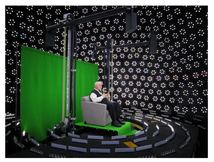
Deep Volumetric Video From Very Sparse Multi-View Performance Capture

Zeng Huang, Tianye Li, Weikai Chen, Yajie Zhao, Jun Xing,
Chloe LeGendre, Linjie Luo, Chongyang Ma, and Hao Li

USC, USC-ICT, Snap Inc., Pinscreen

ECCV 2018 European Conference on Computer Vision

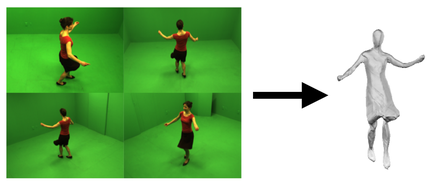
Objective



USC ICT Light Stage

Microsoft Mixed Reality Capture Studio

- Previous performance capture system requires pre-scanned template, large number of cameras or active sensors.



3D body performance capture from very sparse views

- We focus on the task of template-free, per-frame 3D surface reconstruction from very sparse RGB sensors.

Why Challenging?

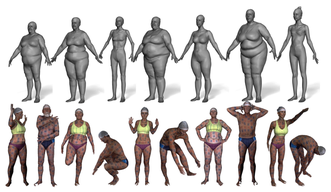


Image from MPI Tübingen

Image from DeepFashion dataset

Complex body shape, pose and appearance, with high variations of clothes.

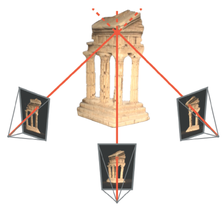


Image from [Goesele et al. 10]

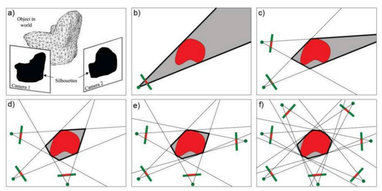


Image by Simon J. D. Prince

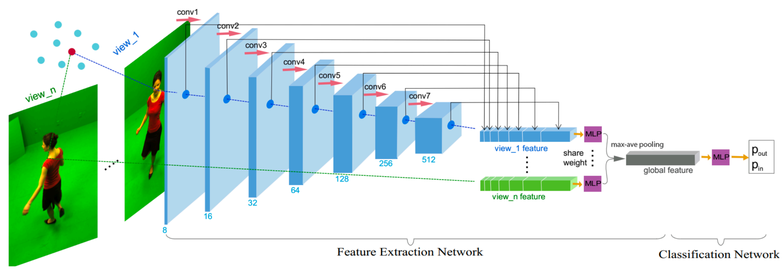
Conventional method, such as multi-view stereo and shape from silhouette, cannot handle very sparse views or textureless region.

DynamicFAUST [Bogo et al. 17]

?

We lack 3D data for **clothed** human with sufficient shape and texture variations.

Our Approach

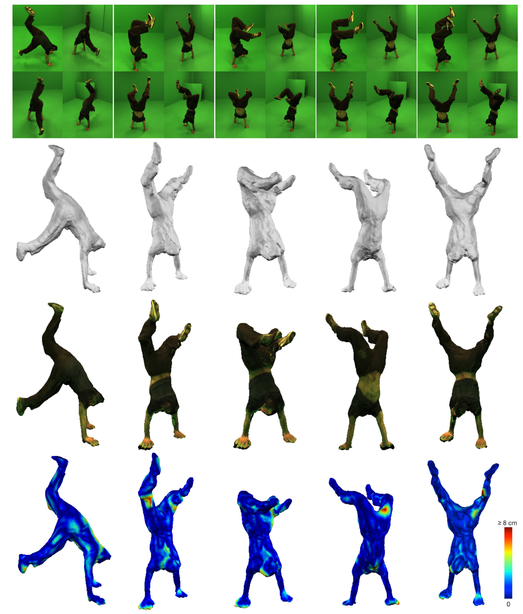


- **Formulation**: Learn to estimate the probability for any 3D point in capture volume if it is inside/outside the reconstructed object.
- **Intuition**: learn the "consensus" of multi-view 2D deep features that are sampled at projected 3D query point.
- **Advantage 1**: reduce the 3D reconstruction problem to a classification problem (easier to learn).
- **Advantage 2**: flexible reconstruction resolution: depends on query point density, instead of limited voxel resolution.

Data

- As far as we know, there is **no** large 3D dataset that covers wide body and clothes variation for full-body reconstruction.
- We build a **synthetic dataset** by rendering 50 rigged and animated human character from Adobe Mixamo, each character in 13 animation sequences.
- By only training on the synthetic data, our network can **generalize** well on real dataset, such as [Vlasic et al. 08].

Results



Video

Paper

Comparison