Spring 2015

CSCI 599: **Digital Geometry Processing**

15.1 **Facial Performance Capture**

Hao Li

http://cs599.hao-li.com

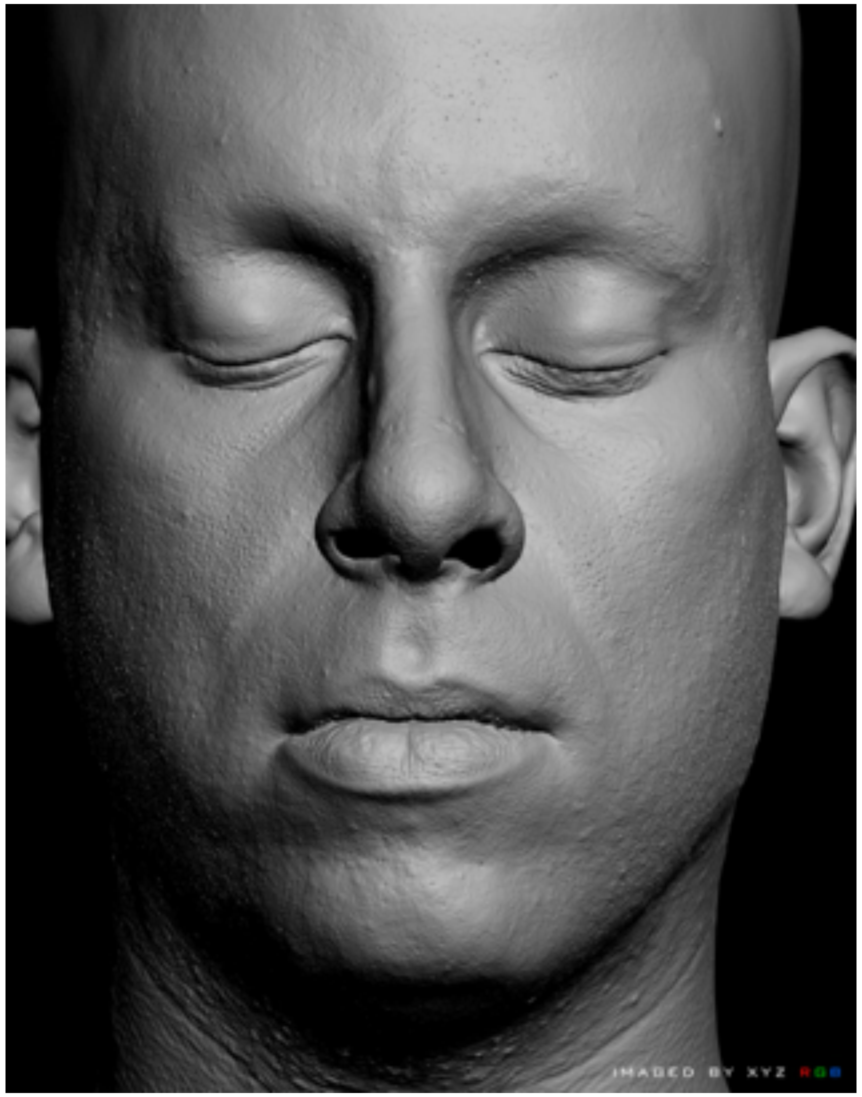
Performance to Facial Animation

Motion Capture

Motion Capture

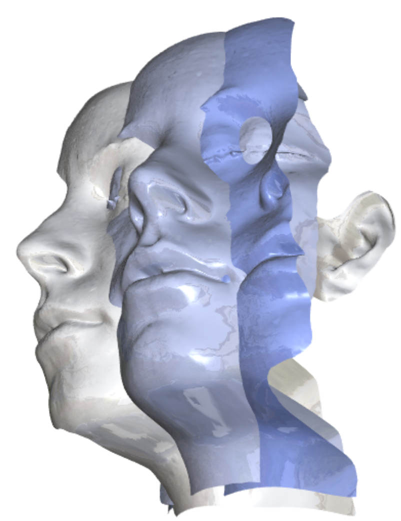
Facial Animation in Films

Facial Modeling and Scanning

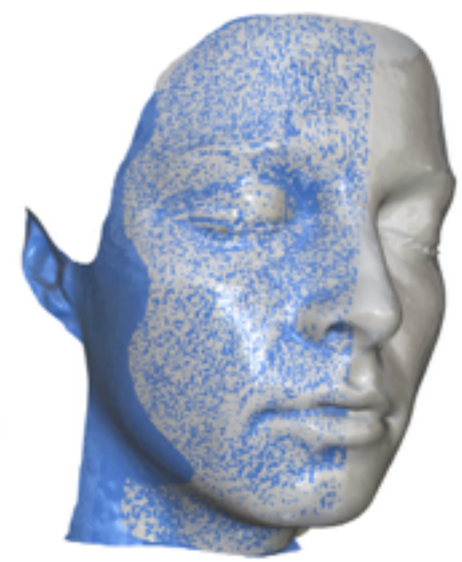


Facial Modeling and Scanning

acquisition



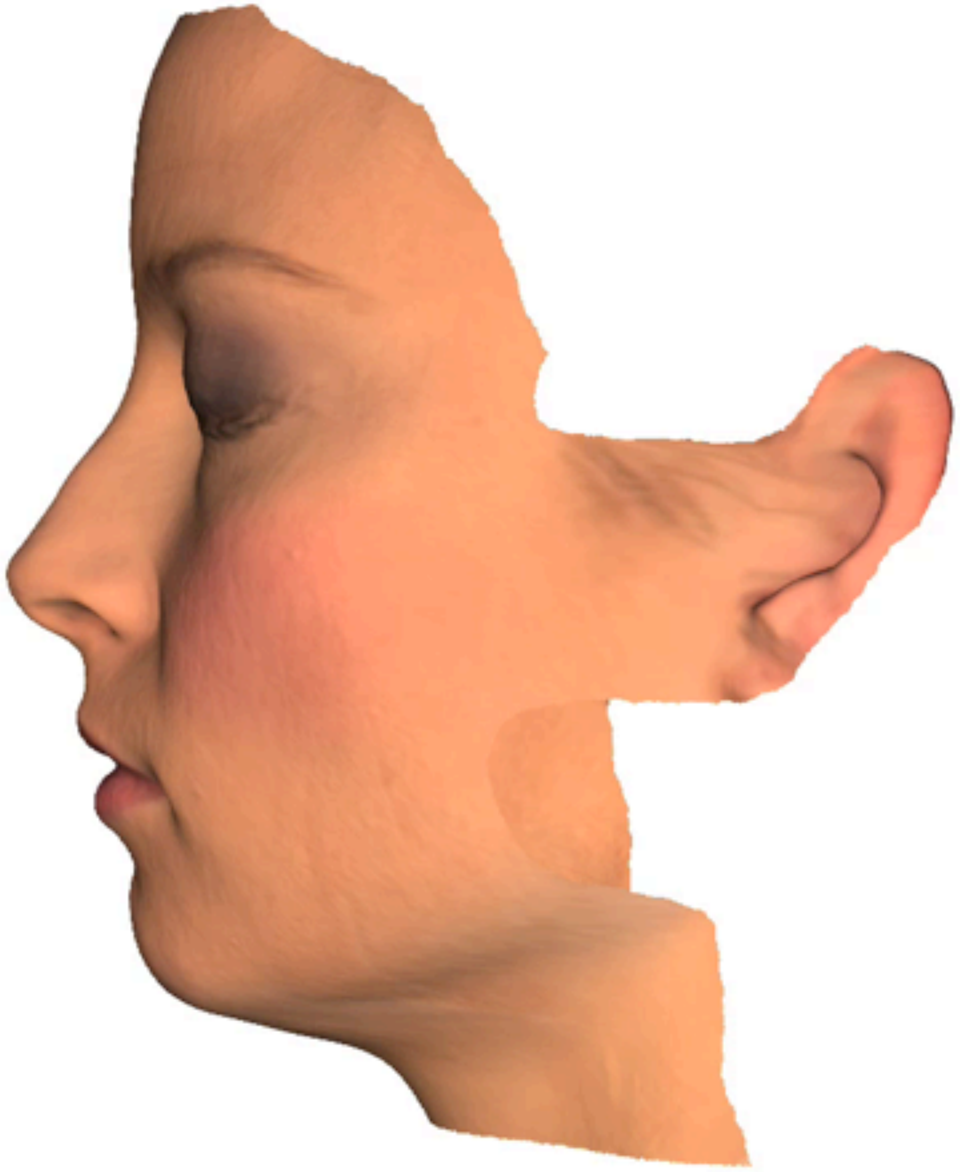
initial
alignment



registration

merging

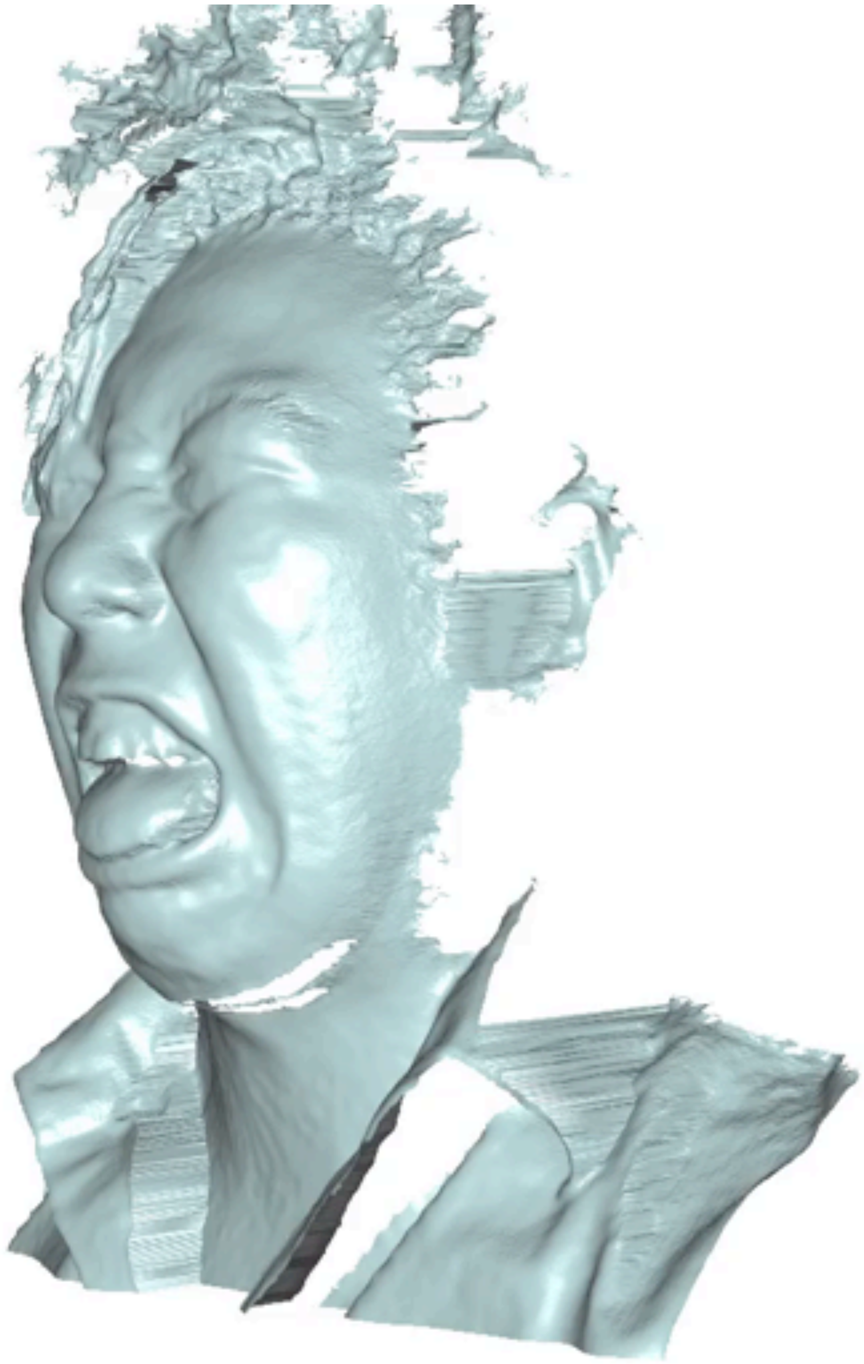
Facial Modeling and Scanning



High-End 3D Scanning

Low-Cost Passive Scanning (AGI Soft)

stereo pair

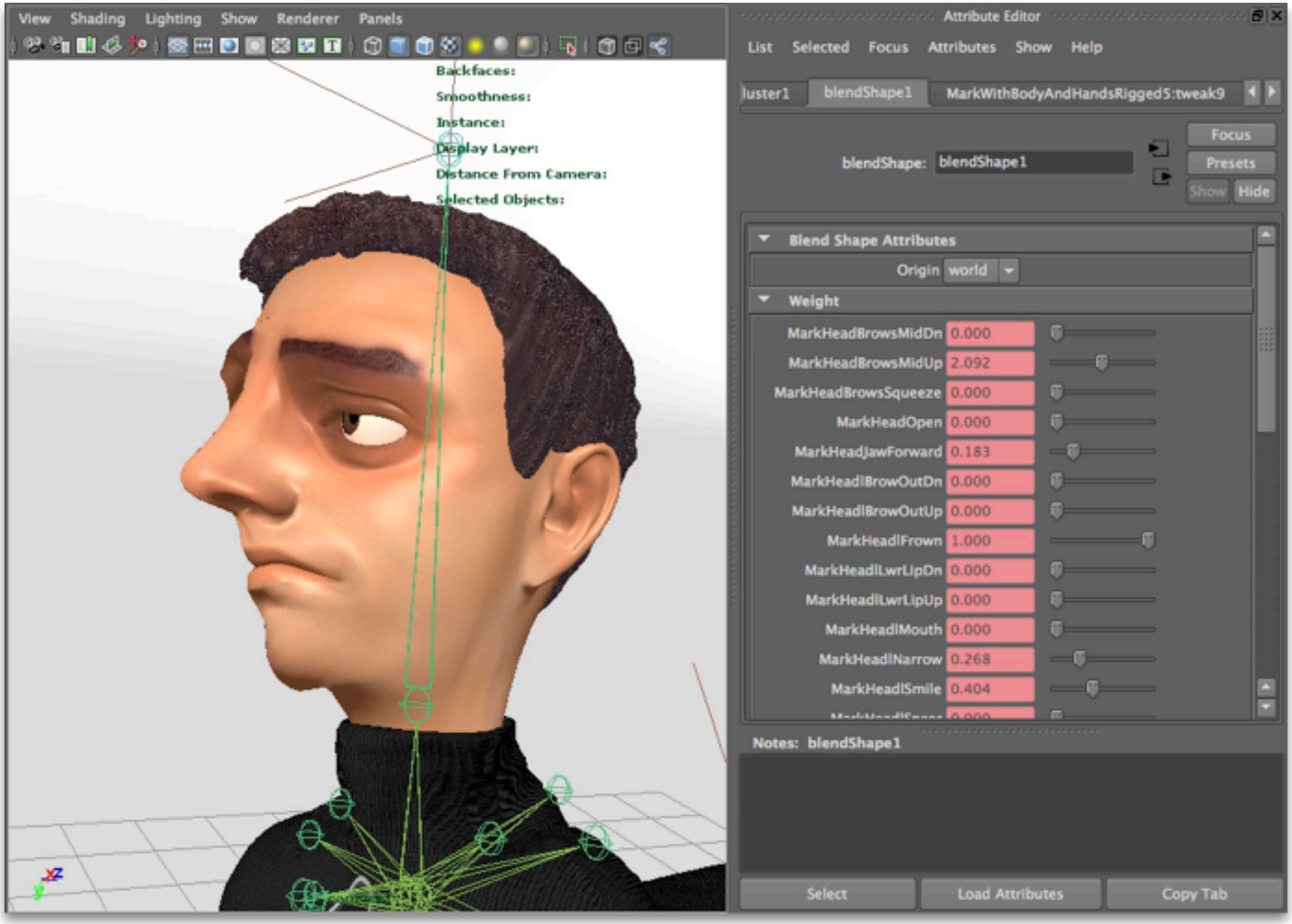


3D scan

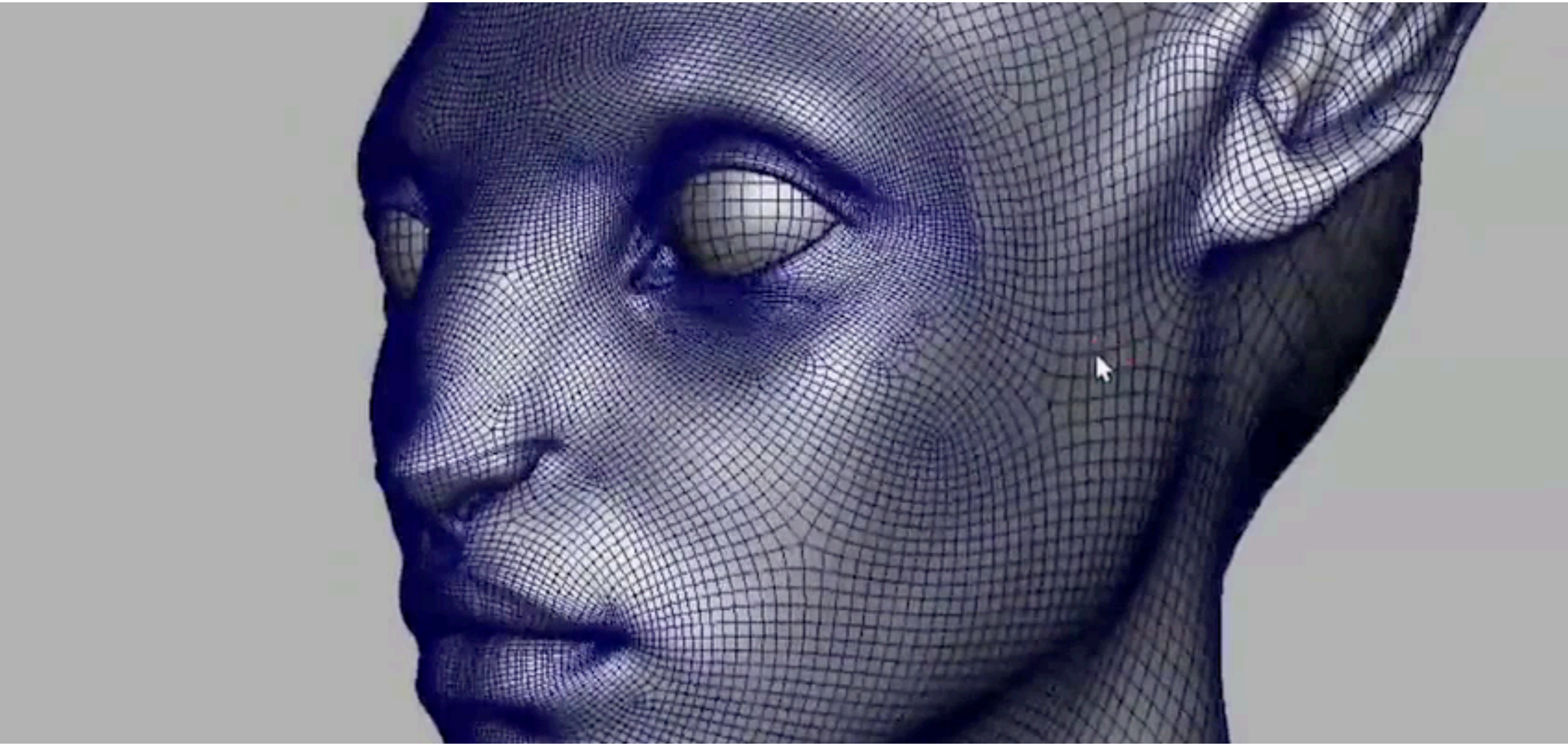
Low-Cost Active Scanning

Microsoft Kinect & Kinect Fusion

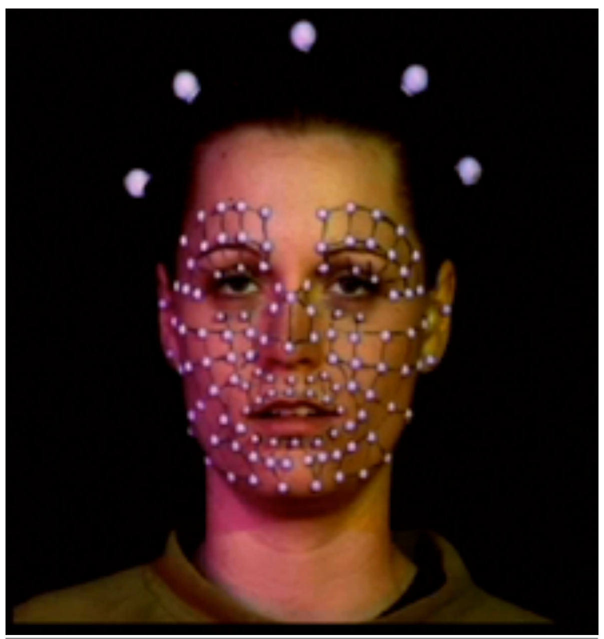
Rigging & Animation



Blendshapes & Correctives for Realism



Motion Capture Technologies



Sparse Markers

Dense Markers

MOVA

Markerless

Image Metrics

Using Markers

input performance

input video
with markers

tracking

retargeting

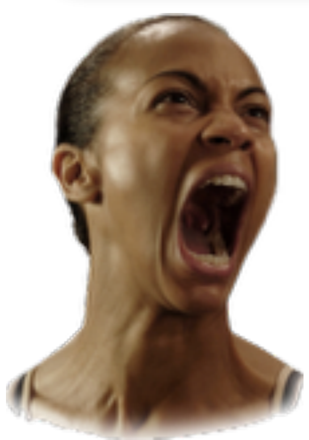
Using Dense Markers

Vision-Based Tracking **& Texturing**

LA Noire

Typical **Animation Workflow in Industry**

3D Scanning → Modeling + Fitting → Light-weight Rigging → Motion Capture → Cleanup & Key-Framing



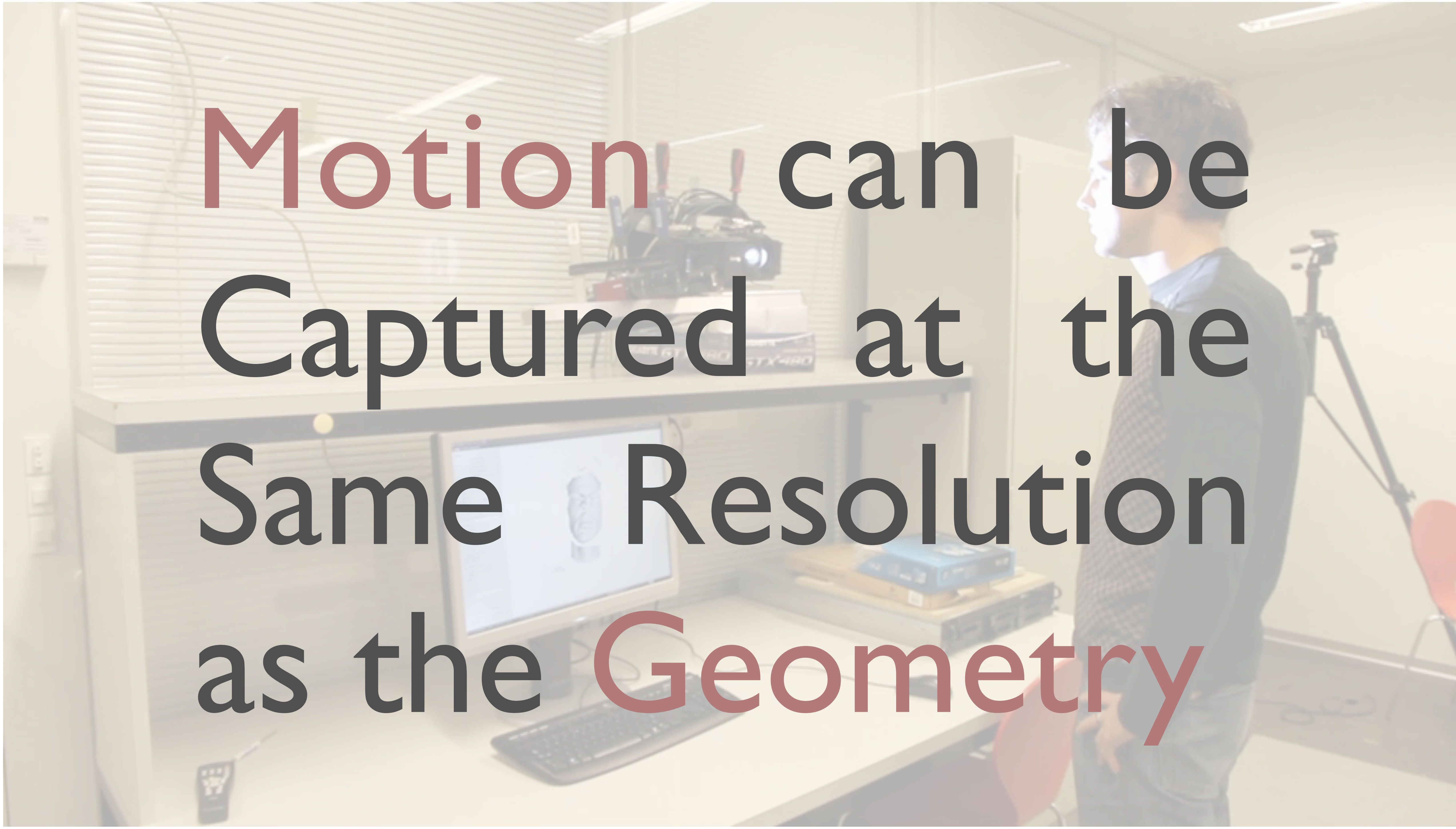
Source Character

Modeling → Complex Rigging → Retargeting → Key-Framing + Proc.+Sim.

Target Character

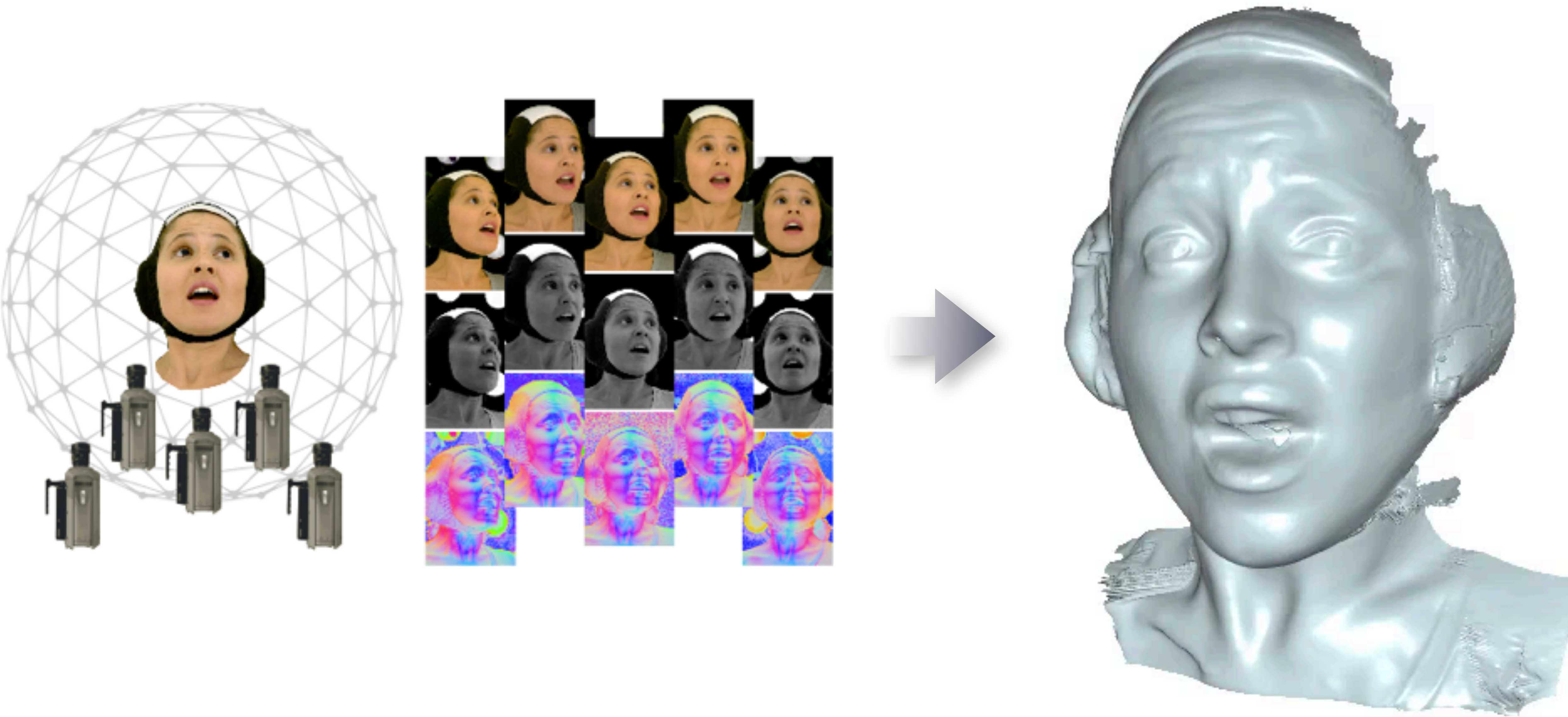
Markerless Facial Capture

3D Range Sensor

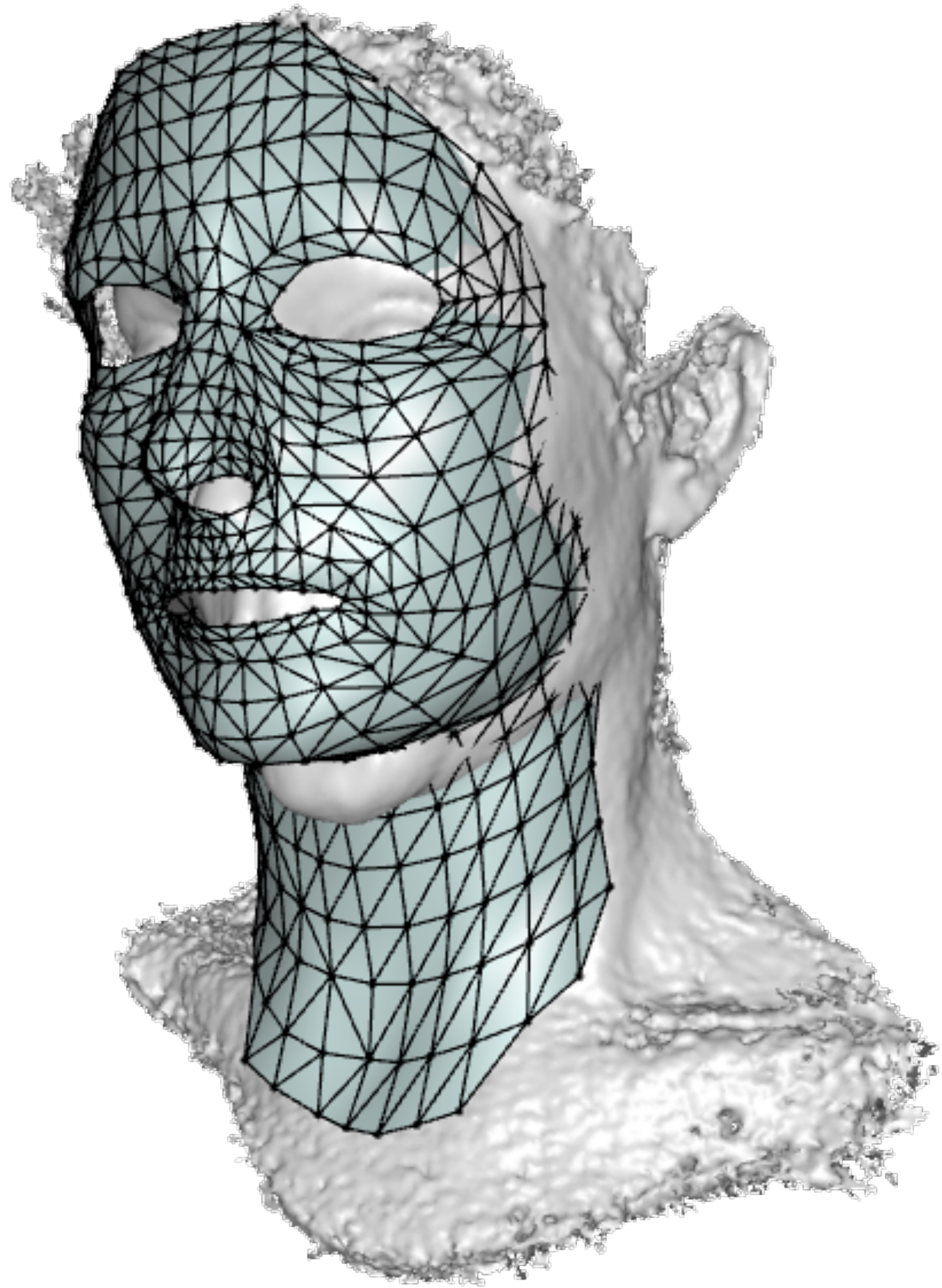


Vapor Ware? (Spatial Phase Imaging)

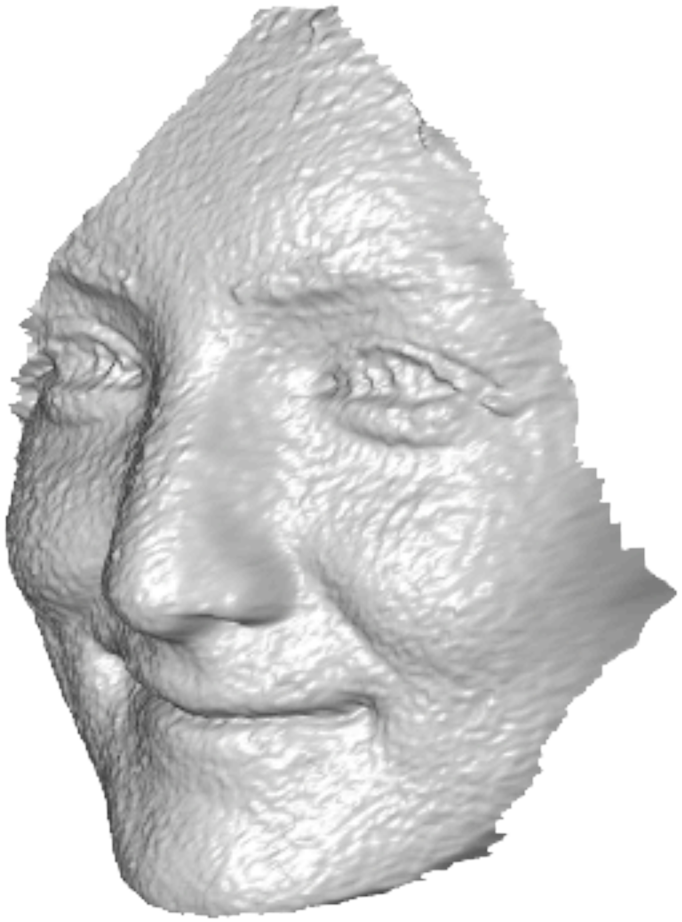
USC ICT Light Stage 5



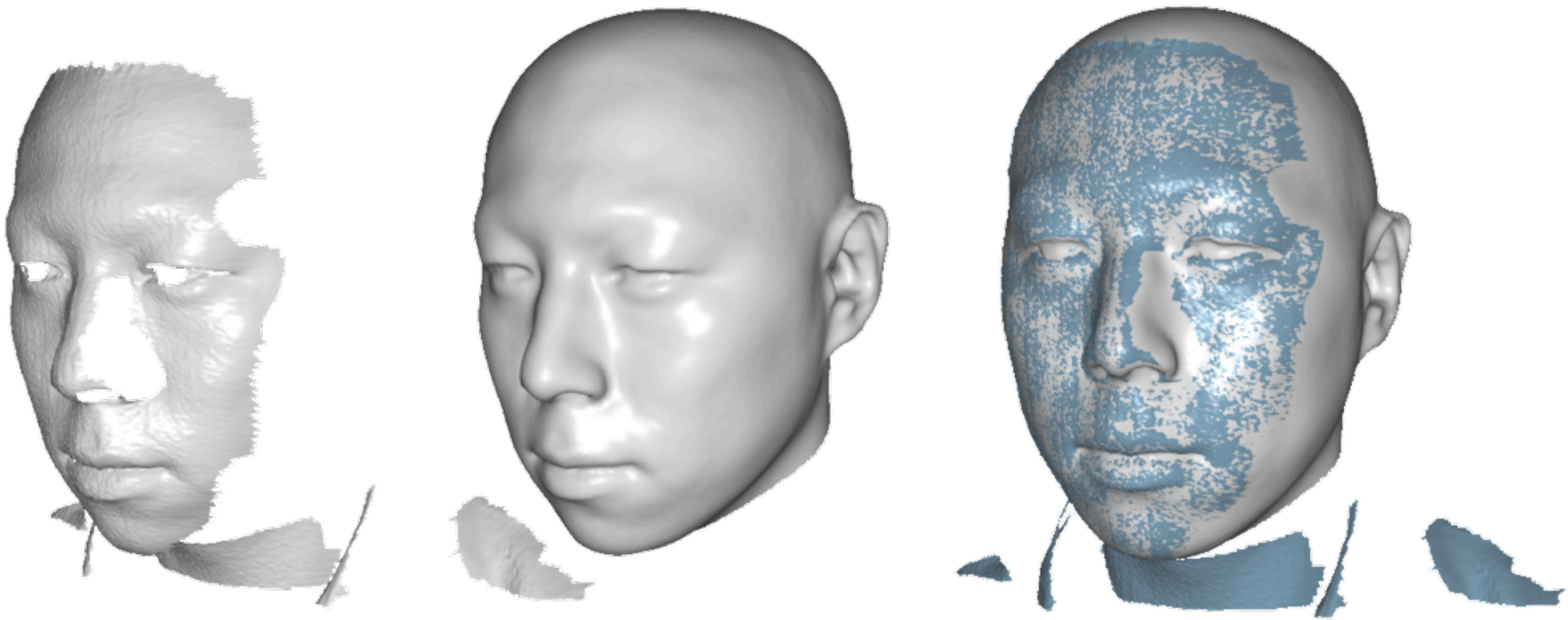
Template Fitting



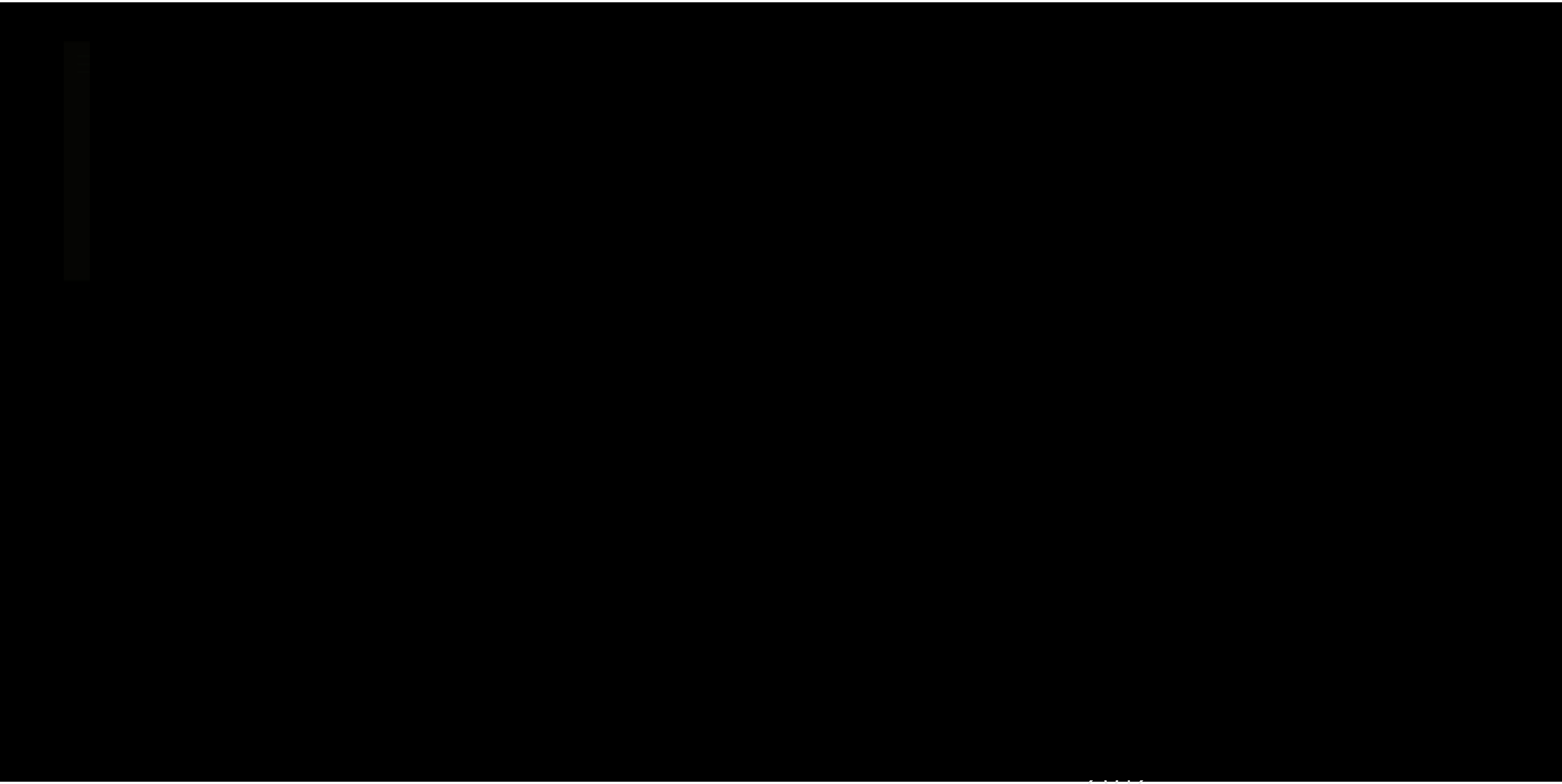
Template Fitting with PCA



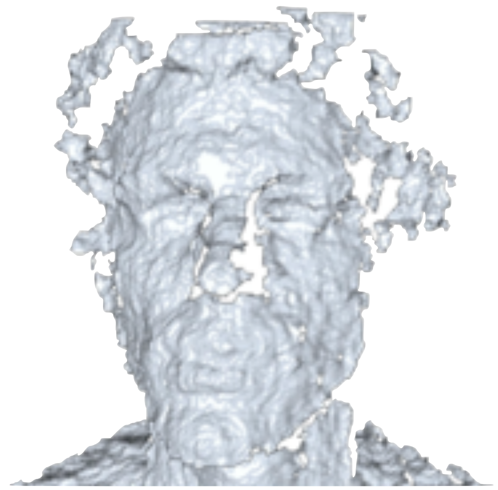
Template Based Tracking

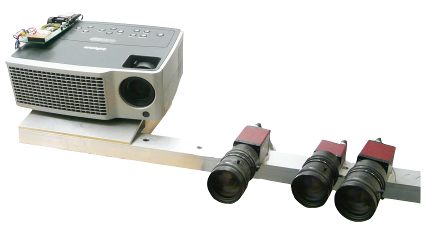


Overview



Requirements for a Practical System





1. Real-time performance
2. Robustness to noise
3. High-level semantics

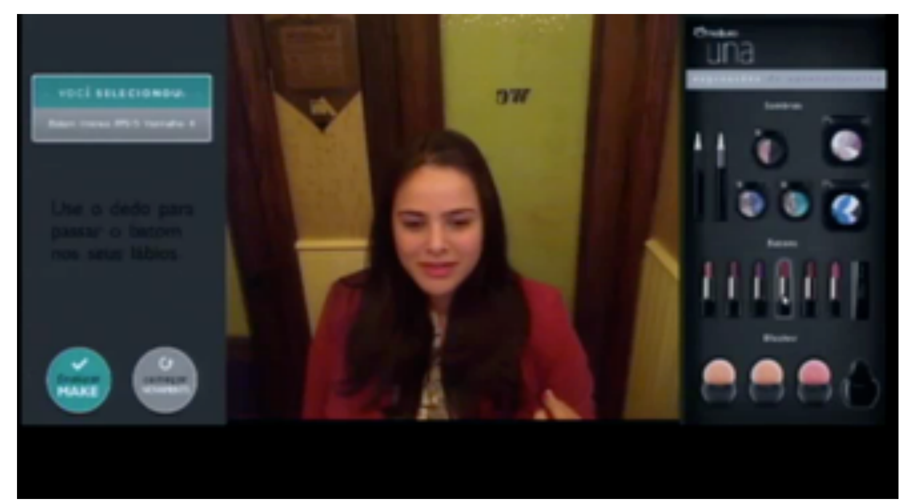
Realtime Facial Capture

Why Realtime?

VFX/Game Production

Virtual Avatars

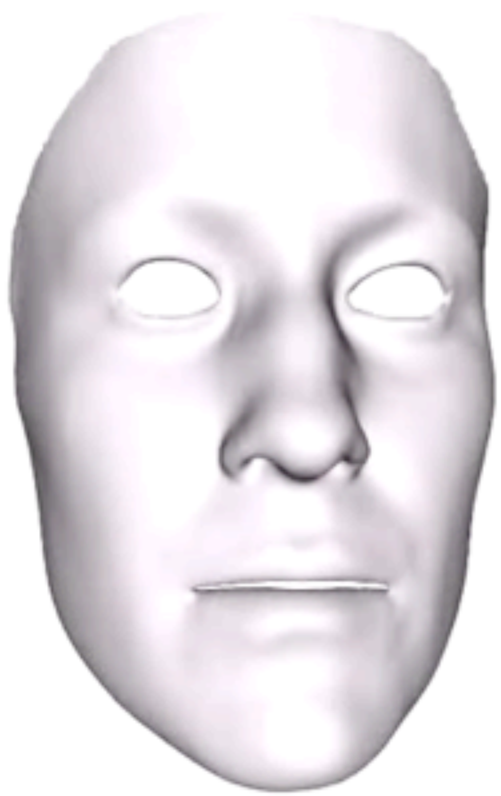
Robotics



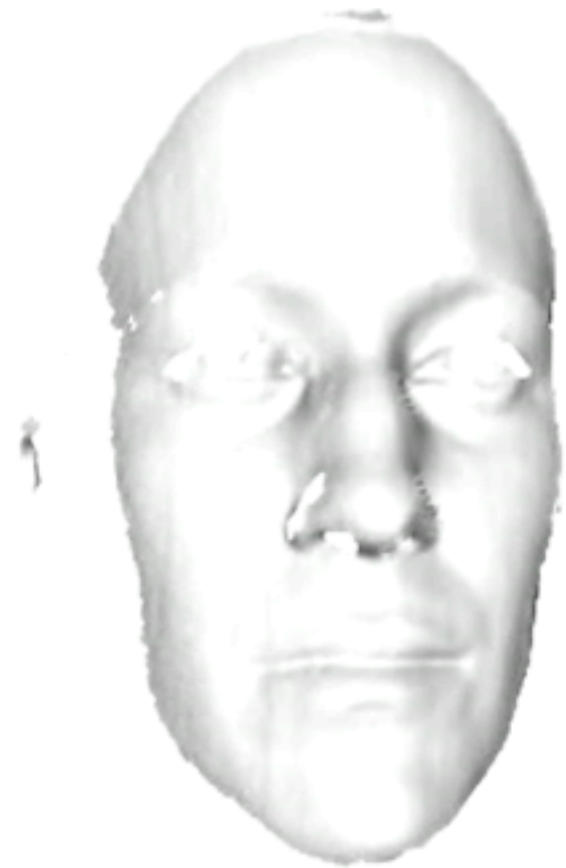
AR/Virtual Mirror

Objective

Building Expression Space

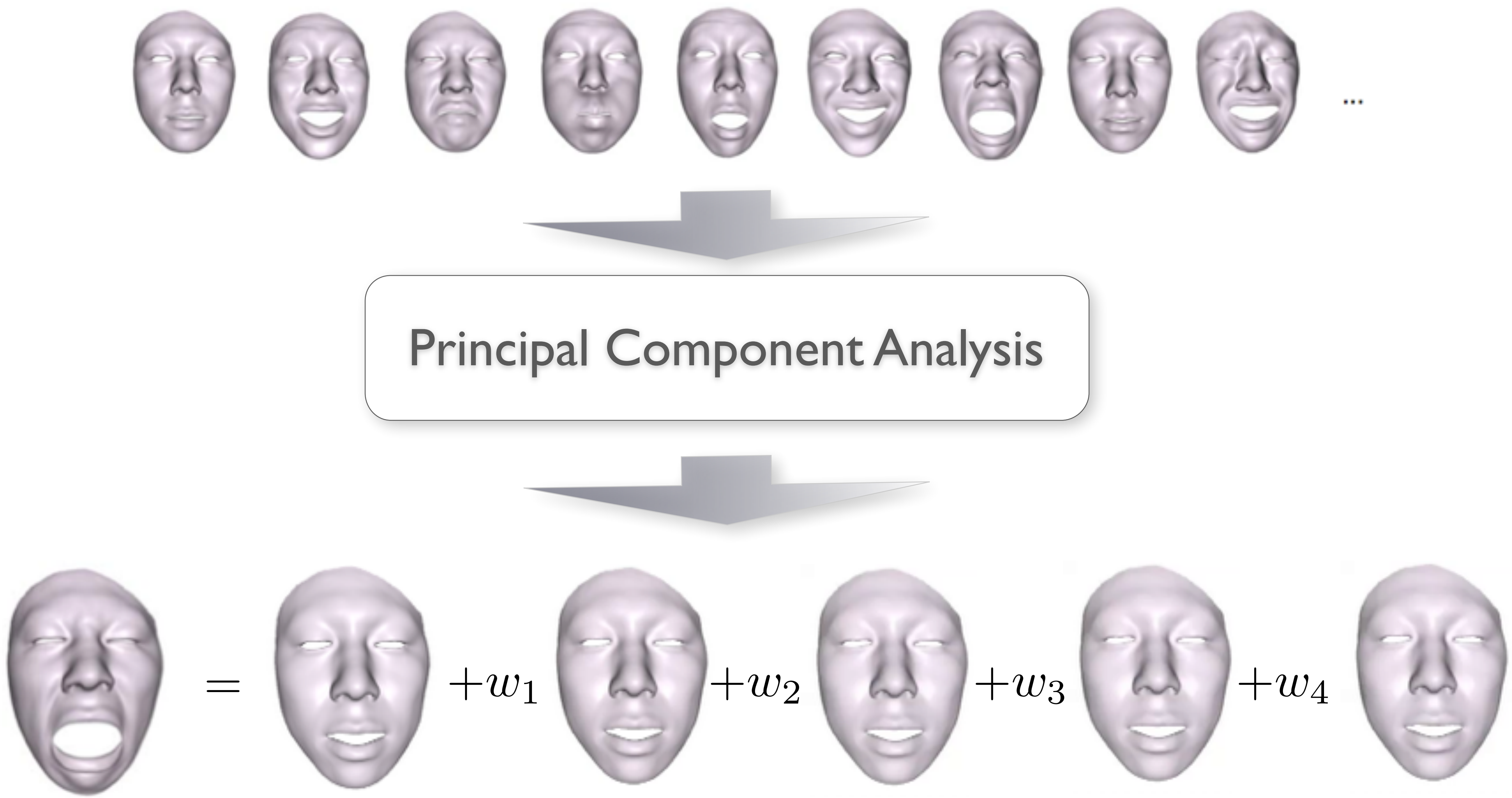


tracked template



input scan

Expression PCA for Reduced Dimension



Realtime Systems

depth sensor as input

with training
Weise et al. SCA 09

with little training
Li et al. Siggraph 2010

little to no training
Weise et al. Siggraph 2011 &
Bouaziz et al. Siggraph 2013

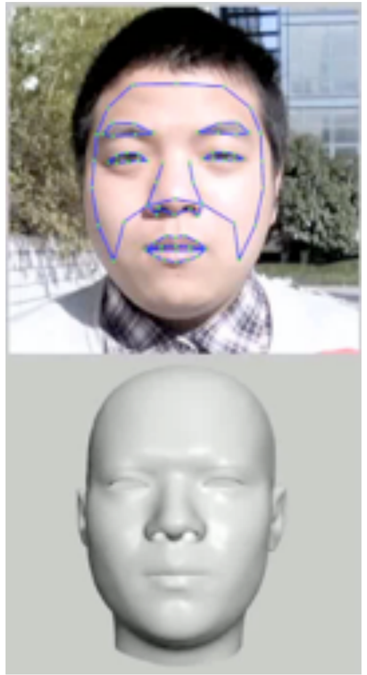
reduced calibration and more accessible

Realtime Systems

video as input

without training
Saragih et al. IJCV 2011

without training
Image Metrics 2011



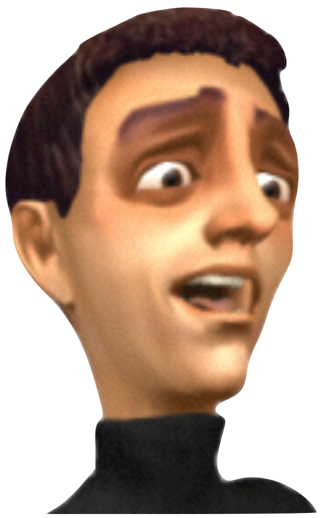
with training
Cao et al. Siggraph 2013

increased accuracy and expressiveness

Automatic Facial Rigging

Blendshape Animation

Blendshape Animation



$$= B_0 + \alpha_1 B_1 + \alpha_2 B_2 + \alpha_3 B_3 + \ldots$$

Blendshape Retargeting

many blendshapes

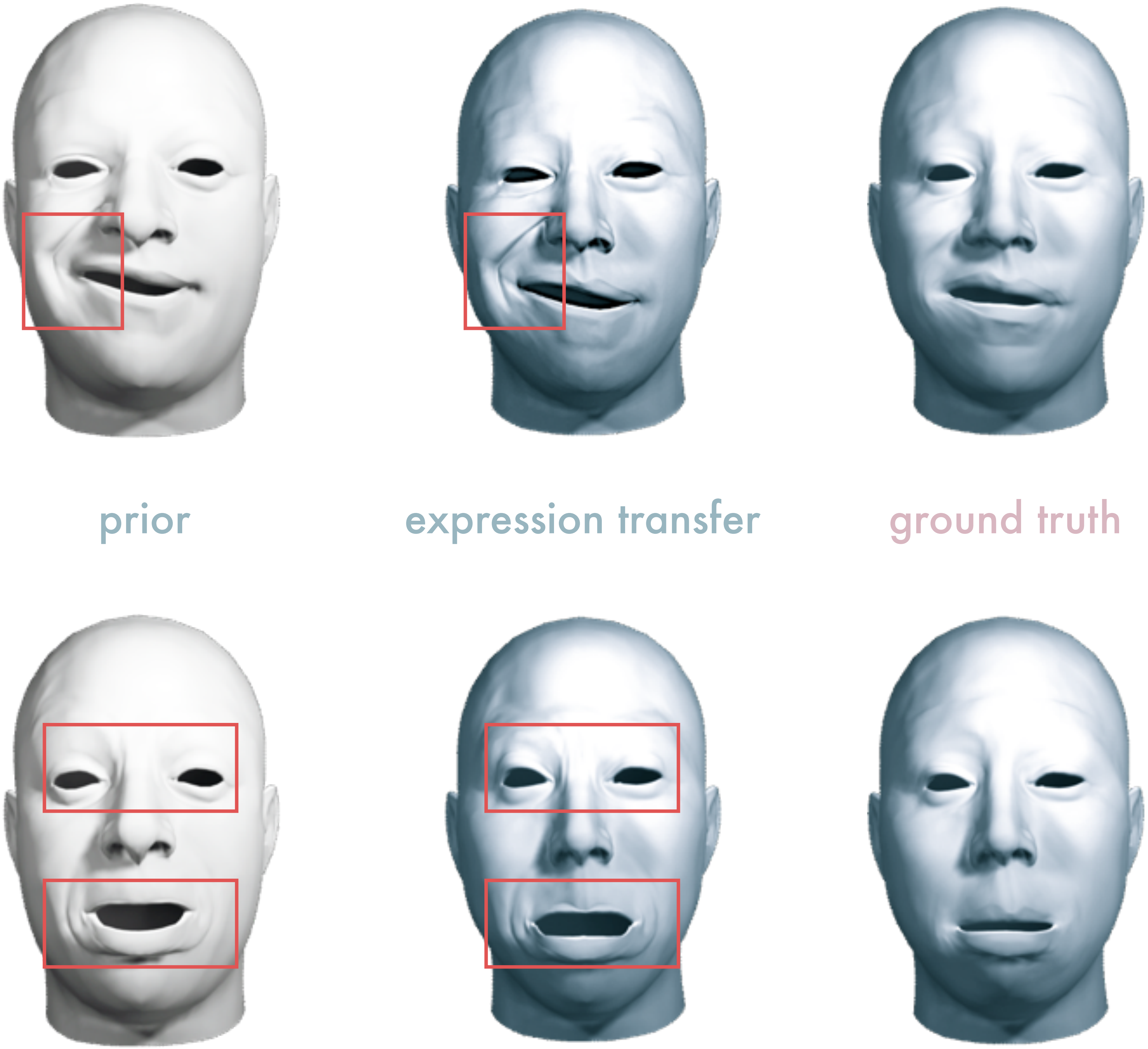
Expression Transfer

prior
blendshapes

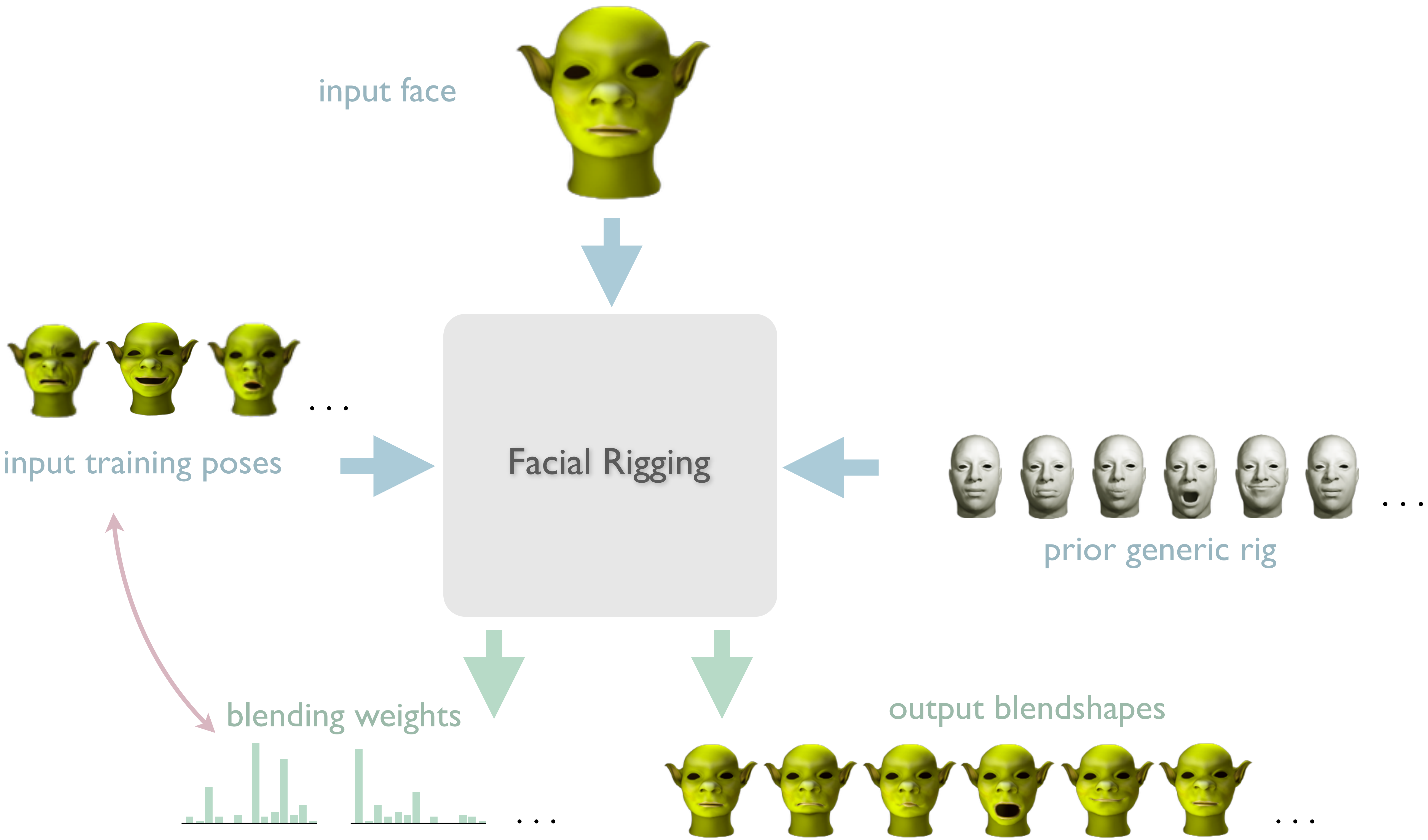
[Noh & Neumann '01]
[Sumner & Popovic '04]

reconstructed
blendshapes

Problems



Example Based-Facial Rigging



Bilinear Problem

input face

input examples

$$B_0 + \sum_{i=1}^{n} \alpha_{ij} B_i \approx S_j$$

blending weights

output blendshapes

Decoupled Optimization

$$B_0 + \sum_{i=1}^{n} \alpha_{ij} B_i \;\approx\; S_j$$

Decoupled Optimization

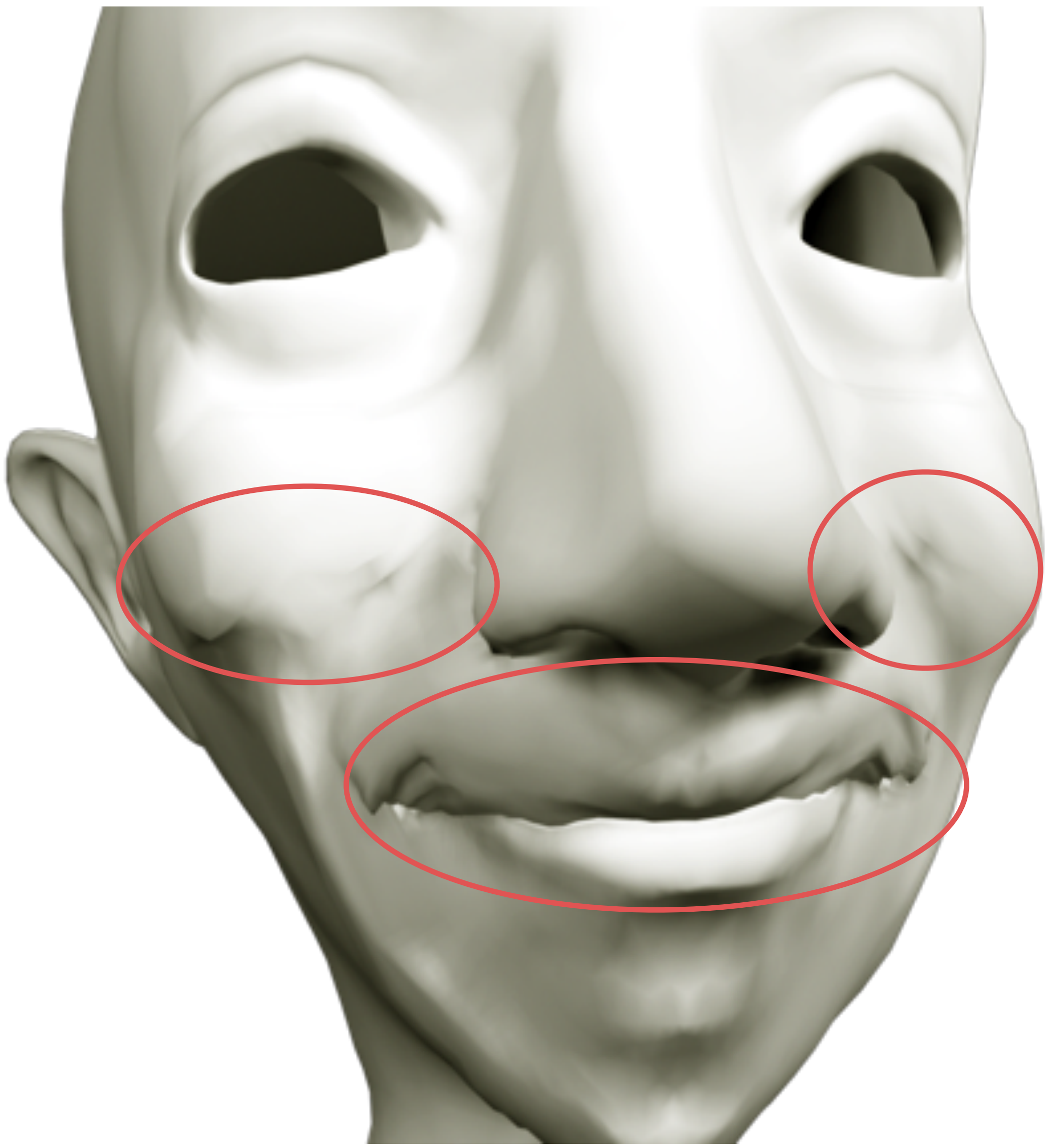
Step A

$$B_0 + \sum_{i=1}^{n} \alpha_{ij} B_i \approx S_j$$

Step B

$$B_0 + \sum_{i=1}^{n} \alpha_{ij} B_i \approx S_j$$

Gradient Domain Optimization



$$\underset{B_i}{\operatorname{argmin}} \| B_0 + \sum_{i=1}^{n} \alpha_{ij} B_i - S_j \|^2 + \beta \| B_i - \tilde{B}_i \|^2$$

$$\underset{M_i}{\operatorname{argmin}} \| M_0 + \sum_{i=1}^{n} \alpha_{ij} M_i - M_j^S \|^2 + \beta \| M_i + M_0 - G_i \cdot M_0 \|^2$$

Comparison

prior | without examples | with 6 examples — whistle

input example — surprise

Directable Facial Animation

3D scans

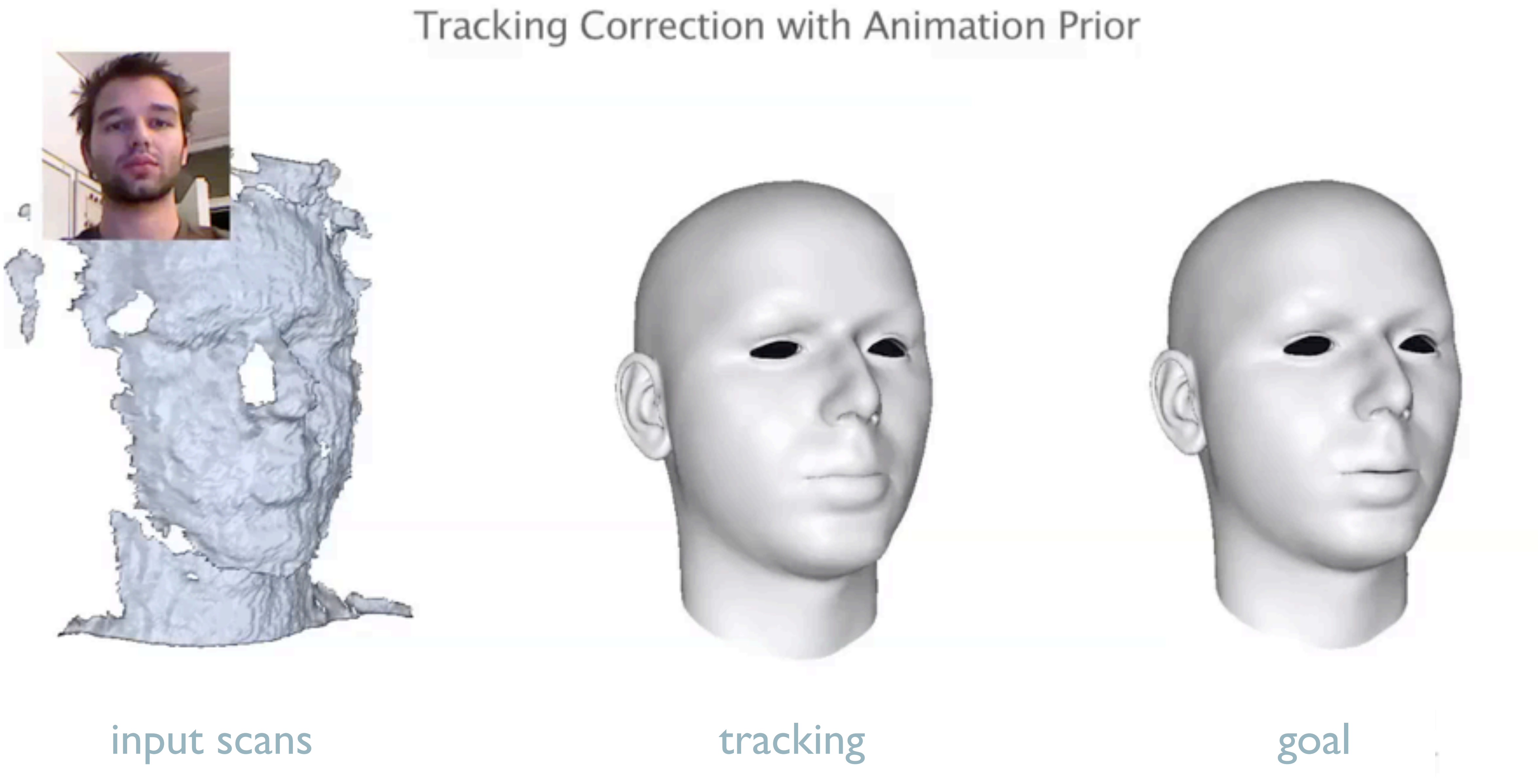
facial tracking

Blendshapes for Tracking

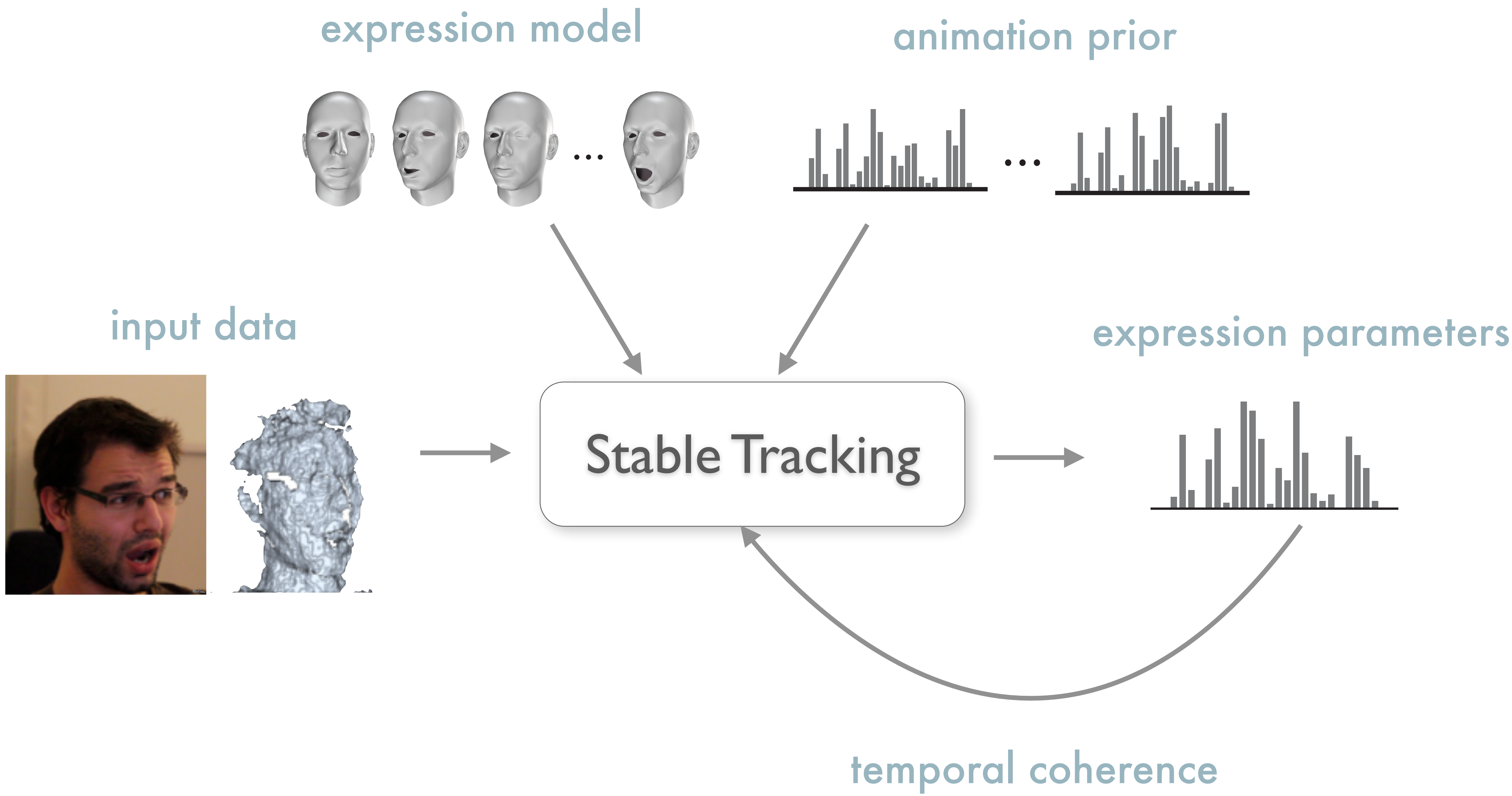
ICP with Blendshapes

Animation Prior

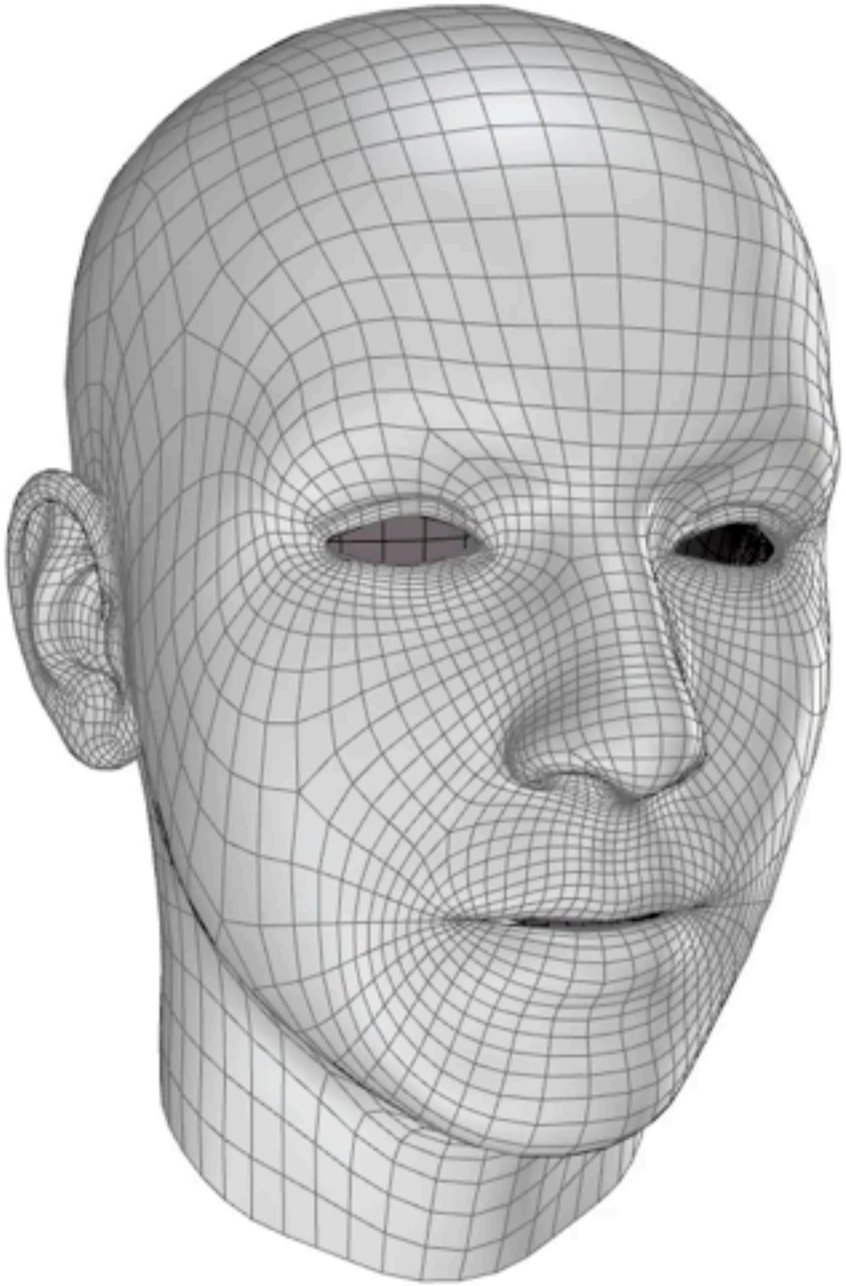
Problem: **Noisy Input**



Performance-Based Facial Tracking



Animation as Prior

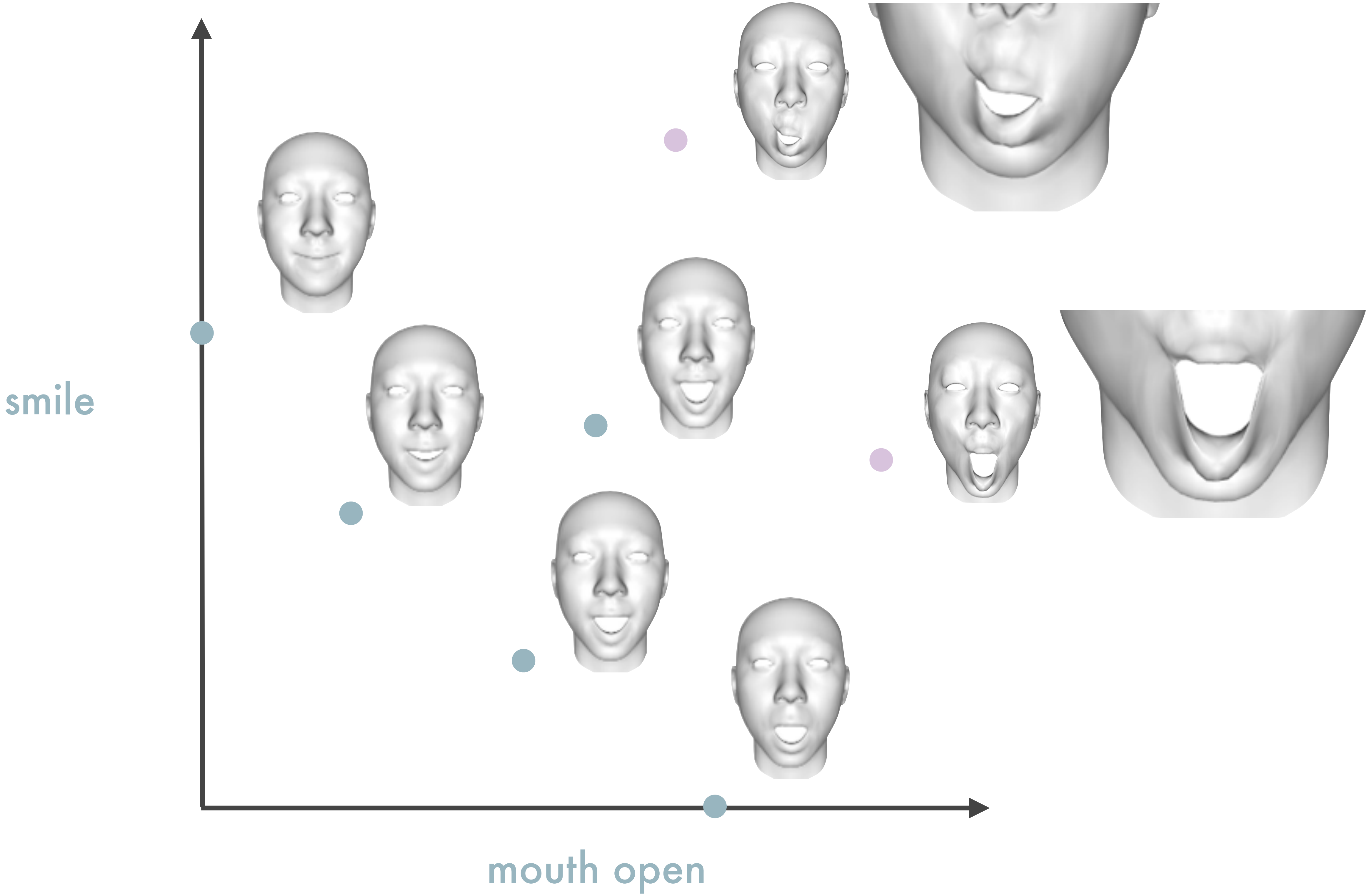


reference video

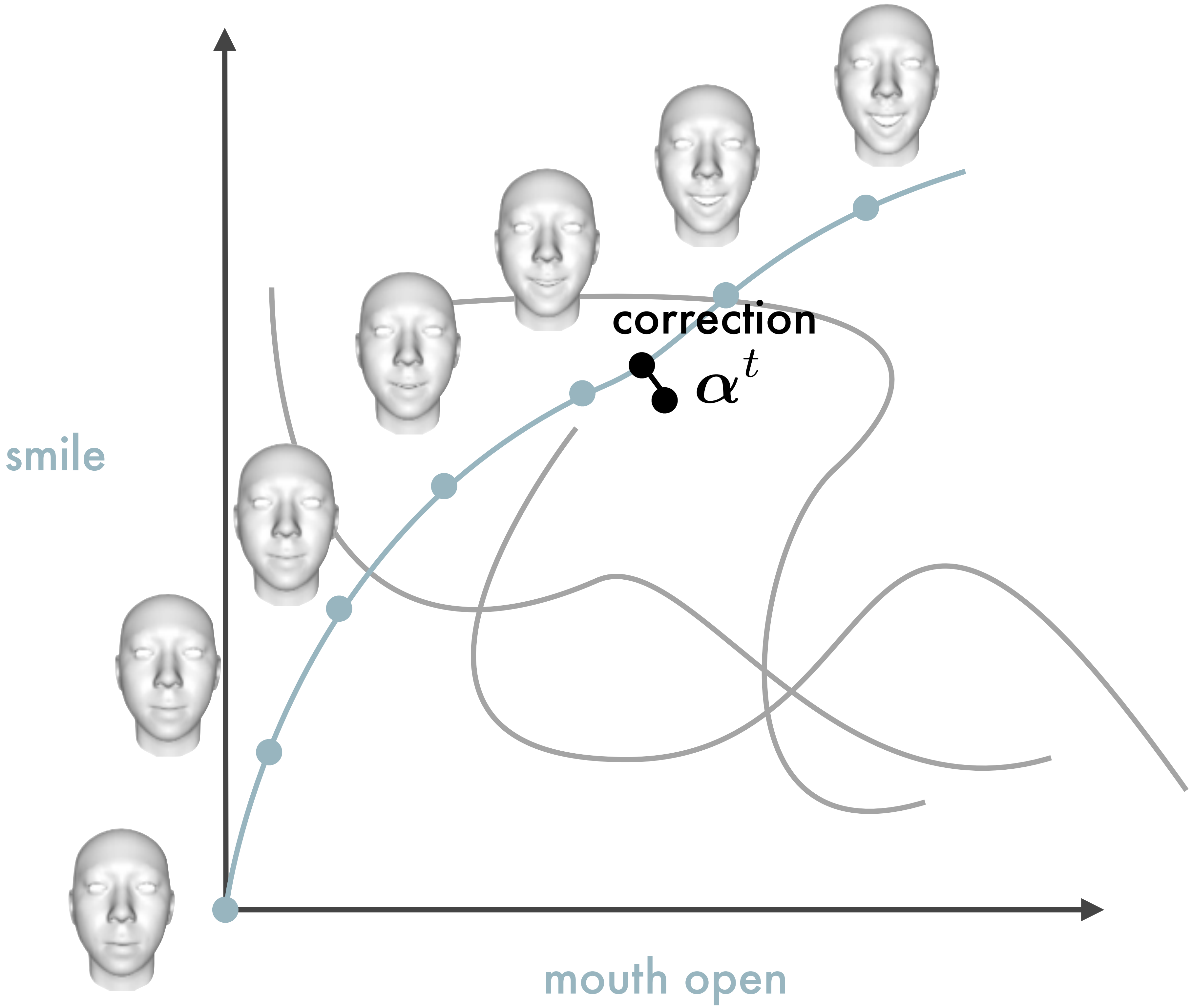
artist

9500 frames

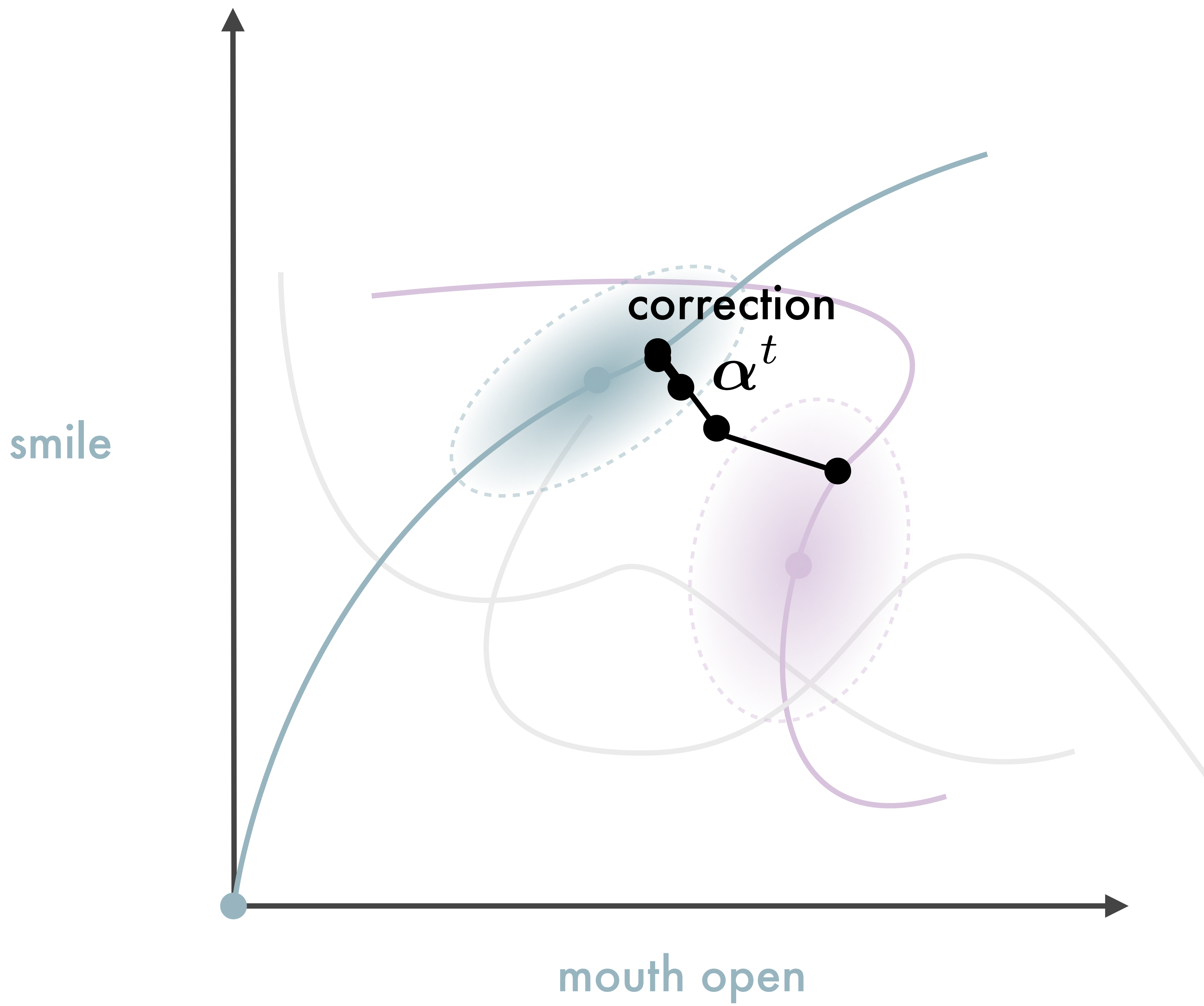
N-Dim Expression Space



Animation Manifold

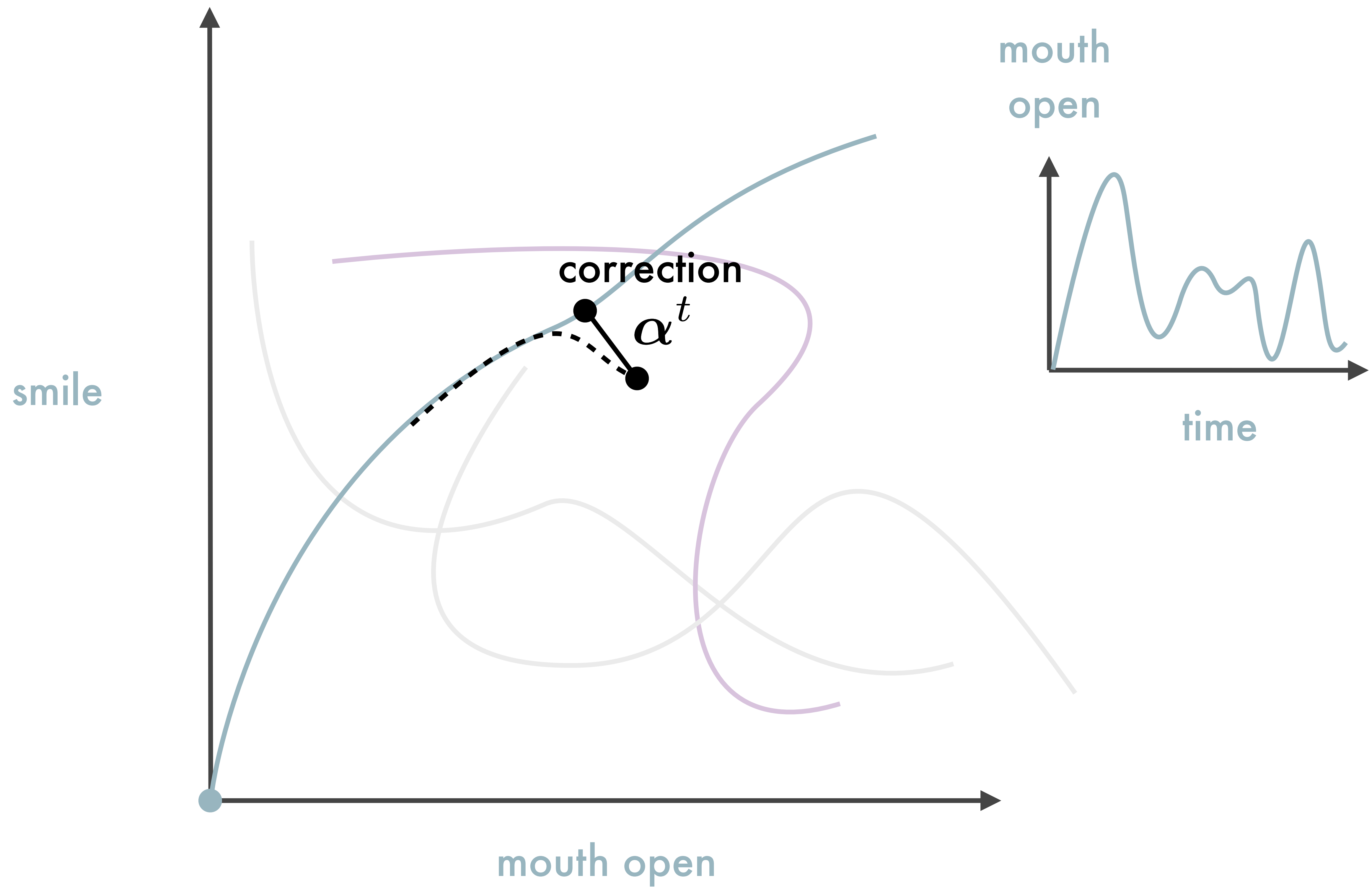


Probabilistic Animation Prior

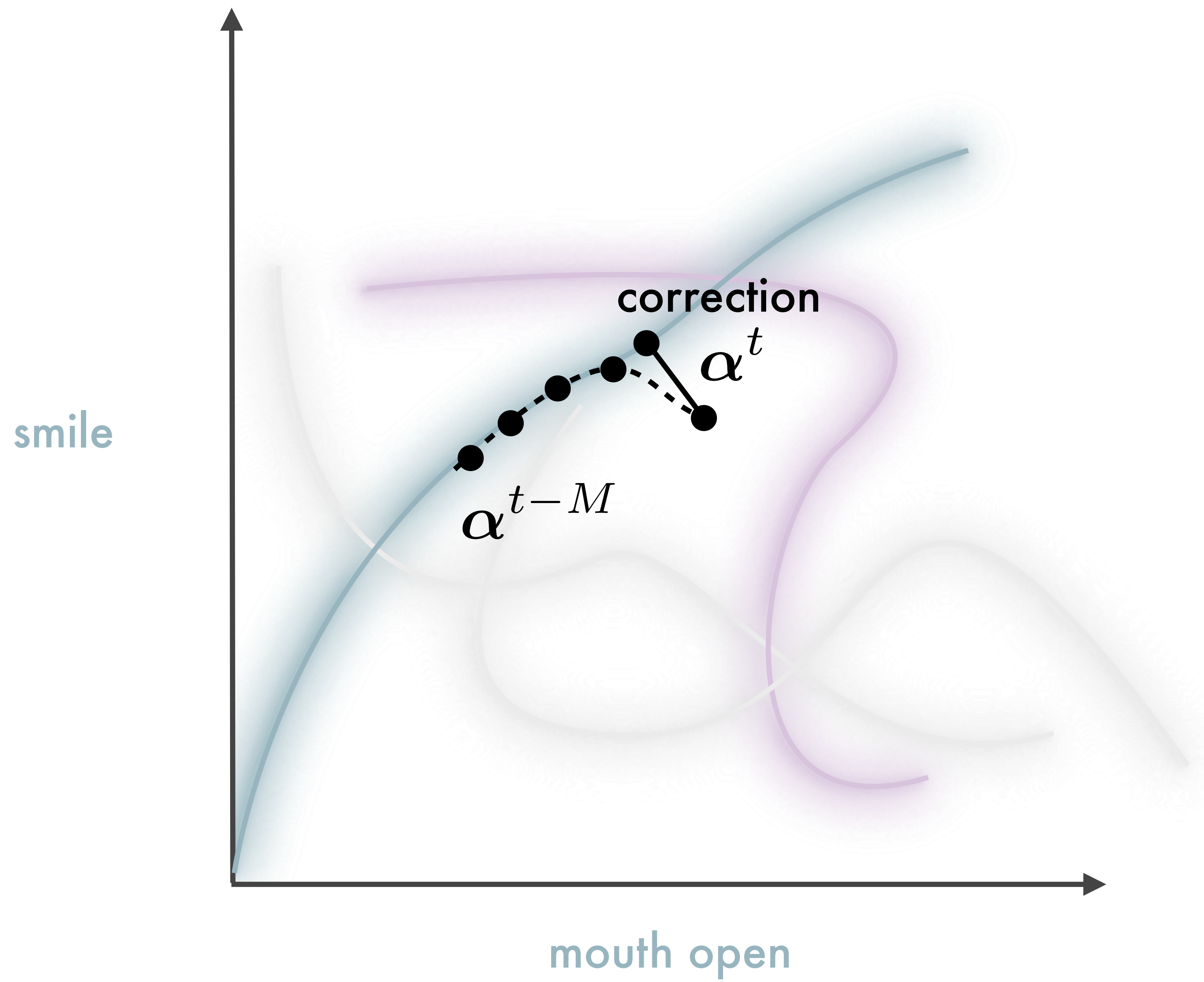


Lau et al. 2009

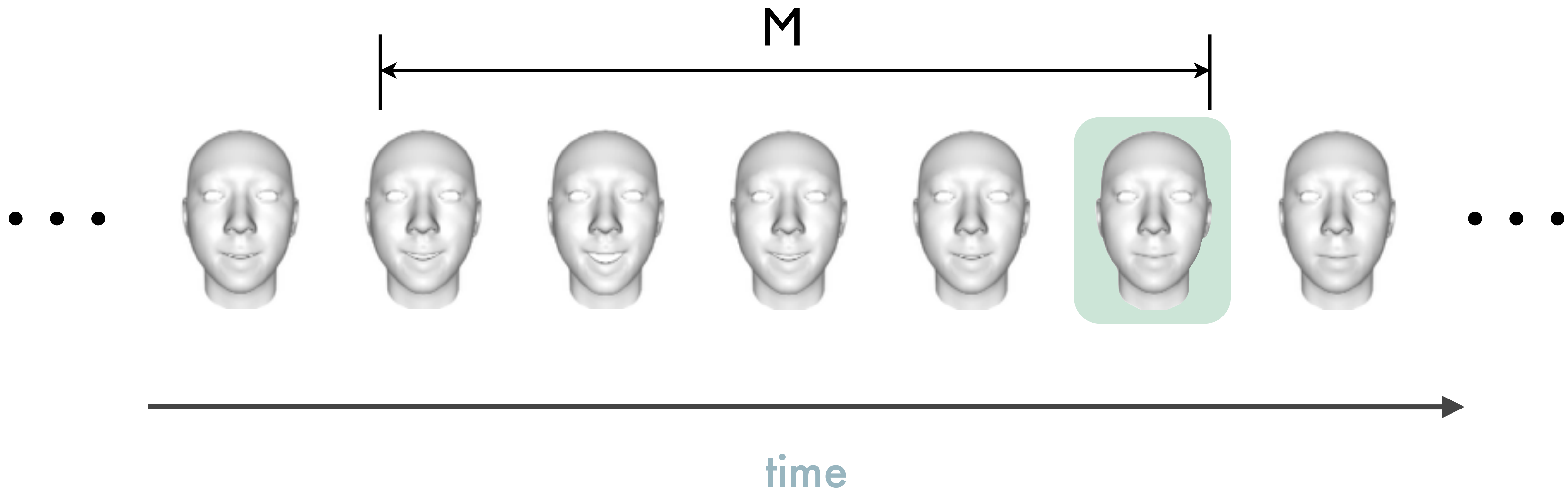
Probabilistic Animation Prior



Probabilistic Animation Prior



Temporal Joint Probabilistic Distribution



$$p(\boldsymbol{\alpha}^t, .., \boldsymbol{\alpha}^{t-M}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\boldsymbol{\alpha}^t, .., \boldsymbol{\alpha}^{t-M} | \boldsymbol{\mu}_k, C_k C_k^T + \sigma_k^2 I).$$

MPPCA model

weights

mean

principal components

Gaussian noise

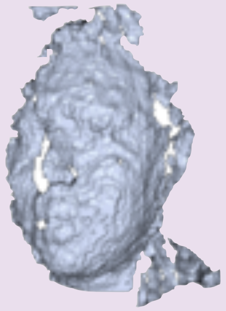
MAP Estimation

$$\boldsymbol{\alpha}^t = \arg\max_{\boldsymbol{\alpha}} p(\boldsymbol{\alpha}|D, \boldsymbol{\alpha}^{t-1}, .., \boldsymbol{\alpha}^{t-M})$$

MPPCA

$$\approx \arg\max_{\boldsymbol{\alpha}} \underbrace{p(D|\boldsymbol{\alpha})}_{\text{likelihood}} \underbrace{p(\boldsymbol{\alpha}, \boldsymbol{\alpha}^{t-1}, .., \boldsymbol{\alpha}^{t-M})}_{\text{prior}}$$

geometry



$$p(G|\mathbf{x}) = \prod_{i=1}^{V} k_{geo} \exp\left(-\frac{||\mathbf{n}_i^T(\mathbf{v}_i - \mathbf{v}_i^*)||^2}{2\sigma_{\text{geo}}^2}\right)$$

texture

$$p(I|\mathbf{x}) = \prod_{i=1}^{V} k_{im} \exp\left(-\frac{||\nabla I_i^T(\mathbf{p}_i - \mathbf{p}_i^*)||^2}{2\sigma_{\text{im}}^2}\right)$$

ILM's Kinect Monster Mirror

Fast **Calibration**

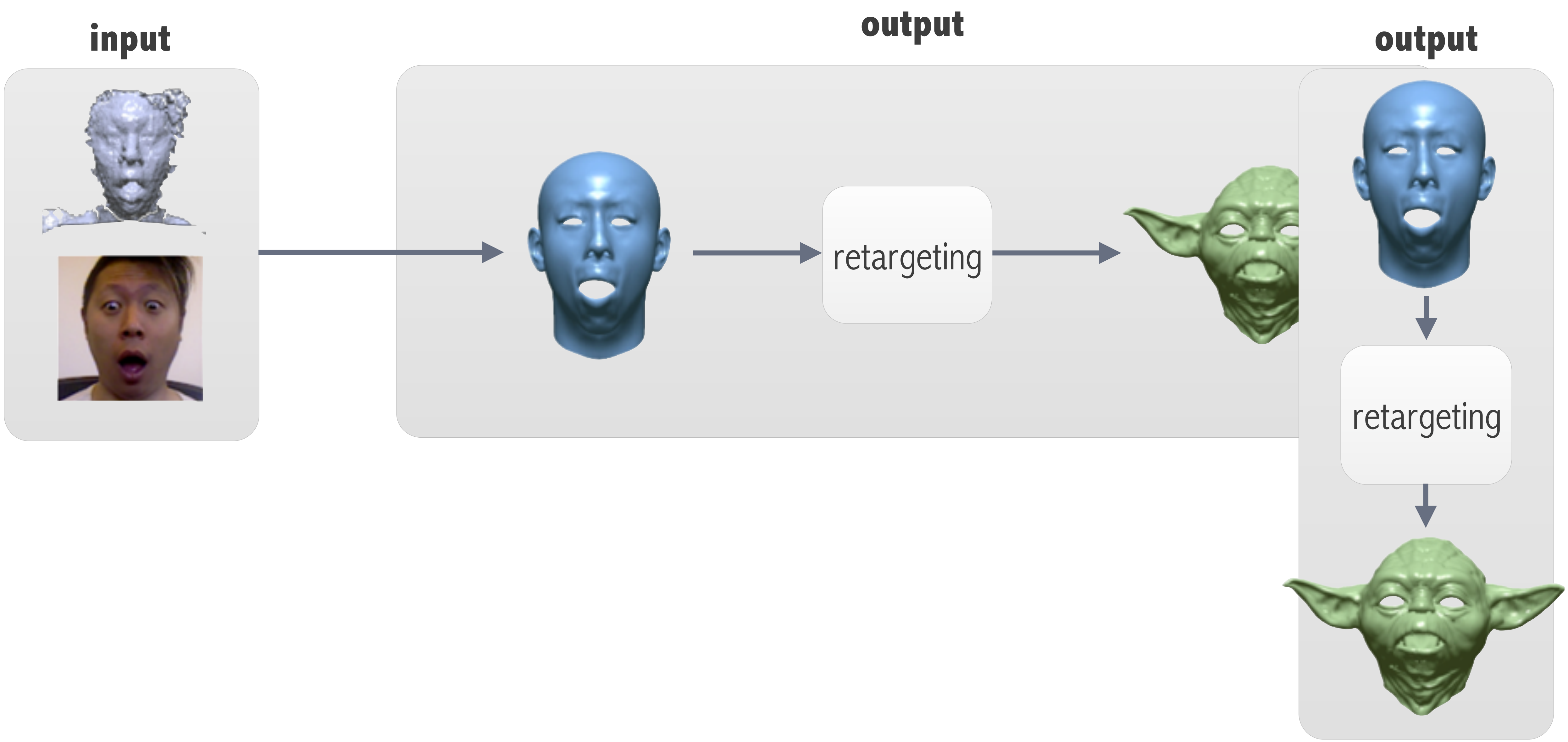
Li et al. SIGGRAPH 2013

Facial Performance Capture

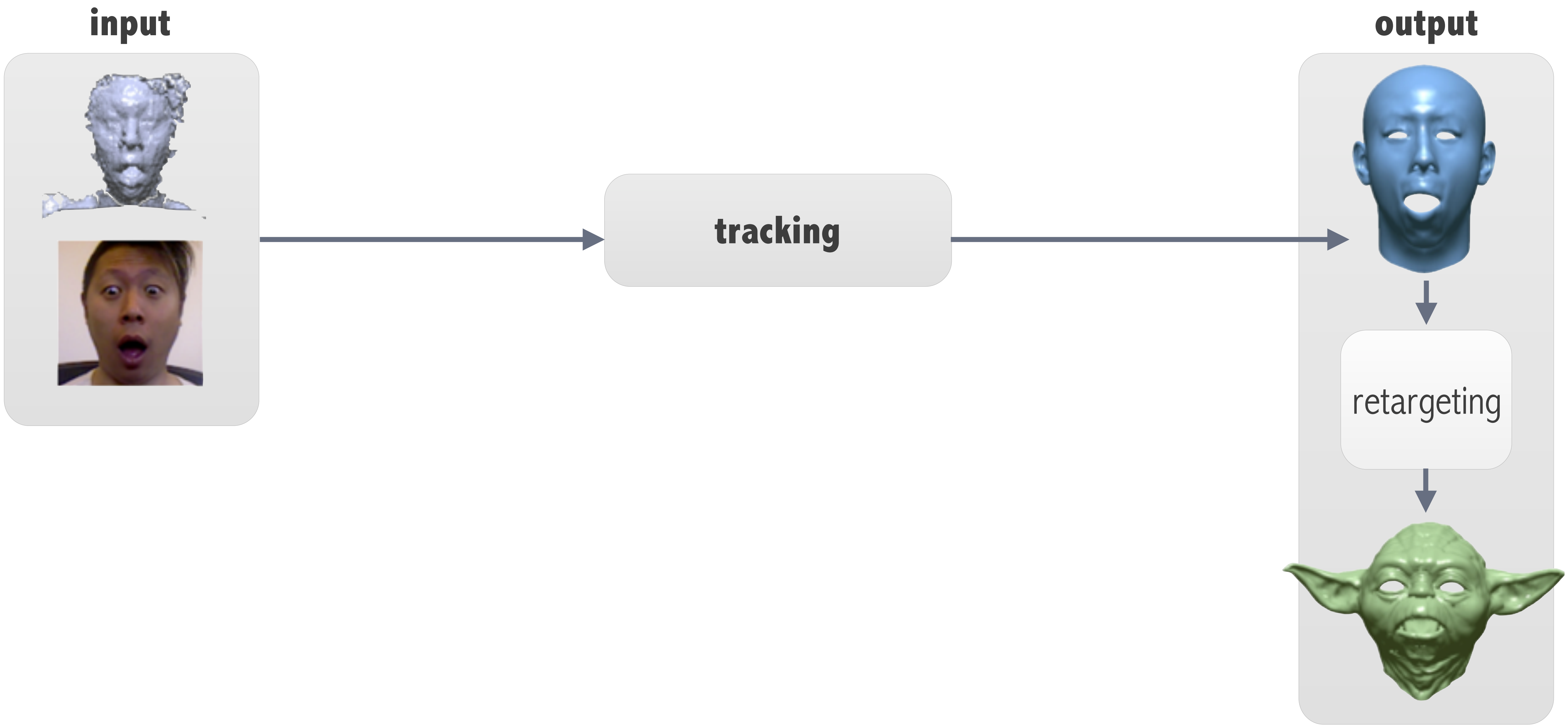
Li et al. SIGGRAPH 2013

Pipeline

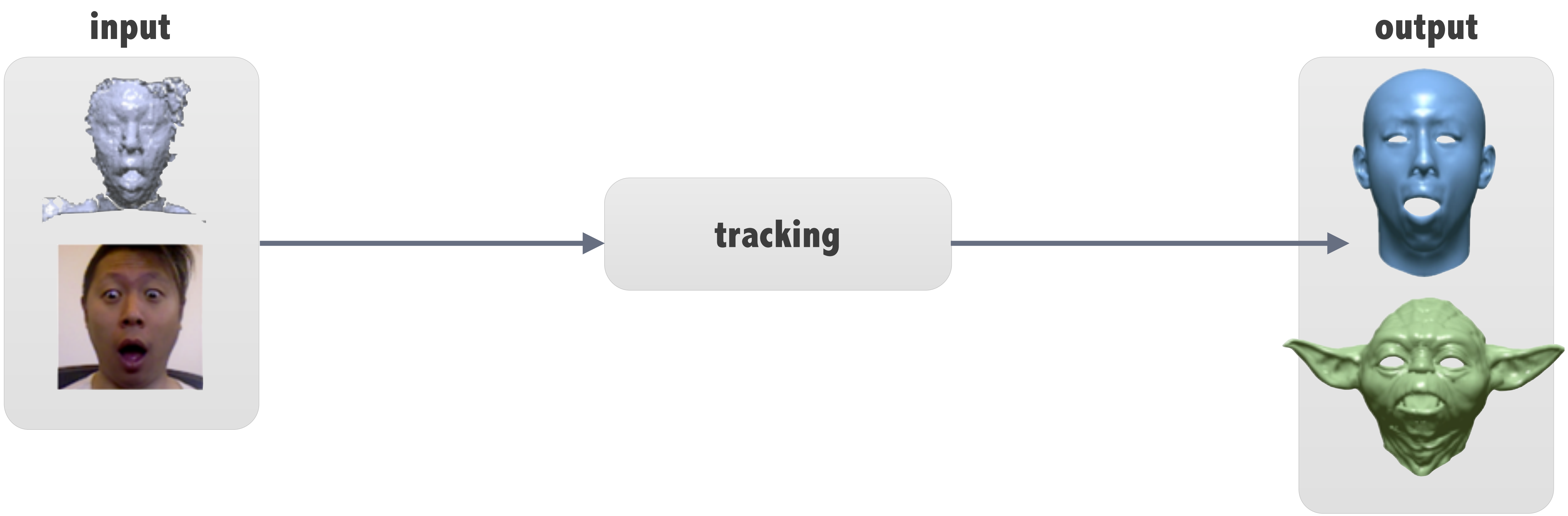
Pipeline Overview



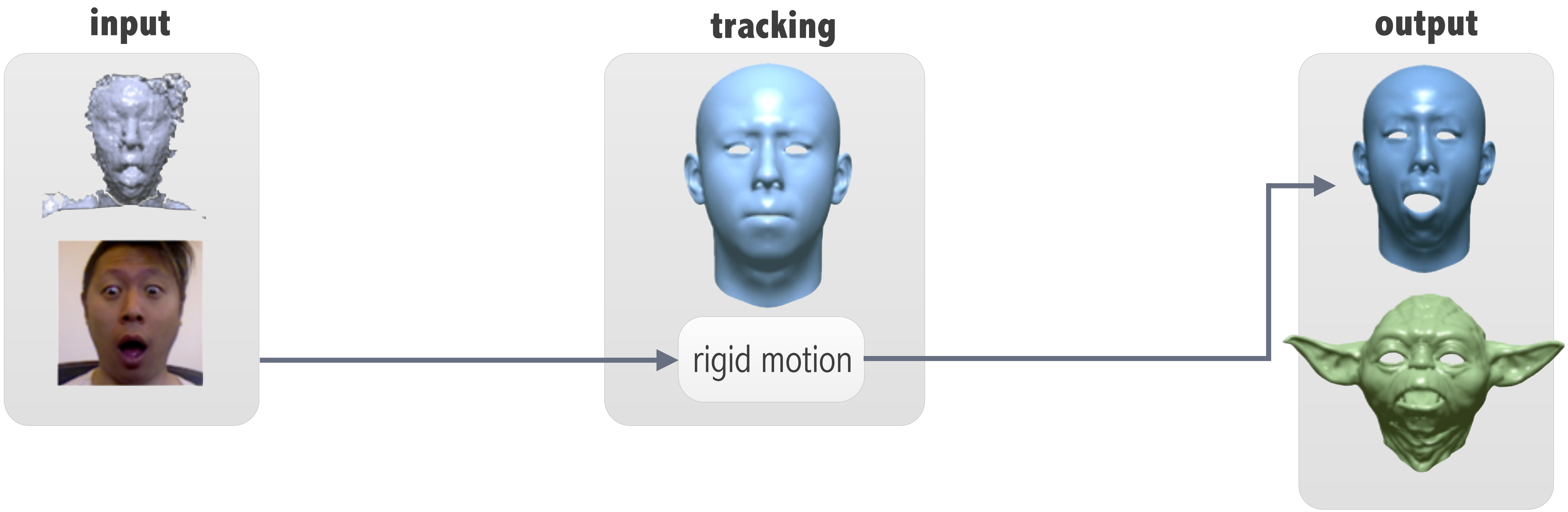
Pipeline Overview



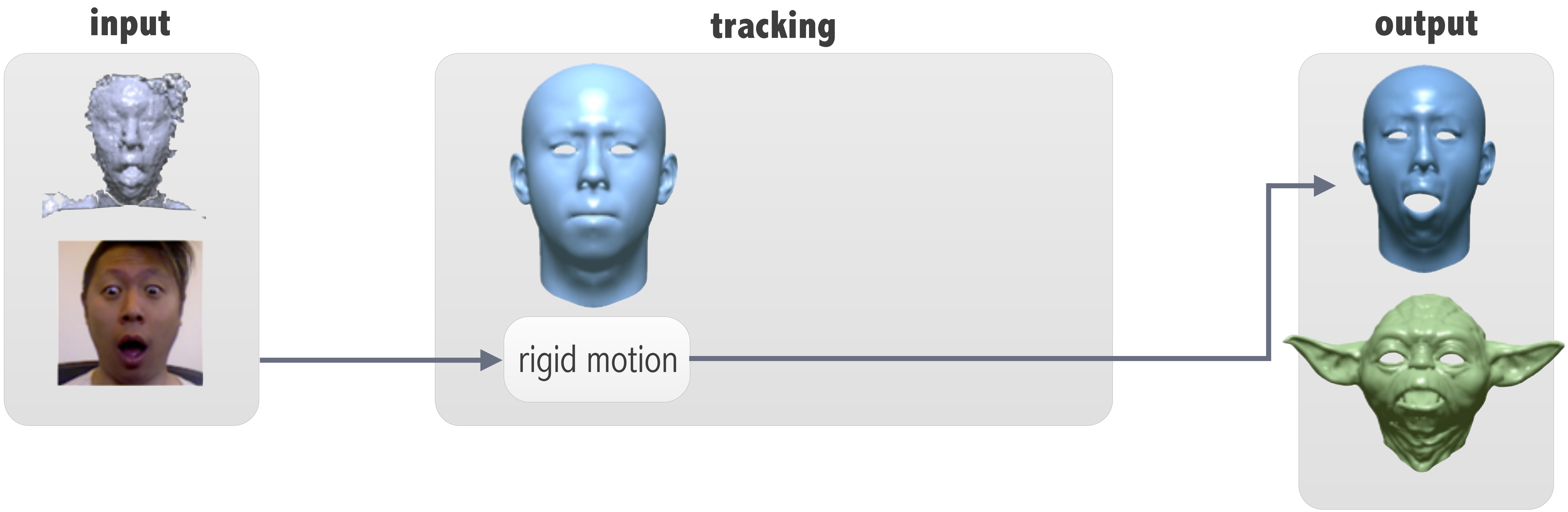
Pipeline Overview



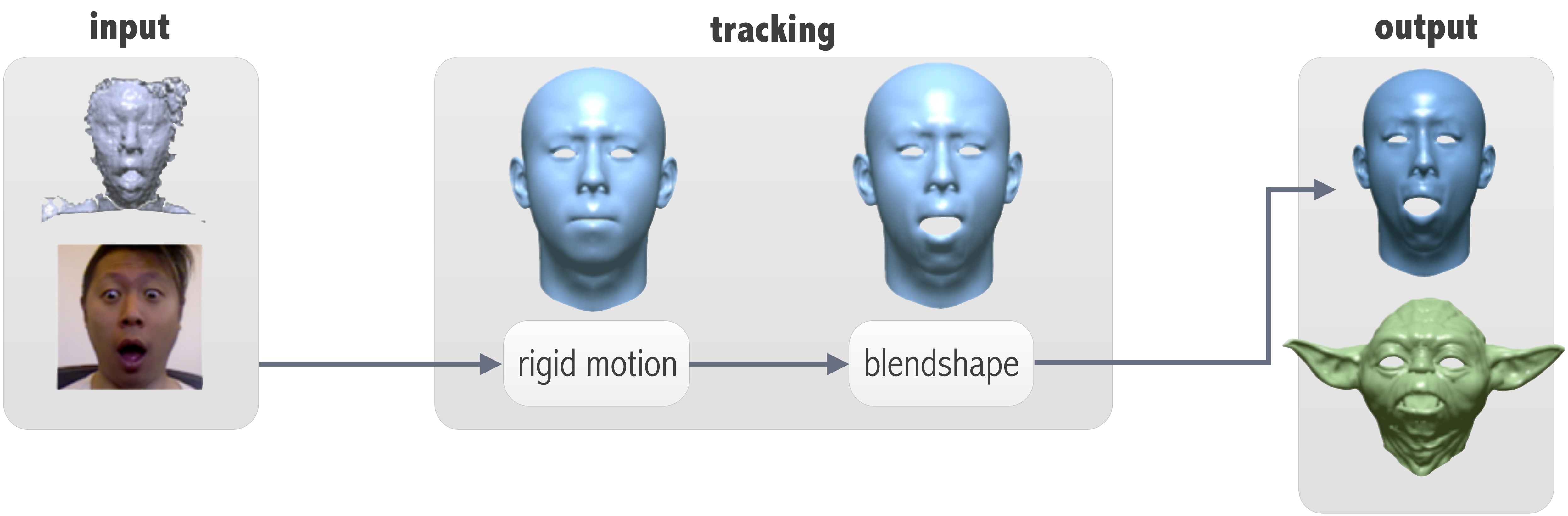
Pipeline Overview



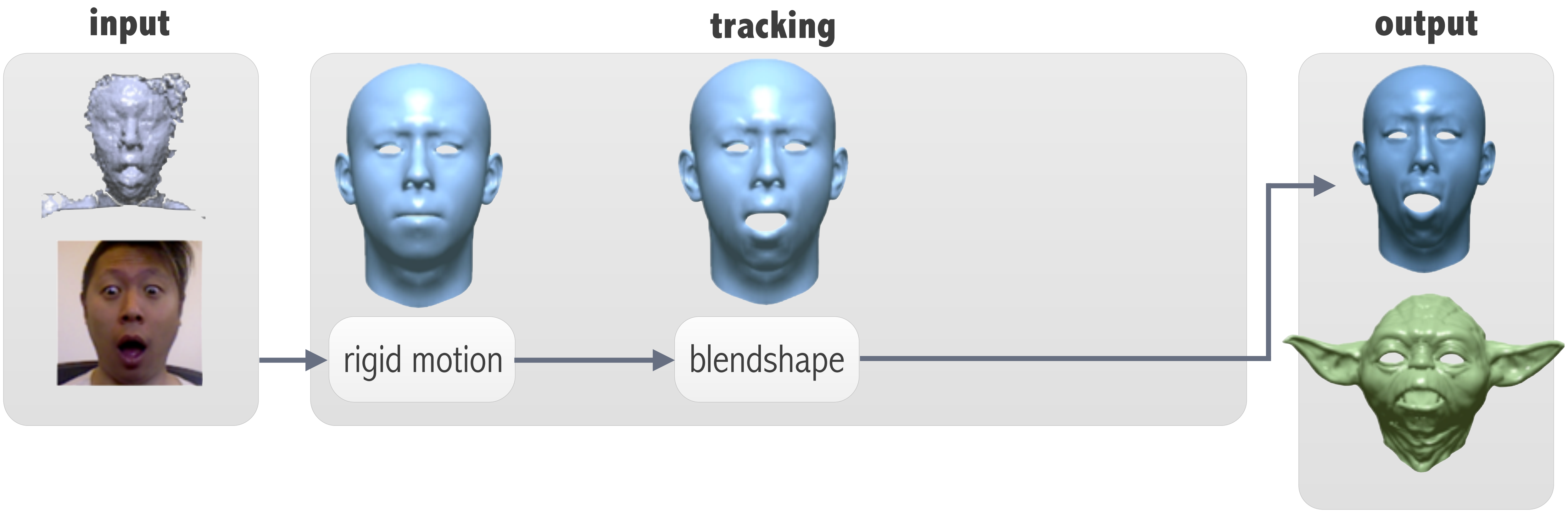
Pipeline Overview



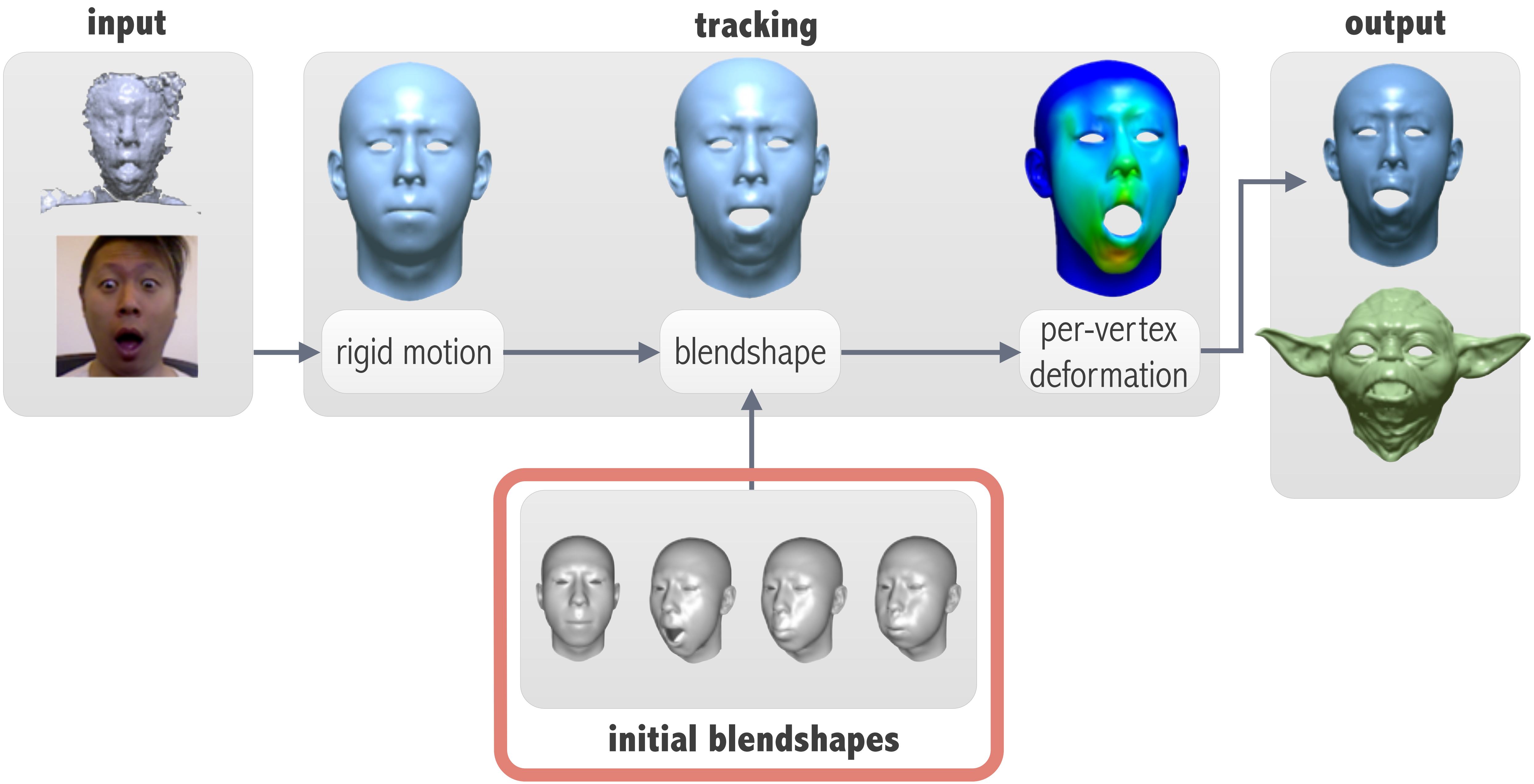
Pipeline Overview



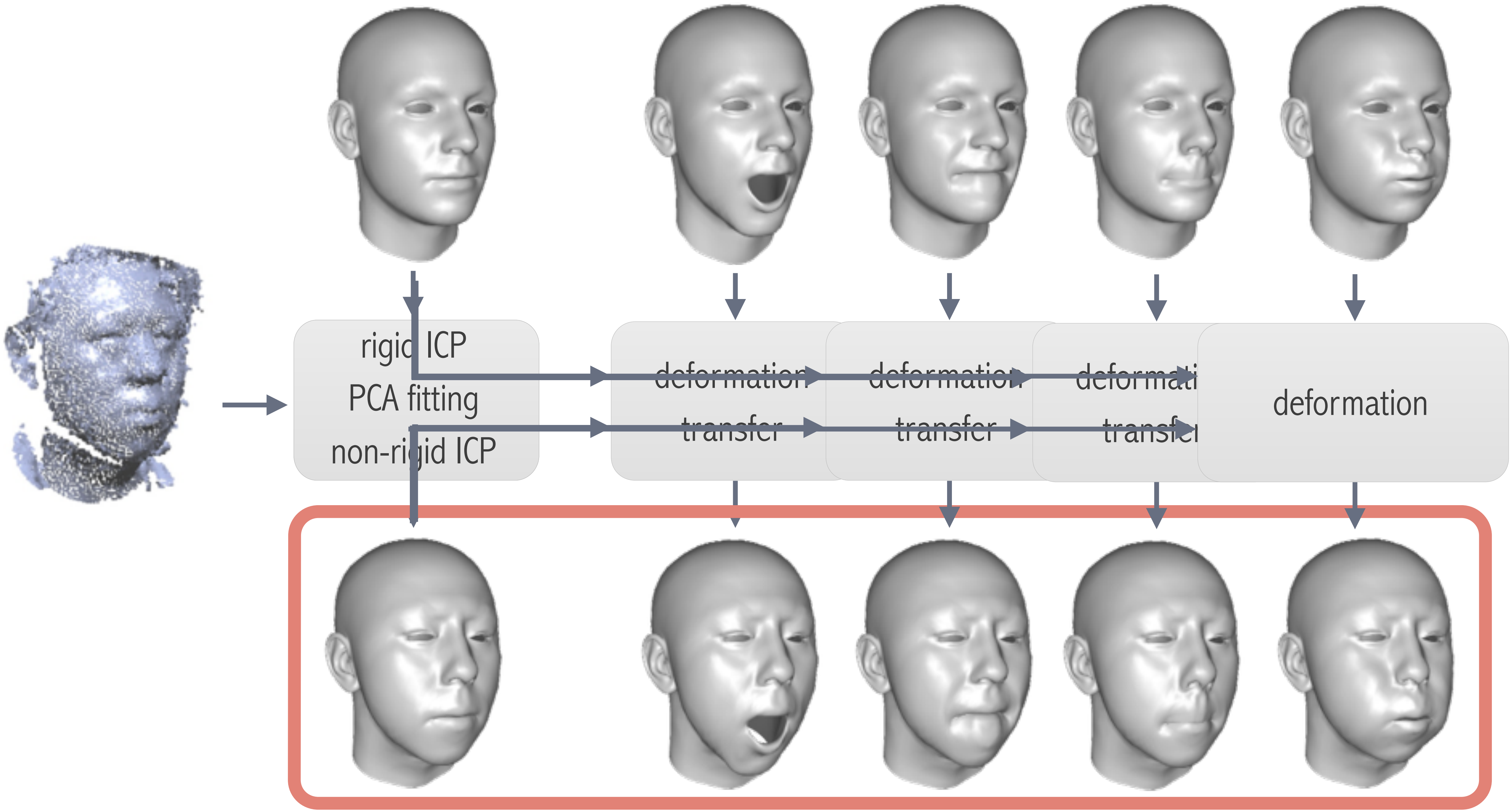
Pipeline Overview



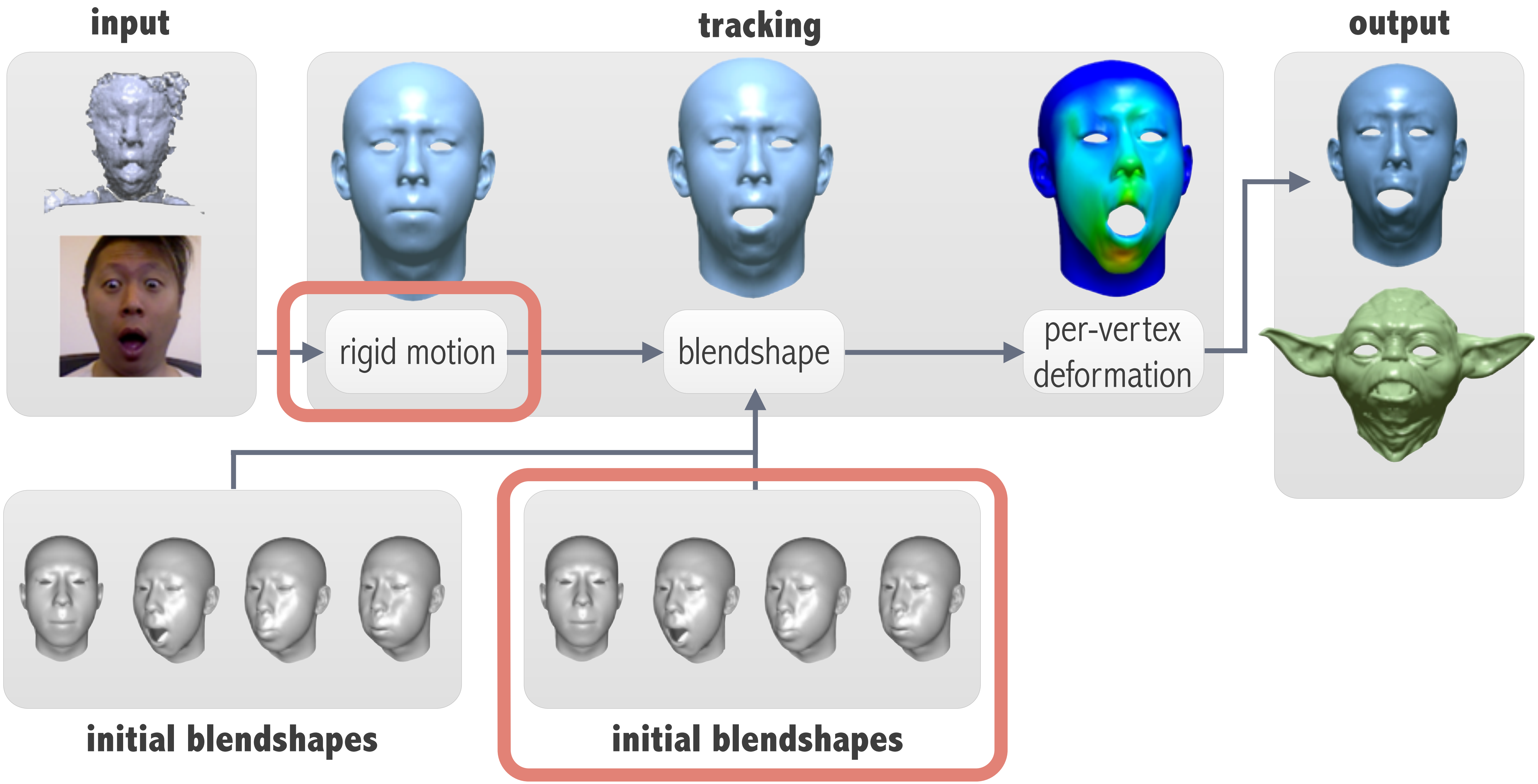
Pipeline Overview



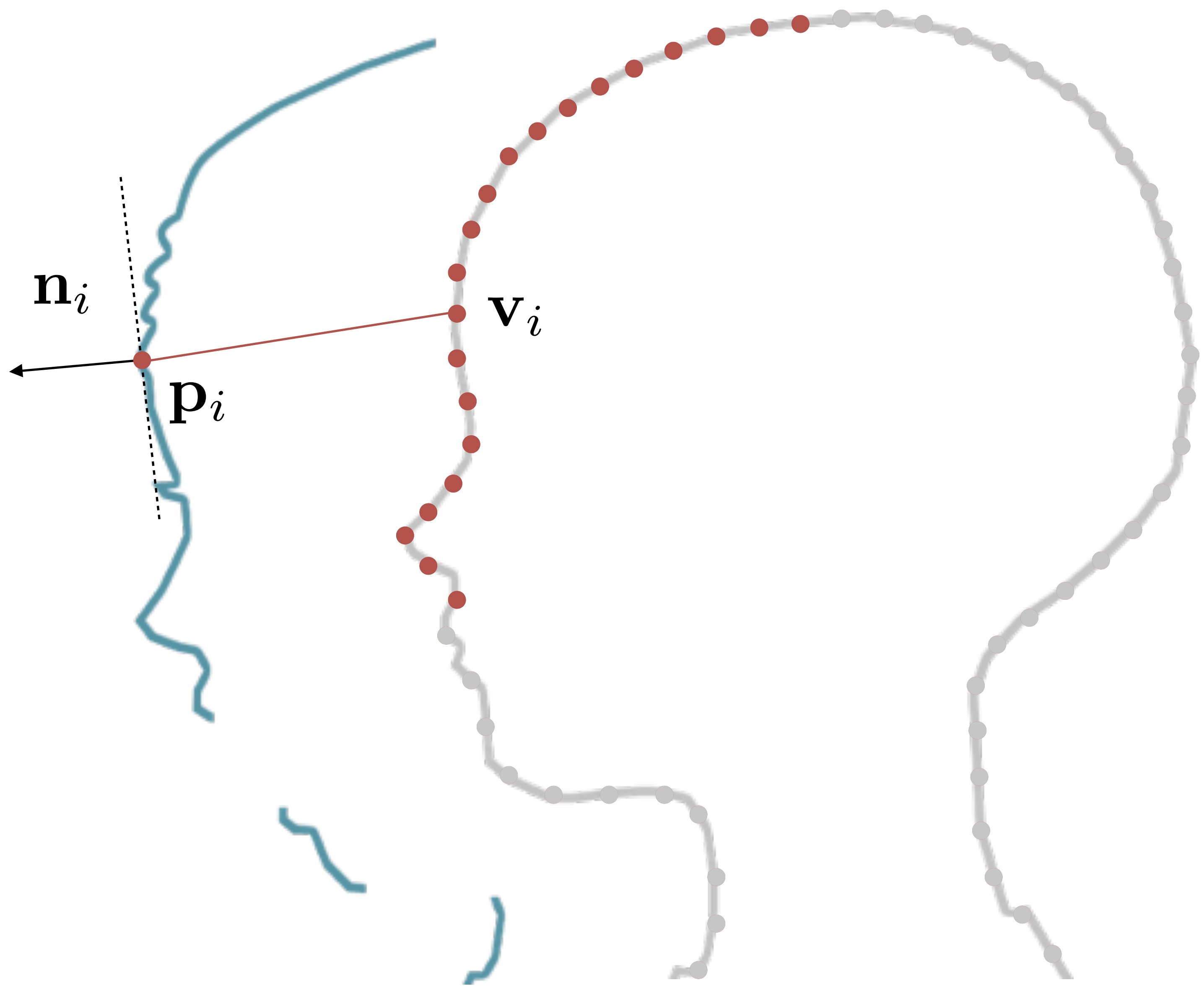
Building Initial Blendshape Model



Pipeline Overview



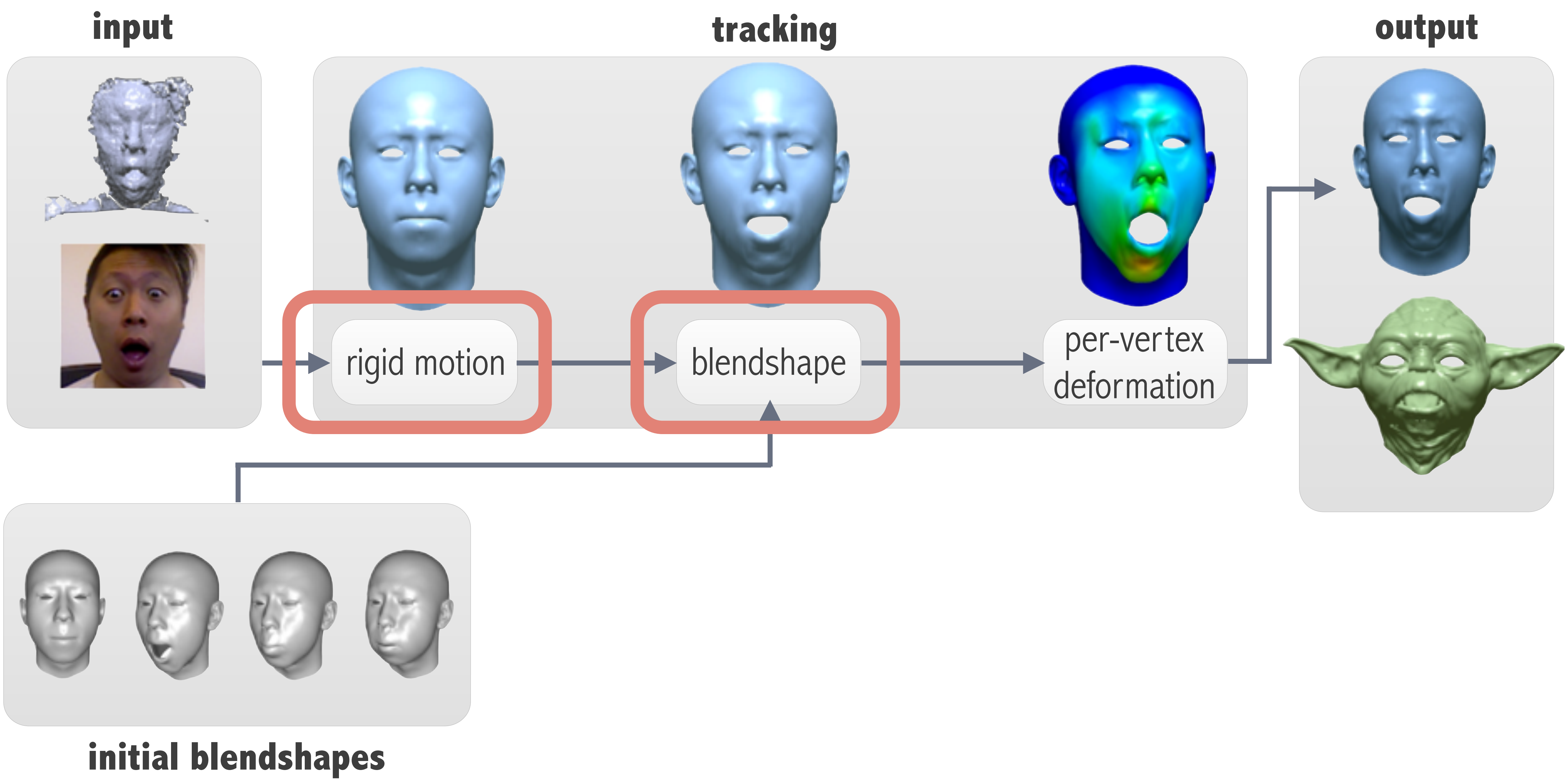
Rigid Motion Tracking



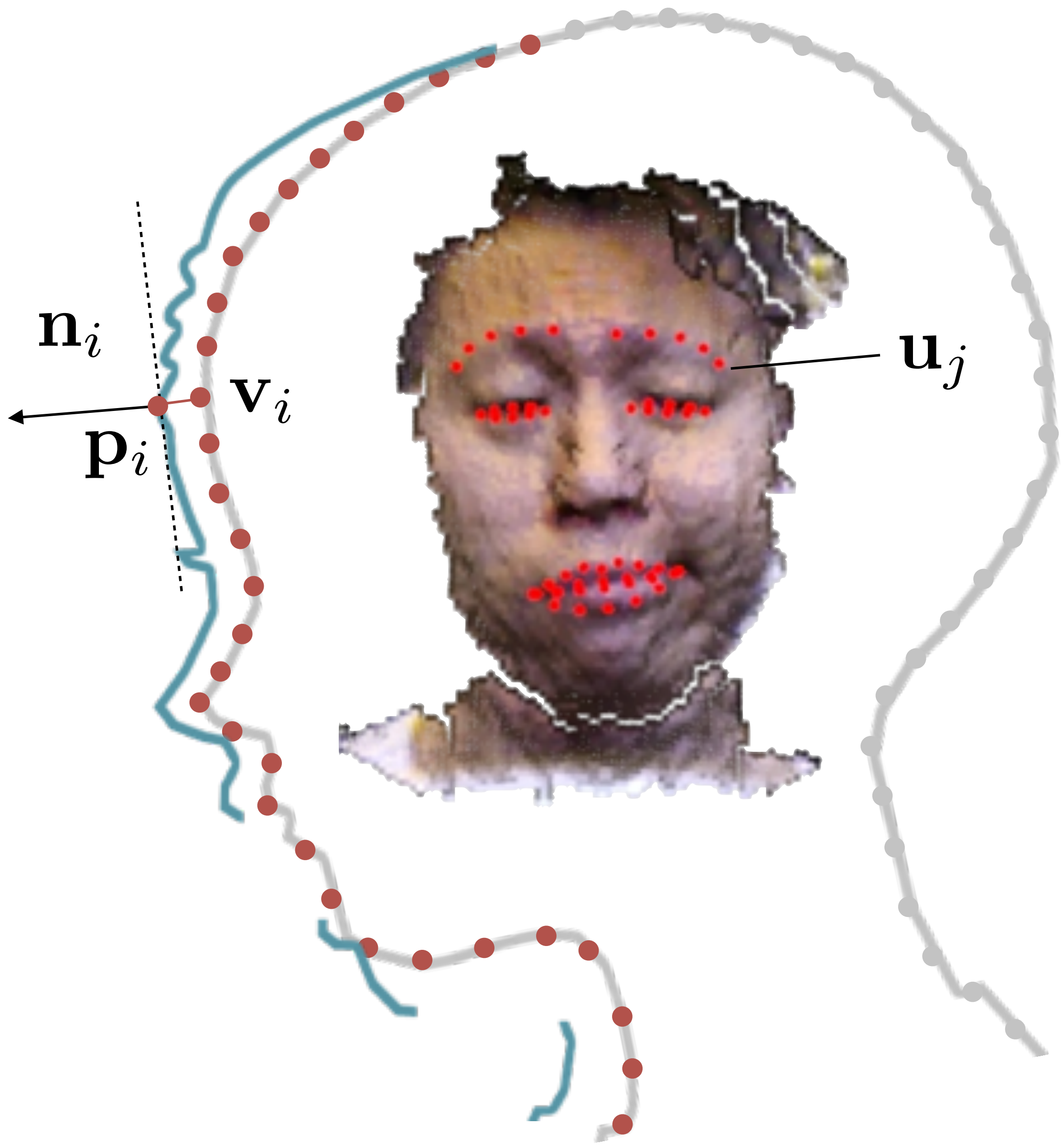
$$c_i^{\mathrm{S}}(\mathbf{R}, \mathbf{t}) = \mathbf{n}_i^{\top}(\mathbf{v}_i(\mathbf{R}, \mathbf{t}) - \mathbf{p}_i)$$

$$E_{\mathrm{rigid}} = \min_{\mathbf{R},\mathbf{t}} \sum_i c_i^{\mathrm{S}}(\mathbf{R}, \mathbf{t})$$

Rigid Motion Tracking



Blendshape Tracking



$$\mathbf{v}_i(\mathbf{x}) = \mathbf{v}_i^{(0)} + \sum_l \mathbf{v}_i^{(l)} x_l$$

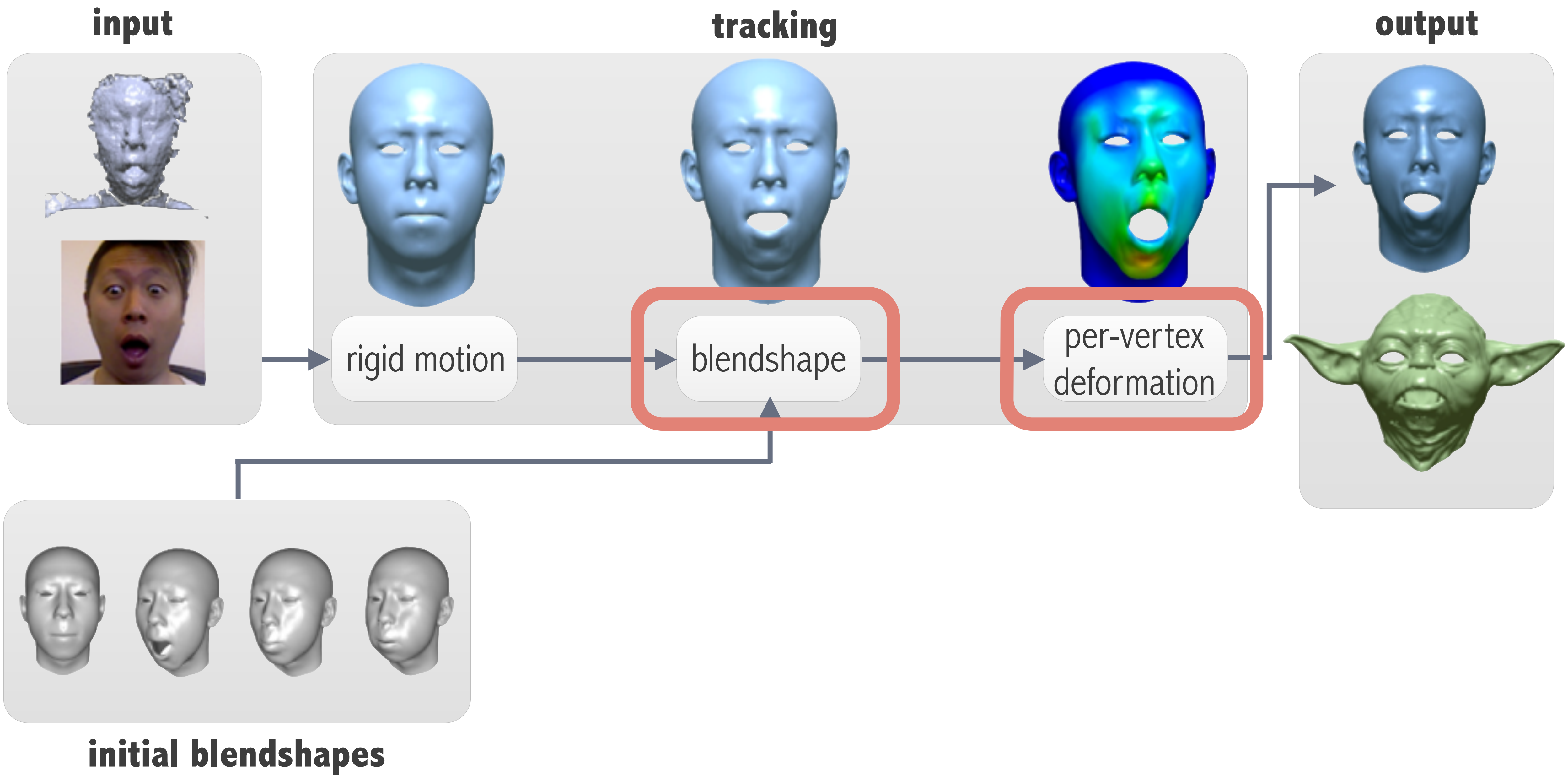
$$c_i^{\mathrm{S}}(\mathbf{x}) = \mathbf{n}_i^\top (\mathbf{v}_i(\mathbf{x}) - \mathbf{p}_i)$$

$$\mathbf{c}_j^{\mathrm{F}}(\mathbf{x}) = \mathbf{H}_j(\mathbf{u}_j)\ \mathbf{P}\ \mathbf{v}_j(\mathbf{x})$$

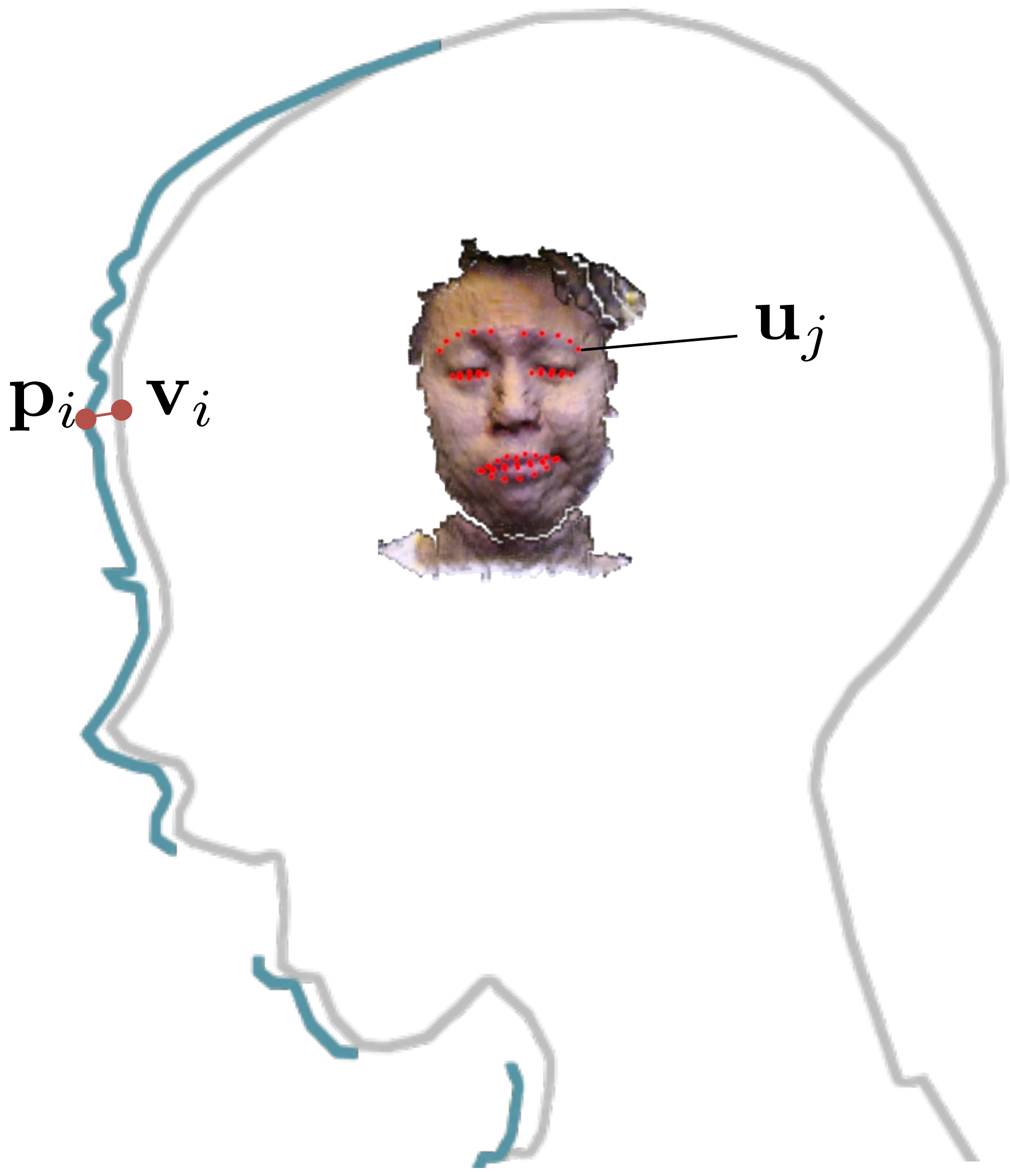
$$E_{\mathrm{bs}} = \min_{\mathbf{x}} \sum_i (c_i^{\mathrm{S}}(\mathbf{x}))^2 + w \sum_j \|\mathbf{c}_j^{\mathrm{F}}(\mathbf{x})\|^2$$

$$x_l \in [0, 1]$$

Pipeline Overview



Per-Vertex Deformation



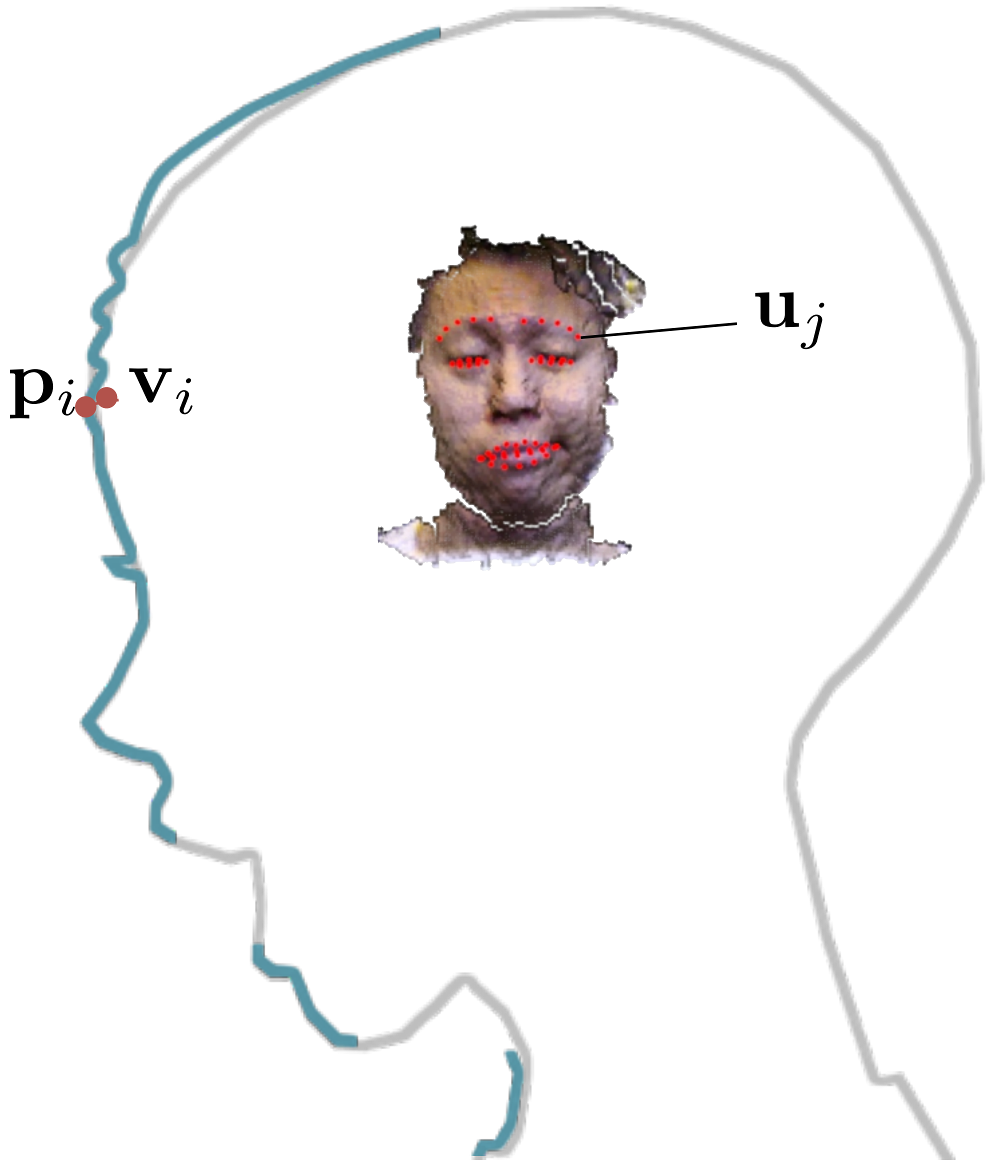
$$\mathbf{c}_i^{\mathrm{P}}(\Delta\mathbf{v}_i) = (\mathbf{p}_i - \mathbf{v}_i) - \Delta\mathbf{v}_i$$

$$\mathbf{c}_j^{\mathrm{W}}(\Delta\mathbf{v}_j) = \mathbf{H}_j(\mathbf{u}_j)\,\mathbf{P}\,\Delta\mathbf{v}_j$$

$$\mathbf{c}^{\mathrm{L}}(\Delta\mathbf{v}) = \mathbf{L}(\mathbf{b}_0)\,\Delta\mathbf{v}$$

$$\mathbf{G}\begin{bmatrix}\mathbf{I}\\ \mathbf{Q}\\ \mathbf{L}\end{bmatrix}\Delta\mathbf{v} = \mathbf{a}$$

Fast Laplacian Deformation



$$\mathbf{G}\begin{bmatrix}\mathbf{I}\\ \mathbf{Q}\\ \mathbf{L}\end{bmatrix}\Delta\mathbf{v} = \mathbf{a}$$

$$\mathbf{G}\ \ \mathbf{K}\ \ \Delta\mathbf{v} = \mathbf{a}$$

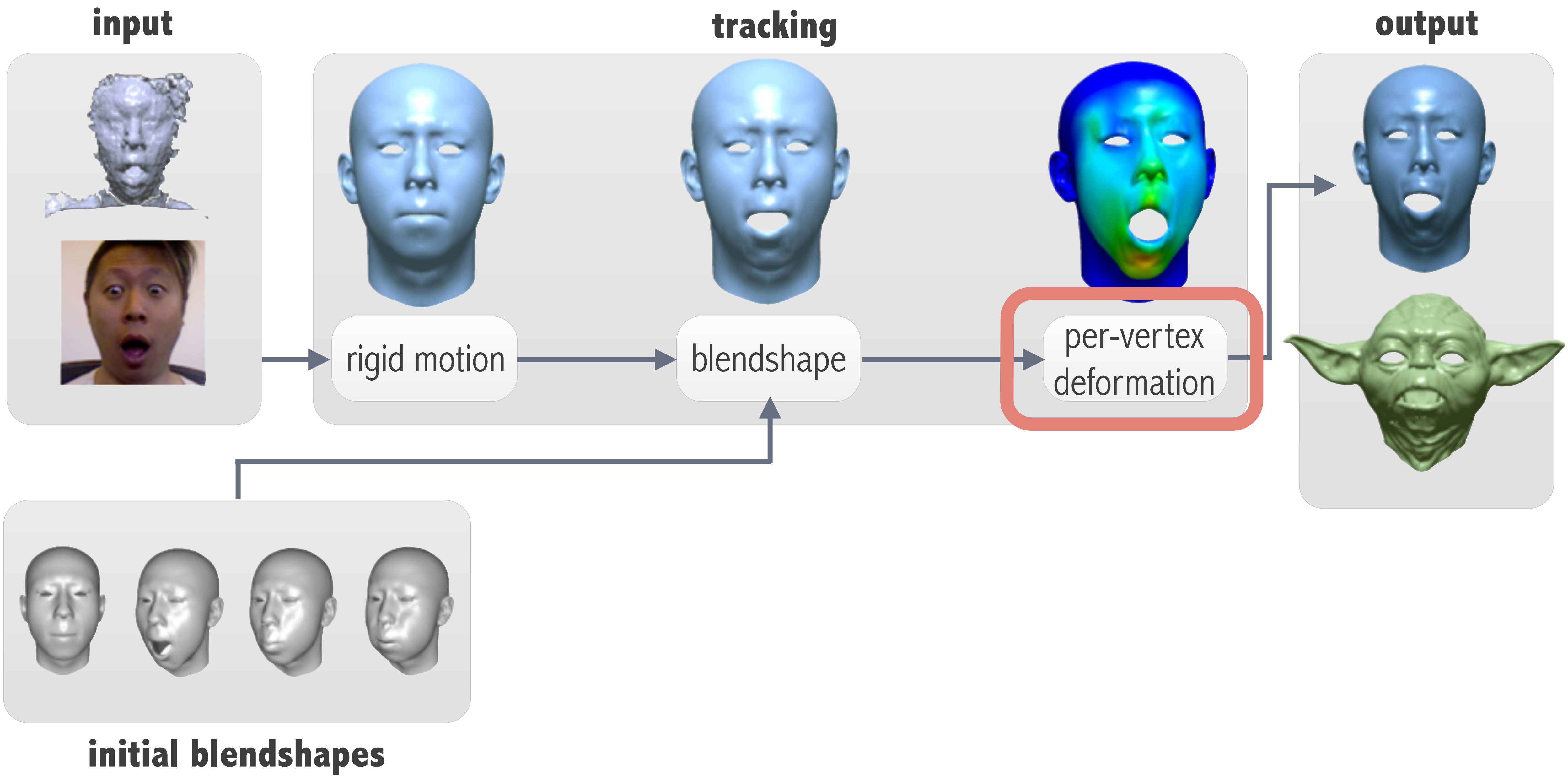
$$\Delta\mathbf{v} = \mathbf{K}^\top[\mathbf{K}\mathbf{K}^\top]^{-1}\mathbf{G}^{-1}\mathbf{a}$$

sparse
constant in time

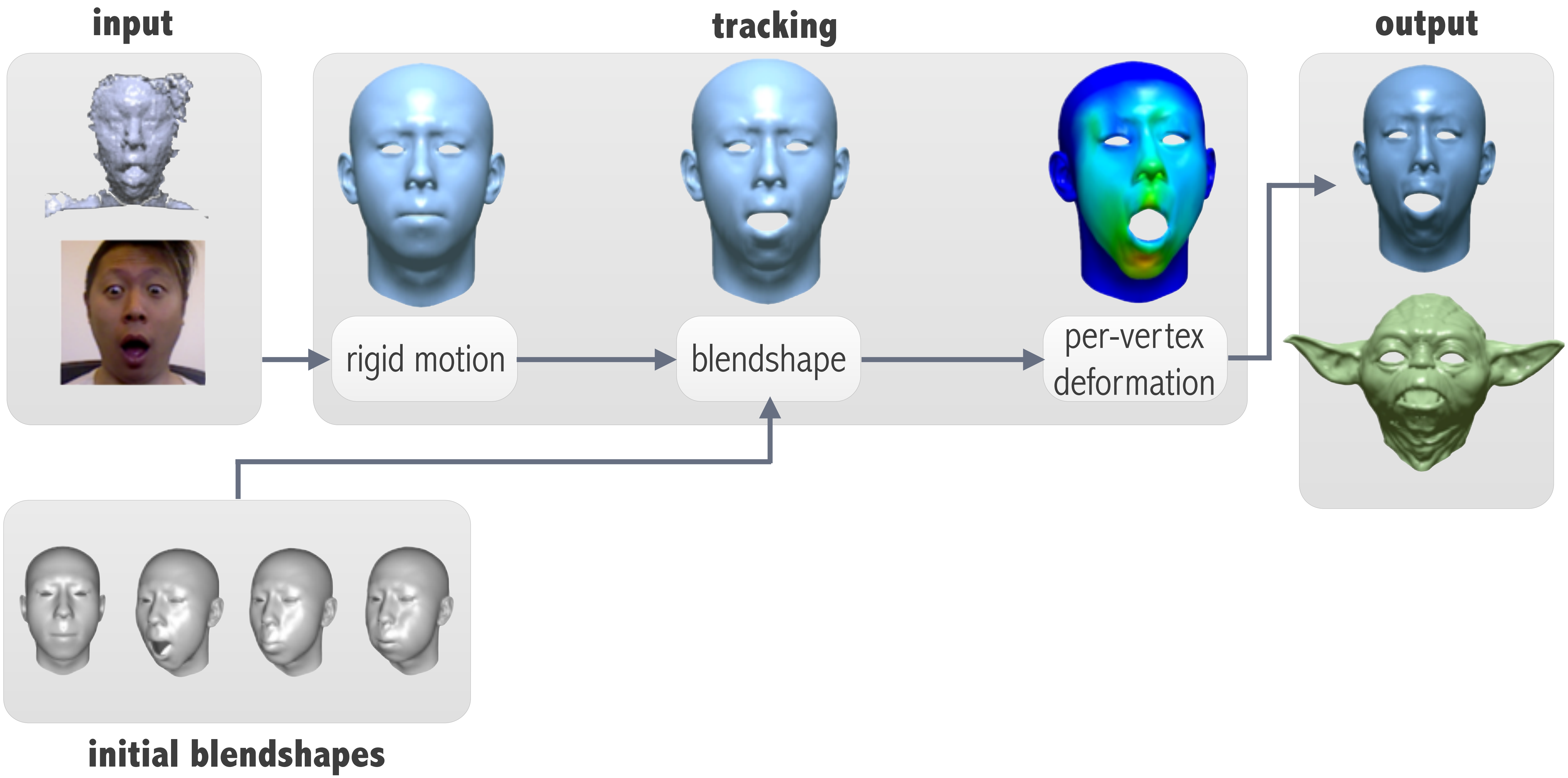
pre-factorized
constant in time

sparse
trivially inversed

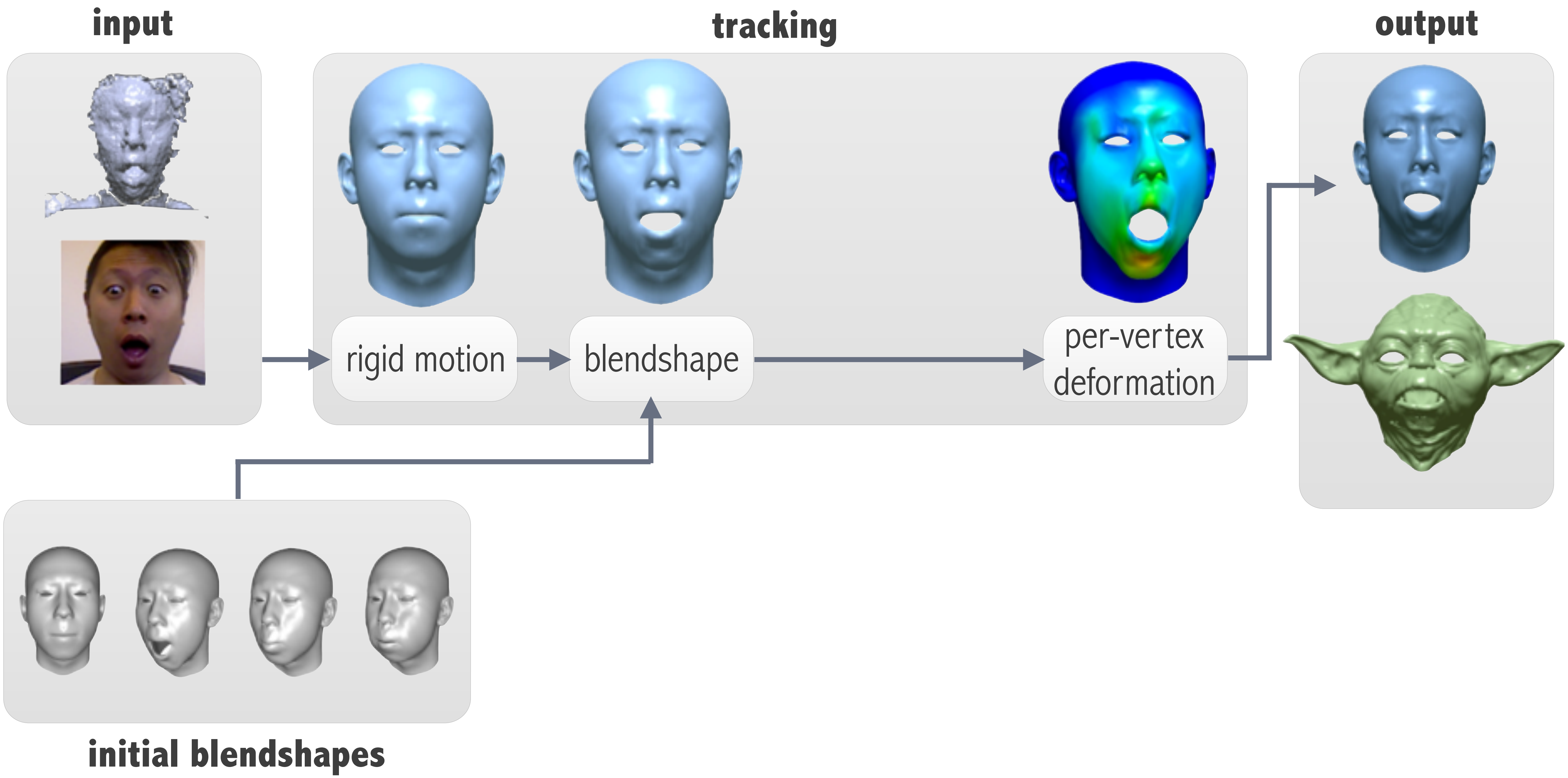
Pipeline Overview



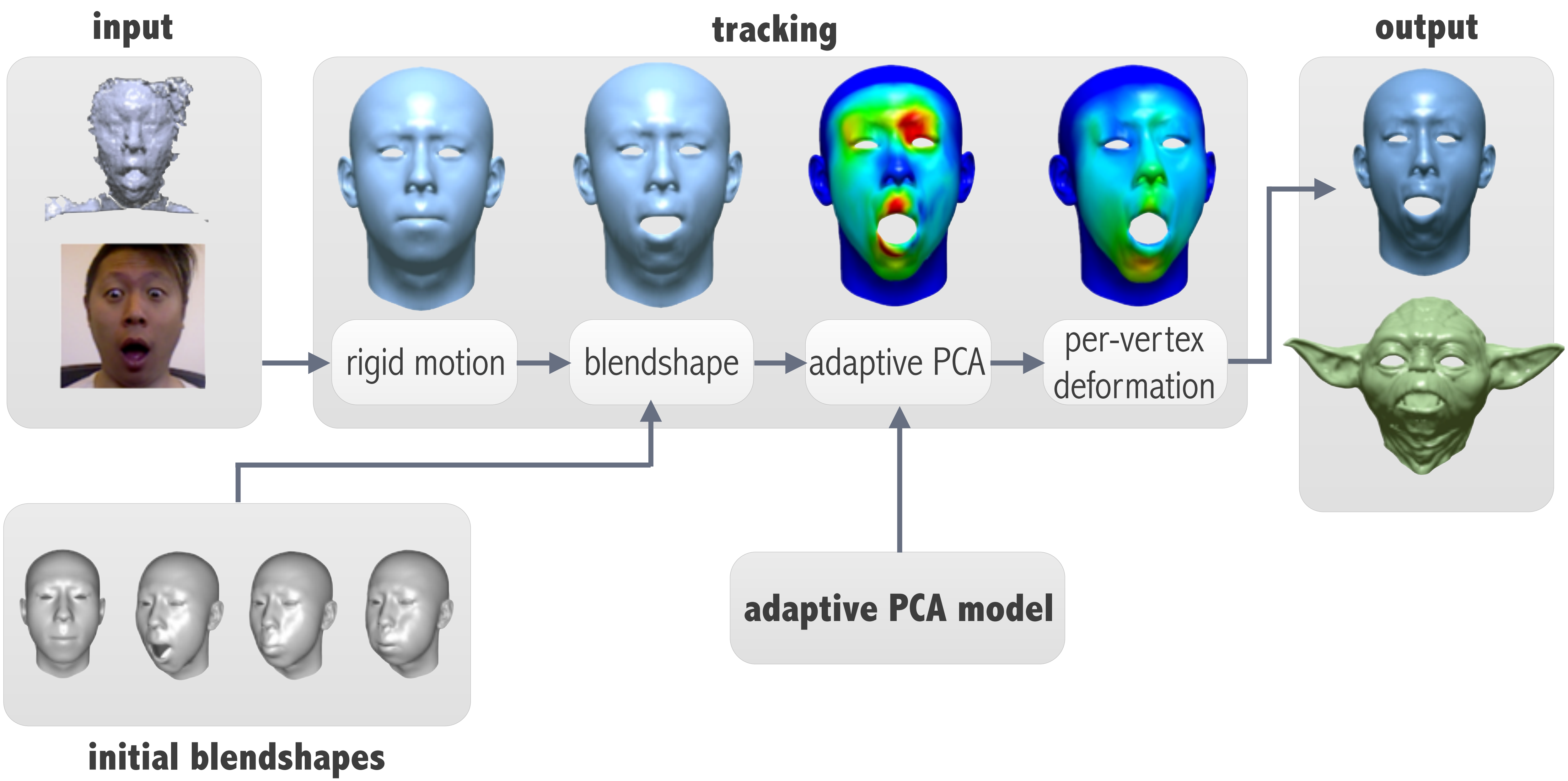
Pipeline Overview



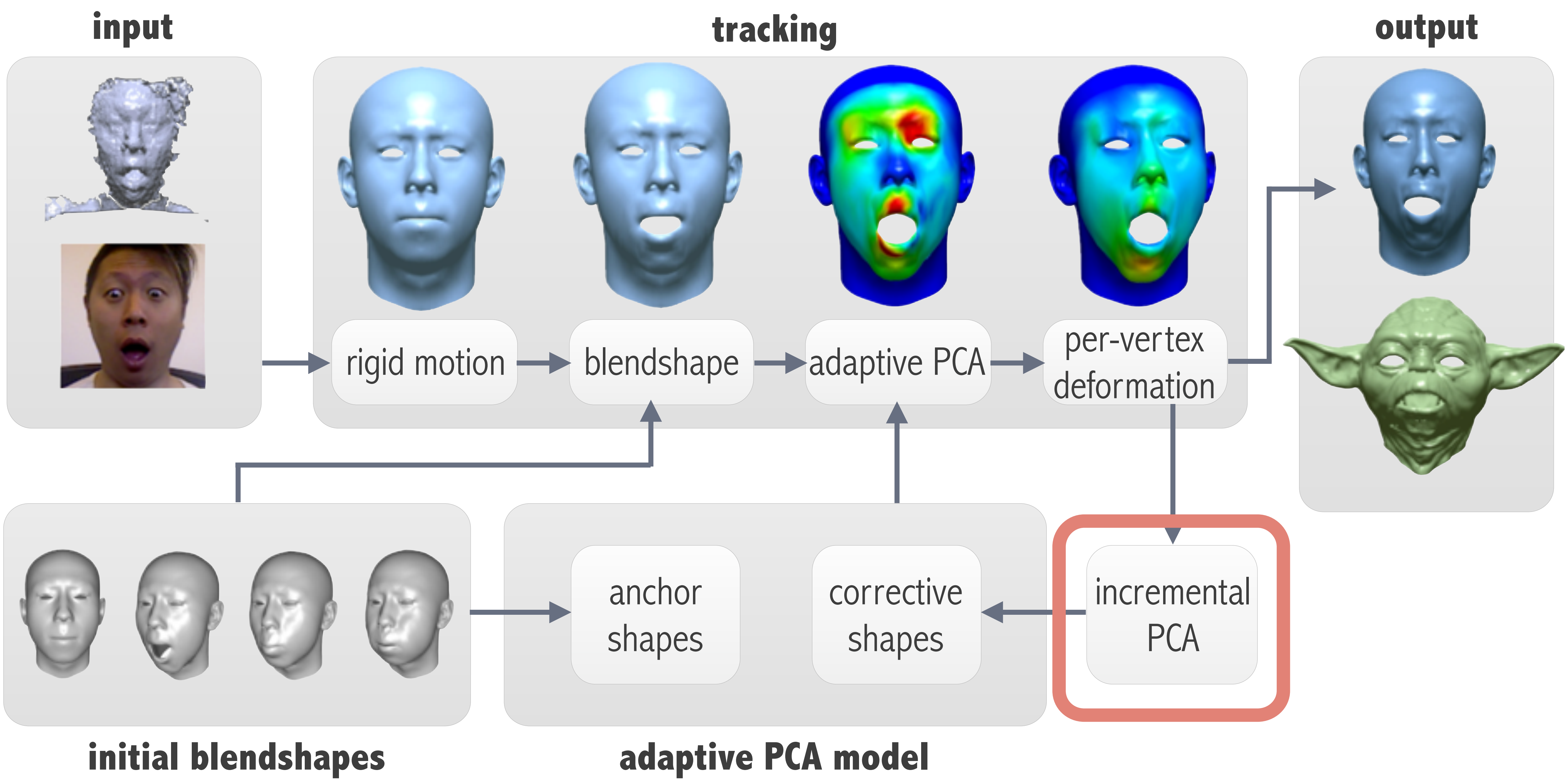
Pipeline Overview



Pipeline Overview

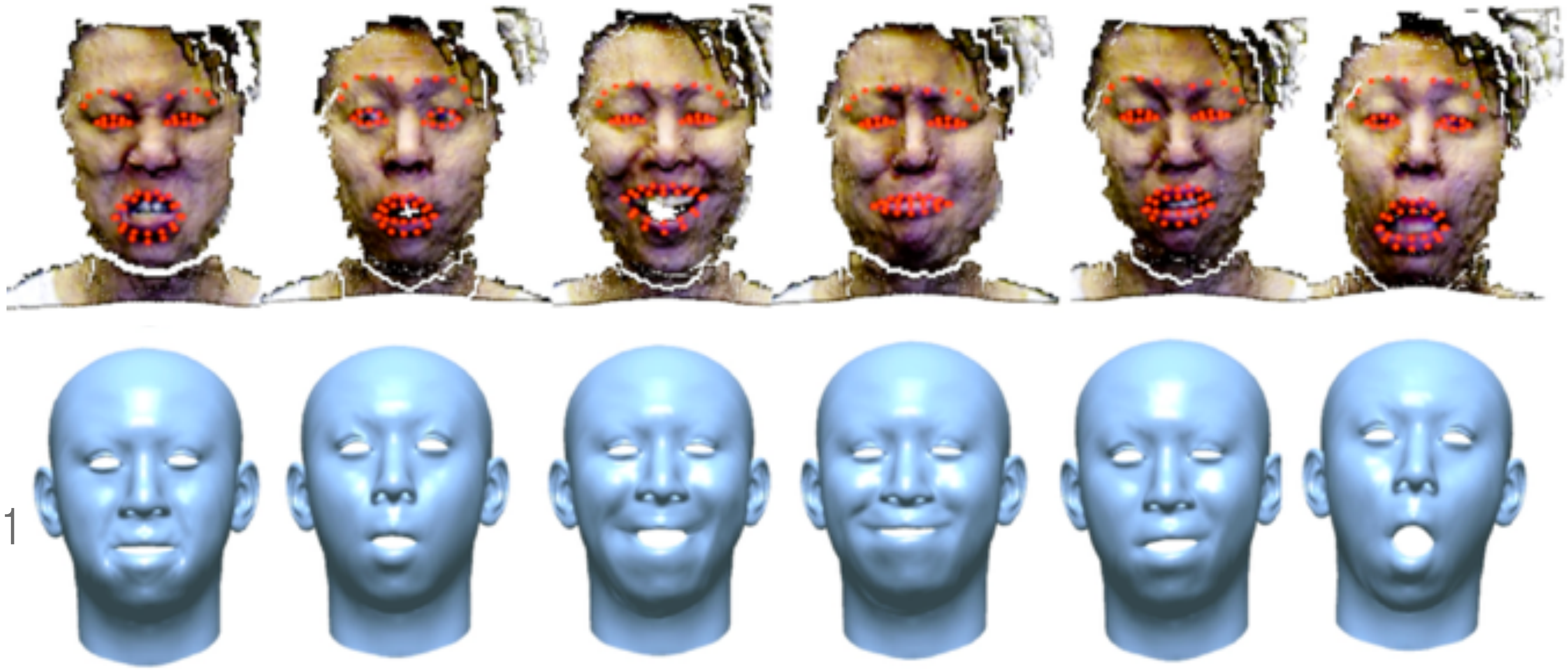


Pipeline Overview

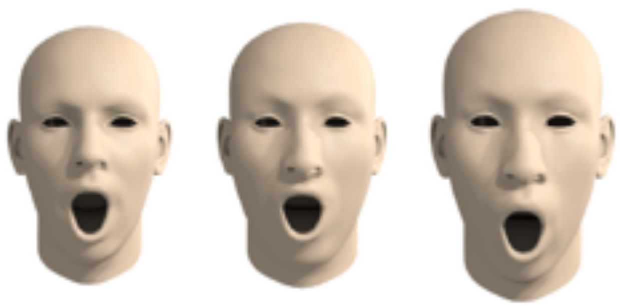
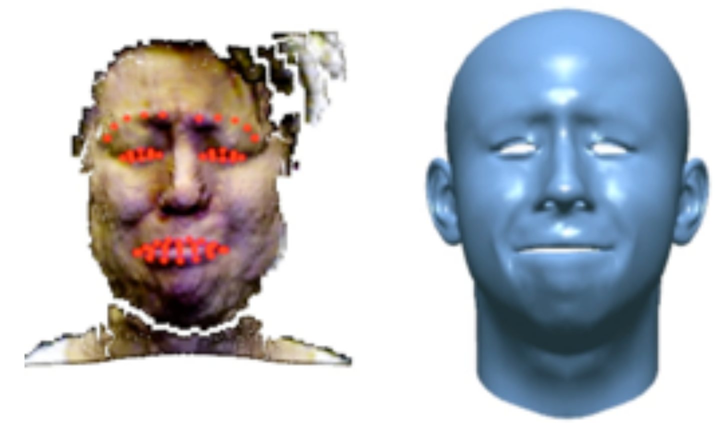


Tracking Comparison

Tracking Basic Emotions



Faces 2014: Discussion

Bouaziz et al. 2013	Cao et al. 2014	Li et al. 2013
		
input 3D/2D	input 2D	input 3D/2D
no calibration	neutral face	neutral face
blendshape	blendshape	per-vertex deformation

Into the Mainstream

FaceX Robustness

input video

face segmentation

FaceX Instant User Switching

input video

face segmentation

FaceX Kids

input video

face segmentation

FaceX Retargeting

Open Problems

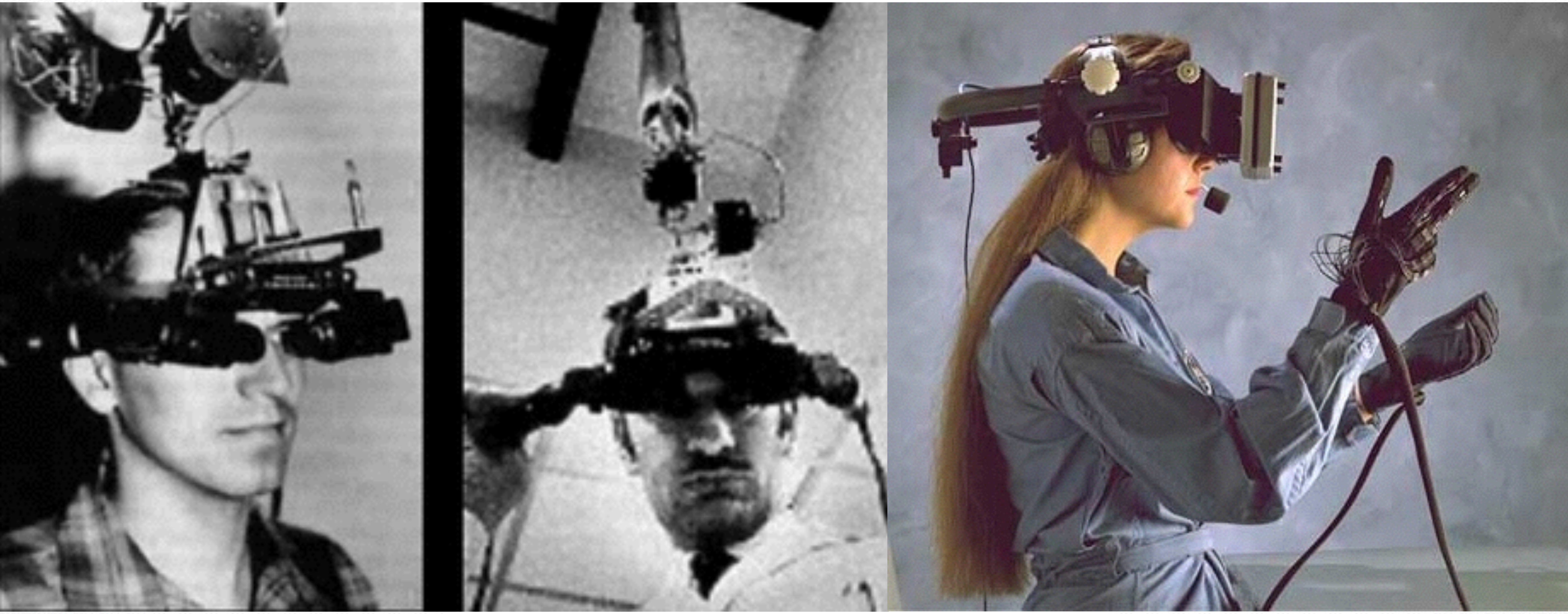
USC/ICT Activision

Open Problems

Melbourne Acting School 2010

Virtual Reality

Once upon a dream



Virtual Reality **Reloaded**

Oculus VR 2012 / Crytek 2014

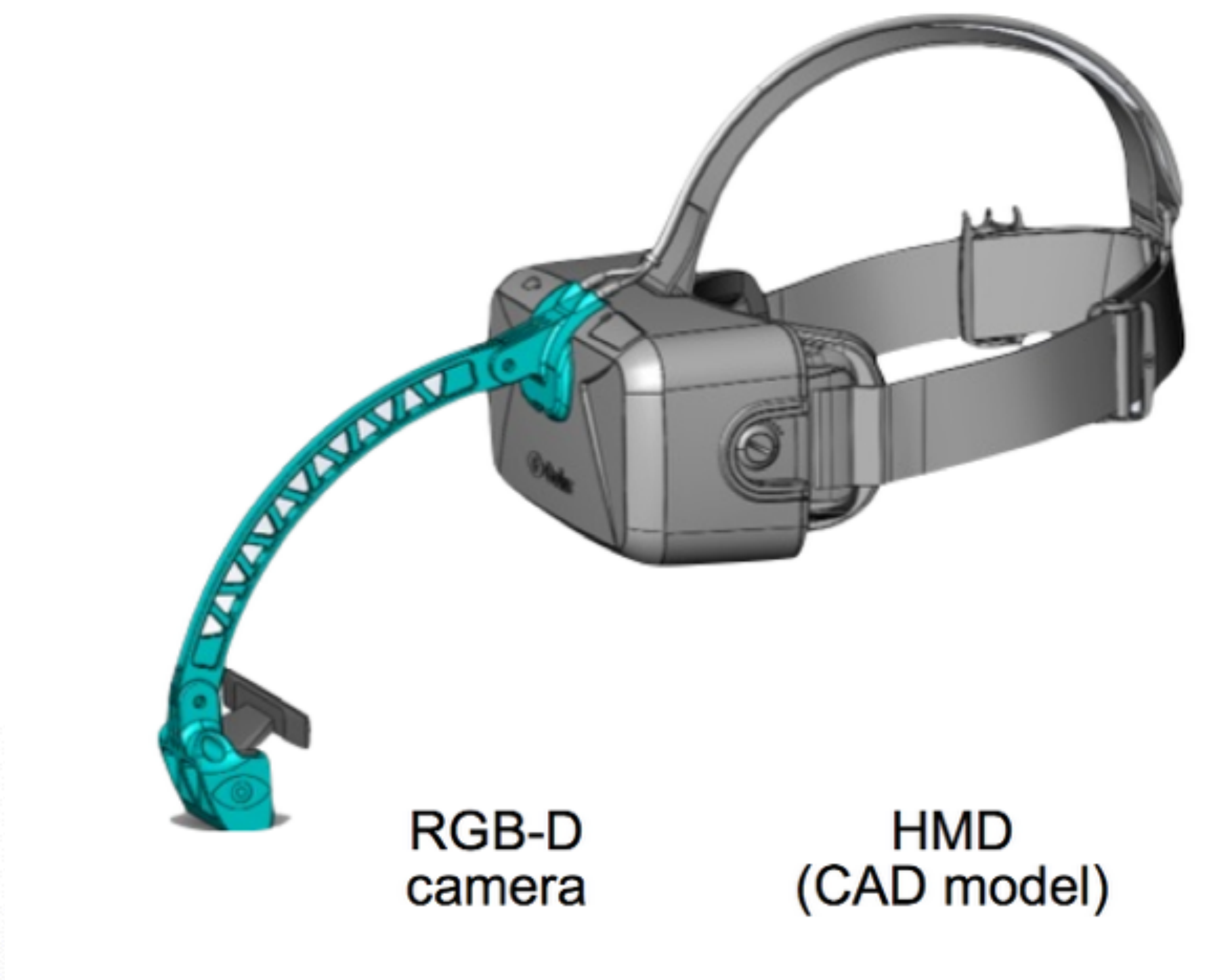
Consuming VR

NextGen Communication Platform

Online Virtual Worlds

Occlusions

Facial Performance Sensing HMD

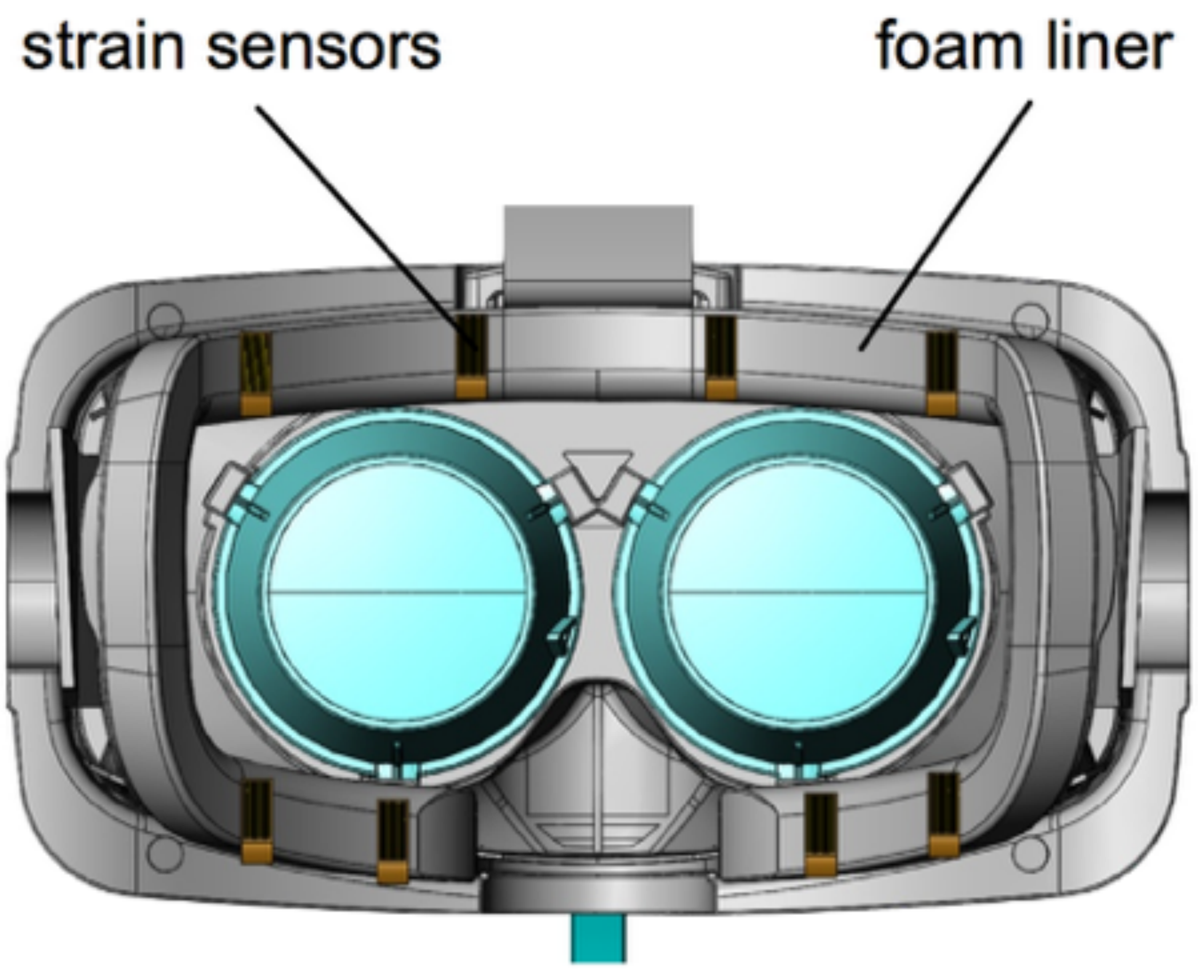


Facial Performance Sensing HMD

OLED display
and cover

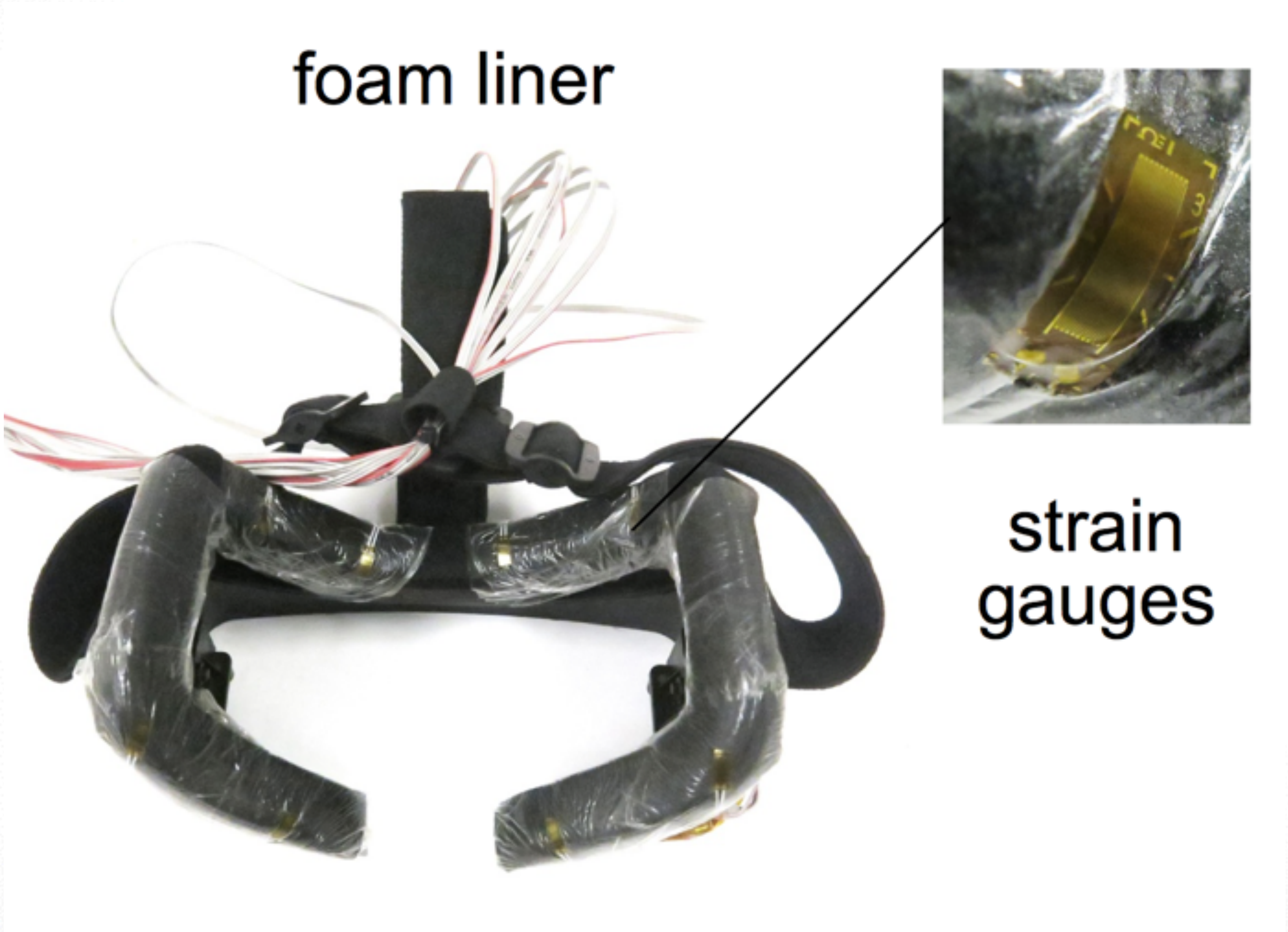
RGB-D camera
(Intel IVCAM)

Facial Performance Sensing HMD

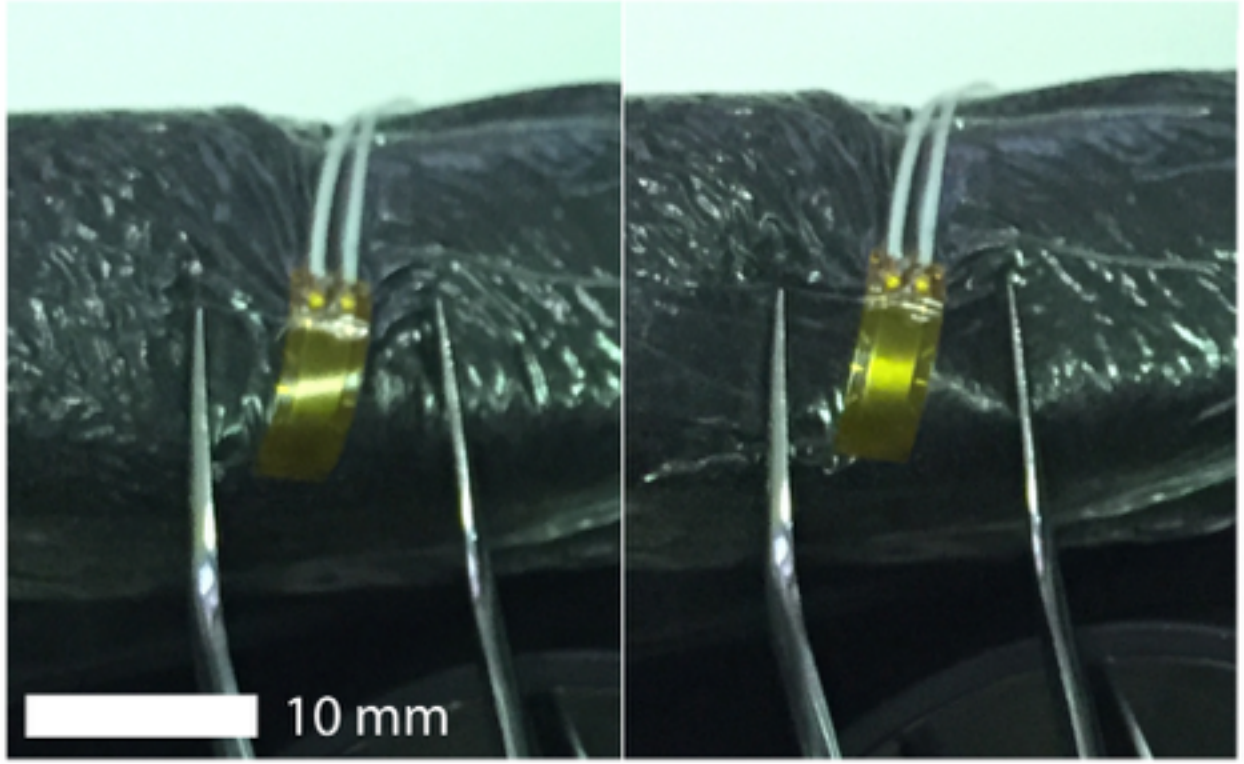


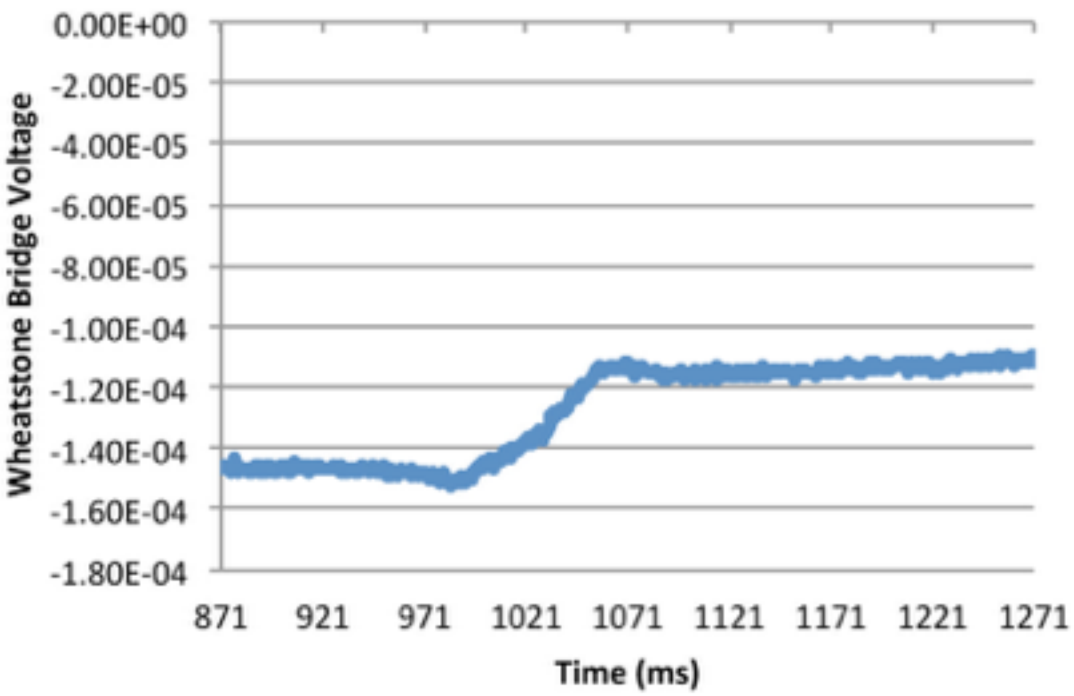
interior
(CAD model)

Facial Performance Sensing HMD

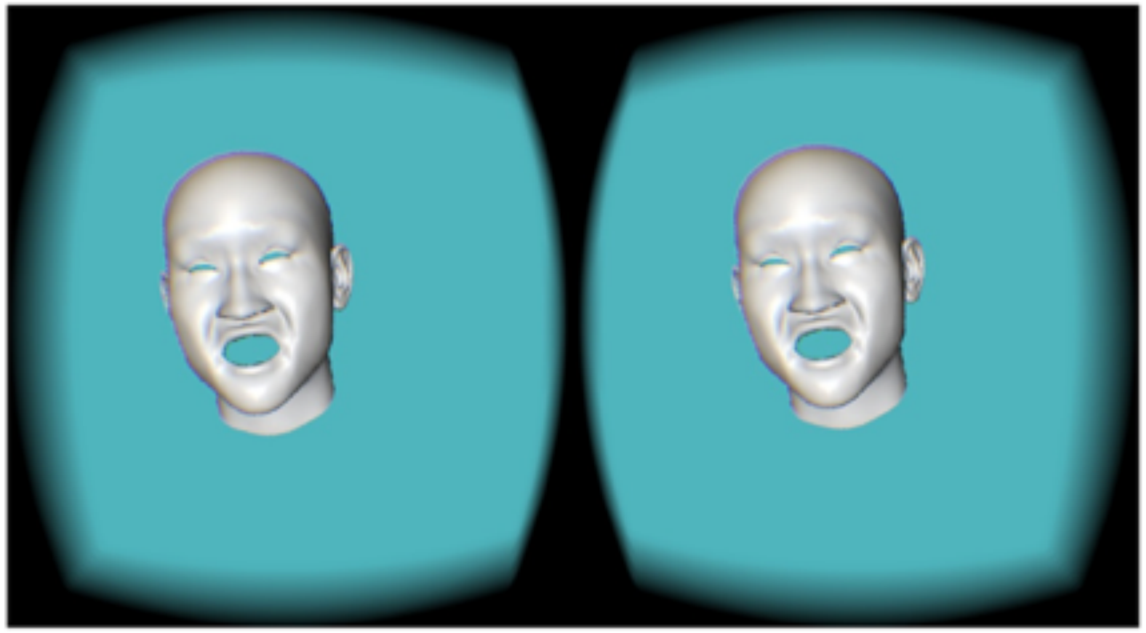


Ultra Thin Flexible Electronic Materials

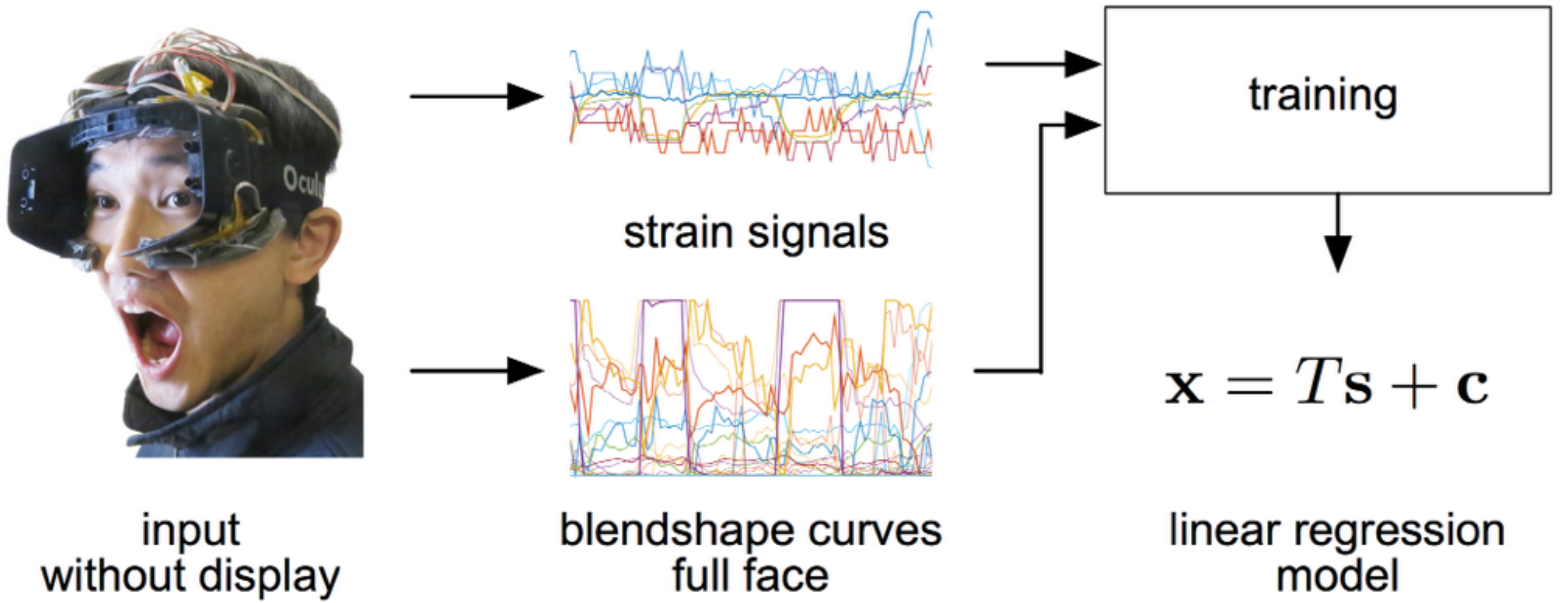




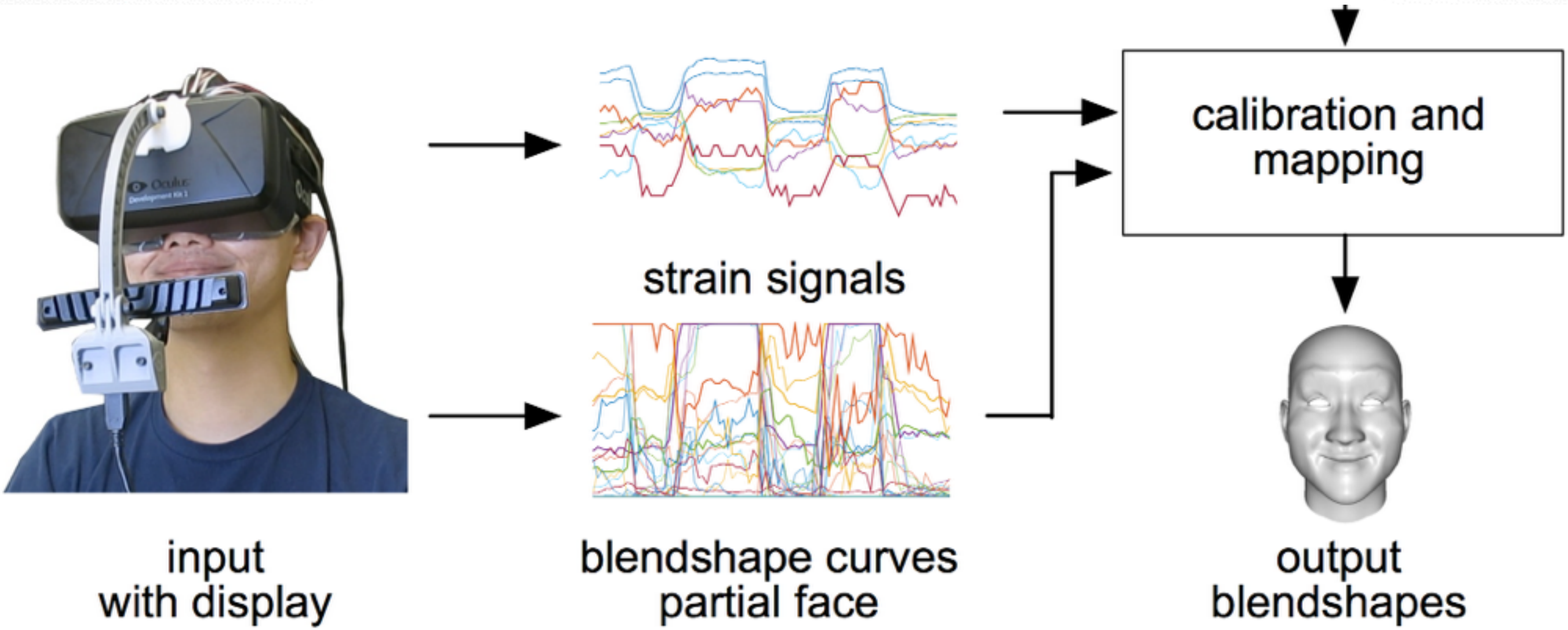
Facial Performance Sensing HMD



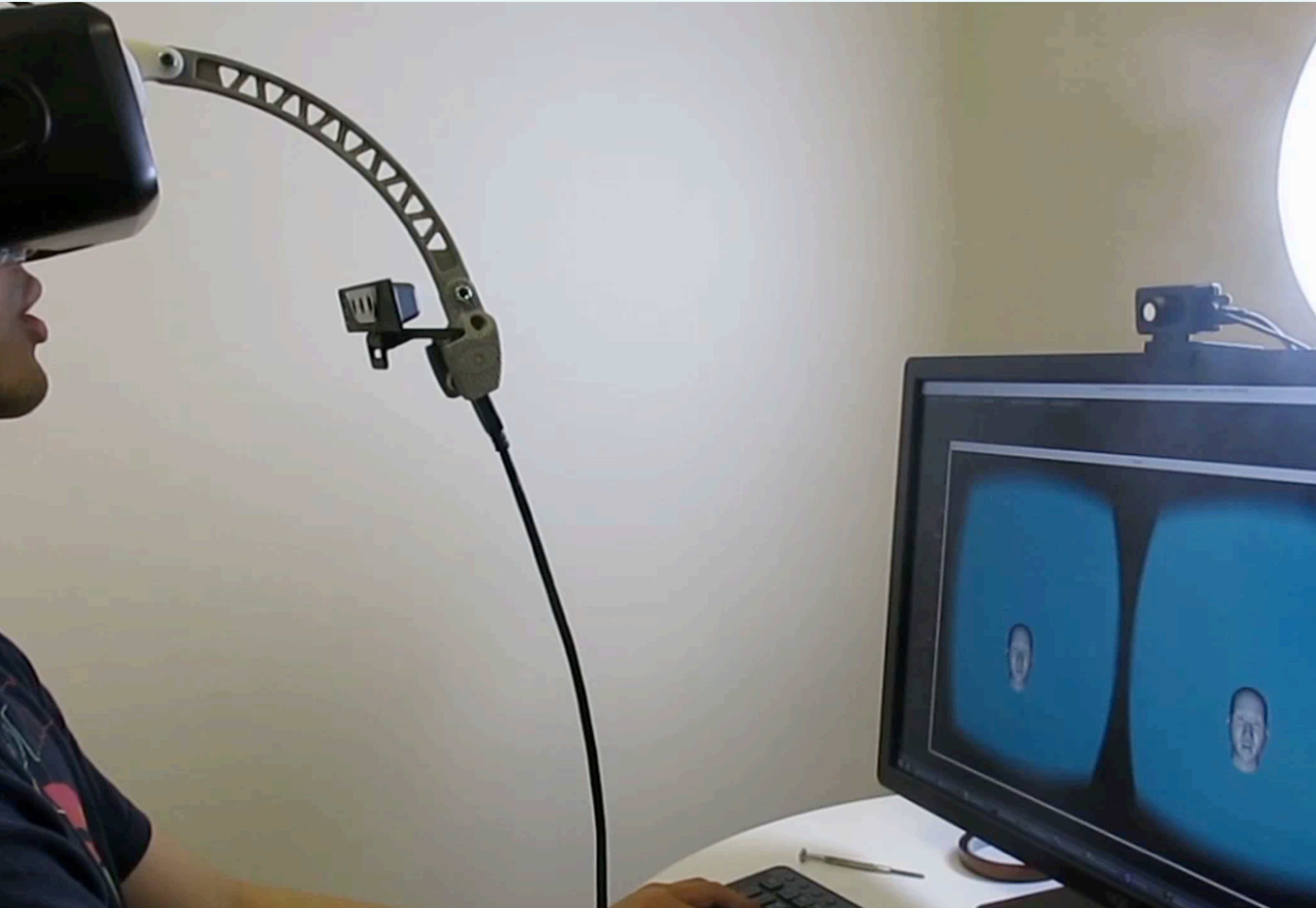
Offline Training



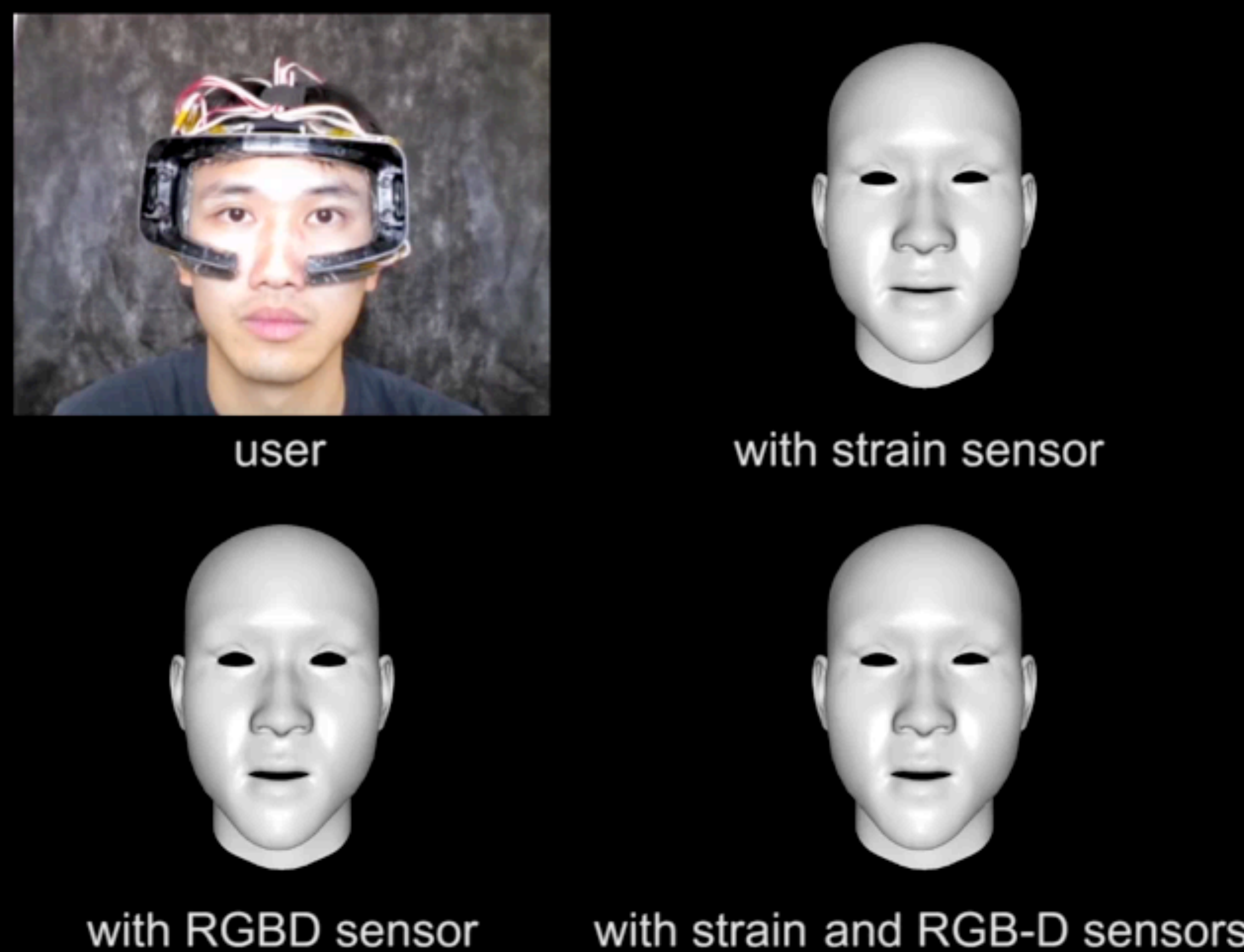
Online Operation



Capture System



Results


user
with strain sensor
with RGBD sensor
with strain and RGB-D sensors

Results

input capture

facial performance capture

http://cs599.hao-li.com

Thanks!