NSF: Neural Surface Fields for Human Modelling from Monocular Depth

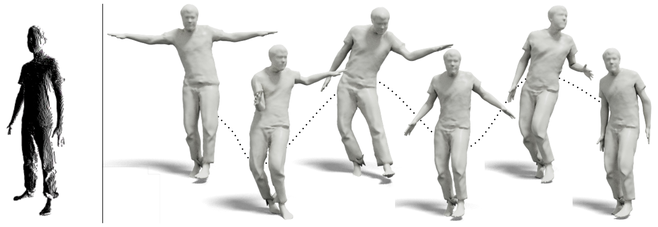
Yuxuan Xue*, Bharat Lal Bhatnagar*, Riccardo Marin, Nikolaos Sarafianos, Yuanlu Xu, Gerard Pons-Moll†, Tony Tung†

University of Tübingen, Tübingen AI Center, Max Planck Institute for Informatics, Meta Reality Labs Research

ICCV23 PARIS

Task

Modelling clothed human from **monocular depth** sequences, can be used for shape completion, texture transfer and re-animation.



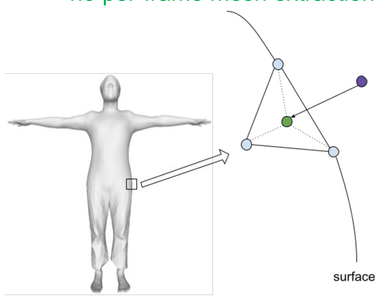
Motivation

Mesh Representation
- existing coherent connectivity
- not flexible in resolution & topology

Point Representation
- flexible
- per-frame Poisson Reconstruction
- incoherent mesh after reconstruction

Neural Implicit Representation
- continuous
- per-frame Marching Cube
- incoherent mesh after extraction

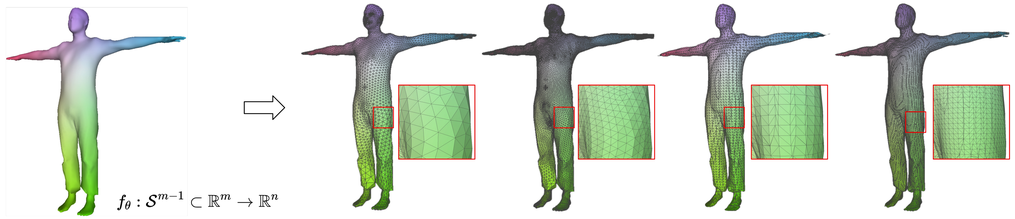
Combine Neural Implicit and Mesh Representation
- continuous
- coherent
- no per-frame mesh extraction



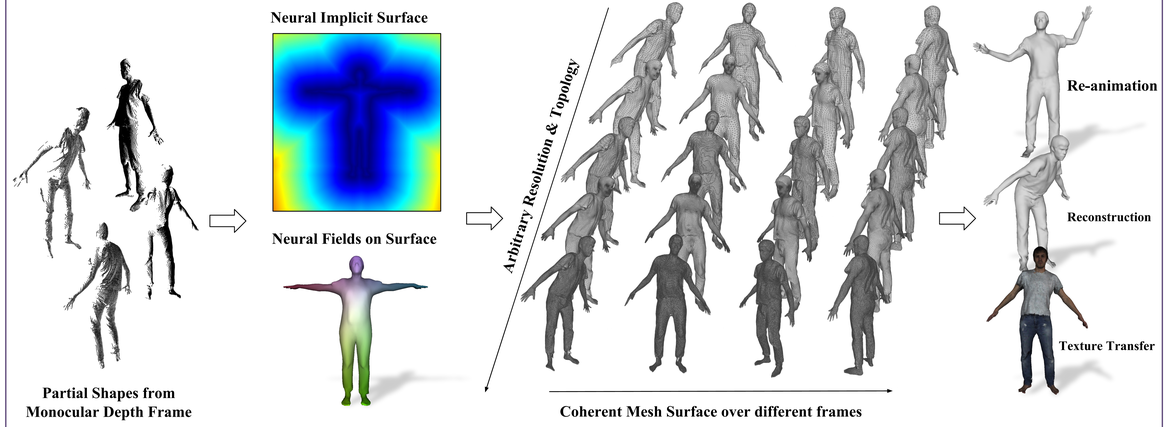
Define a Neural Field on the surface!

Neural Surface Fields (NSF)

We define a neural field solely on the surface, which utilizes the mesh surface ***coherency*** and ***connectivity***. NSF generalizes to the surface in ***arbitrary resolution*** and ***topology*** without retraining.

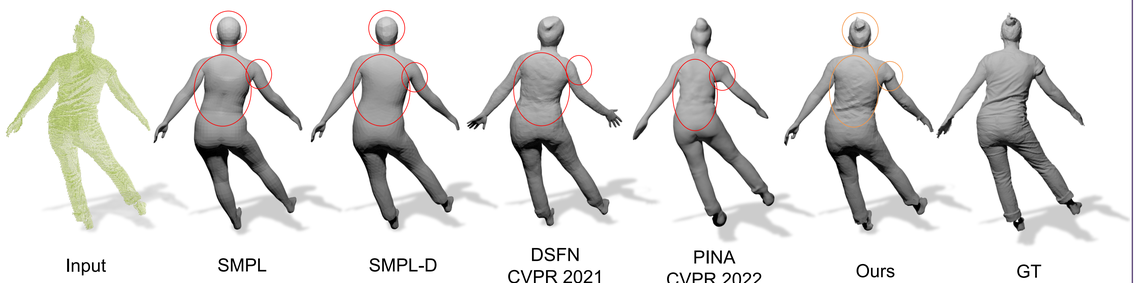


NSF for Human Modelling

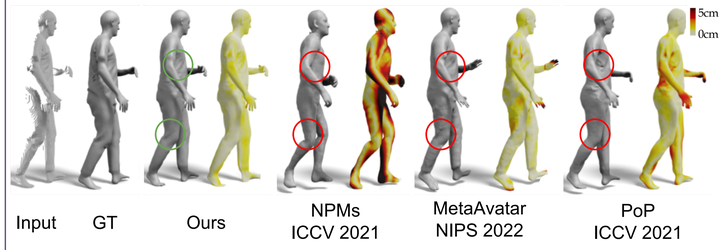


Experimental Results

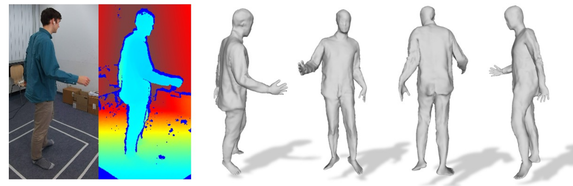
Results on BuFF dataset:



Results On CAPE dataset:



Results on DSFN dataset:



Key Takeaways

- **40x**: NSF is 40x faster than point-based representation, which needs per frame Poisson Reconstruction. (0.039s vs 1.6s)
- **180x**: NSF is 180x faster than implicit representation, which needs per frame Marching Cube surface extraction. (0.039s vs 7s)
- **97.4%**: 97.4% memory compared to volume-based feature representation at resolution of 128^3.
- **86.0%**: 86.0% memory compared to triplane-based feature representation at resolution of 128^3.

Our NSF achieves state-of-the-art accuracy while being more **efficient** and **faster coherent** mesh surface reconstruction!

Summary

- NSF, a novel hybrid surface representation which is
  - efficient, no surface extraction per frame
  - flexible, to arbitrary resolution & topology
  - coherent, across all frames

Code and results available: yuxuan-xue.com/nsf