Computer Vision for Computer Graphics– SS13

Shape capture: Facial performance capture

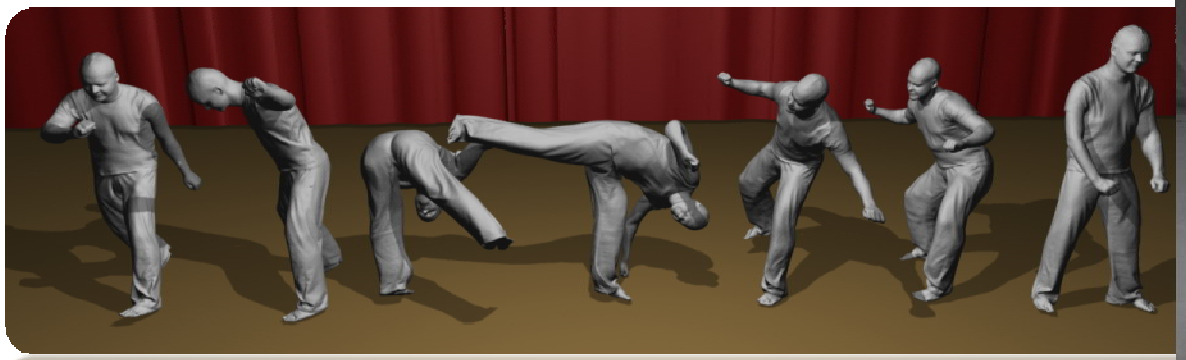
Darya Dedik
Supervisor: Pablo Garrido

June 25, 2013

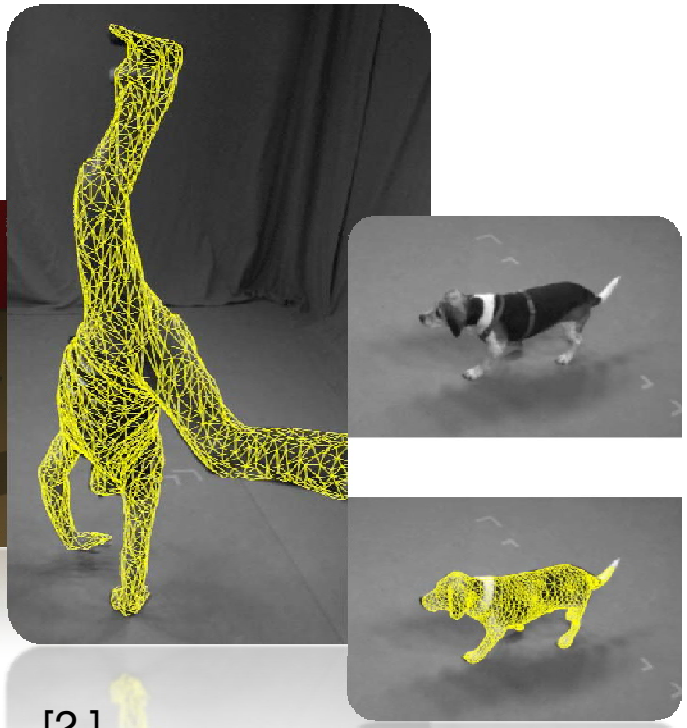
Performance Capture

[1] Performance Capture from Sparse Multi-View Video

[2] Motion Capture Using Joint Skeleton Tracking and Surface Estimation



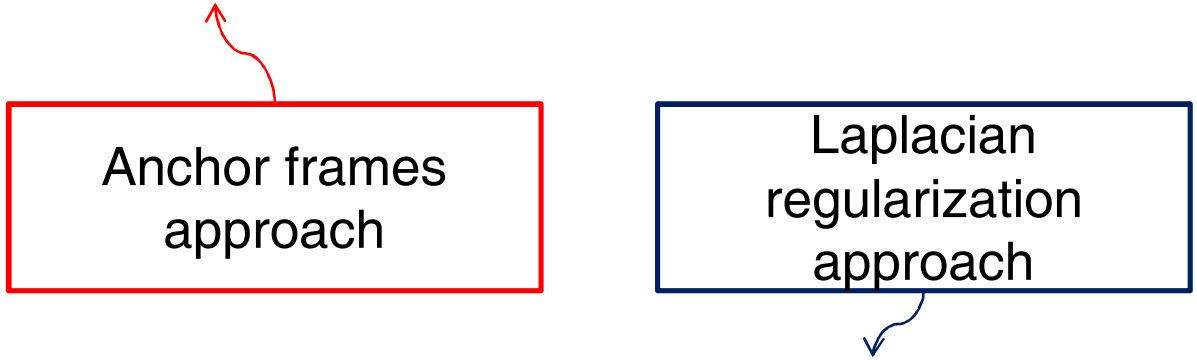
[1]



[2]

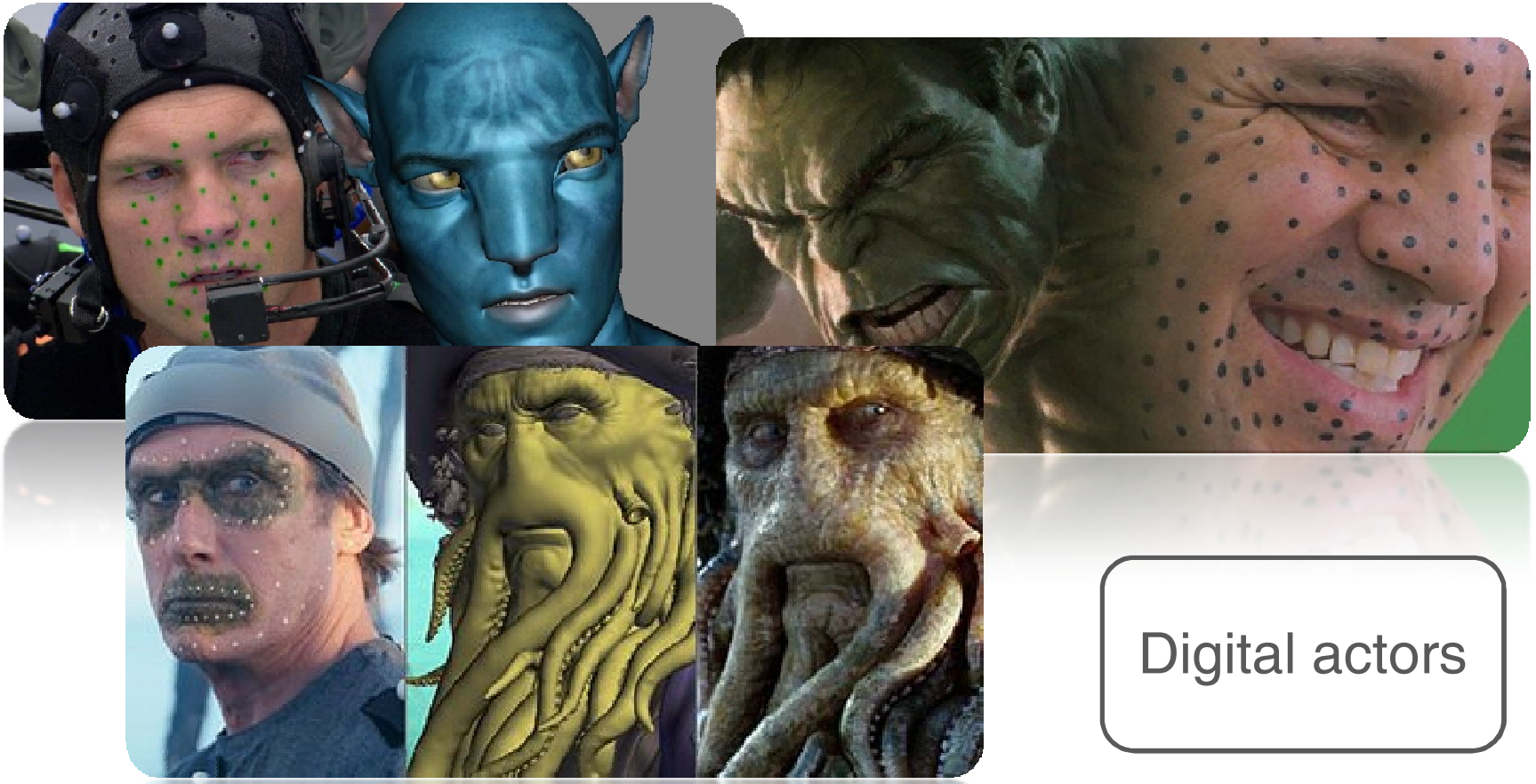
Outline

1. **High-Quality Passive Facial Performance Capture using Anchor Frames**. T. Beeler, F. Hahn, D. Bradley, B. Bickel, P. Beardsley, C. Gotsman, M. Gross



2. **Lightweight Binocular Facial Performance Capture under Uncontrolled Lighting**. Levi Valgaerts, Chenglei Wu, Andrés Bruhn, Hans-Peter Seidel, Christian Theobalt

Motivation



Digital actors

[20th Century Fox, Industrial Light and Magic (ILM)]

Two Approaches

- Goal:
  - reconstruct high-quality highly-detailed face geometry
  - accurately track motion, expression of the face

- Both strategies:
  - track motion of the face in time
  - use different strategies: anchoring and Laplacian regularization in terms of drift regularization

Outline

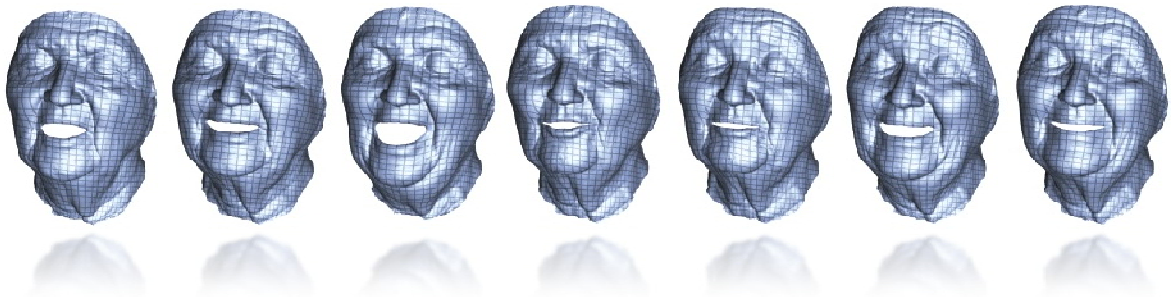
1. **High-Quality Passive Facial Performance Capture using Anchor Frames**. T. Beeler, F. Hahn, D. Bradley, B. Bickel, P. Beardsley, C. Gotsman, M. Gross

2. **Lightweight Binocular Facial Performance Capture under Uncontrolled Lighting**. Levi Valgaerts, Chenglei Wu, Andrés Bruhn, Hans-Peter Seidel, Christian Theobalt

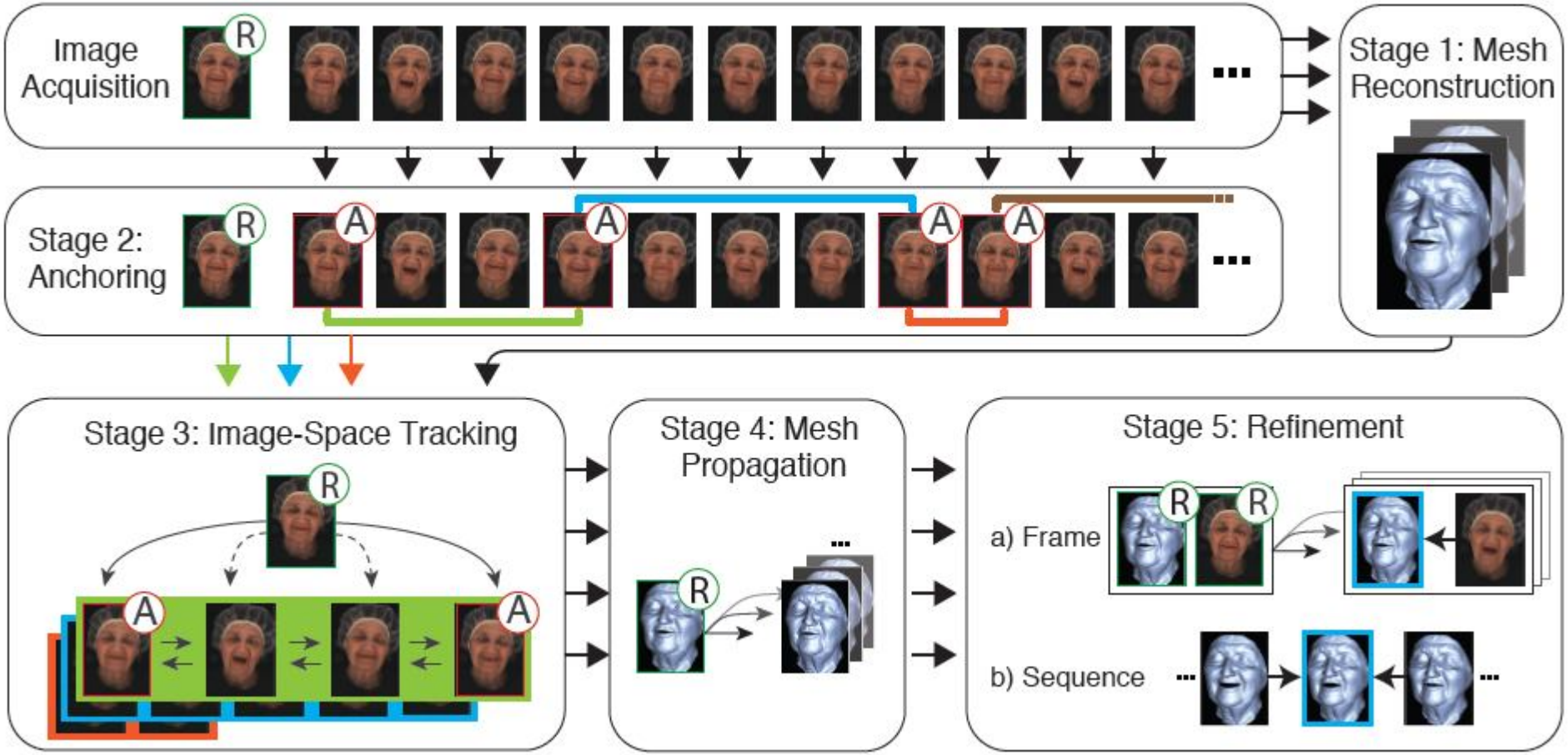
Approach

- Set-up:
  - Multiple cameras
  - Passive illumination
- Input:
  - Sequence of frames of the face

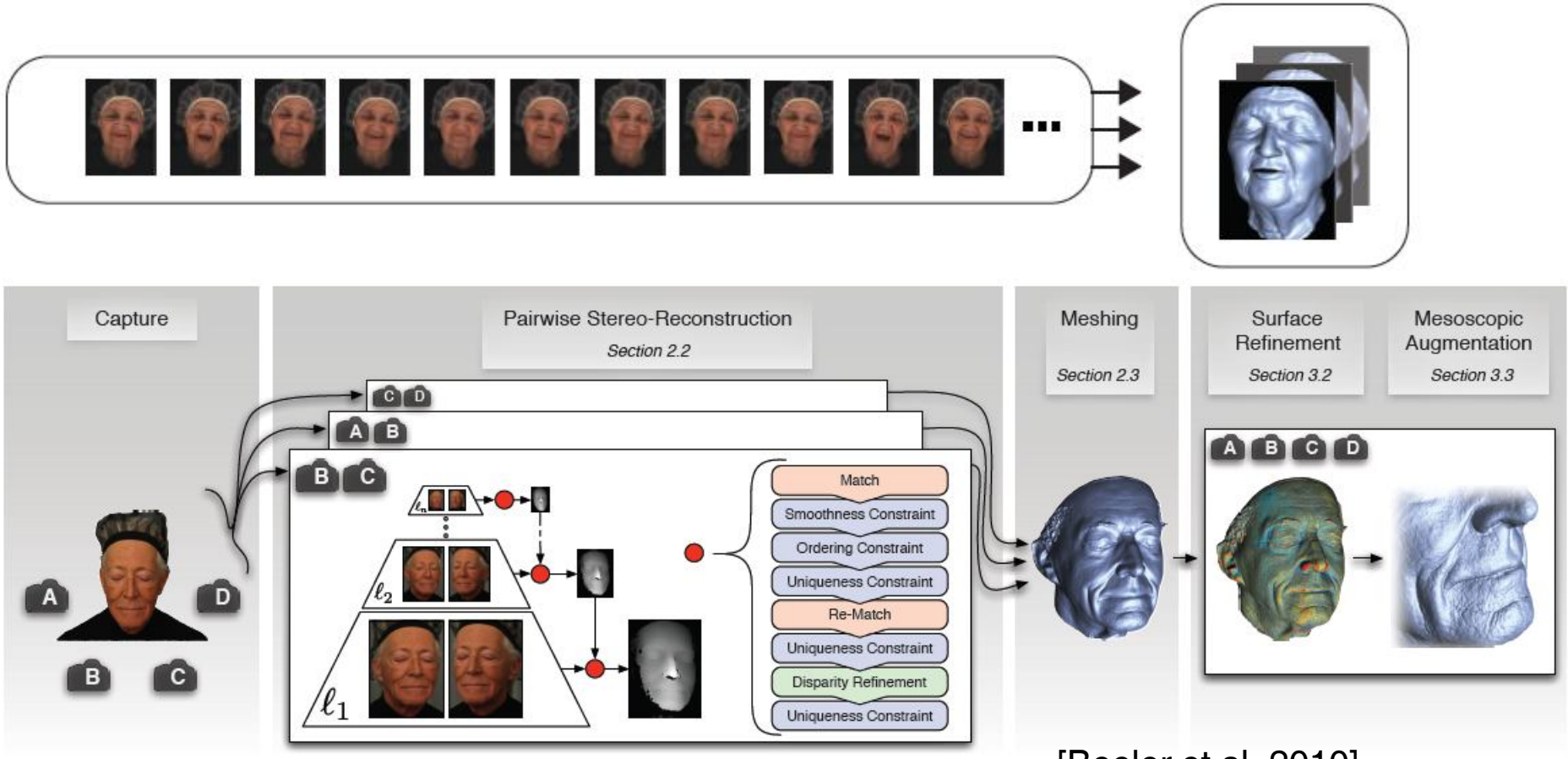
- Output:



5 Stage Method

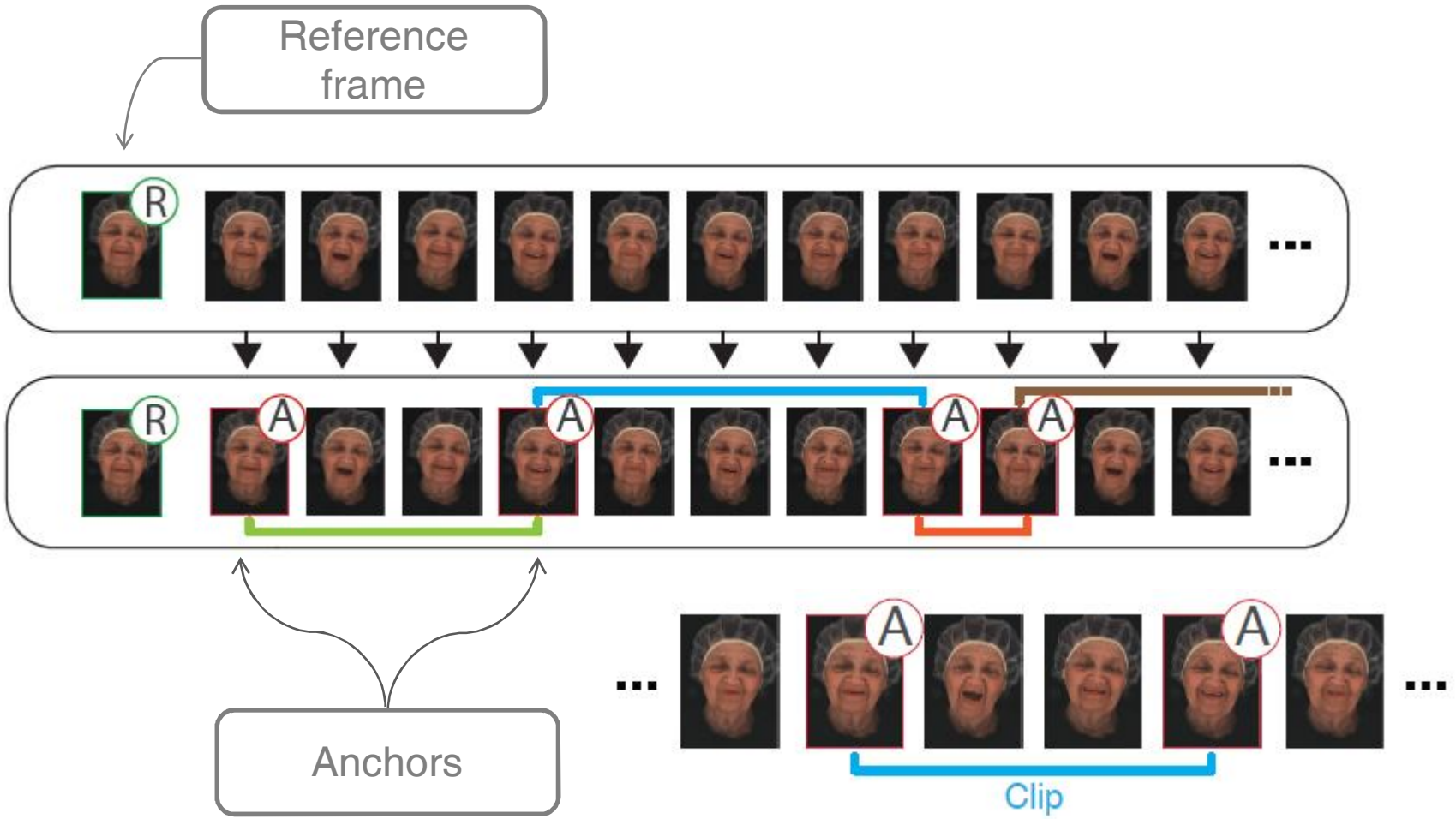


Step1: Stereo Reconstruction



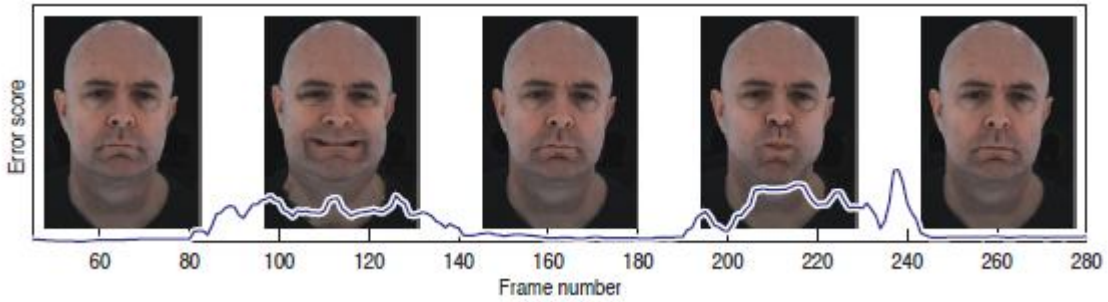
[Beeler et al, 2010]

Step 2: Anchoring

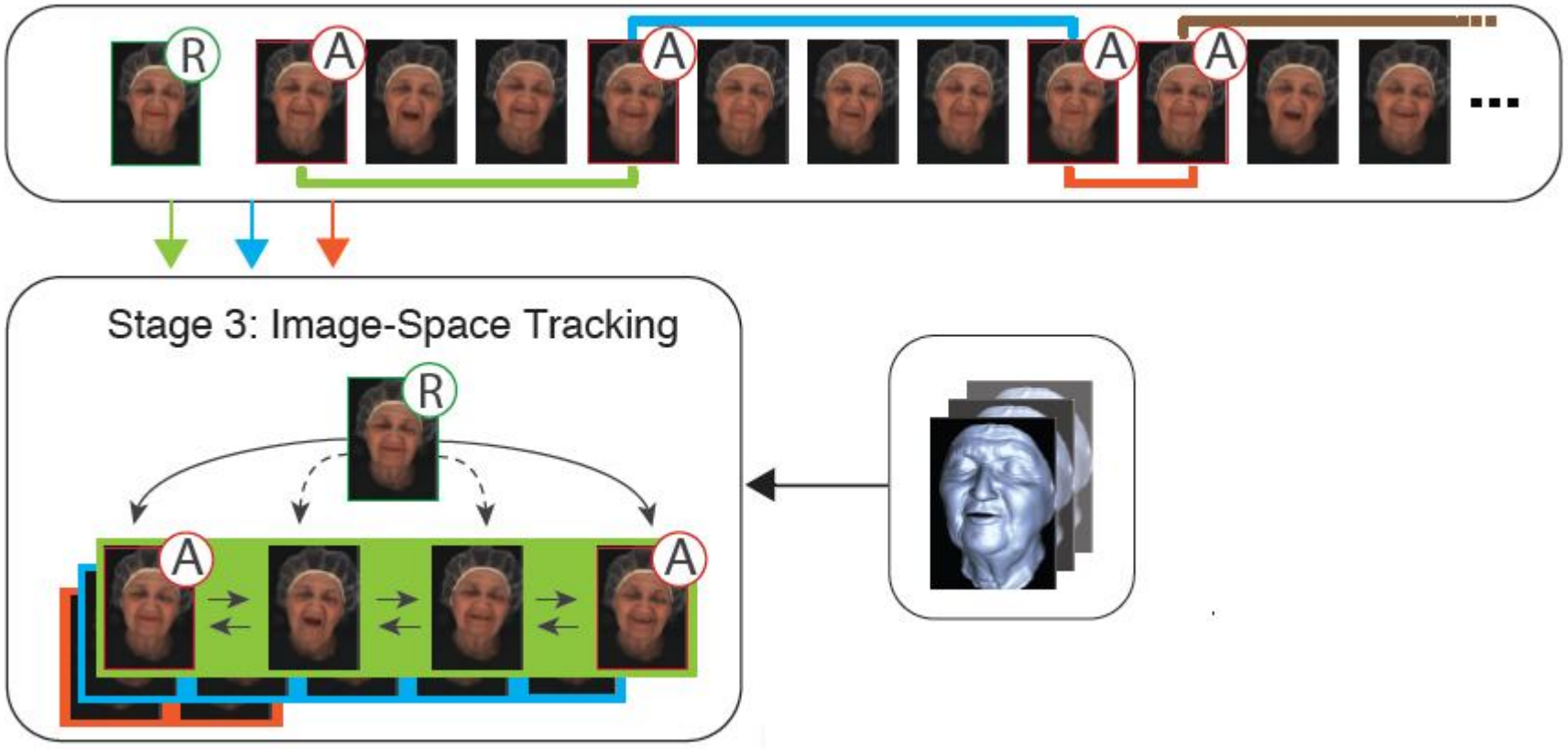


Identifying Anchor Frames

1. Feature set detection S_c in reference frame
2. Correspondence matching by normalized cross-correlation
3. Error score *E* computation, detection anchors

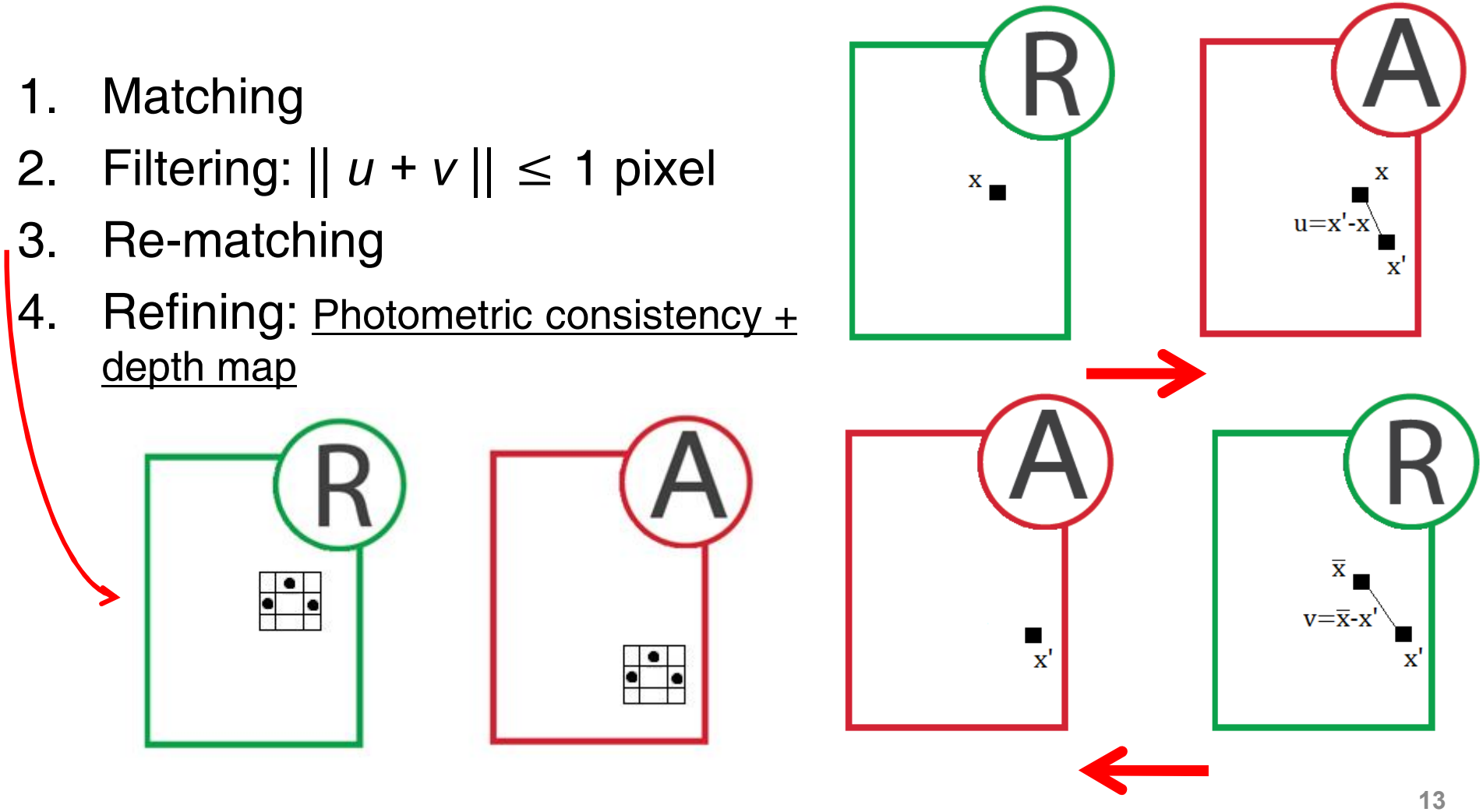


Step 3: Image-Space Tracking

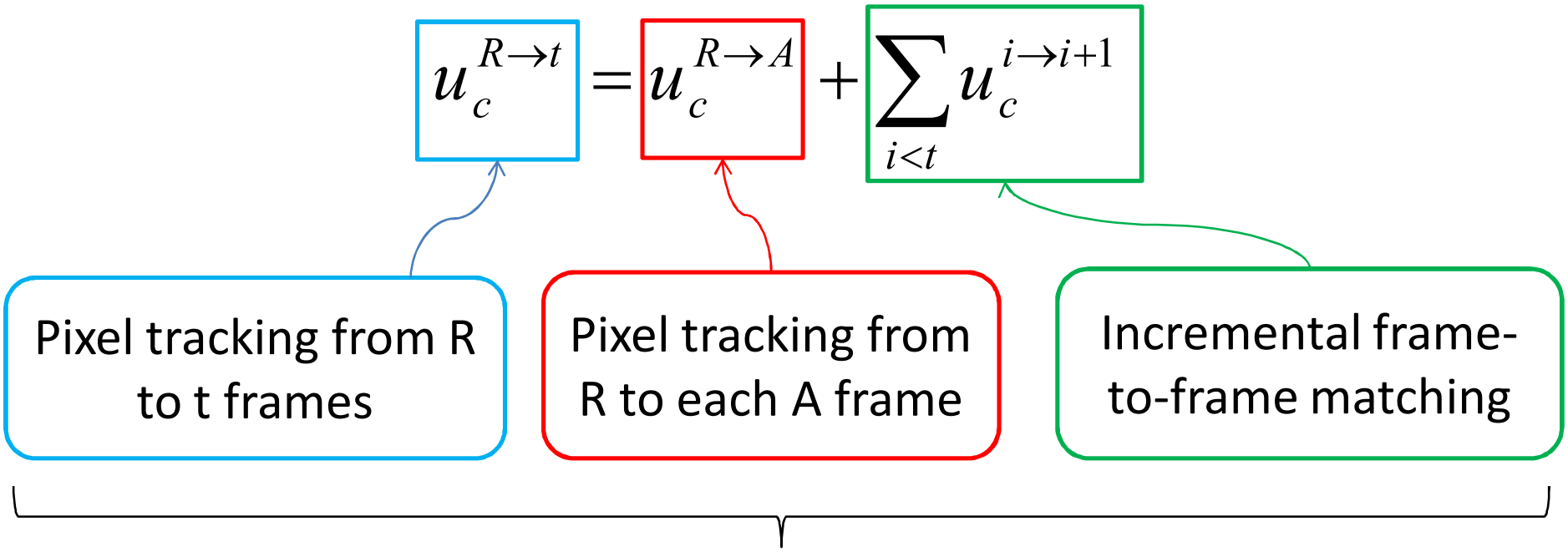


Tracking from R to A frames

1. Matching
2. Filtering: $\| u + v \| \leq 1$ pixel
3. Re-matching
4. Refining: Photometric consistency + depth map

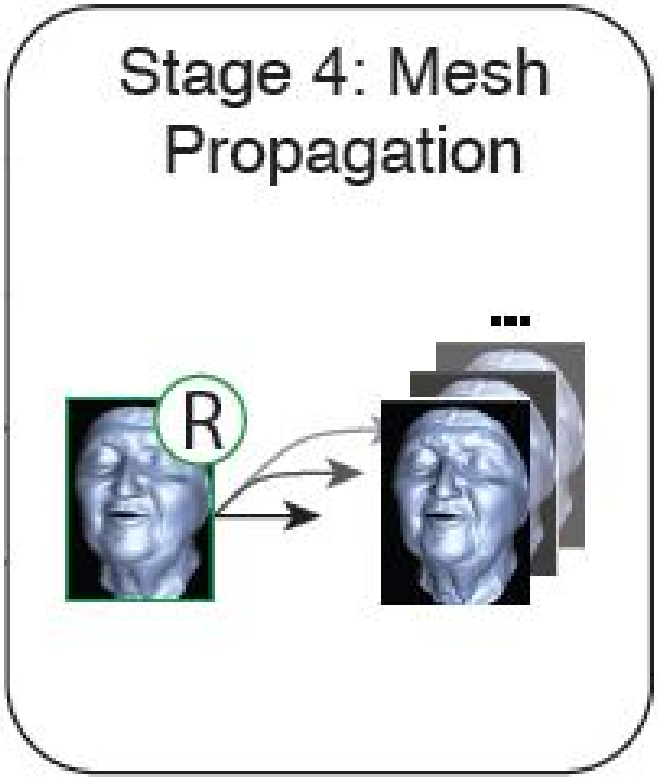


Tracking from R to other frames

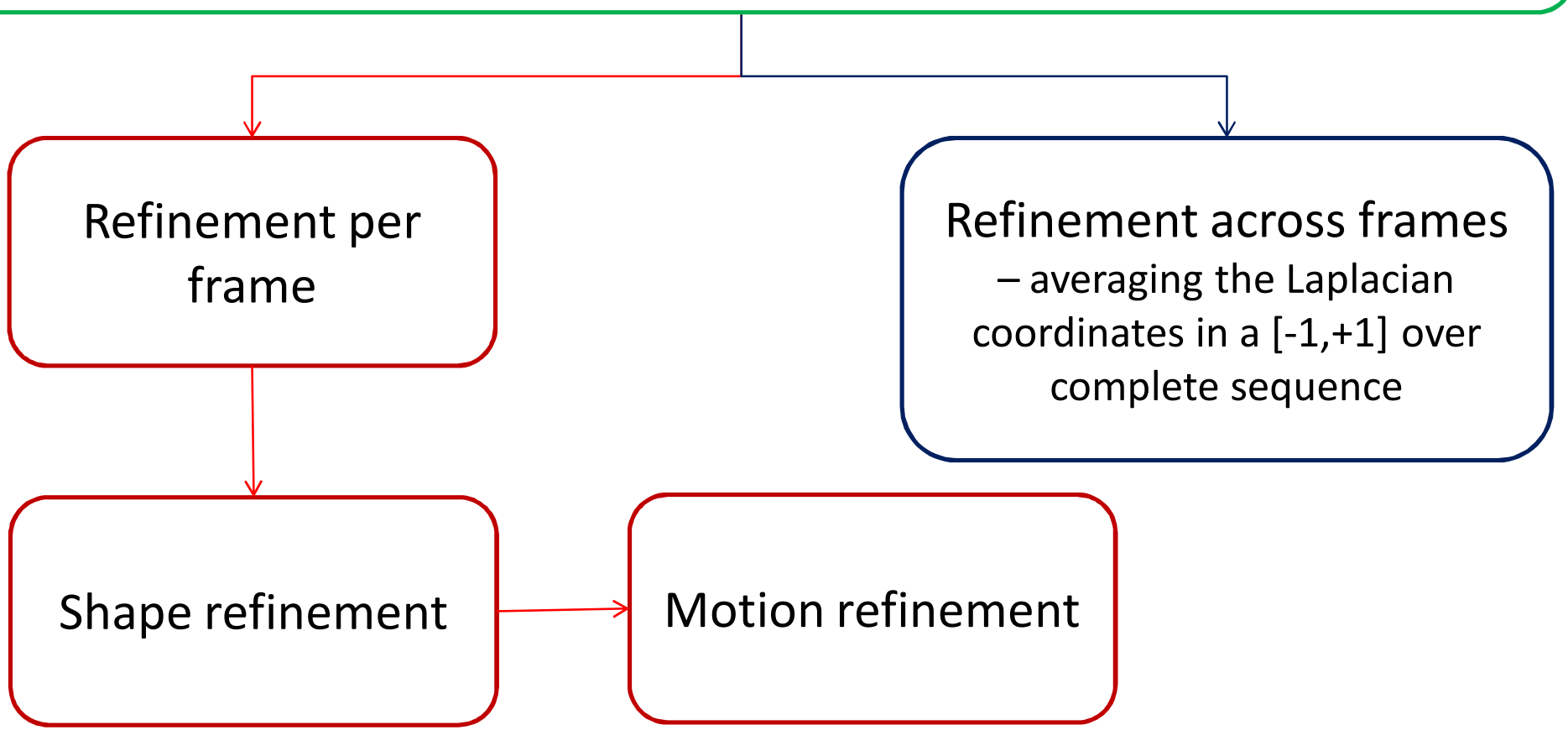


Step 4: Mesh Propagation

- Goal: to find transformed 3D position of each vertex due to the motion and deformation of the face
- Motion fields obtained in the previous step are used



Step 5: Mesh Refinement



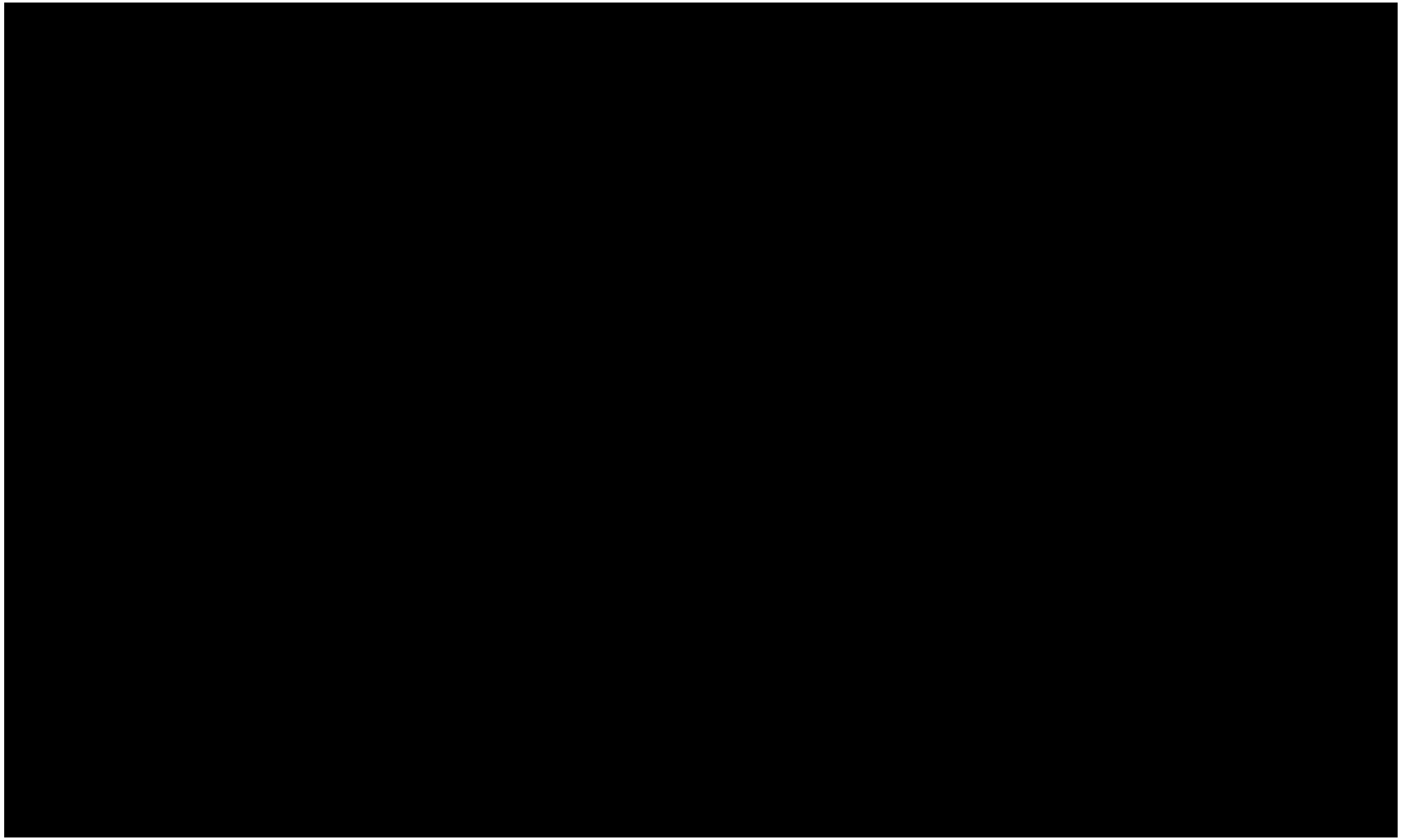
Refinement per Frame

Goal: find for each vertex the position in space that optimizes:

1. Spatial image fidelity
2. Temporal image fidelity
3. Mesh fidelity
4. Geometry smoothness

Find details in [Beeler et al, 2010]

Results



Limitations

1. Controlled lightning
2. Drift
3. Short sequences
4. Partition into clips might be impossible

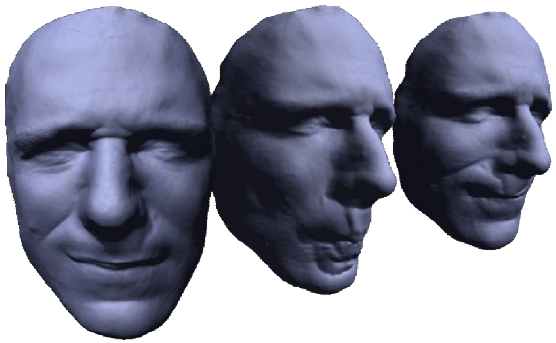
Outline

1. **High-Quality Passive Facial Performance Capture using Anchor Frames**. T. Beeler, F. Hahn, D. Bradley, B. Bickel, P. Beardsley, C. Gotsman, M. Gross

2. **Lightweight Binocular Facial Performance Capture under Uncontrolled Lighting**. Levi Valgaerts, Chenglei Wu, Andrés Bruhn, Hans-Peter Seidel, Christian Theobalt

Approach

- Set-up:
  - 2 cameras
  - Uncontrolled lightning, outdoor lightning
- Input:
  - Stereo video sequence of the face
- Output:
  - High-quality, spatio-temporally coherent face geometry over time



Input

Left stereo view

Right stereo view

Template Reconstruction

Variational stereo method to recover 2D displacement field minimizing energy of the form:

$$E = \int_{\Omega} (E_D + \alpha E_G + \beta E_S) dx$$

Data term | Geometry term | Smoothness term

Stereo Reconstruction

Smoothness

Scene Flow Estimation

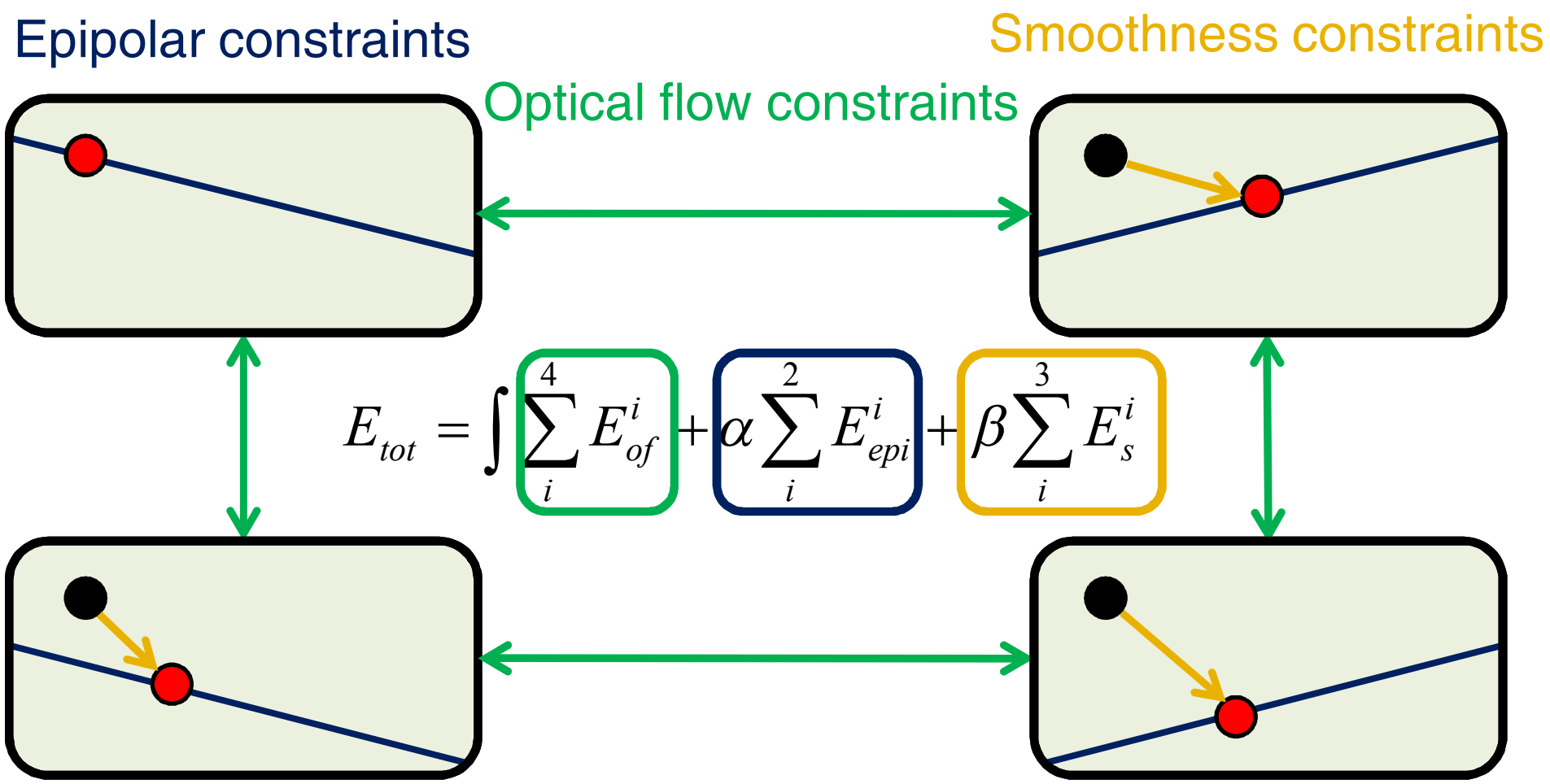
Current
left frame

Current
right frame

Next
left frame

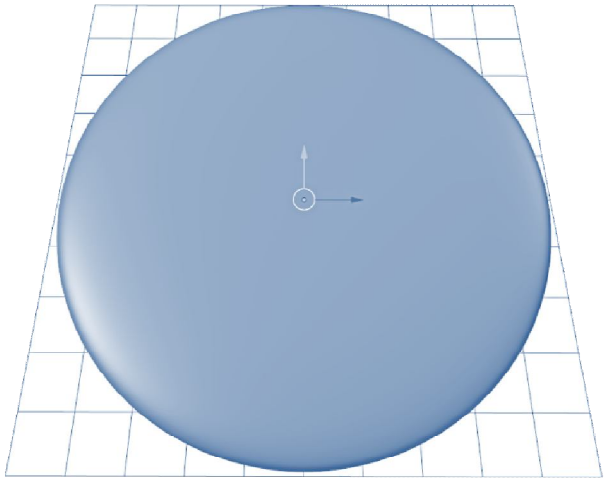
Next
right frame

Scene Flow Estimation

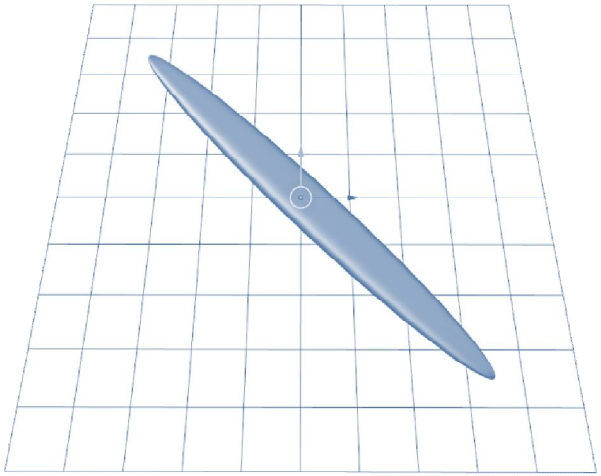


Structure-Aware Regularization

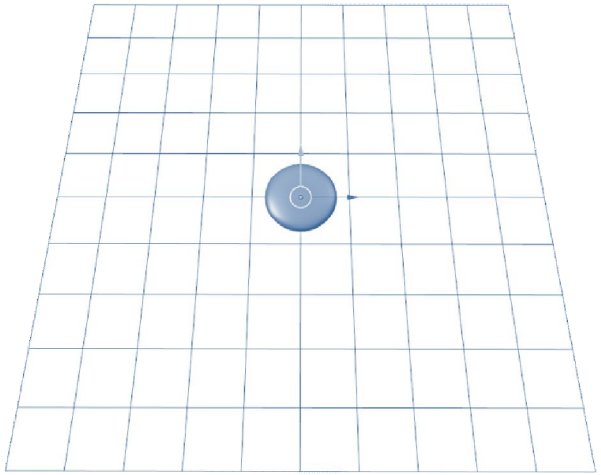
$$E_{tot} = \int \sum_{i}^{4} E_{of}^{i} + \alpha \sum_{i}^{2} E_{epi}^{i} + \beta \sum_{i}^{3} E_{s}^{i}$$



Flat regions

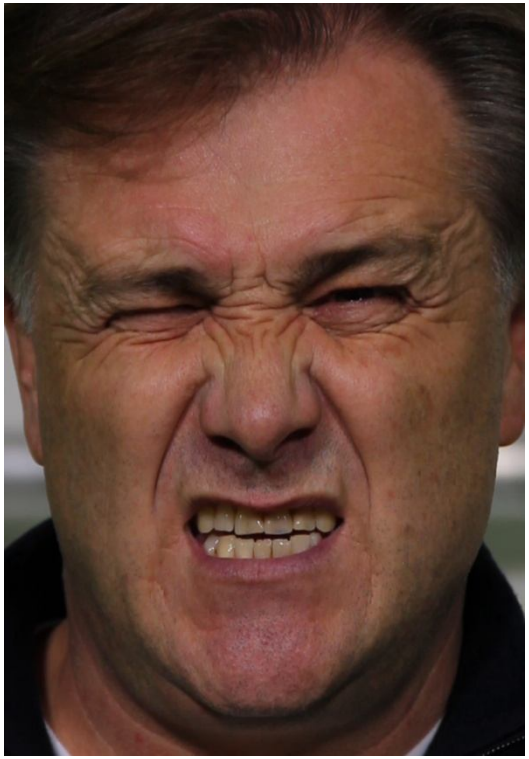


Structure edges

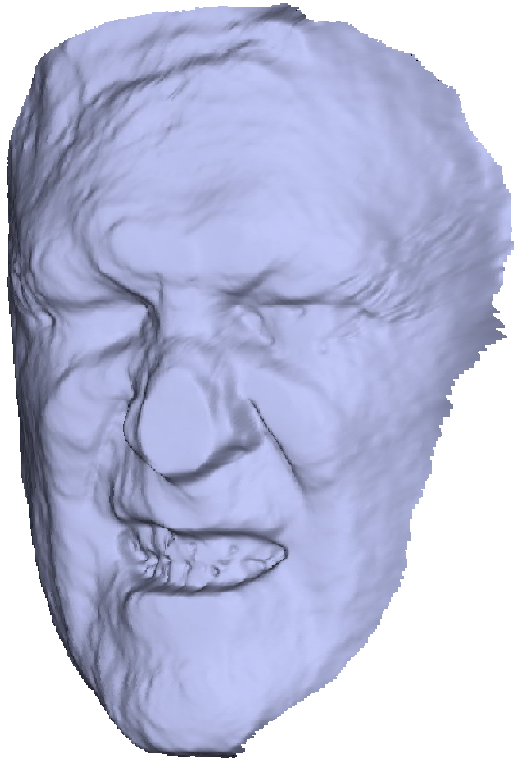


Corners

Structure-Aware Regularization result



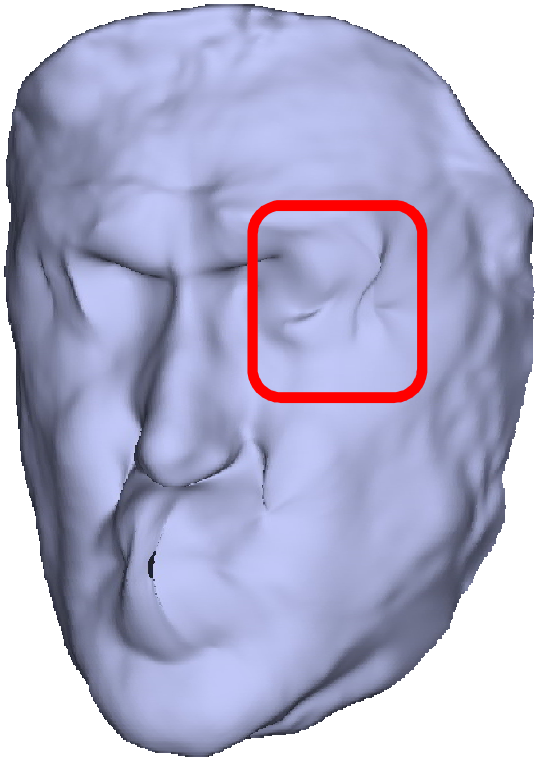
Left frame

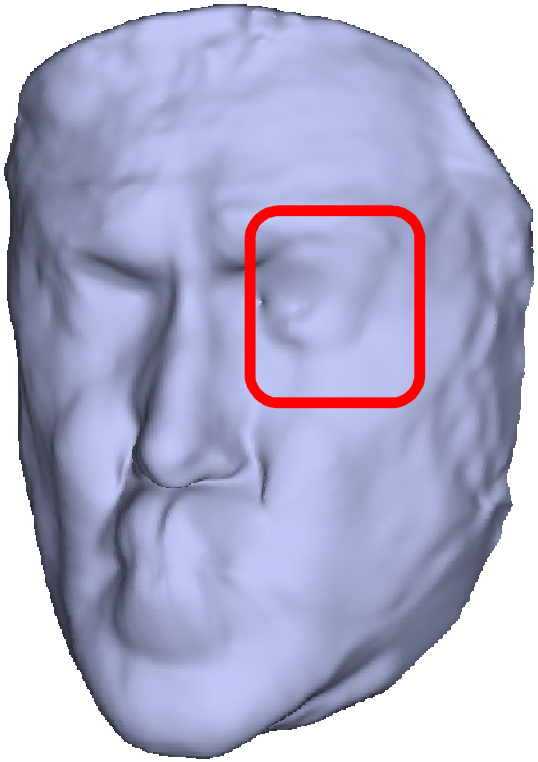


[Valgaerts et al. 2010]

Our approach

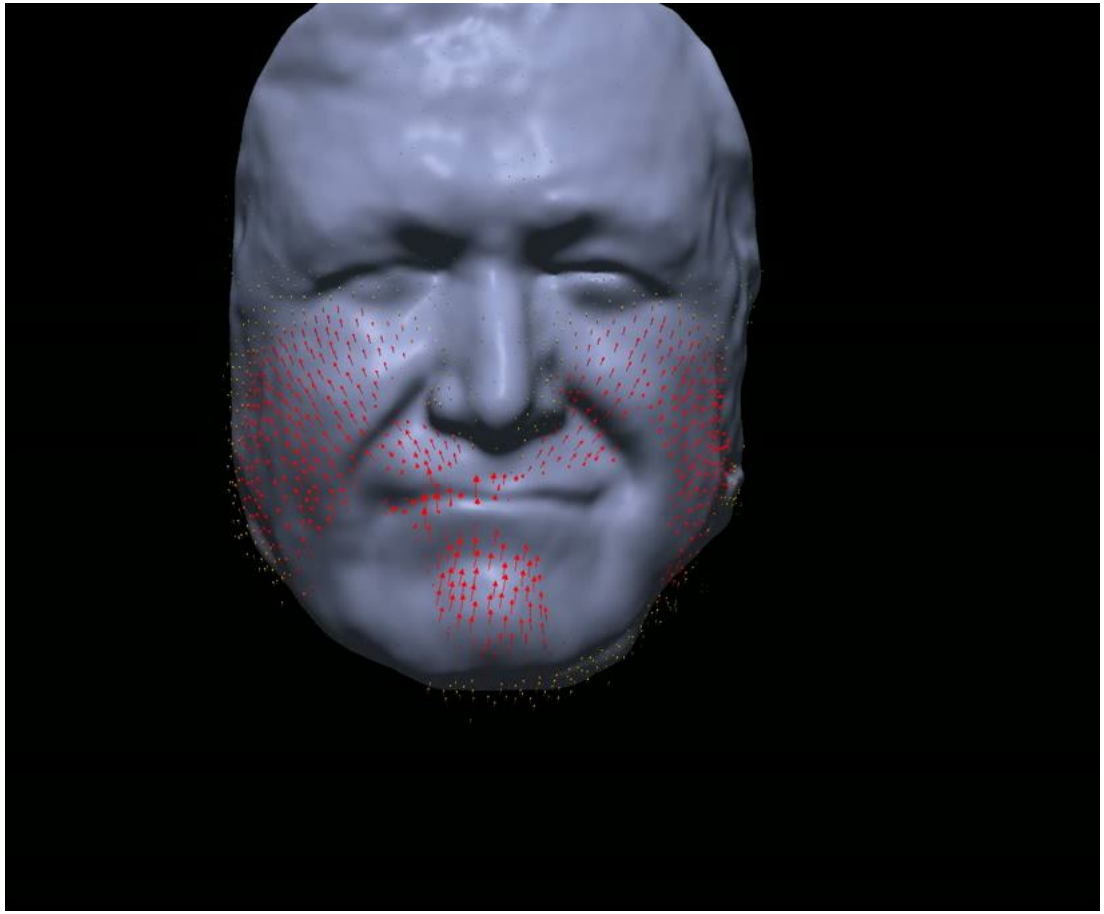
Structure-Aware Regularization result



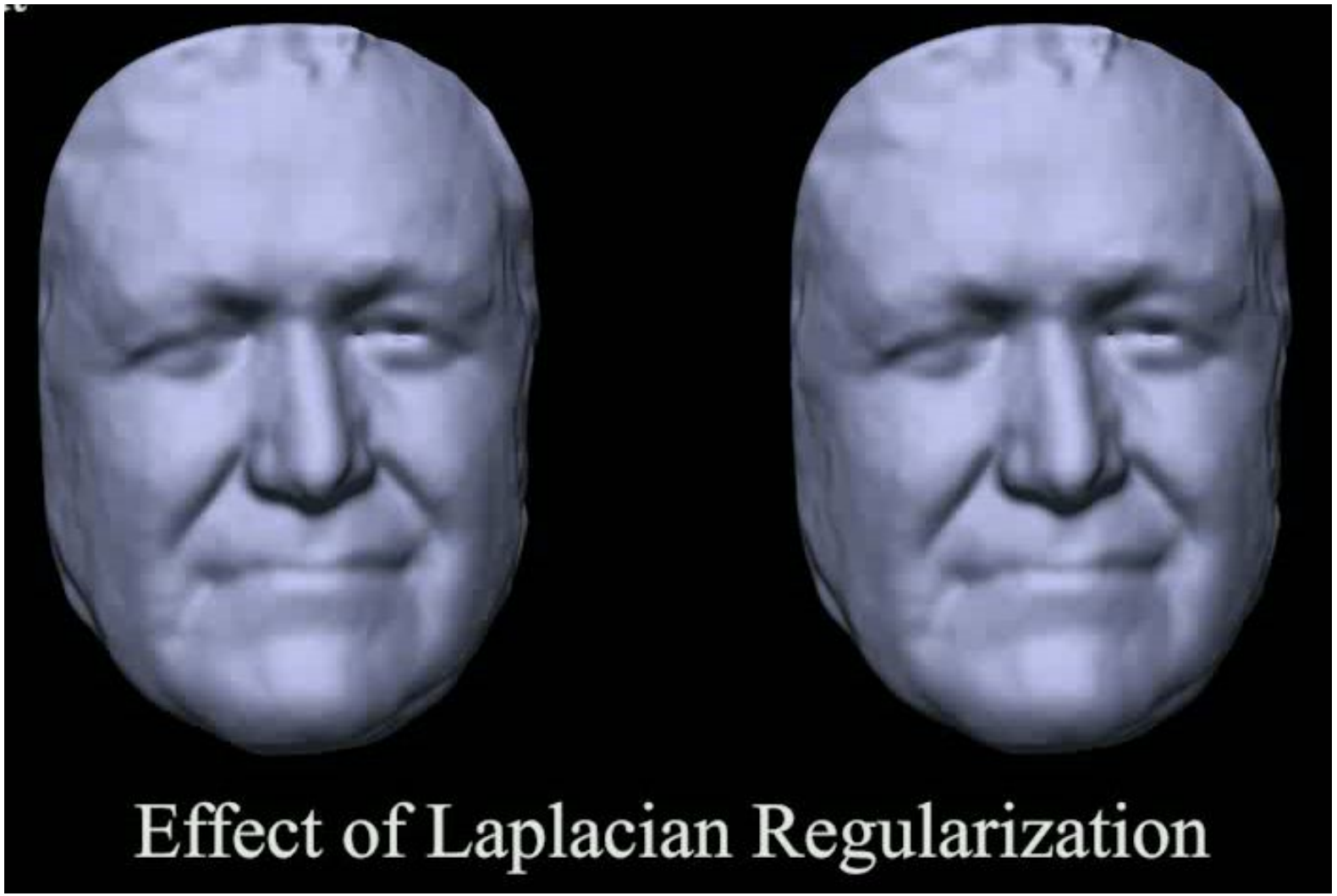


Target frame [Valgaerts et al. 2010] Our approach

Scene Flow Estimation Results



Mesh Tracking



Motion Refinement

Input image

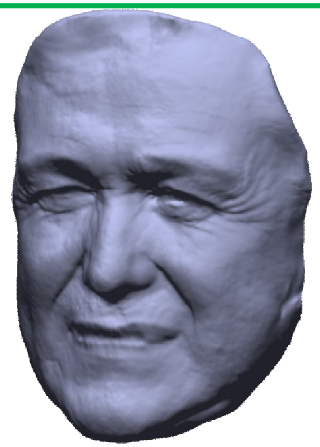
Without motion refinement

With motion refinement

Shape Refinement

Previous frame

Fine scale detail



Current frame

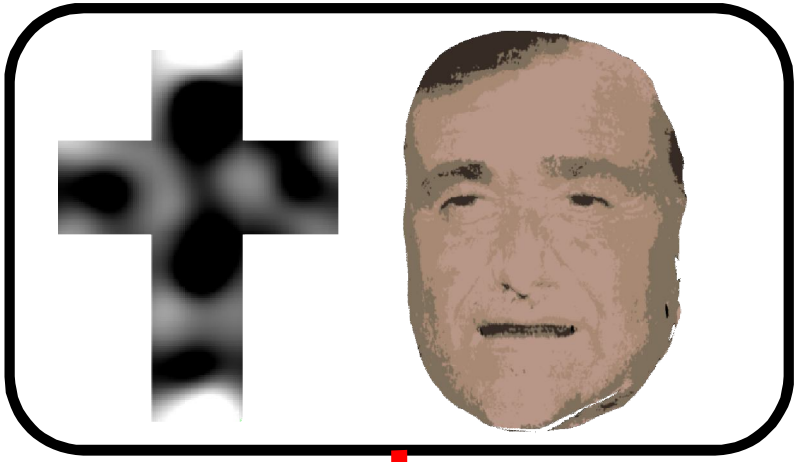
Coarse- scale geometry

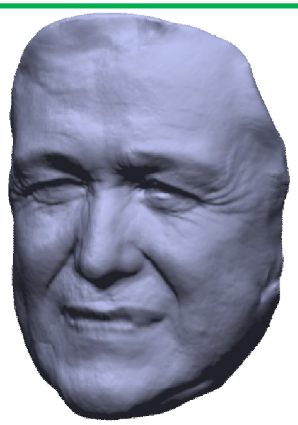
Wanted

Fine scale detail

Shape Refinement Step

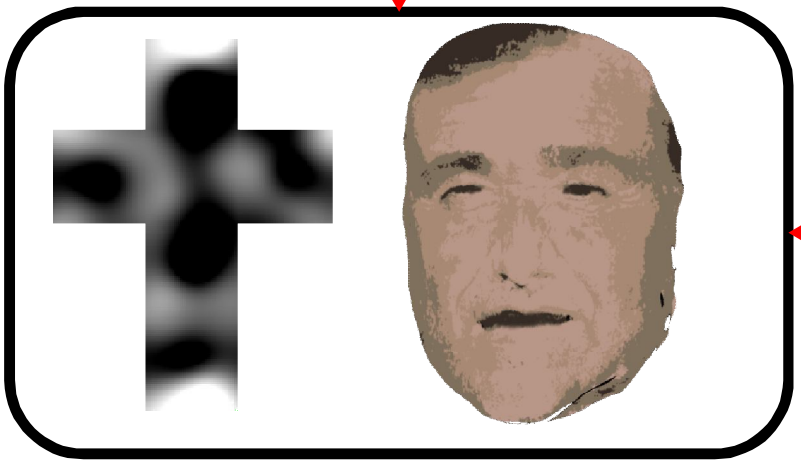
Previous frame





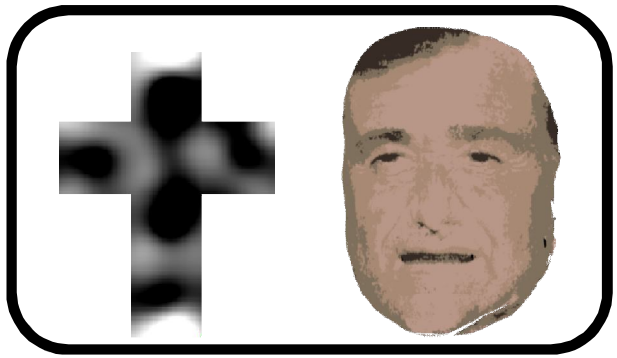
Estimate lighting

Current frame

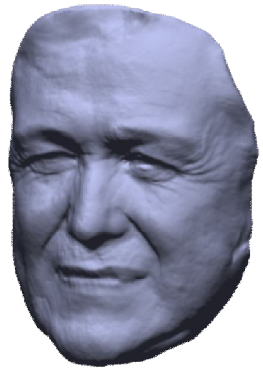


Shape Refinement Step

Previous frame



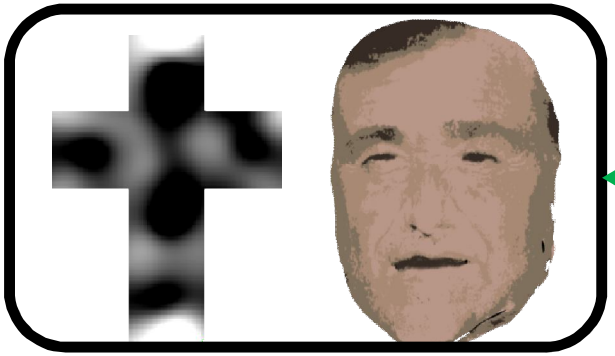
Smoothness constraints



Similarity constraints

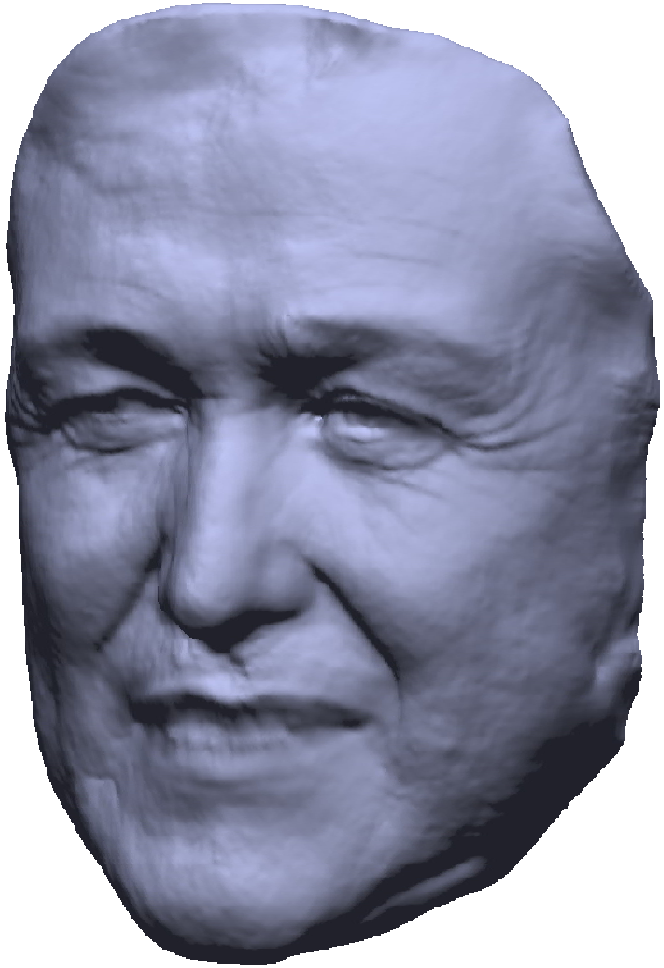
$$E_{tot} = E_{sha} + \alpha E_{sim} + \beta E_{s}$$

Shading constraints

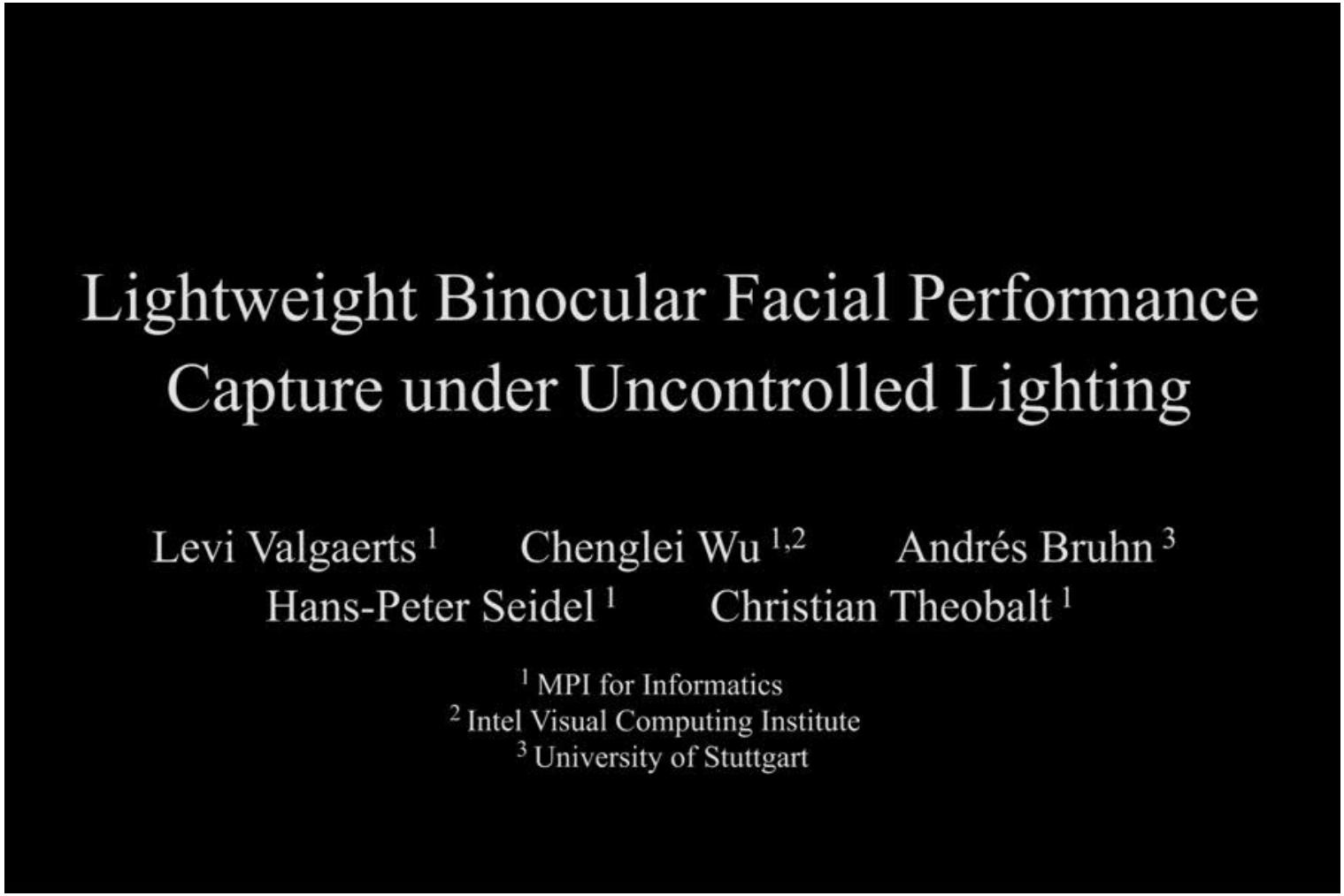


Current frame

Refinement Results



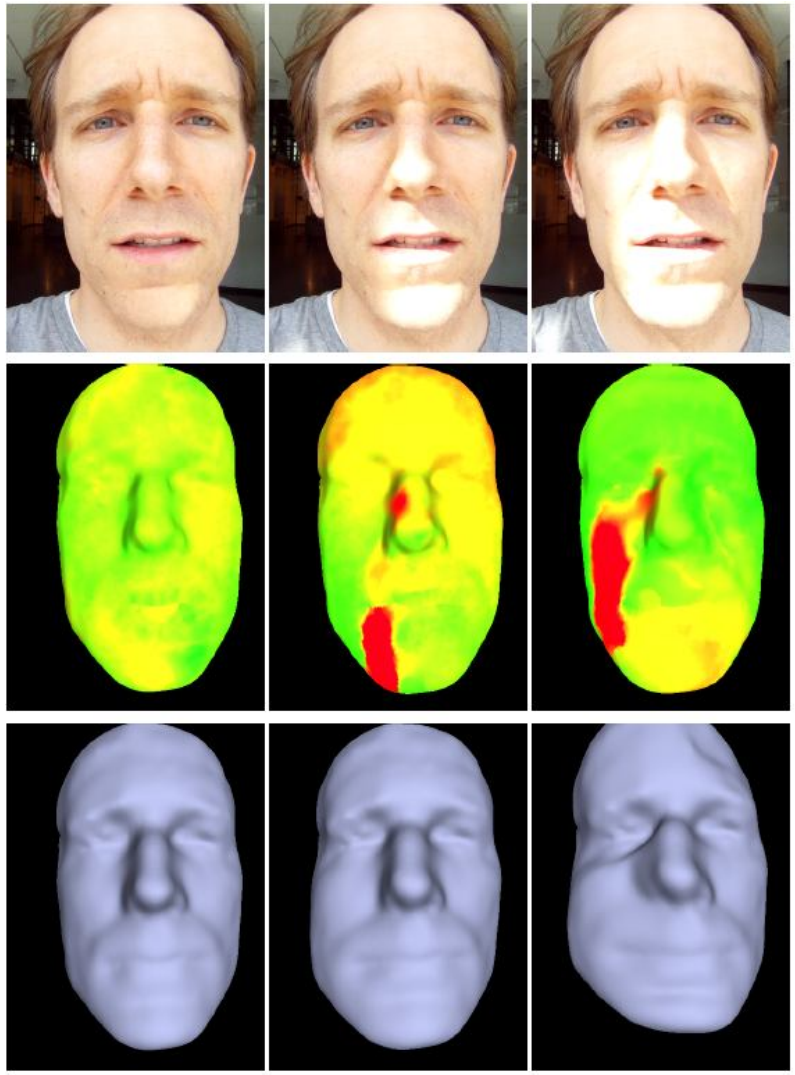
Face Performance Capture results



Limitations

- Stereo data set with small baseline
- Only two-view case, not all the face geometry is captured
- Strong self shadows

Failure Case



Summary

1. High-Quality Passive Facial Performance Capture using Anchor Frames

- Future work:
  - From controlled to uncontrolled lightning
  - More faithful reconstruction of the eye geometry
  - Using multiple reference frames simultaneously
  - Anchor clips for 3D reconstruction

2. Lightweight Binocular Facial Performance Capture under Uncontrolled Lighting

- Future work:
  - Photometric invariant scene flow
  - Better lighting models for shape refinement

References

- VALGAERTS, L., BRUHN, A., ZIMMER, H., WEICKERT, J., STOLL, C., AND THEOBALT, C. 2010. Joint estimation of motion, structure and geometry from stereo sequences. In *Proc. ECCV, Springer LNCS, vol. 6314, 568–581.*
- BEELER, T., BICKEL, B., SUMNER, R., BEARDSLEY, P., AND GROSS, M. 2010. High-quality single-shot capture of facial geometry. ACM Trans. Graphics (Proc. SIGGRAPH), 40.

Thanks to Pablo Garrido!

Questions ?