

*Spring 2019*

# CSCI 621: **Digital Geometry Processing**

## 15 **Facial Performance Capture**

Hao Li

<http://cs621.hao-li.com>





# Performance to Facial Animation

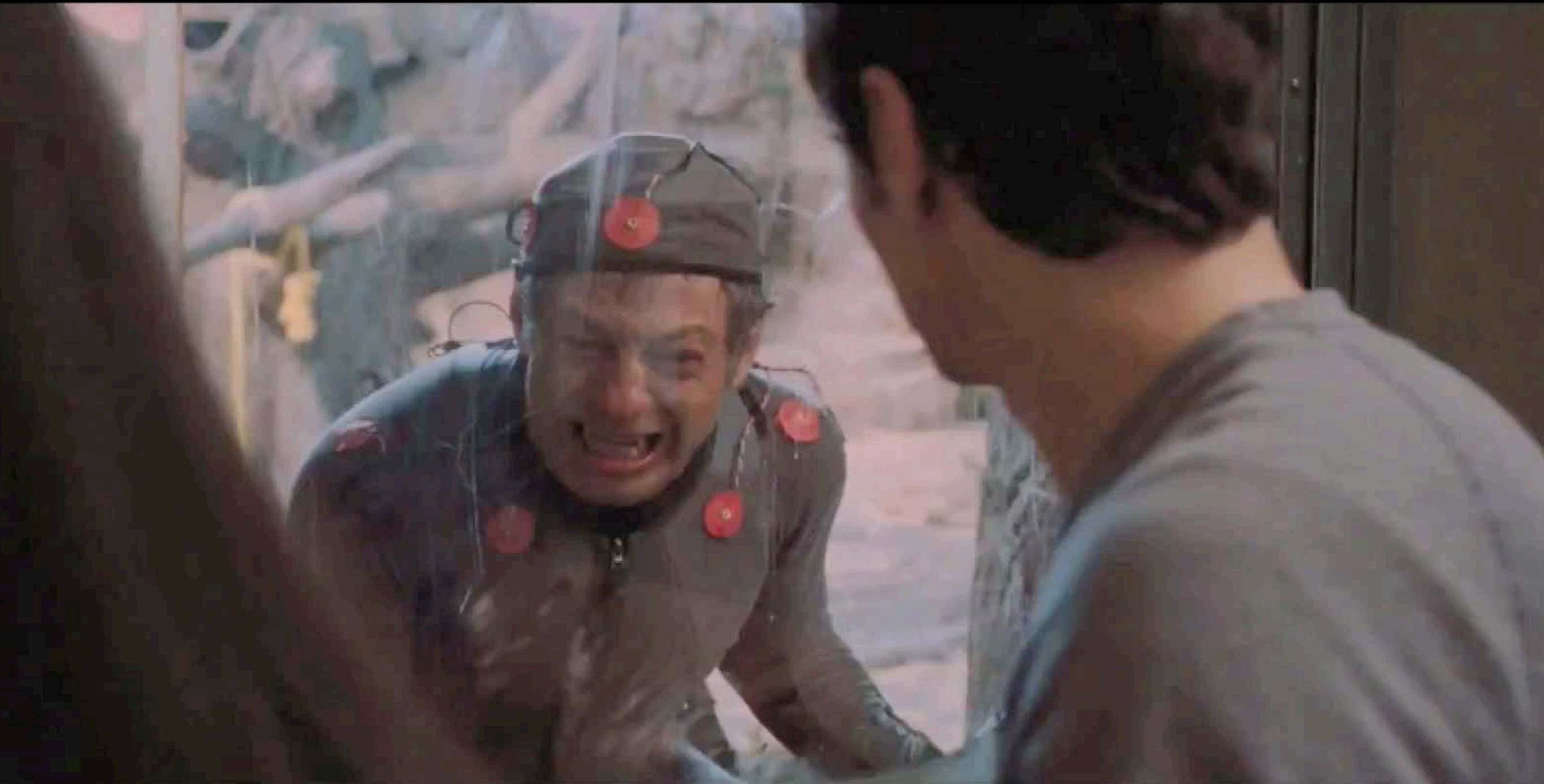




# Motion Capture



# Motion Capture





# Facial Animation in Films







Weta Digital



# Facial Modeling and Scanning

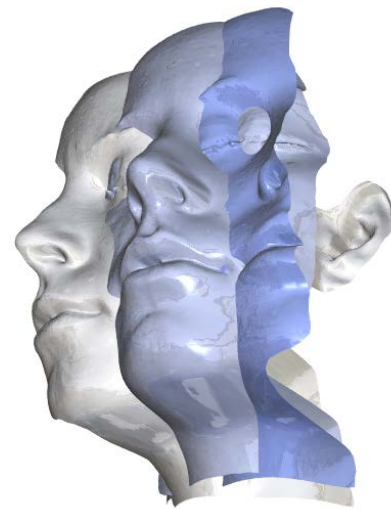




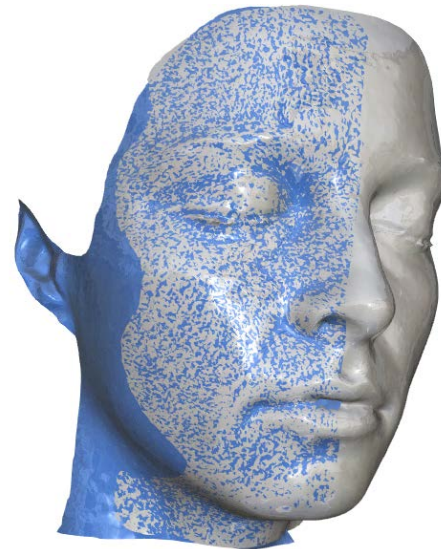
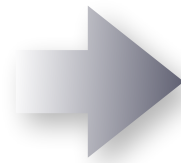
# Facial Modeling and Scanning



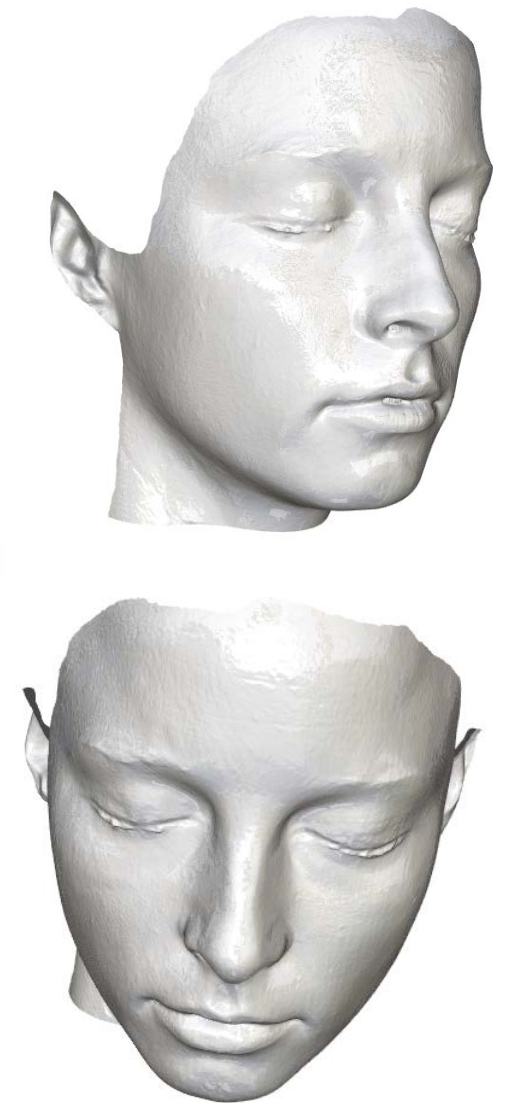
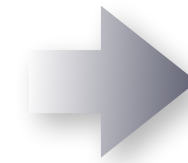
acquisition



initial  
alignment



registration

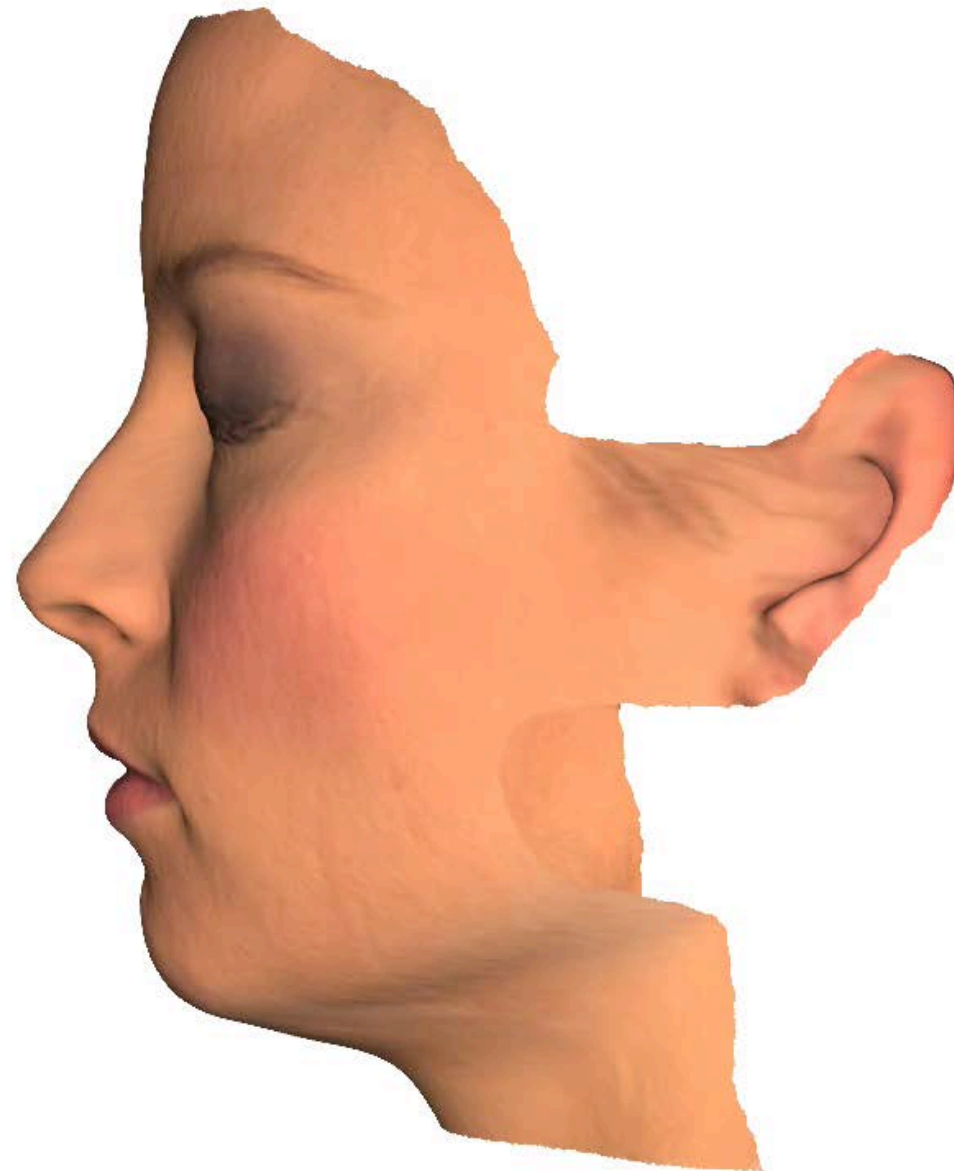


merging

copyright Paramount Pictures



# Facial Modeling and Scanning





# High-End 3D Scanning

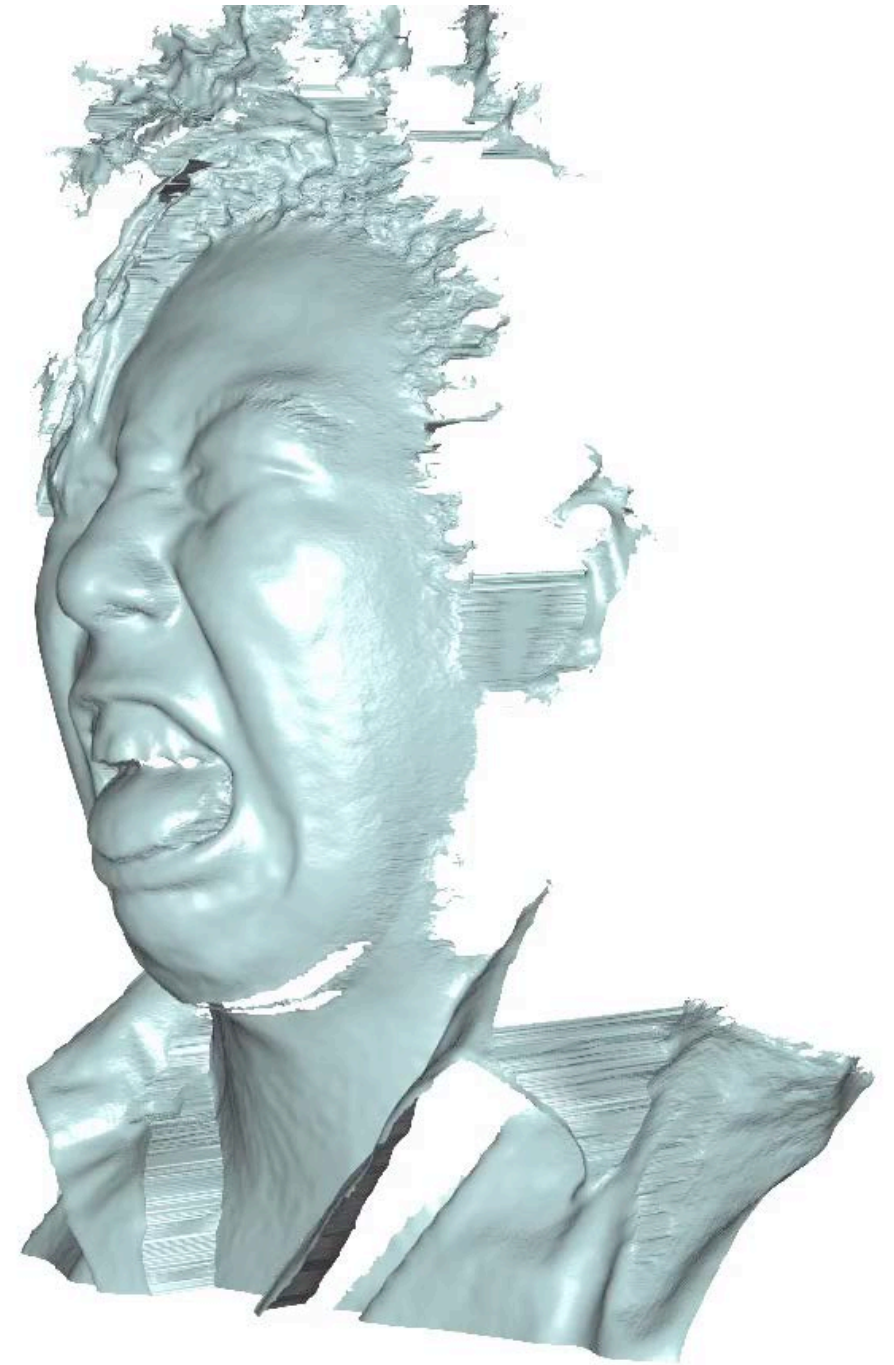




# Low-Cost Passive Scanning (AGI Soft)



stereo pair



3D scan

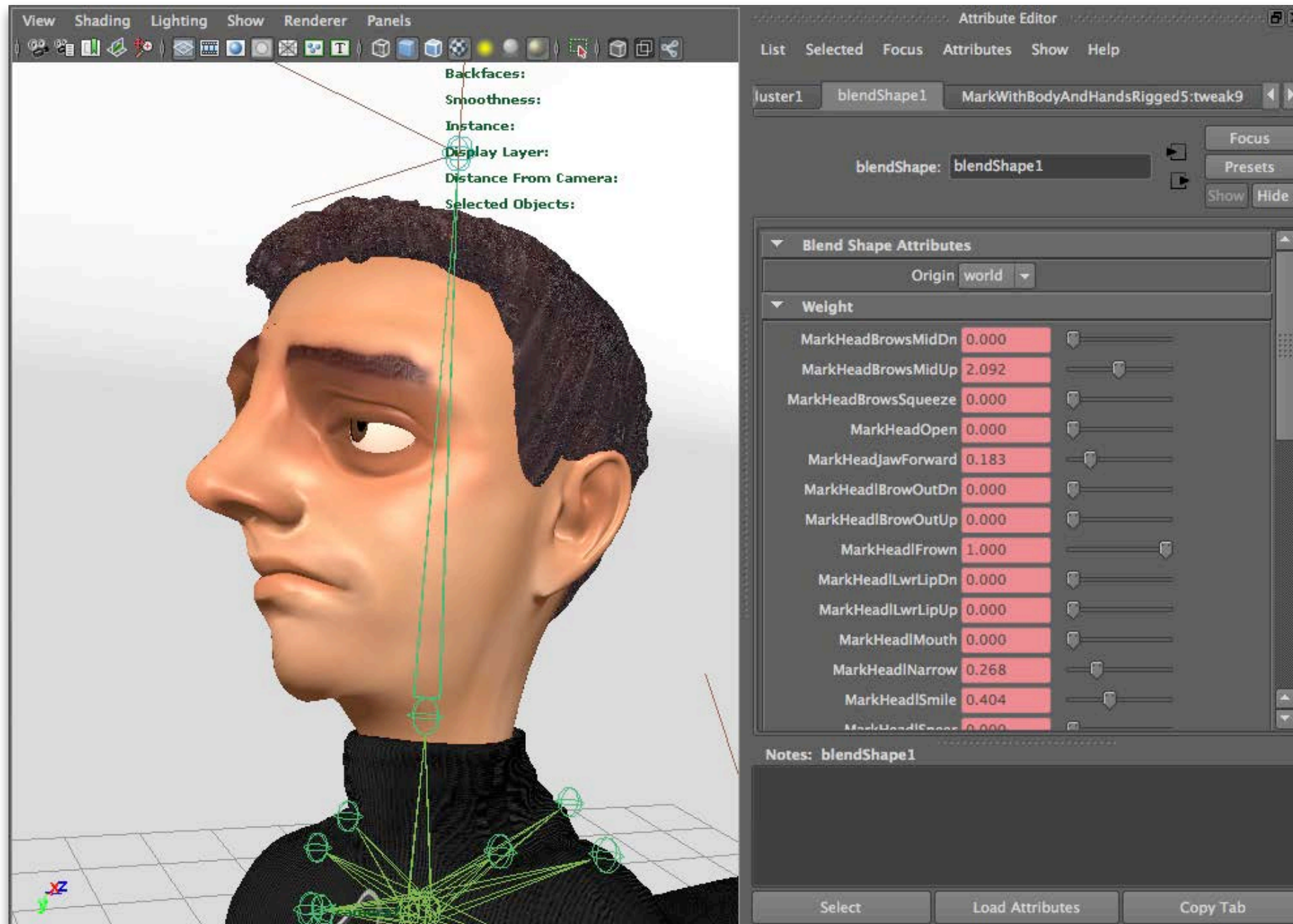


# Low-Cost Active Scanning



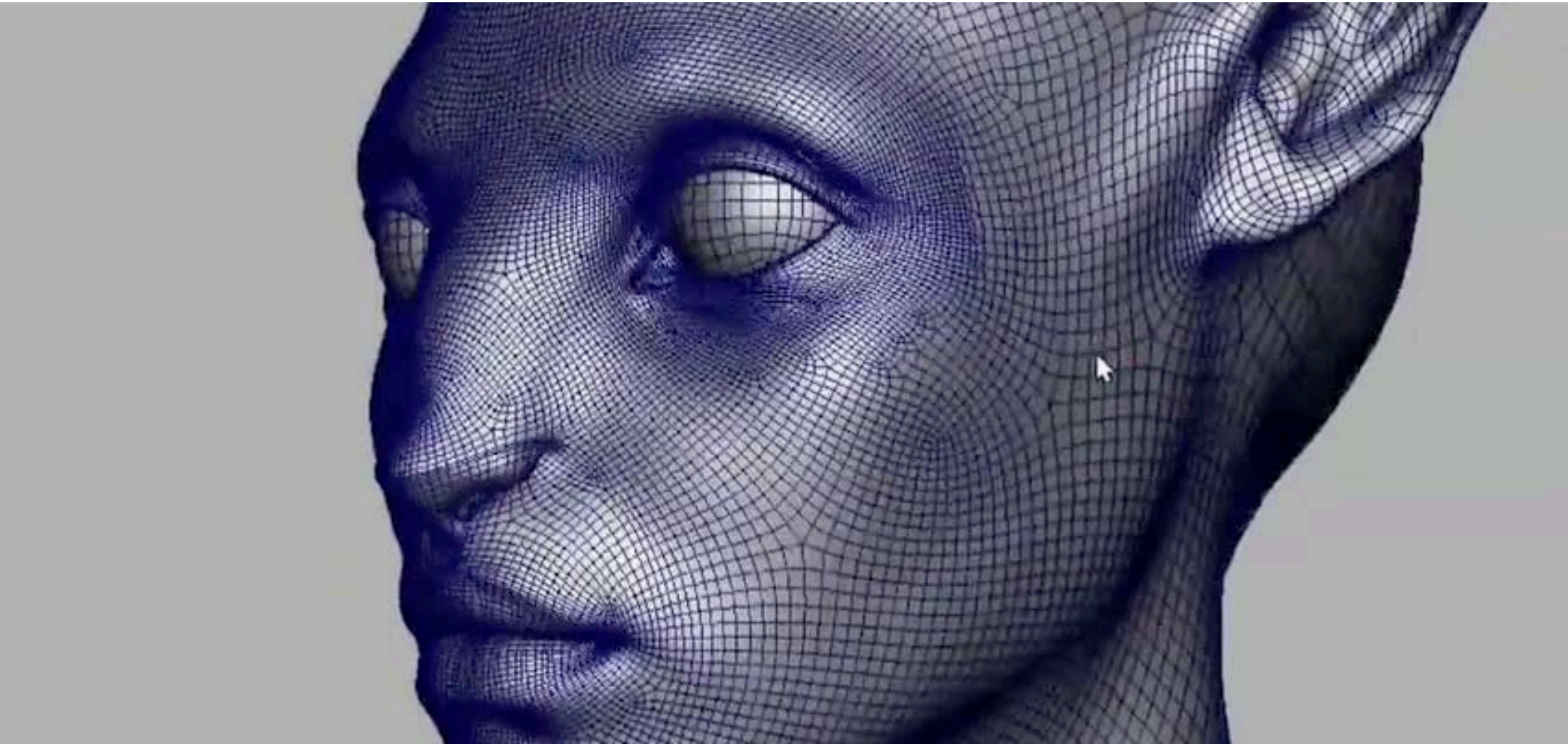
Microsoft Kinect & Kinect Fusion

# Rigging & Animation



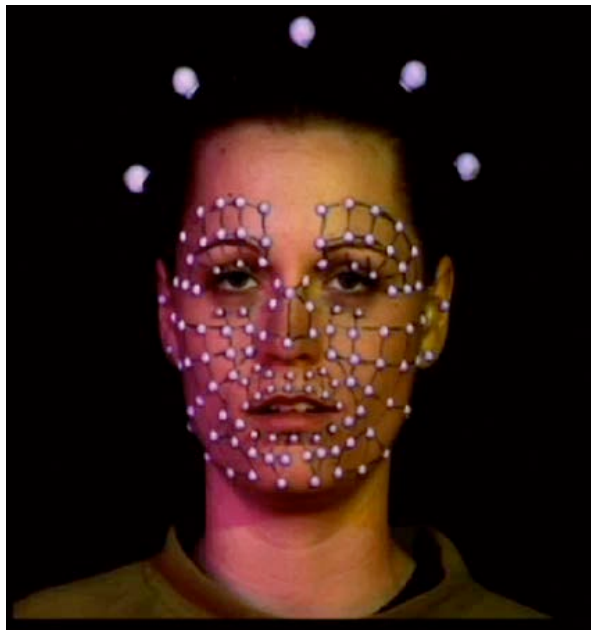


# Blendshapes & Correctives for Realism





# Motion Capture Technologies



Sparse Markers



Dense Markers  
MOVA



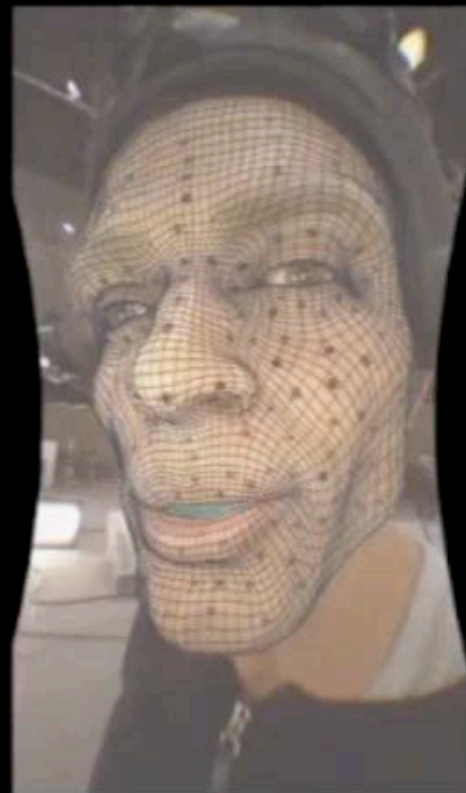
Markerless  
Image Metrics



# Using Markers



input performance



input video  
with markers



tracking



retargeting

# Using Dense Markers

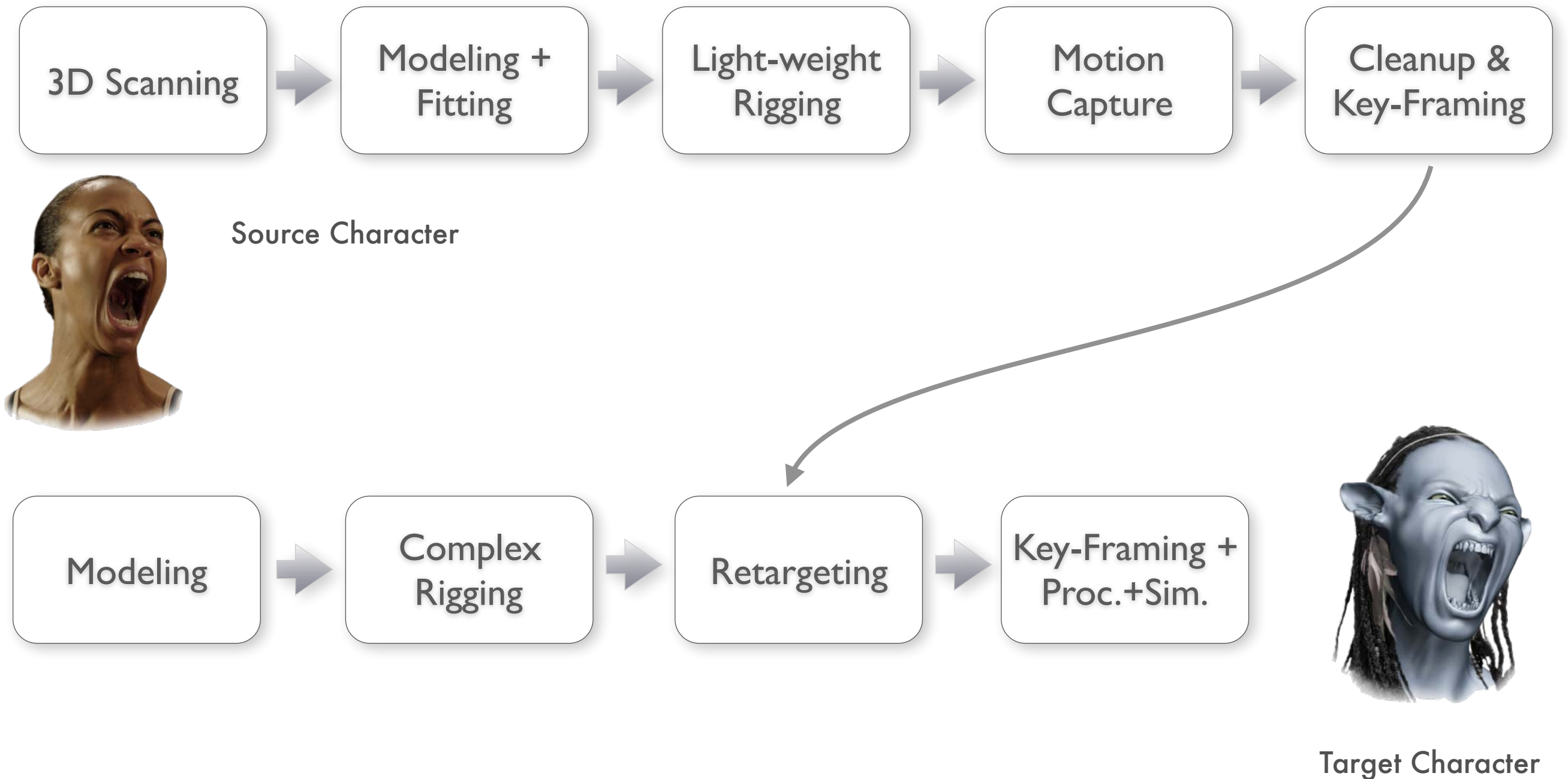




# Vision-Based Tracking & Texturing



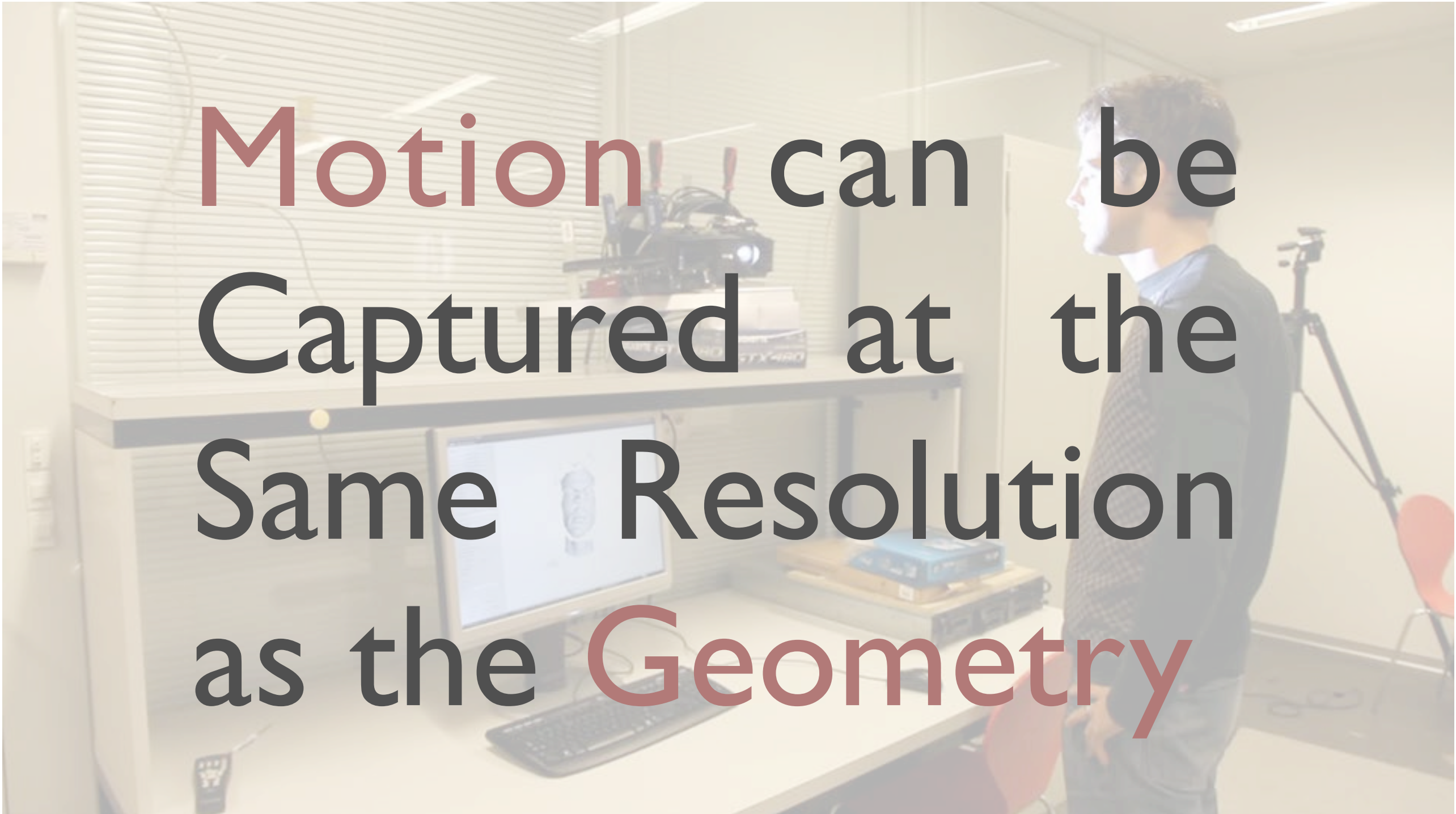
# Typical Animation Workflow in Industry





# Markerless Facial Capture

# 3D Range Sensor



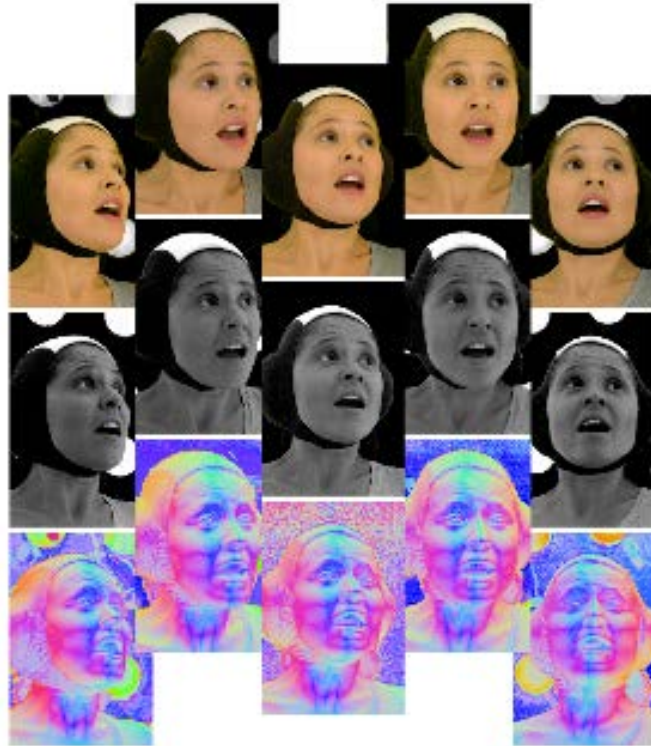
Motion can be  
Captured at the  
Same Resolution  
as the Geometry



# Vapor Ware? (Spatial Phase Imaging)

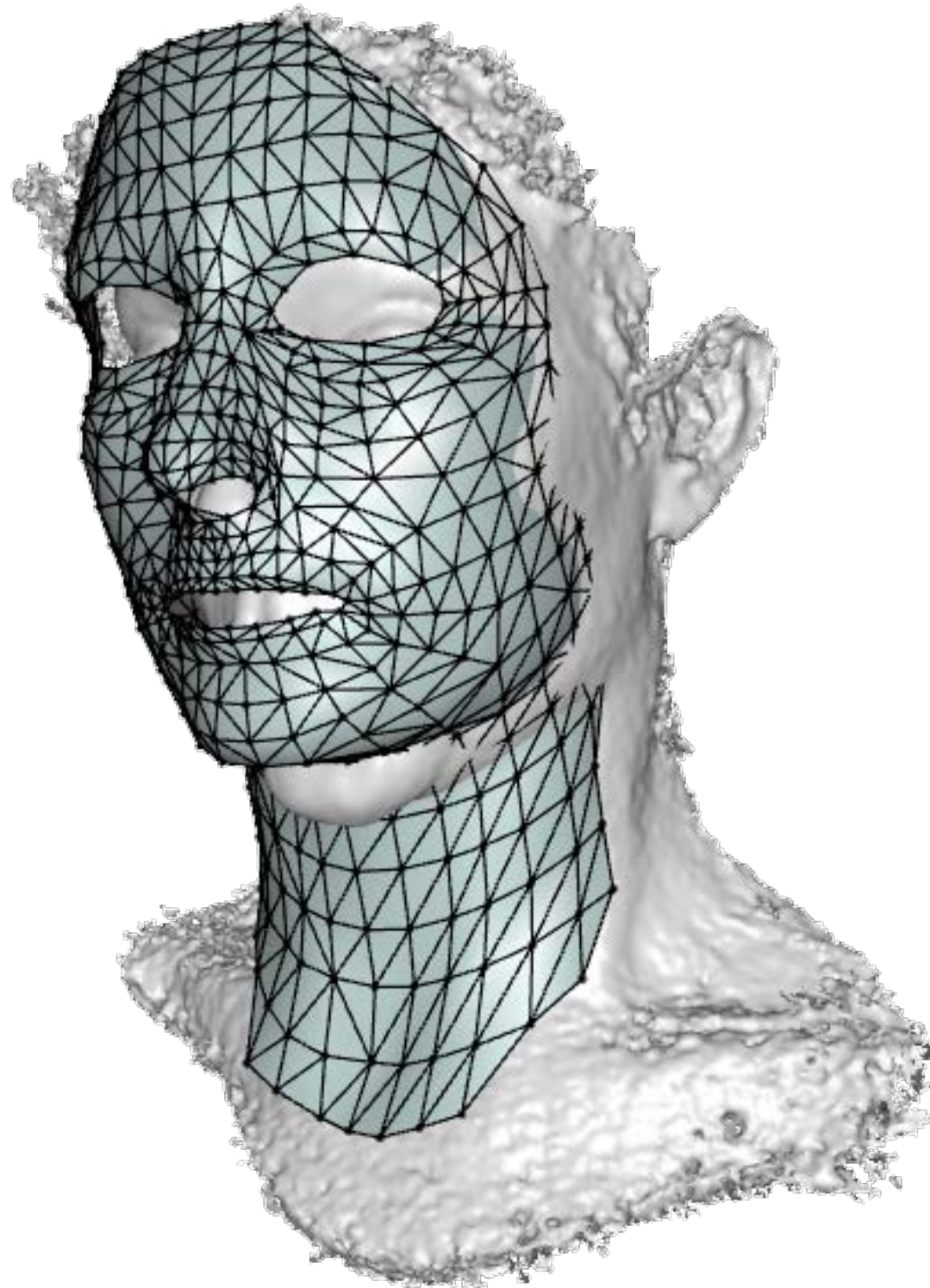


# USC ICT Light Stage 5

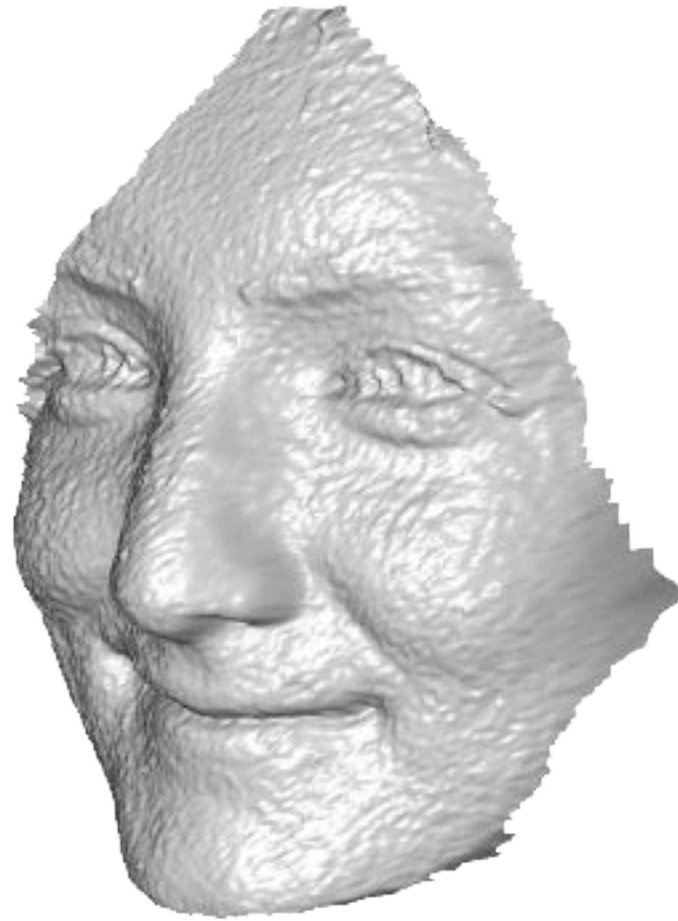




# Template Fitting

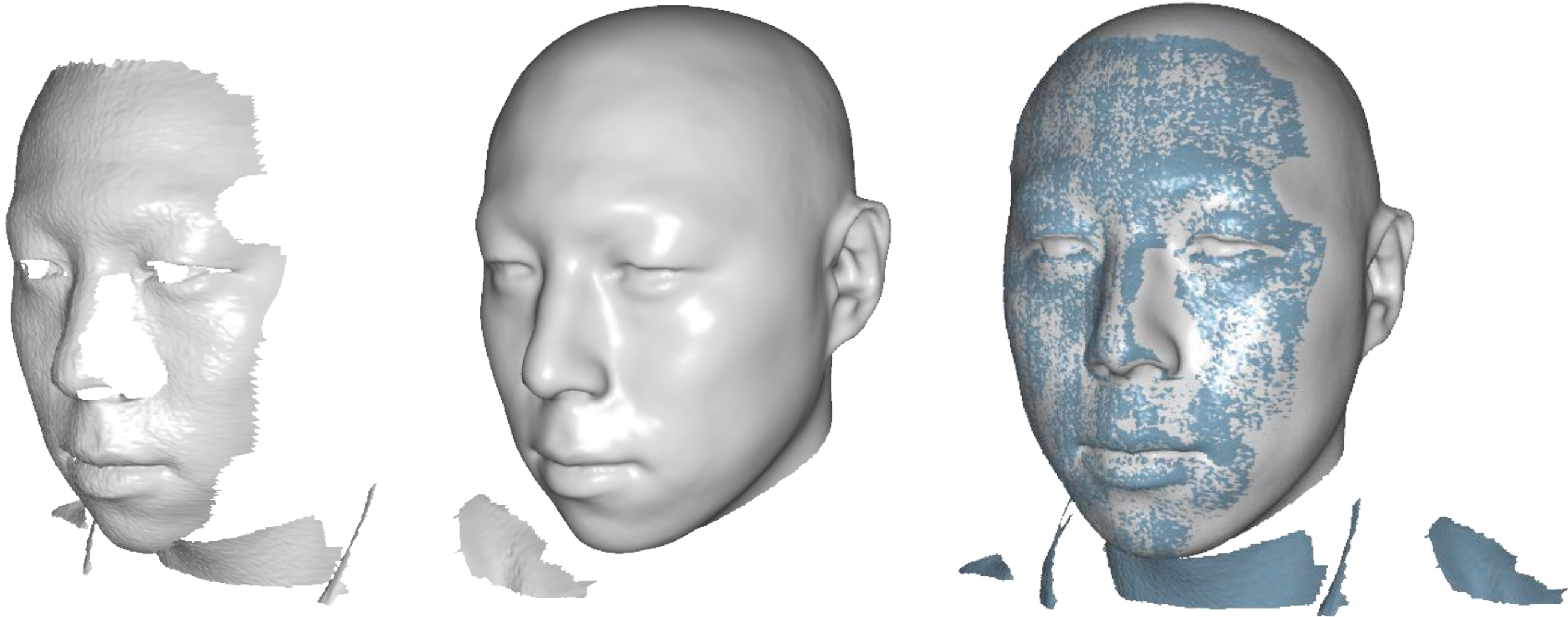


# Template Fitting with PCA





# Template Based Tracking



# Overview



# Using Light Stage X



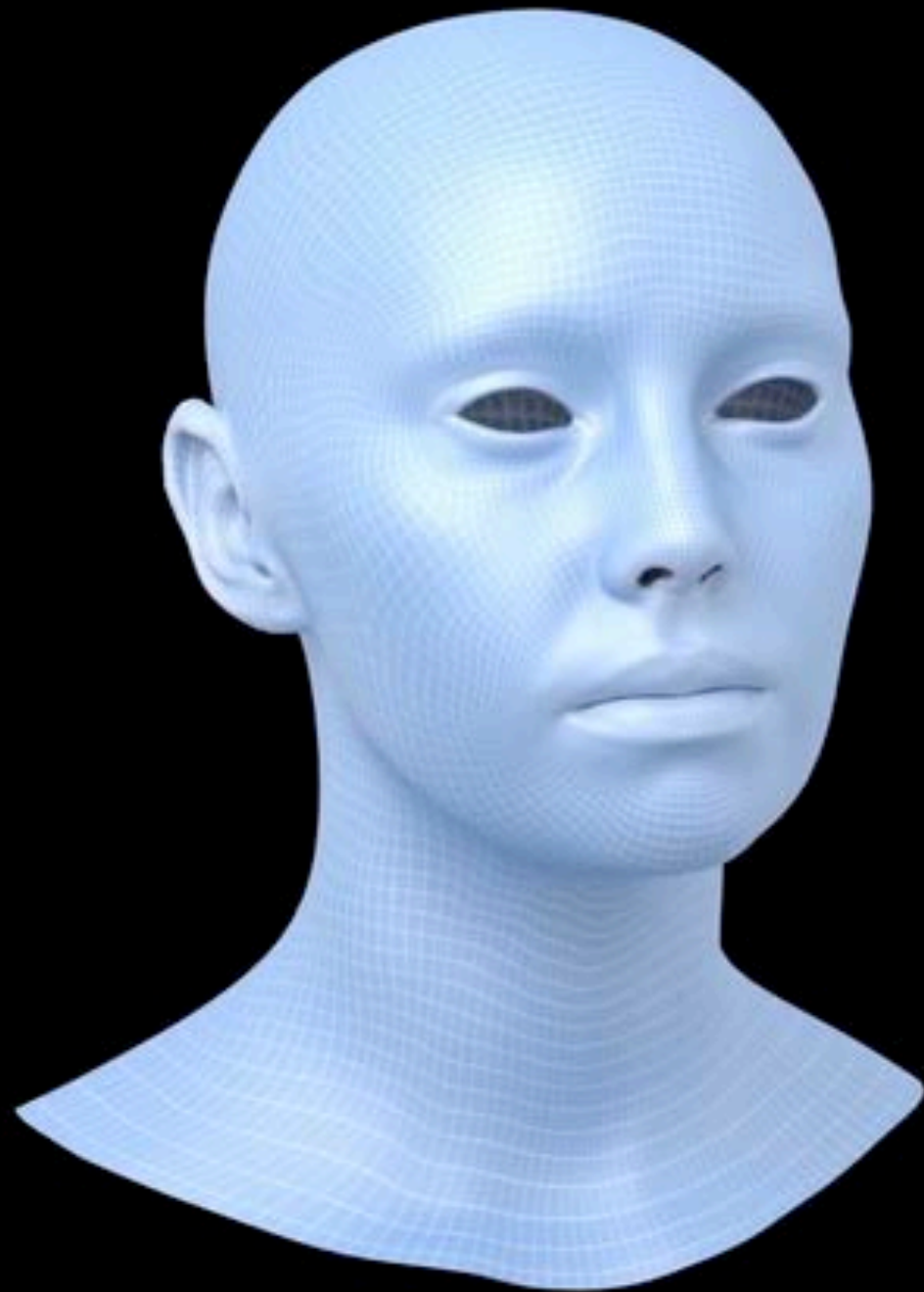
# Using Light Stage X



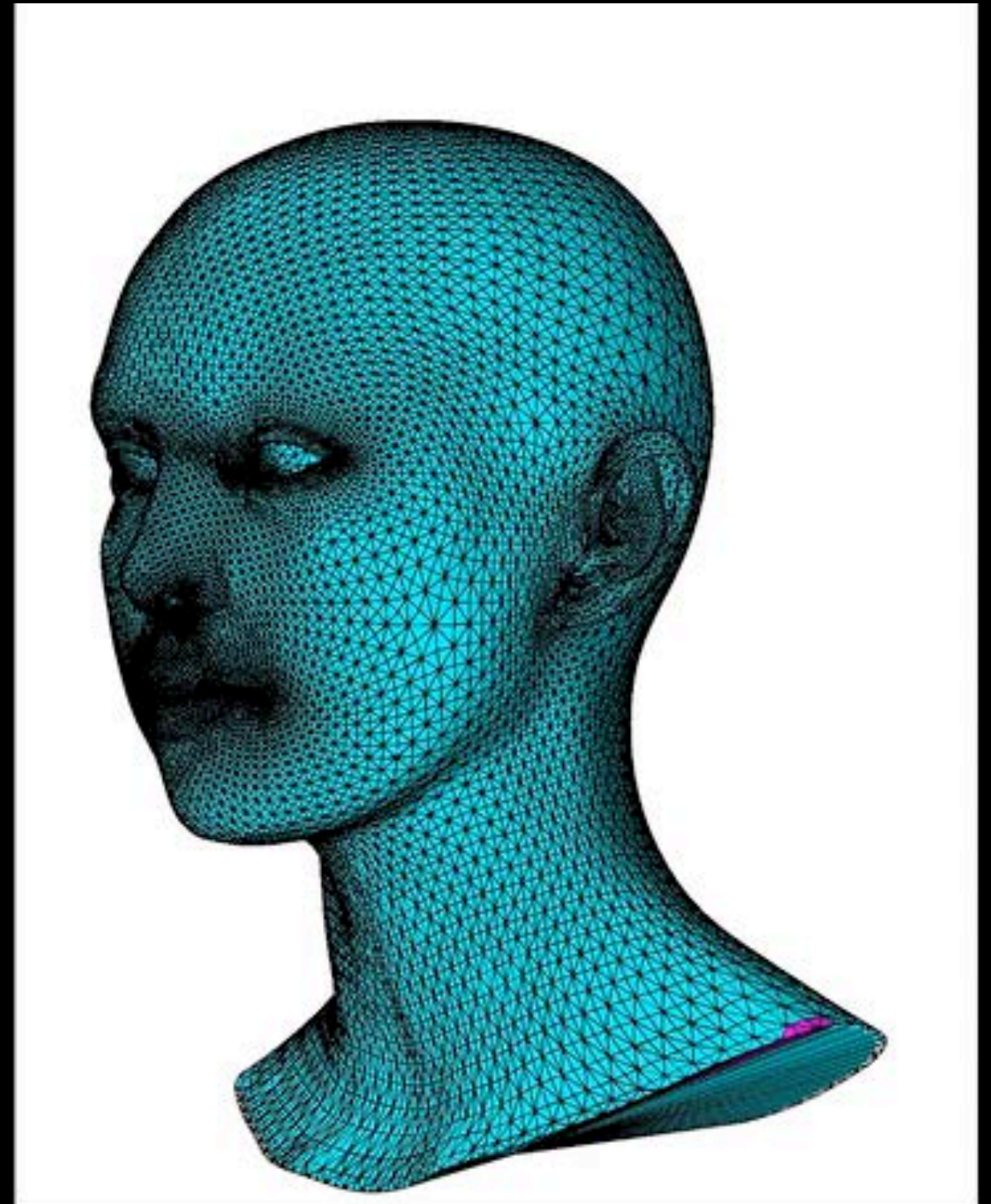
Facial Performance Captured under static Fullon BLUE LEDs



# Using Light Stage X



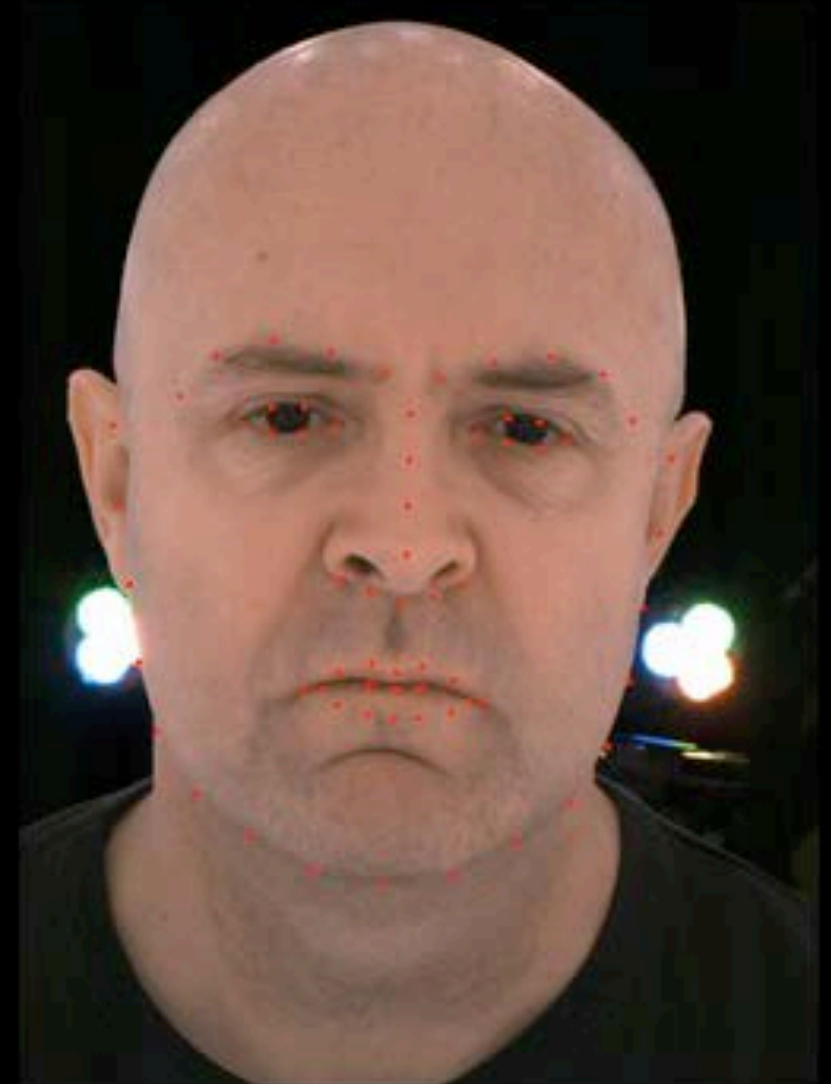
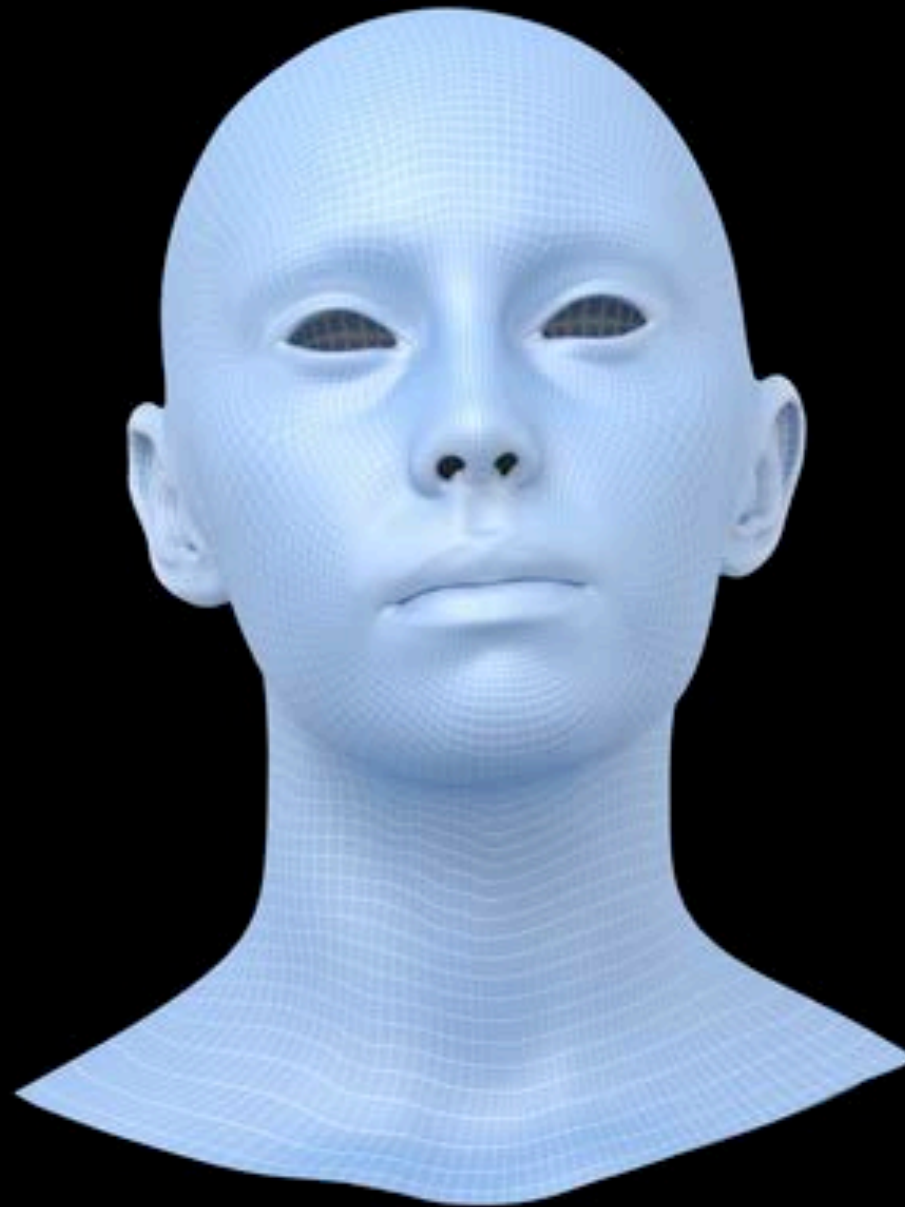
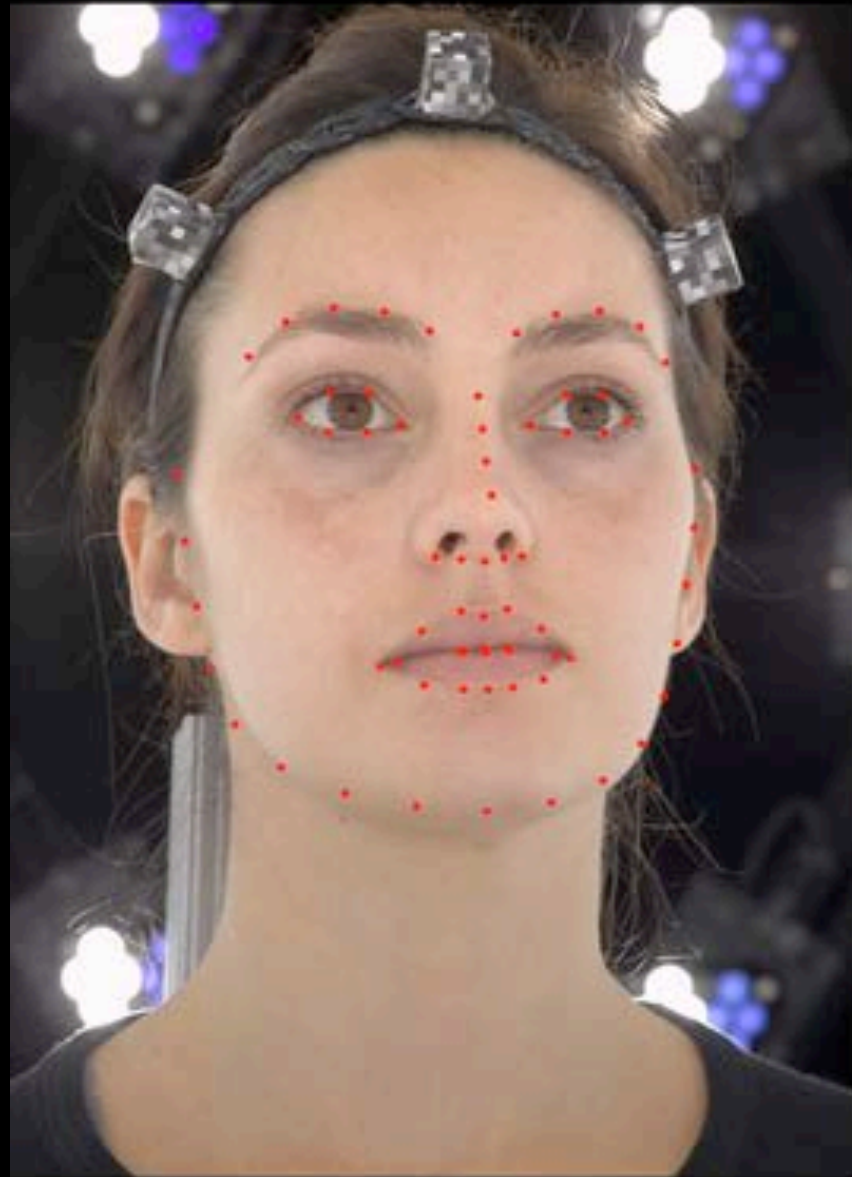
Artist Quality Template



Tetrahedral Mesh Constructed From the Surface



# Using Light Stage X



Initial Deformation Based on Facial andmarks

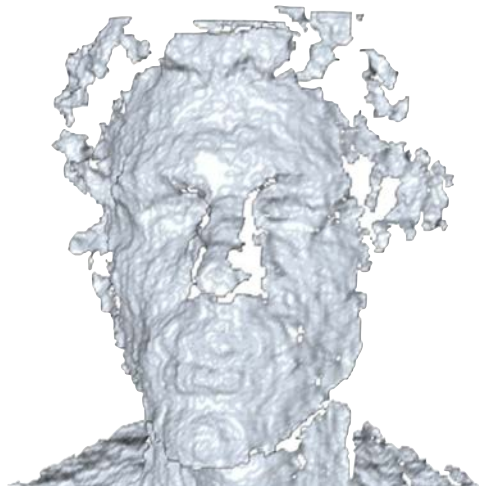


# Using Light Stage X



Performance Tracking of Multiple Subject using a Shared Template

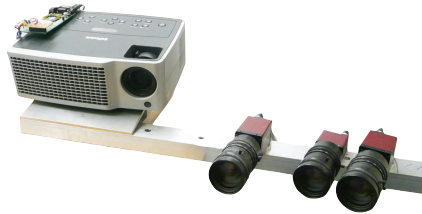
# Requirements for a Practical System



1. Real-time performance

2. Robustness to noise

3. High-level semantics



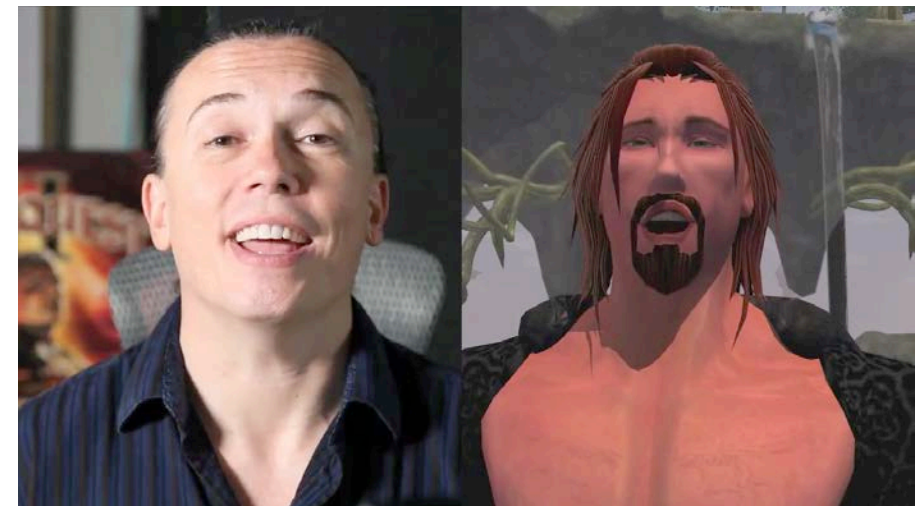


# Realtime Facial Capture

# Why Realtime?



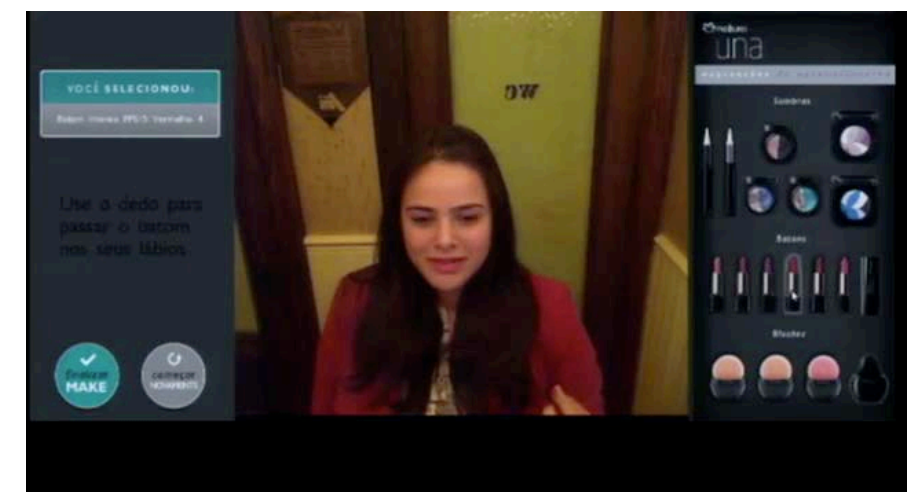
VFX/Game Production



Virtual Avatars



Robotics



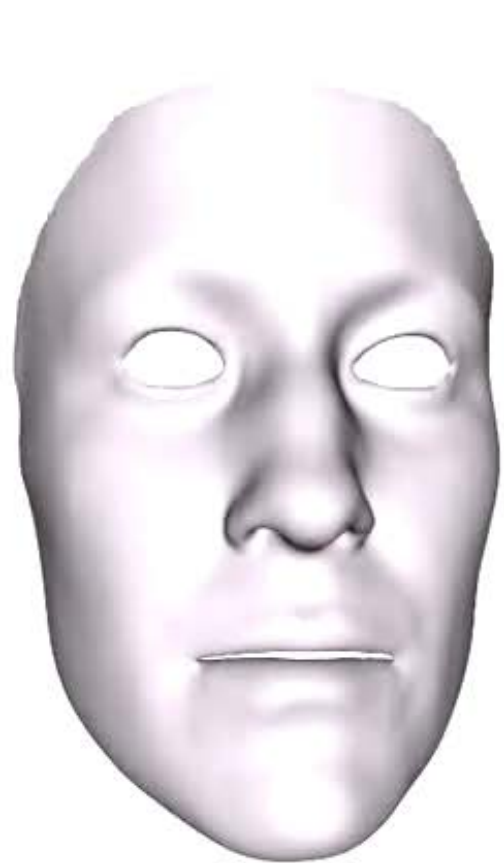
AR/Virtual Mirror



# Objective



# Building Expression Space



tracked template



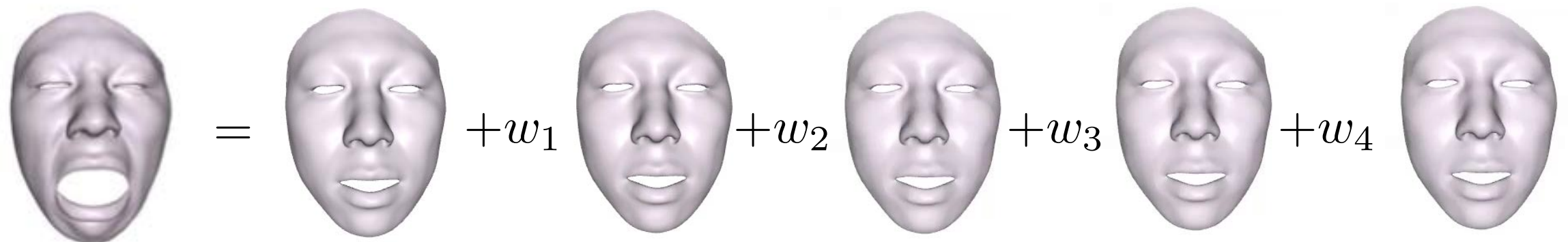
input scan



# Expression PCA for Reduced Dimension

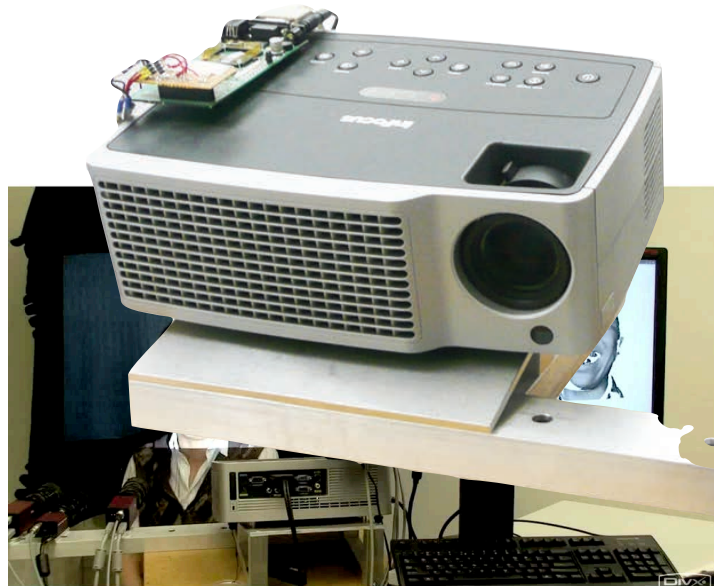


Principal Component Analysis



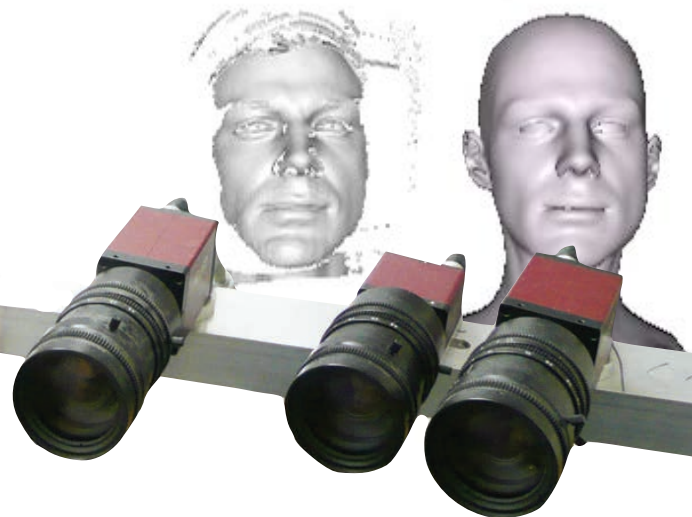
# Realtime Systems

## depth sensor as input



**with training**

Weise et al. SCA 09



**with little training**

Li et al. Siggraph 2010



**little to no training**

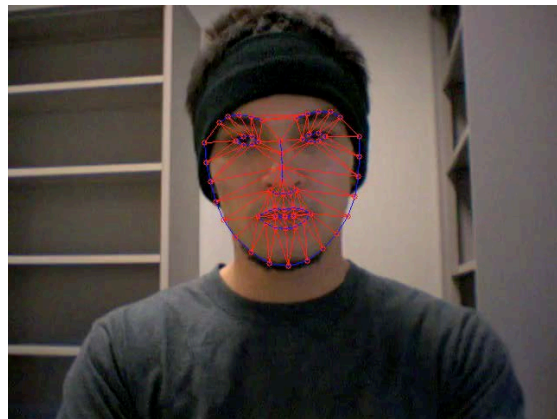
Weise et al. Siggraph 2011 &  
Bouaziz et al. Siggraph 2013

**reduced calibration and more accessible**



# Realtime Systems

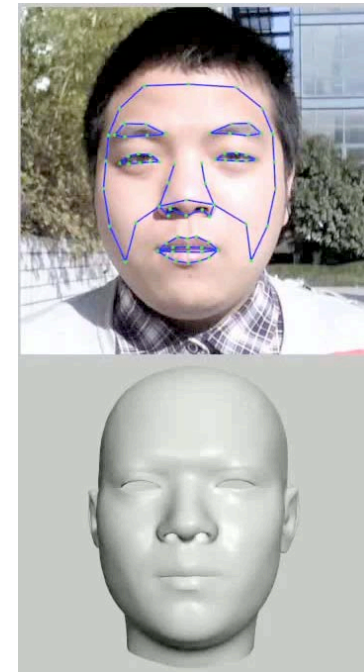
**video as input**



**without training**  
Saragih et al. IJCV 2011



**without training**  
Image Metrics 2011



**with training**  
Cao et al. Siggraph 2013

**increased accuracy and expressiveness**

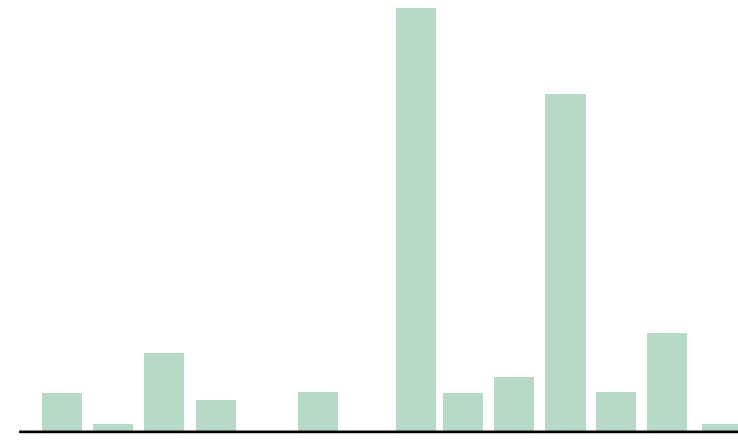
# Automatic Facial Rigging



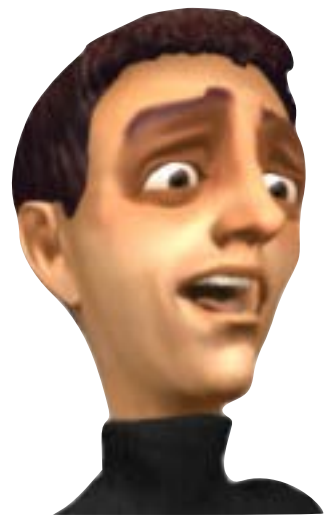
# Blendshape Animation

## Blendshape Animation

blending weights



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



laughing



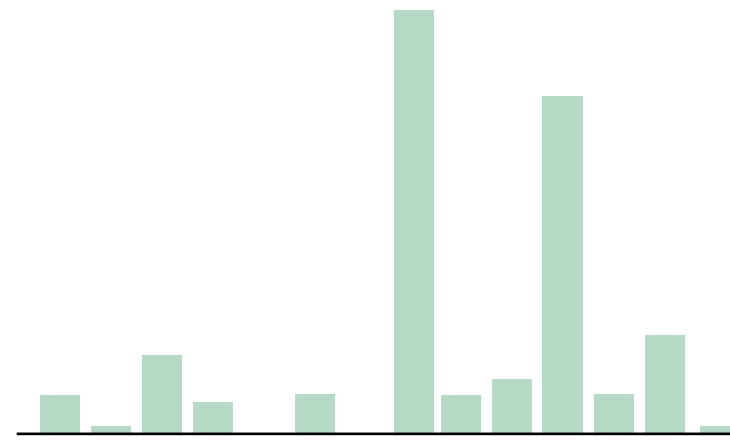
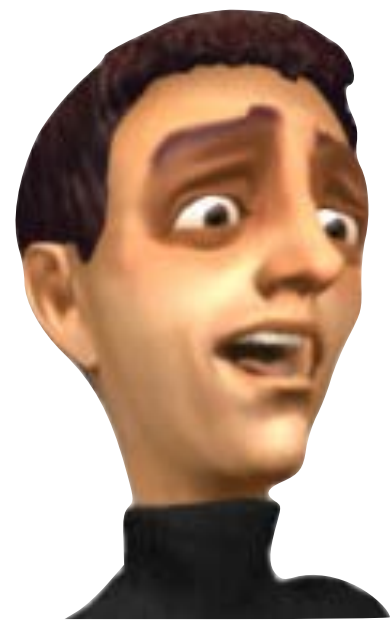
neutral face



blendshapes



# Blendshape Retargeting



laughing



many blendshapes

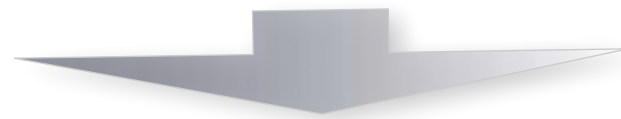


# Expression Transfer

prior  
blendshapes



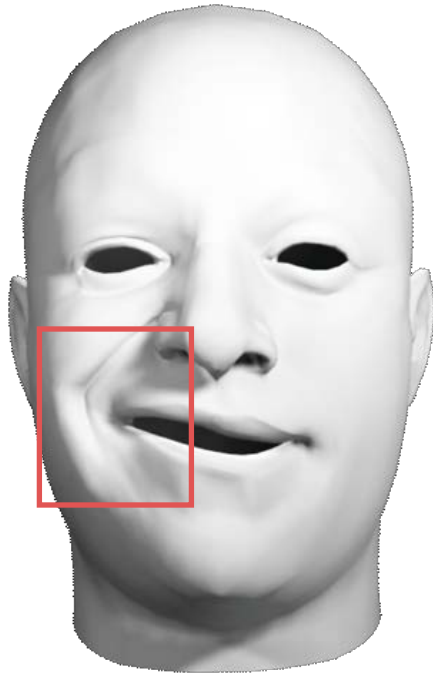
[Noh & Neumann '01]  
[Sumner & Popovic '04]



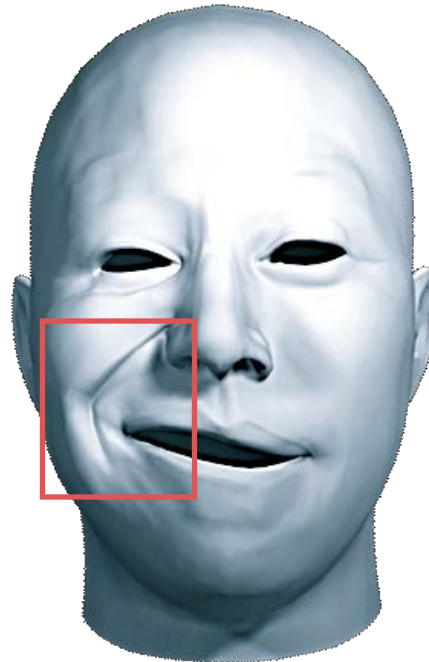
reconstructed  
blendshapes



# Problems



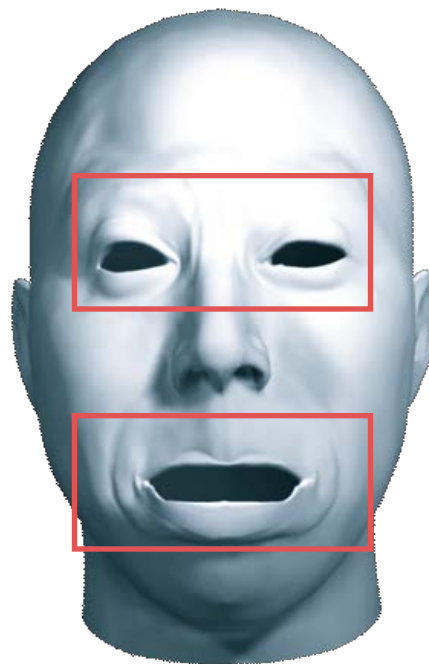
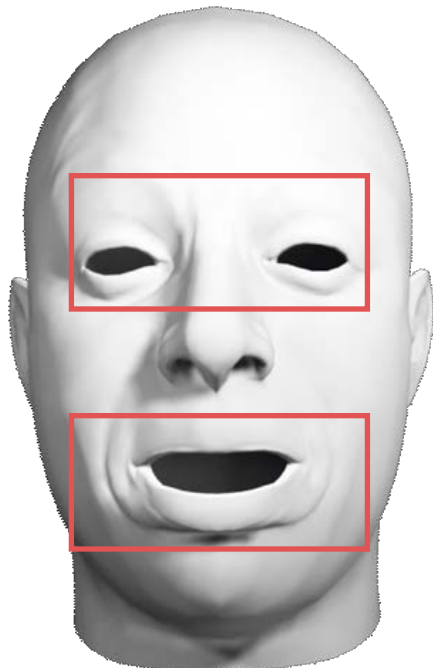
prior



expression transfer

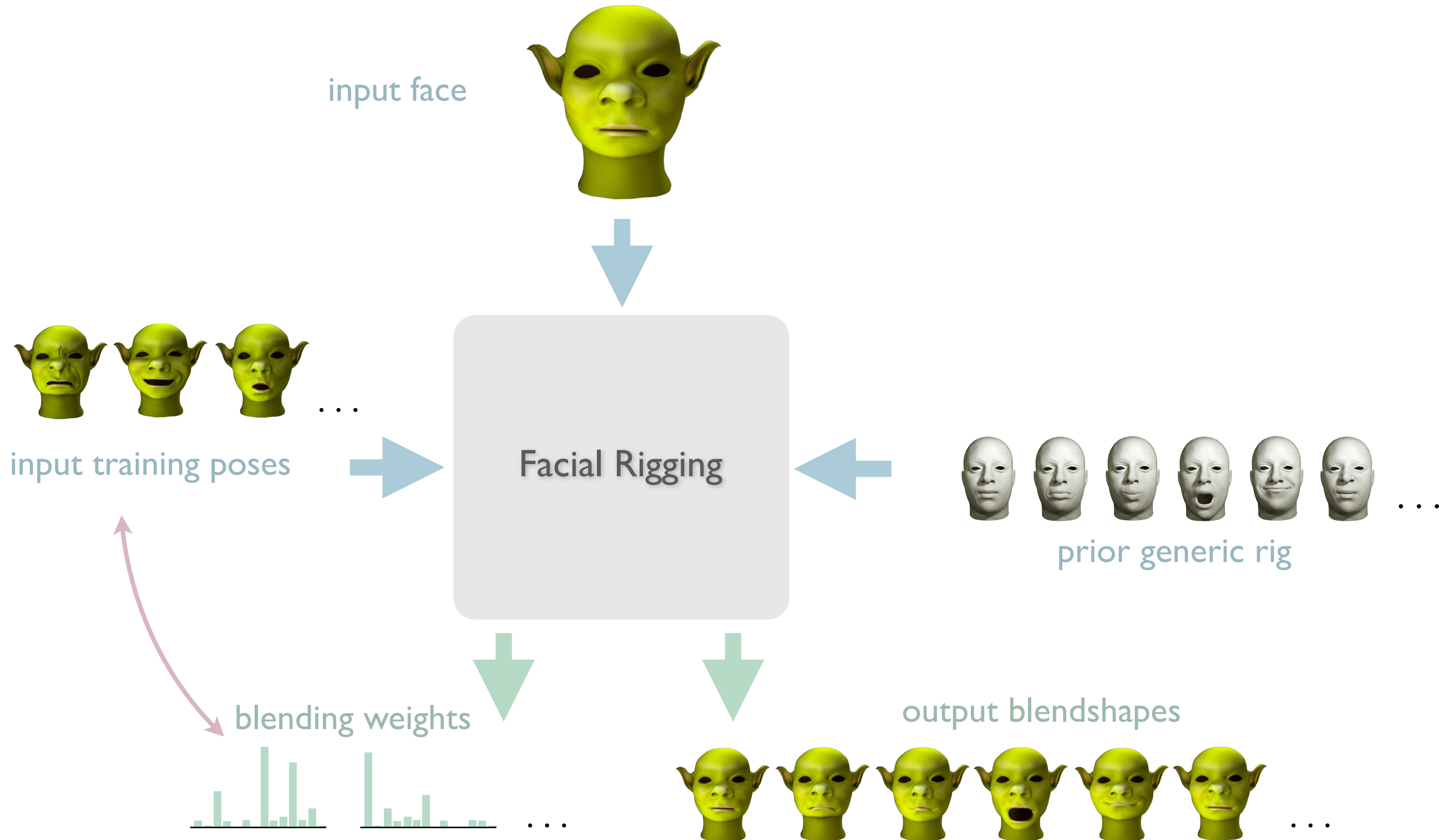


ground truth

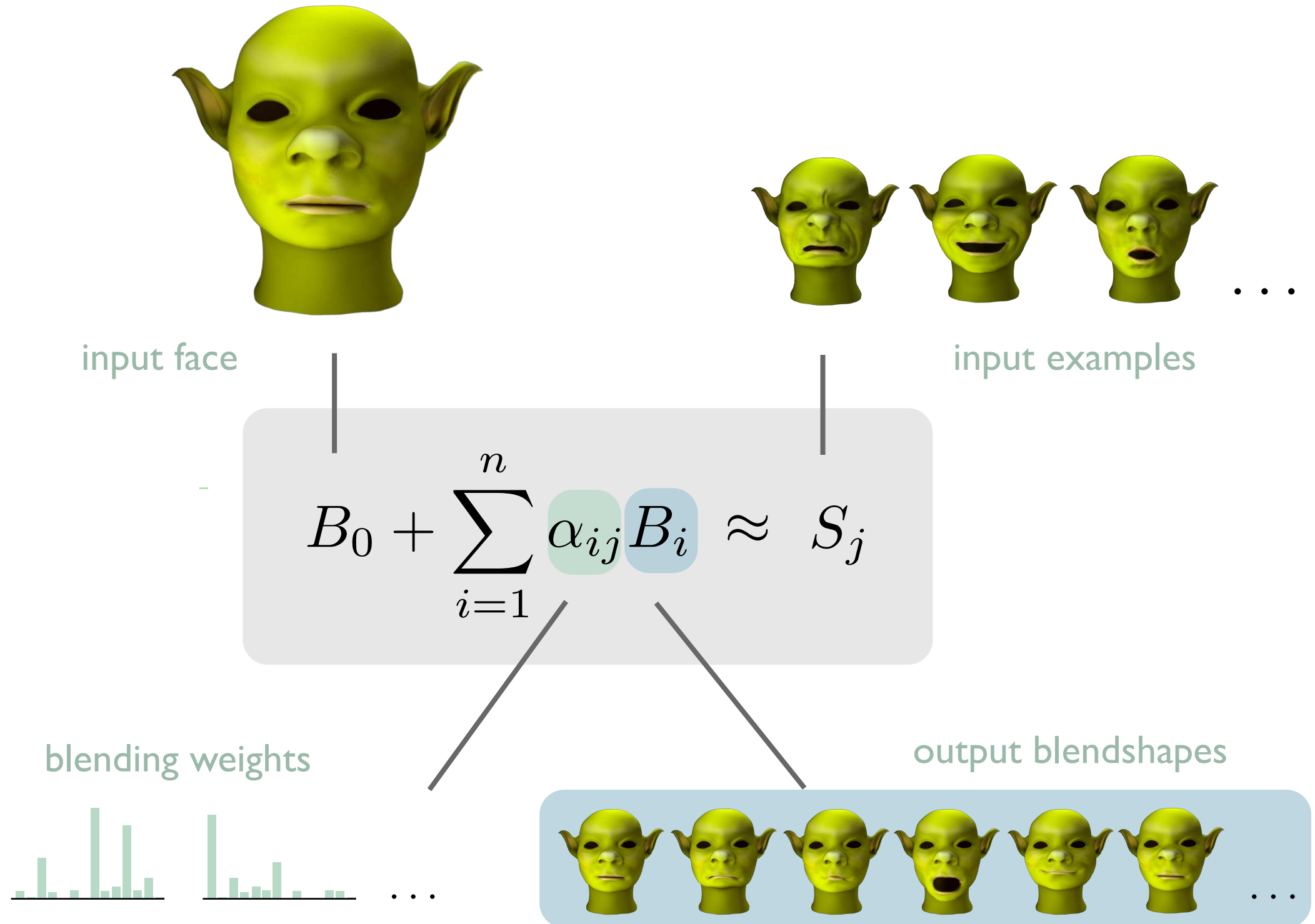




# Example Based-Facial Rigging



# Bilinear Problem



# Decoupled Optimization

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



# Decoupled Optimization

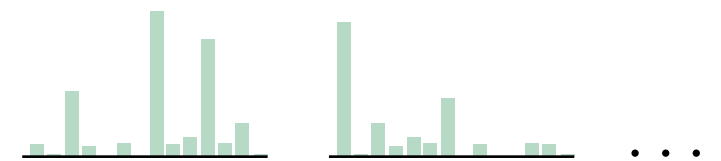
Step A

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

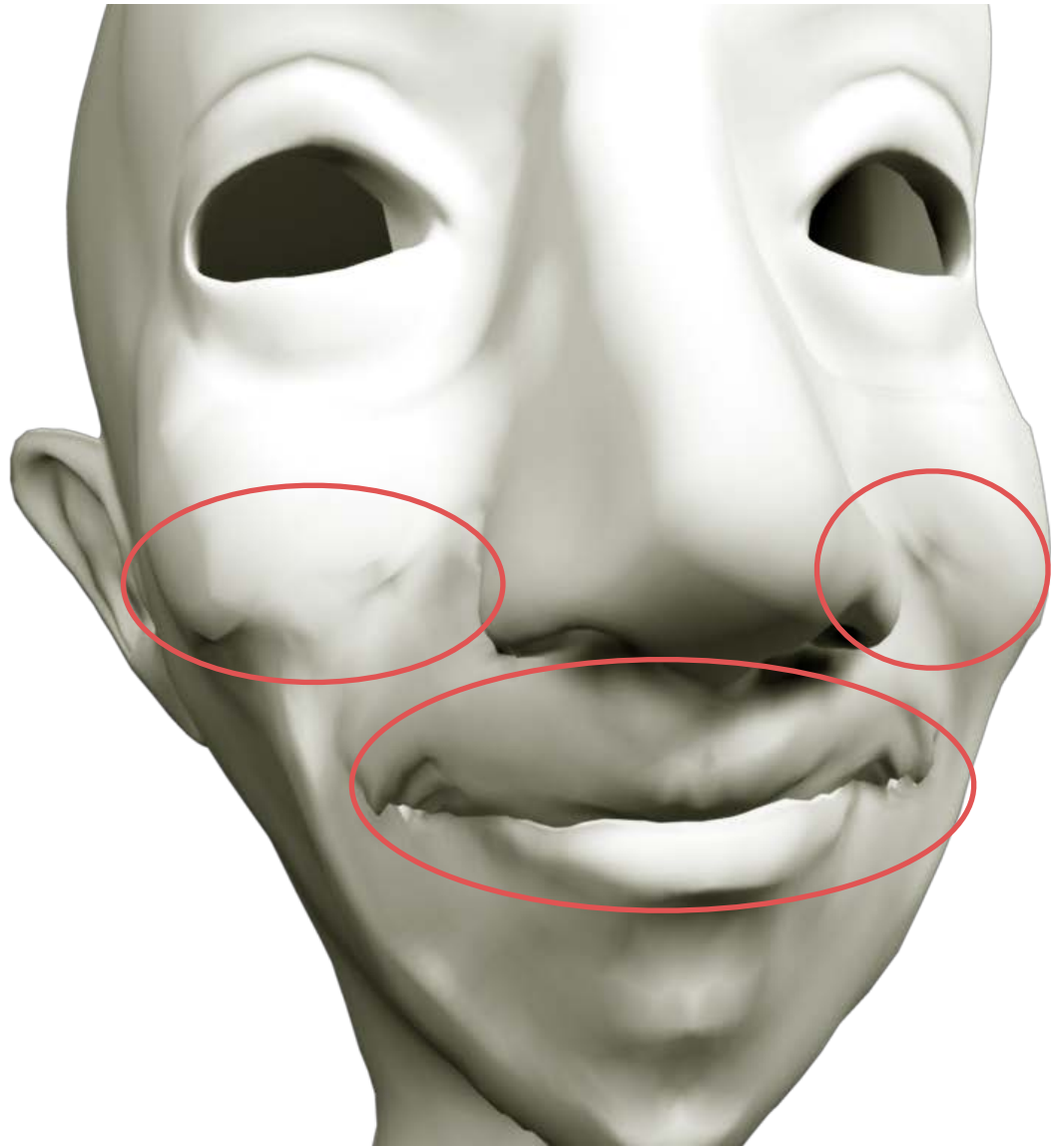


Step B

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



# Gradient Domain Optimization



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

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

# Comparison



prior



without examples



with 6 examples



input example

whistle

surprise



# Directable Facial Animation



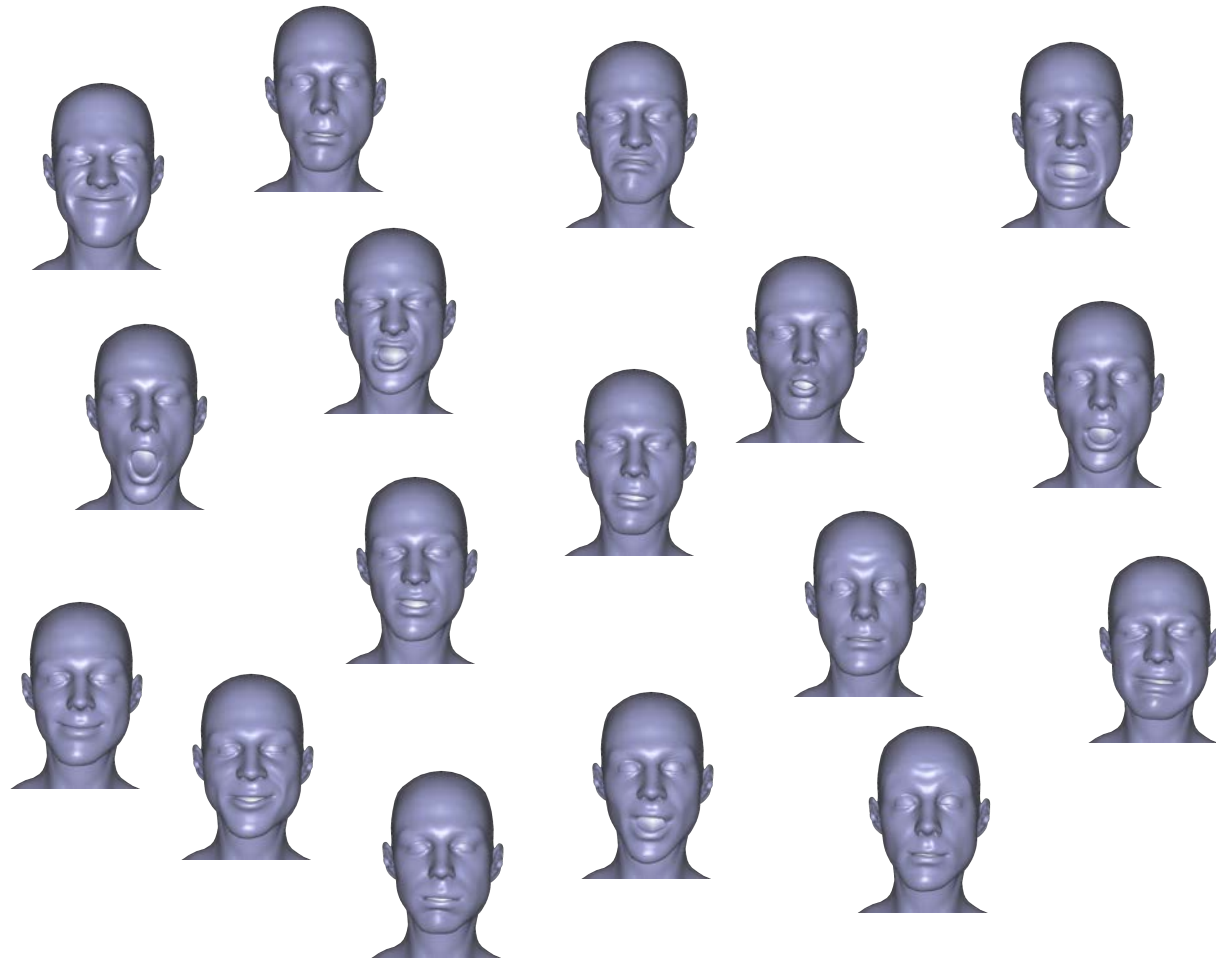
3D scans



facial tracking

# Blendshapes for Tracking

ICP with Blendshapes



# Animation Prior



# Problem: **Noisy Input**

Tracking Correction with Animation Prior



input scans



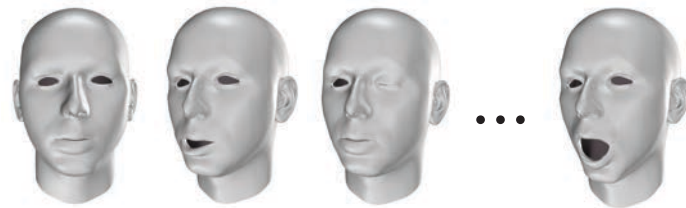
tracking



goal

# Performance-Based Facial Tracking

expression model



animation prior

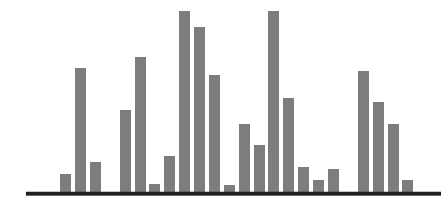


input data



Stable Tracking

expression parameters

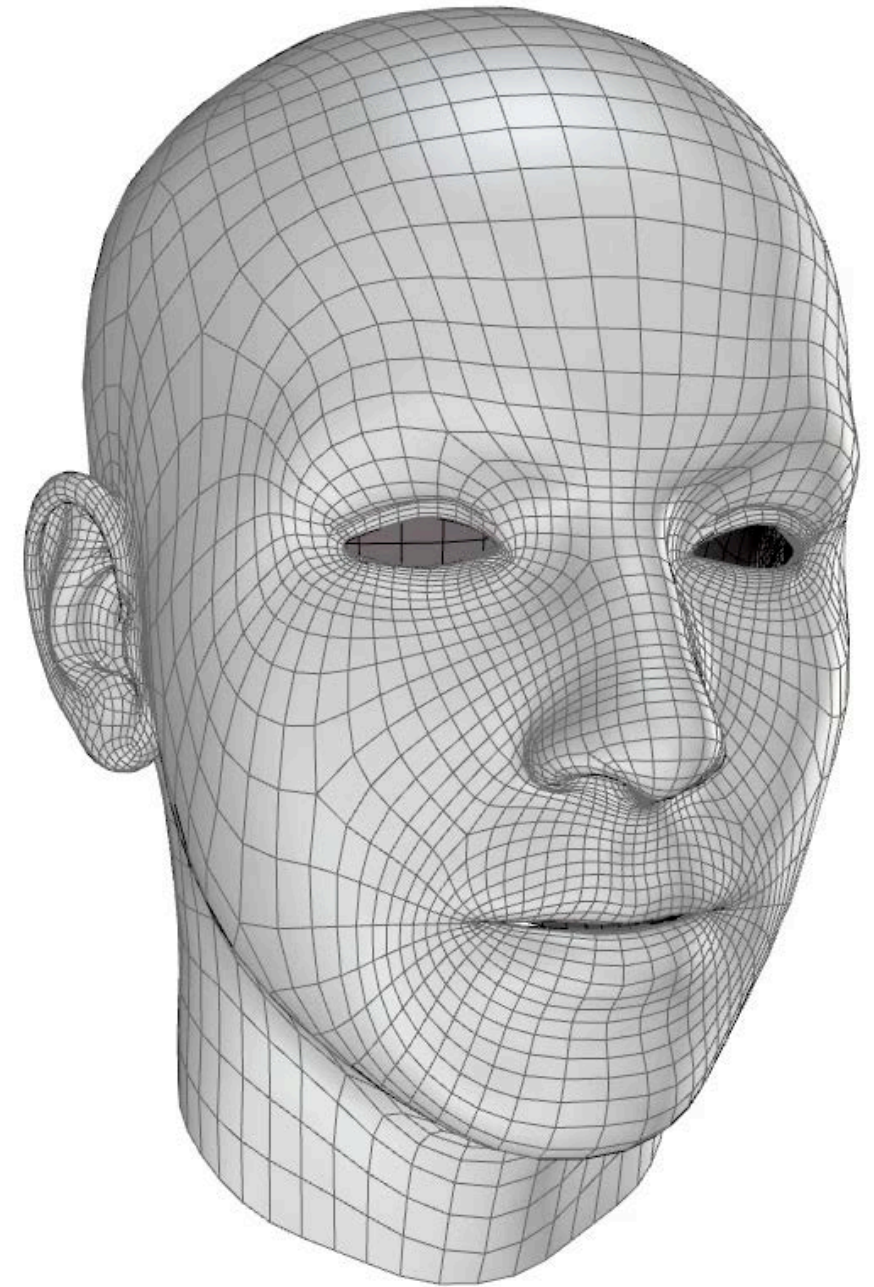


temporal coherence

# Animation as Prior



reference video

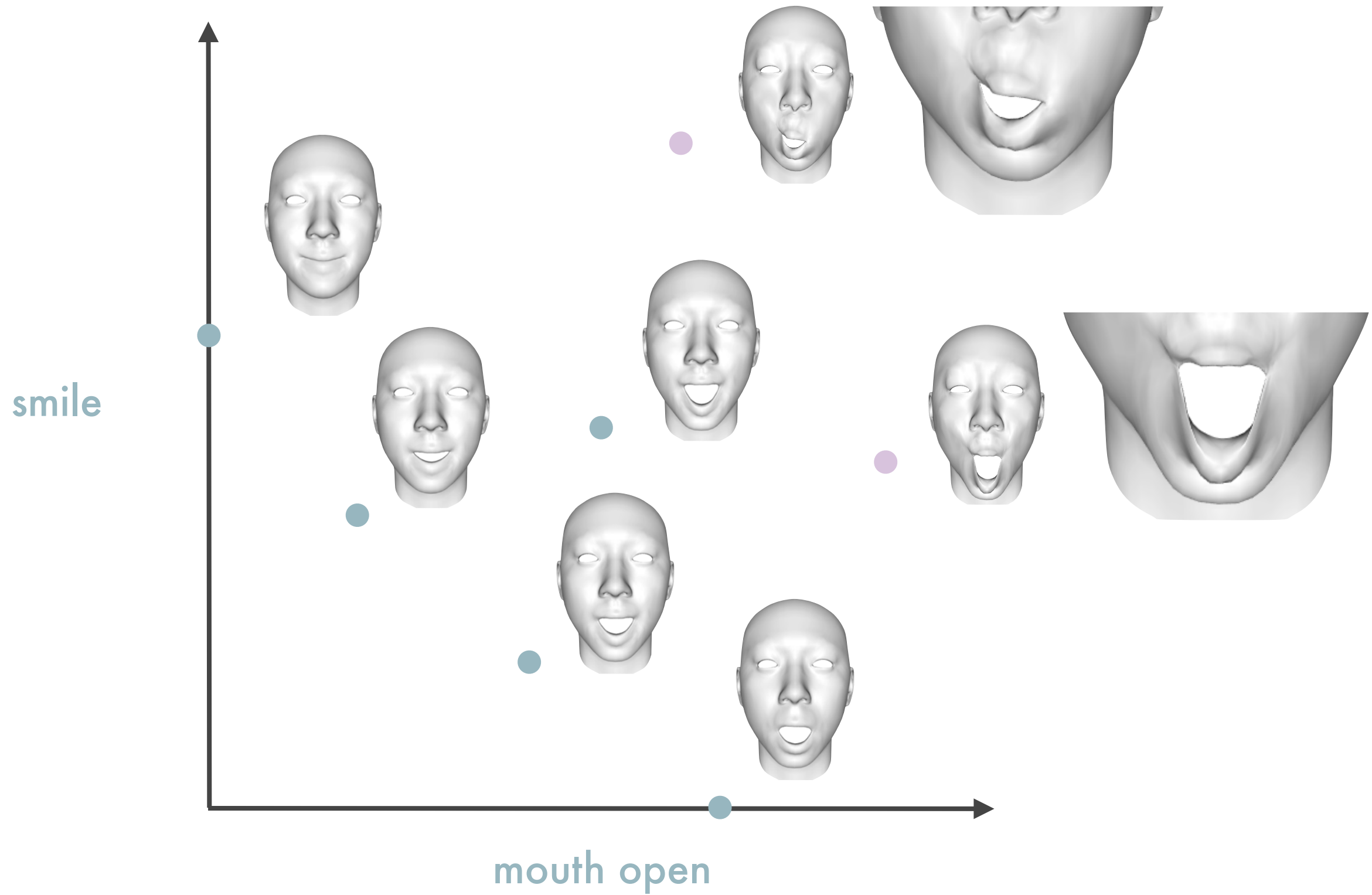


9500 frames

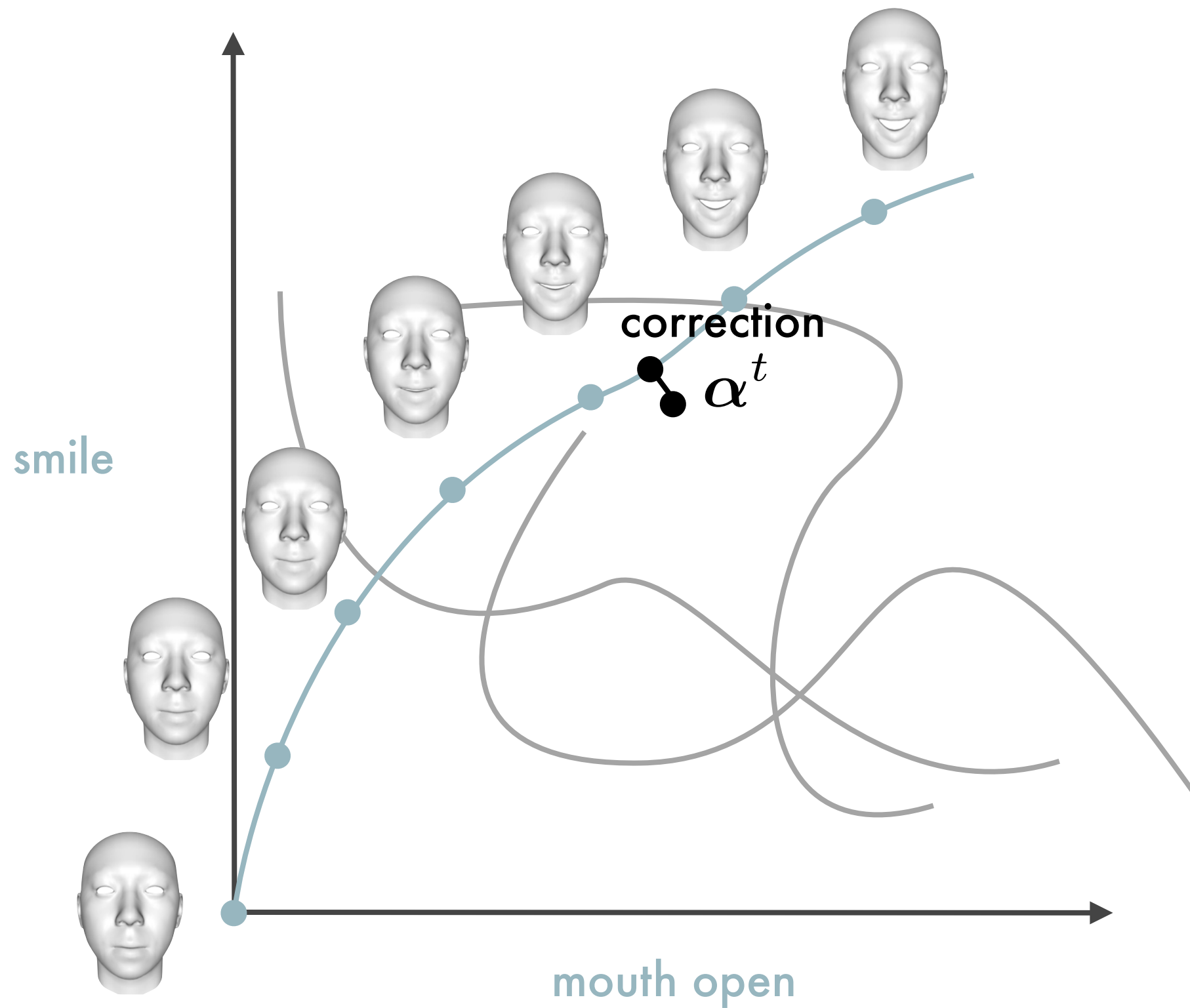




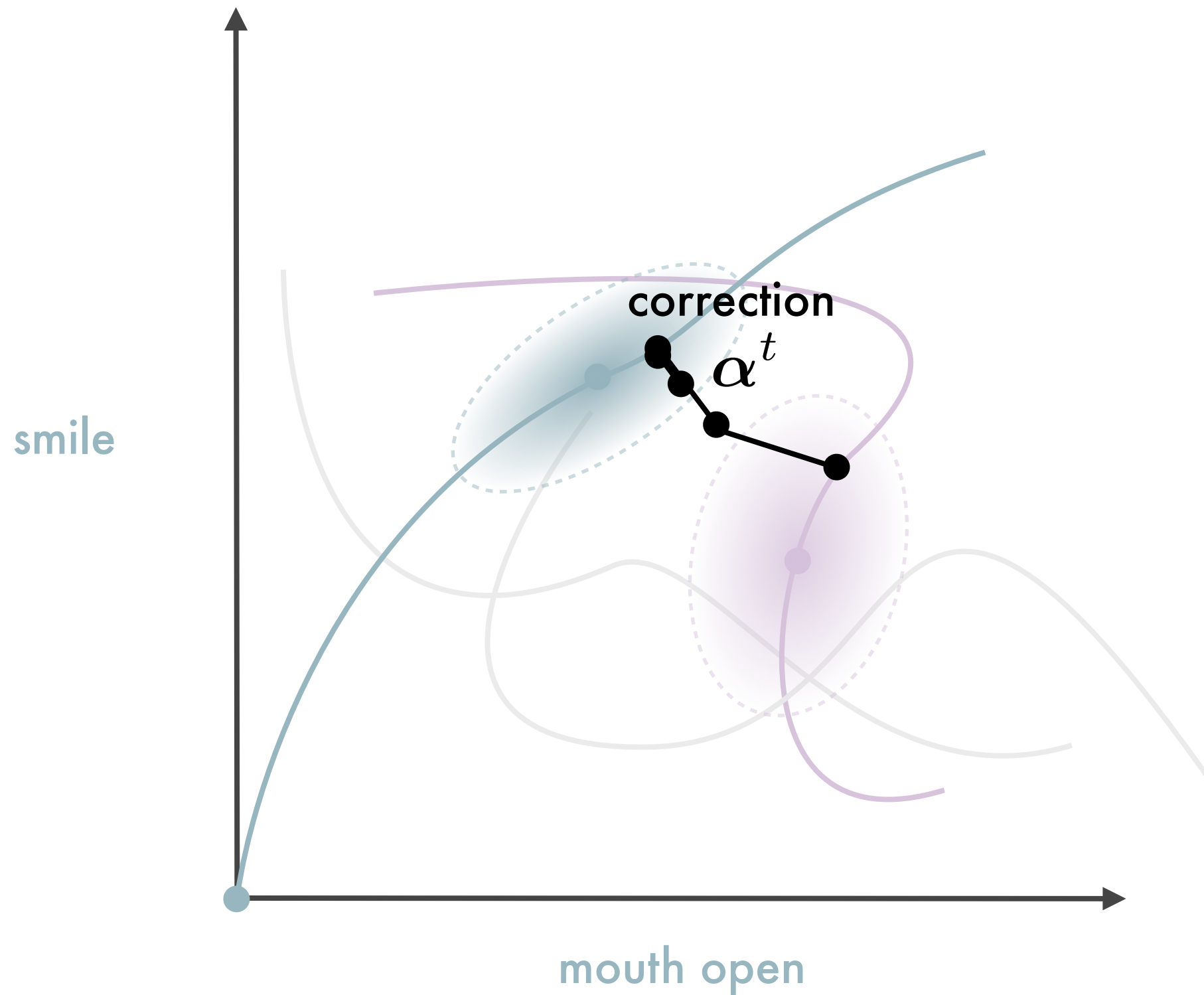
# N-Dim Expression Space



# Animation Manifold



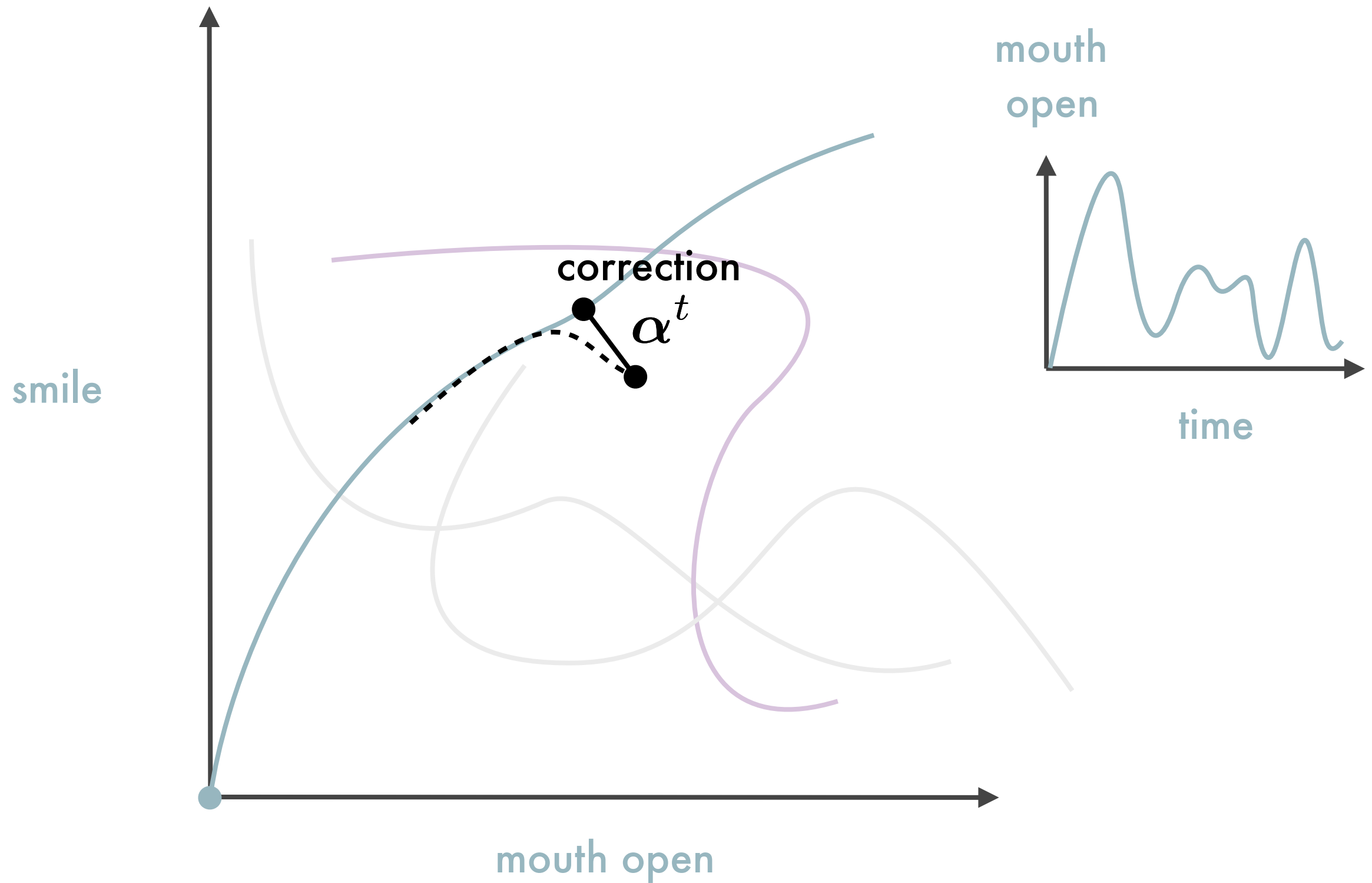
# Probabilistic Animation Prior



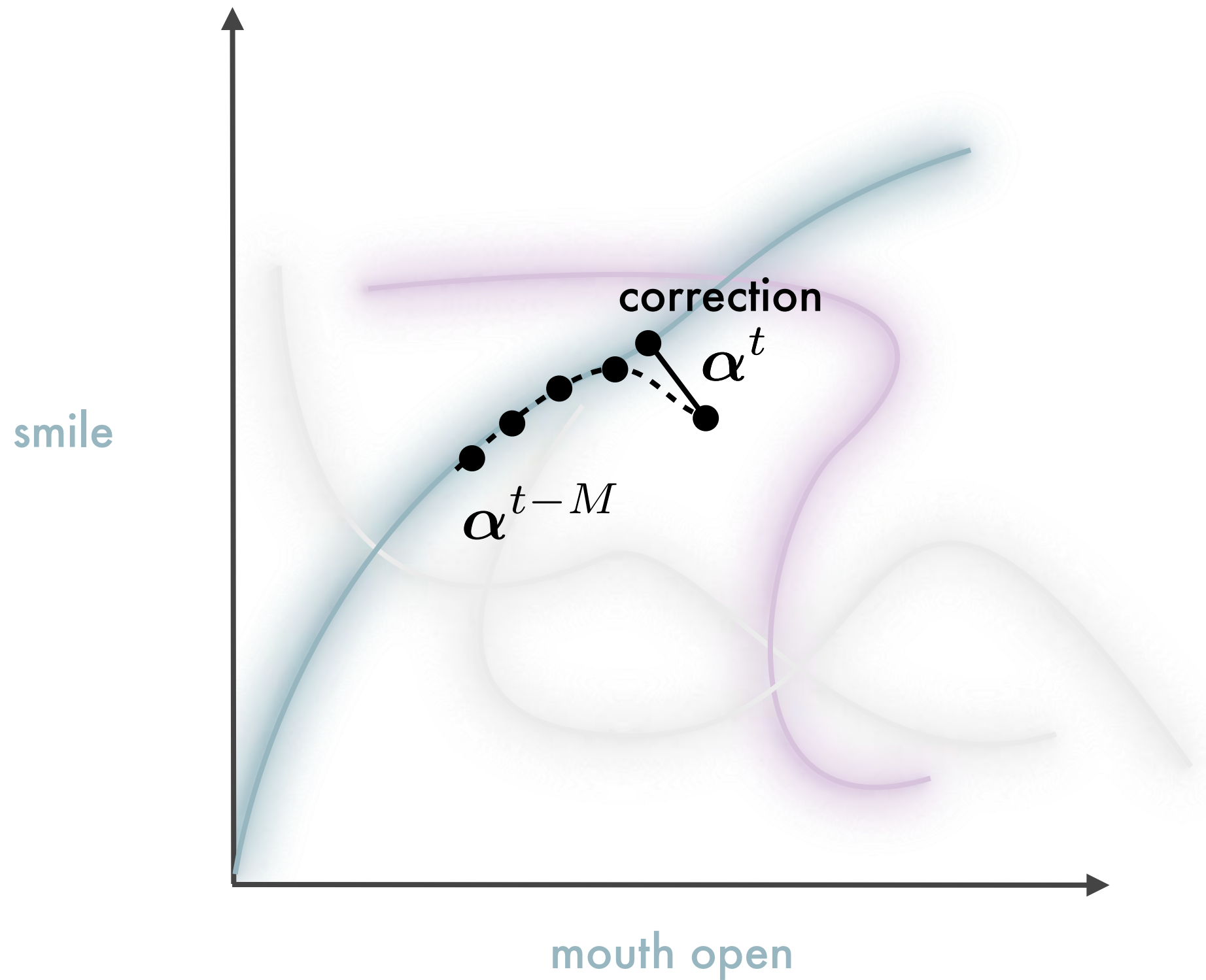
Lau et al. 2009



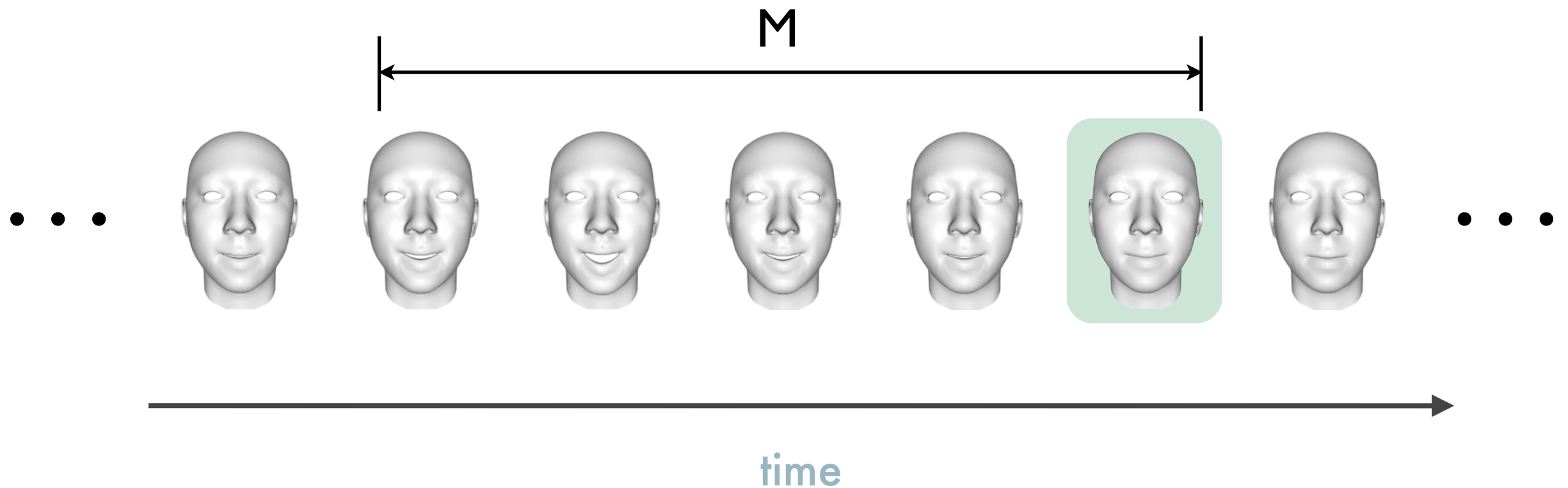
# Probabilistic Animation Prior



# Probabilistic Animation Prior



# Temporal Joint Probabilistic Distribution



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

MPPCA model

weights

mean

principal components

Gaussian noise



# MAP Estimation

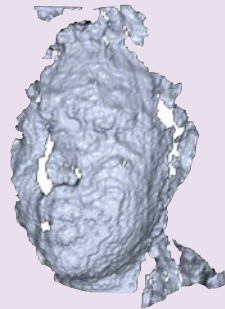


$$\alpha^t = \arg \max_{\alpha} p(\alpha | D, \alpha^{t-1}, \dots, \alpha^{t-M})$$

MPPCA

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

geometry



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

texture



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

# **ILM's Kinect Monster Mirror**

# Fast Calibration

Li et al. SIGGRAPH 2013





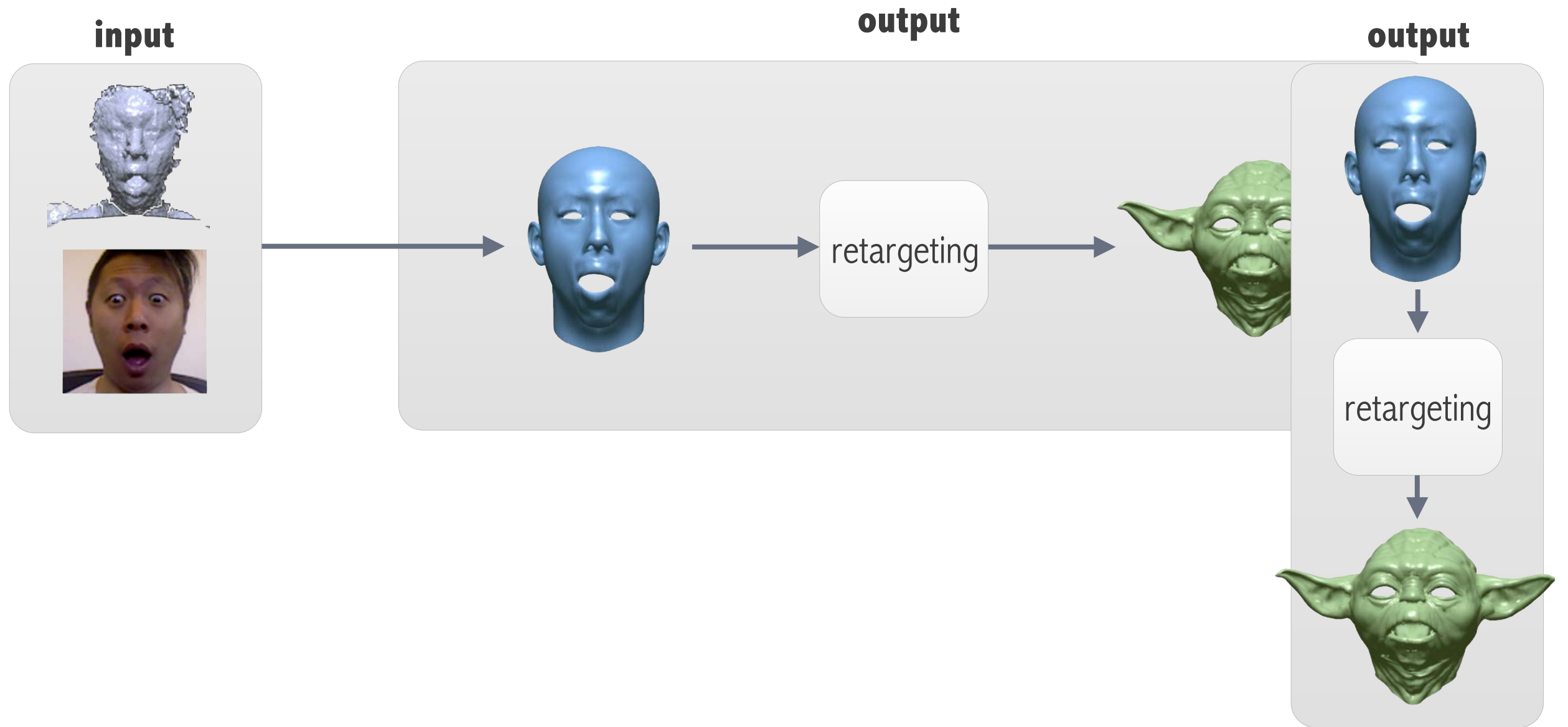
# Facial Performance Capture

Li et al. SIGGRAPH 2013



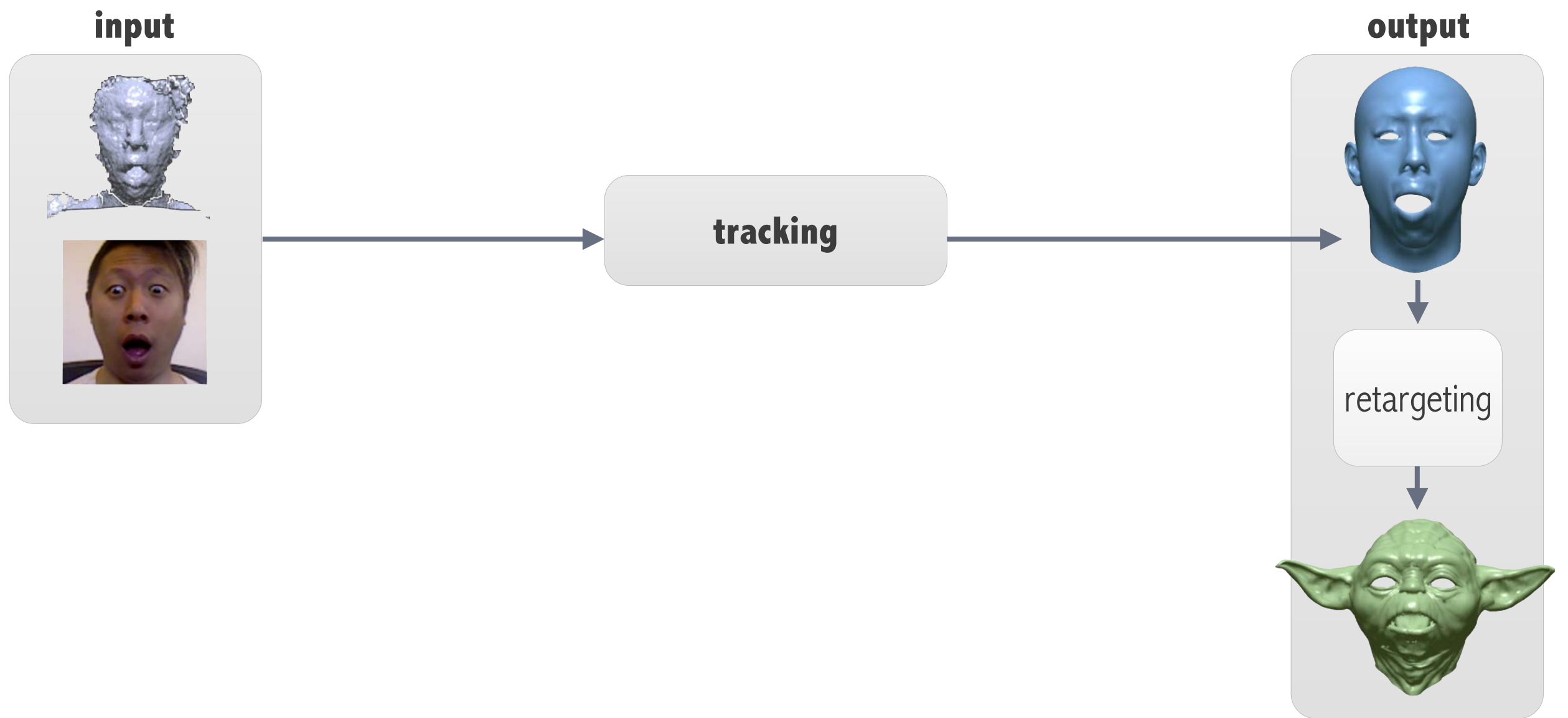
# Pipeline

# Pipeline Overview

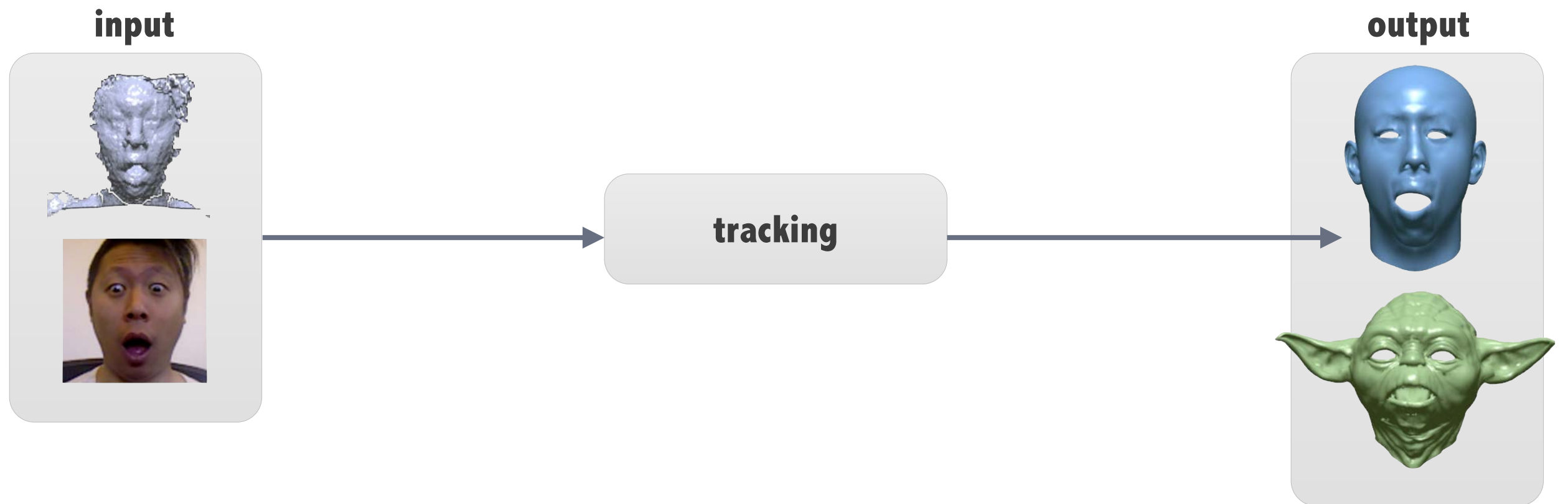




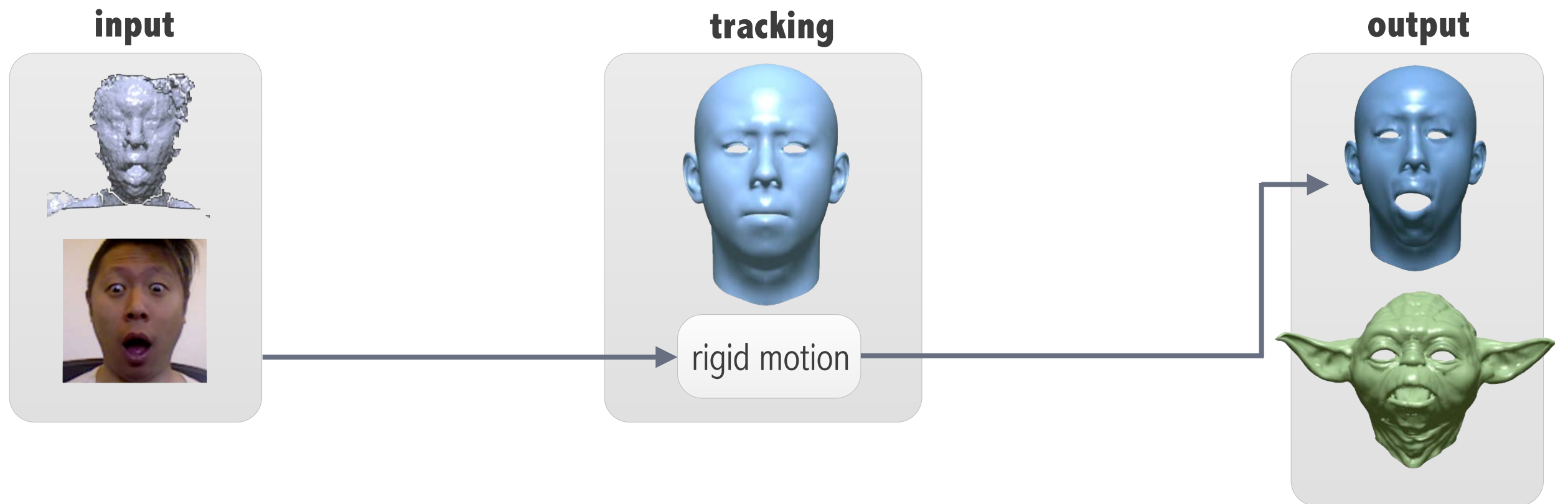
# Pipeline Overview



# Pipeline Overview

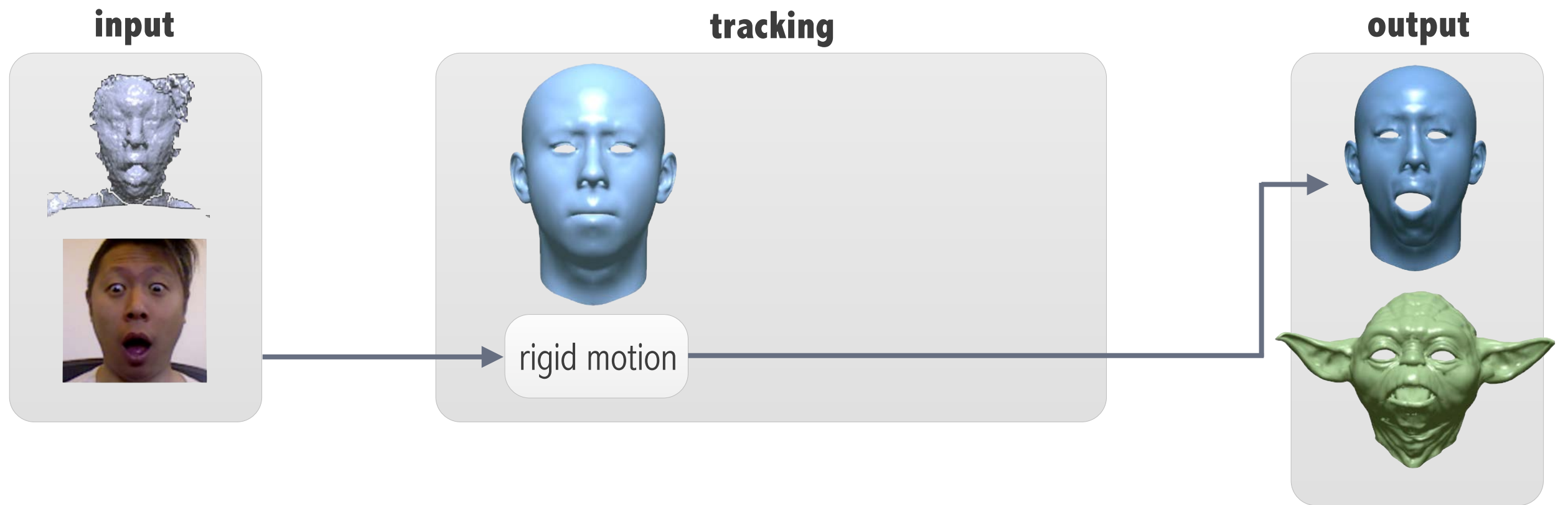


# Pipeline Overview

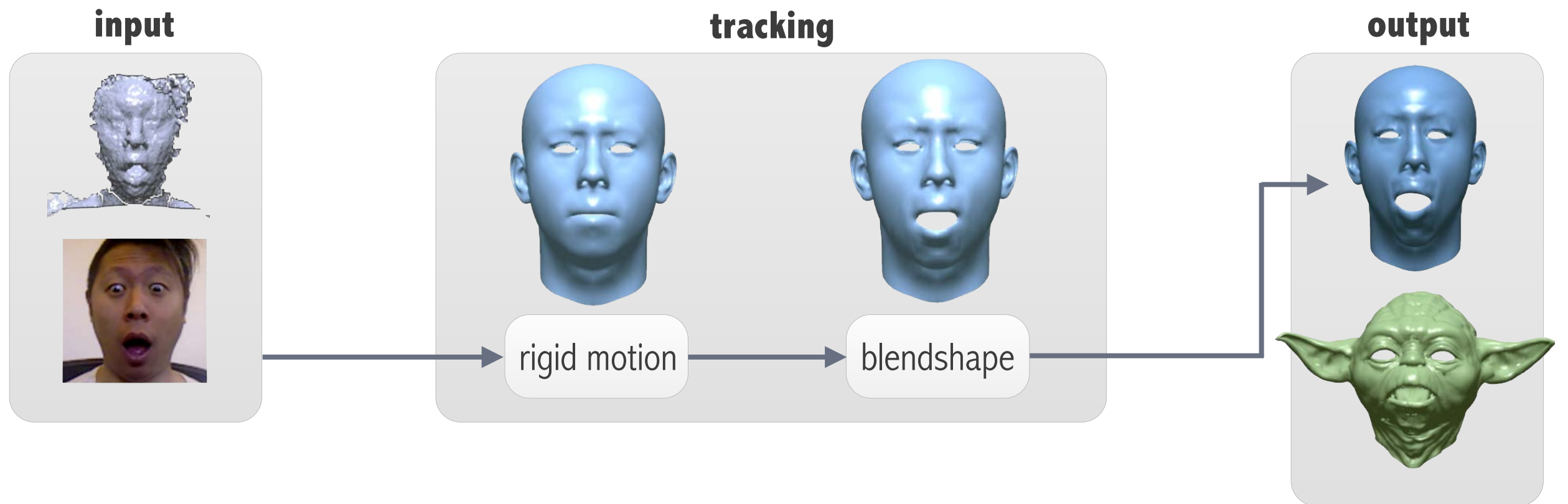




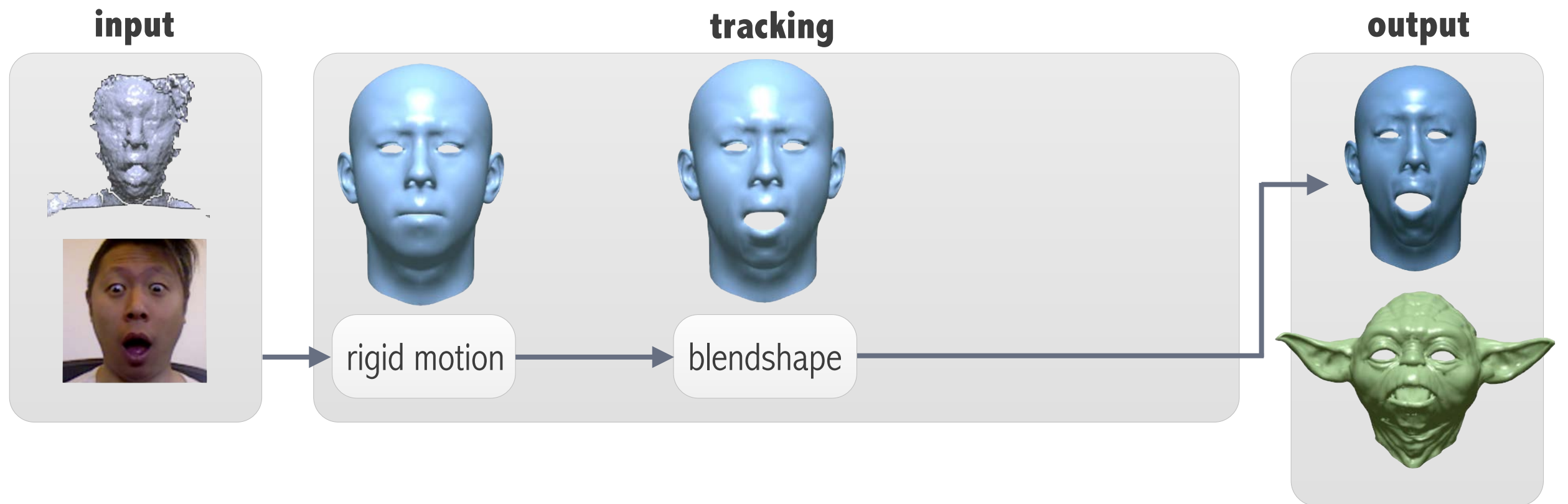
# Pipeline Overview



# Pipeline Overview

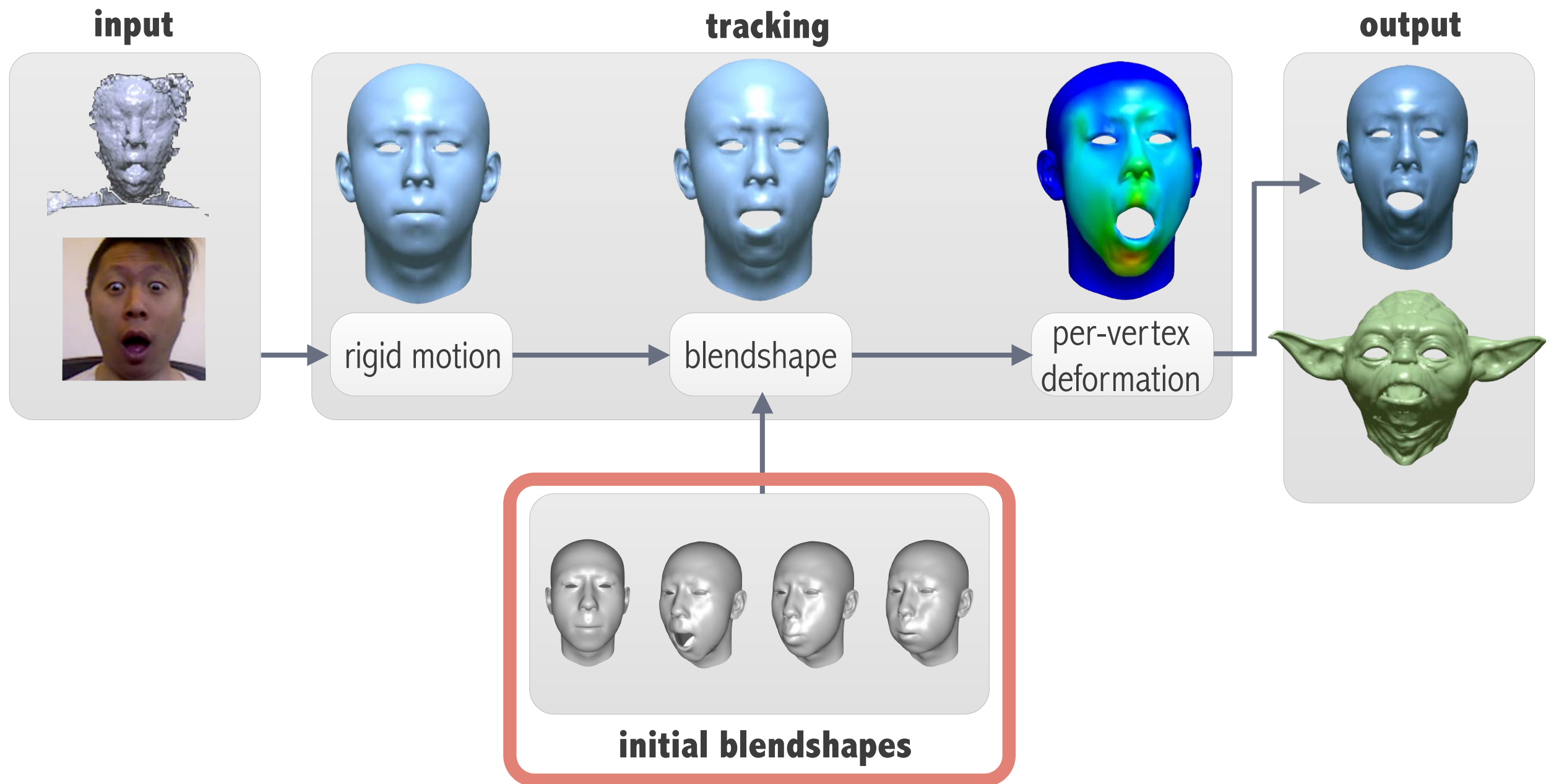


# Pipeline Overview

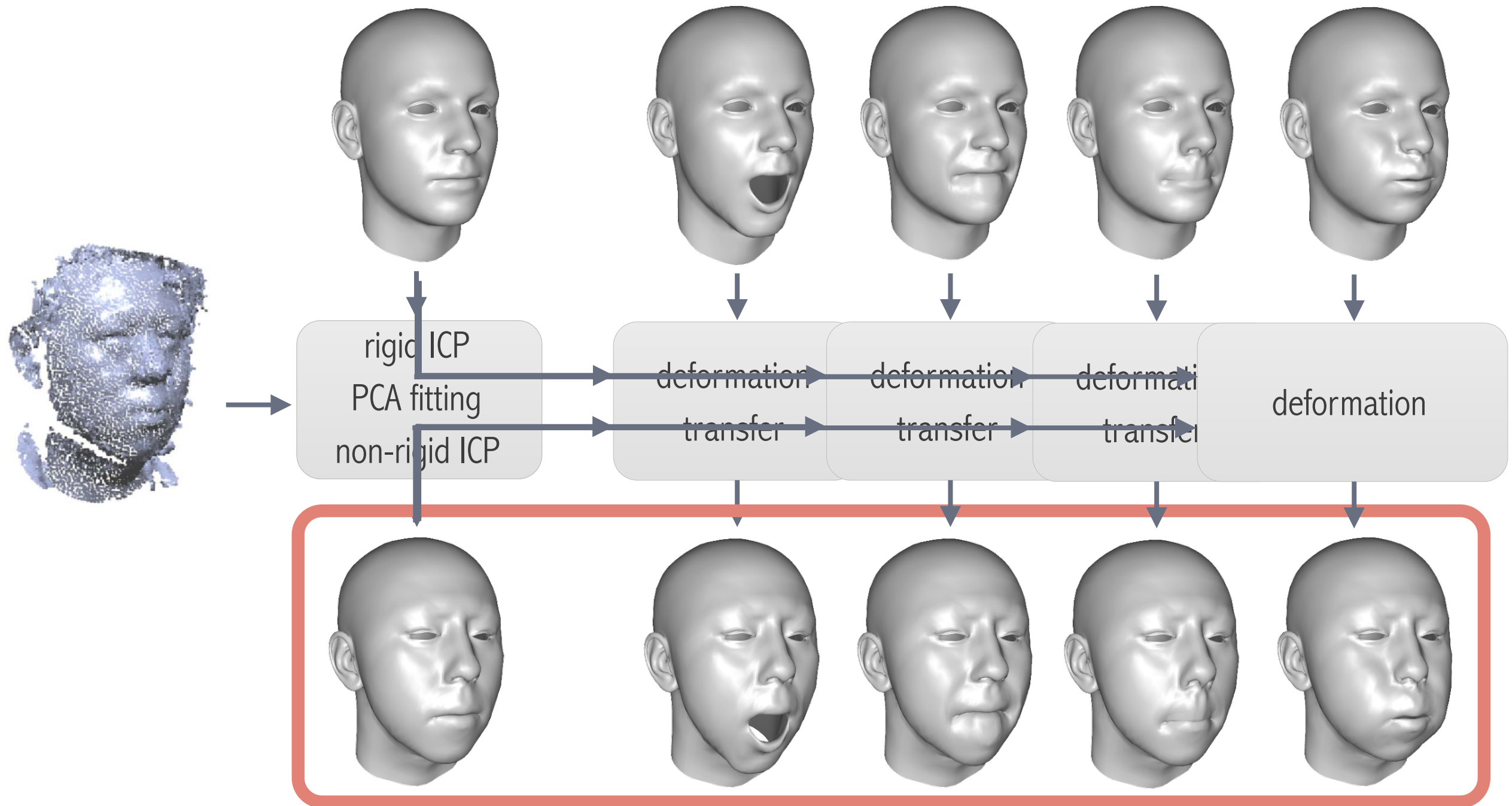




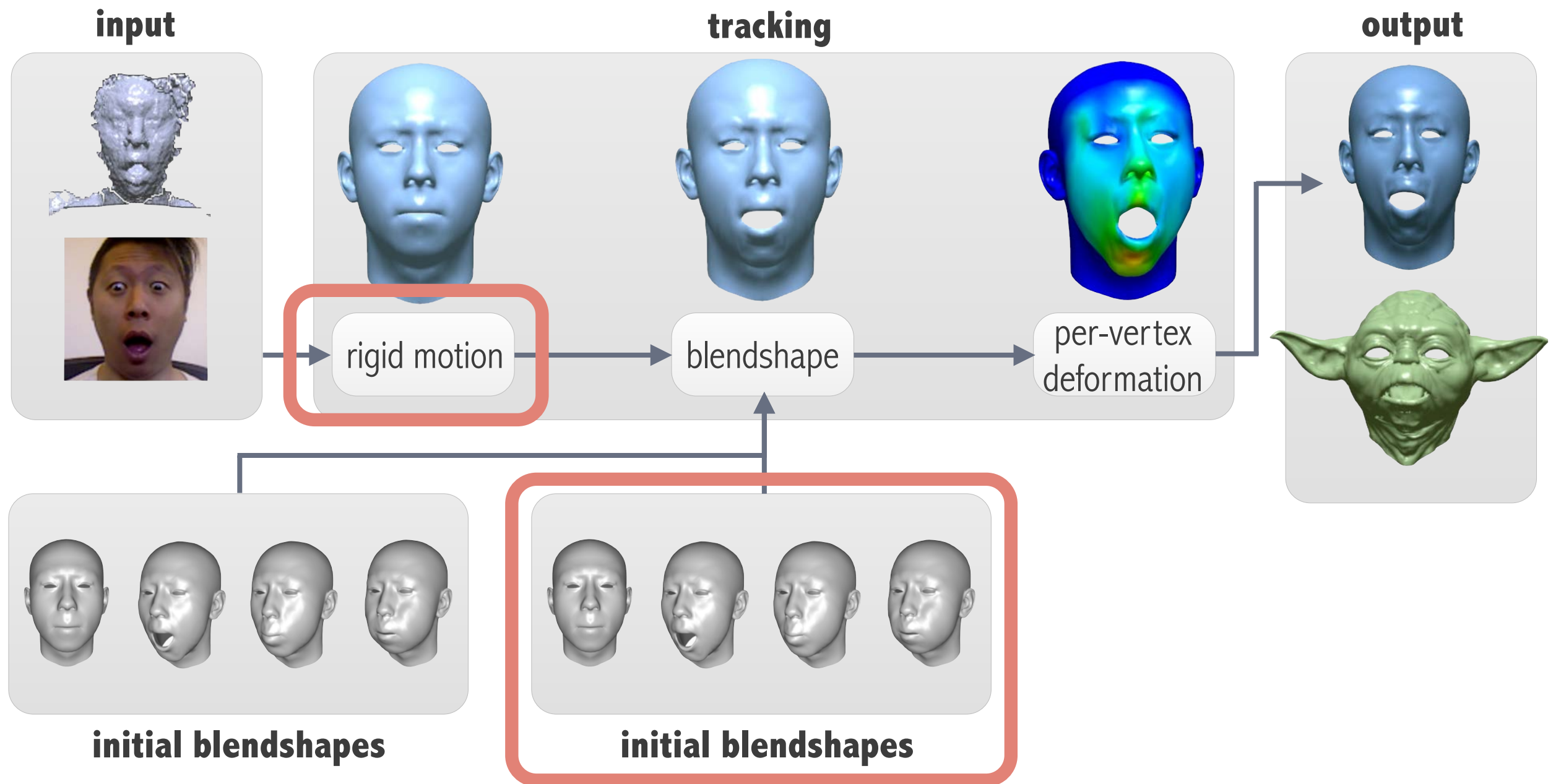
# Pipeline Overview



# Building Initial Blendshape Model

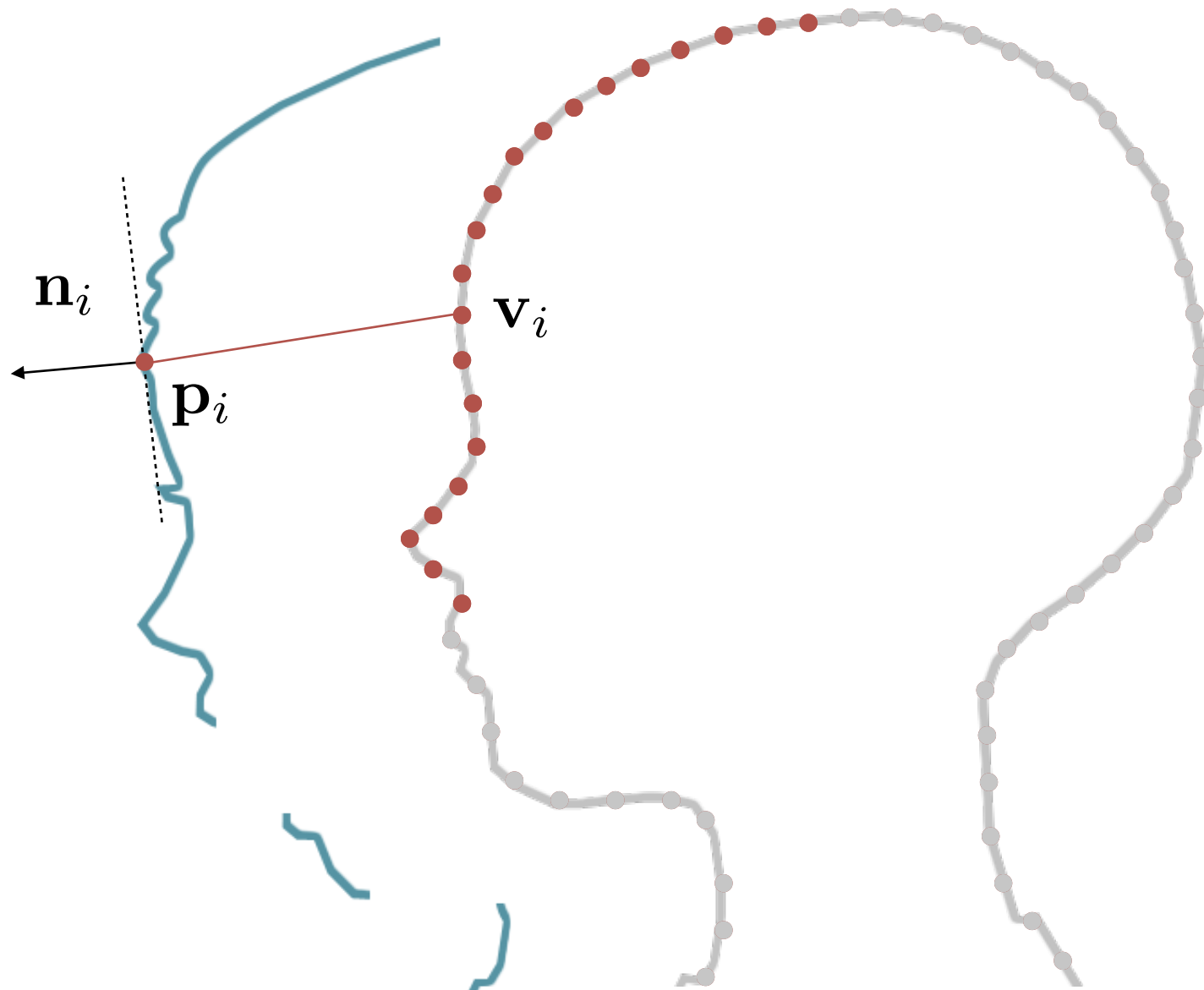


# Pipeline Overview





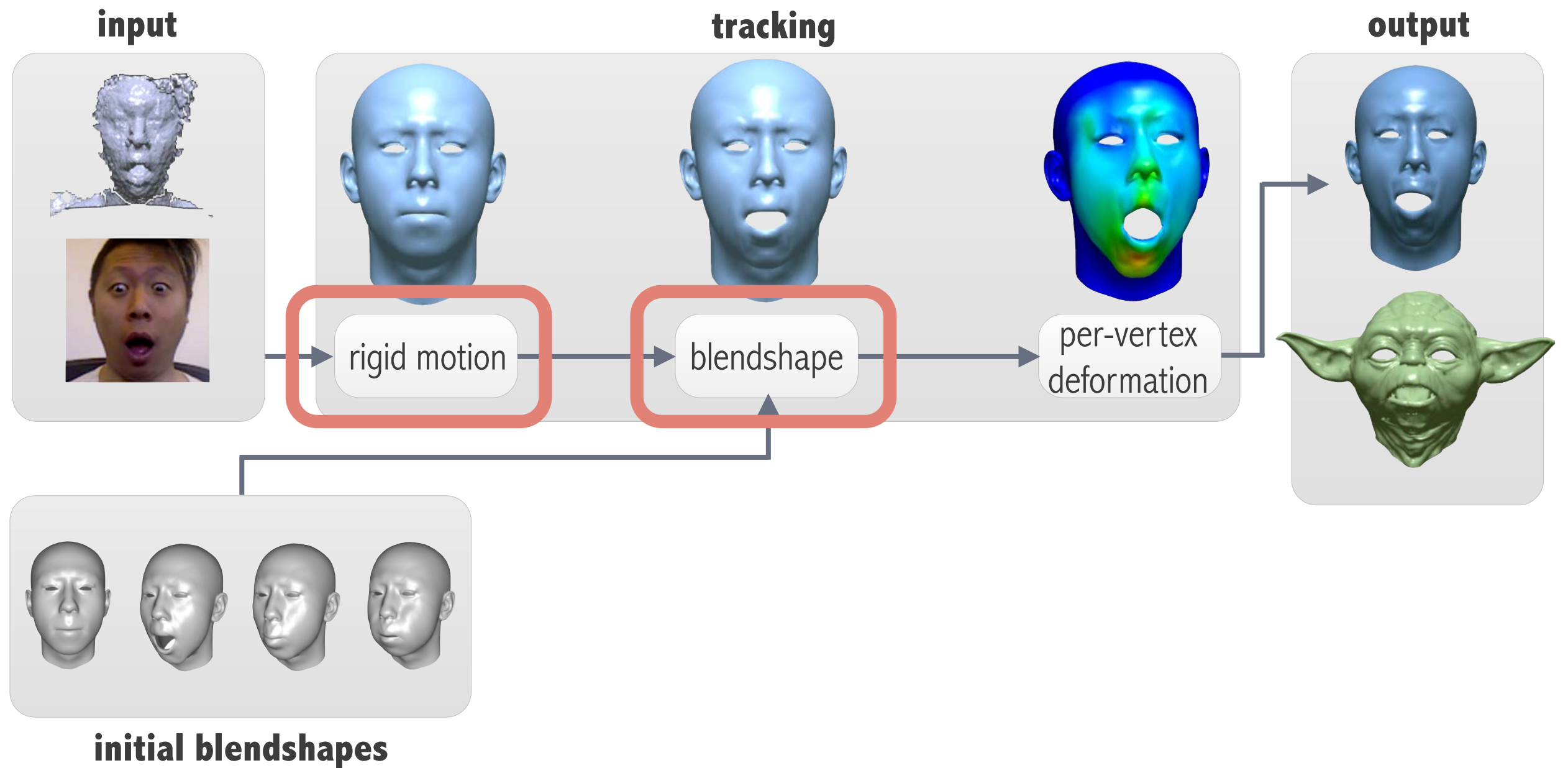
# Rigid Motion Tracking



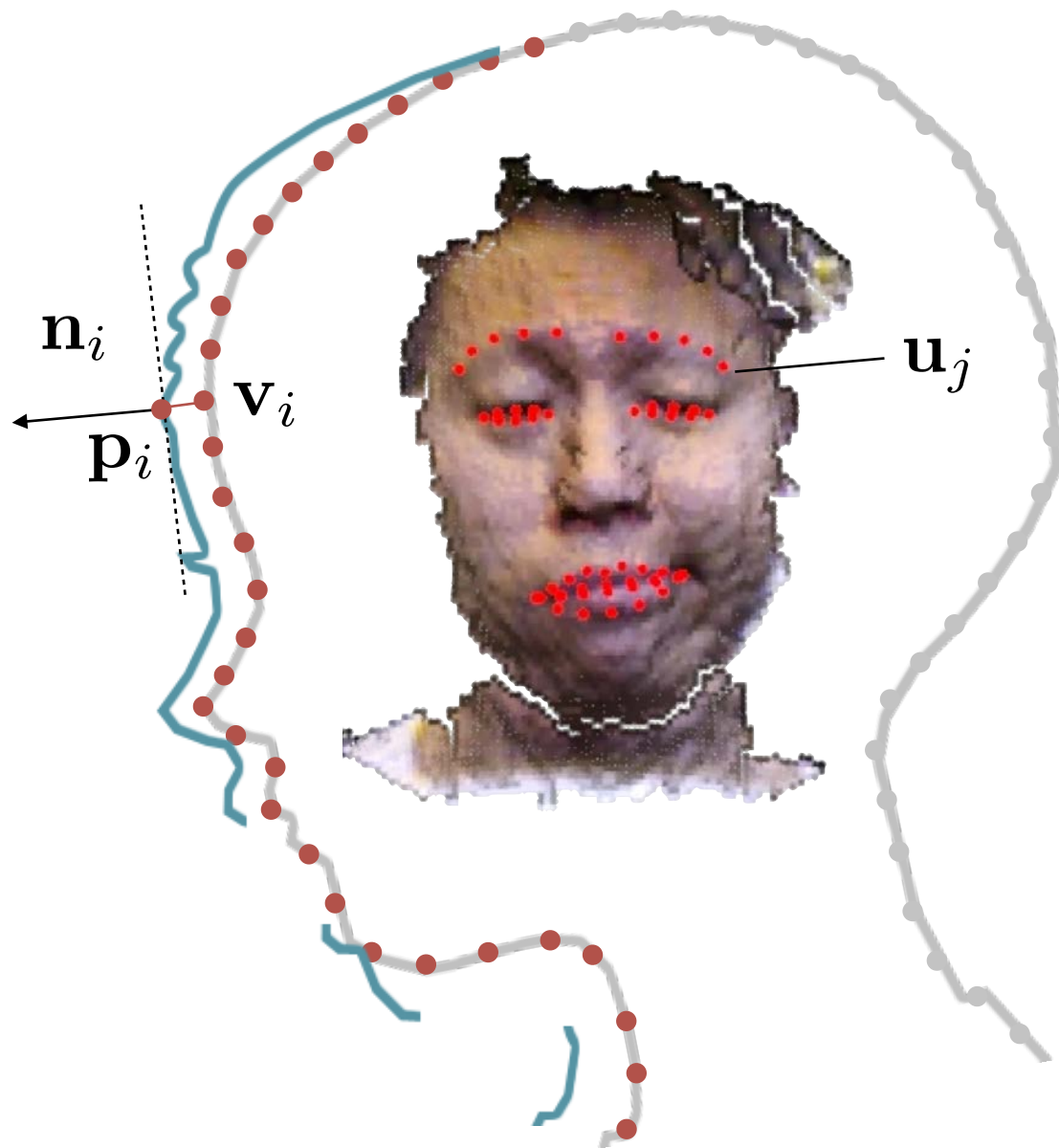
$$c_i^S(\mathbf{R}, \mathbf{t}) = \mathbf{n}_i^\top (\mathbf{v}_i(\mathbf{R}, \mathbf{t}) - \mathbf{p}_i)$$

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

# Rigid Motion Tracking



# Blendshape Tracking



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

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

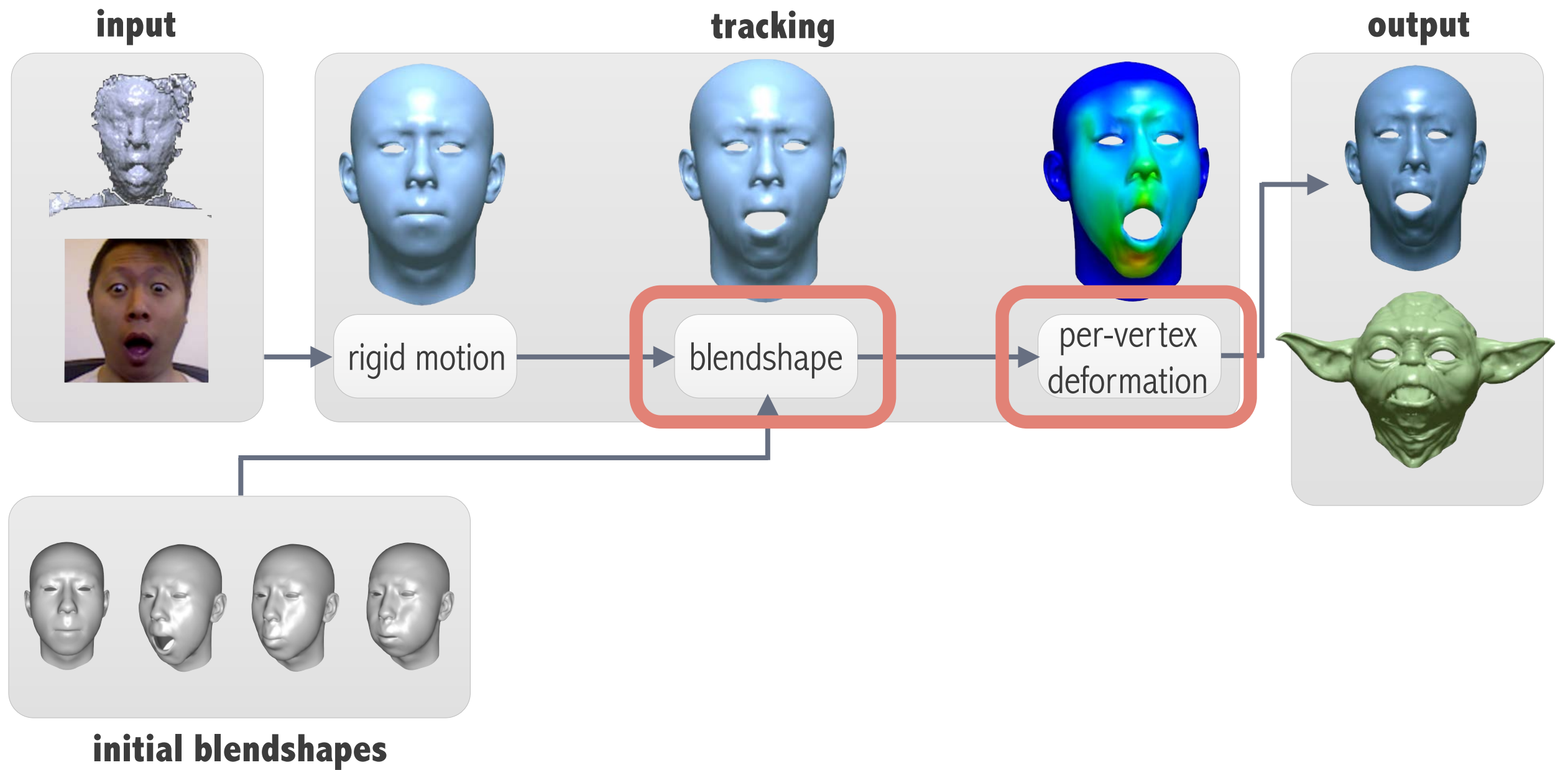
$$\mathbf{c}_j^F(\mathbf{x}) = \mathbf{H}_j(\mathbf{u}_j) \mathbf{P} \mathbf{v}_j(\mathbf{x})$$

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

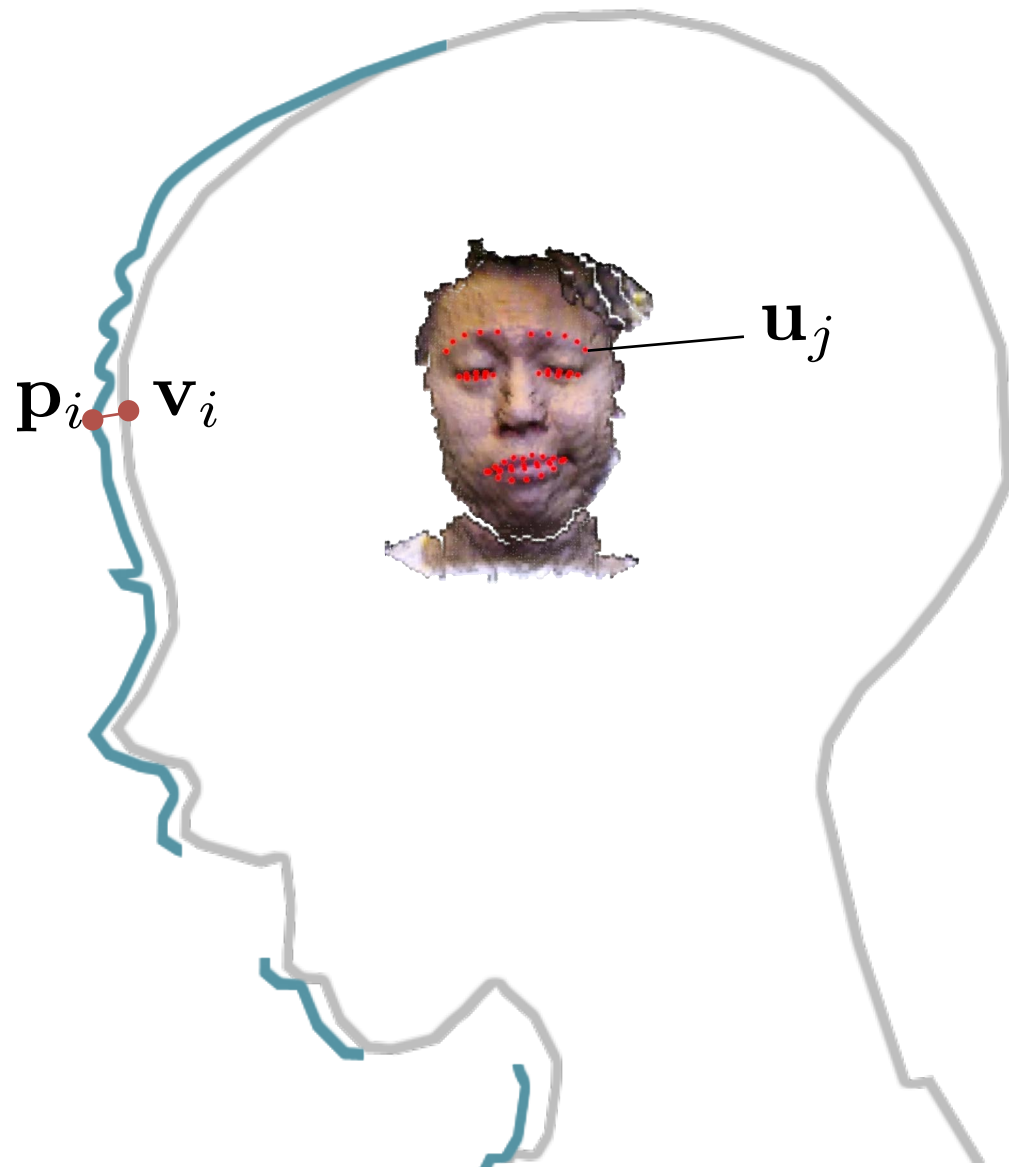
$$x_l \in [0, 1]$$



# Pipeline Overview



# Per-Vertex Deformation



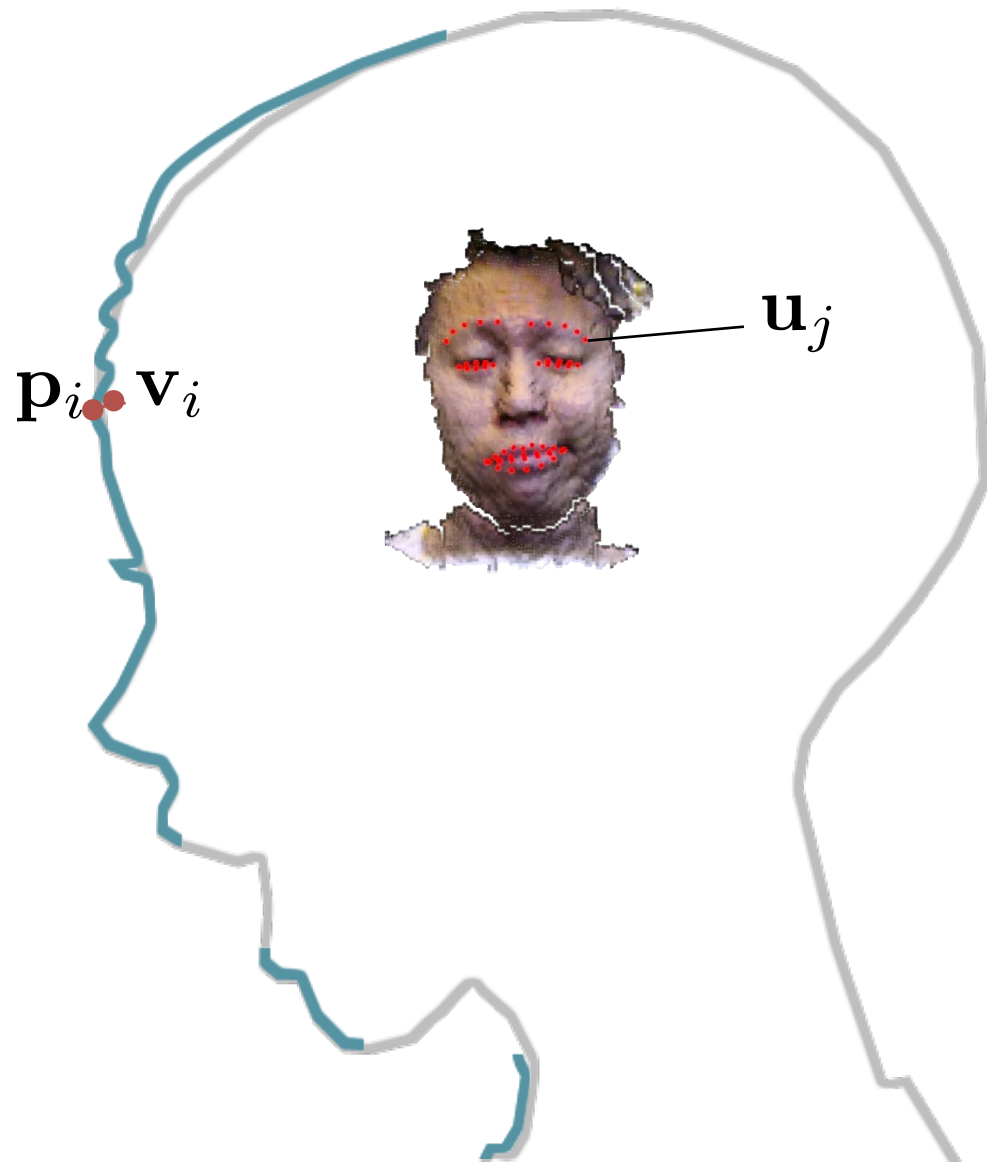
$$\mathbf{c}_i^{\mathbf{P}}(\Delta \mathbf{v}_i) = (\mathbf{p}_i - \mathbf{v}_i) - \Delta \mathbf{v}_i$$

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

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

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

# Fast Laplacian Deformation



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

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

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

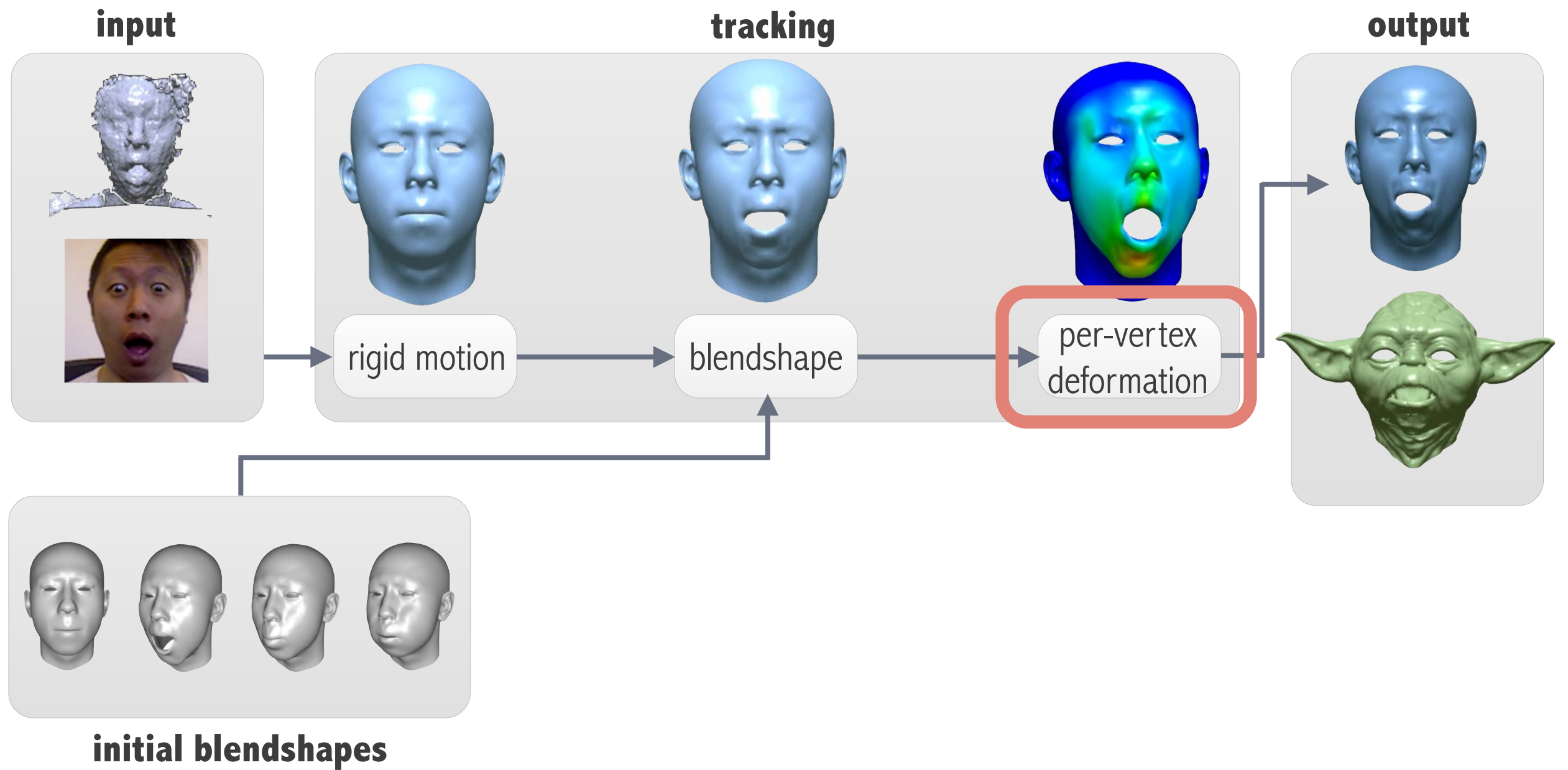
sparse  
constant in time

pre-factorized  
constant in time

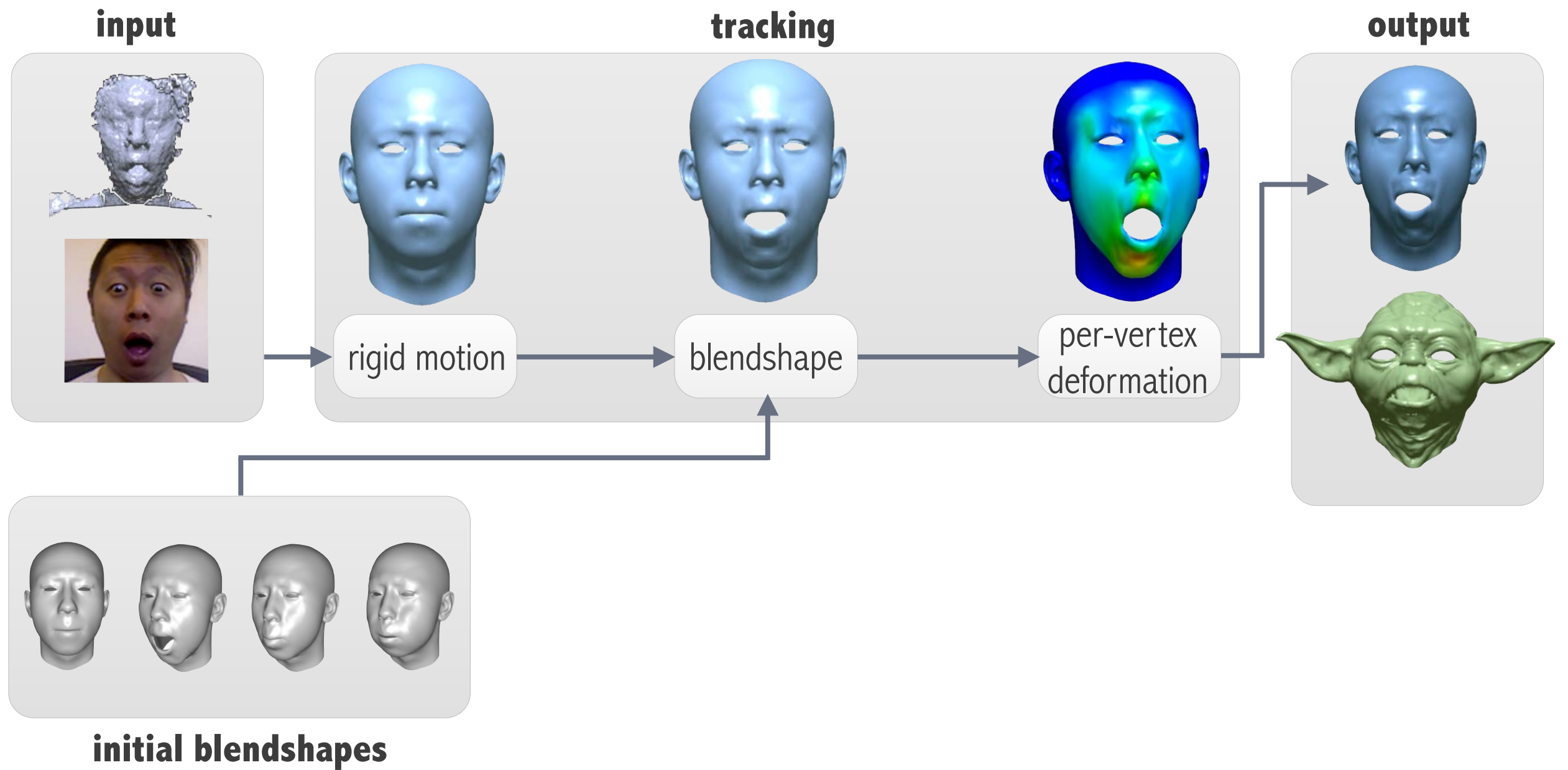
sparse  
trivially inverted



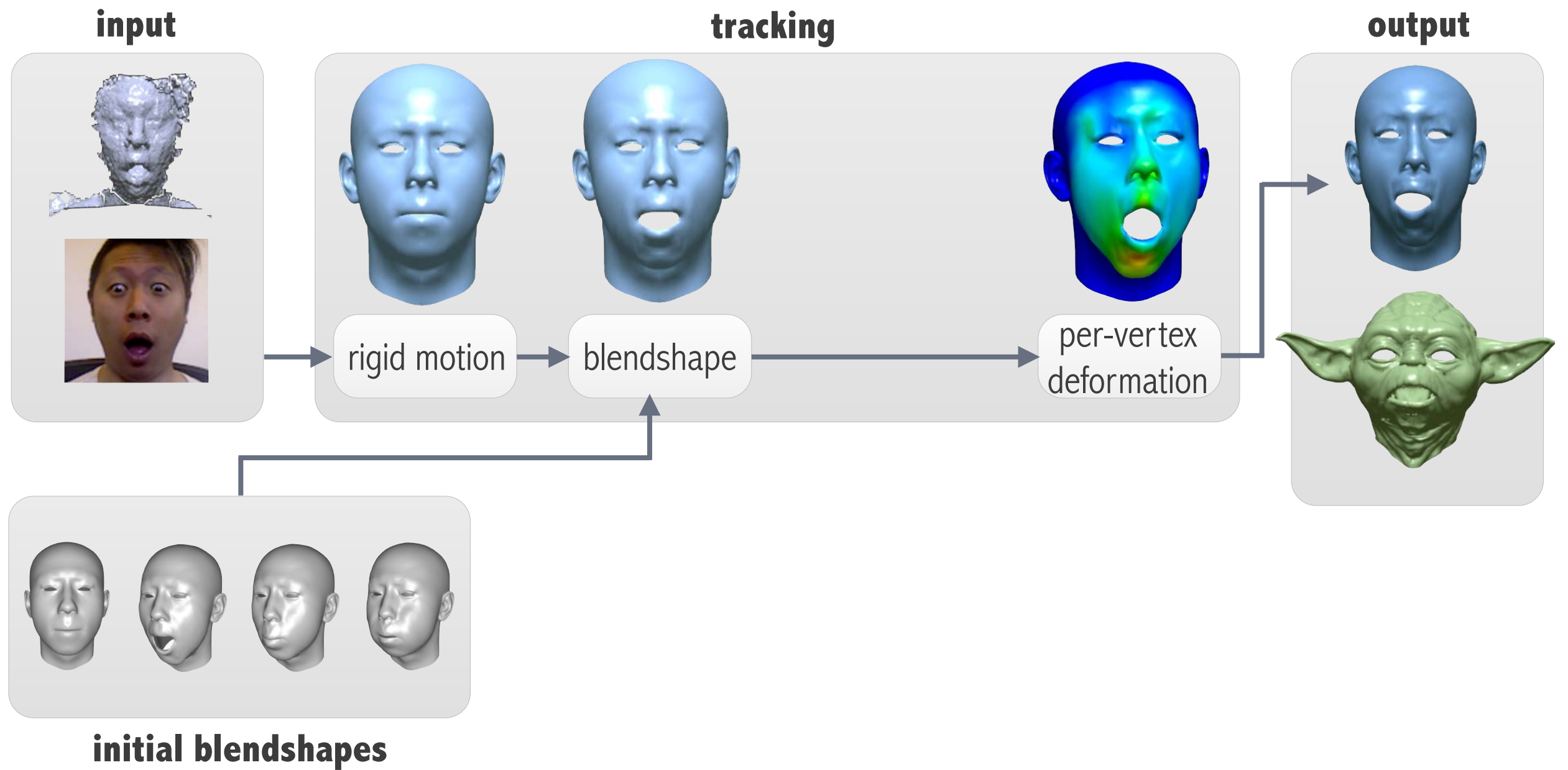
# Pipeline Overview



# Pipeline Overview

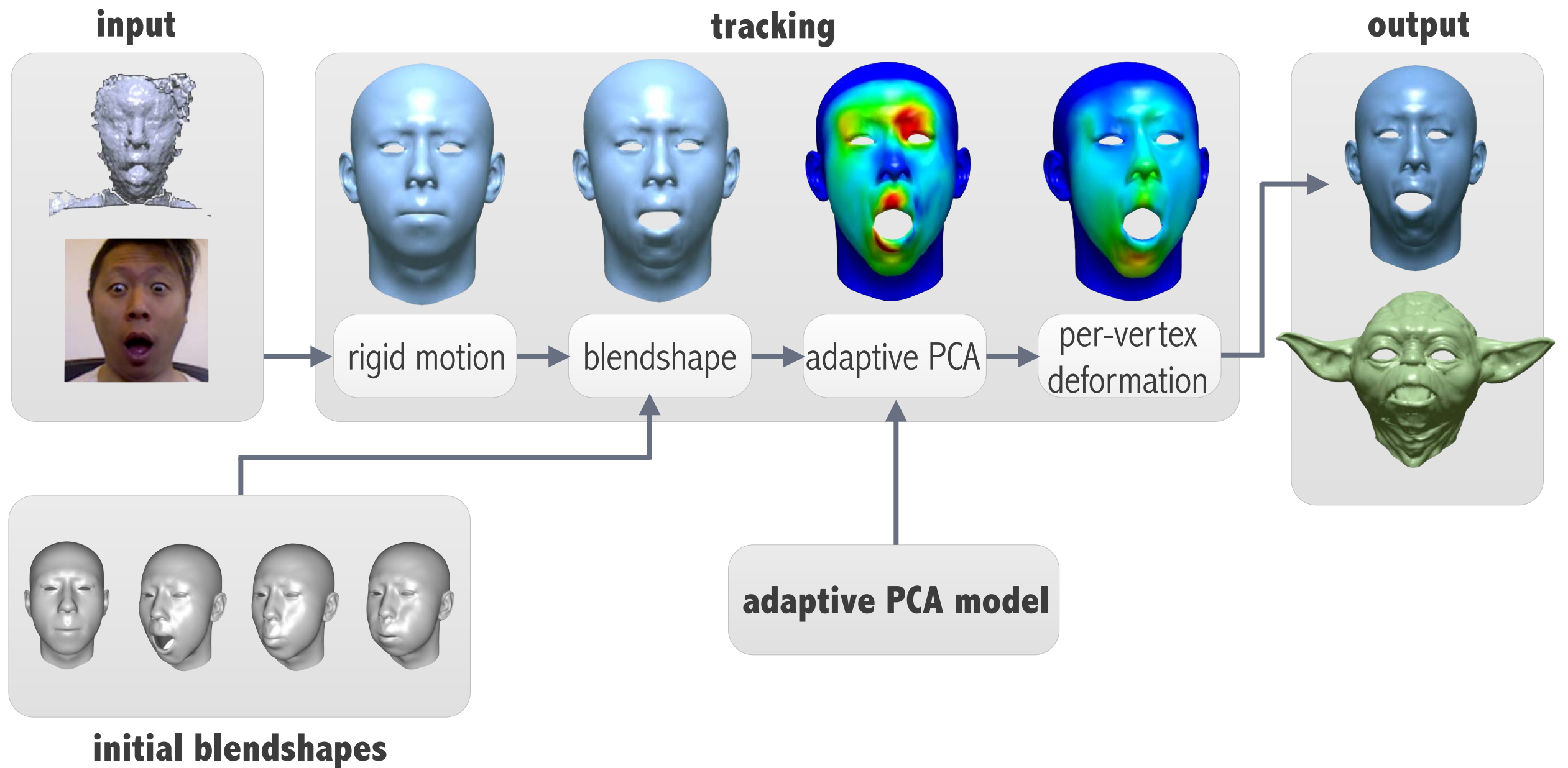


# Pipeline Overview

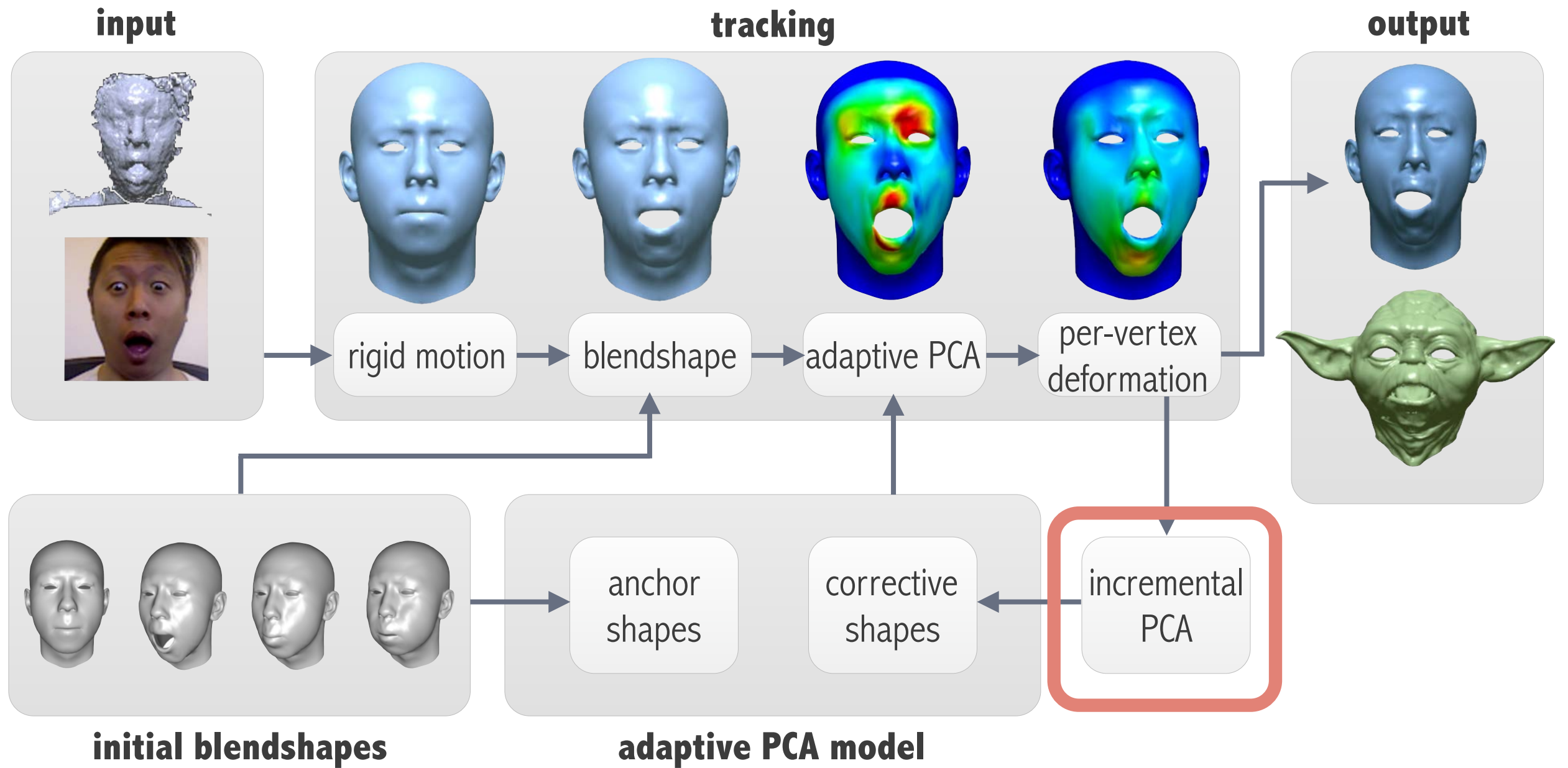




# Pipeline Overview



# Pipeline Overview



# Tracking Comparison

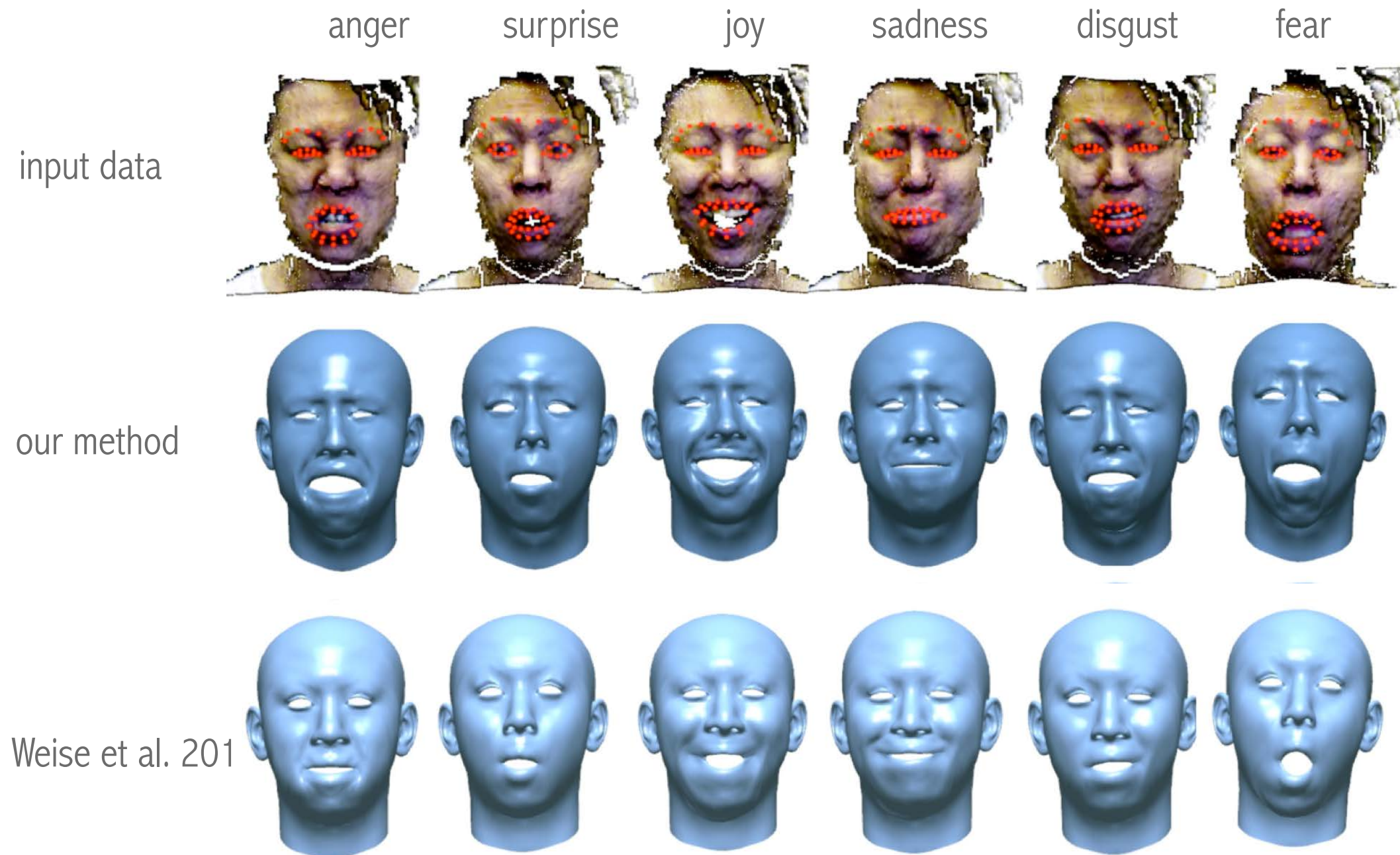


depth map &  
2D features



# Tracking Basic Emotions

Li et al. SIGGRAPH 2013



# Faces 2014: Discussion

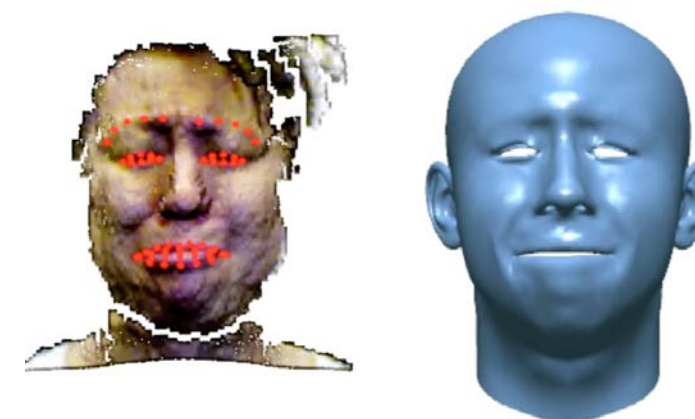
Bouaziz et al. 2013



Cao et al. 2014



Li et al. 2013



input 3D/2D

no calibration

blendshape

input 2D

neutral face

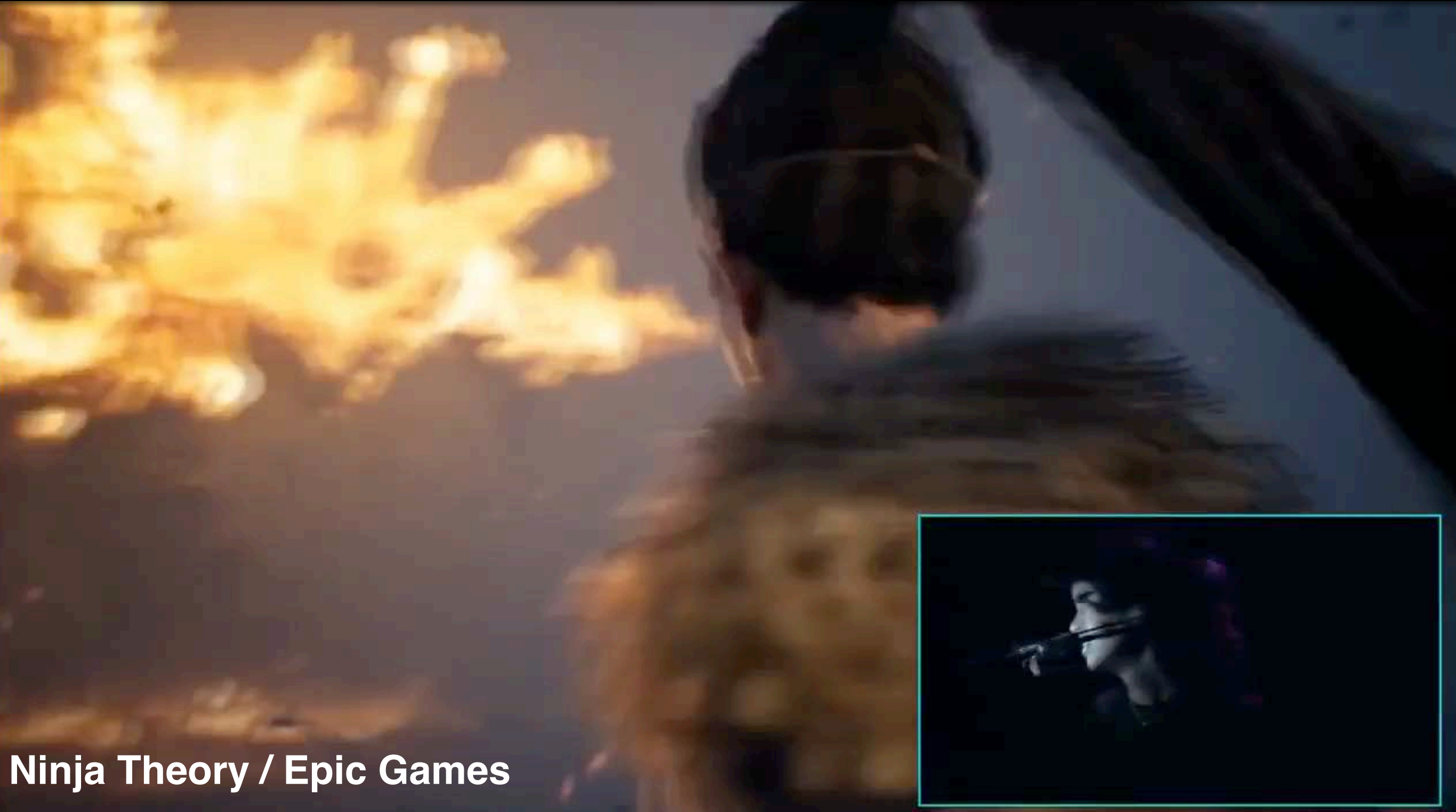
blendshape

input 3D/2D

neutral face

per-vertex deformation

# **Into the Mainstream**

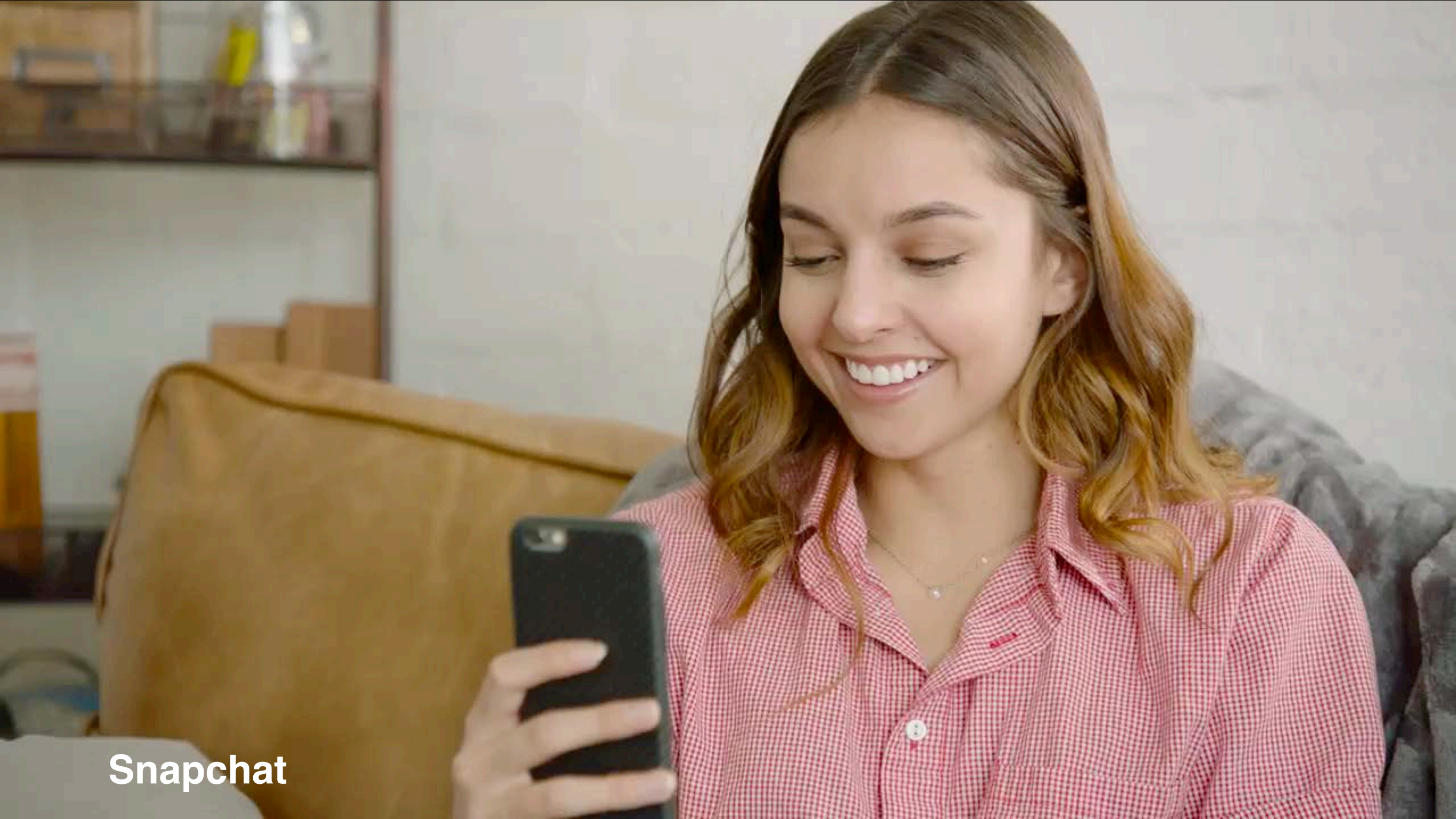


**Ninja Theory / Epic Games**



YourN p





**Snapchat**





MSQRD

Facebook / MSQRD

# FaceX Robustness



input video



face segmentation



# FaceX Instant User Switching



input video



face segmentation

# FaceX Kids



input video



face segmentation

# State-of-the-Art in Real-Time Facial Tracking





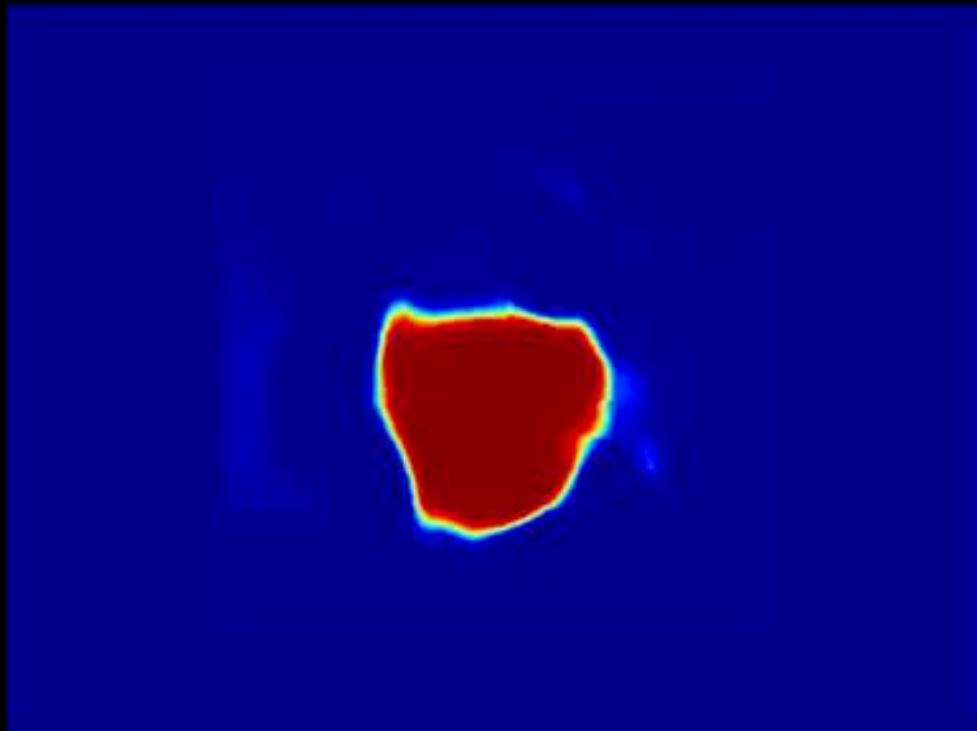
# Preliminary Findings: Segmentation



input video



facial segmentation/tracking

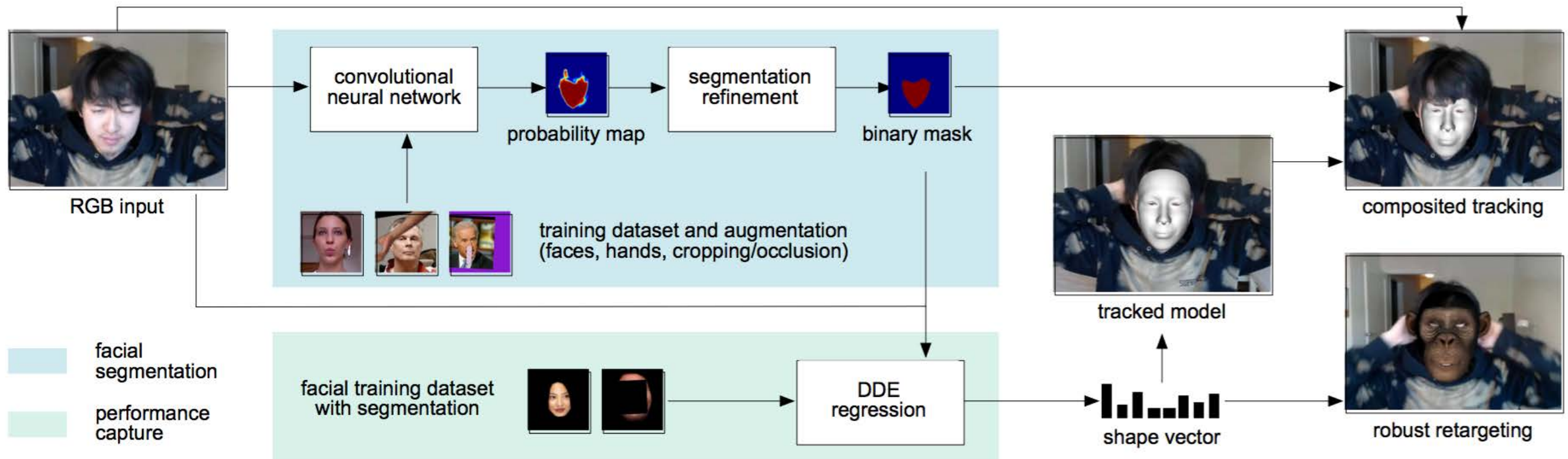


probability map

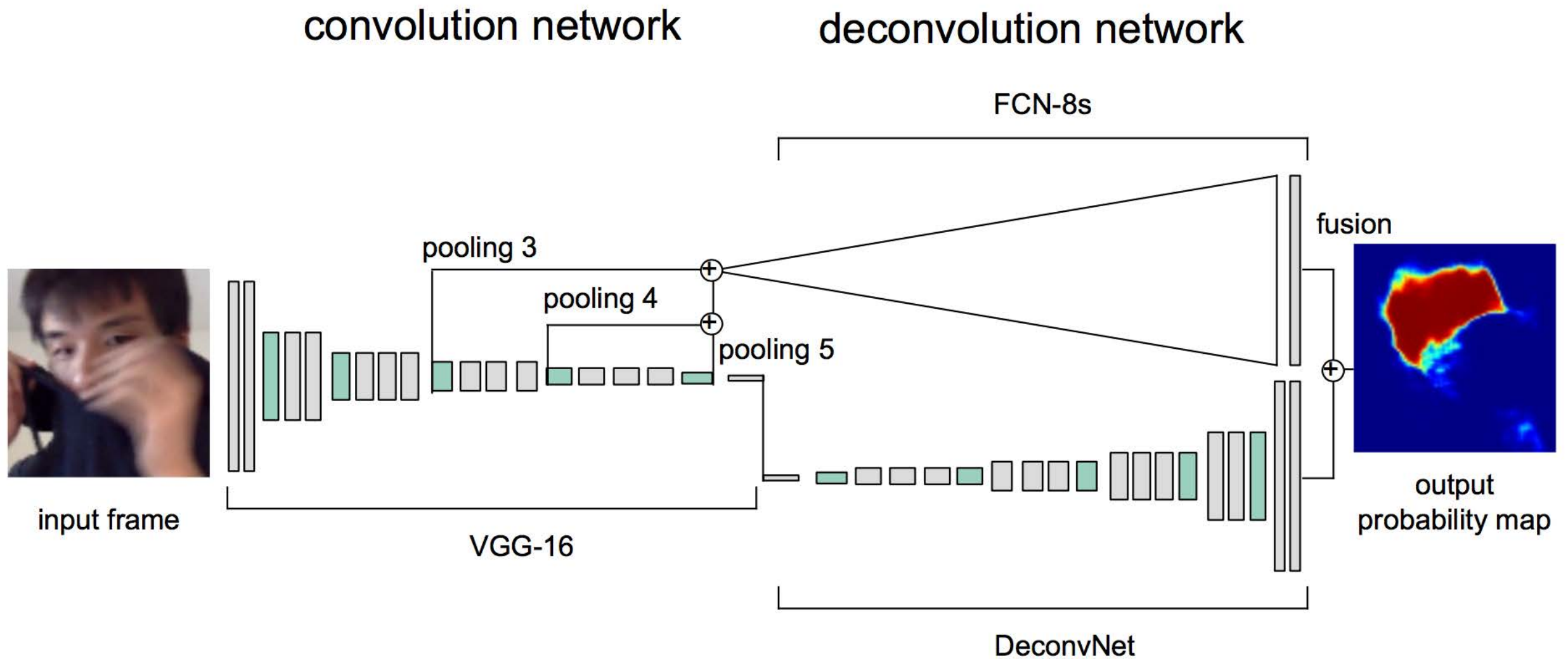


composited result

# Pipeline



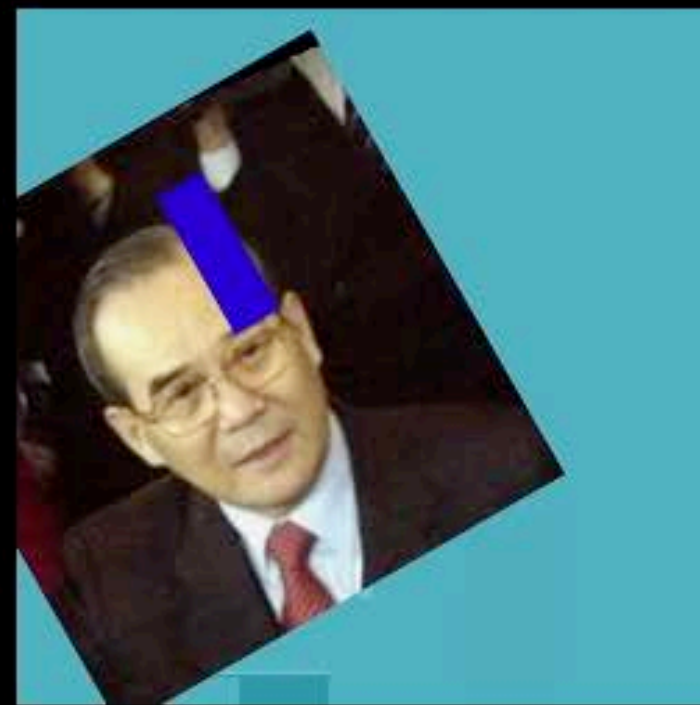
# Two-Stream Deconvolution Network







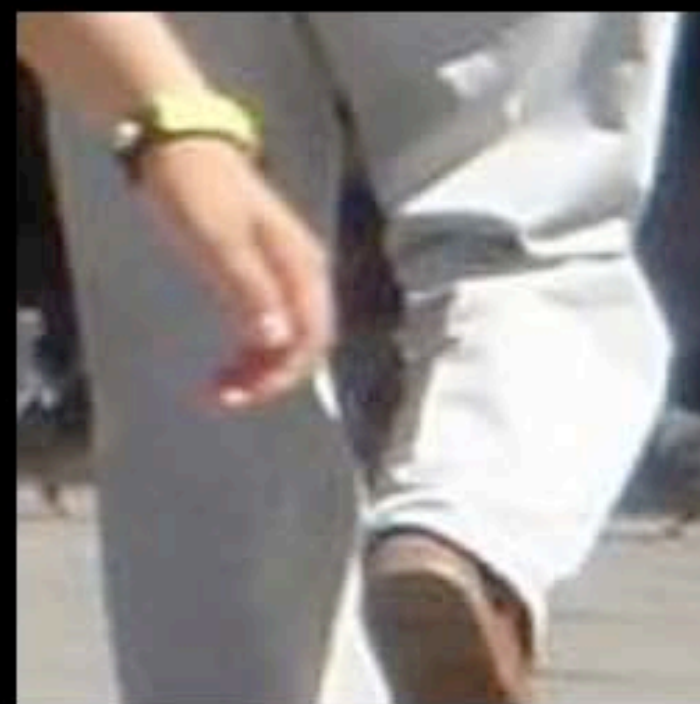
face data



occlusion / cropping



hand over face compositing



negative hand data

# Comparison



input video



Cao et al. 2014



our method

# 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







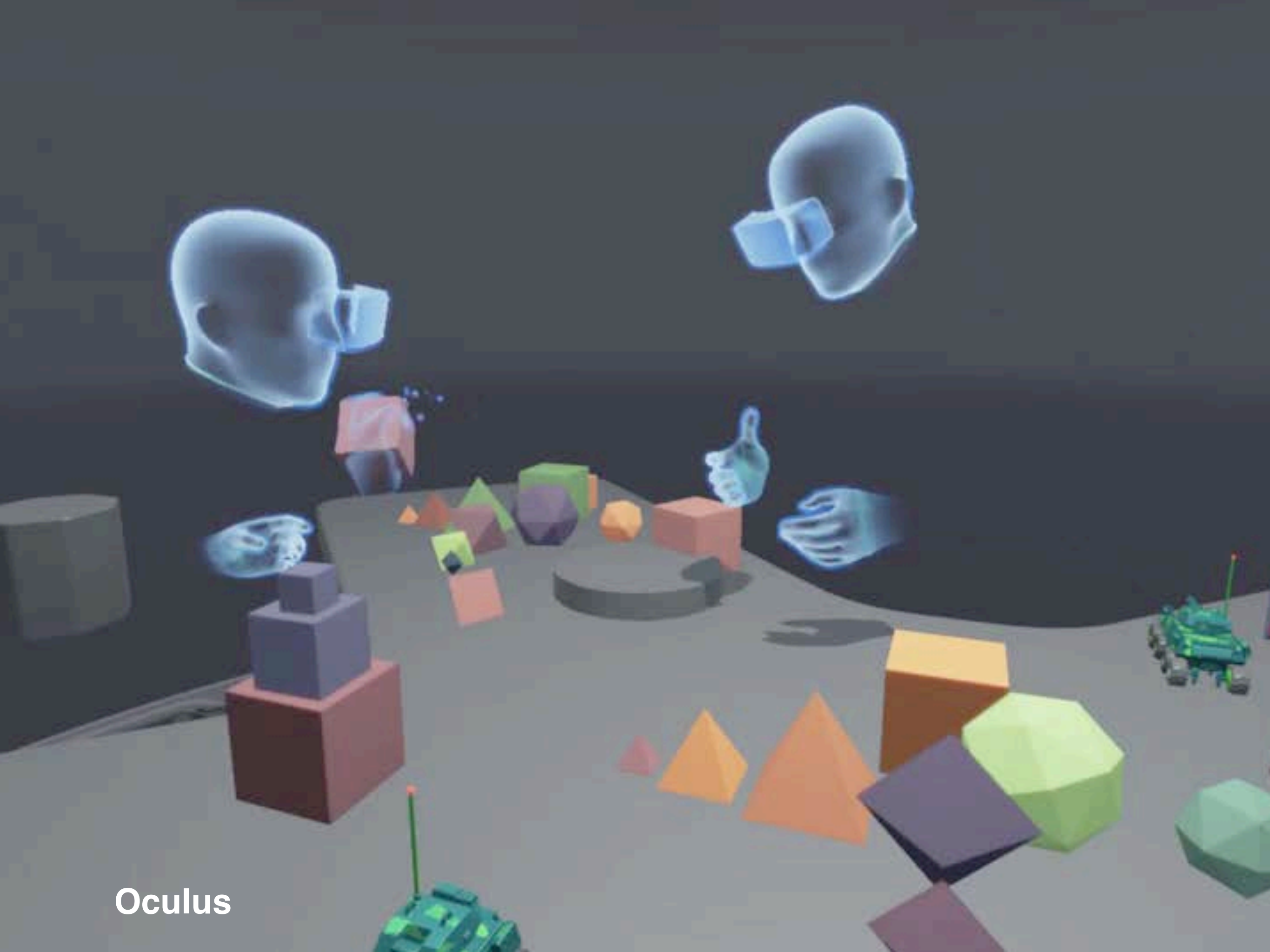


# **NextGen Communication Platform**

# Online Virtual Worlds





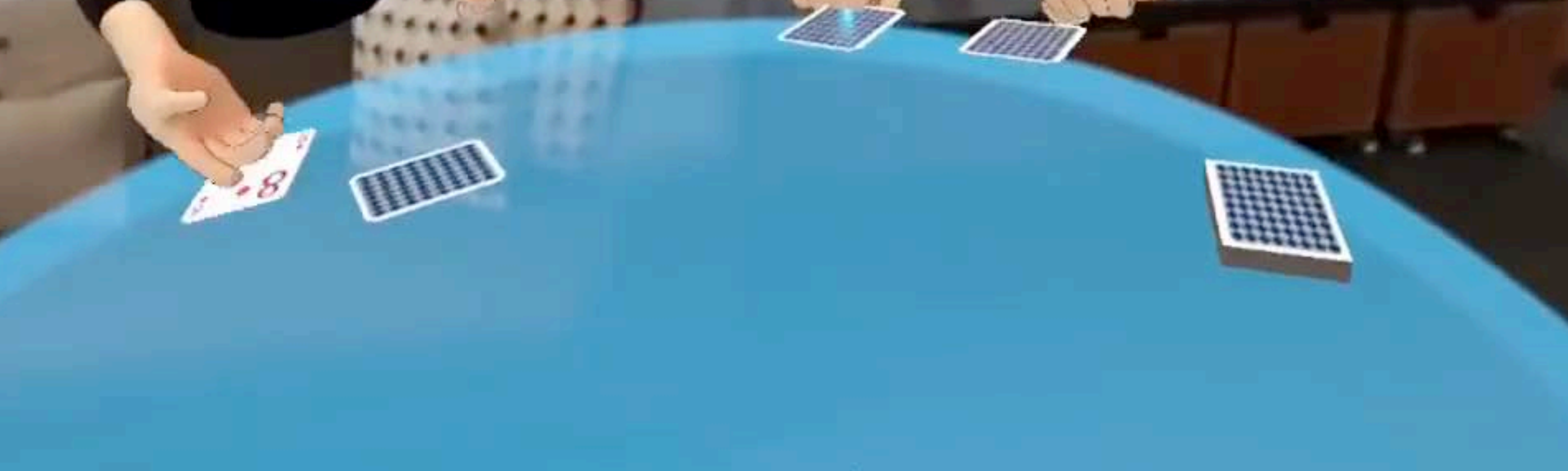


Oculus



Lucy

Michael





# Virtual Training, Simulation, Education





# Challenges



2016



2021

- Accurate simulation/training
- Building trust, communicate feelings, resolving conflicts



# Occlusions

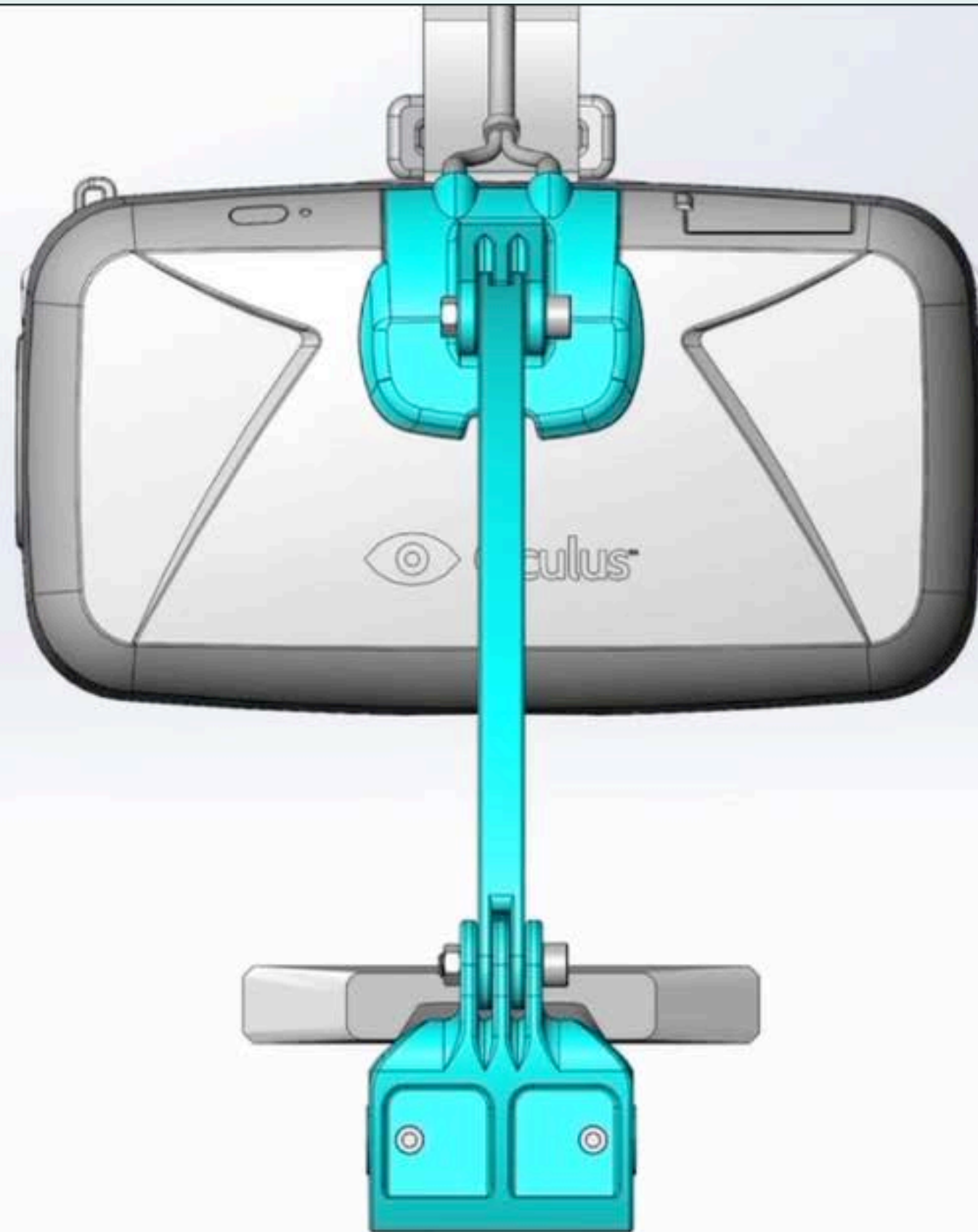




# Preliminary Findings: Segmentation



# Facial Performance Sensing HMD

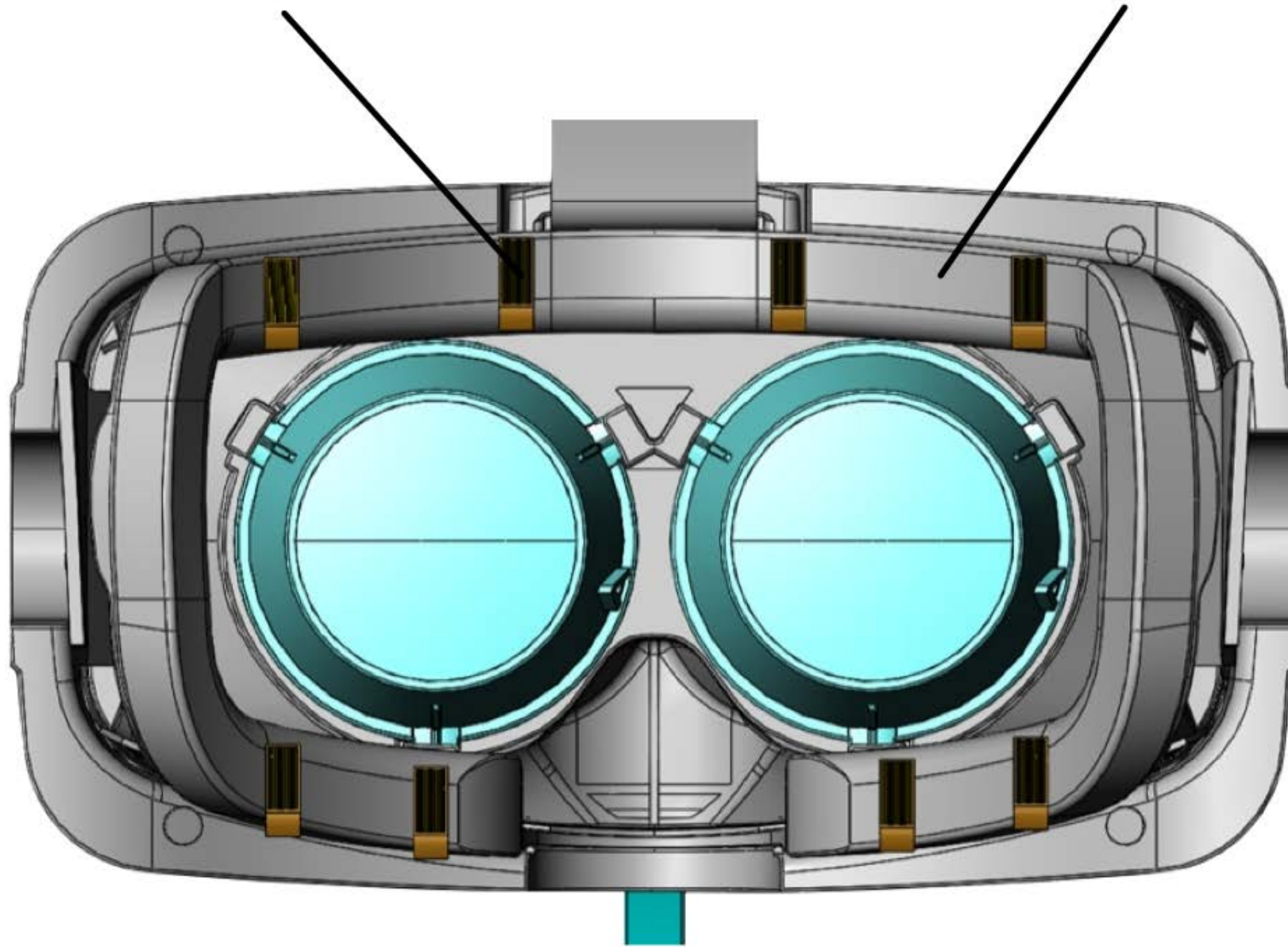




# Facial Performance Sensing HMD

strain sensors

foam liner



interior  
(CAD model)

# Facial Performance Sensing HMD



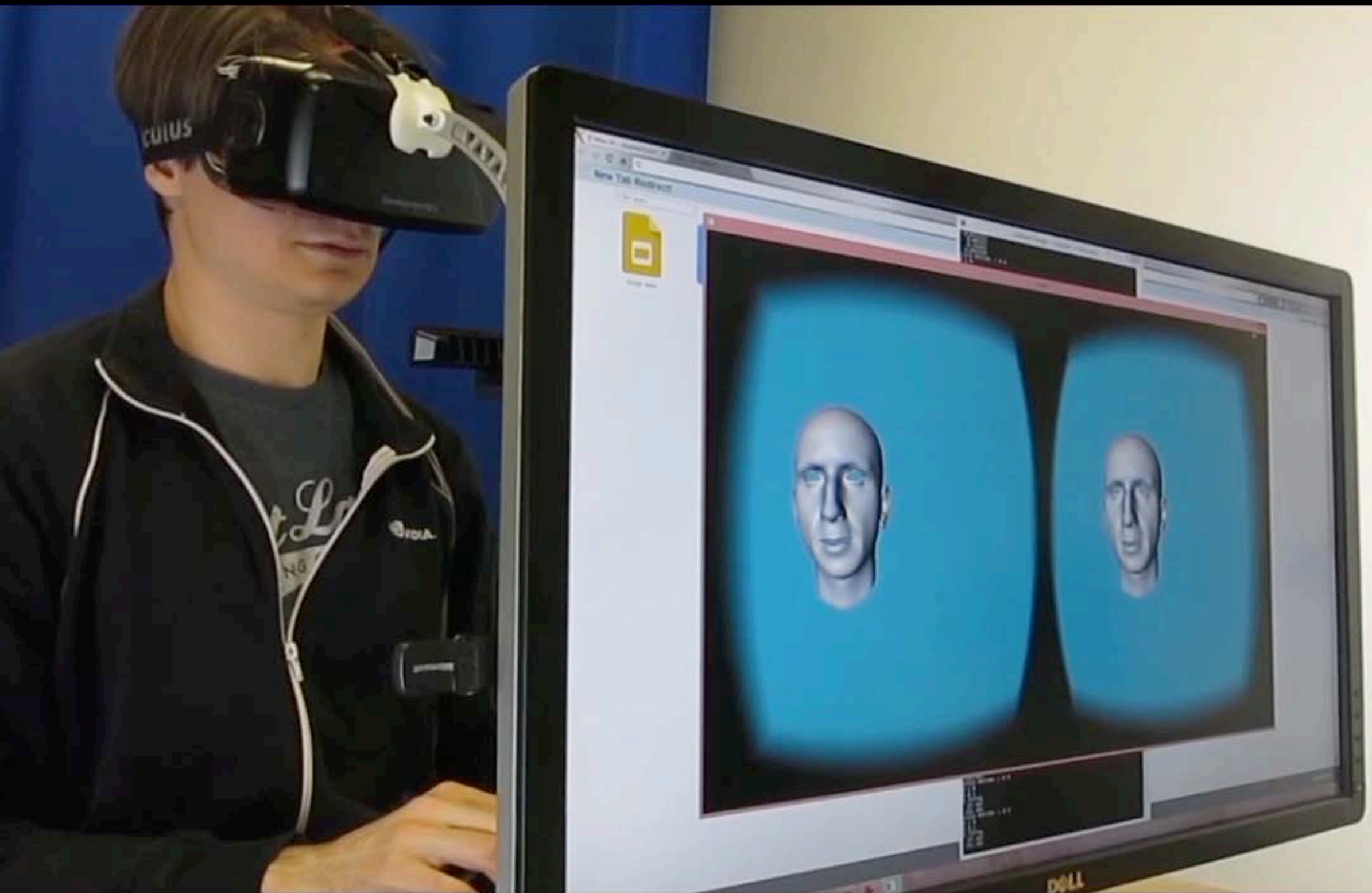


# Ultra Thin Flexible Electronic Materials

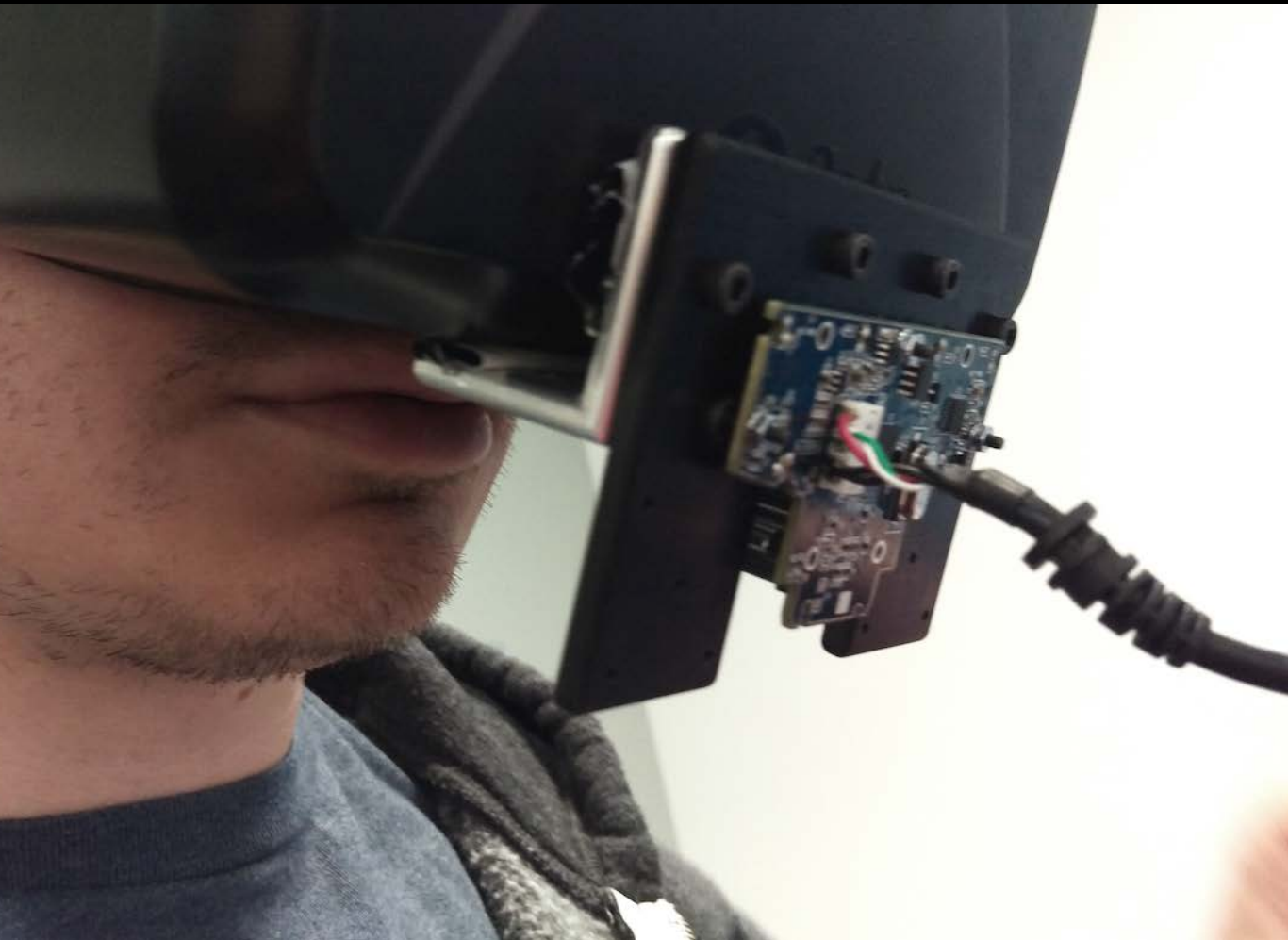




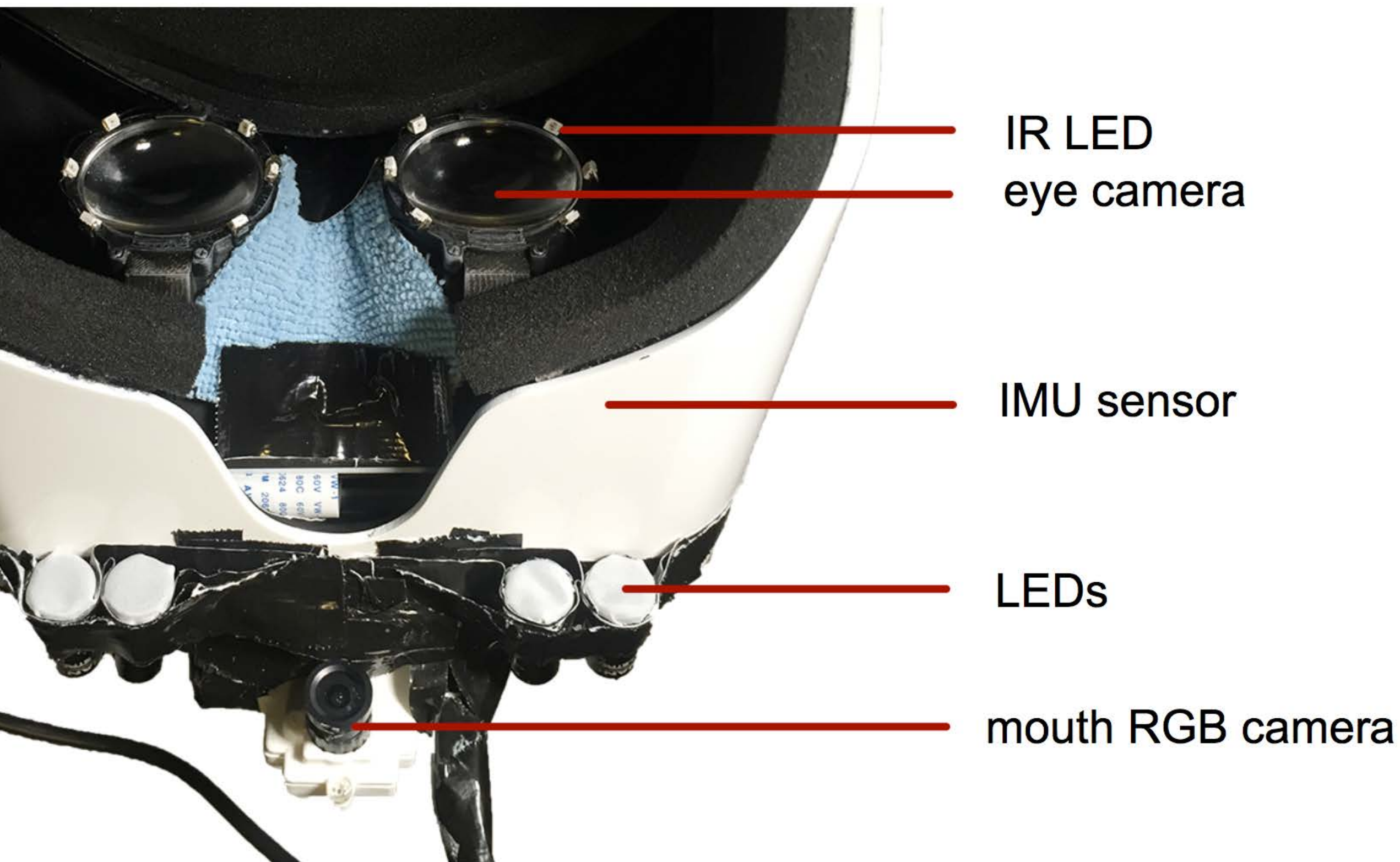
# Live Demo



# HMD Prototype Design







IR LED  
eye camera

IMU sensor

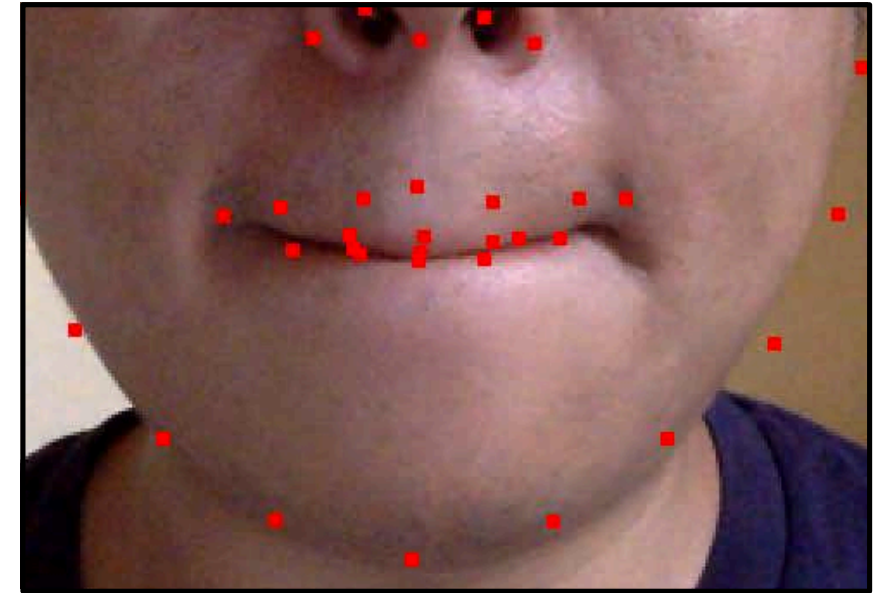
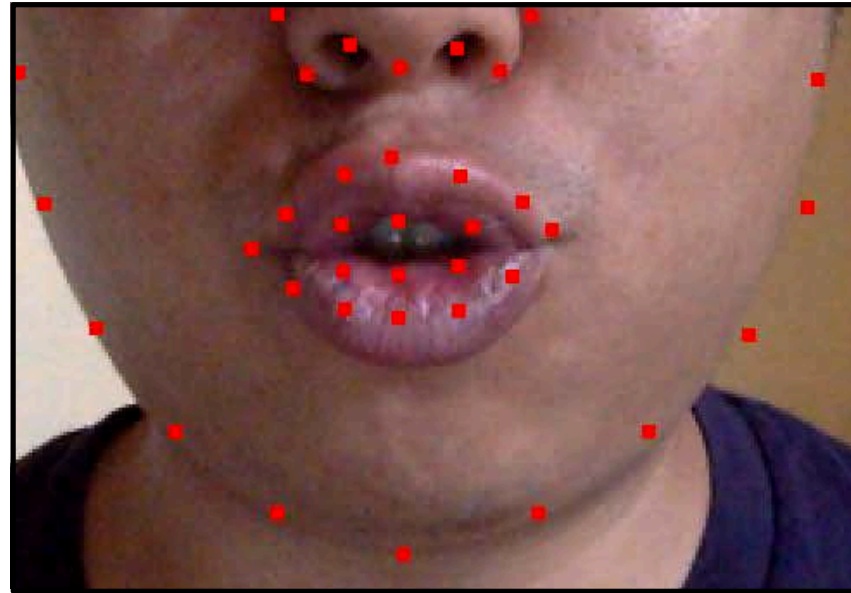
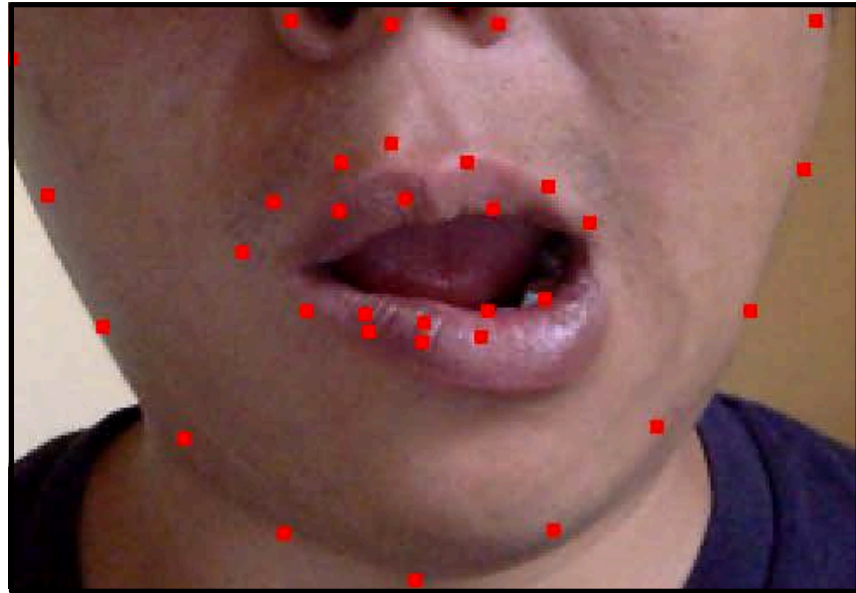
LEDs

mouth RGB camera

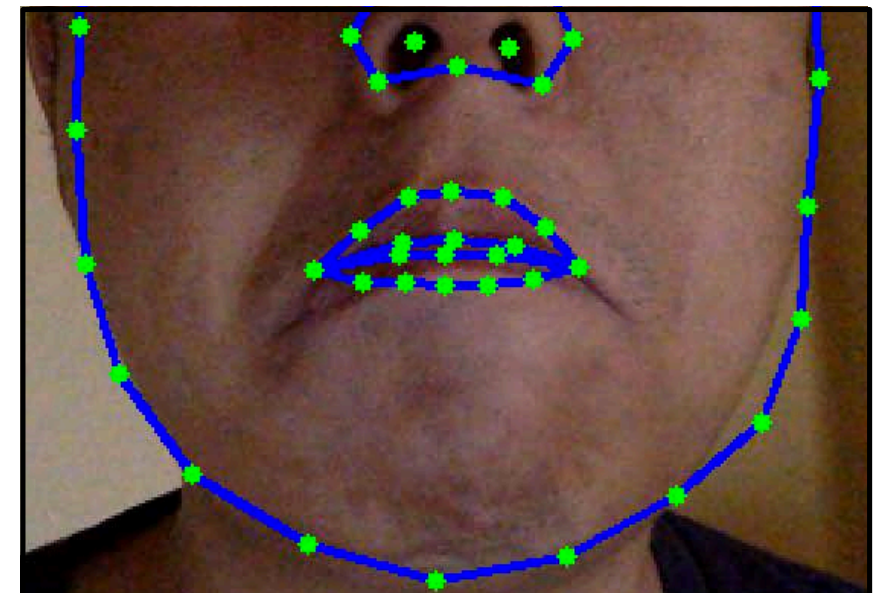
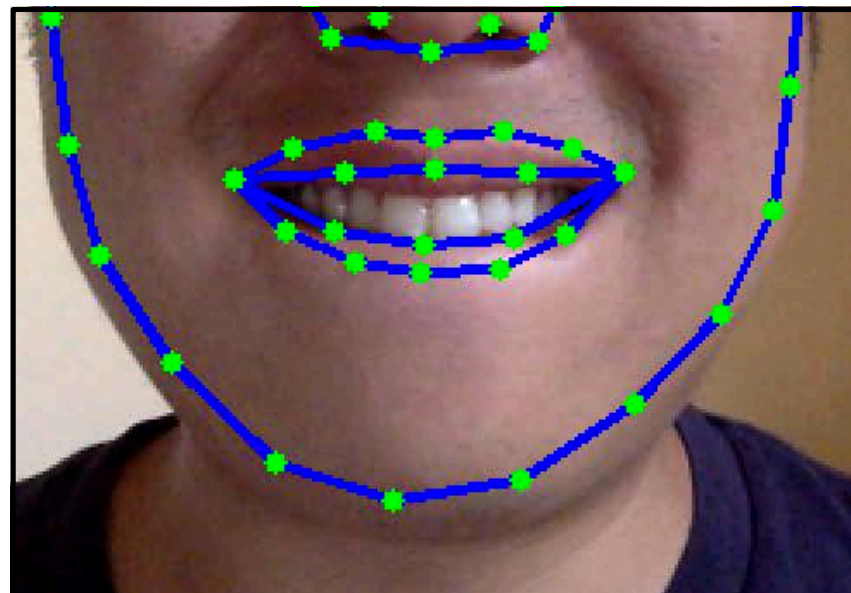
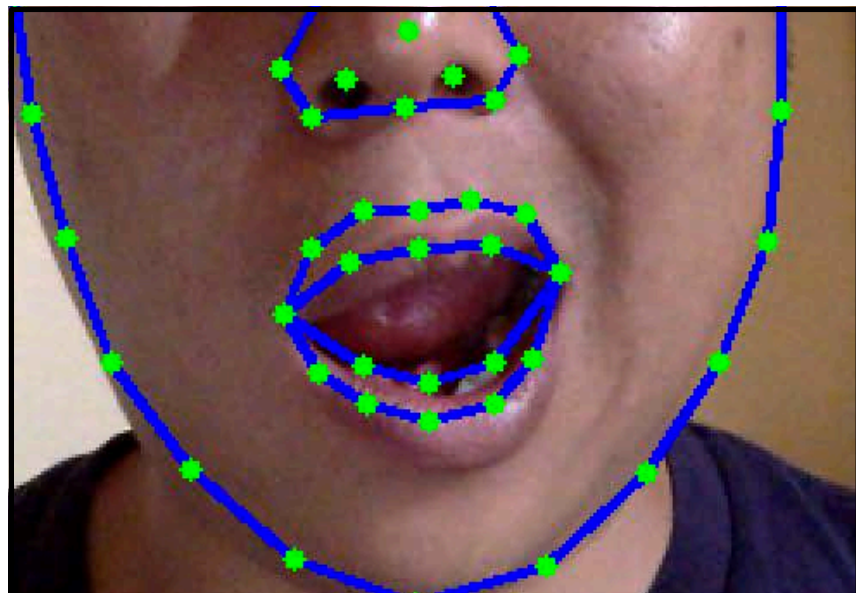




# Feature-Based Tracking



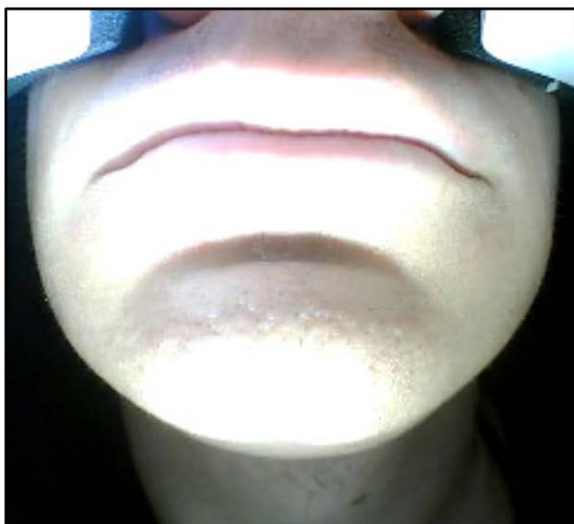
Kazemi and Sullivan, CVPR 2014



Cao et al., SIGGRAPH 2014



# High Dimensionality & Non-Linearity



occlusions, lighting, expressions, identity

sticky lips, biting, visemes, ...



# Deep Learning Model for Facial Expressions

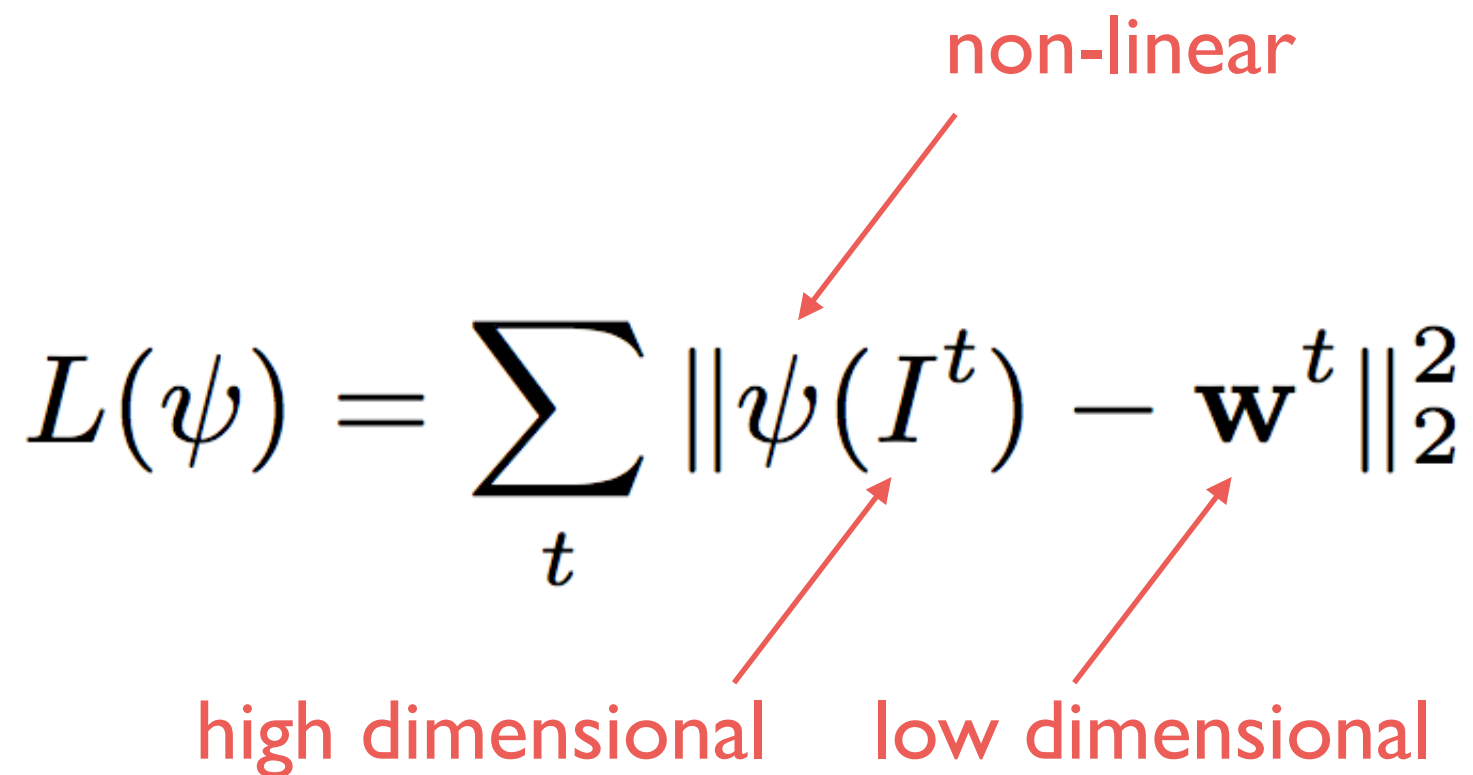
$$f^t = \mathbf{b}_0 + \sum_i^N \mathbf{w}_i^t (\mathbf{b}_i - \mathbf{b}_0)$$

$$L(\psi) = \sum_t \|\psi(I^t) - \mathbf{w}^t\|_2^2$$

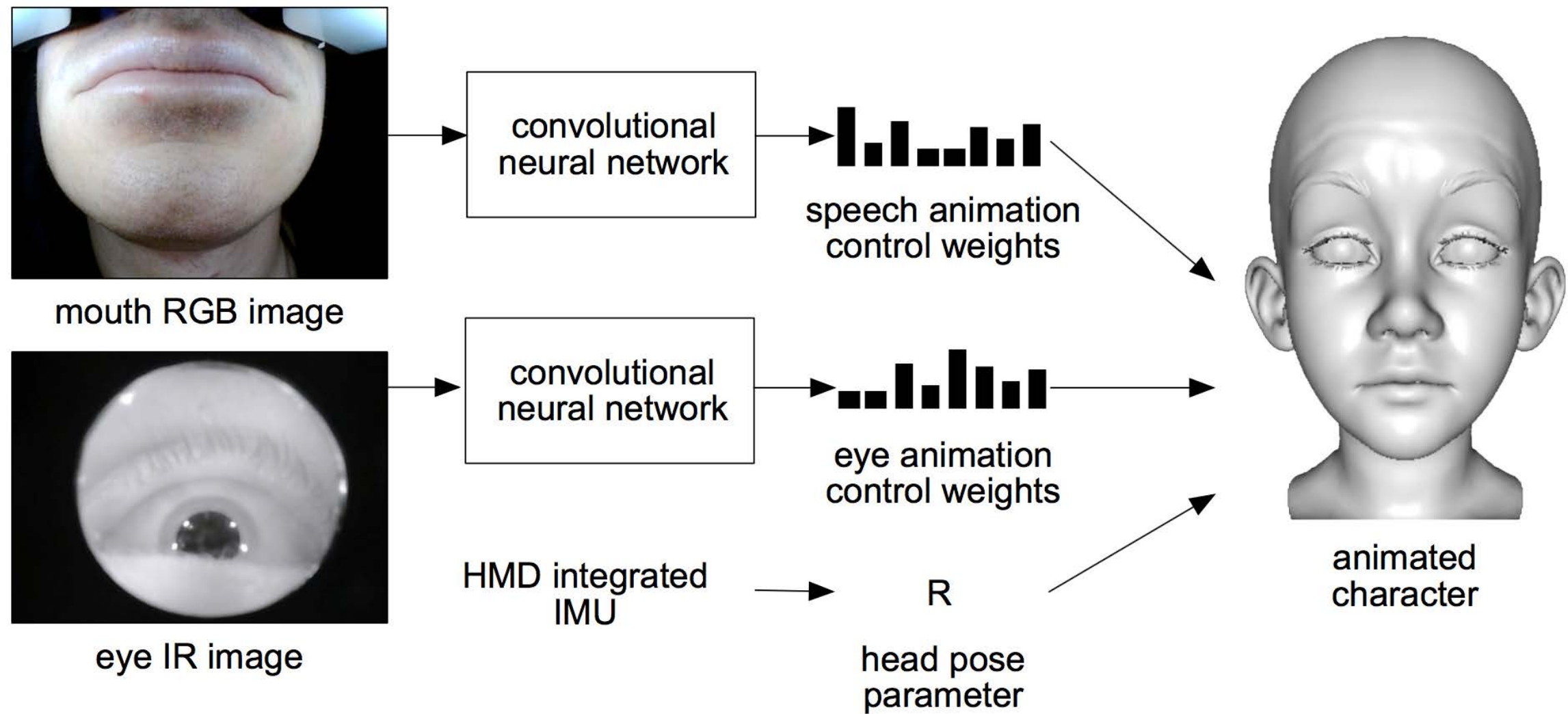
non-linear

high dimensional

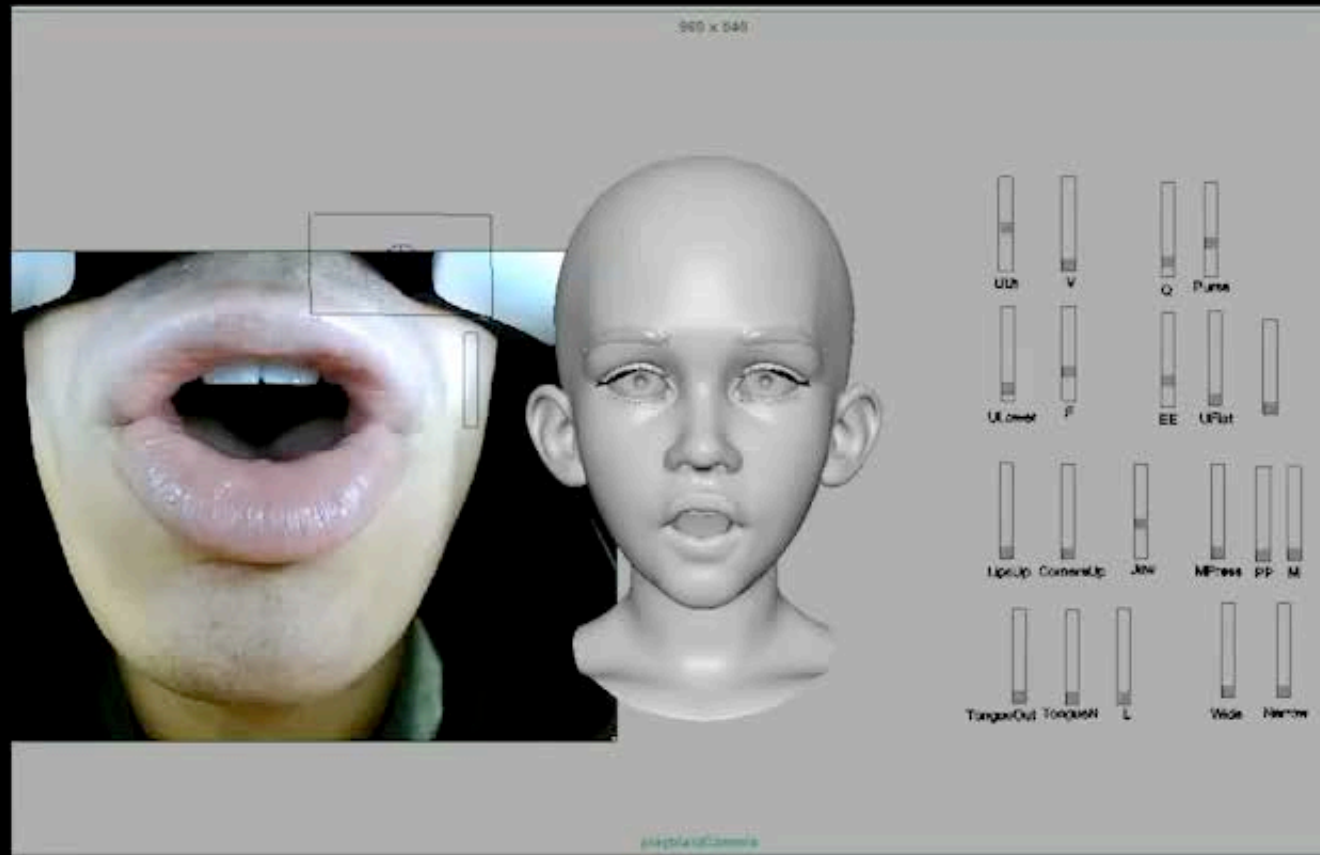
low dimensional



# Online Operation



# Label Transfer via Audio Alignment



reference  
data

reference  
animation

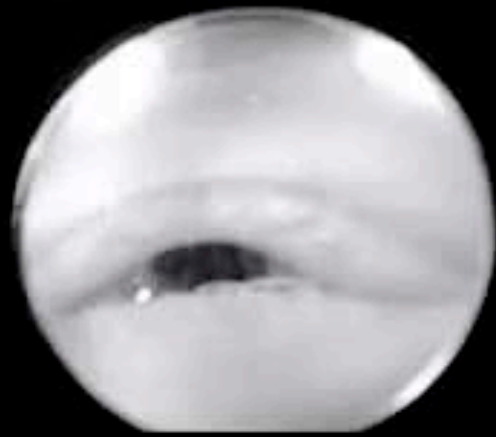


dynamic time warped  
training data





# Eye Animation Results



input video



output animation

# Retargeting



input performance

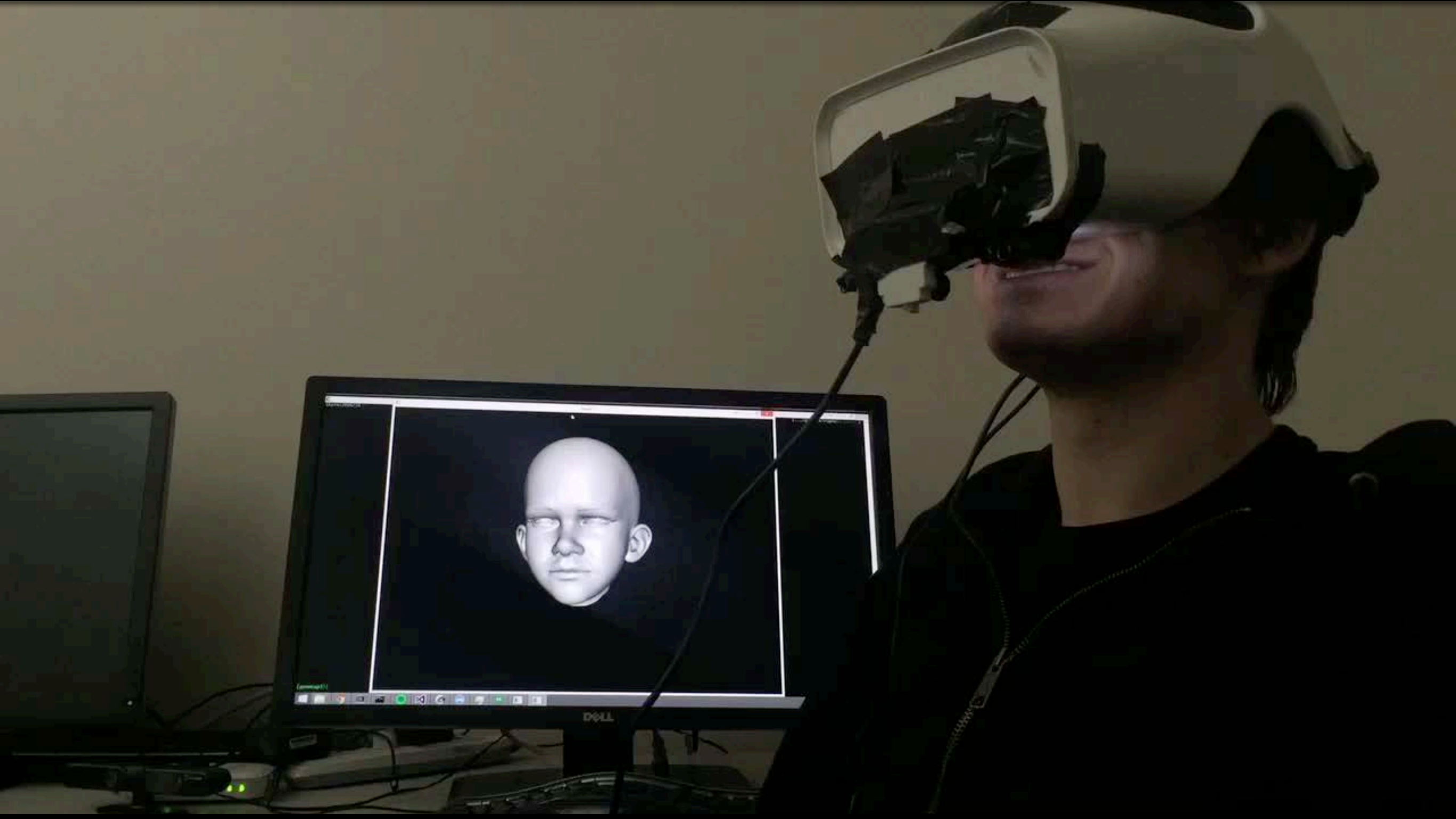


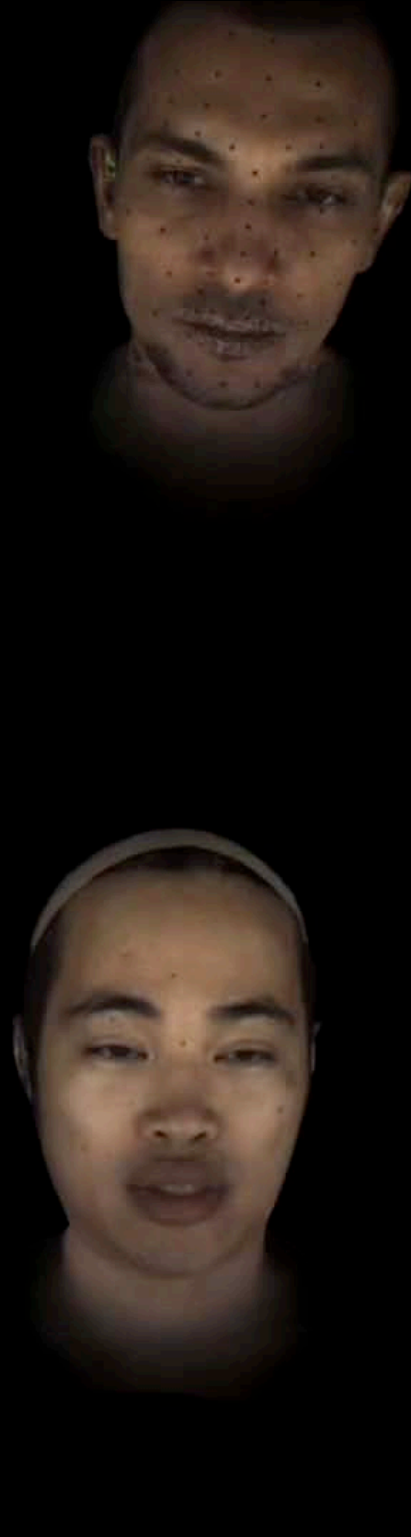
mouth camera



real-time  
facial animation

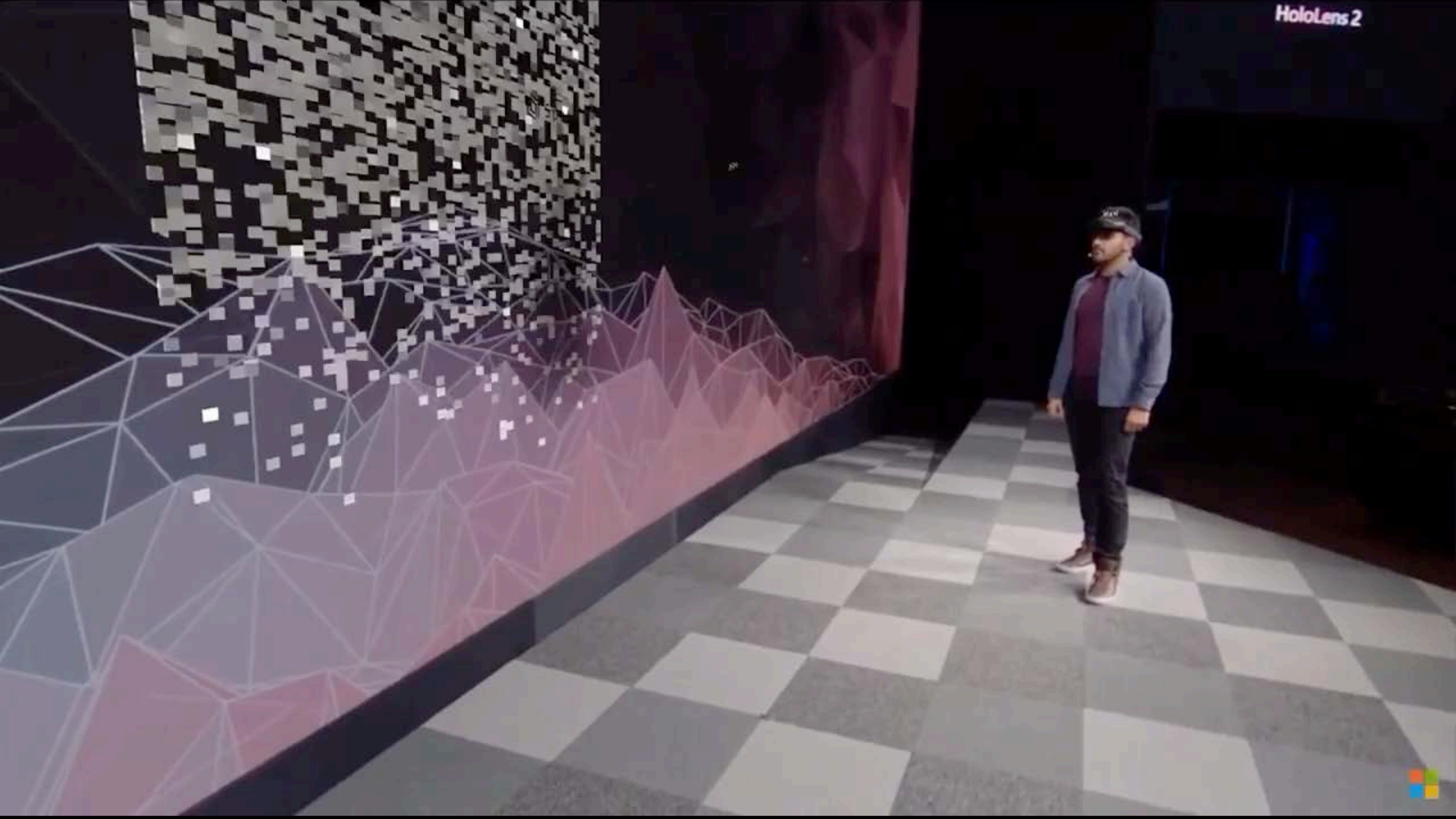








HoloLens 2

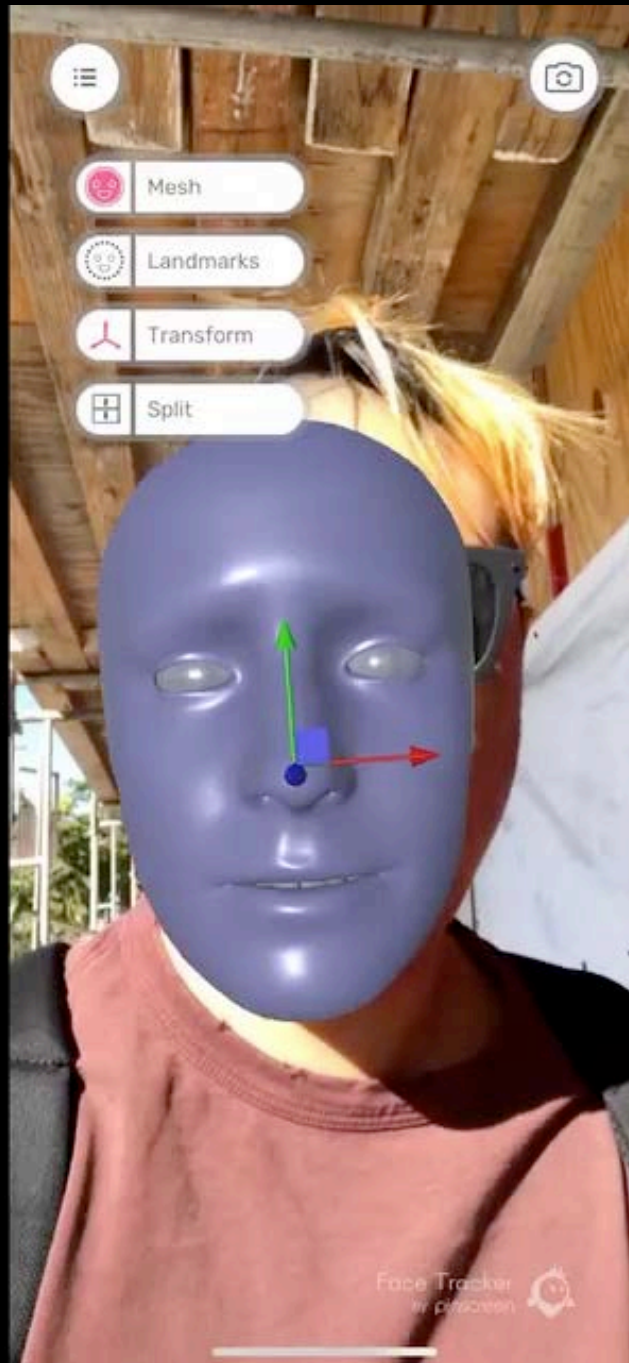




# Full Facial Performance Capture

Character 1  
Validation shot 1

# Pinscreen Facial Tracker (2018)





**Pinscreen (2017)**



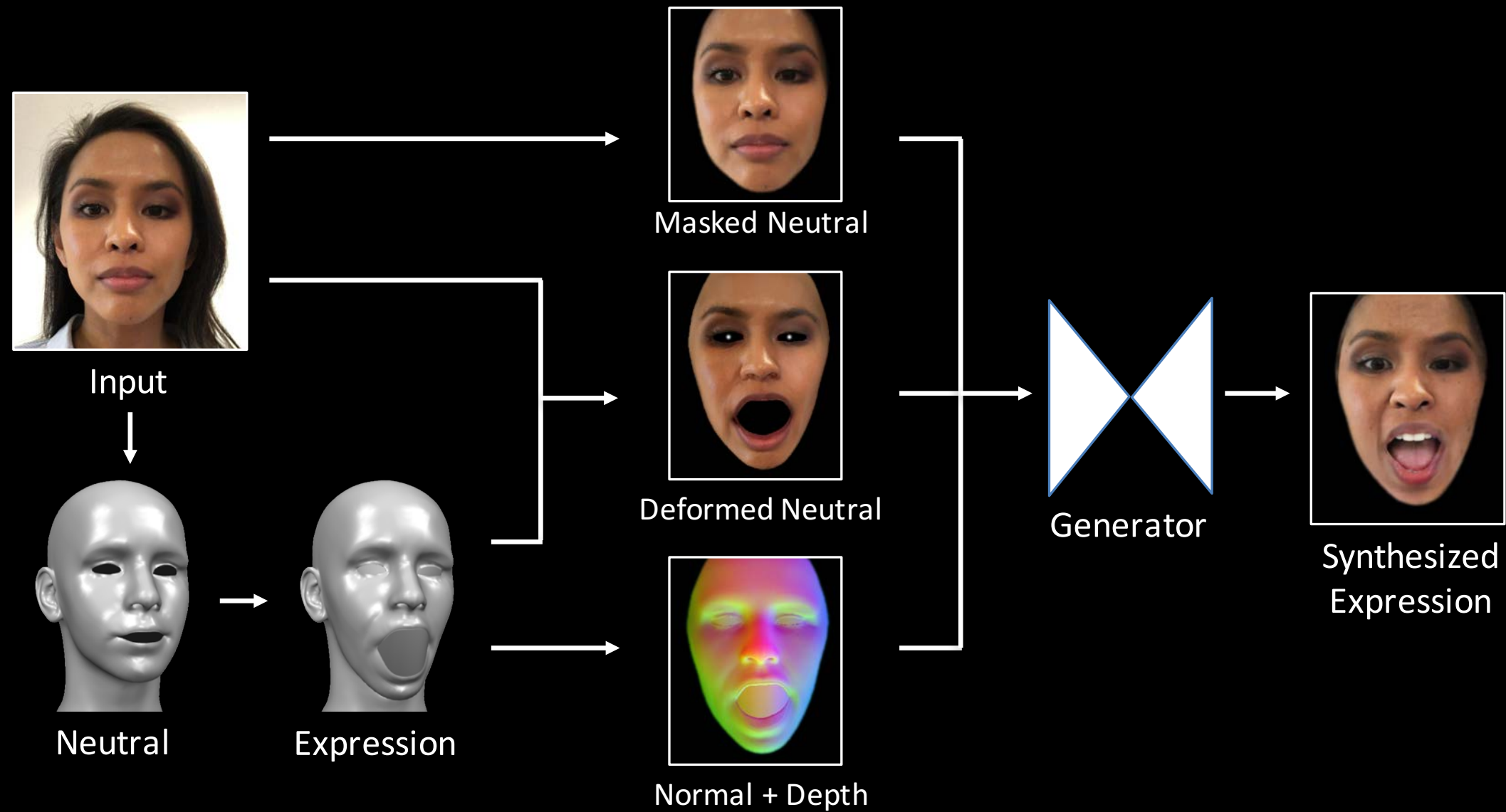
# Real-Time Lighting Estimation



# Deep Learning-Based Face Synthesis



# Deep Learning-Based Face Synthesis

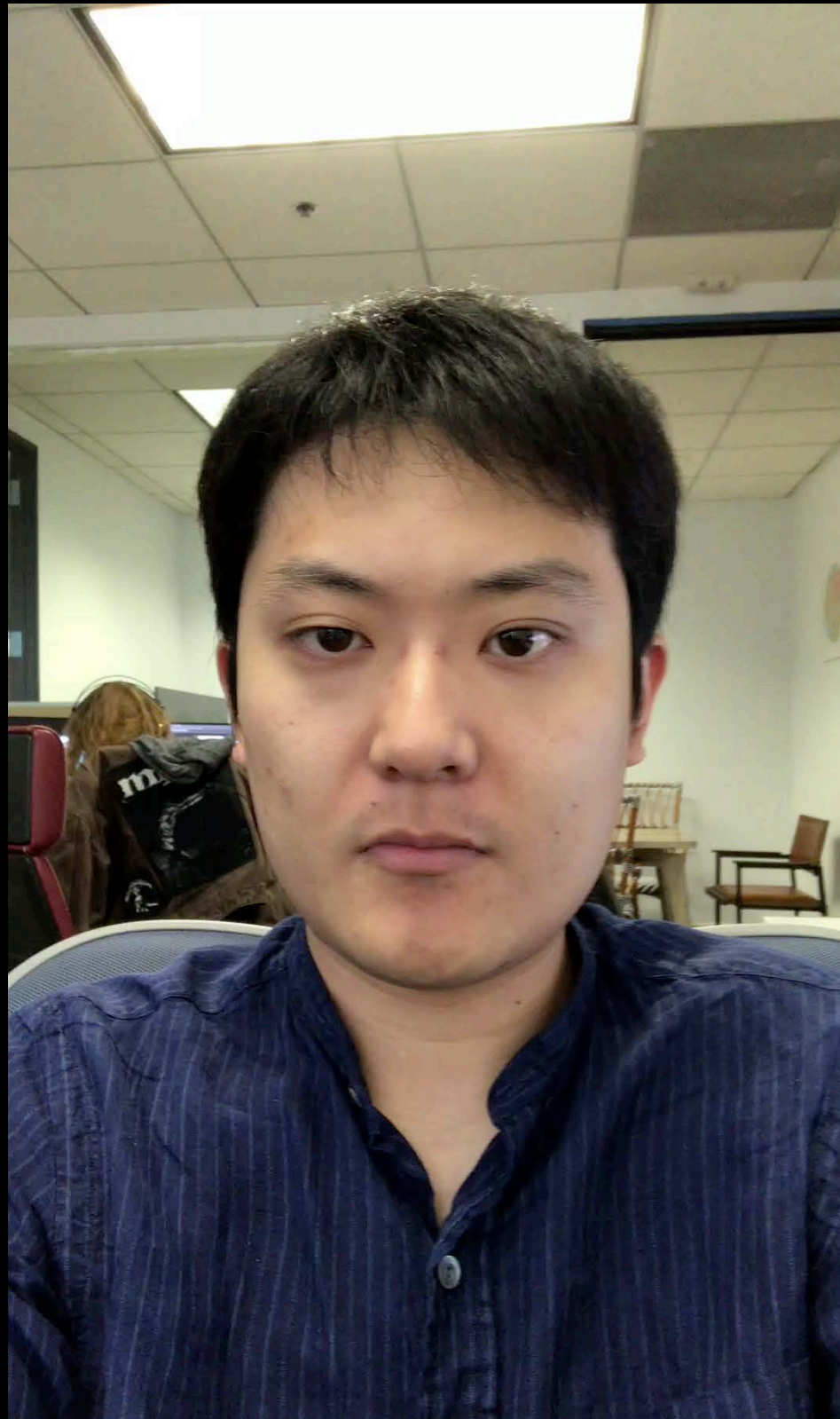




# Conditional GAN



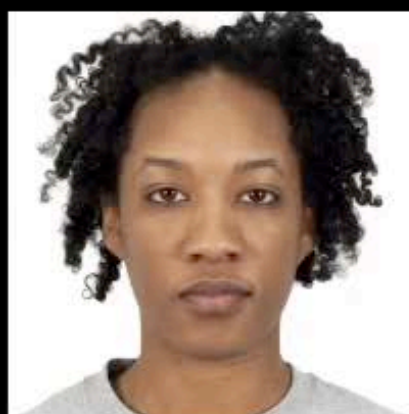
# Deep Learning-Based Face Synthesis



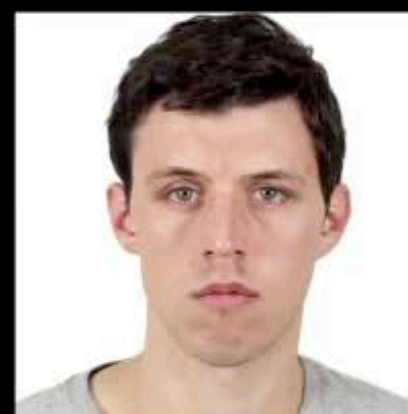
# Deep Learning-Based Face Synthesis



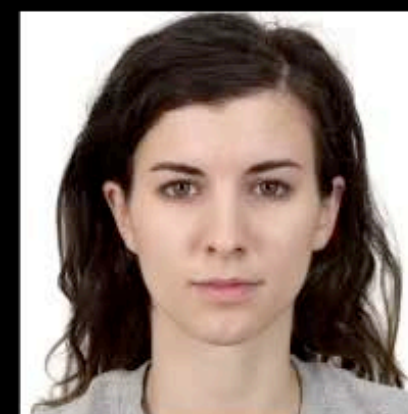
source video



subject A



subject B



subject C

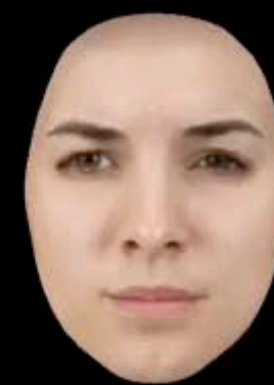
input image



our result



our result



our result



# Deep Learning-Based Face Synthesis



source video



subject A



subject B



subject C



subject D



subject E



subject F









# NextGen Photoreal Avatars

confidential









*The New York Times*

Here Come the  
Fake Videos, Too



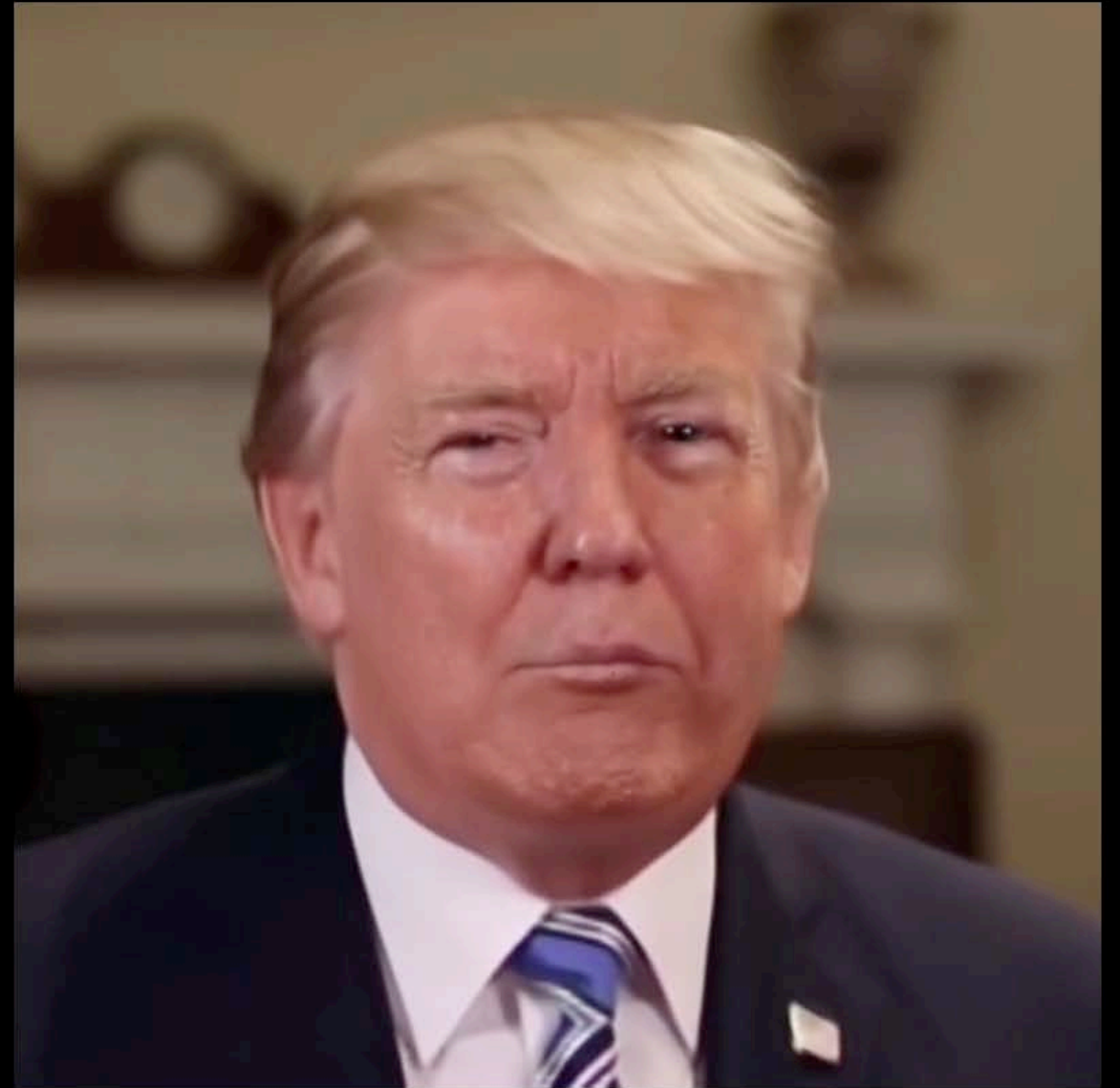
GOLDEN  
GLOBE  
AWARDS



GOLDEN  
GLOBE  
AWARDS



# Full Video Inference







# Real-time Face Replacement



# Real-time Retargeting and Illumination Estimation



# Real-time Face Replacement



# Real-time Retargeting (More Subjects)

# Interaction With 3D Avatars



**Blade Runner 2049 (2017)**



<http://cs621.hao-li.com>

# Thanks!

