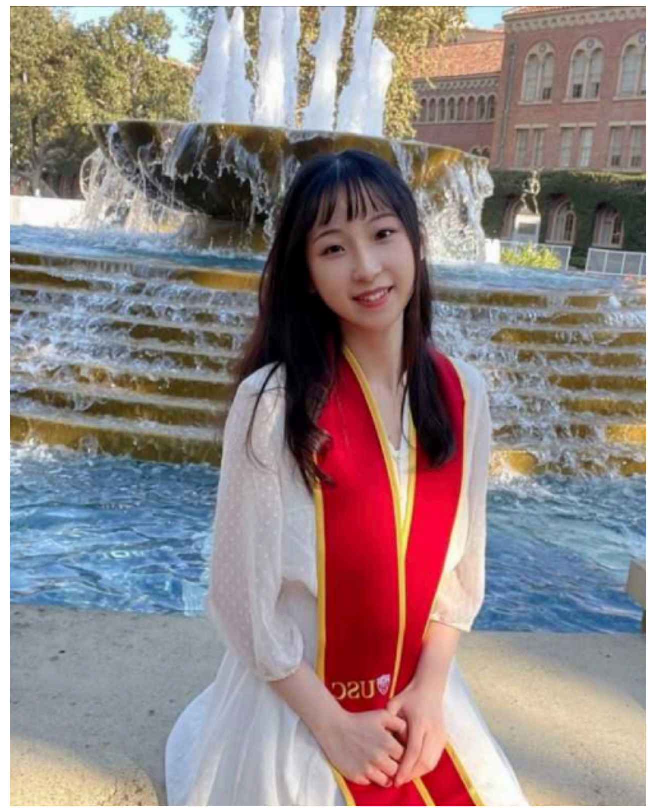
Topologically Consistent Multi-View Face Inference Using Volumetric Sampling

ICCV 2021

Tianye Li[1,2] Shichen Liu[1,2] Timo Bolkart[3] Jiayi Liu[1,2] Hao Li[1,2] Yajie Zhao[2]

USC

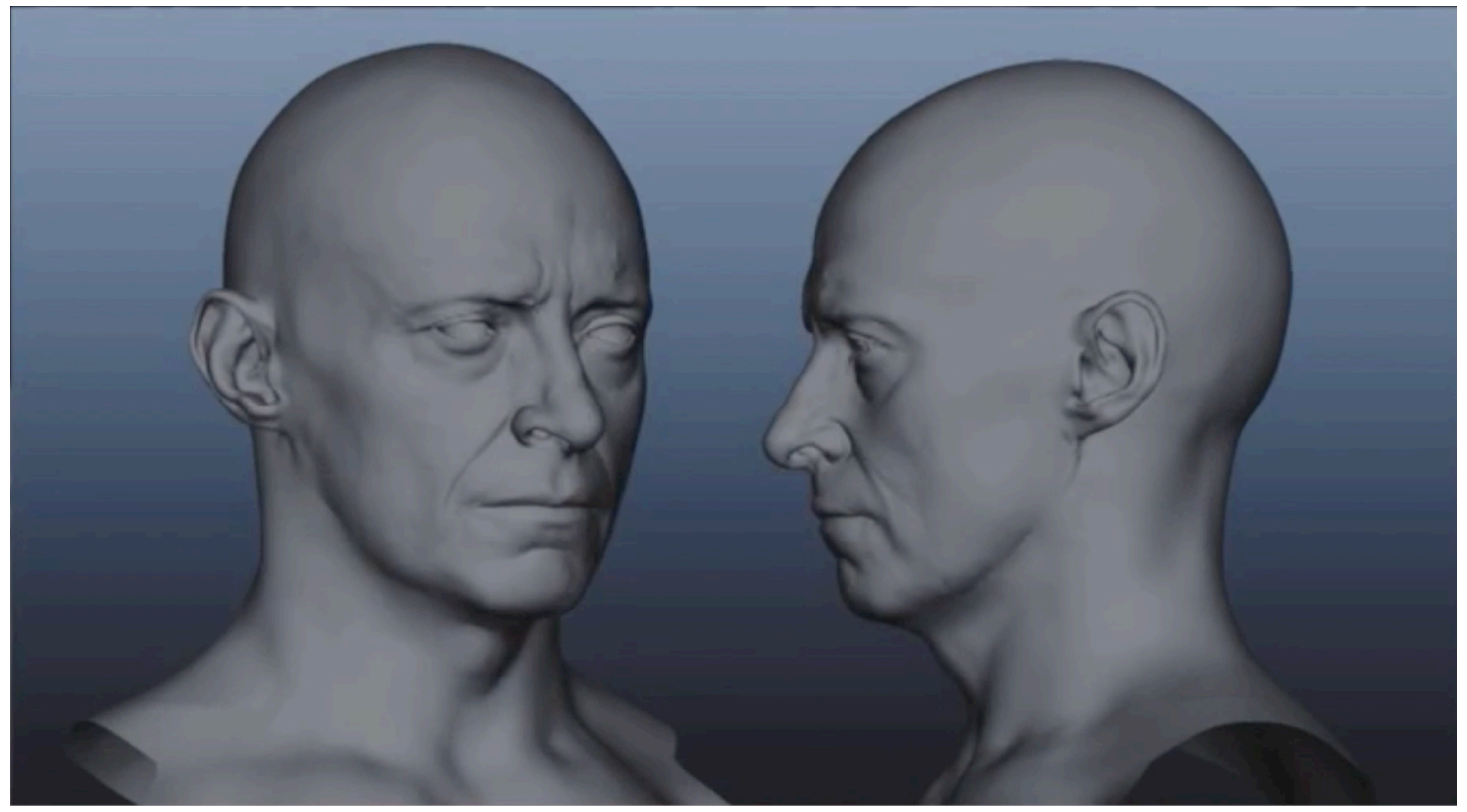
USC Institute for Creative Technologies

Max Planck Institute for Intelligent Systems

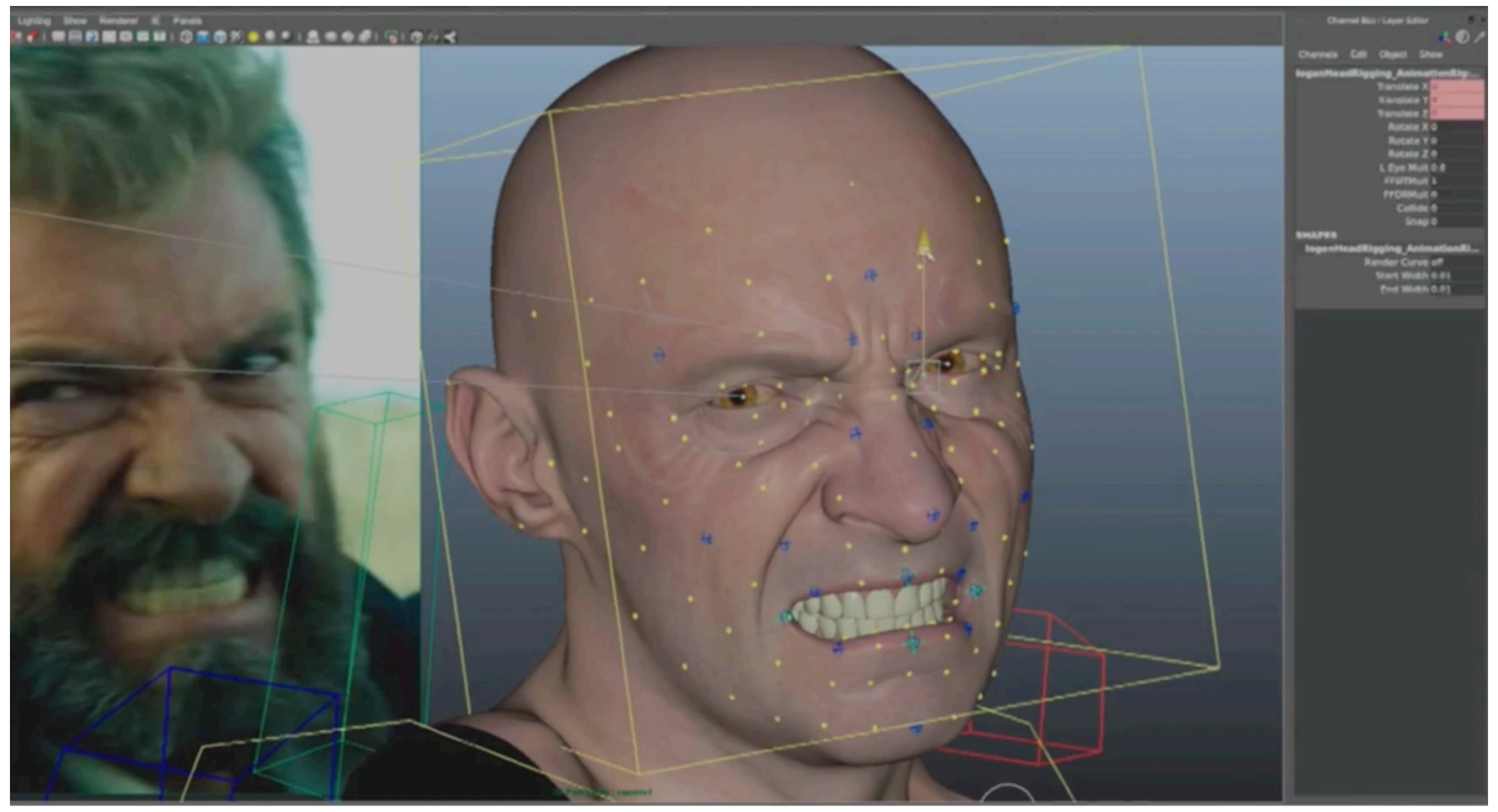
Virtual Character for VFX

Logan (Twentieth Century Fox / Image-Engine Design 2017)

Virtual Character for VFX



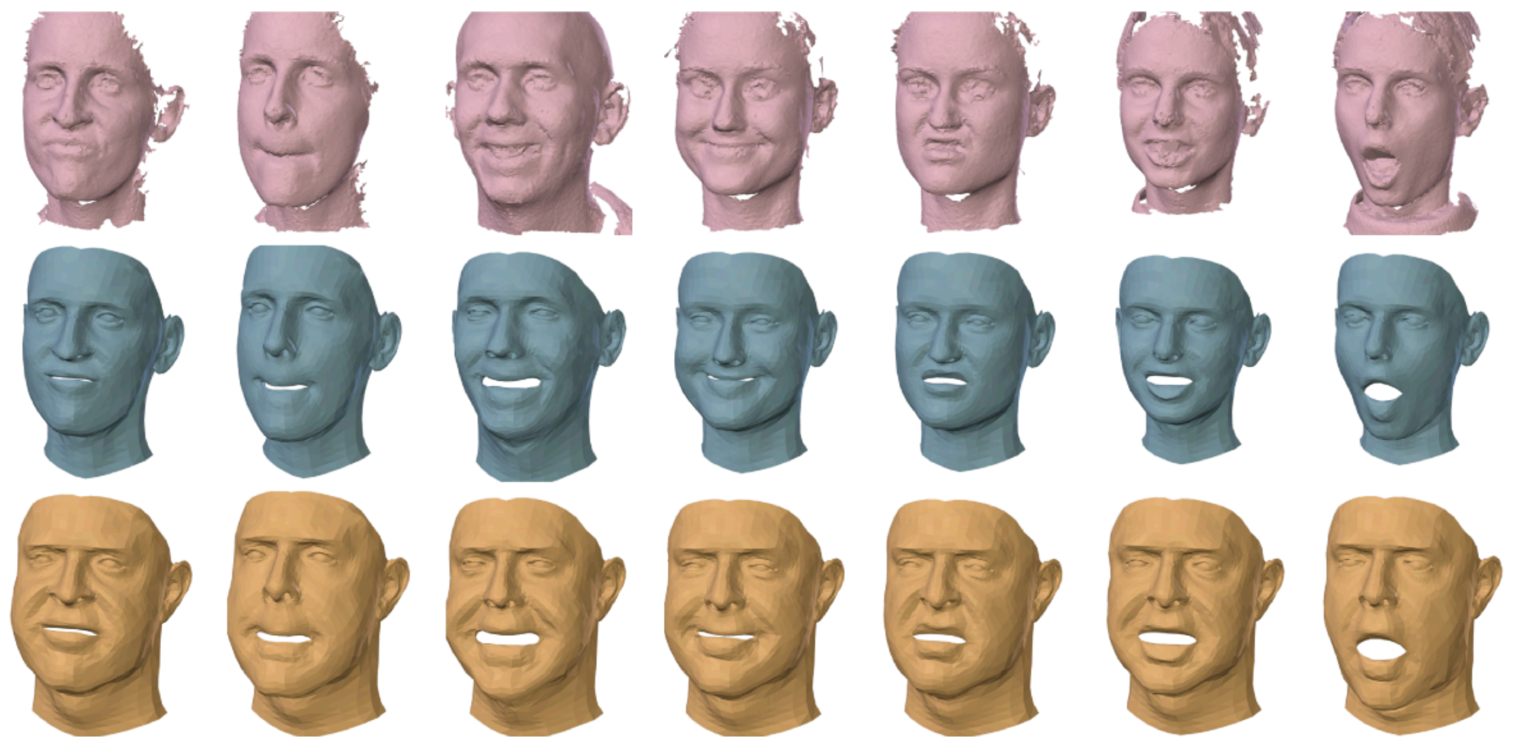
Personalized Model



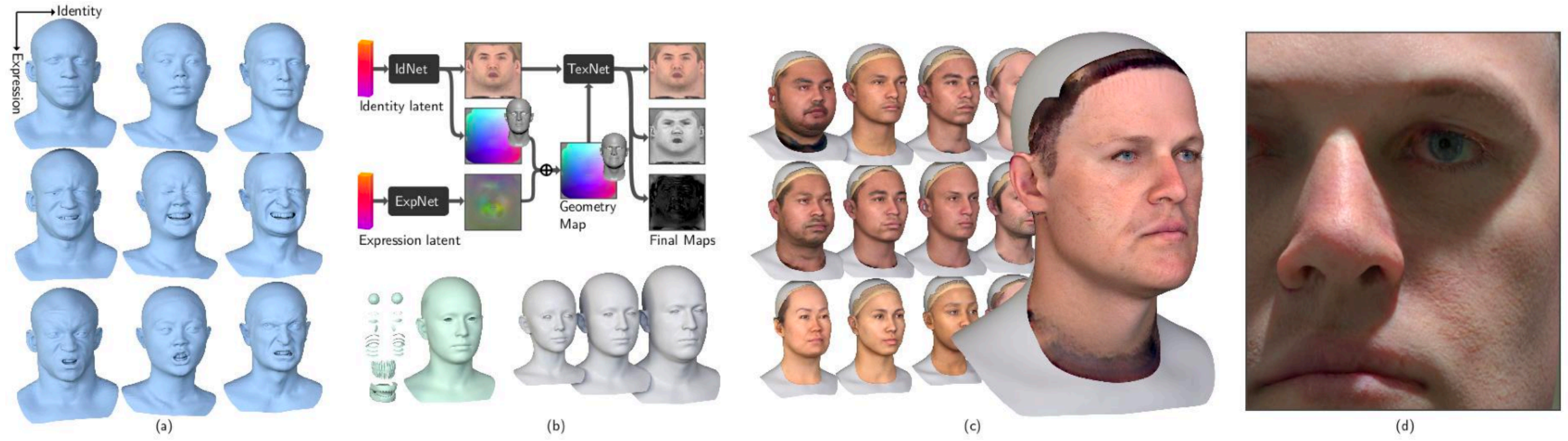
Animation

Appearance Capture and Rendering

Face Dataset Capture



Li and Bolkart et al. SIGGRAPH Asia 2017



Li, Bladin and Zhao et al. CVPR 2020

Require $10^3 \sim 10^5$ meshes

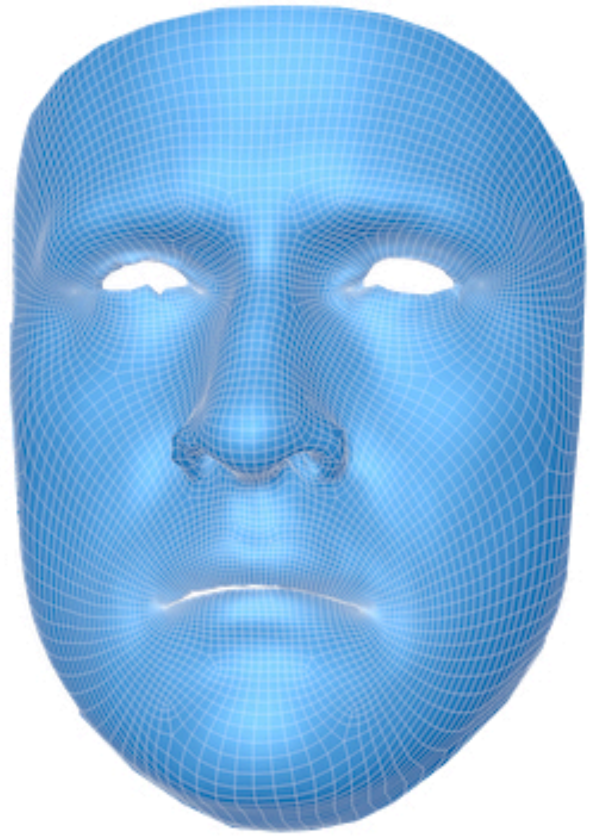
~10 min per mesh

$10^2 \sim 10^4$ days on a single machine

+ manual work

Goal: Face Capture and Registration

Calibrated multi-view images



Meshes in *consistent topology*
(“Registrations” or “Alignments”)

Appearance
and detail maps

Traditional Systems

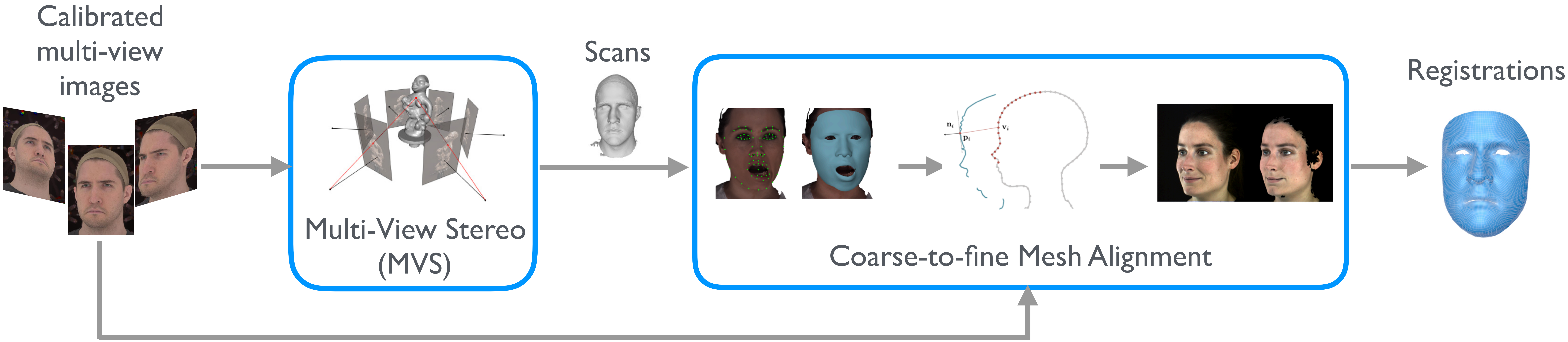
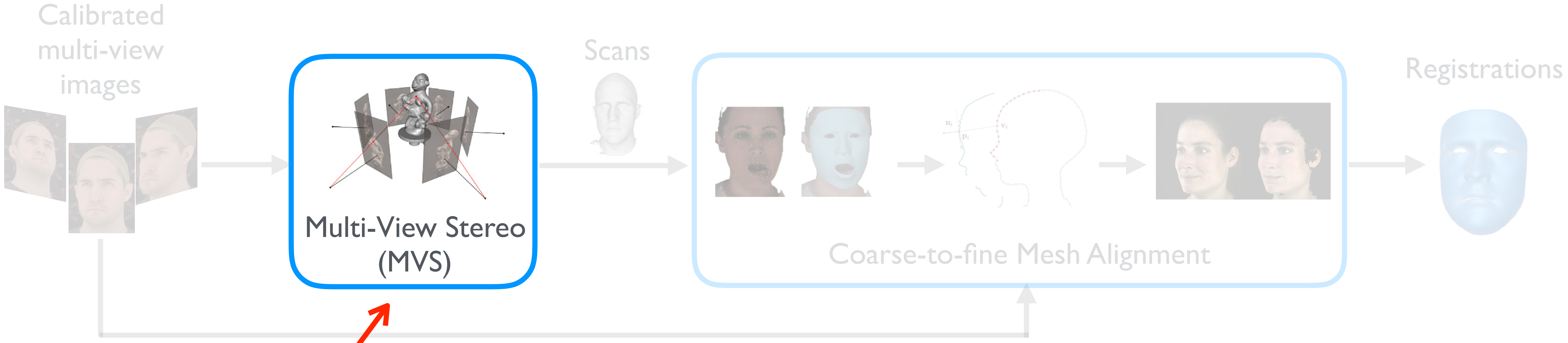


Image credits:
MVS: Furukawa and Hernández, "Multi-View Stereo: A Tutorial", Found. Trends Comput. Graph. Vis. 2015
Mesh alignments: Li et al. "Learning a model of facial shape and expression from 4D scans." SIGGRAPH Asia 2017;
Hao Li, "Animation reconstruction of deformable surfaces". Diss. ETH Zurich 2010

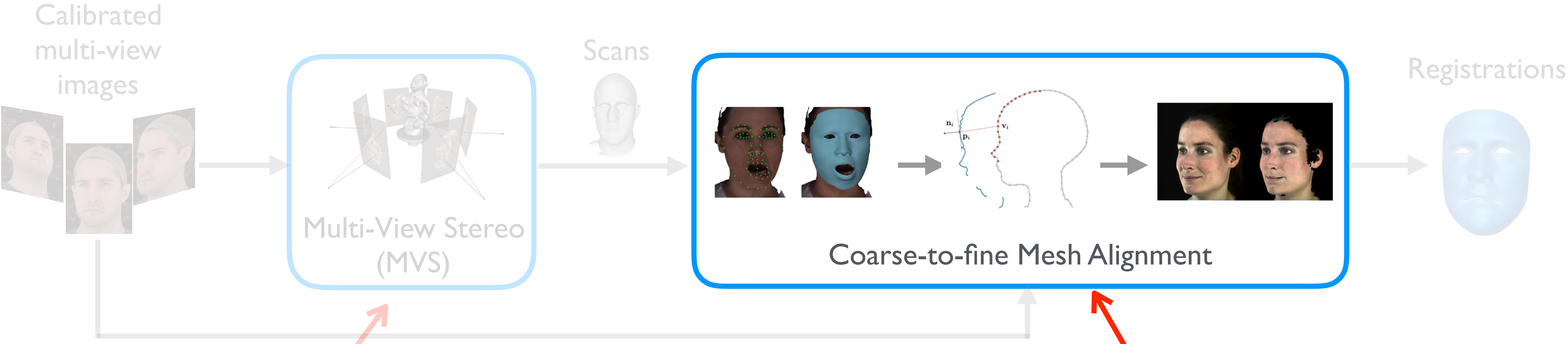
Traditional Systems



Issues:
- Slow
- Prone to errors

  Texture-less, transparent, or reflective surfaces

Traditional Systems

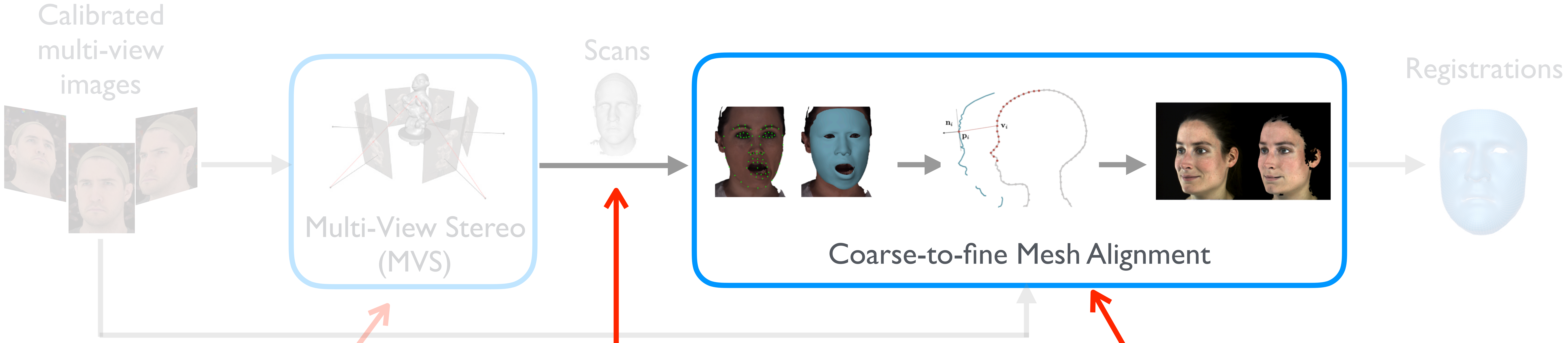


- Slow
- Requires manual work

Issues:
- Slow
- Prone to errors

Texture-less, transparent, or reflective surfaces

Traditional Systems



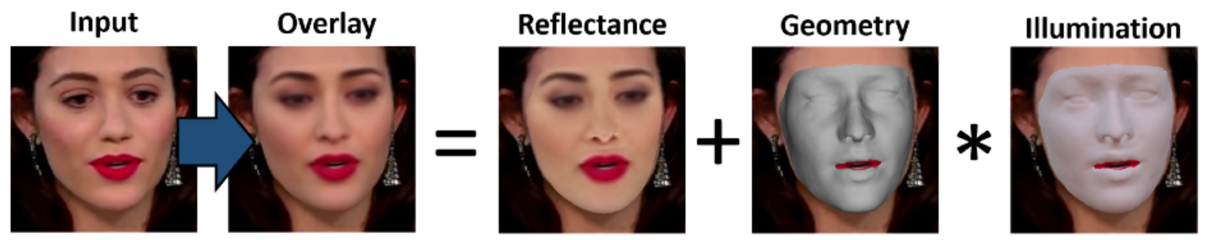
Issues:
- Slow
- Prone to errors

Texture-less, transparent, or reflective surfaces

MVS errors affect registration quality

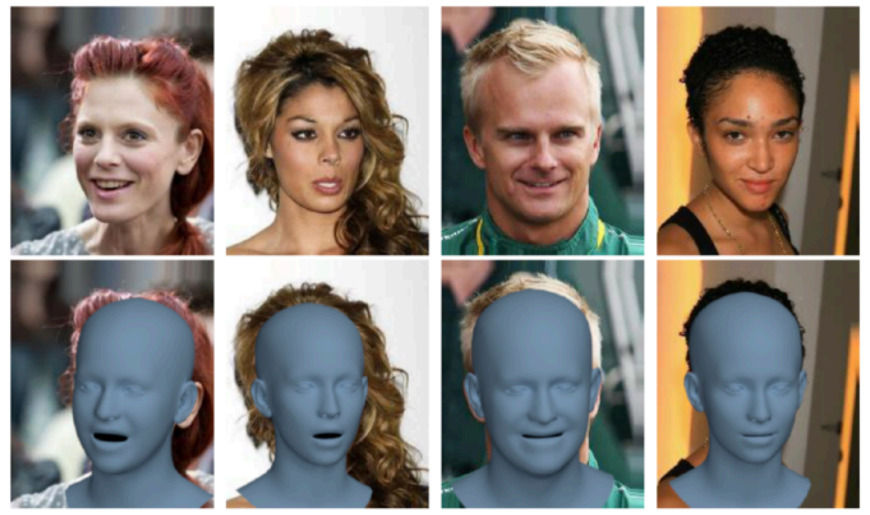
- Slow
- Requires manual work

Previous Work: Faces

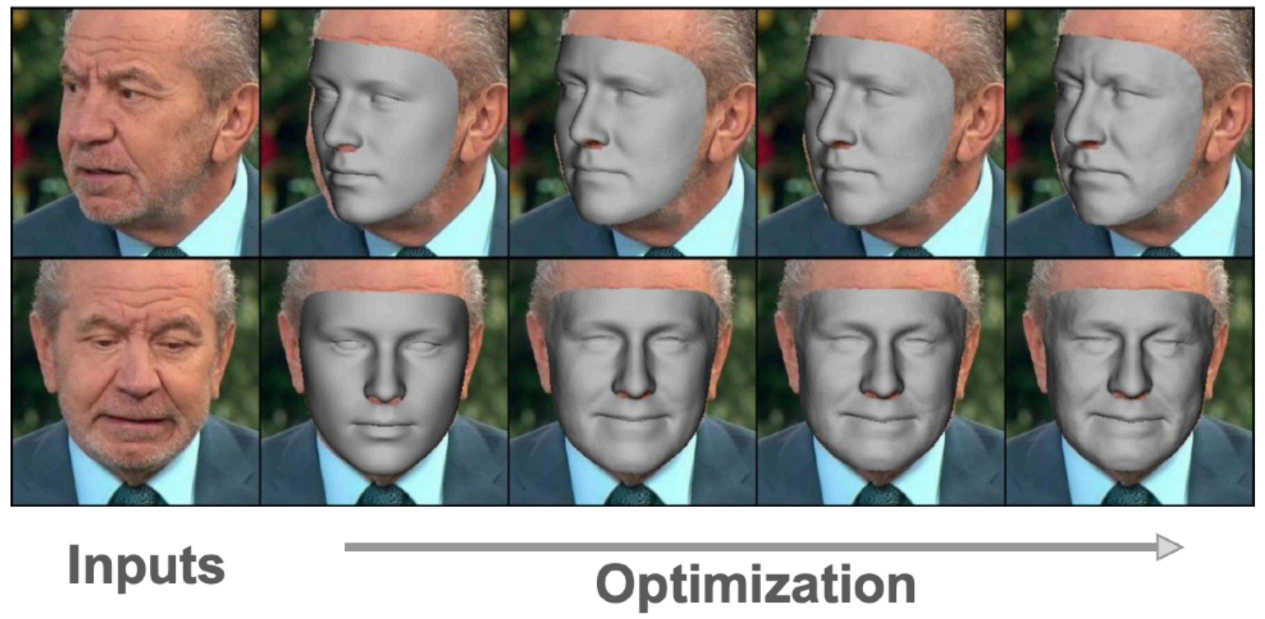


FML
Tewari et al. CVPR 19

PRNet
Feng et al. ECCV 18



RingNet
Sanyal et al. CVPR 19



DFNRMVS
Bai et al. CVPR 20

Can work on unconstrained inputs

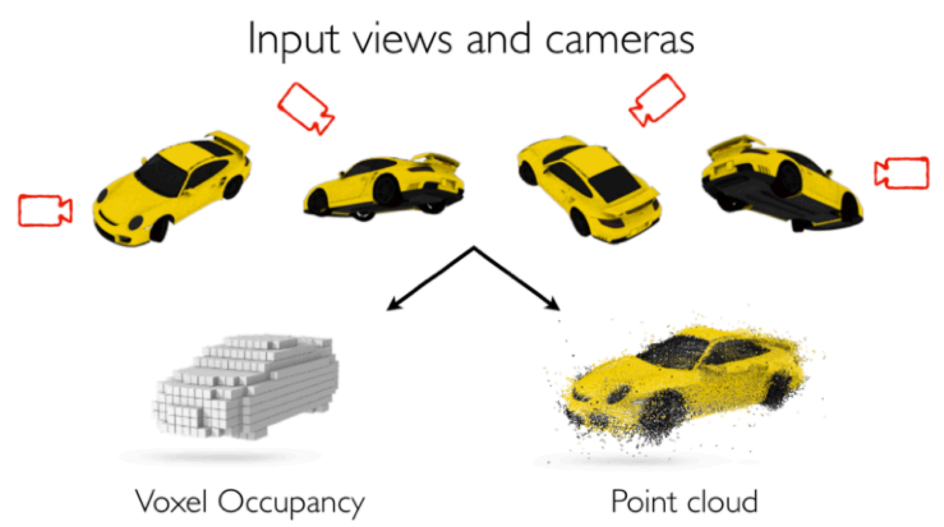
Not *metrically* accurate

- "Memorization"
- Cannot explicitly utilize 3D info (if naively adapted to multi-view)

Challenging for extreme expression

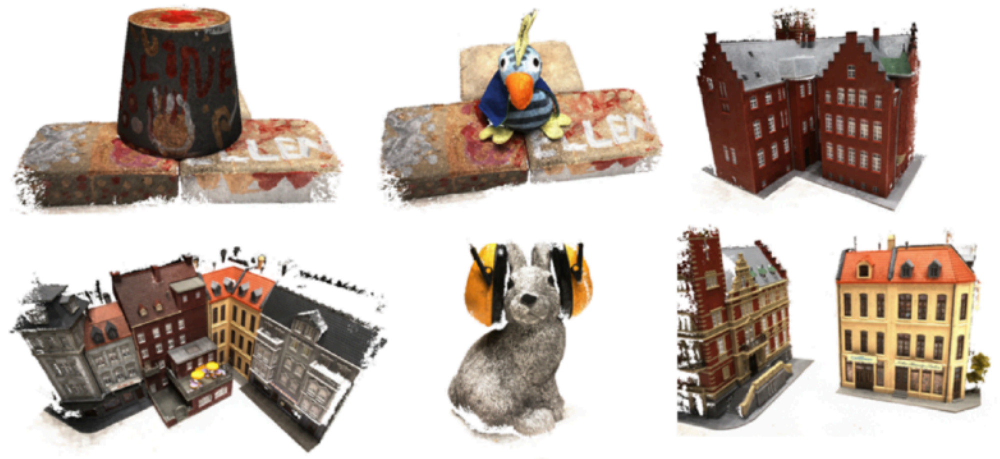
- Constrained by 3DMM

Previous Work: Beyond Faces



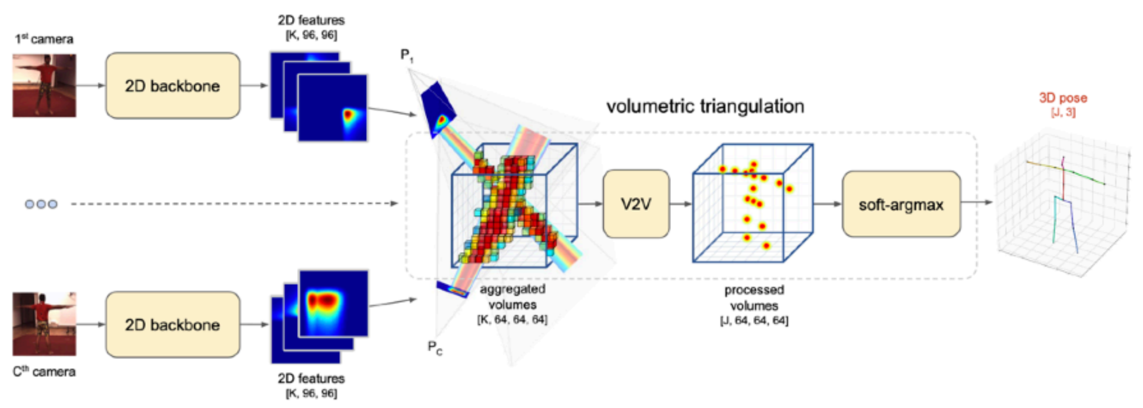
Learning MVS Machine
Kar et al. NeurIPS 17

Deep Volumetric Video
Huang et al. ECCV 18



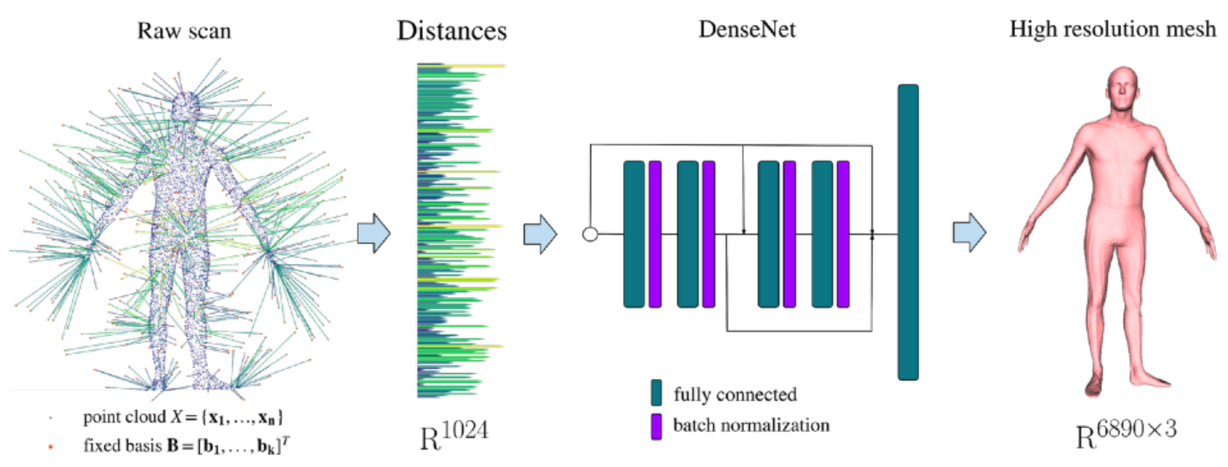
CasMVSNet
Gu et al. CVPR 20

No dense correspondence



Learnable Triangulation
Iskakov et al. ICCV 19

Only sparse correspondences



Basis Point Set (BPS)
Prokudin et al. ICCV 19

Still needs scans (from MVS)

The ToFu Framework

Topologically consistent **F**ace inference from m**u**lti-view

Images from multi-view

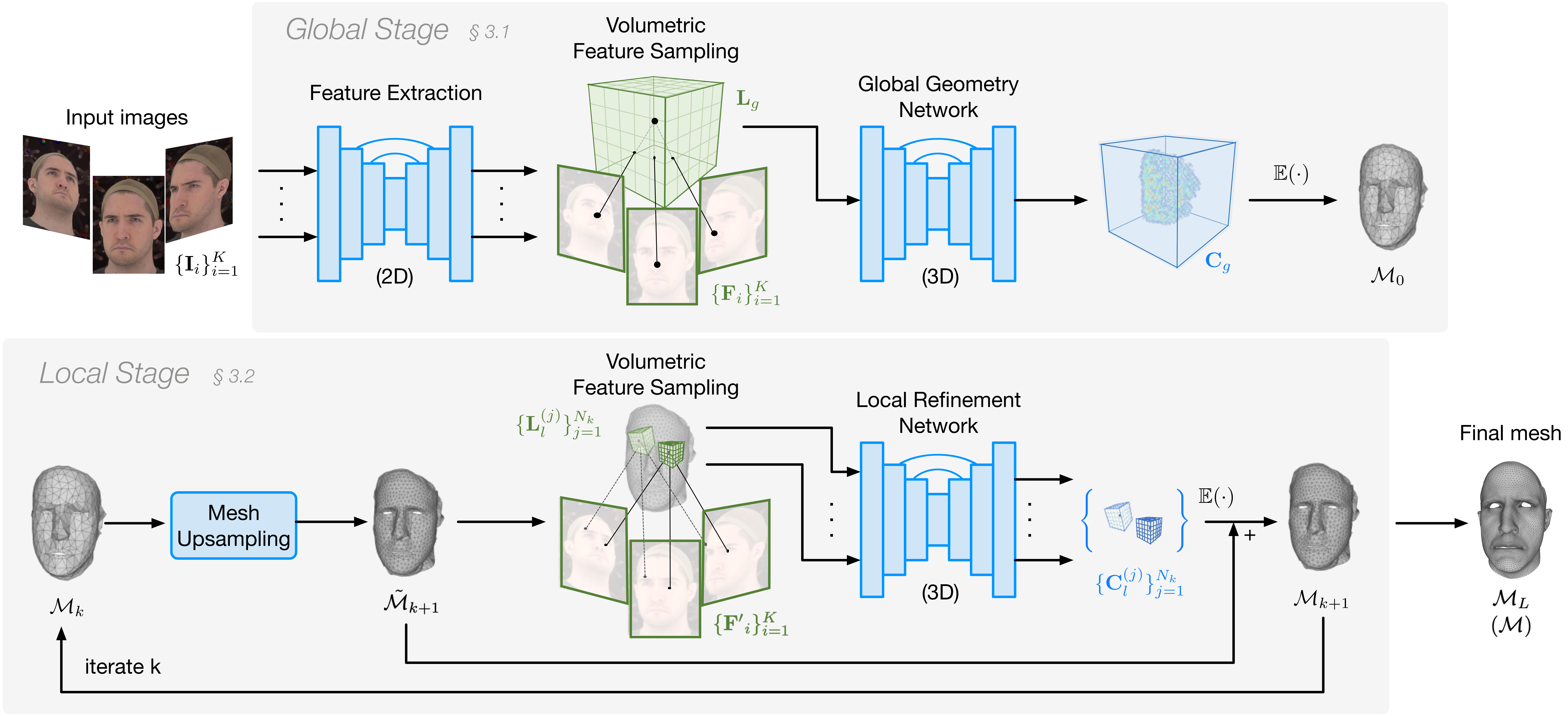
Mesh in consistent topology inferred in 0.385 seconds

Rendering with captured facial details and appearances

The ToFu Framework

Topologically consistent **F**ace inference from m**u**lti-view

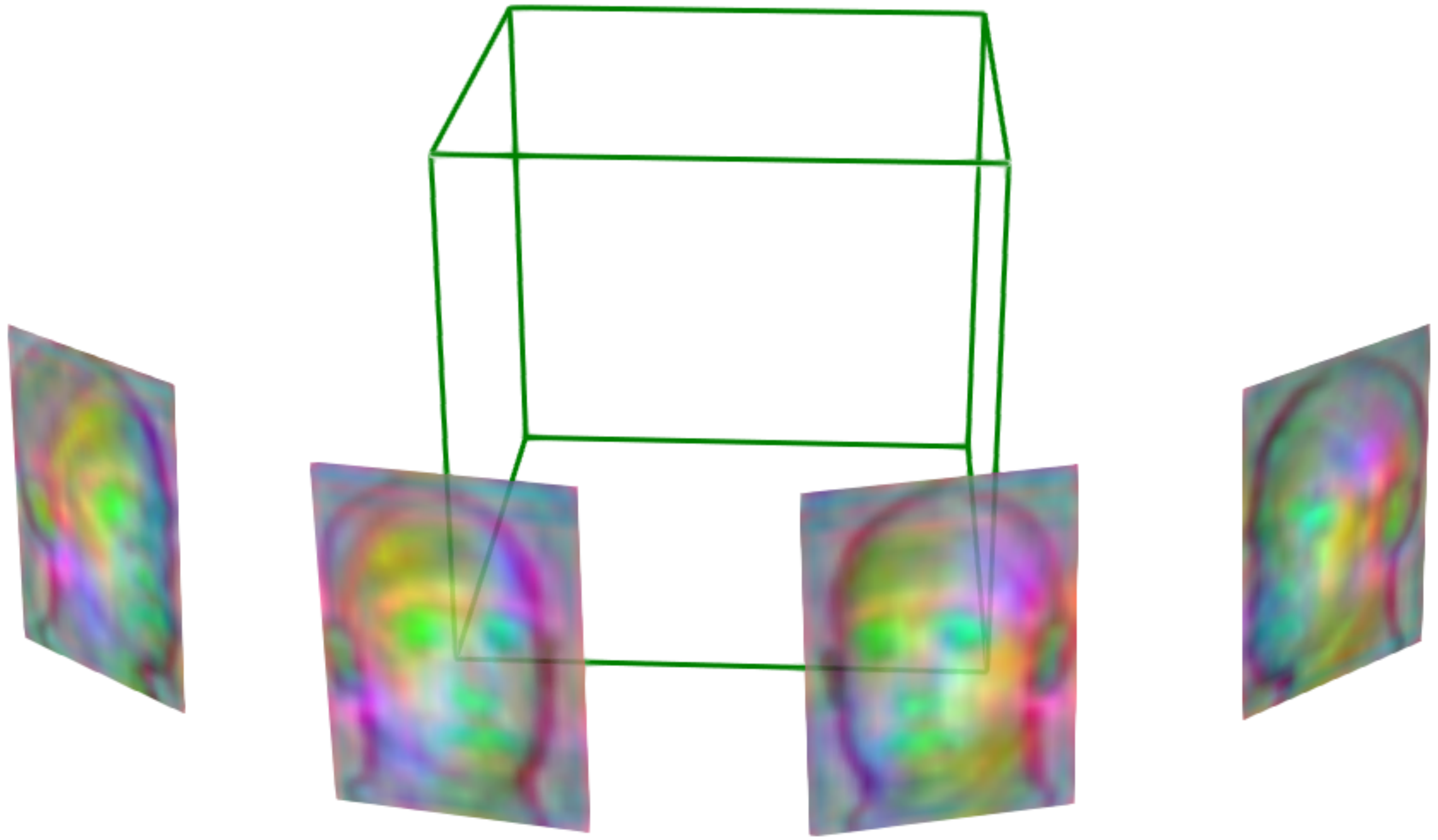
Progressive Mesh Generation



Global Stage

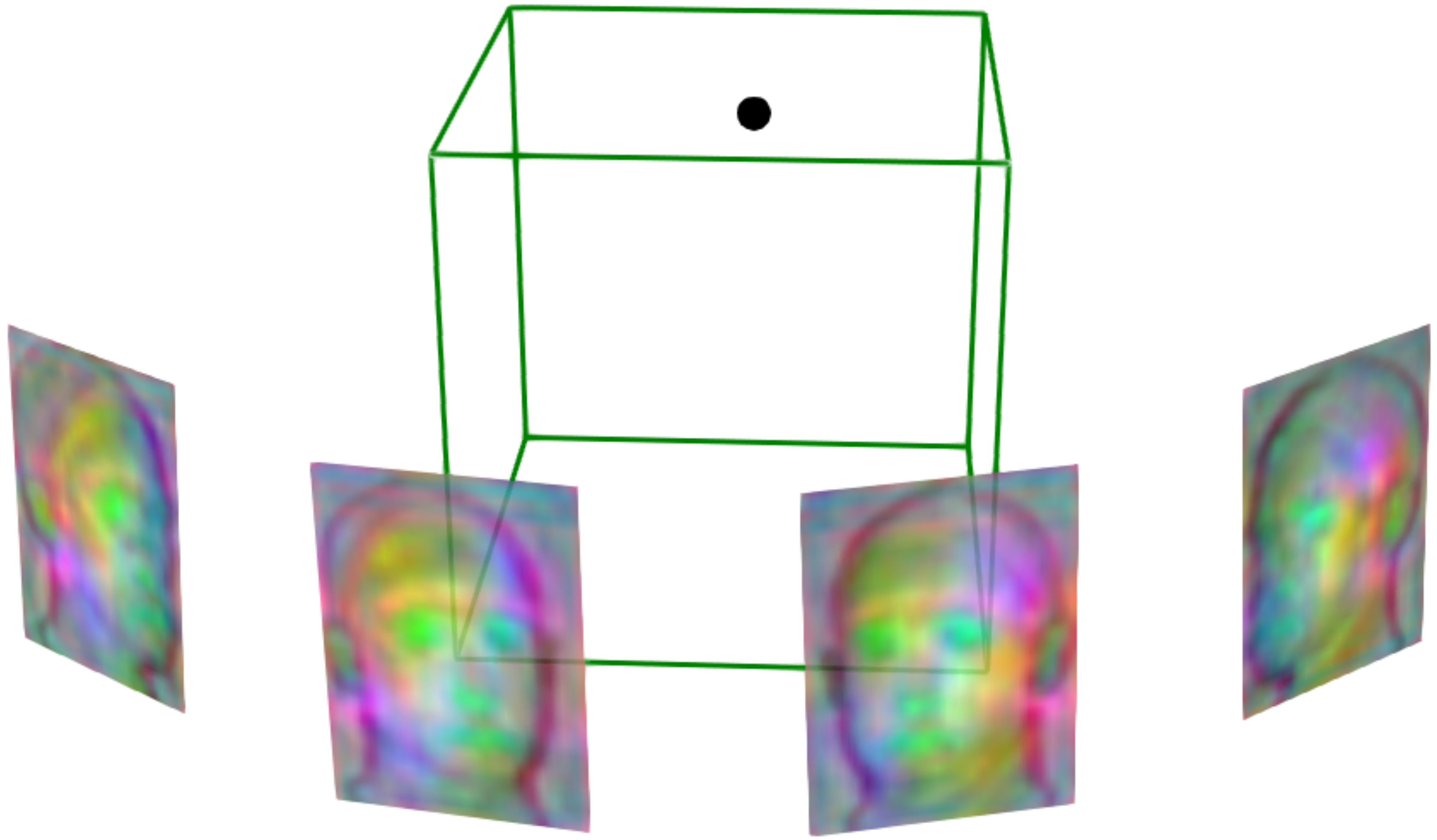
Input images

Global Stage: Feature Maps



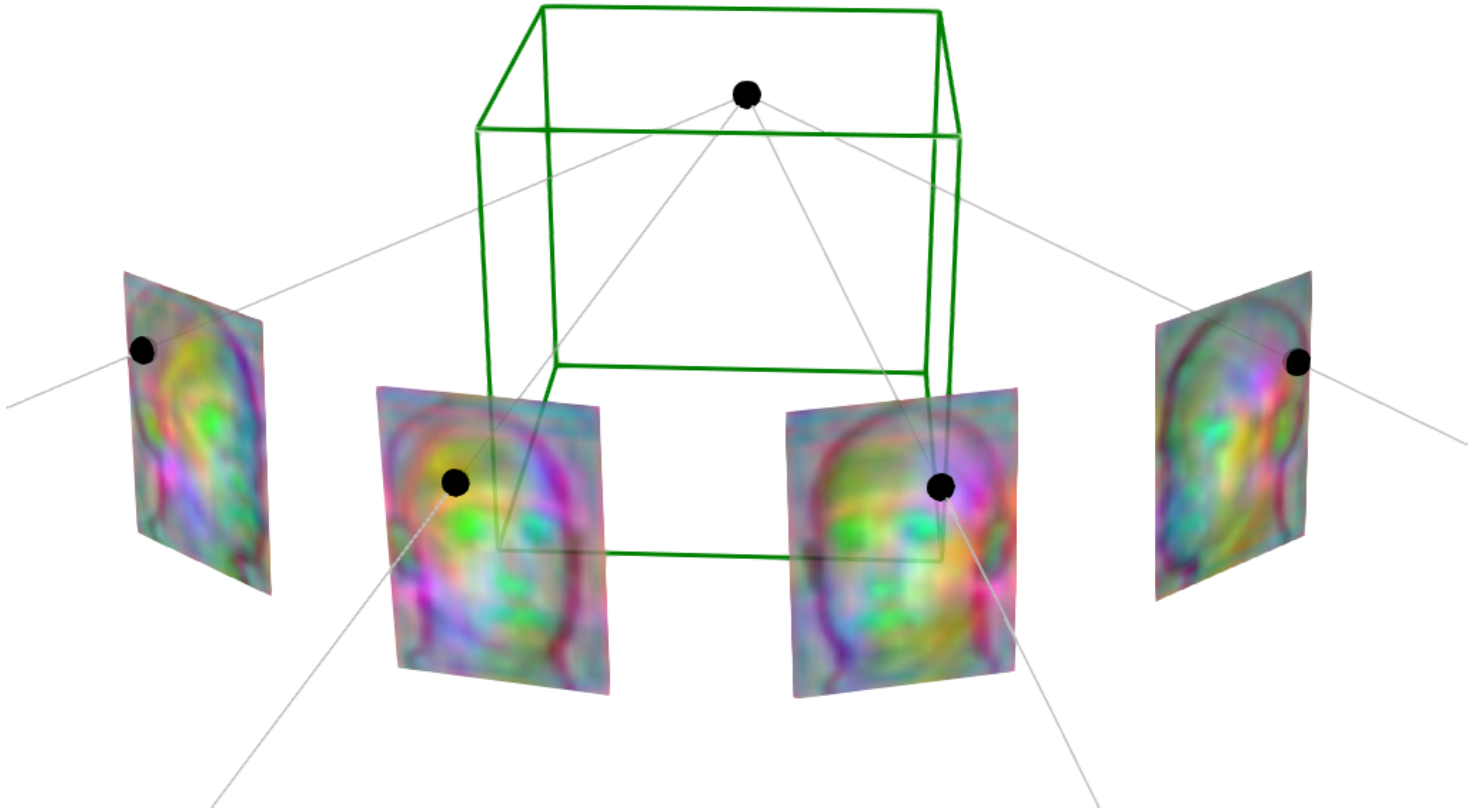
Inferred high-dim. feature maps

Global Stage: Volumetric Feature Sampling



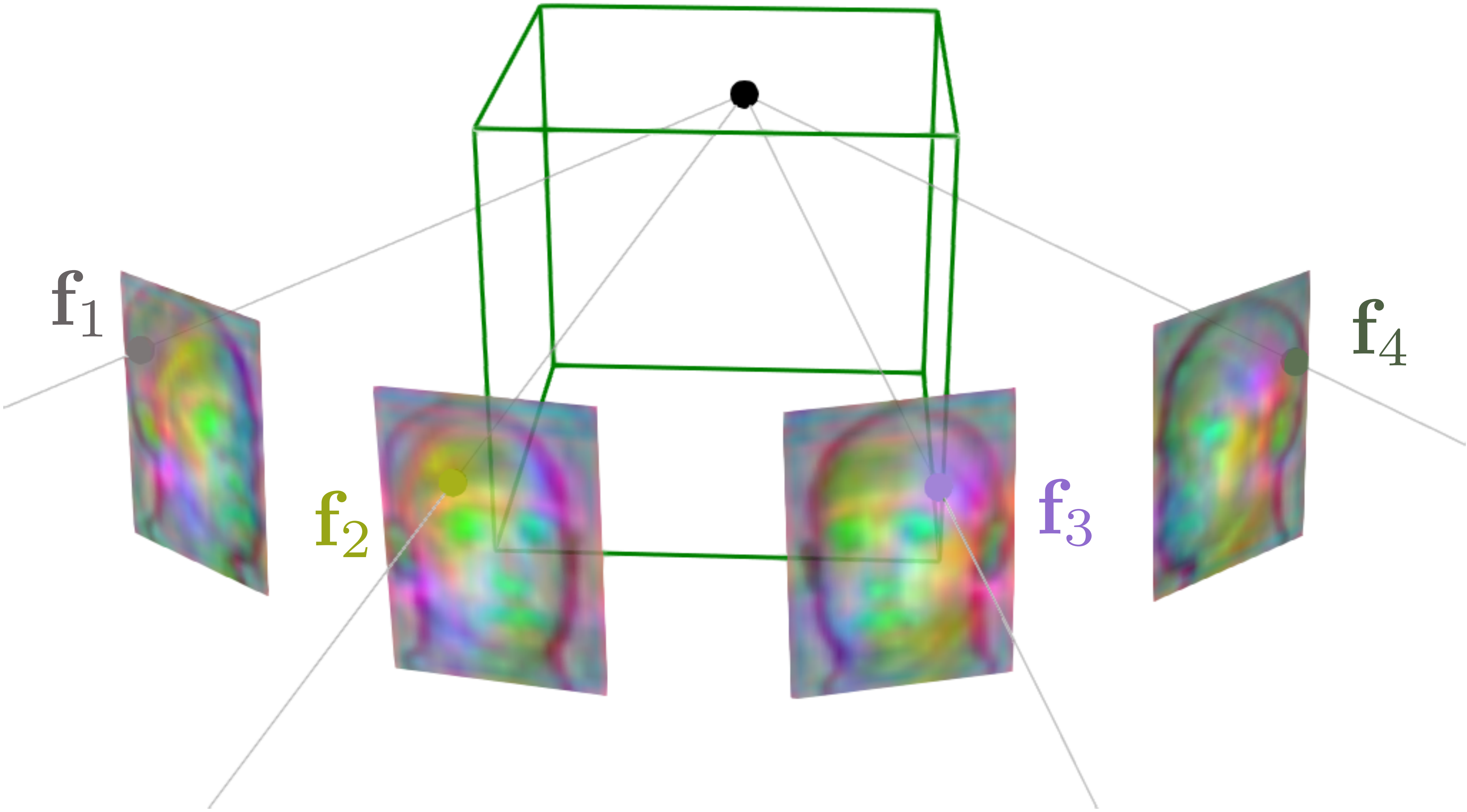
Volumetric feature sampling

Global Stage: Volumetric Feature Sampling



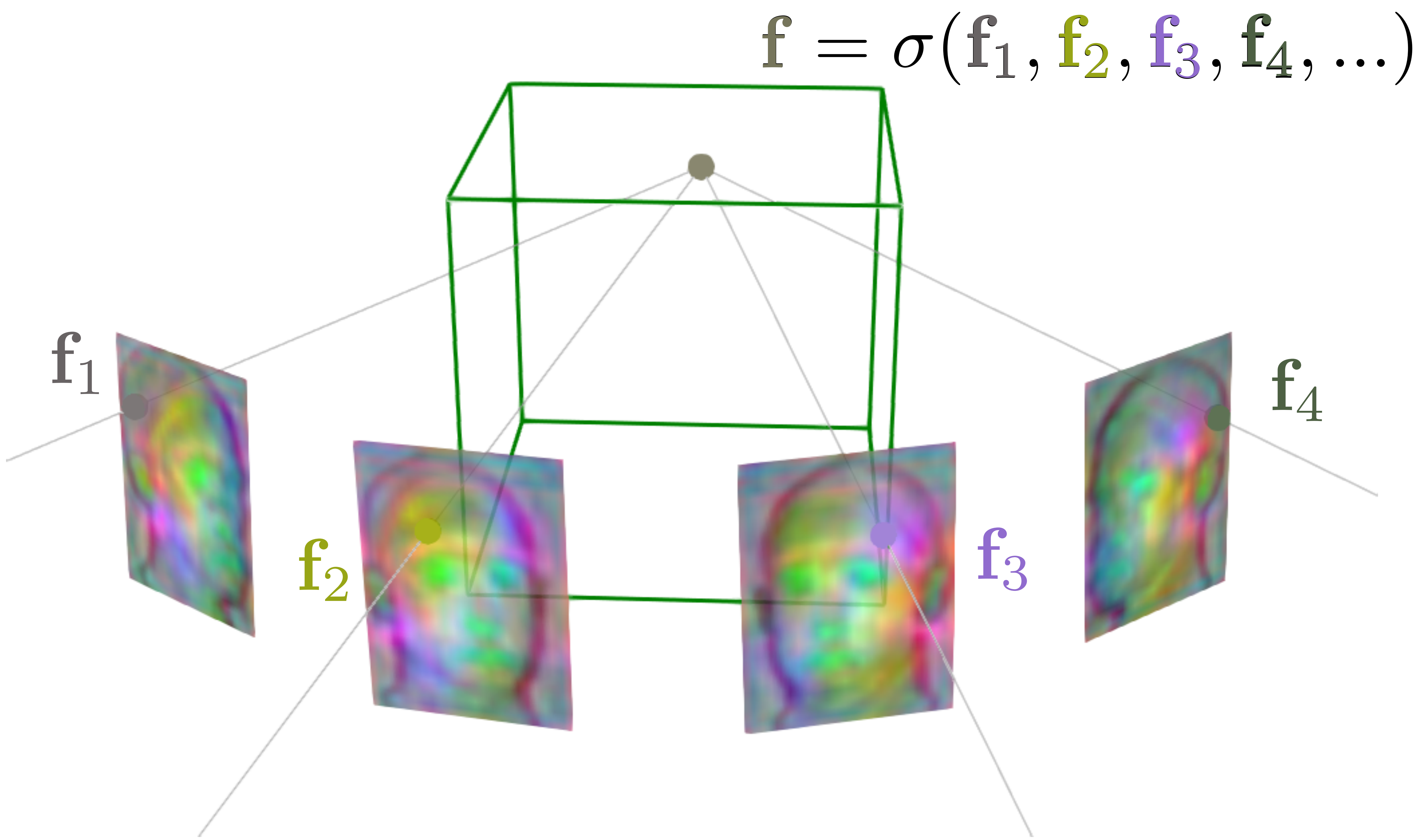
Volumetric feature sampling

Global Stage: Volumetric Feature Sampling



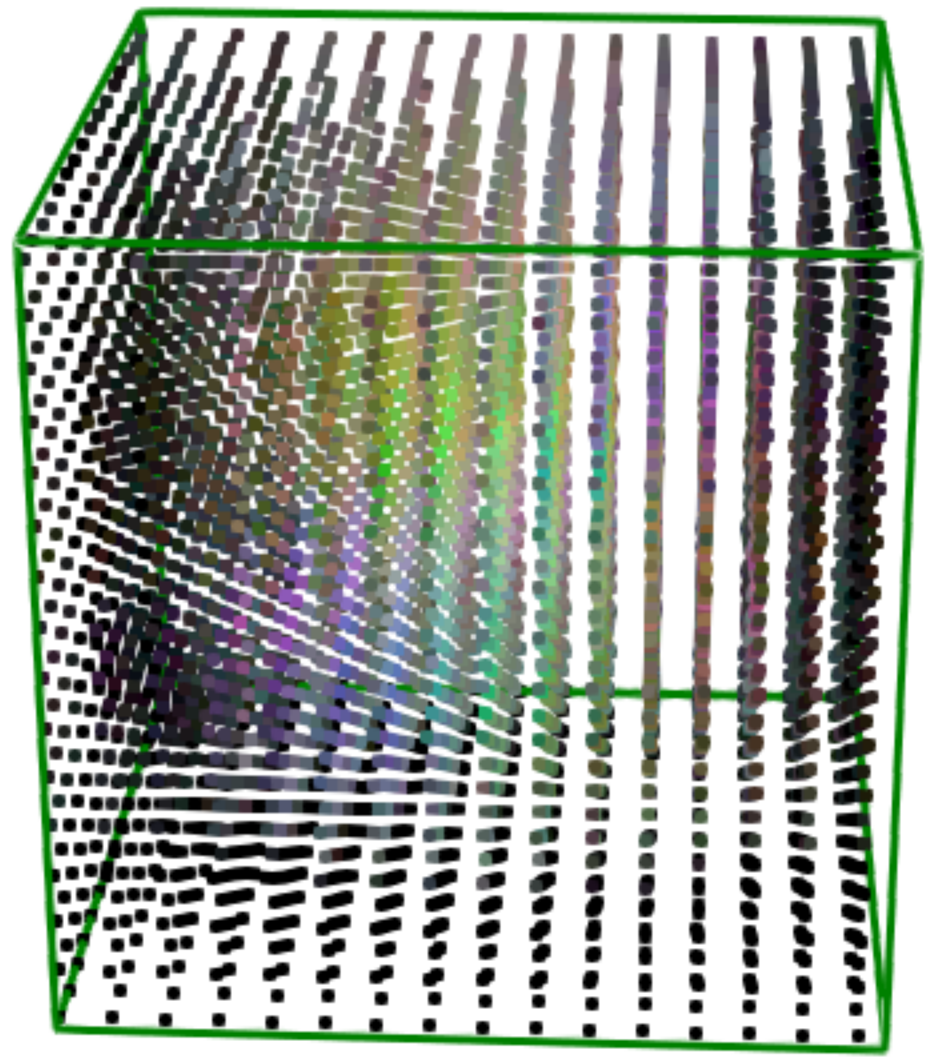
Volumetric feature sampling

Global Stage: Volumetric Feature Sampling



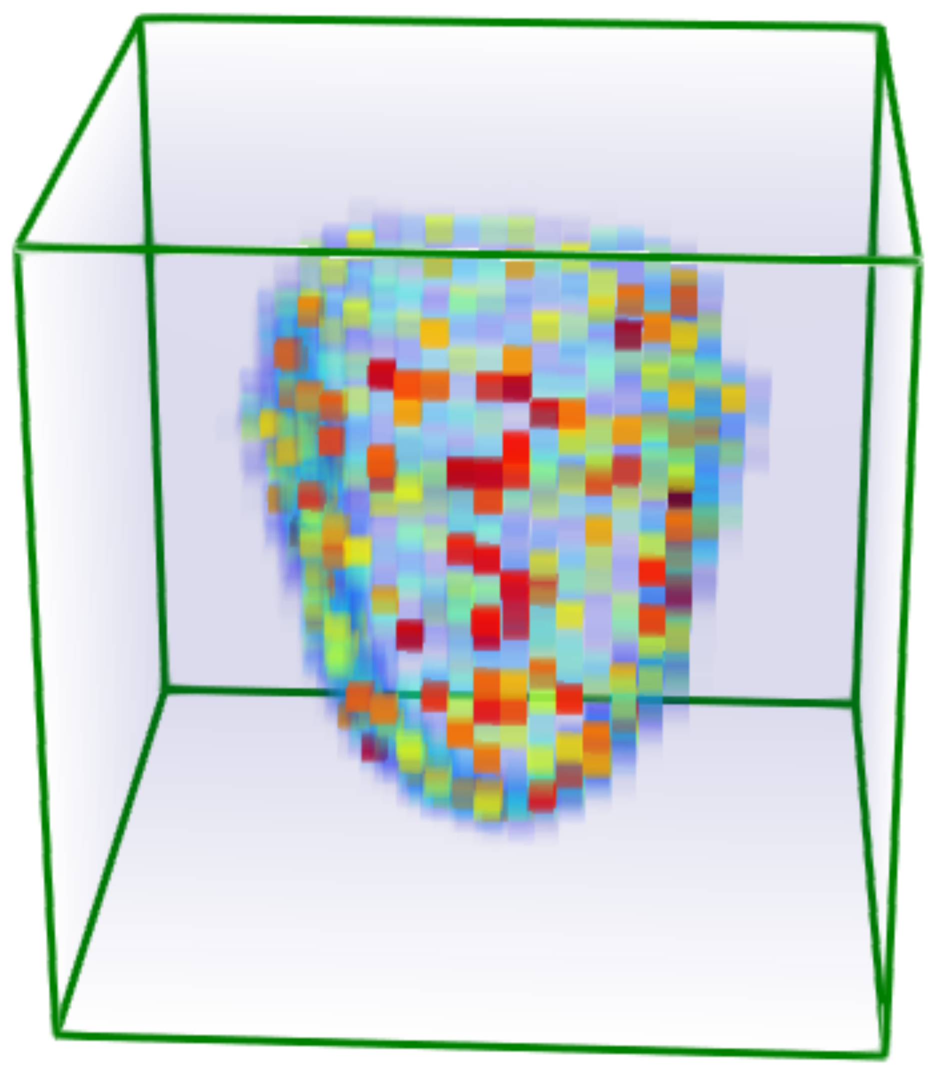
Volumetric feature sampling

Global Stage: Volumetric Feature Sampling



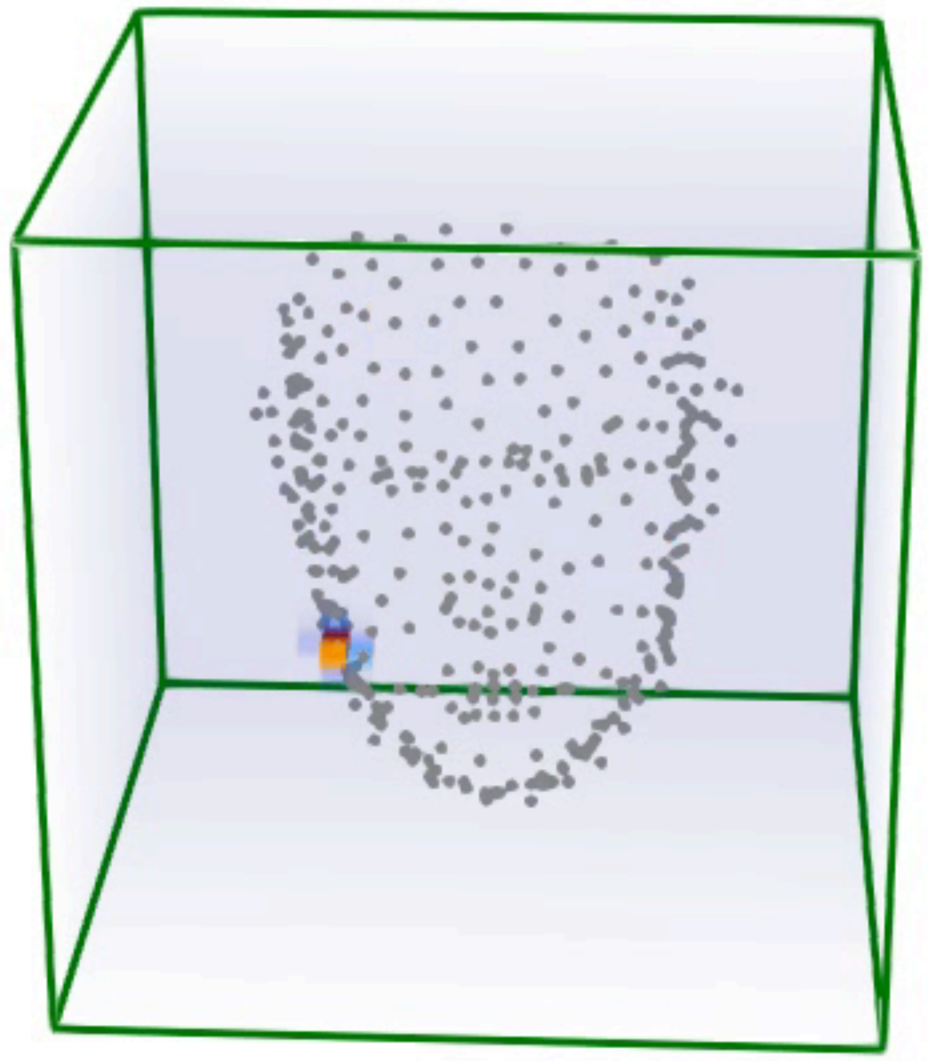
Feature volume

Global Stage: Mesh in Consistent Topology



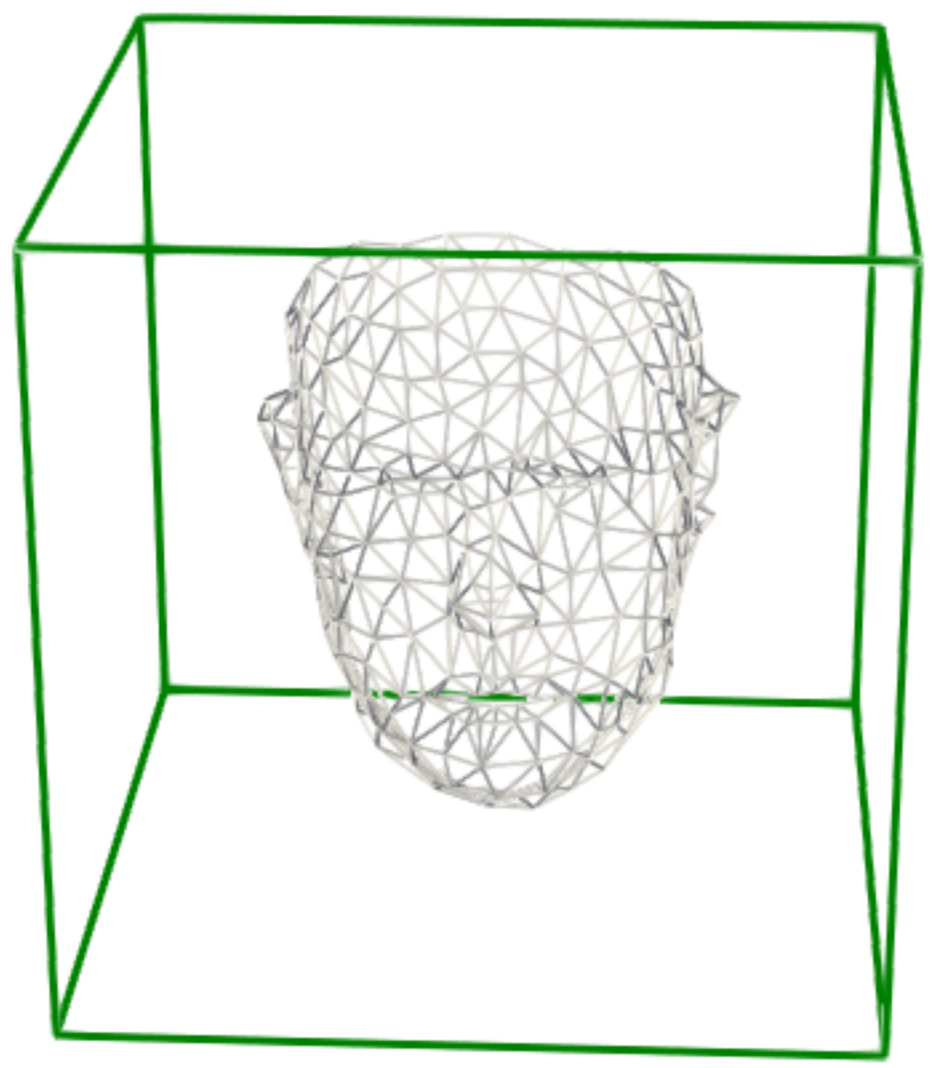
Inferred probability volume

Global Stage: Mesh in Consistent Topology



Extract vertex positions
in designed topology

Global Stage: Mesh in Consistent Topology

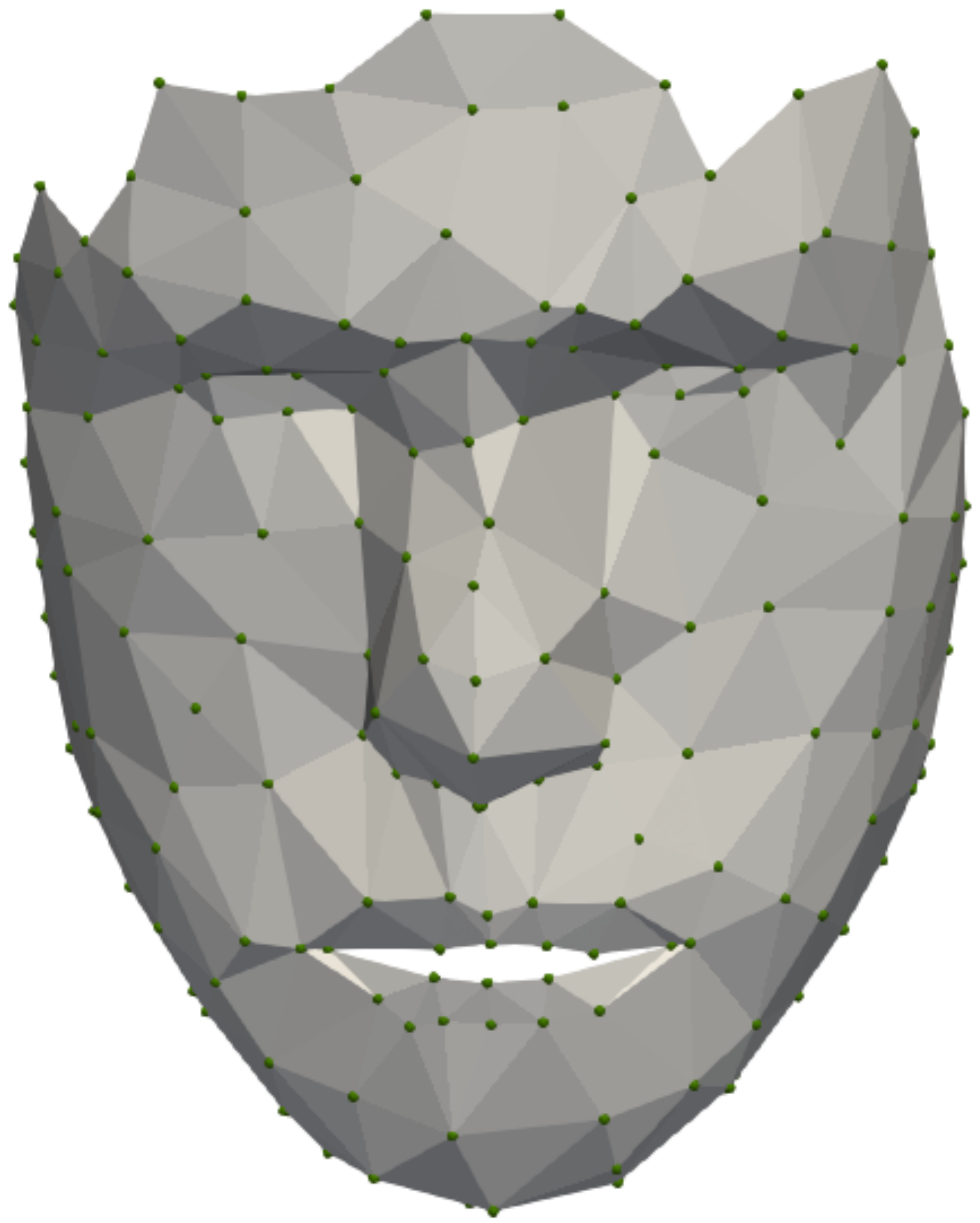


Inferred mesh

Local Stage

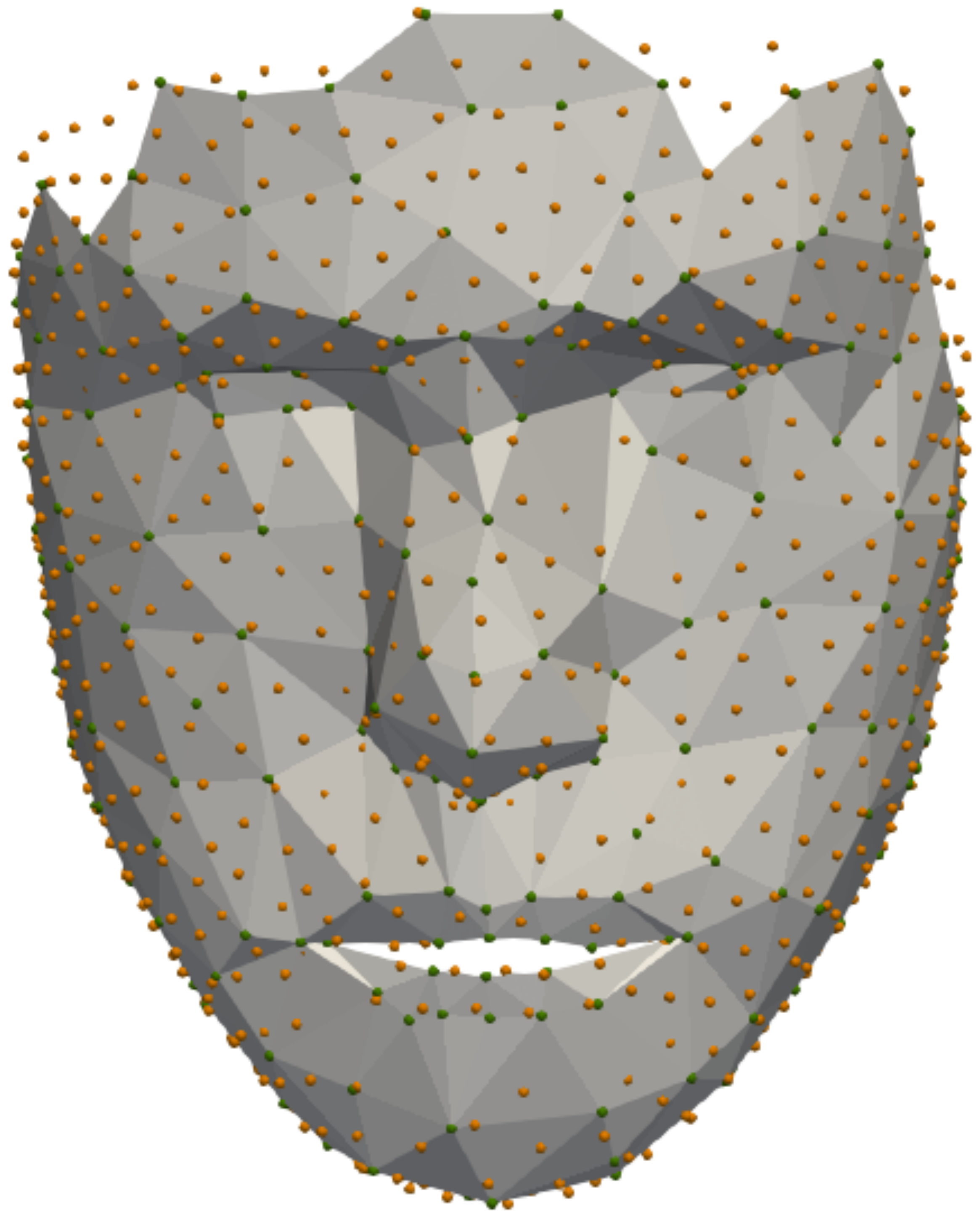
Inferred mesh

Local Stage: Mesh Upsampling



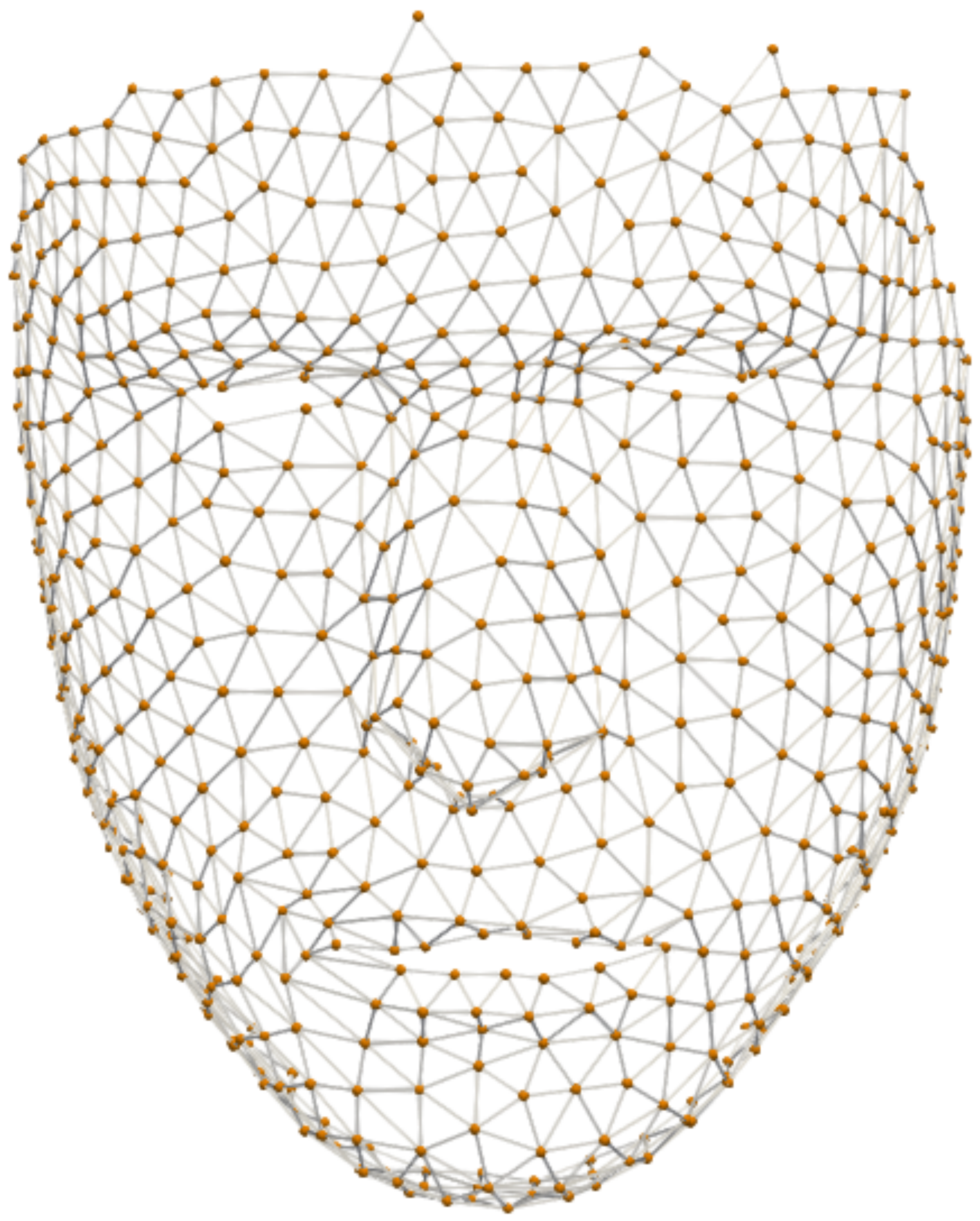
Mesh upsampling

Local Stage: Mesh Upsampling



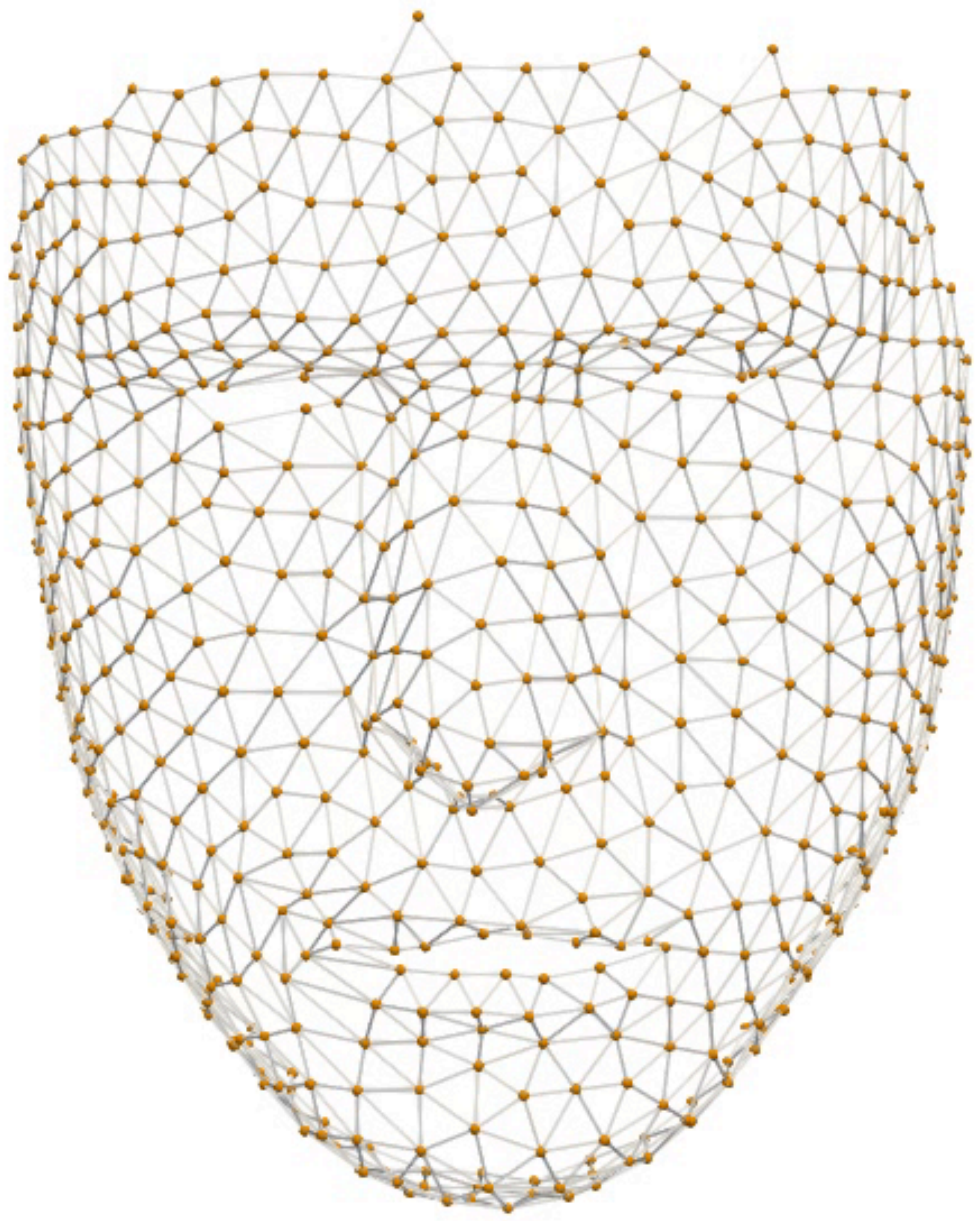
Mesh upsampling

Local Stage: Local Refinement

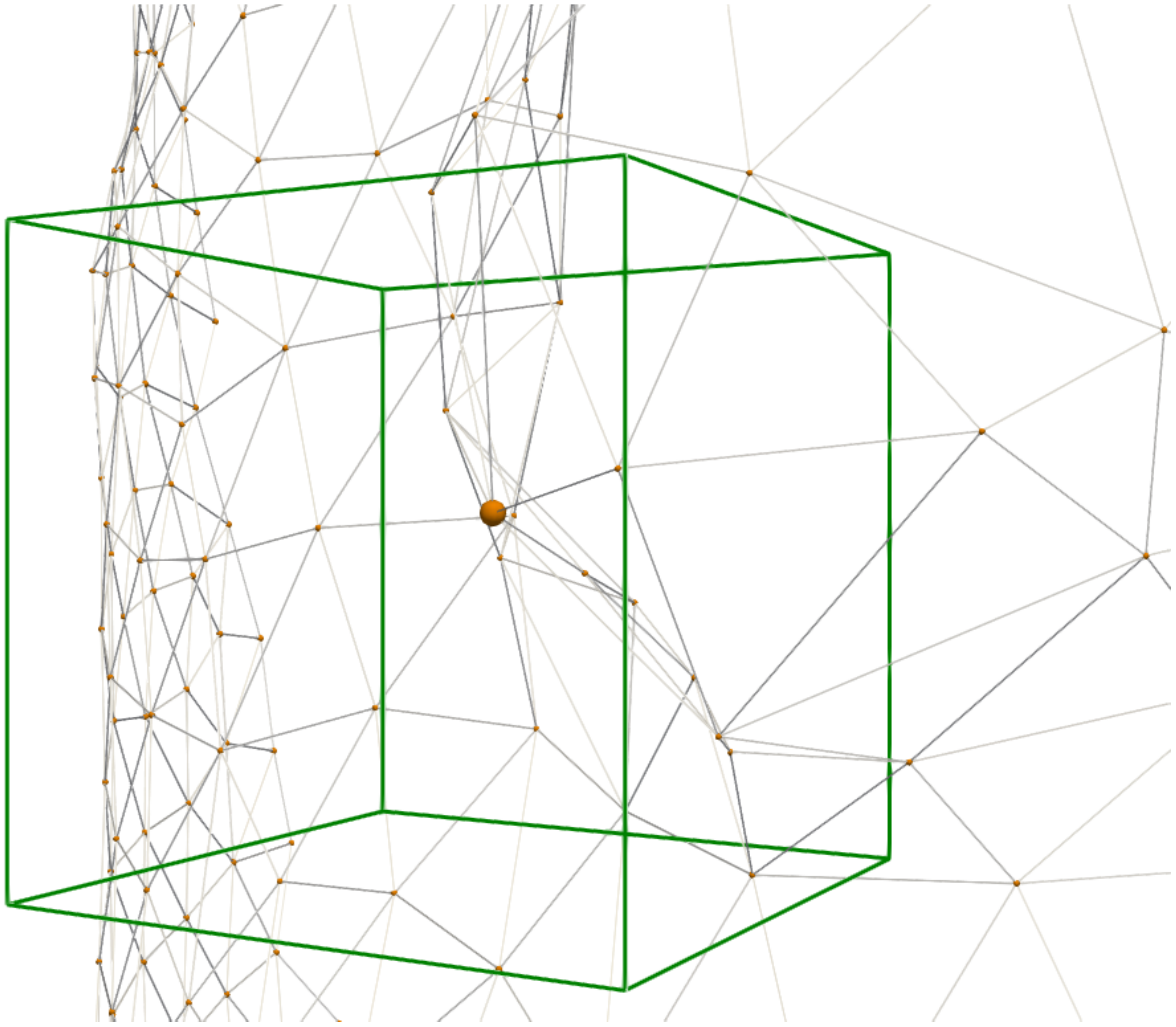


Mesh upsampling

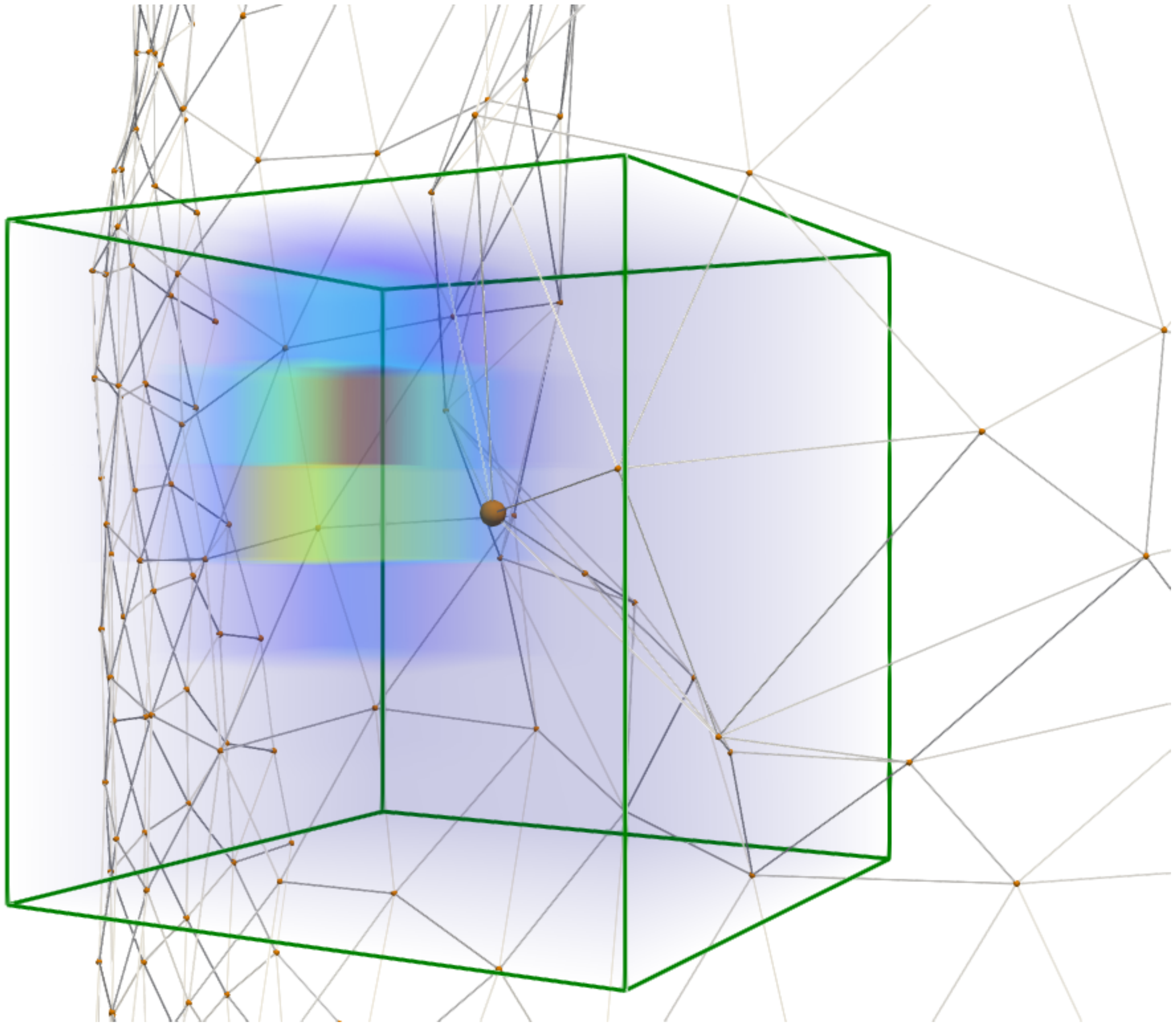
Local Stage: Local Refinement



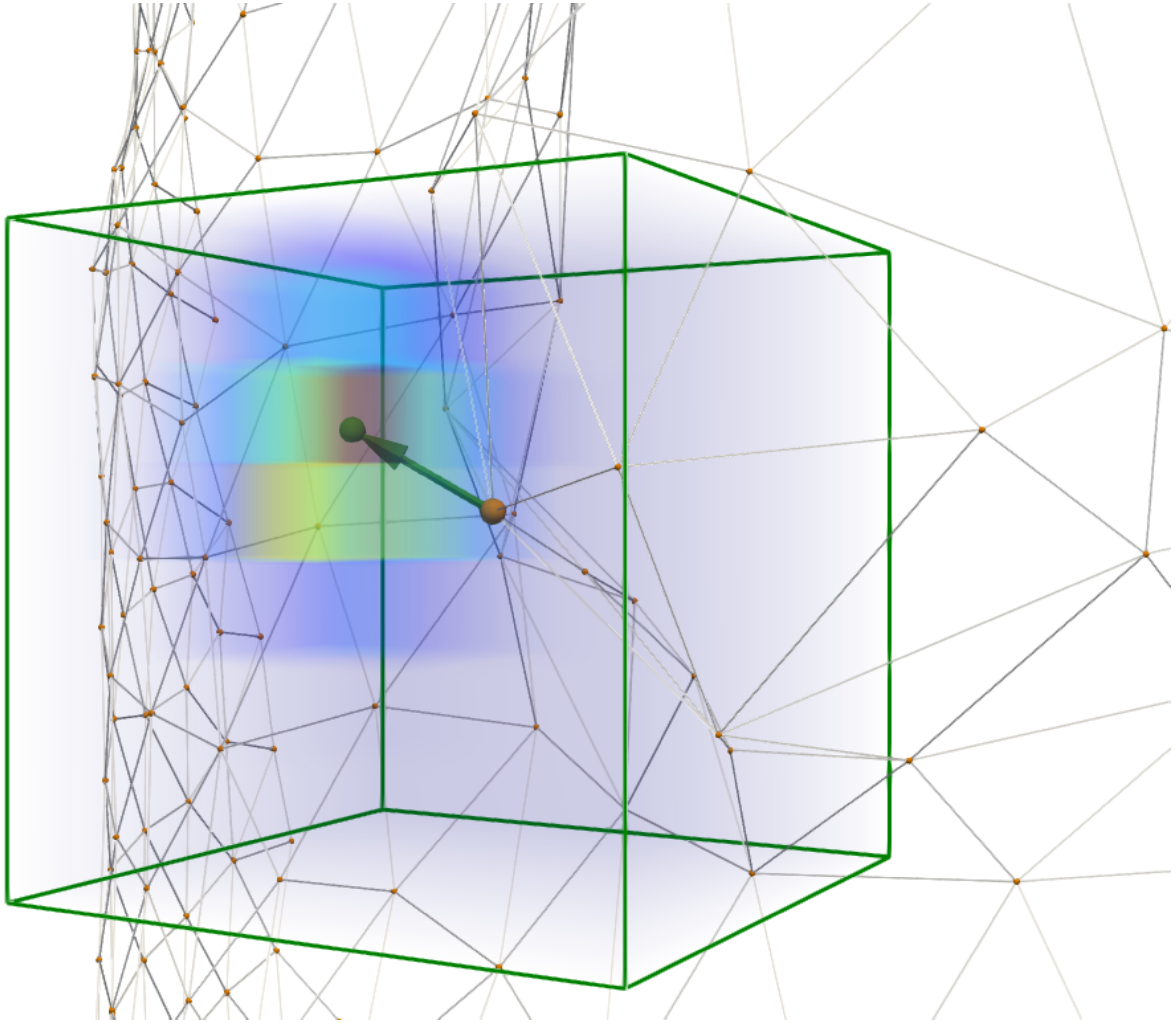
Local Stage: Local Refinement



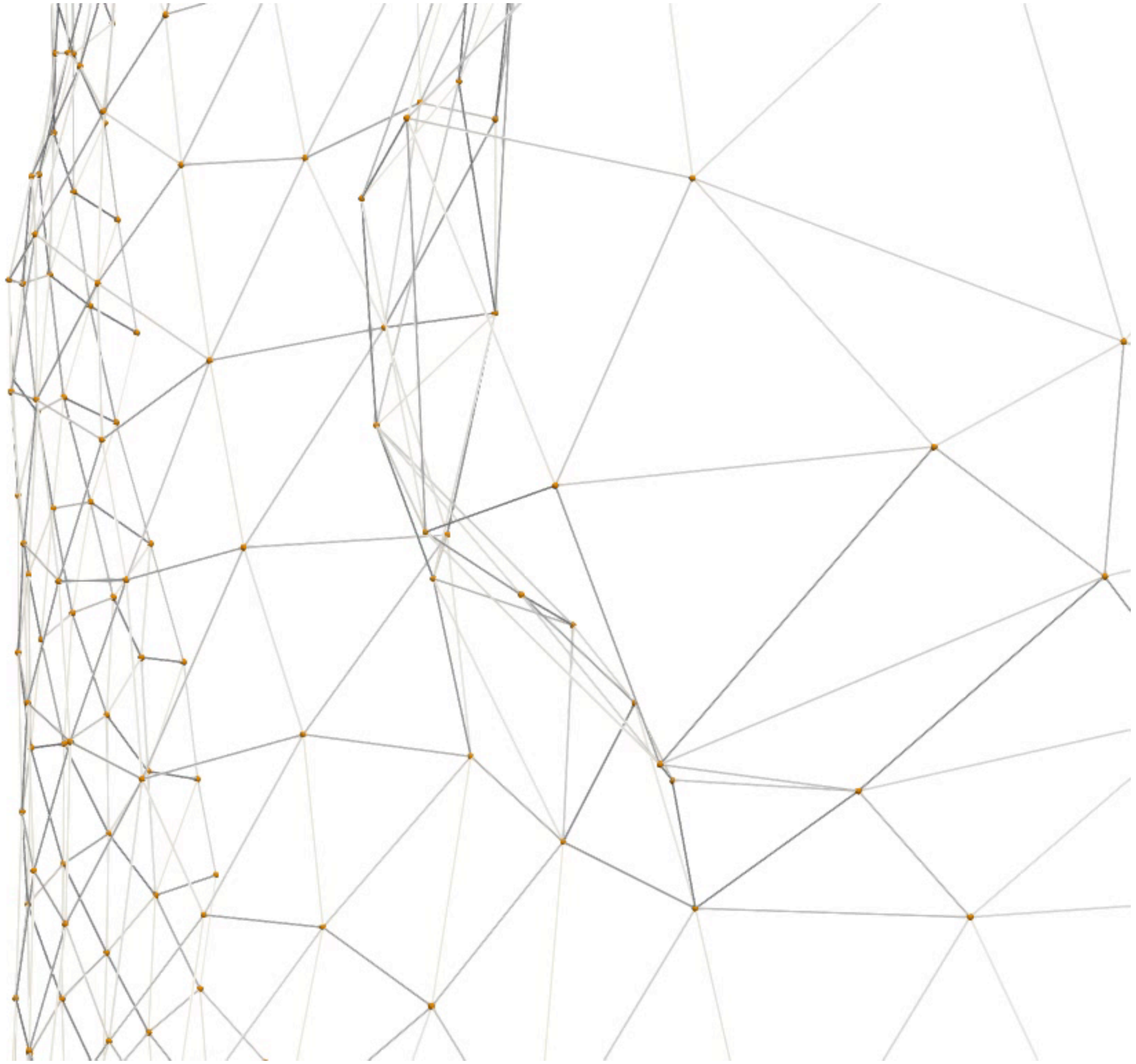
Local Stage: Local Refinement



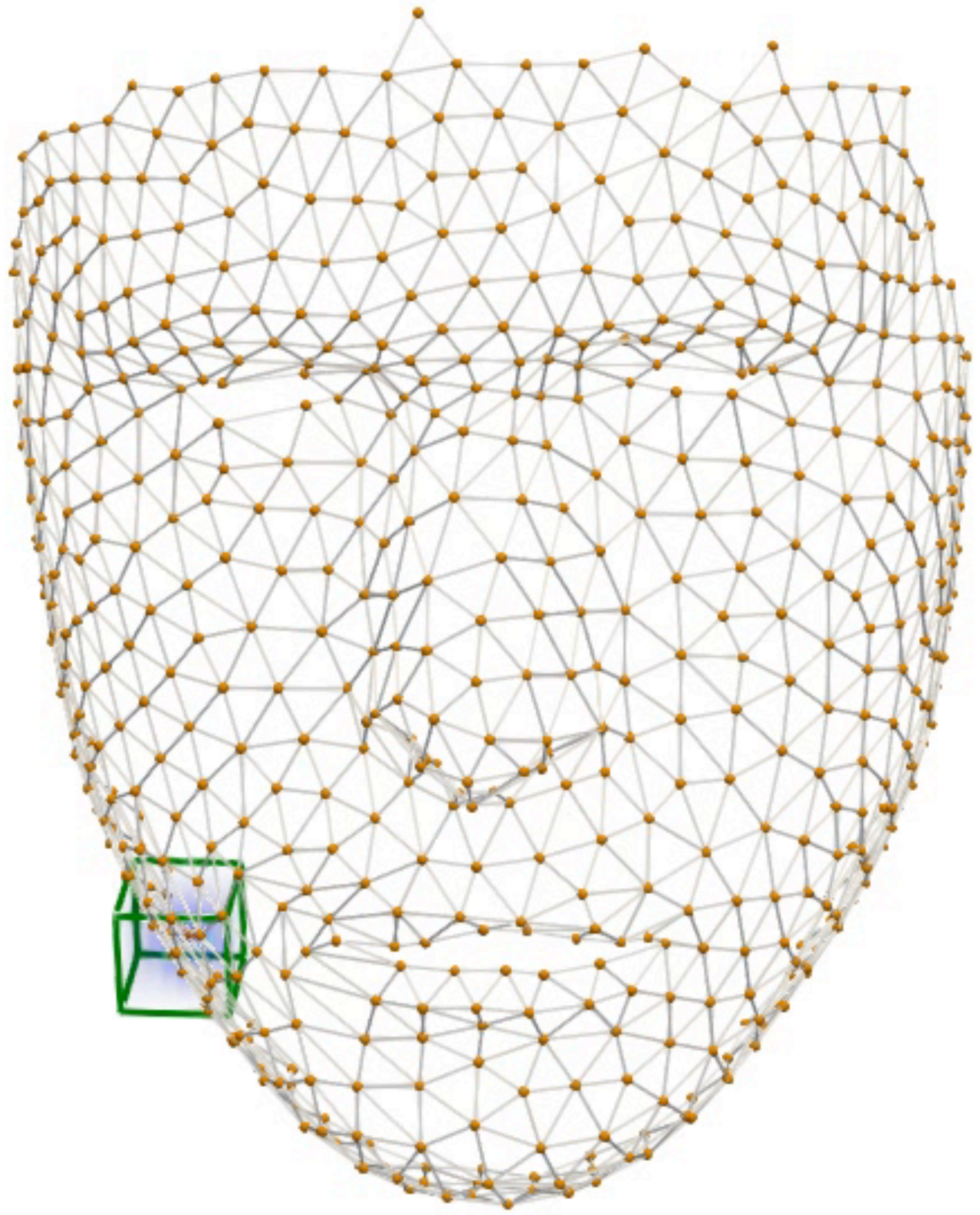
Local Stage: Local Refinement



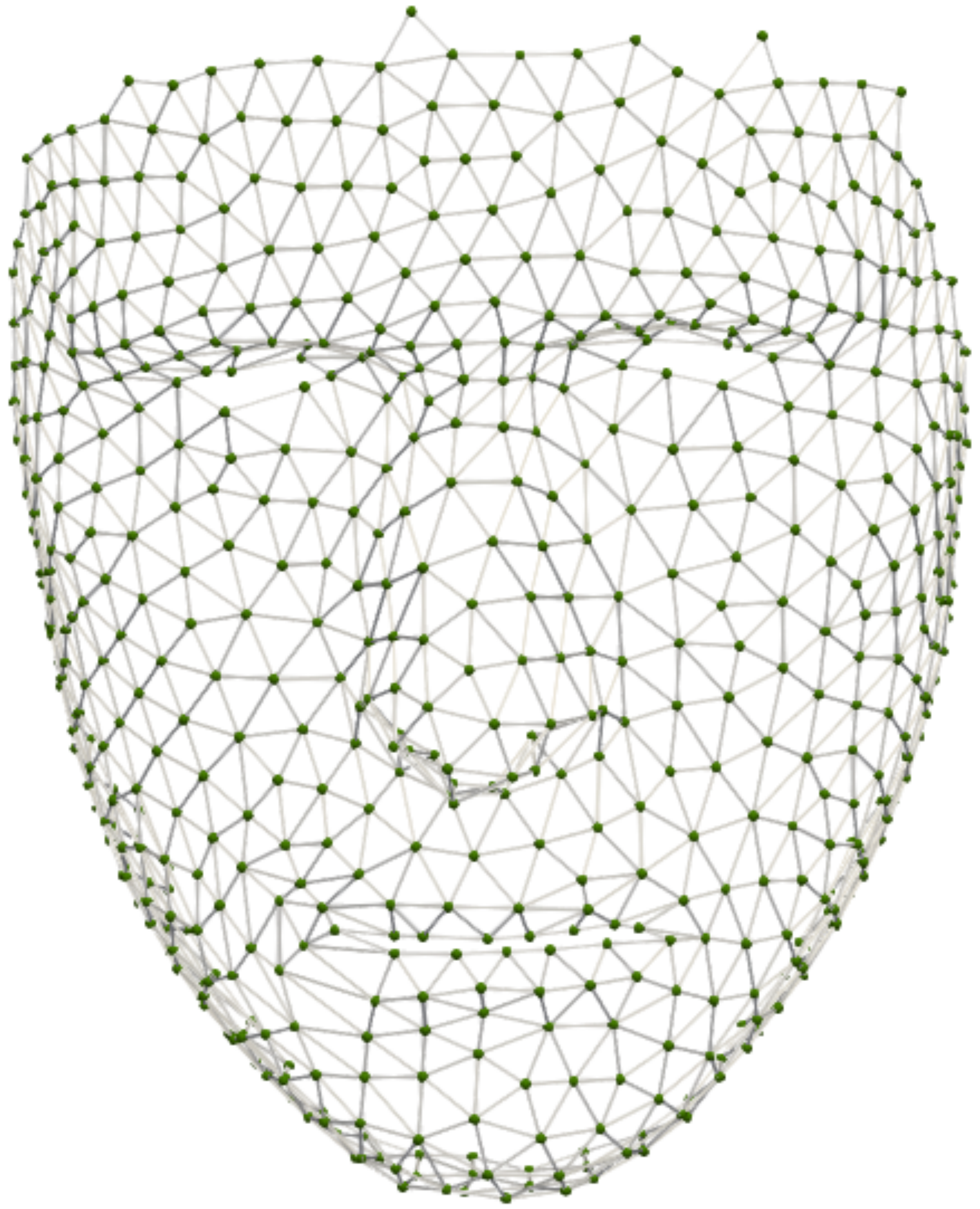
Local Stage: Local Refinement



Local Stage: Local Refinement

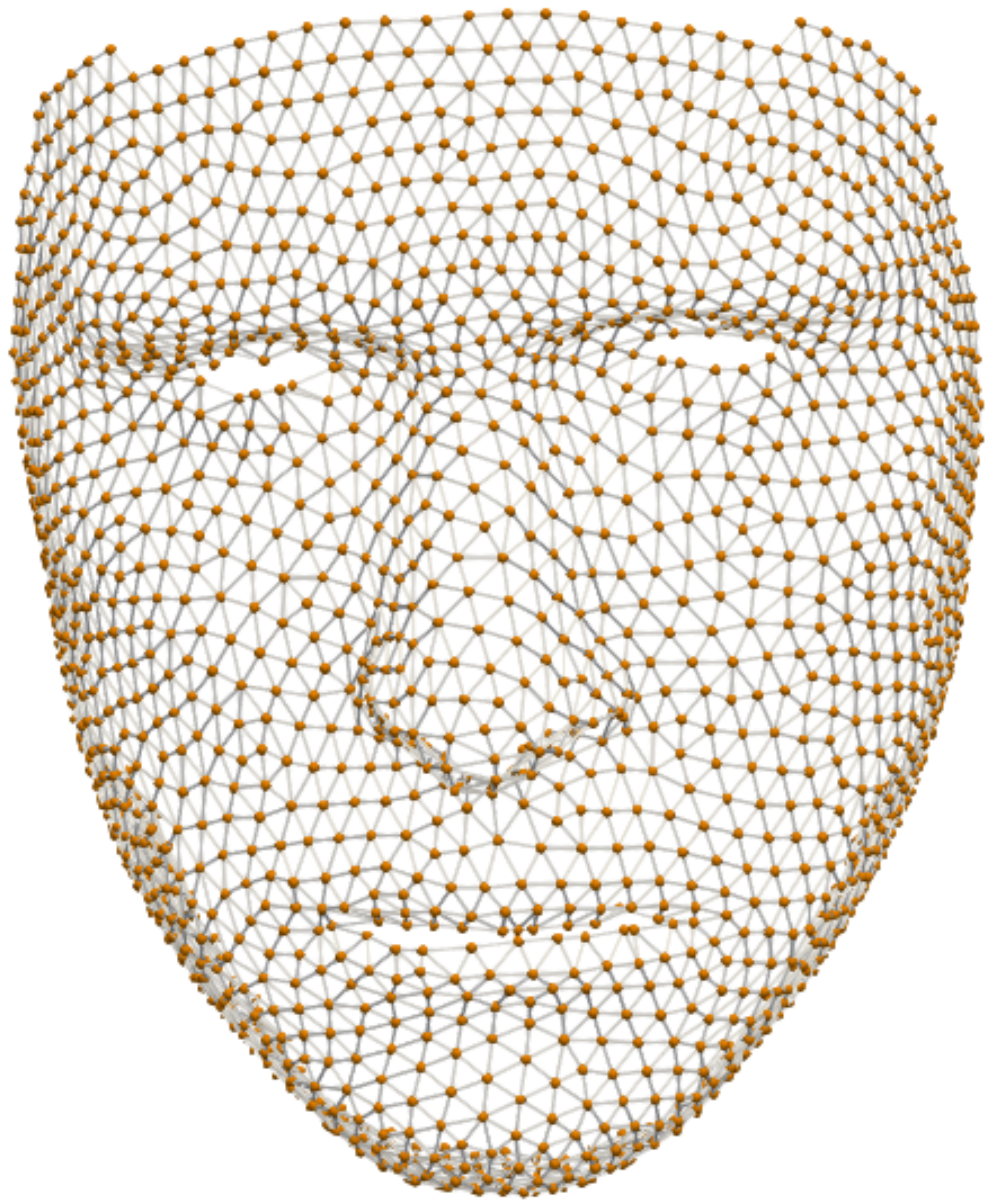


Local Stage: Local Refinement



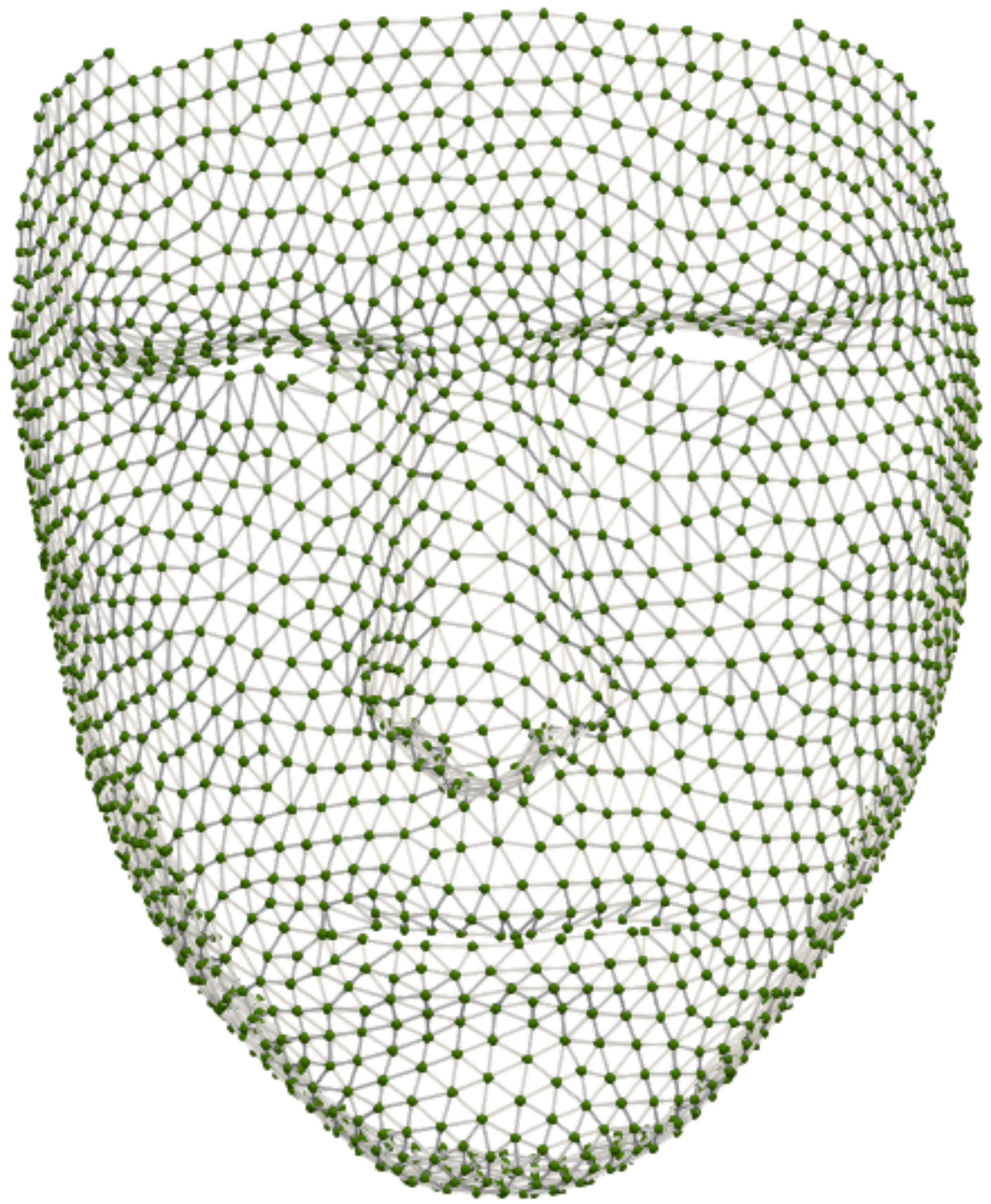
Refined mesh at level 1

Progressive Mesh Generation



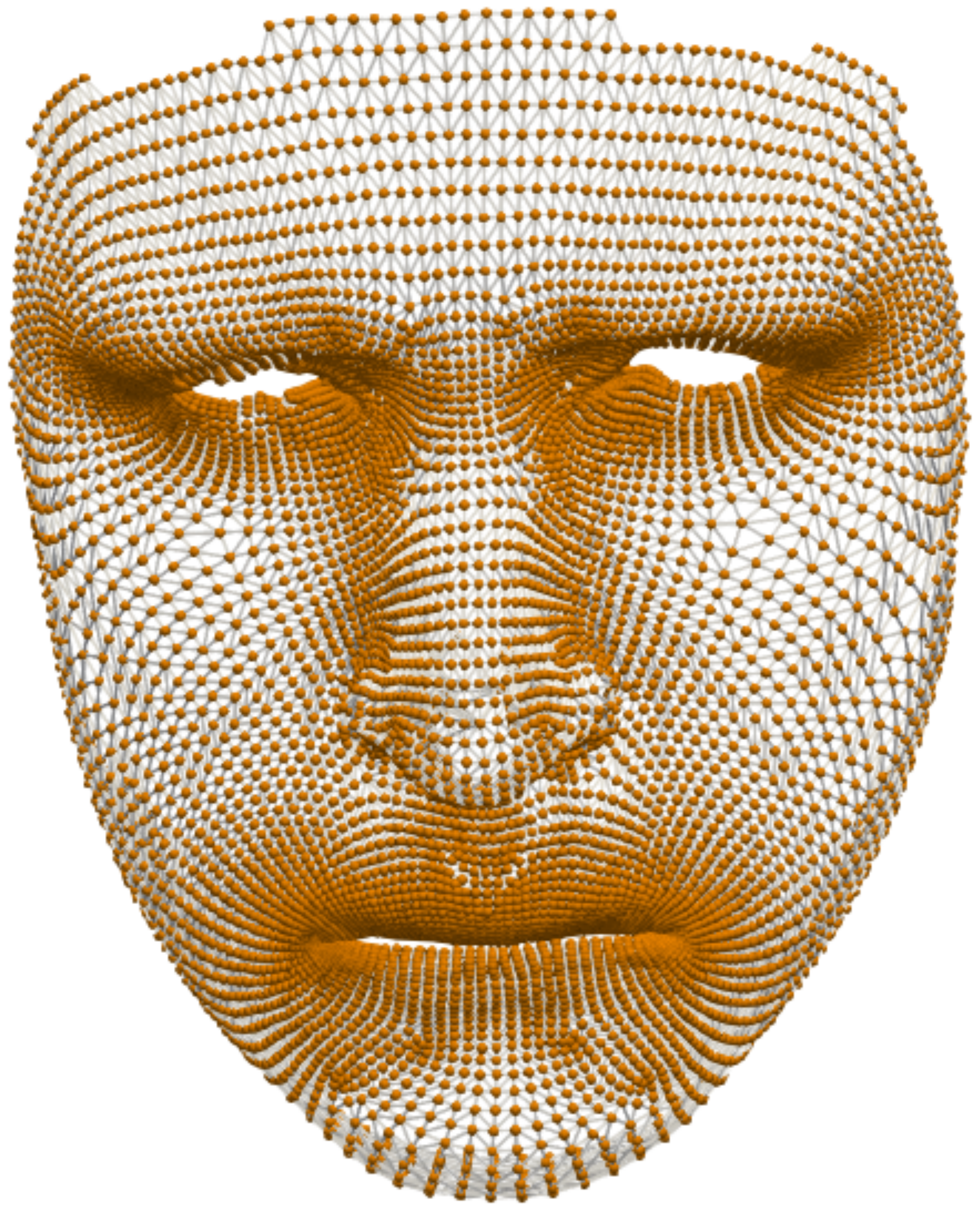
Upsampled mesh at level 2

Progressive Mesh Generation



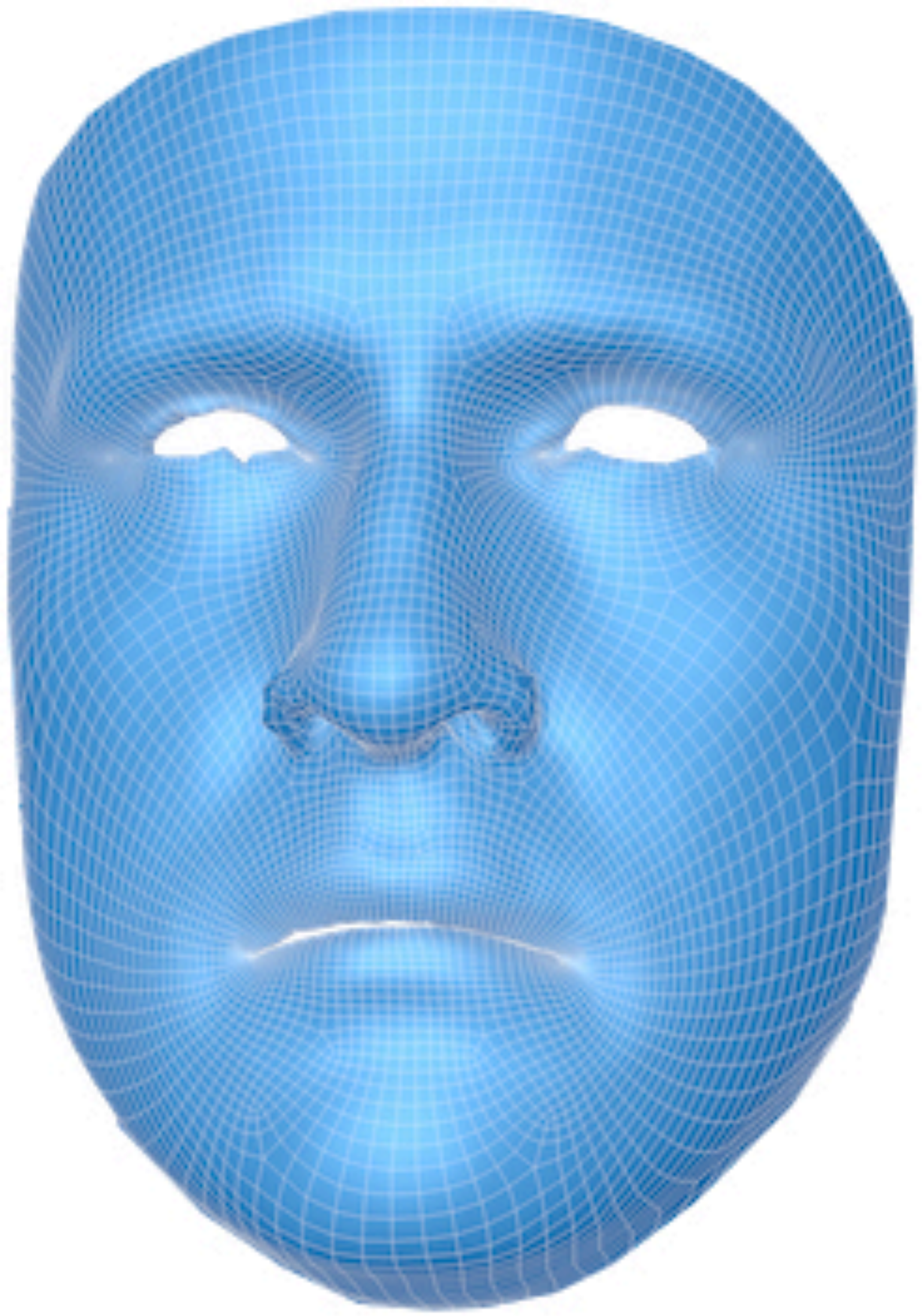
Refined mesh at level 2

Progressive Mesh Generation



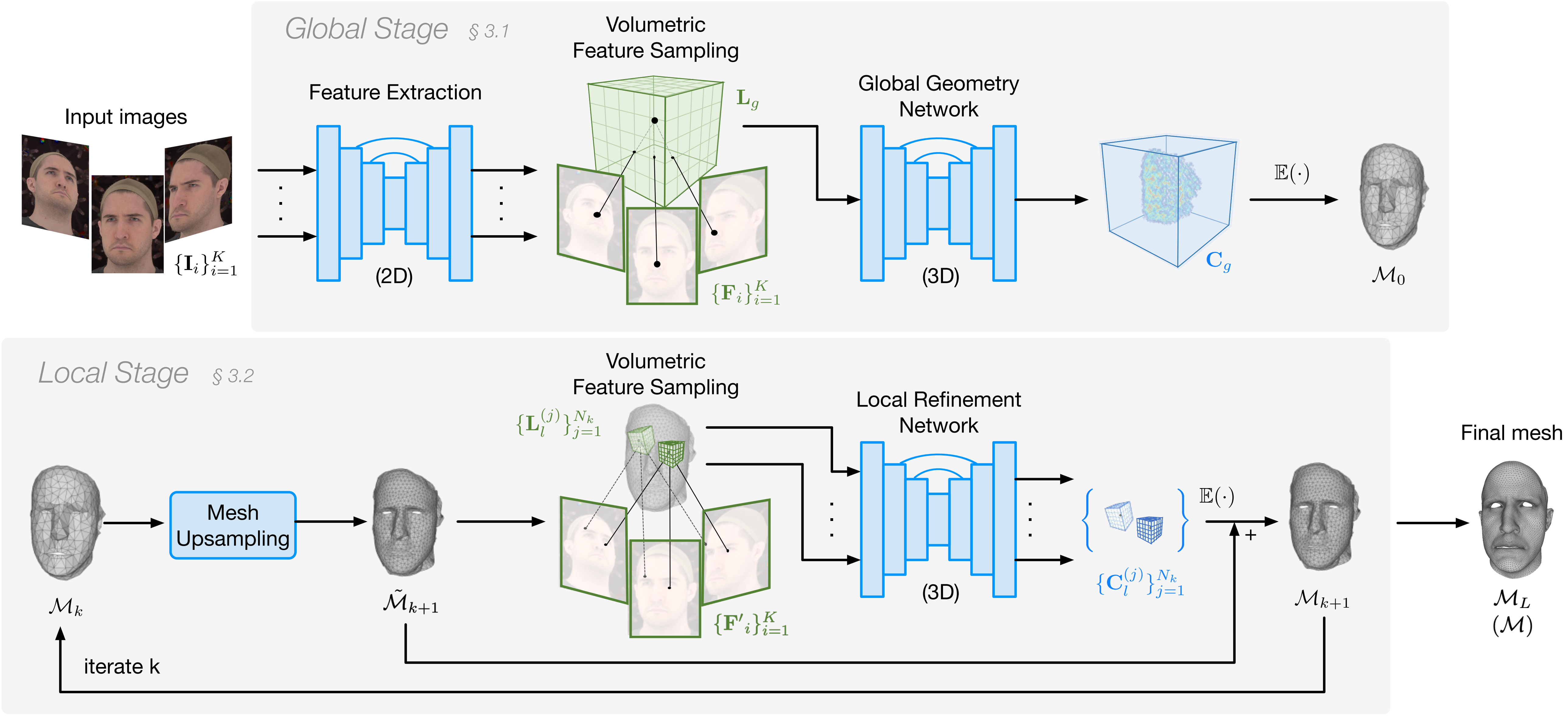
Upsampled mesh at level 3

Progressive Mesh Generation



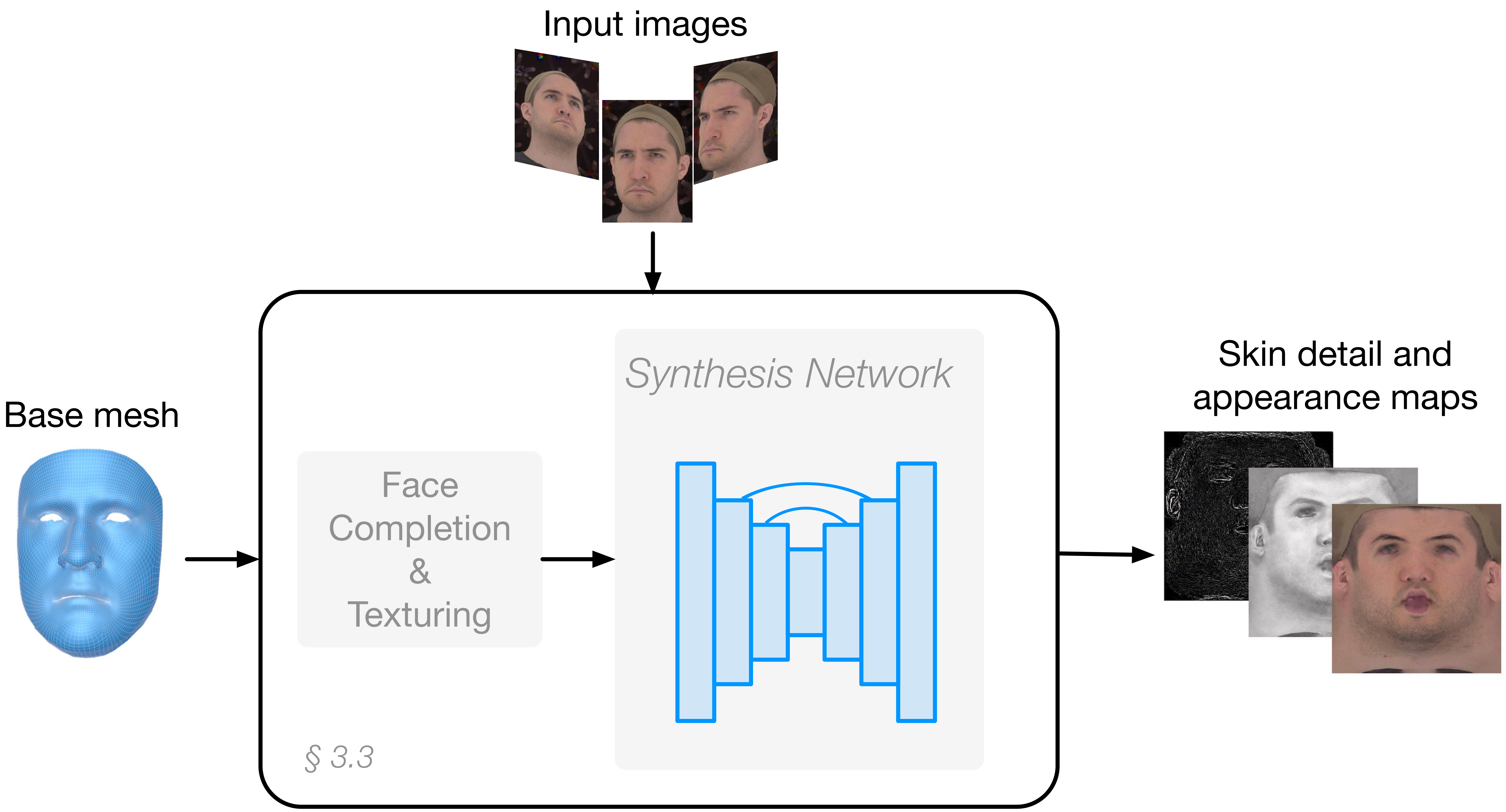
Final prediction

Training

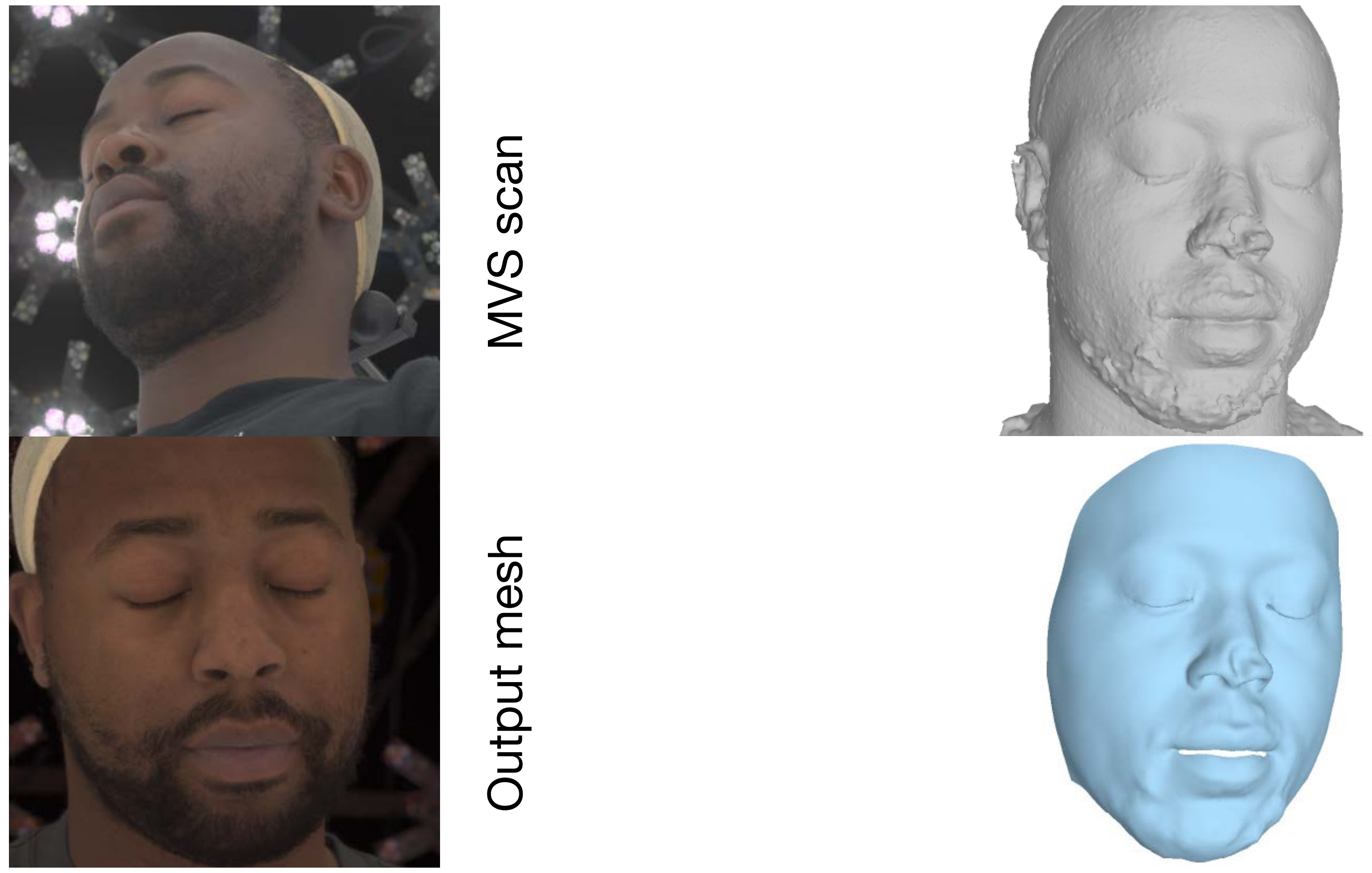


- The progressive mesh generation is differentiable end-to-end
- Fully supervised training
- Data: >1,200 curated ground truth (Li, Bladin and Zhao et al. CVPR 20)

Appearance and Detail Capture

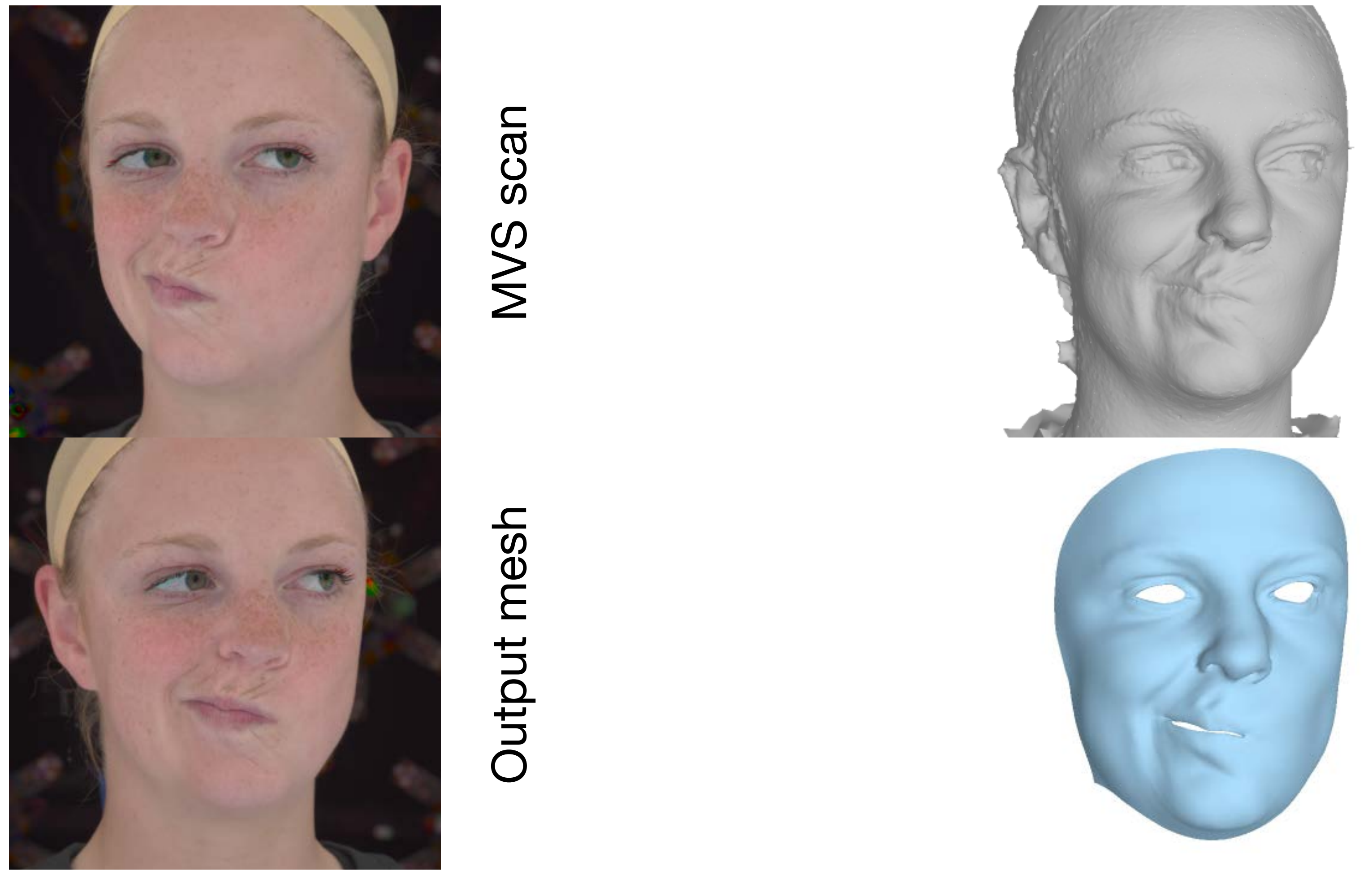


Results: Robustness



Input images
(2 of 15)

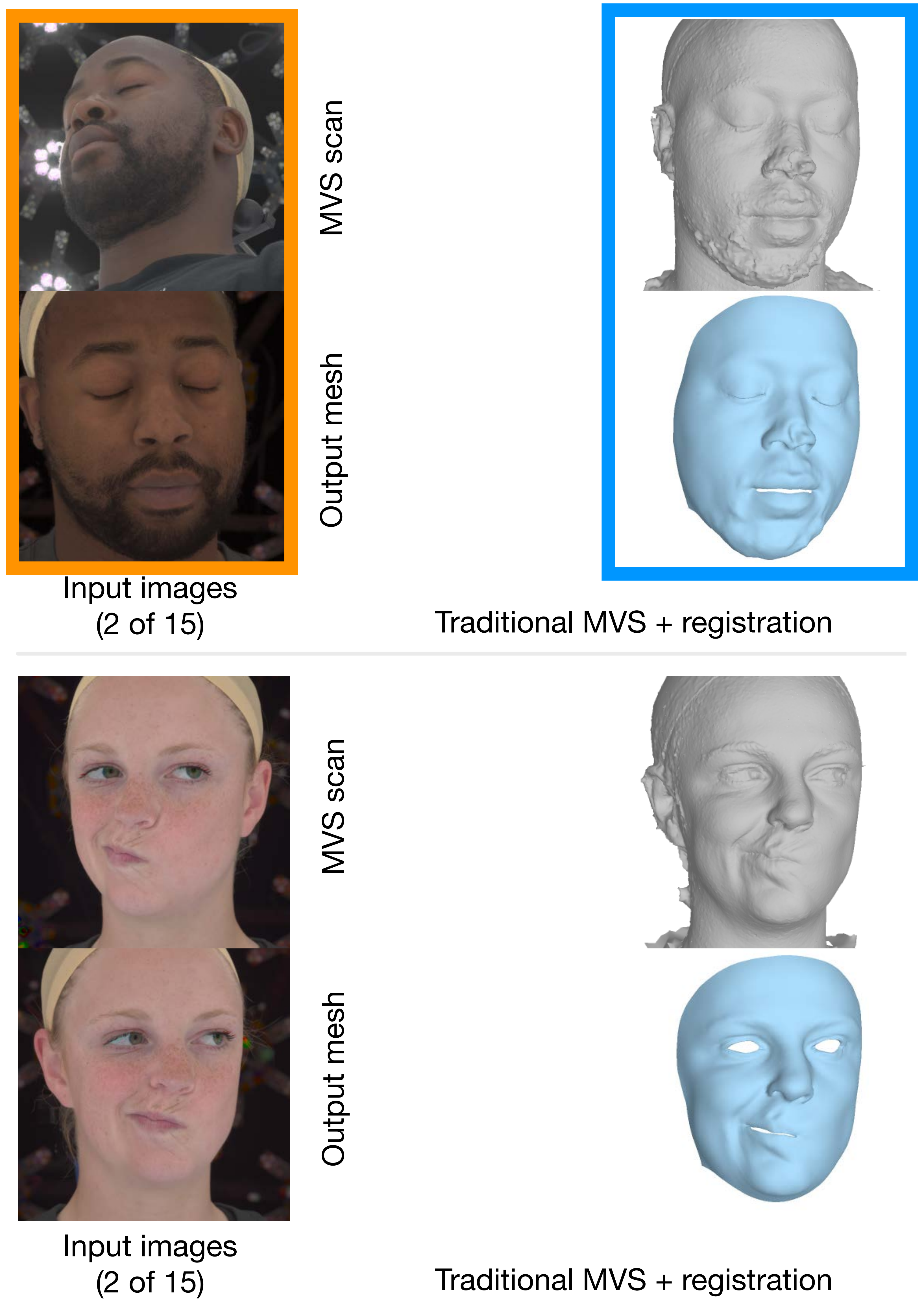
Traditional MVS + registration

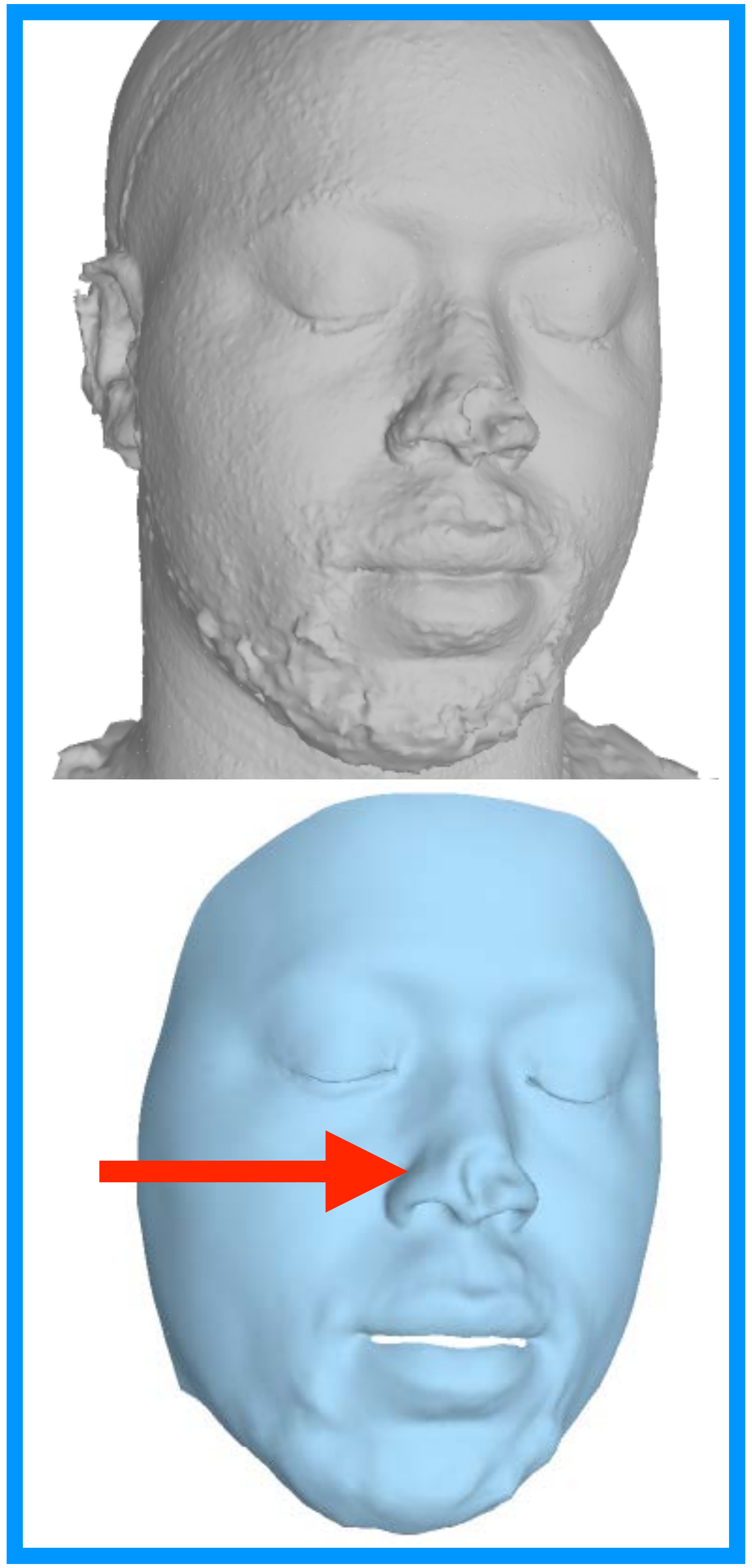


Input images
(2 of 15)

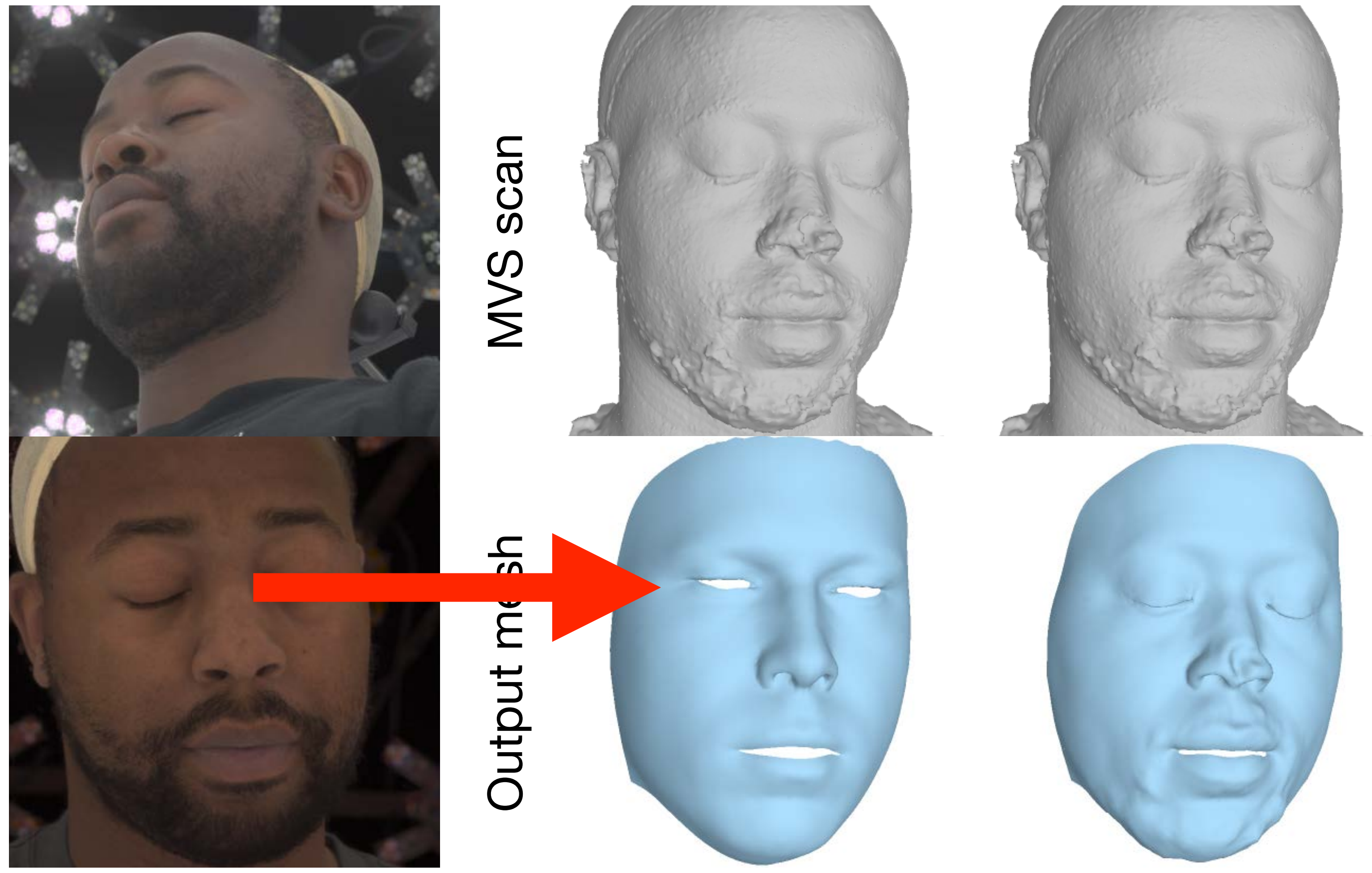
Traditional MVS + registration

Results: Robustness

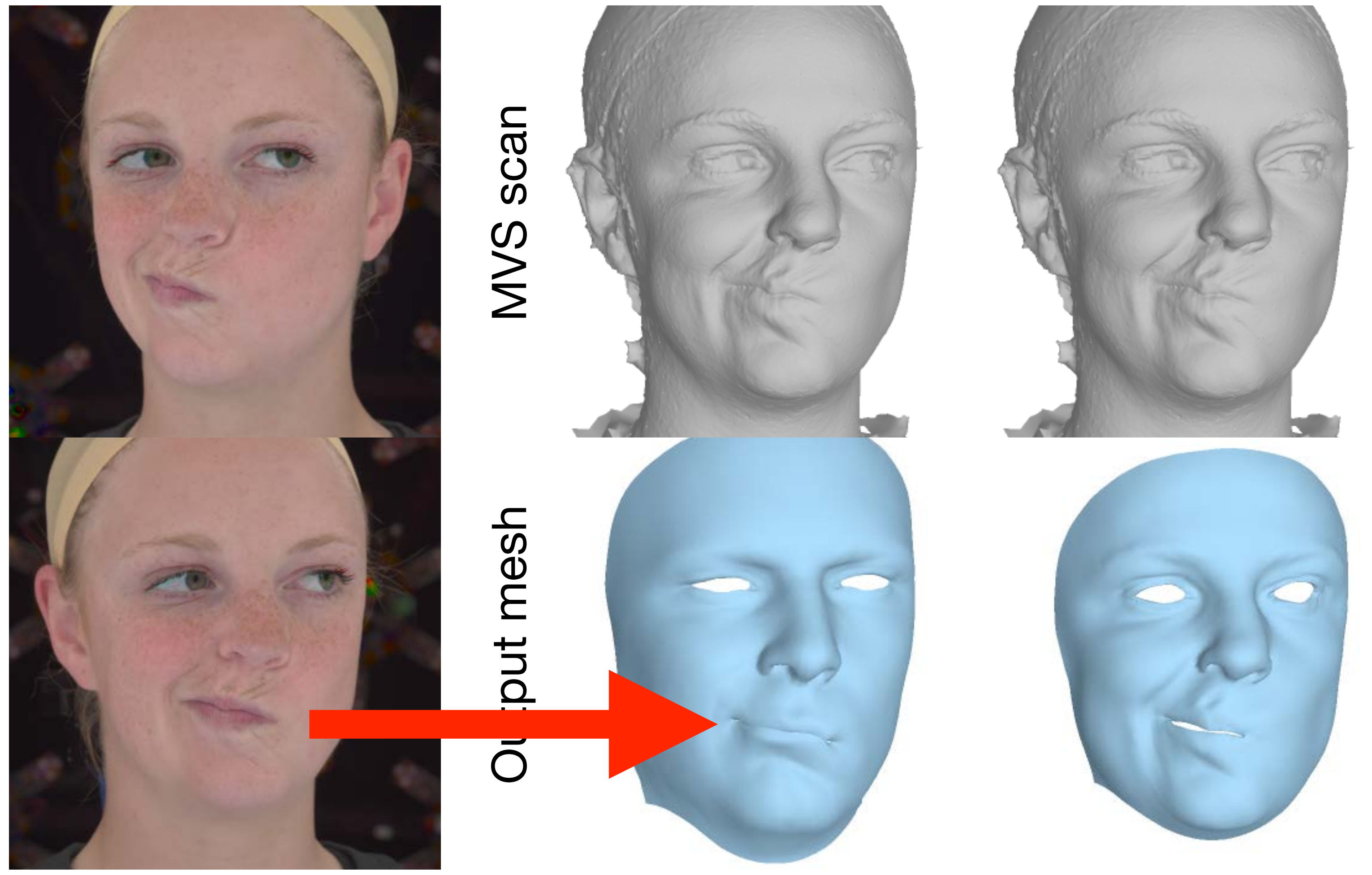




Results: Robustness

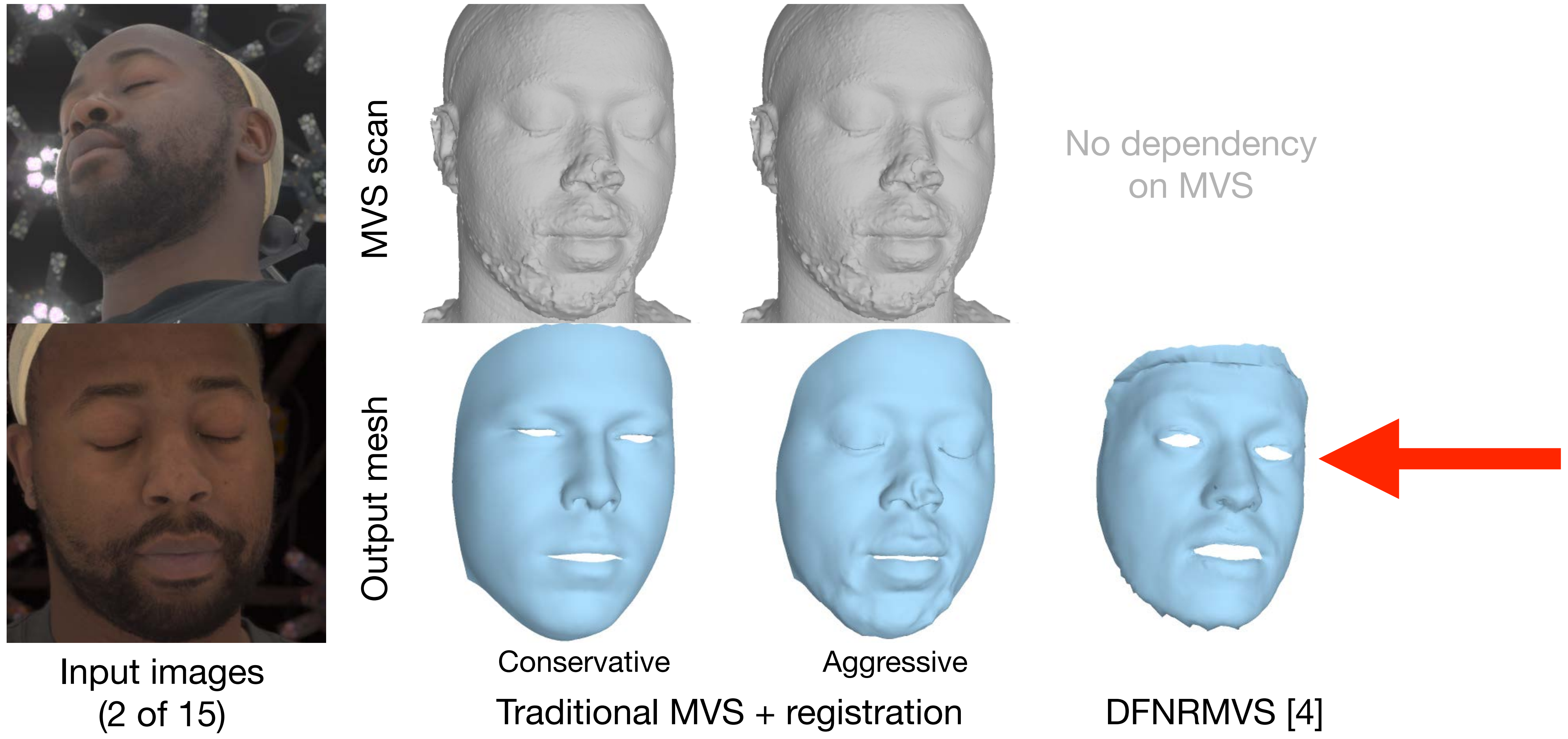


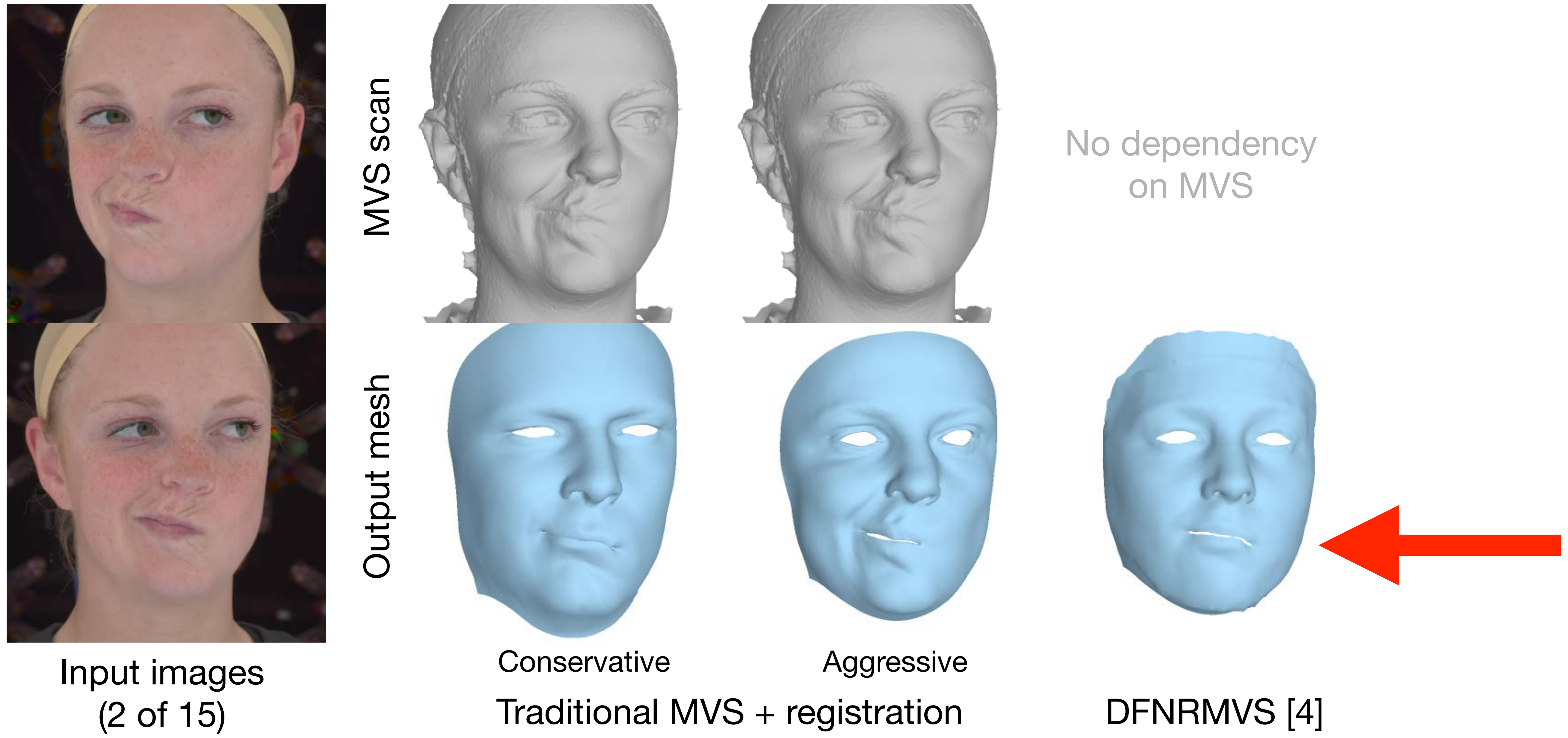
Input images
(2 of 15)
Conservative
Aggressive
Traditional MVS + registration



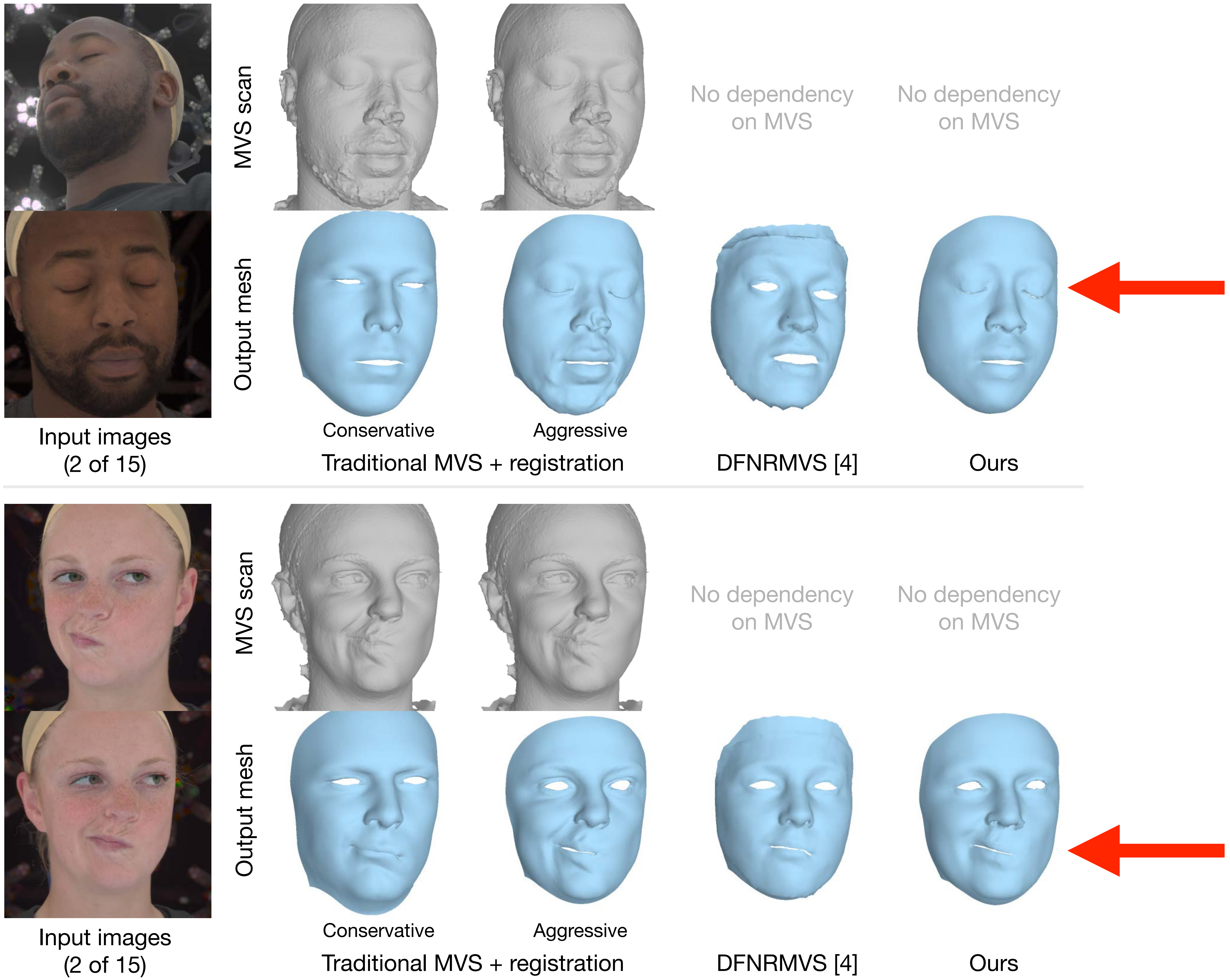
Input images
(2 of 15)
Conservative
Aggressive
Traditional MVS + registration

Results: Robustness

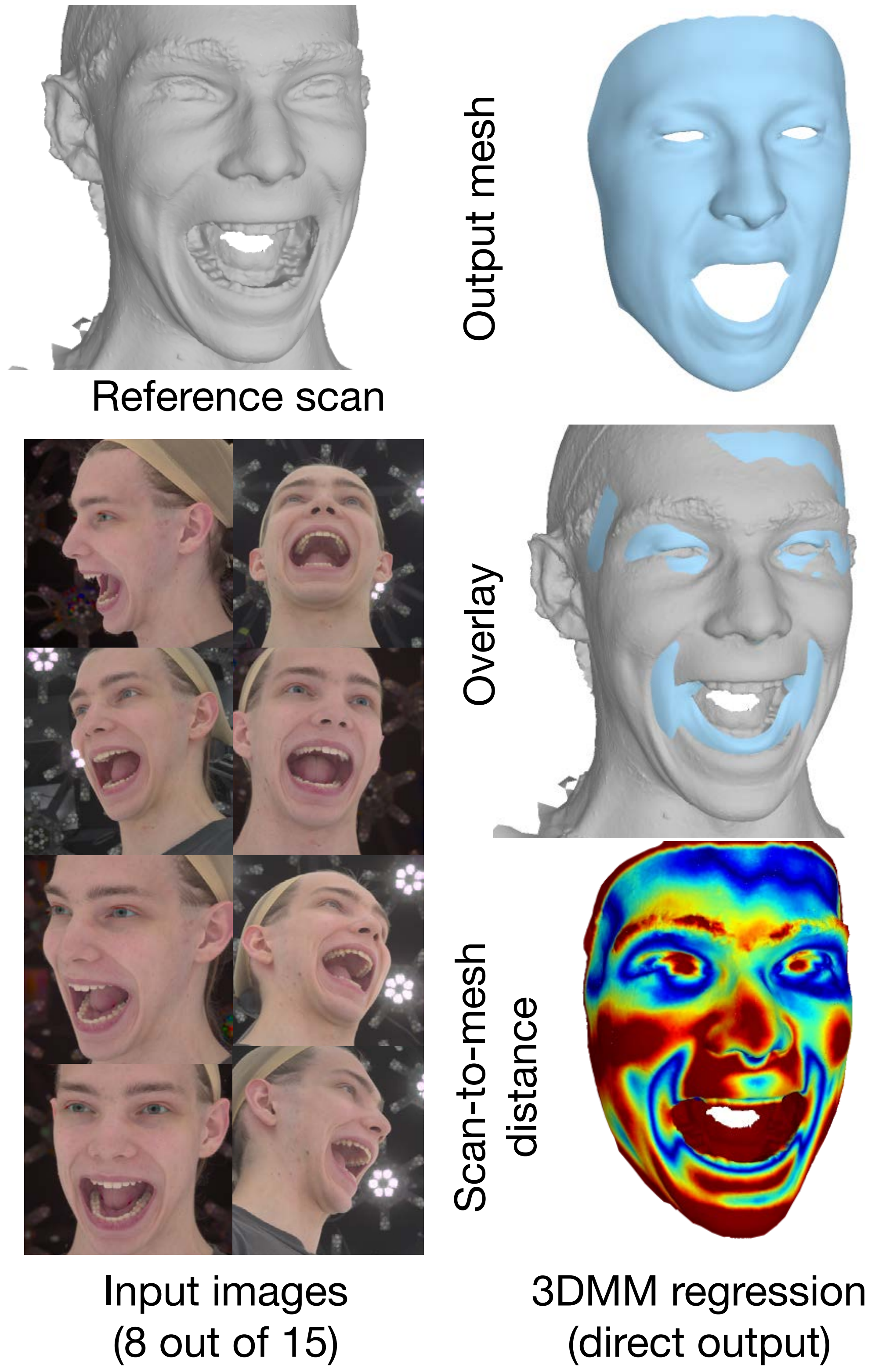


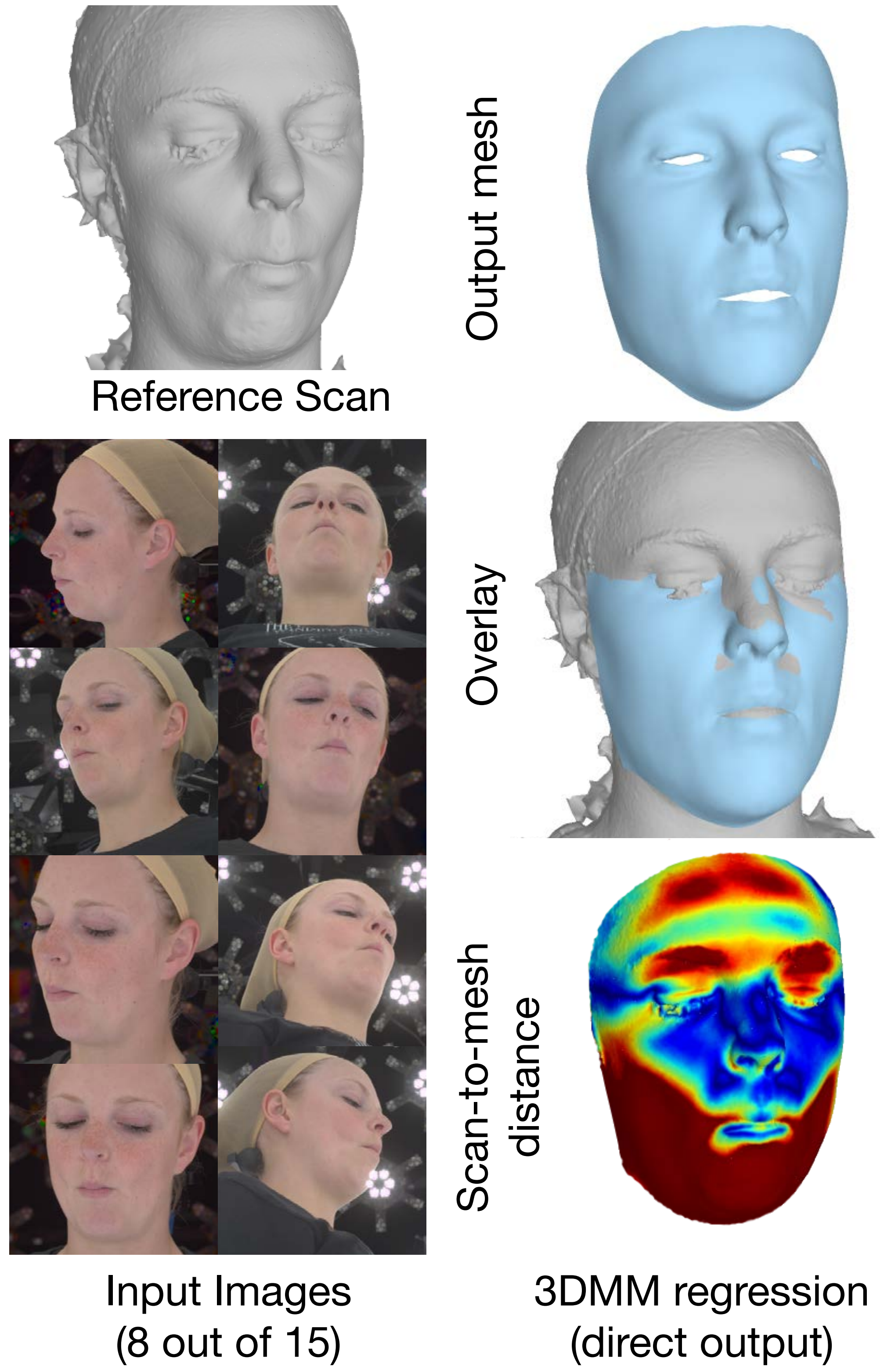


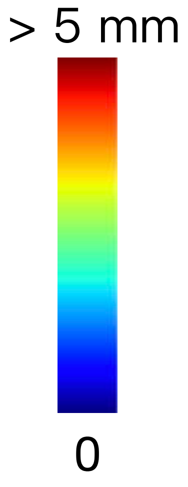
Results: Robustness



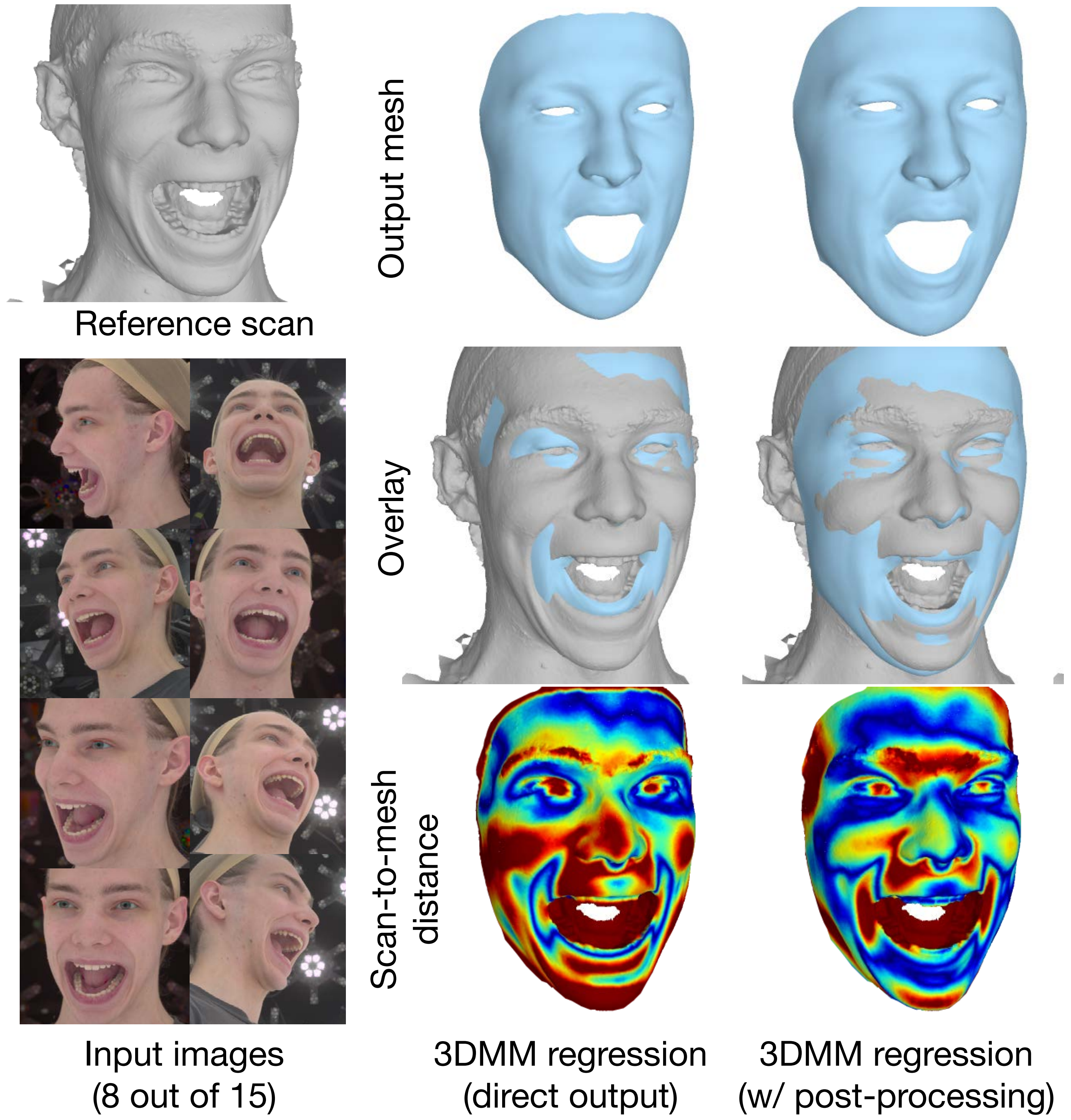
Results: Geometric Accuracy

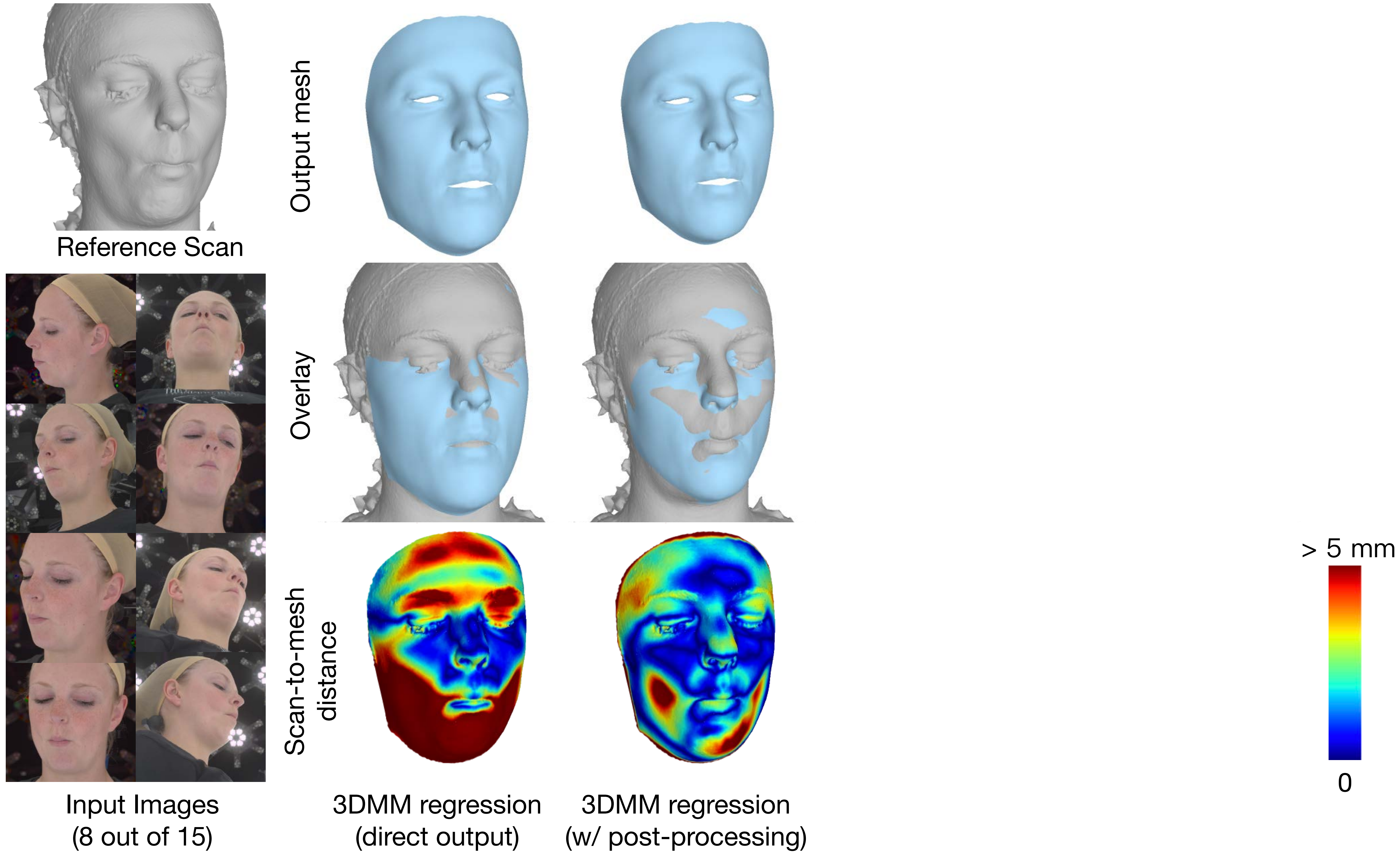




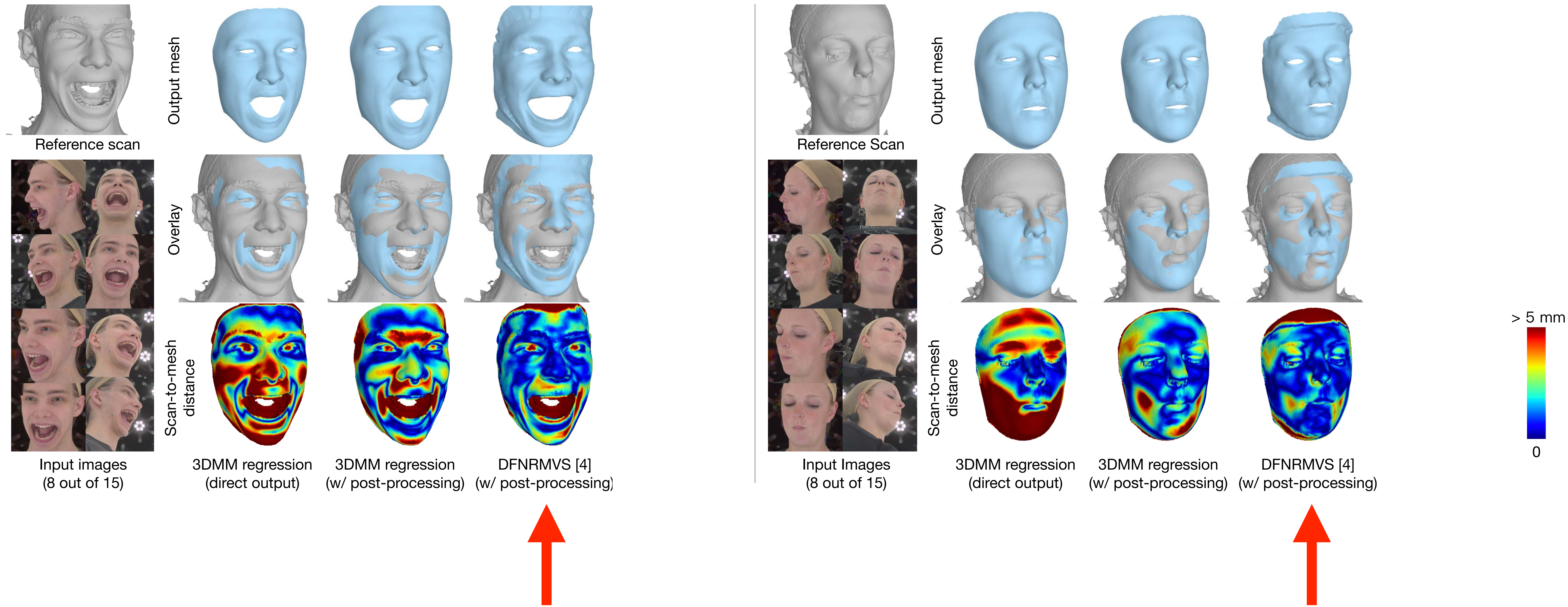


Results: Geometric Accuracy

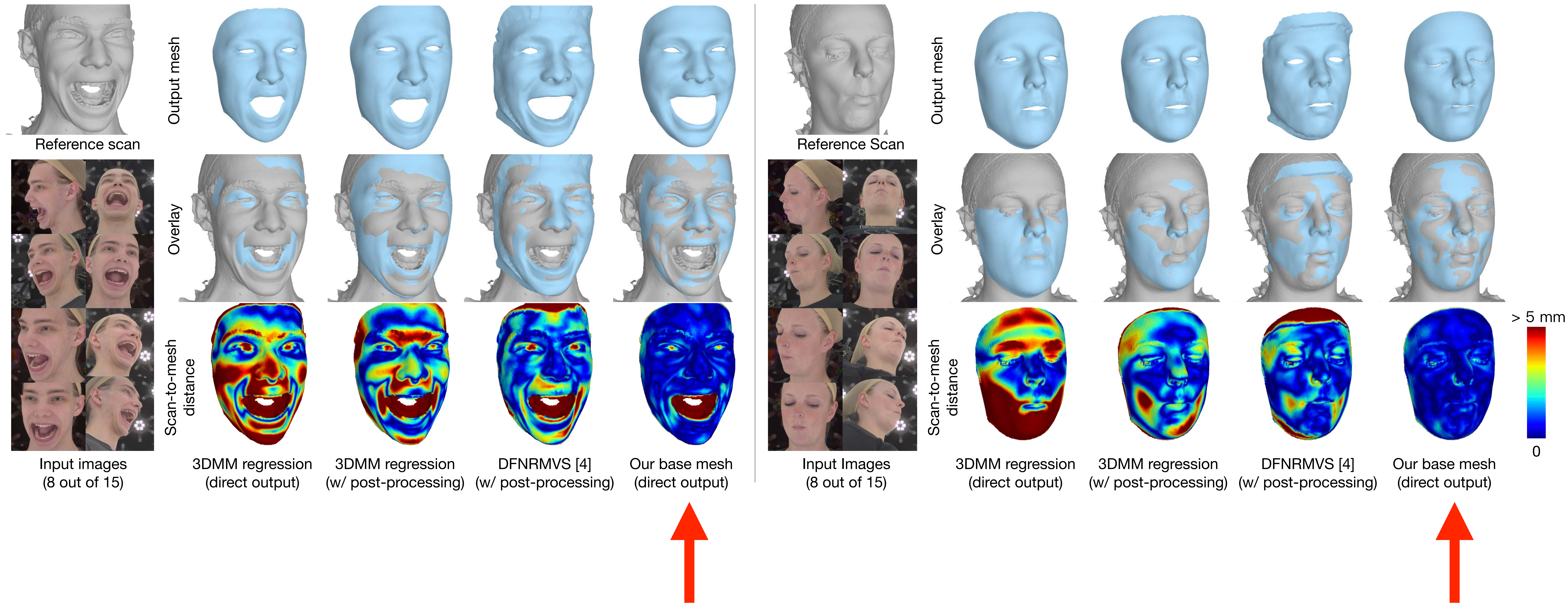




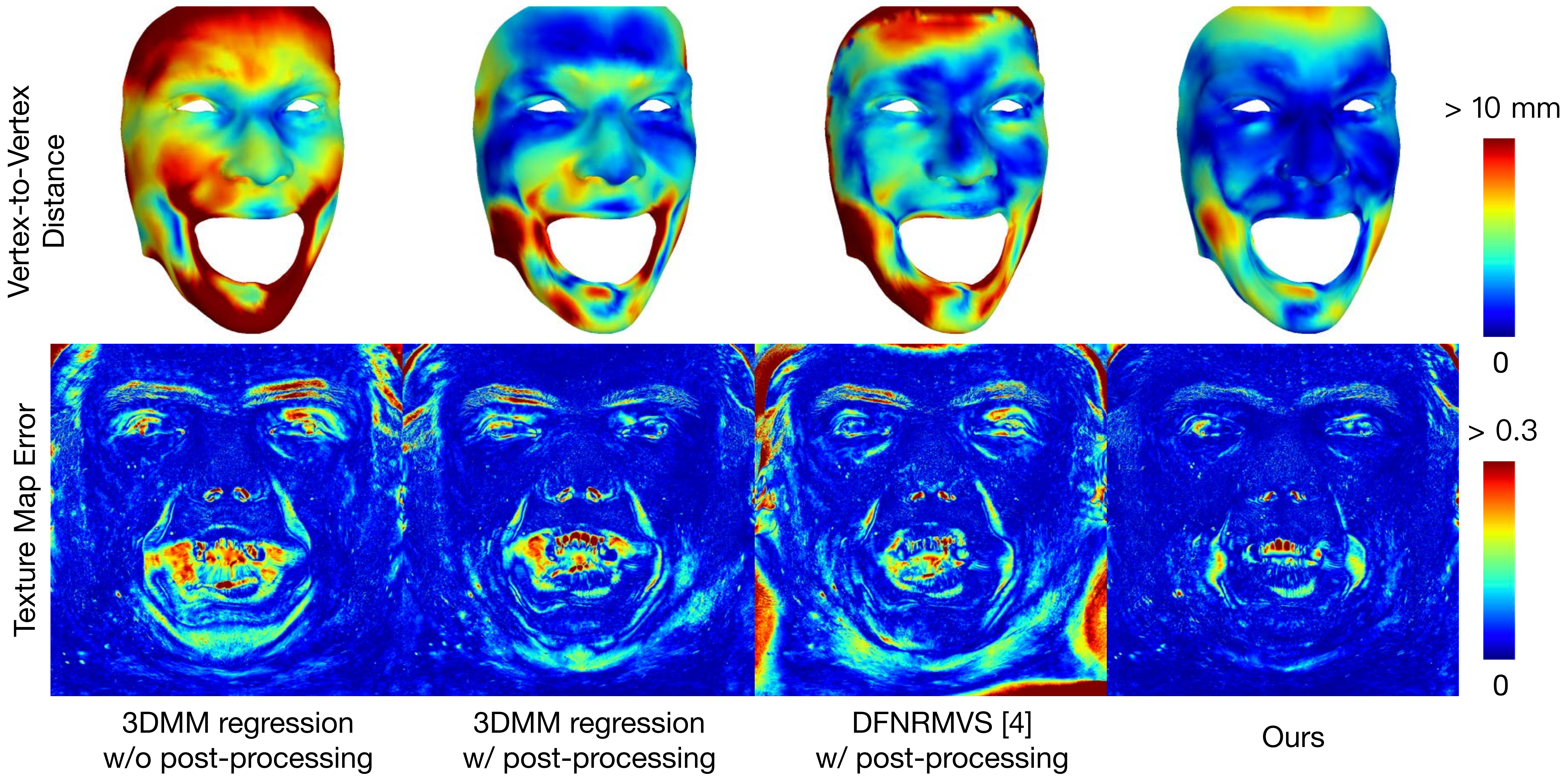
Results: Geometric Accuracy



Results: Geometric Accuracy



Results: Correspondence Accuracy



Results: Speed

Methods	Time	Automatic
Traditional pipeline	600+	✗
DFNRMVS [4]	4.5	✓
ToFu (base mesh)	0.385	✓

(Measured in seconds)

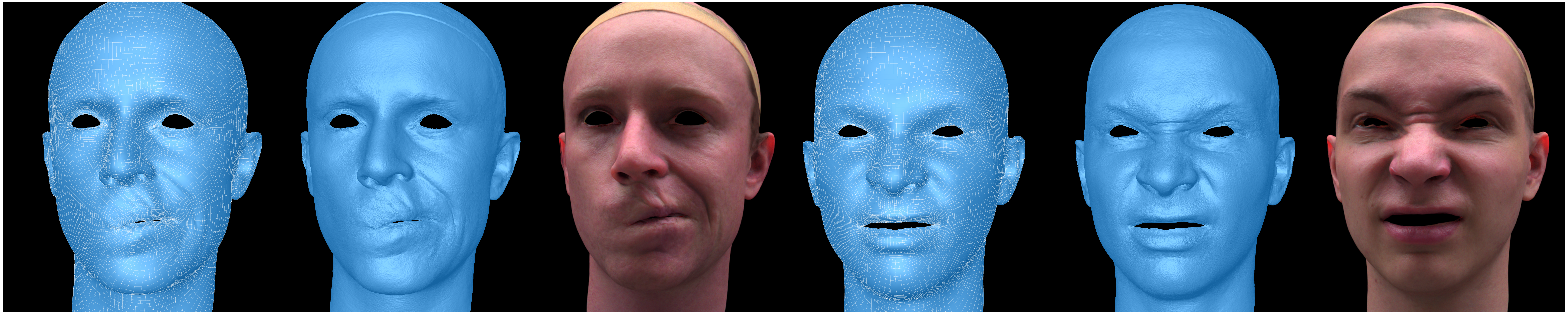
Our method achieves 2~3 orders of magnitude faster runtime.

Results: Dynamic Performance Capture

Inferred topologically consistent meshes (output)

Per-frame results without temporal smoothing

Results: Detailed Appearance



Base mesh Detailed mesh With full attributes Base mesh Detailed mesh With full attributes

Limitation and Future Work

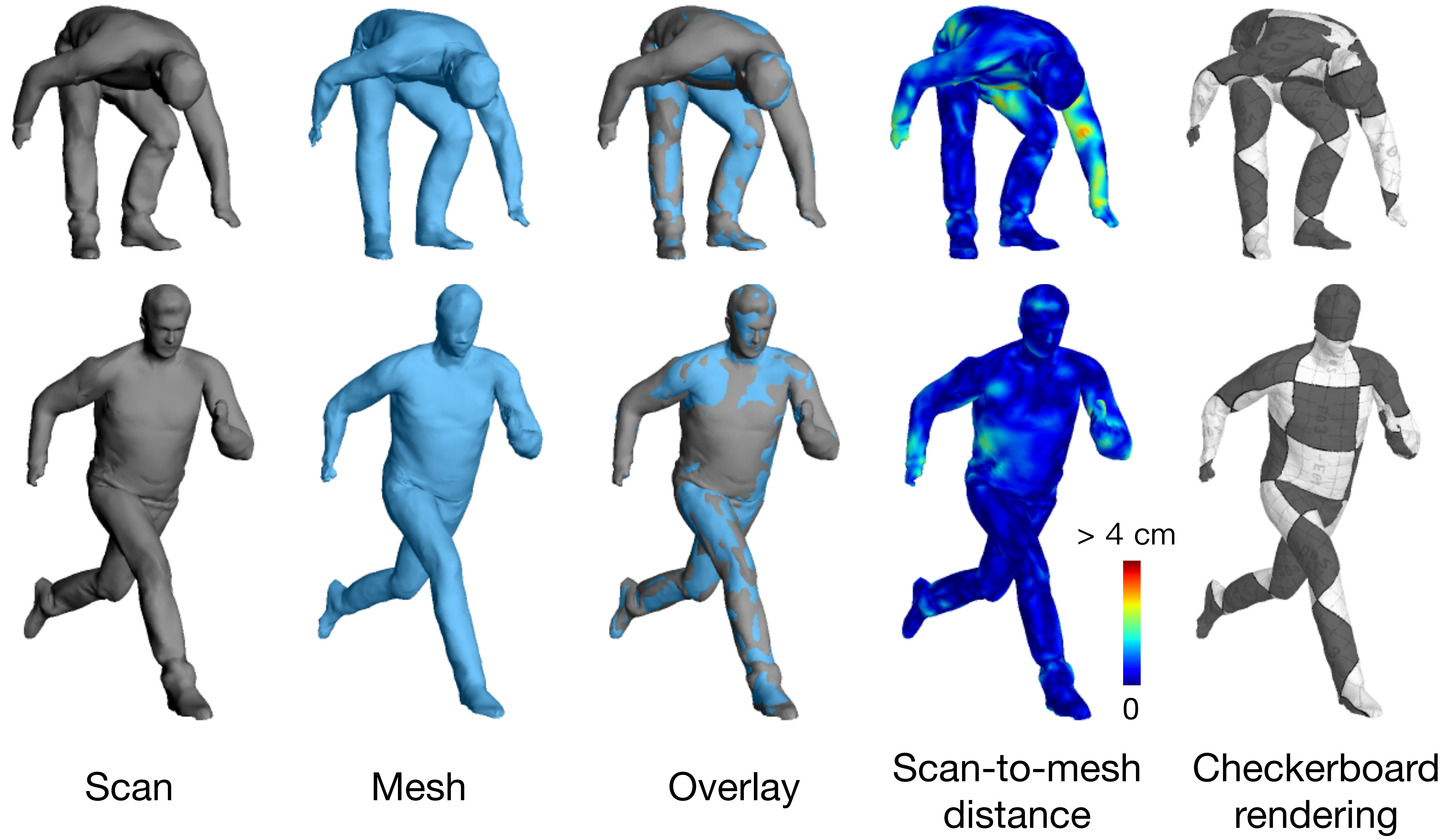
Our method can work on a new capture setup (CoMA datasets) with fine-tuning.

Fewer requirements on data and capture system

Image credit:
Right (dataset): Li, Bladin and Zhao et al. "Learning formation of physically based face attributes", CVPR 2020

Limitation and Future Work

Complete head model
(with hair, eyeballs, etc.)



Other objects
(e.g. clothed humans or general scenes)

Image credit:
Left: Lombardi et al,. Learning Dynamic Renderable Volumes from Images, SIGGRAPH 19

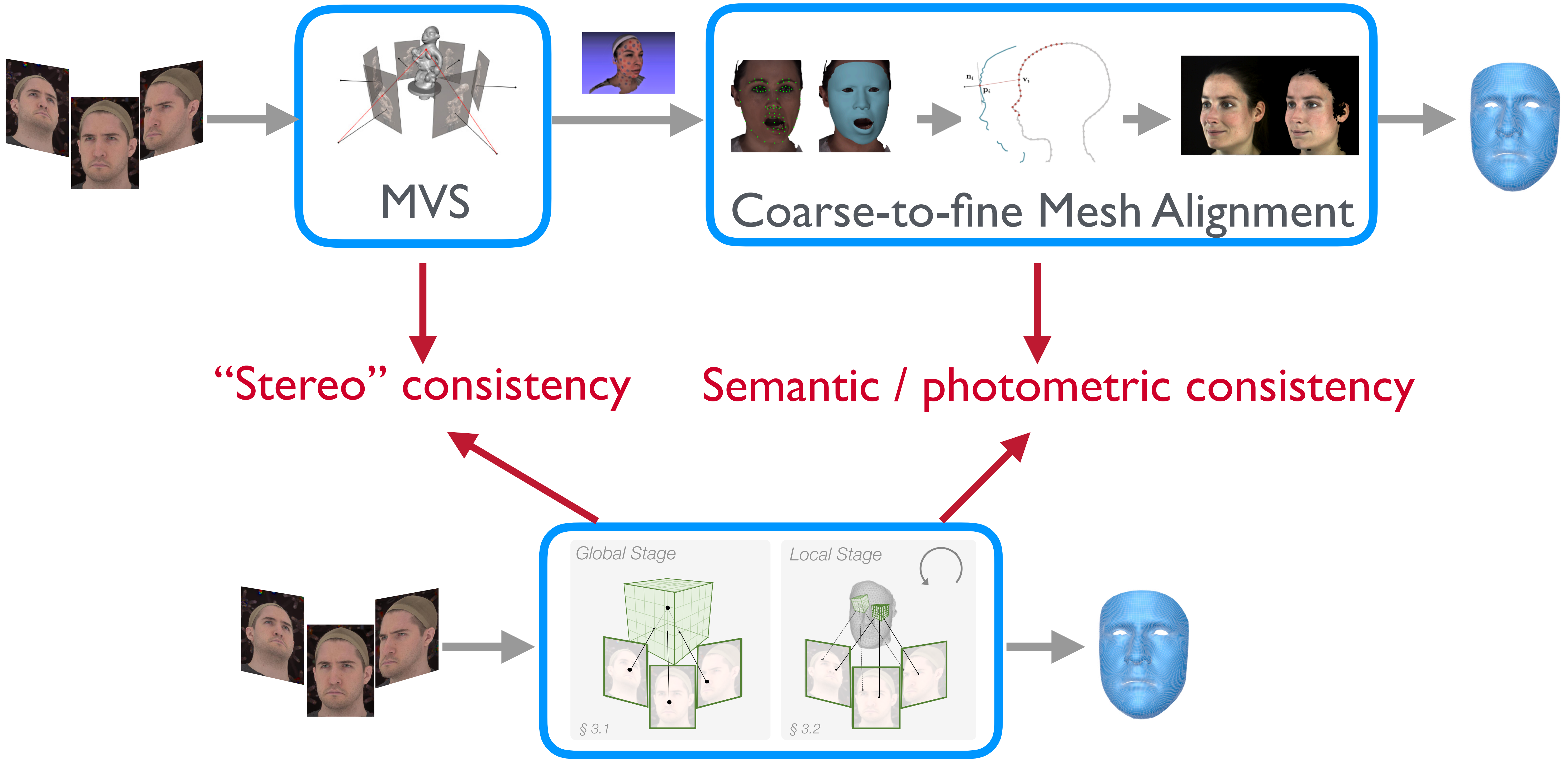
Conclusion: System

- Effective architecture to infer topologically consistent face meshes
  - volumetric feature sampling
  - volumetric networks
  - coarse-to-fine design
- Flexible yet robust inferences
- Accurate registration meshes inferred from images in 0.385 seconds

Conclusion: Methodology

Traditional systems

ToFu - base mesh



- Explore the synergy between reconstruction and registration
- Dense correspondence can be learnt in volumetric space

Thanks!

Comparison and Discussions

Rendering and Data Capture

More information

- Paper and supp. materials
- Video
- Code

GitHub

PyTorch

https://tianyeli.github.io/tofu