Real-time Motion Capture Facial Animation

Catarina Runa Miranda
Verónica Costa Orvalho

PORTO INTERACTIVE CENTER

U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

instituto de telecomunicações

Overview

- Introduction
- MoCap Fundamental Science
- Facial MoCap Tracking
- MoCap Facial Animation
- MoCap VR Methods
- Contributions
- Conclusion

Introduction

- Main results
- Motivation
- Problem Statement
- Goal
- Framework

Goal

Research and develop methods for

non-expert users to **recognize** facial movements **non-intrusively** and **map** them to a 3D character **on-the-fly**

Motivation

- LIFEisGAME and VERE projects' goal:

Markerless and Real-time Facial Animation of 3D characters using off-the-shelf hardware

Problem Statement

Realistic Facial animation

labor-intensive & **expert dependent**

Ryse Son of Rome, Crytek (2013)

Problem Statement

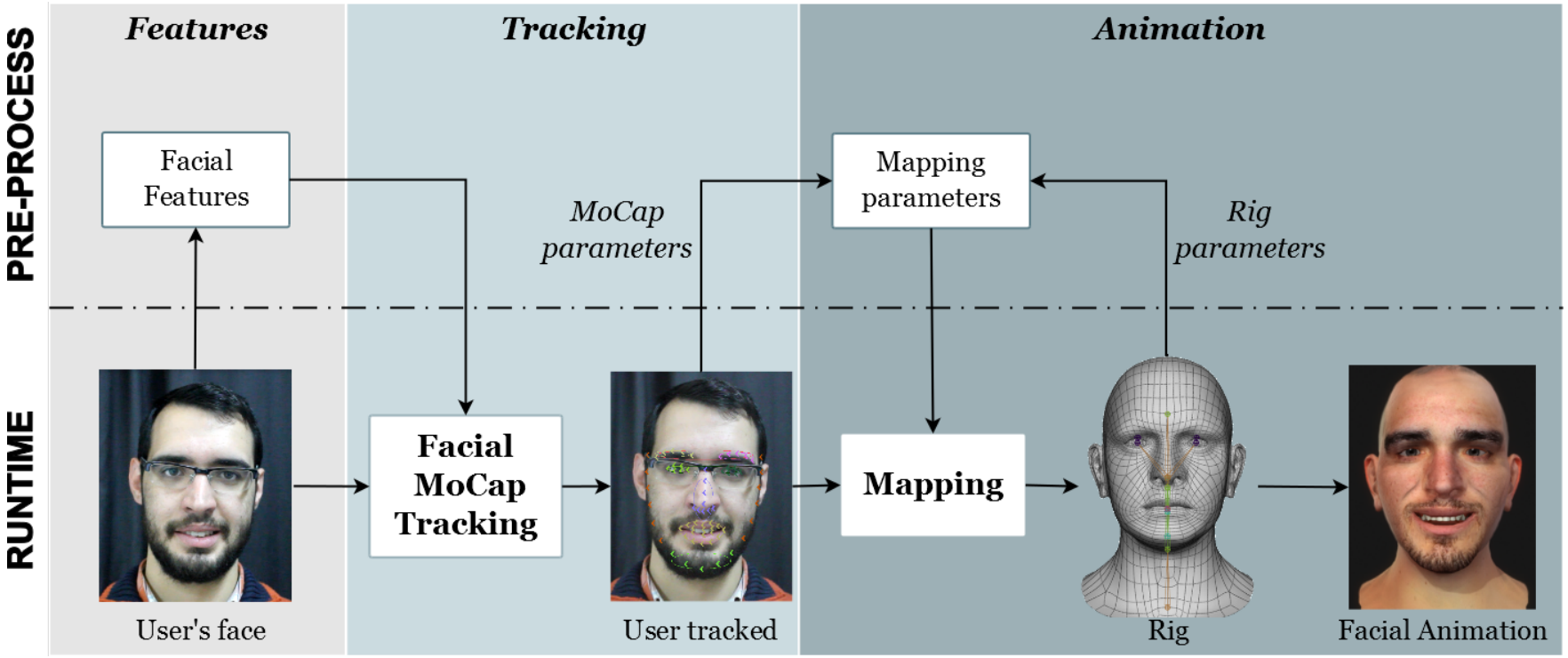
MoCap Facial animation solutions are

not suitable for general user

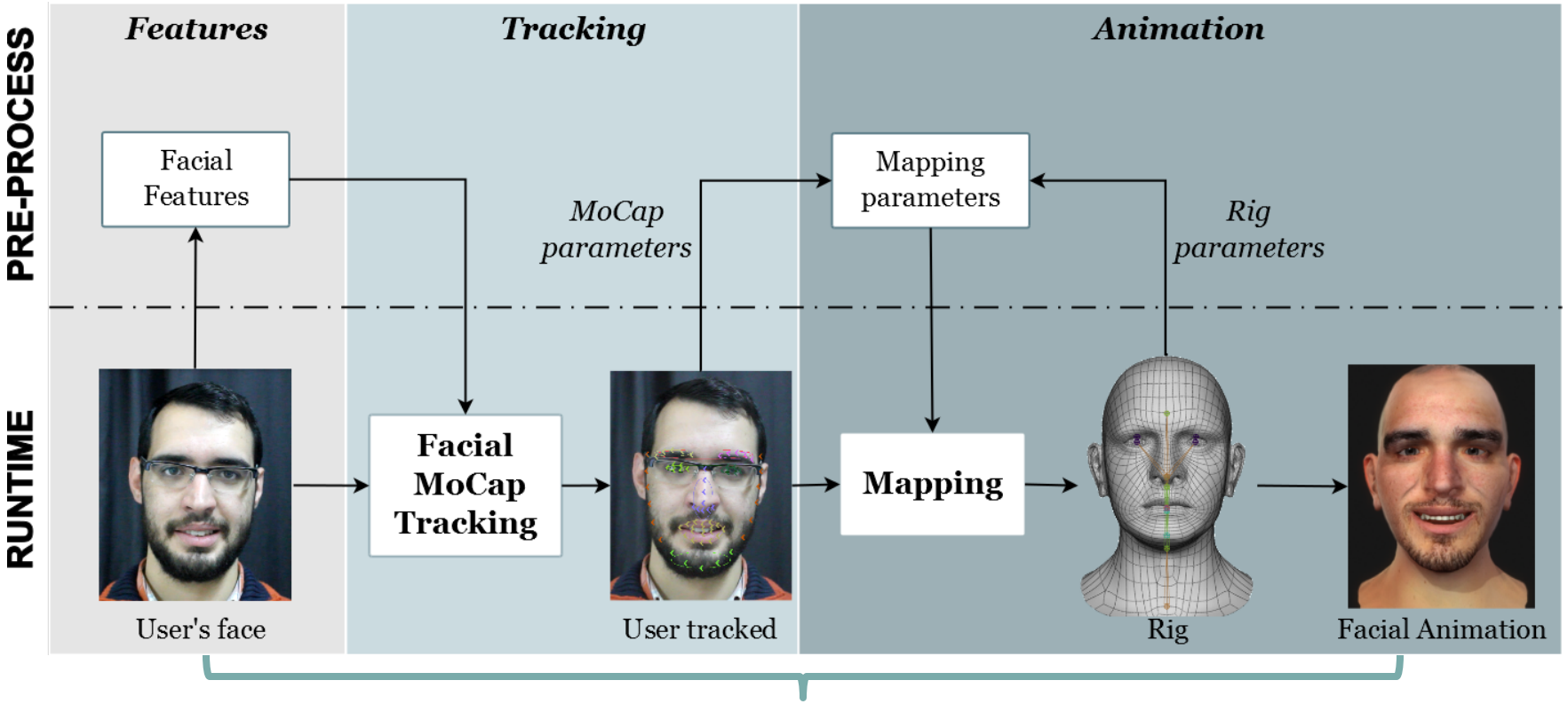
Graham Fyffe et al. Driving High-Resolution Facial Scans with Video Performance Capture (2014)

expensive setups & **complex calibrations** & **not compatible to VR environments**

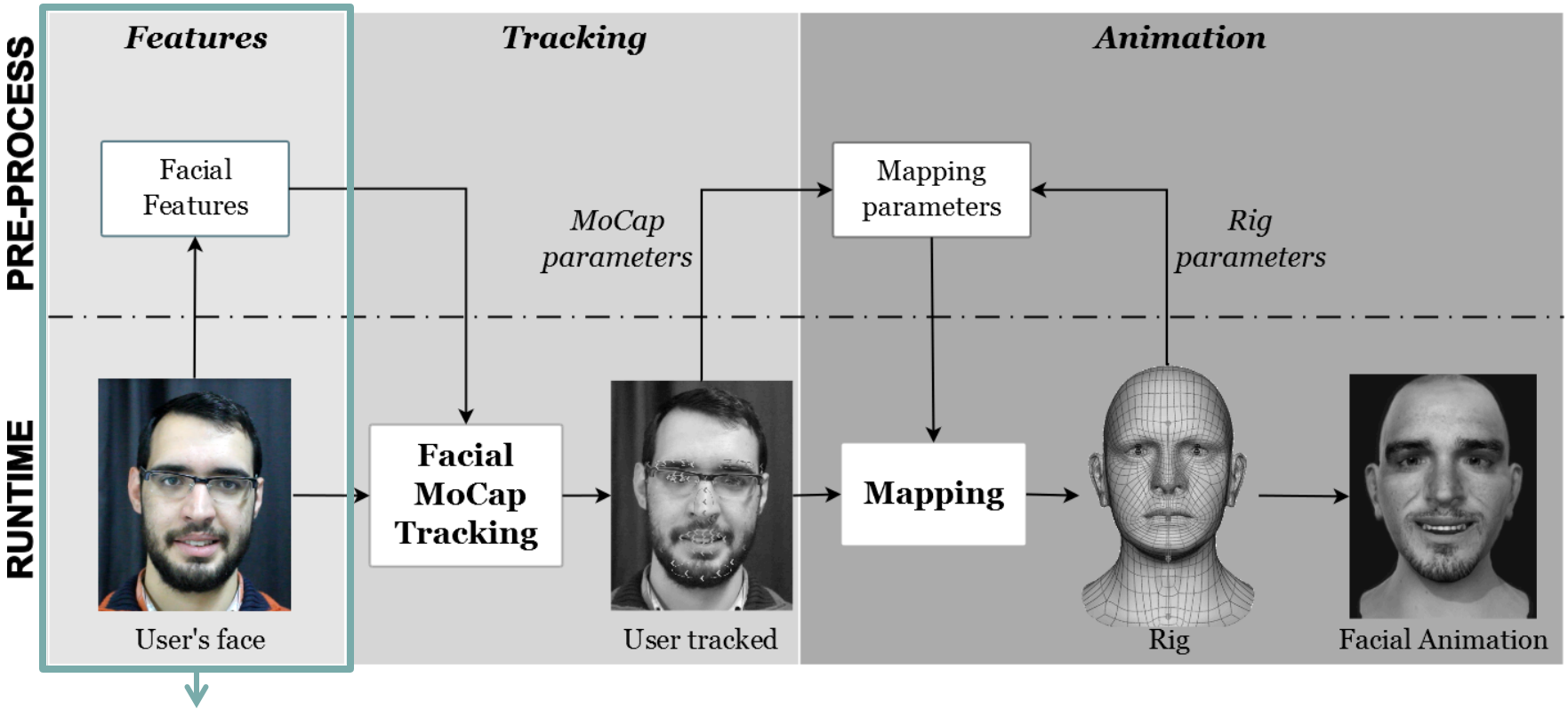
Framework


Features
Tracking
Animation
PRE-PROCESS
RUNTIME
Facial Features
MoCap parameters
Mapping parameters
Rig parameters
Facial MoCap Tracking
Mapping
User's face
User tracked
Rig
Facial Animation

Framework



Independent solutions
for real-time MoCap facial animation



Define **which features** need to be tracked and mapped to 3D character

MoCap Fundamental Science

- Face Image Task: Self-perception of Facial Features
- Real-time Emotion Recognition

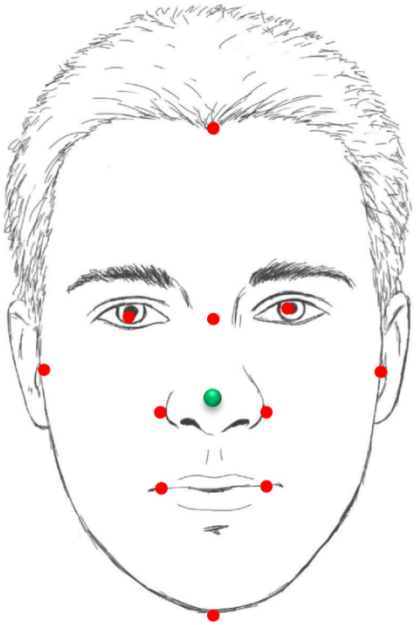
Study 1:

Face Image Task

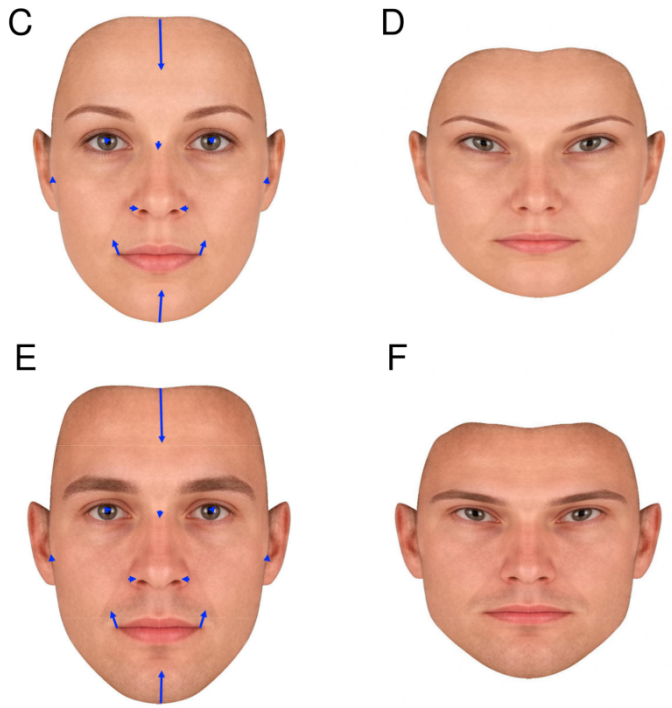
*To understand **how individuals perceive their own facial structure** through the evaluation of their knowledge about the position of key facial features*

Face Image Task:
Experiment overview

50 participants indicated the location of key features (red) of their own face relative to anchor point (green)



Face Image Task: Conclusions



average horizontal (x) and vertical (y) error

- Human's spacial **perception** of his own face using 11 key features is **poor**
- High loadings on the upper face accompanied by low loadings of the lower face, or vice versa

MoCap Tracker
VR MoCap Methods

Study 2:

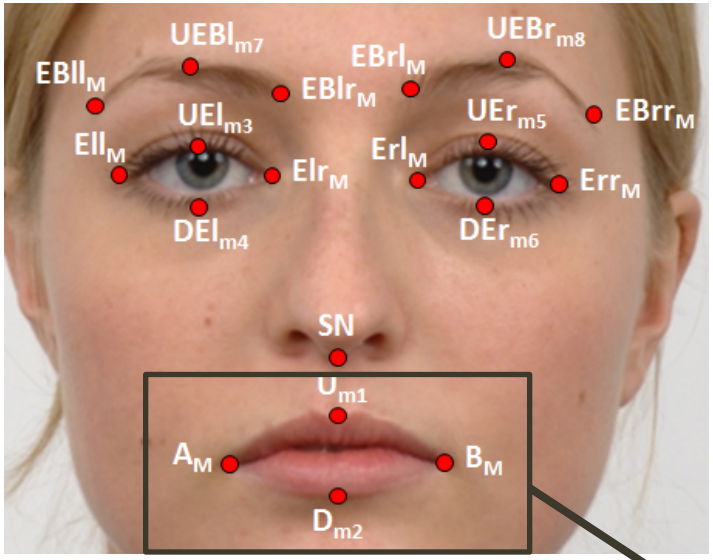
Real-time Emotion Recognition:

*which **facial features characterize the six universal emotions***

+

*Real-time geometric features **extraction** and emotion **classification***

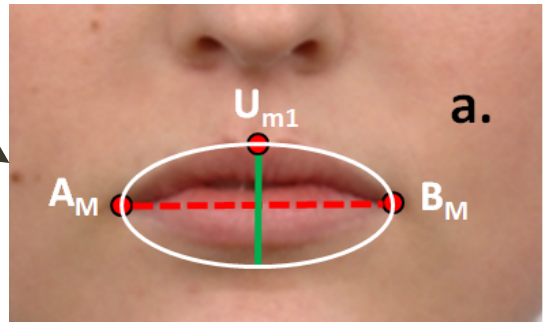
Real-time Emotion Recognition
Geometrical Features Extraction method

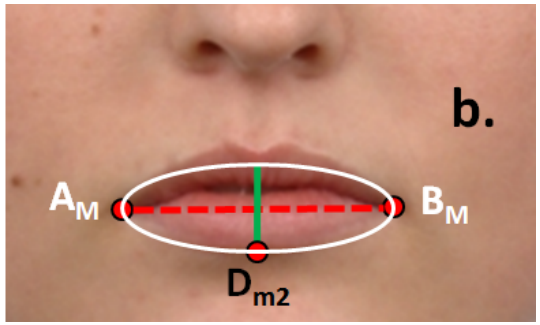


Eccentricity features

$$e = \frac{\sqrt{a^2 - b^2}}{a} \qquad \in [0,1]$$

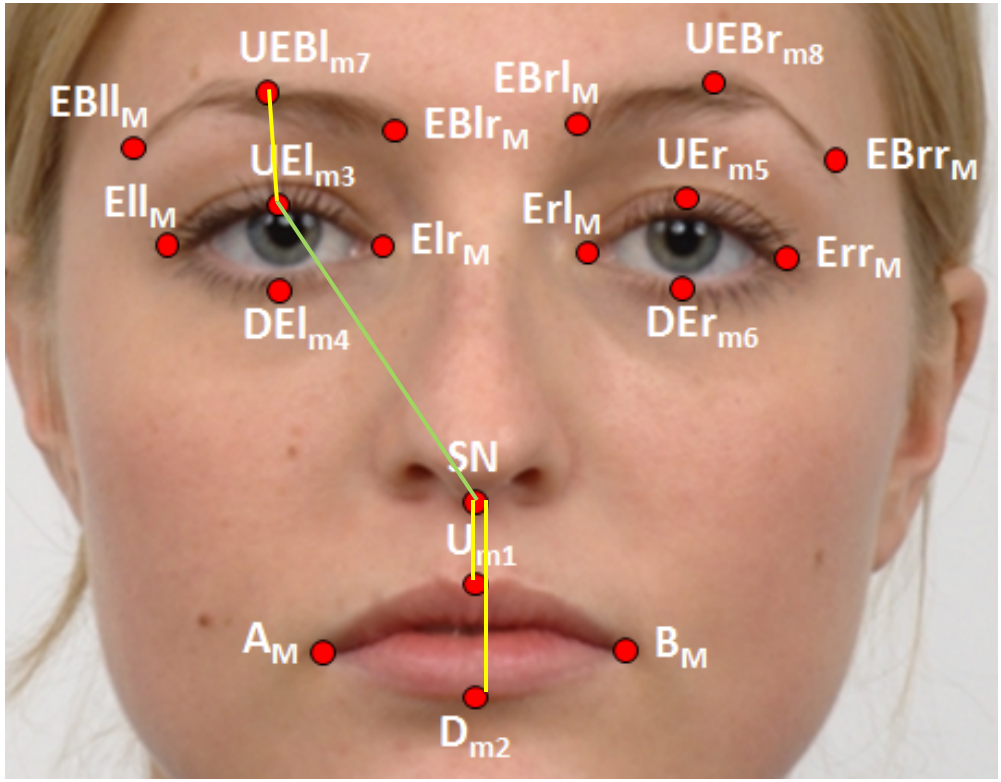
$$a = \frac{B_{Mx} - A_{Mx}}{2} \text{ and } b = A_{My} - U_{m1y}$$





Real-time Emotion Recognition
Geometrical Features Extraction method

Linear features



Real-time Emotion Recognition Results

Comparison with state of the art methods:

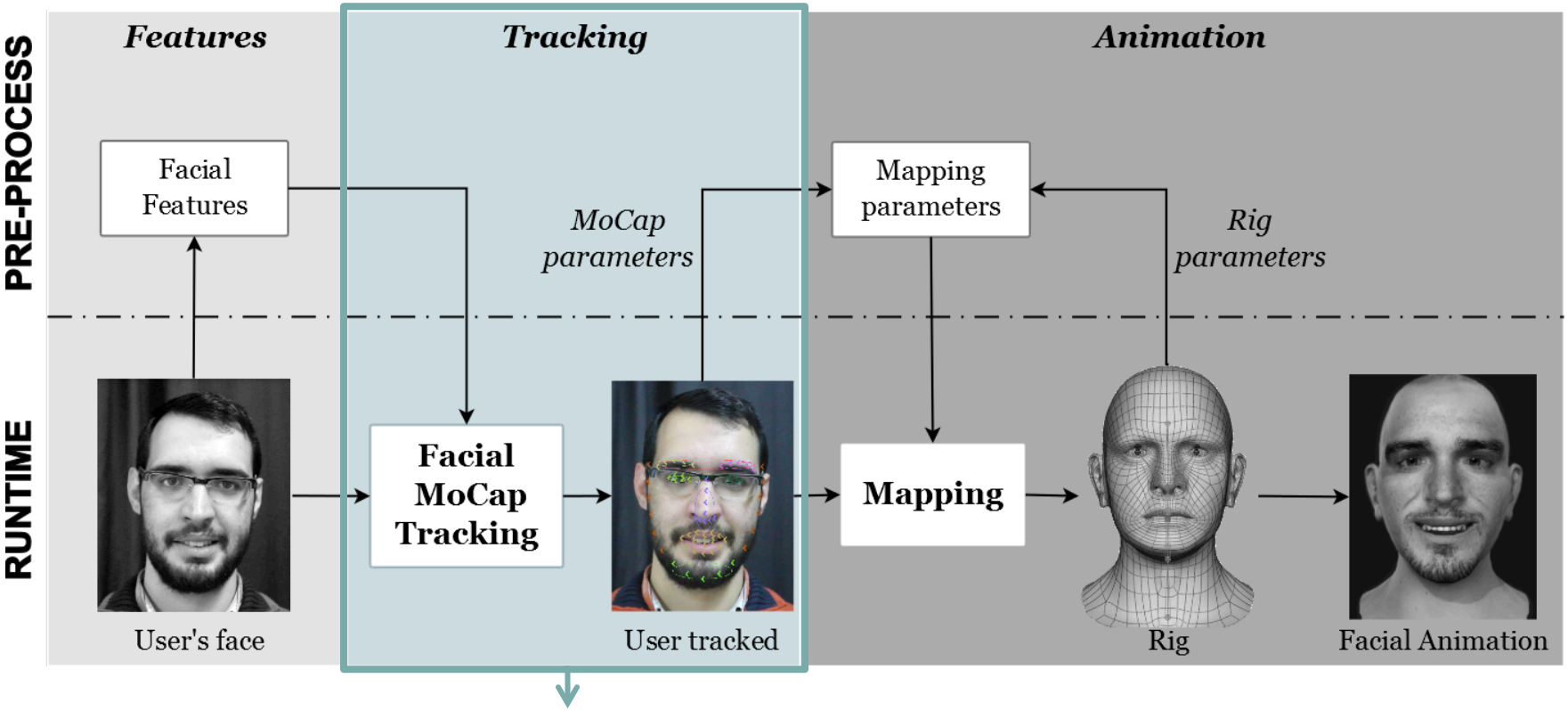
Method	Differential	Accuracy[%]
Michel et al. (Michel and El Kaliouby, 2003)	No	72
Pardas et al. (Pardàs and Bonafonte, 2002)	No	84
Bartlett et al. (Bartlett et al., 2003)	No	84
Our method *S3*	No	**89**
Michel et al. (Michel and El Kaliouby, 2003)	Yes	84
Cohen et al. (Cohen et al., 2003)	Yes	88
Wang et al.(Wang and Yin, 2007)	Yes	93
Our method *S5*	Yes	**94**

Real-time Emotion Recognition
Conclusions

Our geometric method:

- Allows **real-time** feature extraction
- Recognizes 6 universal emotions with **94%** of Presents **higher accuracy than state of the art methods**

VR MoCap Methods



To **track unique facial features**

reducing user manual intervention

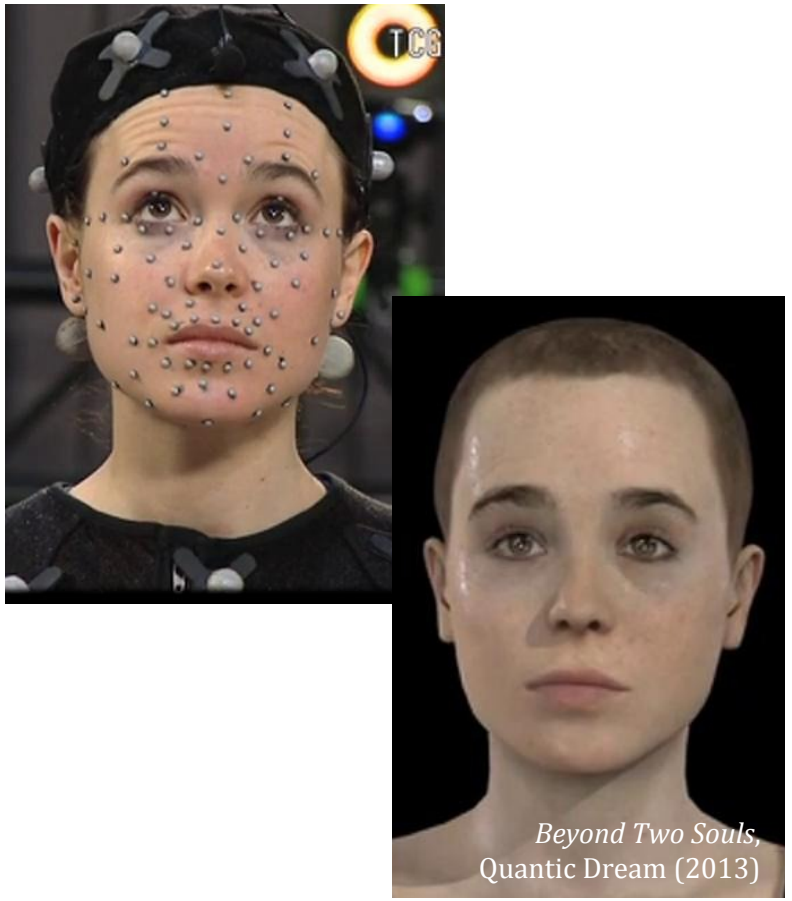
Facial MoCap Tracking

- Background
- Methodology
- Results
- Conclusions

Background

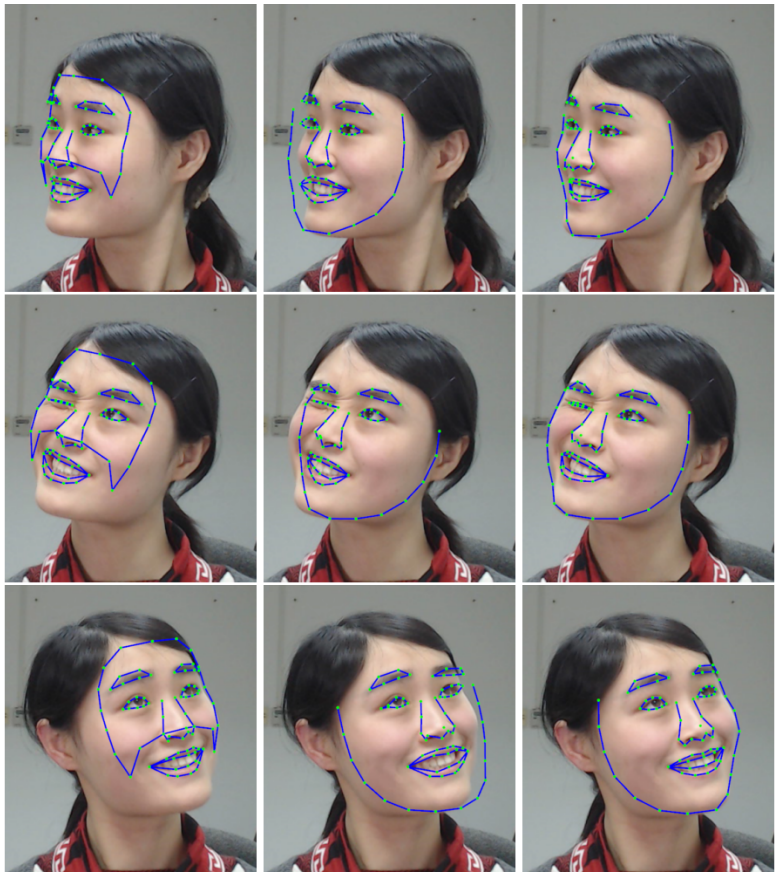
Equipment- based

- Intrusive
- Expert dependent
- Time consuming
- Offline fine tuning



Beyond Two Souls,
Quantic Dream (2013)

Background



Chen Cao et al. Displaced dynamic expression regression (2014)

Markerless

- Less intrusive
- Manual and tedious calibrations
- Model fitting in each frame limits facial movements detected
- Locate only semantic facial features like eyes, mouth, nose, etc
- Not compatible with persistent partial occlusions

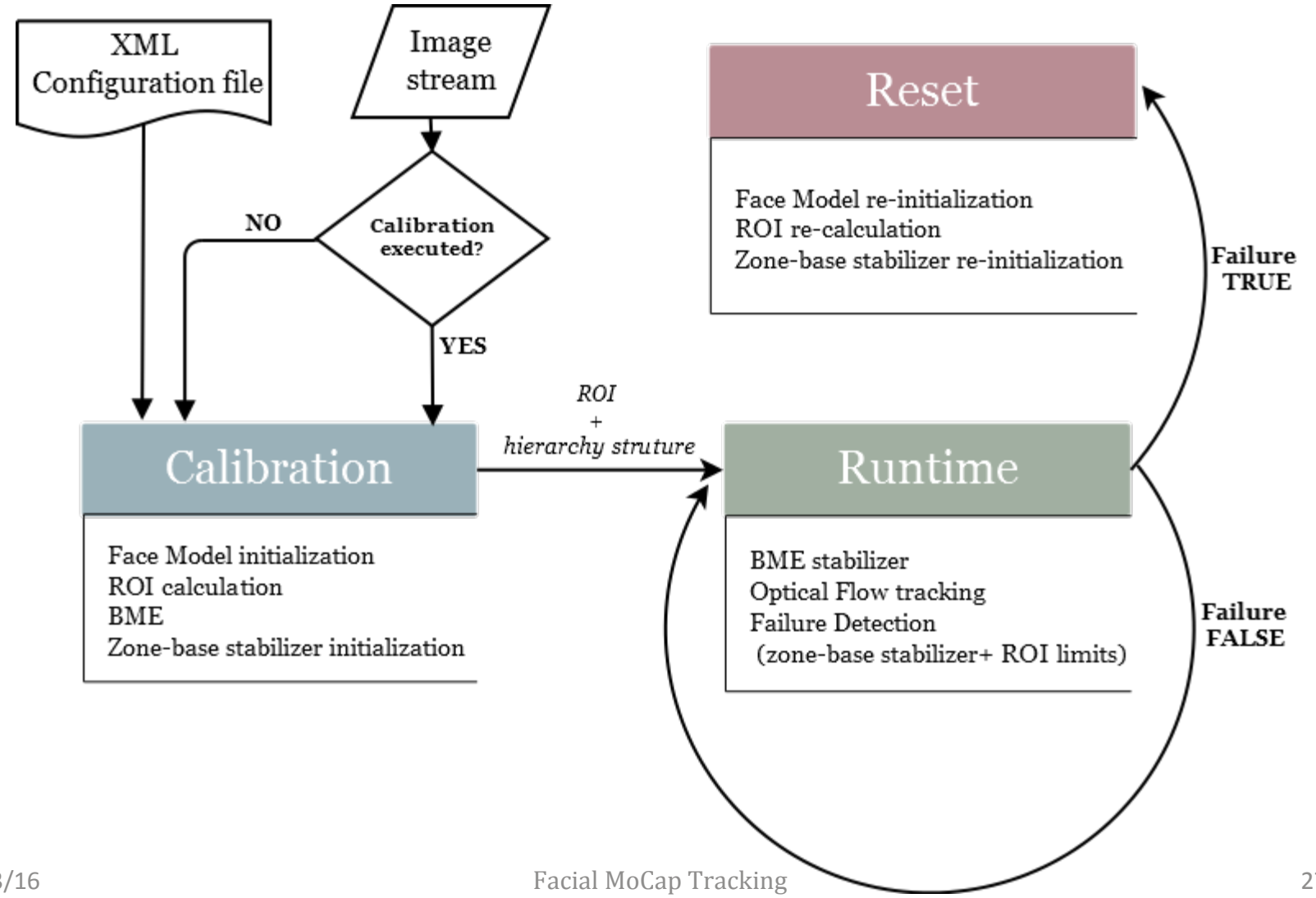
Goal

Markerless *tracking of* ***unique*** *facial features movements, such as cheeks or forehead movements and asymmetrical movements, using* ***off-the-shelf*** *hardware.*

Hypothesis

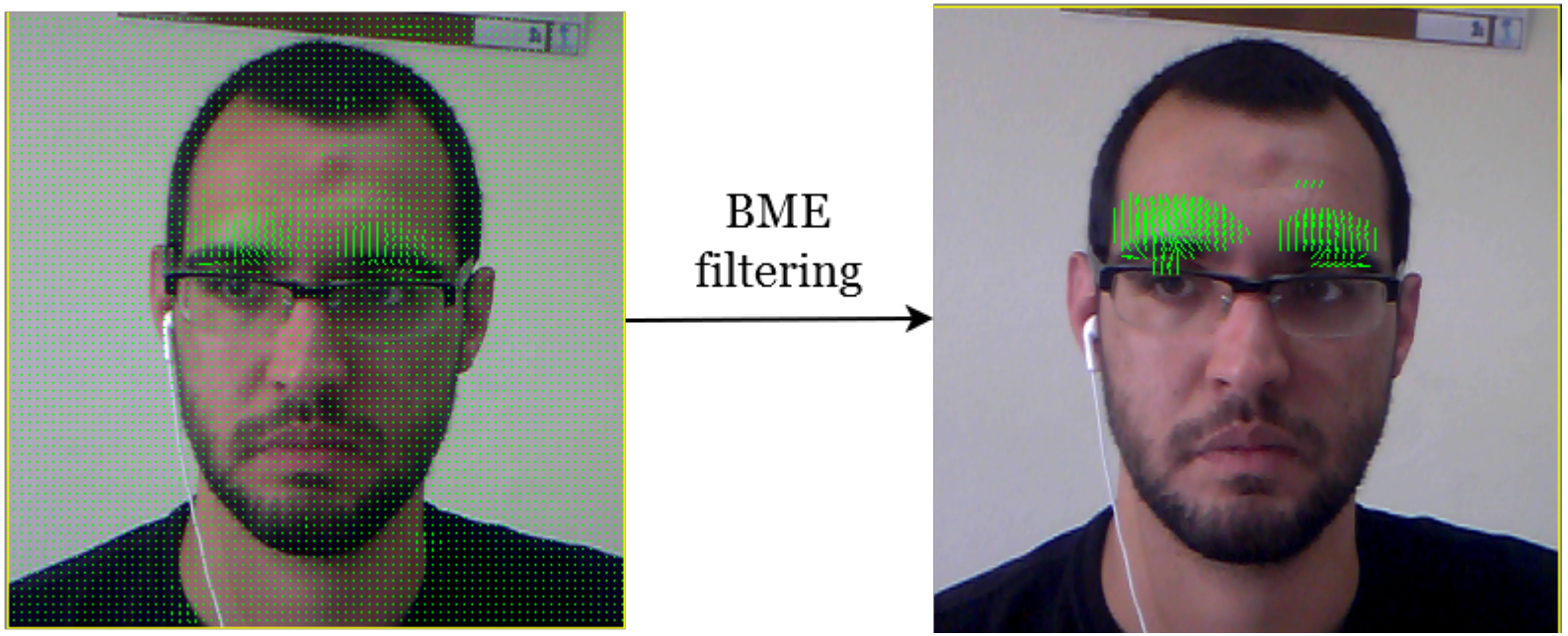
To prove that we can use the sensitivity of **Optical Flow** algorithms to track subtle and unique facial movements

Method


XML
Configuration file
Image
stream
Calibration
executed?
NO
YES
Reset
Face Model re-initialization
ROI re-calculation
Zone-base stabilizer re-initialization
Failure
TRUE
ROI
+
hierarchy struture
Calibration
Face Model initialization
ROI calculation
BME
Zone-base stabilizer initialization
Runtime
BME stabilizer
Optical Flow tracking
Failure Detection
(zone-base stabilizer+ ROI limits)
Failure
FALSE

Calibration:
Stabilization methods

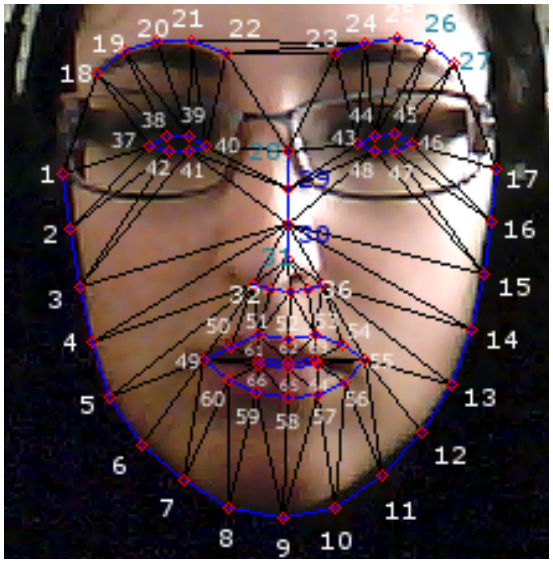
BME - Baseline Movement Estimation

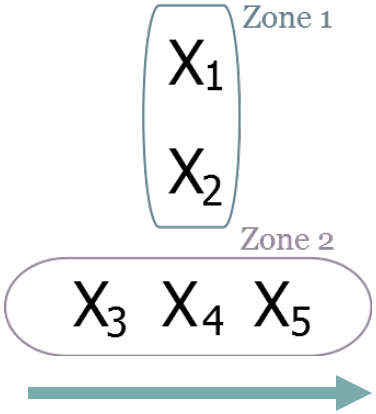


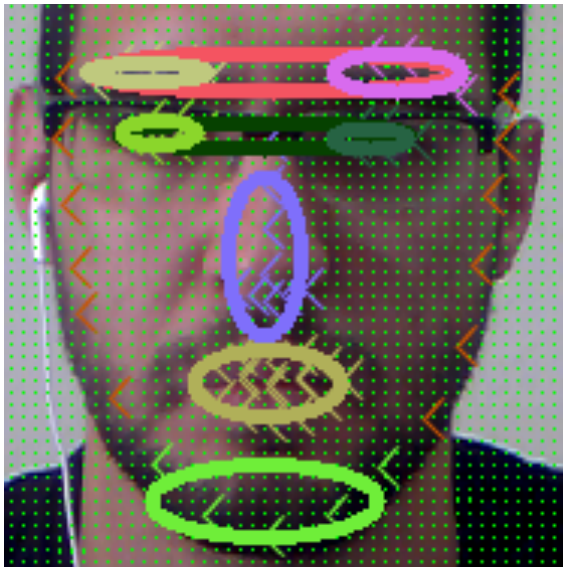
Calibration:
Stabilization methods

Zone-based stabilization

loads a hierarchy of facial zones and landmarks that define a certain facial model





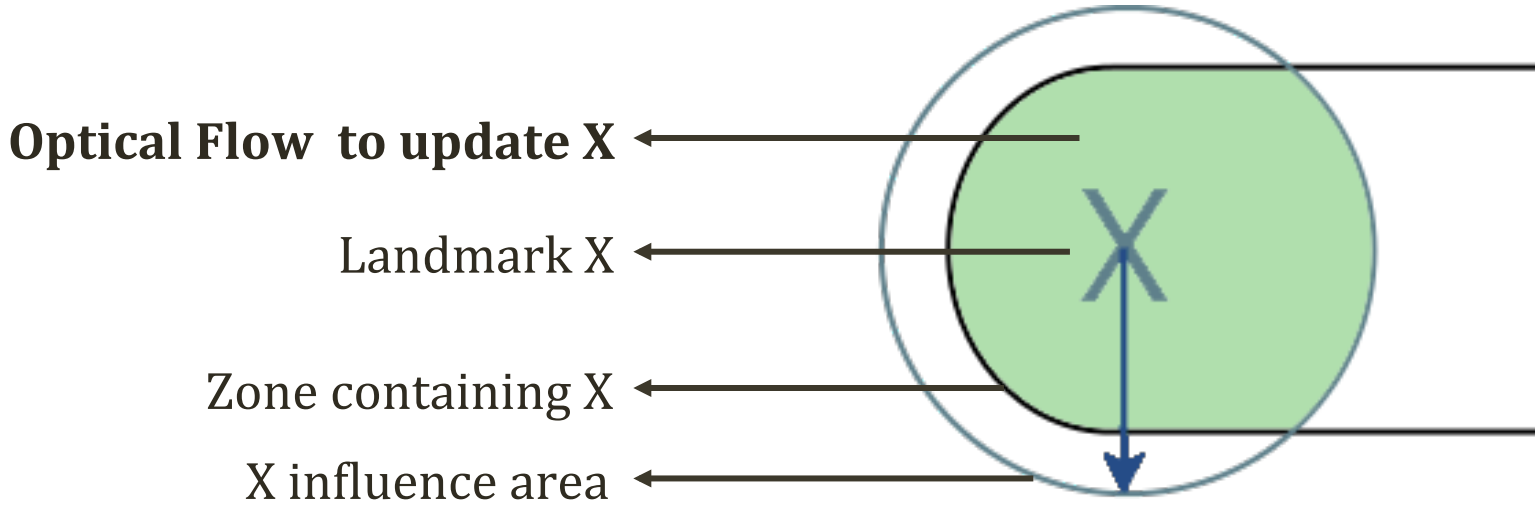


J.M. Saragih et al. Real-time avatar animation from a single image (2011)

Runtime
Tracking movements

1) Update landmarks:

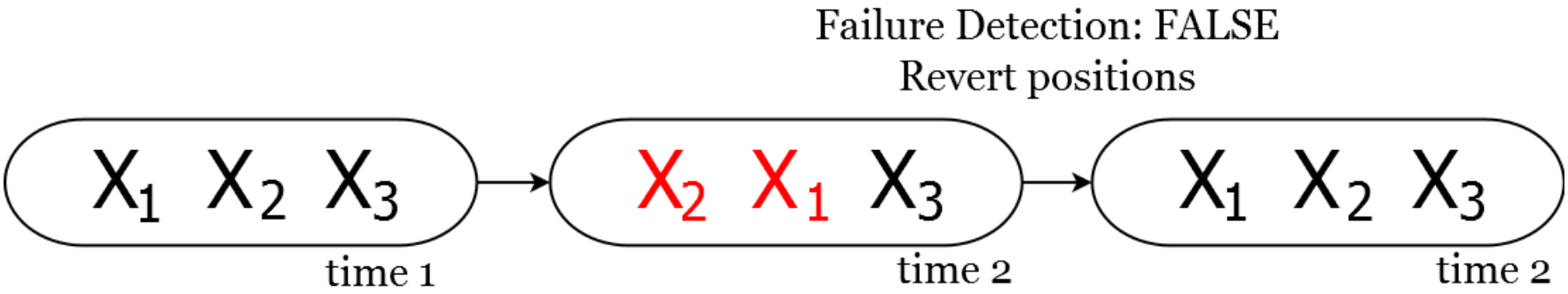
Optical Flow + Zone–based stabilization
zone limits (black line) and influence ratio (blue ratio)



Runtime
Zone-based stabilization

2) Failure Check:

Hierarchy **structure is maintained**

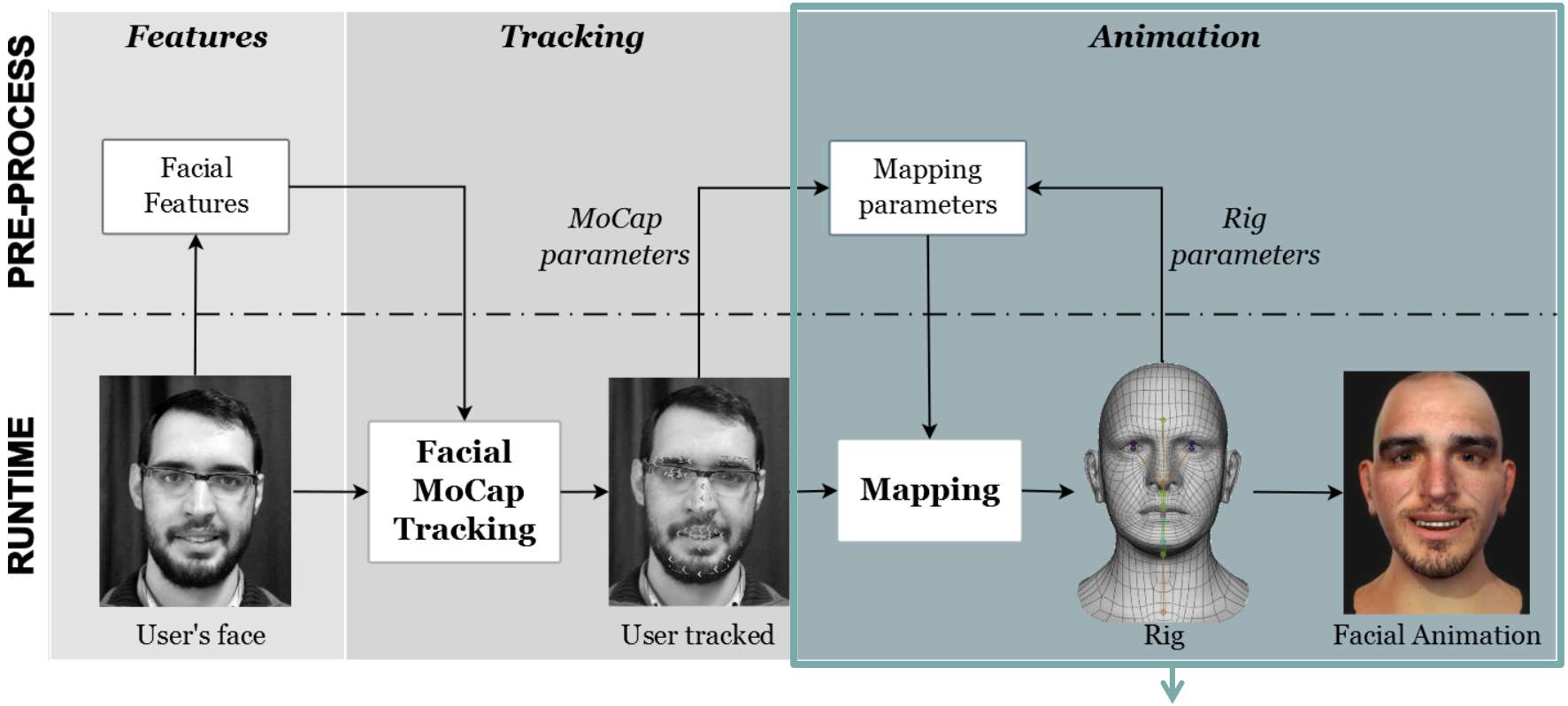


Results

Conclusions

Our method:

- Allows **unsupervised real-time tracking of uncommon facial features**, such cheeks movements
- Performs less accuractly than recent SotA methods under extreme environmental changes or during presence of more than one participant



To **automatically transfer movements** tracked to 3D character creating facial animation

MoCap Facial Animation

- Background
- Methodology
- Results
- Conclusions

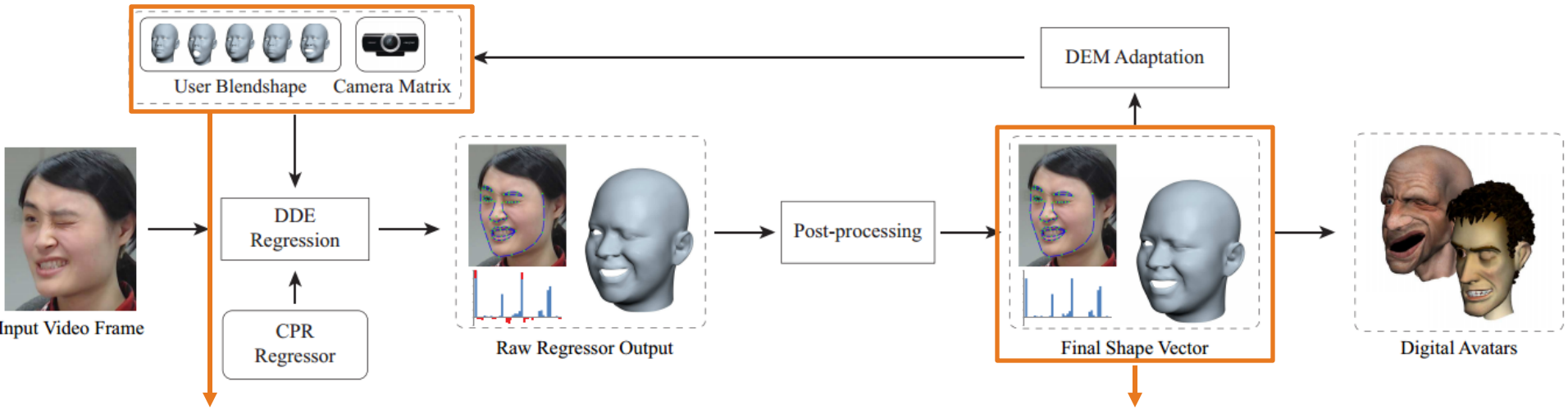
Background

- Example-based algorithms: Digital-Ira

Graham Fyffe et al. Driving High-Resolution Facial Scans with Video Performance Capture (2014)

Background

- Example-based algorithms:



User-dependent
long calibrations
+ Model Learning

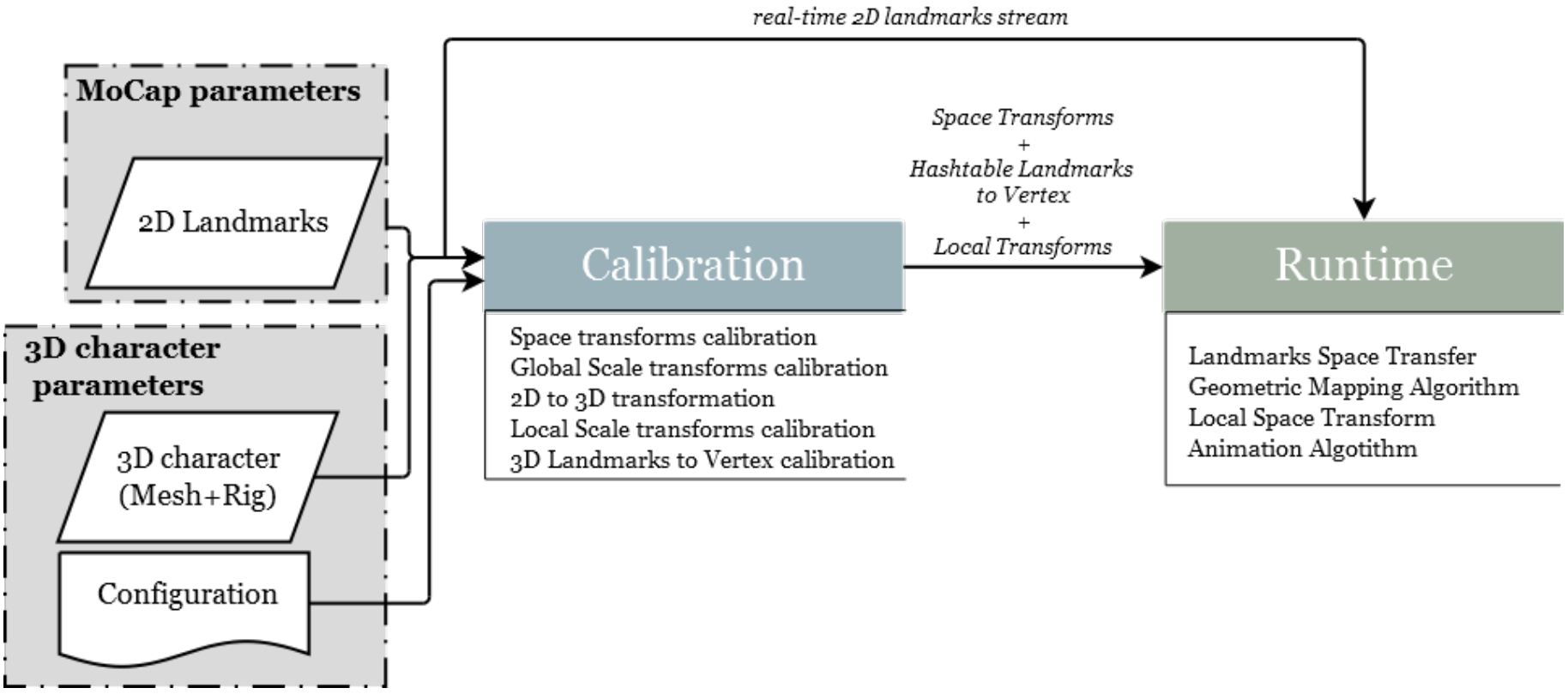
Blendshape Rig
Limited facial expressions

Chen Cao et al. Displaced dynamic expression regression (2014)

Goal

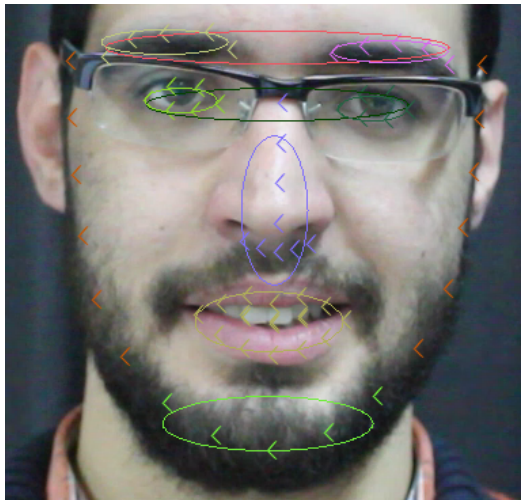
to create a mapping method that ***adapts to user-choice MoCap tracking algorithm*** *and* ***reduces user-dependent calibration*** *requirements.*

Mapping method



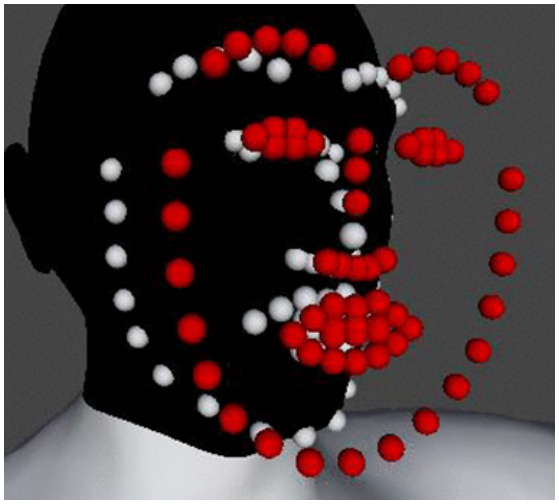
Mapping method
Calibration

1) Global + Local Transform between MoCap tracking and 3D character's rig



2D landmarks
MoCap tracking space

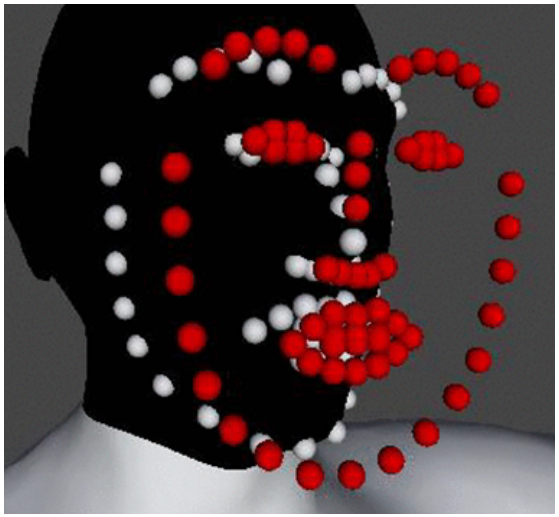
≠ spaces



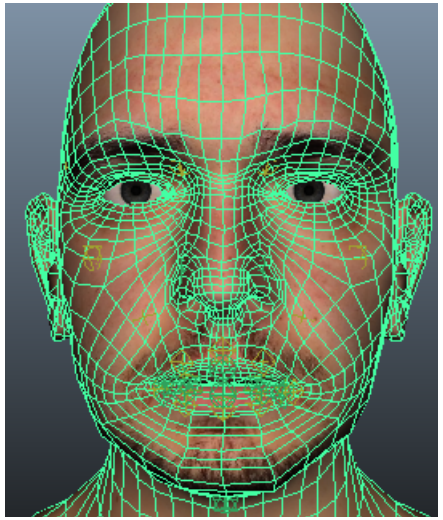
3D landmarks
3D character's space

Mapping method
Calibration

2) Hashtable with connection between 3D landmarks and vertex in the 3D character's mesh



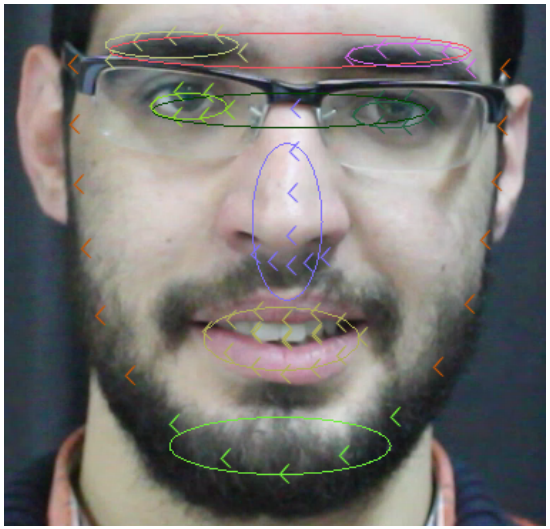
3D landmarks
3D character's space



Vertex in the Mesh
3D character's space

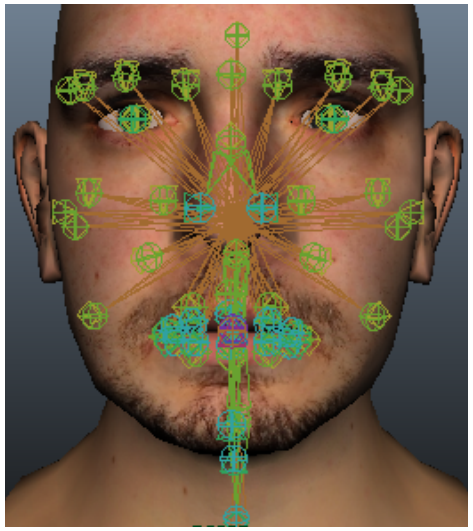
Mapping method
Runtime

+ Apply Global and Local **transform**
+ **Geometric Mapping** between vertex and bones
+ calculate intensity of bone's movements **to create animation**



2D landmarks
MoCap tracking space

≠ topology



Rig's bones
3D character's space

Mapping method
Runtime: **Geometric Mapping**

correspondence between vertex (in the mesh) and the **bones** in the 3D character's rig.

To each vertex:

1) Main-bone translation:

2D Main – bone translation

$$B_{Mxyz} = \frac{L_{xyz} - Li_{xyz}}{W_{main}} + Bi_{Mxyz}$$

associated Landmark Movement

Sum of Weights of bone and childs

Main – bone Initial position

Mapping method
Runtime: **Geometric Mapping**

correspondence between vertex (in the mesh) and the **bones** in the 3D character's rig.

To each vertex:

1) Main-bone translation:

$$B_{Mxyz} = \frac{L_{xyz} - Li_{xyz}}{W_{main}} + Bi_{Mxyz}$$

2) Secondary bones translation:

Sec – bone initial translation

Bone translation

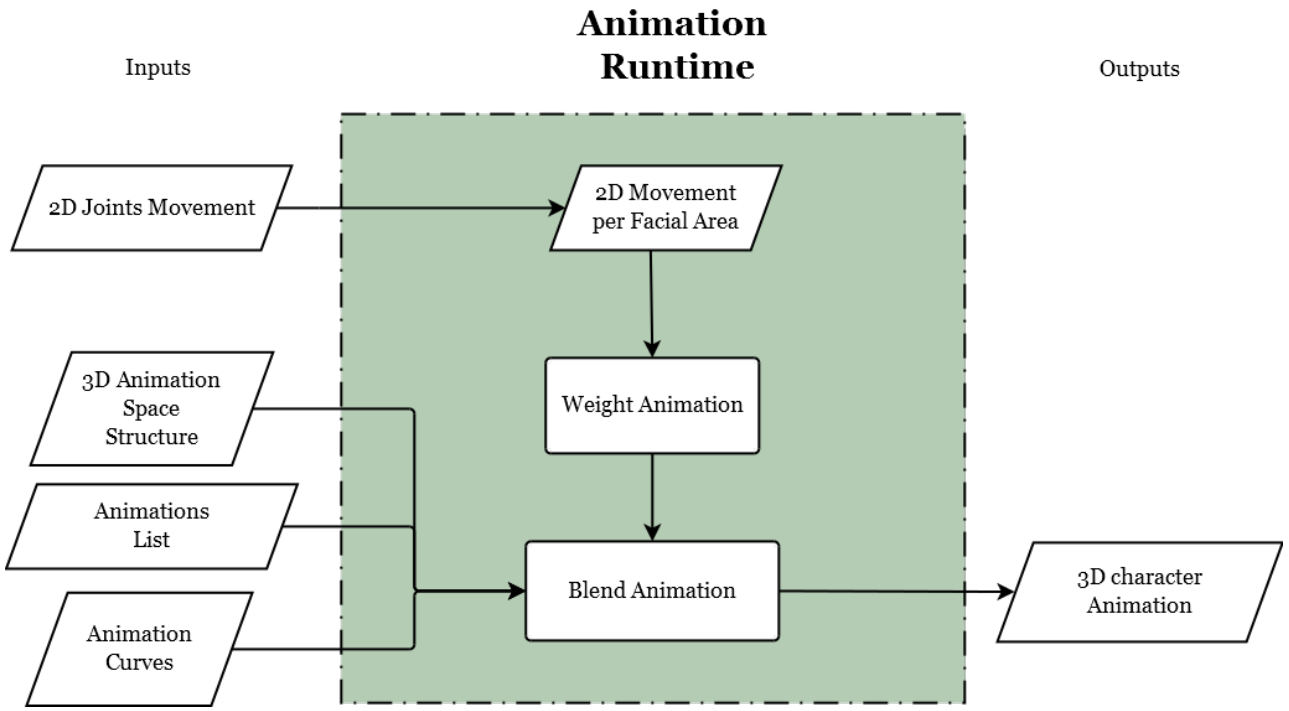
2D Sec – bone translation

$$B_{sec-xyz} = (B_{Mxyz} - Bi_{Mxyz}) * W_{sec} + Bi_{sec-xyz}$$

Sum of weights of Bsec and childs

Mapping method
Runtime: **Animation**

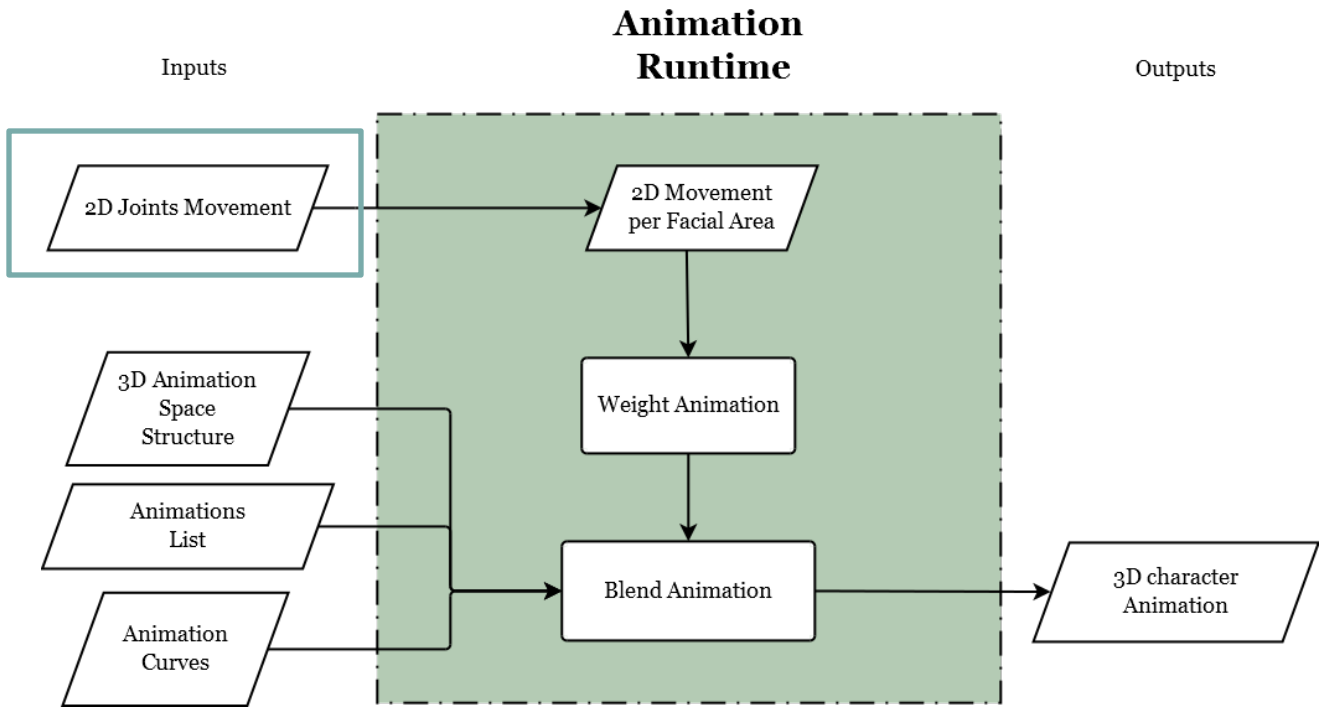
calculates the **intensity of deformation** produced by the translation of each bone.



Mapping method
Runtime: **Animation**

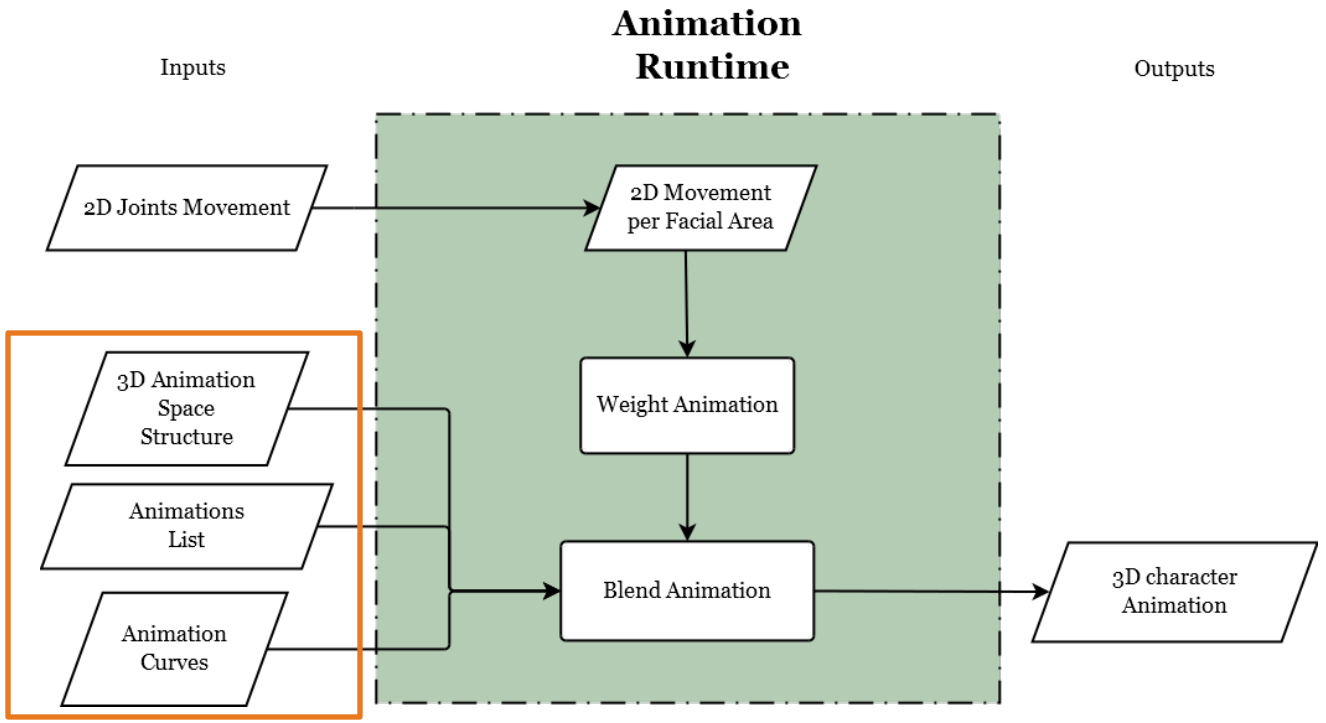
calculates the **intensity of deformation** produced by the translation of each bone.

Output: **Geometric mapping**



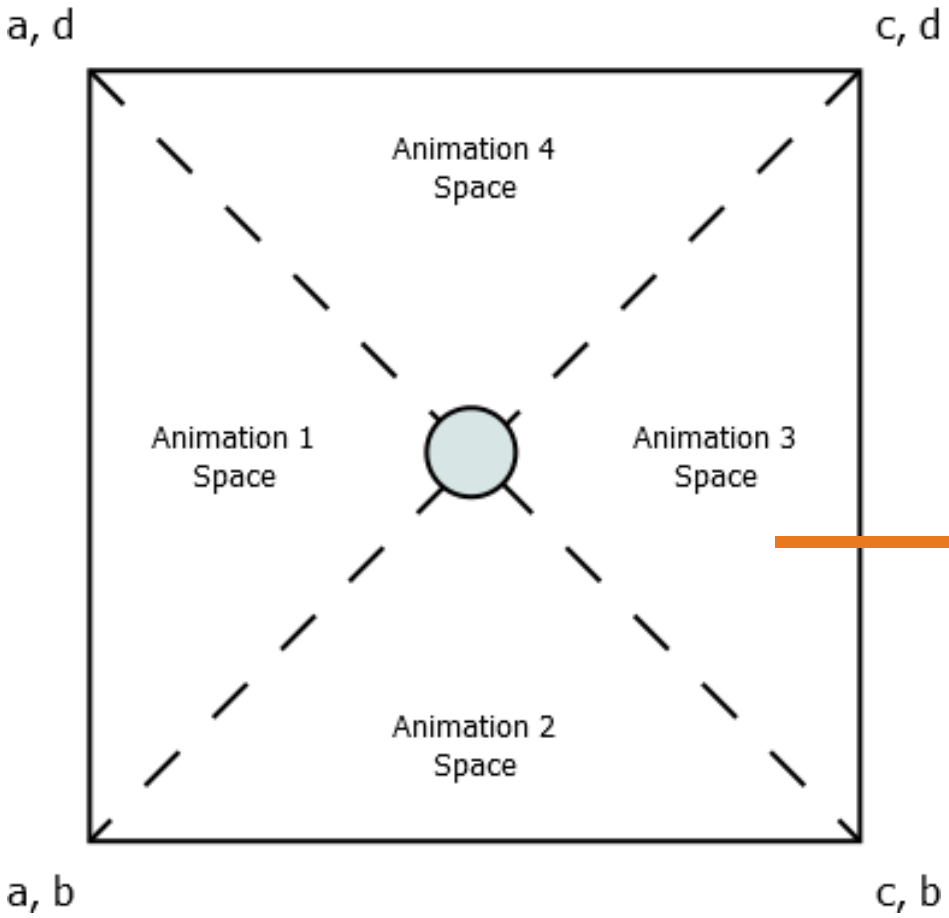
Mapping method
Runtime: **Animation**

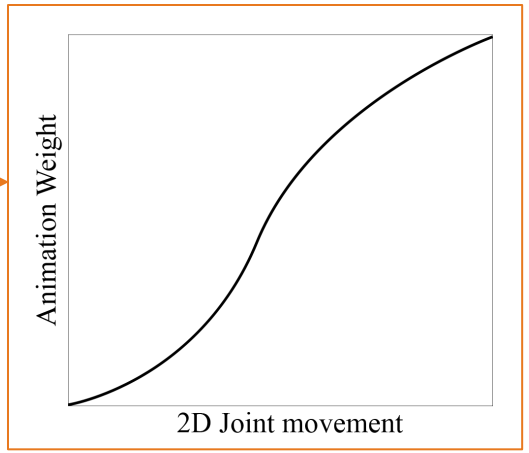
calculates the **intensity of deformation** produced by the translation of each bone.



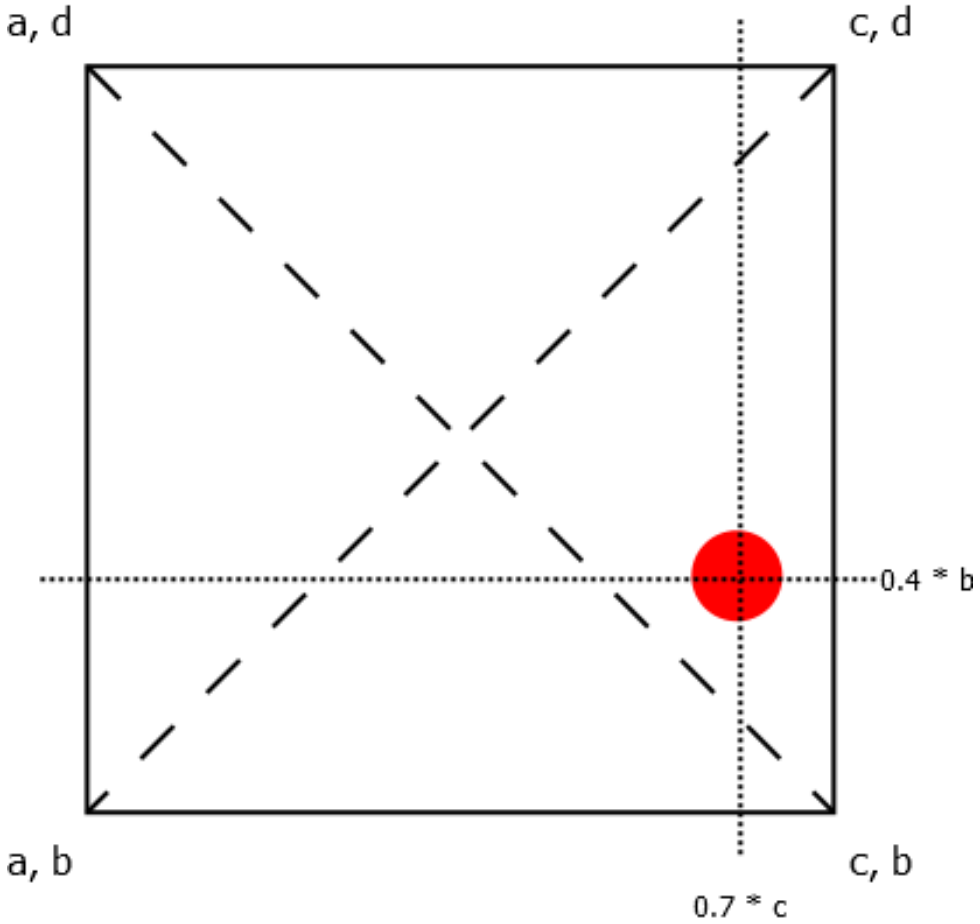
Rig setup by artist

Mapping method
Runtime: **Animation**





Mapping method
Runtime: **Animation**



Bone Animation =
Animation 2 with weight 0.4*b
Animation 3 with weight 0.7*c

MoCap

Real-time Facial Animation

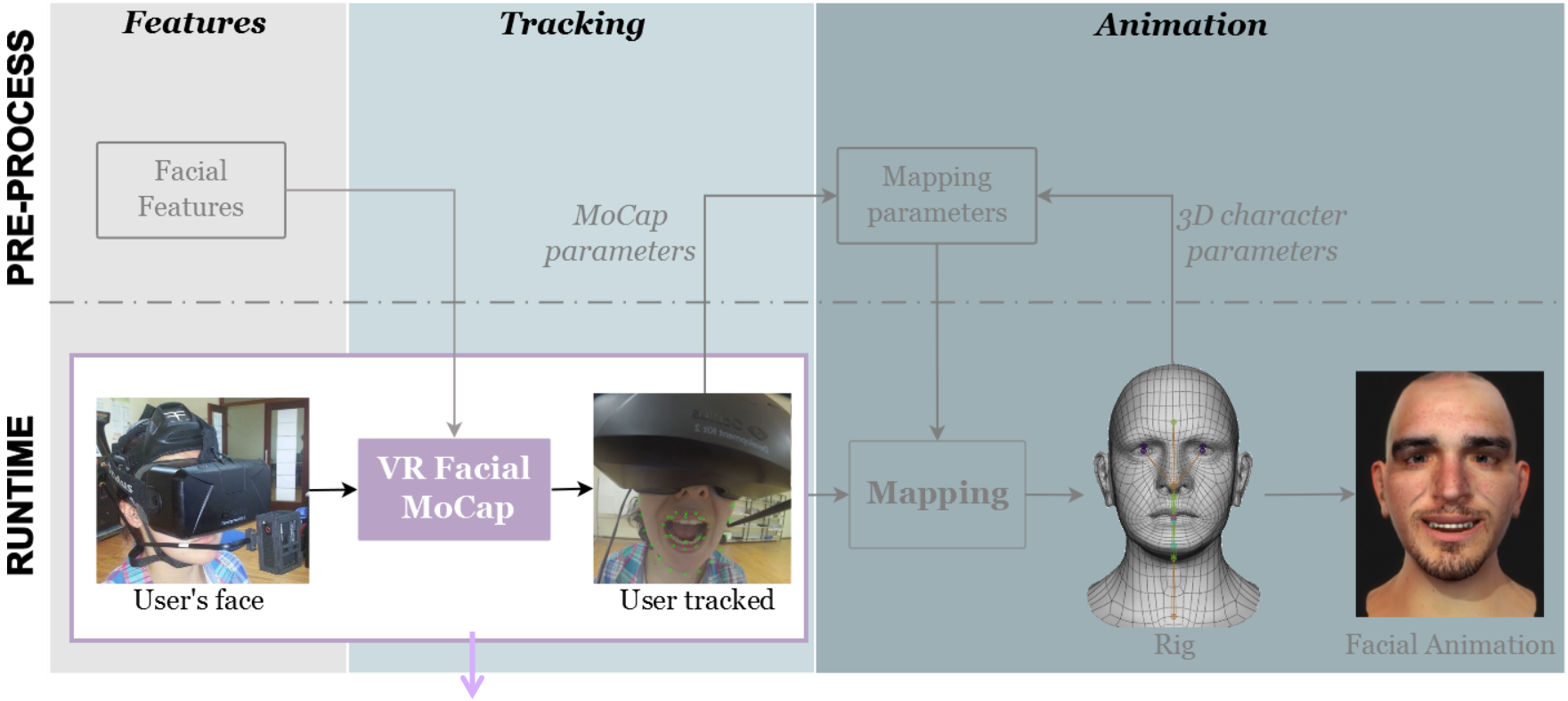
Test Case 1

Real-time MoCap Tracker

Conclusions

Geometric mapping algorithm:

- Adapts to **different MoCap tracking** systems
- Allow **real-time animation** without complex calibrations
- Reproduces **asymmetrical facial movements**



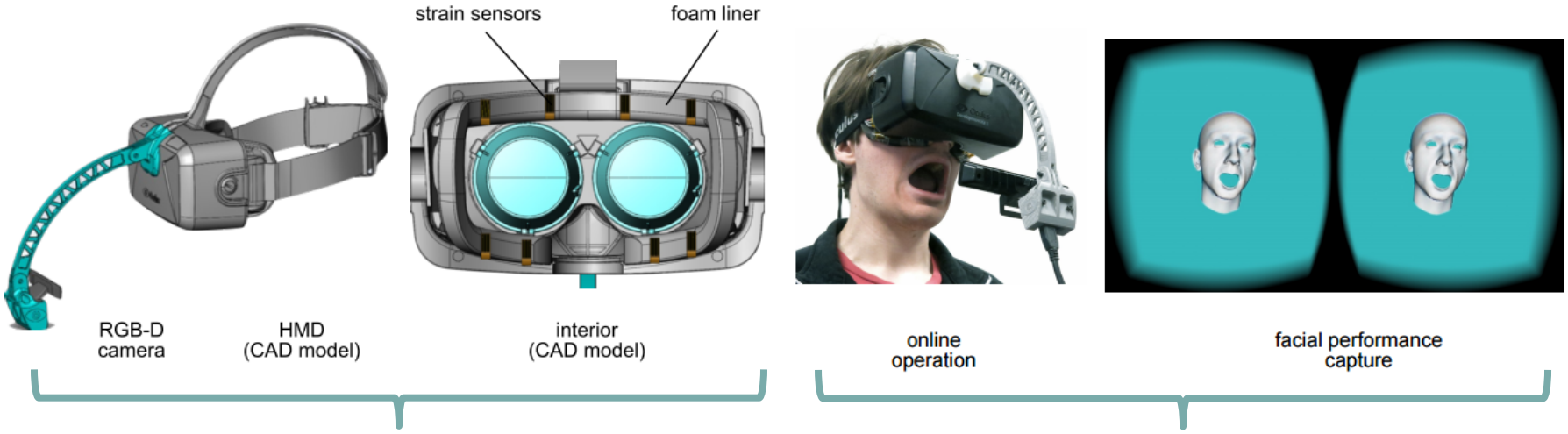
To create **MoCap tracking** systems compatible with **VR** environments

MoCap VR Methods

- Background
- Methodology
- Results
- Conclusions

Background

- **VR scenario:** Persistent partial occlusions

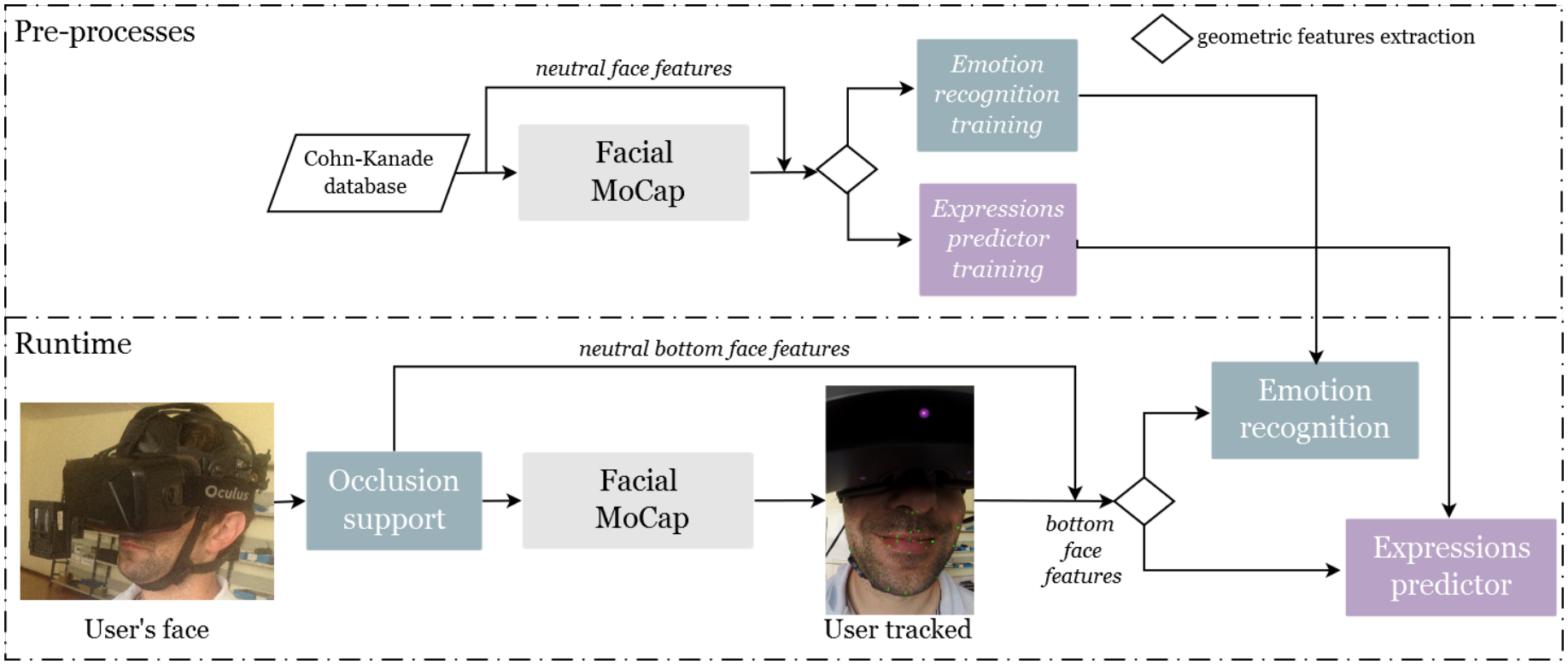


Hao Li et al, Facial performance sensing head-mounted display (2015)

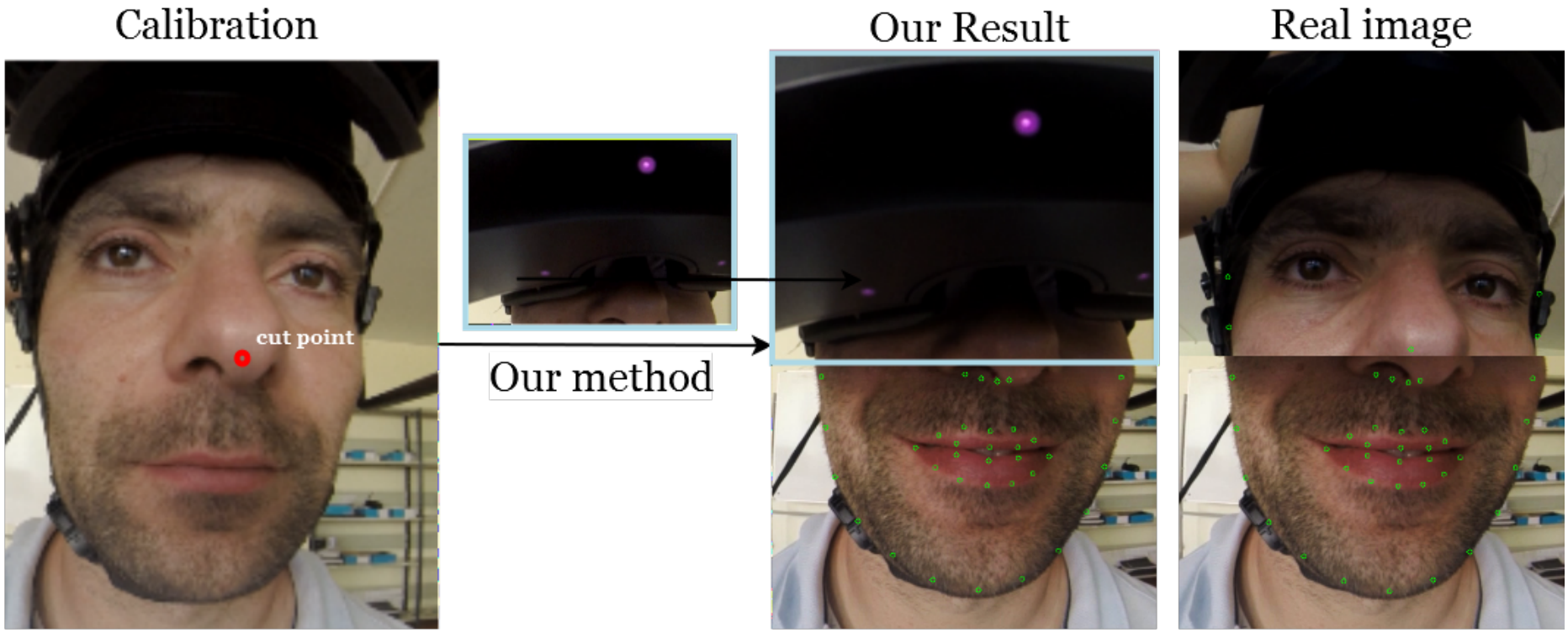
Goal

*to create methods to **estimate facial expressions** of upper part of the face and **predicts emotions** using movements tracked from bottom of the face.*

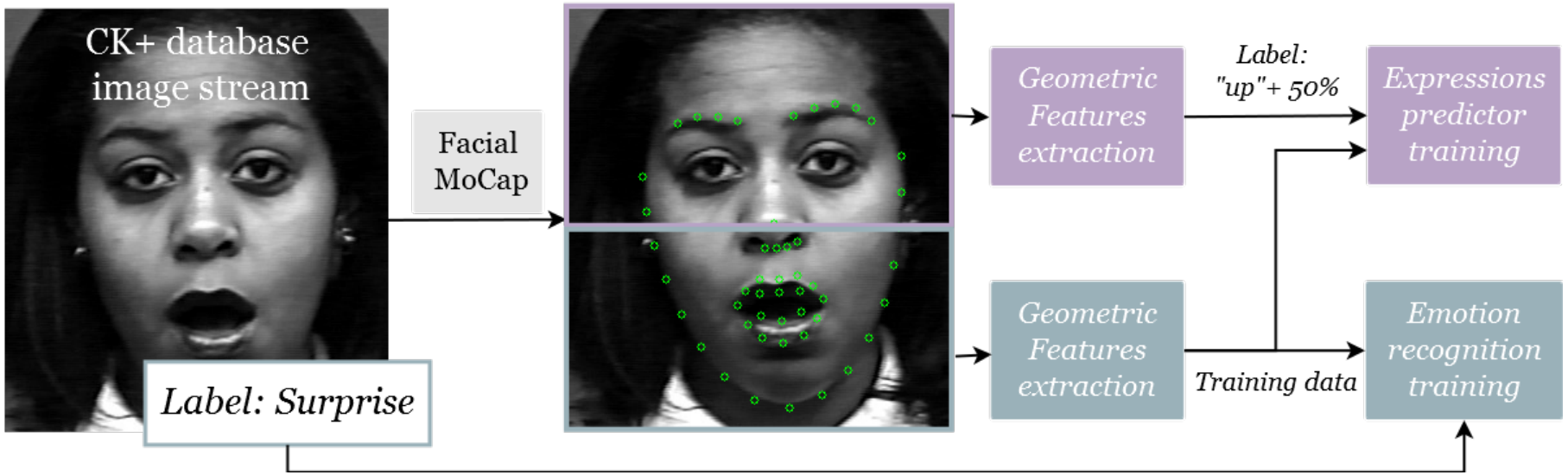
MoCap VR methods



MoCap VR methods
Persistent Partial Occlusions



MoCap VR methods
Assessing Facial Expressions



Geometric features extraction:

$$D(p_i, p_c) = \sqrt{\frac{((p_i(x) - p_c(x))^2 + (p_i(y) - p_c(y))^2}{\|p_i - p_c\|}}$$

Statistical Validation

Emotions	K-Fold Accuracy (%)
six [EF78]	64.80
four [JJ13]	69.07

Eyebrows movements	K-Fold Accuracy(%)
Up *S1*	91.47
Up *S2*	87.02
Down *S1*	70.63
Down *S2*	69.13

Statistical Validation

Emotions	K-Fold Accuracy (%)
six [EF78]	64.80
four [JJ13]	69.07

Eyebrows movements	K-Fold Accuracy(%)
Up *S1*	91.47
Up *S2*	87.02
Down *S1*	70.63
Down *S2*	69.13

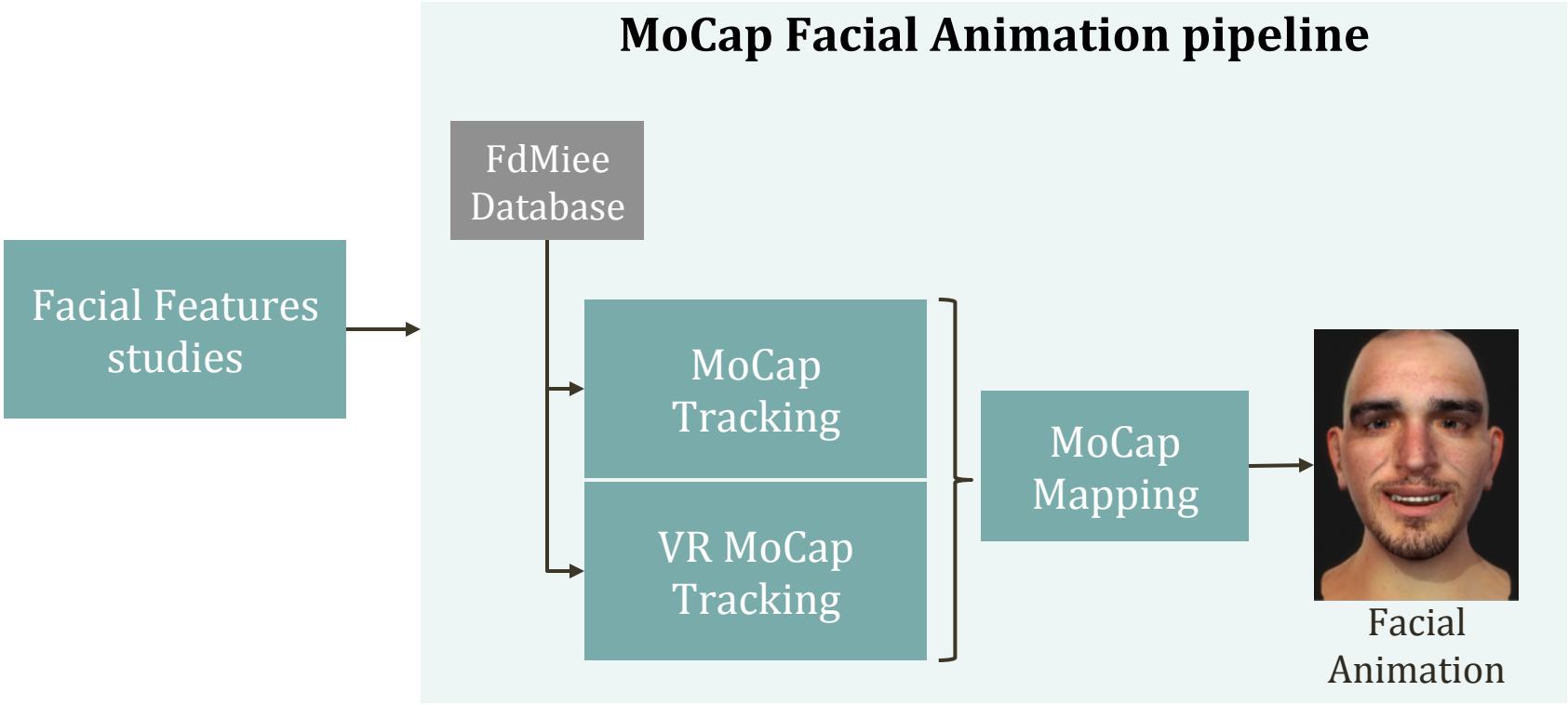
VR MoCap Tracking

Conclusions

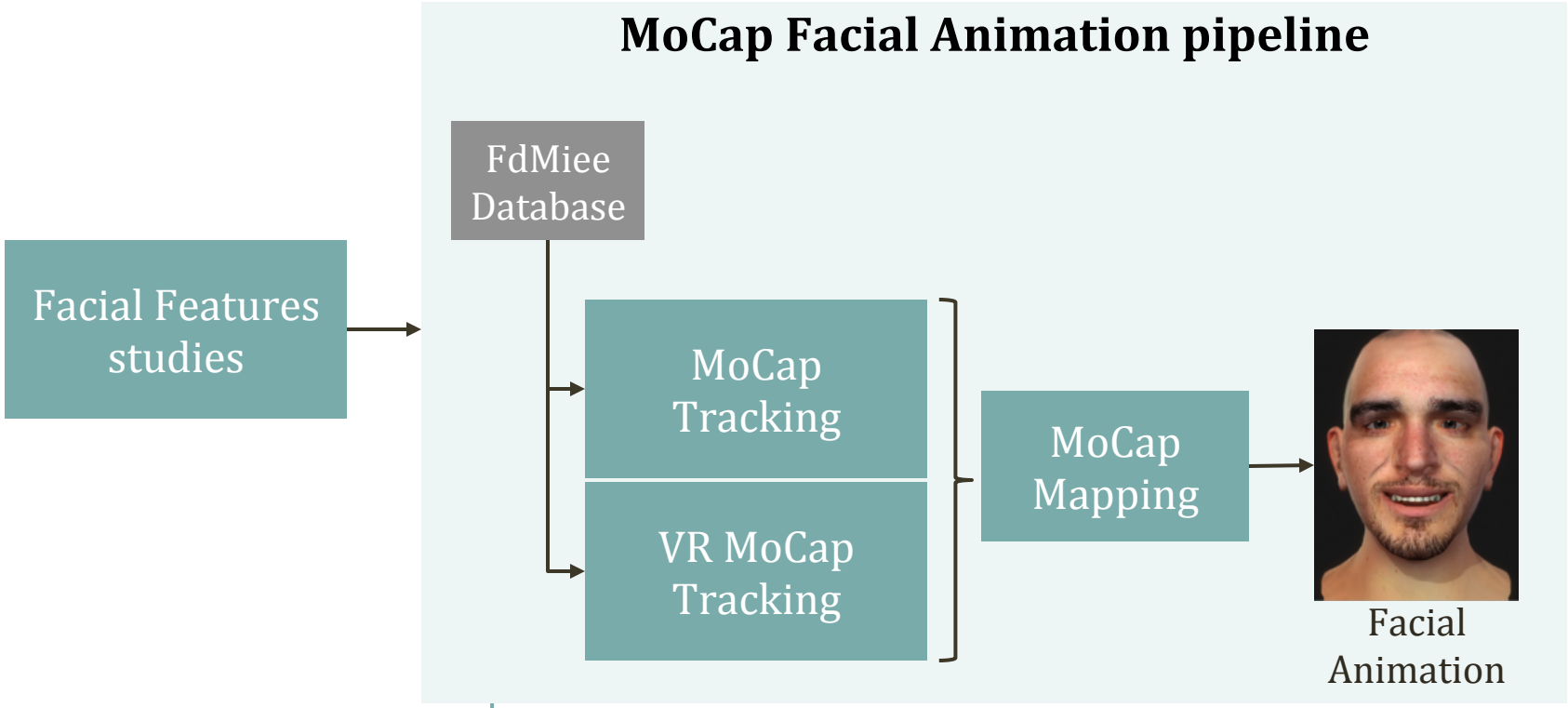
MoCap VR methods:

- Make generic MoCap tracker systems **compatible with persistent partial occlusions** in VR environments
- **Predict six universal emotions**
- **Estimate eyebrows' movements**

Contributions



Contributions



Real-time
Modular
Non-intrusive

Reduce manual intervention
Usable by non-experts
Off-the-shelf hardware

Dissemination

- 2 articles accepted
- 3 articles submitted
- 1 Eurographics course submitted
- 2 best idea/concept Award
- 1 EU Project Workshop
- 5 invited talks

Conclusion

Facial Features studies

- **F**ace perception
- **F**acial behaviors
- **P**sychology of emotions

MoCap tracking

MoCap Mapping

VR MoCap Tracking

Conclusion

Facial Features studies	MoCap tracking	MoCap Mapping	VR MoCap Tracking
Face perception **F**acial behaviors **P**sychology of emotions	**OF** to track unique facial traits **B**iometrics **S**ecurity		

Conclusion

Facial Features studies	MoCap tracking	MoCap Mapping	VR MoCap Tracking
Face perception **F**acial behaviors **P**sychology of emotions	**OF** to track unique facial traits **B**iometrics **S**ecurity	**A**daptive animation algorithms **U**ser friendly applications	

Conclusion

Facial Features studies

- **F**ace perception
- **F**acial behaviors
- **P**sychology of emotions

MoCap tracking

- **OF** to track unique facial traits
- **B**iometrics
- **S**ecurity

MoCap Mapping

- **A**daptive animation algorithms
- **U**ser friendly applications

VR MoCap Tracking

- **VR** tracking of emotions and facial expressions
- **H**ardware free approach
- **L**earning algorithms for expressions predictio

Take-home message

Facial Animation created

by anyone for everyone!

Thank you!

FdMiee's protocol

Protocol to create facial databases under a wide range of environemnt and behavior changes

FdMiee database:

- 6 participants
- 3 capture systems
- 6 Fixed Parameters