Topologically Consistent Multi-View Face Inference Using Volumetric Sampling

ICCV 2021
Virtual Character for VFX

Logan  (Twentieth Century Fox / Image-Engine Design 2017)
Virtual Character for VFX

Personalized Model

Animation

Appearance Capture and Rendering

Credit: Image-Engine Design / USC ICT
Face Dataset Capture

Li and Bolkart et al. SIGGRAPH Asia 2017

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

~10 min per mesh

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

+ manual work

Li, Bladin and Zhao et al. CVPR 2020
Goal: Face Capture and Registration

Calibrated multi-view images

Meshes in consistent topology ("Registrations" or "Alignments")

Appearance and detail maps
Traditional Systems

Calibrated multi-view images → Multi-View Stereo (MVS) → Scans → Coarse-to-fine Mesh Alignment → Registrations

Image credits:
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

Calibrated multi-view images → Multi-View Stereo (MVS) → Scans → Coarse-to-fine Mesh Alignment → Registrations

Issues:
- Slow
- Prone to errors
- Texture-less, transparent, or reflective surfaces
Traditional Systems

Calibrated multi-view images → Multi-View Stereo (MVS) → Scans → Coarse-to-fine Mesh Alignment → Registrations

Issues: • Slow
• Prone to errors
Texture-less, transparent, or reflective surfaces

• Slow
• Requires manual work
Traditional Systems

Calibrated multi-view images

Multi-View Stereo (MVS)

Scans

Coarse-to-fine Mesh Alignment

Issues:
- Slow
- Prone to errors
- Texture-less, transparent, or reflective surfaces

MVS errors affect registration quality

- Slow
- Requires manual work

Registrations
Previous Work: Faces

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 Face inference from multi-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 Face inference from multi-view

Input images

Global Stage

Local Stage

Progressive Mesh Generation

Base mesh

Synthesis Network

Face Completion & Texturing

Appearance and Detail Capture

Skin detail and appearance maps

Detailed mesh

$\{I_i\}_{i=1}^{K}$
Progressive Mesh Generation

**Global Stage** § 3.1

- Input images
- Feature Extraction
- Volumetric Feature Sampling
- Global Geometry Network

**Local Stage** § 3.2

- Mesh Upsampling
- Volumetric Feature Sampling
- Local Refinement Network

Iterate k

Final mesh
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
Global Stage: Volumetric Feature Sampling

Volumetric feature sampling
Global Stage: Volumetric Feature Sampling

\[ f = \sigma(f_1, f_2, f_3, f_4, \ldots) \]
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
Inferred mesh
Local Stage: Mesh Upsampling
Local Stage: 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
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

Input images

Base mesh

Face Completion & Texturing

§ 3.3

Synthesis Network

Skin detail and appearance maps

Input images

Skin detail and appearance maps
Results: Robustness

Traditional MVS + registration

Ours

DFNRMVS [4]
Results: Robustness

Traditional MVS + registration

Ours

DFNRMVS [4]

Input images
(2 of 15)

Output mesh

MVS scan

Conservative

Aggressive

No dependency

on MVS

Output mesh

MVS scan

Conservative

Aggressive

No dependency

on MVS

Input images
(2 of 15)

Traditional MVS + registration
Results: Robustness

Input images (2 of 15)

Conservative
Traditional MVS + registration

Aggressive

Input images (2 of 15)

Conservative
Traditional MVS + registration

Aggressive

Traditional MVS + registration

Ours

DFNRMVS [4]

Output mesh

MVS scan

Output mesh

MVS scan

No dependency on MVS

No dependency on MVS
Results: Robustness

Traditional MVS + registration

Ours

DFNRMVS [4]

Input images (2 of 15)

Output mesh

Conservative

Aggressive

Traditional MVS + registration

DFNRMVS [4]

No dependency on MVS
Results: Robustness

Input images (2 of 15)

Traditional MVS + registration
DFNRMVS [4]
Ours

Output mesh

MVS scan

No dependency on MVS No dependency on MVS

Conservative Aggressive

Traditional MVS + registration DFNRMVS [4] Ours
Results: Geometric Accuracy

- Reference scan
- 3DMM regression (w/ post-processing)
- DFNRMVS [4] (w/ post-processing)
- Our base mesh (direct output)
- Input images (8 out of 15)
- Output mesh
- Overlay
- Scan-to-mesh distance

- Reference Scan
- 3DMM regression (direct output)
- Input Images (8 out of 15)
- Output mesh
- Overlay
- Scan-to-mesh distance

> 5 mm

0
Results: Geometric Accuracy
Results: Geometric Accuracy
Results: Geometric Accuracy
Results: Correspondence Accuracy

Vertex-to-Vertex Distance

Texture Map Error

3DMM regression w/o post-processing
3DMM regression w/ post-processing
DFNRMVS [4] w/ post-processing
Ours

> 10 mm
0
> 0.3
0
## Results: Speed

Our method achieves \textbf{2~3 orders of magnitude faster} runtime.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Time</th>
<th>Automatic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Traditional pipeline</td>
<td>600+</td>
<td>✗</td>
</tr>
<tr>
<td>ToFu (base mesh)</td>
<td>0.385</td>
<td>✓</td>
</tr>
</tbody>
</table>

(Measured in seconds)
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)

Scan
Mesh
Overlay
Scan-to-mesh distance
Checkerboard rendering

Scan
Mesh
Overlay
Scan-to-mesh distance
Checkerboard rendering

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!

More information
- Paper and supp. materials
- Video
- Code

GitHub
PyTorch
https://tianyeli.github.io/tofu