Stabilized Real-time Face Tracking via a Learned Dynamic Rigidity Prior

Real-time 3D Face Tracking and Animation for Everyone

Chen Cao

Senior Research Scientist, Snap Inc.

Agenda

- Background of 3D face tracking and animation
- Reviews of my previous work
- Stabilized Real-time Face Tracking via a Learned Dynamic Rigidity Prior

What is 3D Face Tracking and Animation

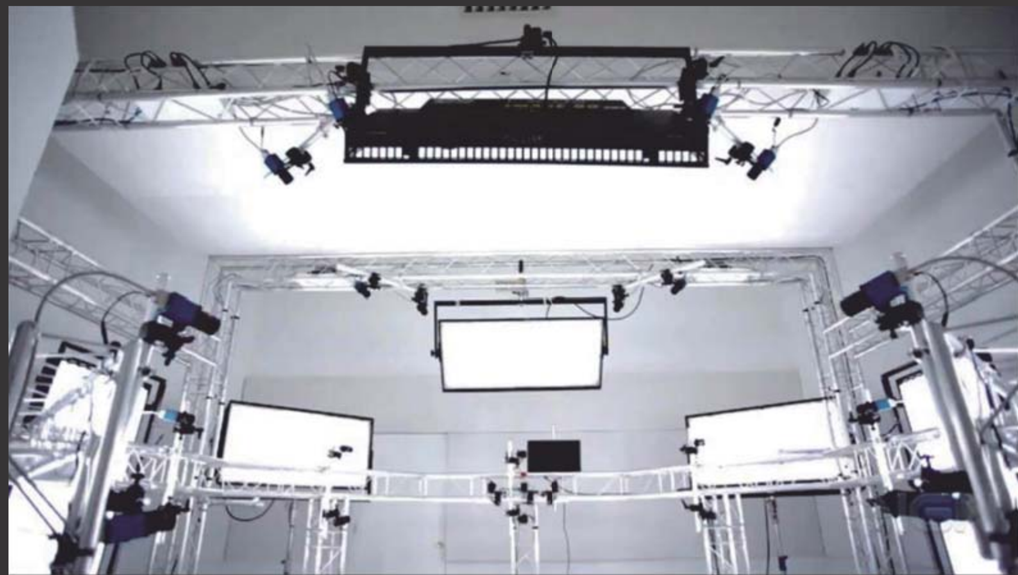
- Create animation of human face
  - Realistic, expressive
- Performance-based facial animation
  - Use the human face itself as the "input device"

A facial animation example

http://www.youtube.com/watch?v=-wtv4bsLWvw

Background: applications in movie and game

Movie: Avatar

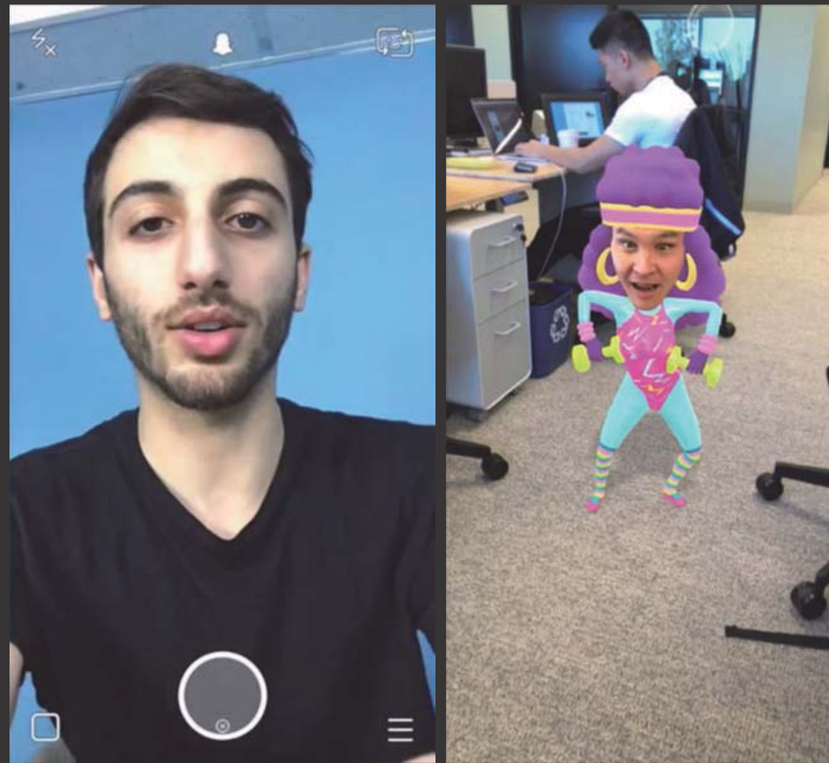


Game: L.A.Noire

Background: applications in AR/VR

Microsoft HoloLens

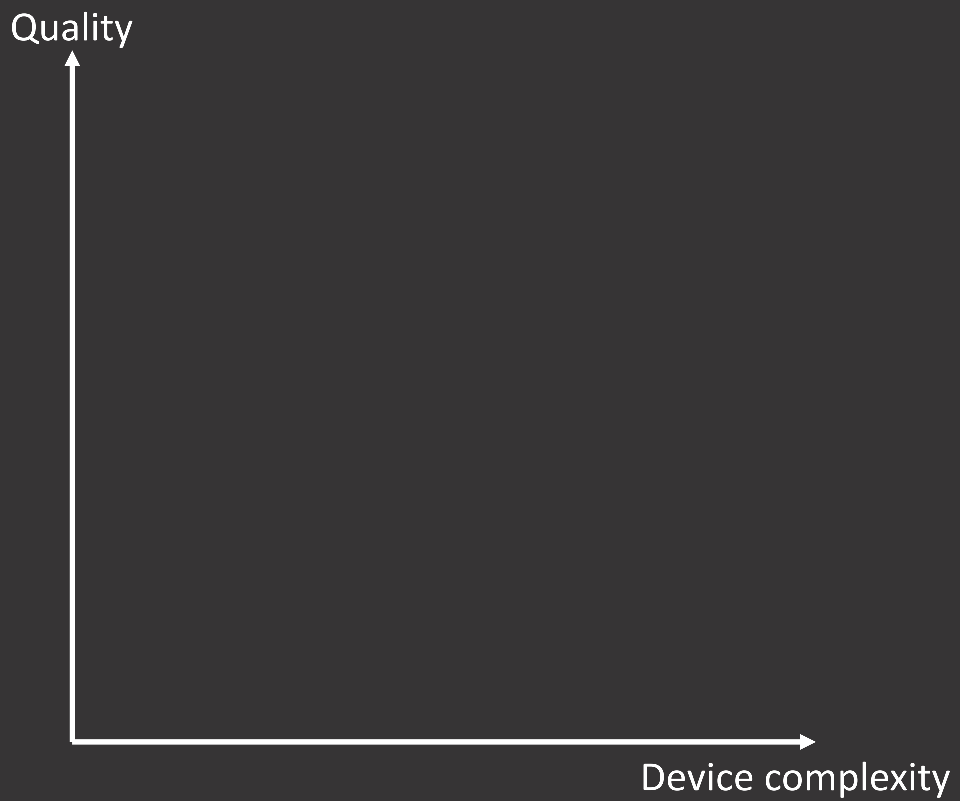
Holo



Snapchat

Related Work

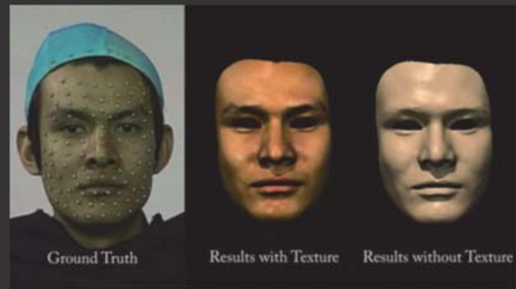
- Previous work: trade-off between device complexity and capture quality



Related Work

- Previous work: high quality offline facial performance capture

- ➢ Special equipment
- ➢ Offline
- ➢ For studio usage



[Huang et al. 2011]

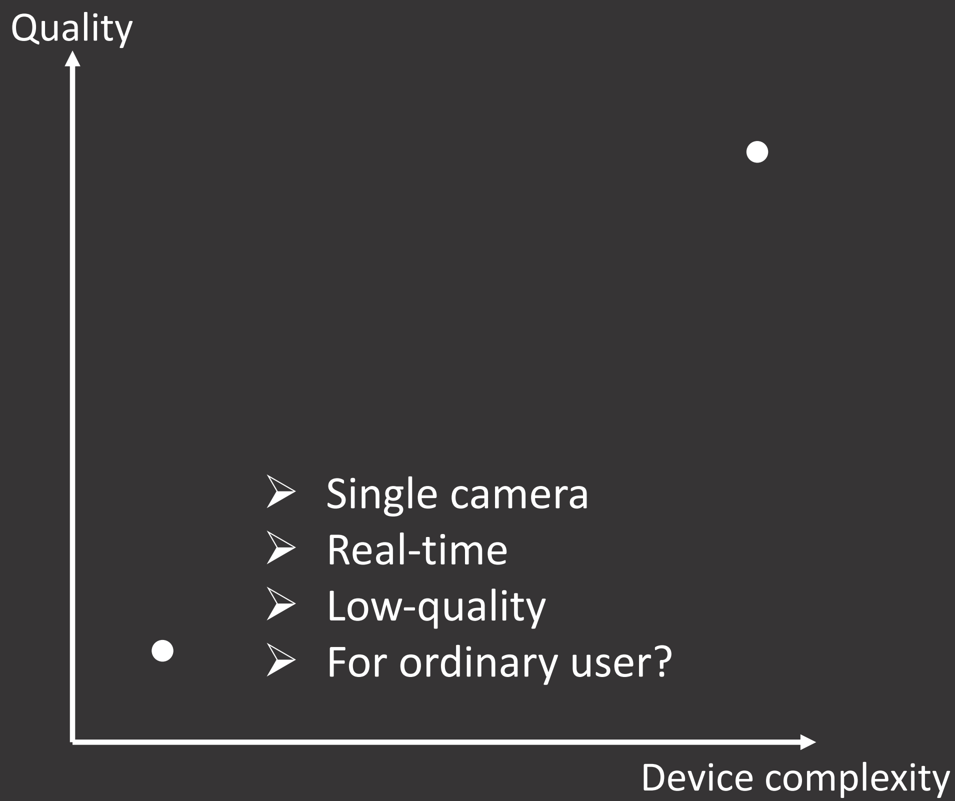
[Bradley et al. 2010]

[Beeler et al. 2011]

[Fyffe et al. 2015]

Related Work

- Previous work: real-time face animation based on single camera



- ➢ Single camera
- ➢ Real-time
- ➢ Low-quality
- ➢ For ordinary user?

[Vlasic et al. 2005]

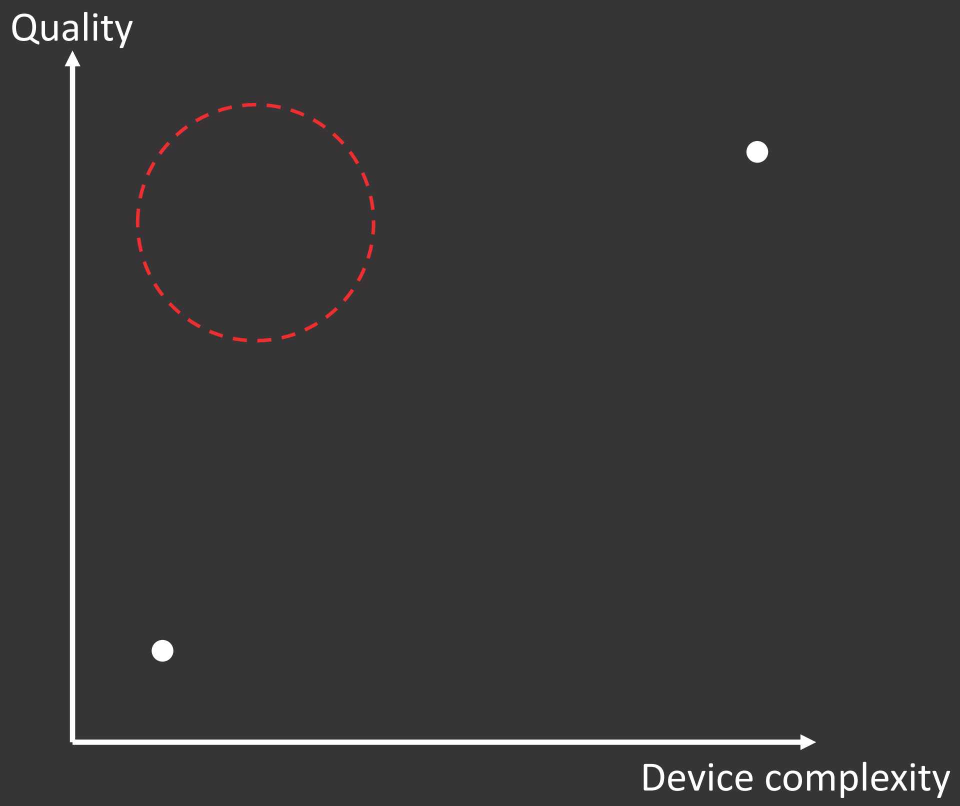
[Saragih et al. 2011]

[Tewari et al. 2017]

[Tewari et al. 2018]

Related Work

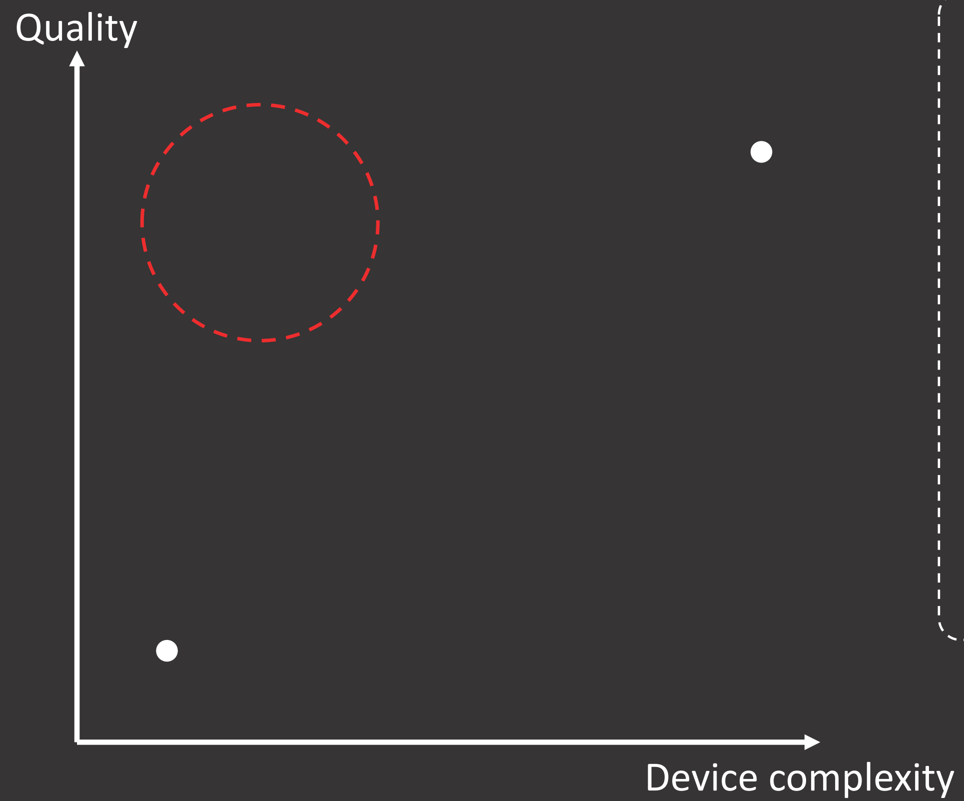
- Our goal: *real-time high-quality* 3D face tracking and animation for *everyone*



- Ease of use
  - Ordinary RGB camera (RGB-D camera), no any special equipment
  - Mobile device: limited computation power and memory
  - Real-time performance
- Robustness
  - Different users
  - Fast motions, large head pose, exaggerated facial expressions
  - Lighting/background changes
- Accurate and Realistic
  - Avoid uncanny valley

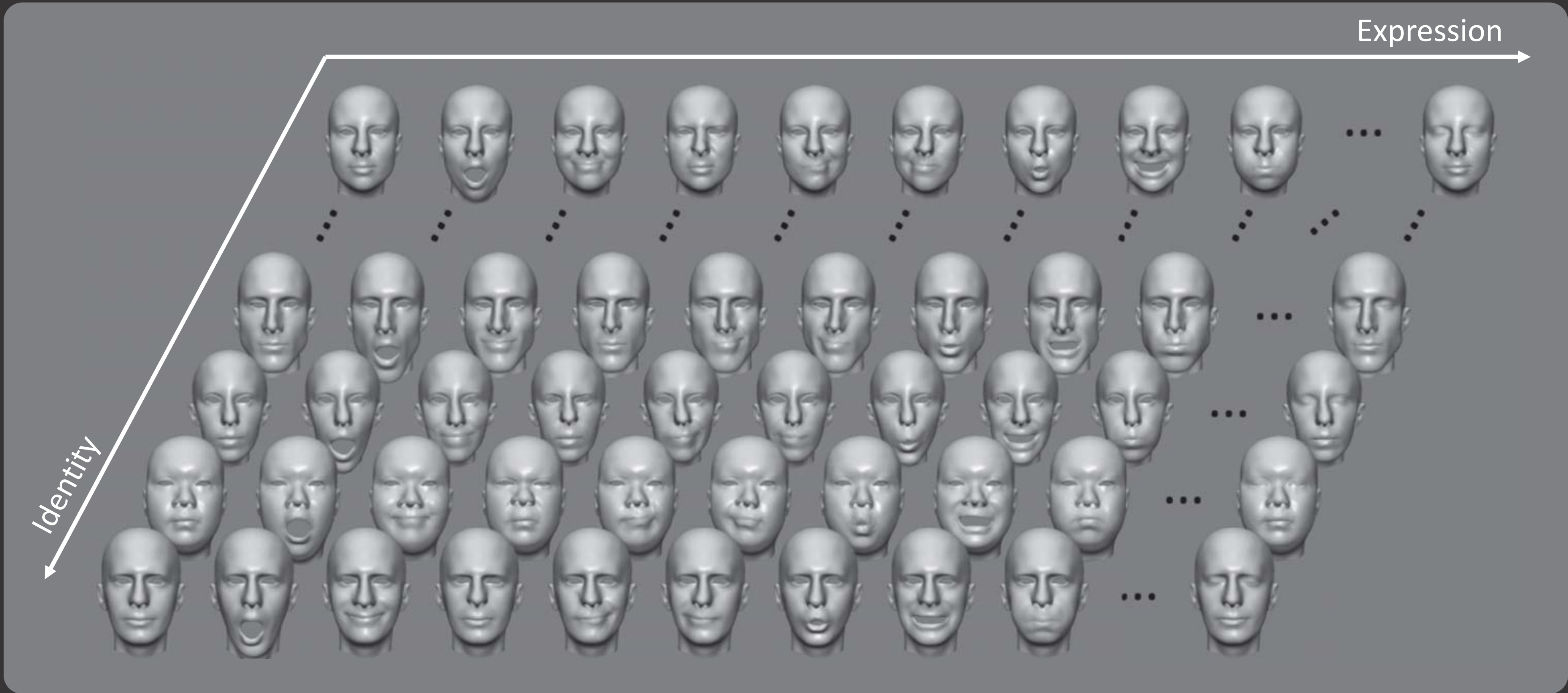
Related Work

- Our goal: *real-time high-quality* 3D face tracking and animation for *ordinary users*



- FaceWarehouse: A 3D Facial Expression Database for Visual Computing (TVCG 2014)
- 3D Shape Regression for Real-time Facial Animation (SIGGRAPH 2013)
- Displaced Dynamic Expression Regression for Real-time Facial Tracking and Animation (SIGGRAPH 2014)
- Real-Time High-Fidelity Facial Performance Capture (SIGGRAPH 2015)
- **Stabilized Real-time Face Tracking via a Learned Dynamic Rigidity Prior** (SIGGRAPH Asia 2018)
- Joint Audio-Video Driven Facial Animation (ICASSP 2018)

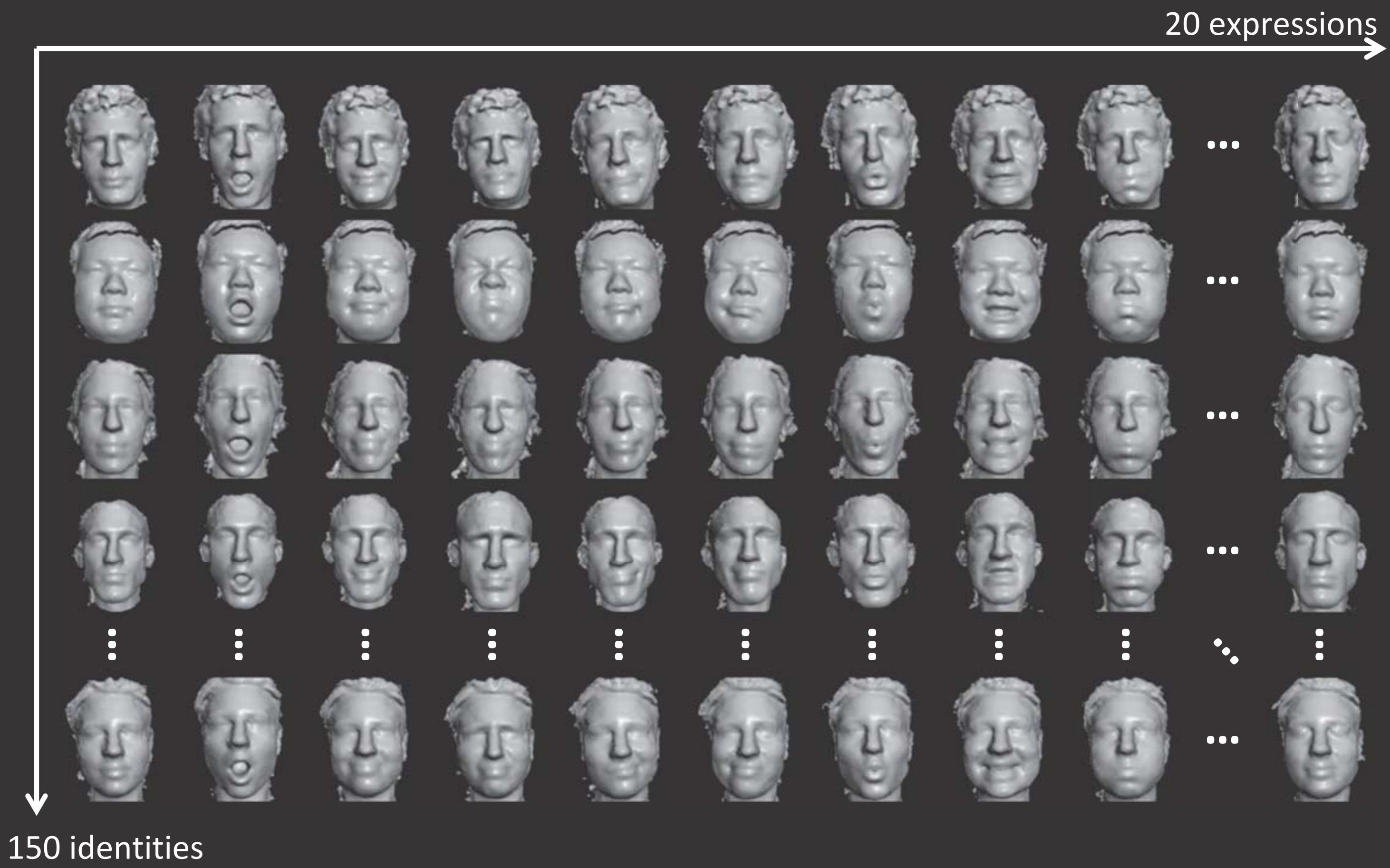
FaceWarehouse: A 3D Facial Expression Database for Visual Computing



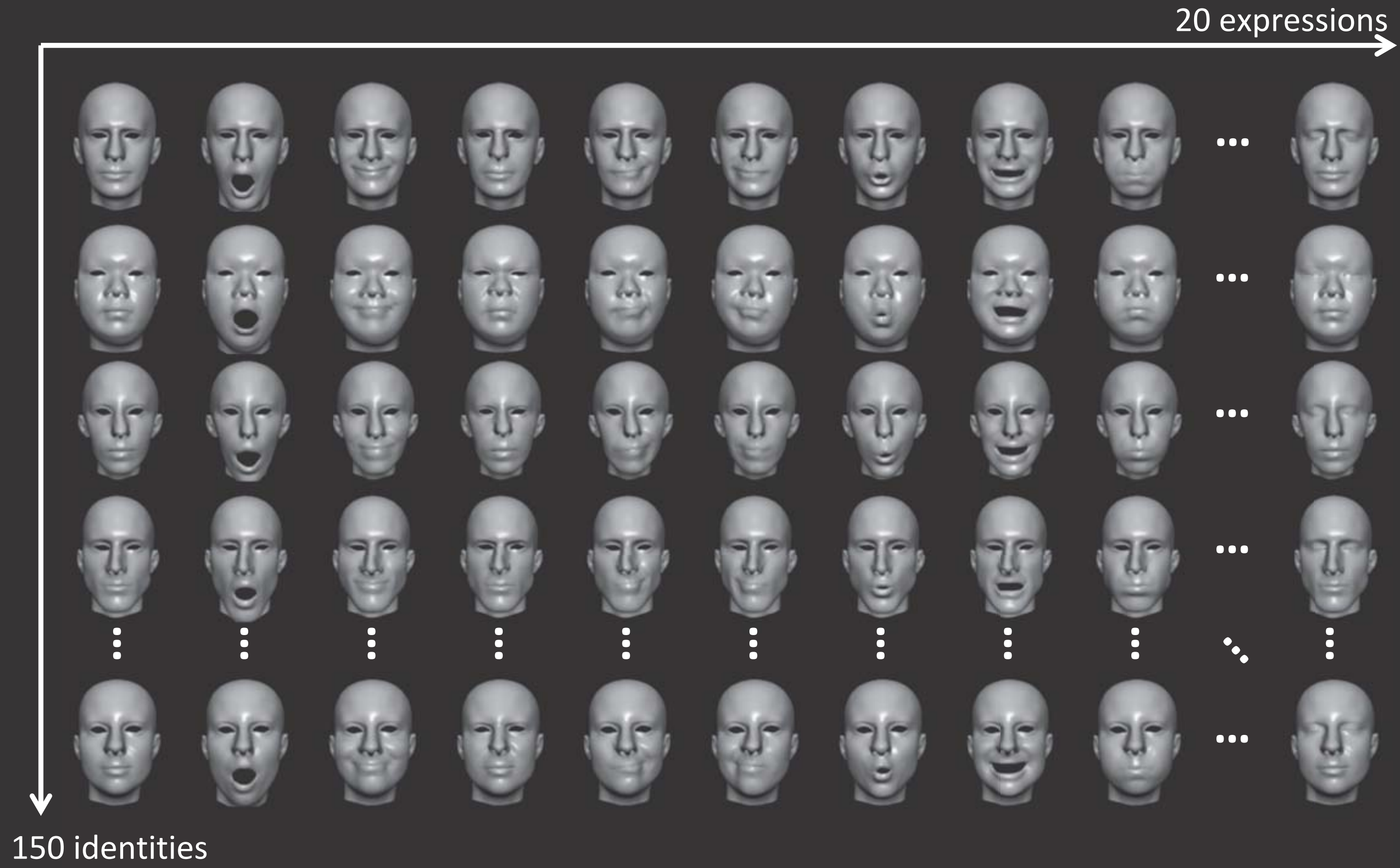
* **Chen Cao**, Yanlin Weng, Shun Zhou, Yiying Tong, Kun Zhou, FaceWarehouse: a 3D Facial Expression Database for Visual Computing, IEEE Transactions on Visualization and Computer Graphics, 20(3), March 2014.

FaceWarehouse: A 3D Facial Expression Database for Visual Computing

FaceWarehouse: A 3D Facial Expression Database for Visual Computing



FaceWarehouse: A 3D Facial Expression Database for Visual Computing



FaceWarehouse: A 3D Facial Expression Database for Visual Computing

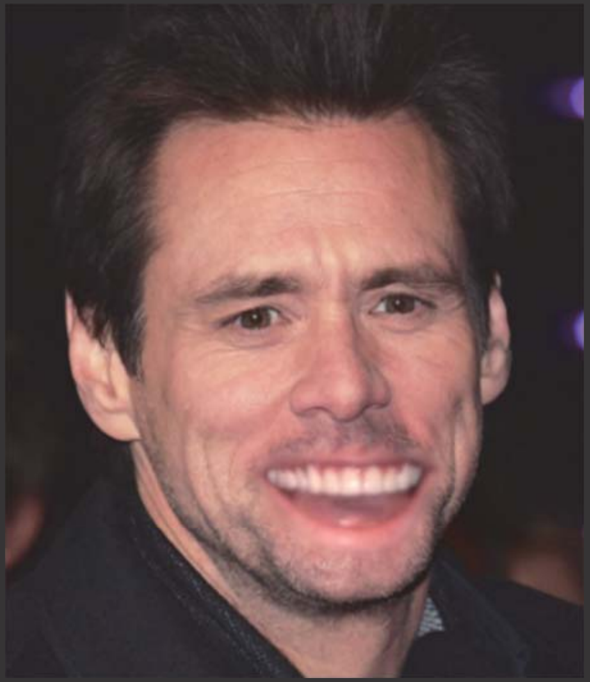
- Compared with other public face database

Database	[Blanz et al. 1999]	[Vlasic et al. 2005]	[Zhang et al. 2013]	[Min et al. 2014]	[Aly et al. 2015]	[Heseltine et al. 2005]	Ours
Identity Number	200	15	100	52	32	350	150
Expression Number	1	10	7	3	7	4	47

FaceWarehouse: A 3D Facial Expression Database for Visual Computing

Application #1: Facial image manipulation

FaceWarehouse: A 3D Facial Expression Database for Visual Computing

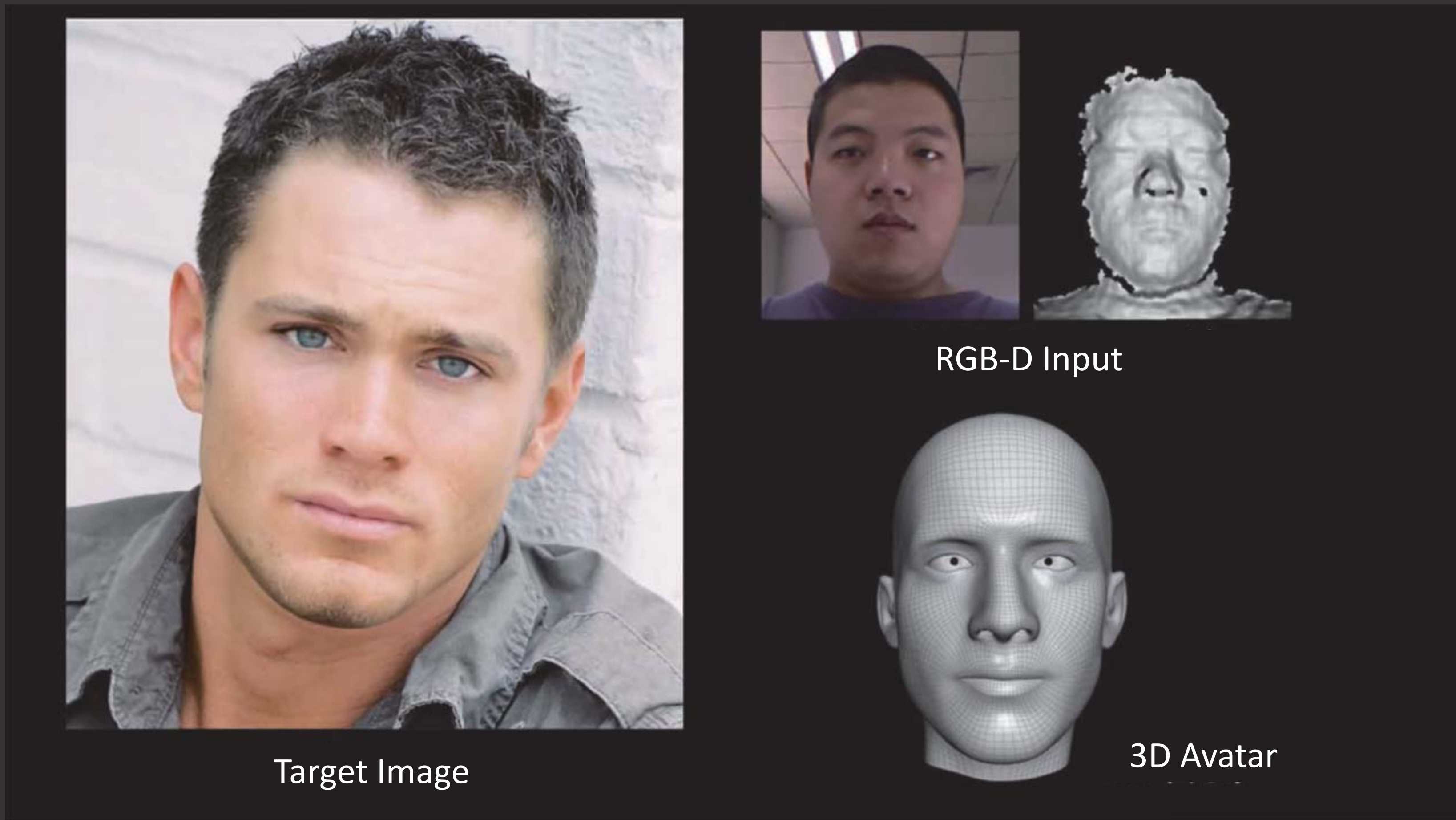


Application #2: Face component transfer

FaceWarehouse: A 3D Facial Expression Database for Visual Computing

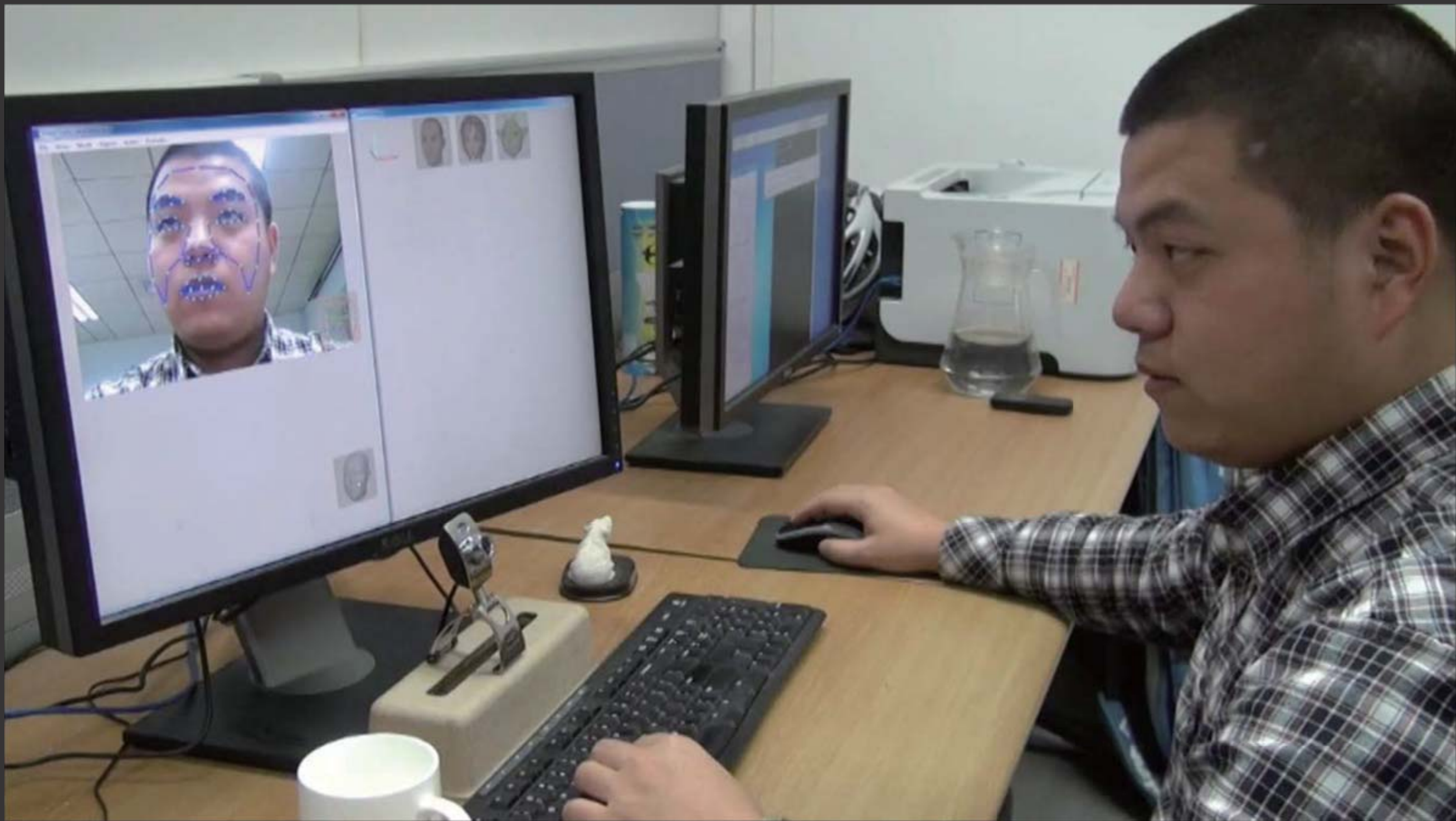
Application #2: Face component transfer

FaceWarehouse: A 3D Facial Expression Database for Visual Computing



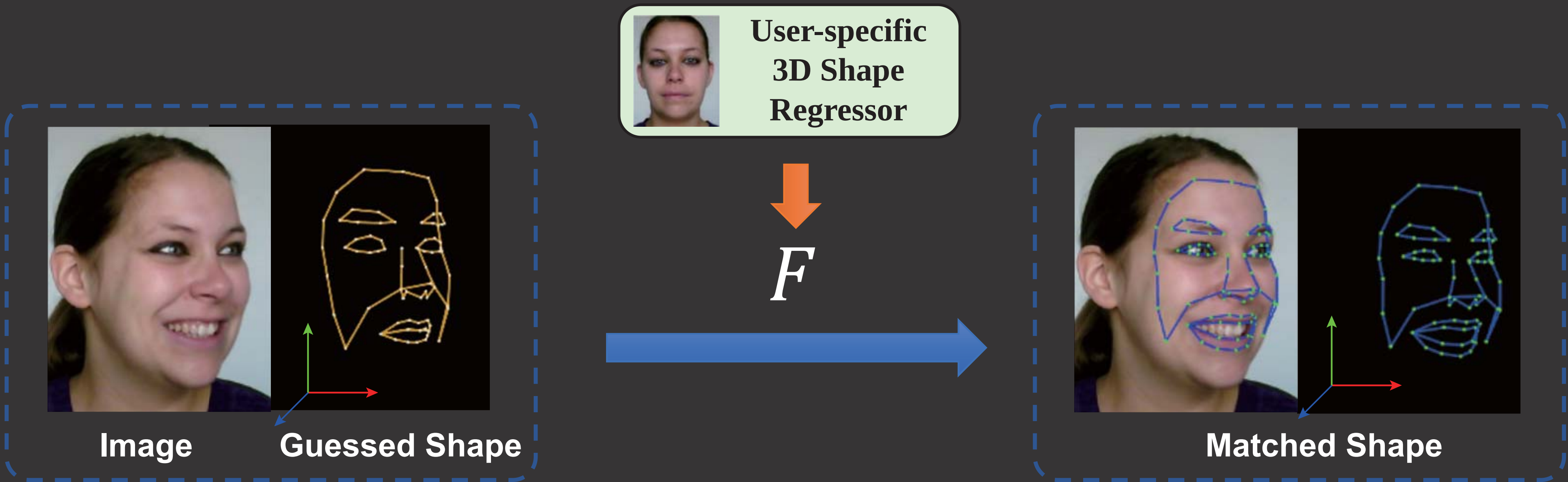
Application #3: Real-Time Performance-Based Facial Image Animation

3D Shape Regression for Real-time Facial Animation

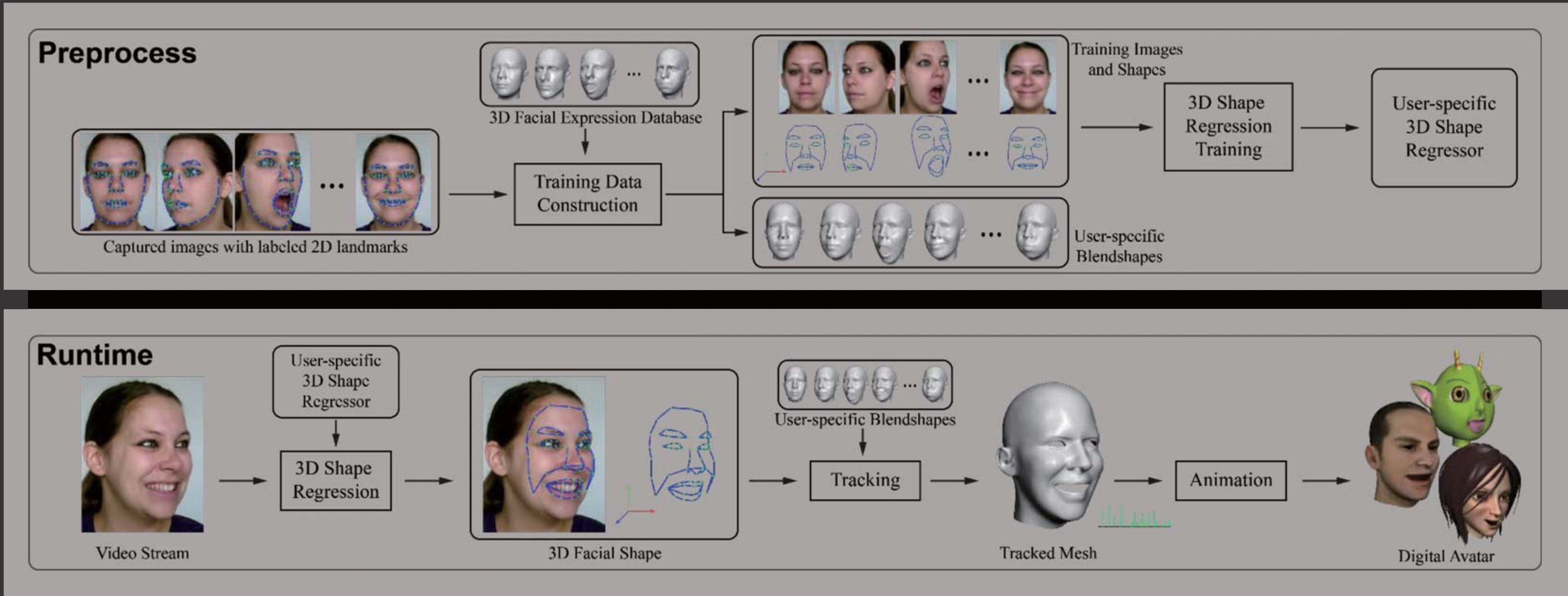


* **Chen Cao**, Yanlin Weng, Stephen Lin, Kun Zhou, 3D Shape Regression for Real-time Facial Animation, ACM TOG (Siggraph 2013), 32(4), 2013

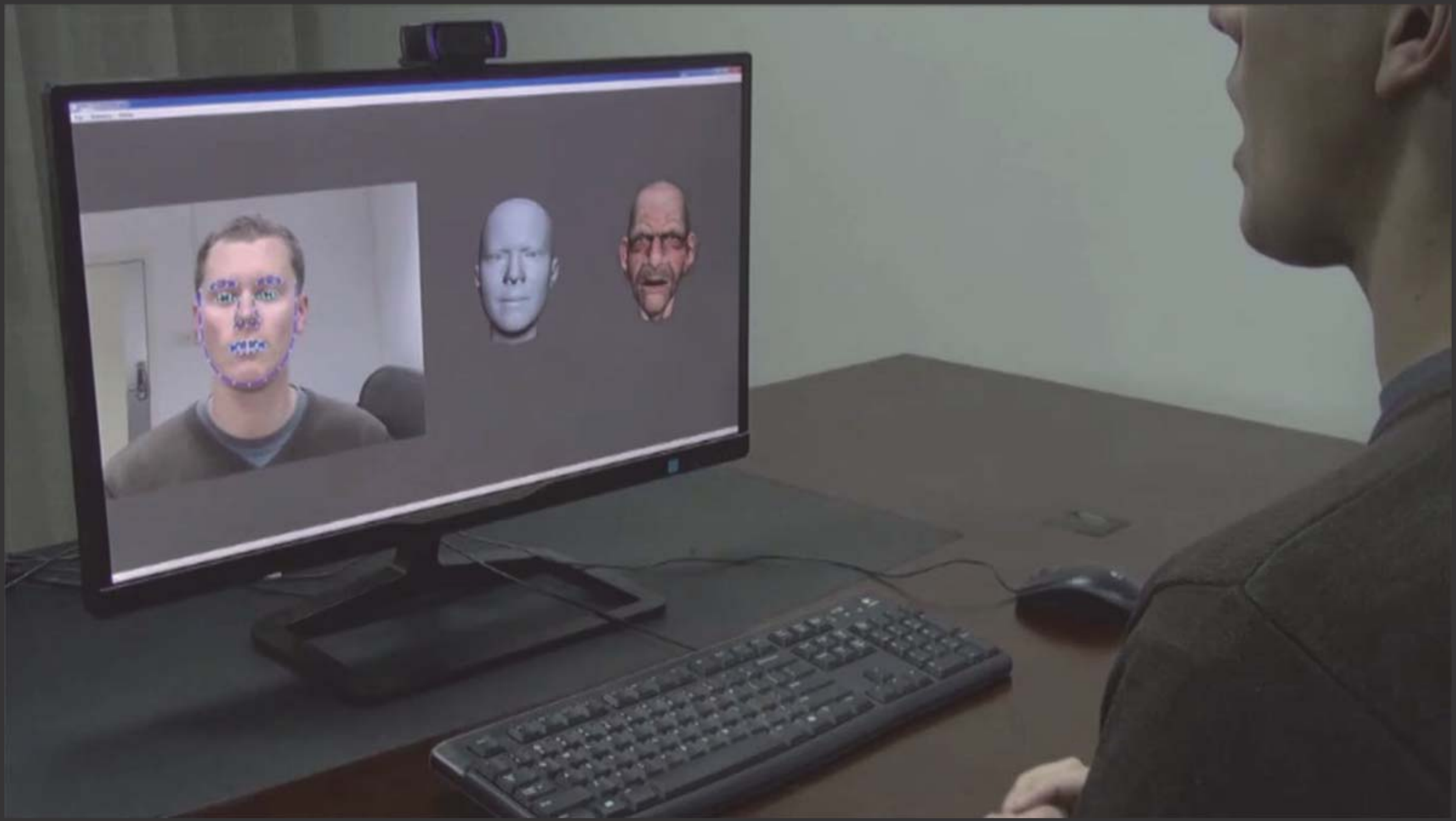
3D Shape Regression for Real-time Facial Animation



3D Shape Regression for Real-time Facial Animation



Displaced Dynamic Expression Regression for Real-time Facial Tracking and Animation



* **Chen Cao**, Qiming Hou, Kun Zhou, Displaced Dynamic Expression Regression for Real-time Facial Tracking and Animation, ACM TOG (Siggraph 2014), 33(4), 2014

Displaced Dynamic Expression Regression for Real-time Facial Tracking and Animation

- **DDE**: simultaneously represent the 3D face shape and the 2D facial landmarks

$$S = \mathrm{DDE}(\mathbf{Q}, \mathbf{u}; \mathbf{e}, \mathbf{R}, \mathbf{t}, \mathbf{D})$$

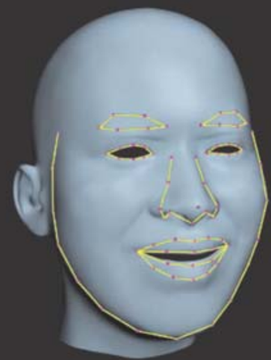
$\mathbf{Q}$: camera matrix

$\mathbf{u}$: user identity coefficeints

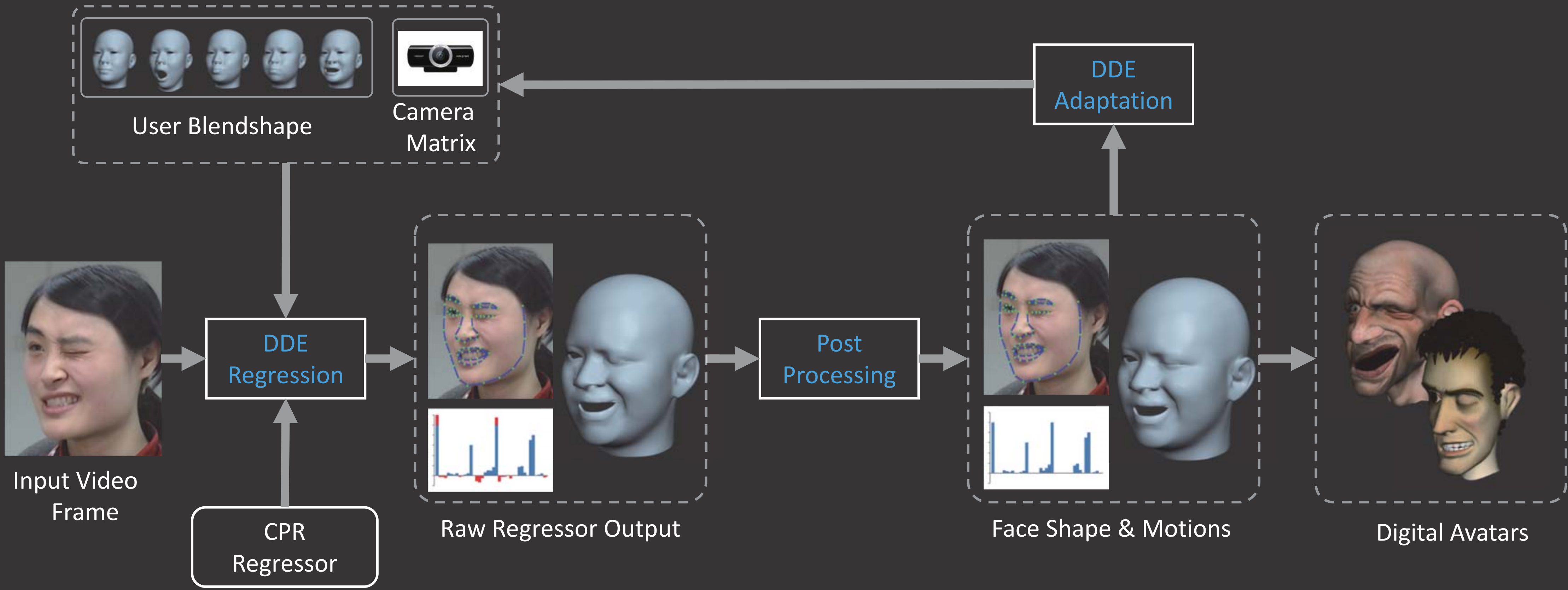
$\mathbf{e}$: expression coefficients

$\mathbf{R}, \mathbf{t}$: rotation, translation

$\mathbf{D}$: 2D landmark displacements



Displaced Dynamic Expression Regression for Real-time Facial Tracking and Animation



Real-Time High-Fidelity Facial Performance Capture

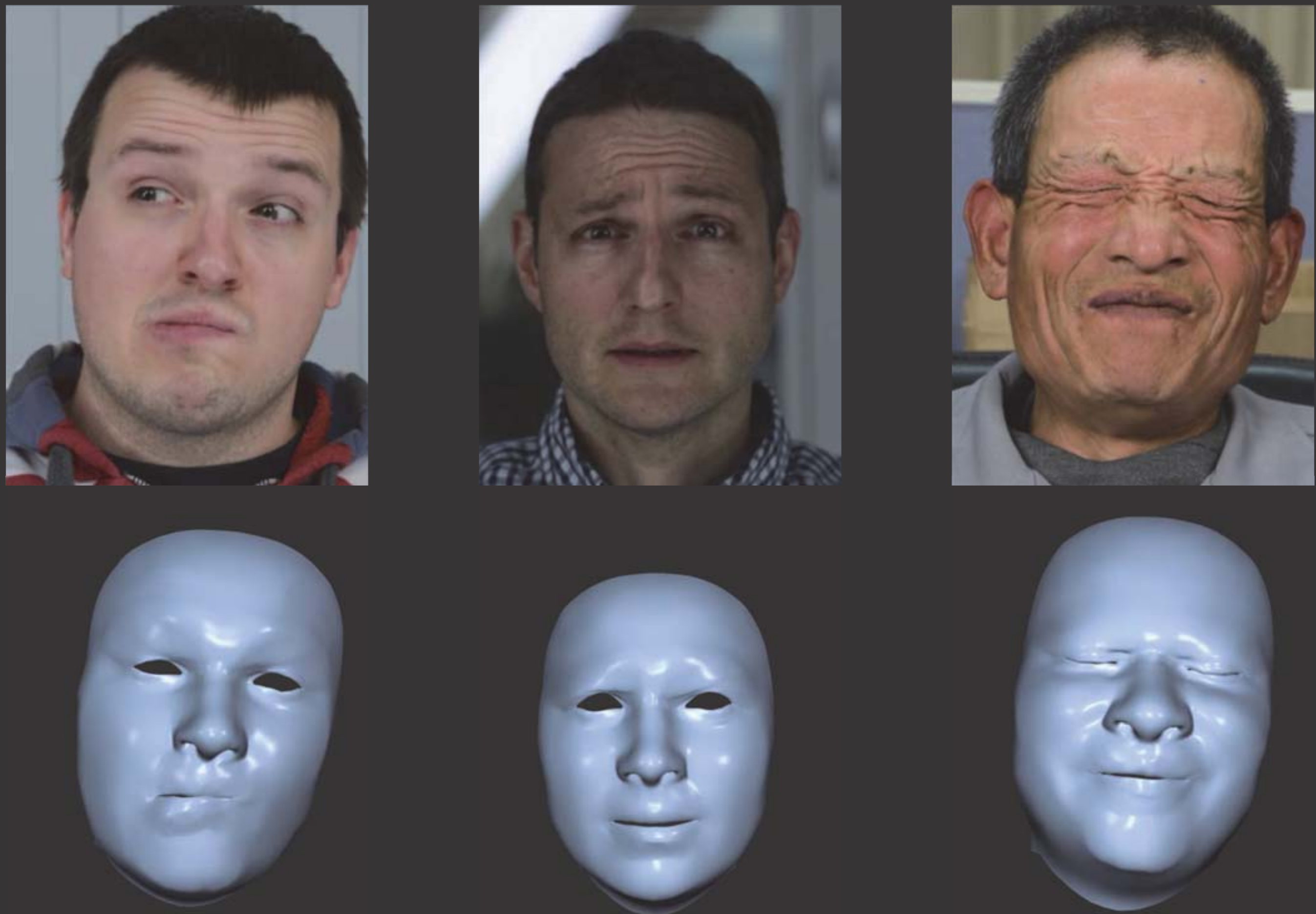
Input video | Coarse mesh | High-fidelity mesh

* **Chen Cao**, Derek Bradley, Kun Zhou, Thabo Beeler, Real-Time High-Fidelity Facial Performance Capture, ACM TOG (Siggraph 2015), 34(4), 2015

Real-Time High-Fidelity Facial Performance Capture

- **[Beeler et al. 2011]**: Offline high-quality face capture

Real-Time High-Fidelity Facial Performance Capture



- **Different** persons
- **Similar** shape
- **Generic** face model

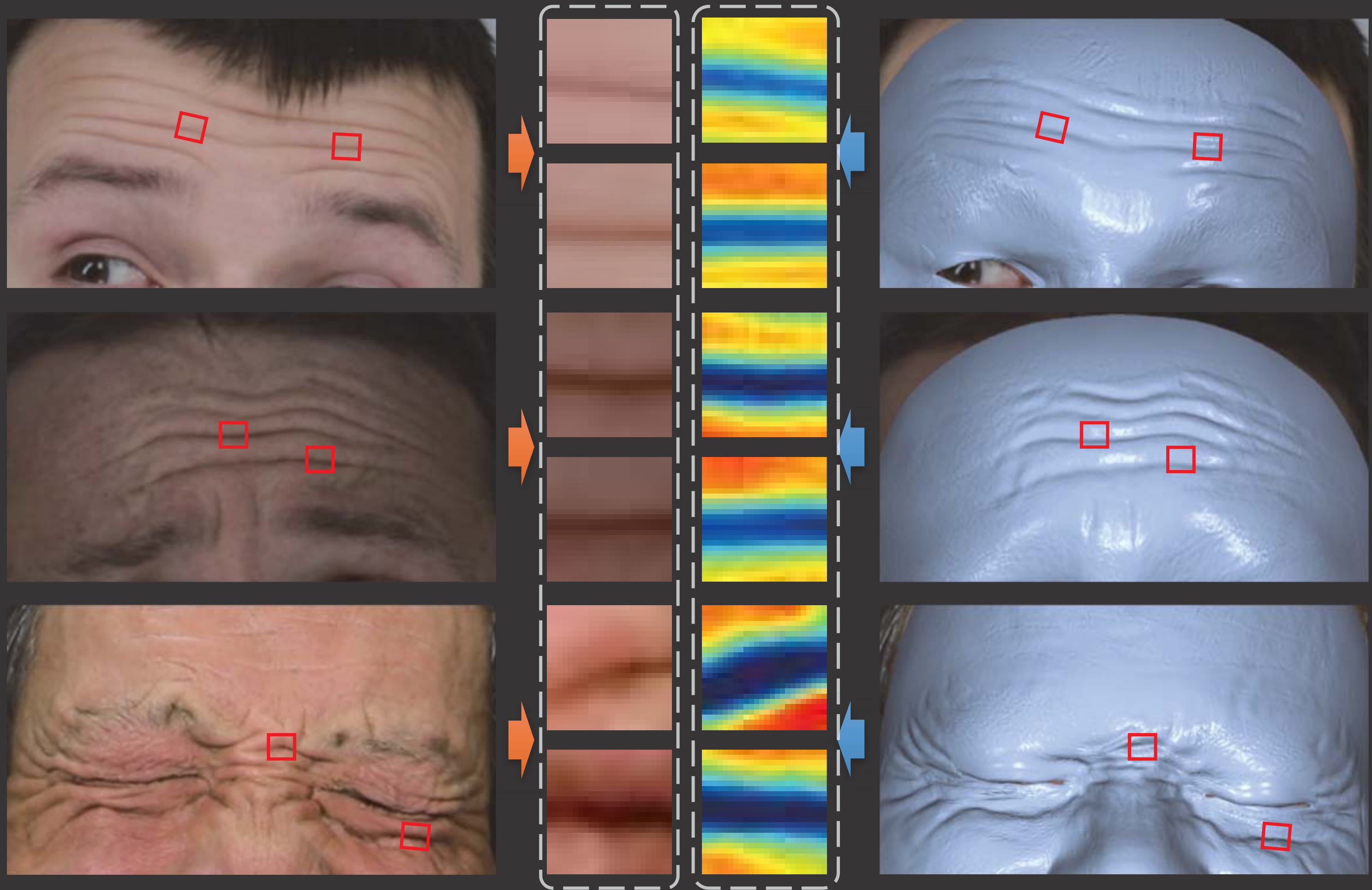
Real-Time High-Fidelity Facial Performance Capture

- **Different** locations
- **Different** scales
- **Different** orientations

- **Personal-specific** wrinkles

Real-Time High-Fidelity Facial Performance Capture

Real-Time High-Fidelity Facial Performance Capture



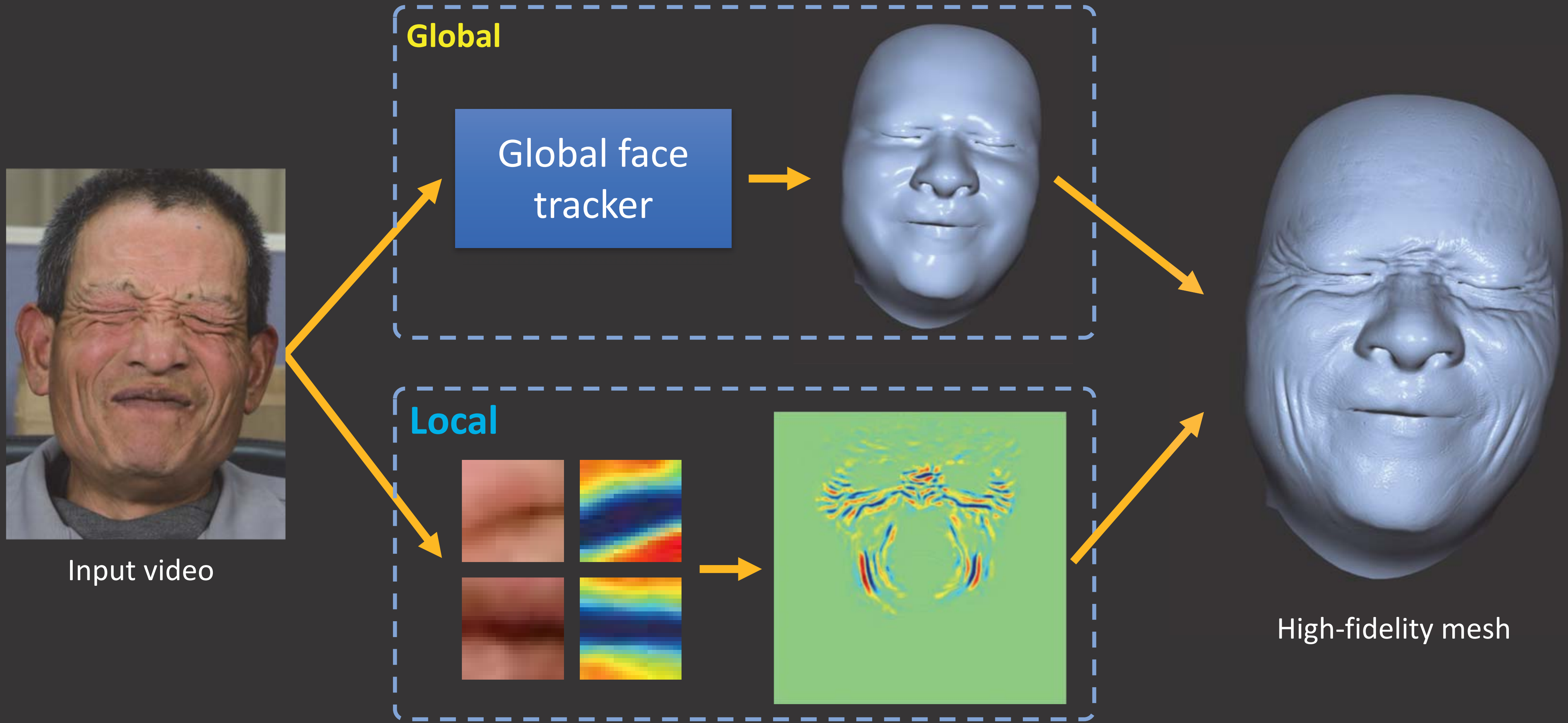
- **Locally** self-similar

Visual **Appearance**

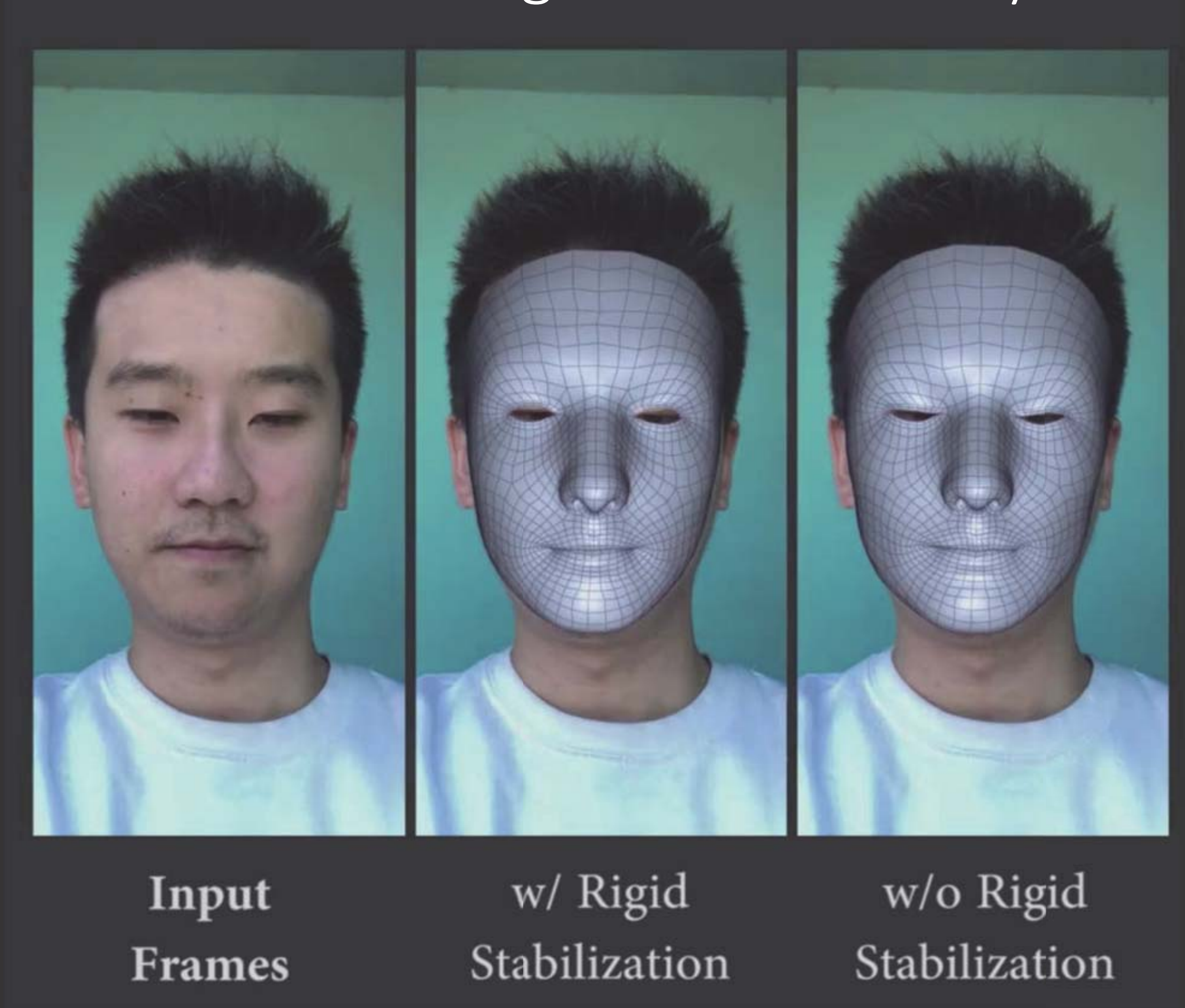
↕

Local **Shape**

Real-Time High-Fidelity Facial Performance Capture



Stabilized Real-time Face Tracking via a Learned Dynamic Rigidity Prior



* **Chen Cao**, Menglei Chai, Oliver Woodford, Linjie Luo, Stabilized Real-time Face Tracking via a Learned Dynamic Rigidity Prior, ACM TOG (Siggraph Asia 2018)

Problem Description

$$\mathbf{F}(\boldsymbol{\beta}) = \mathbf{B}_0 + \sum_i \beta_i \mathbf{B}_i$$

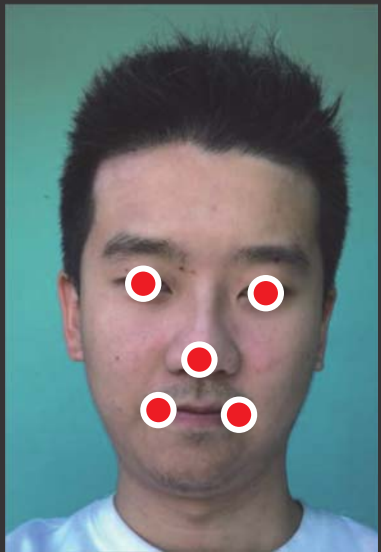
$$\mathbf{P}(\mathrm{T}, \boldsymbol{\beta}) = \boldsymbol{proj}\left(\mathrm{T}\ \mathbf{F}(\boldsymbol{\beta})\right)$$

T: rigid transformation

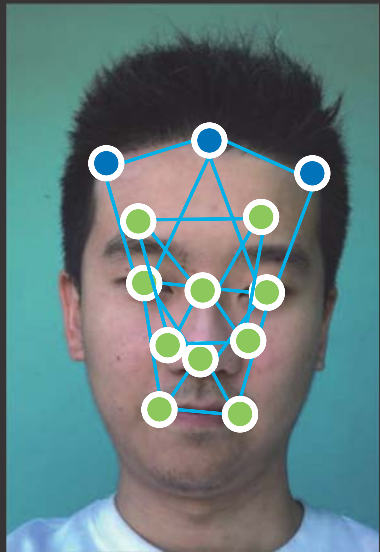
$\boldsymbol{\beta}$: non-rigid expression coefficients

$$\mathrm{argmin}_{\mathrm{T},\boldsymbol{\beta}} \sum_i \left\| \mathbf{P}(\mathrm{T}, \boldsymbol{\beta})^{(v_i)} - L_i \right\|^2$$

Frame #1



2D Landmarks



Fitted Mesh

Problem Description

$$\mathbf{F}(\boldsymbol{\beta}) = \mathbf{B}_0 + \sum_i \beta_i \mathbf{B}_i$$

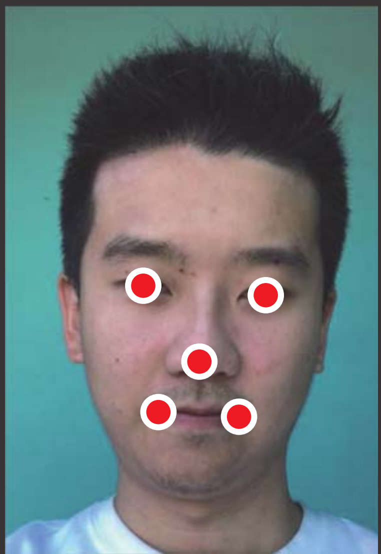
$$\mathbf{P}(\mathbf{T}, \boldsymbol{\beta}) = \boldsymbol{proj}(\mathbf{T}\,\mathbf{F}(\boldsymbol{\beta}))$$

$\mathbf{T}$: rigid transformation

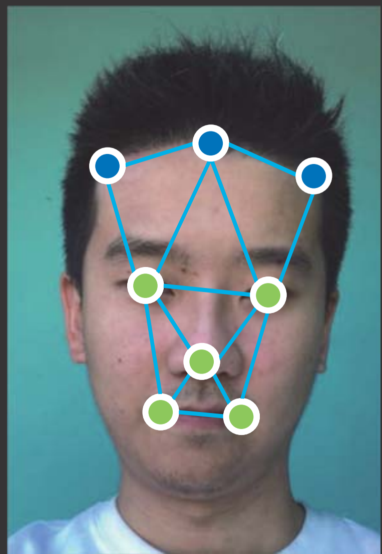
$\boldsymbol{\beta}$: non-rigid expression coefficients

$$\mathrm{argmin}_{\mathbf{T},\boldsymbol{\beta}} \sum_i \left\| \mathbf{P}(\mathbf{T}, \boldsymbol{\beta})^{(v_i)} - L_i \right\|^2$$

Frame #1

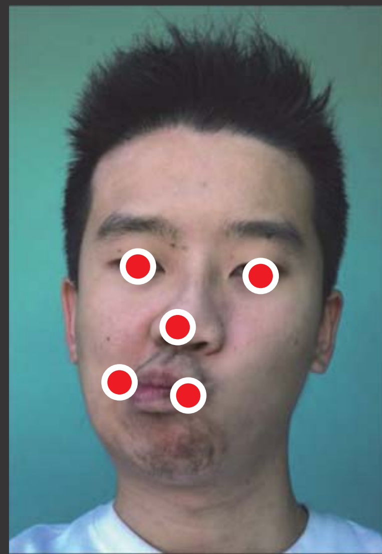


2D Landmarks

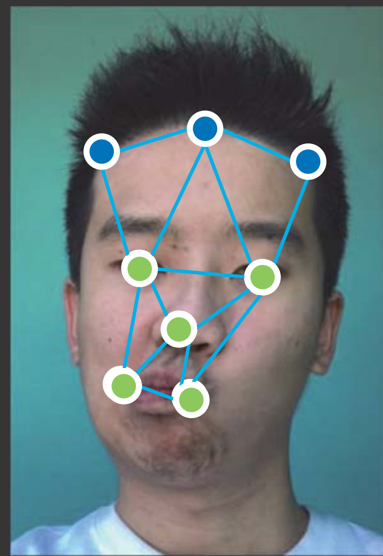


Fitted Mesh

Frame #2



2D Landmarks



Fitted Mesh

Problem Description

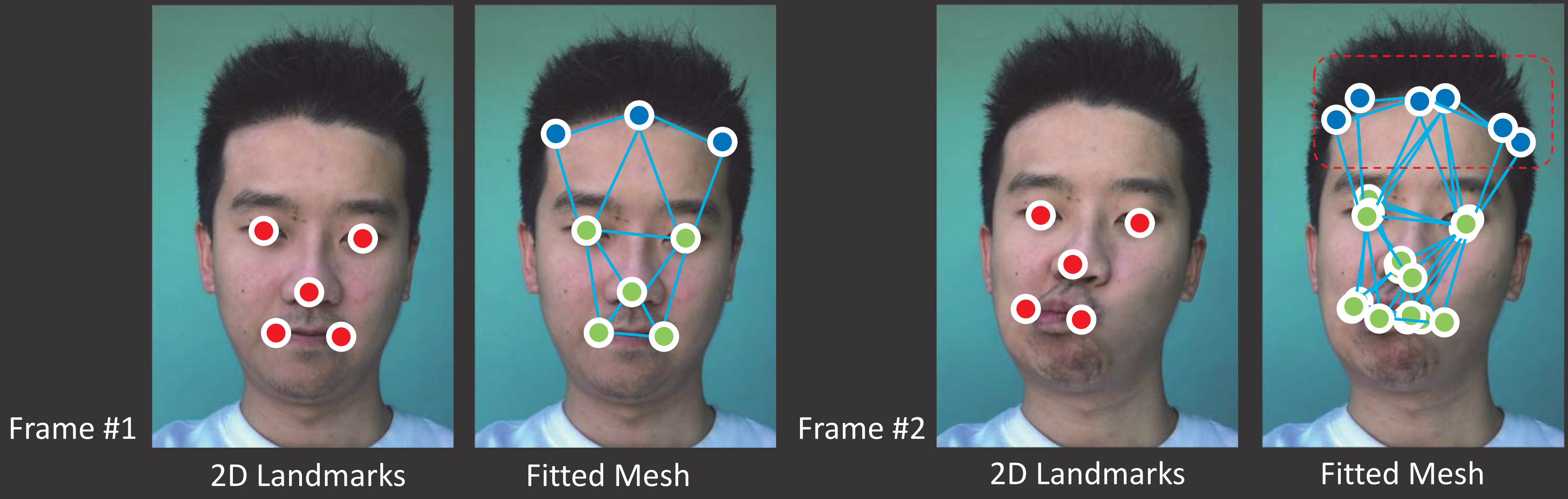
$$\mathbf{F}(\boldsymbol{\beta}) = \mathbf{B}_0 + \sum_i \beta_i \mathbf{B}_i$$

$$\mathbf{P}(\mathbf{T}, \boldsymbol{\beta}) = \boldsymbol{proj}\big(\mathbf{T}\,\mathbf{F}(\boldsymbol{\beta})\big)$$

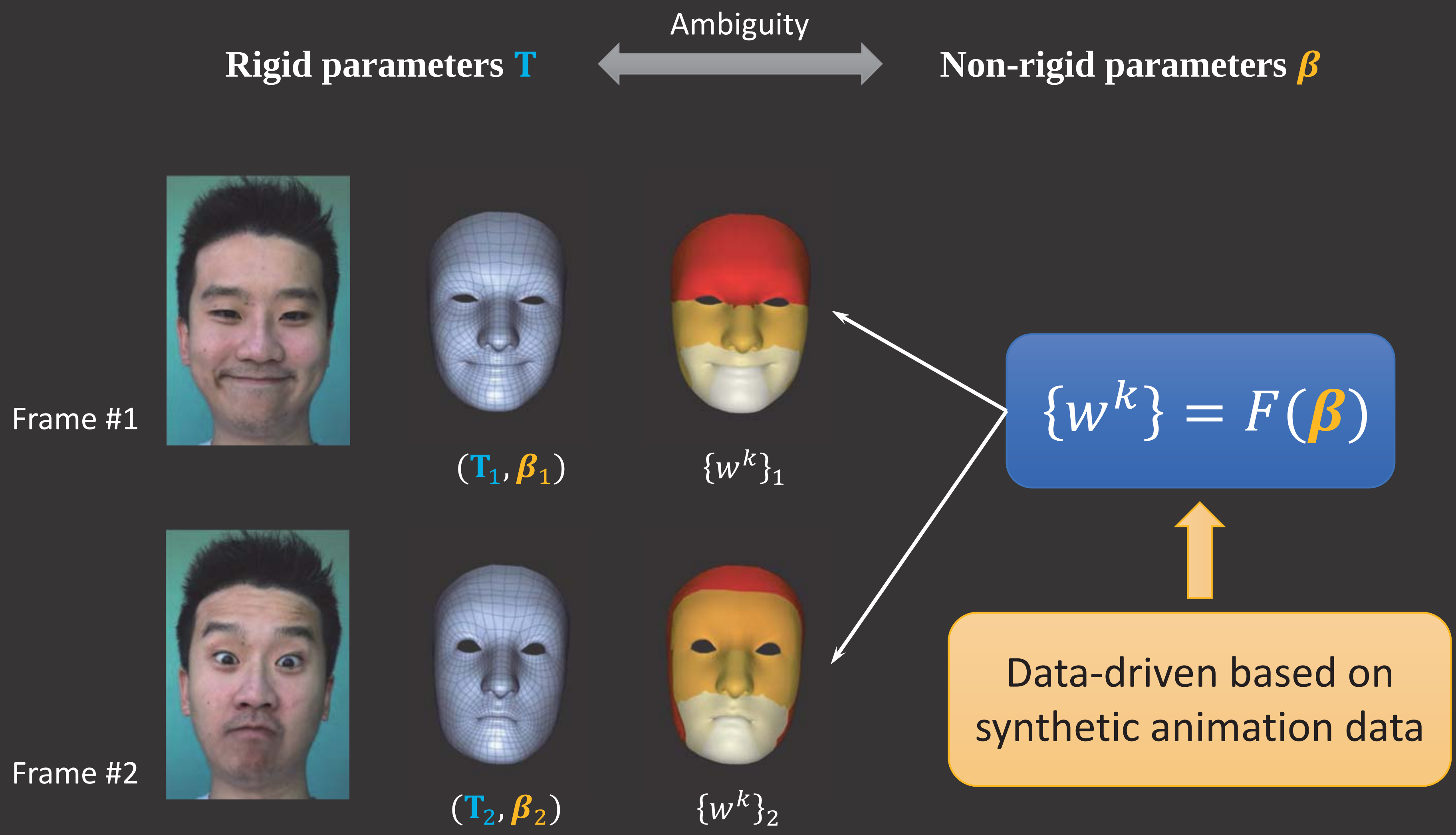
$\mathbf{T}$: rigid transformation

$\boldsymbol{\beta}$: non-rigid expression coefficients

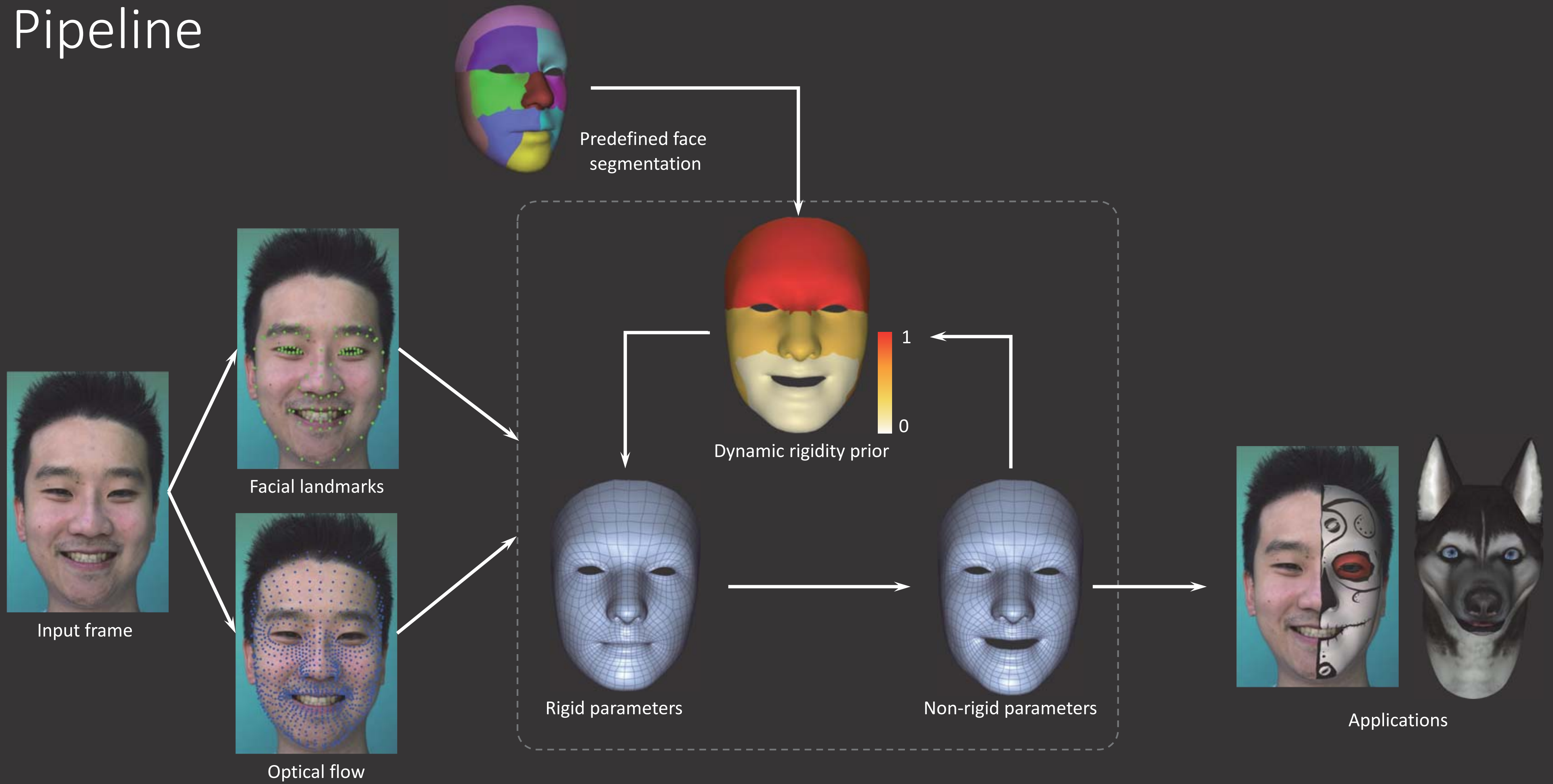
$$\operatorname{argmin}_{\mathbf{T},\boldsymbol{\beta}} \sum_i \left\| \mathbf{P}(\mathbf{T}, \boldsymbol{\beta})^{(v_i)} - L_i \right\|^2$$



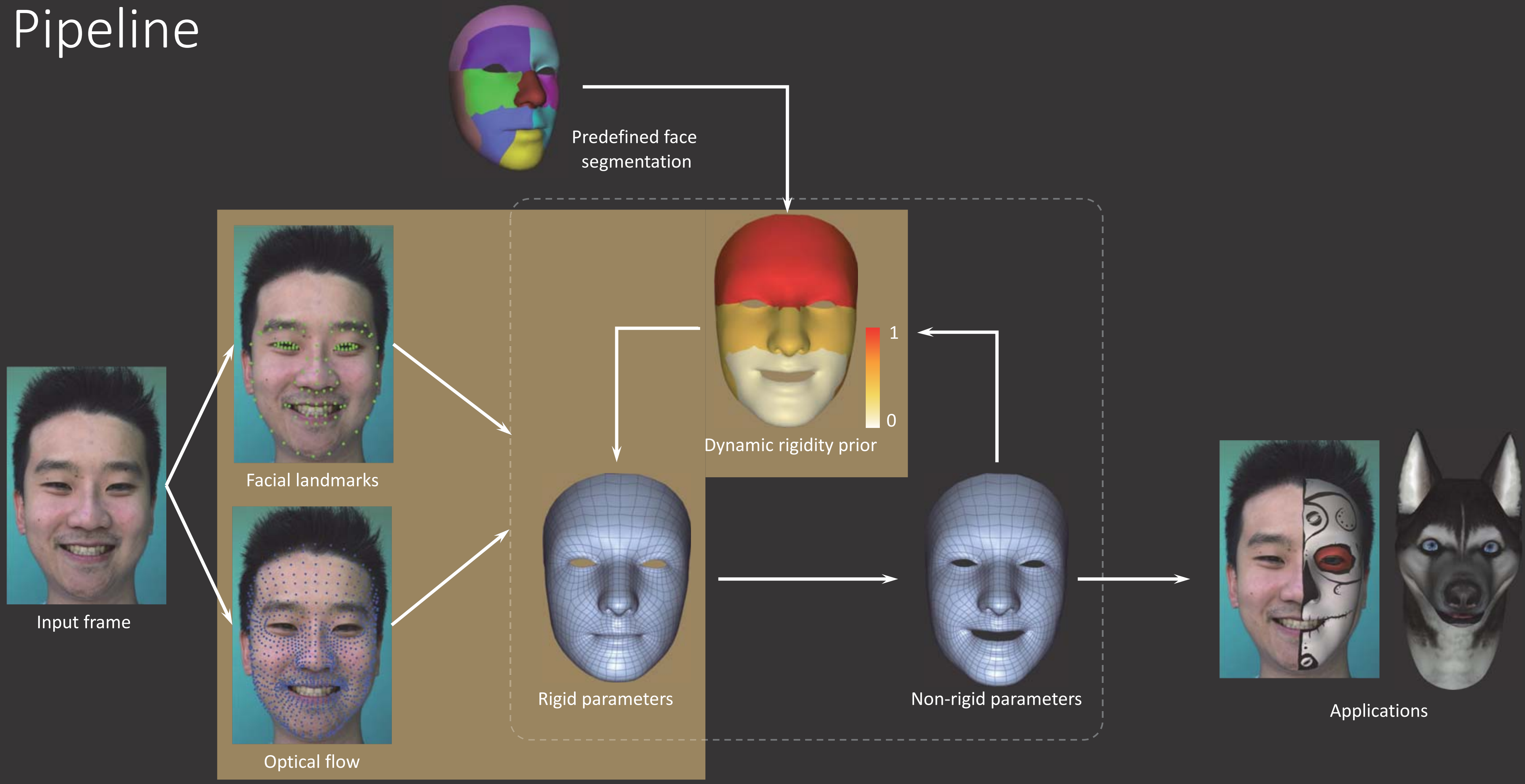
Our Solution



Pipeline


Predefined face
segmentation
Dynamic rigidity prior
1
0
Facial landmarks
Input frame
Optical flow
Rigid parameters
Non-rigid parameters
Applications

Pipeline



Predefined face segmentation
Input frame
Facial landmarks
Optical flow
Dynamic rigidity prior
1
0
Rigid parameters
Non-rigid parameters
Applications

Rigid estimation

$$\mathbf{P}(\mathbf{T}, \boldsymbol{\beta}) = \boldsymbol{proj}\big(\mathbf{T}\, \mathbf{F}(\boldsymbol{\beta})\big)$$

$\mathbf{F}(\boldsymbol{\beta})$

$\{L_i\}$

Facial landmarks

$$\mathrm{argmin}_{\mathbf{T}} \sum_i \left\| \mathbf{P}(\mathbf{T}, \boldsymbol{\beta})^{(l_i)} - L_i \right\|^2$$

$\{D_i\}$

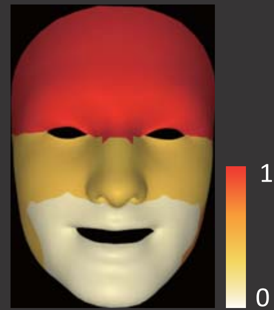
Optical flow

$$\mathrm{argmin}_{\mathbf{T}} \sum_i \left\| \mathbf{P}(\mathbf{T}, \boldsymbol{\beta})^{(d_i)} - D_i \right\|^2$$

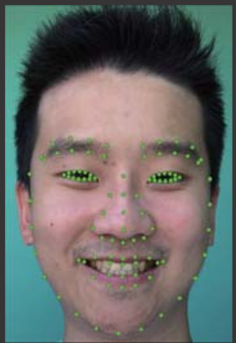
Rigid estimation

$$\mathbf{P}(\mathbf{T}, \boldsymbol{\beta}) = \boldsymbol{proj}\big(\mathbf{T}\ \mathbf{F}(\boldsymbol{\beta})\big)$$

$\mathbf{F}(\boldsymbol{\beta})$



$\{w^k\}$



$\{L_i\}$

Facial landmarks

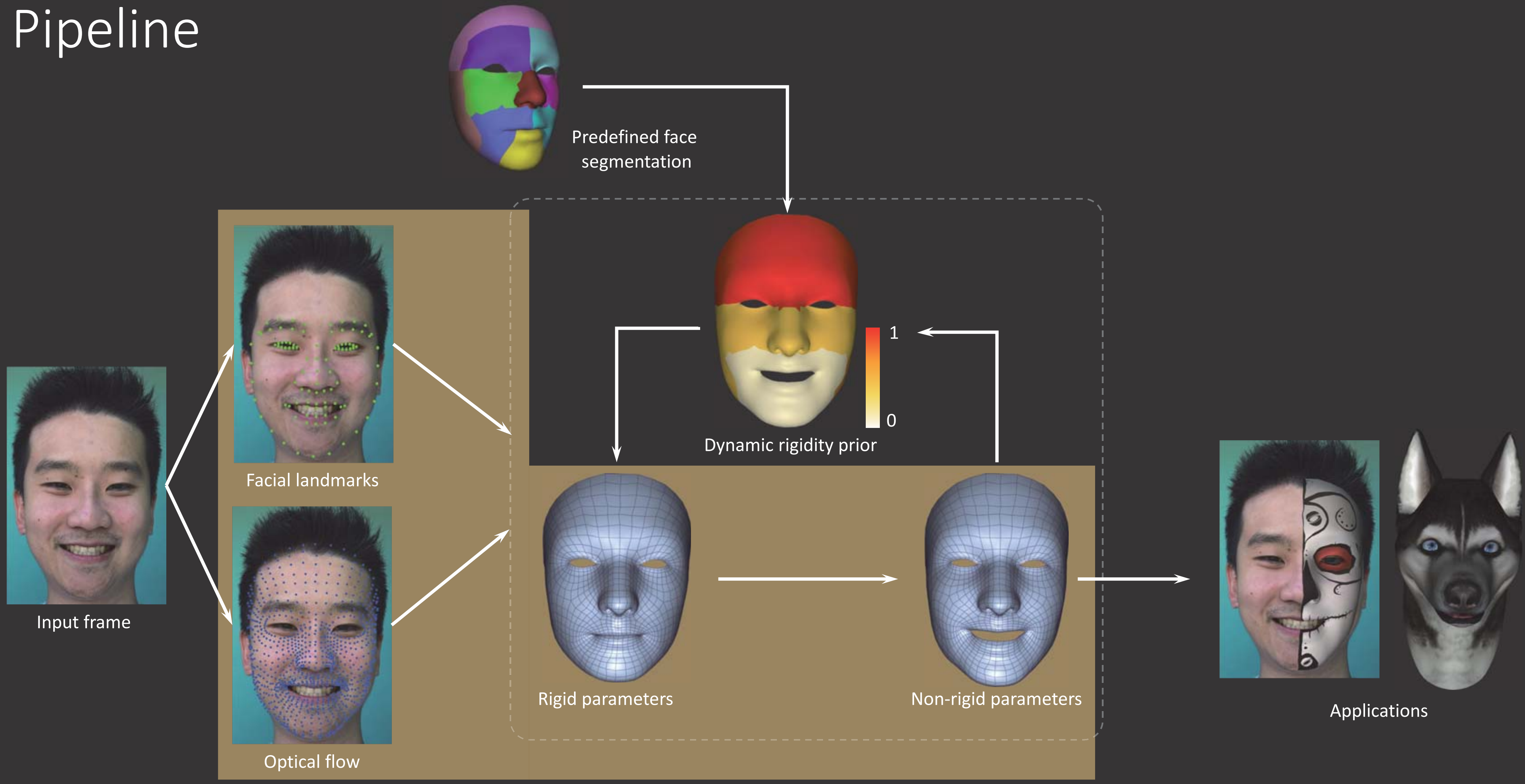
$$\mathrm{argmin}_{\mathbf{T}} \sum_k w^k \sum_{i \in \mathcal{L}^k} \left\| \mathbf{P}(\mathbf{T}, \boldsymbol{\beta})^{(\boldsymbol{l}_i)} - L_i \right\|^2$$

$\{D_i\}$

Optical flow

$$\mathrm{argmin}_{\mathbf{T}} \sum_k w^k \sum_{i \in \Gamma^k} \left\| \mathbf{P}(\mathbf{T}, \boldsymbol{\beta})^{(\boldsymbol{d}_i)} - D_i \right\|^2$$

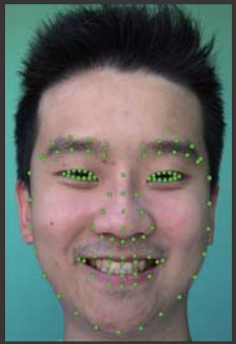
Pipeline


Predefined face segmentation
Dynamic rigidity prior
1
0
Input frame
Facial landmarks
Optical flow
Rigid parameters
Non-rigid parameters
Applications

Non-rigid estimation

$$\mathbf{P}(\mathbf{T}, \boldsymbol{\beta}) = \boldsymbol{proj}\big(\mathbf{T}\ \mathbf{F}(\boldsymbol{\beta})\big)$$

$\mathbf{F}(\boldsymbol{\beta})$



$\{L_i\}$

Facial landmarks

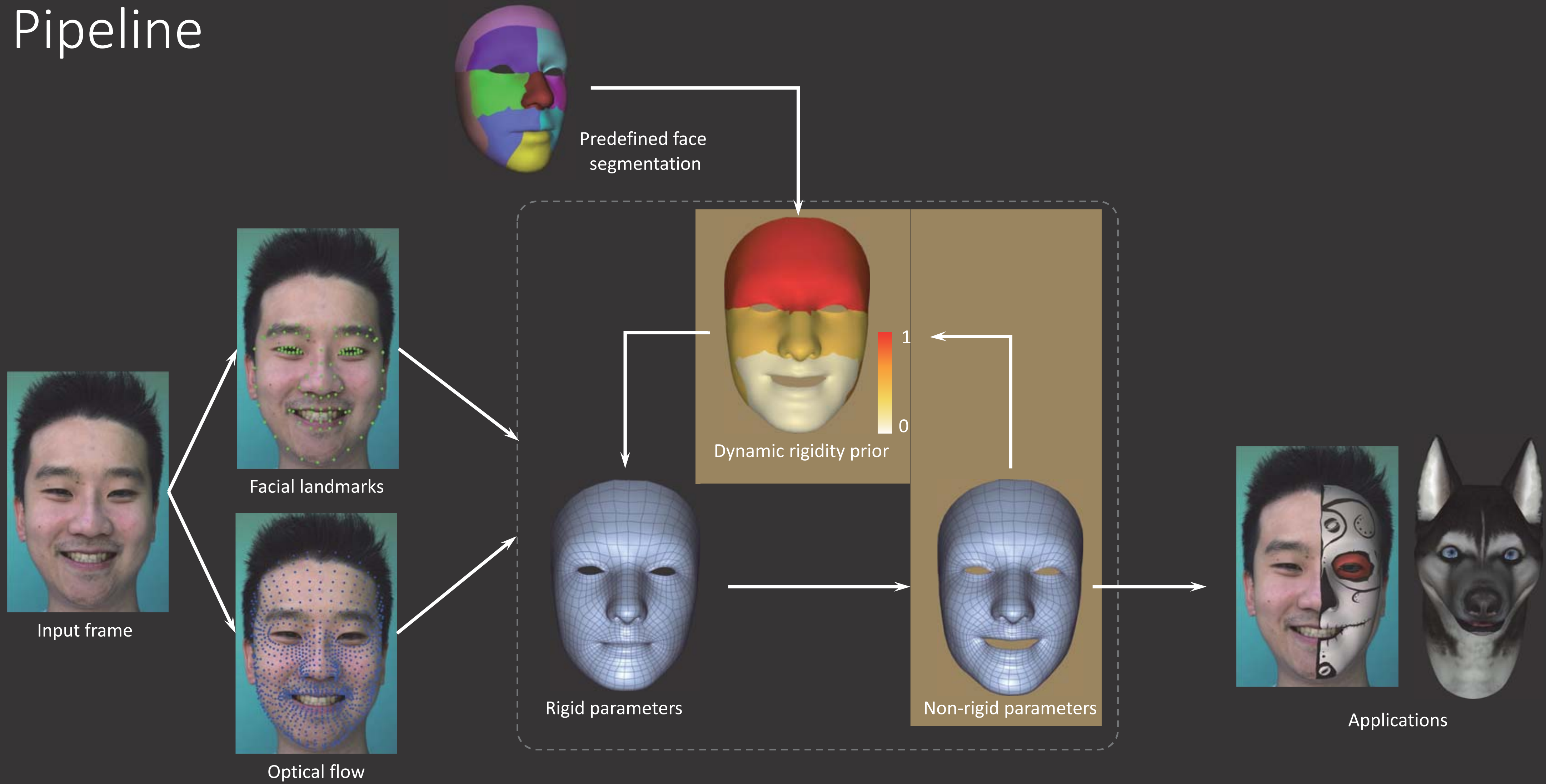
$$\mathrm{argmin}_{\boldsymbol{\beta}} \sum_i \left\| \mathbf{P}(\mathbf{T}, \boldsymbol{\beta})^{(\boldsymbol{l_i})} - L_i \right\|^2$$

$\{D_i\}$

Optical flow

$$\mathrm{argmin}_{\boldsymbol{\beta}} \sum_i \left\| \mathbf{P}(\mathbf{T}, \boldsymbol{\beta})^{(\boldsymbol{d_i})} - D_i \right\|^2$$

Pipeline

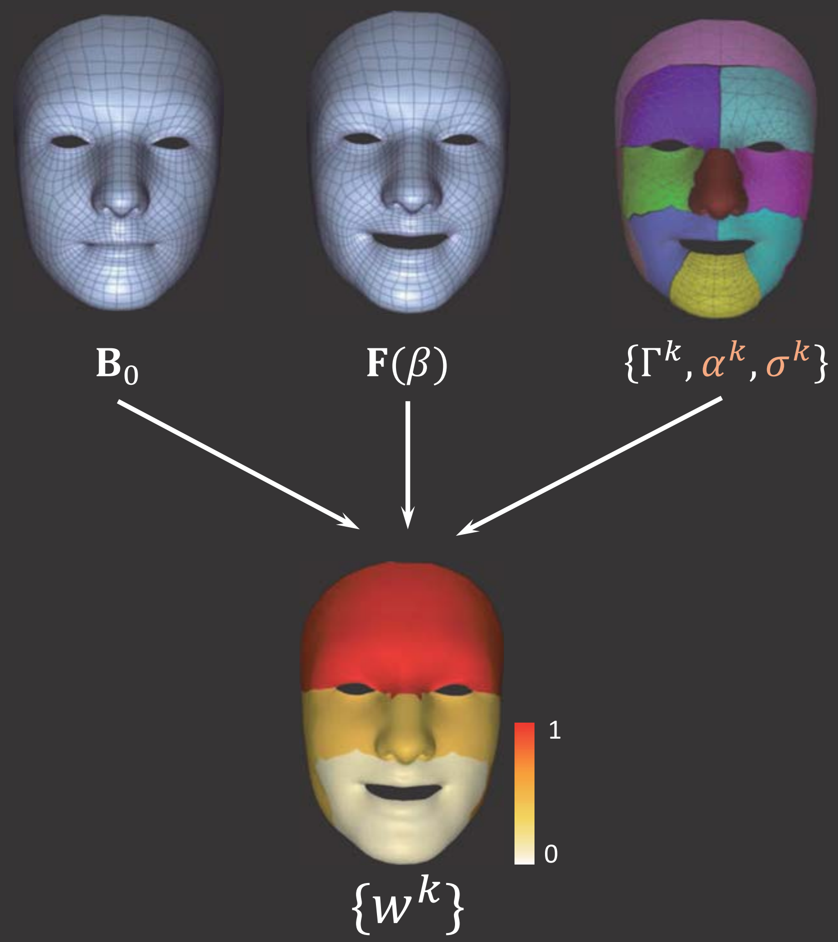

Predefined face segmentation
Dynamic rigidity prior
1
0
Facial landmarks
Input frame
Optical flow
Rigid parameters
Non-rigid parameters
Applications

Dynamic Rigidity Prior

- Formulation
  - Based on the motion of each region $\mathbf{F}^k(\boldsymbol{\beta})$ compared to its neutral expression base shape $\mathbf{B}_0^k$

$$w^k = \alpha^k \exp\left(-\frac{\left\|\mathbf{F}^k(\boldsymbol{\beta}) - \mathbf{B}_0^k\right\|^2}{(\sigma^k)^2|\Gamma^k|}\right)$$

- Learned hyper-parameters $\{\alpha^k, \sigma^k\}$
  - Data-driven method



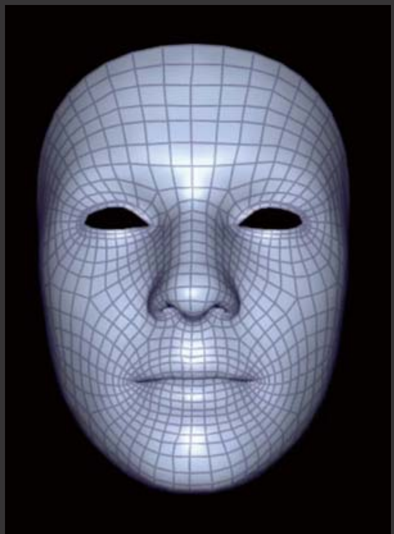
Dynamic Rigidity Prior

- Learning hyper-parameters $\{\alpha^k, \sigma^k\}$
  - Training data with ground truth $\mathbf{T}^*$ and $\boldsymbol{\beta}^*$
  - Synthesized data

$$\arg\min_{\alpha^k, \sigma^k} \|\mathbf{T} - \mathbf{T}^*\|^2$$

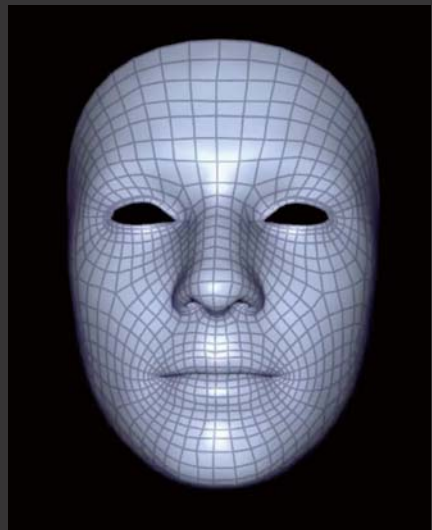
Estimated $\mathbf{T}, \boldsymbol{\beta}$

Estimate



$\mathbf{F}(\boldsymbol{\beta}^*)$

$\mathbf{T}^*$



$\mathbf{T}^*\mathbf{F}(\boldsymbol{\beta}^*)$

Project

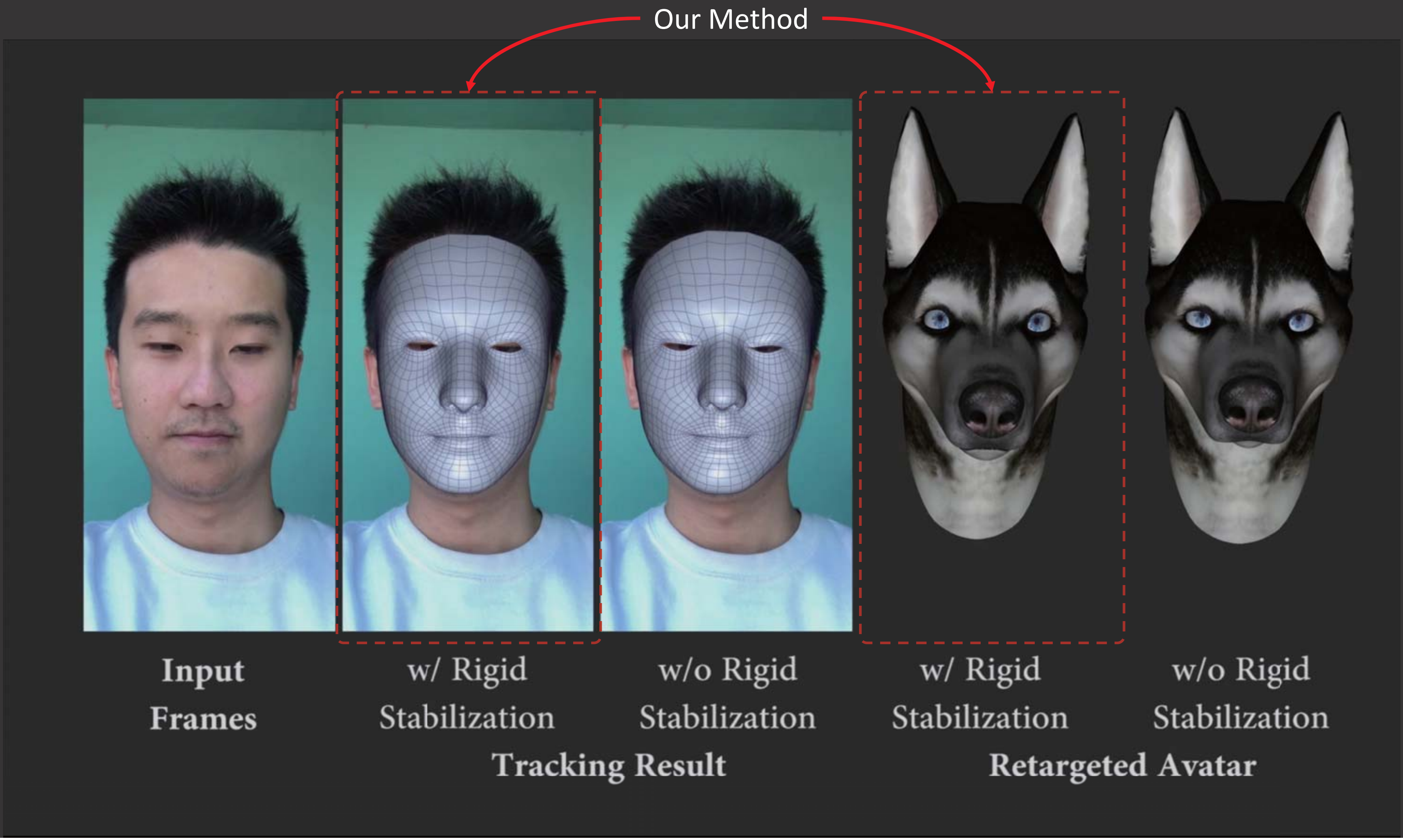
Perturb

Facial landmarks $\{L_i\}$

Optical flow $\{D_i\}$

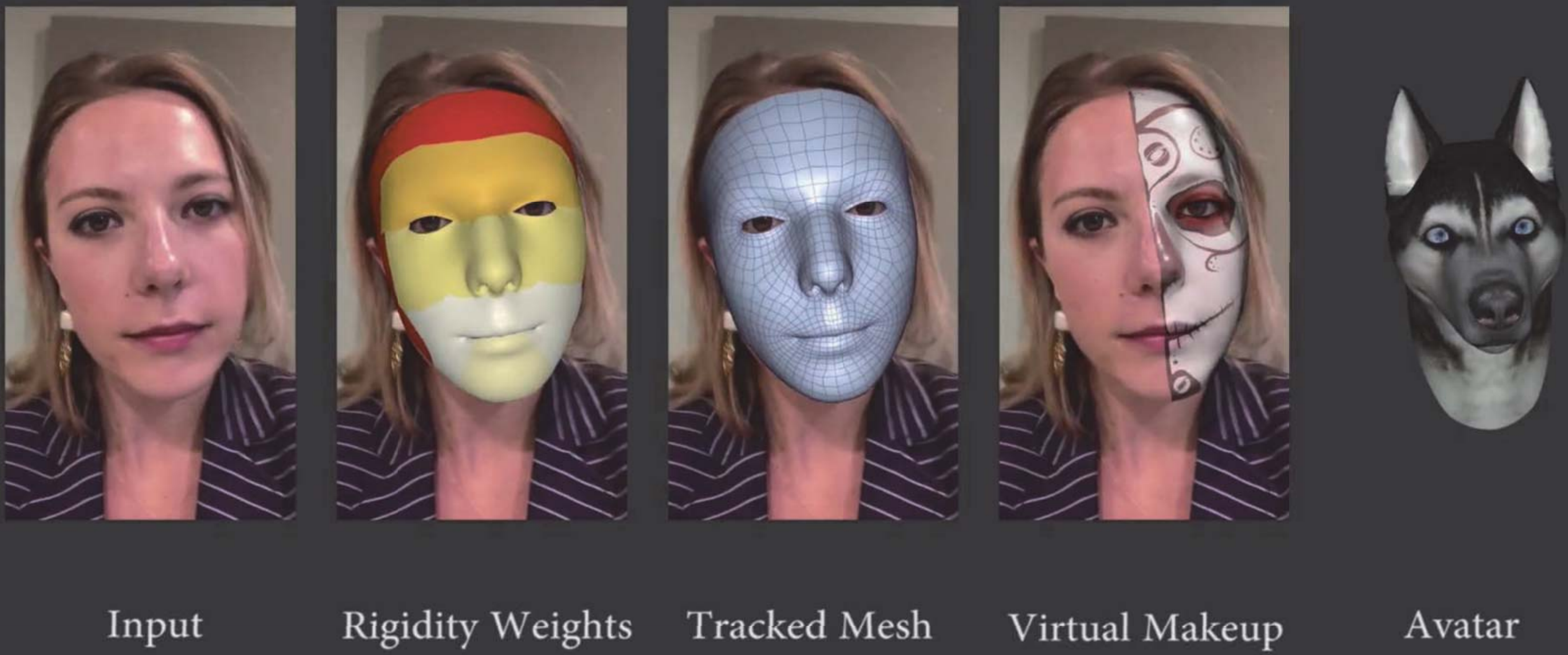
Perturbed parameters $\{\tilde{\mathbf{T}}, \tilde{\boldsymbol{\beta}}\}$

Results



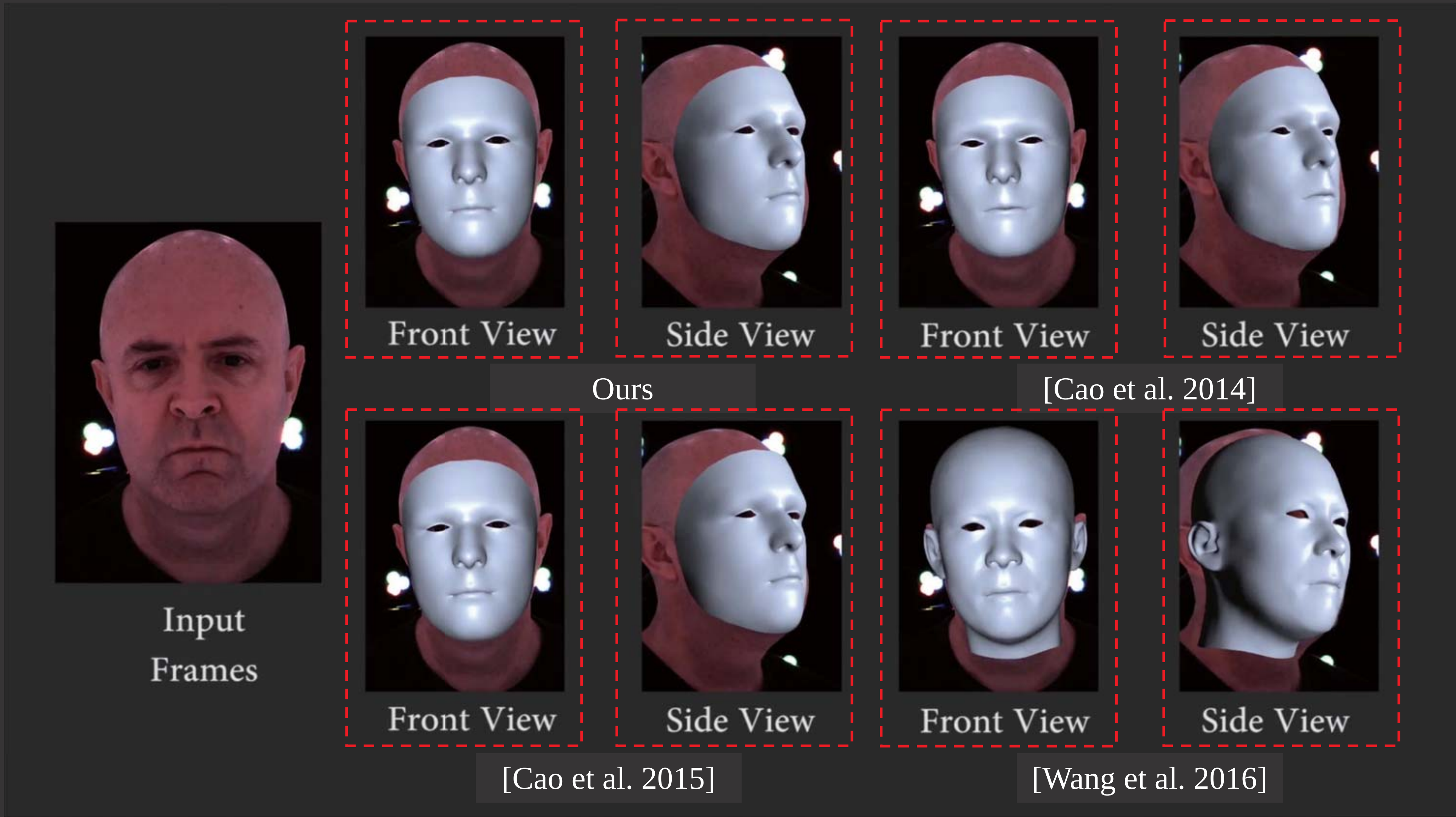
Our Method
Input
Frames
w/ Rigid
Stabilization
w/o Rigid
Stabilization
Tracking Result
w/ Rigid
Stabilization
w/o Rigid
Stabilization
Retargeted Avatar

Results

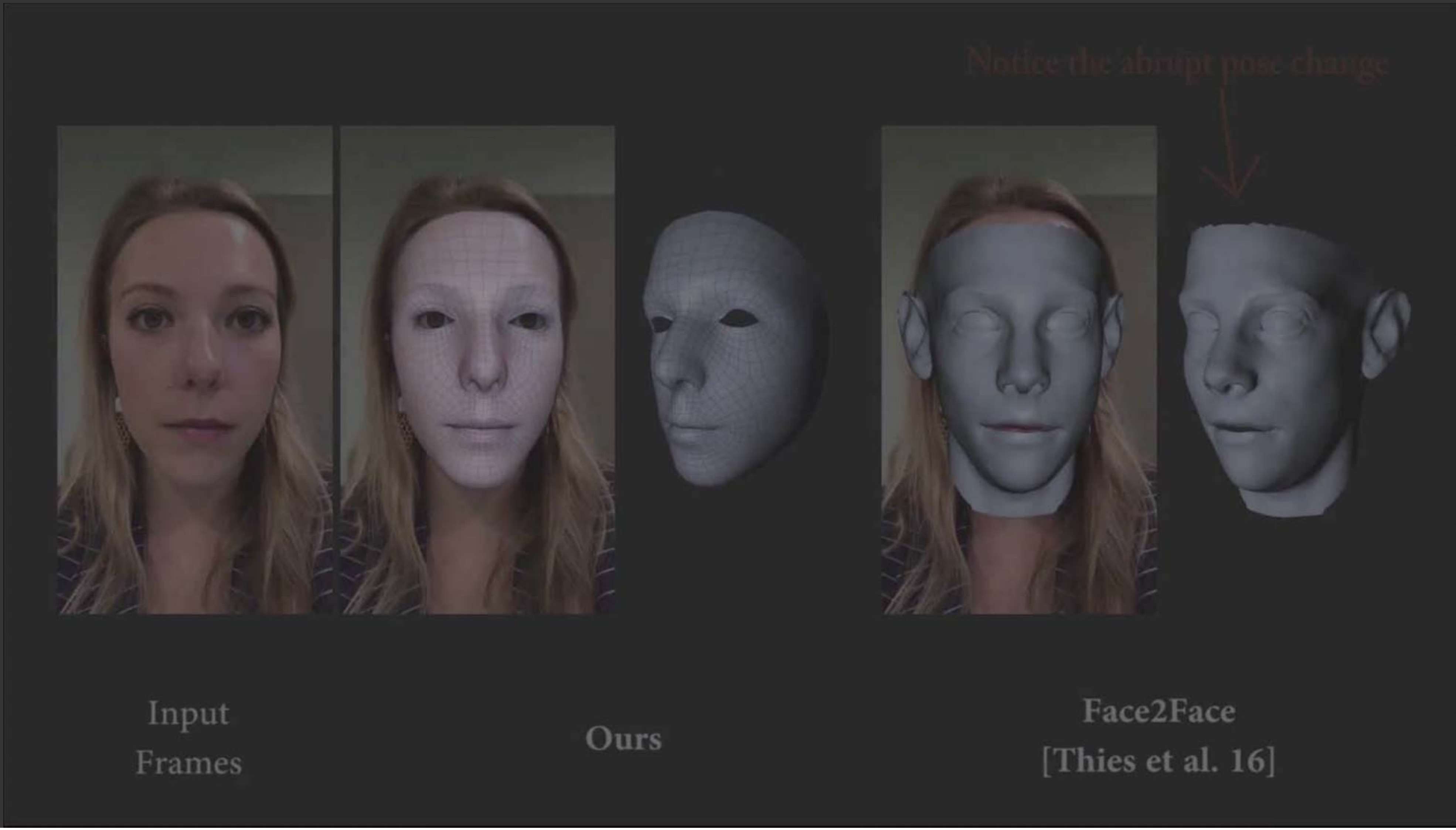

Input
Rigidity Weights
Tracked Mesh
Virtual Makeup
Avatar

Live Demo

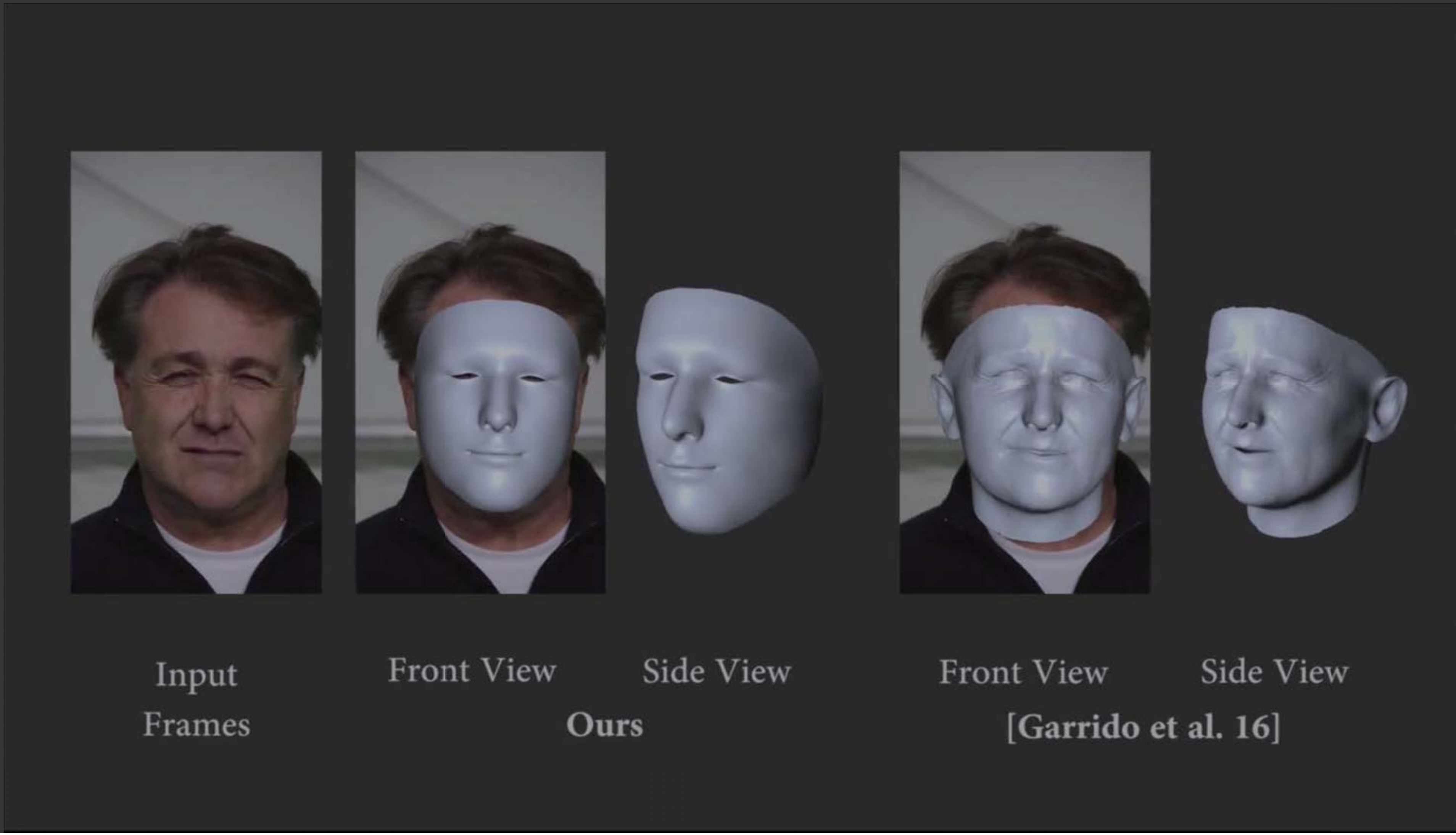
Results: comparisons with other methods



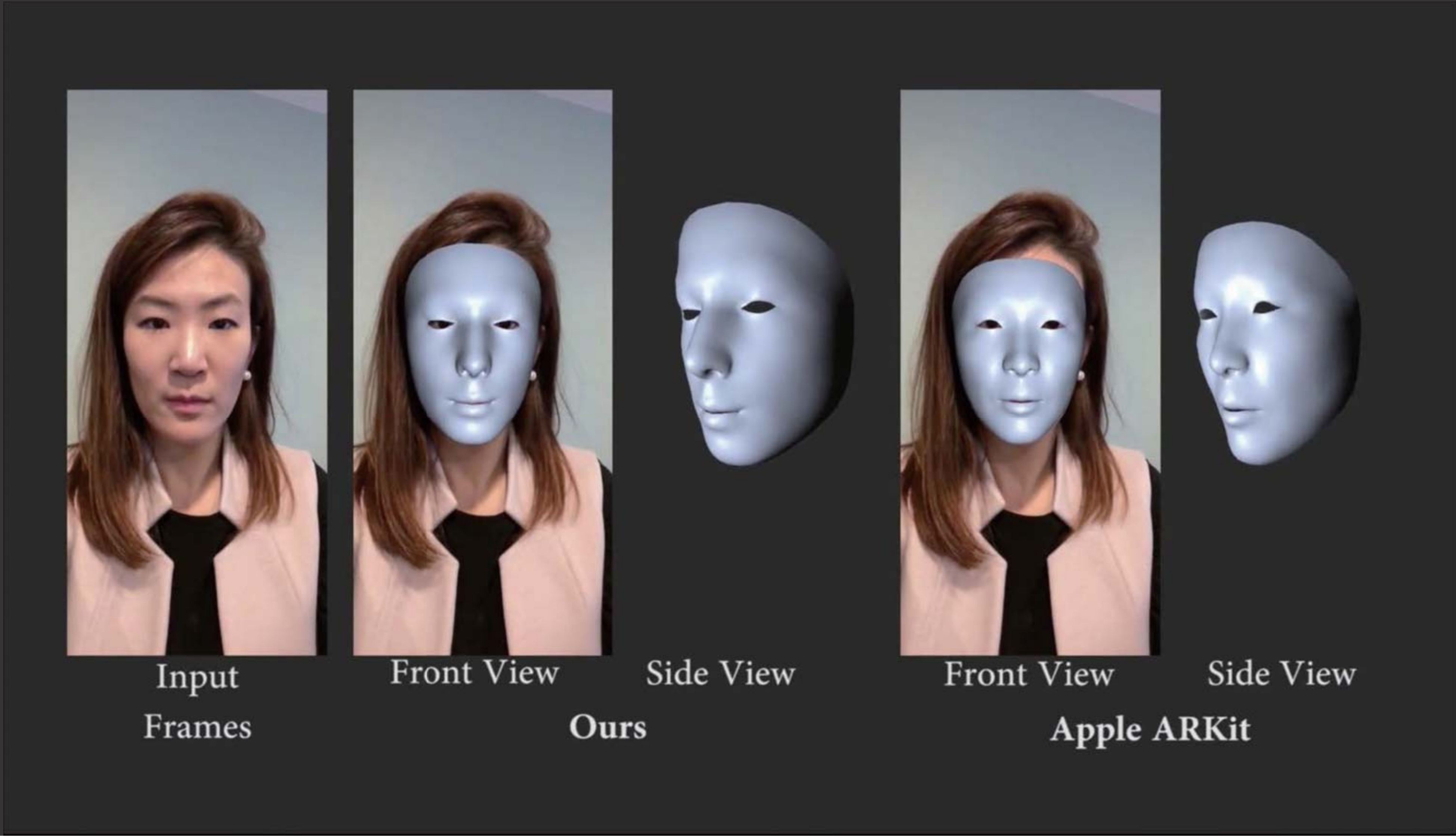
Results: comparisons with [Thies et al. 2016] (Face2Face)



Results: comparisons with [Garrido et al. 2016]



Results: comparisons with Apple ARKit



Conclusions

- Goal: *Real-time* 3D face tracking and animation for *everyone*
  - Ease of use
    - Ordinary RGB camera (RGB-D camera), no any special equipment
    - Mobile device: limited computation power and memory
    - Real-time performance
  - Robustness
    - Different users
    - Fast motions, large head pose, exaggerated facial expressions
    - Lighting/background changes
  - Accurate and Realistic
    - Avoid uncanny valley

Thank you!

2019 intern opportunity in Los Angeles
Please send resume to: zju.caochen@gmail.com