Faces and Image-Based Lighting

Digital Visual Effects, Spring 2009

Yung-Yu Chuang

2009/6/11

with slides by Richard Szeliski, Steve Seitz, Alex Efros, Li-Yi Wei and Paul Debevec

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

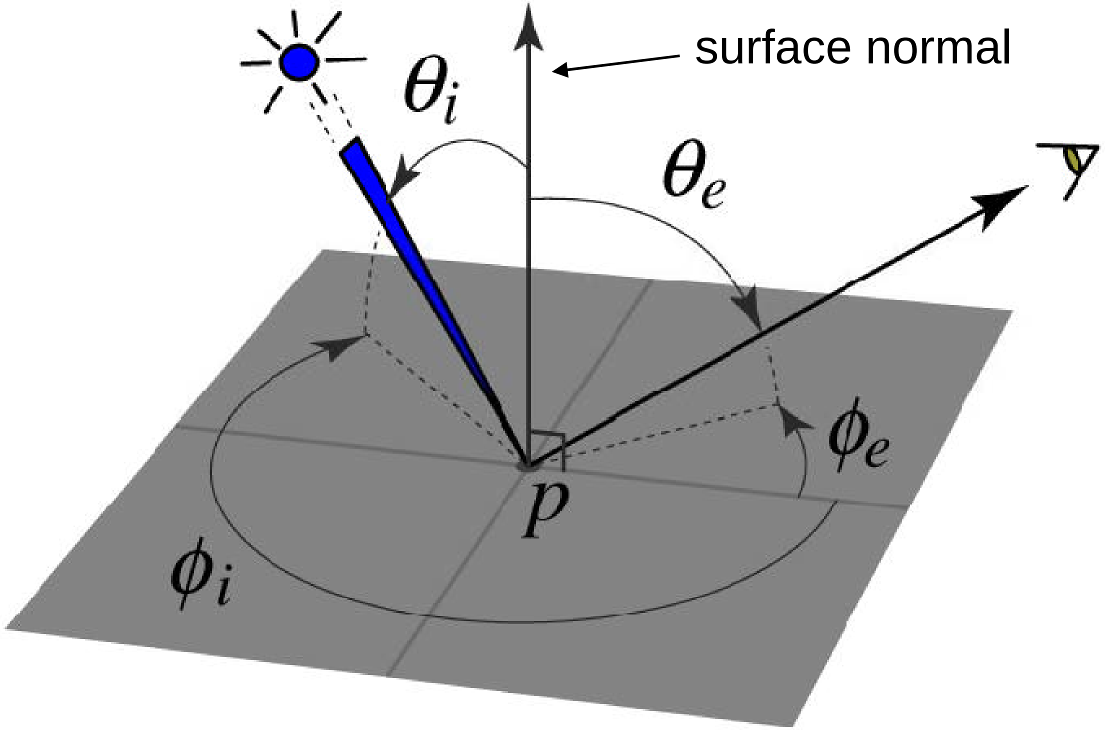
Image-based lighting

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?

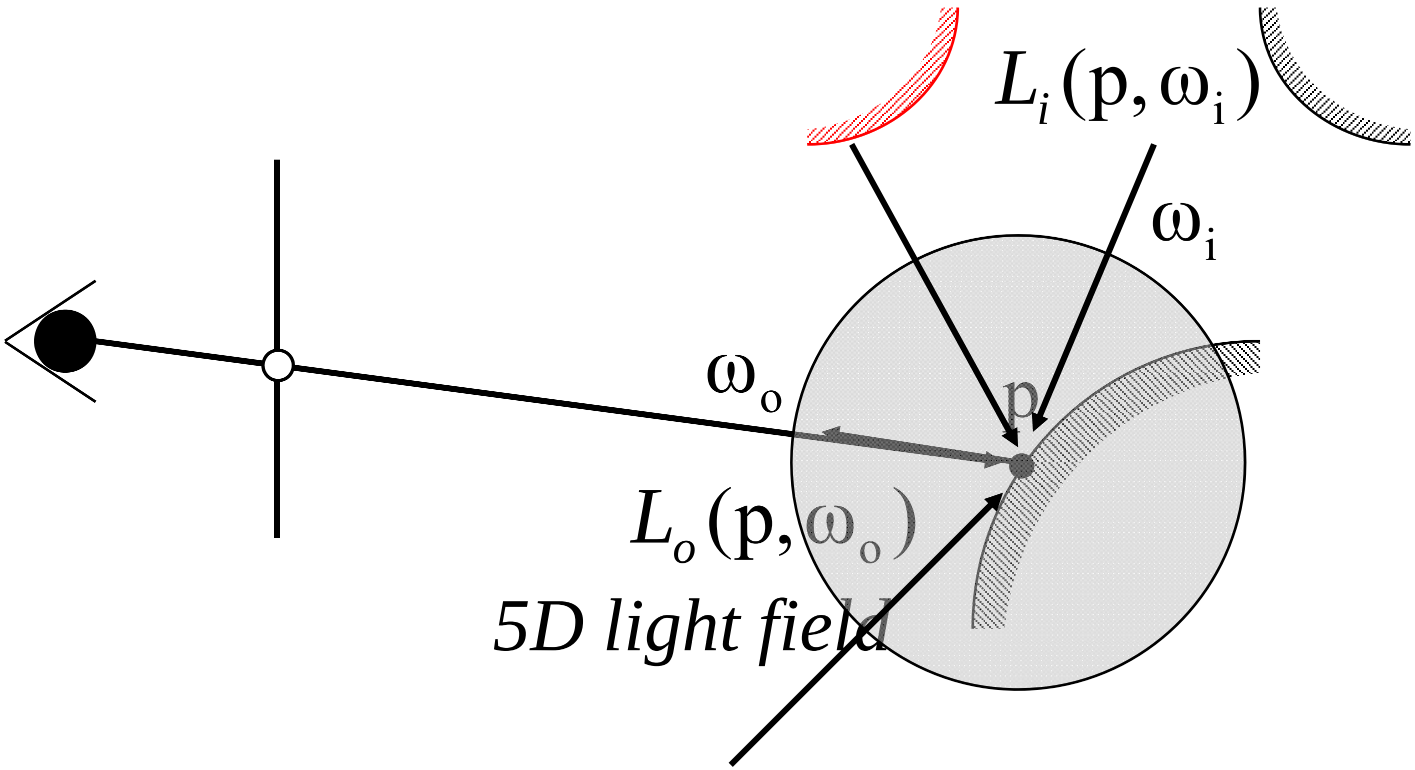
Reflectance

- 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



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

Rendering equation



$$L_o(\mathrm{p}, \omega_\mathrm{o}) = L_e(\mathrm{p}, \omega_\mathrm{o}) + \int_{S^2} \rho(\mathrm{p}, \omega_\mathrm{o}, \omega_\mathrm{i}) L_i(\mathrm{p}, \omega_\mathrm{i}) \left|\cos\theta_\mathrm{i}\right| d\omega_\mathrm{i}$$

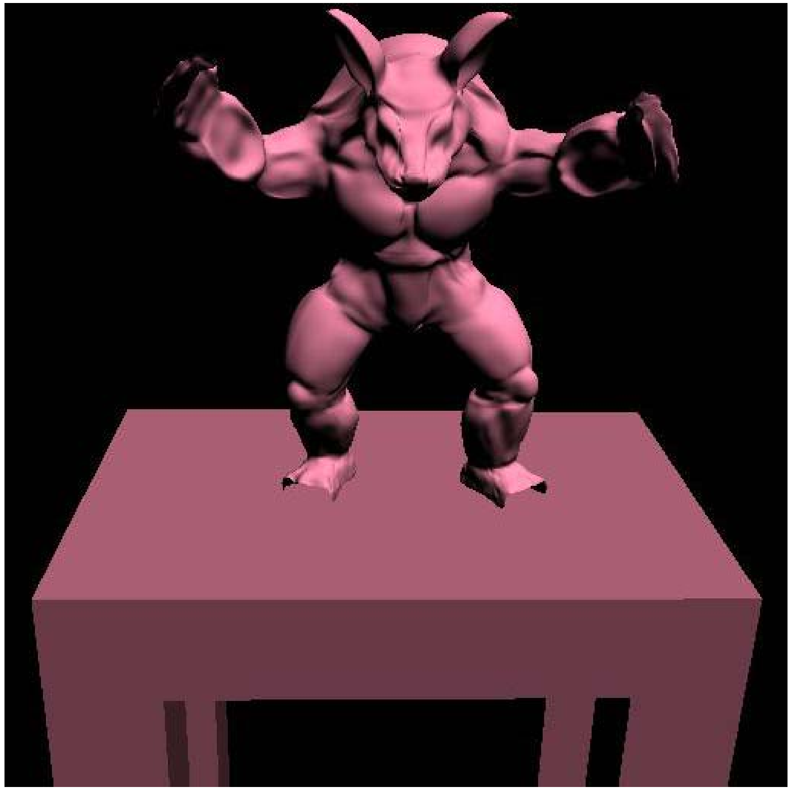
Complex illumination

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

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

Point lights

Classically, rendering is performed assuming point light sources



directional source

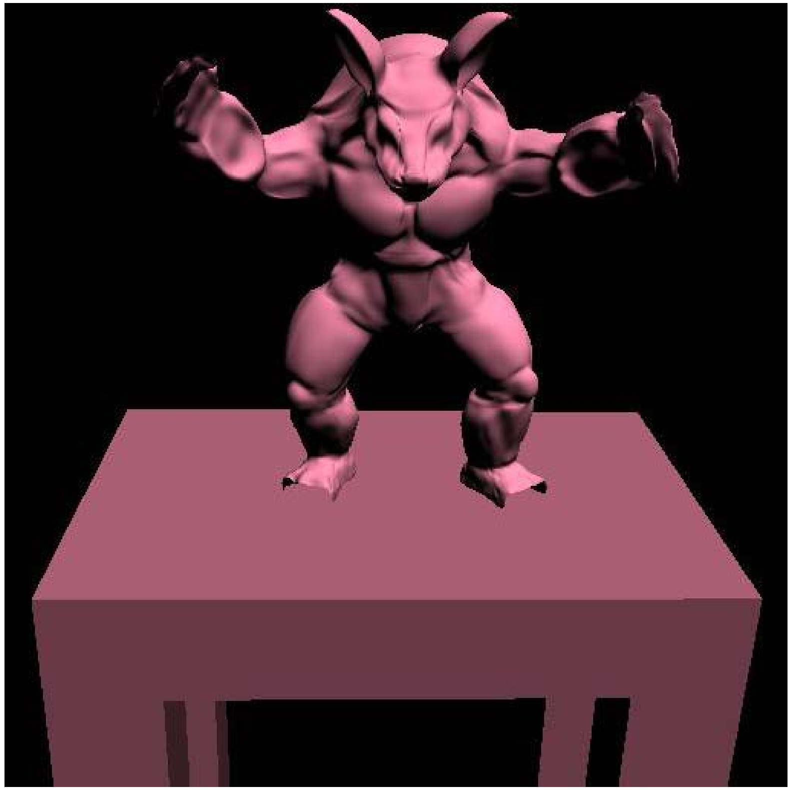
Natural illumination

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

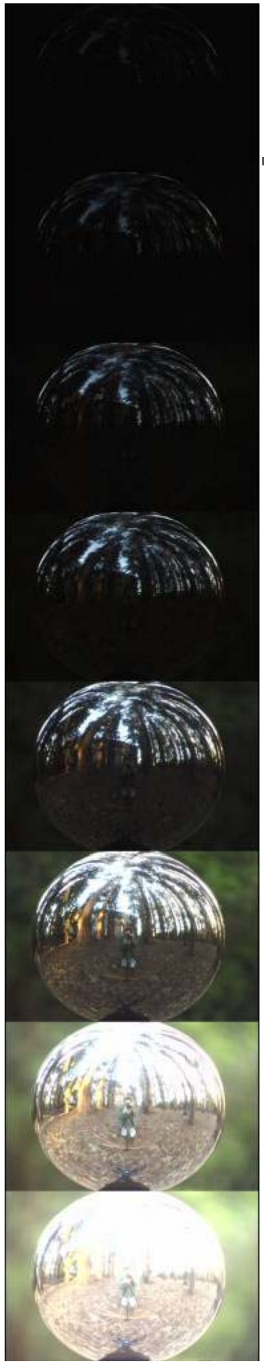


directional source

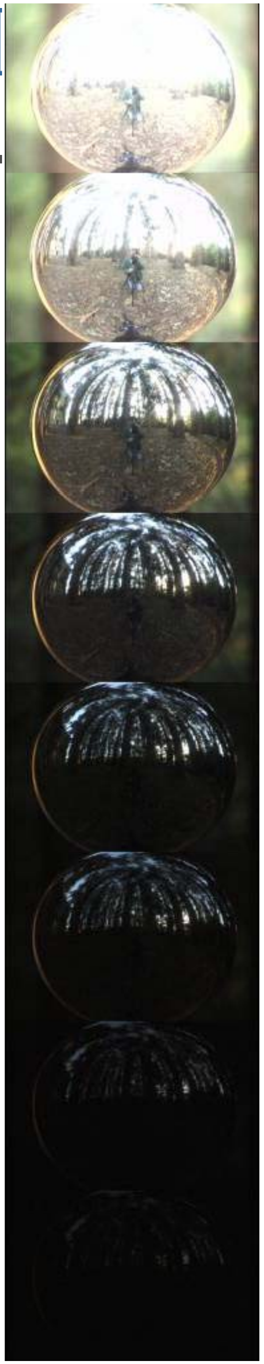
natural illumination

Environment maps

Miller and Hoffman, 1984

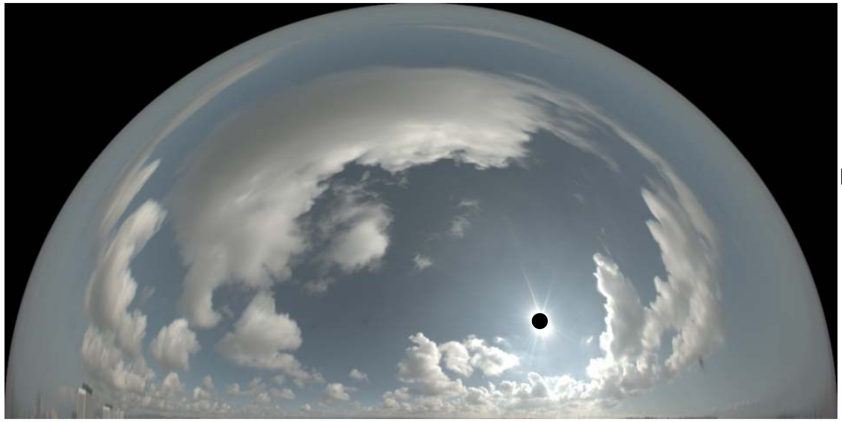


Acquiring the Light Probe

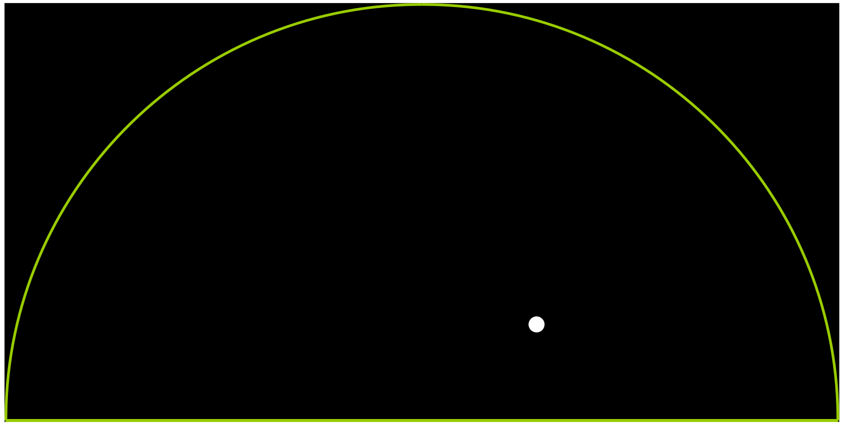


HDRI Sky Probe

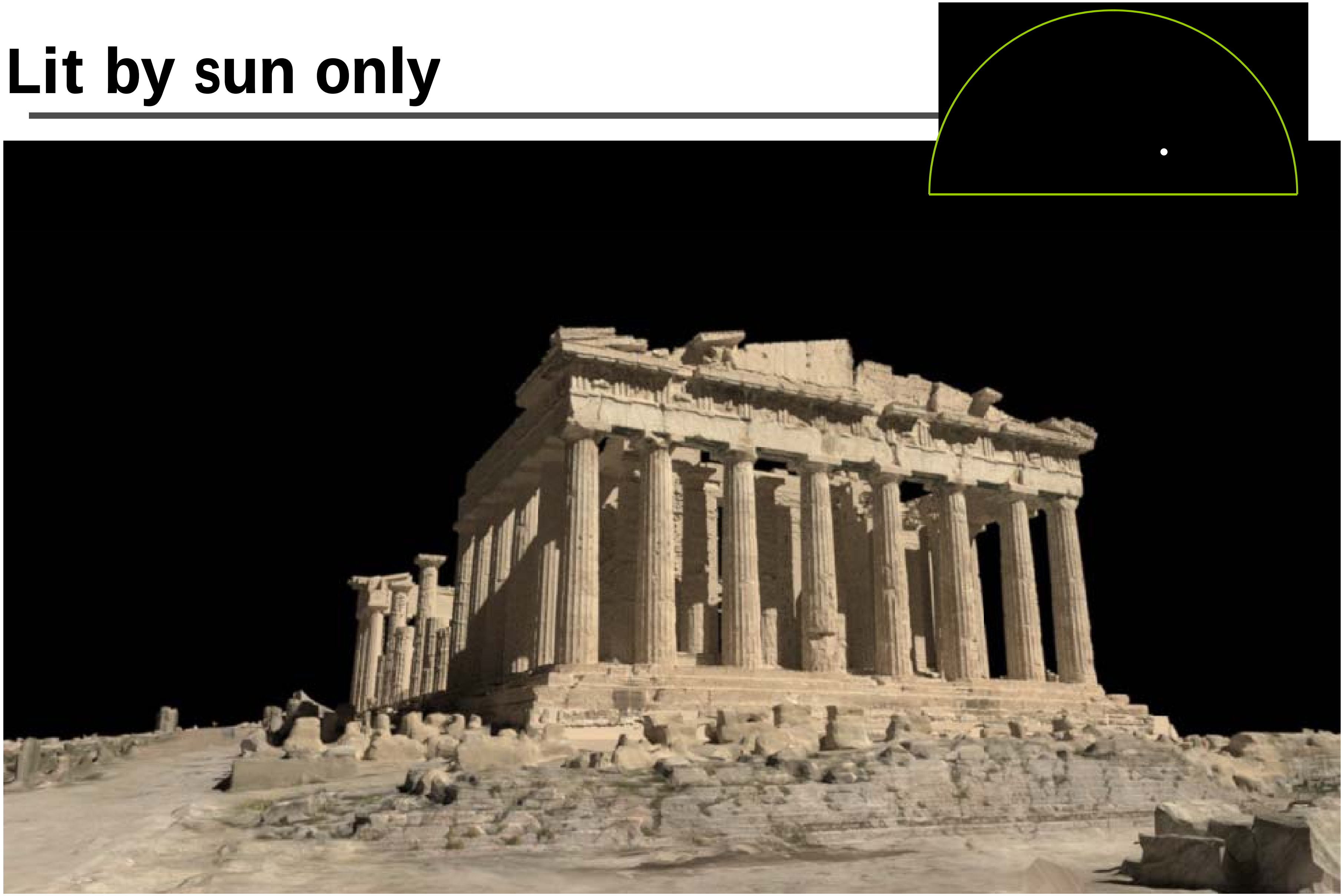
Clipped Sky + Sun Source



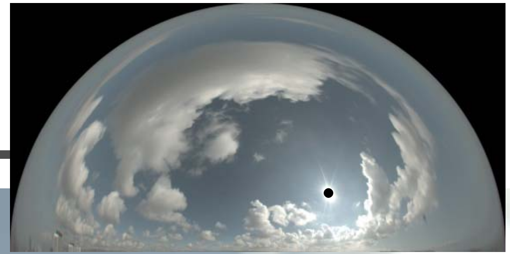
+



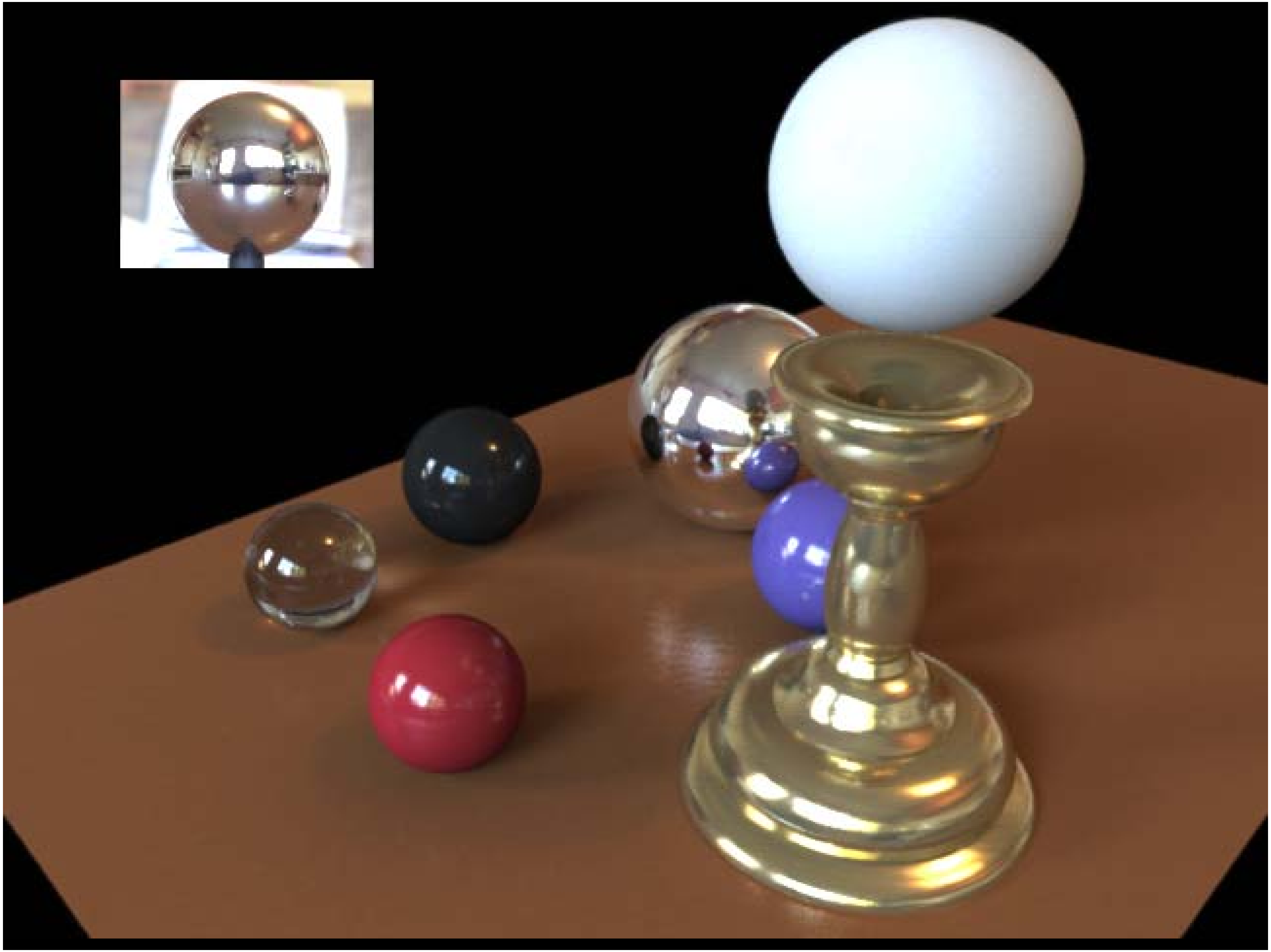
Lit by sun only

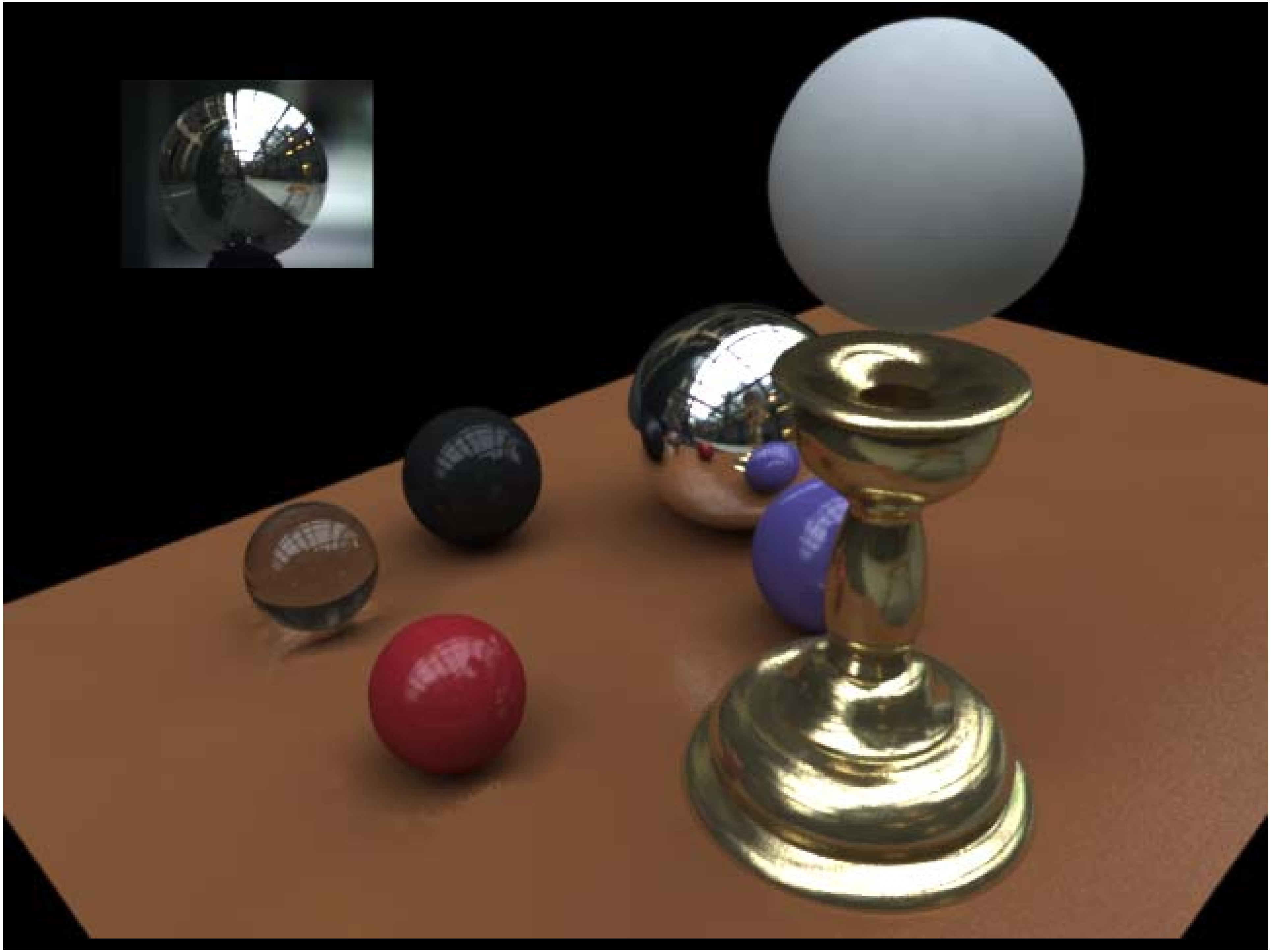


Lit by sky only



Lit by sun and sky

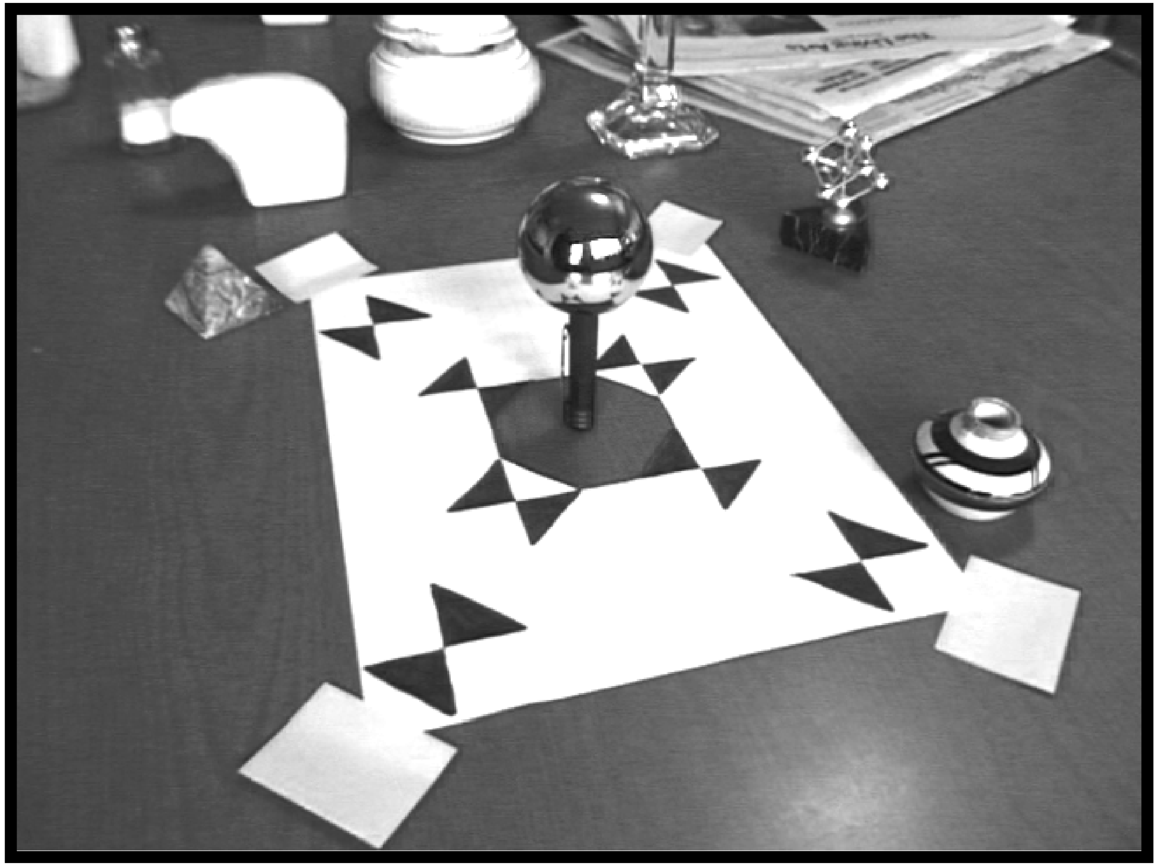




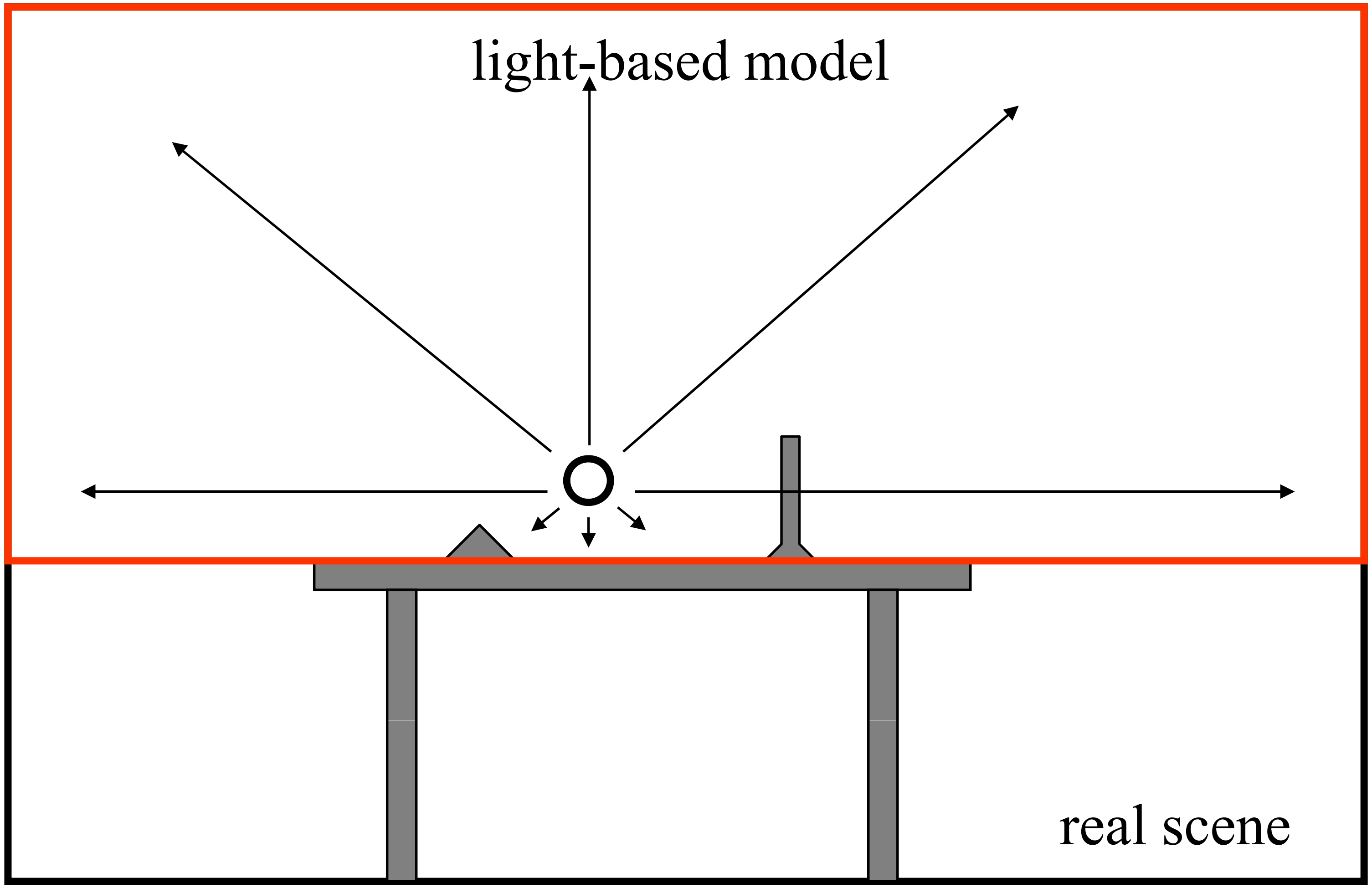
Real Scene Example

- Goal: place synthetic objects on table

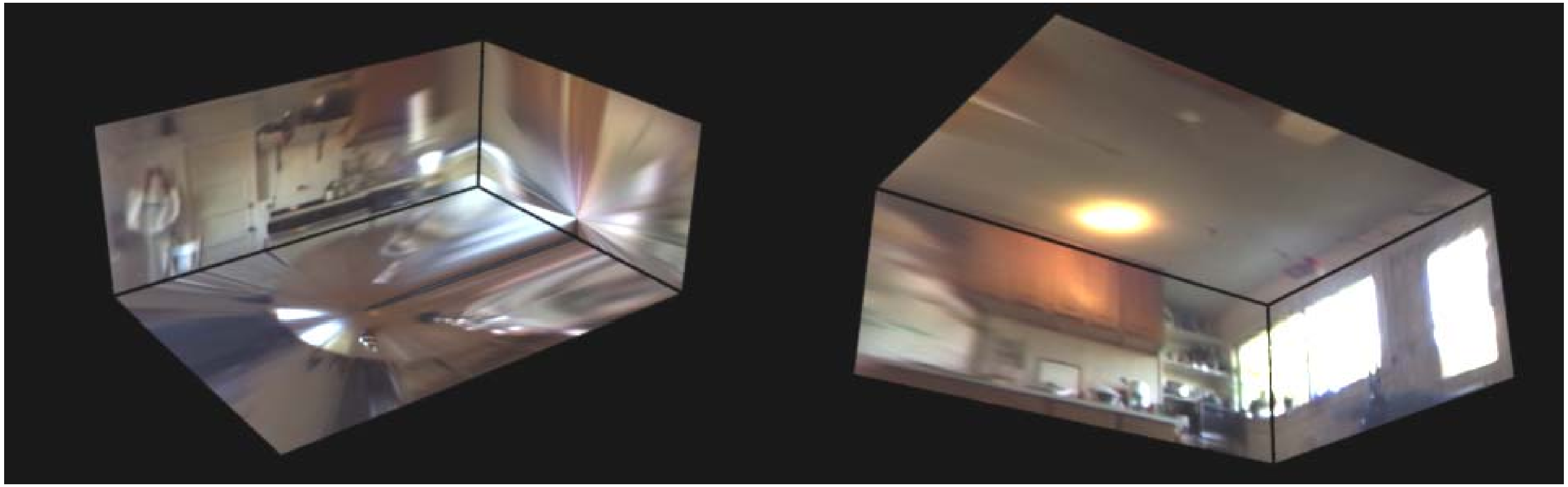
Light Probe / Calibration Grid

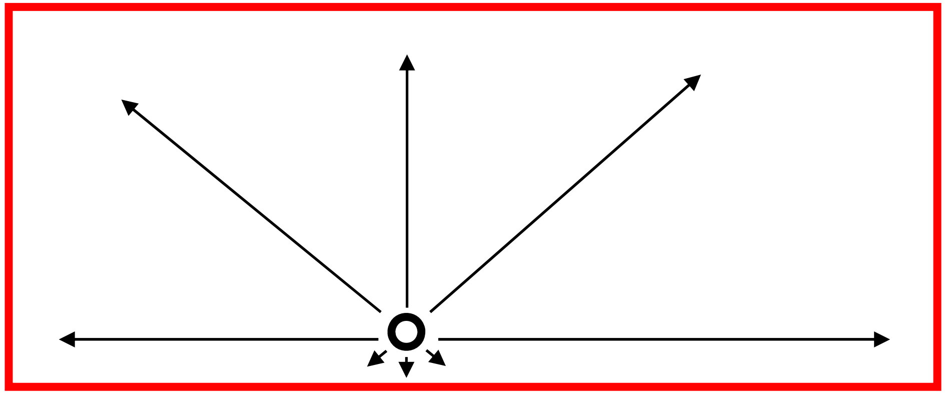


Modeling the Scene



The *Light-Based* Room Model





Rendering into the Scene

- Background Plate

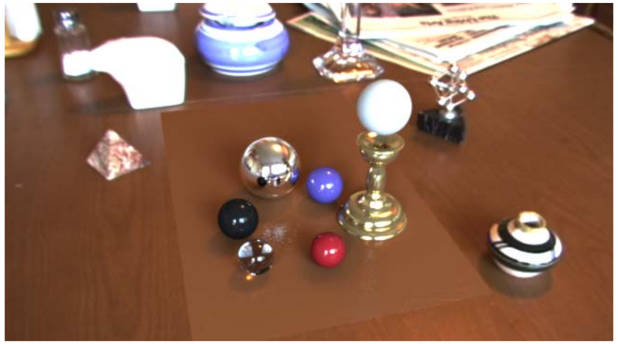
Rendering into the scene

- Objects and Local Scene matched to Scene

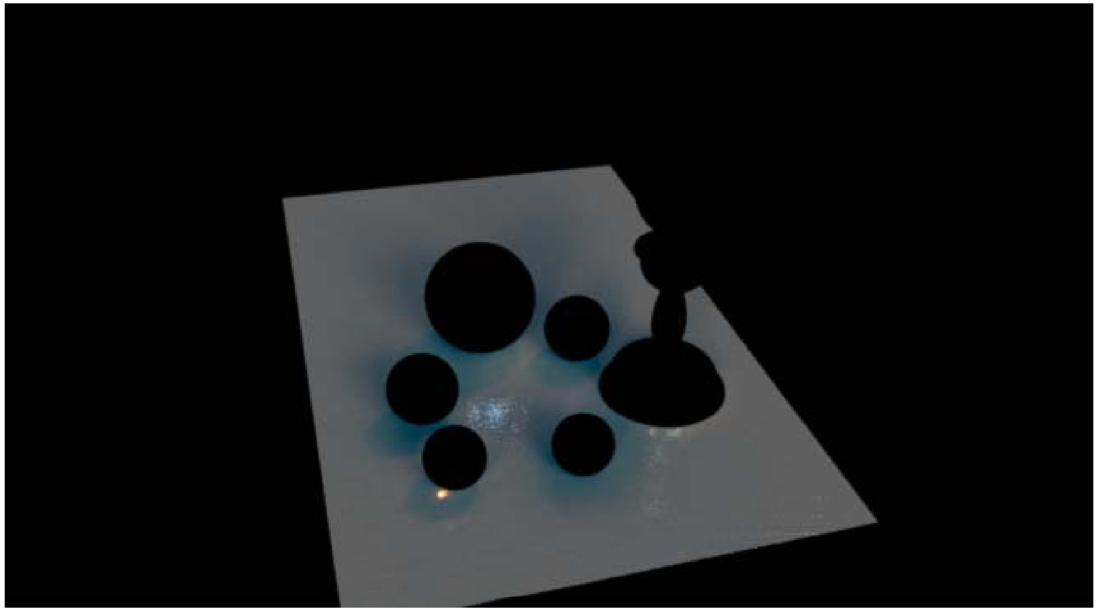
Differential rendering

- Local scene w/o objects, illuminated by model

Differential rendering

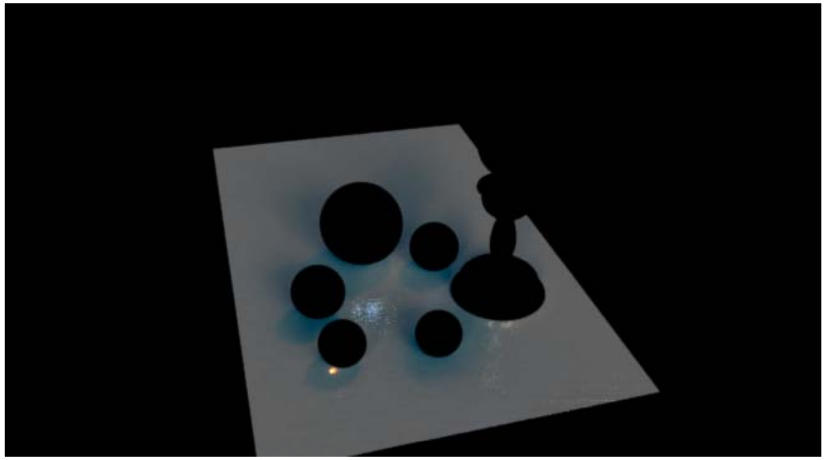


-



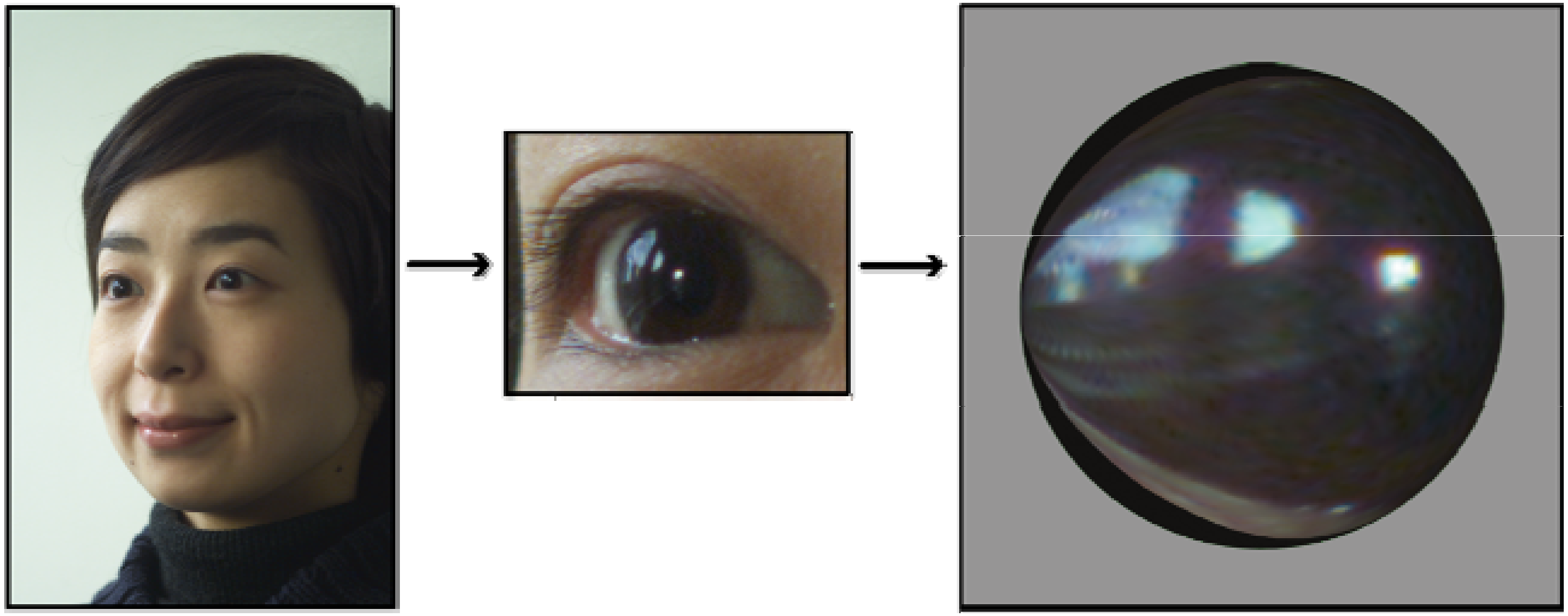
Differential rendering

+



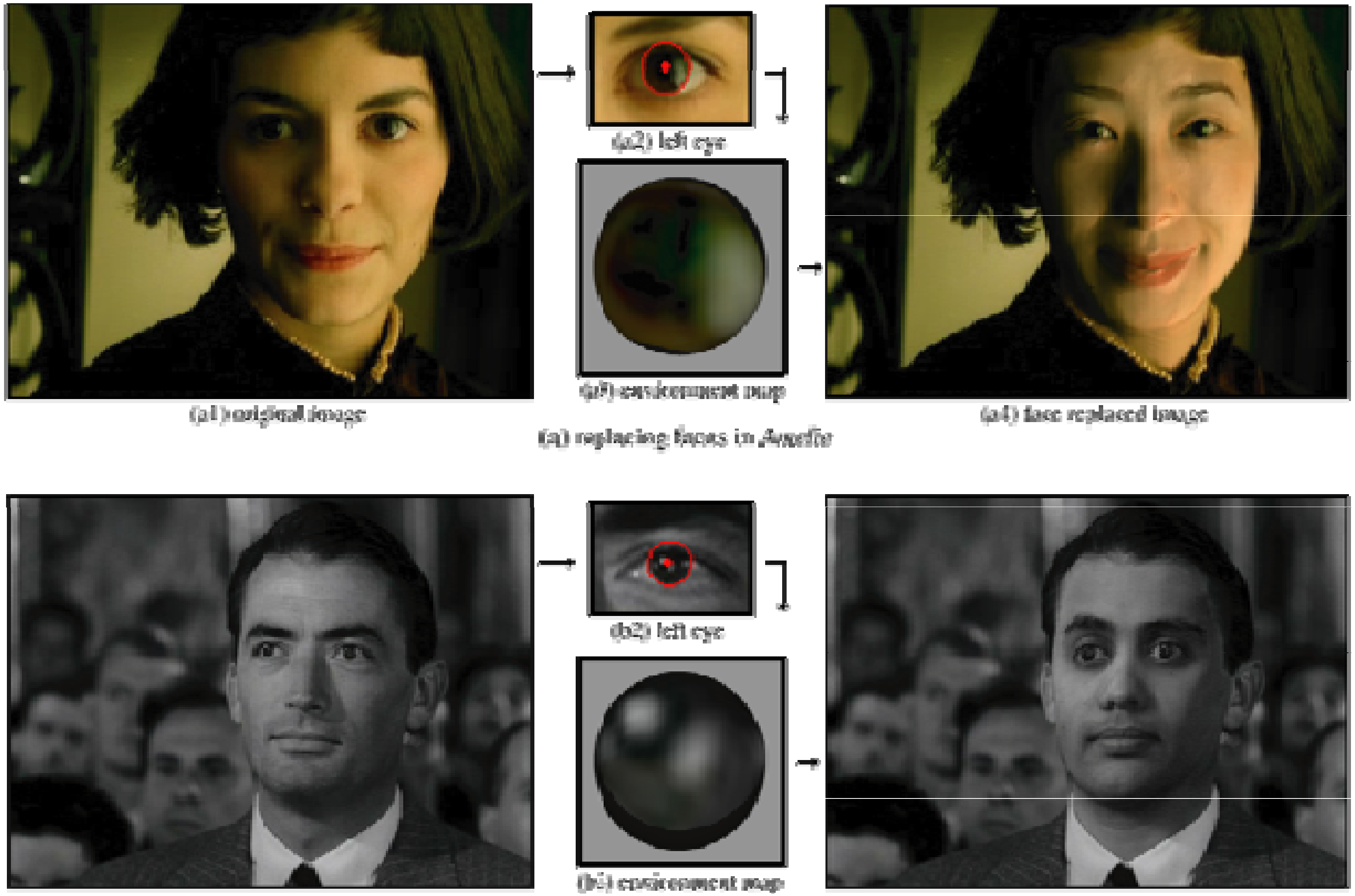
Environment map from single image?

Eye as light probe! (Nayar et al)



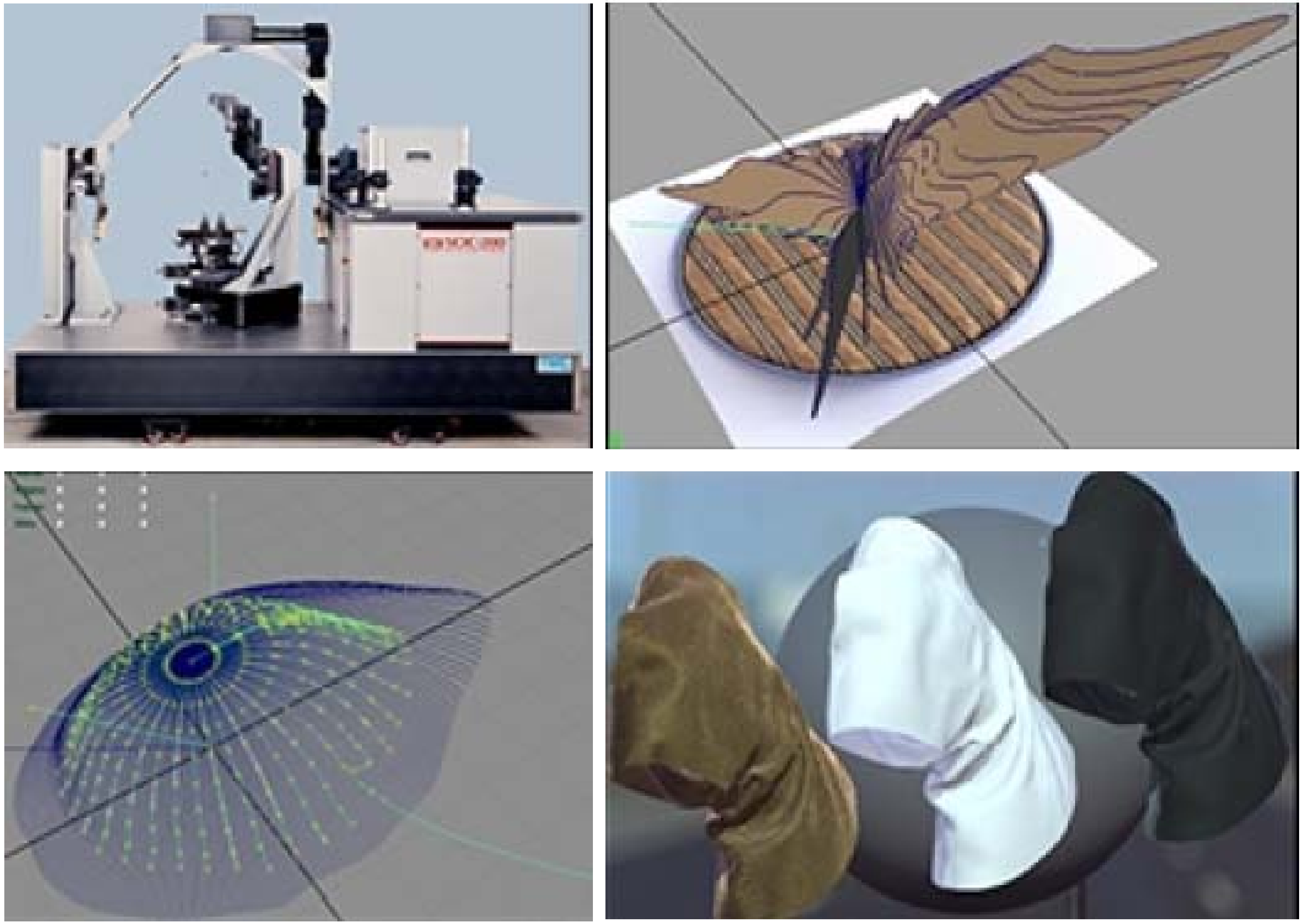
Results

DigiVFX



Application in “Superman returns”

Capturing reflectance



Application in "The Matrix Reloaded"

3D acquisition for faces

Cyberware scanners

face & head scanner

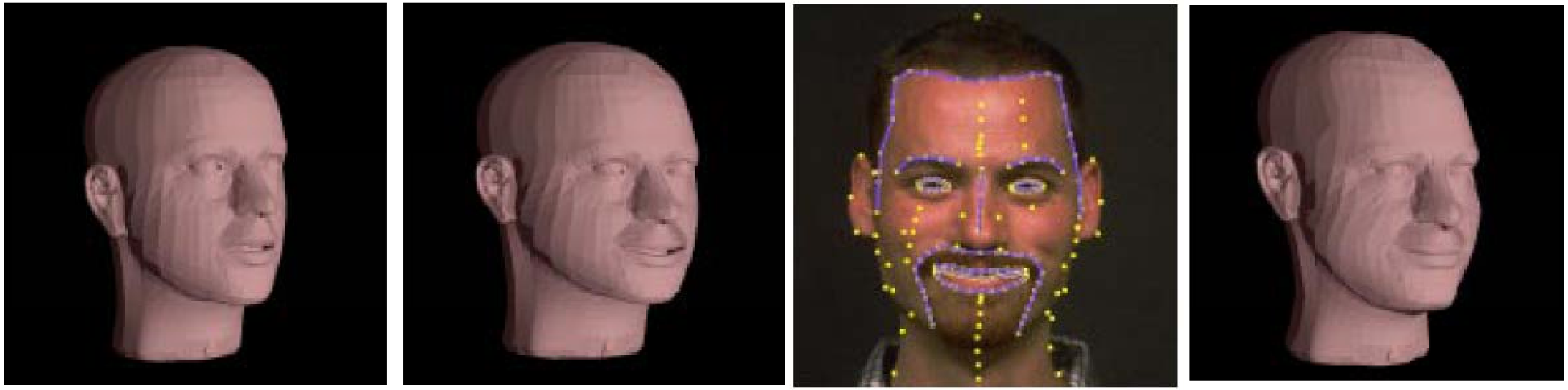
whole body scanner

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

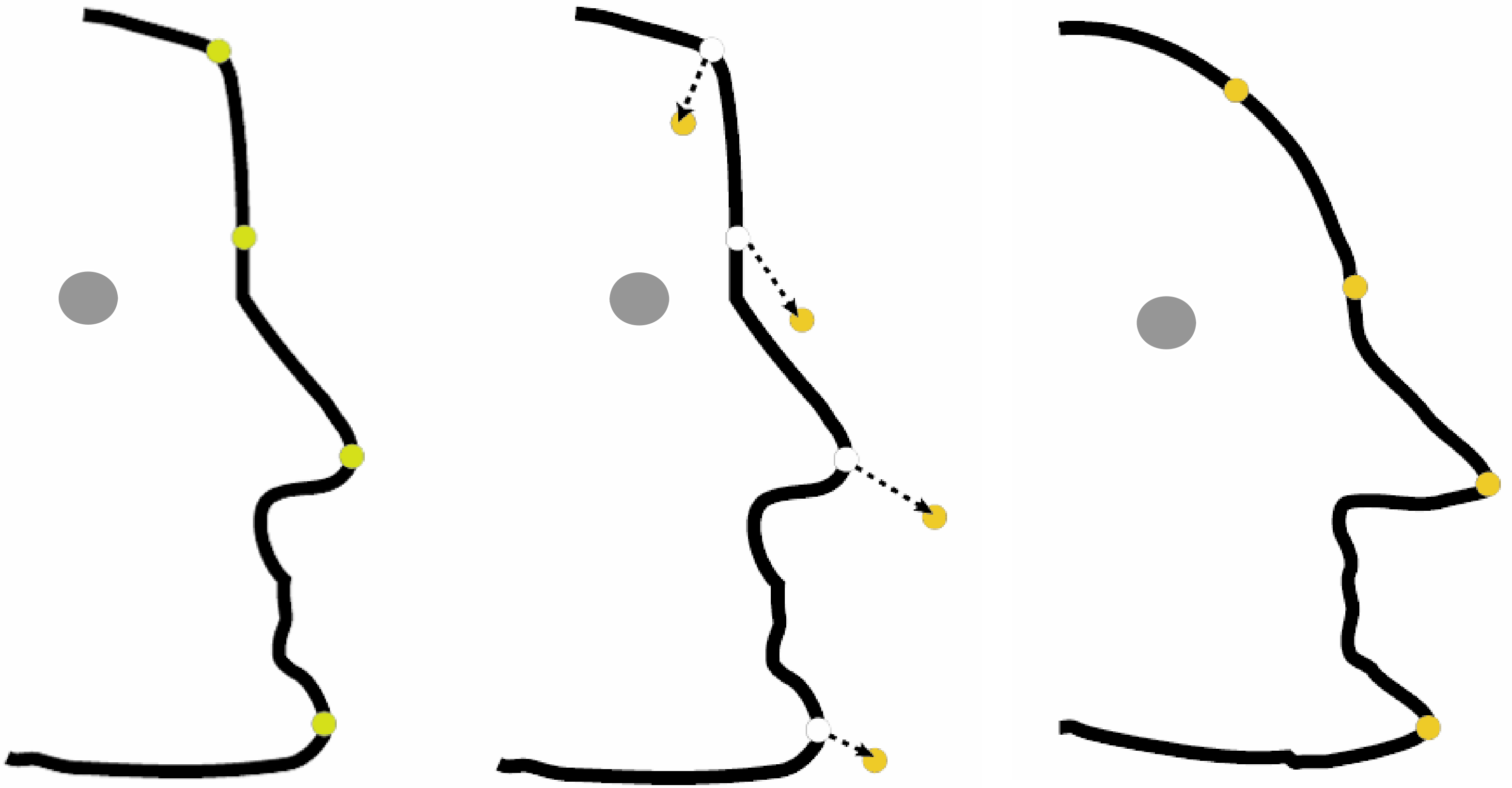
pose estimation

more features

deformed model

Mesh deformation

- Compute displacement of feature points
- Apply scattered data interpolation

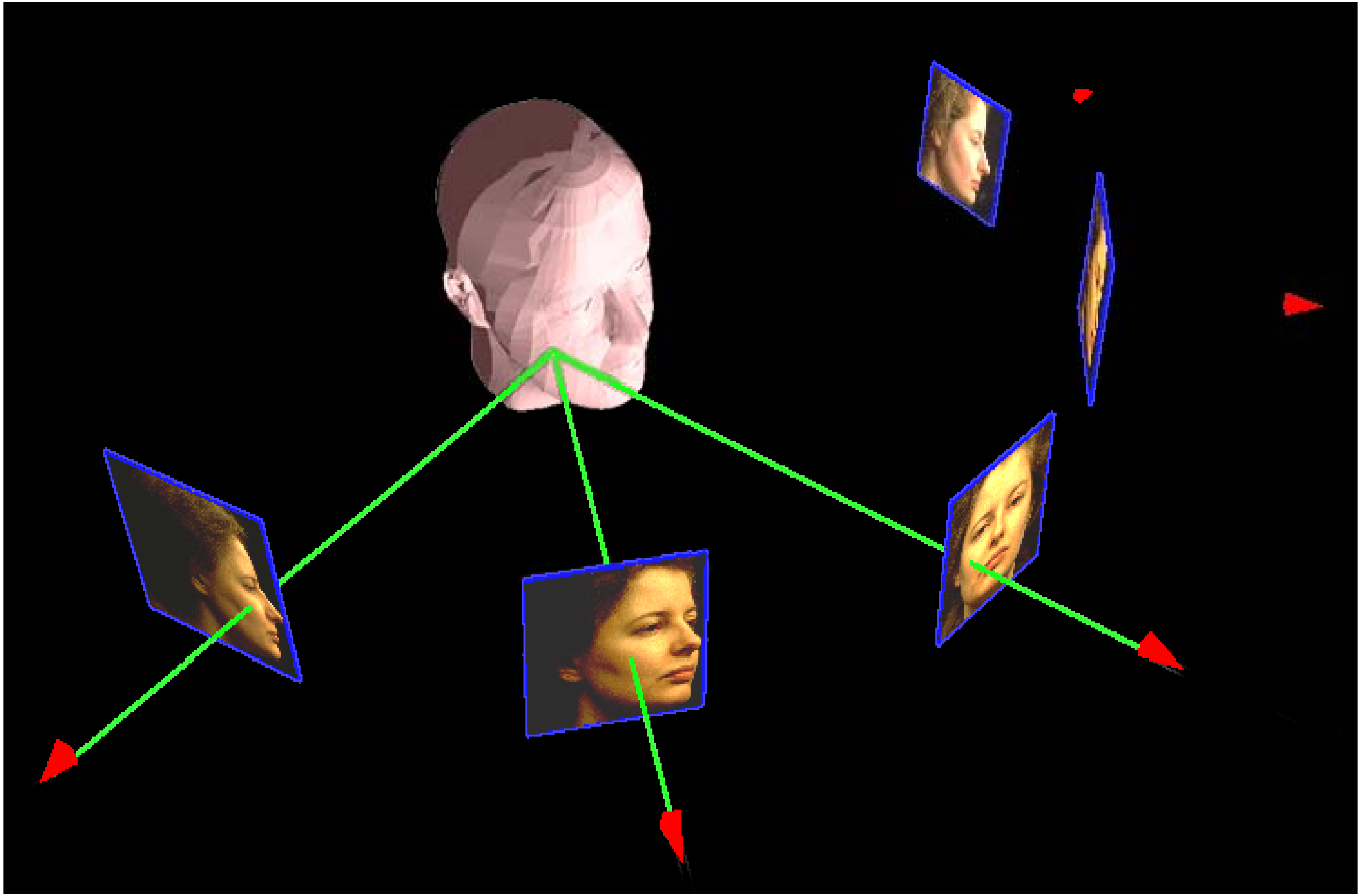


generic model displacement deformed model

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

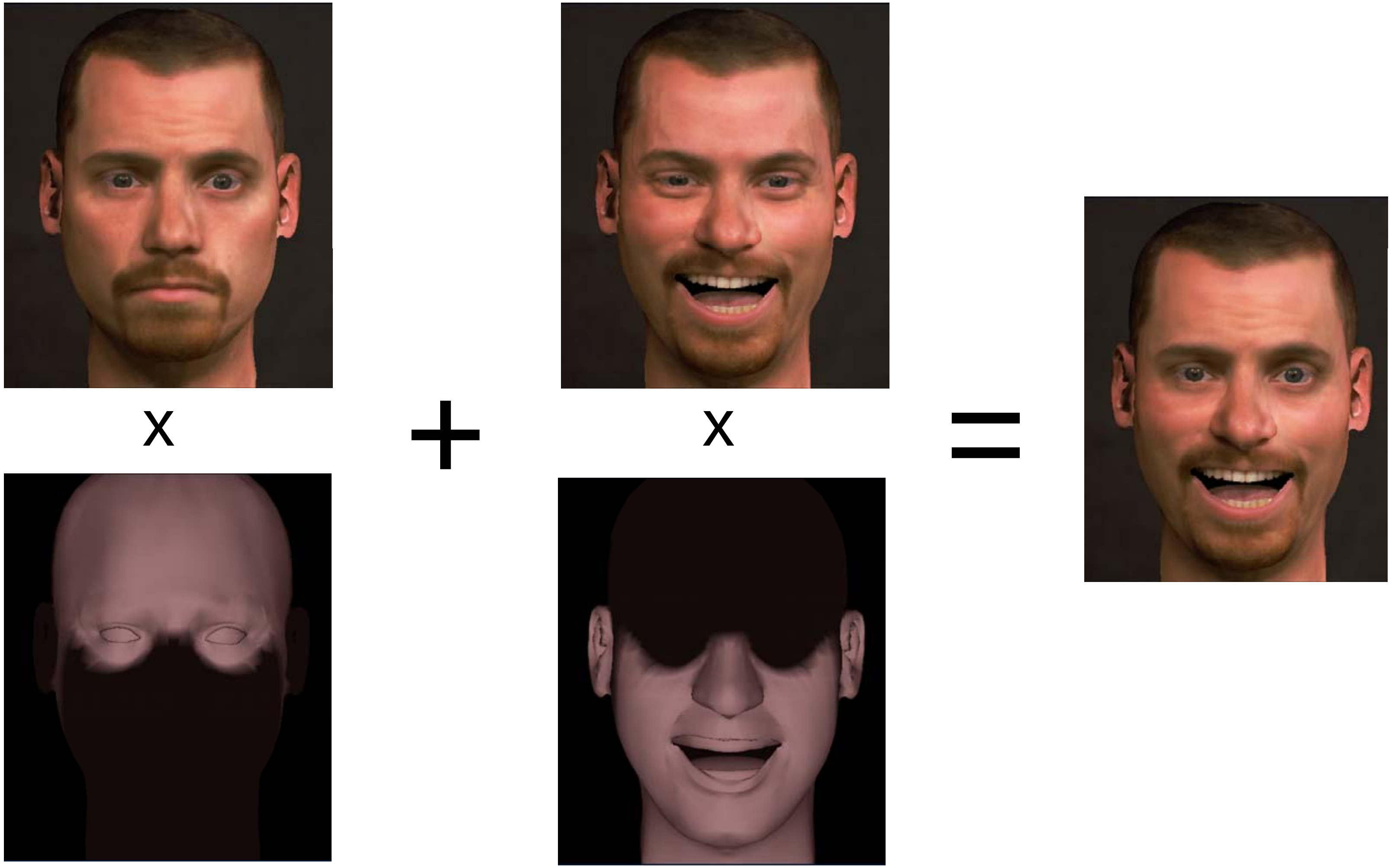
New expressions are created with 3D morphing:

/2 +

/2 =

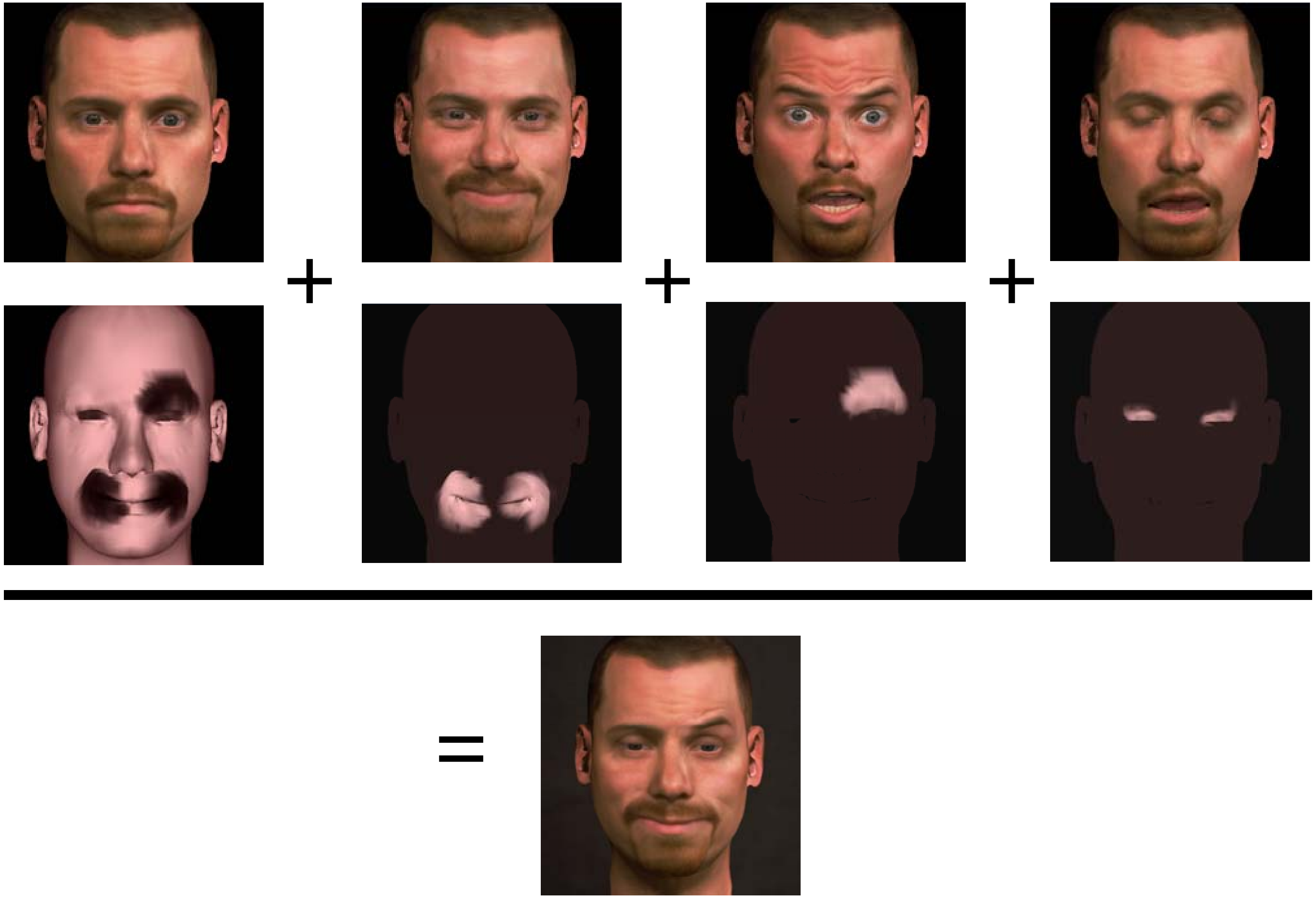
Applying a global blend

Creating new expressions



Applying a region-based blend

Creating new expressions



Using a painterly interface

Drunken smile

Animating between expressions

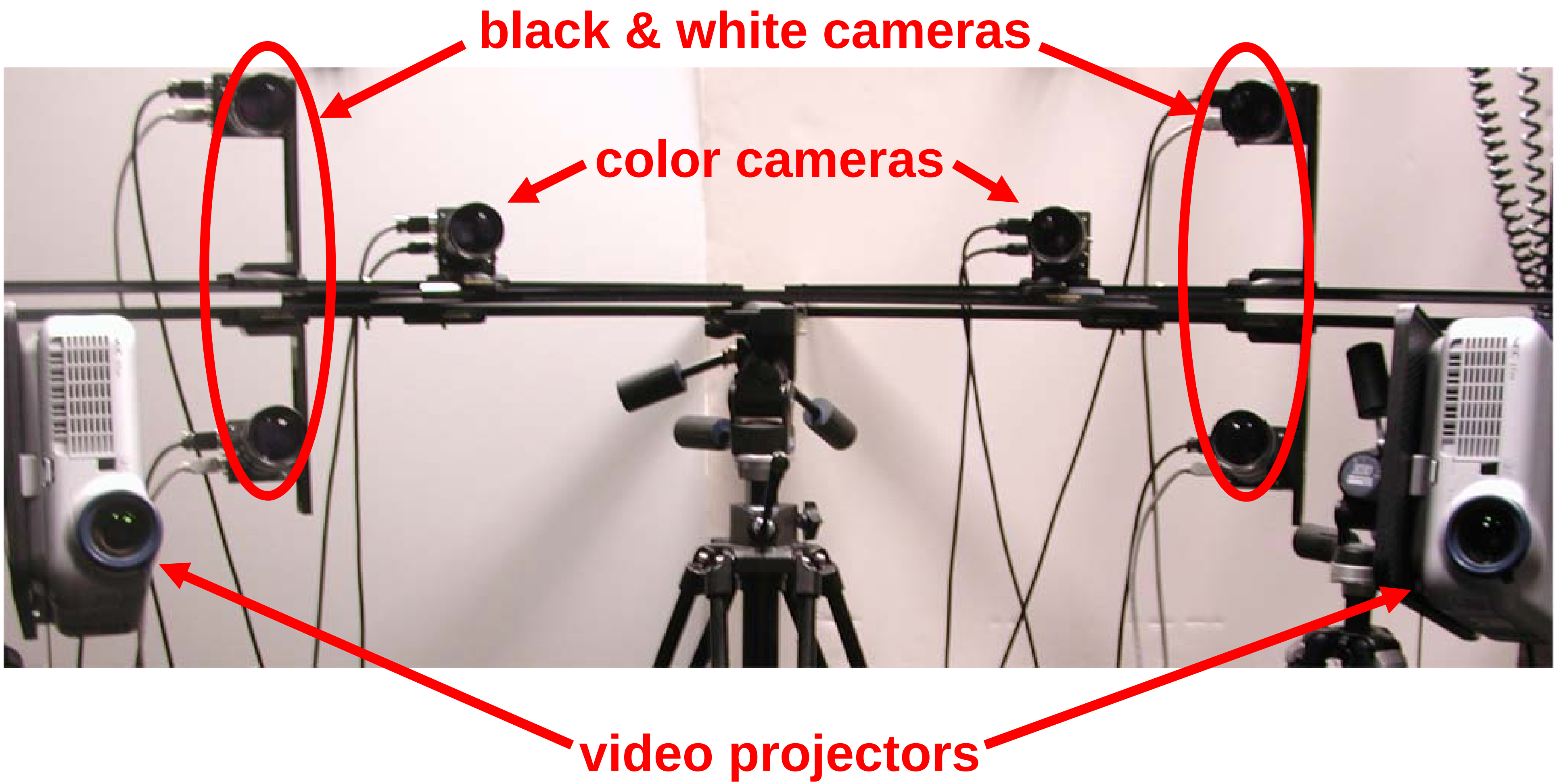
Morphing over time creates animation:

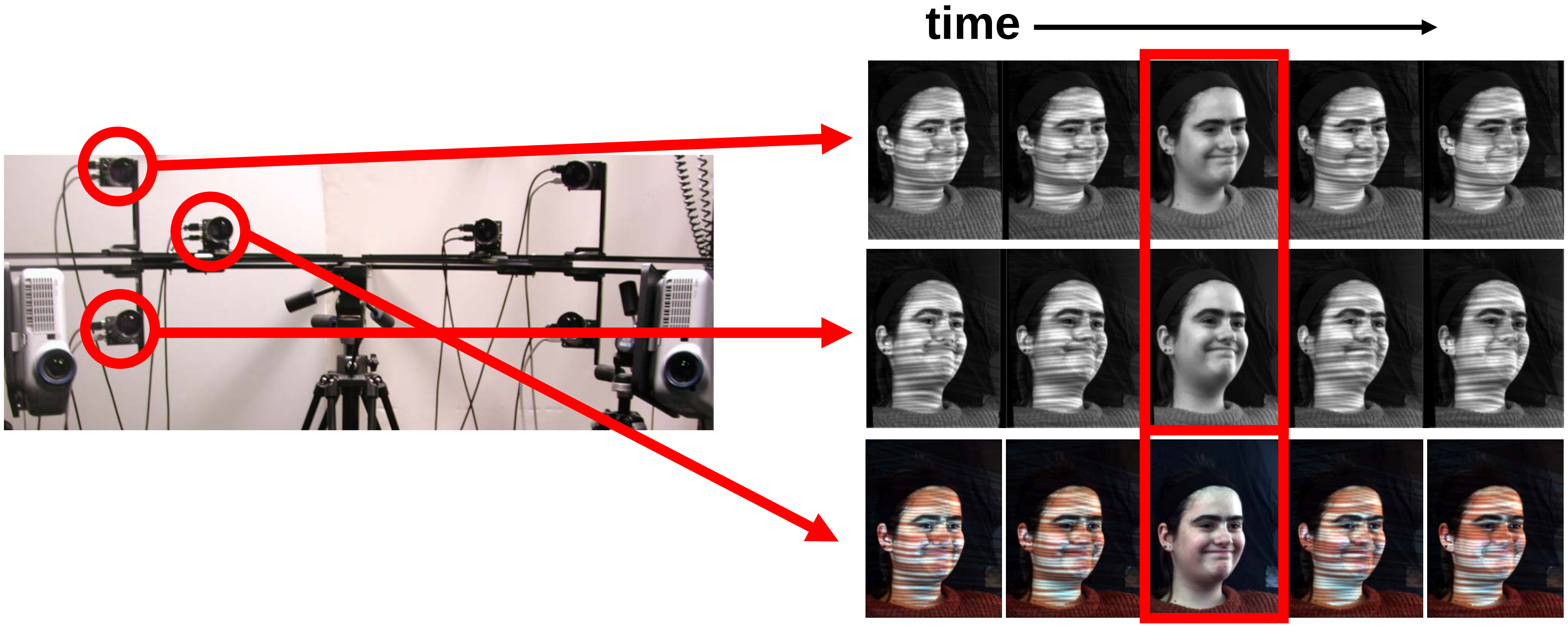
"neutral" → "joy"

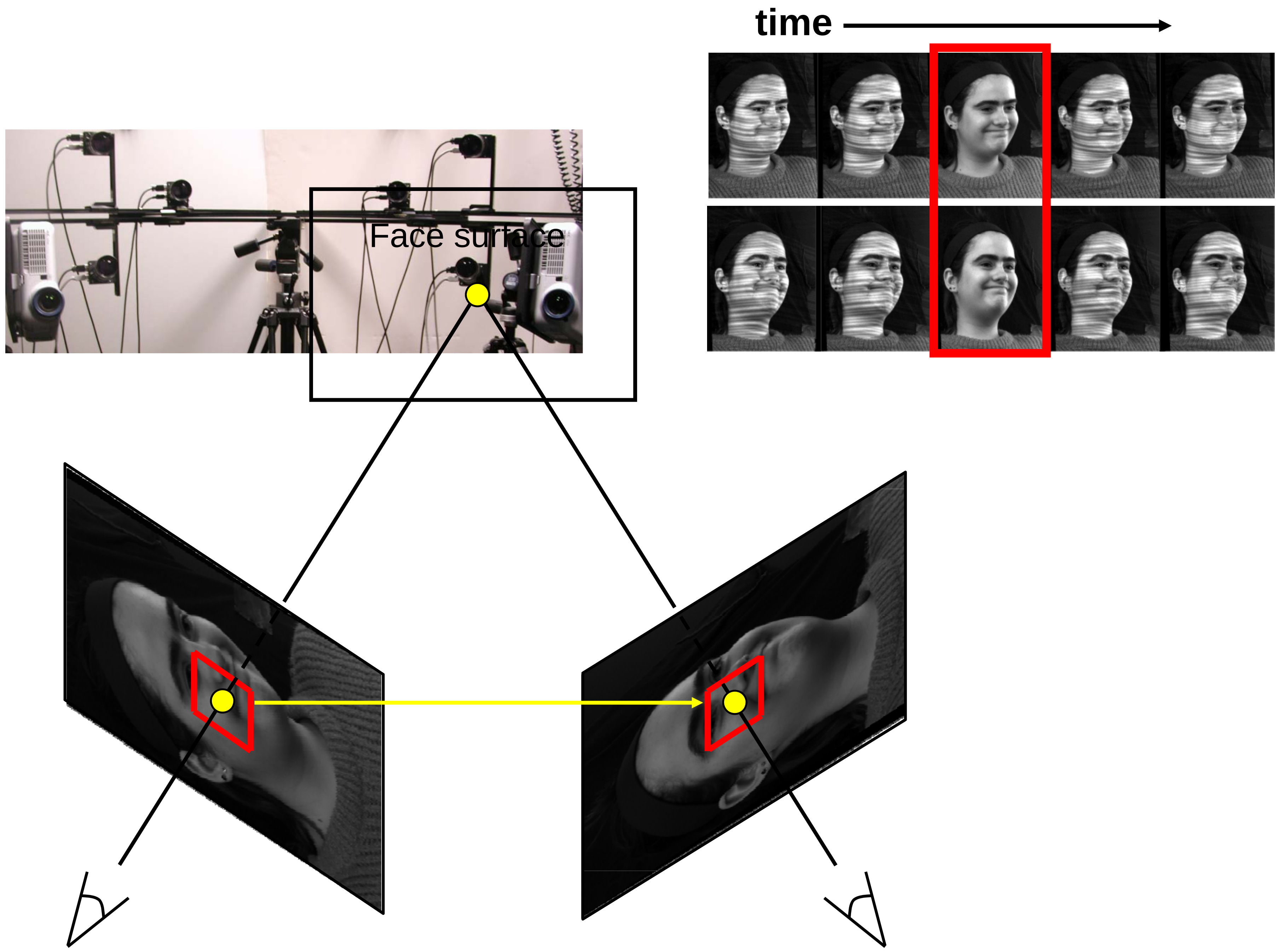
Video

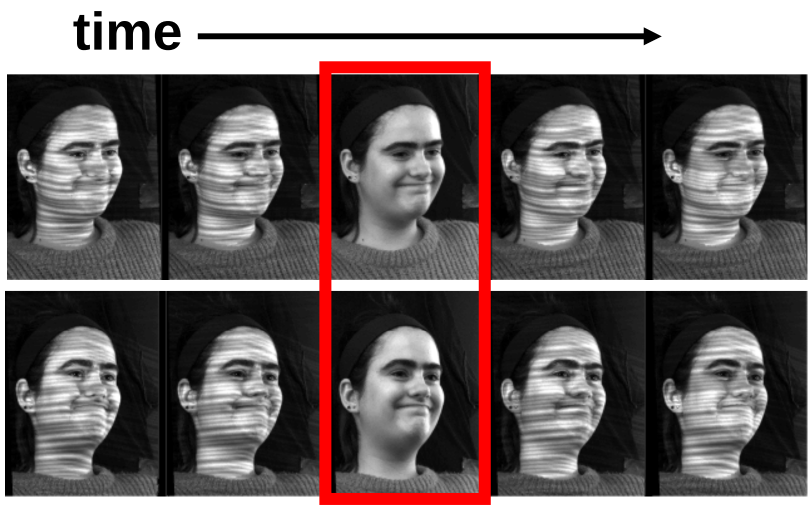
Spacetime faces

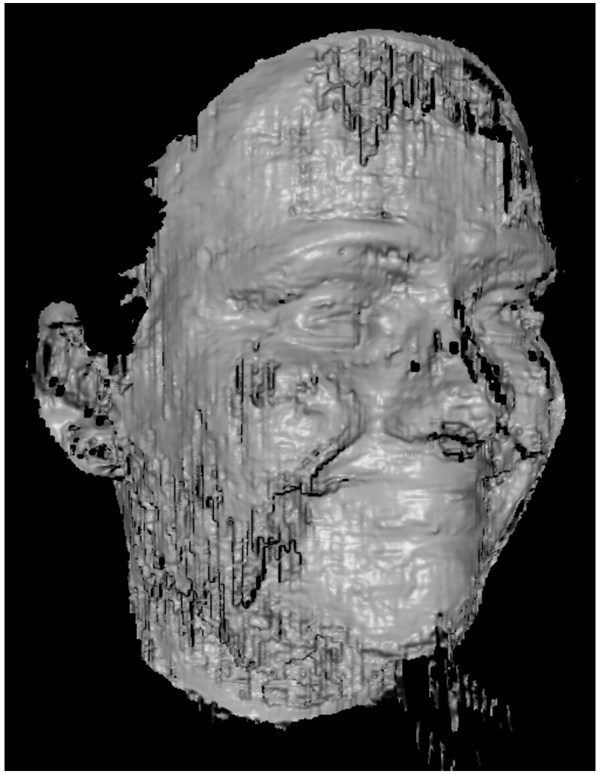
Spacetime faces



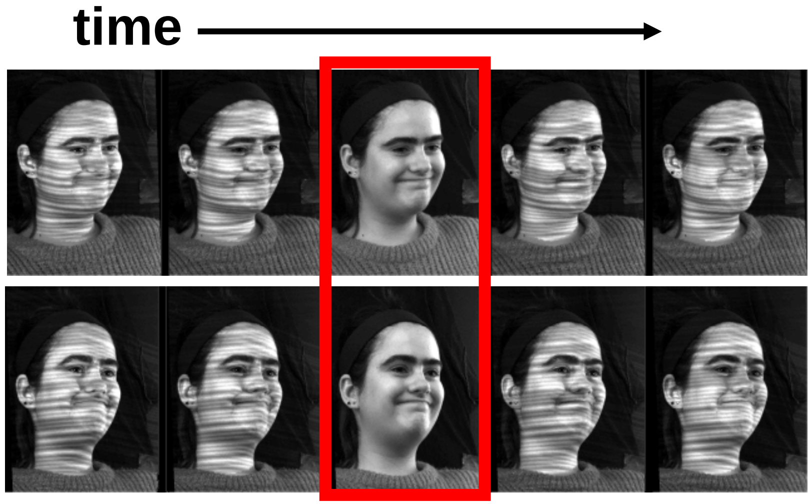

time

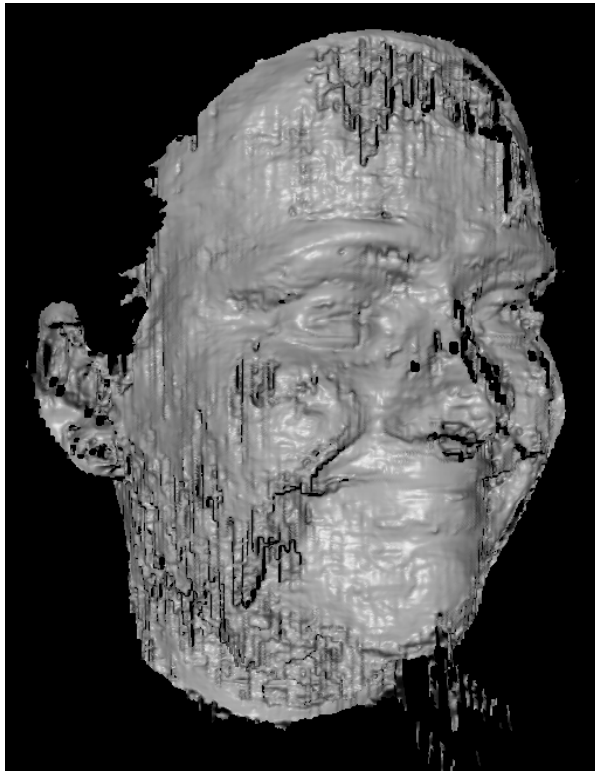

time
Face surface



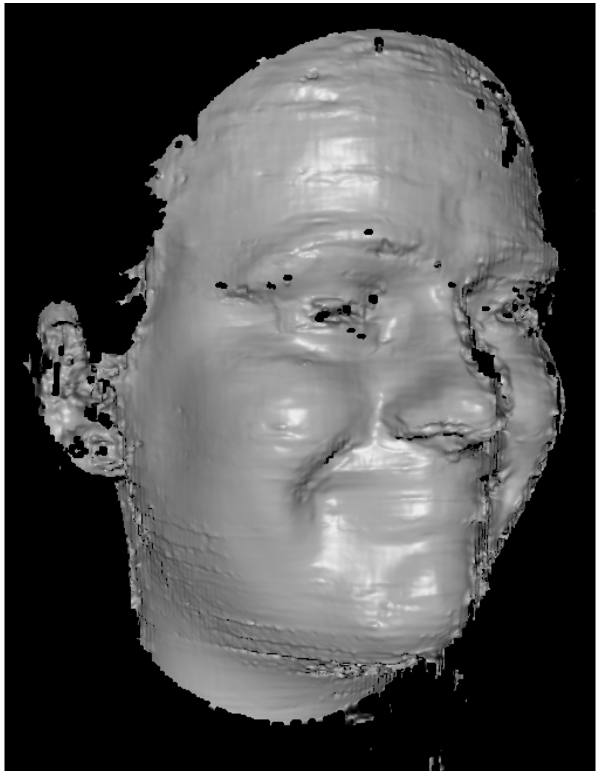


stereo

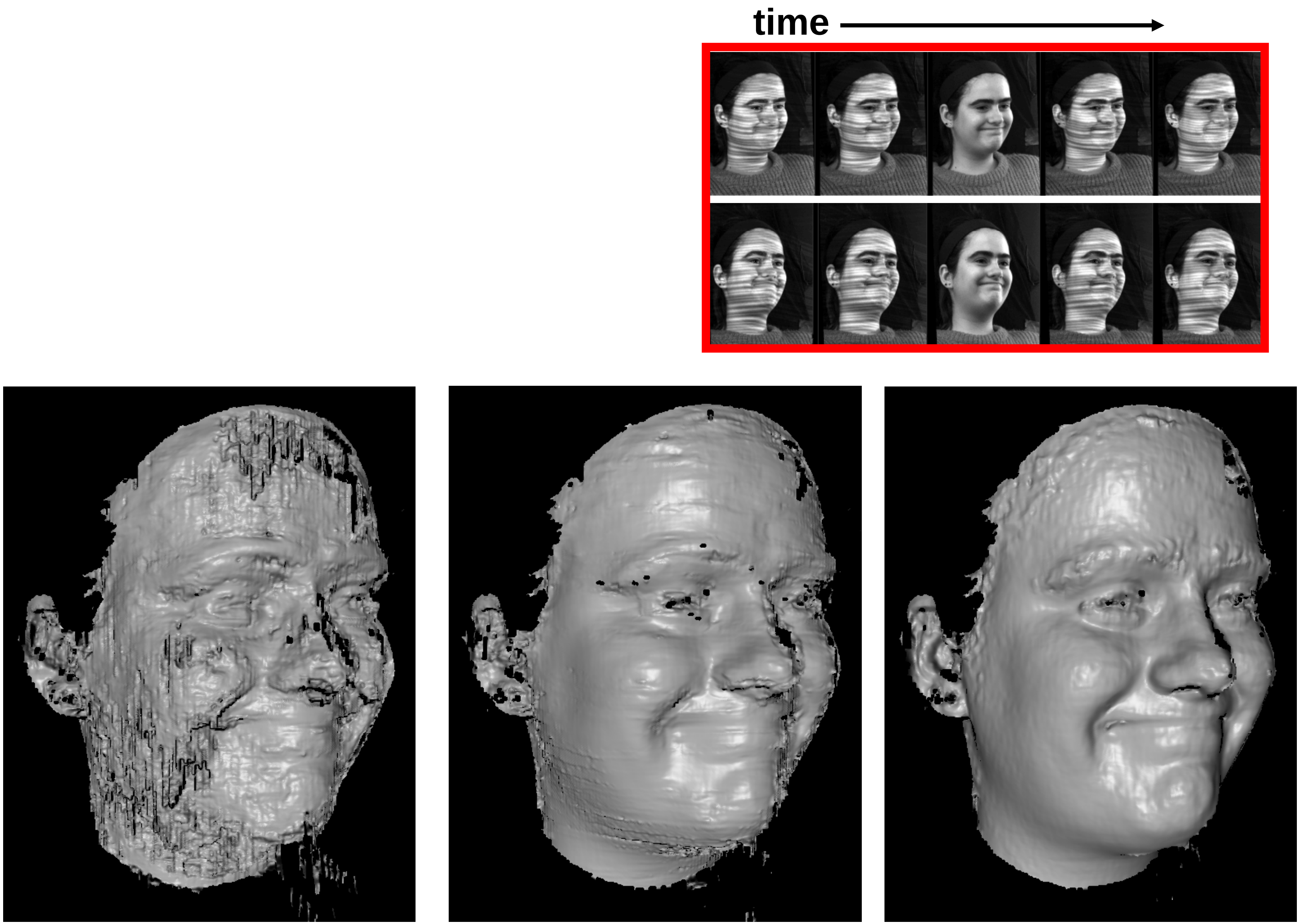




stereo

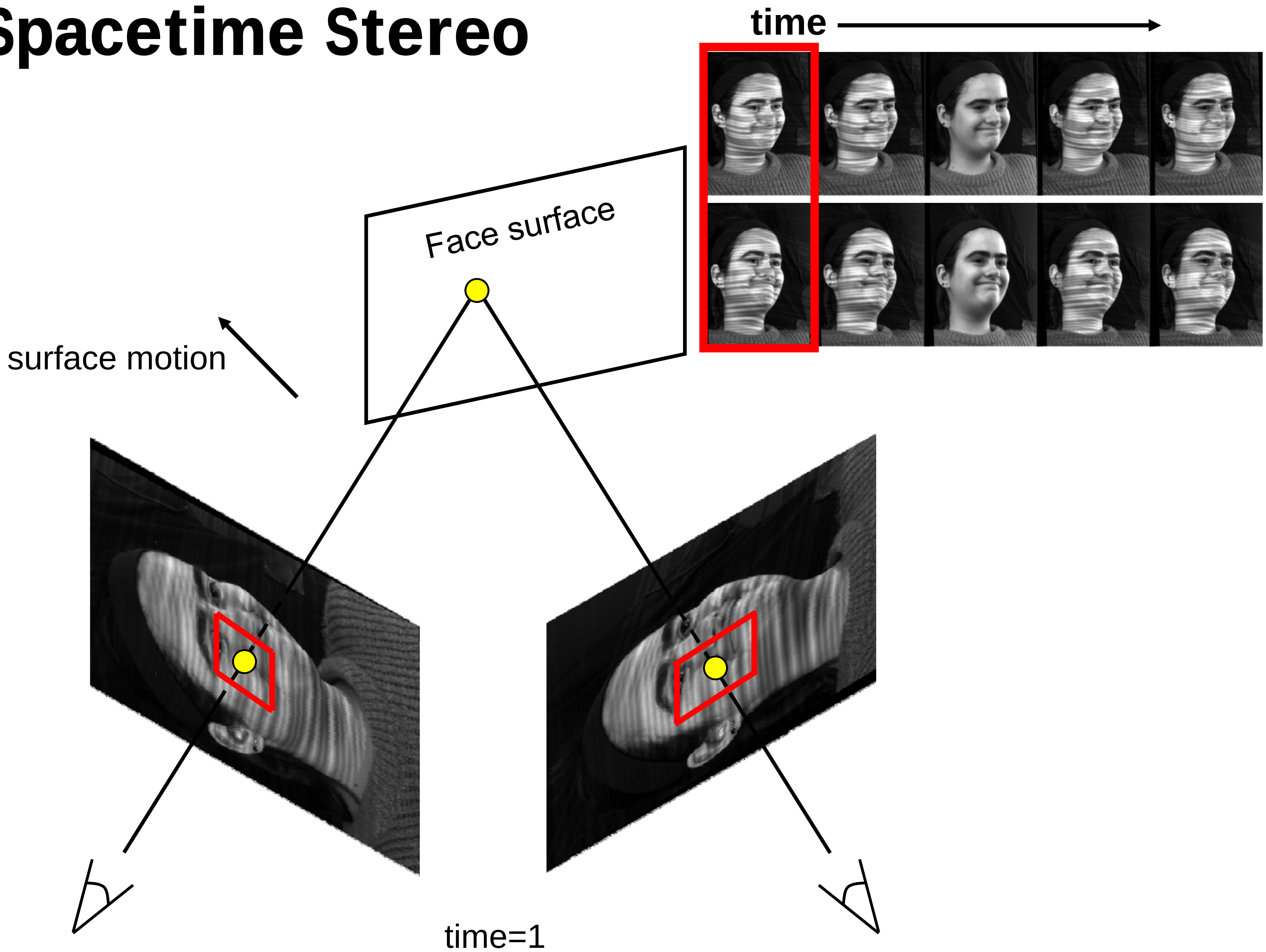


active stereo

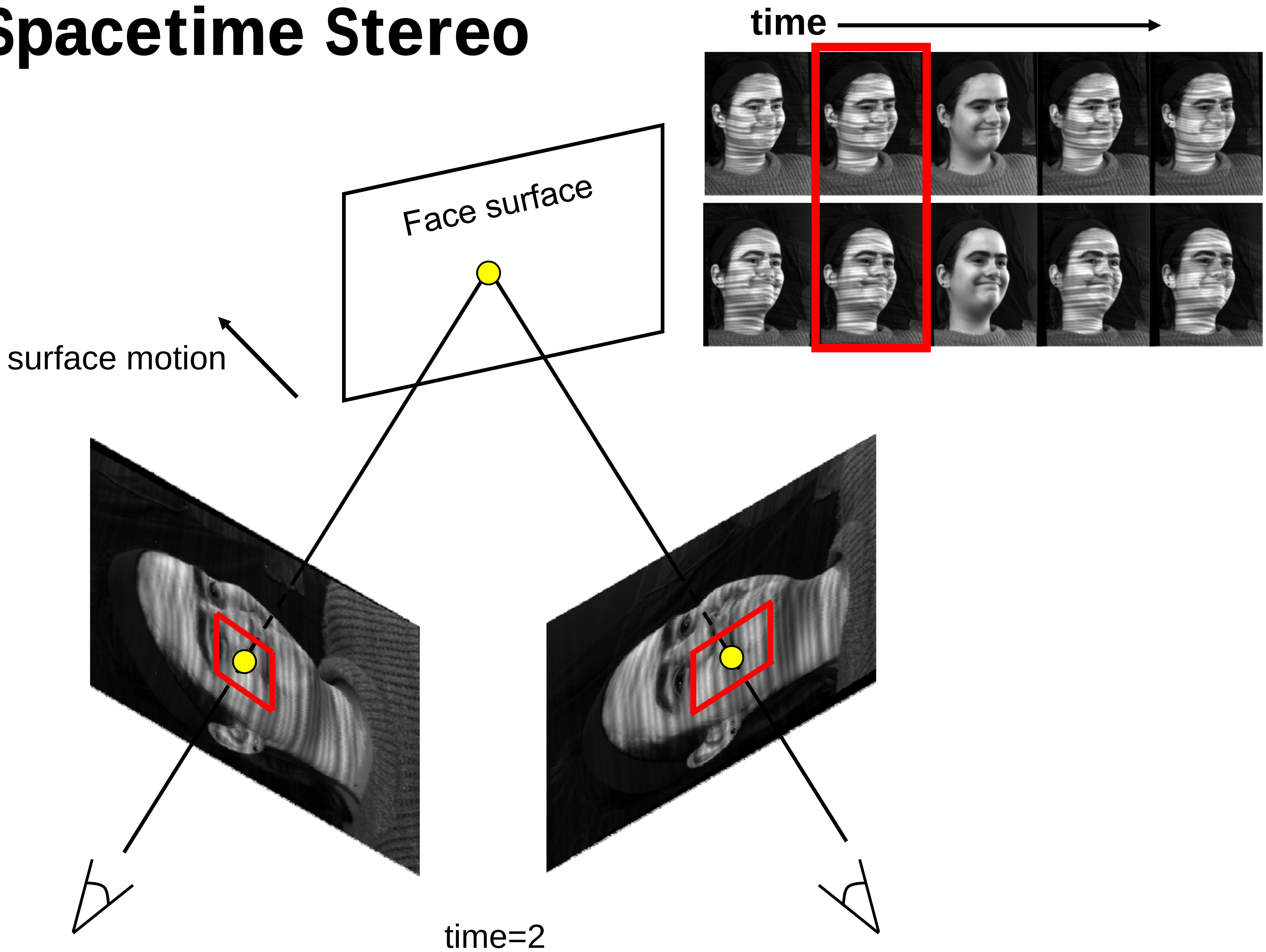


stereo | active stereo | spacetime stereo

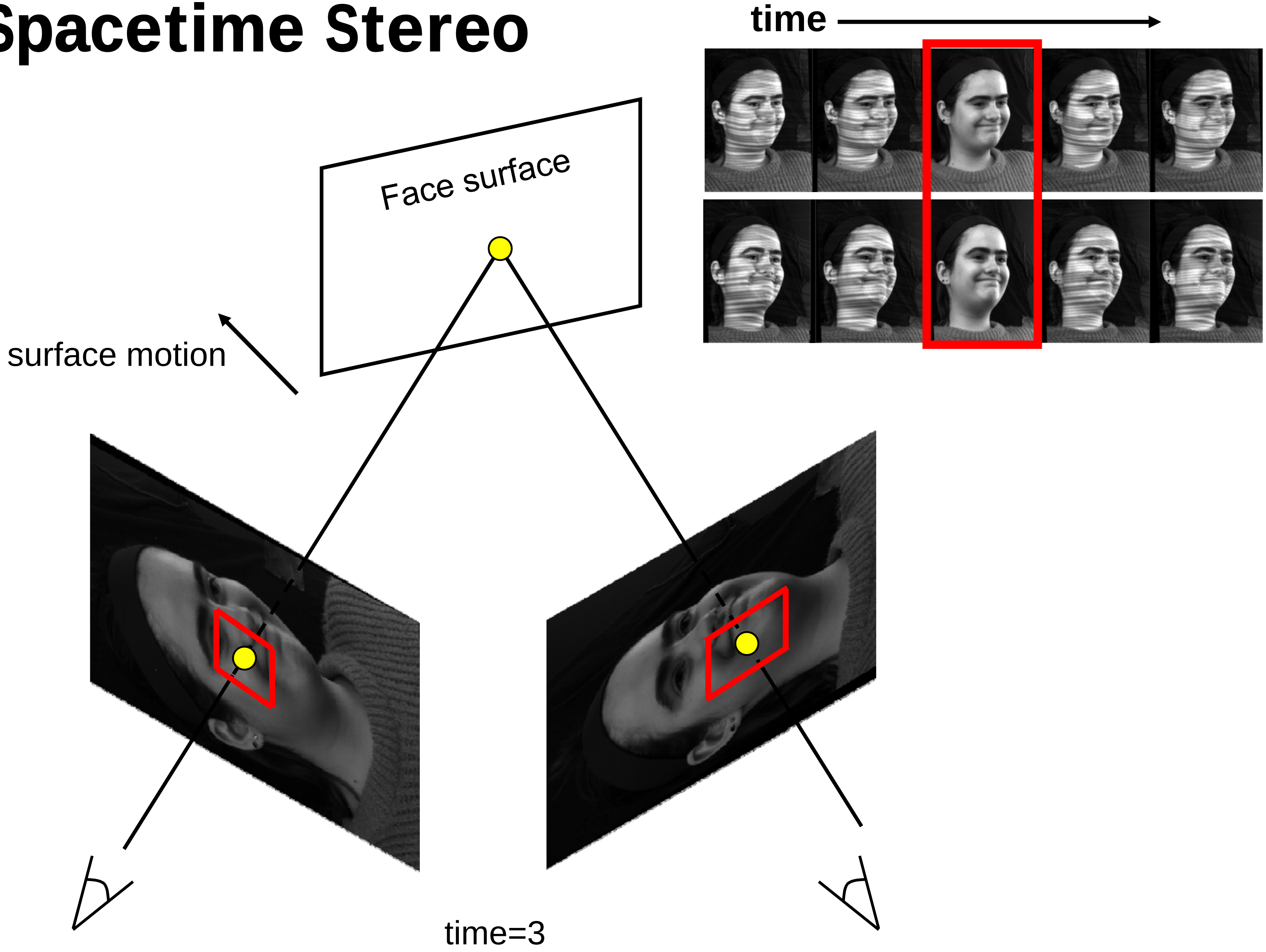
Spacetime Stereo



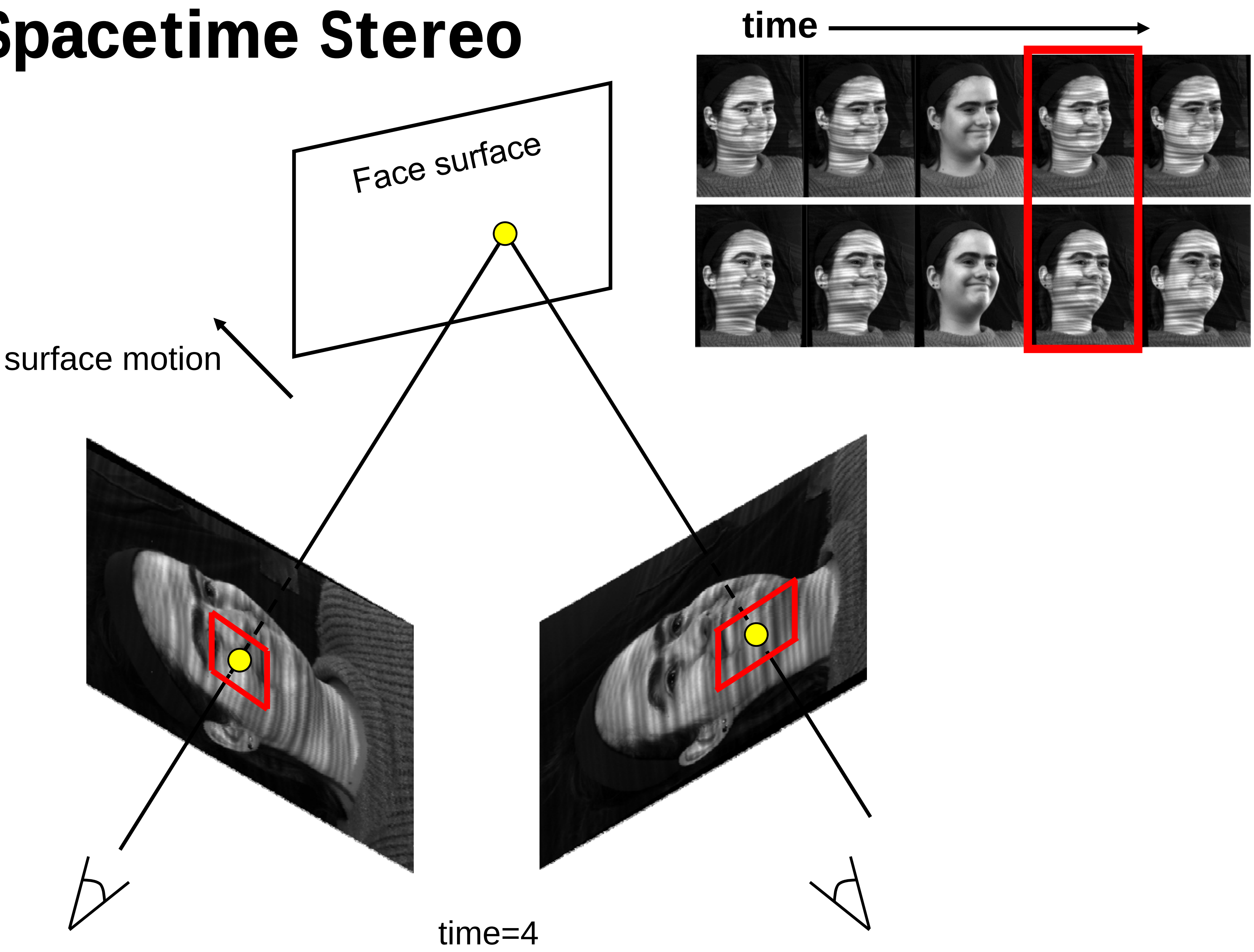
Spacetime Stereo



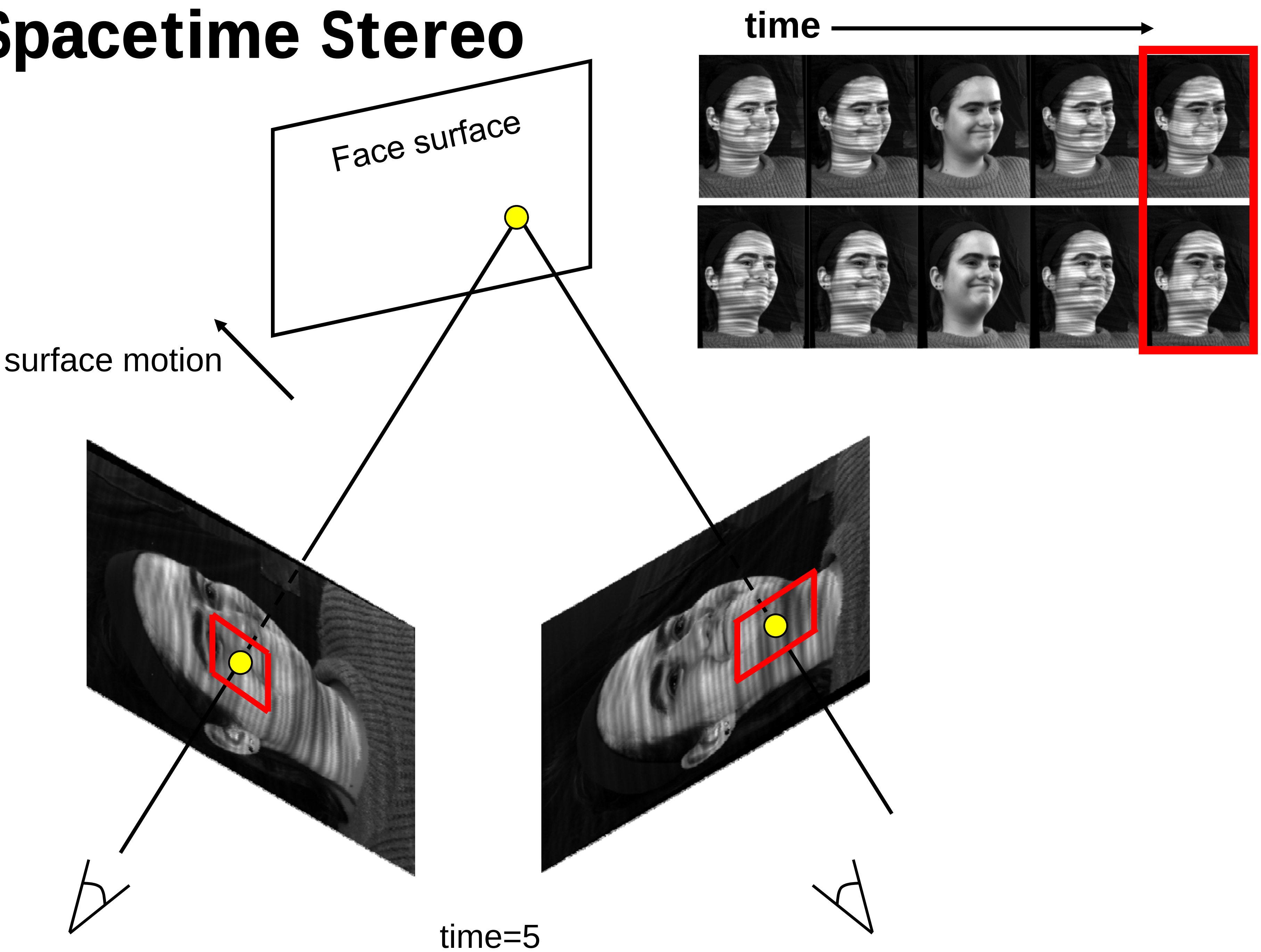
Spacetime Stereo



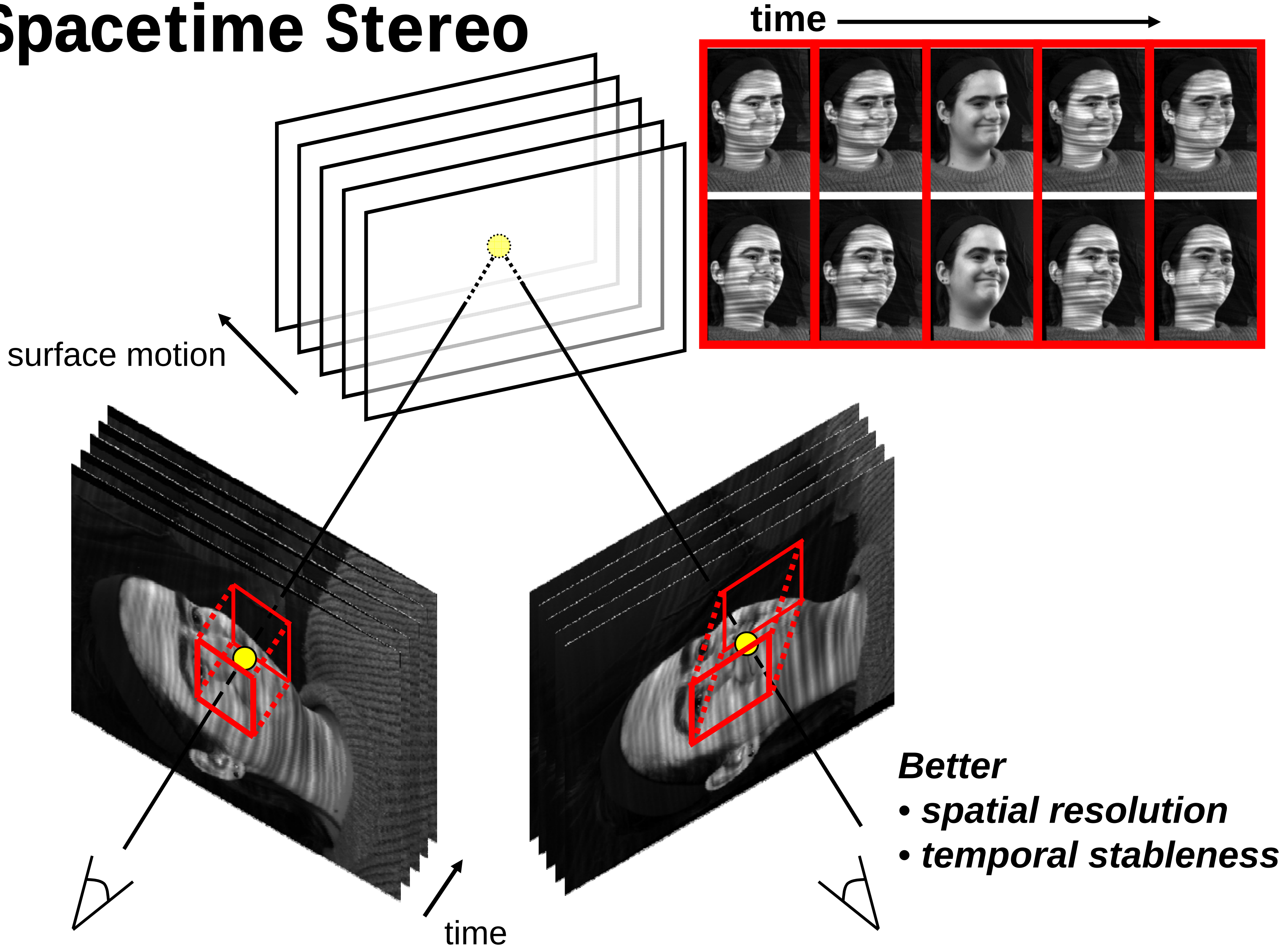
Spacetime Stereo



Spacetime Stereo

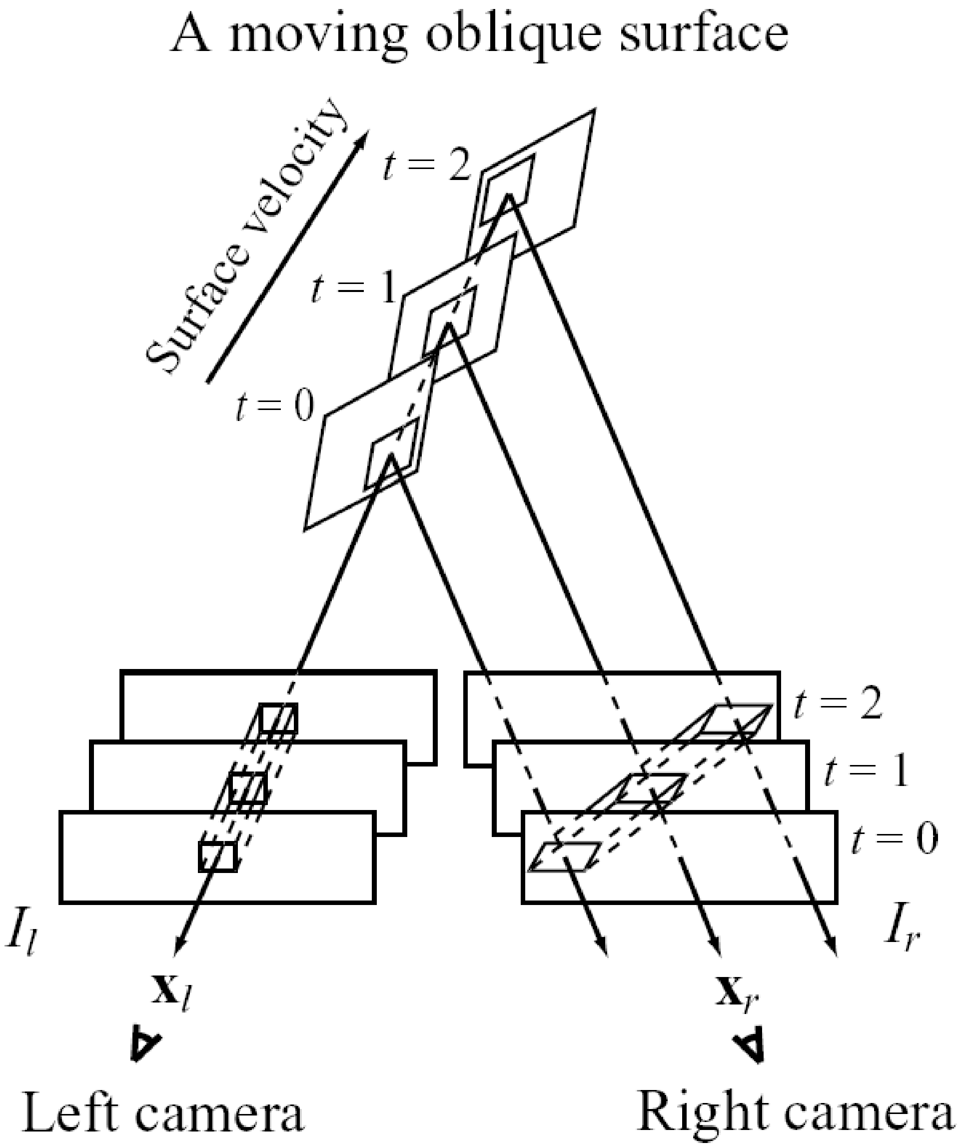


Spacetime Stereo

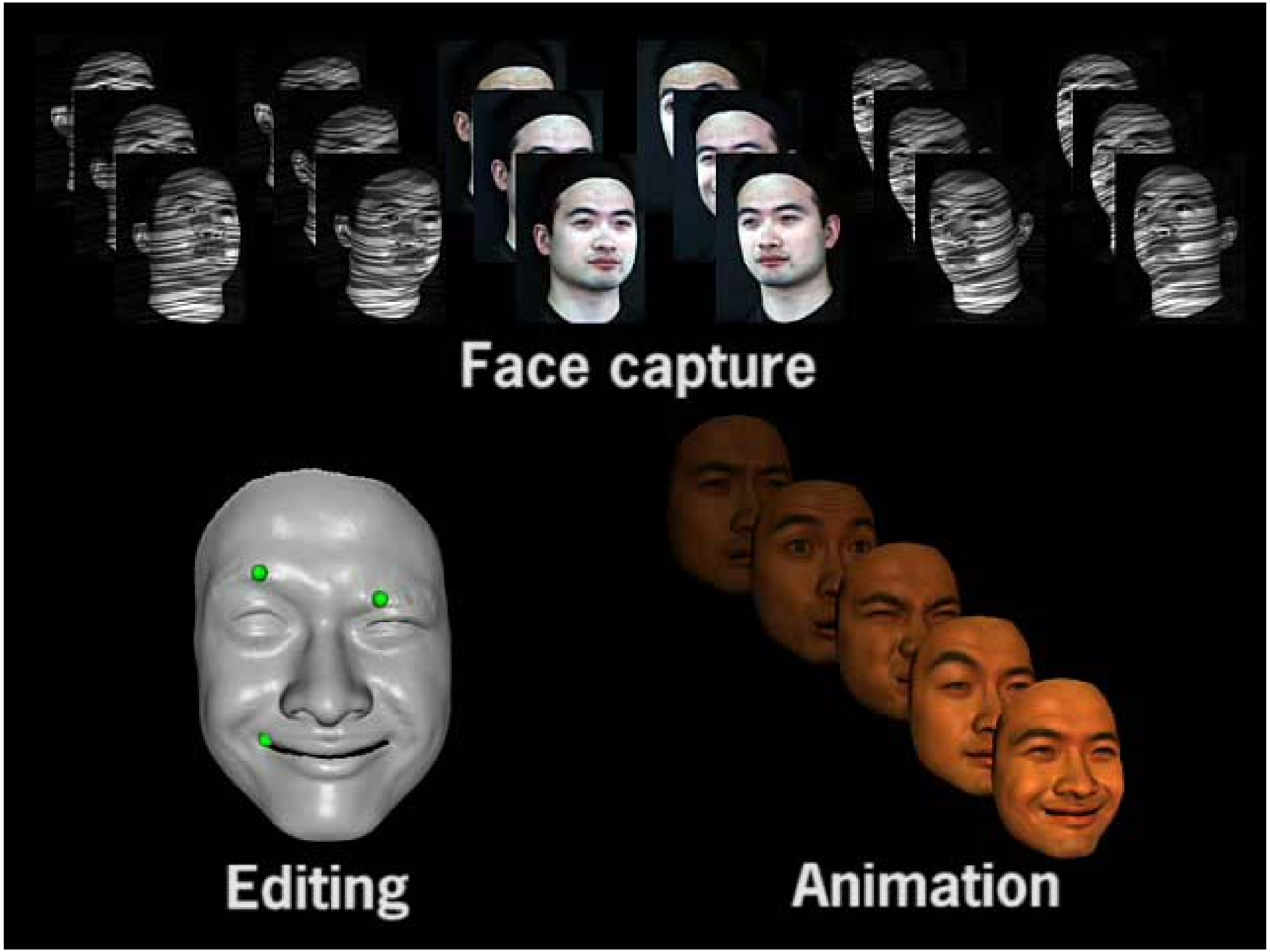


Better
- ***spatial resolution***
- ***temporal stableness***

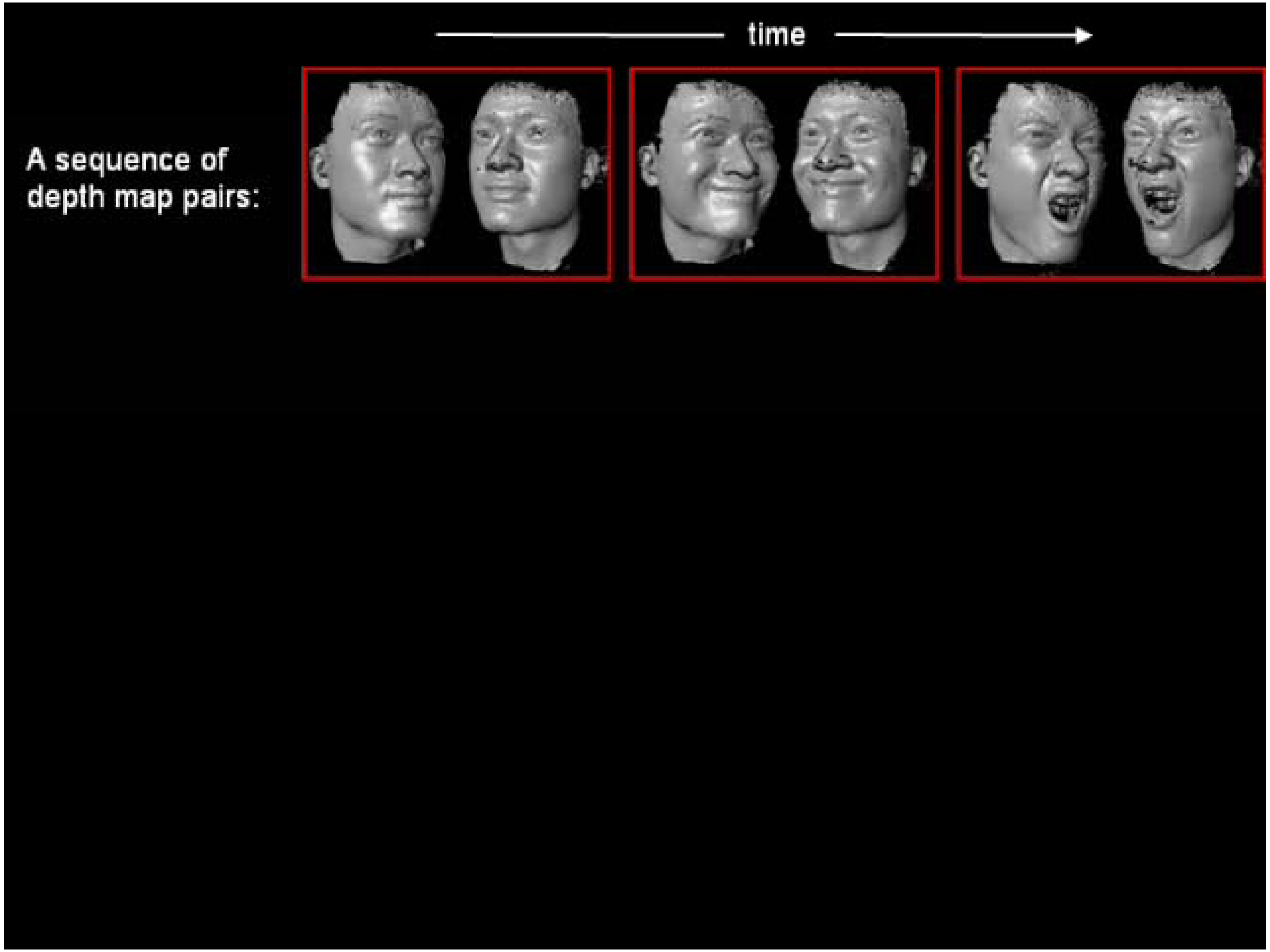
Spacetime stereo matching



Video

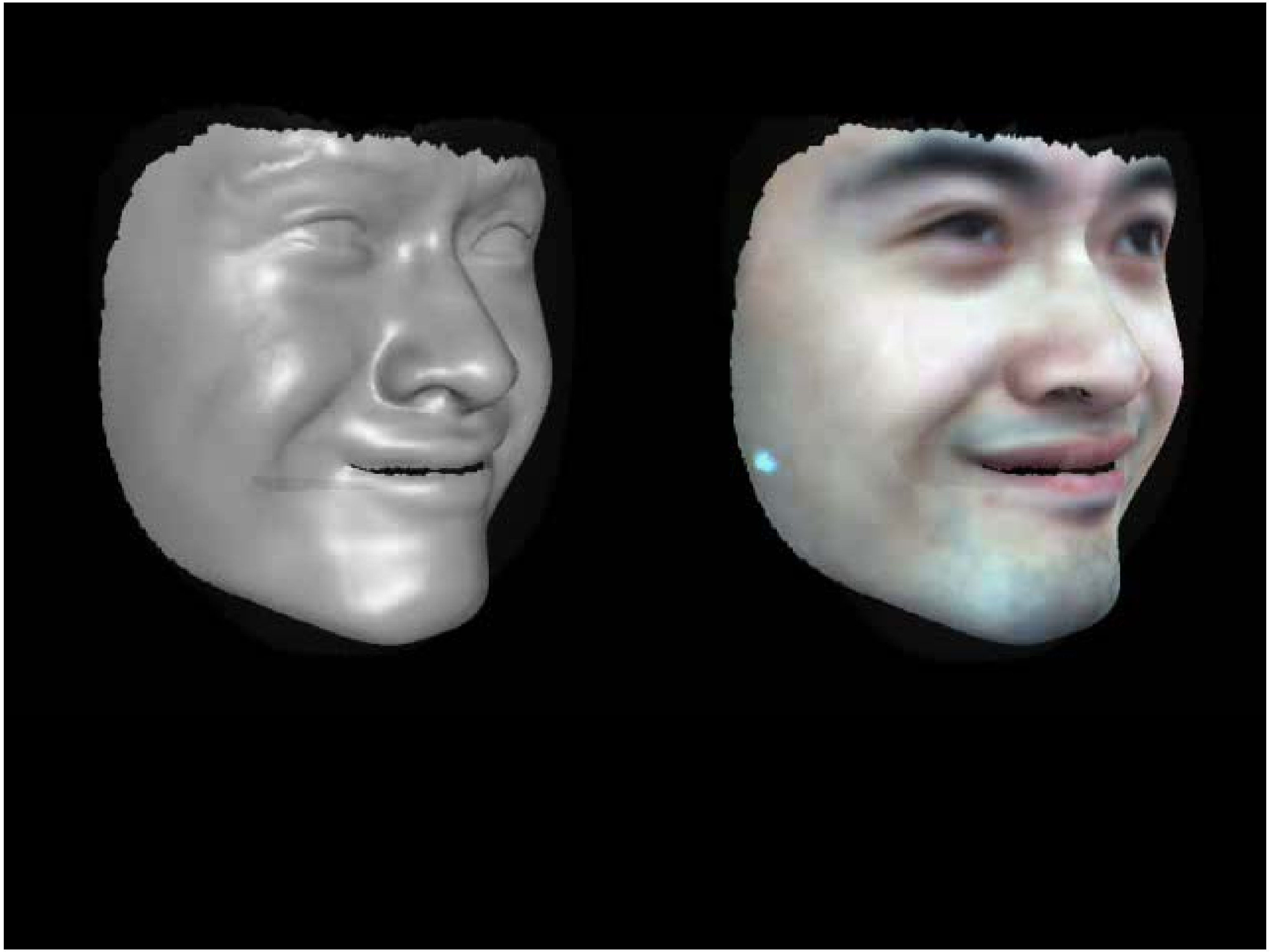


Fitting

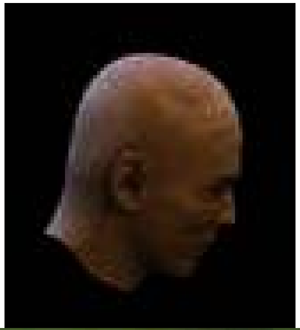


FaceIK

Animation



3D face applications: The one



3D face applications: Gladiator

extra 3M

Statistical methods

Statistical methods

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

$$z^* = \max_z P(z \mid y)$$

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

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

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

Statistical methods

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 8×10^{11} blocks per day per person.

Generic priors

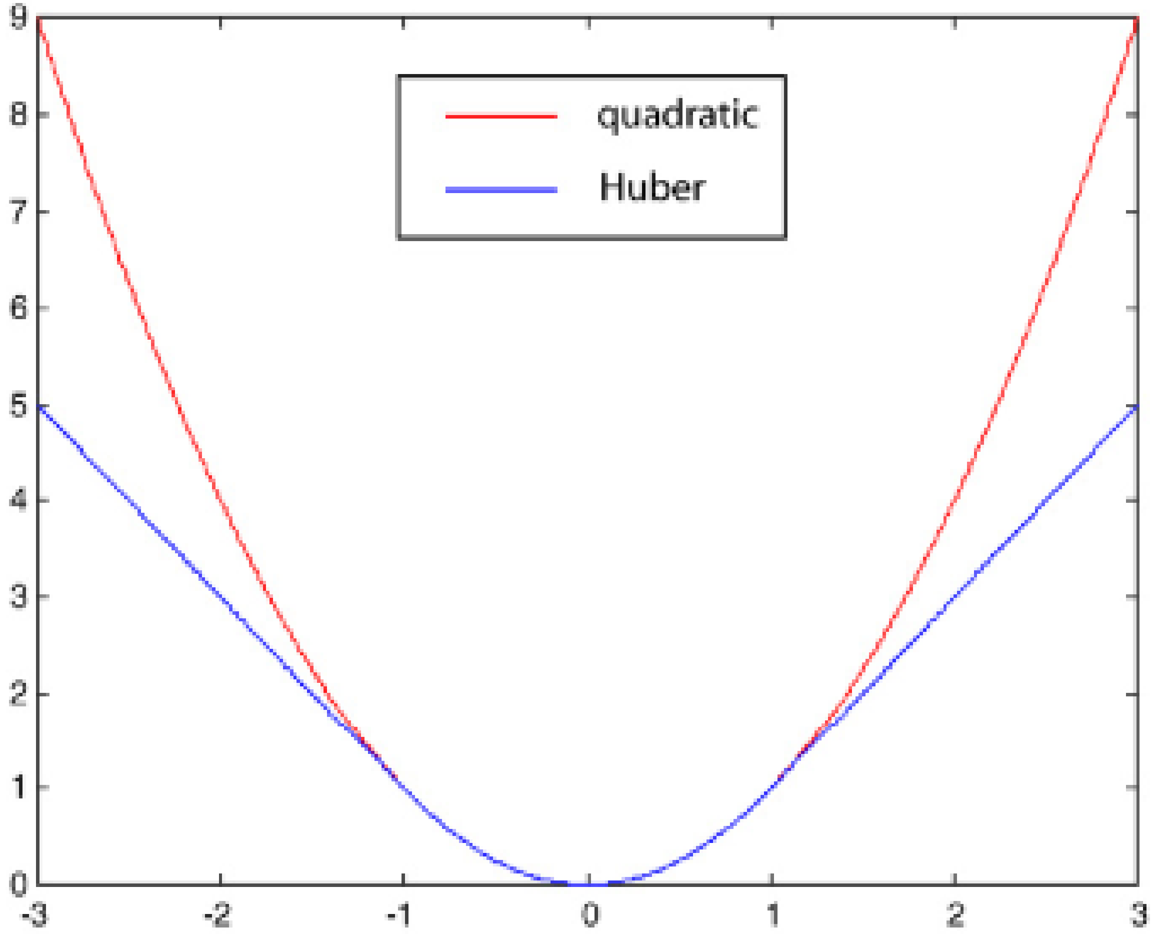
"Smooth images are good images."

$$L(z) = \sum_{x} \rho(V(x))$$

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

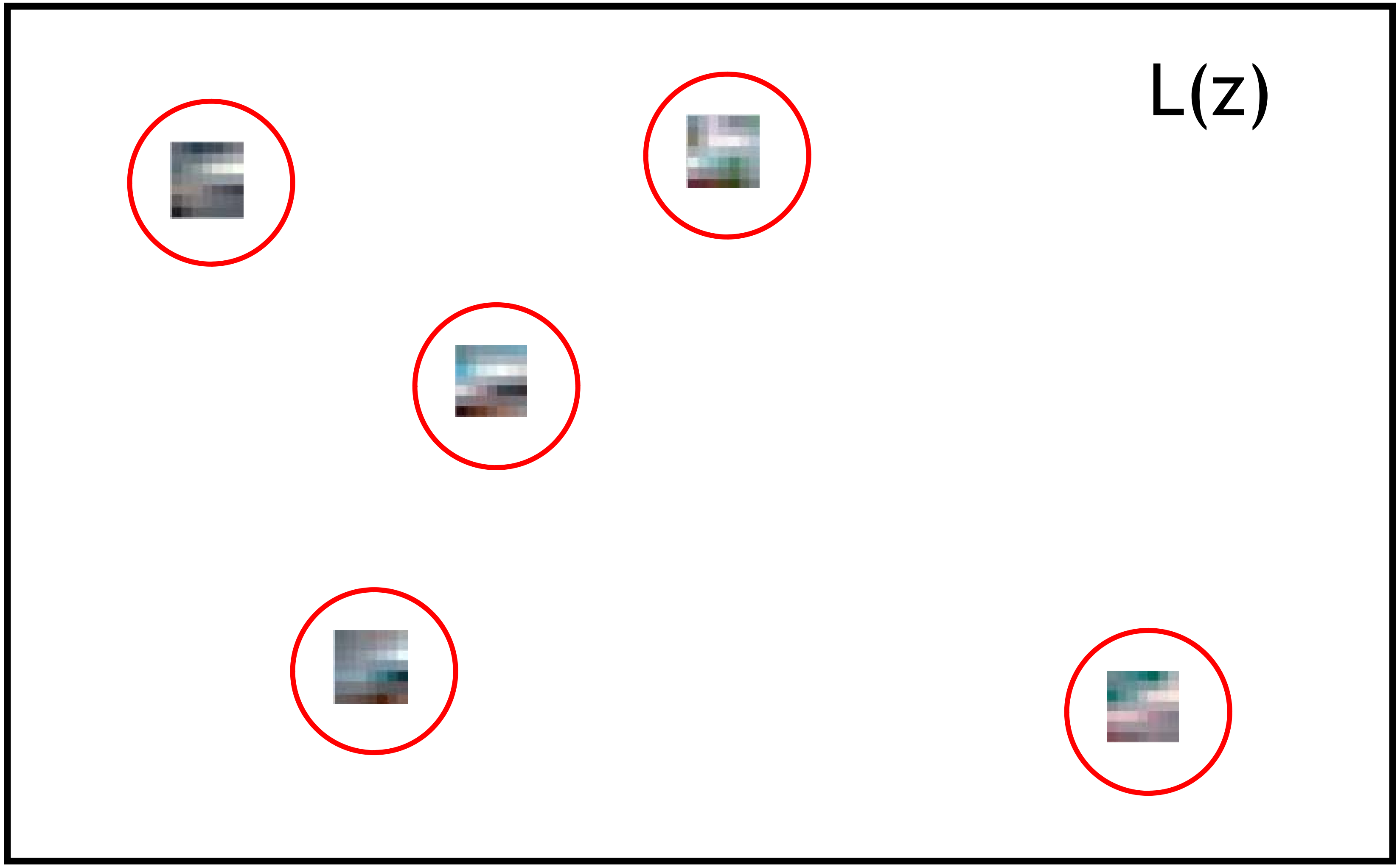


Example-based priors

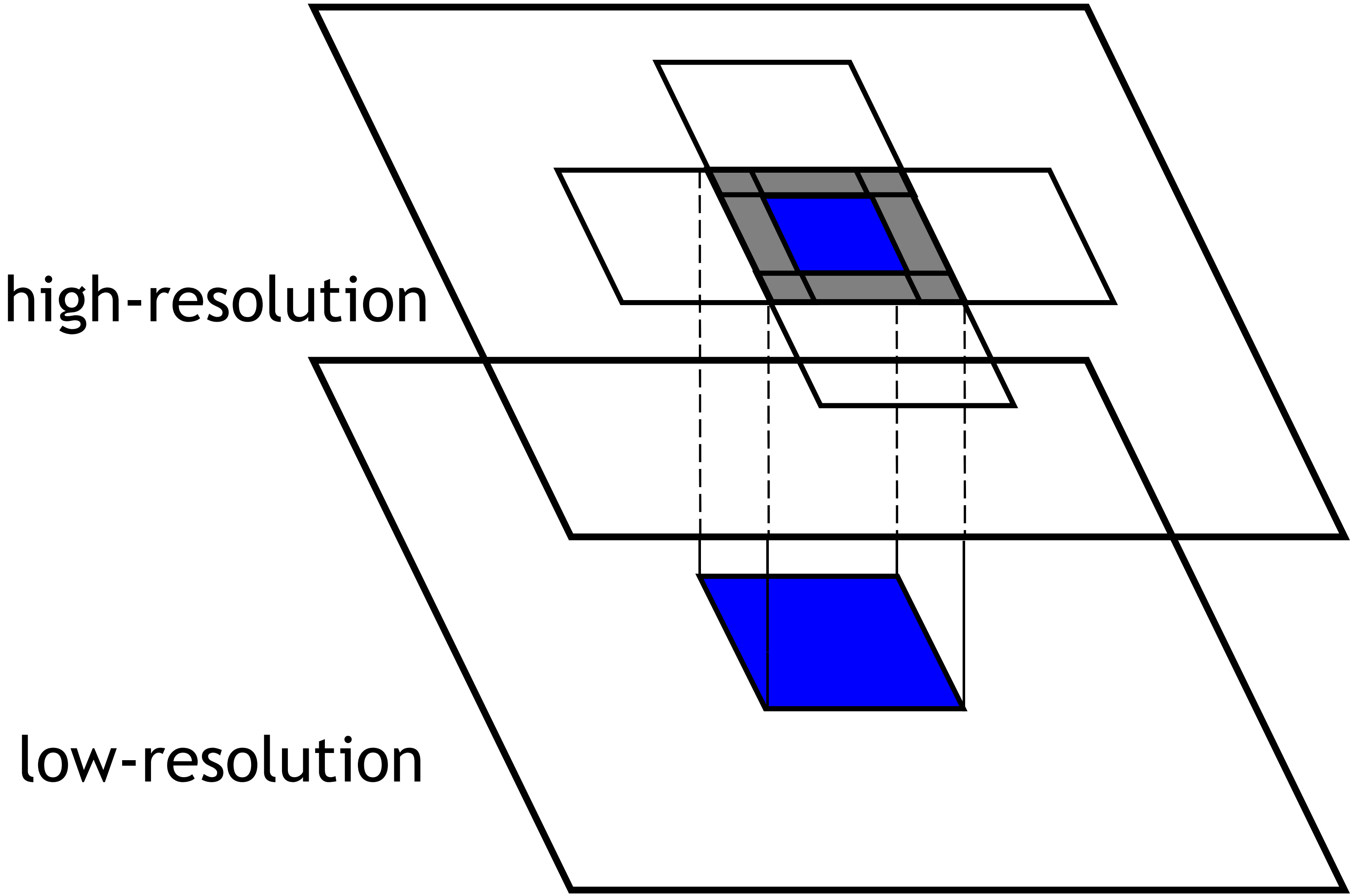
"Existing images are good images."

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 …"

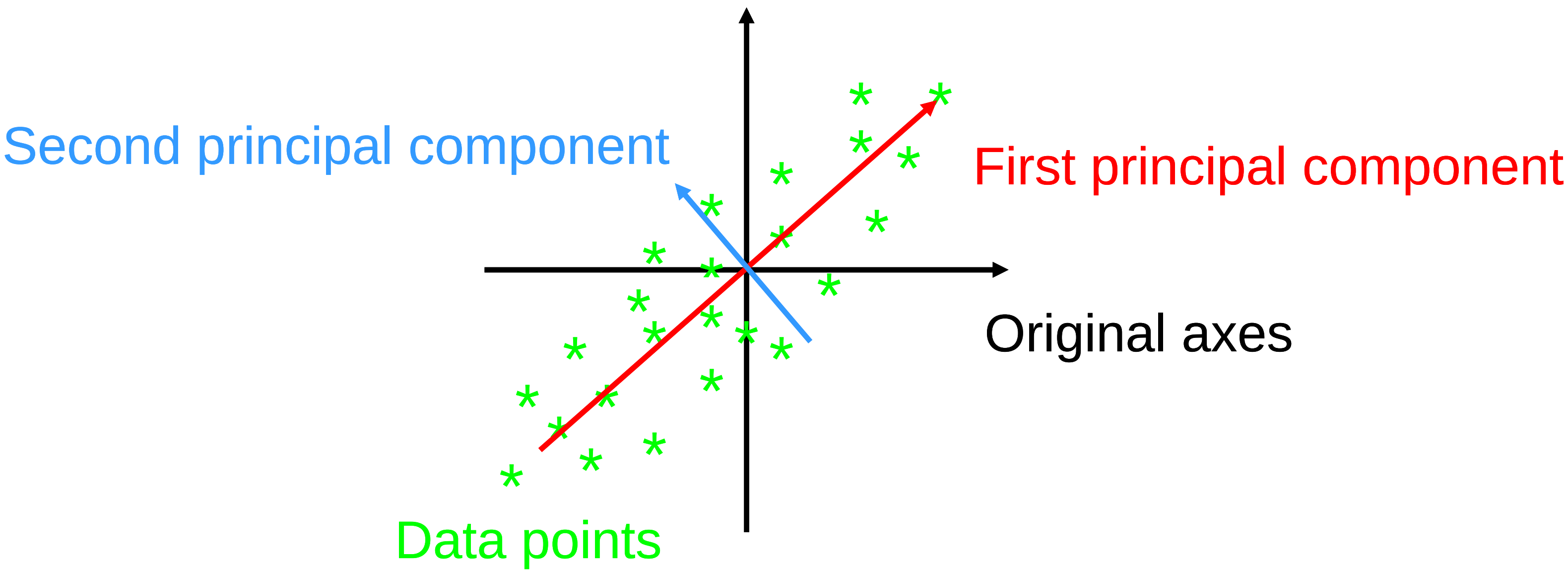
Parametric model $\quad Z=WX+\mu \quad L(X)$

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

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

PCA

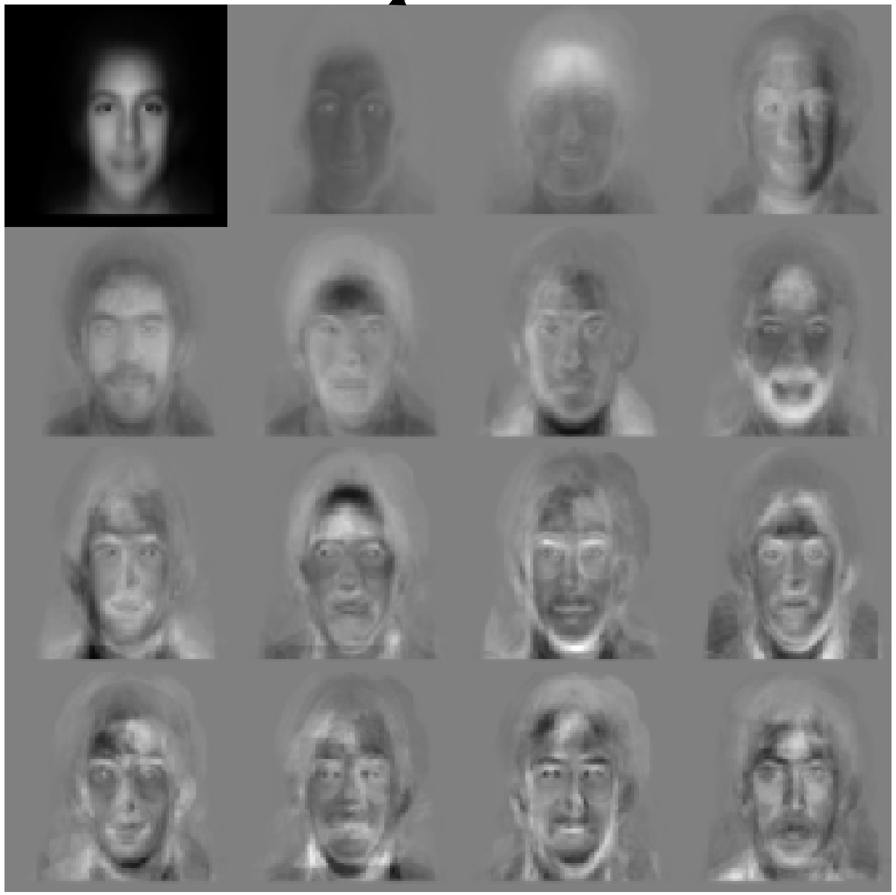
- Principal Components Analysis (PCA): approximating a high-dimensional data set with a lower-dimensional subspace



PCA on faces: “eigenfaces”

Average face

First principal component



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

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

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

Super-resolution

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

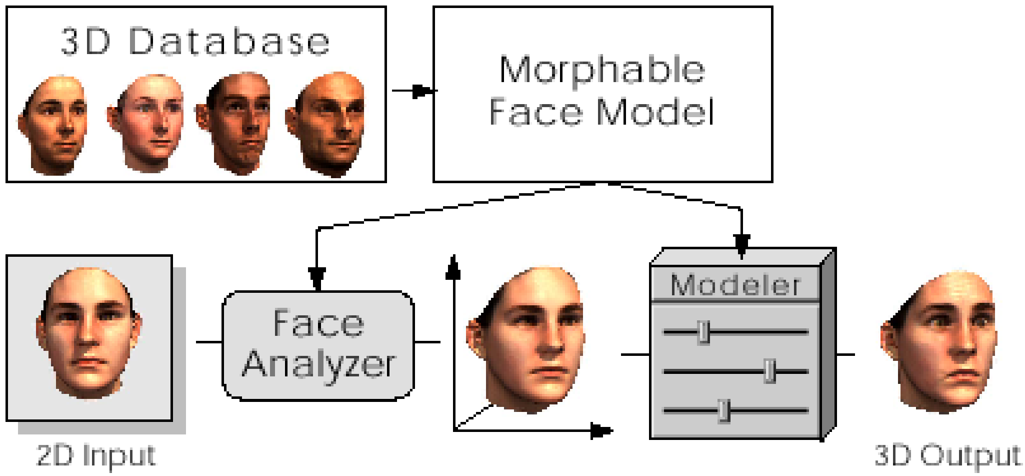
(a) Input low 24×32 (b) Our results (c) Cubic B-Spline

(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

shape examplars texture examplars

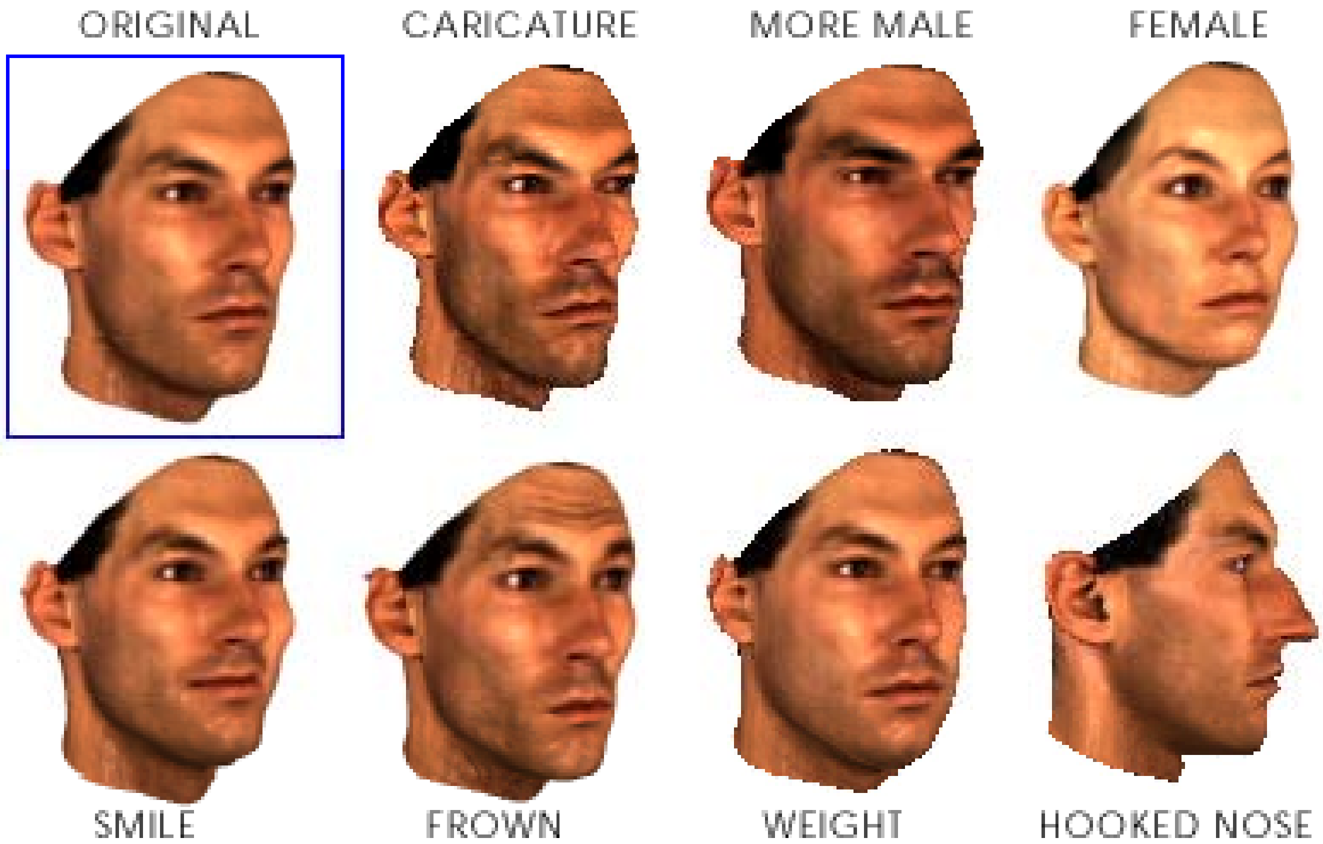
$$S_{model} = \overline{S} + \sum_{i=1}^{m-1} \alpha_i s_i \,, \quad T_{model} = \overline{T} + \sum_{i=1}^{m-1} \beta_i t_i \,, \qquad (1)$$

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

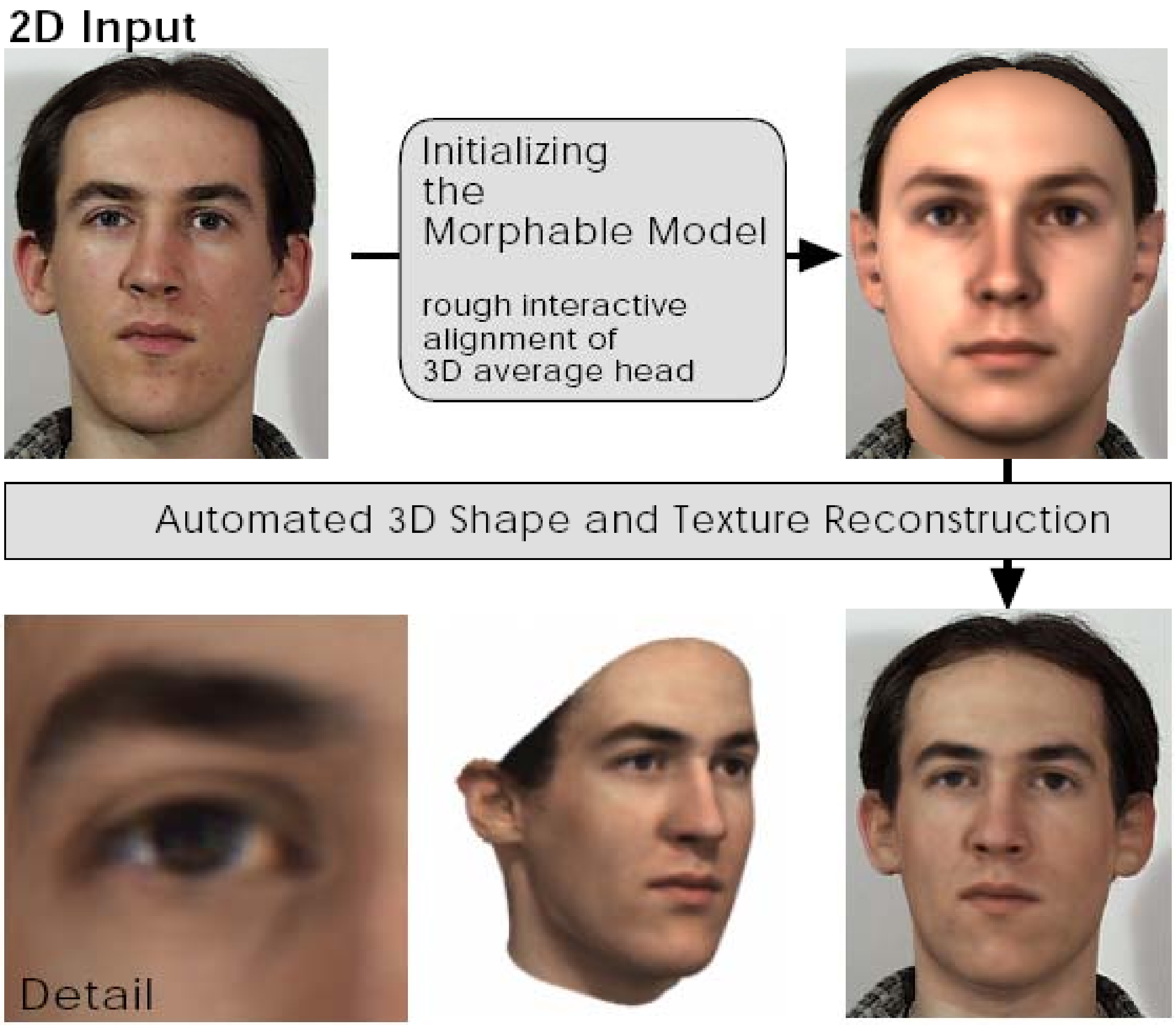
$$p(\vec{\alpha}) \sim exp[-\frac{1}{2} \sum_{i=1}^{m-1} (\alpha_i / \sigma_i)^2], \qquad (2)$$

Morphable model of 3D faces

- Adding some variations



Reconstruction from single image



$\alpha_j \quad \beta_j$

Rendering must be similar to the input if we guess right

Reconstruction from single image

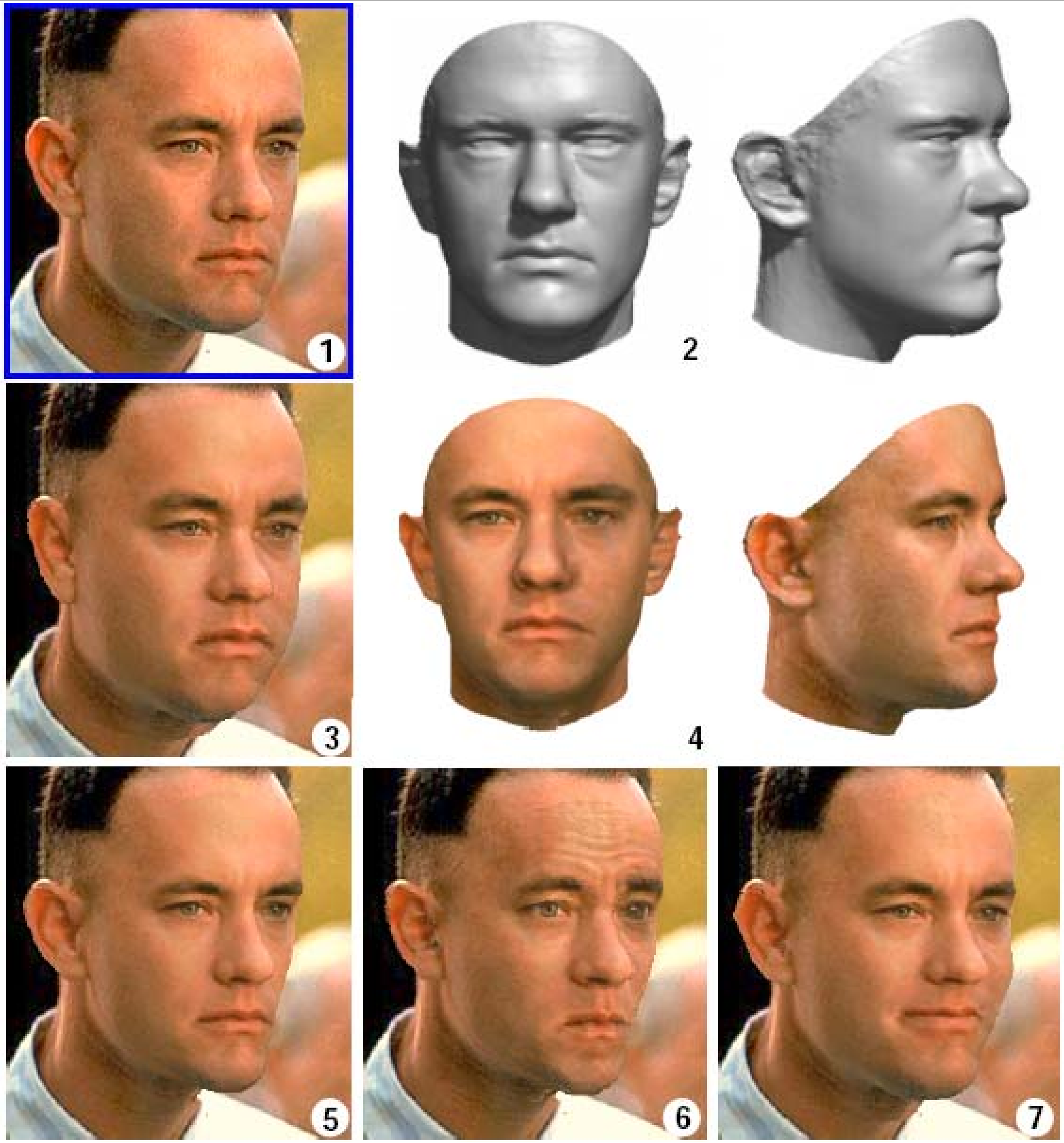
$$E = \frac{1}{\sigma_N^2}\boxed{E_I} + \boxed{\sum_{j=1}^{m-1}\boxed{\frac{\alpha_j^2}{\sigma_{S,j}^2}} + \sum_{j=1}^{m-1}\boxed{\frac{\beta_j^2}{\sigma_{T,j}^2}} + \sum_j \boxed{\frac{(\rho_j - \bar{\rho}_j)^2}{\sigma_{\rho,j}^2}}} \text{ prior}$$

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

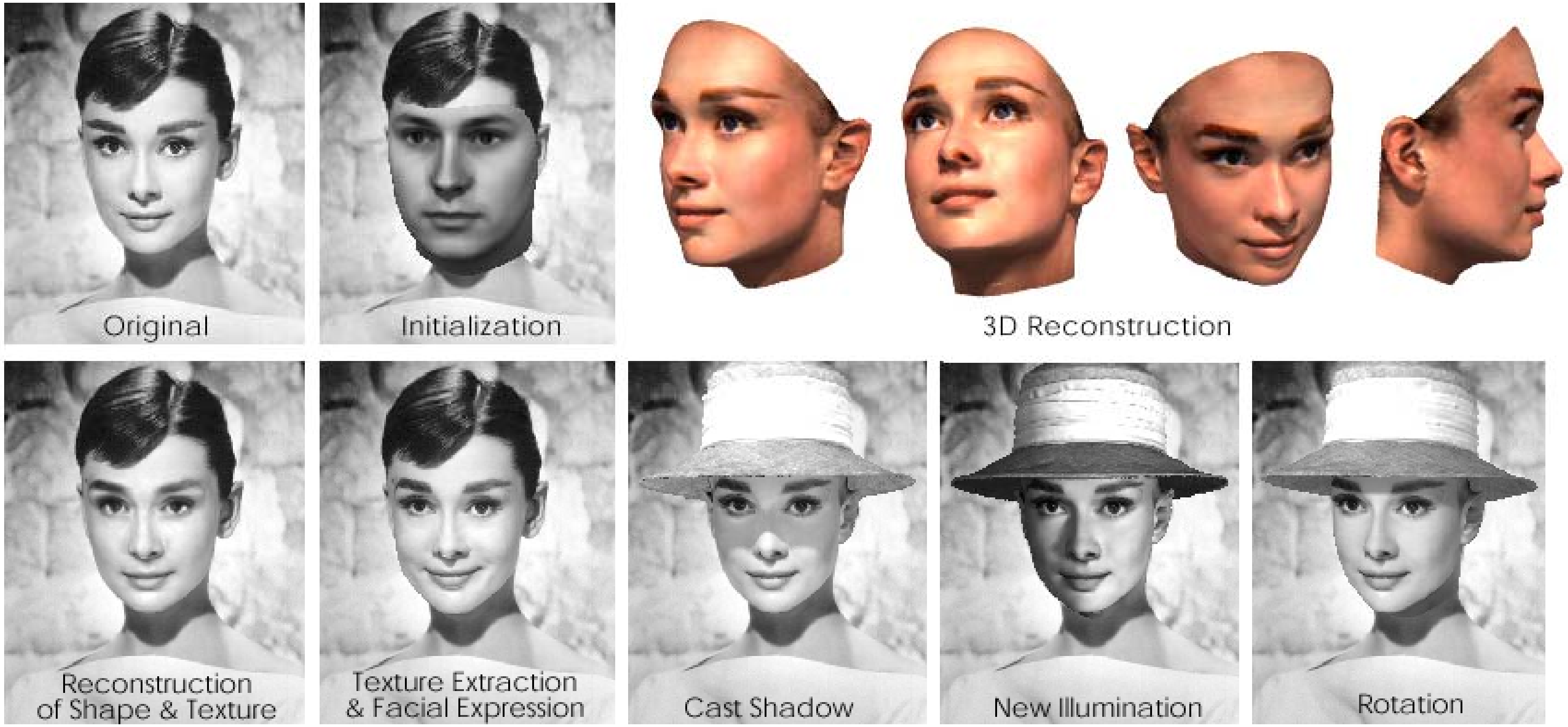
shape and texture priors are learnt from database

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

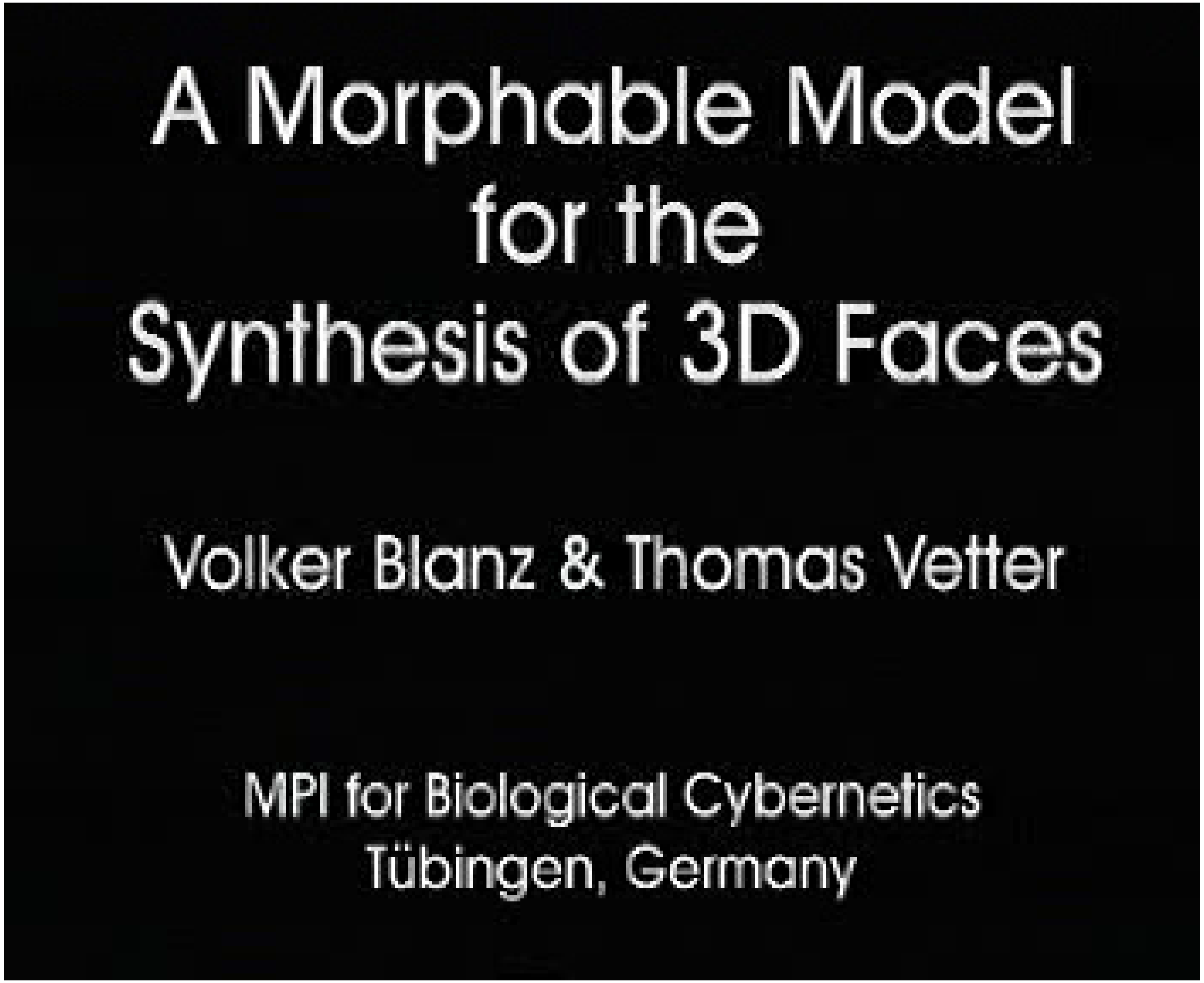
Modifying a single image



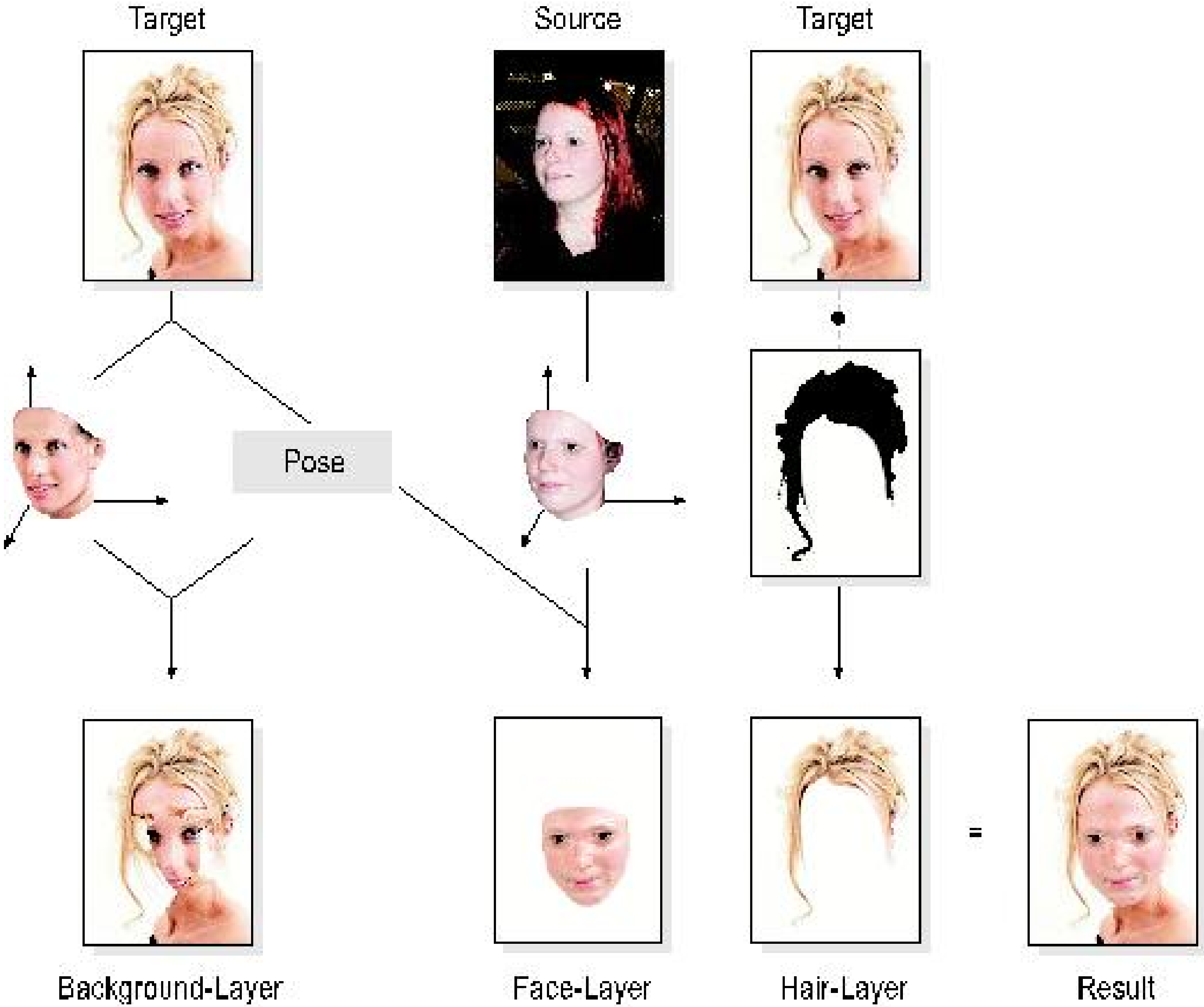
Animating from a single image



Video


A Morphable Model
for the
Synthesis of 3D Faces
Volker Blanz & Thomas Vetter
MPI for Biological Cybernetics
Tübingen, Germany

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

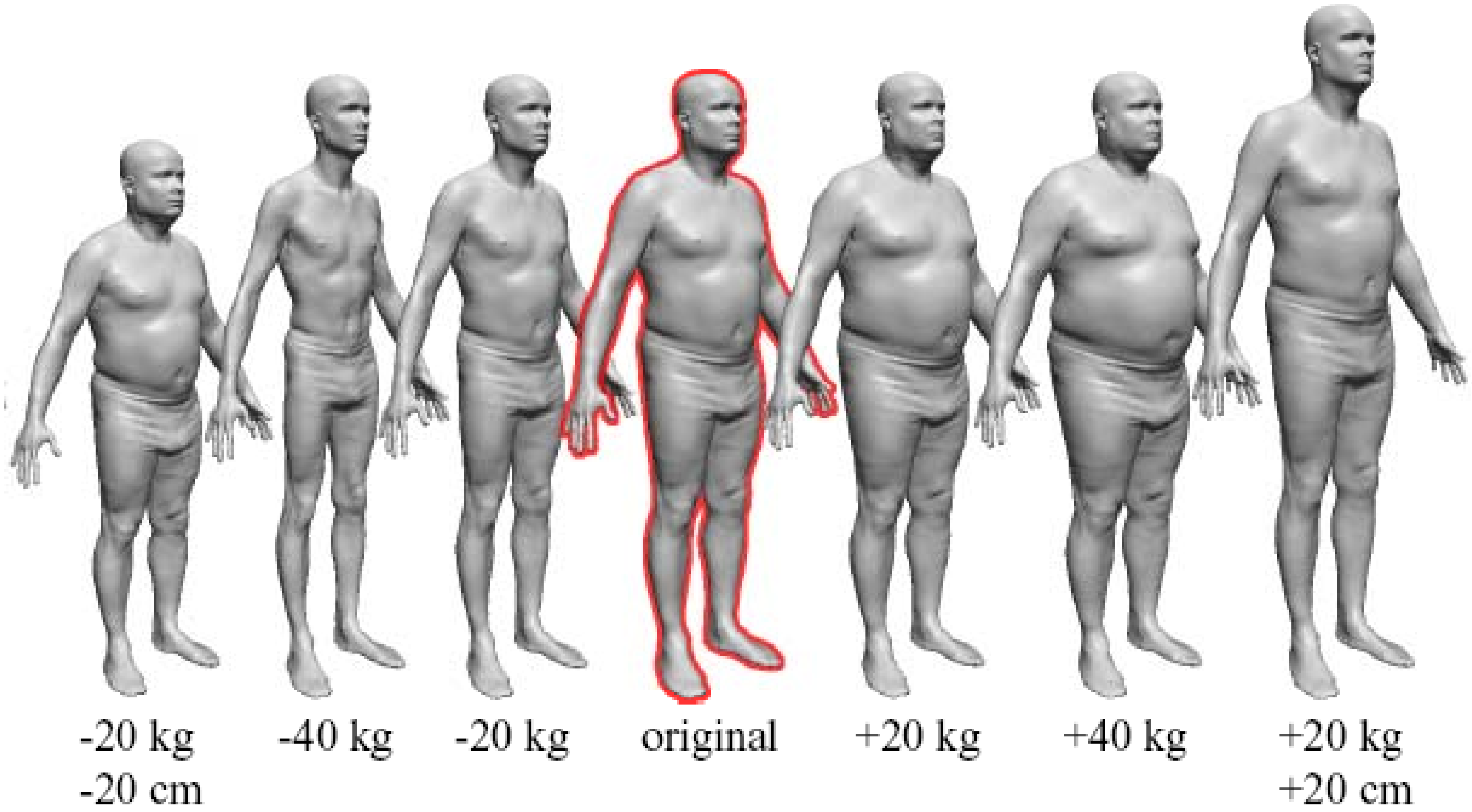
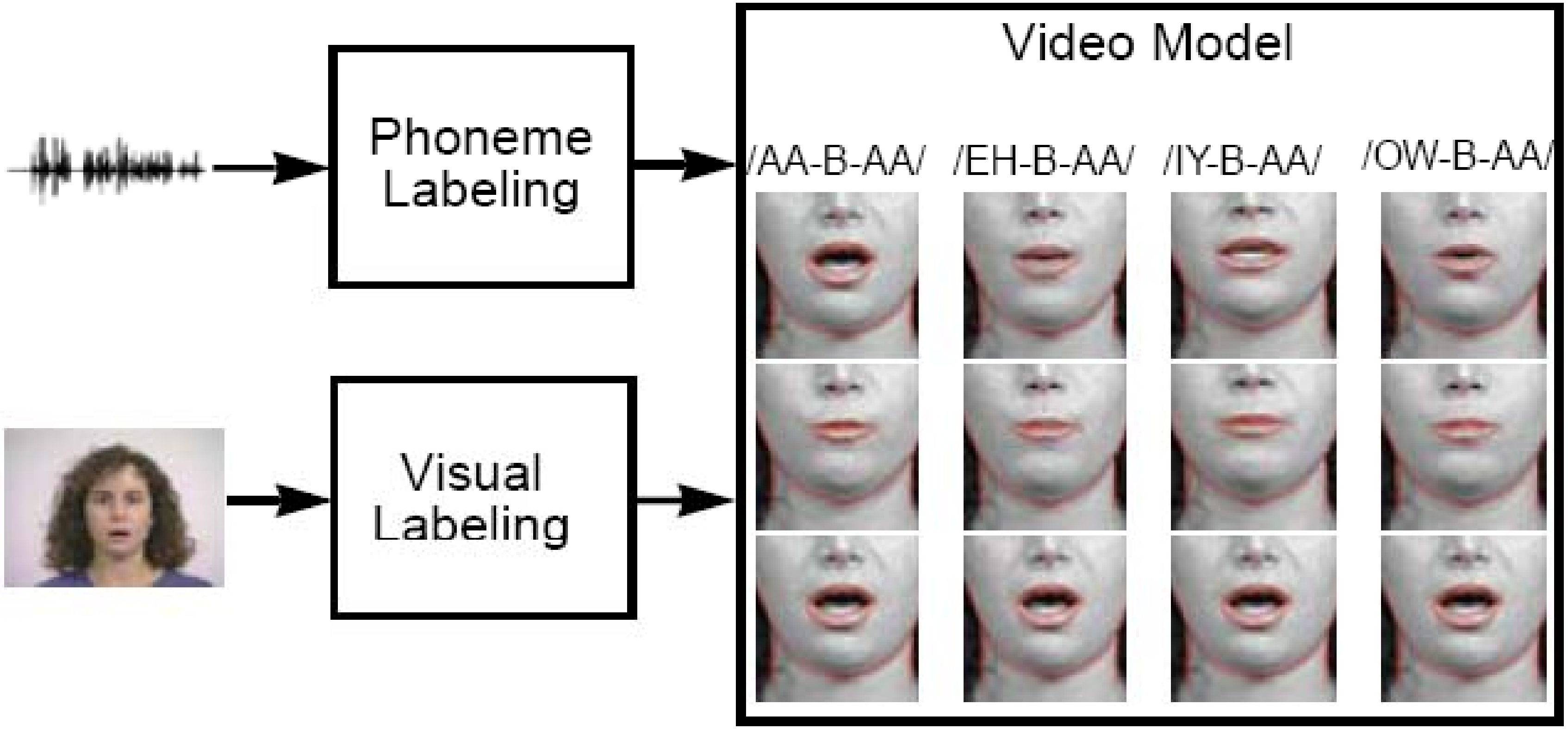
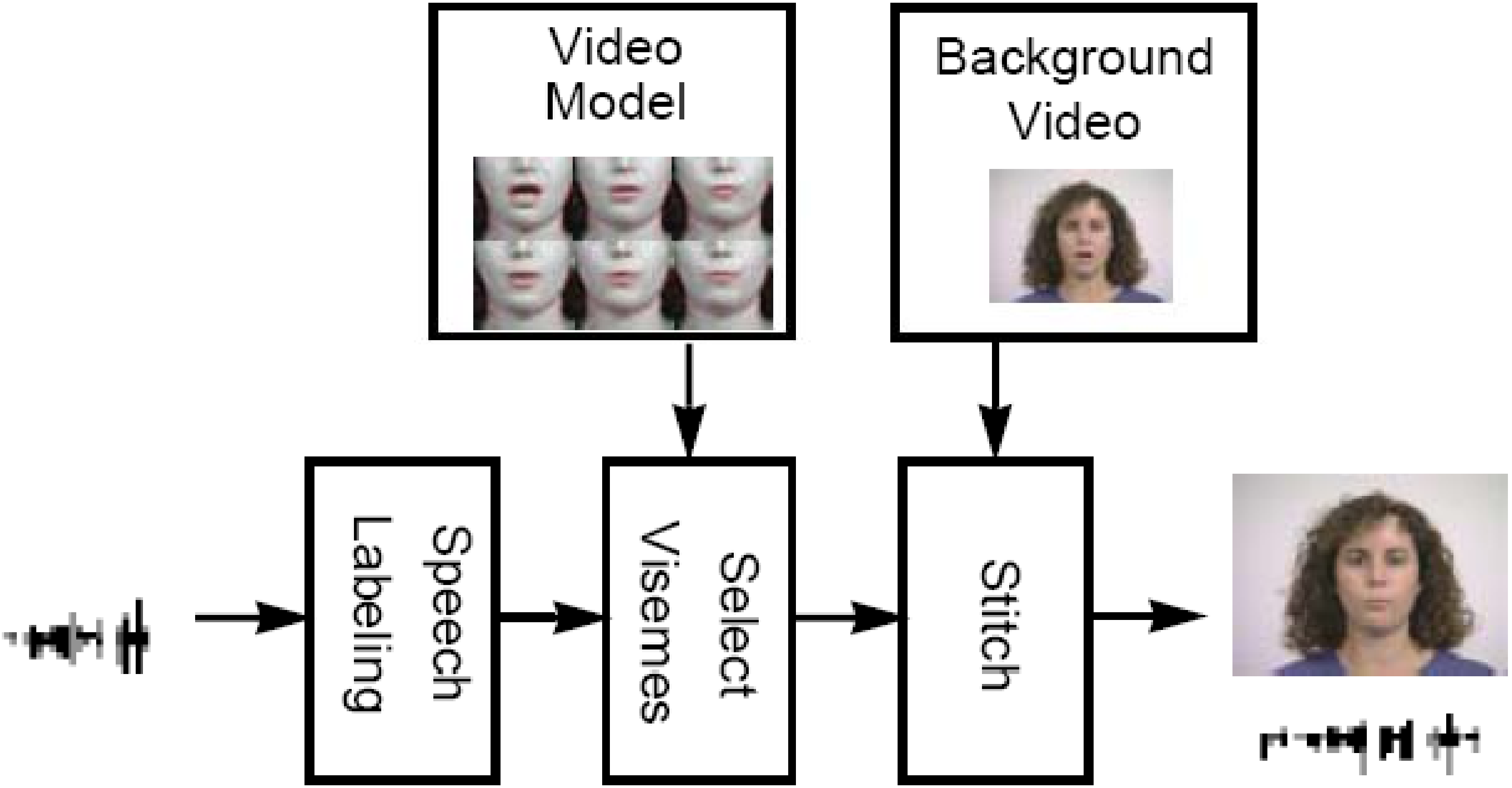


Image-based faces (lip sync.)

Video rewrite (analysis)



Video rewrite (synthesis)



Results

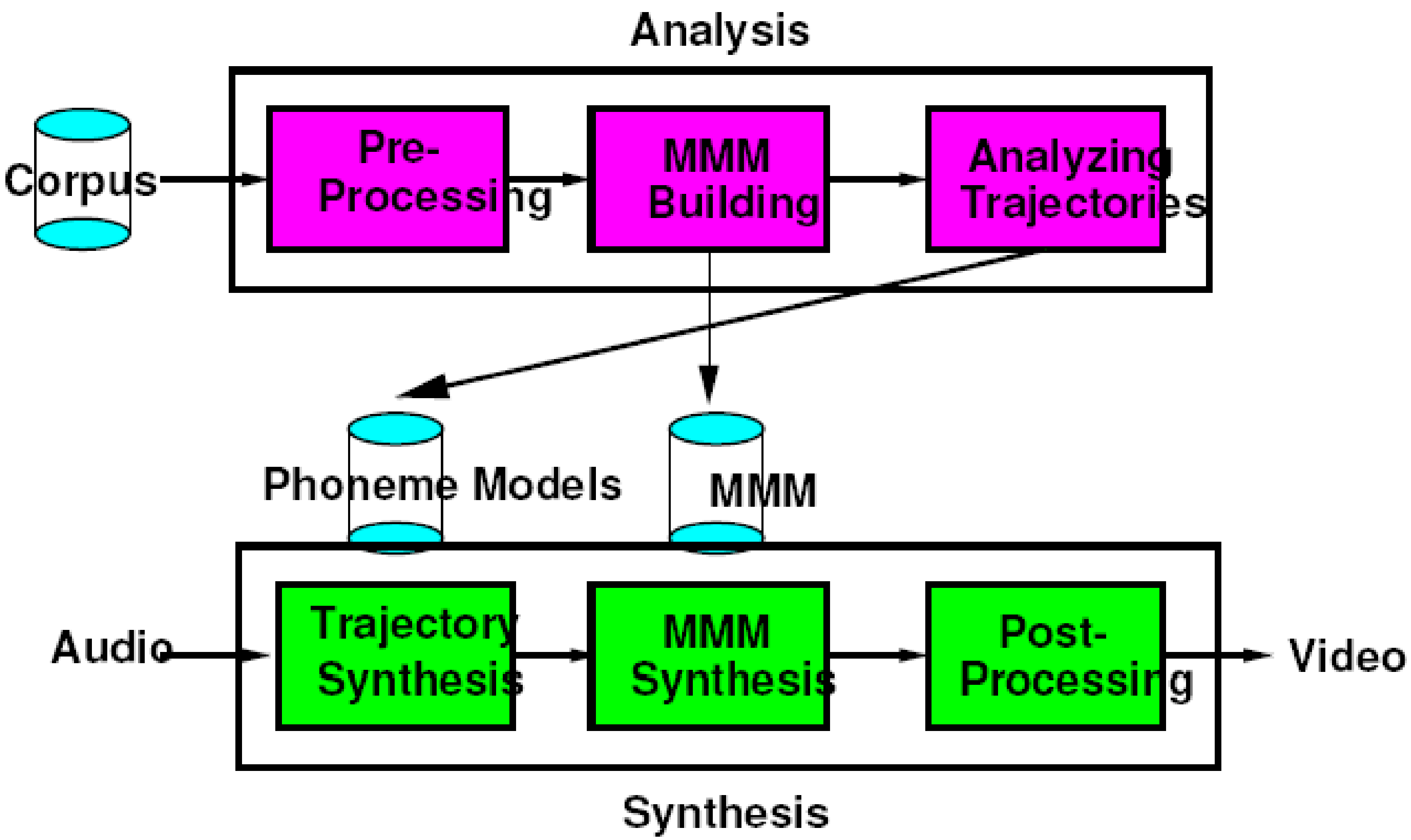
- Video database
  - 2 minutes of JFK
    - Only half usable
    - Head rotation

training video

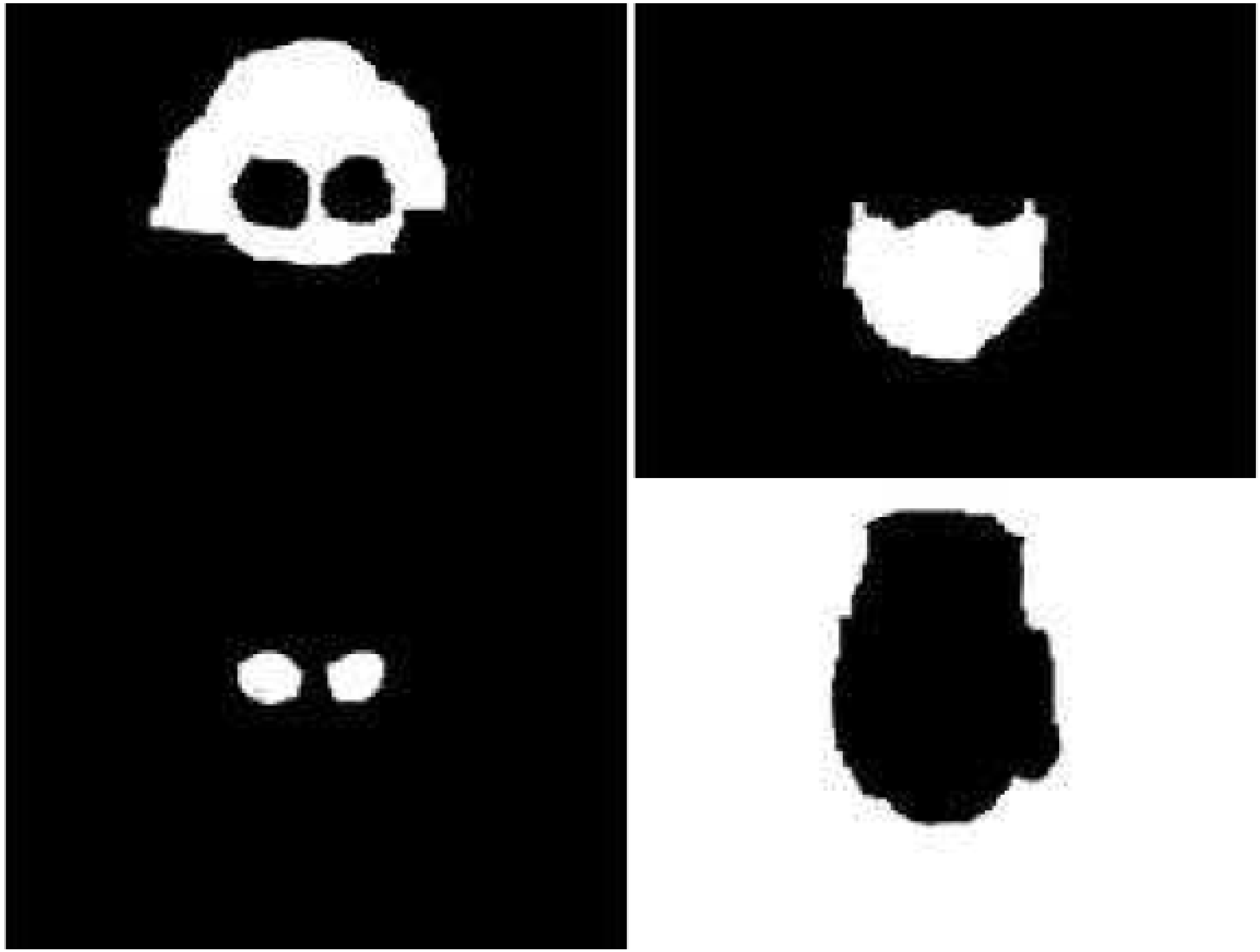
Read my lips.

I never met Forest Gump.

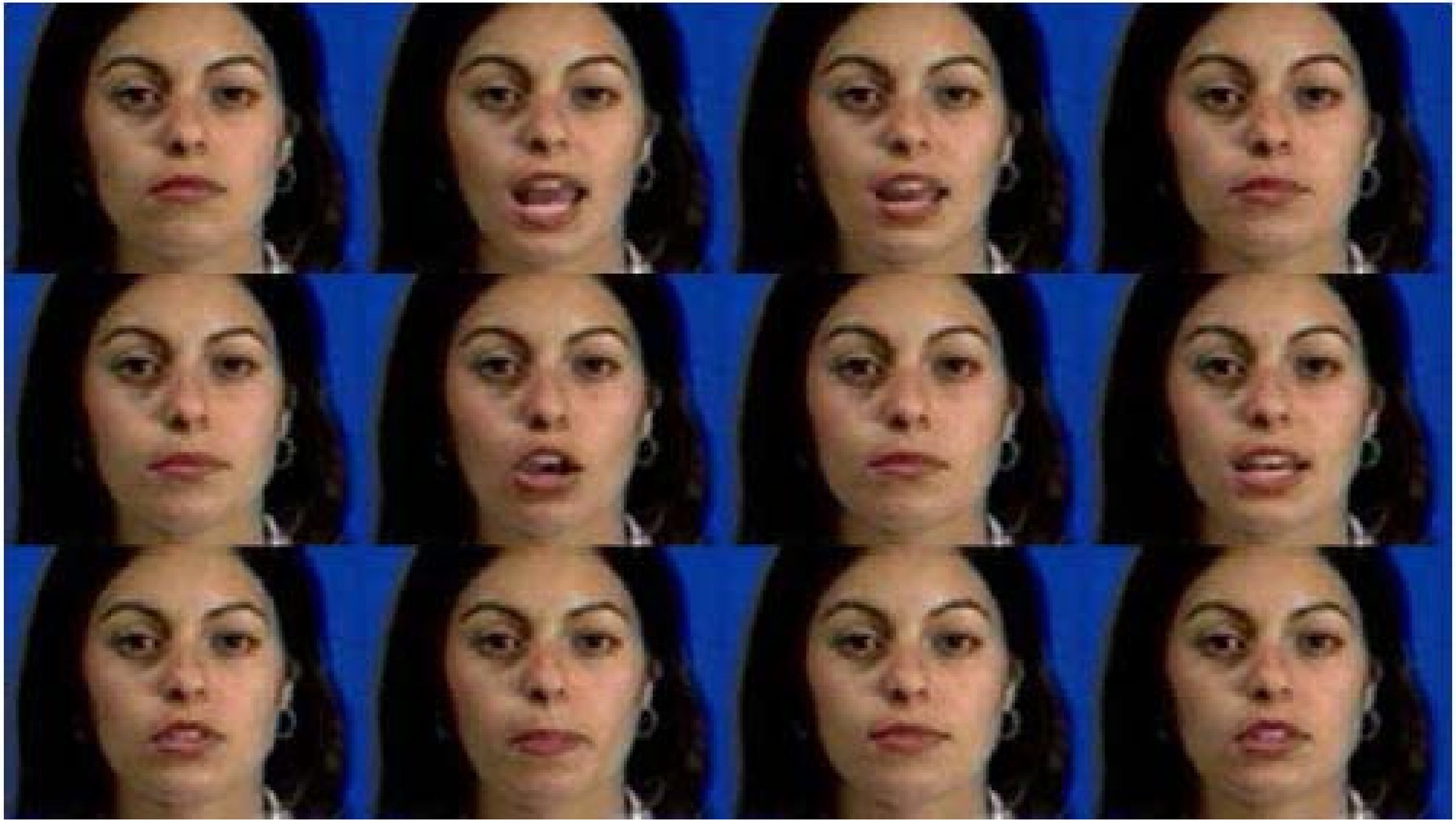
Morphable speech model



Preprocessing



Prototypes (PCA+k-mean clustering)



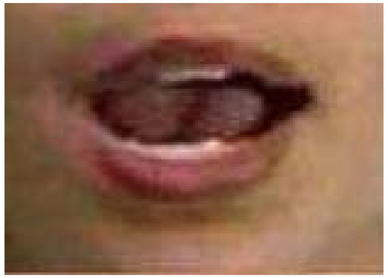
We find I_i and C_i for each prototype image.

Morphable model

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

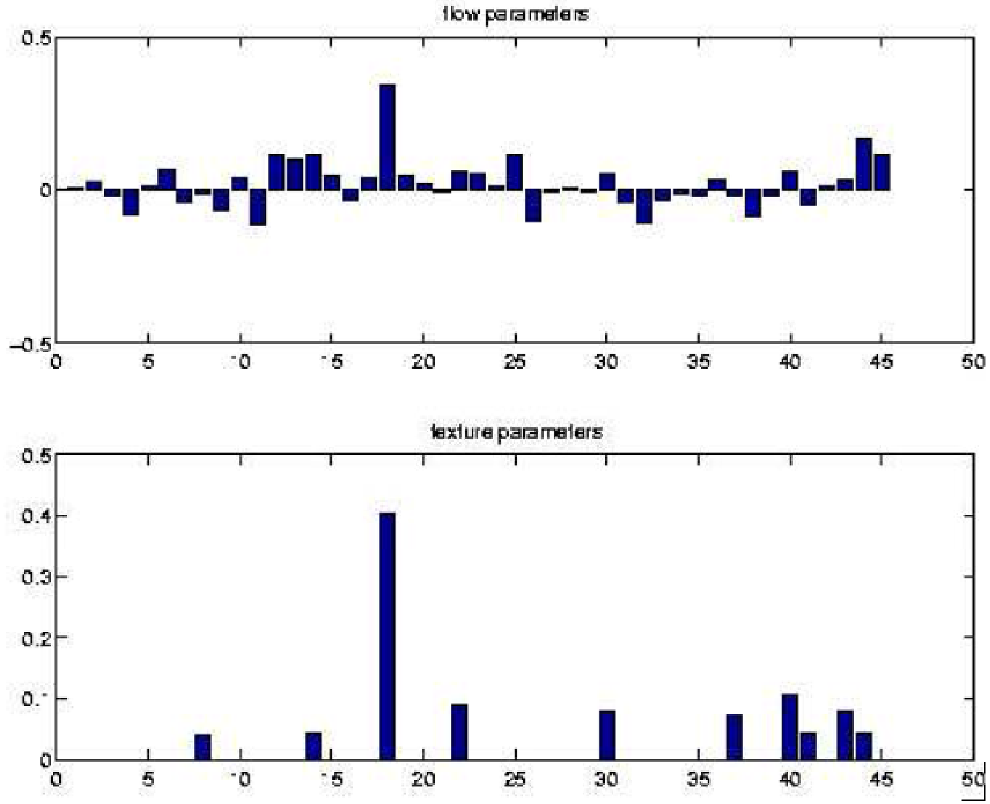
$$\mathbf{I} \underset{\text{synthesis}}{\overset{\text{analysis}}{\rightleftarrows}} \boldsymbol{\alpha\beta}$$

Morphable model



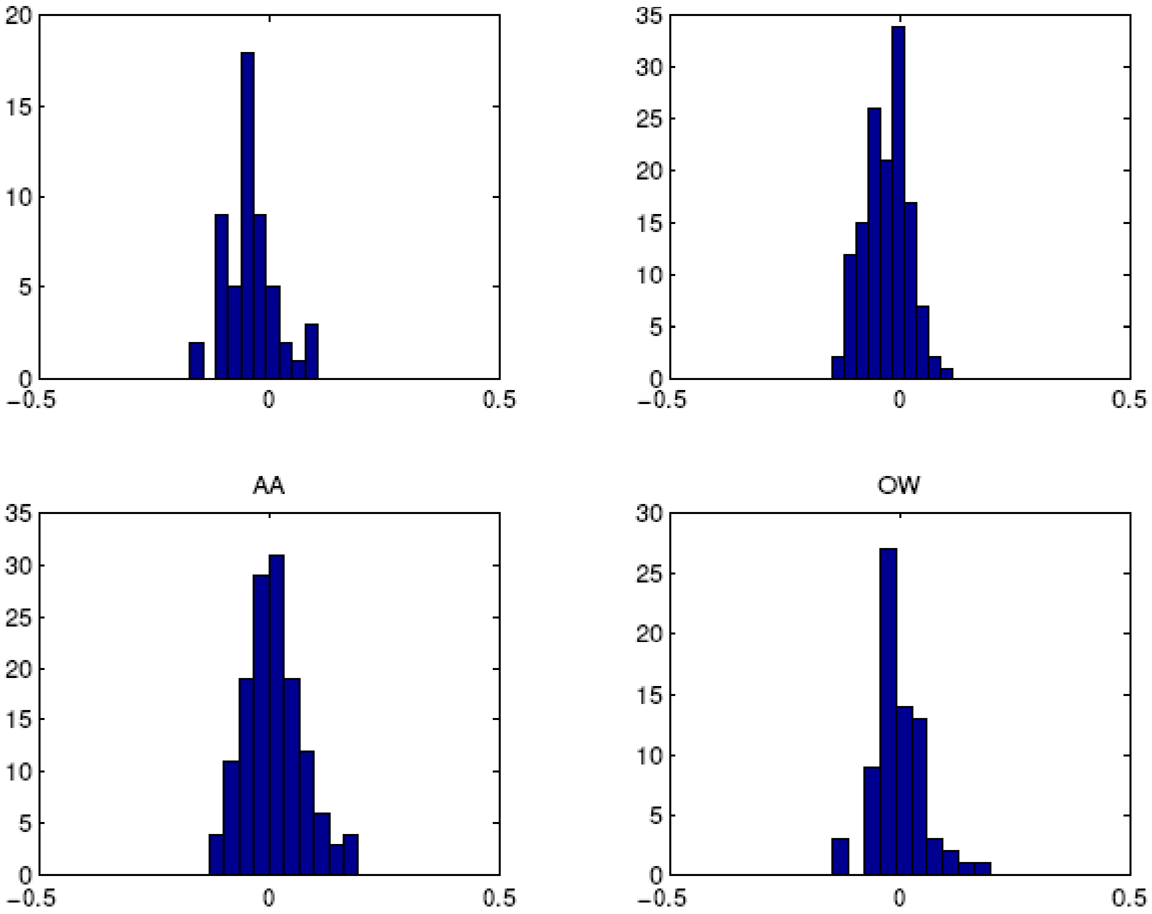
analysis

synthesis



Synthesis

$$E = \underbrace{(y-\mu)^T D^T \Sigma^{-1} D(y-\mu)}_{target\ term} + \lambda \underbrace{y^T W^T W y}_{smoothness}$$

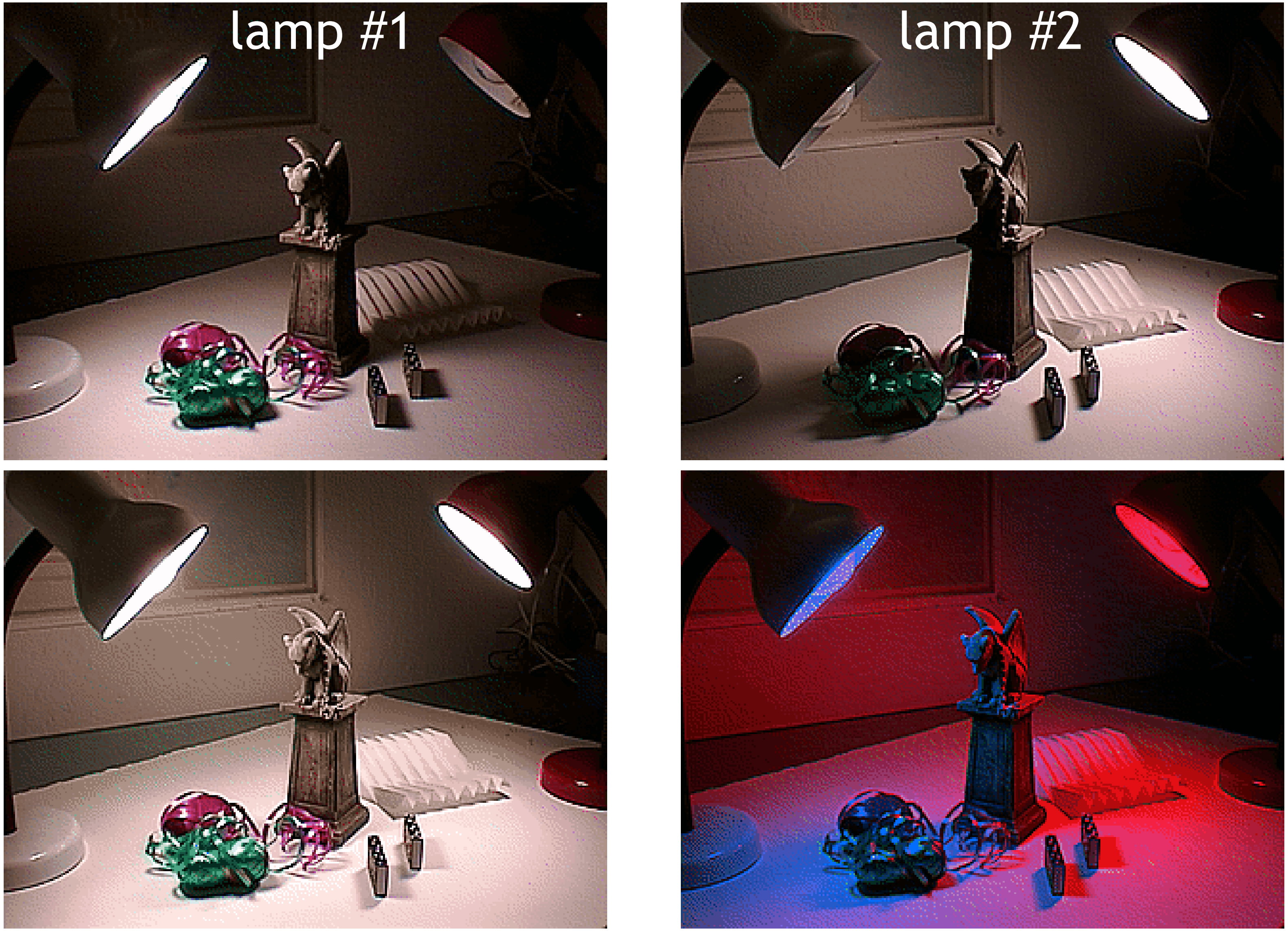


Results

Results

Relighting faces

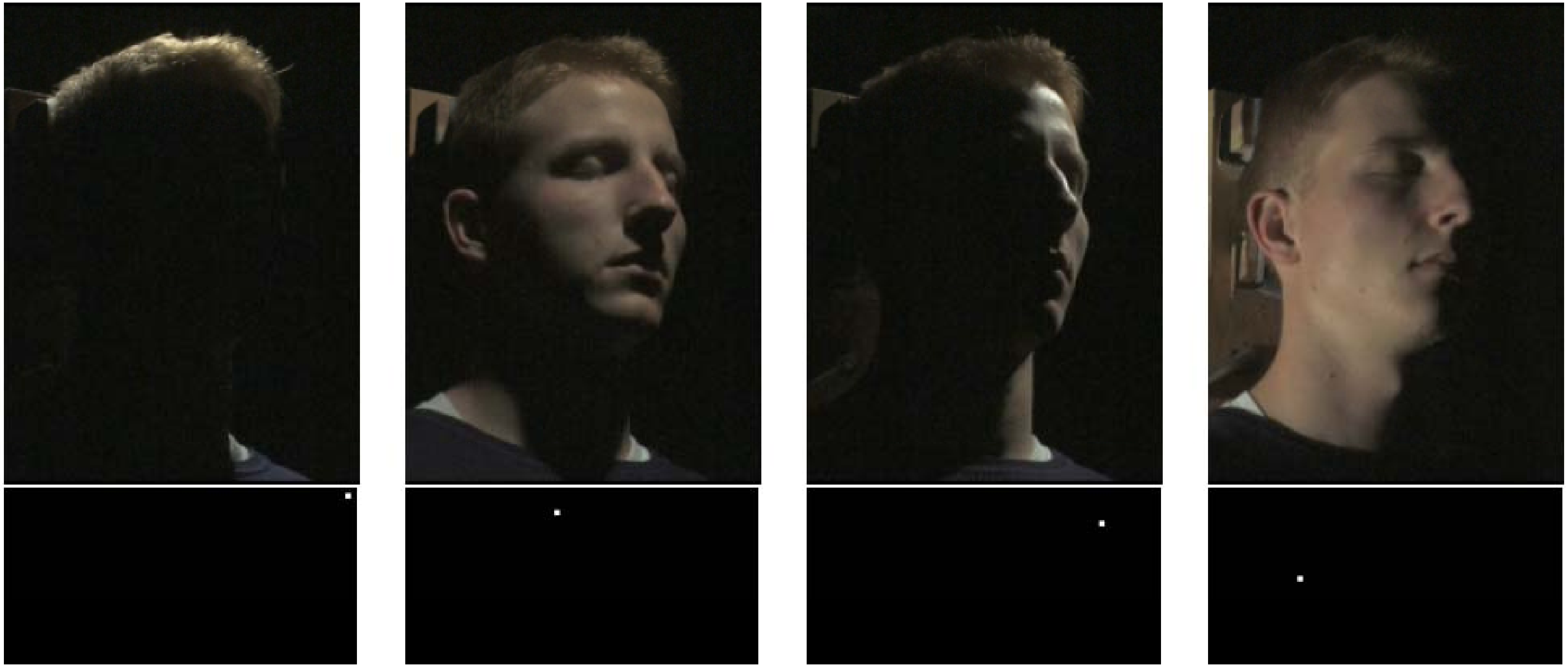
Light is additive



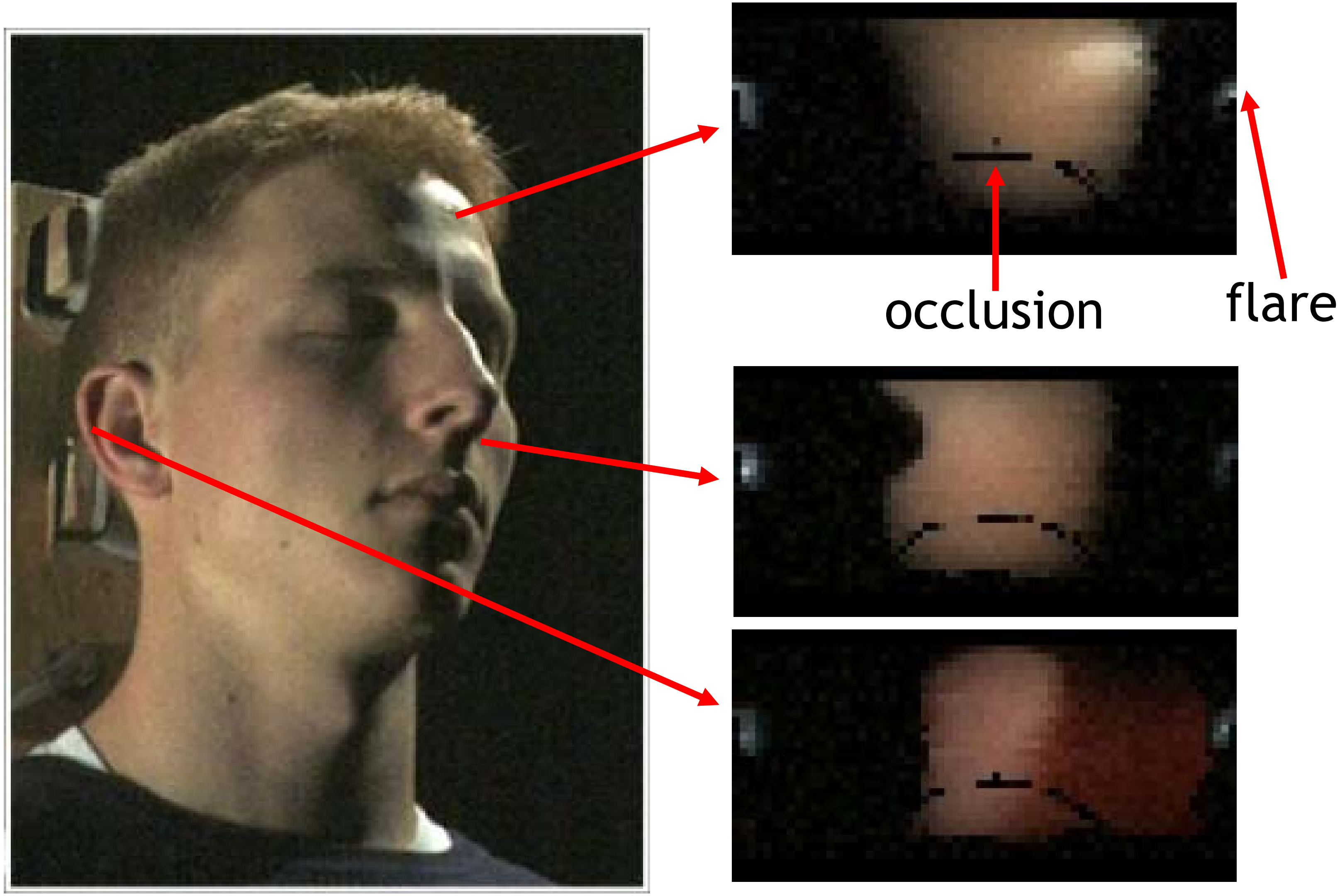
Light stage 1.0

Light stage 1.0

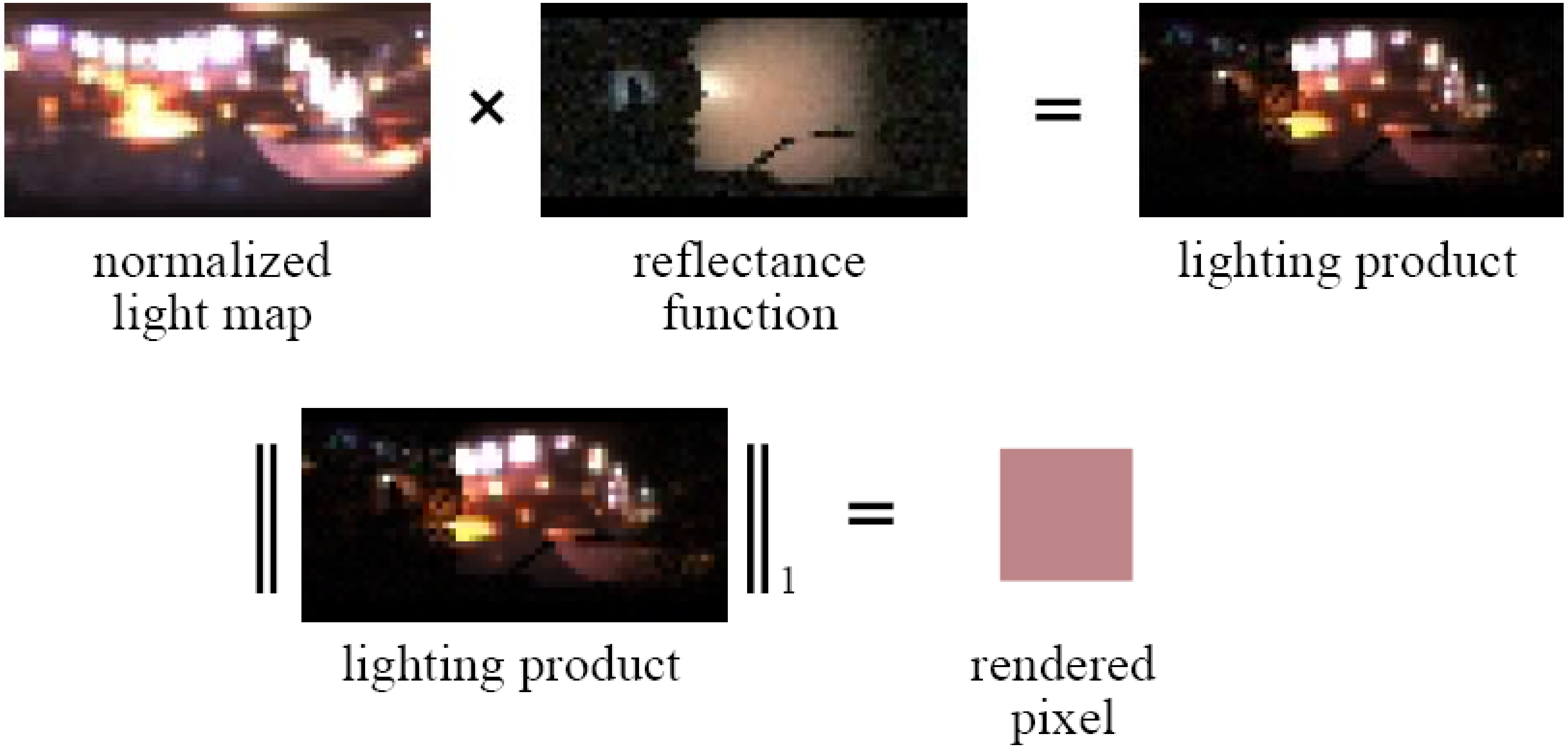
Input images



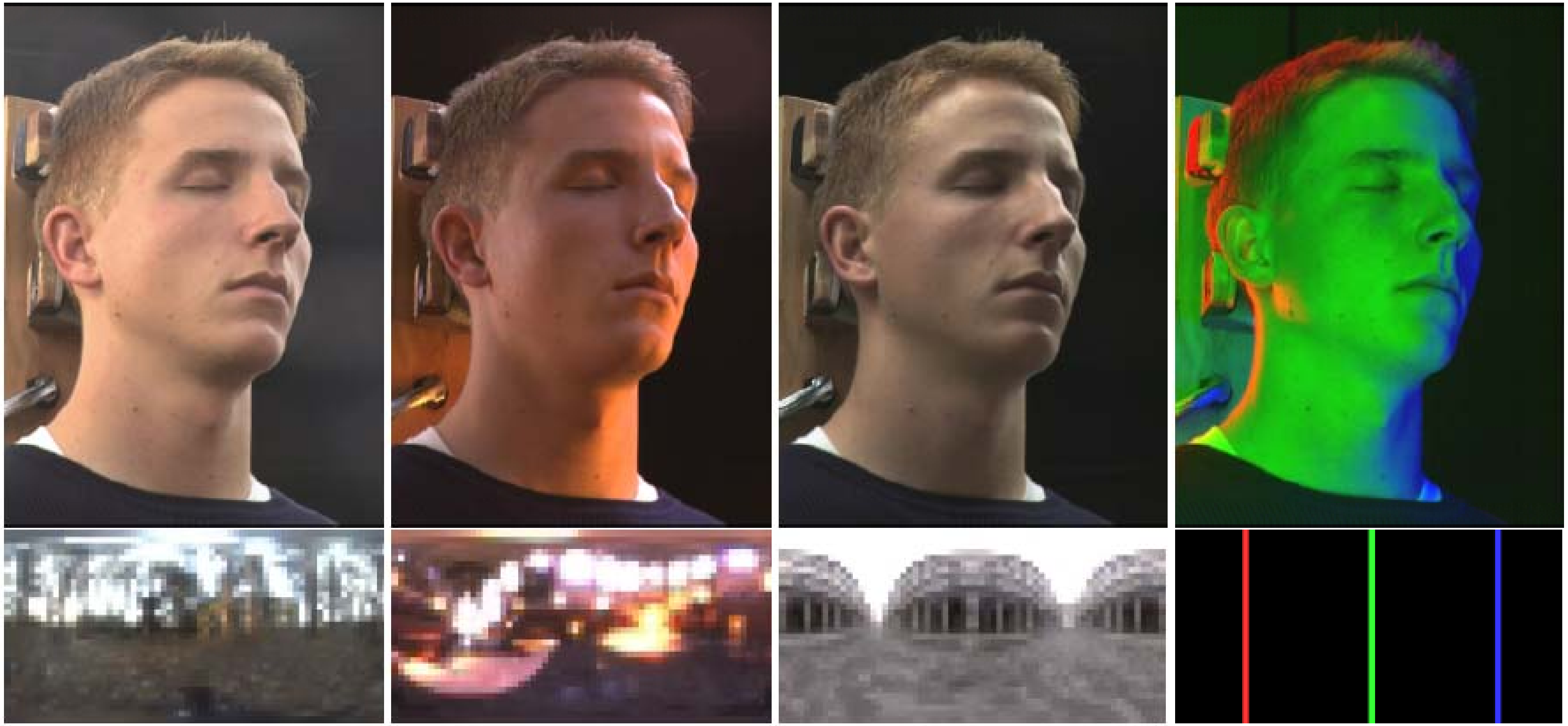
Reflectance function



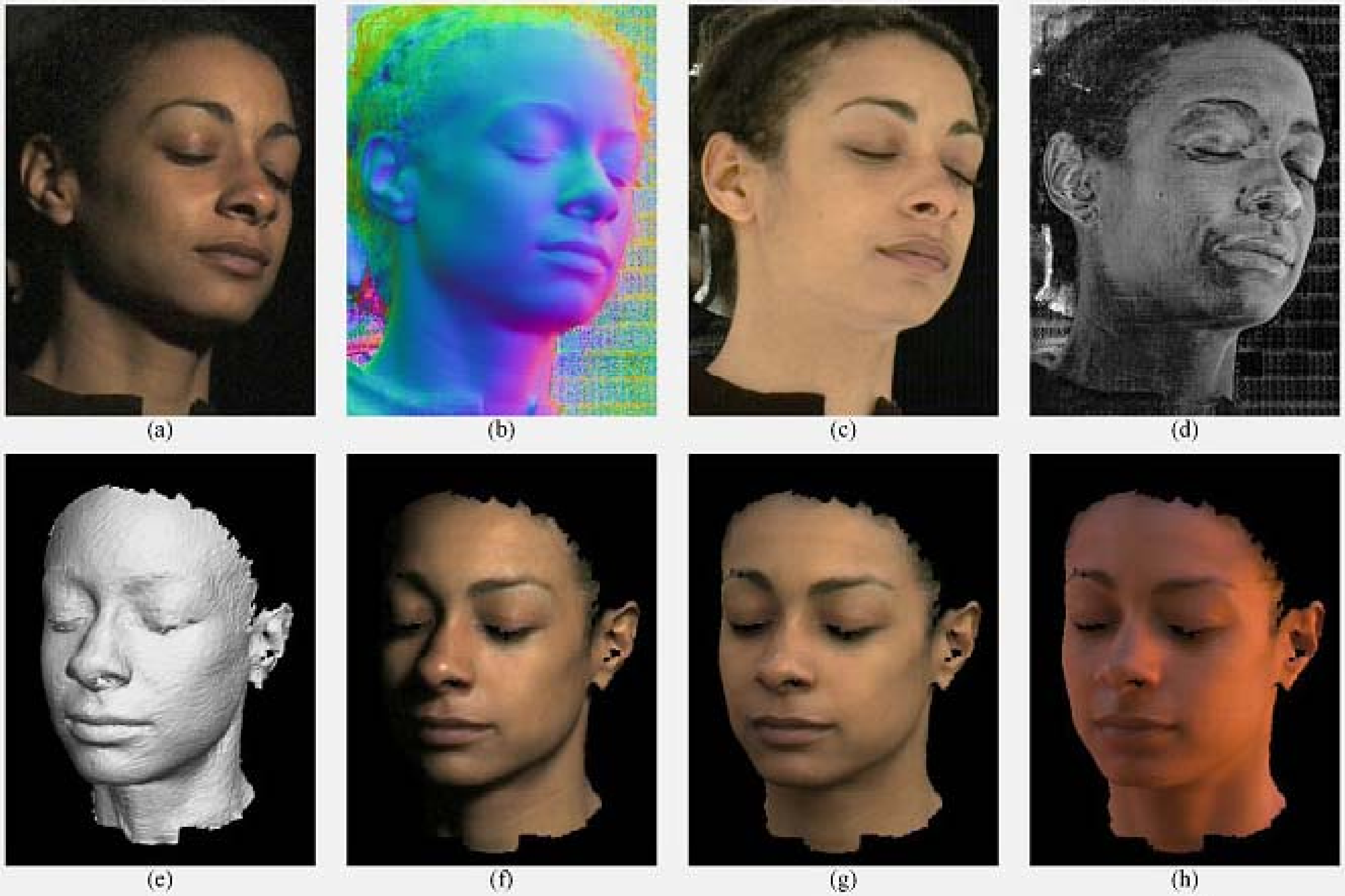
Relighting


×
=
normalized light map
reflectance function
lighting product
=
1
lighting product
rendered pixel

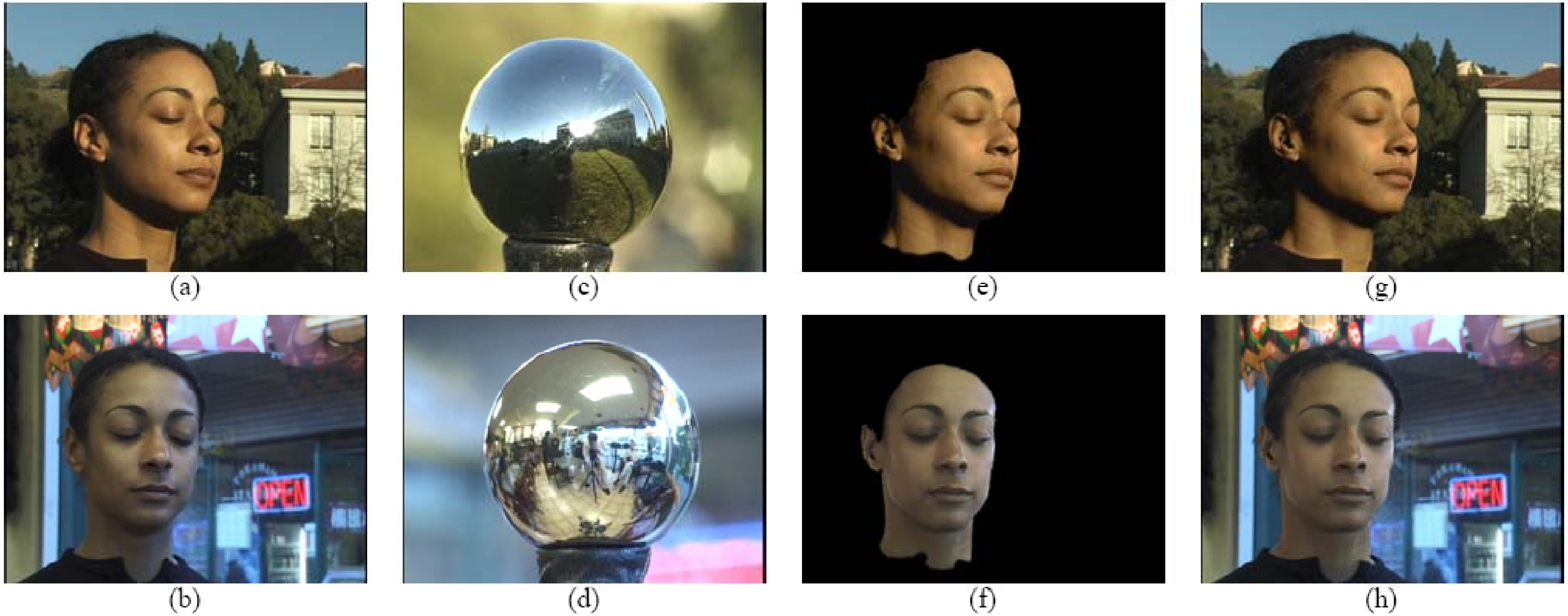
Results



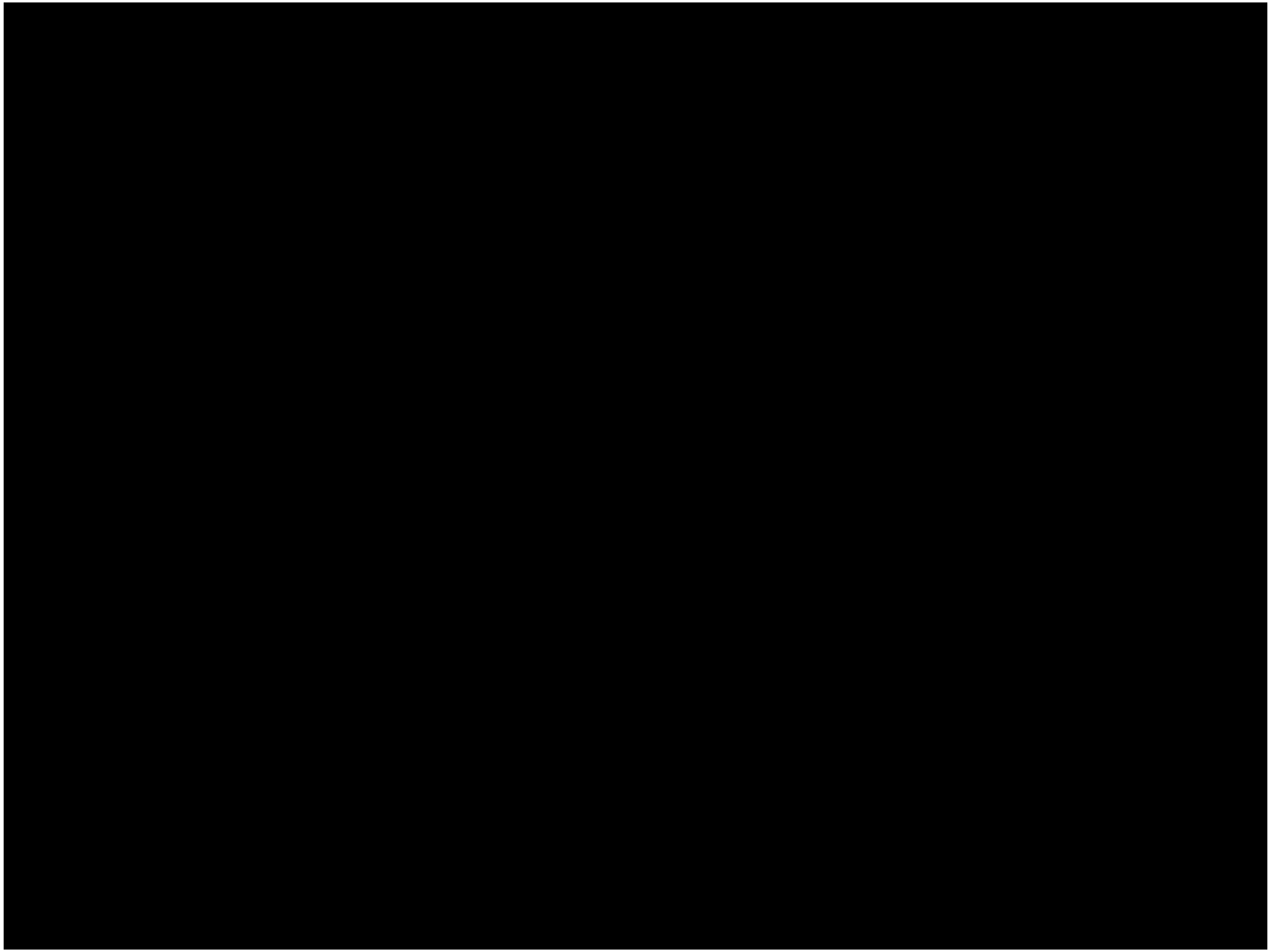
Changing viewpoints


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

Results



Video



3D face applications: Spiderman 2

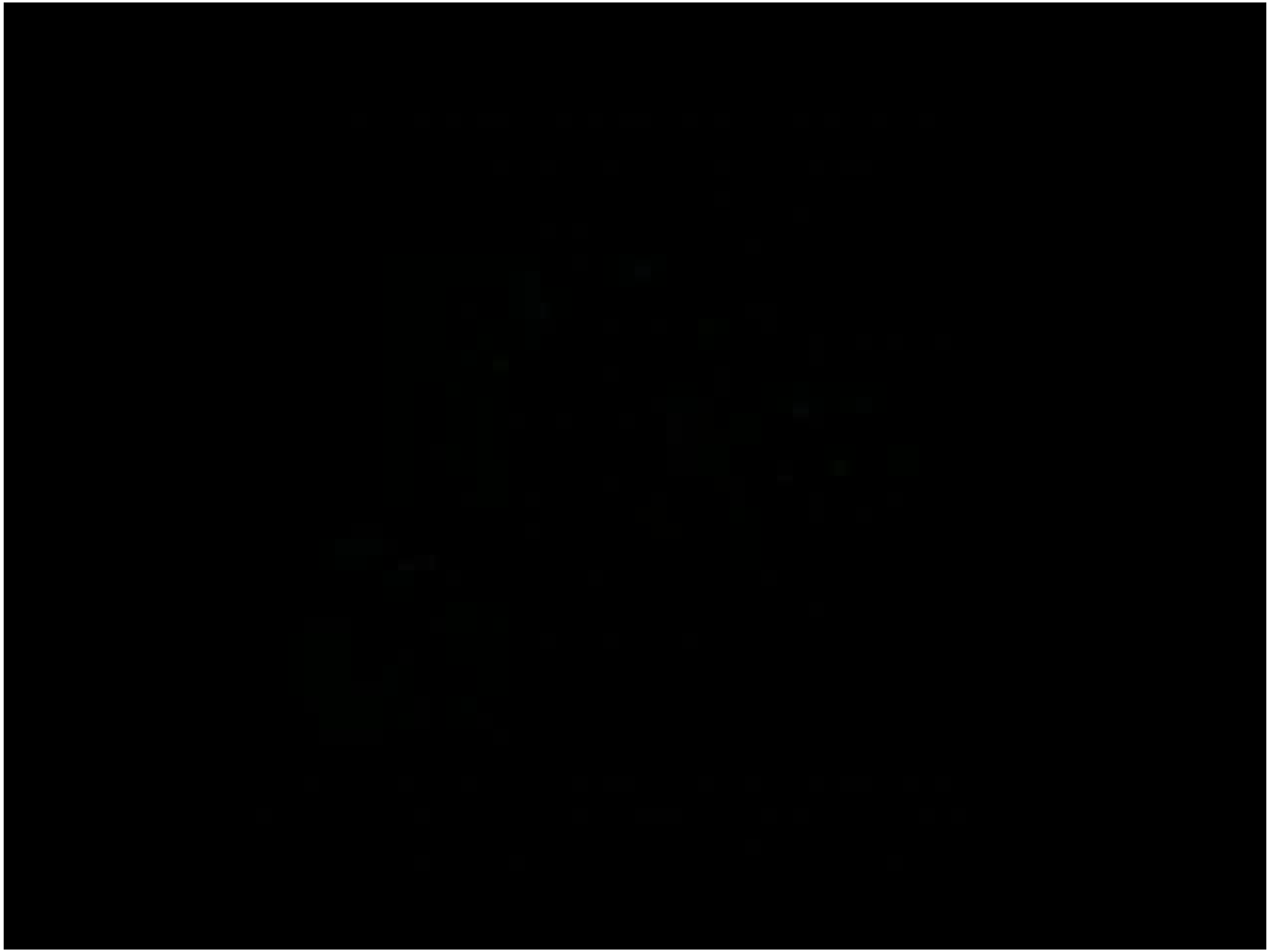
Spiderman 2

real synthetic

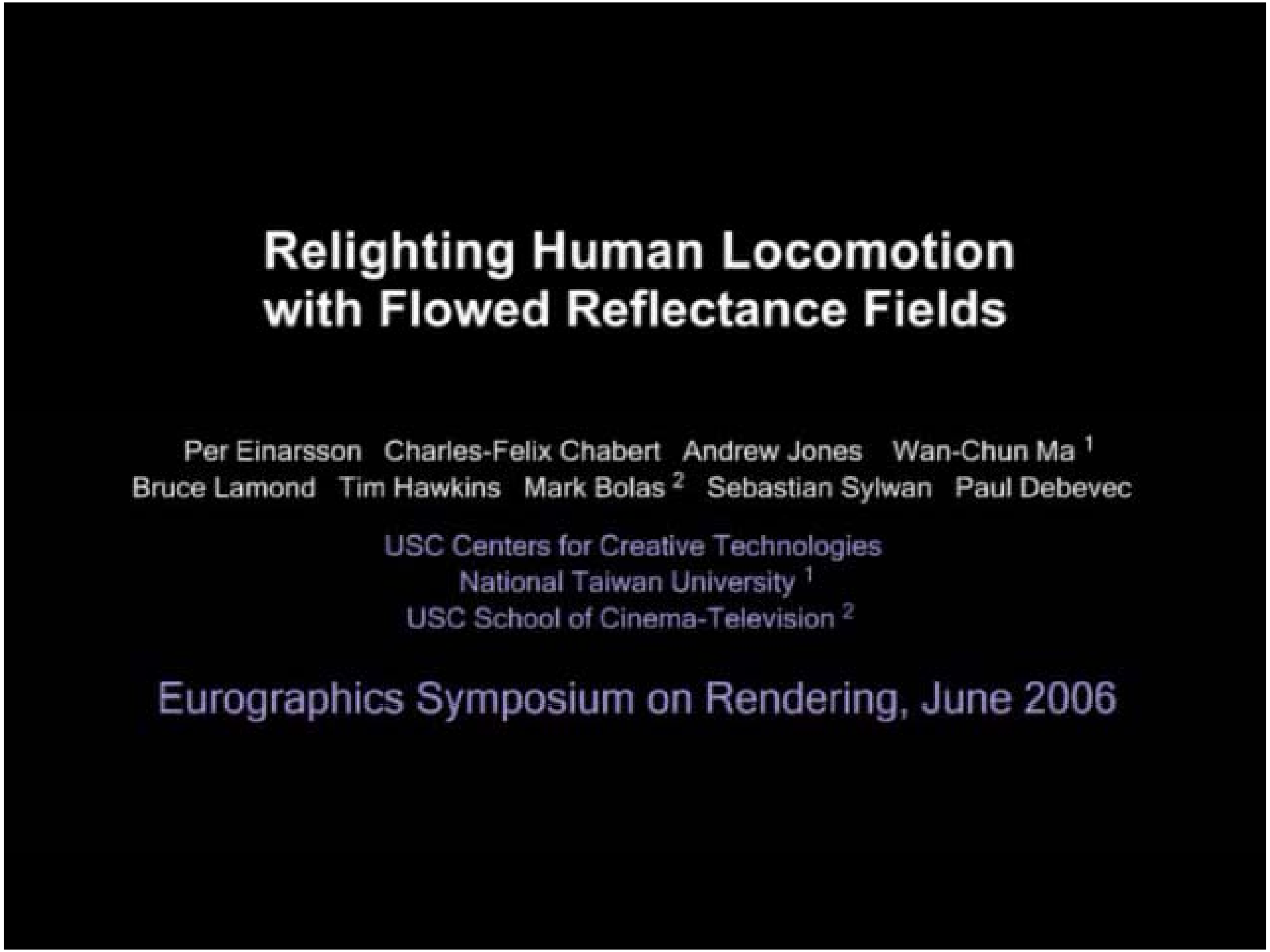
Spiderman 2

video

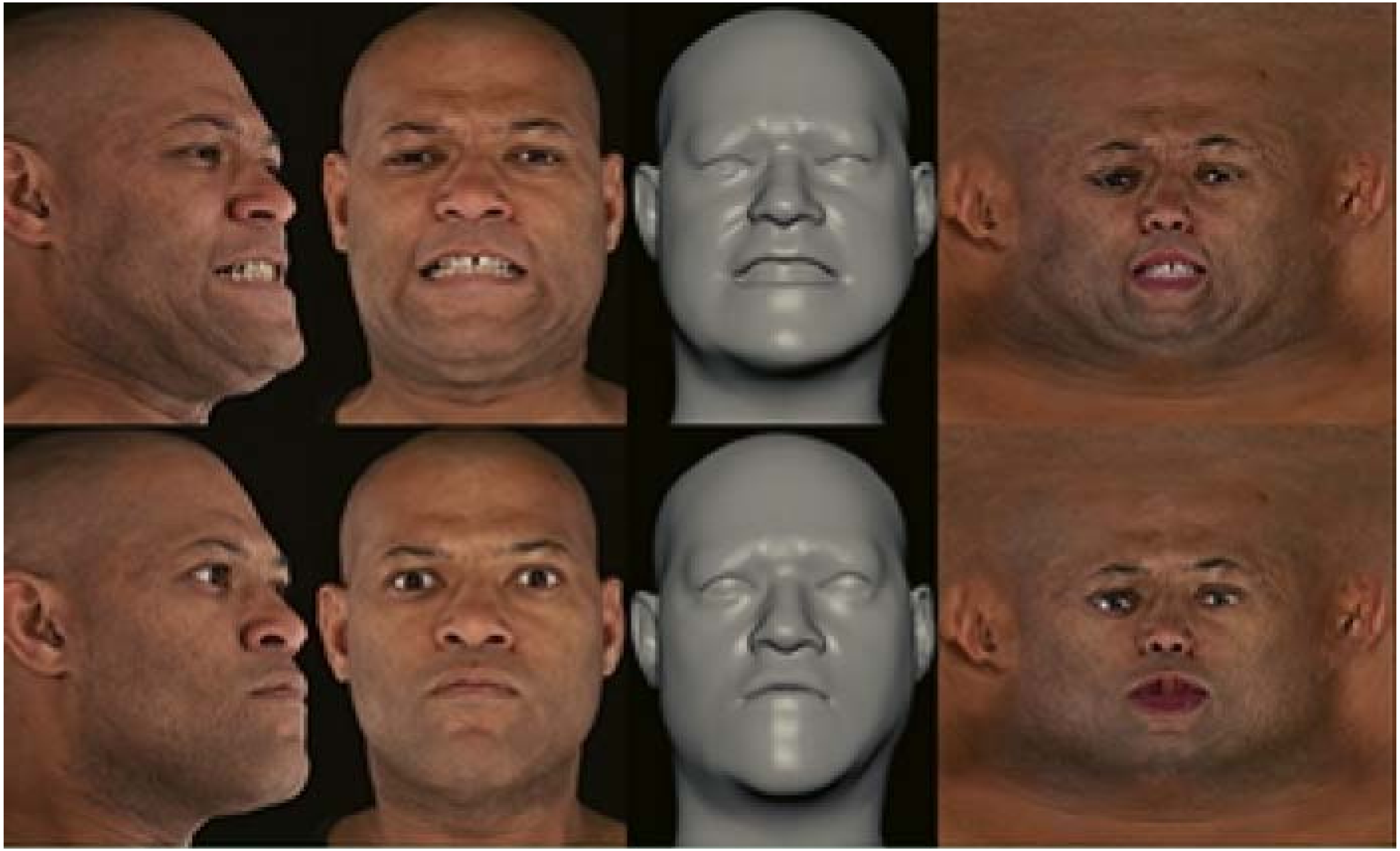
Light stage 3



Light stage 6



Application: The Matrix Reloaded



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.