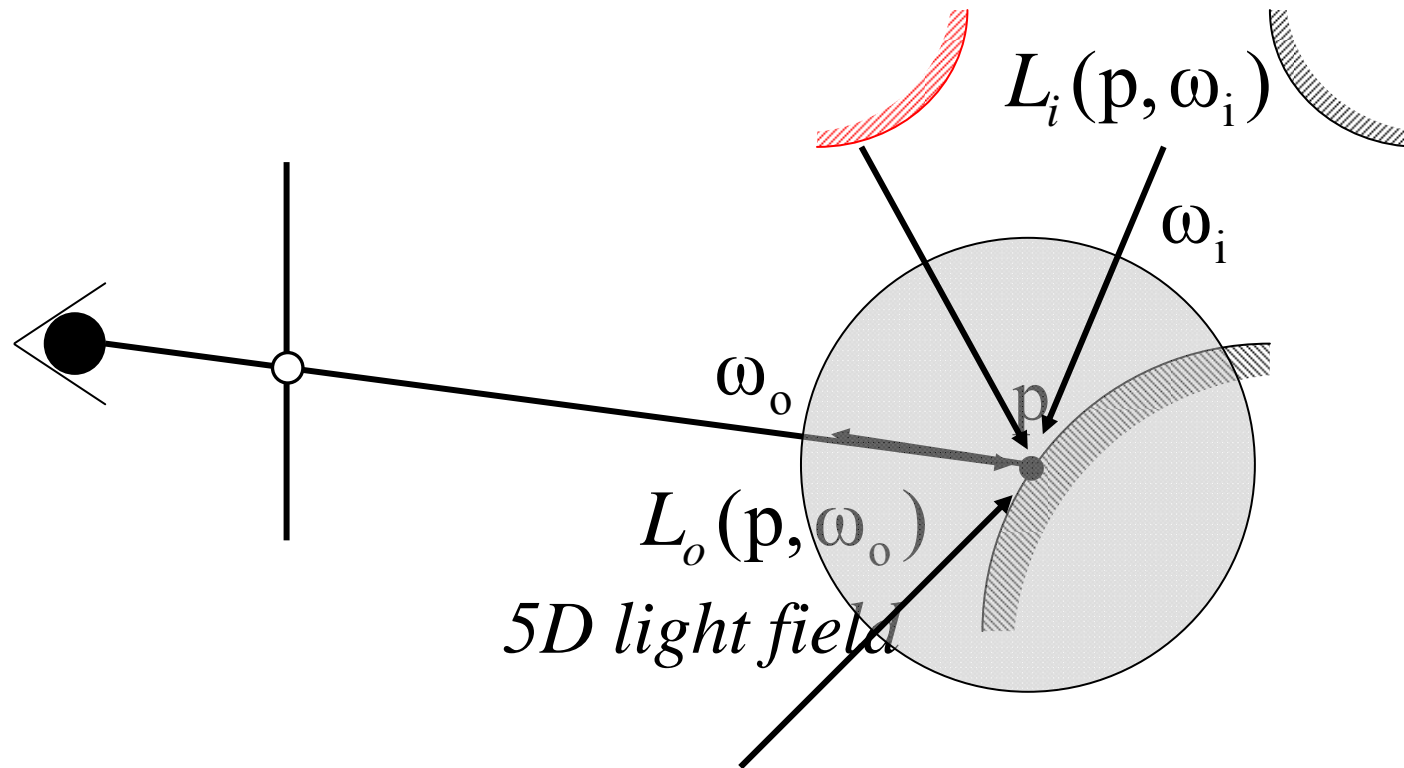- How to capture lighting and reflectance?

Reflectance

- The Bidirectional Reflection Distribution Function
  - Given an incoming ray (θ_i, ϕ_i) and outgoing ray (θ_e, ϕ_e) what proportion of the incoming light is reflected along out

surface normal
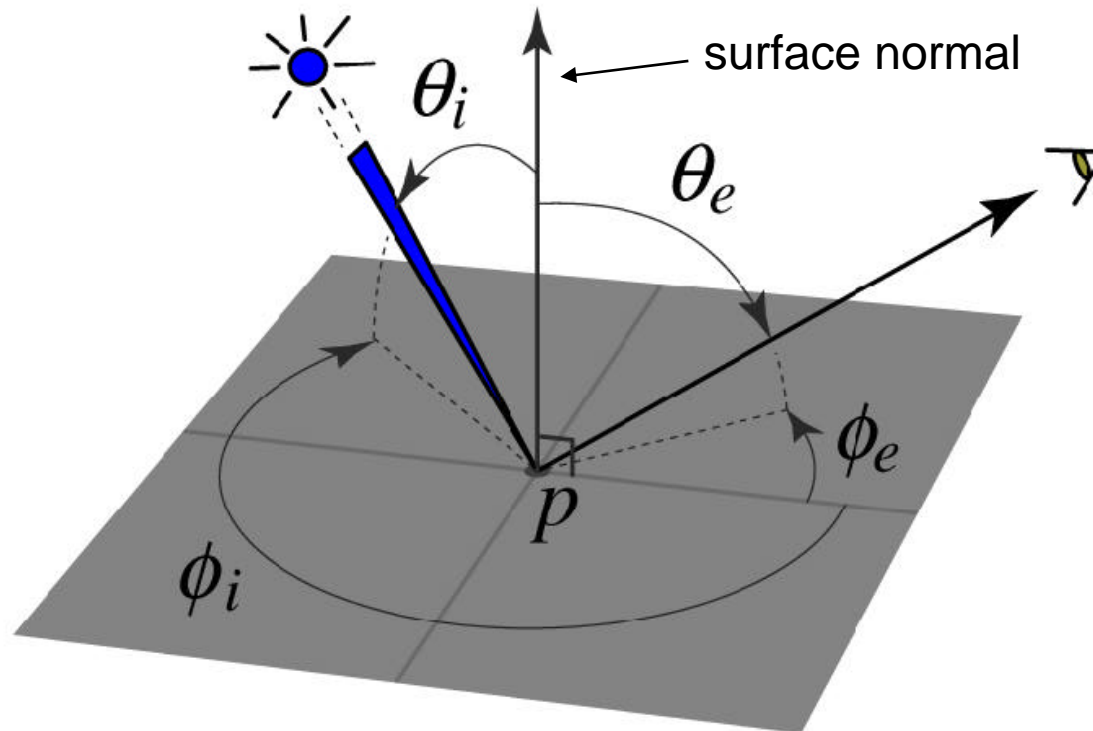
Answer given by the BRDF: $\rho(\theta_i, \phi_i, \theta_e, \phi_e)$

Rendering equation

$L_i(\mathrm{p},\omega_i)$

ω_i

ω_o

$L_o(\mathrm{p},\omega_o)$

p

5D light field

$$L_o(\mathrm{p},\omega_o) = L_e(\mathrm{p},\omega_o)$$

$$+ \int_{s^2} \rho(\mathrm{p},\omega_o,\omega_i) L_i(\mathrm{p},\omega_i) \left| \cos\theta_i \right| d\omega_i$$

Complex illumination

DigiVFX

$$L_o(\text{p}, \omega_o) = L_e(\text{p}, \omega_o)$$

$$+ \int_{s^2} f(\text{p}, \omega_o, \omega_i) L_i(\text{p}, \omega_i) \left| \cos \theta_i \right| d\omega_i$$

$$B(\text{p}, \omega_o) = \int_{s^2} f(\text{p}, \omega_o, \omega_i) L_d(\text{p}, \omega_i) \left| \cos \theta_i \right| d\omega_i$$

reflectance   lighting

Natural illumination

People perceive materials more easily under natural illumination than simplified illumination.

Images courtesy Ron Dror and Ted Adelson

Natural illumination

Rendering with natural illumination is more expensive compared to using simplified illumination
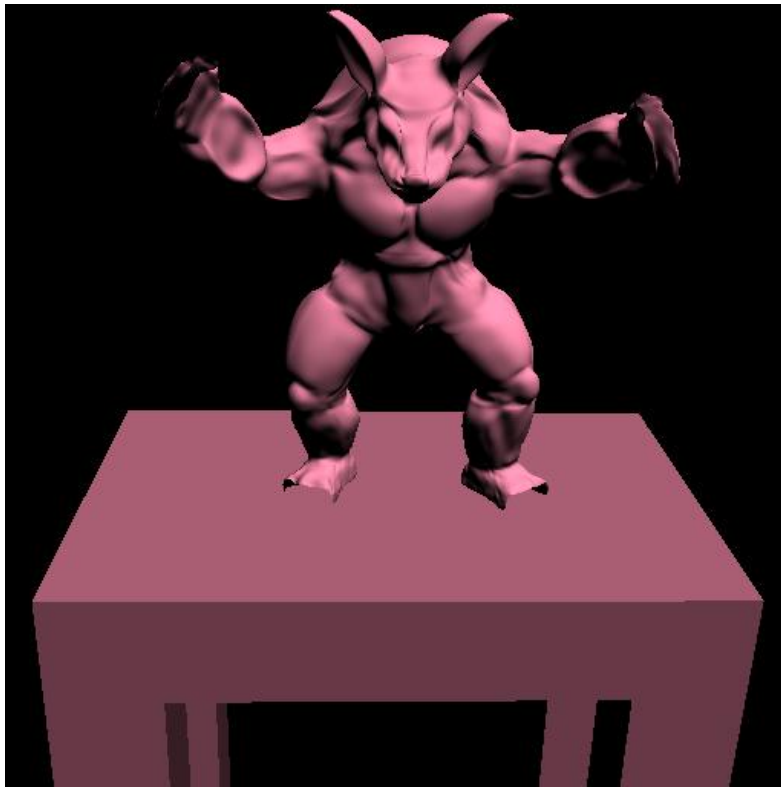
directional source

natural illumination

Environment maps

Miller and Hoffman, 1984
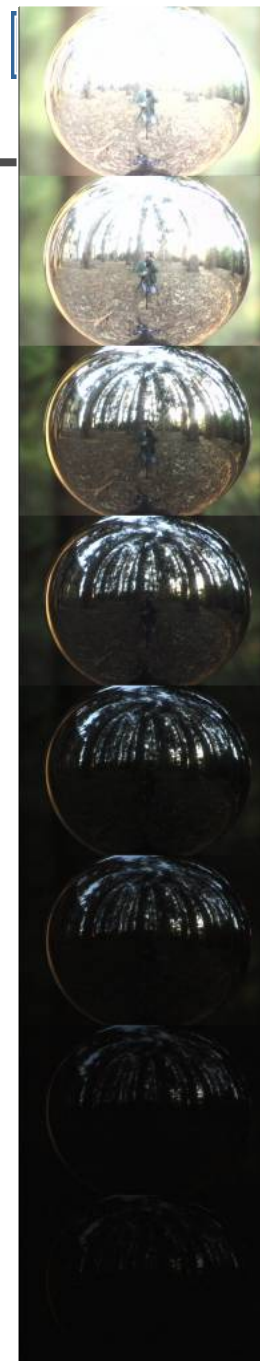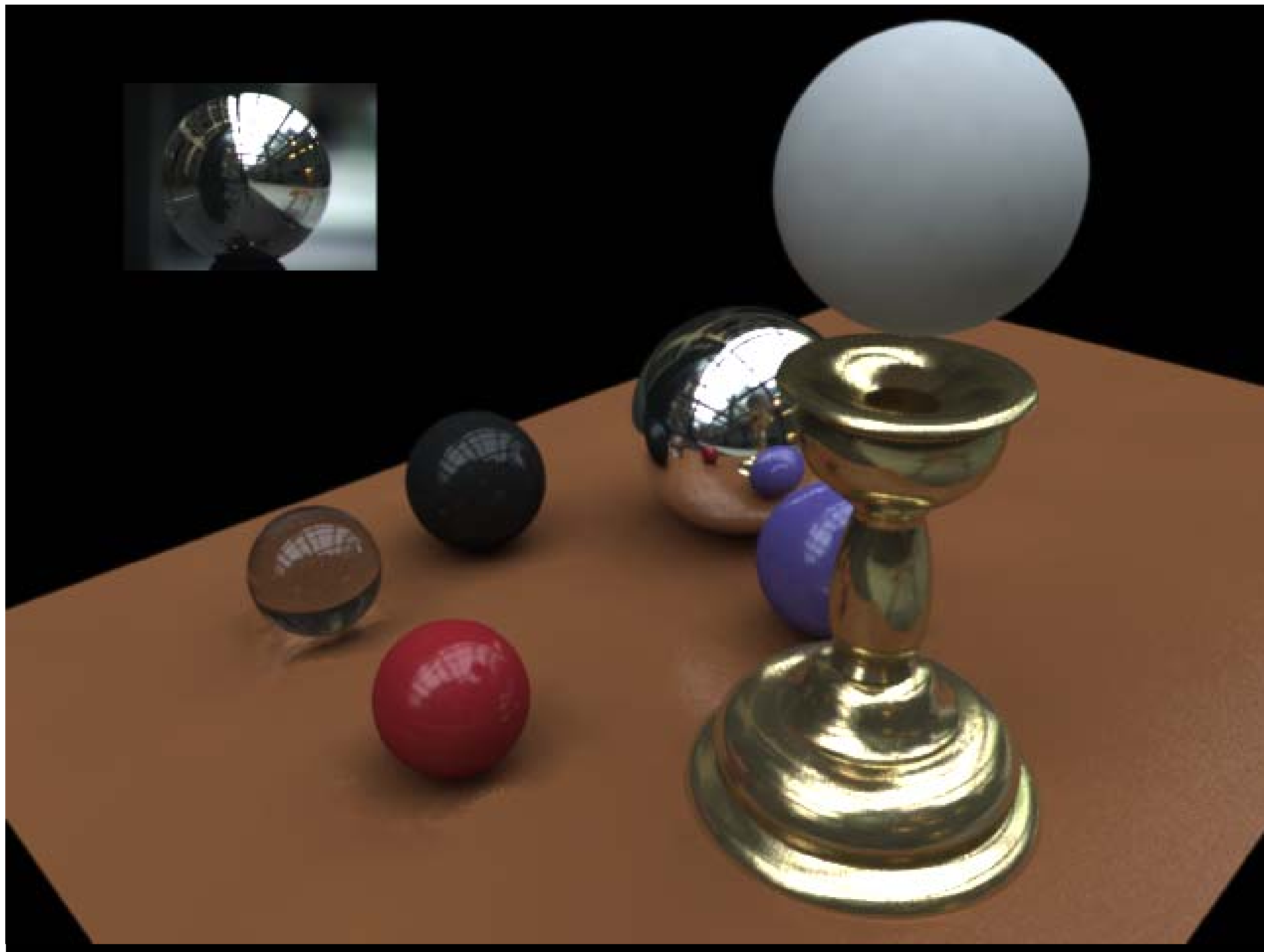
HDRI Sky Probe
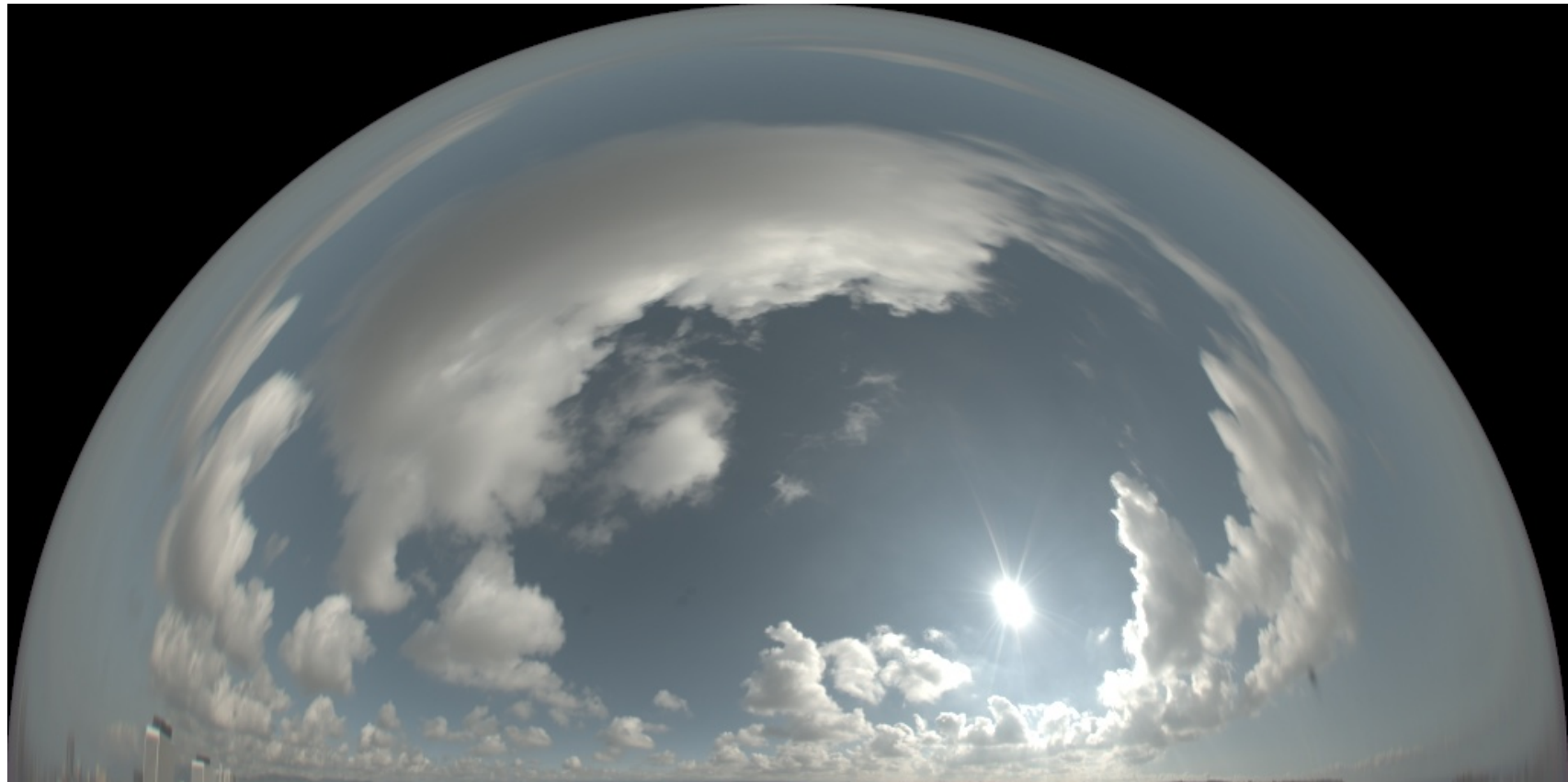
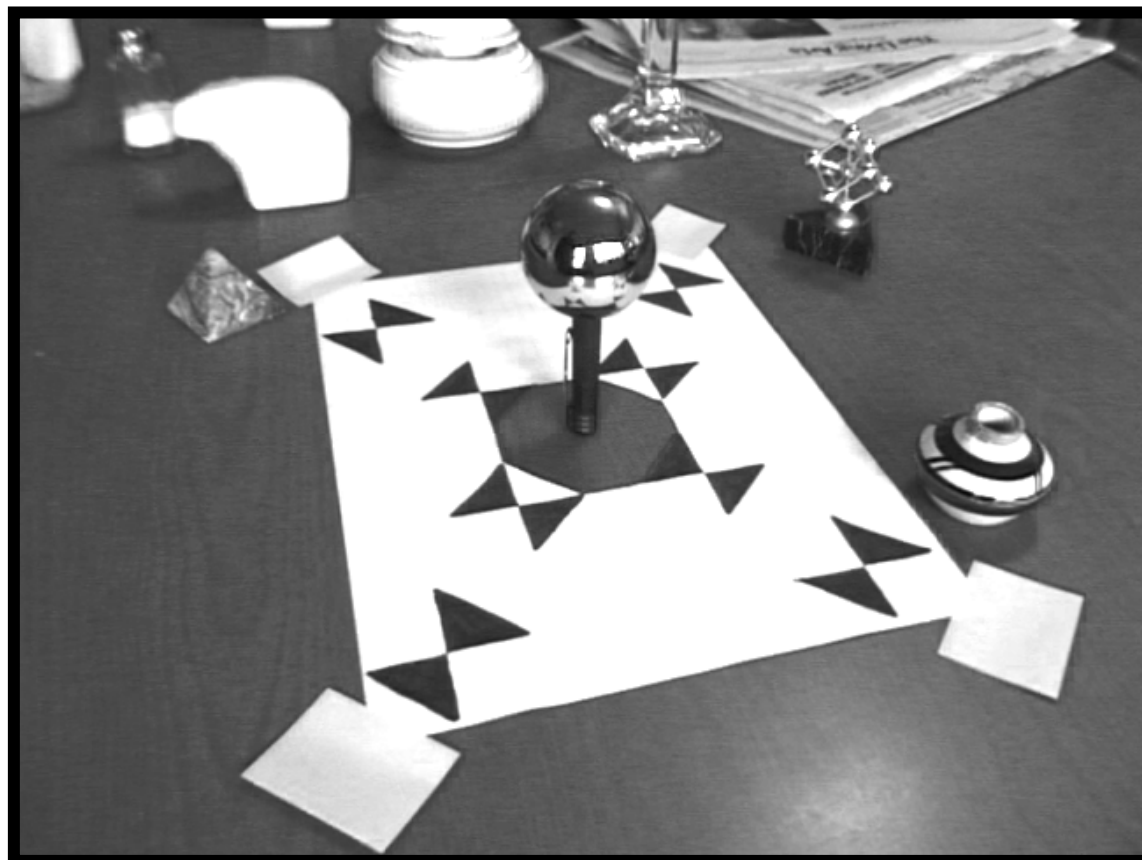# Clipped Sky + Sun Source
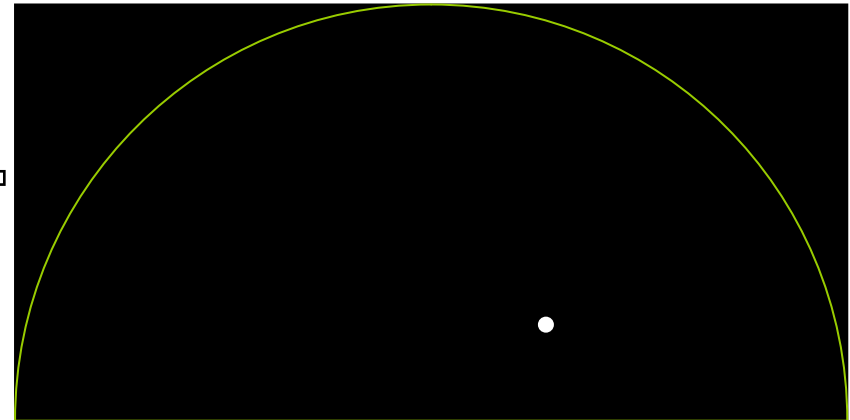
Lit by sun only
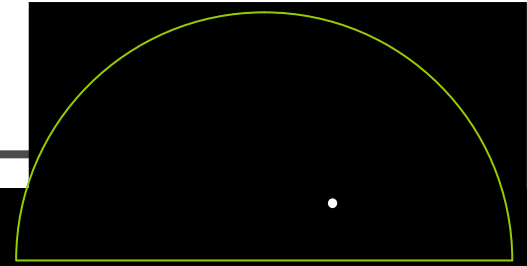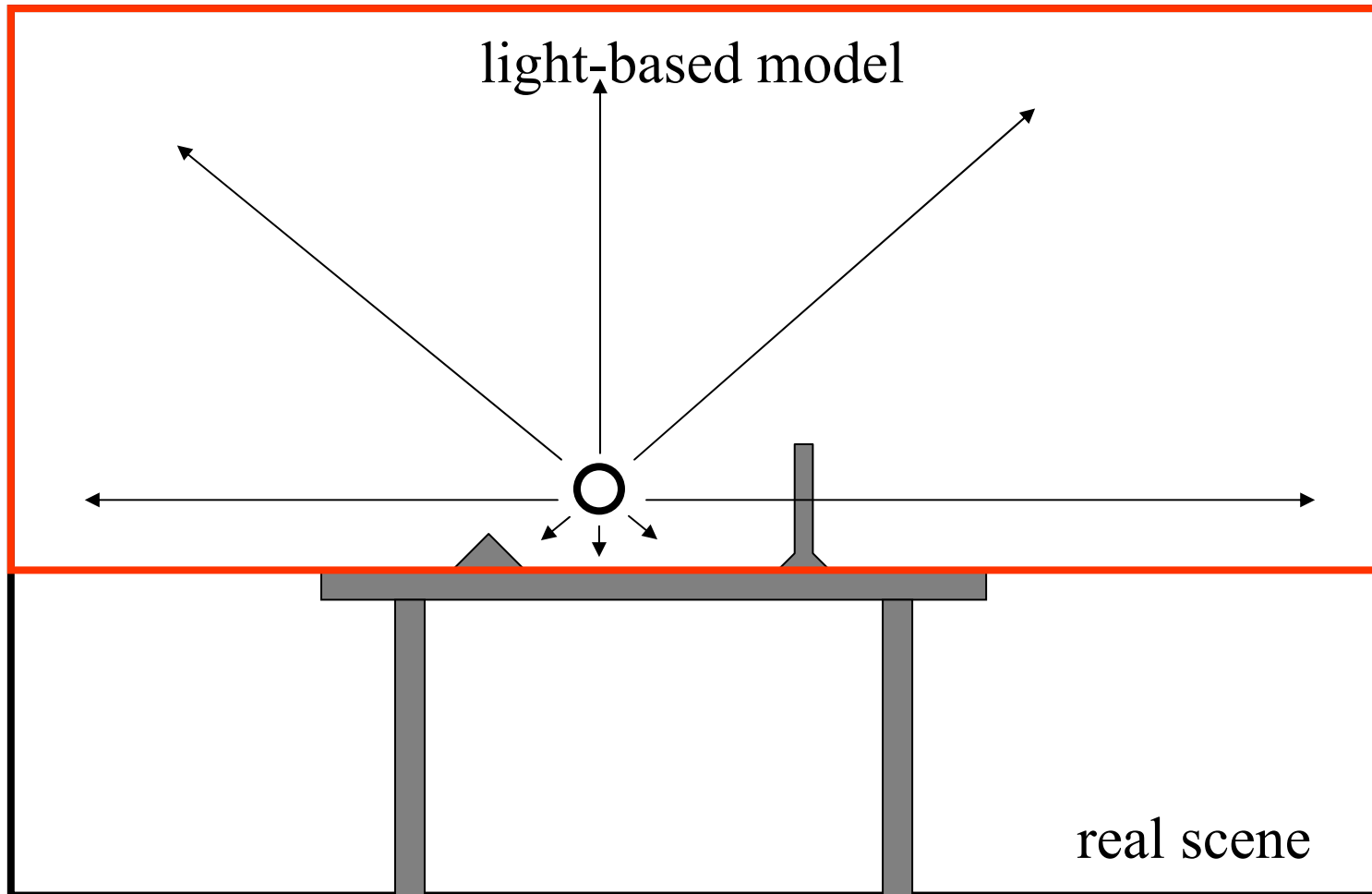
Lit by sky only

Lit by sun and sky
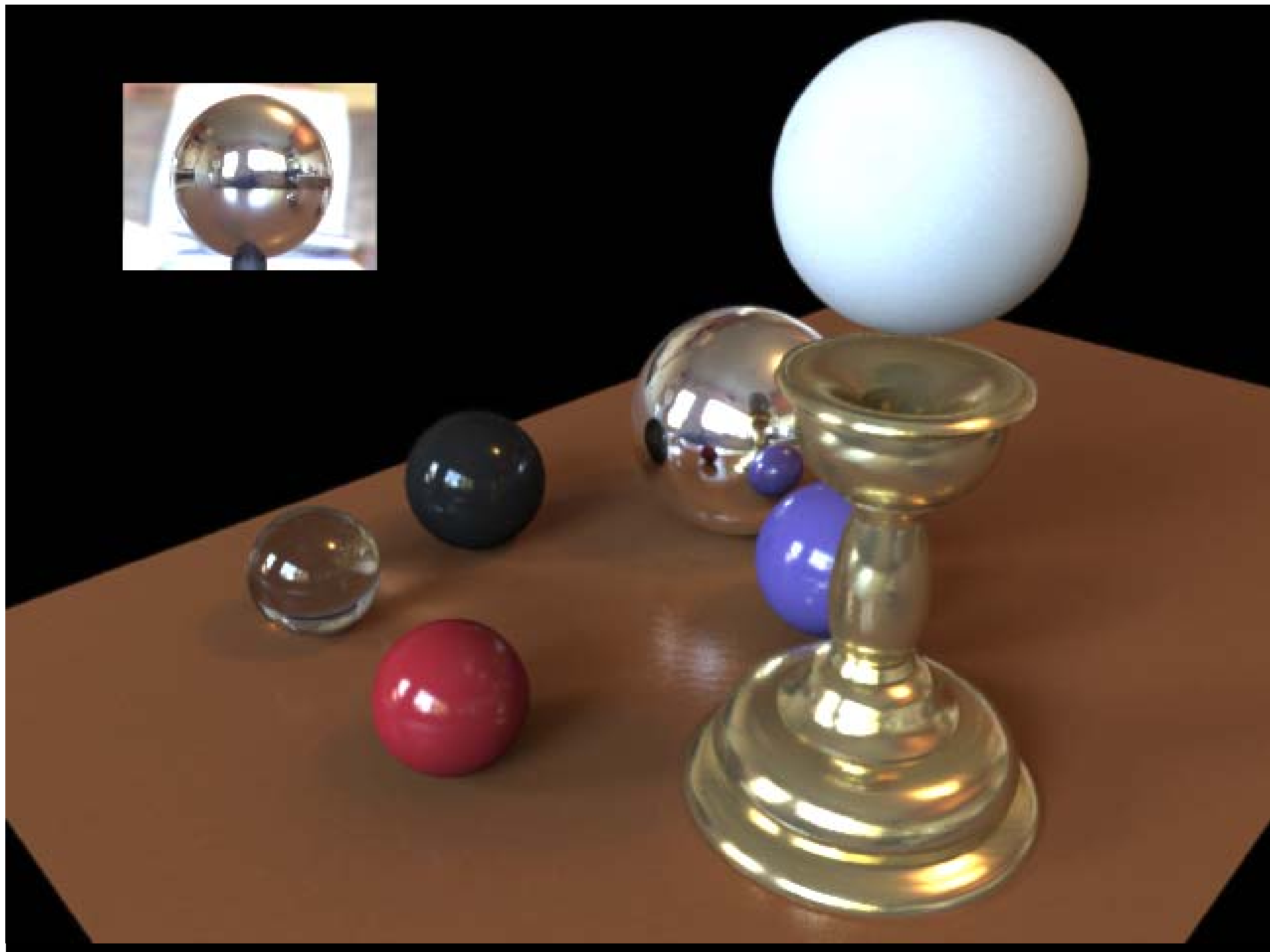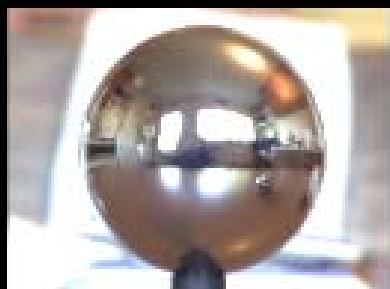
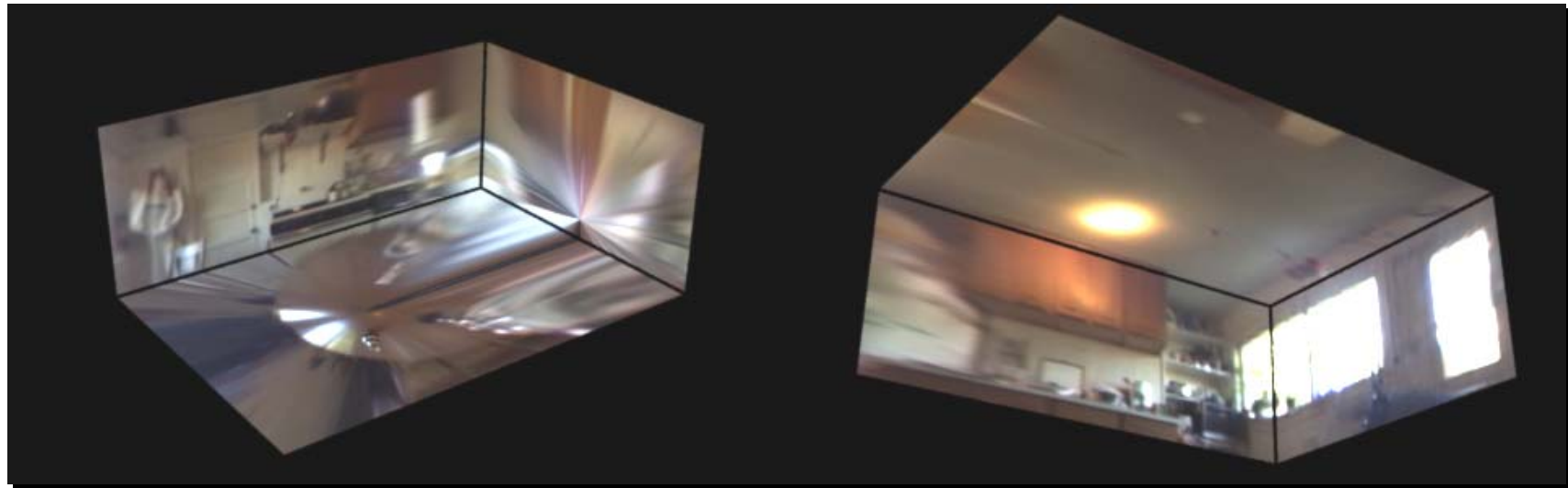# Real Scene Example

- Goal: place synthetic objects on table

Light Probe / Calibration Grid

Modeling the Scene

light-based model

real scene

The *Light-Based* Room Model
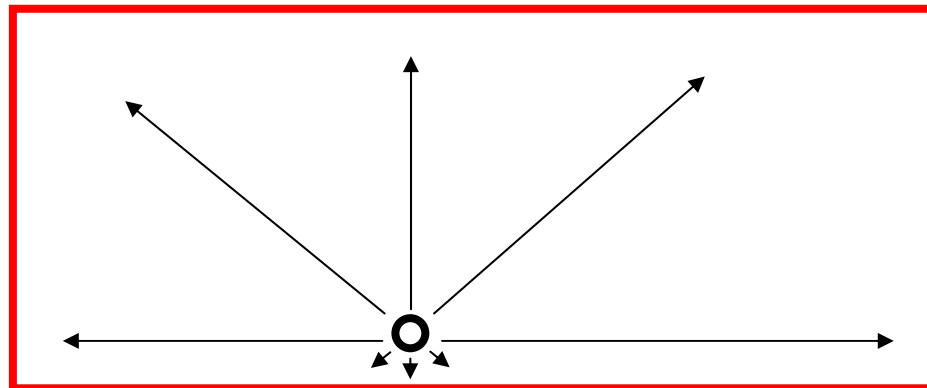
Rendering into the Scene

- Background Plate

Rendering into the scene

- Objects and Local Scene matched to Scene

Differential rendering
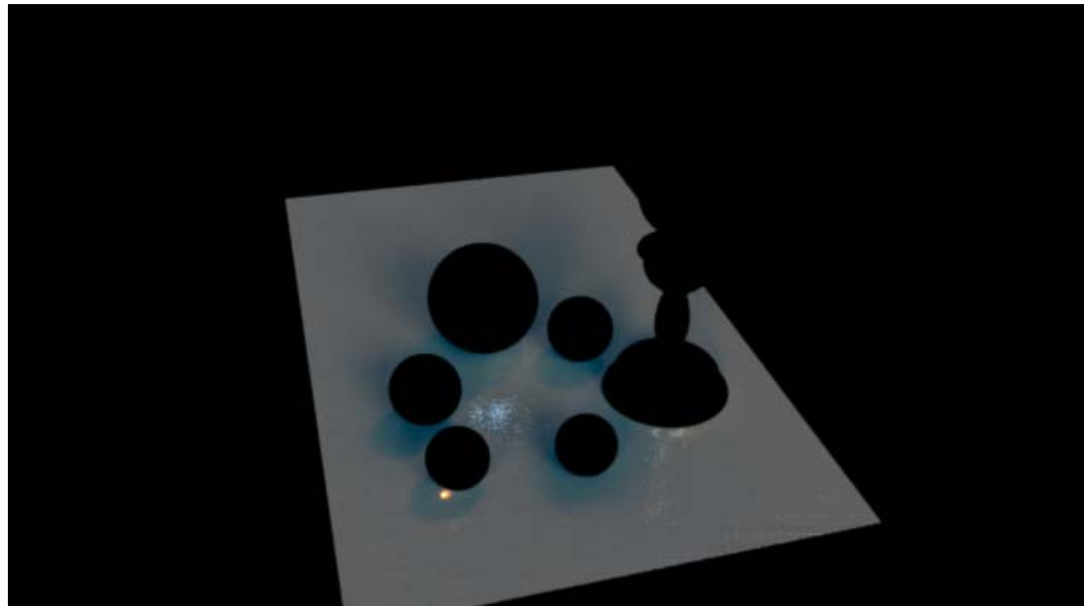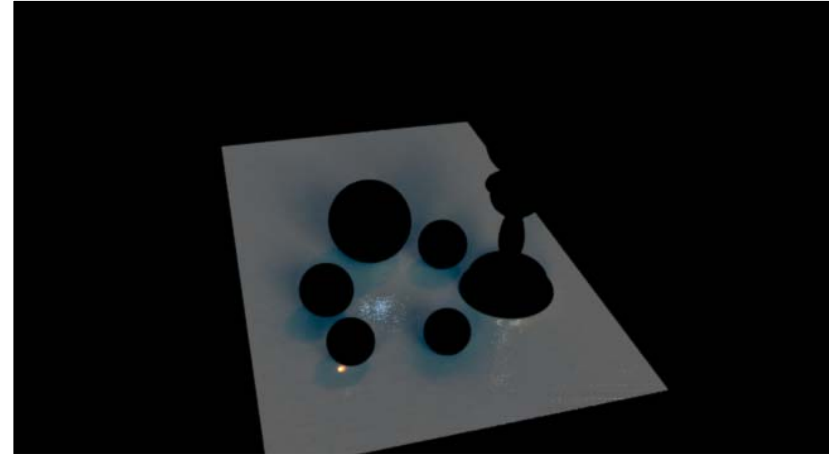
- Local scene w/o objects, illuminated by model

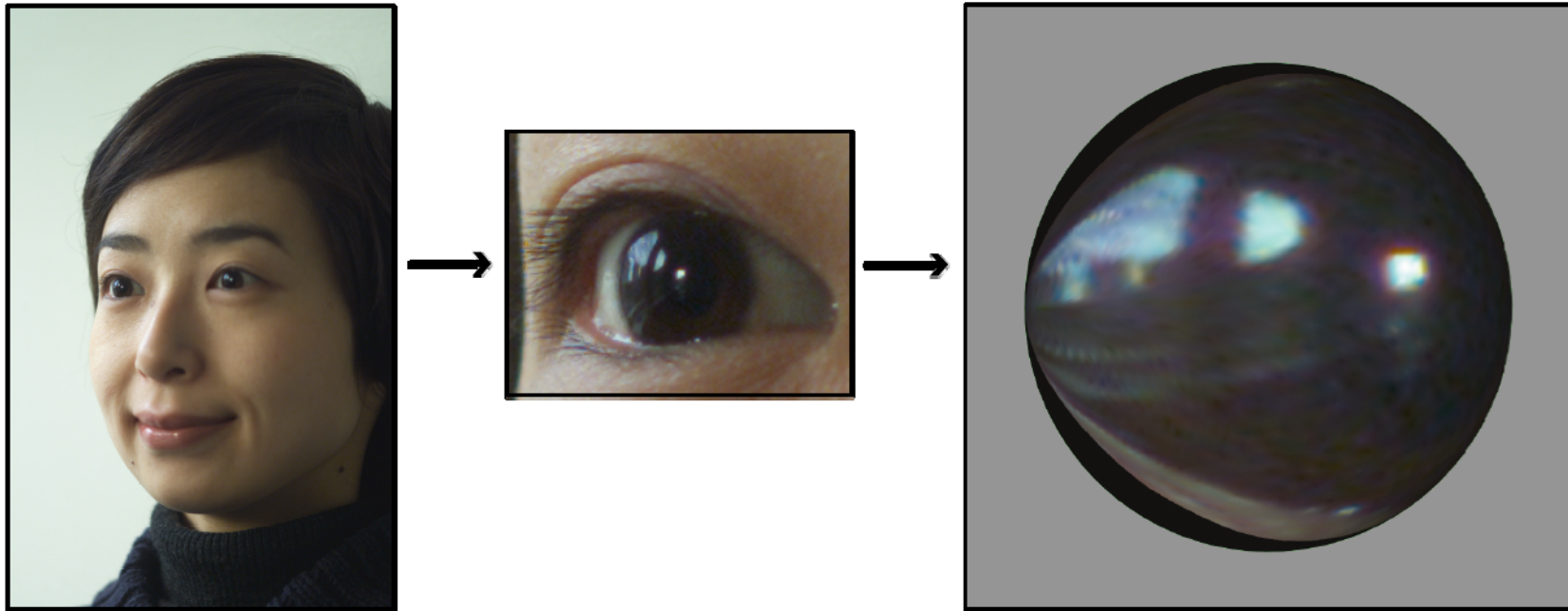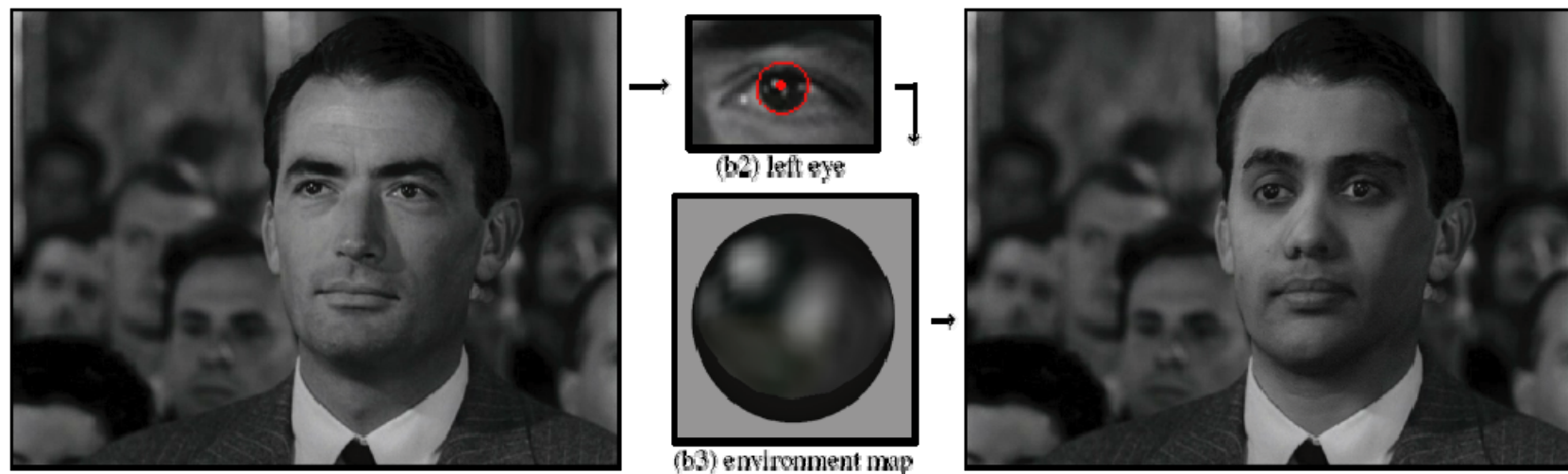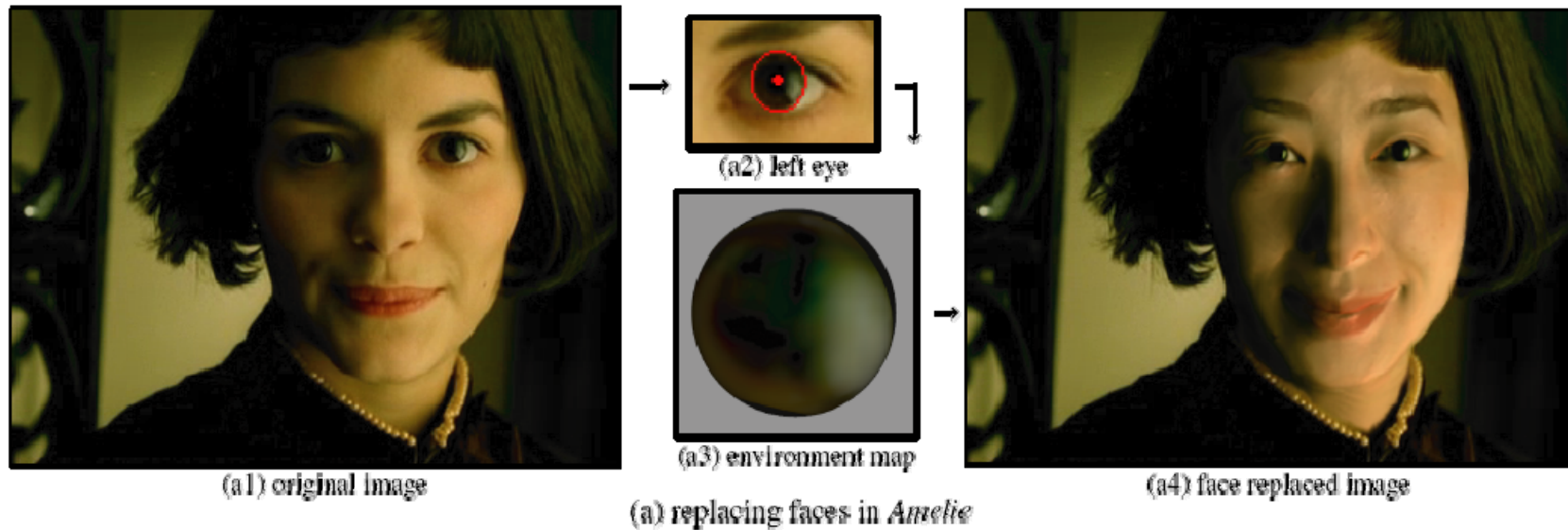# Differential rendering

Differential rendering

Environment map from single image?

Eye as light probe! (Nayar et al)

Results

(a1) original image

(a2) left eye

(a3) environment map

(a4) face replaced image

(a) replacing faces in *Amelie*
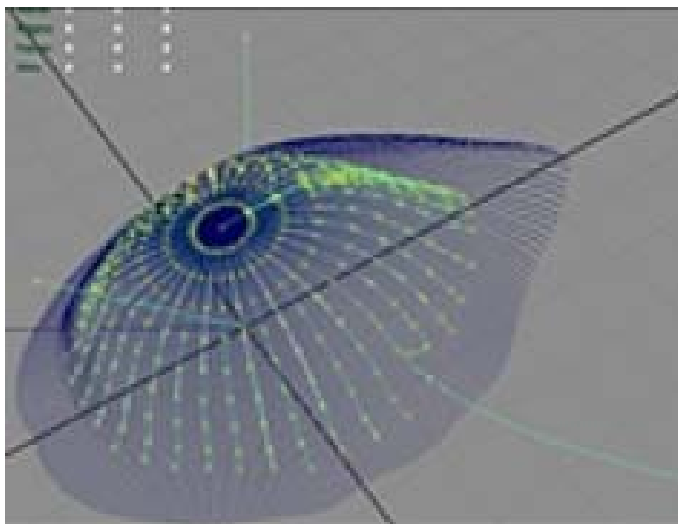
(b2) left eye

(b3) environment map

Application in "Superman returns"

Capturing reflectance
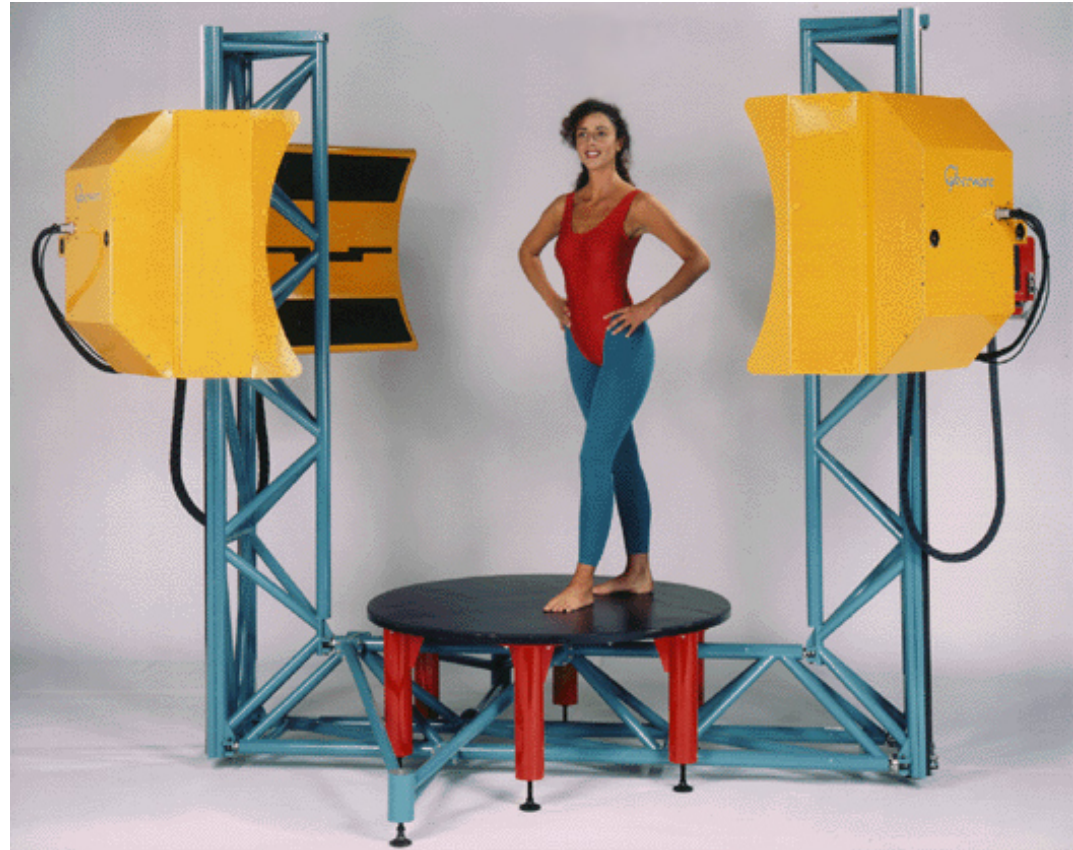
Application in "The Matrix Reloaded"

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 the generic face model to fit points
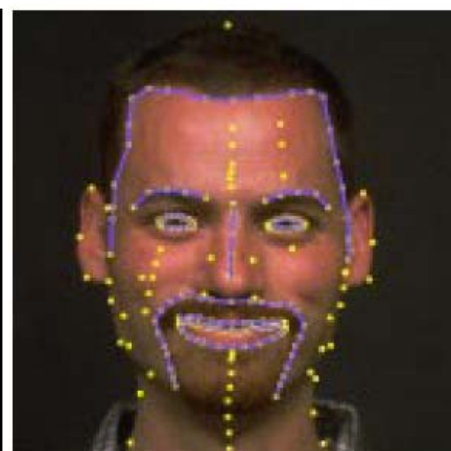 5. Extract textures from photos

Reconstruct a 3D model

input photographs
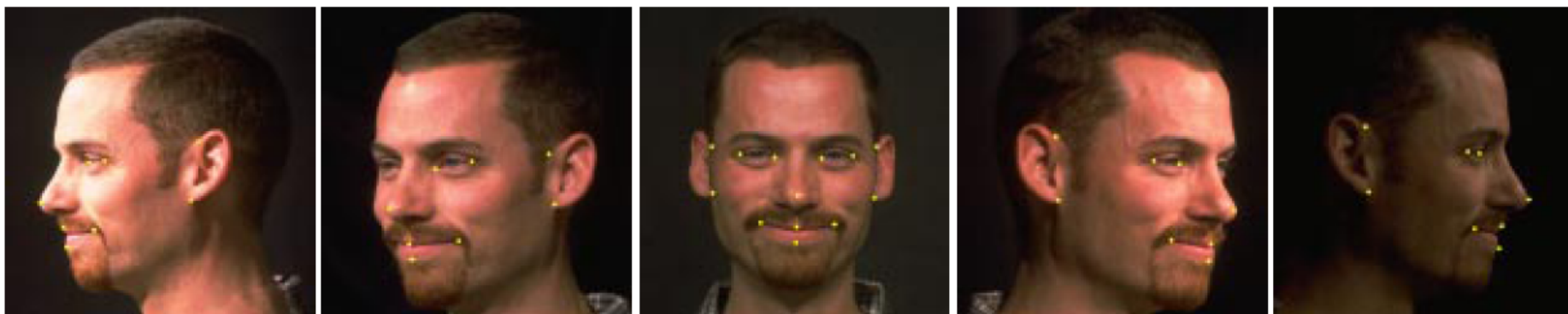
generic 3D
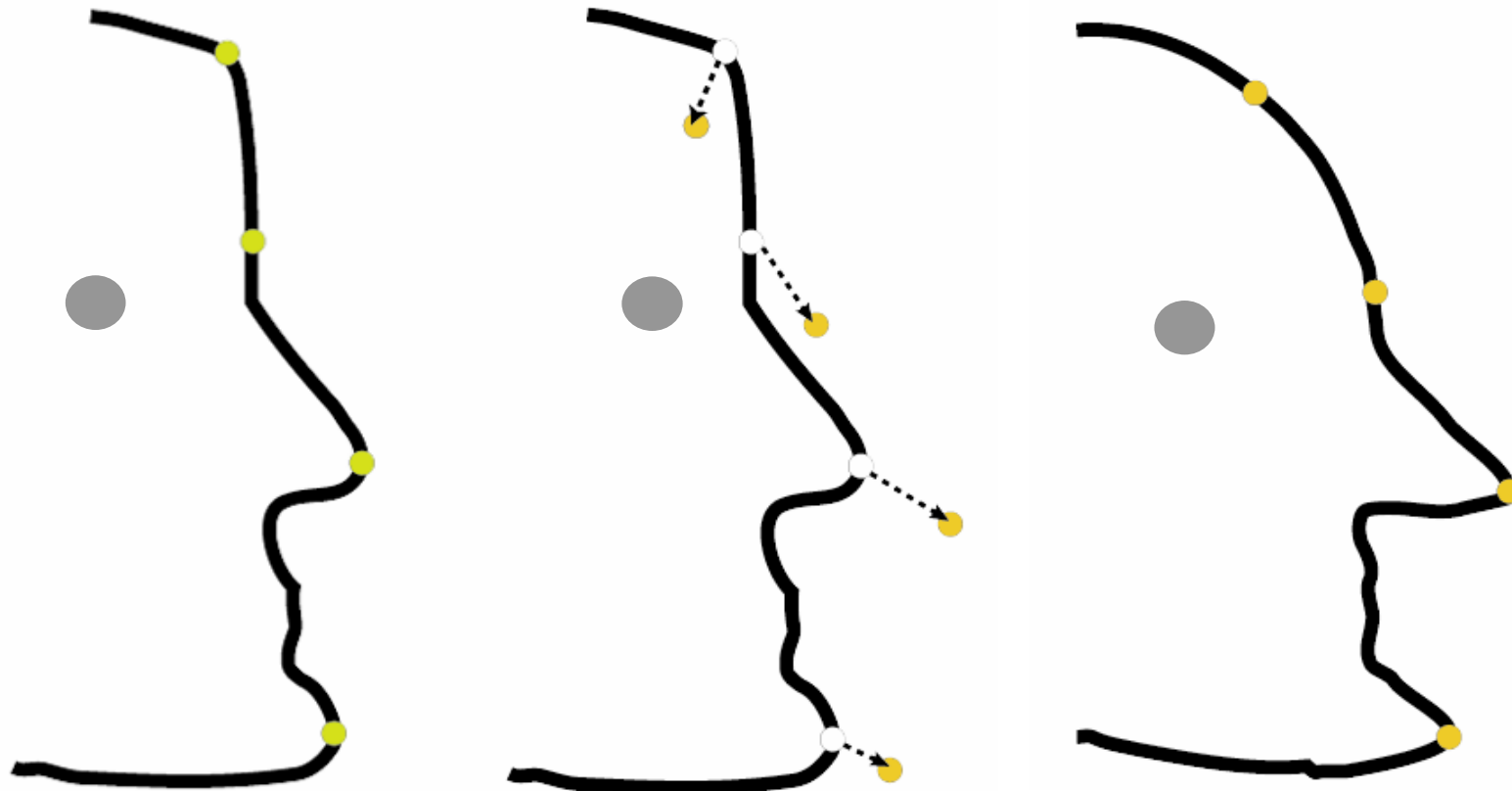face model

pose
estimation

more
features

deformed
model

Mesh deformation

- Compute displacement of feature points
- Apply scattered data interpolation
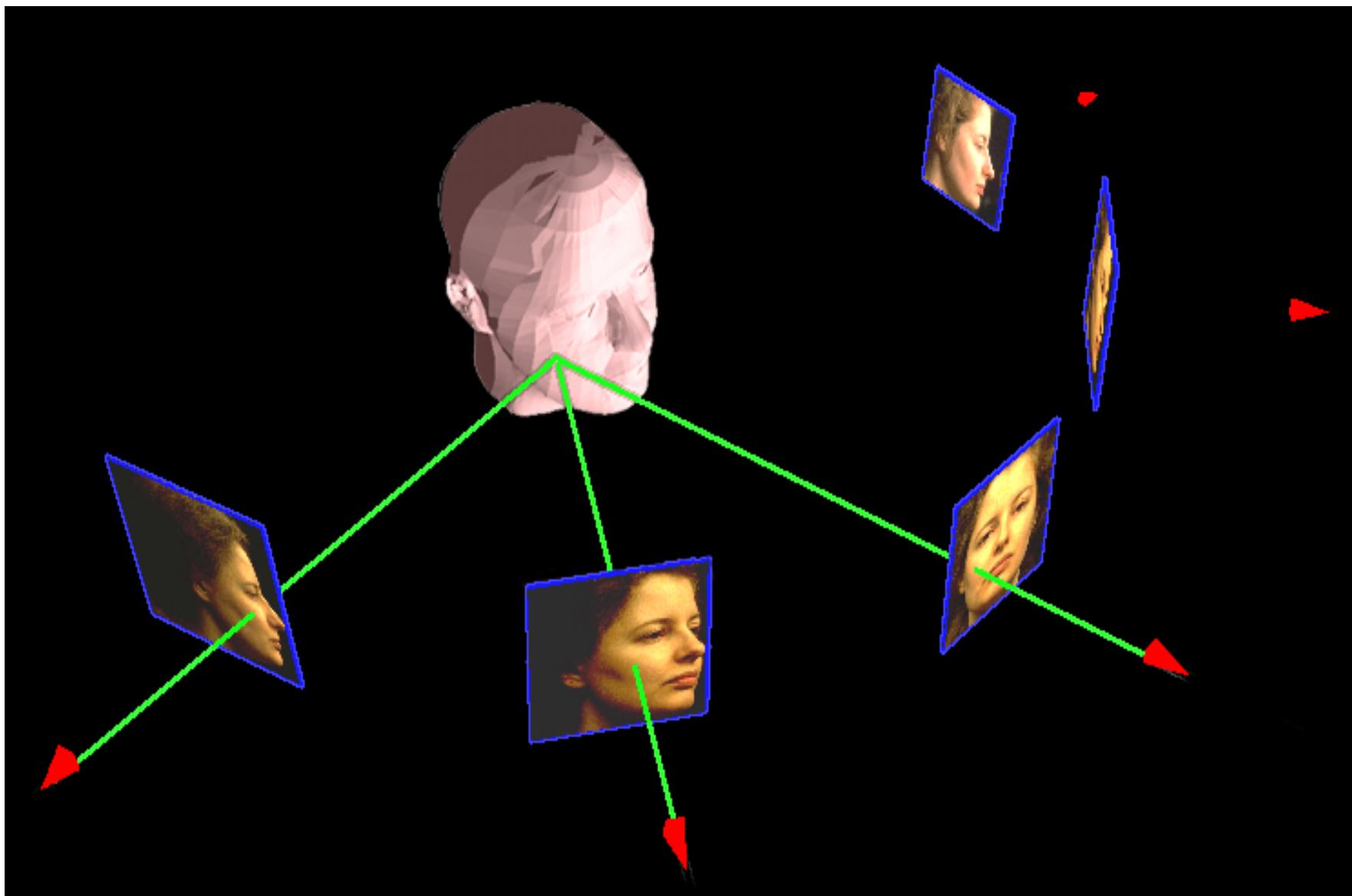

generic model   displacement   deformed model

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

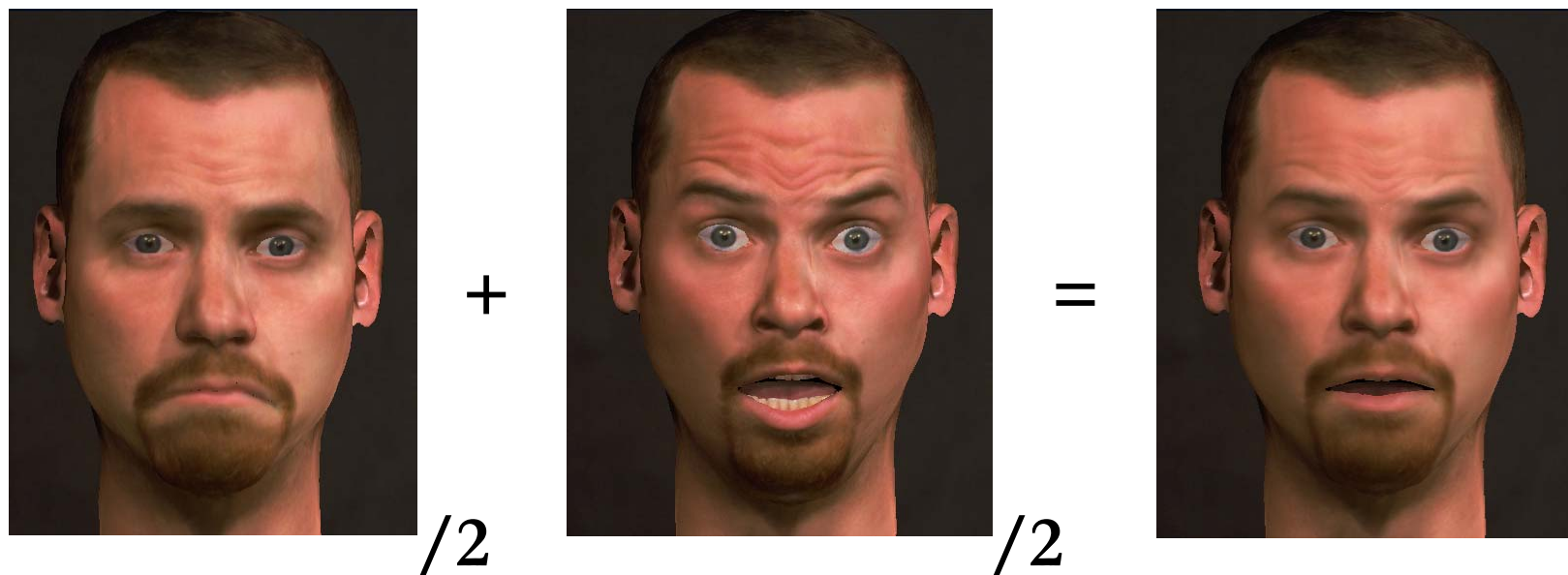view-independent   view-dependent

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:

+ /2   = /2
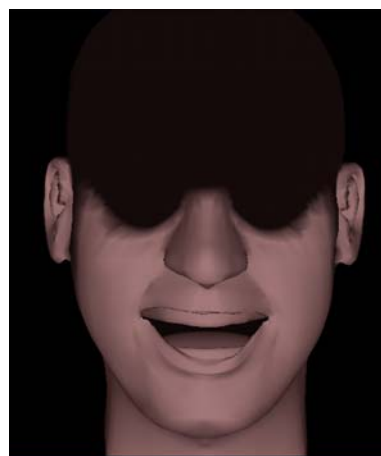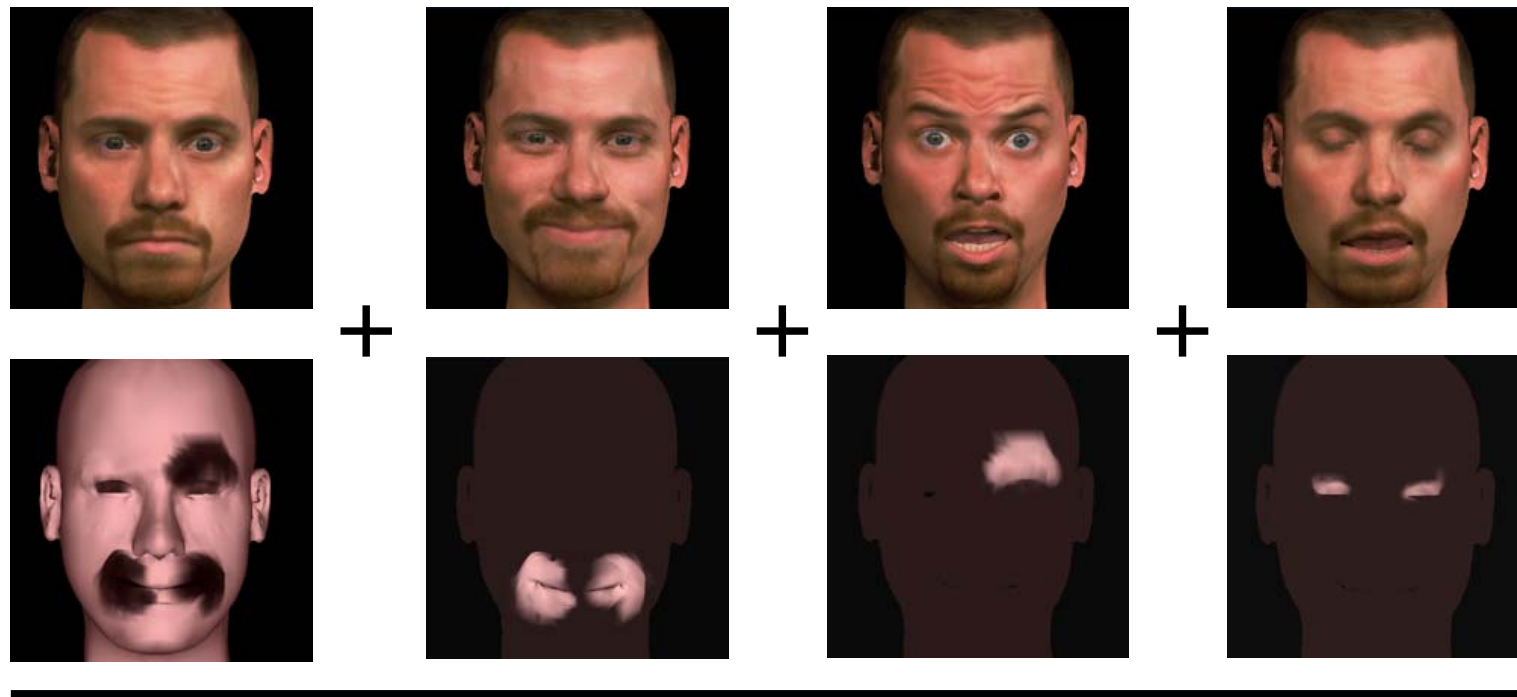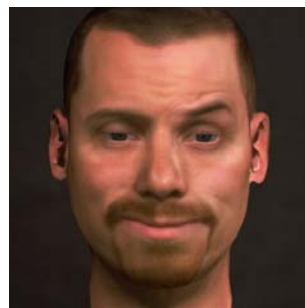
Applying a global blend

Creating new expressions

Applying a region-based blend

Creating new expressions

Using a painterly interface

Drunken smile

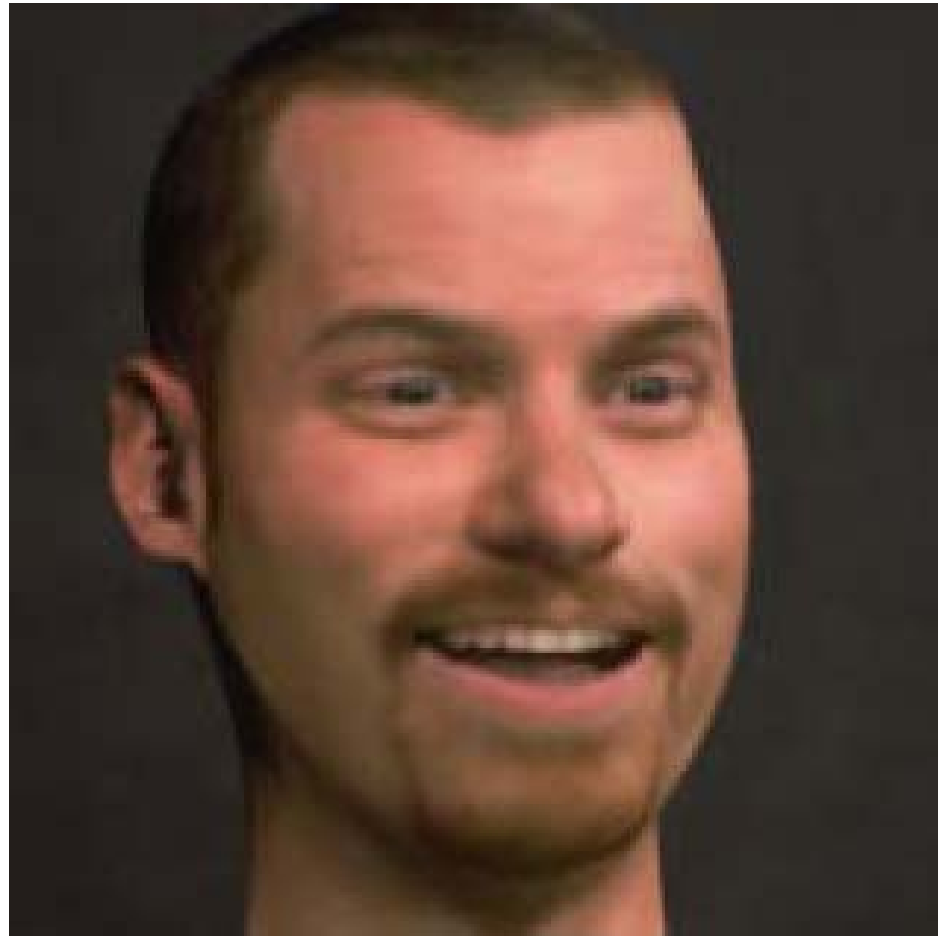# Animating between expressions
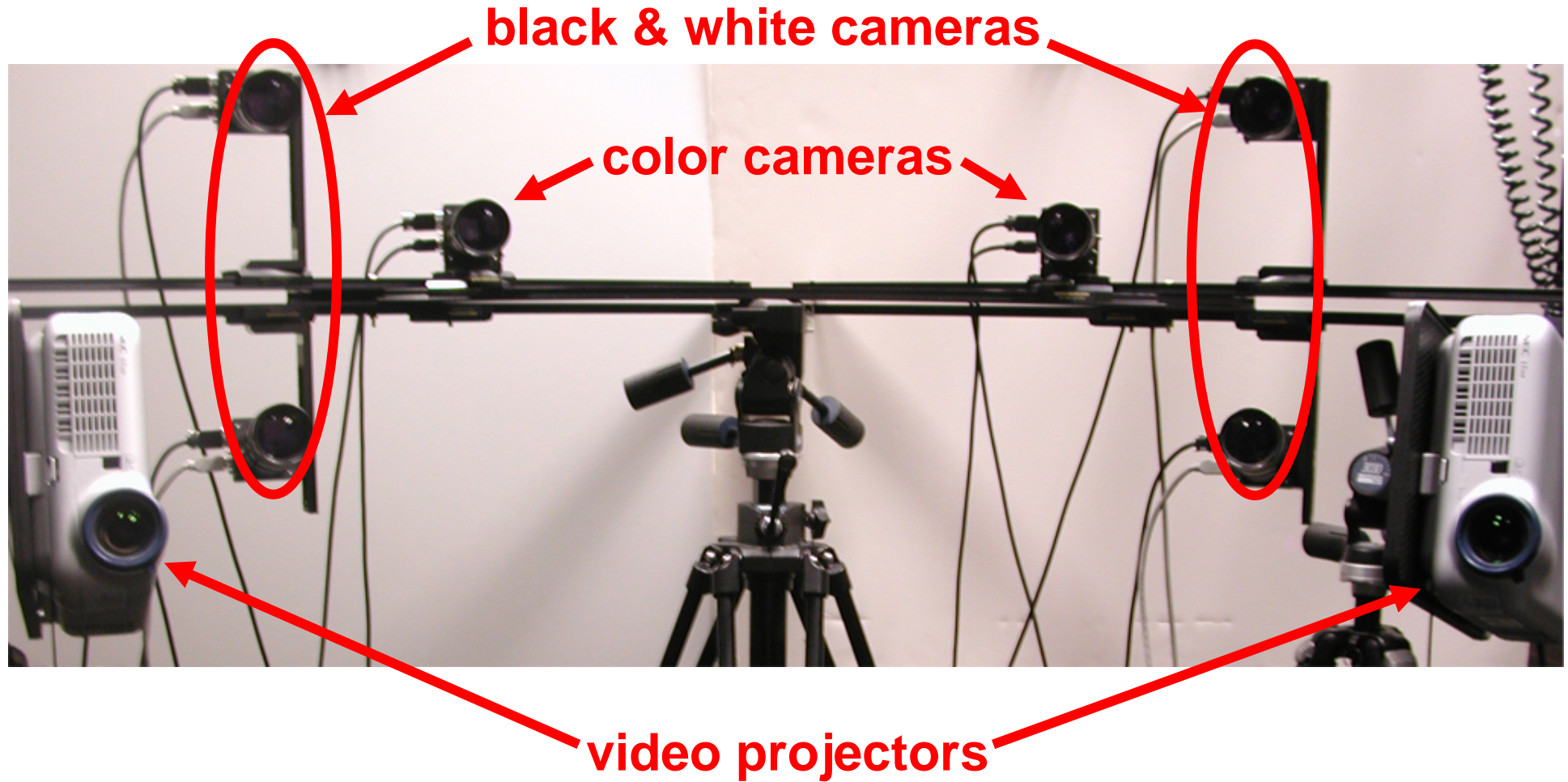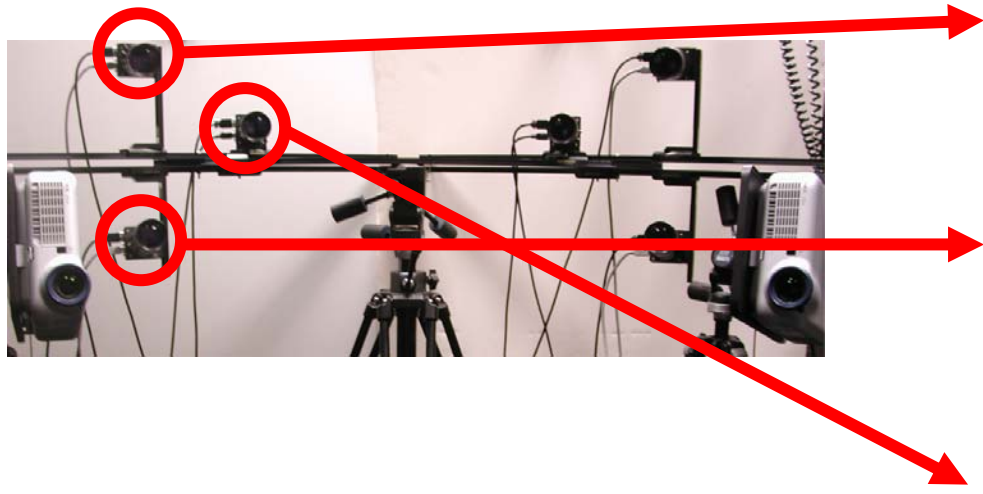
Morphing over time creates animation:

"neutral" ⟶ "joy"

Video

Spacetime faces

Spacetime faces
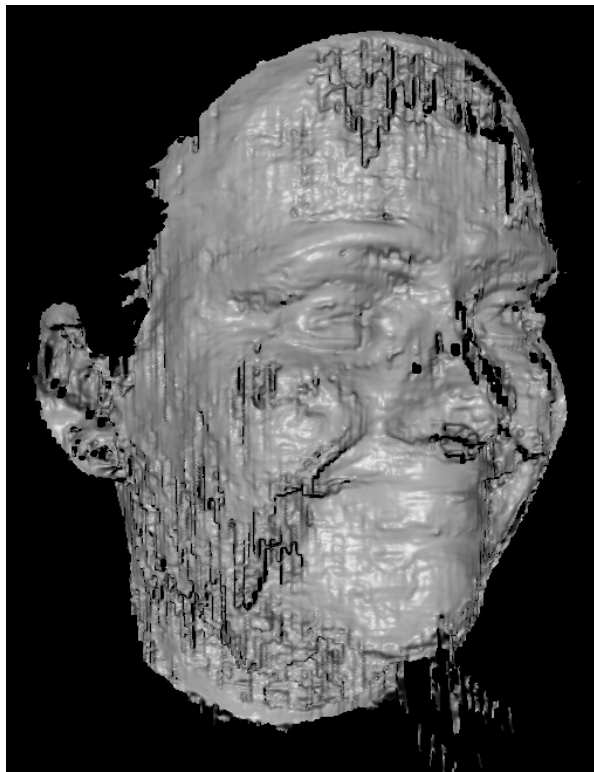
black & white cameras

color cameras

video projectors
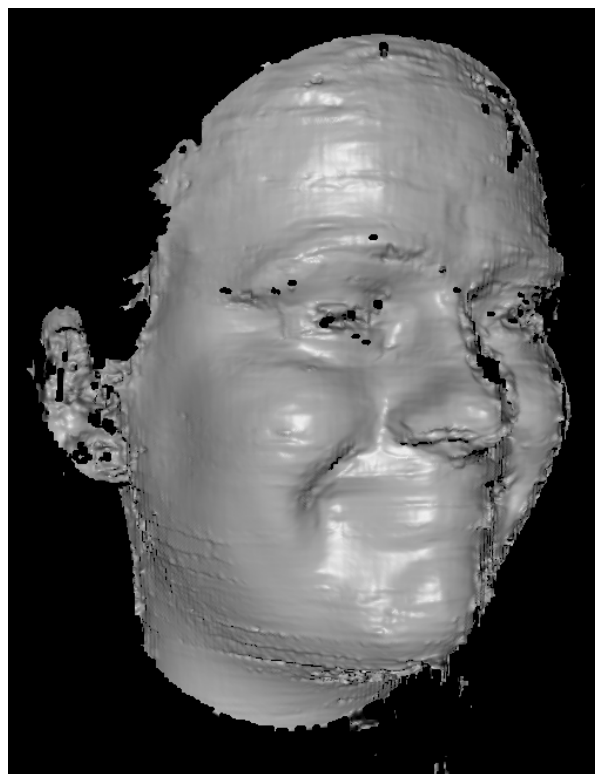
time

time →
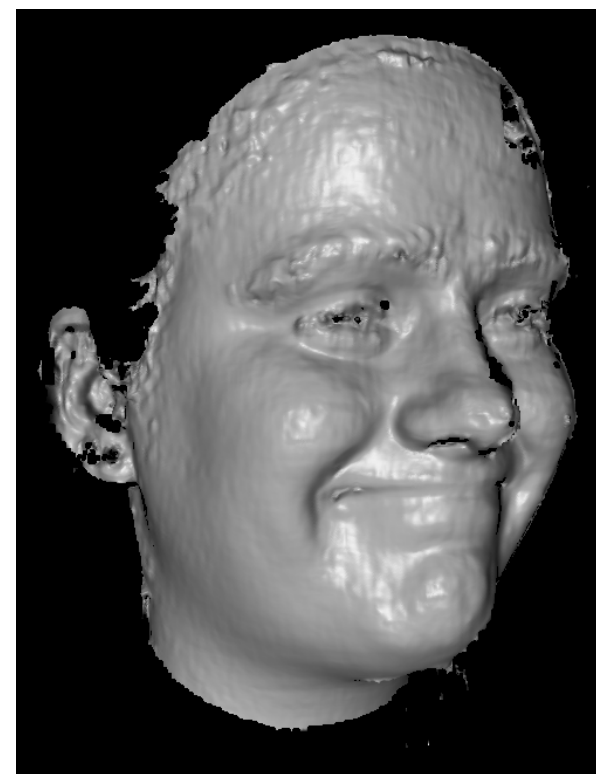
Face surface
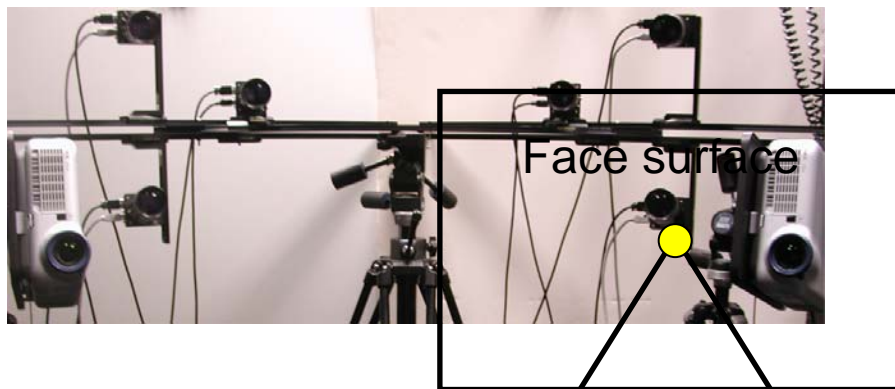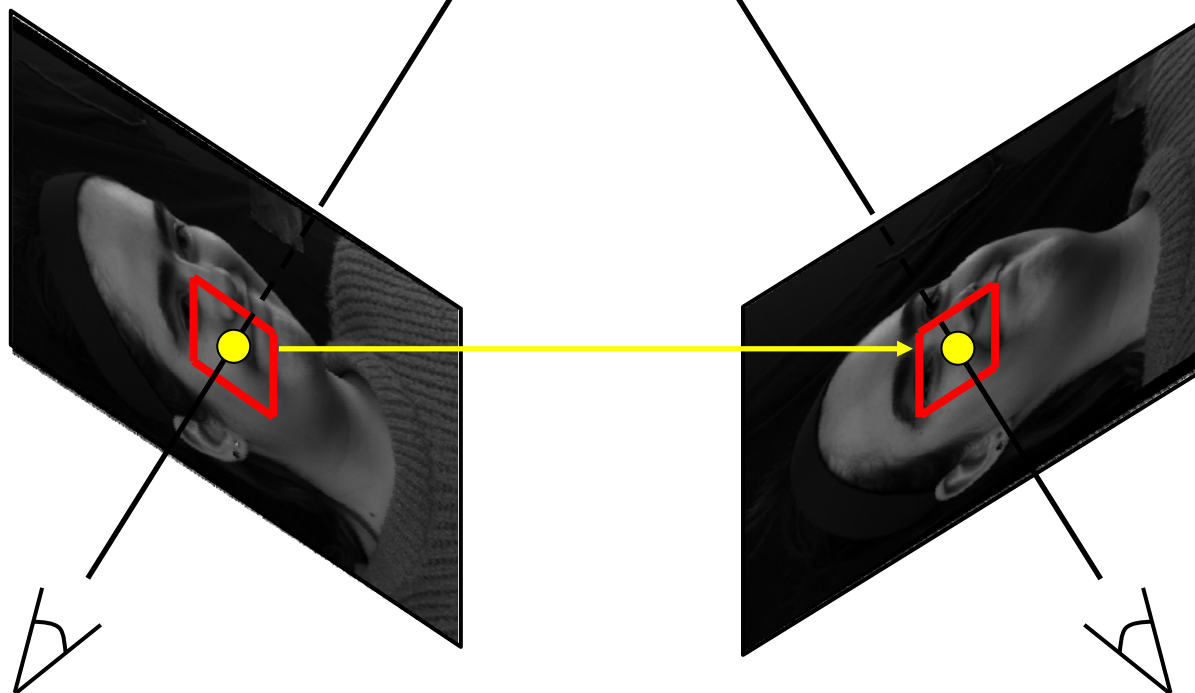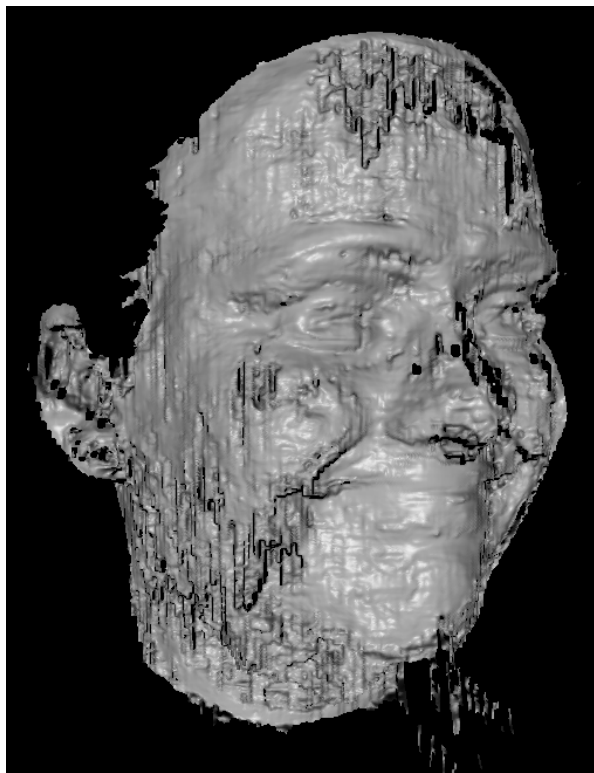
time
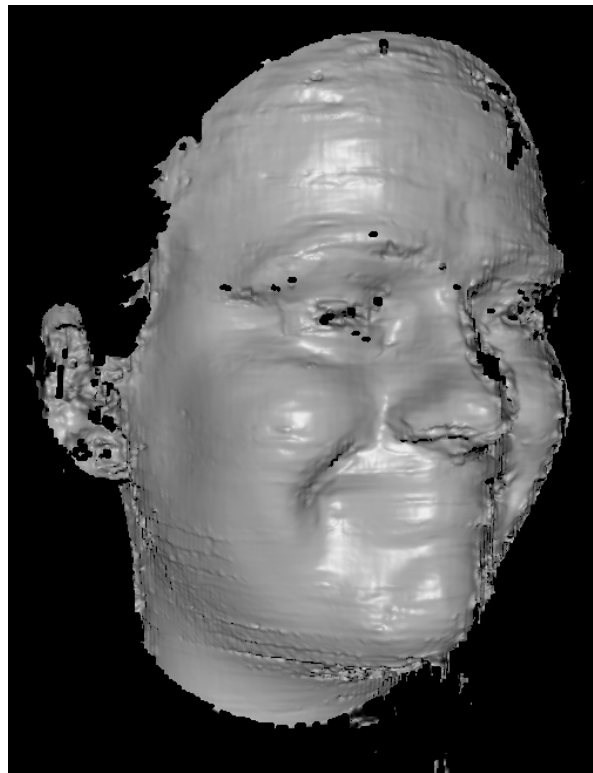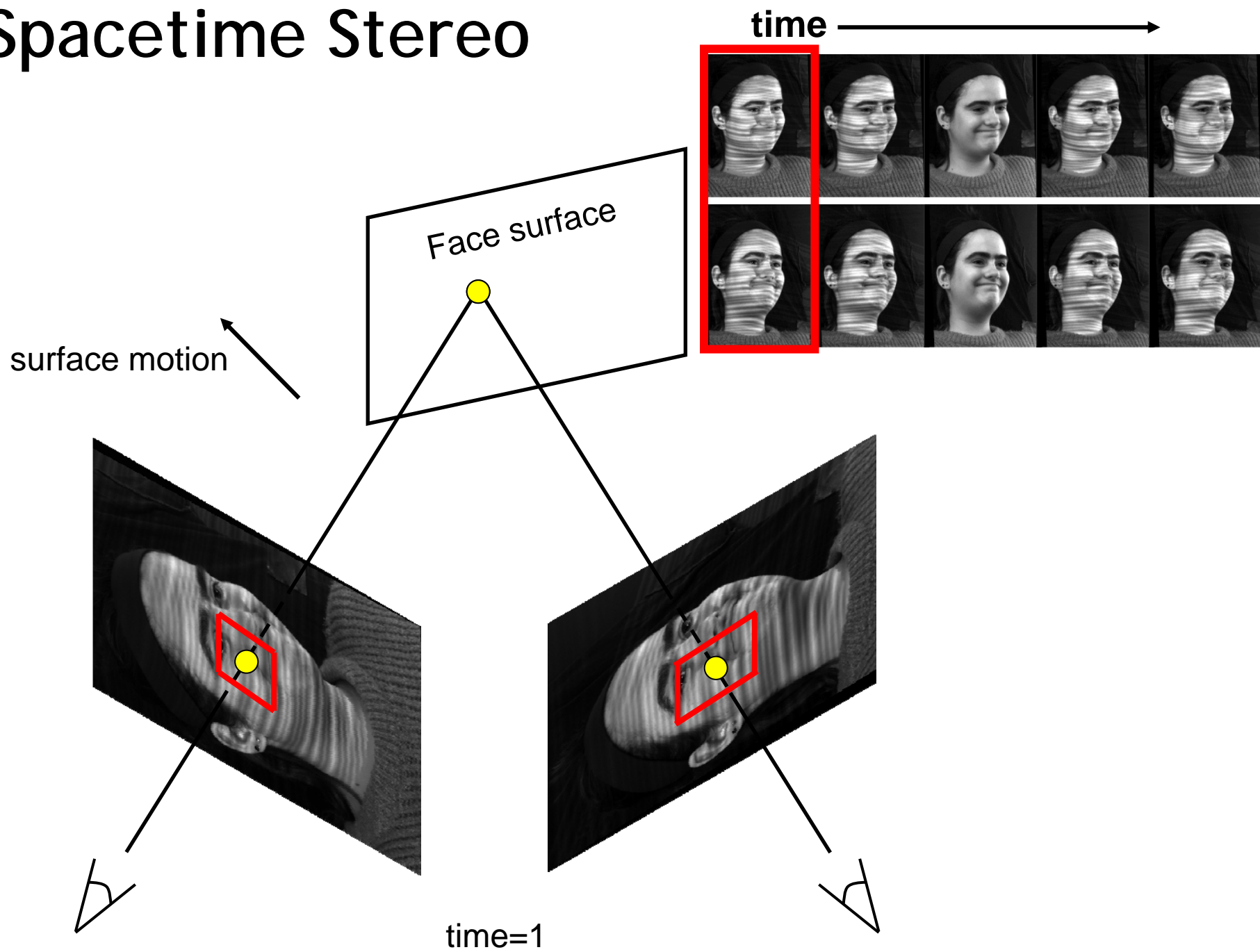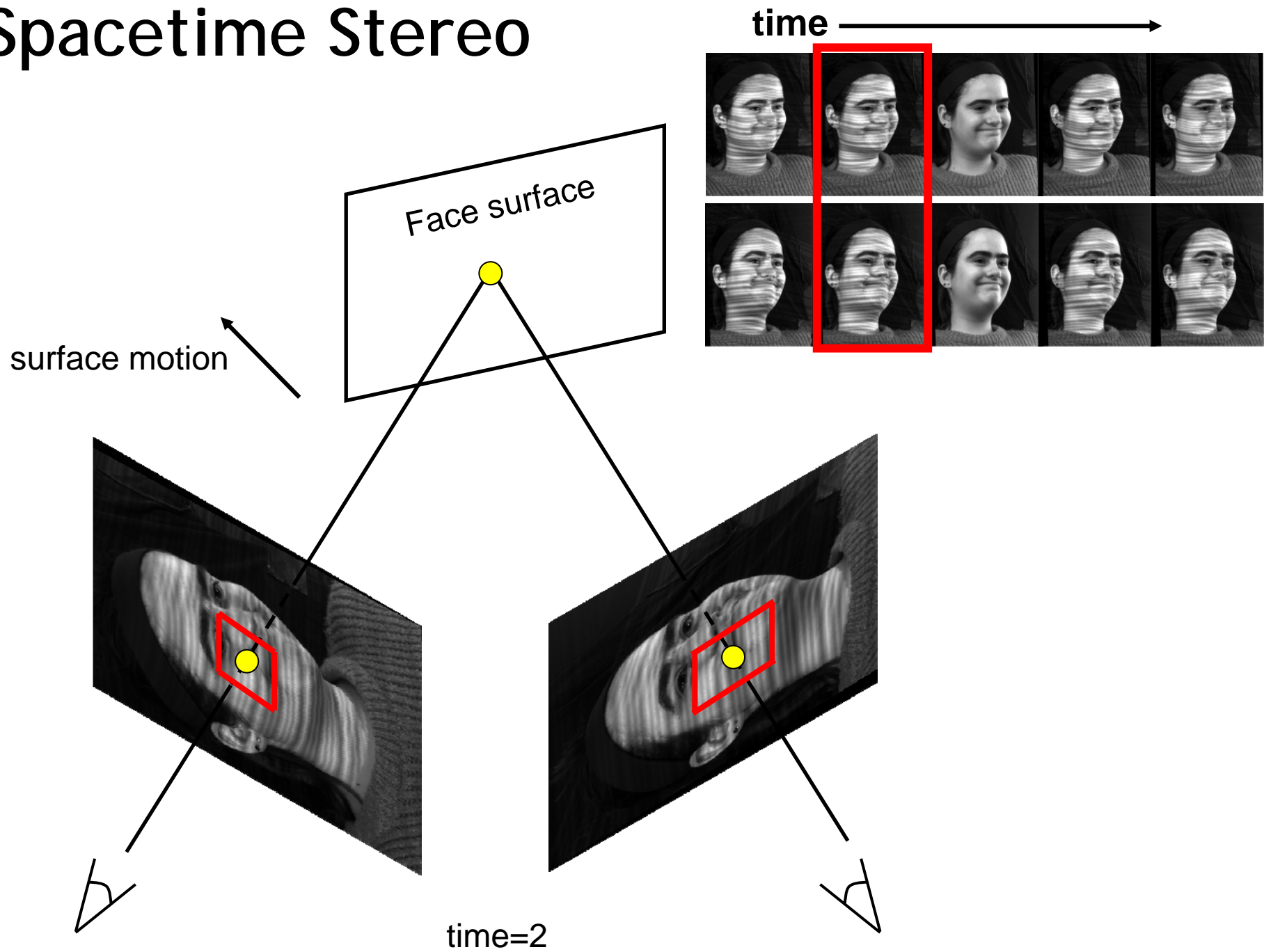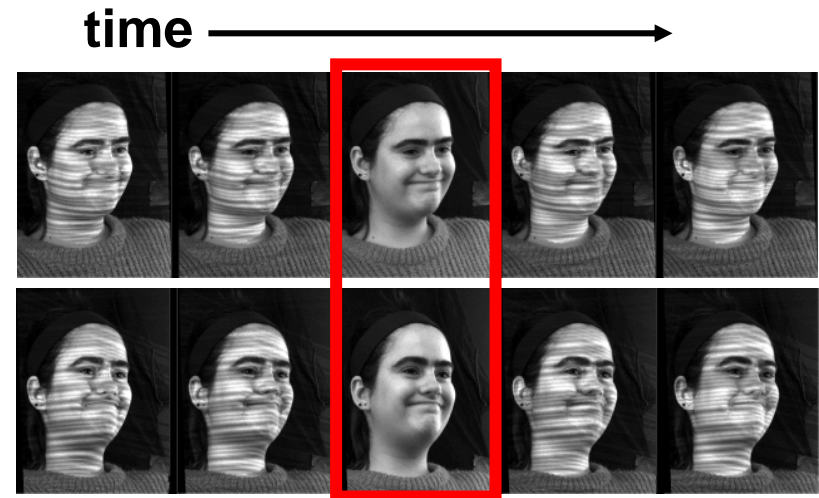
stereo

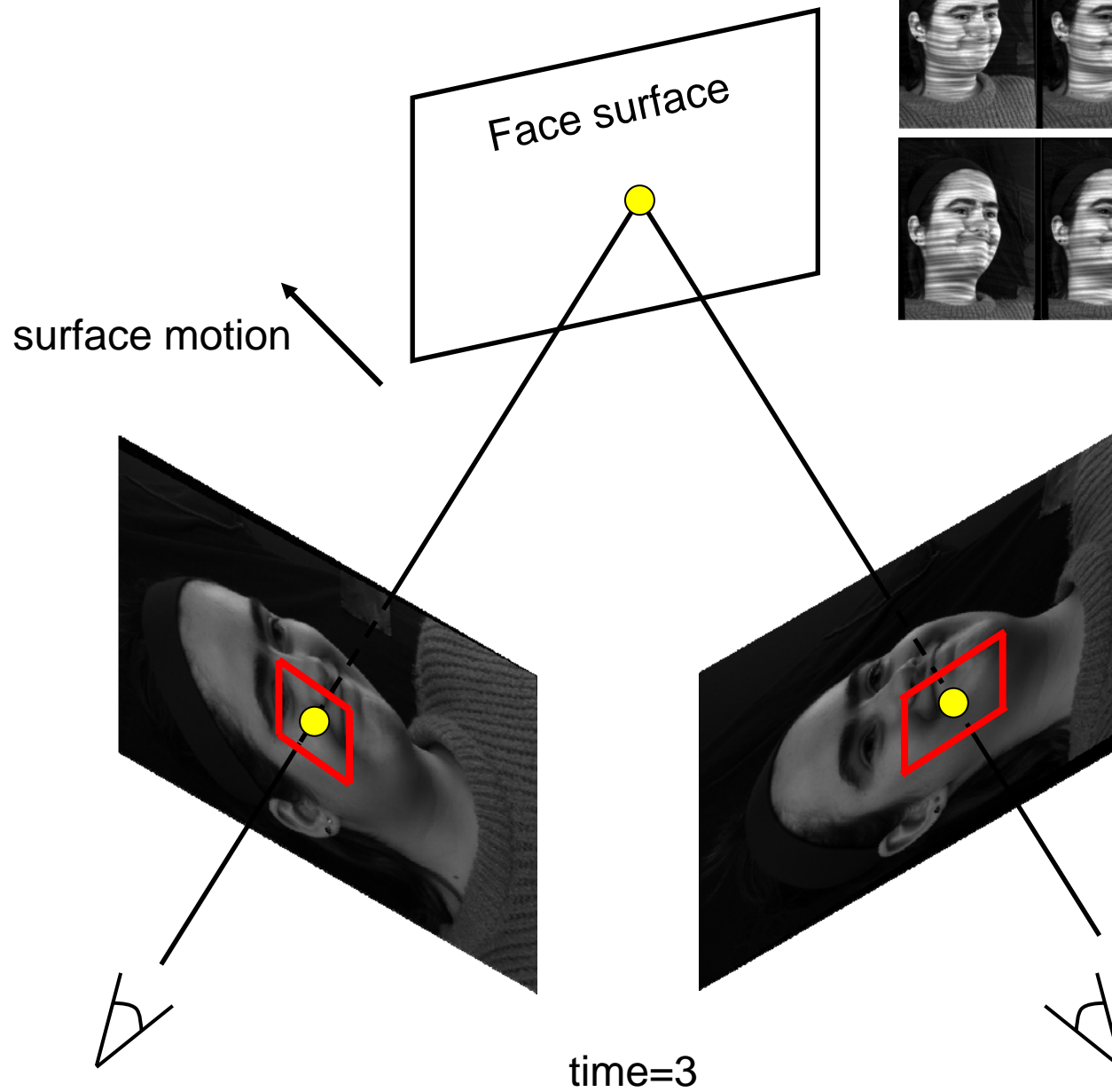time ⟶

stereo

active stereo

time →

stereo | active stereo | spacetime stereo
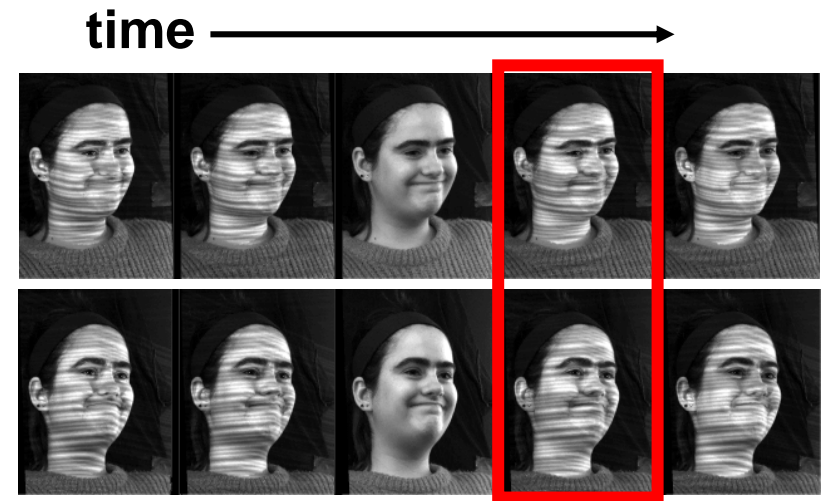
Spacetime Stereo

time

Face surface

surface motion

time=1

Spacetime Stereo

time

Face surface

surface motion

time=2

Spacetime Stereo

Face surface

surface motion

time

time=3

Spacetime Stereo

time

Face surface

surface motion

time=4

Spacetime Stereo

time

Face surface

surface motion

time=5

Spacetime Stereo

time ⟶

surface motion

time

Better
- *spatial resolution*
- *temporal stableness*
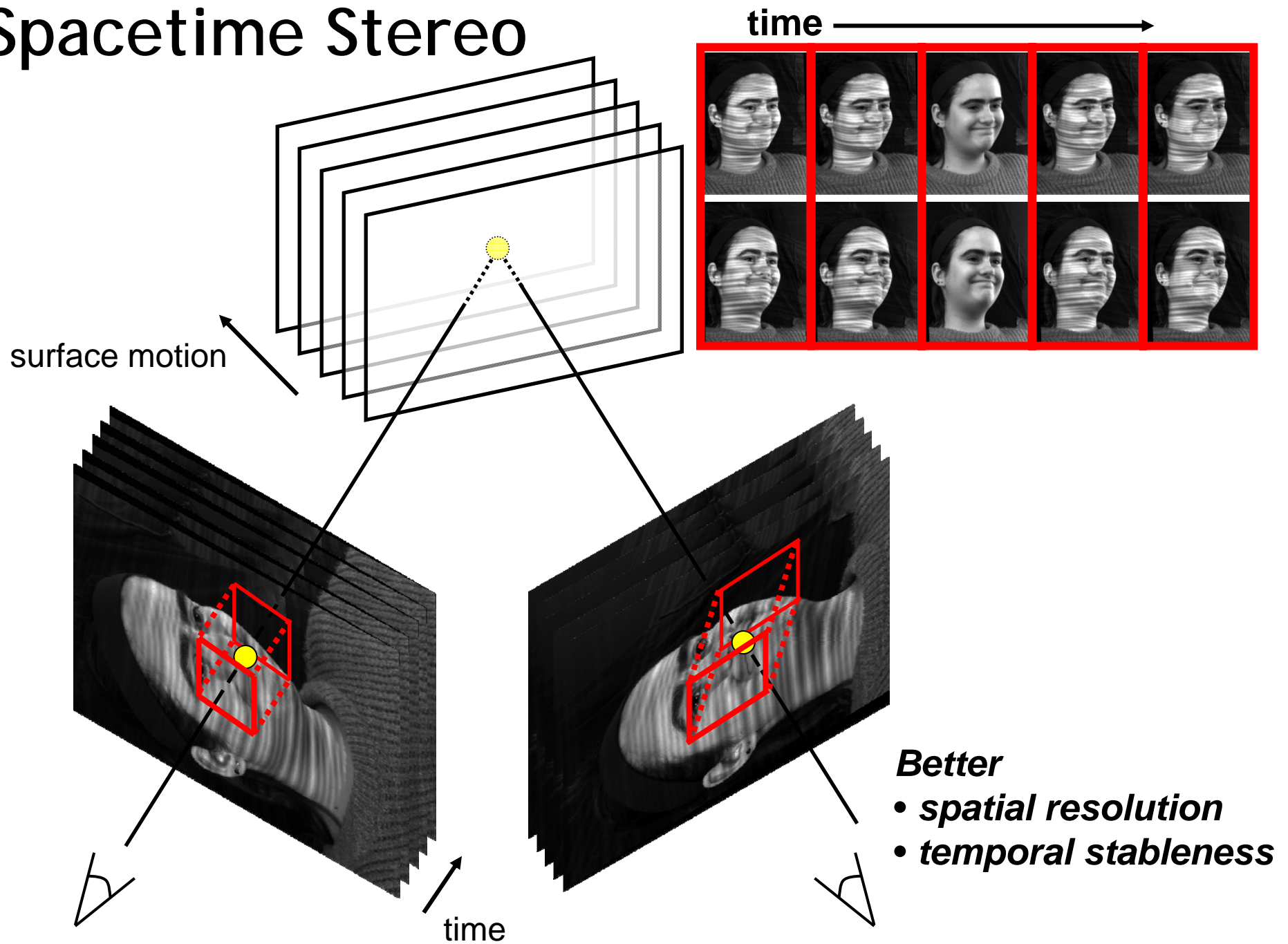
Spacetime stereo matching

A moving oblique surface

Video

Face capture
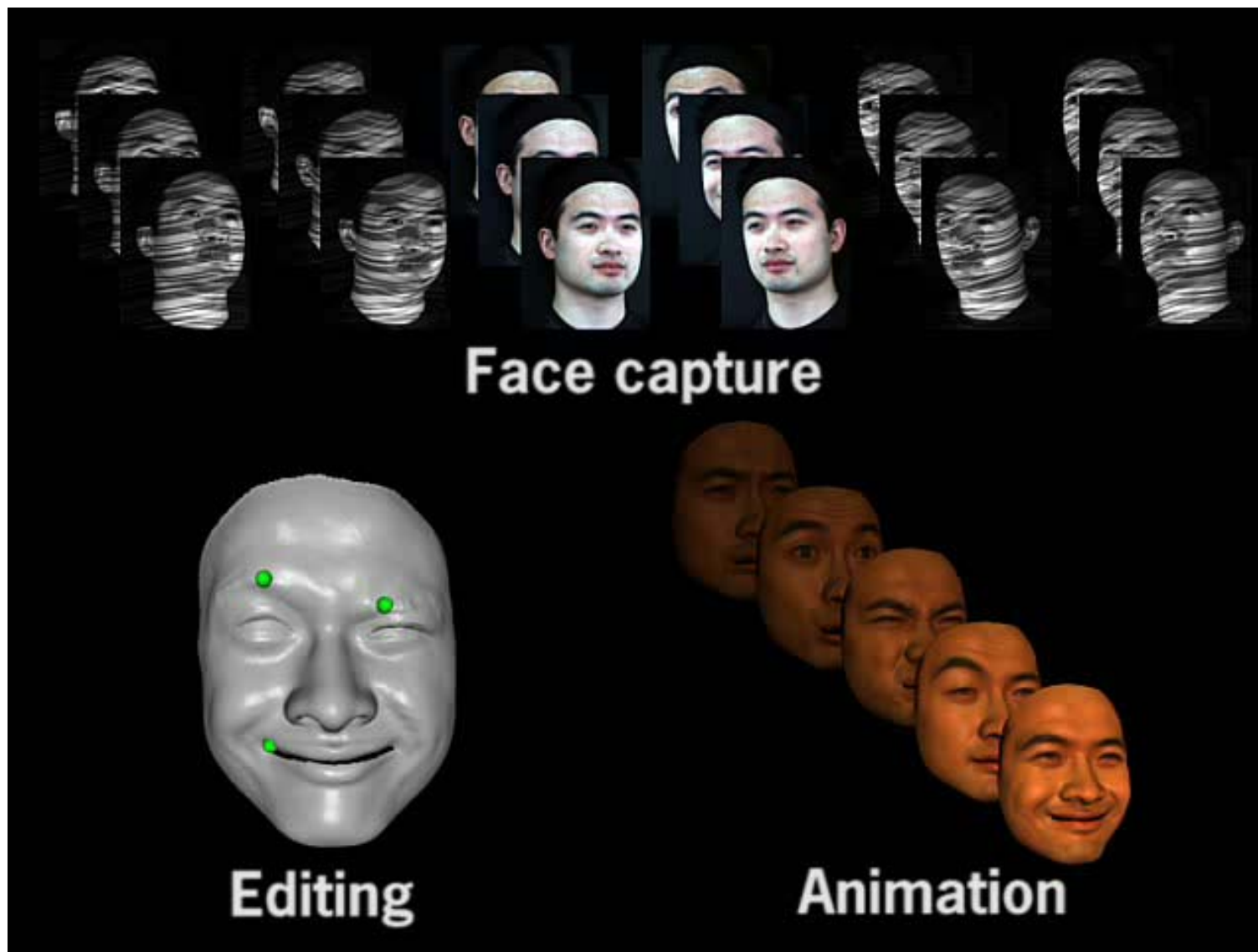
Editing

Animation

Fitting

Face Editing

Animation
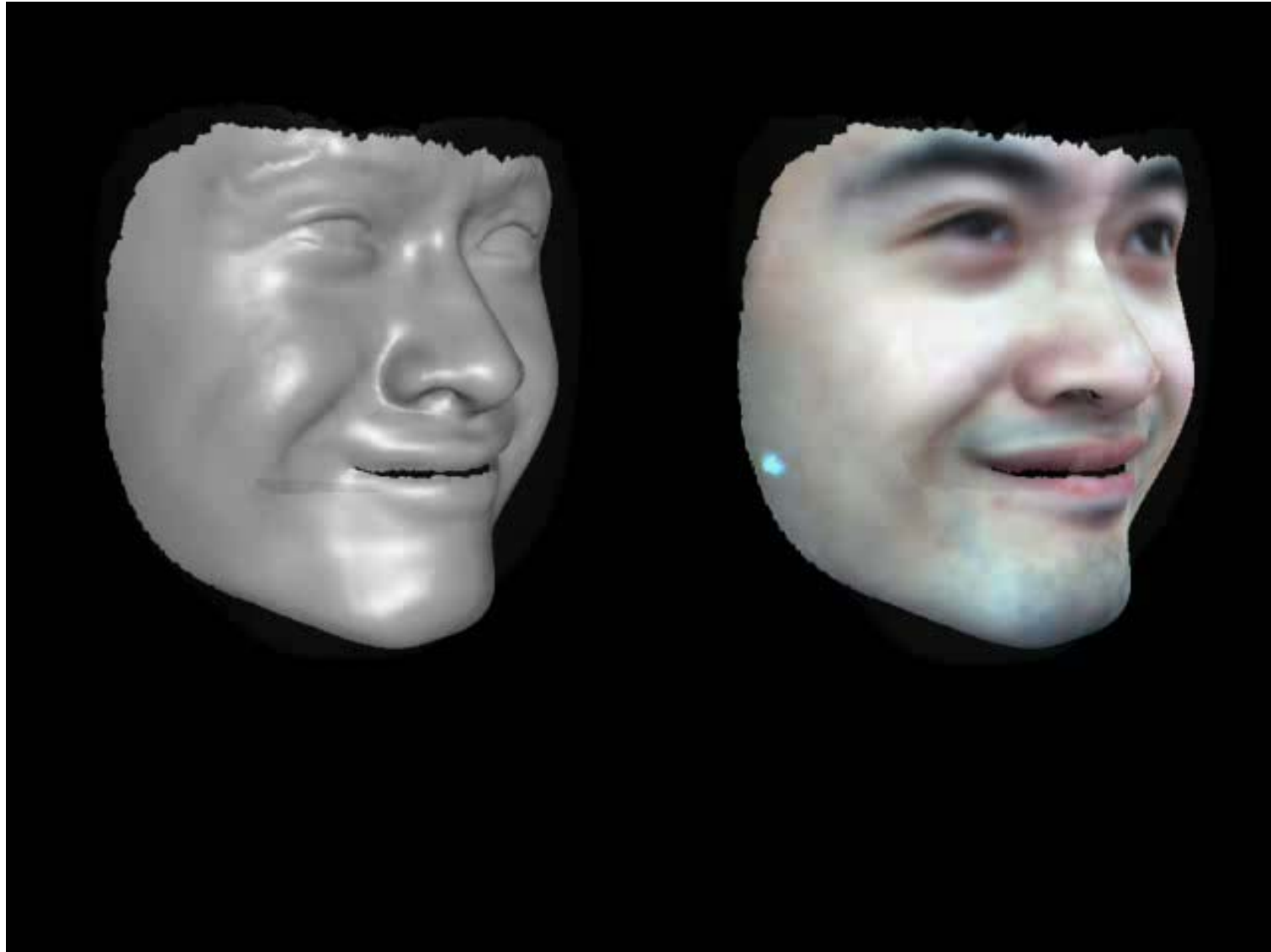
3D face applications: The one

3D face applications: Gladiator
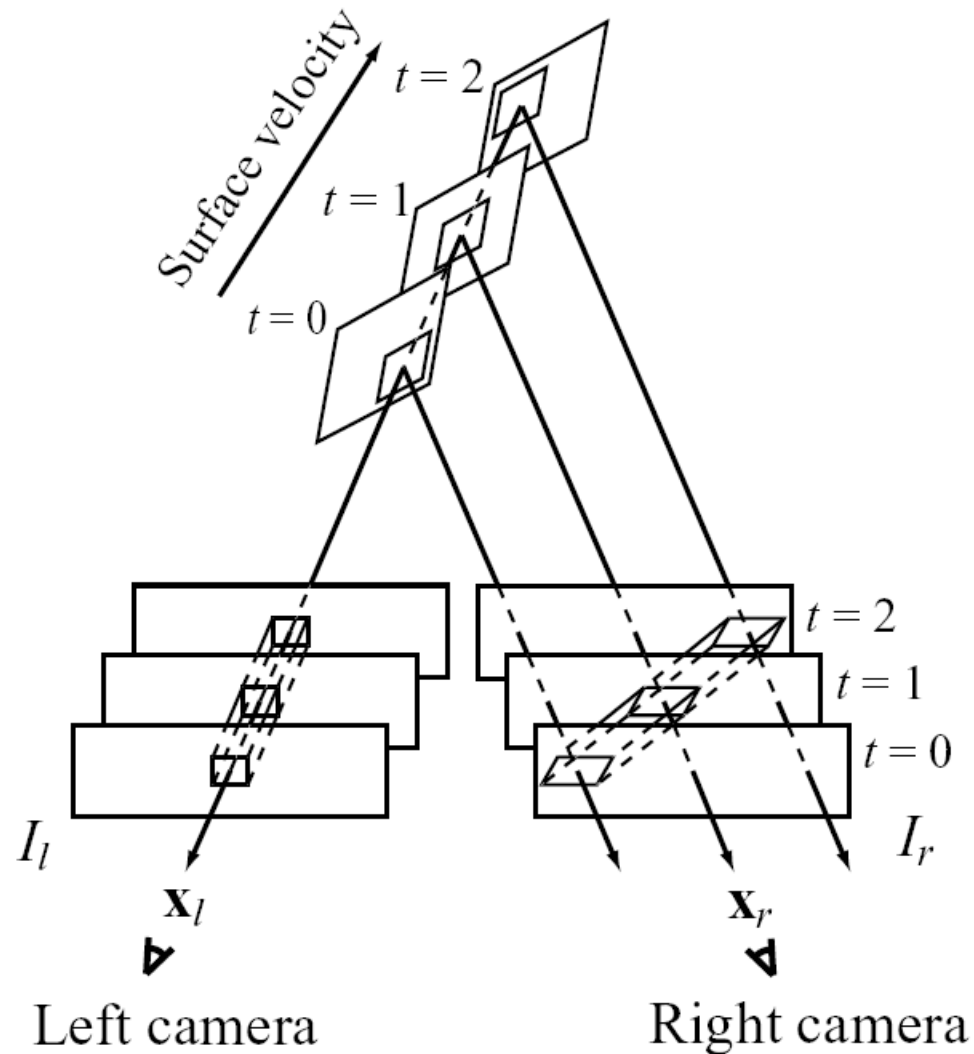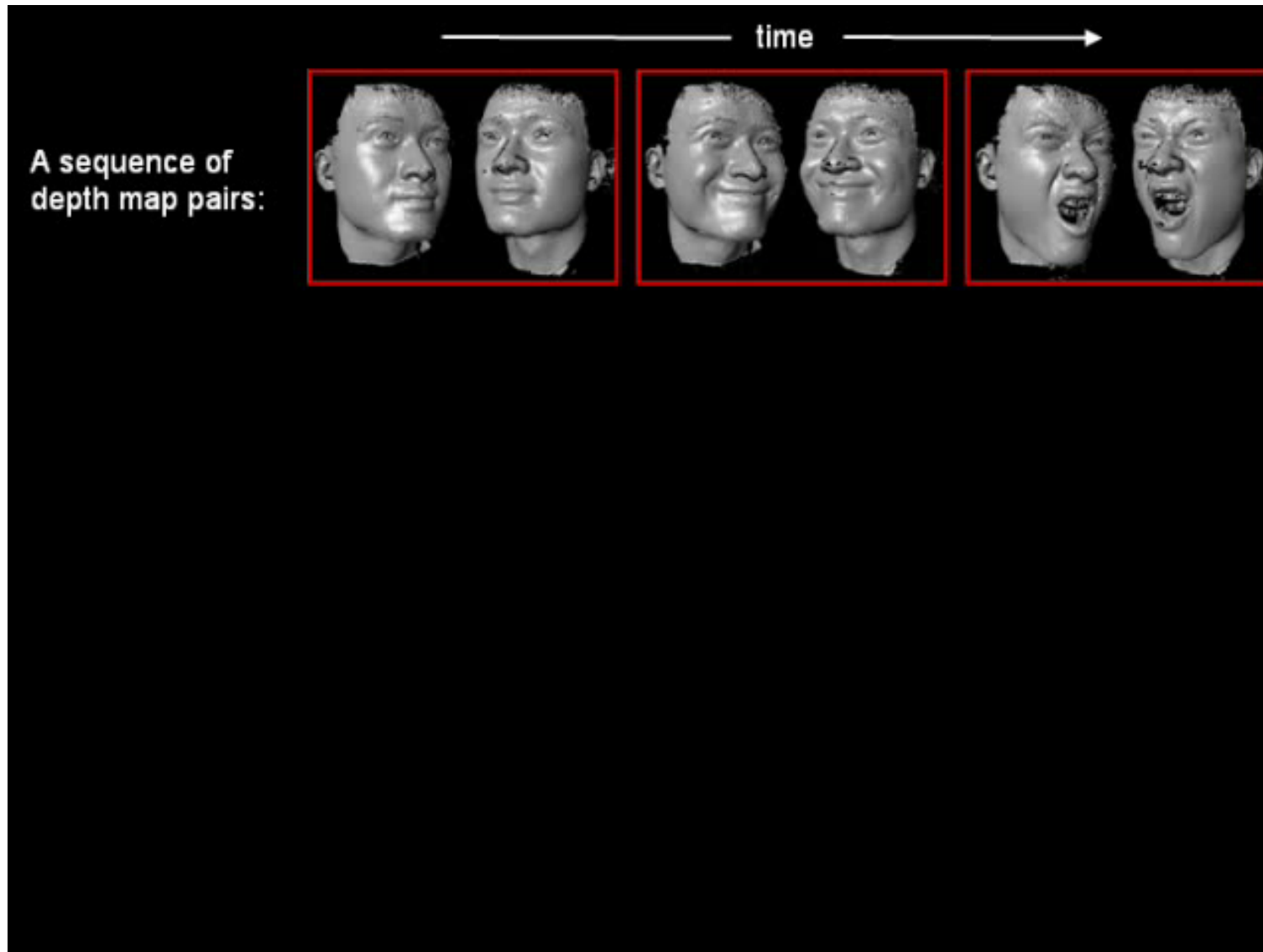

extra 3M

Statistical methods

Statistical methods

para-meters $\quad z \longrightarrow \boxed{\text{f(z)}+\varepsilon} \longrightarrow y \quad$ observed signal
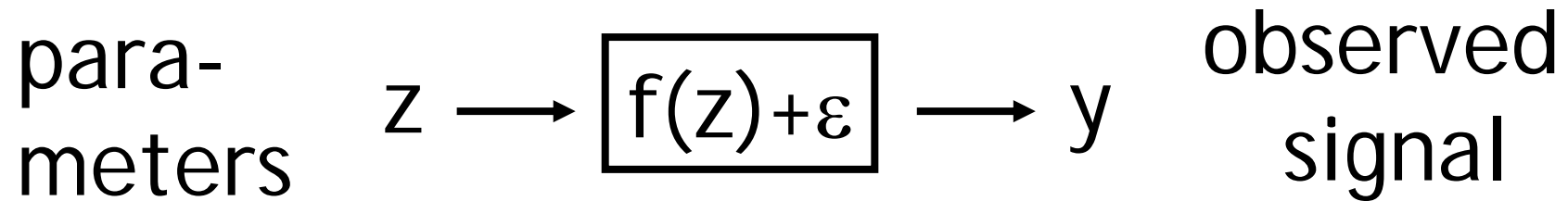
$$z^* = \max_z P(z \mid y)$$

$$= \max_z \frac{P(y \mid z)P(z)}{P(y)}$$

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

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

Statistical methods

DigiVFX

para-meters
$z \longrightarrow \boxed{f(z) + \varepsilon} \longrightarrow y$
observed signal
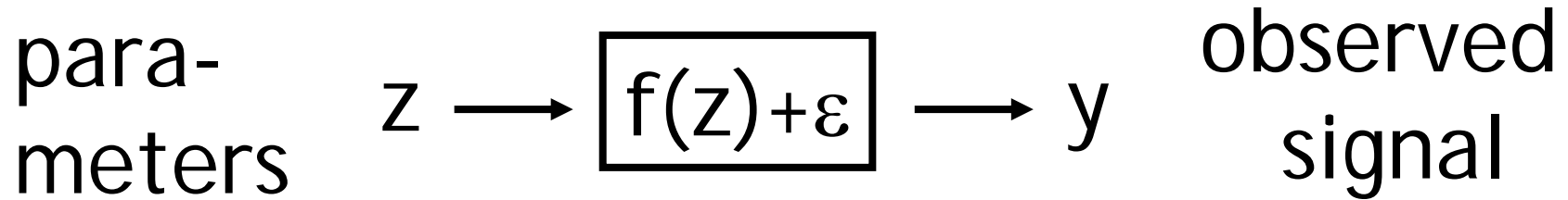
$$z^* = \min_z L(y \mid z) + L(z)$$

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

a-priori knowledge

Statistical methods

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

Takeo Kanade

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

Generic priors

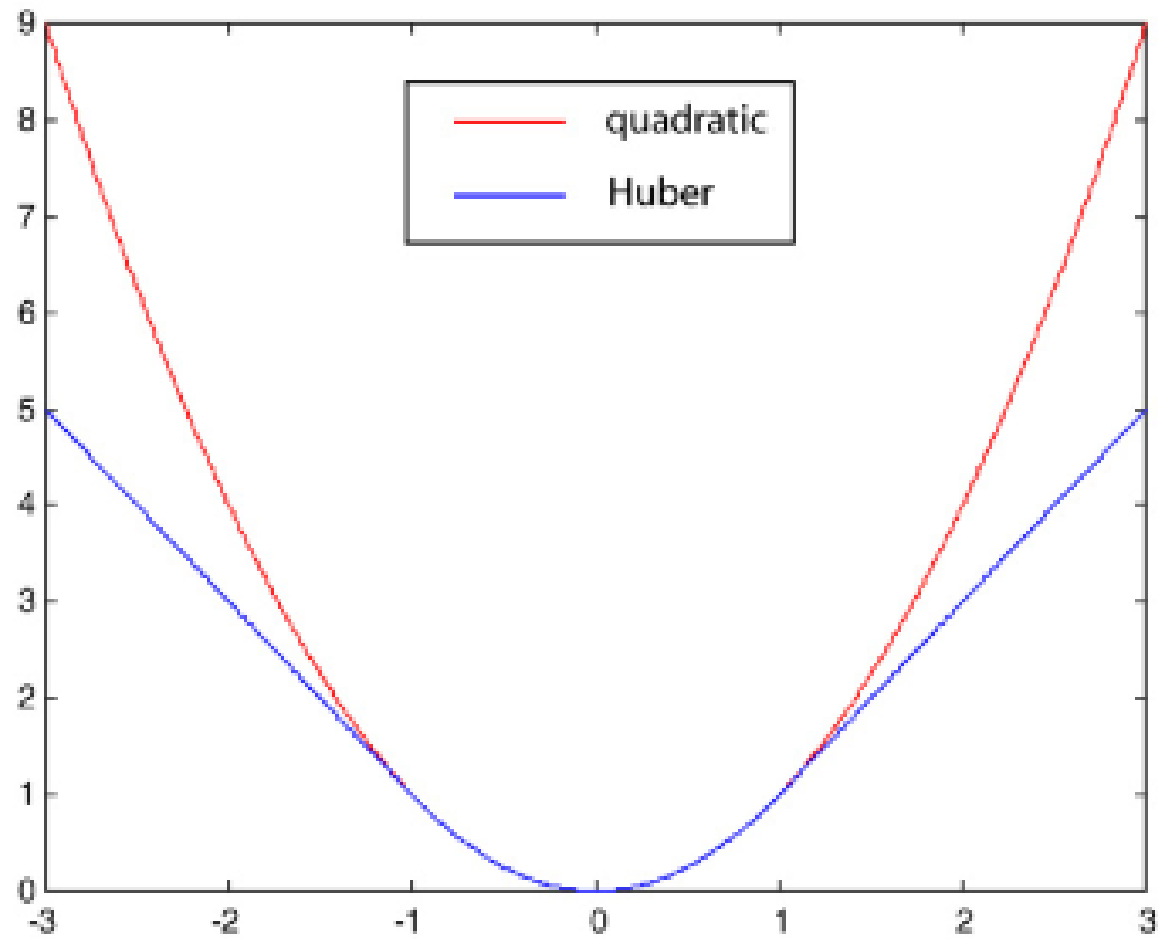"Smooth images are good images."

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

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

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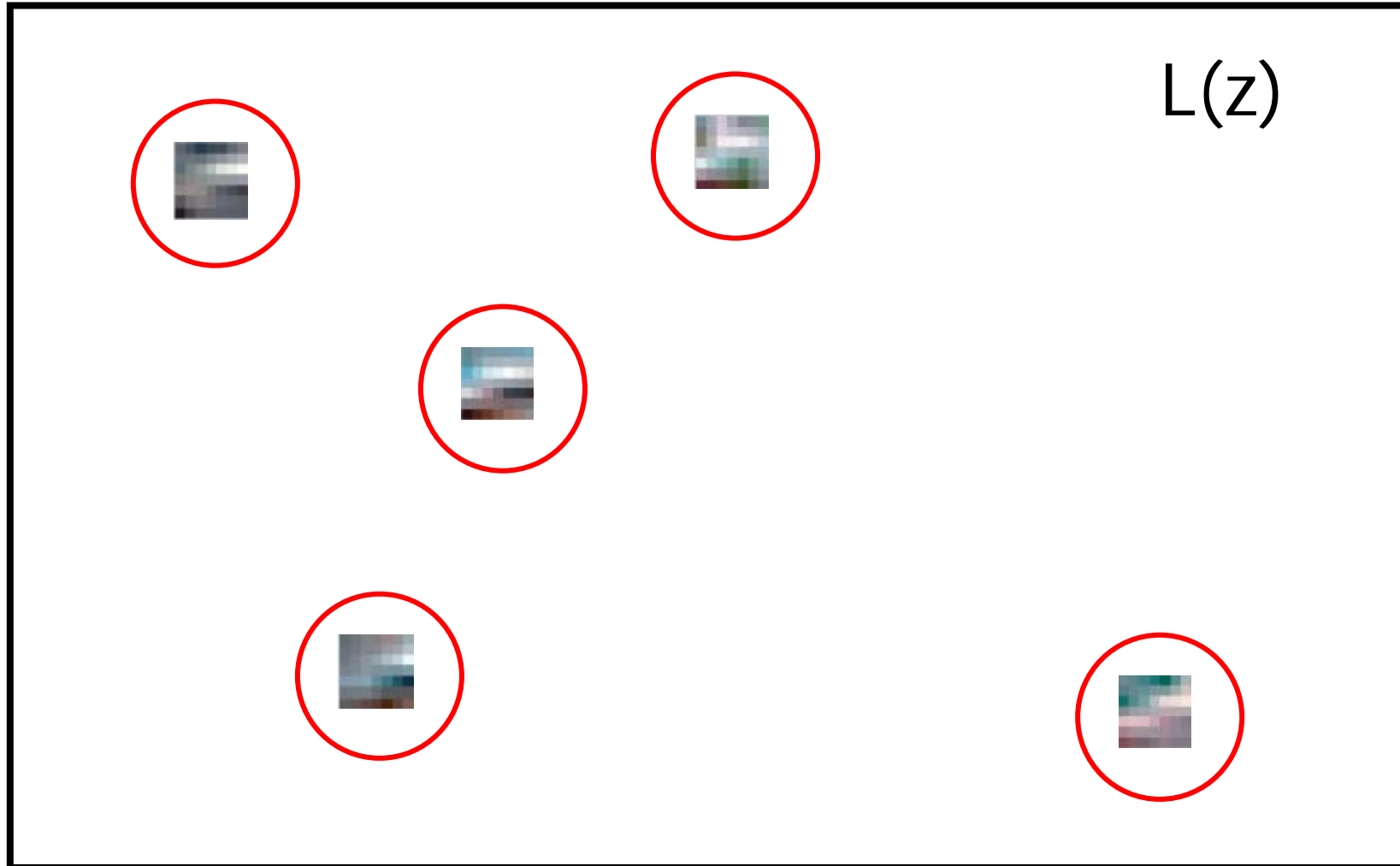
Generic priors

Example-based priors
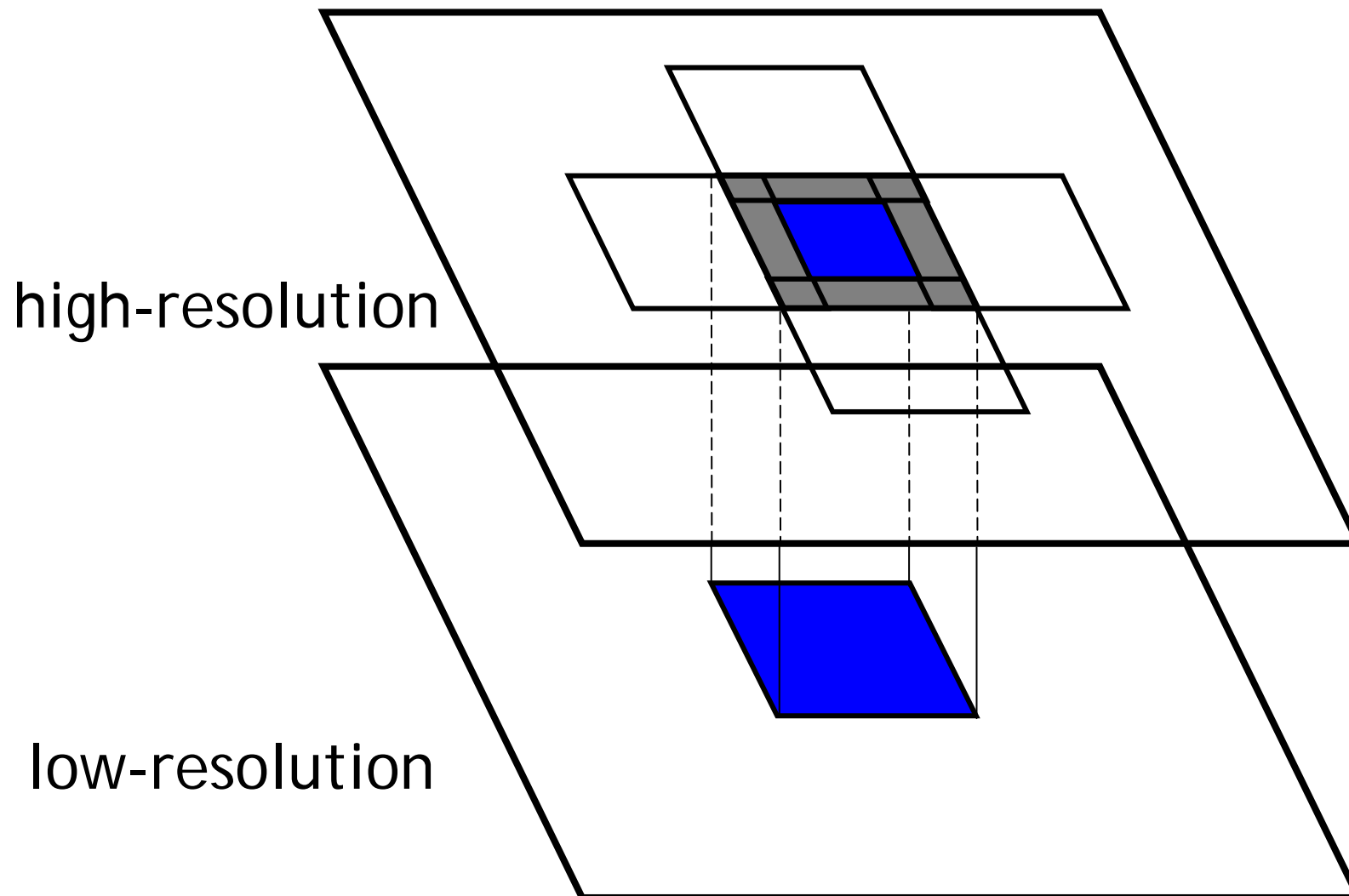
"Existing images are good images."

six 200×200
Images $\Rightarrow$
2,000,000
pairs

Example-based priors

Example-based priors

high-resolution

low-resolution

Model-based priors

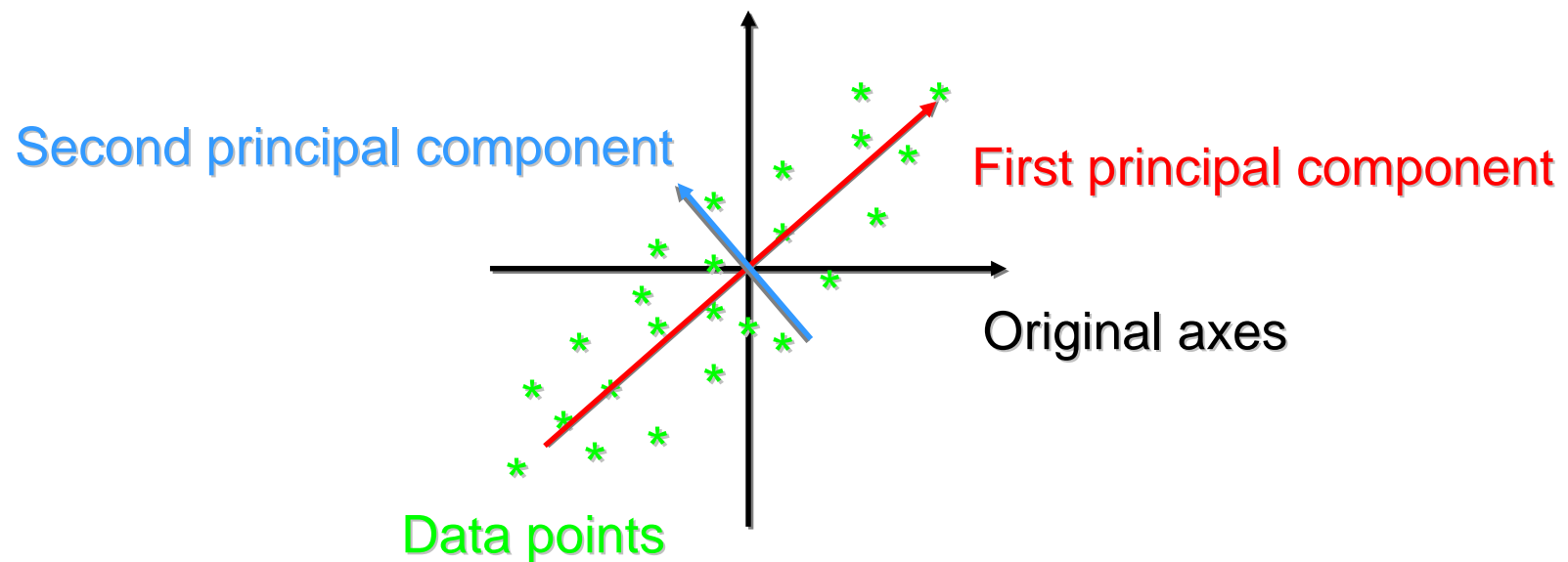"Face images are good images when working on face images ..."

Parametric model

$Z=WX+\mu$
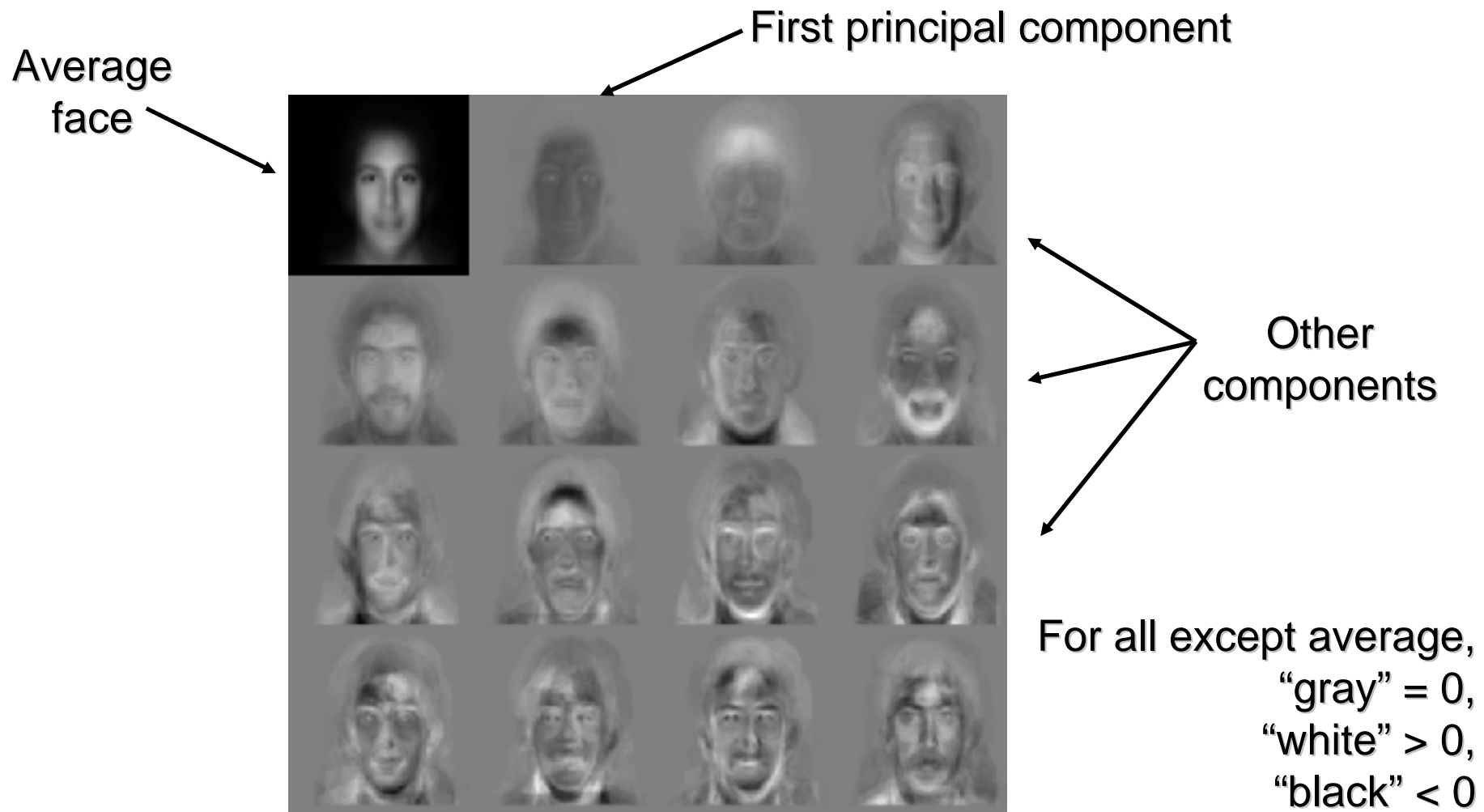
$L(X)$

$$z^* = \min_{z} L(y \mid z) + L(z)$$

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

PCA

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

Second principal component

First principal component

Original axes

Data points

PCA on faces: "eigenfaces"

DigiVFX

First principal component

Average face

Other components

For all except average,
"gray" = 0,
"white" > 0,
"black" < 0

Model-based priors

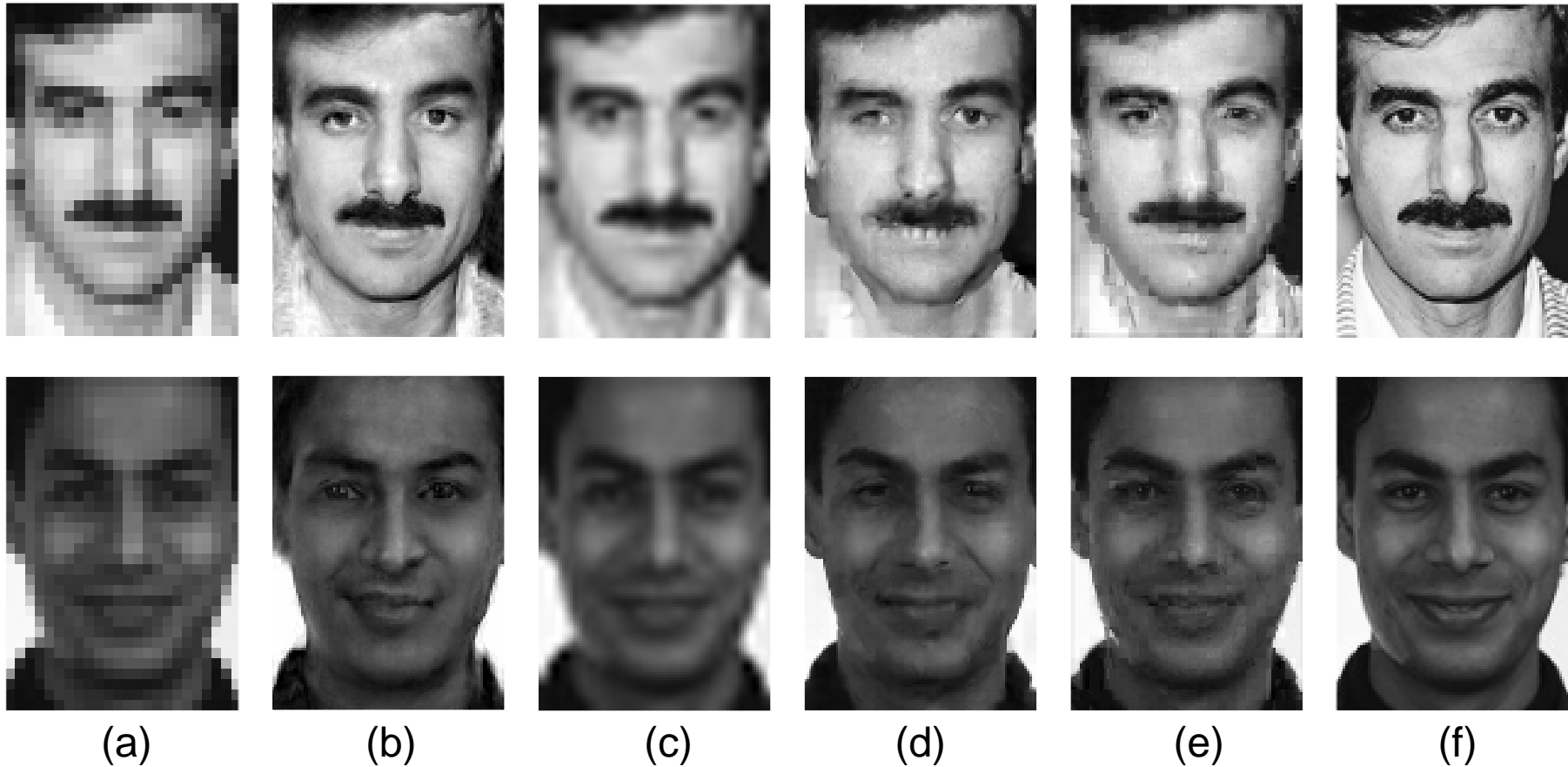"Face images are good images when working on face images ..."

Parametric model

$$Z = WX + \mu \qquad L(X)$$

$$z^* = \min_{z} L(y \mid z) + L(z)$$

$$\begin{cases} X^* = \min_{x} L(y \mid WX + \mu) + L(X) \\ z^* = WX^* + \mu \end{cases}$$
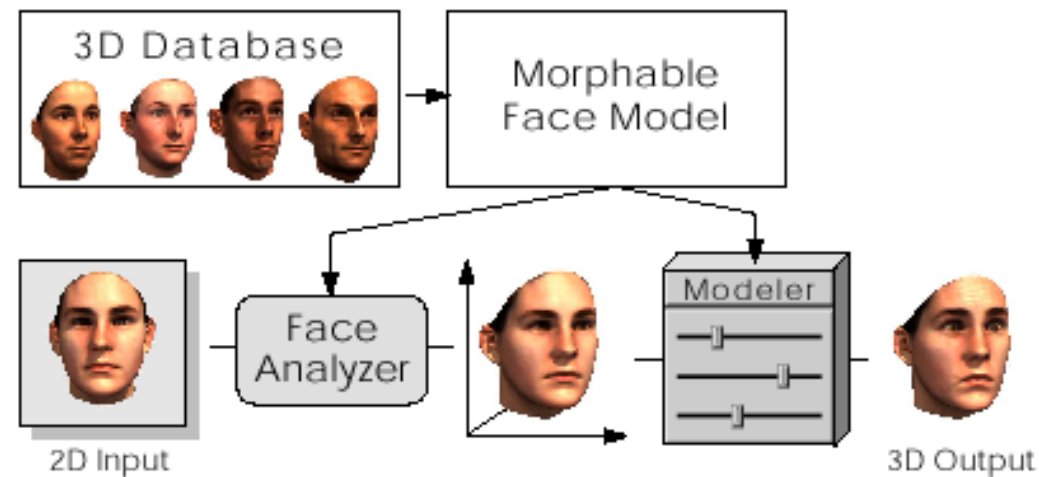
Super-resolution

(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
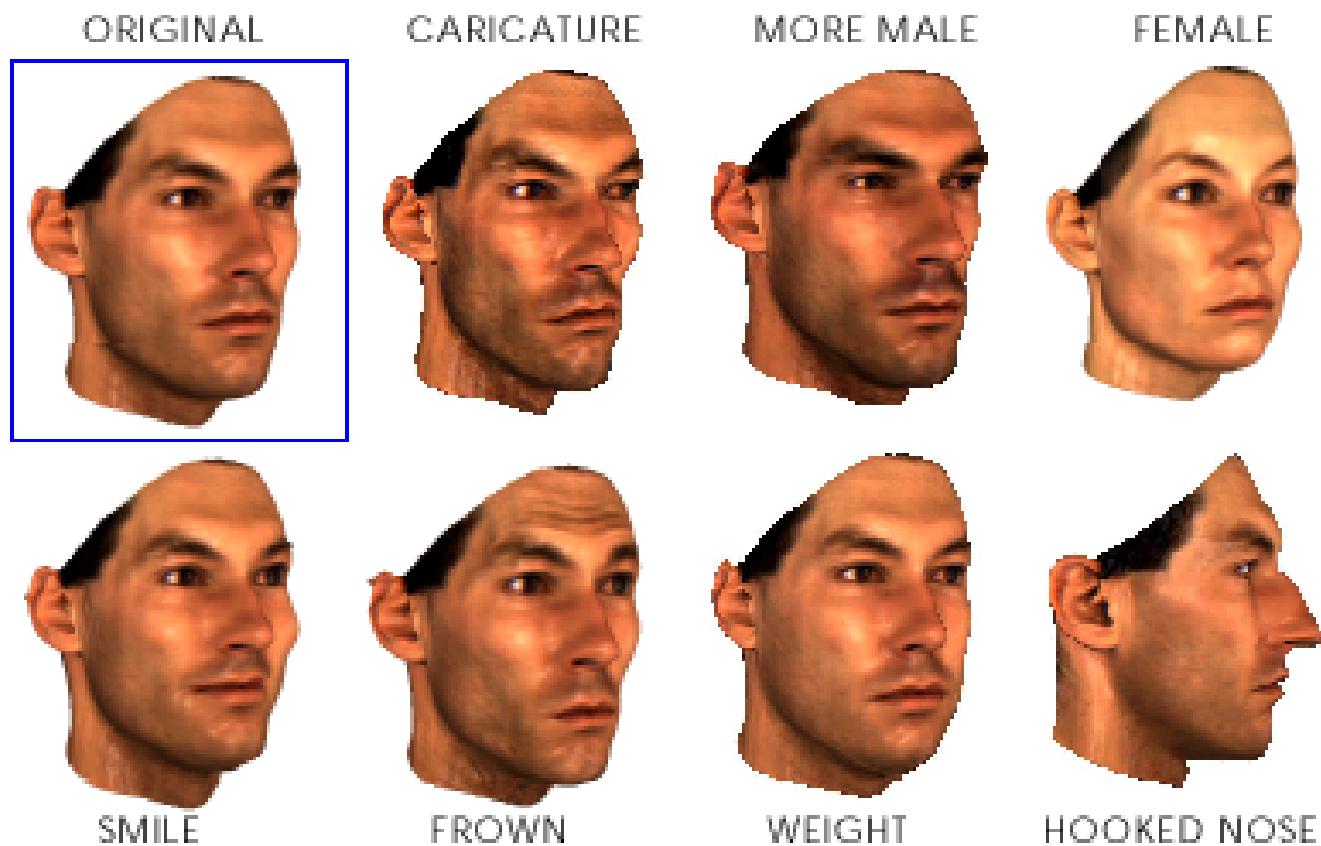
Morphable model

shape examplars   texture examplars

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

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

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

Morphable model of 3D faces

- Adding some variations

Reconstruction from single image
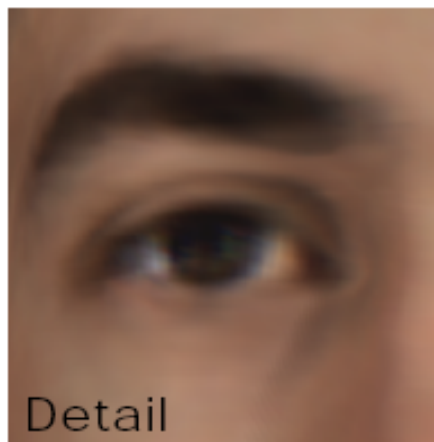
Reconstruction from single image

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

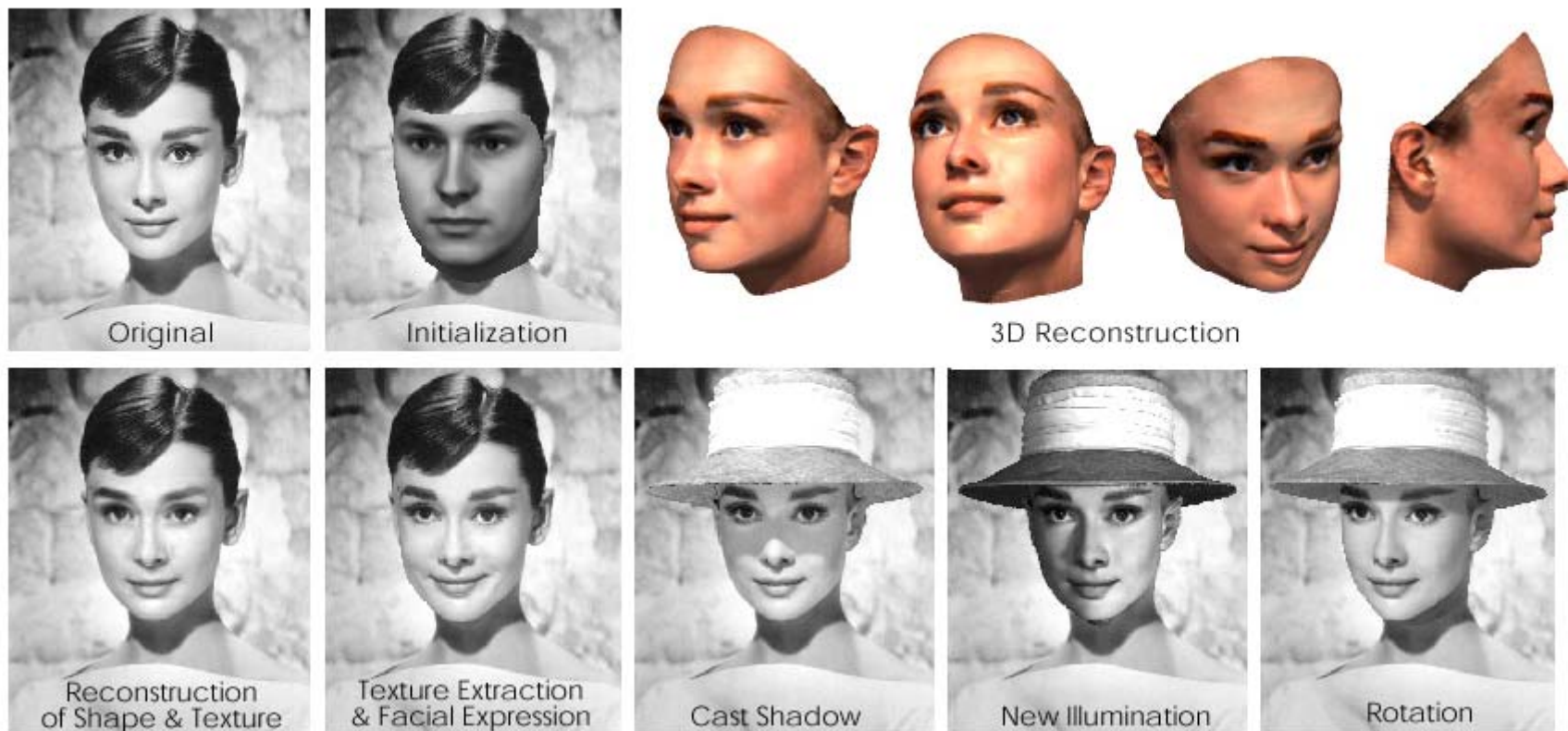$$E_I = \sum_{x,y} \|\mathbf{I}_{input}(x,y) - \mathbf{I}_{model}(x,y)\|^2$$

shape and texture priors are learnt from database

ρ is the set of parameters for shading including camera pose, lighting and so on

Modifying a single image
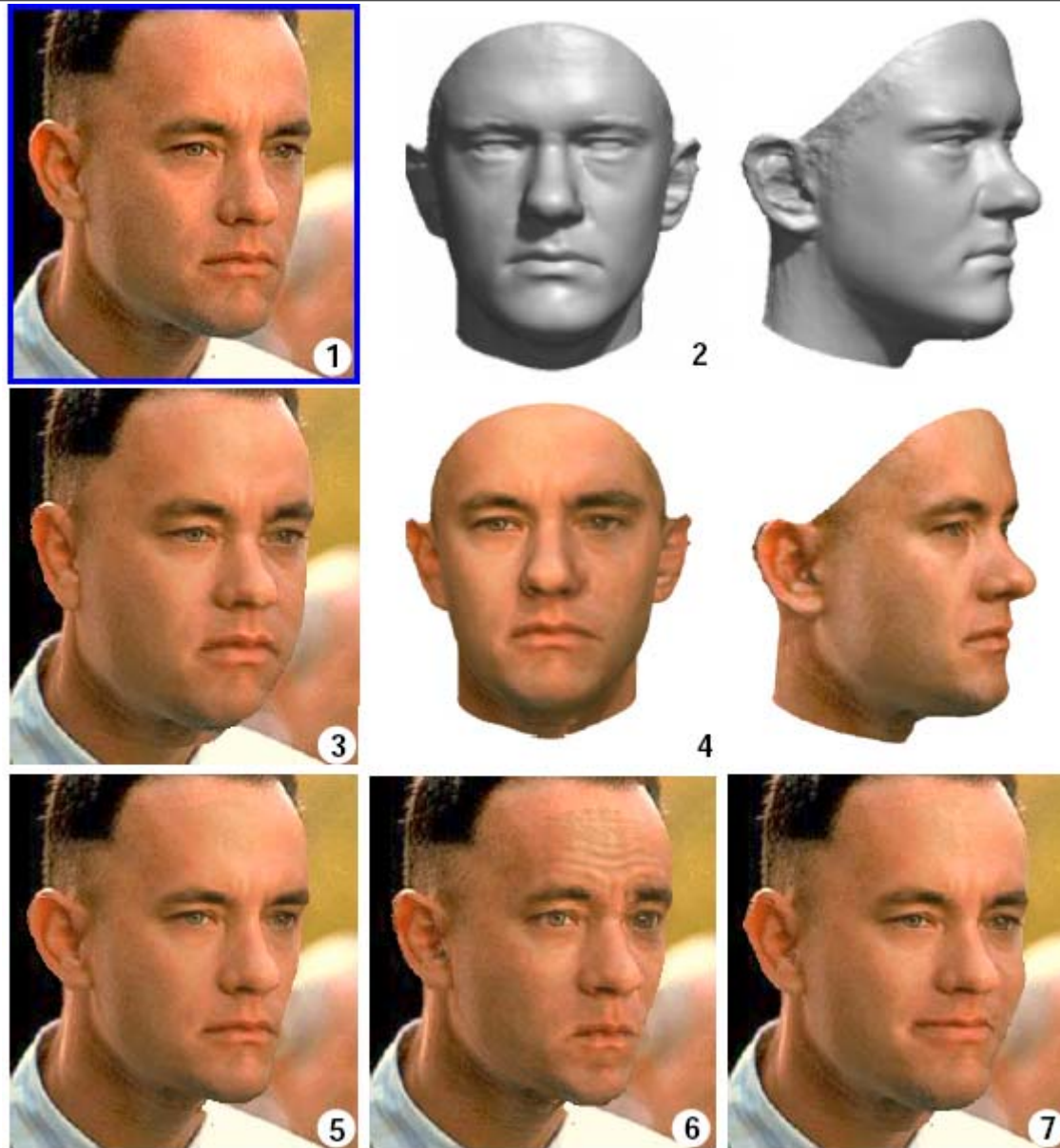
Animating from a single image

Video

Exchanging faces in images
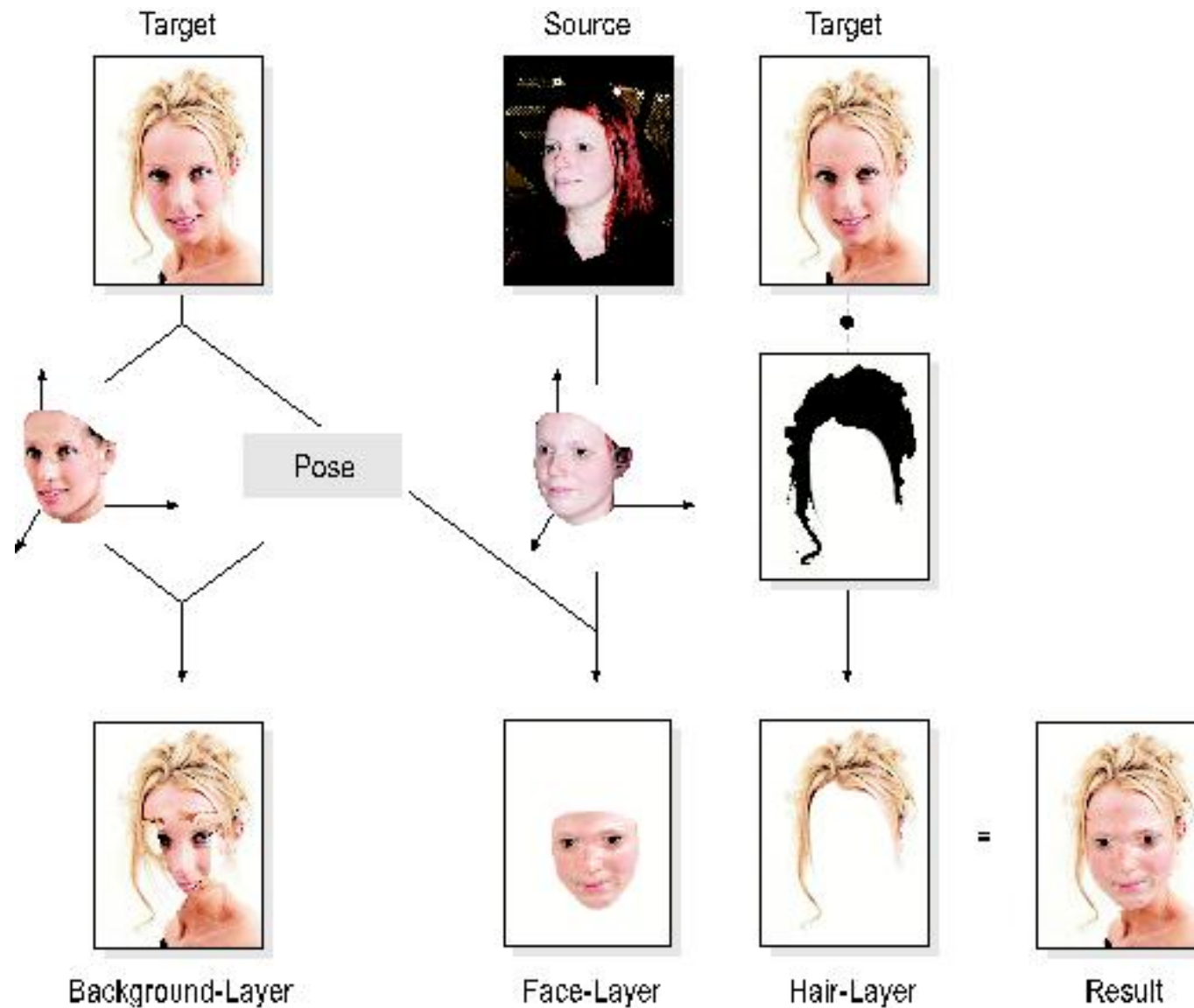
Exchange faces in images
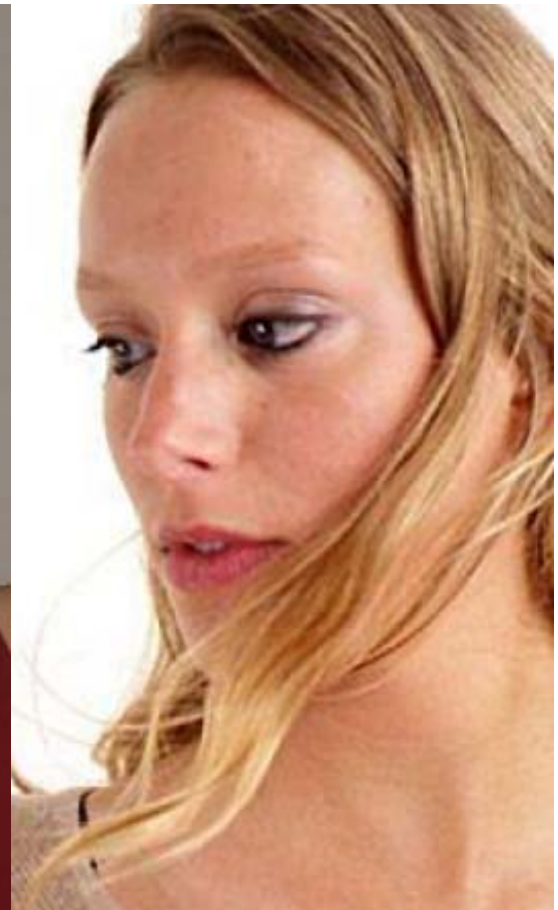
Exchange faces in images

Exchange faces in images

Exchange faces in images
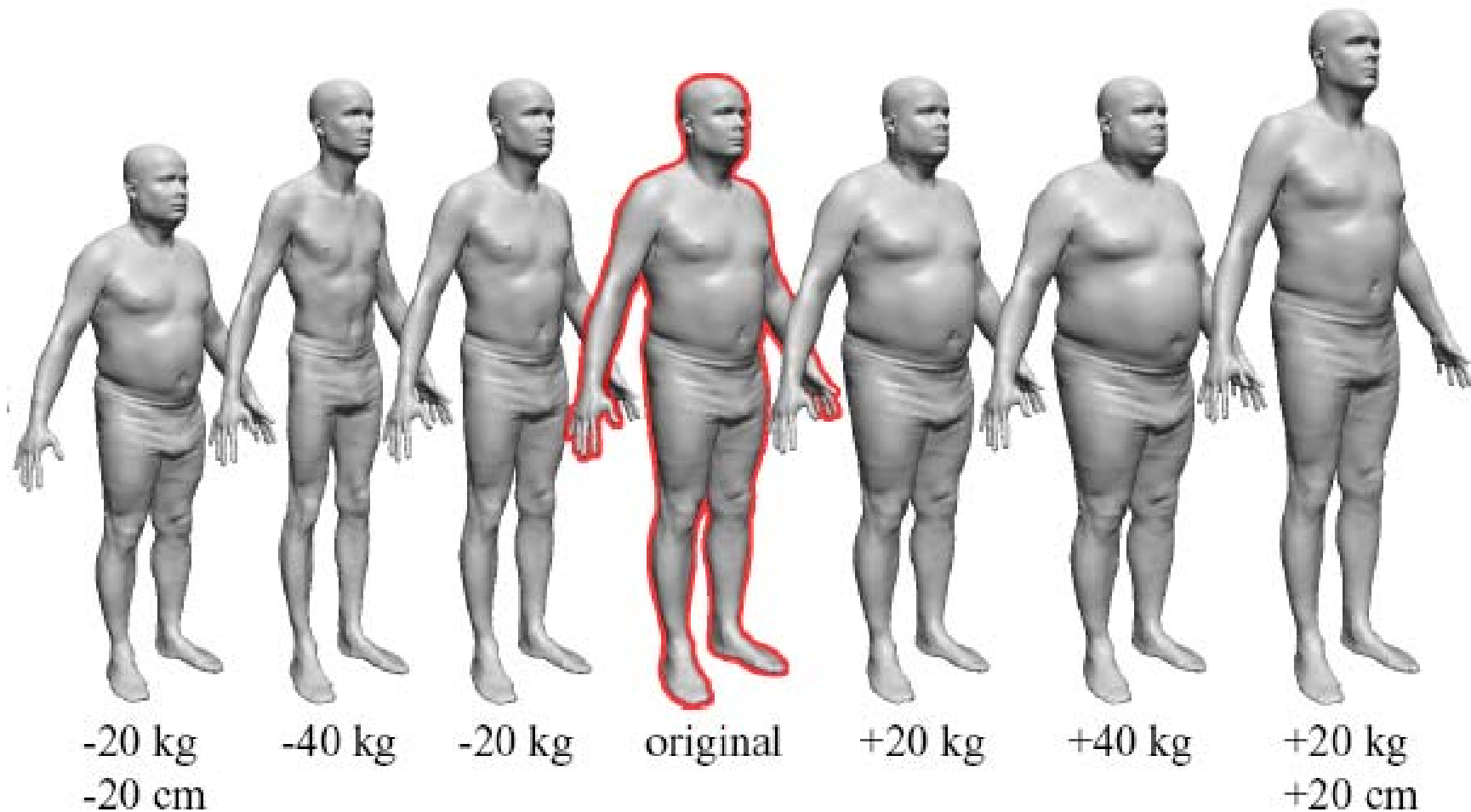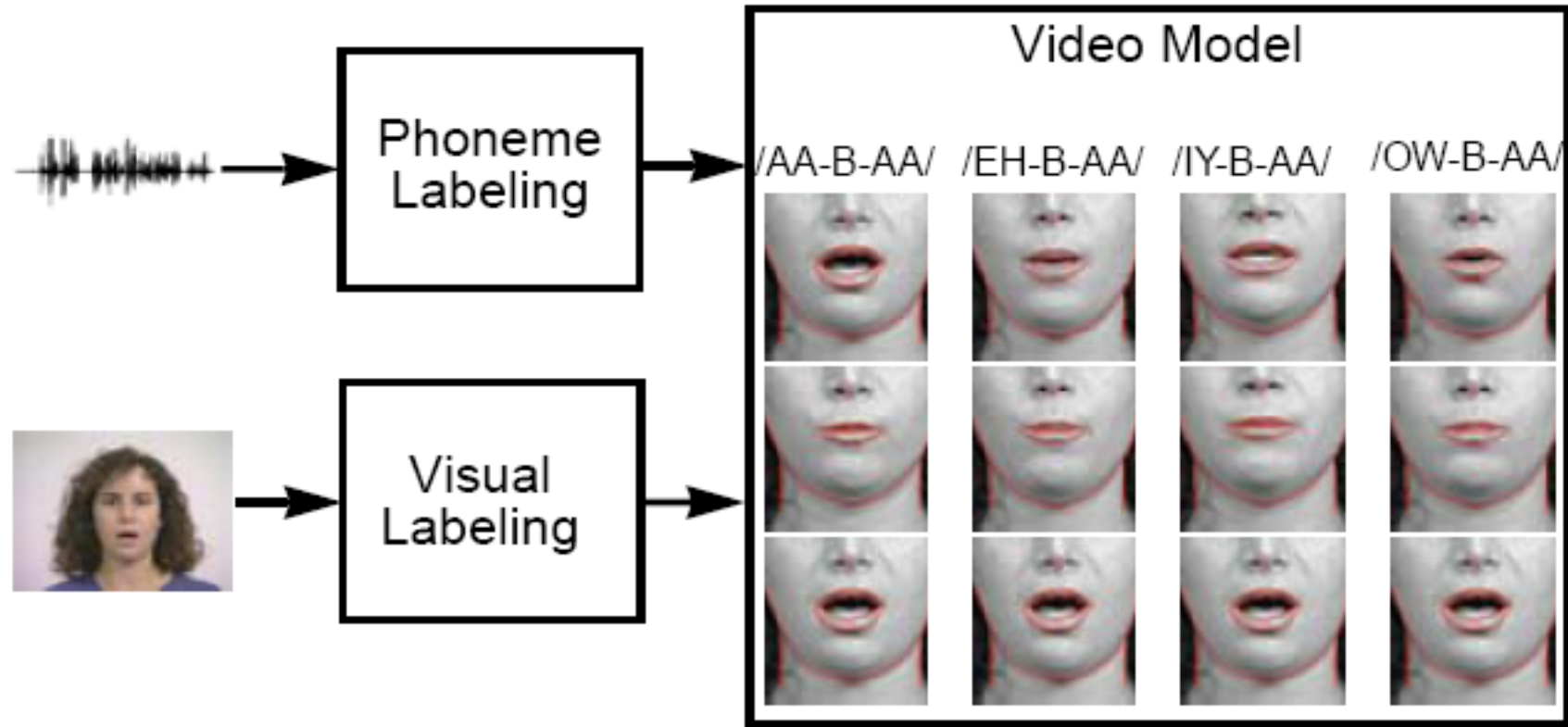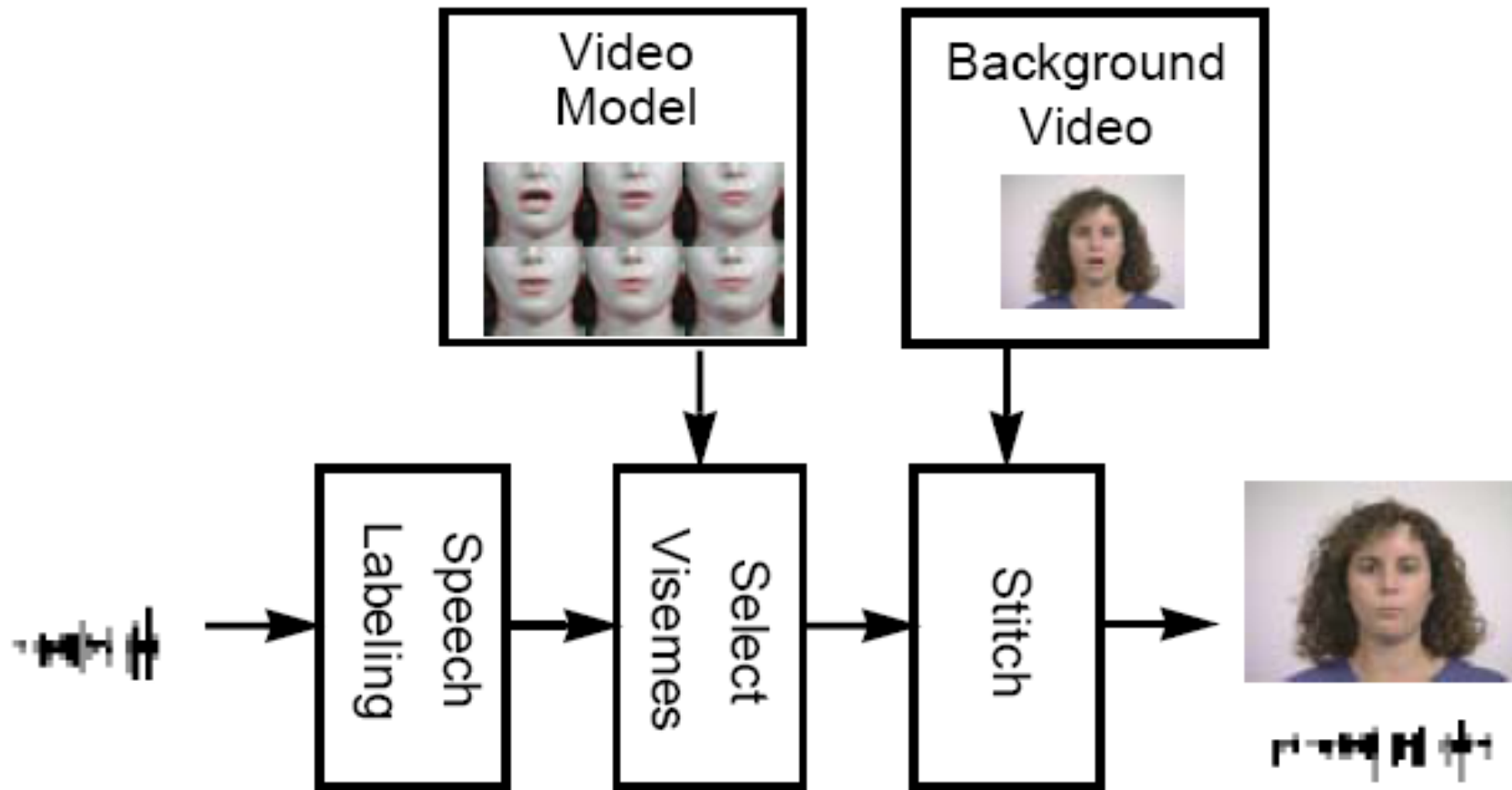
Morphable model for human body

-20 kg -40 kg -20 kg original +20 kg +40 kg +20 kg
-20 cm   +20 cm

Image-based faces
(lip sync.)

Video rewrite (analysis)

Video rewrite (synthesis)

Results

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

training video

Read my lips.

I never met Forest Gump.

Morphable speech model
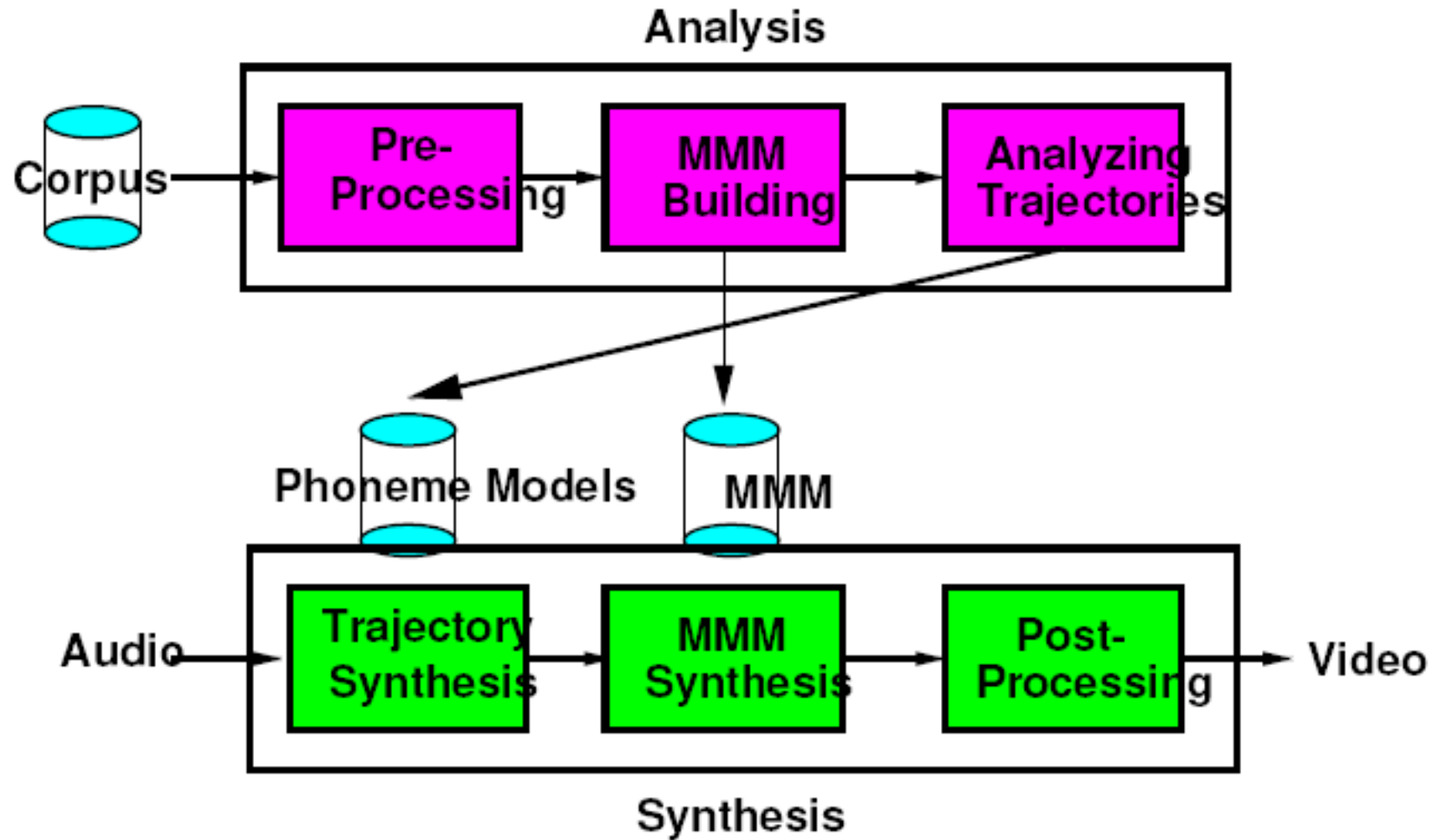
Preprocessing
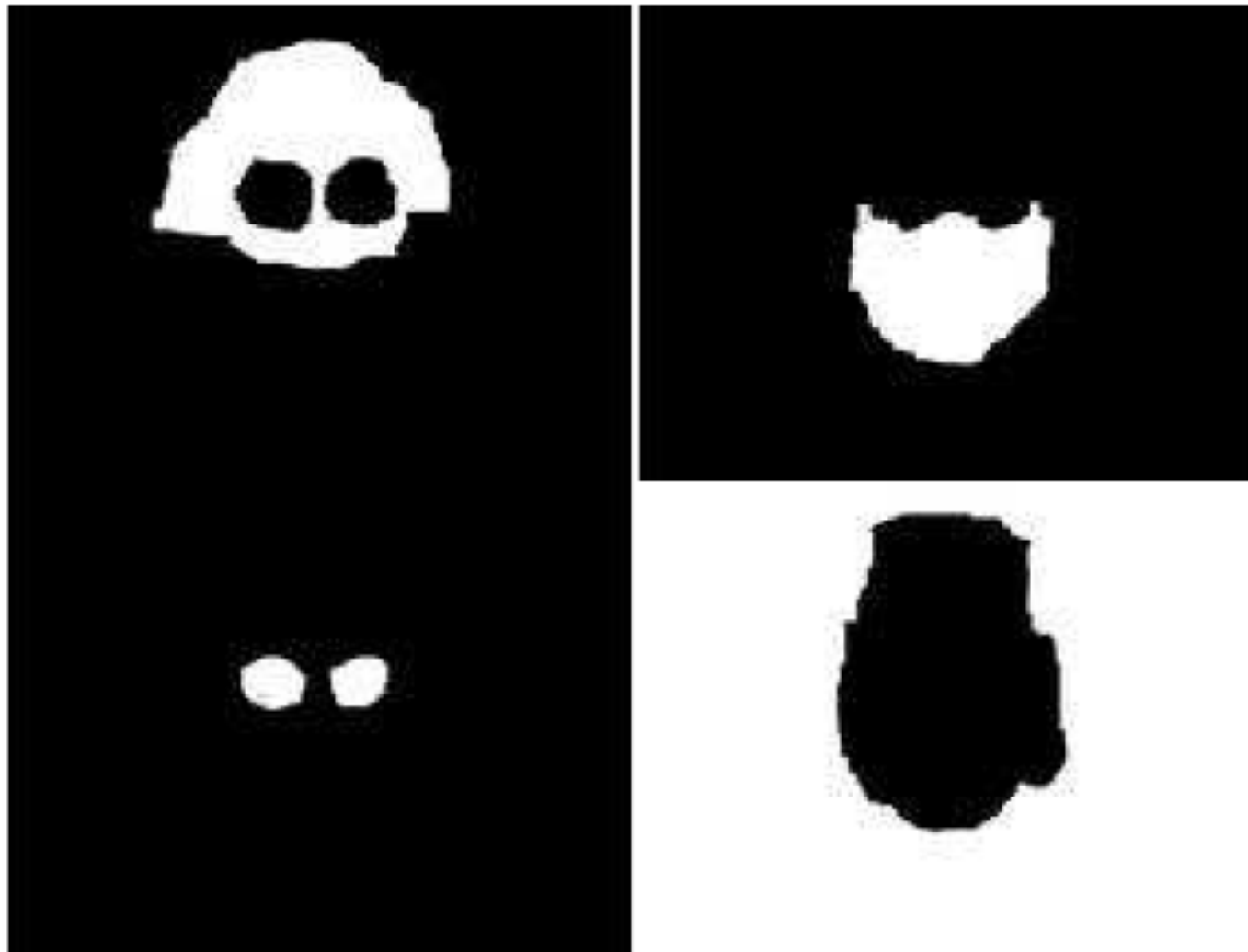
Prototypes (PCA+k-mean clustering)
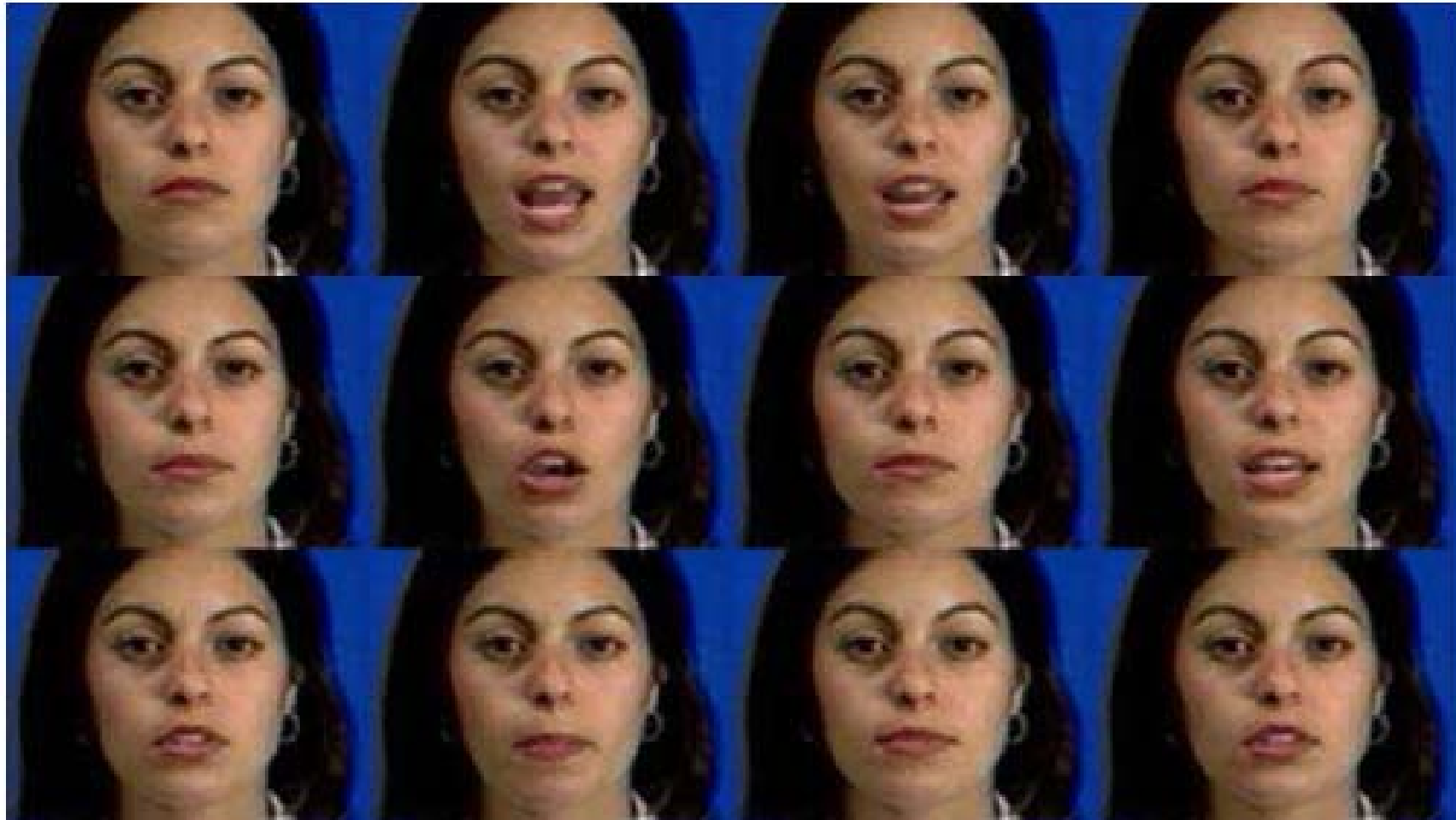
We find I_i and C_i for each prototype image.

Morphable model

$$I^{morph}(\alpha, \beta) = \sum_{i=1}^{N} \beta_i \mathbf{W}(I_i, \mathbf{W}(\sum_{j=1}^{N} \alpha_j C_j - C_i, C_i))$$
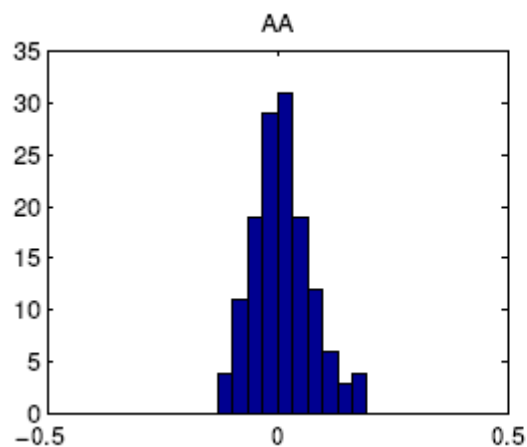
analysis

$$\mathbf{I} \rightleftharpoons \alpha \ \beta$$
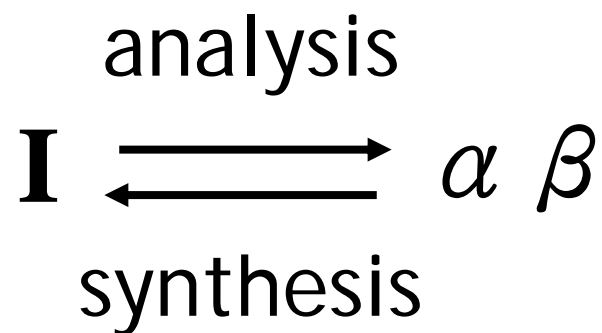
synthesis

Morphable model

analysis

synthesis
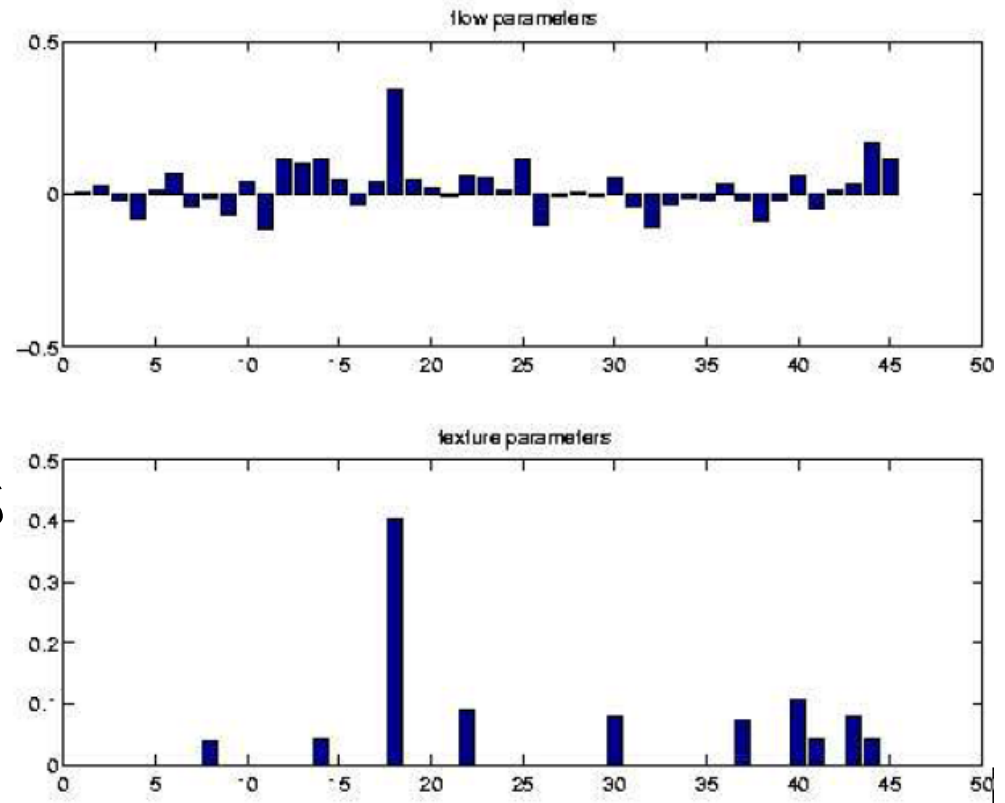
Synthesis

$$E = (y - \mu)^T D^T \Sigma^{-1} D (y - \mu) + \lambda \, y^T W^T W y$$

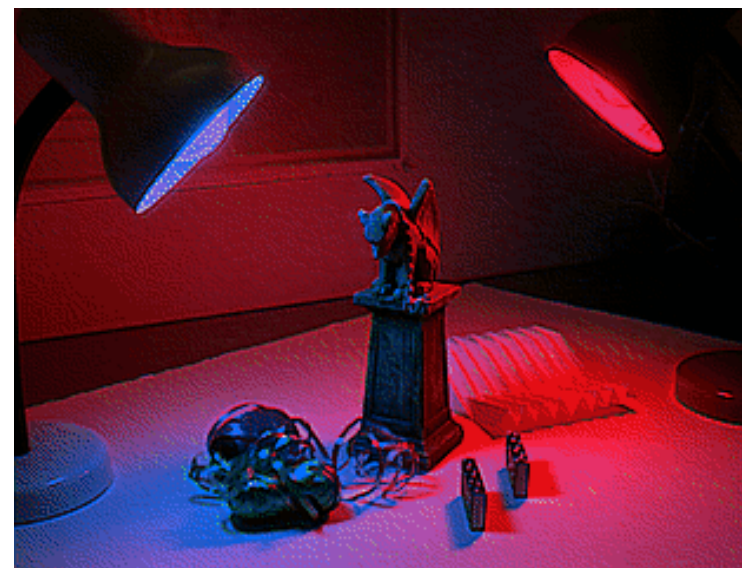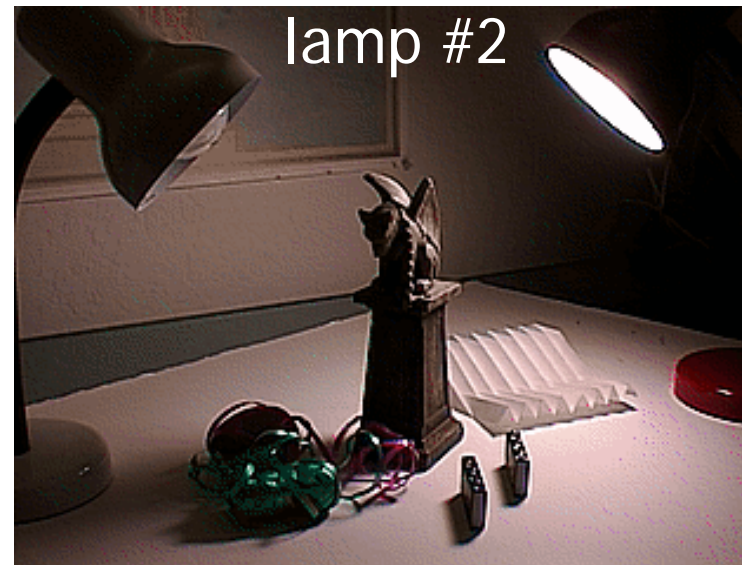$\underbrace{\qquad\qquad}_{target\ term}$ $\underbrace{\qquad}_{smoothness}$
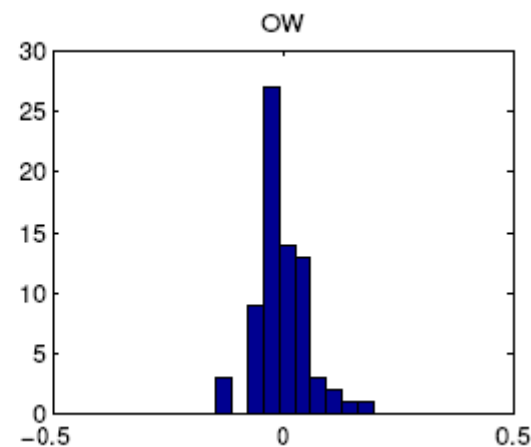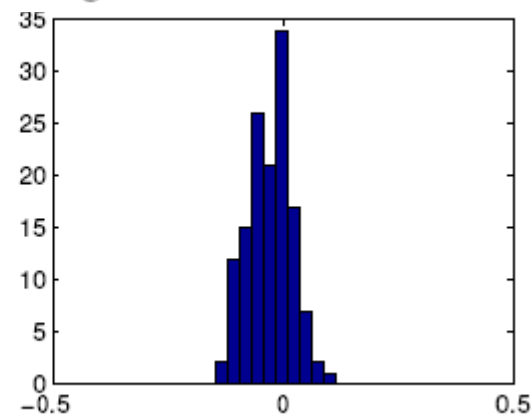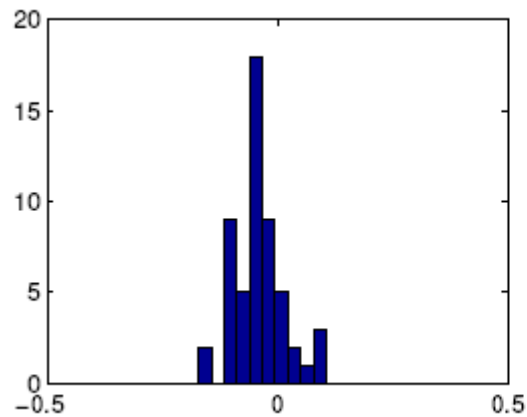
Results

Results

Relighting faces

Light is additive

lamp #1

lamp #2

Light stage 1.0
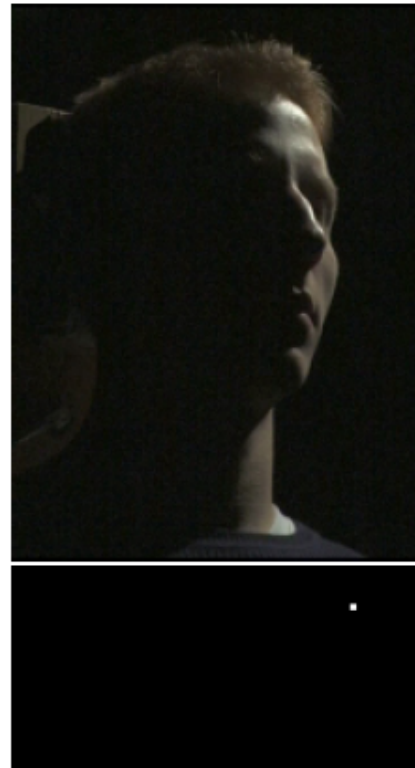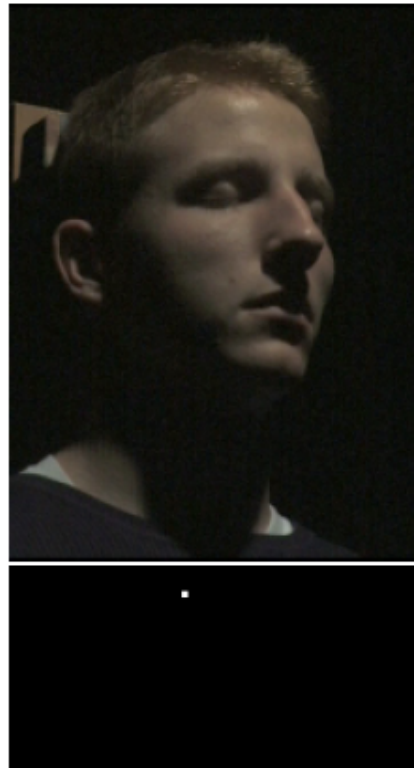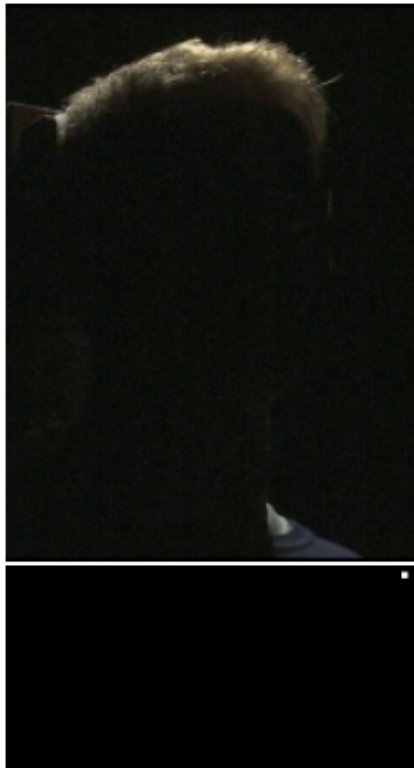
Light stage 1.0

64x32 lighting directions

Input images
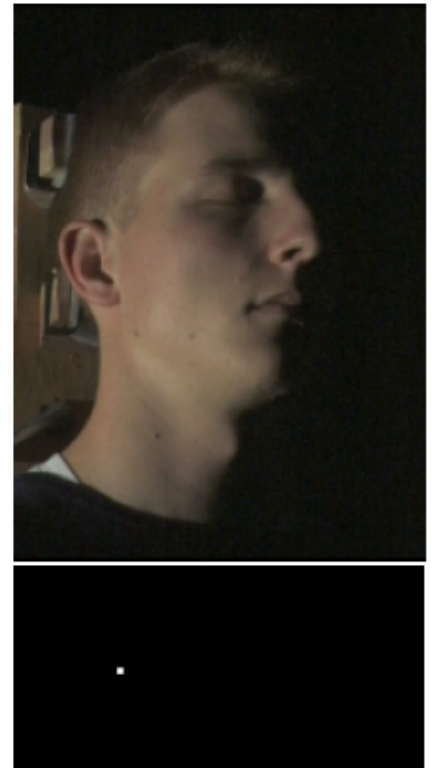
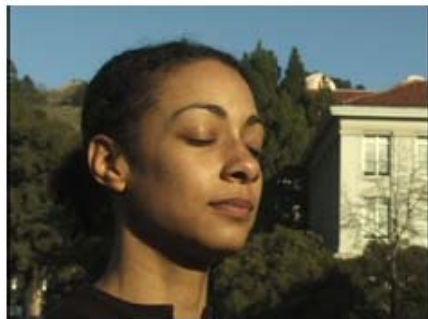# Reflectance function
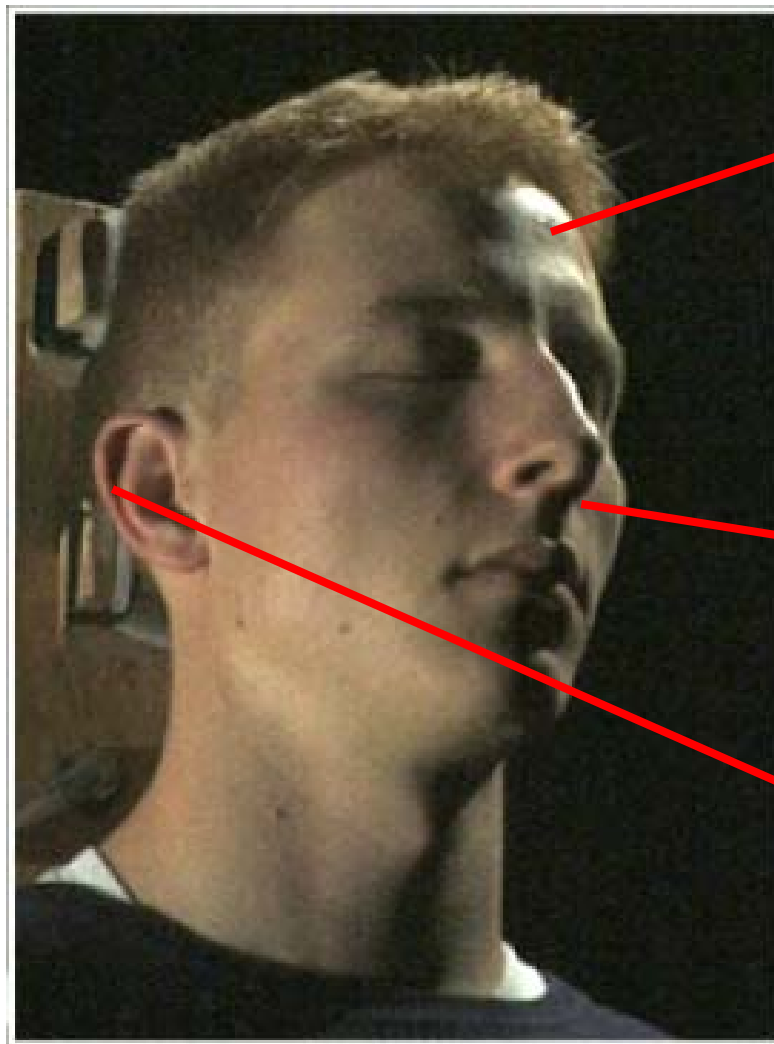
occlusion   flare

Relighting

normalized
light map

×

reflectance
function

=

lighting product

$\left\| \quad \right\|_1$

lighting product

=

rendered
pixel
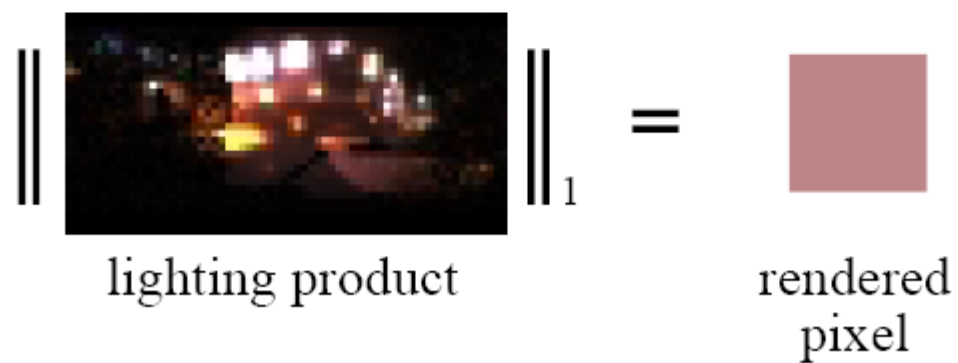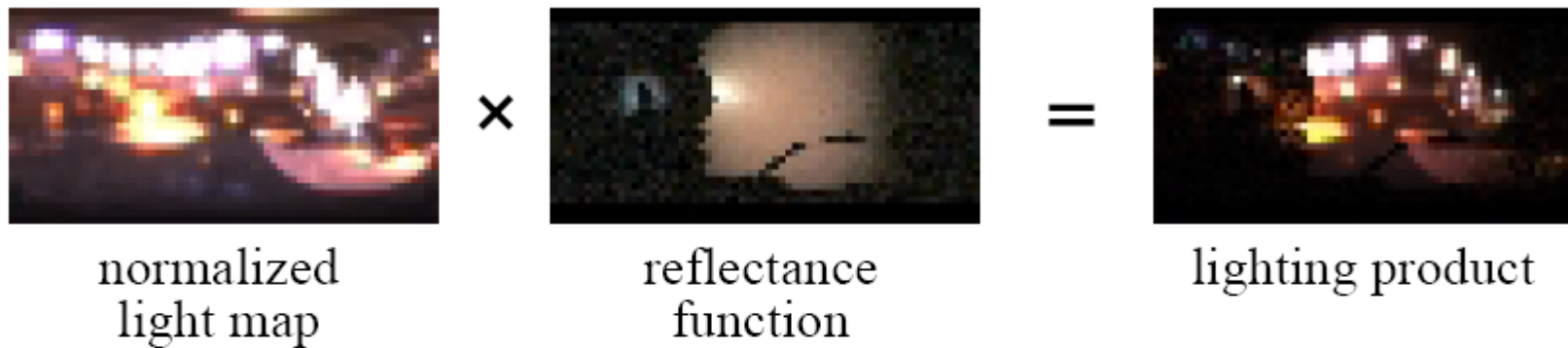
Results

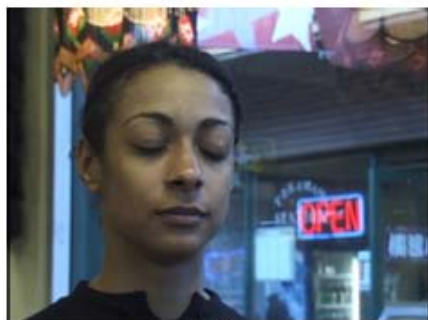# Changing viewpoints

(a)

(b)

(c)

(d)

(e)

(f)

(g)
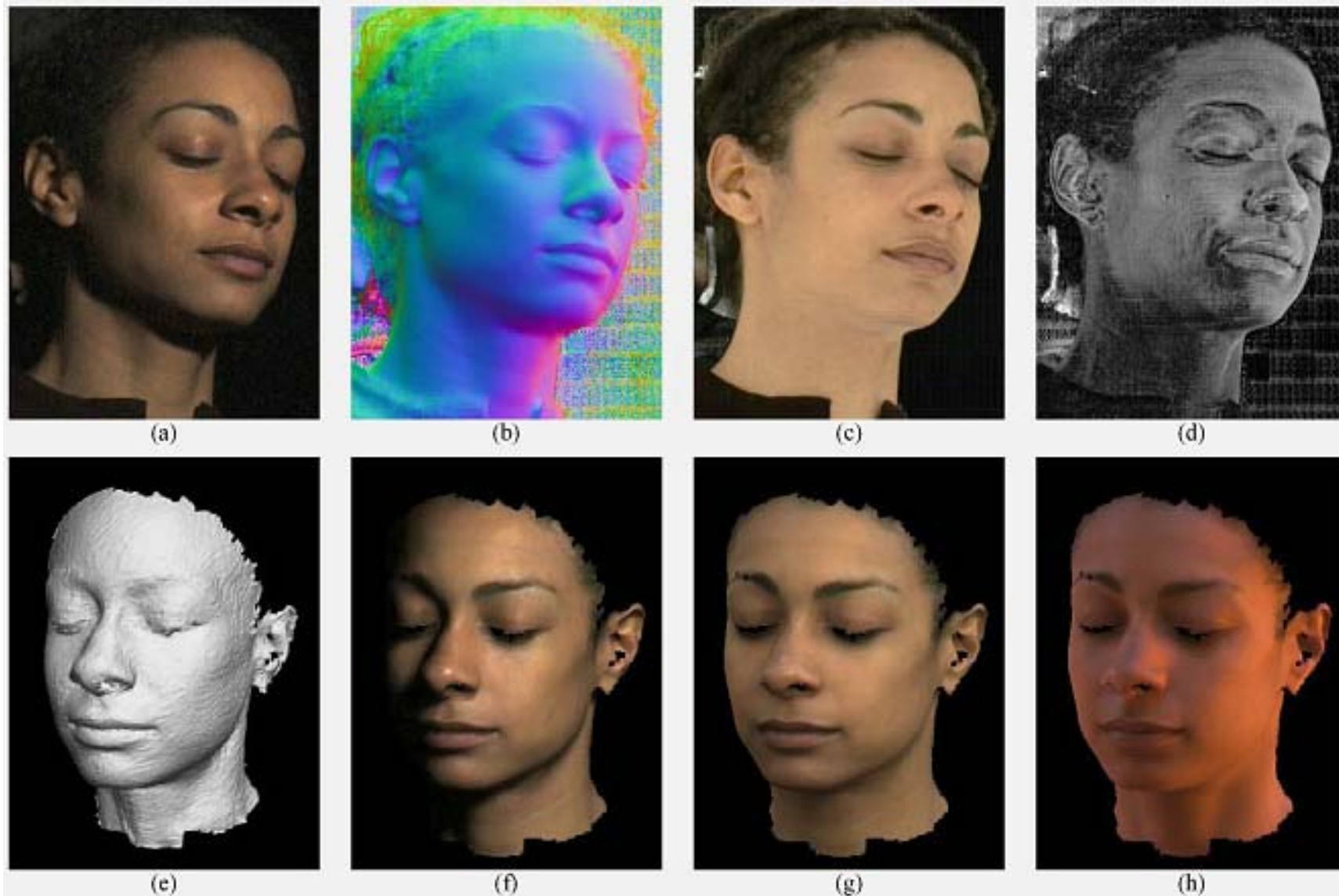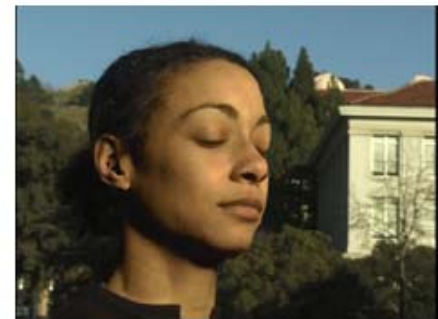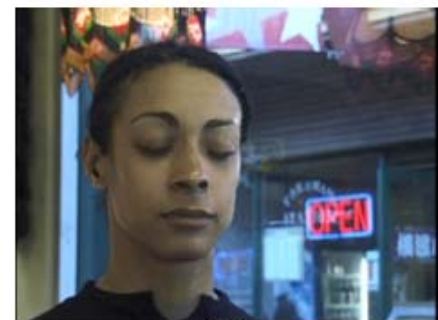
(h)
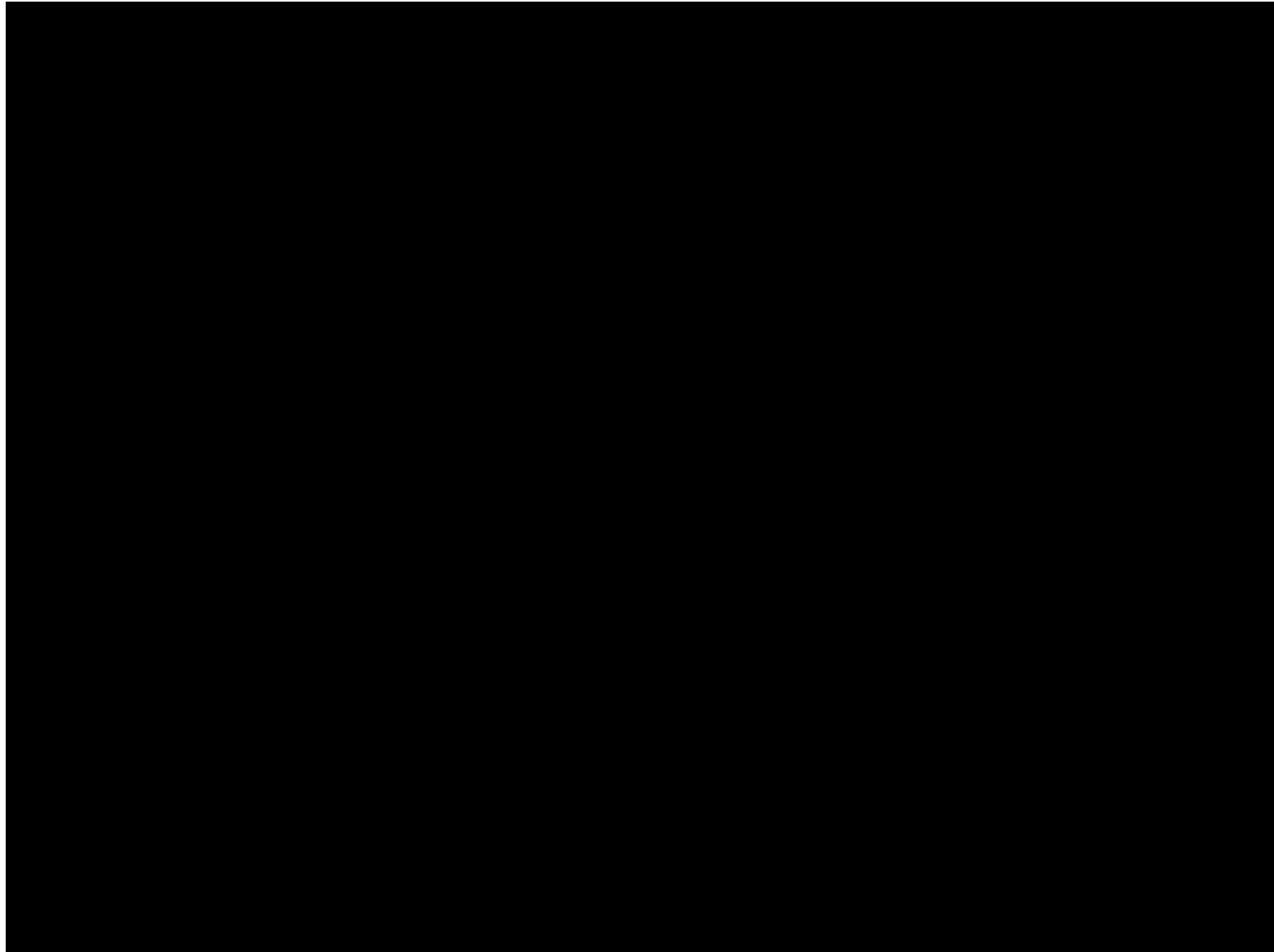
Video

3D face applications: Spiderman 2

Spiderman 2

real   synthetic

Spiderman 2

video

Light stage 3

Light stage 6

Application: The Matrix Reloaded

References

- Paul Debevec, Rendering Synthetic Objects into Real Scenes: Bridging Traditional and Image-based Graphics with Global Illumination and High Dynamic Range Photography, SIGGRAPH 1998.

- F. Pighin, J. Hecker, D. Lischinski, D. H. Salesin, and R. Szeliski. Synthesizing realistic facial expressions from photographs. SIGGRAPH 1998, pp75-84.

- 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.

- Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, Mark Sagar, Acquiring the Reflectance Field of a Human Face, SIGGRAPH 2000.

- 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.