Lifelike Humans: Detailed Reconstruction of Expressive Human Faces

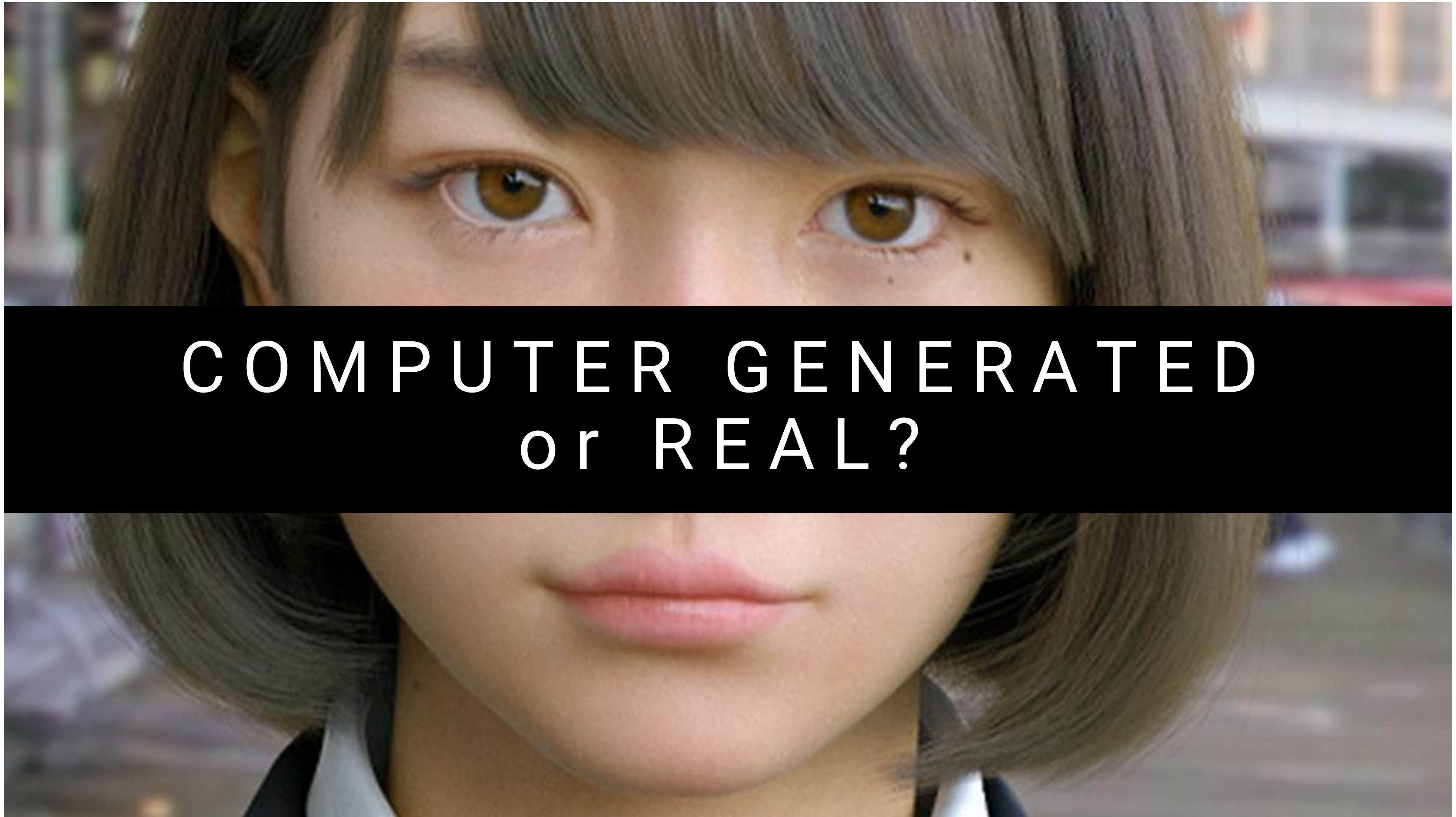
A dissertation submitted by **Gemma Rotger Moll** at Universitat Autònoma de Barcelona to fulfil the degree of **Doctor of Philosophy**.

Advisors: Felipe Lumbreras, Antonio Agudo

Bellaterra, January 12, 2021

CVC Computer Vision Center
UAB Universitat Autònoma de Barcelona
Institut de Robòtica i Informàtica Industrial

INTRODUCTION



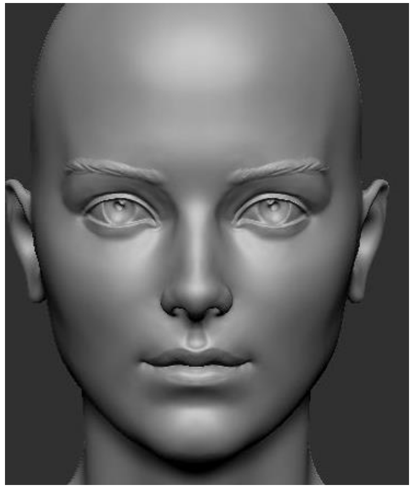
INTRODUCTION

KEY ASPECTS ON FACIAL REALISM

1. SKIN WRINKLES
2. FACIAL HAIR
3. FACIAL MICROEXPRESSIONS

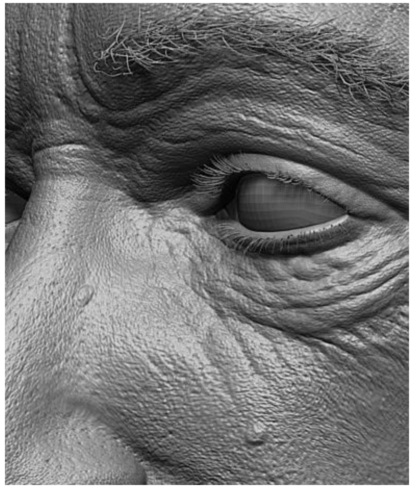
OUTLINE

PART 0



INTRODUCTION

PART 1



DETAIL

CONTRIBUTIONS

1. Code the facial wrinkle and hair spaces without any training data nor user interaction
2. Fast and accurate recovery of facial wrinkles and wounds
3. Handle detailed reconstruction under uncontrolled general lighting

SKIN DETAIL

CONTRIBUTIONS

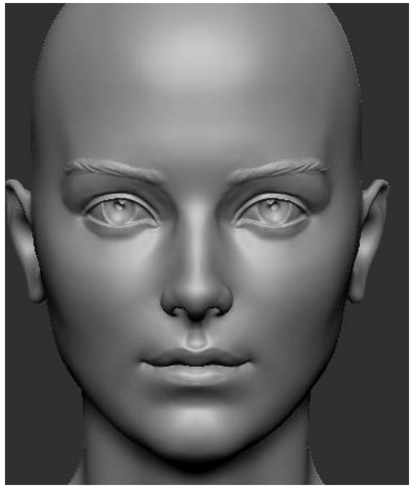
4 Robust parametrizable model for facial hair

5 New formulation to retrieve facial hair geometry from a single RGB image

FACIAL HAIR

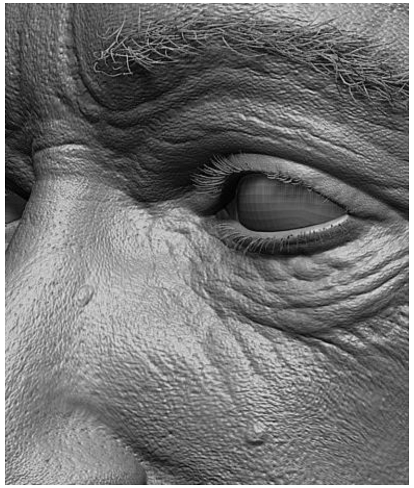
OUTLINE

PART 0



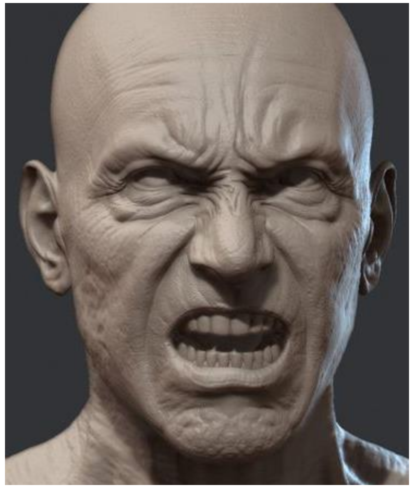
INTRODUCTION

PART 1



DETAIL

PART 2



EXPRESSION

CONTRIBUTIONS

6 Transfer expressions from 2D-to-3D in a dense manner, without the use of a user-specific low-rank model

7 Transferring new expression wrinkles adapted to the source geometry

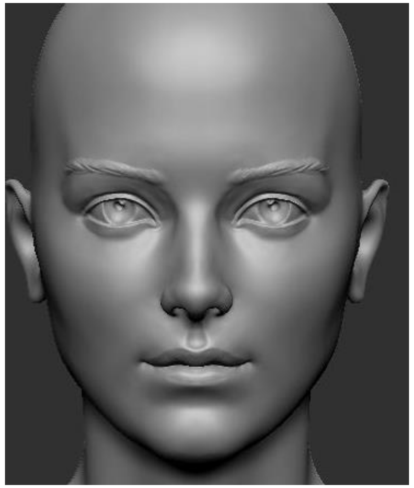
8 Holding the facial detail consistency on expression transfer

9 Animate facial hair in a simple manner

FACIAL EXPRESSIONS

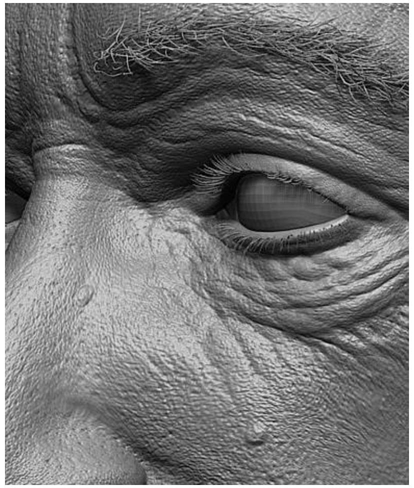
OUTLINE

PART 0



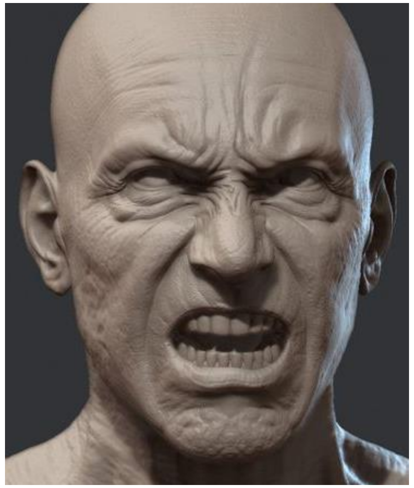
INTRODUCTION

PART 1



DETAIL

PART 2

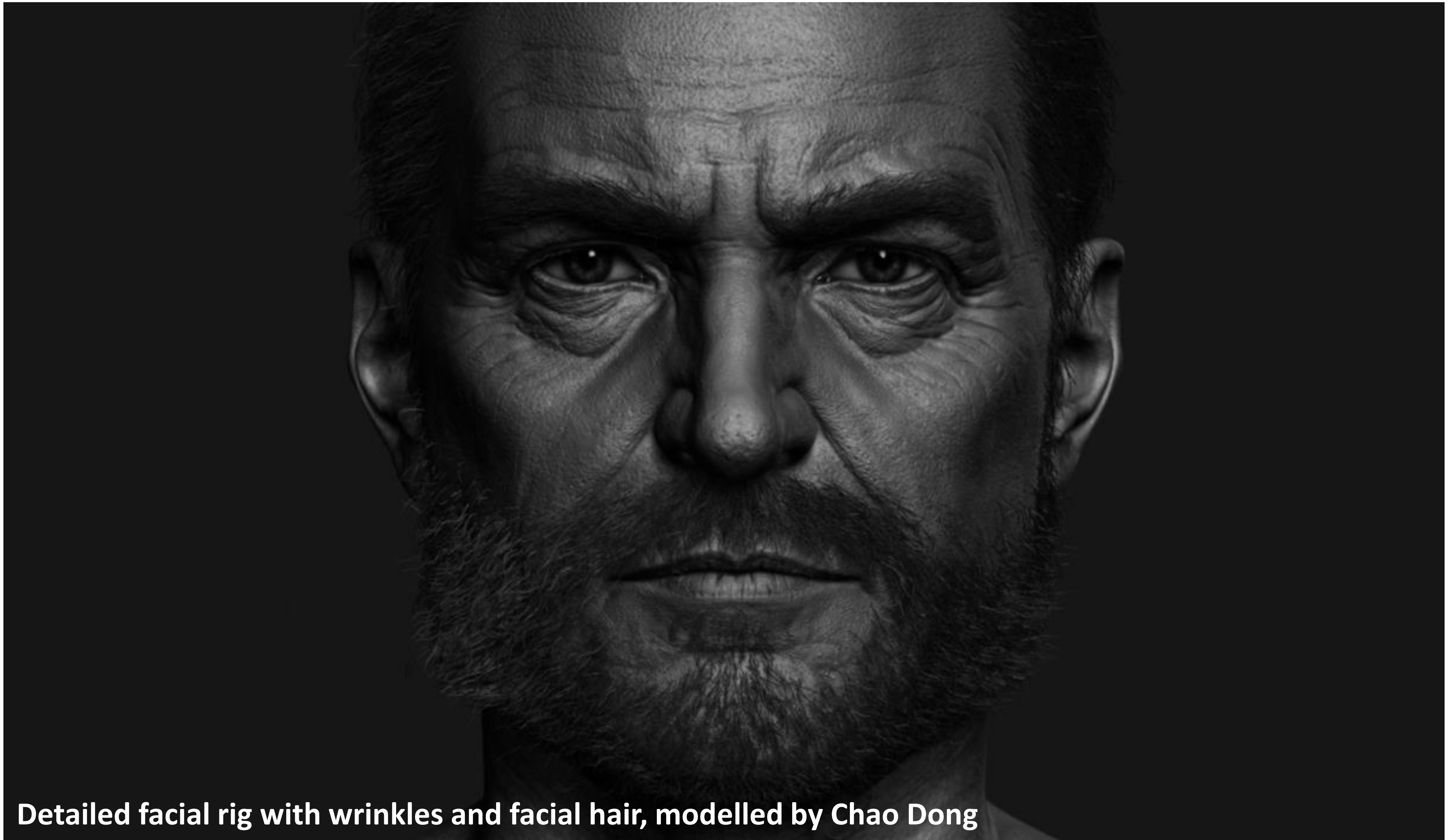


EXPRESSION

PART 3

CLAUSULA

DETAIL PART I



Detailed facial rig with wrinkles and facial hair, modelled by Chao Dong

1

SKIN WRINKLES

RELATED WORK

MODEL BASED: A priori information required, >time

Romdhani05, Huber16, Jiang18

DATA DRIVEN: A priori data required, low rank

Suwajanakorn14, Agudo17, Agudo19

DEEP LEARNING: Even more data required, training setups

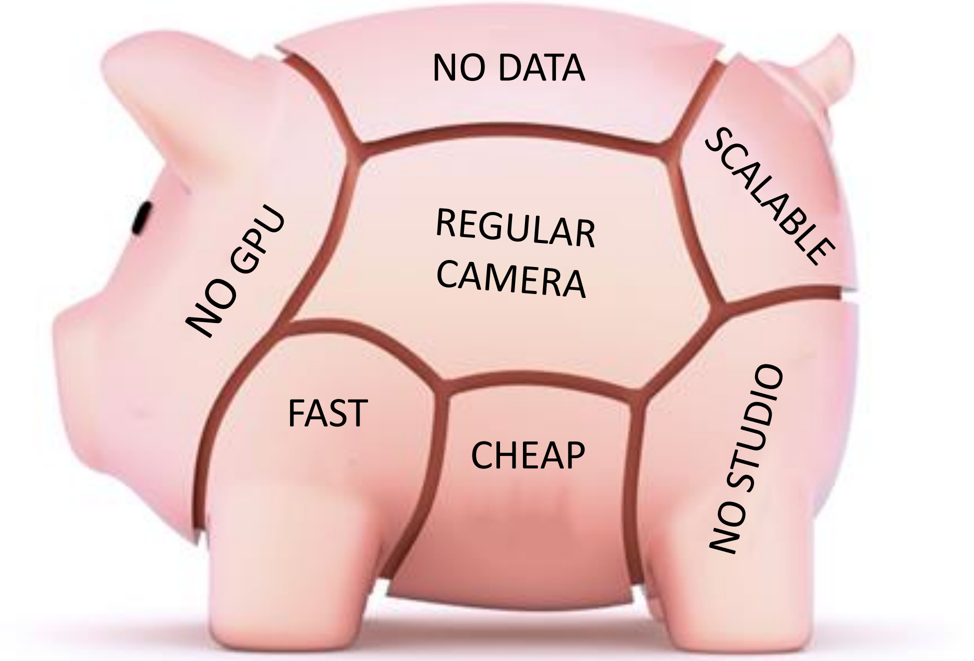
Richardson16, Richardson17, Jackson17, Chinaev18, Tuan Tran18

OUR METHOD: Unsupervised, does not require a priori information nor data

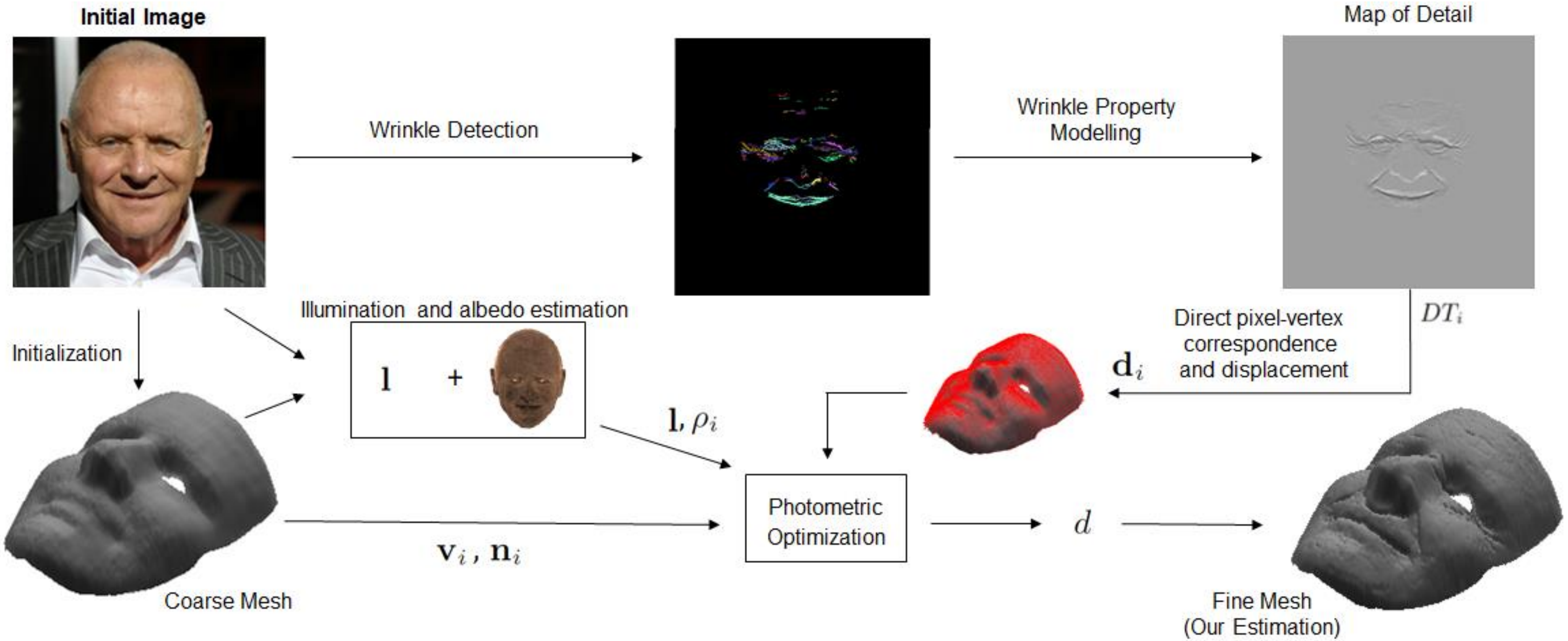
PURPOSE

1. FAST
2. EASY
3. UNSUPERVISED
4. NO TRAINING REQUIRED
5. NO GPU
6. WITH ANY REGULAR CAMERA
7. ANY ILLUMINATION CONTROL

How much far
can we achieve?



SCHEME

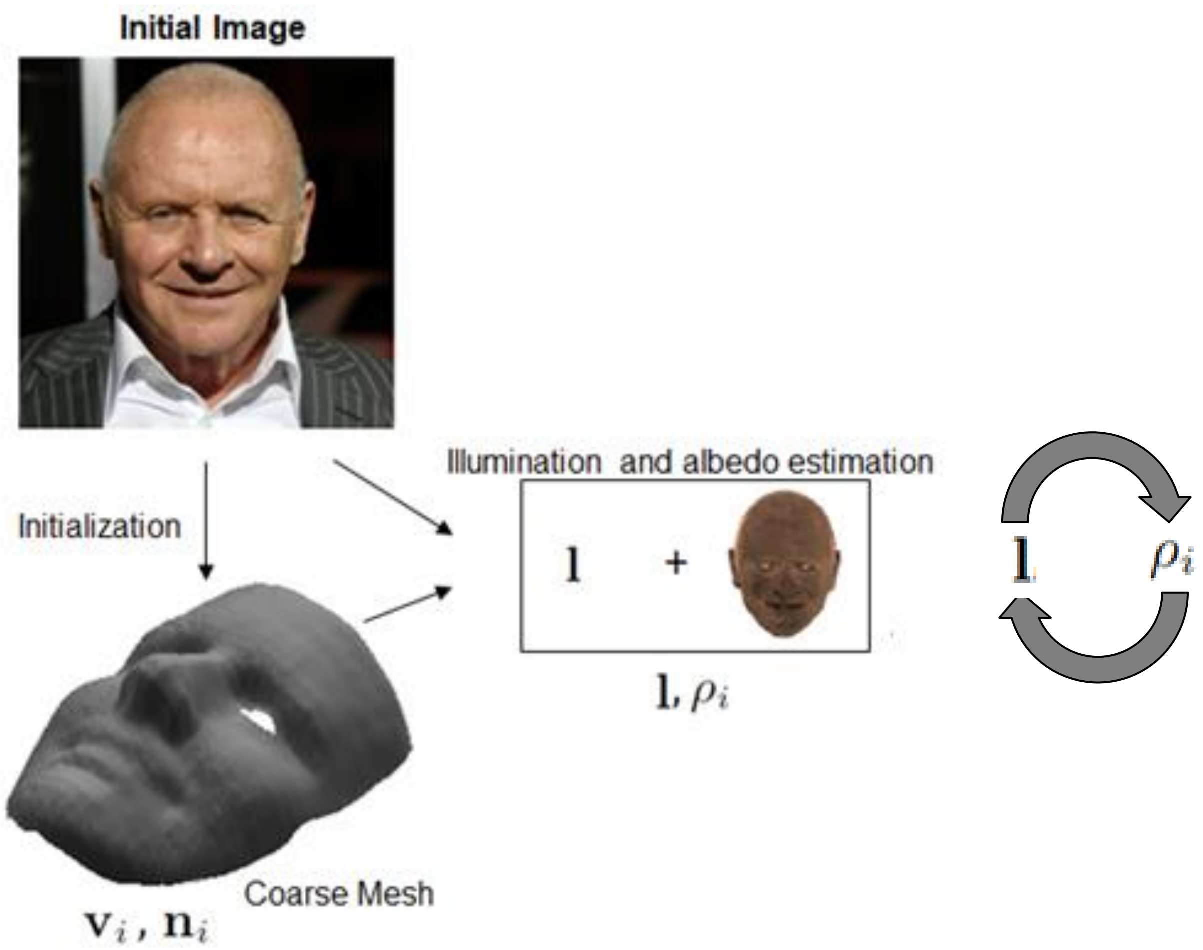

Initial Image
Wrinkle Detection
Wrinkle Property
Modelling
Map of Detail
Illumination and albedo estimation
Initialization
l +
Direct pixel-vertex
correspondence
and displacement
DT_i
$\mathbf{d}_i$
$\mathbf{l}, \rho_i$
Photometric
Optimization
d
$\mathbf{v}_i$, $\mathbf{n}_i$
Coarse Mesh
Fine Mesh
(Our Estimation)

INITIALIZATION

1. ρ_i as MEAN COLOR
2. RECOMPUTE $\boldsymbol{l}$
3. RECOMPUTE ρ_i

Repeat until convergence

Jackson17



WRINKLE DETECTION

1

INITIAL IMAGE

Input Image

2

DERIVATIVE RESPONSE

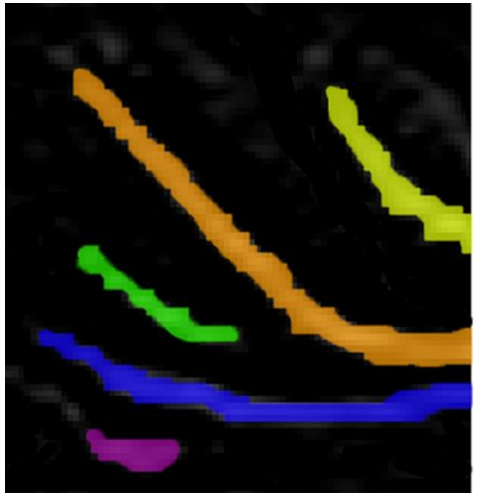
dx dy

3

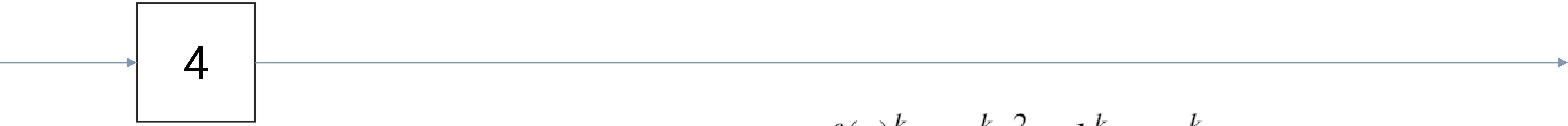
WRINKLE SEGMENTATION

Thresholding

8-connections

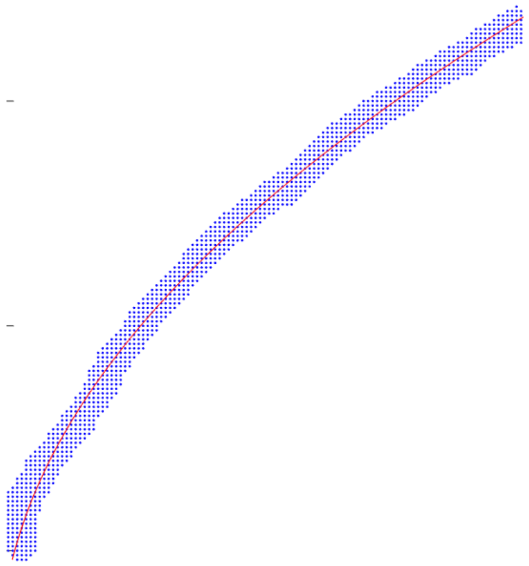


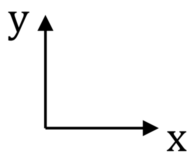
WRINKLE MODELLING



POLYNOMIAL FITTING

$$f(x)^k = a^k x^2 + b^k x + c^k \ ,$$



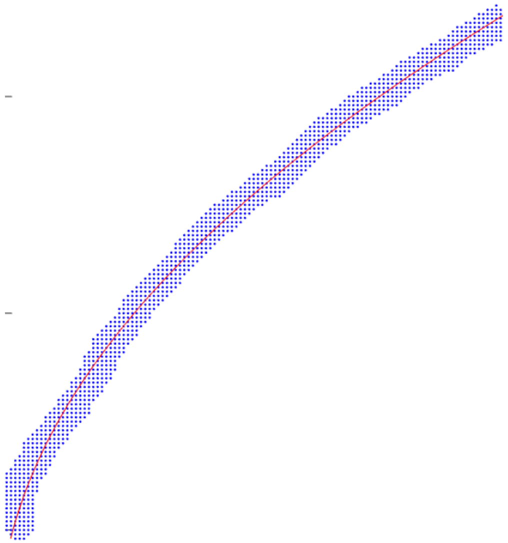


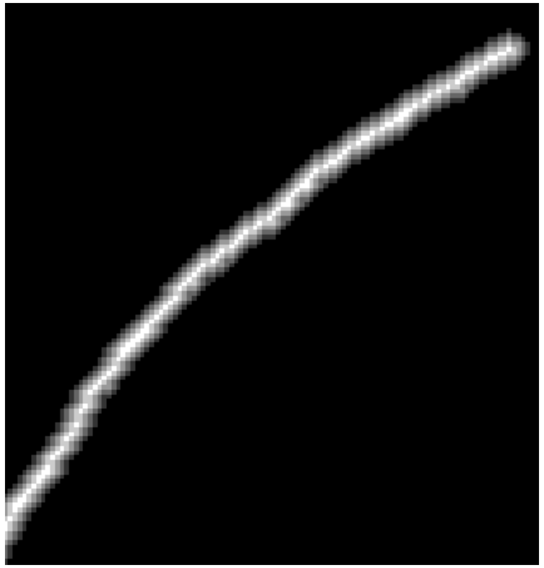
WRINKLE MODELLING

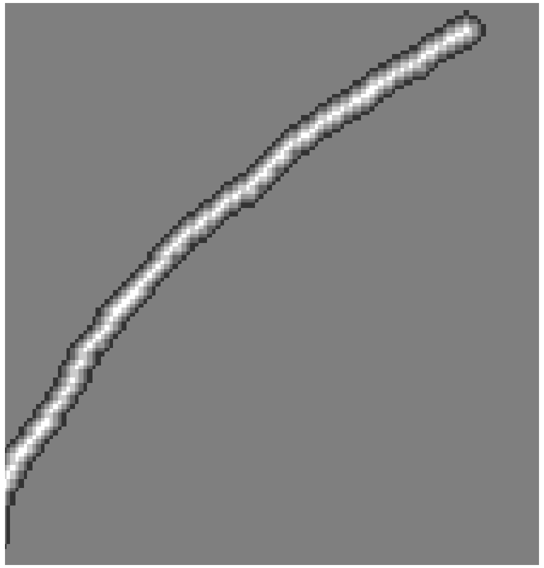
5

WRINKLE PROPERTIES

$$DT_i = \left(1 - \frac{dC_i}{h}\right) \cdot \left(-exp\left(\frac{-dP_i}{w}\right)\right) .$$







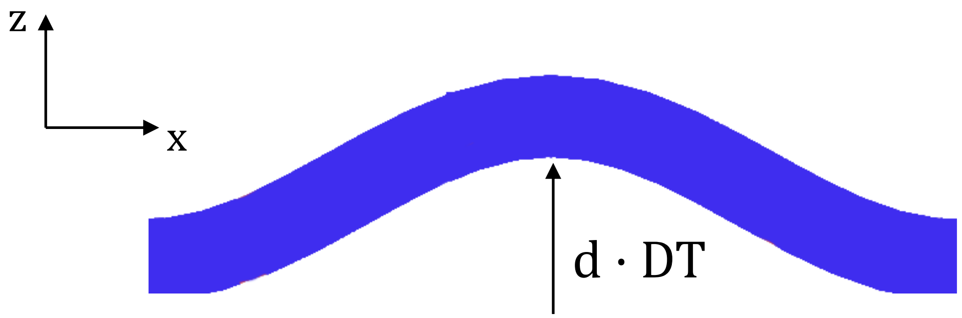
WRINKLE MODELLING

6

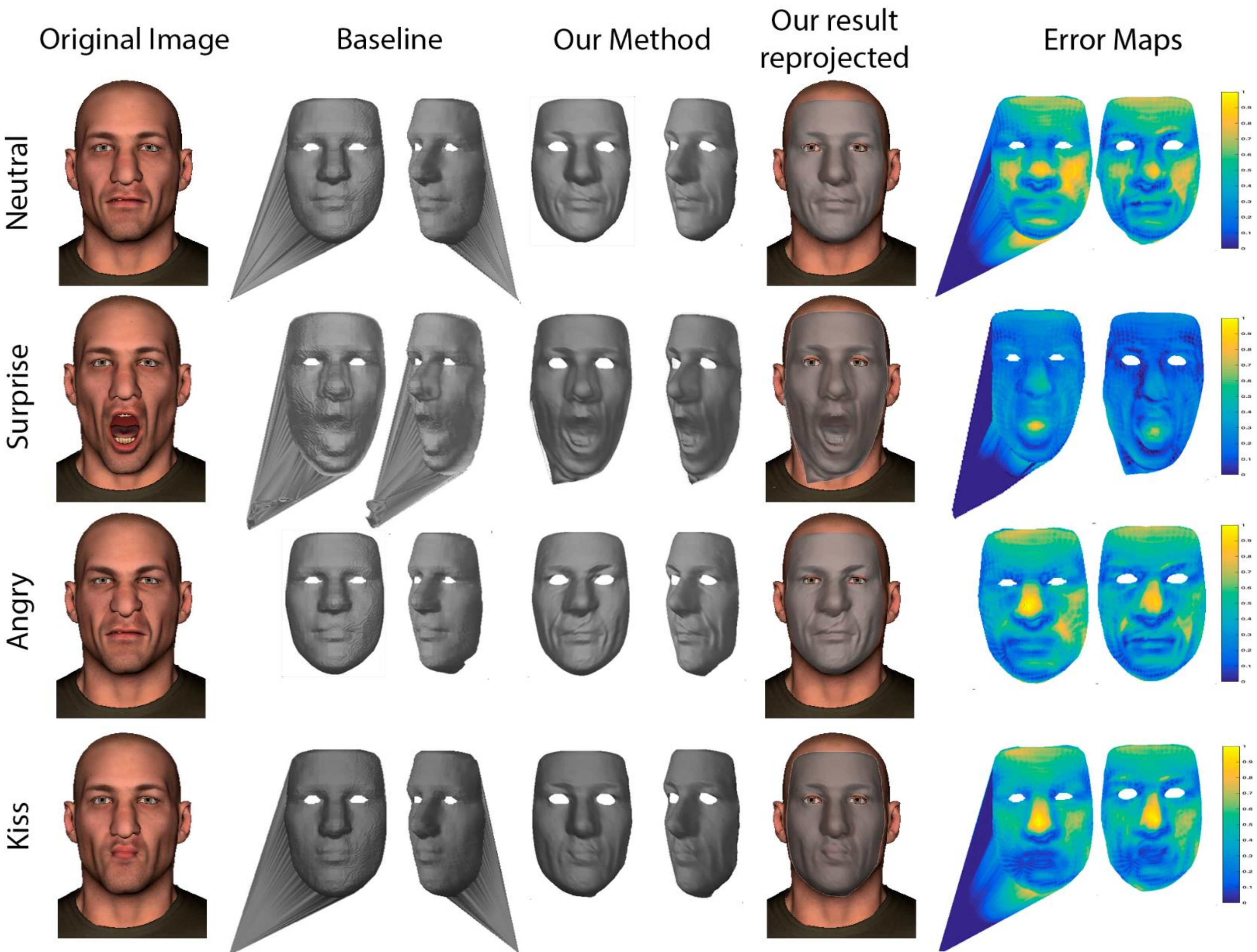
PHOTOMETRIC DEPTH ESTIMATION

$$\mathscr{E}(d) = \arg\min_{d} \sum_{i=1}^{I} \|\mathbf{I}_i - \rho_i \cdot (\mathbf{l} \cdot \mathbf{H}(\mathbf{n}_i(\mathbf{v}_i + \mathbf{d}_i)))\|_2^2$$

$$\text{subject to} \quad \mathbf{d}_i = \mathbf{n}_i \cdot d \cdot DT_i$$

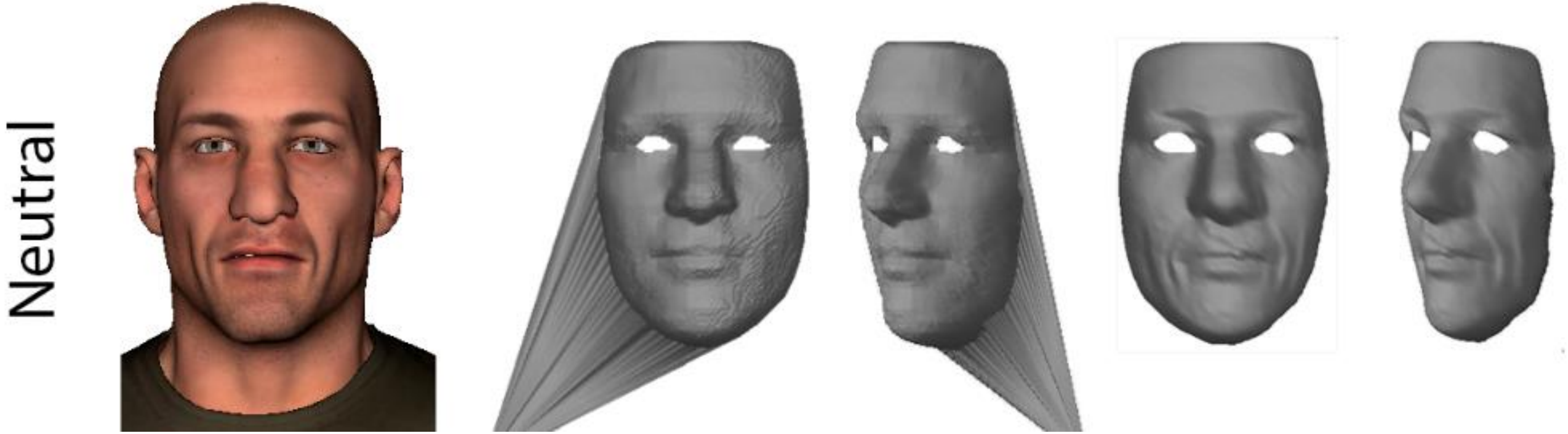


RESULTS


Original Image
Baseline
Our Method
Our result reprojected
Error Maps
Neutral
Surprise
Angry
Kiss

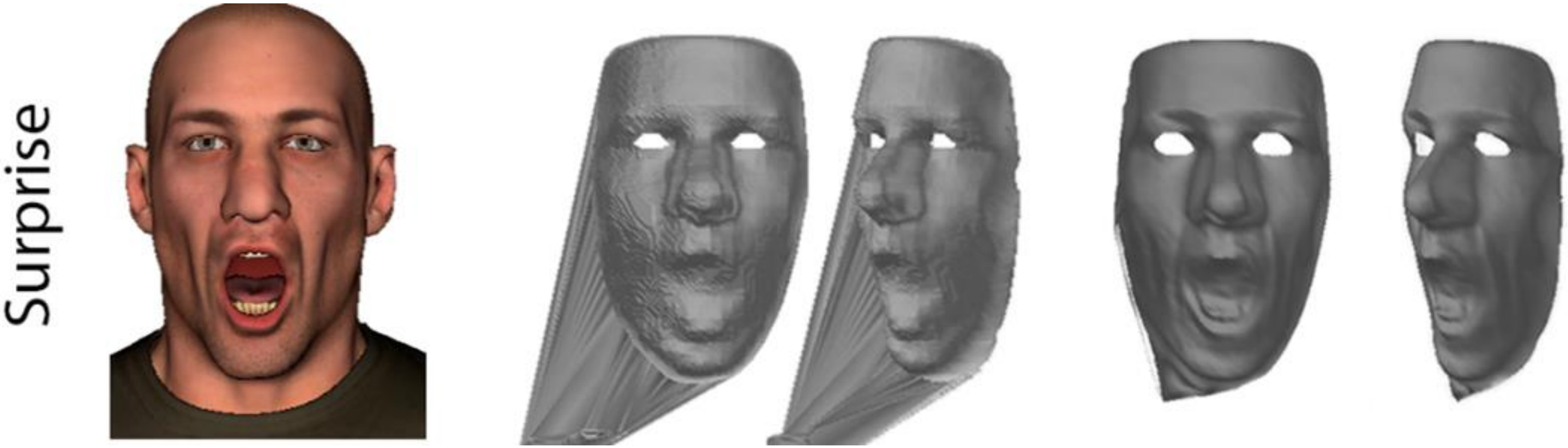
RESULTS

Close Up



RESULTS

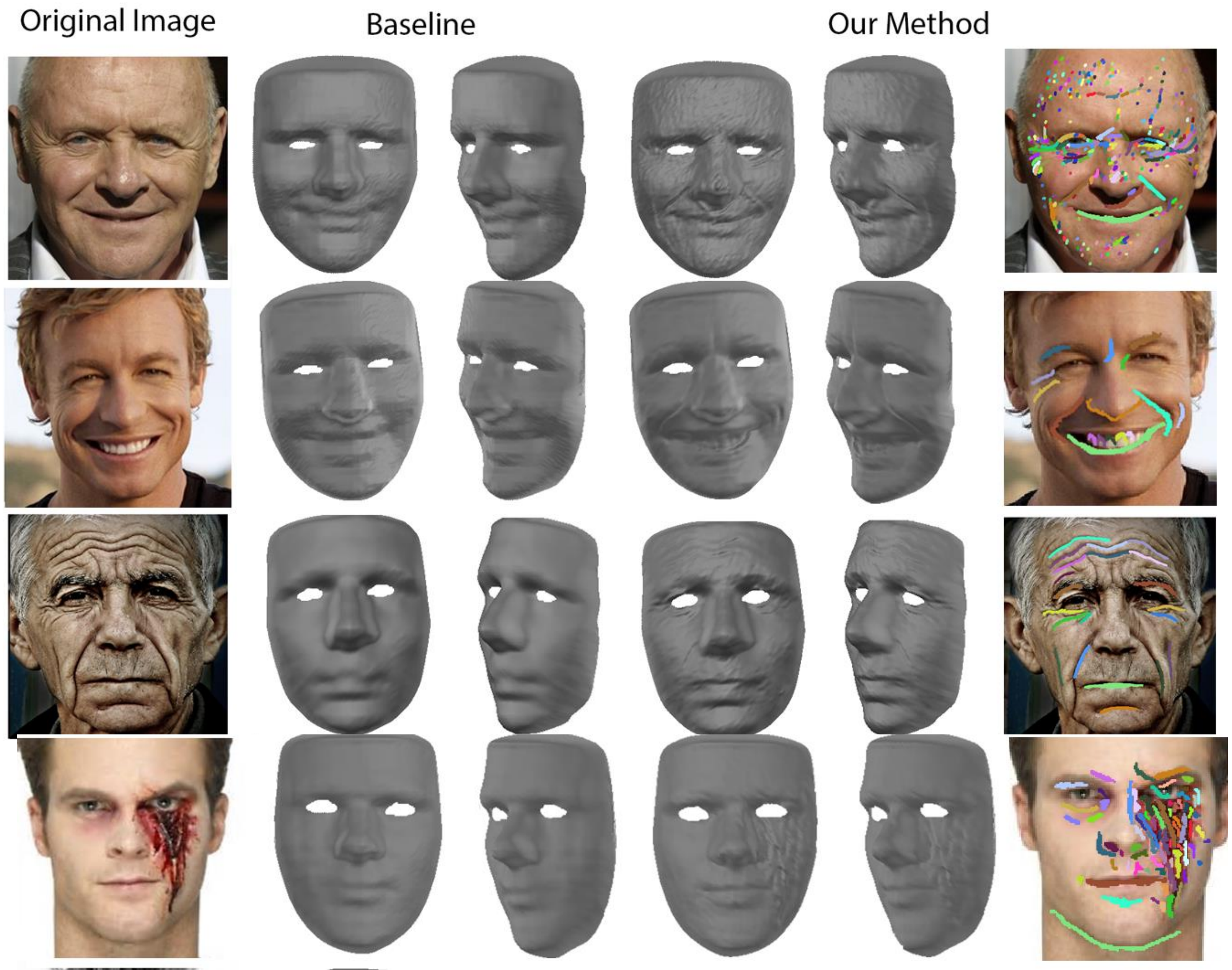
Close Up



RESULTS

Expre.		Neutral	Sad	Angry	Pain	Surprise	Kiss	Happy	Sing	Average
Jackson17	$\varepsilon_{3Dtotal}$	0.022	0.040	0.039	0.037	0.064	0.050	0.055	0.059	0.046
	$\varepsilon_{3Dwrink.}$	0.808	0.811	0.814	0.814	0.613	0.820	0.811	0.813	0.788
	$\mathscr{E}(d)_{total}$	0.393	0.518	0.261	0.494	0.259	0.422	0.240	0.133	0.340
	$\mathscr{E}(d)_{wrink.}$	0.157	0.155	0.160	0.154	0.171	0.154	0.163	0.149	0.158
Ours	$\varepsilon_{3Dtotal}$	**0.021**	**0.034**	**0.036**	**0.036**	**0.042**	**0.044**	**0.051**	**0.054**	**0.039**
	$\varepsilon_{3Dwrink.}$	**0.791**	**0.797**	**0.801**	**0.801**	**0.602**	**0.801**	**0.797**	**0.799**	**0.774**
	$\mathscr{E}(d)_{total}$	0.389	0.508	0.253	0.486	0.254	0.414	0.233	0.126	0.332
	$\mathscr{E}(d)_{wrink.}$	0.125	0.123	0.127	0.128	0.151	0.122	0.134	0.136	0.131
Ours	nW	21	28	27	23	27	22	22	19	23.6
	$t(s)$	4.996	2.573	2.570	2.436	7.282	7.306	2.593	2.491	4.031

RESULTS



RESULTS

RESULTS

Close Up

Garrido13

TEMPORAL PRIORS

200K Points

OUR METHOD

SINGLE IMAGE

20K Points

CONCLUSIONS

1. Reconstruct wrinkles from a single image
2. Easy and fast wrinkle modelling
3. No training data is required not a priori information
4. Our method may fail with strong cast shadows, severe occlusions, and strong textures

2

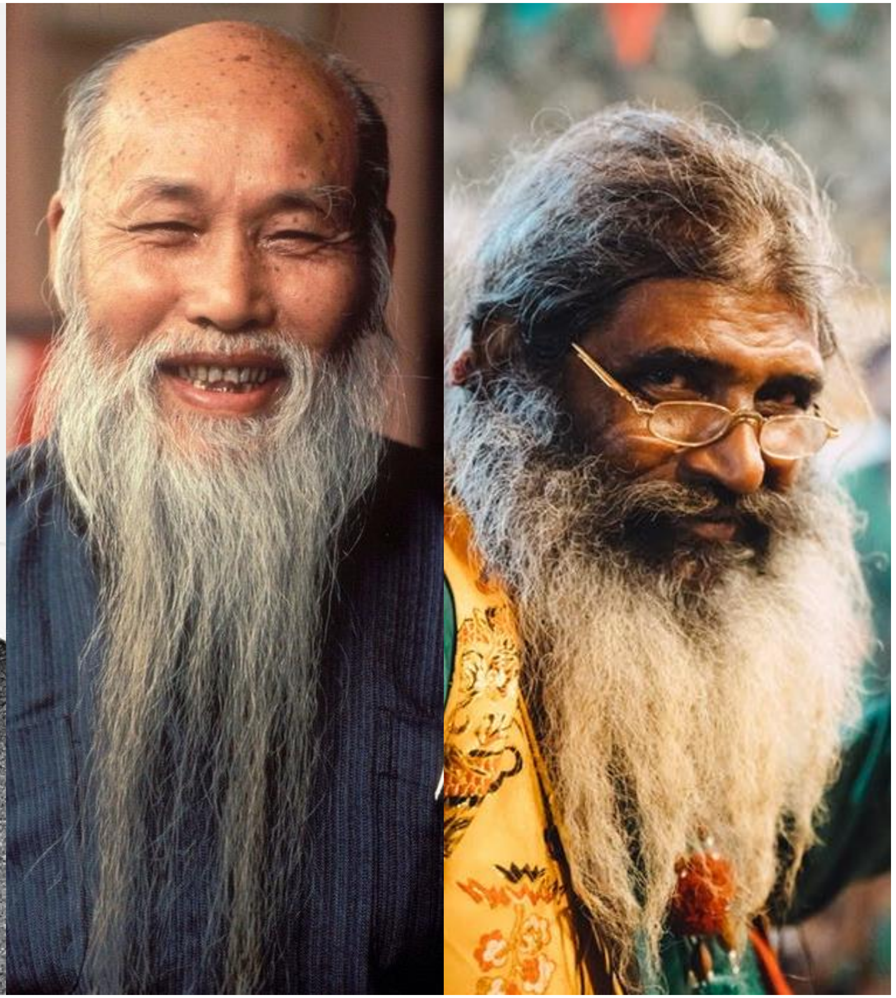
FACIAL HAIR

BEARDS ARE A DISTINCTIVE FEATURE

HISTORICAL

BEARDS ARE A DISTINCTIVE FEATURE

FASHION



CULTURAL

HOWEVER THEY ARE NOT QUITE POPULAR

HAIR / HAIRSTYLE RECONSTRUCTION

Paris04

Paris05

Wei05

Jakob09

Chai13

Luo13

Hu14

Yu14

Hu15

Chai15

Chai16

Zhang17

Chen18

Hachmann18

Zhou18

FACIAL HAIR RECONSTRUCTION

Beeler12

RELATED WORK

STEREO: Calibrated Multicamera Setup

Paris04, Paris08, Wei05, Jakob09, Beeler12

ACTIVE LIGHT: Requires Specific Setup

Chen18, Hachmann18

DATA DRIVEN: Requires large amounts of data, low-rank

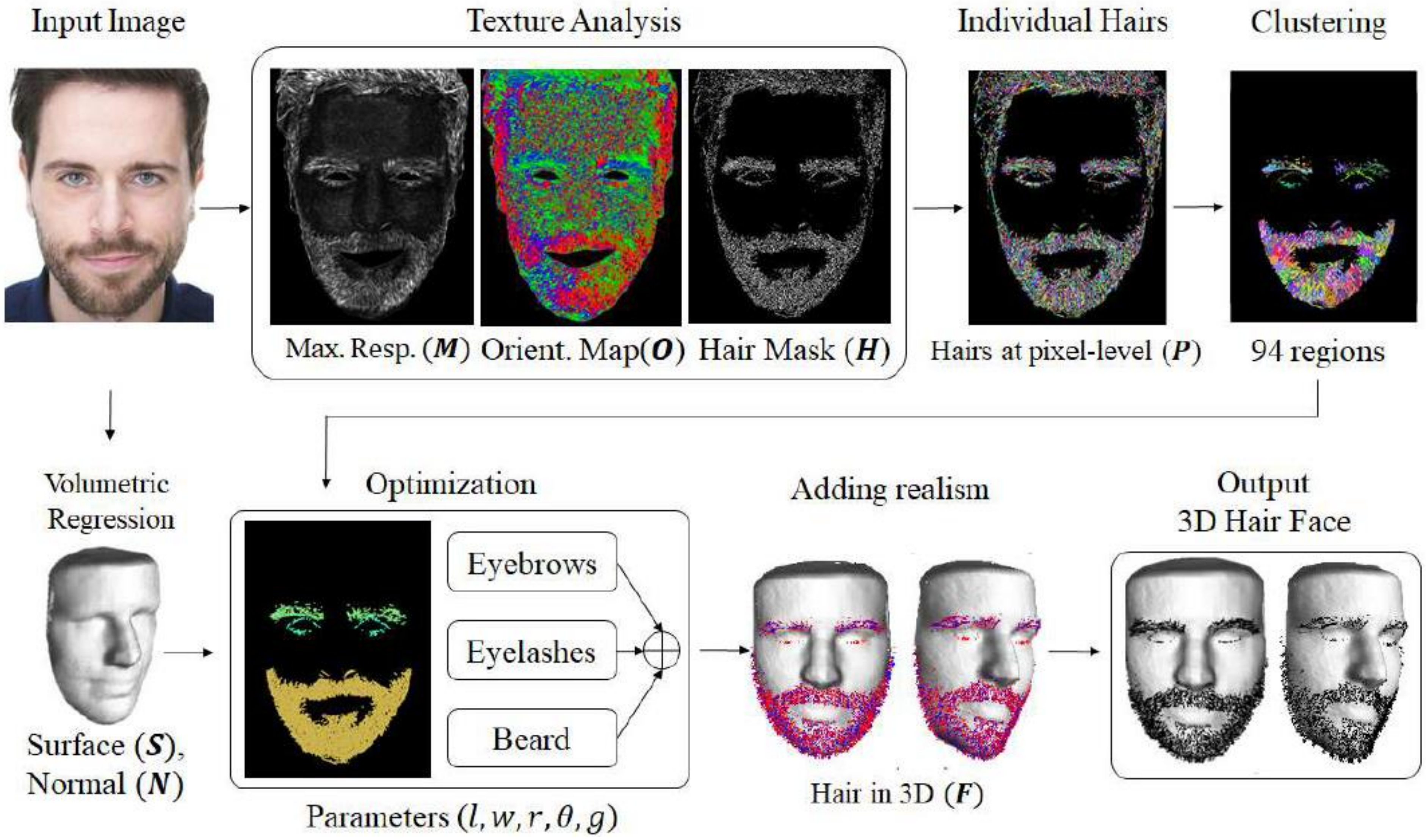
Hu14, Hu15, Luo13, Zhang17

DEEP LEARNING: Even more data required, training setups

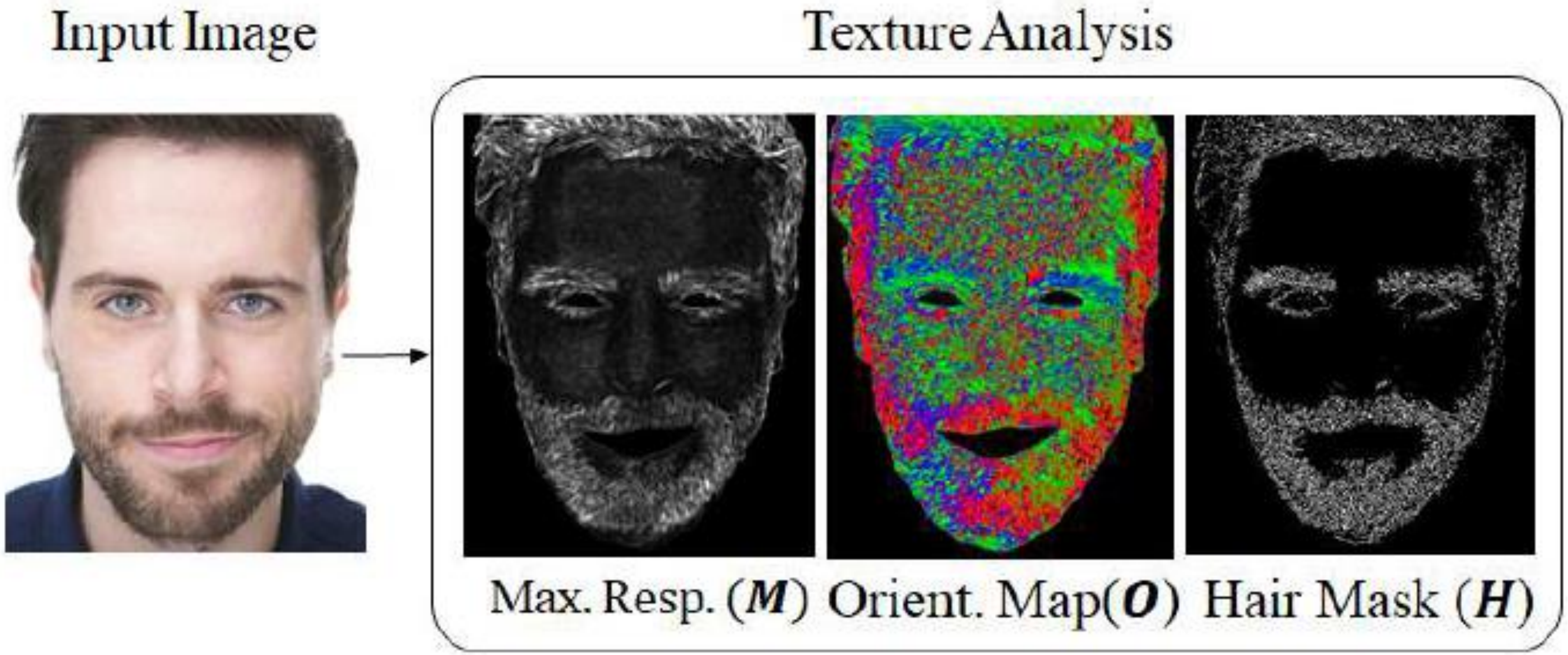
Chai13, Chai15, Chai16, Zhou18

OUR METHOD: Single Image, does not require any training data

SCHEME

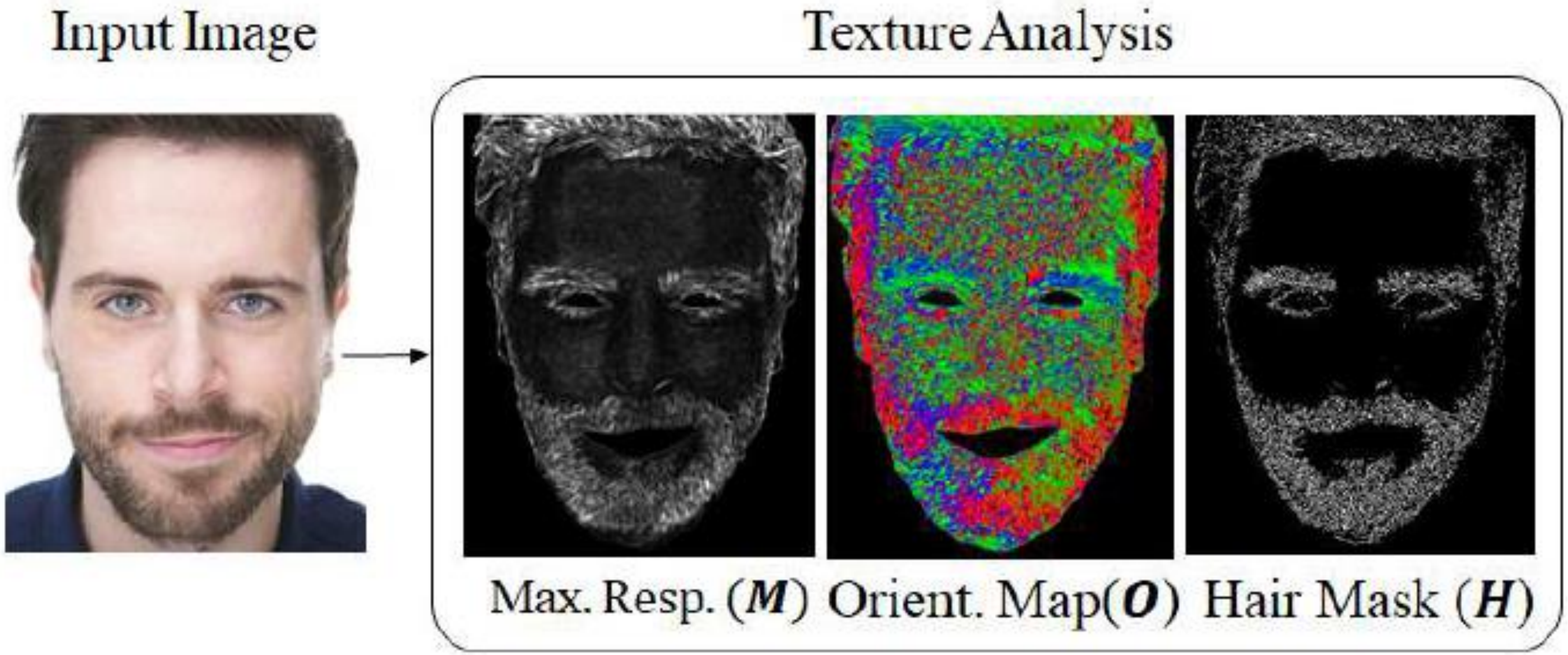

Input Image
Texture Analysis
Individual Hairs
Clustering
Max. Resp. (M)
Orient. Map(O)
Hair Mask (H)
Hairs at pixel-level (P)
94 regions
Volumetric Regression
Surface (S), Normal (N)
Optimization
Eyebrows
Eyelashes
Beard
Parameters (l, w, r, θ, g)
Adding realism
Hair in 3D (F)
Output 3D Hair Face

TEXTURE ANALYSIS



$$F(x, y) = |K_\theta * V|_{(x,y)} + |K_\theta * S|_{(x,y)}$$

TEXTURE ANALYSIS

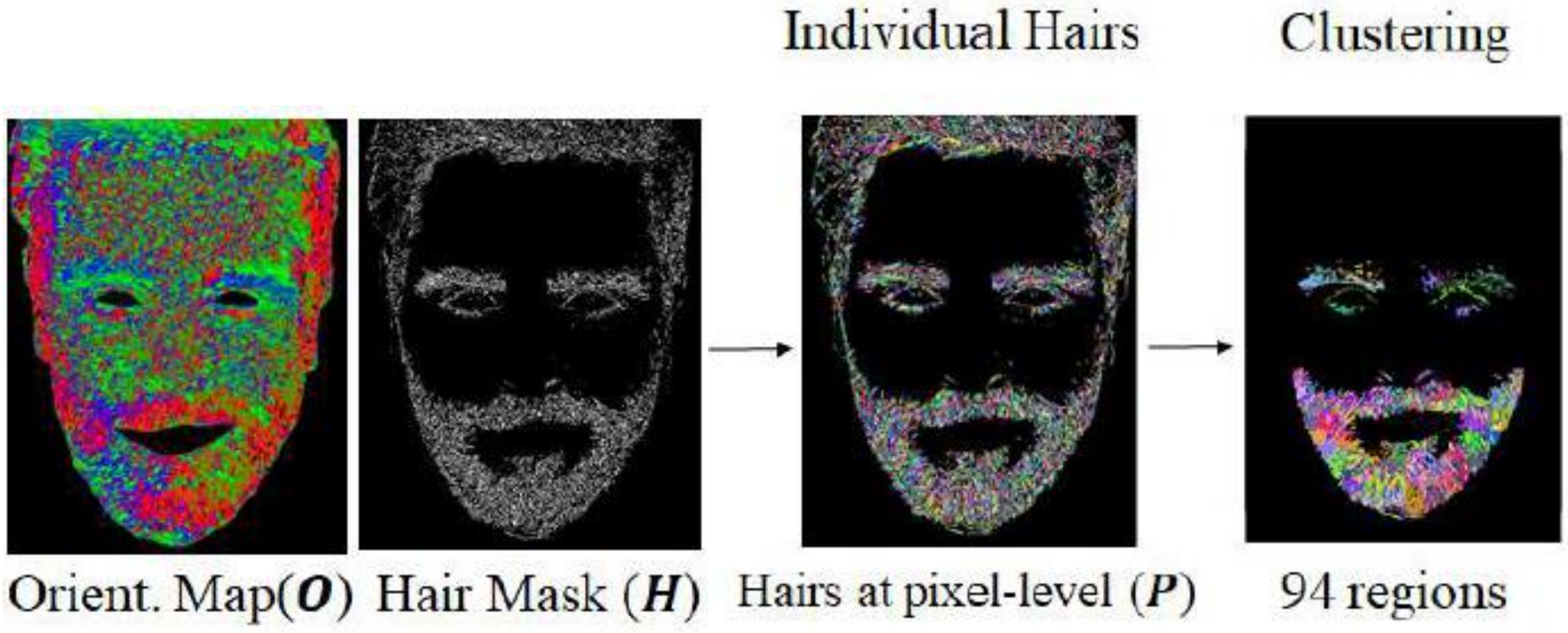


$$\boldsymbol{M}(x, y) = \max(F(x, y)) \, ,$$

$$\boldsymbol{O}(x, y) = \theta_{\max(F(x,y))} \, ,$$

$$\boldsymbol{H}(x, y) = M(x, y) > \tau \, .$$

HAIR TRACES + CLUSTERS



PARAMETRIC HAIR MODEL

1 SIZE (LENGHT + WIDTH)

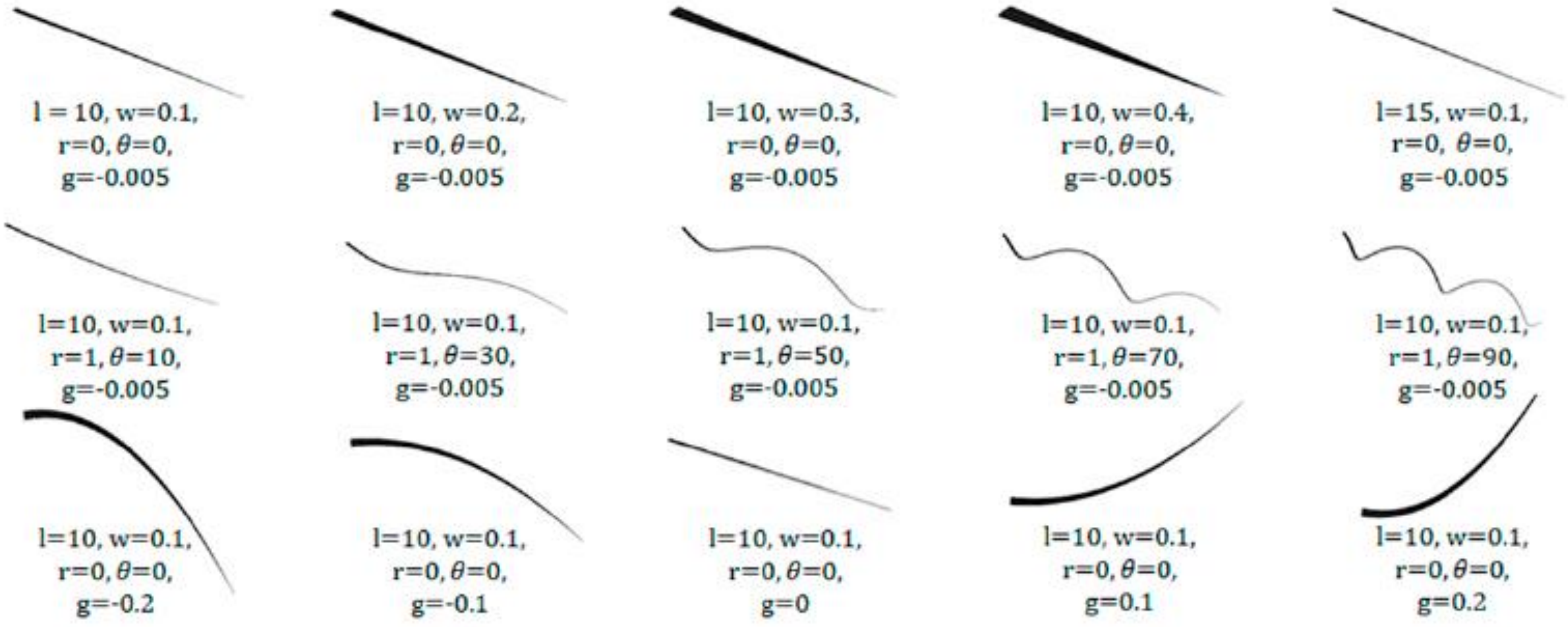
$[l, w]$

2 CURVINESS

$[r, \theta]$

3 GRAVITY-LIKE EFFECT

$[g]$



PARAMETRIC HAIR MODEL

$$\boldsymbol{F}_i^h(\boldsymbol{p}^h, \boldsymbol{n}^h, l, w, r, \theta, g) = \boldsymbol{p}_0^h + \boldsymbol{R}(\boldsymbol{n}^h)\frac{(i-1)\cdot l}{s} Hx(r, \theta, i) - Gy(g, i)\,,$$

3D-Helix ***Gravity-like effect***

3D-Helix

$$Hx(r, \theta, i) = (i-1, r\cdot\sin(\theta\cdot(i-1)), r\cdot\cos(\theta\cdot(i-1)))\,,$$

Gravity-like effect

$$Gy(g, i) = \boldsymbol{v}\cdot(i-1)\cdot\left(\cos(\alpha), -\frac{g(i-1)^2}{2}, \sin(\alpha)\right)\,,$$

Parameters: $l, w, r, \theta, g, (s)$

ENERGY MINIMIZATION

$$\varepsilon_{\text{total}} = \varepsilon_{\text{len}} + \varepsilon_{ori} + \varepsilon_{tip} + \varepsilon_{cur}$$

ENERGY MINIMIZATION

length

$$\varepsilon_{\text{total}} = \varepsilon_{\text{len}} + \varepsilon_{ori} + \varepsilon_{tip} + \varepsilon_{cur}$$

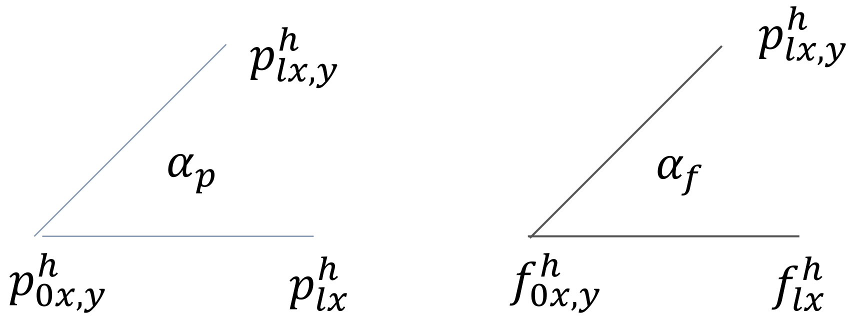
$$\varepsilon_{len} = \sum_{h} ||(p_l^h - p_0^h) - (f_l^h - f_0^h)||_2^2$$

ENERGY MINIMIZATION

orientation

$$\varepsilon_{\text{total}} = \varepsilon_{\text{len}} + \varepsilon_{ori} + \varepsilon_{tip} + \varepsilon_{cur}$$

$$\varepsilon_{ori} = \sum_{h} ||tan^{-1}\left(\frac{p_{ly}^{h} - p_{0y}^{h}}{p_{lx}^{h} - p_{0x}^{h}}\right) - tan^{-1}\left(\frac{f_{ly}^{h} - f_{0y}^{h}}{f_{lx}^{h} - f_{0x}^{h}}\right)||_{2}^{2}$$



ENERGY MINIMIZATION

tip-to-tip

$$\varepsilon_{\text{total}} = \varepsilon_{\text{len}} + \varepsilon_{ori} + \varepsilon_{tip} + \varepsilon_{cur}$$

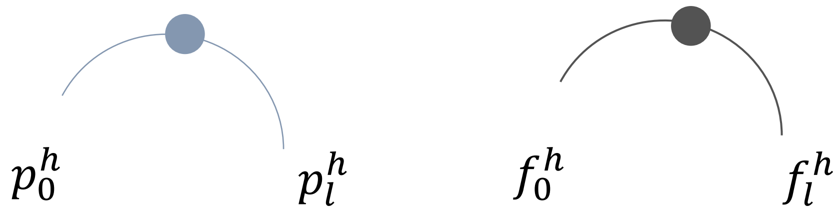
$$\varepsilon_{tip} = \sum_{h} ||\left(p_l^h - f_l^h\right)||_2^2$$

ENERGY MINIMIZATION

curviness

$$\varepsilon_{\mathrm{total}} = \varepsilon_{\mathrm{len}} + \varepsilon_{ori} + \varepsilon_{tip} + \varepsilon_{cur}$$

$$\varepsilon_{cur} = \sum_{h} \sum_{i} \left\| \frac{\left|\left(p_l^h - p_0^h\right) - \left(p_0^h - p_i^h\right)\right|}{\left(p_l^h - p_0^h\right)} - \frac{\left|\left(f_l^h - f_0^h\right) - \left(f_0^h - f_i^h\right)\right|}{\left(f_l^h - f_0^h\right)} \right\|_2^2$$



ADD FURTHER REALISM

Appending Hair Density

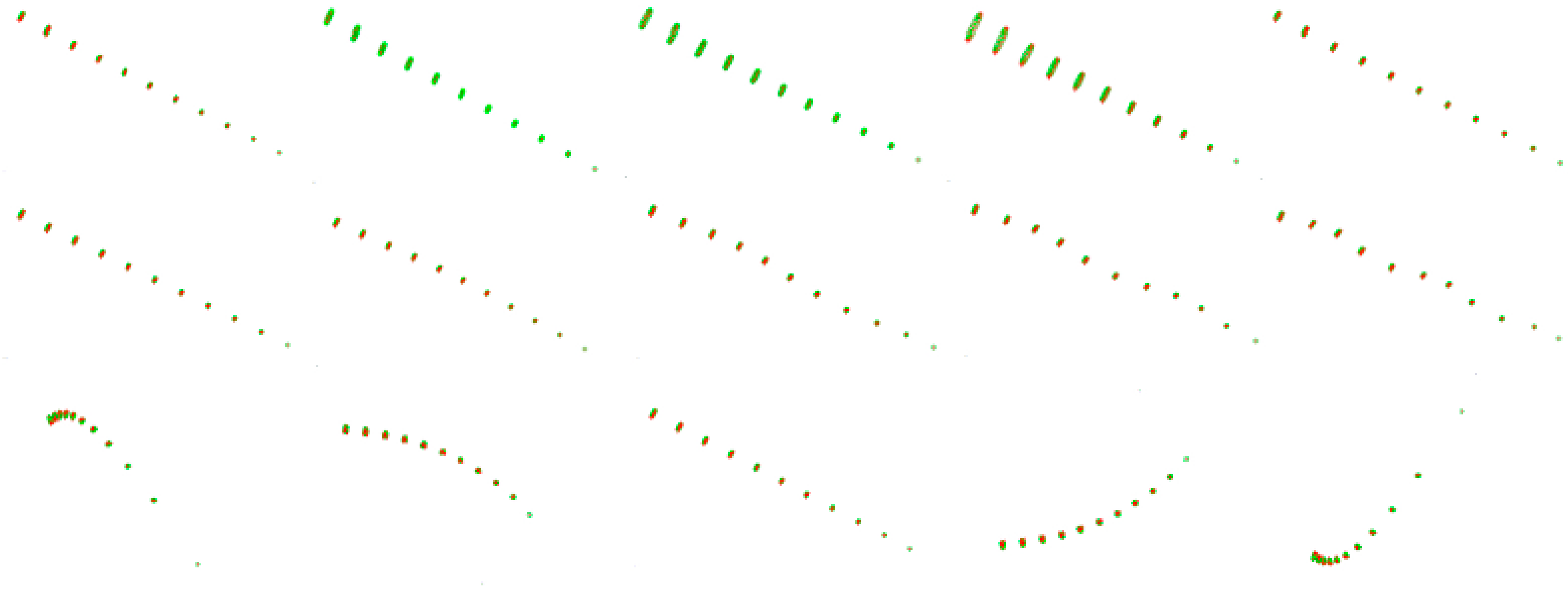
$$G = \sum_i \sum_j (\boldsymbol{H}(i,j) - \pi(\boldsymbol{F})(i,j)) \cdot (\pi(\boldsymbol{F}) \oplus \boldsymbol{D})(i,j)\,,$$

Small Random Variations

$$\boldsymbol{F}^h(\boldsymbol{p}^h, \boldsymbol{n}^h + \boldsymbol{\lambda}_r, l + \lambda_l, w, r, \theta, g)\,,$$

RESULTS

Synthetic Experiment



RESULTS

Synthetic Experiment

(row,col)	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	Average
error 3D	0.027	0.004	0.006	0.014	0.017	0.014	0.013	0.017	0.015	0.013	0.002	0.002	0.006	0.003	0.008	0.011
time (s)	40.443	40.936	50.613	51.977	40.996	52.167	52.873	62.767	51.463	51.853	40.443	59.267	61.289	61.311	64.152	52.170

RESULTS

Real Data

RESULTS

Real Data

RESULTS

Real Data

RESULTS

Image						
im res.	849×1273	1024×768	1200×825	1600×2400	4000×6000	393×588
hairs up	766(61)	62(2)	72(6)	181(19)	804(6)	168(13)
hairs lo	10251(510)	9575(155)	17874(408)	4647(41)	0(0)	0(0)

Image						
im res.	1265×1920	845×650	1760×2640	1750×1168	3999×3999	5075×5760
hairs up	551(63)	609(55)	66(9)	334(67)	7132(238)	162(7)
hairs lo	6250 (212)	7332(746)	7733(440)	9384(361)	0(0)	0(0)

RESULTS

Real Data- Close Up

Eyelashes

Soft Beard

Frida Kahlo
Eyebrows

Medium density
mustache

RESULTS

Comparison with Beeler12

Beeler12 Ours

CONCLUSIONS

1. Easy and intuitive modelling
2. Any data is required nor a priori information
3. May fail with strong orientation changes, and depends on the quality of the individual hair tracing

EXPRESSIONS PART II

Blending shapes expressions developed by the digital artist Francesco Lupo

3

EXPRESSION TRANSFER

INTRODUCTION

RELATED WORK – MOTION CAPTURE

MARKERS SETUPS

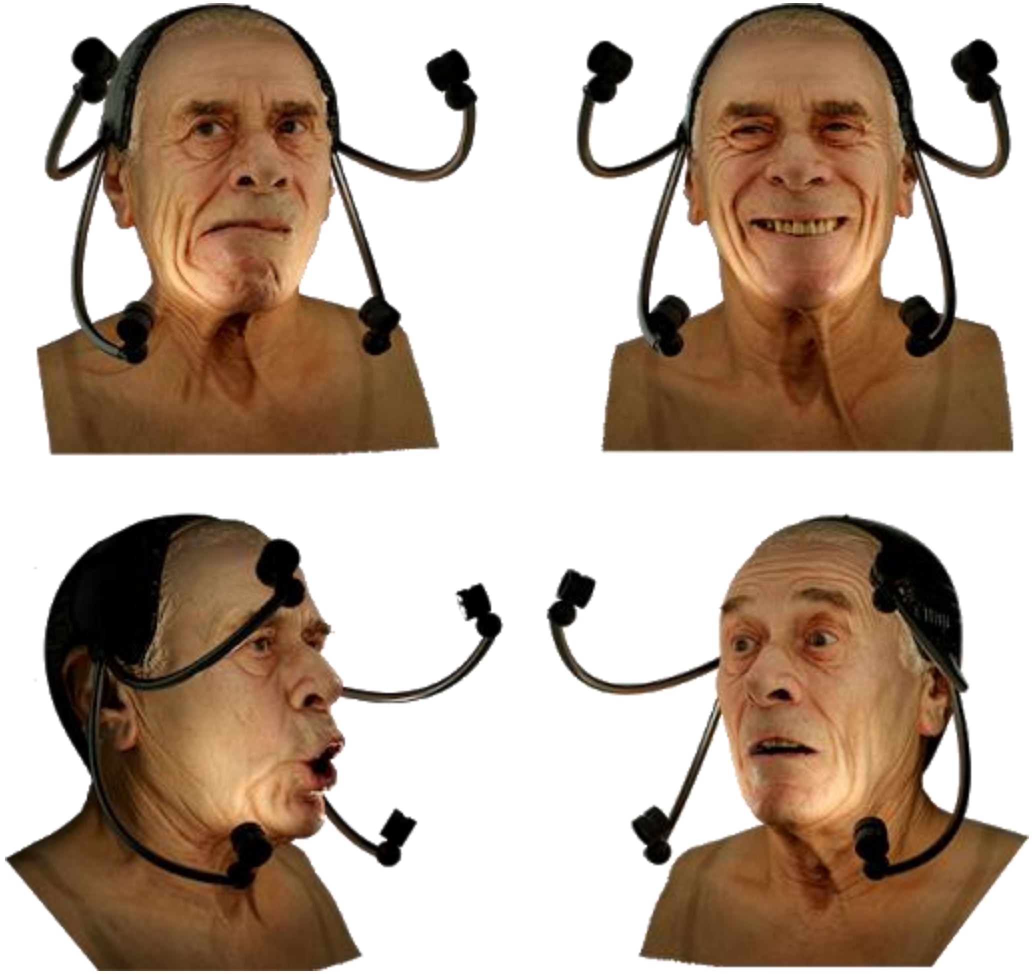
OPTICAL MARKERS:

Bickel07

LIGHT PATTERNS:

Weise11

RELATED WORK – MOTION CAPTURE



MARKERLESS SETUPS

LINEAR OPTIMIZATION

Cao13

MULTILINEAR OPTIMIZATION

Vlasic05, Cao14

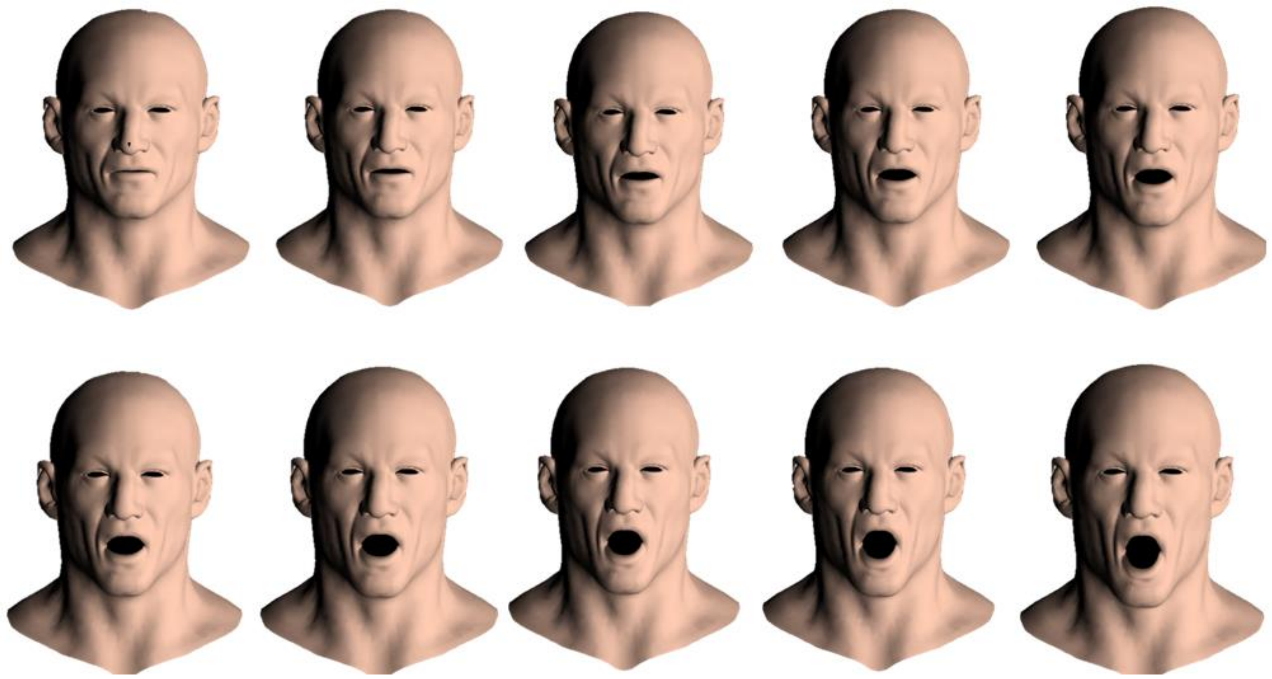
MULTIVIEW STEREO

Beeler10, Bradley10

RGB-D

Bouaziz13, Thies15

RELATED WORK – EXPRESSION TRANSFER



2D-to-2D

Thies15, Averbuch-Elor17

3D-to-3D

Sumner04

2D-to-3D

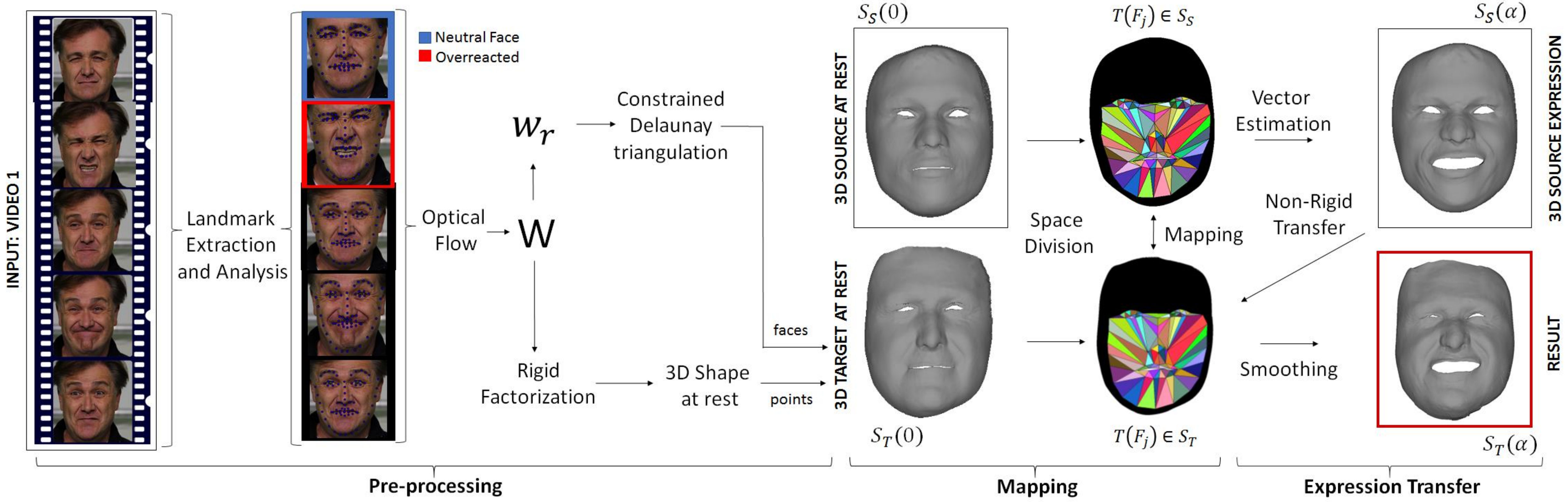
Thies16 (low-rank, no detail)

OUR METHOD (2D-to-3D)

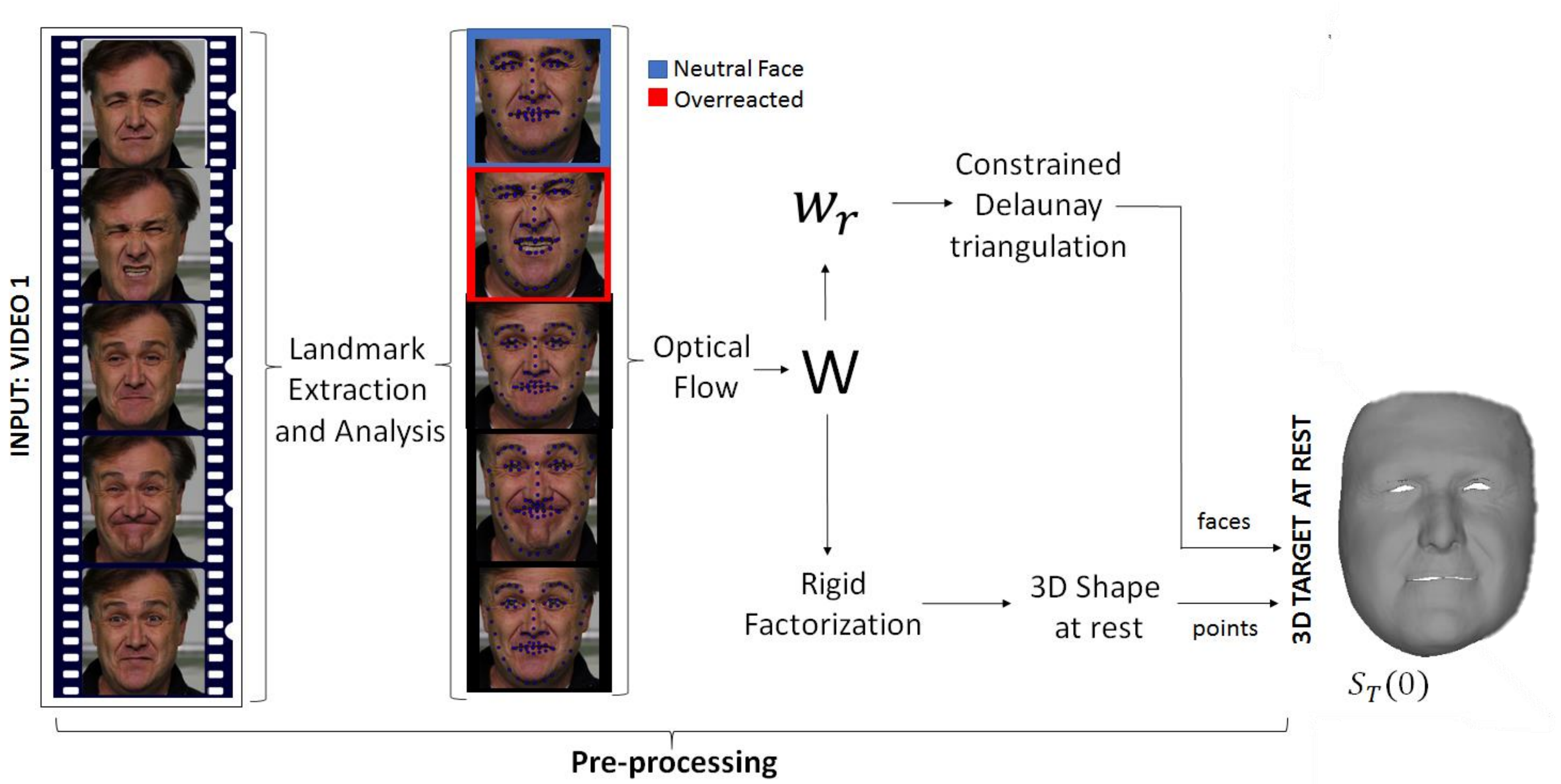
No low-rank restrictions

Full detail

SCHEME


INPUT: VIDEO 1
Landmark Extraction and Analysis
Neutral Face
Overreacted
Optical Flow
W
W_r
Constrained Delaunay triangulation
Rigid Factorization
3D Shape at rest
faces
points
$S_S(0)$
3D SOURCE AT REST
3D TARGET AT REST
$S_T(0)$
$T(F_j) \in S_S$
Space Division
Mapping
$T(F_j) \in S_T$
Vector Estimation
Non-Rigid Transfer
Smoothing
$S_S(\alpha)$
3D SOURCE EXPRESSION
RESULT
$S_T(\alpha)$
Pre-processing
Mapping
Expression Transfer

SCHEME


INPUT: VIDEO 1
Landmark
Extraction
and Analysis
Neutral Face
Overreacted
Optical
Flow
W
w_r
Constrained
Delaunay
triangulation
Rigid
Factorization
3D Shape
at rest
faces
points
3D TARGET AT REST
$S_T(0)$
Pre-processing

PRE-PROCESSING

From RGB to 3D model

1 → **2** → **3** →

1. INITIAL VIDEO
2. OPTICAL FLOW

   W

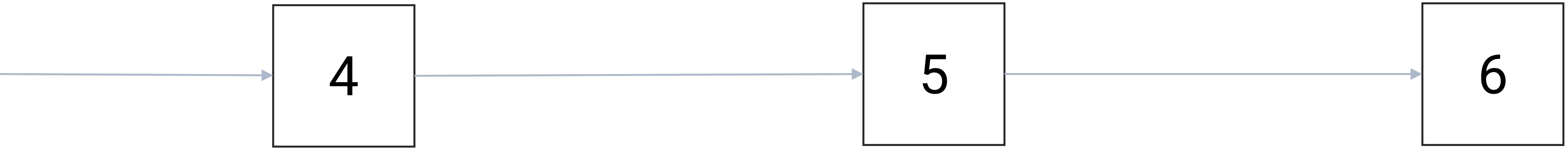
   $$W = \begin{bmatrix} w_1^1 & \cdots & w_1^N \\ \vdots & \ddots & \vdots \\ w_{2F}^1 & \cdots & w_{2F}^N \end{bmatrix},$$

3. OVERREACTED FRAMES REMOVAL

   $W \rightarrow W_r$

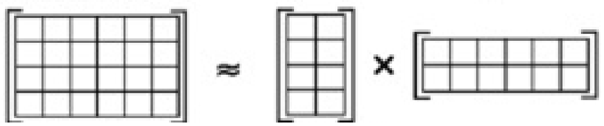
PRE-PROCESSING

From RGB to 3D model

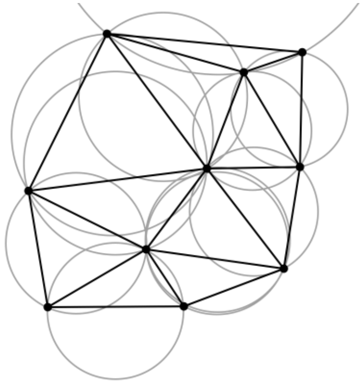


MATRIX FACTORIZATION

$$Wr = RS_T(0)$$

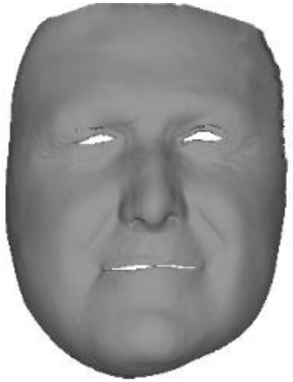


DELAUNAY TRIANGULATION

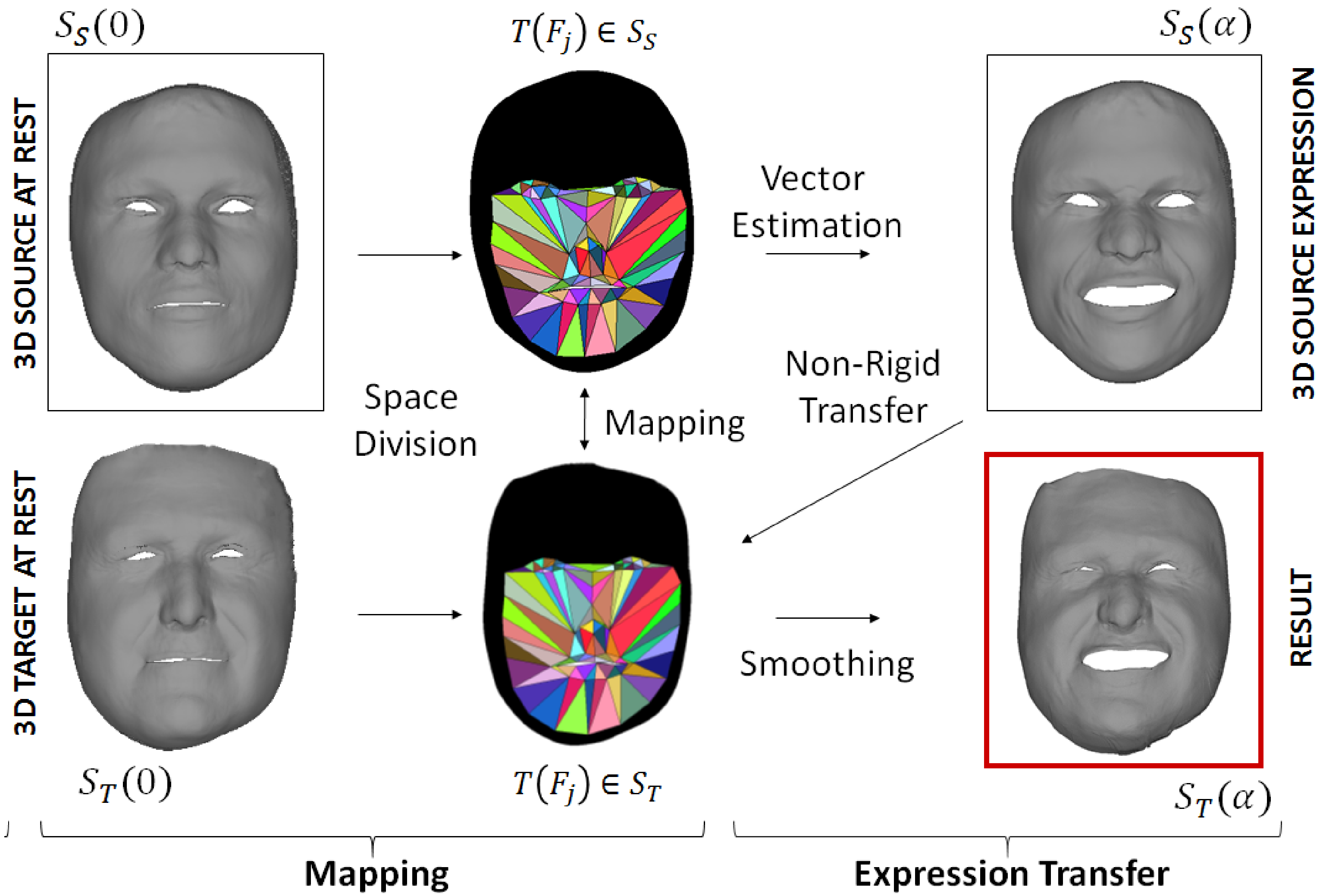


SHAPE MESH

$$S_T(0)$$



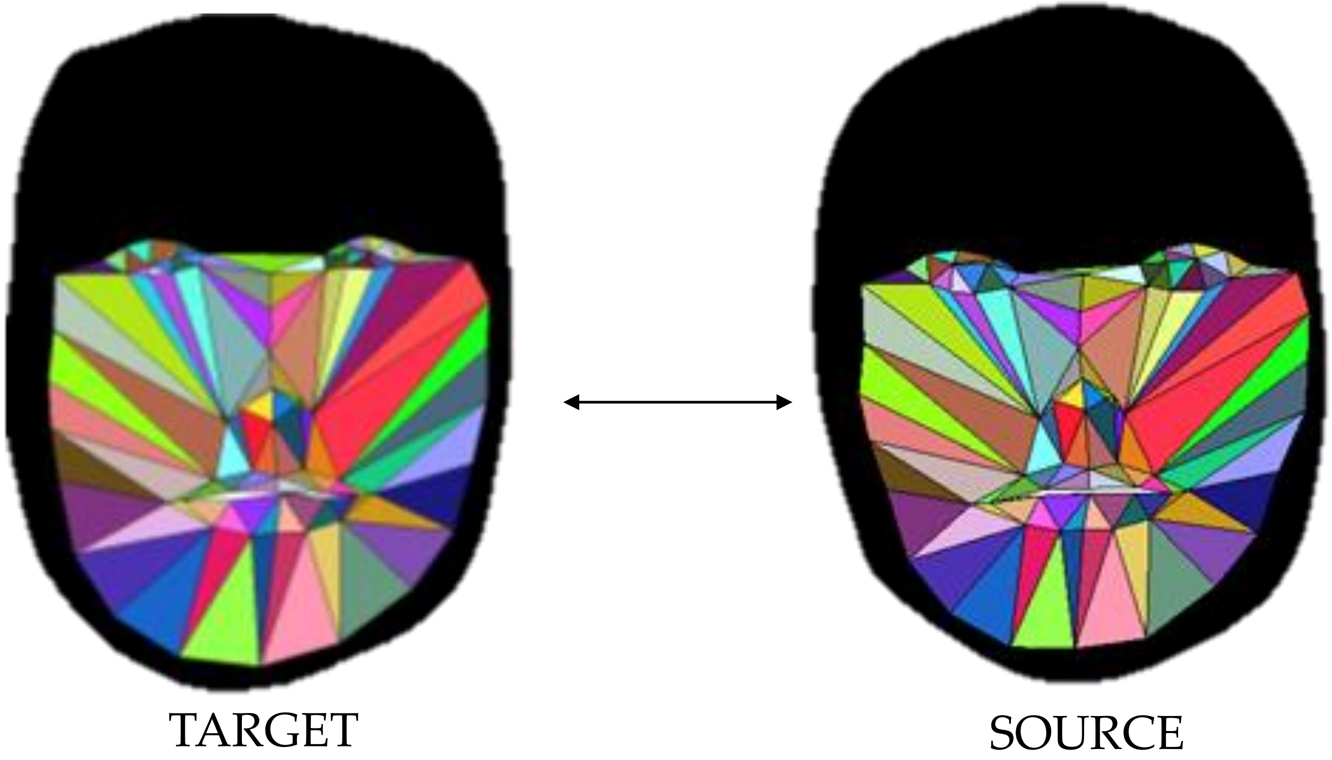
SCHEME



$S_S(0)$
$T(F_j) \in S_S$
$S_S(\alpha)$
3D SOURCE AT REST
Vector Estimation
3D SOURCE EXPRESSION
Non-Rigid Transfer
Space Division
Mapping
3D TARGET AT REST
Smoothing
RESULT
$S_T(0)$
$T(F_j) \in S_T$
$S_T(\alpha)$
Mapping
Expression Transfer

MAPPING PROCES

Low Resolution Correspondence

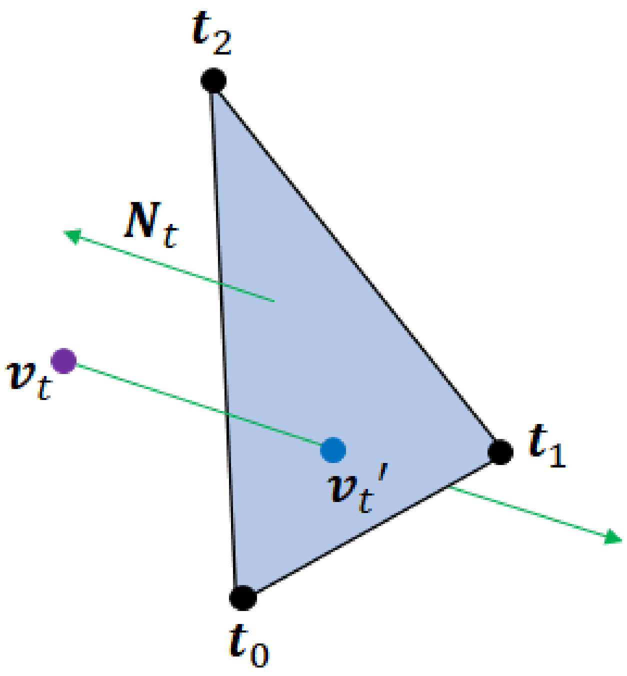


68 FACIAL LANDMARKS

101 TRIANGULAR REGIONS

SUB-REGION CLASSIFICATION

Inside-Outside Function



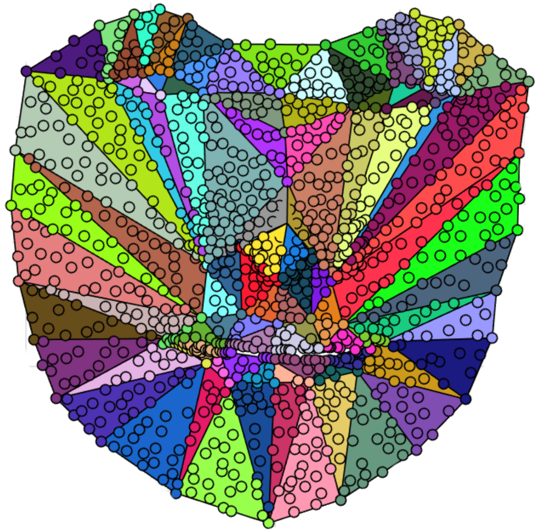
$$\mu(v_t^i, t_j^n) = \begin{cases} 1, if\ \boldsymbol{N}((t_j^{n+1} - t_j^n) \times (v_t^i - t_j^n)) \\ 0, otherwise \end{cases}$$

SUB-REGION CLASSIFICATION

Graphical Representation

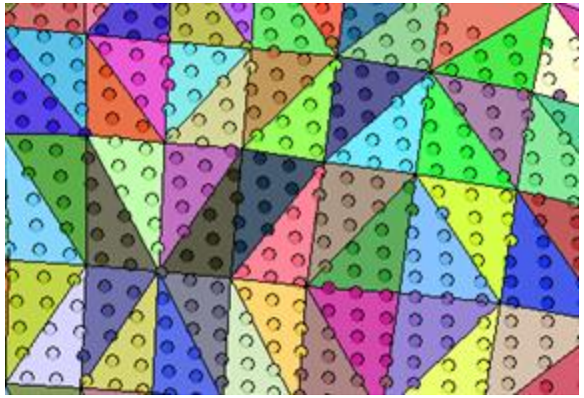
COARSE CLASSIFICATION

101 regions



FINE CLASSIFICATION

k regions



SUB-REGION CLASSIFICATION

Mathematical Representation

COARSE CLASSIFICATION

101 regions

FINE CLASSIFICATION

k regions

$$\boldsymbol{T}^{\mathcal{F}|} = \begin{bmatrix} t_{x0}^{j} - t_{x2}^{j} & t_{x1}^{j} - t_{x2}^{j} \\ t_{y0}^{j} - t_{y2}^{j} & t_{y1}^{j} - t_{y2}^{j} \\ t_{z0}^{j} - t_{z2}^{j} & t_{z1}^{j} - t_{z2}^{j} \end{bmatrix}$$

$$\boldsymbol{T}^{C_T^k} = \begin{bmatrix} c_{x0}^{k} - c_{x2}^{k} & c_{x1}^{k} - c_{x2}^{k} \\ c_{y0}^{k} - c_{y2}^{k} & c_{y1}^{k} - c_{y2}^{k} \\ c_{z0}^{k} - c_{z2}^{k} & c_{z1}^{k} - c_{z2}^{k} \end{bmatrix}$$

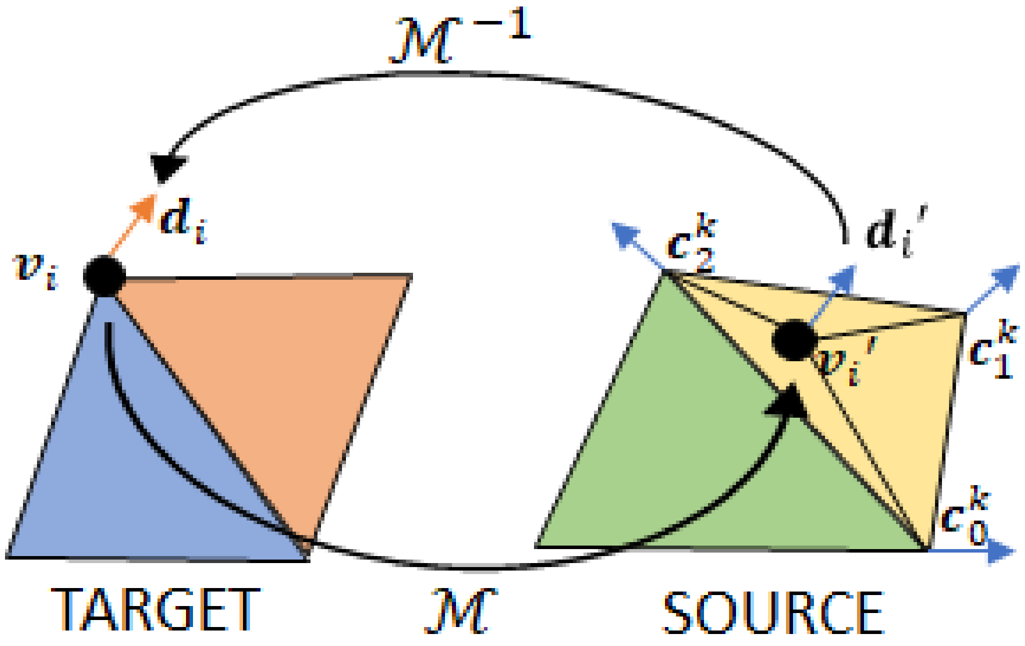
MAPPING FUNCTIONS

REGULAR

$$\mathcal{M}: \boldsymbol{v}_s^{i\,\prime} = \left(\left(\boldsymbol{T}^{C_T^k}\left(\mathcal{E}\left(\boldsymbol{T}^{\mathcal{F}|}\right)^{-1}\left(\boldsymbol{v}_t^i - \boldsymbol{t}_3\right)\right) + \boldsymbol{c}_3^k\right) + \boldsymbol{d}\right)(-\boldsymbol{N})$$

INVERSE

$$\mathcal{M}^{-1}: \boldsymbol{v}_t^{i\,\prime} = \left(\mathcal{E}\left(\boldsymbol{T}^{\mathcal{F}|}\right)\left((\boldsymbol{T}^{C_T^k})^{-1}\left(\boldsymbol{d}_i'(\boldsymbol{N}) - \boldsymbol{c}_3^k - \boldsymbol{d}\right)\right) + \boldsymbol{t}_3\right)$$

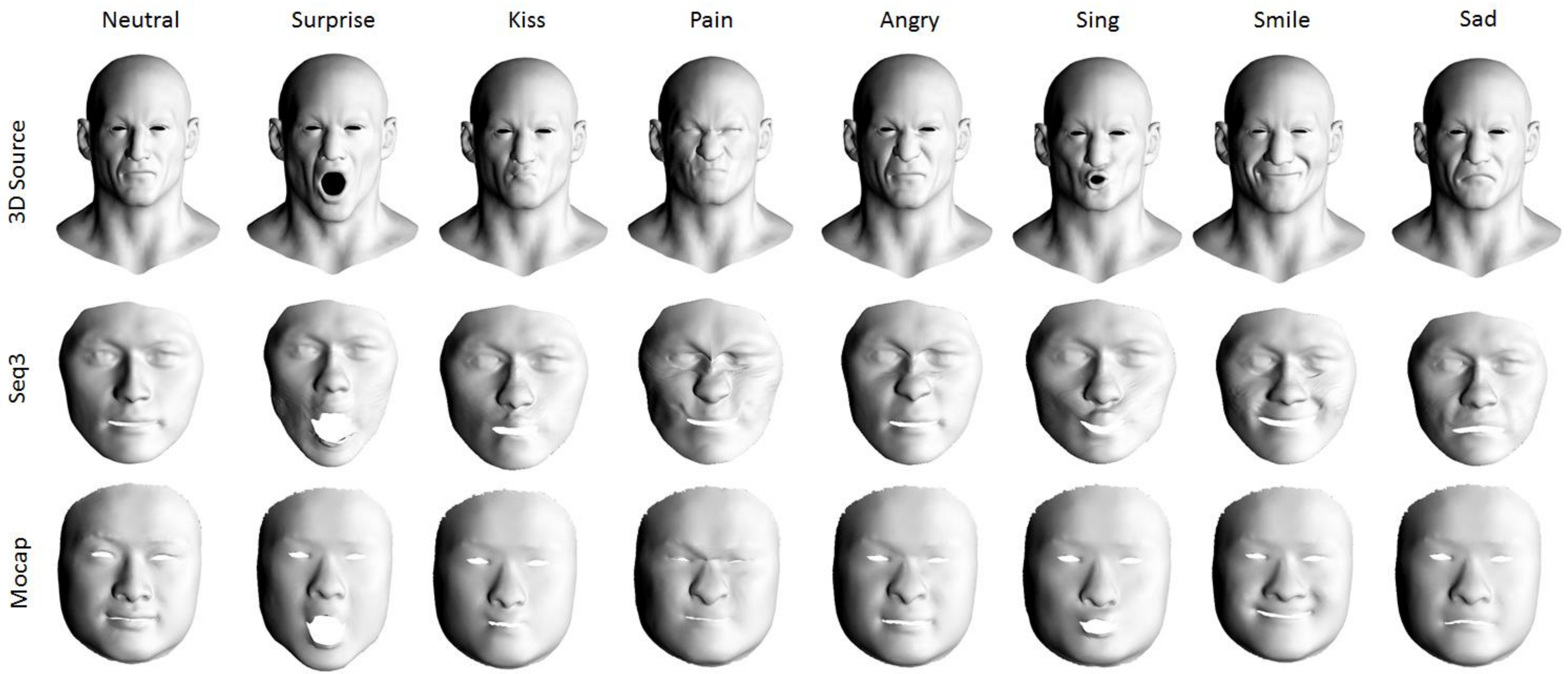


EXPRESSION TRANSER

Smoothing Energy

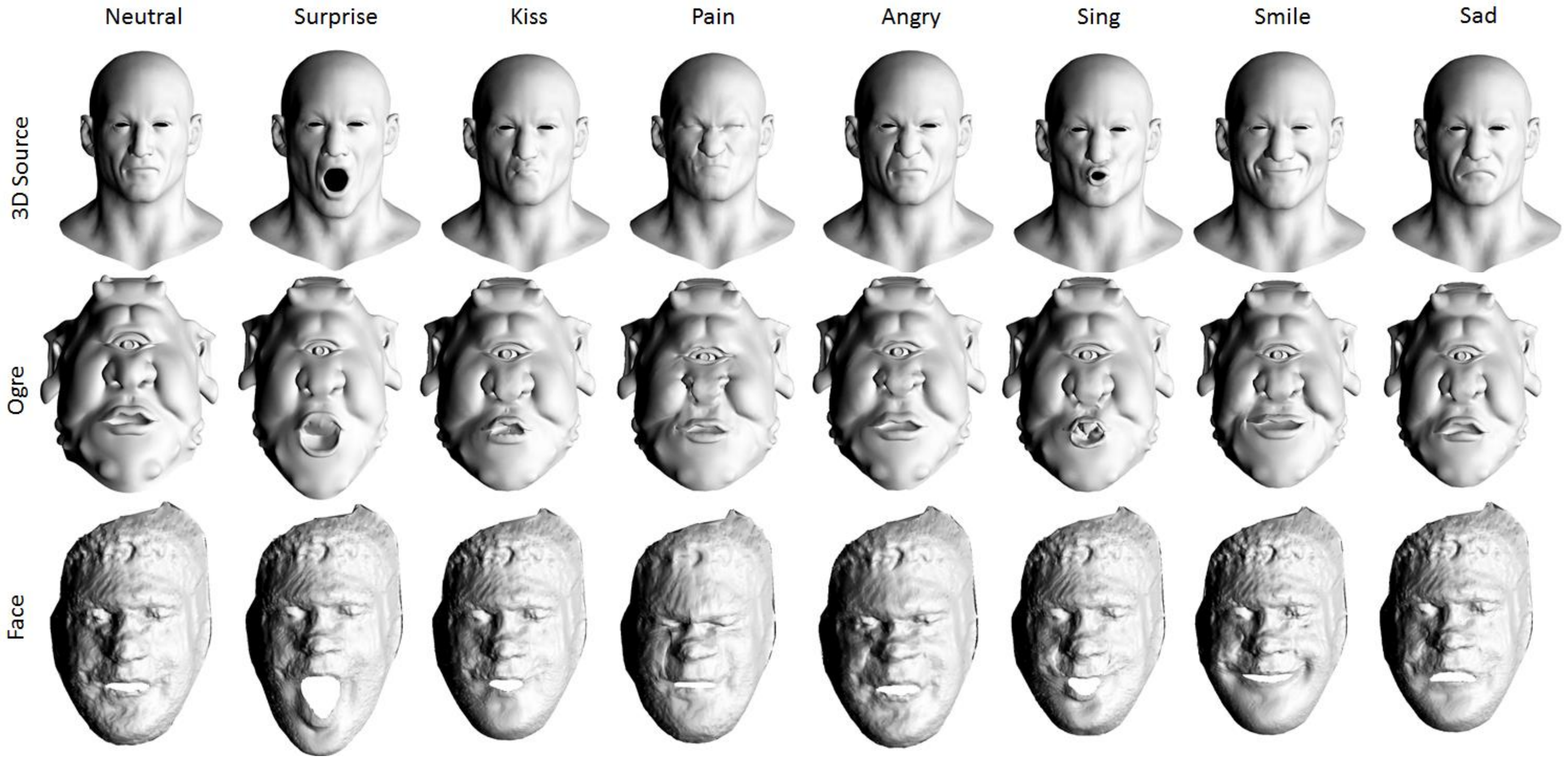
$$E = \sum_{l=1}^{L} \frac{1}{\omega_l \omega_d} \sum_{i \in V} \sum_{j \in N_i} \left\| s(\alpha)^i - s(\alpha)^j \right\|_2$$

RESULTS

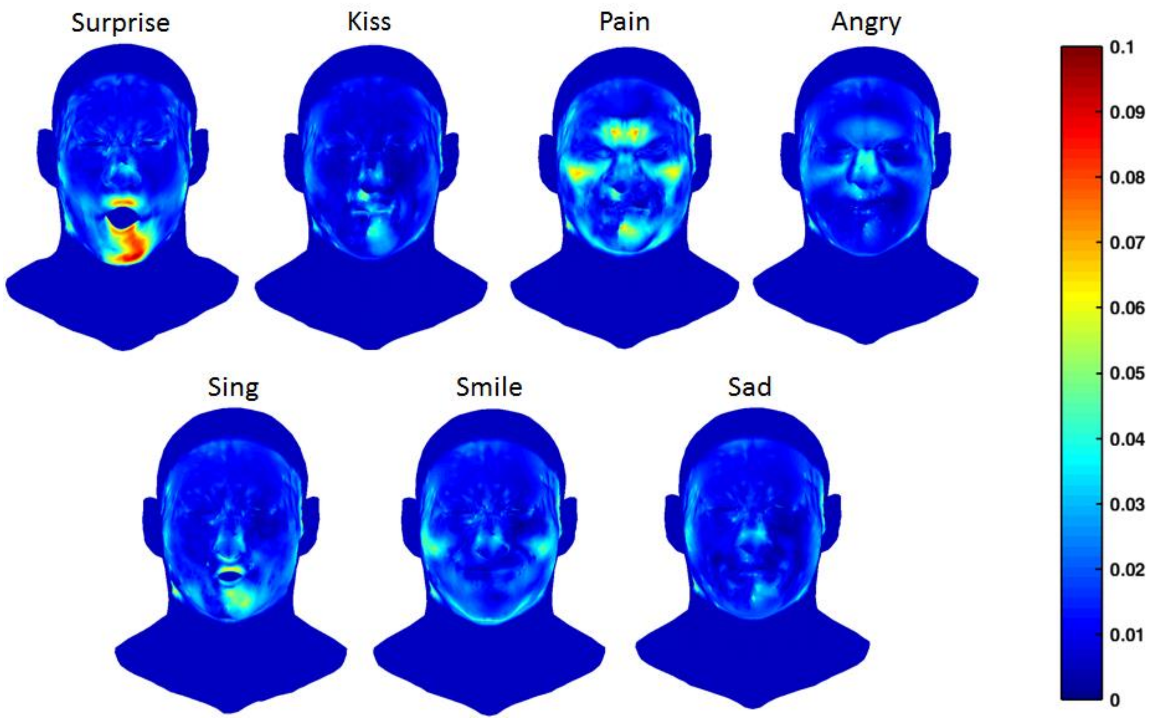


Neutral
Surprise
Kiss
Pain
Angry
Sing
Smile
Sad
3D Source
Seq3
Mocap

RESULTS

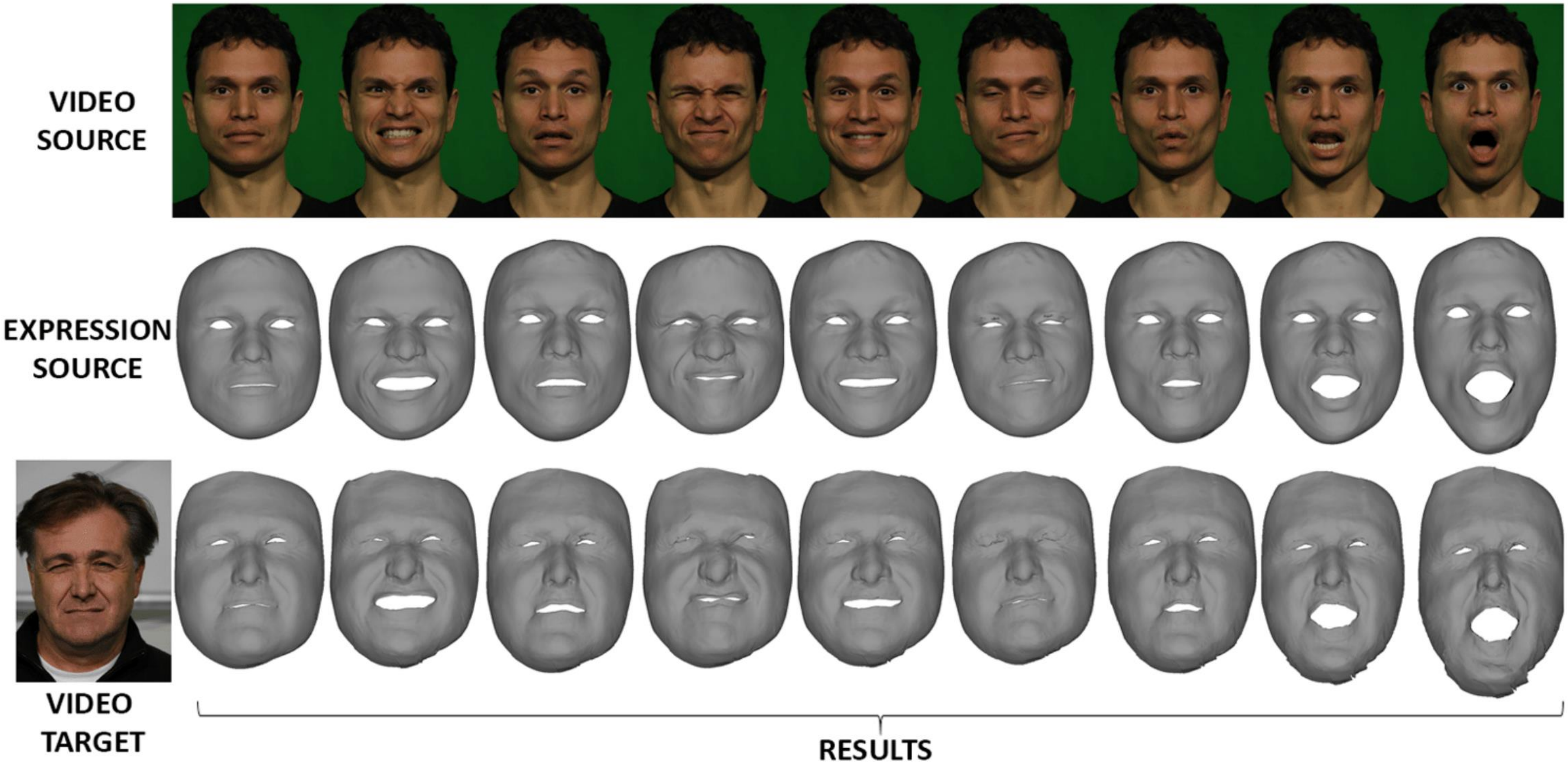

Neutral
Surprise
Kiss
Pain
Angry
Sing
Smile
Sad
3D Source
Ogre
Face

RESULTS

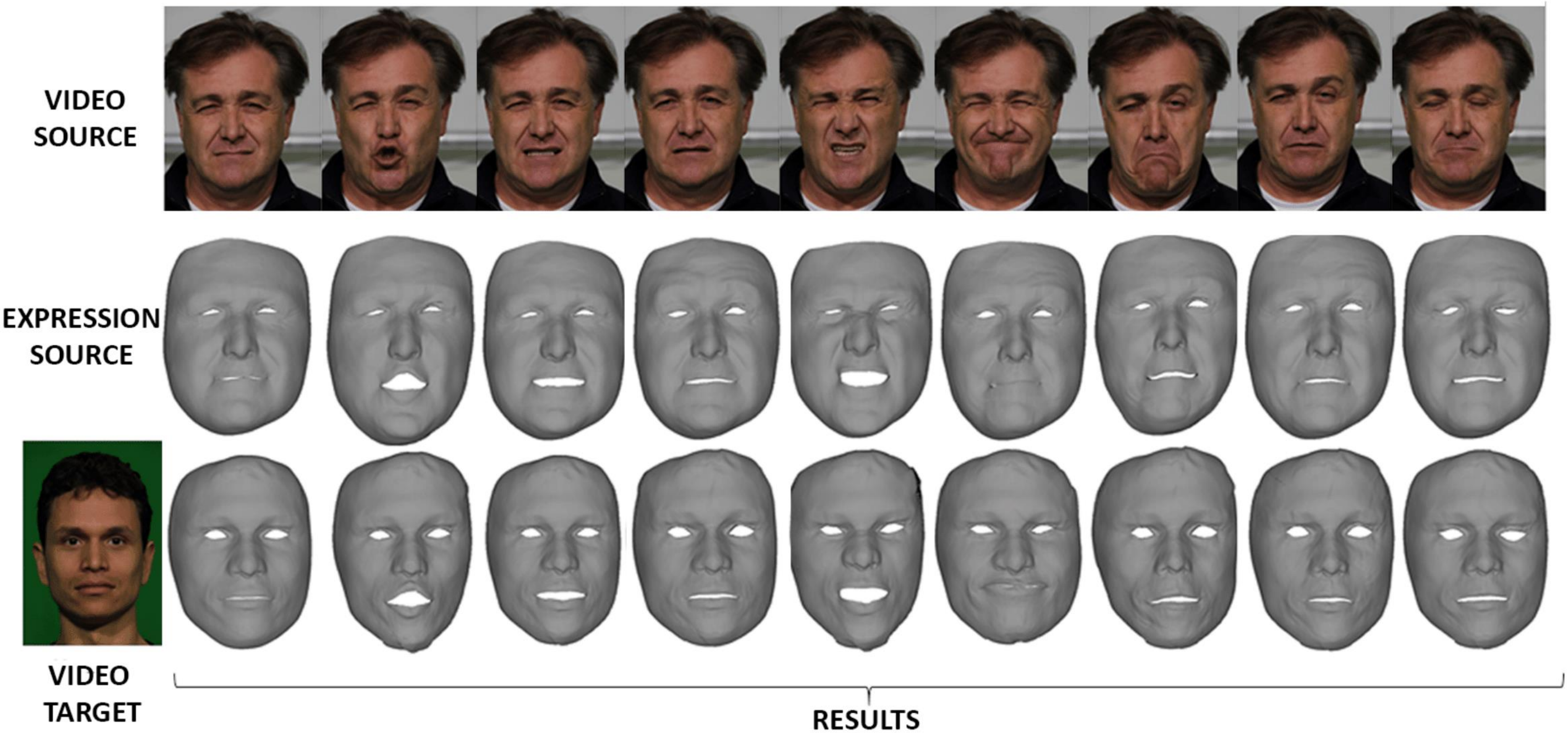


Dataset	Surprise	Kiss	Pain	Angry	Sing	Smile	Sad	$\bar{e}_{3D}$
Seq3 [26]	7.81	2.50	6.32	1.21	3.82	3.28	1.76	3.81
Mocap [21]	6.30	2.53	3.79	1.43	2.80	2.46	1.42	2.96
Ogre [27]	6.55	2.93	3.82	1.20	5.30	2.46	1.27	3.36
Face [26]	7.01	4.56	5.32	4.42	4.92	5.03	4.40	5.09
$\bar{e}_{3D}$	6.92	3.13	4.81	2.07	4.21	3.31	2.21	3.81

RESULTS

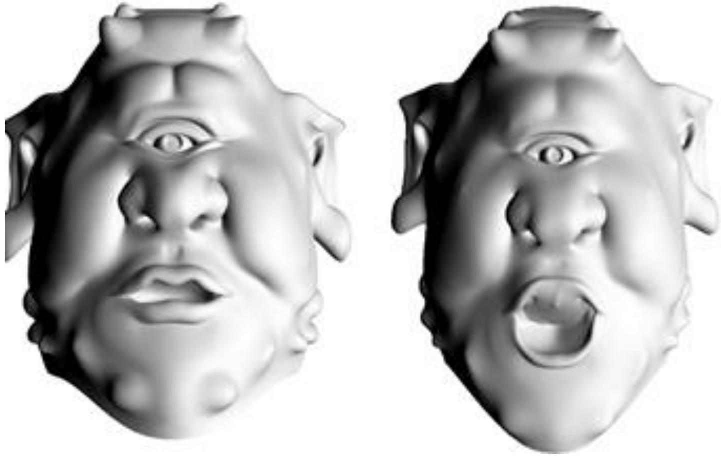


RESULTS



CONCLUSIONS

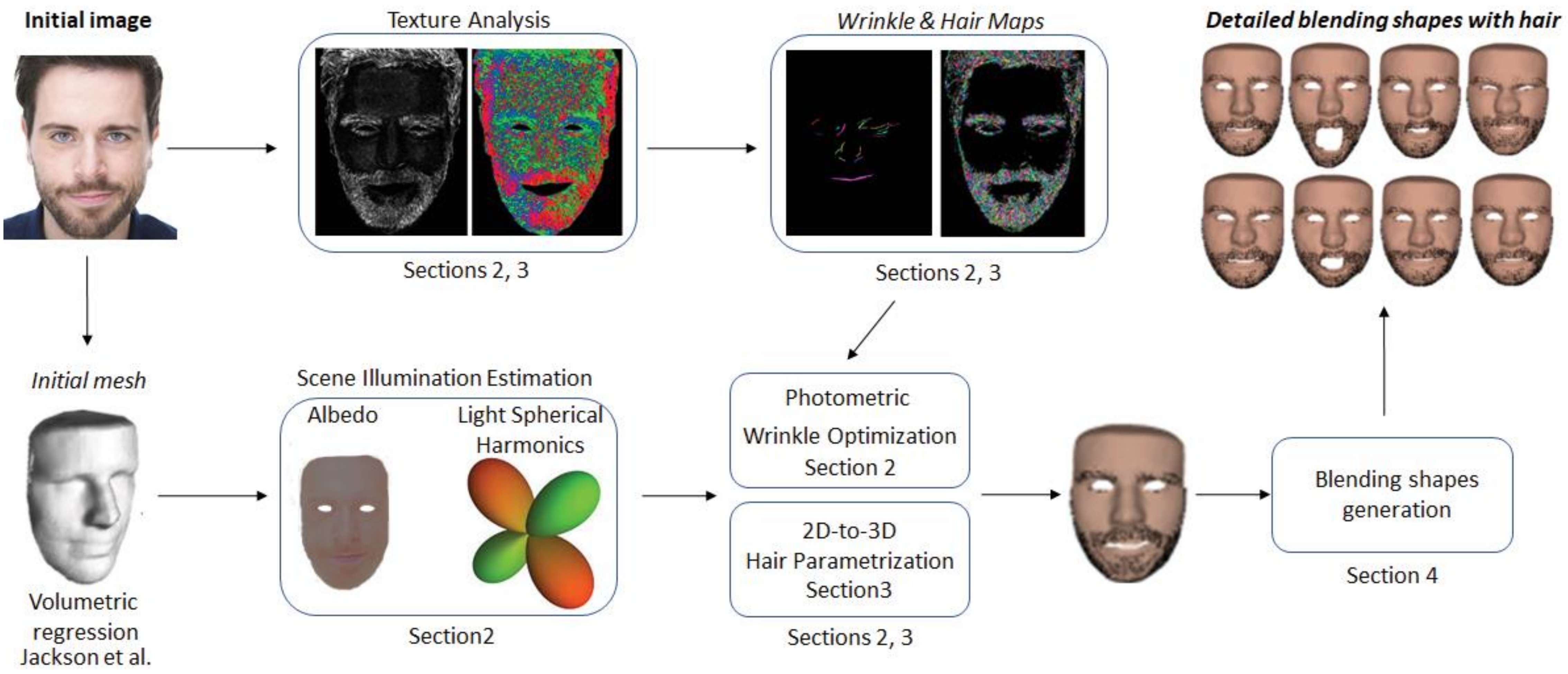
1. 2D-to-3D facial expression transfer
2. Unsupervised approach, robust, efficient, and dense
3. It can handle topology changes, different mesh resolution, and noisy data
4. No training data is required and it has no unreachable expression



4

THE COMPLETE PIPELINE

INTRODUCTION



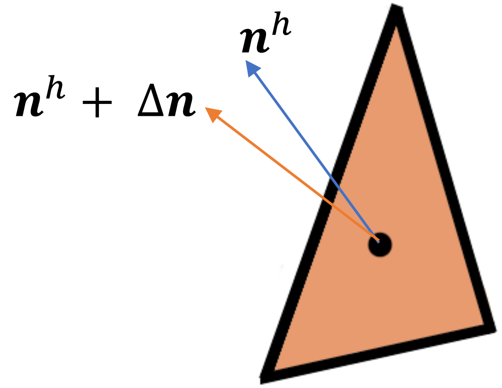
CONCERNS

1 Wrinkle preservation on expression and appending of further expression wrinkles

→ *Solved on Expression Transfer Chapter*

2 Hair recovery and the effect of orientation on details

→ *Link hair root to corresponding fine-triangle*



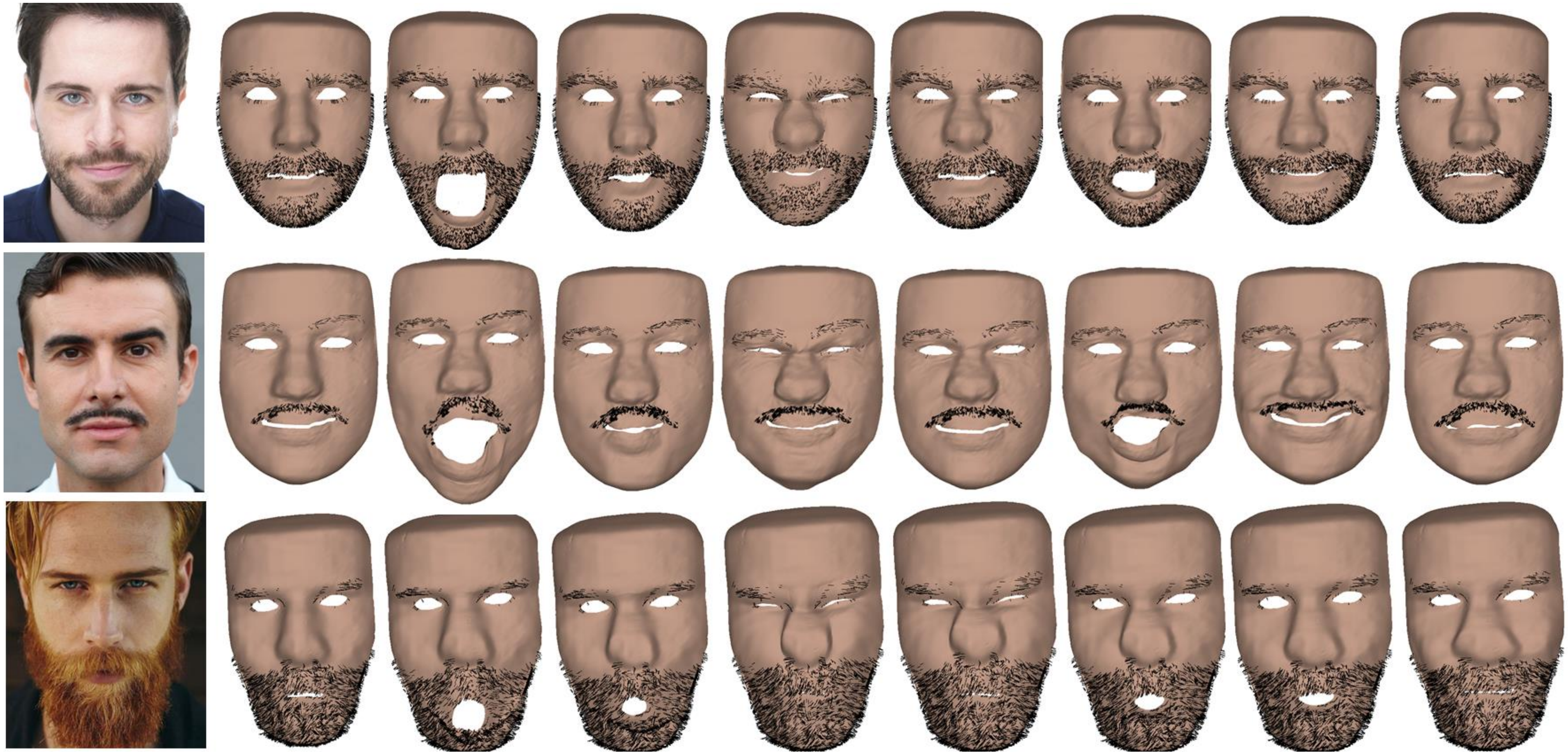
3 Hair animation on expression

→ *Extract displacement vector and orientation variation from the mesh*

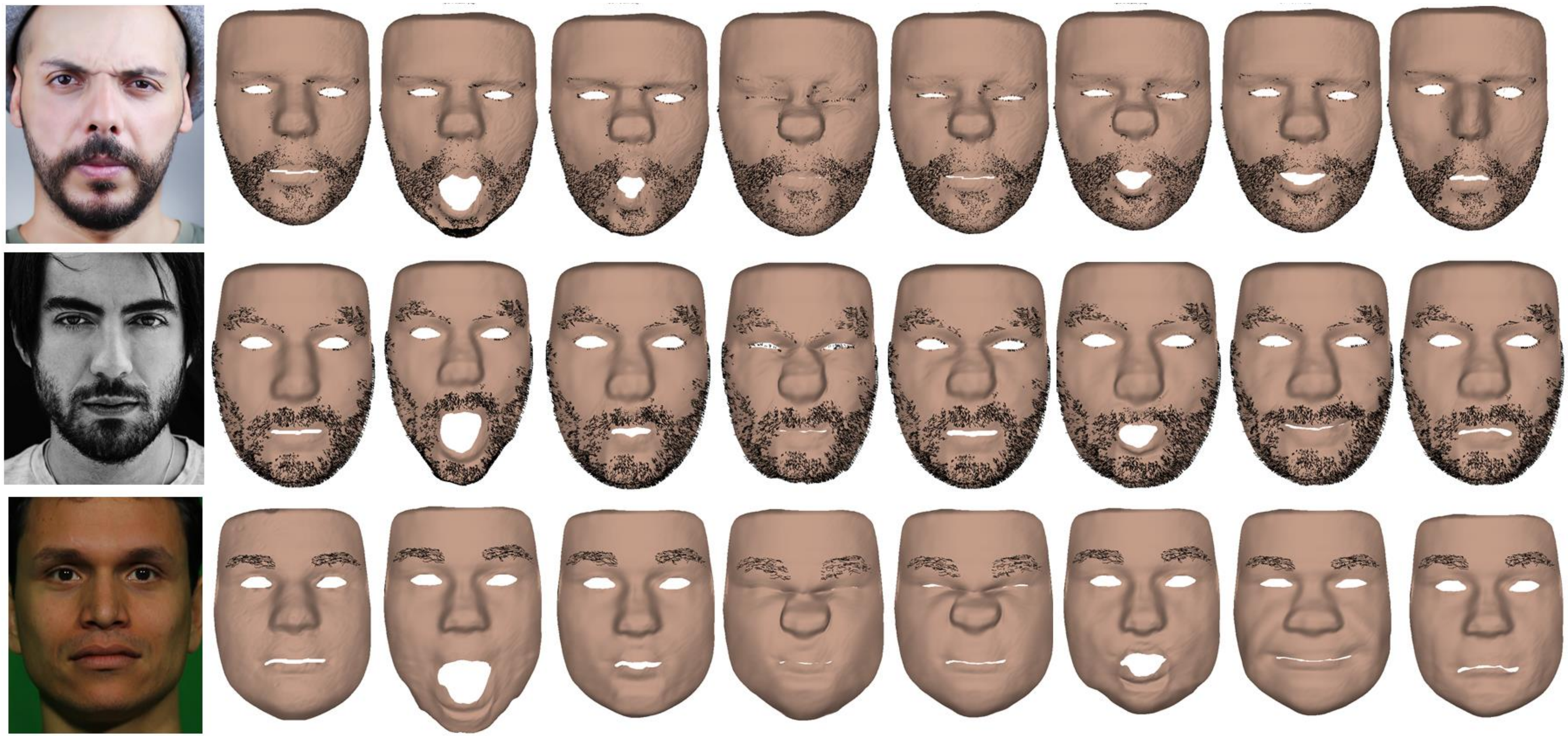
$$\boldsymbol{F}^h(\boldsymbol{p}^h + \Delta\boldsymbol{p}, \boldsymbol{n}^h + \Delta\boldsymbol{n}, l, w, r, \theta, g),$$

Synthetic Experiment

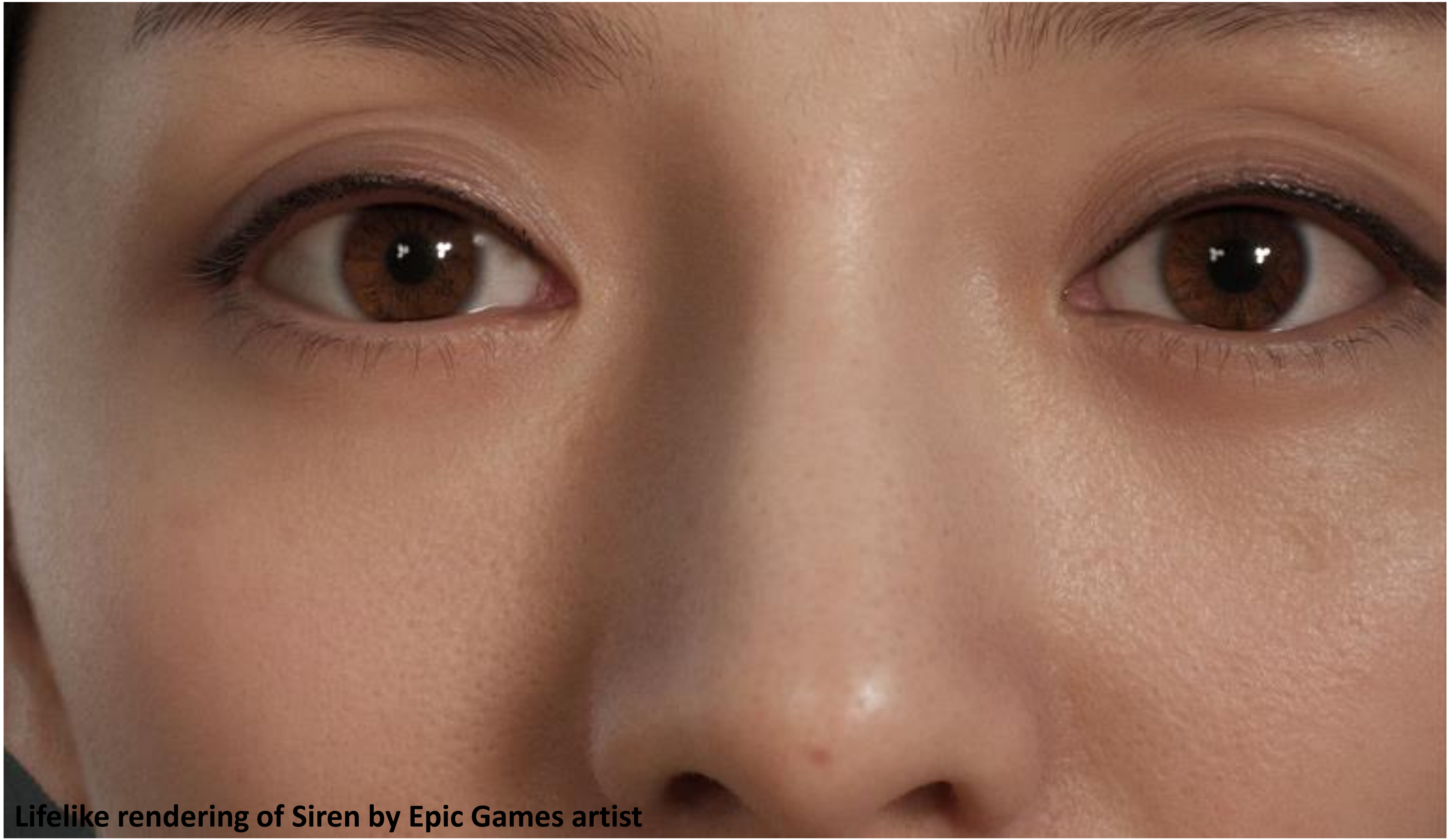
RESULTS



RESULTS



Results on Real Data



Lifelike rendering of Siren by Epic Games artist

CONCLUSIONS

1. Reconstruct detailed faces from a single RGB image (including **WRINKLES** and **FACIAL HAIR**)

2. Retrieve **person-specific** details (scars, aging, characteristic facial hair, etc.)

3. **Any data is required** nor a priori information

4. Efficient, fast, and easy to use/implement (works with **few parameters**)

CONCLUSIONS

5 **Unsupervised 2D-to-3D** facial expression transfer method

6 It can handle **topology changes**, different mesh resolution, and noisy data

7 It keeps **original detail** and animates **facial hair fibers**

8 **Any training data** is required and it has **any unreachable expression**

FUTURE PERSPECTIVE

1. Generate synthetic faces through reconstructing faces from the wild (combine detail, facial hair, and expressions)
2. Achieve better representations of facial detail holding the speed and time budget performance (higher order + skin texture)
3. Naturalness of hair density and crossings with credible representations. Especially on eyebrows and eyelashes
4. Build a transferable set of micro expressions centered on blinking wrinkles and eyebrow hair animation

THANK YOU