2021 ICCV OCTOBER 11-17 VIRTUAL

Topologically Consistent Multi-View Face Inference Using Volumetric Sampling

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

[1] USC [2] USC Institute for Creative Technologies [3] Max Planck Institute for Intelligent Systems

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

Background


Personalized Model

Animation

Appearance Capture and Rendering


Li and Bolkart et al. SIGGRAPH Asia 17


Li, Bladin, Zhao et al. CVPR 20

- Capture 3D face shapes in meshes of consistent topology are crucial for the VFX pipeline.
- Building high-quality 3D morphable models (3DMM) requires sufficient amount of meshes (typically 10^3 ~10^5)
- Existing systems take tremendous amount of processing time (10^2 ~10^4 days on a single machine) + required manual works

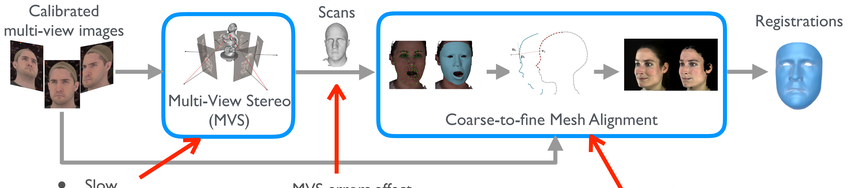
Goal: Face Capture and Registration


Multi-view images with camera calibrations


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


Appearance and detail maps

Traditional Systems



- Slow
- Prone to errors

MVS errors affect registration quality

- Slow
- Requires manual work

Texture-less, transparent, or reflective surfaces

Learning-based Methods


FML
Tewari et al. CVPR 19


RingNet
Sanyal et al. CVPR

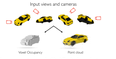
DFNRMVS
Bai et al. CVPR 20

Not *metrically* accurate

- "Memorization"
- Cannot explicitly

Challenging for extreme expression

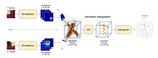
- Constrained by 3DMM


Learning MVS Machine
Kar et al. NeurIPS 17

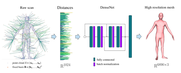

Deep Volumetric Video
Huang et al. ECCV 18


CasMVSNet
Gu et al. CVPR

No dense correspondence


Learnable Triangulation
Iskakov et al. ICCV 19

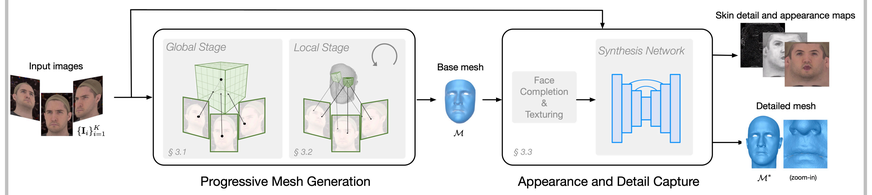
Only sparse correspondences


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

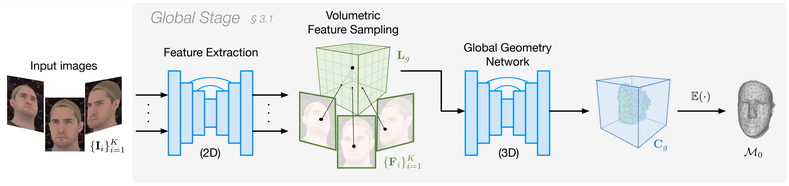
Still needs scans (from MVS)

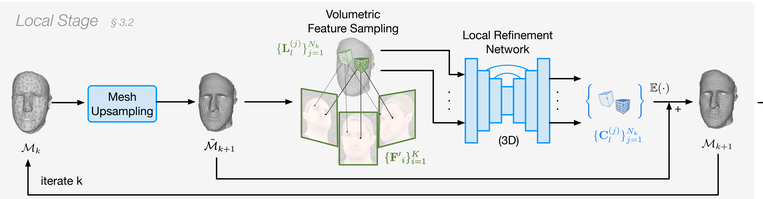
The ToFu Framework

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



Progressive Mesh Generation





Data and 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)

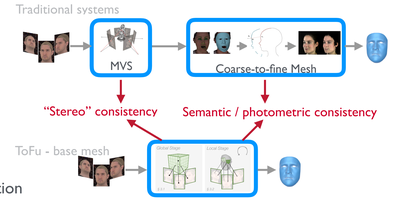
Conclusion

System:

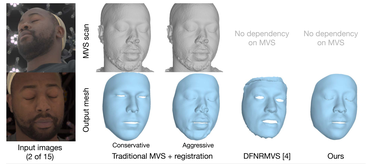
- Effective architecture to infer topologically consistent face meshes
  - volumetric feature sampling and networks
  - coarse-to-fine design
- Accurate, flexible yet robust inferences from images in 0.385 seconds

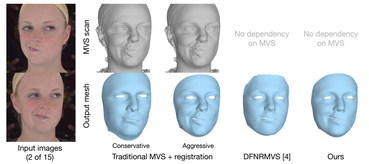
Methodology:

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



Results

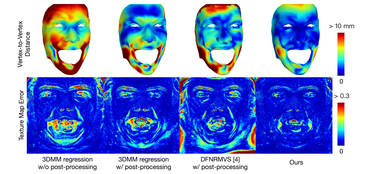




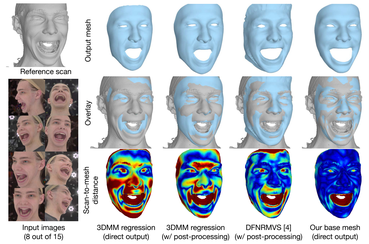
ToFu shows robustness to facial hairs and specularity, which were baffling the traditional methods

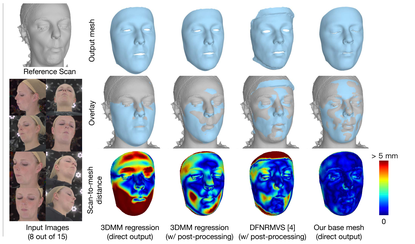
Methods	Time	Automatic
Traditional pipeline	600+	✗
DFNRMVS [4]	4.5	✓
ToFu (base mesh)	0.385	✓

ToFu achieves 2~3 orders of magnitude faster runtime compared to the existing methods



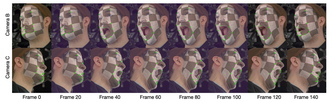
ToFu outperforms existing method in dense correspondence quality





ToFu outperforms existing method in geometric accuracy

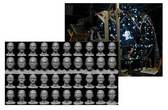
ToFu infers skin detail and appearance maps, enabling high-quality rendering

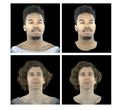


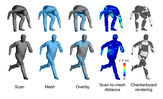
ToFu infers temporally stable registrations although trained and tested frame-by-frame.

Limitation and Future Work

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


Fewer requirement


A complete head model



General objects

References

- Li, Tianye, et al. "Learning a model of facial shape and expression from 4D scans."SIGGRAPH Asia 2017
- Li, Ruilong, et al. "Learning formation of physically-based face attributes." CVPR 2020.
- Tewari, Ayush, et al. "Fml: Face model learning from videos." CVPR 2019.
- Sanyal, Soubhik, et al. "Learning to regress 3D face shape and expression from an image without 3D supervision." CVPR 2019.
- Bai, Ziqian, et al. "Deep facial non-rigid multi-view stereo." CVPR 2020.
- Kar, Abhishek, Christian Häne, and Jitendra Malik. "Learning a multi-view stereo machine." NeurIPS 17
- Huang, Zeng, et al. "Deep volumetric video from very sparse multi-view performance capture." ECCV 2018.
- Gu, Xiaodong, et al. "Cascade cost volume for high-resolution multi-view stereo and stereo matching." CVPR 2020.
- Iskakov, Karim, et al. "Learnable triangulation of human pose." ICCV 2019.
- Prokudin, Sergey, Christoph Lassner, and Javier Romero. "Efficient learning on point clouds with basis point sets." CVPR 2019.
- Ranjan, Anurag, et al. "Generating 3D faces using convolutional mesh autoencoders." ECCV 2018.
- Image credits:
  - VFX pipeline: Image-Engine Design / USC ICT
  - MVS: Furukawa and Hernández, "Multi-View Stereo: A Tutorial",
  - Mesh alignment image credit: 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