Learning 3D Digital Humans from Images, Videos and Scans

Emmy Noether-Programm
DFG Deutsche Forschungsgemeinschaft

Gerard Pons-Moll
Max Planck Institute for Informatics

SIC Saarland Informatics Campus

max planck institut informatik

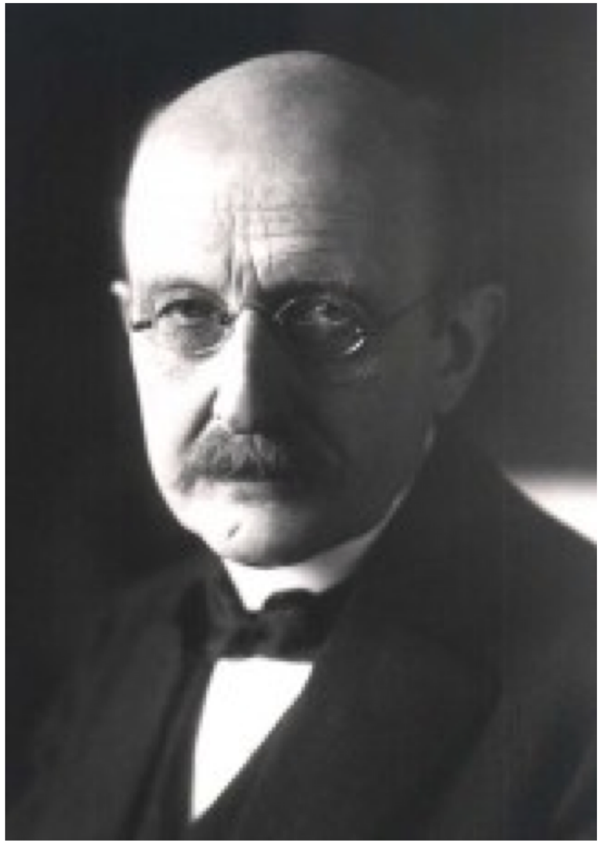
Max Planck (1858 – 1947)

- German theoretical physicist considered to be the *founder* of <u>quantum theory</u>

- Light and electromagnetic waves emitted in discrete “packets” of energy (quanta) → Nobel prize in 1918

- President of the Kaiser-Wilhelm society → in 1938 resigned to protest when nazis took over

- At age 85 still fit enough to climb 3000m peaks ☺



Max Planck Society

rooted in Kaiser Wilhelm Society 1911- 1945
and re-founded 1948

85 institutes (MPIs)
ca. 300 directors
ca. 6500 scientists
incl. 5000 doct. students

33 Nobel prizes (Planck, Hahn, Heisenberg, …, Hell)

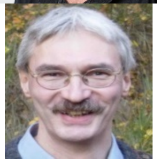
MPI for Informatics (founded in 1990)

D1: Algorithms and Complexity
(Kurt Mehlhorn, 1990-2019)

D2: Computer Vision and Machine Learning
(Bernt Schiele, 2010-2035)

D3: Internet Architecture
(Anja Feldmann, 2018-2033)

D4: Computer Graphics
(Hans-Peter Seidel, 1999-2026)



D5: Databases and Information Systems
(Gerhard Weikum, 2003-2023)

RG1: Automation of Logics
(Christoph Weidenbach, 2005-2033)

5 Scientific directors,
1 Emeritus,
27 Senior researchers (3 tenured)
30 Postdocs,
96 Doctoral students

BIO: Computational Biology
(Thomas Lengauer)

Publications

CSRankings: Computer Science Rankings

CSRankings is a metrics-based ranking of top computer science institutions around the world. **Click on a triangle** (▶) to expand areas or institutions. **Click on a name** to go to a faculty member's home page. **Click on a pie** (the after a name or institution) to see their publication profile as a pie chart. **Click on a Google Scholar icon** to see publications, and **click on the DBLP logo** to go to a DBLP entry.

Rank institutions in Europe by publications from 2009 to 2019

All Areas [off | on]

AI [off | on]

- ▶ Artificial intelligence
- ▶ Computer vision
- ▶ Machine learning & data mining
- ▶ Natural language processing
- ▶ The Web & information retrieval

Systems [off | on]

- ▶ Computer architecture
- ▶ Computer networks
- ▶ Computer security
- ▶ Databases
- ▶ Design automation
- ▶ Embedded & real-time systems
- ▶ High-performance computing
- ▶ Mobile computing

#	Institution	Count	Faculty
1	▶ ETH Zurich	7.7	35
2	▶ Max Planck Institute	5.2	31
3	▶ Technion	5.0	77
4	▶ EPFL	4.7	46
5	▶ University of Edinburgh	4.4	69
6	▶ Tel Aviv University	4.1	42
7	▶ University College London	3.7	55
8	▶ University of Oxford	3.6	50
9	▶ TU Munich	3.4	40
10	▶ TU Darmstadt	3.3	27
11	▶ Hebrew University of Jerusalem	3.0	35
12	▶ Imperial College London	2.9	39
13	▶ University of Cambridge	2.8	38

Publications

CSRankings: Computer Science Rankings

CSRankings is a metrics-based ranking of top computer science institutions around the world. **Click on a triangle** (▶) to expand areas or institutions. **Click on a name** to go to a faculty member's home page. **Click on a pie** (the after a name or institution) to see their publication profile as a pie chart. **Click on a Google Scholar icon** to see publications, and **click on the DBLP logo** to go to a DBLP entry.

Rank institutions in Europe by publications from 2009 to 2019

All Areas [off | on]

AI [off | on]

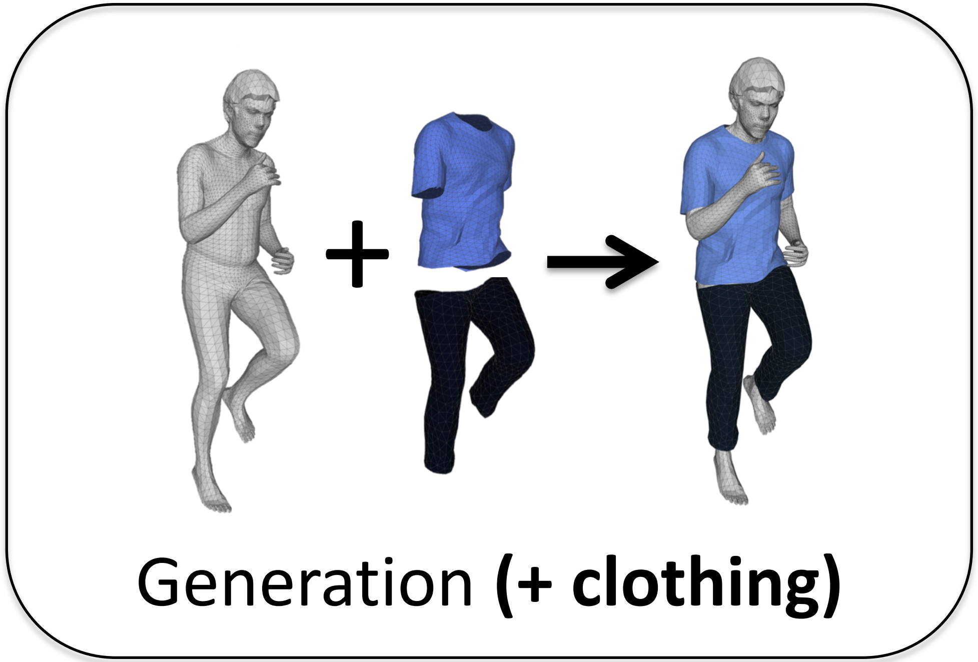
- ▶ Artificial intelligence ☐
- ▶ Computer vision ☑
- ▶ Machine learning & data mining ☐
- ▶ Natural language processing ☐
- ▶ The Web & information retrieval ☐

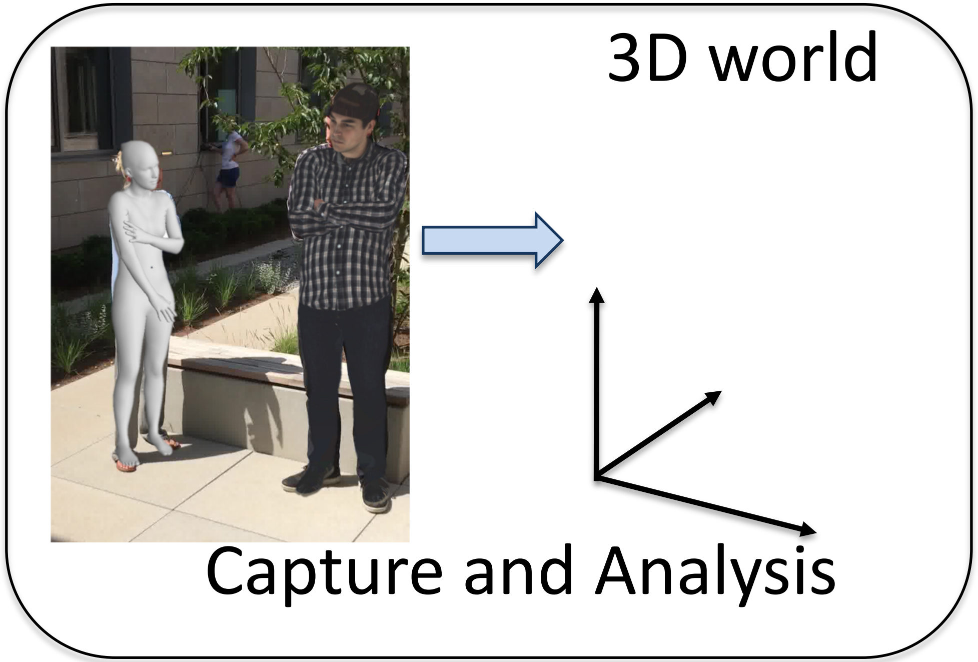
Systems [off | on]

- ▶ Computer architecture ☐
- ▶ Computer networks ☐
- ▶ Computer security ☐
- ▶ Databases ☐
- ▶ Design automation ☐
- ▶ Embedded & real-time systems ☐
- ▶ High-performance computing ☐
- ▶ Mobile computing ☐
- ▶ Measurement & perf. analysis ☐

#	Institution	Count	Faculty
1	▶ Max Planck Institute	52.6	7
2	▶ ETH Zurich	43.1	5
3	▶ University of Surrey	41.6	12
4	▶ TU Munich	37.3	9
5	▶ EPFL	36.2	7
6	▶ Graz University of Technology	34.4	5
7	▶ Ecole Normale Superieure	30.1	5
8	▶ Imperial College London	29.4	10
9	▶ Technion	25.7	12
10	▶ University of Edinburgh	23.5	4
11	▶ TU Darmstadt	15.4	4
12	▶ Queen Mary University of London	14.6	4
13	▶ University College London	14.0	9

Goal: Realistic virtual humans

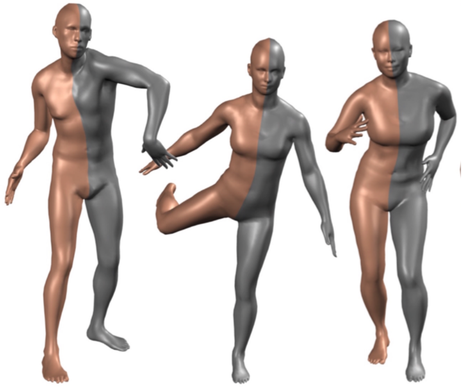




Realistic 3D people models:
- Move and look like real people
- Easy to control and animate
- Easy to fit to data

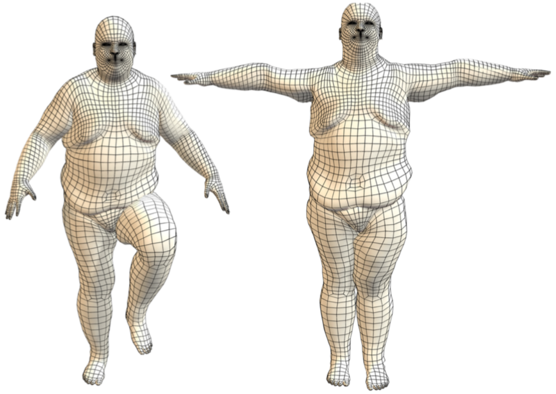
Reconstruction from images:
- Accurate
- Efficient
- Robust

VIRTUAL HUMANS - MENTAL MODEL



SMPL: Pose and Shape

Loper et al.
SiggAsia'15



DYNA/DMPL: Soft-tissue

Pons-Moll et al.
Siggraph'15

ClothCap: Clothing

Pons-Moll et al.
Siggraph'17

“You need 3 things to win a war”

- Money
- Money
- And more Money

Gian Giacomo Trivulzio

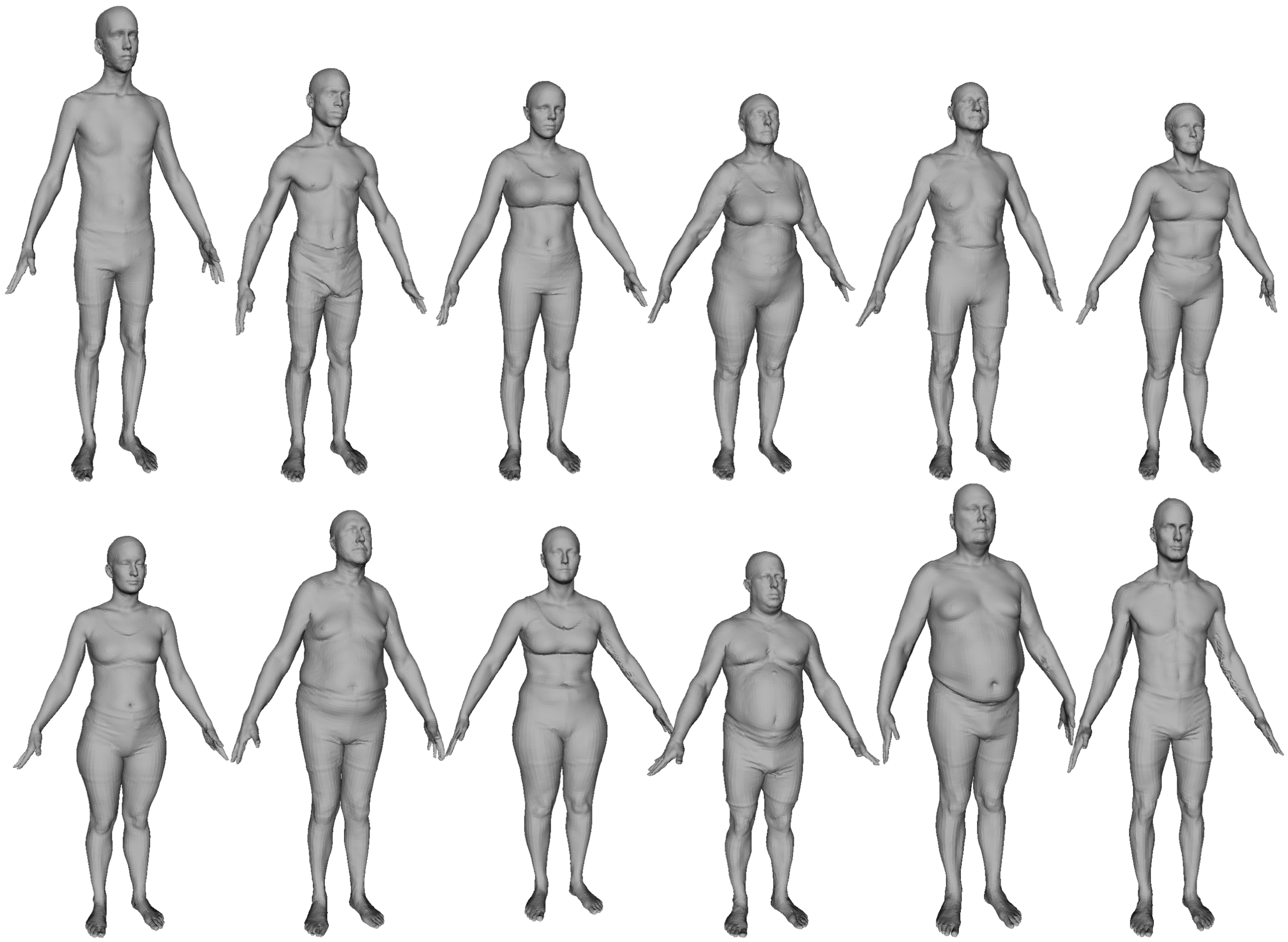
Slide adapted from Vladlen Koltun

You need 3 things to solve an AI problem

- Data
- Data
- And more Data

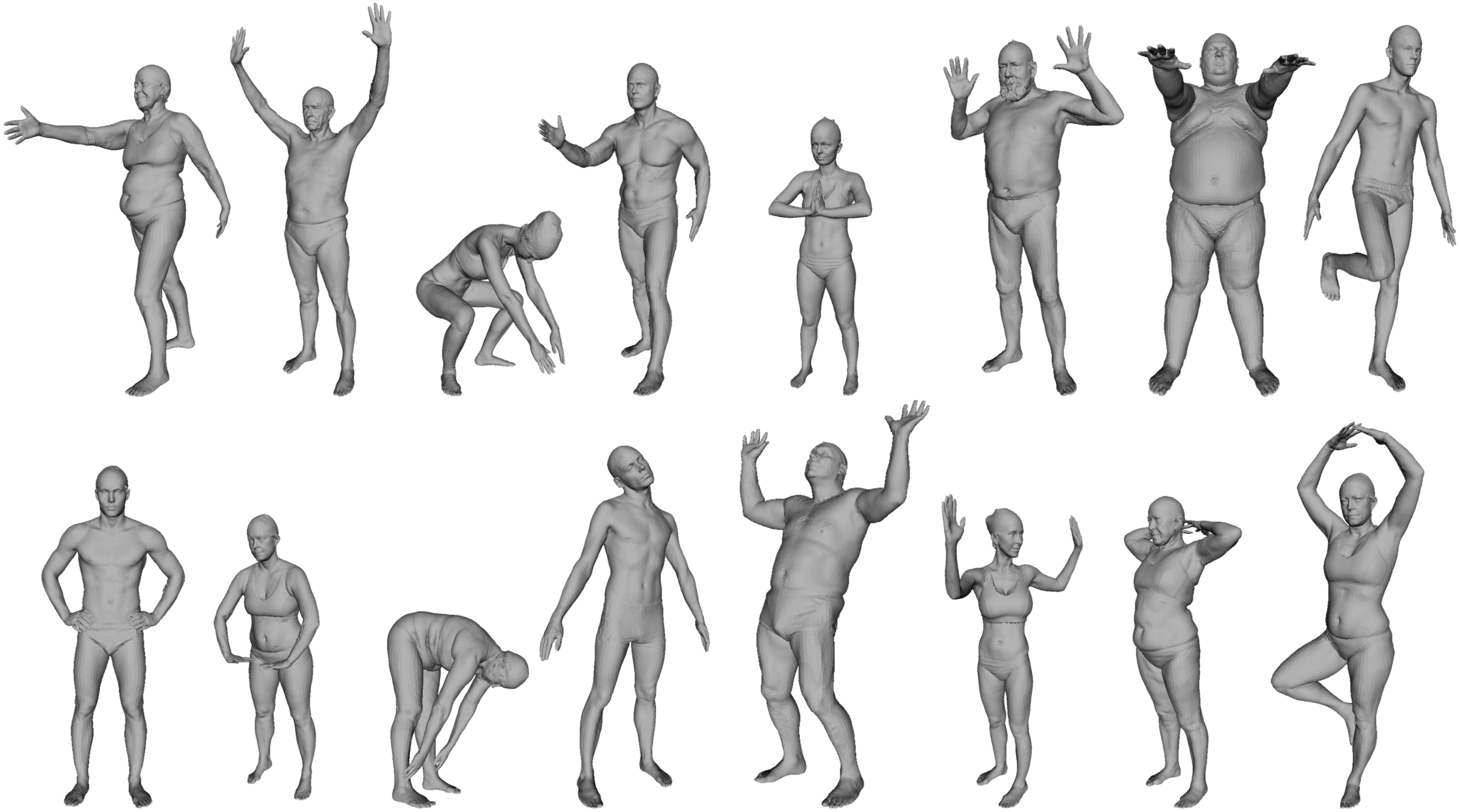
Geoff Hinton

Idea: Collect 3D scans from



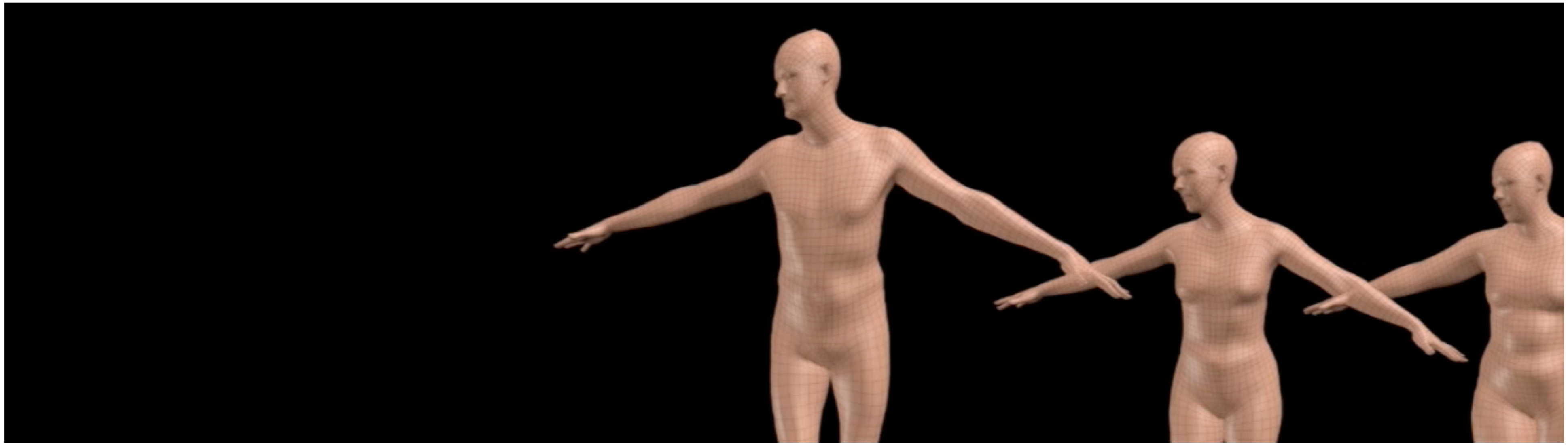
thousands of people…

and thousands of poses



1000’s of high-resolution scans of different shapes and poses

SMPL: A model of pose and shape



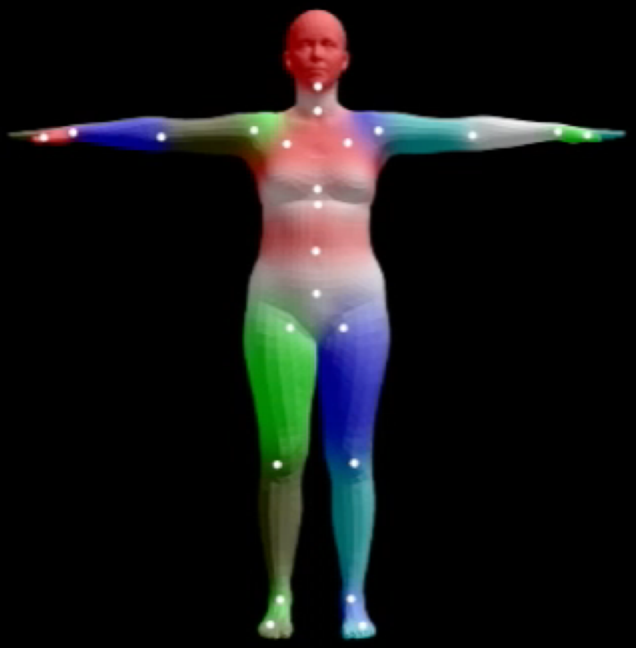
$$M(\theta, \beta; \mathbf{w}) : \mathbb{R}^{|\theta|+|\beta|} \mapsto \mathbb{R}^{3N}$$

$$\text{Latent parameters} \mapsto \text{vertices}$$

SMPL Model

M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, M. Black **SIGGRAPH Asia‘15**

SMPL Model Pipeline



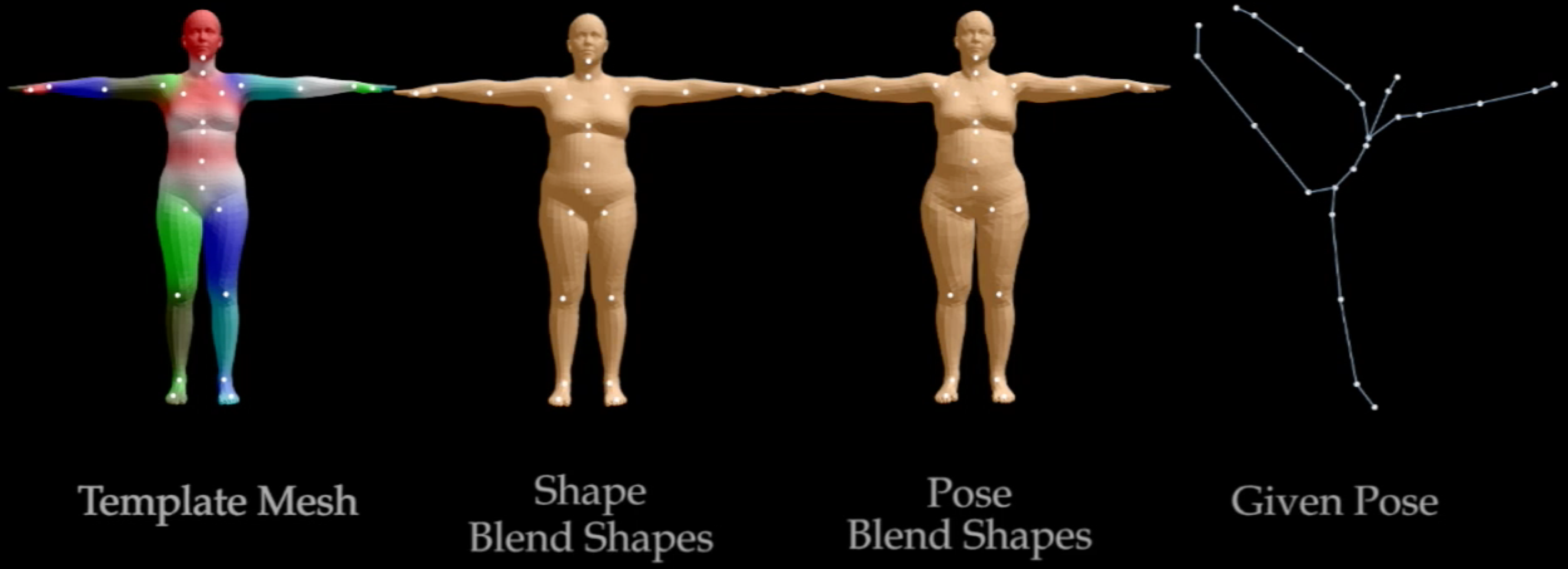
Template Mesh

SMPL Model Pipeline

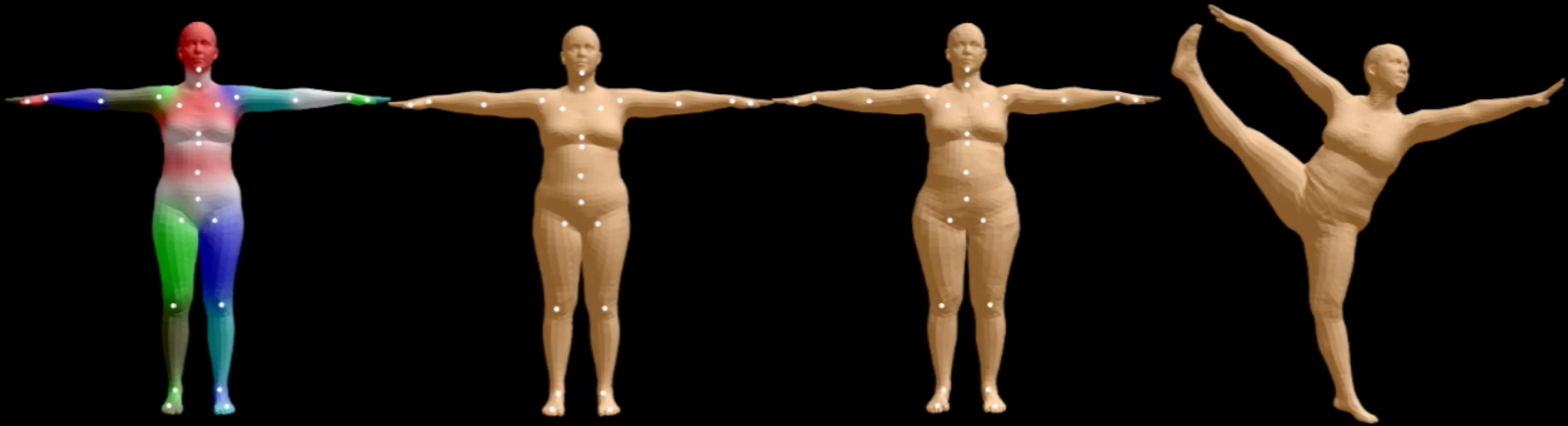
Template Mesh

Shape
Blend Shapes

SMPL Model Pipeline

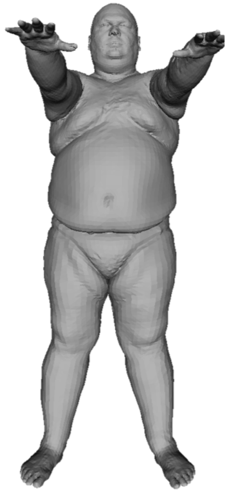


SMPL Model Pipeline

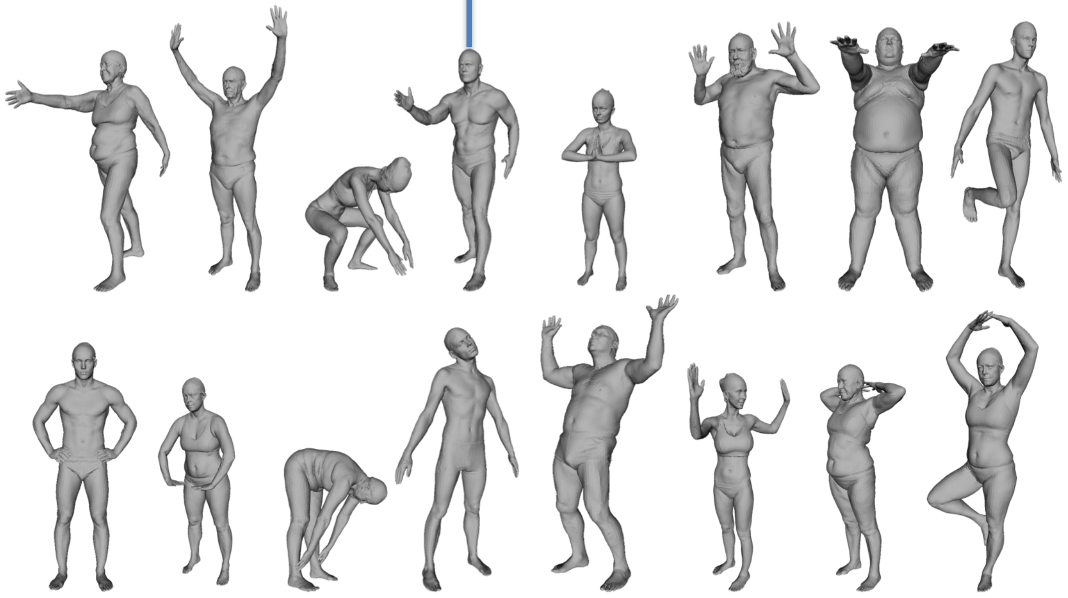


Model Training

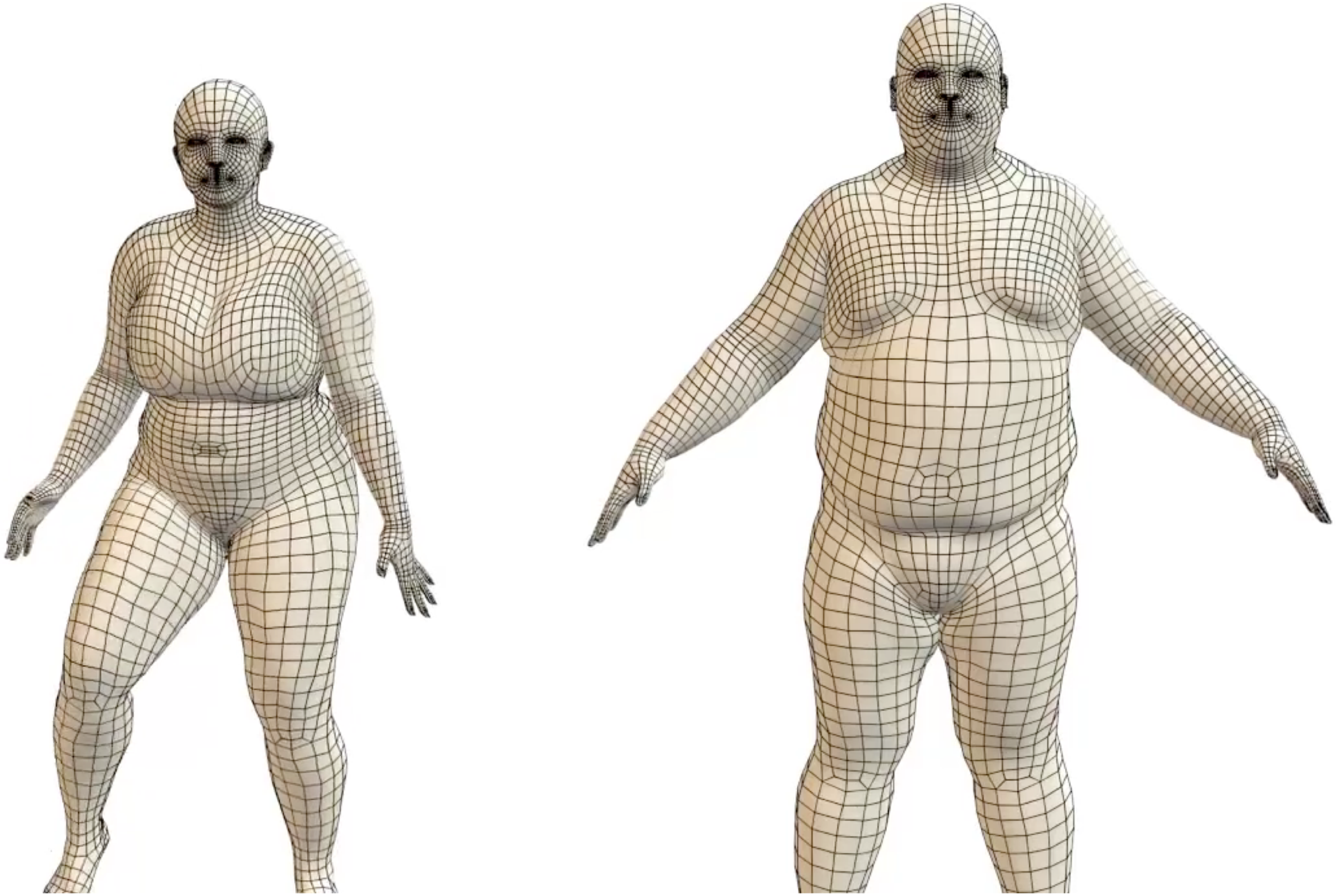
$$\mathbf{w} = \arg\min_{\mathbf{w}} \sum_{j} \| M(\vec{\theta}, \vec{\beta}; \mathbf{w}) -$$



$$\|^2$$

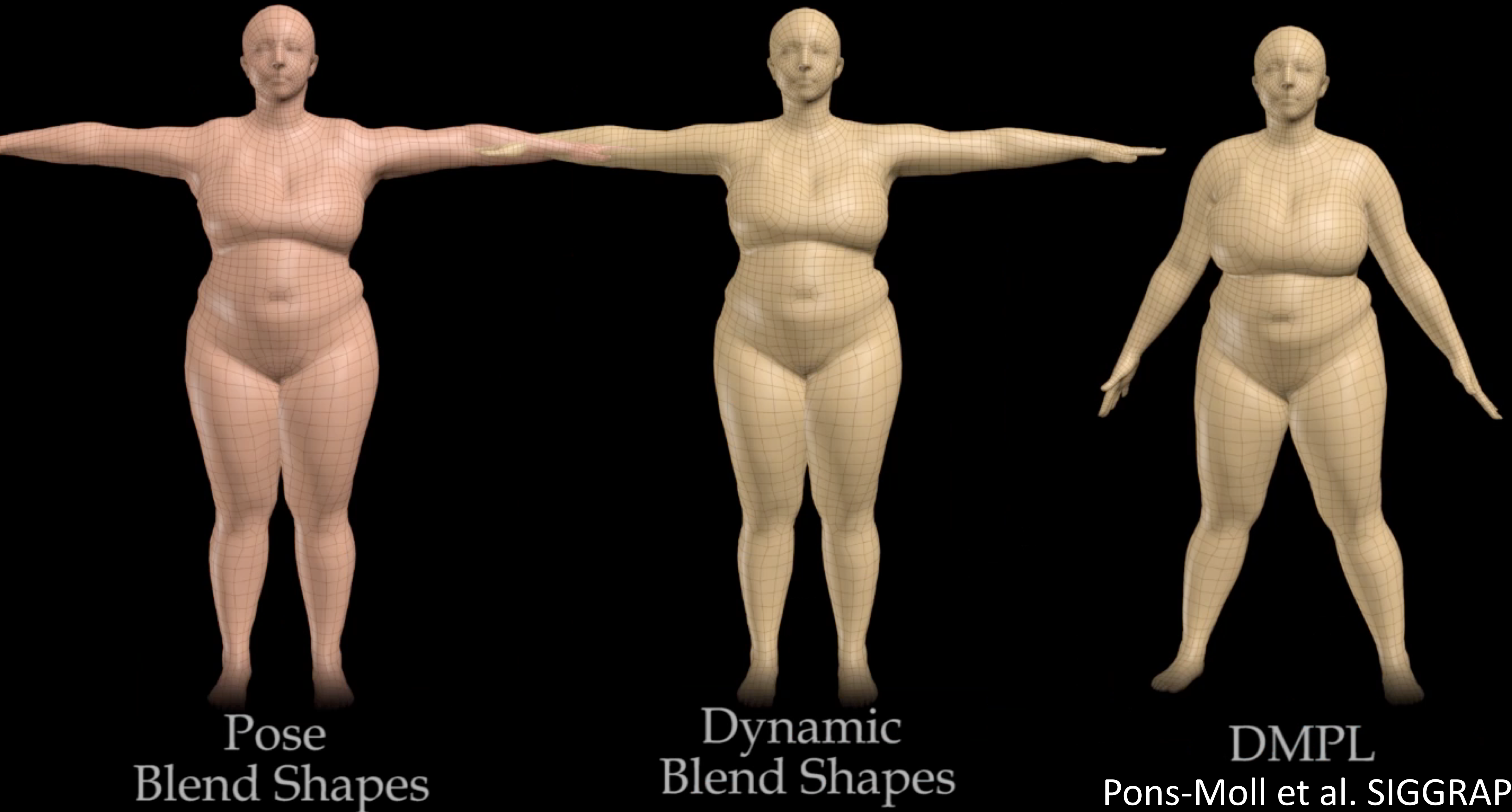


Dyna: A model of how we jiggle



G. Pons-Moll,
J. Romero,
N. Mahmood,
M. Black
SIGGRAPH'15

DYNA: A model soft-tissue

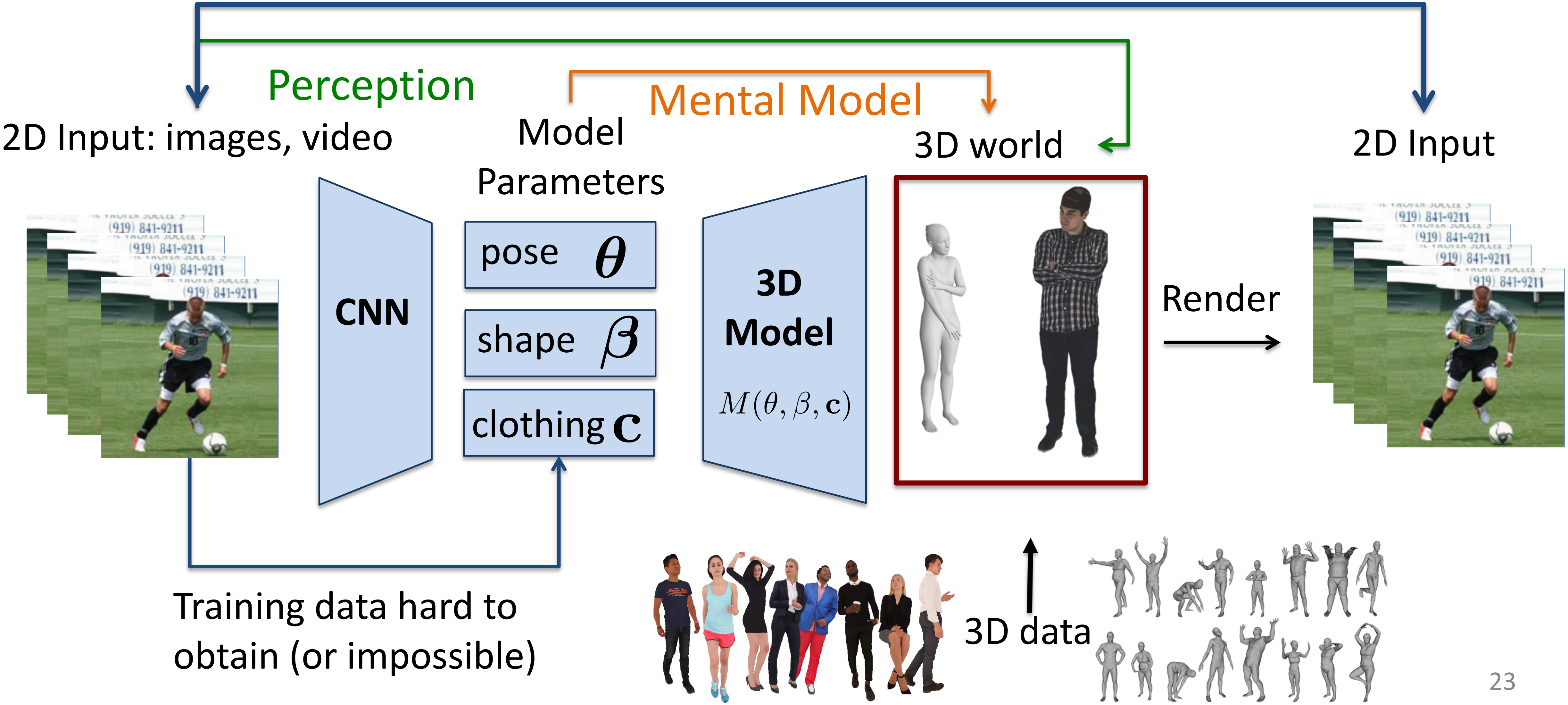


Pons-Moll et al. SIGGRAPH'15

DOES NOT SCALE TO THE REAL WORLD

Vision

Computer Vision + Computer Graphics + Learning

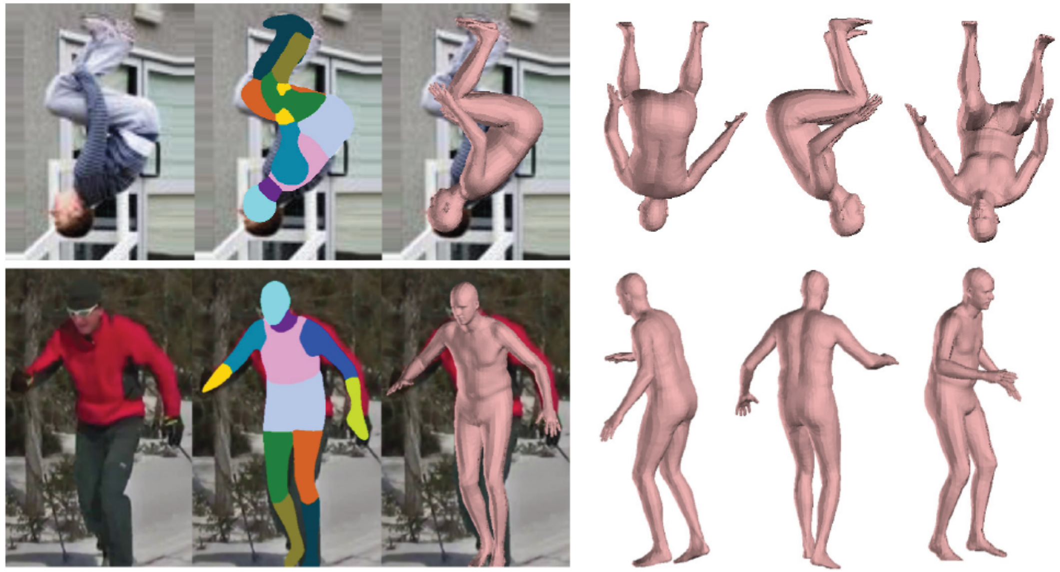


Neural Body Fitting:
Body Pose and Shape from a Single Image





M. Omran C. Lassner G. Pons-Moll P. Gehler B. Schiele

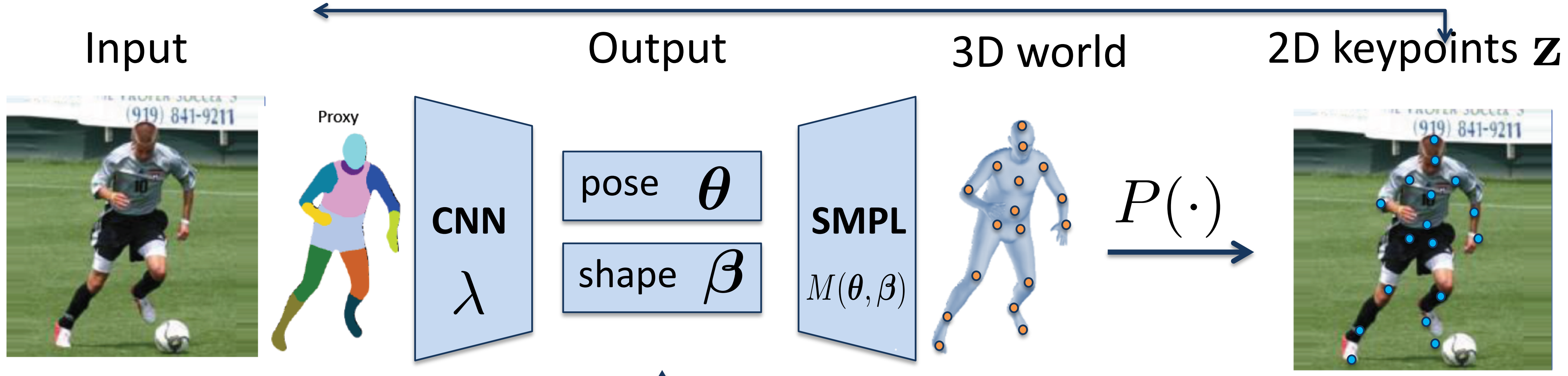


3DV 2018

Best Student Paper Award

Neural Body Fitting
Body Pose and Shape from 1 Image

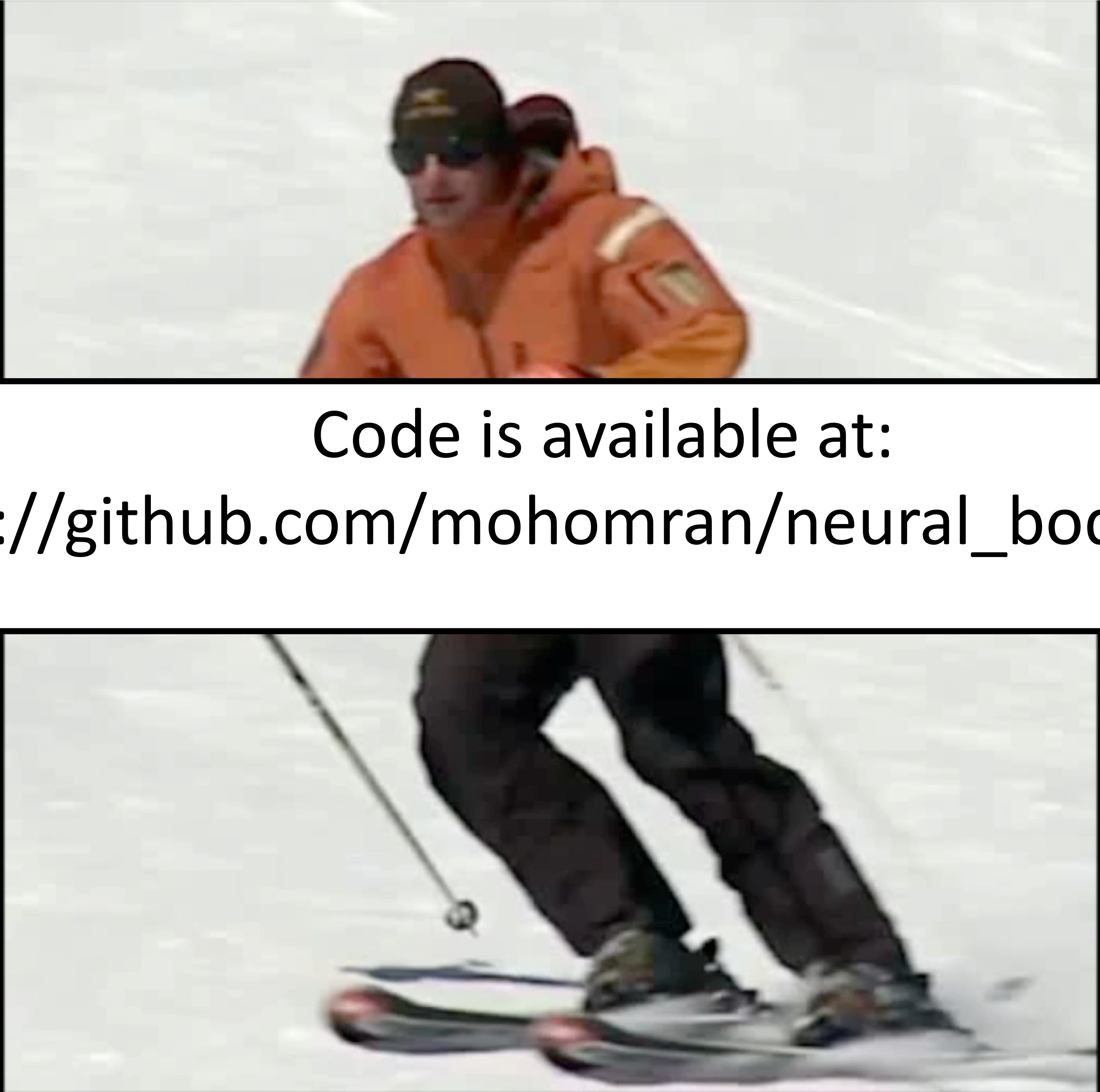
$$\arg\min_{\lambda} \|\hat{\mathbf{z}}(\mathbf{I}, \lambda) - \mathbf{z}\|$$



Training data hard to obtain!

Omran et al.

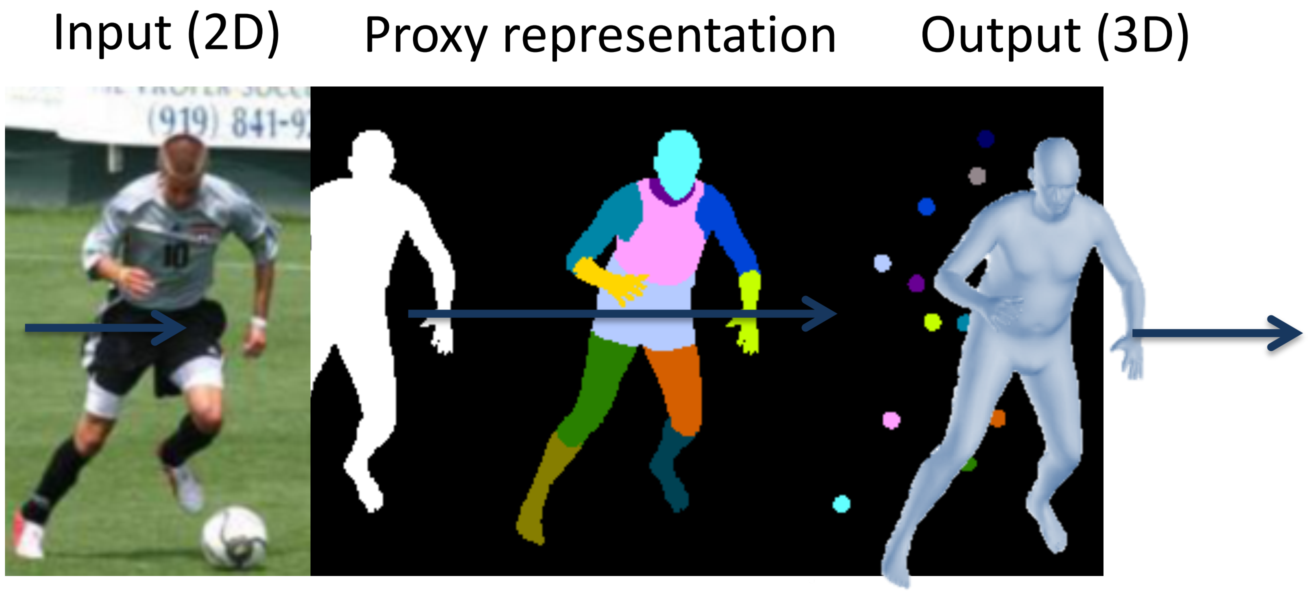
3DV 2018 **Best student paper award**



Code is available at:

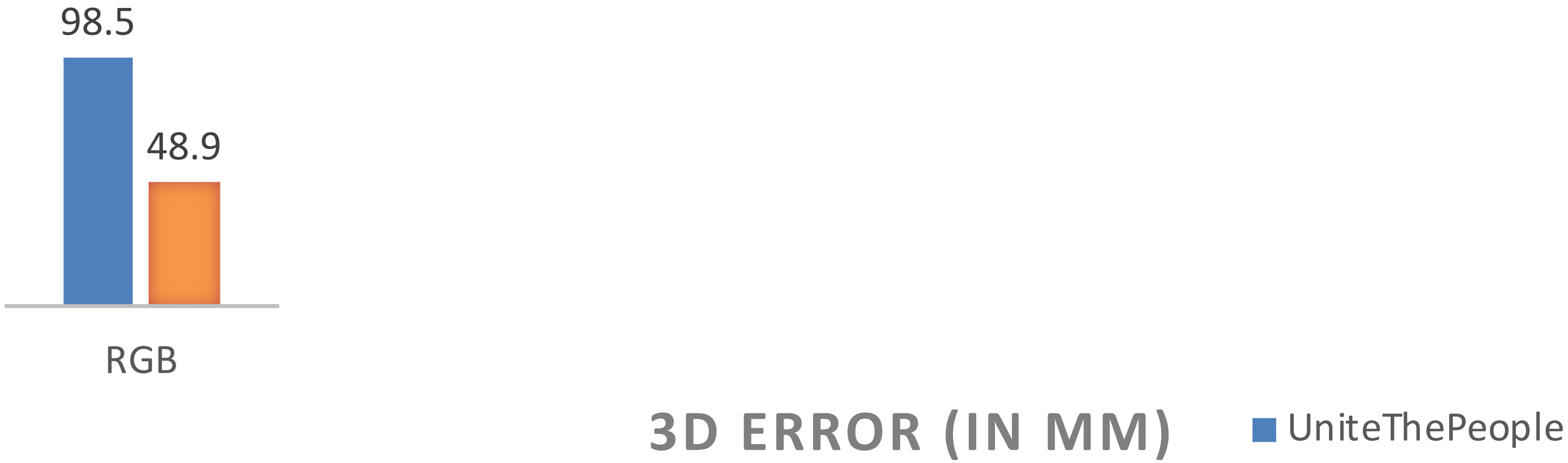
https://github.com/mohomran/neural_body_fitting

Input Representation

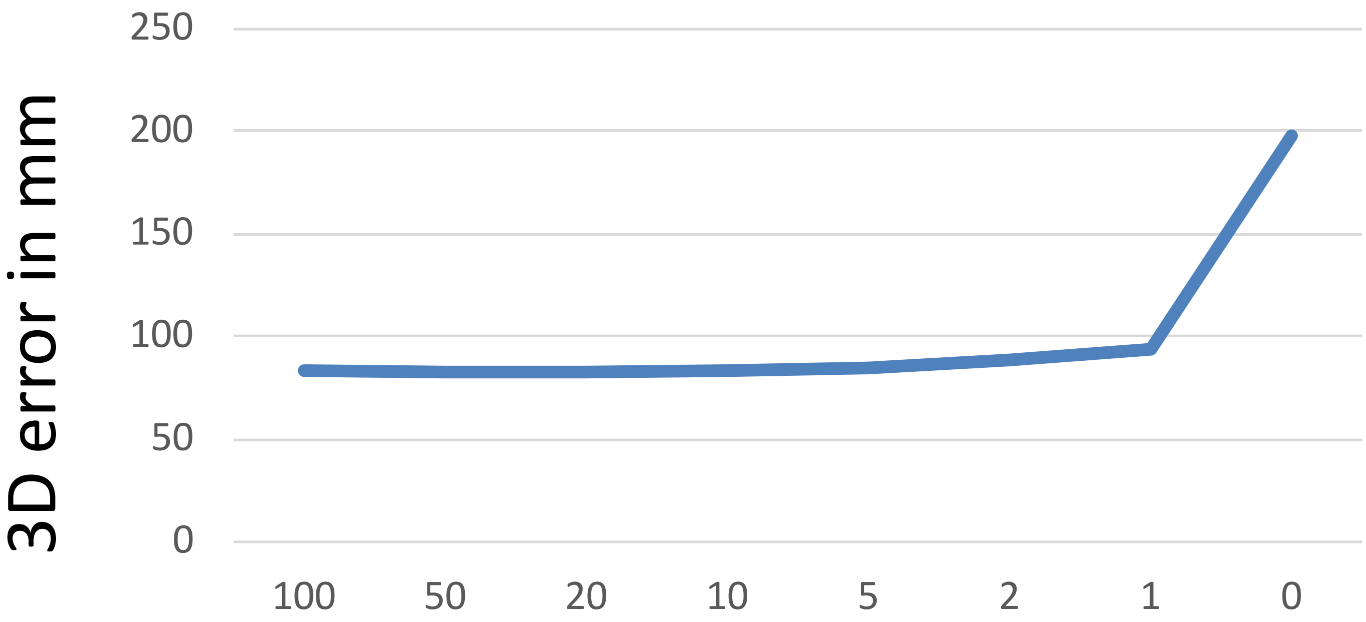


Would an intermediate representation help?
If yes, which?

Input representation



How much 3D data is needed?



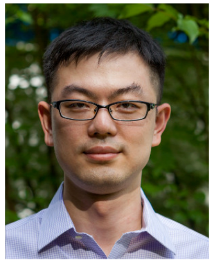
% of training data with 3D ground truth (besides 2D)

Clothing

Video-Based Reconstruction of 3D People Models


T. Alldieck


M. Magnor


W. Xu


C. Theobalt


G. Pons-Moll

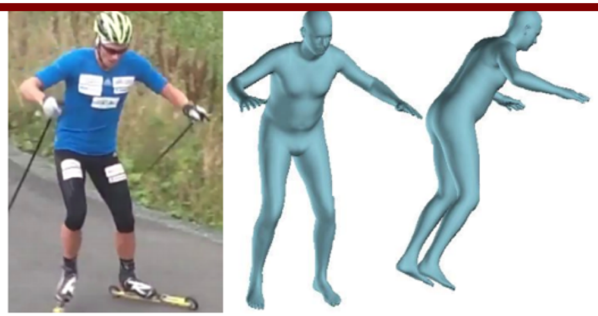
CVPR'18

[spotlight]

Previous Work



No clothing, no personalization!

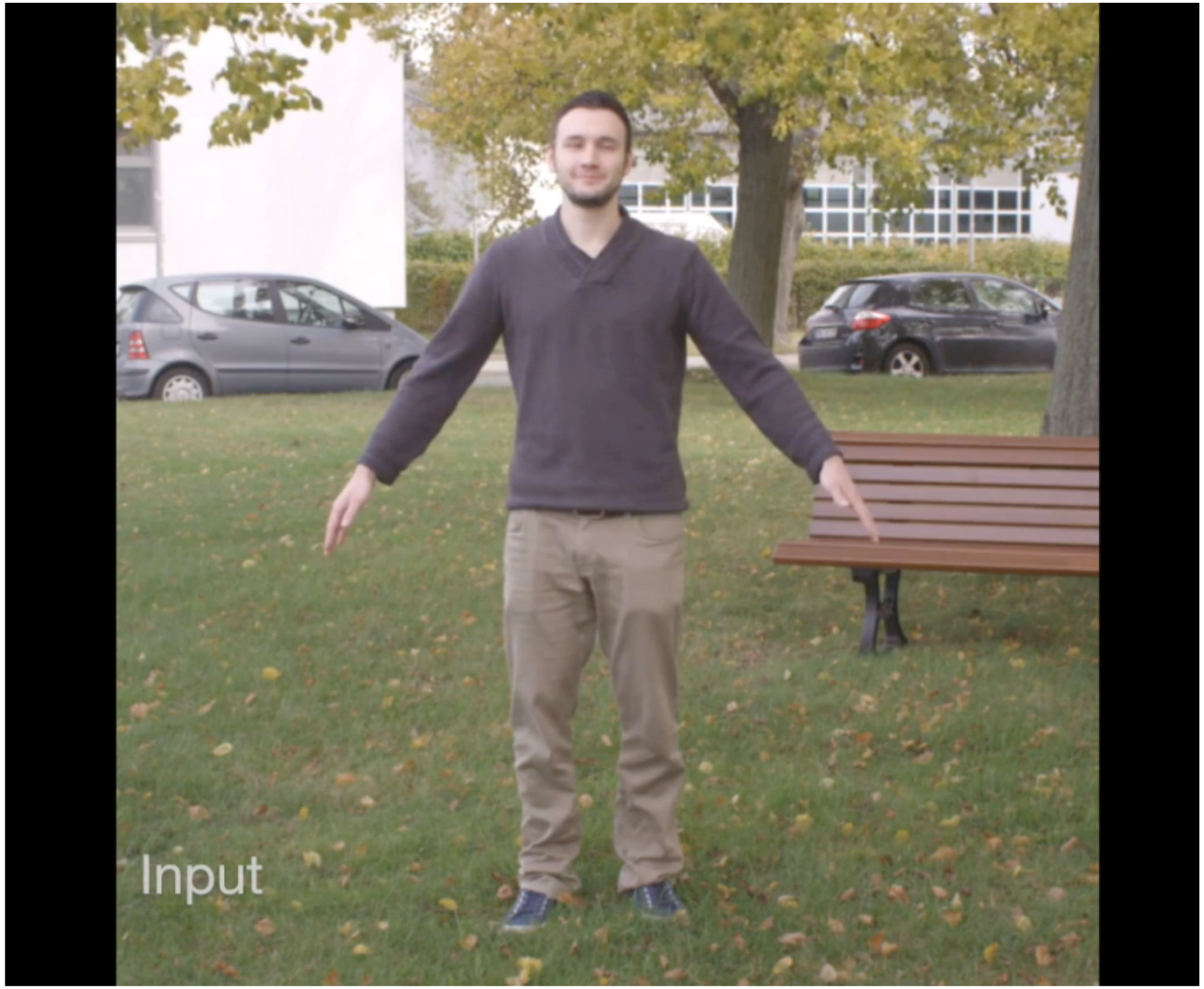


[Pavlakos et al. ’18]

[Kanazawa et al. ’18]

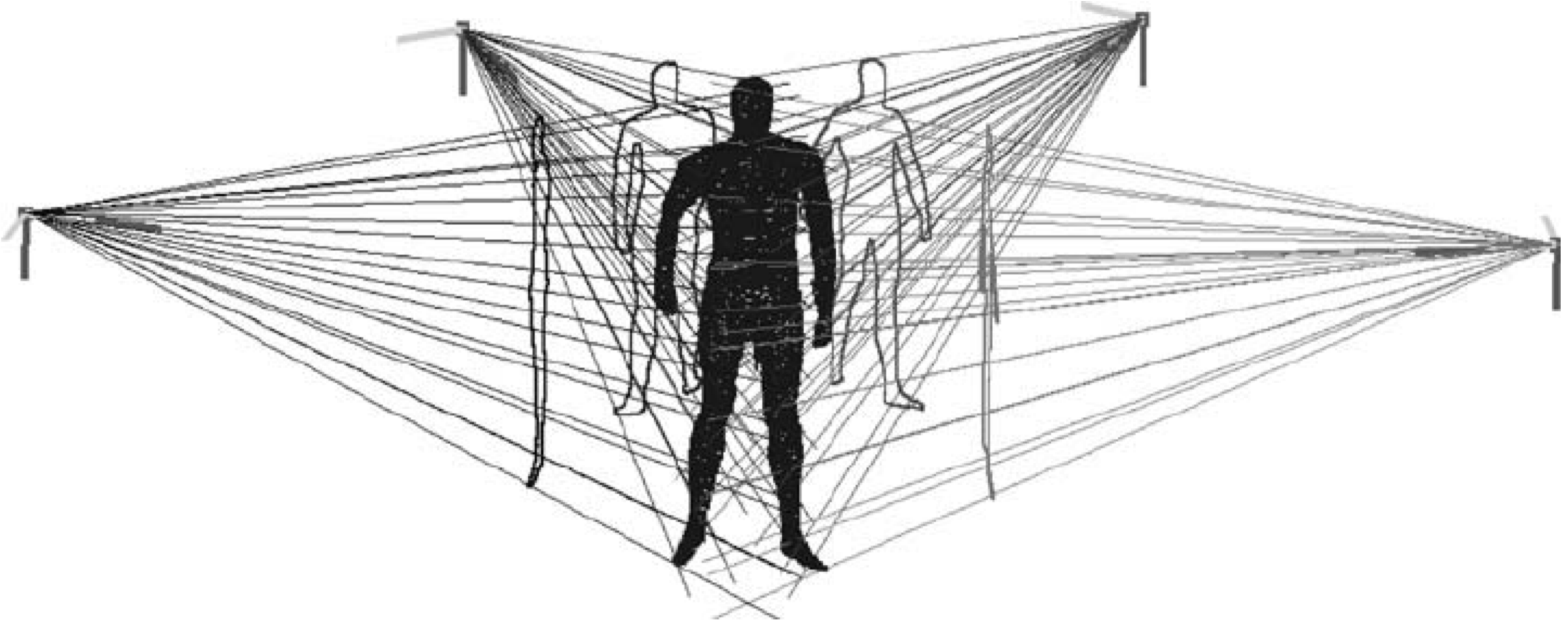
[Bogo et al. ’15]

Goal: 3D Reconstruction of People from a Single Video



Key Idea: Extend Visual Hulls to Dynamic Human Motion

Problem: standard visual hull requires a **static** object captured by multiple views

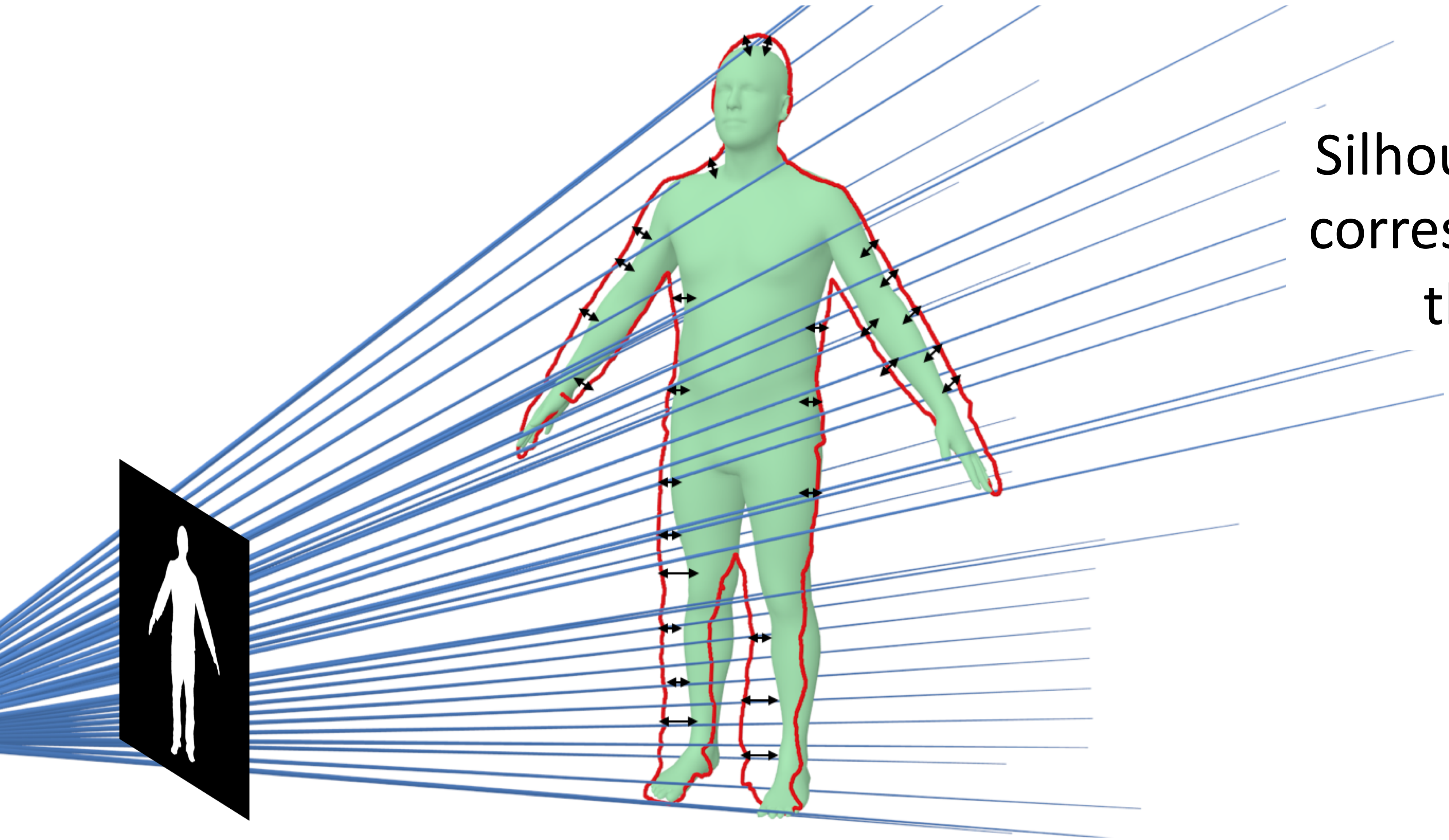


How Can We Generalize It to Dynamic Human Motion ?

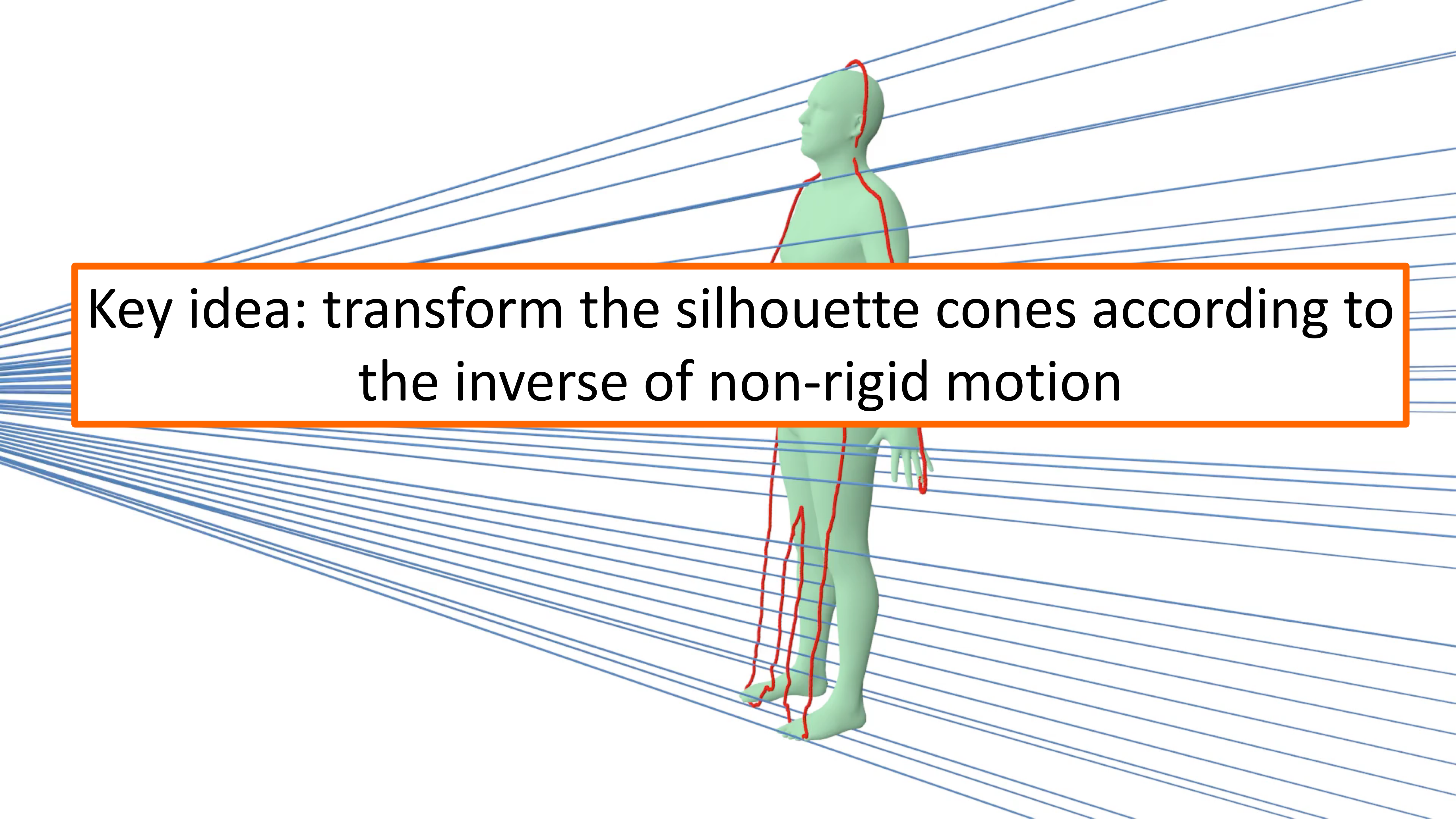
Person is moving!

How Can We Generalize It to Dynamic Human Motion ?

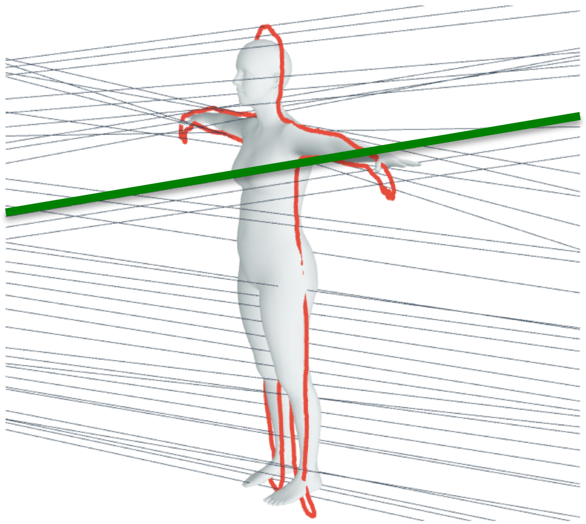
Estimate the 3D human pose and shape per frame

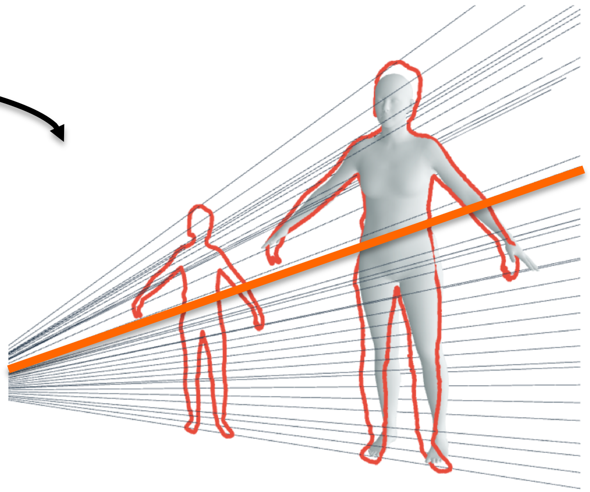


Silhouette rays with correspondences on the surface



Key idea: transform the silhouette cones according to the inverse of non-rigid motion





$$\boxed{\mathbf{r}} = \underbrace{\left(\sum_{k=1}^{K} w_{k,i} G_k(\boldsymbol{\theta}, \boldsymbol{J}_{\boldsymbol{\beta}})\right)^{-1}}_{\text{Inverse of Articulated Motion}} \boxed{\mathbf{r}'} - b_{P,i}(\boldsymbol{\theta}).$$

Ray in Canonical Frame

Ray

Optimize a Single Shape to Fit all *Unposed* Silhouette Cones

$$\boxed{\arg\min_{\boldsymbol{\beta},\mathbf{d}} E_{\text{cons}}(\boldsymbol{\beta},\mathbf{d})}$$

$$E_{\text{data}} = \sum_{(\mathbf{v},\mathbf{r})\in\mathcal{M}} \rho(\mathbf{v}\times\mathbf{r}_n - \mathbf{r}_m)$$

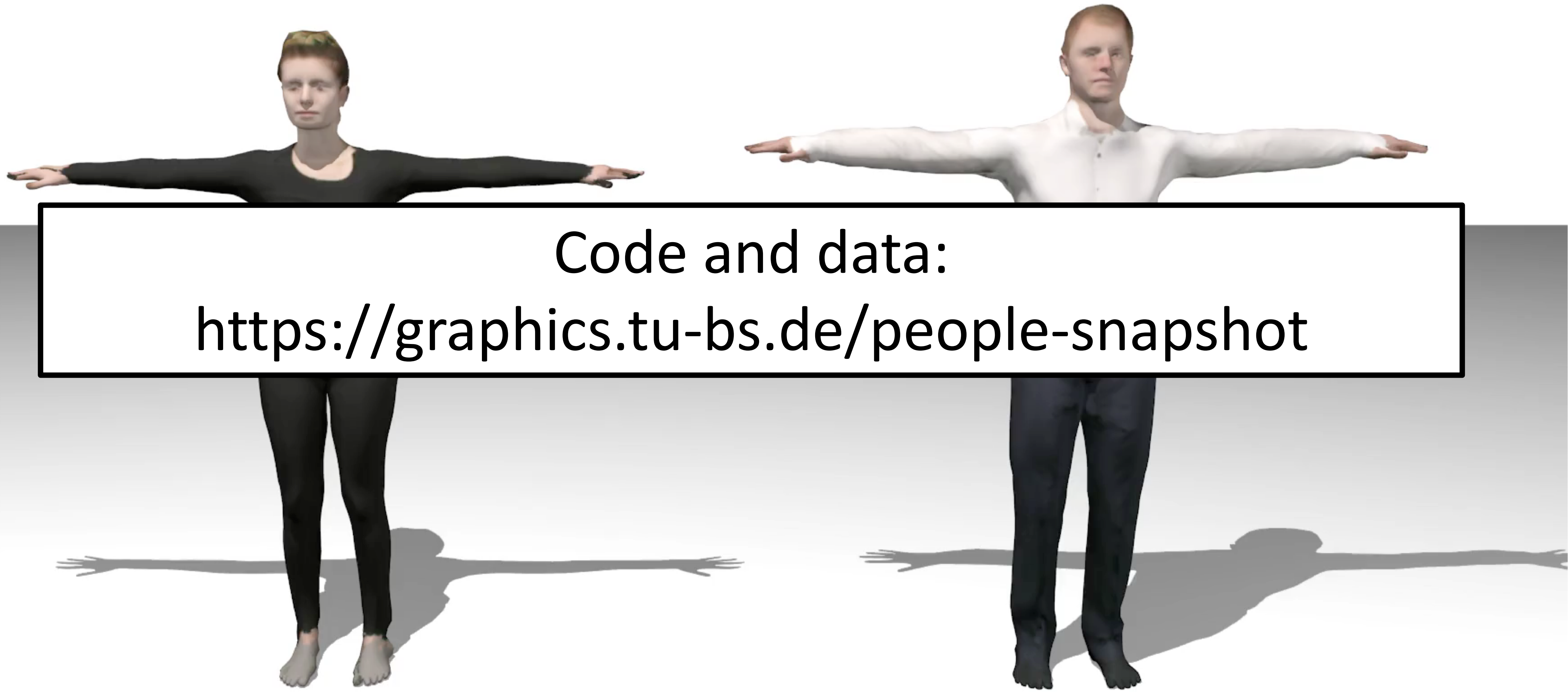
Sum of **point to line** distances

Prior Terms:

- Symmetry
- Prior on Shape
- Surface Smoothness

Alldieck et al. 3DV '18


AUTO



Code and data:

https://graphics.tu-bs.de/people-snapshot

Limitations

- Need to optimize the 3D pose at each frame – slow process
- Requires multiple frames ~100
- Optimization susceptible to local minima

Learning to Reconstruct People in Clothing from a Single RGB Camera

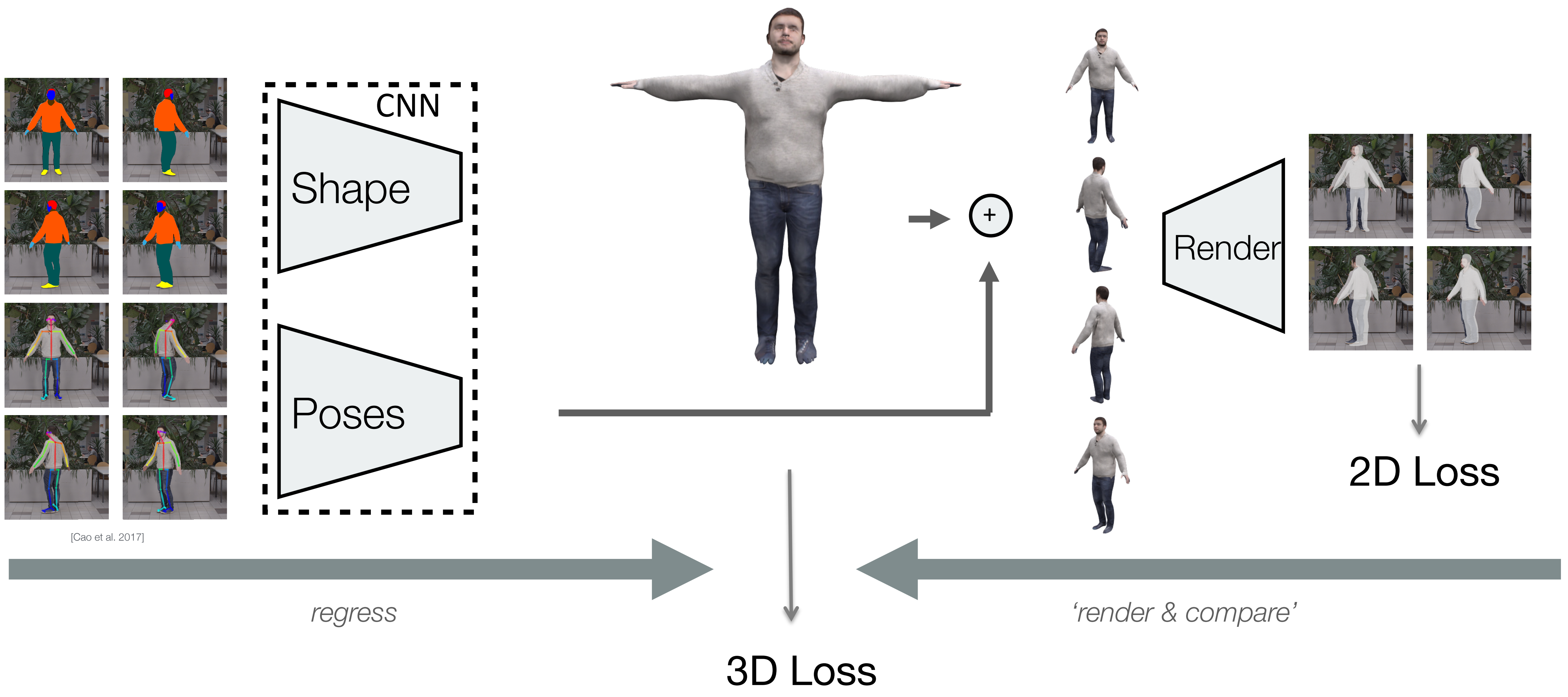




T. Alldieck M. Magnor B. Bhatnagar C. Theobalt G. Pons-Moll

CVPR'19

Octopus



CNN
Shape
Poses
+
Render
2D Loss
[Cao et al. 2017]
regress
'render & compare'
3D Loss

Dataset and Training

Results

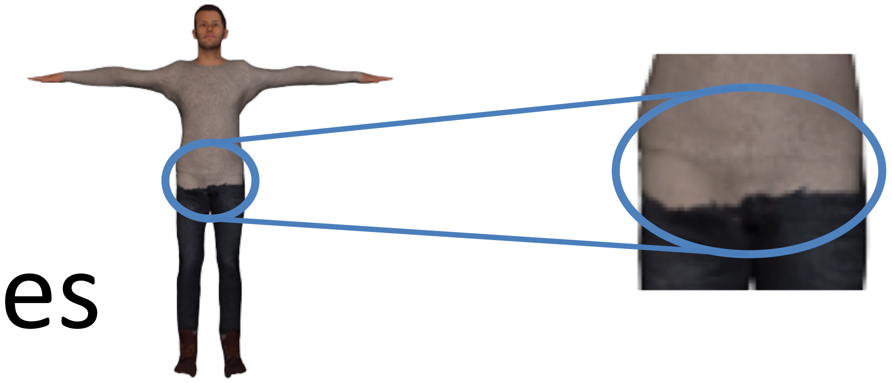
Amount of 3D vs 2D supervision

	Before optimization	After optimization
100%	4.47 ±4.41	3.17 ±3.41
50%	4.57 ±4.52	3.19 ±3.43
20%	4.74 ±4.65	3.29 ±3.53
10%	4.73 ±4.56	3.46 ±3.62

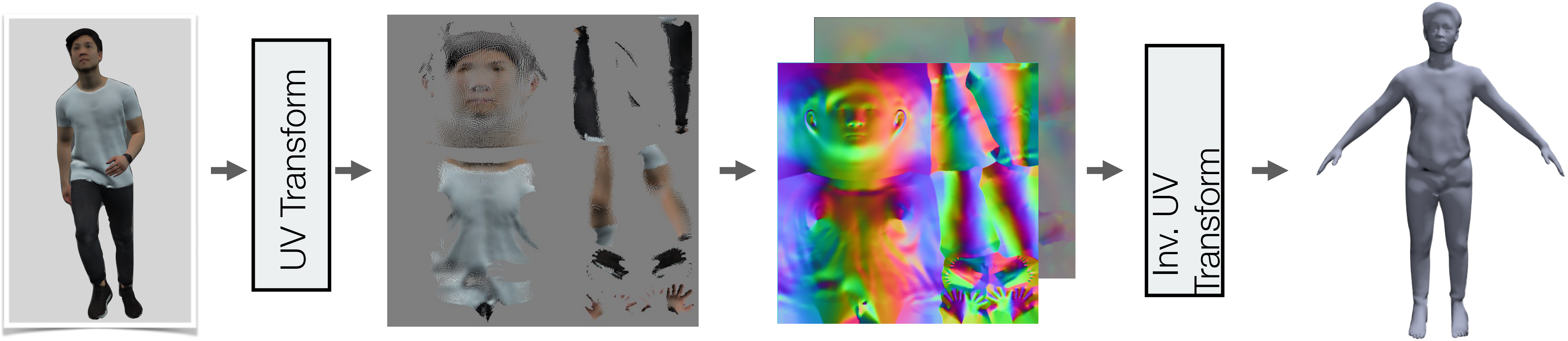
Small Gap

Remaining Problems (with the Representation)

- Lack of geometric detail
- Un-sharp clothing boundaries
- Can not retarget to new bodies
- Limited to clothing with “body topology”



Tex2Shape: Detailed Full Human Body Geometry from a Single Image



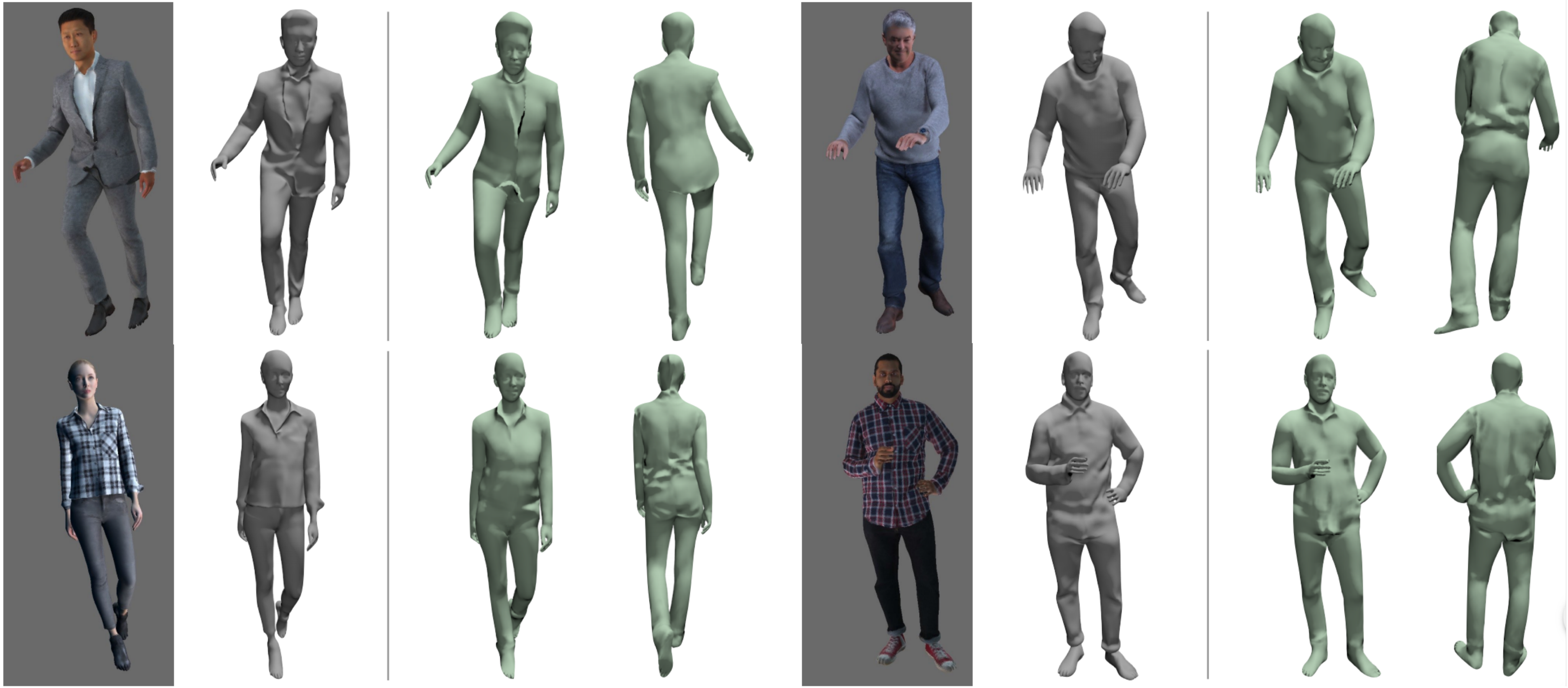
Input

UV Unwrap

Normals and 3D displacements

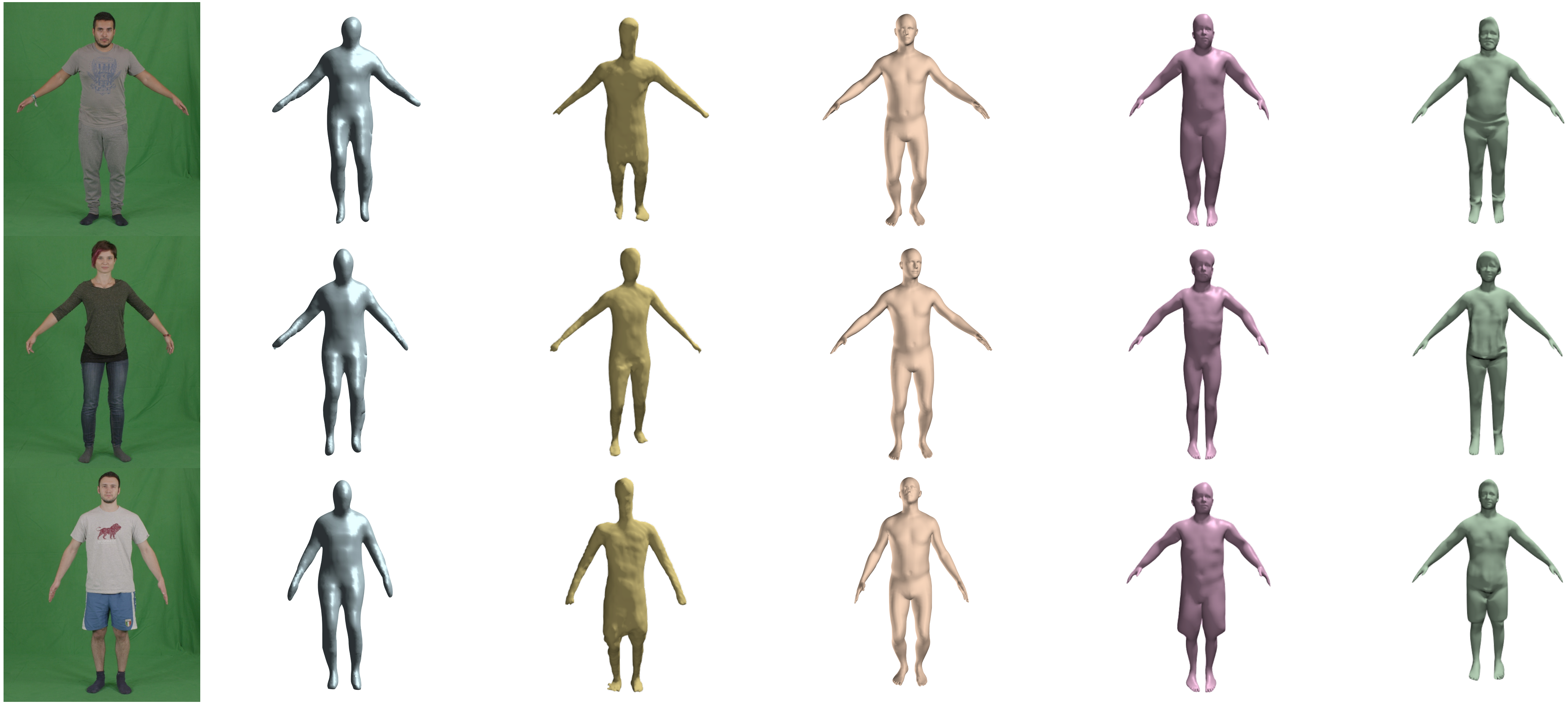
Output

Alldieck et al. ICCV'19



Results

Comparisons



BodyNet
Gül et al. ECCV'18

SciCloPe
Natsume et al.
CVPR'19

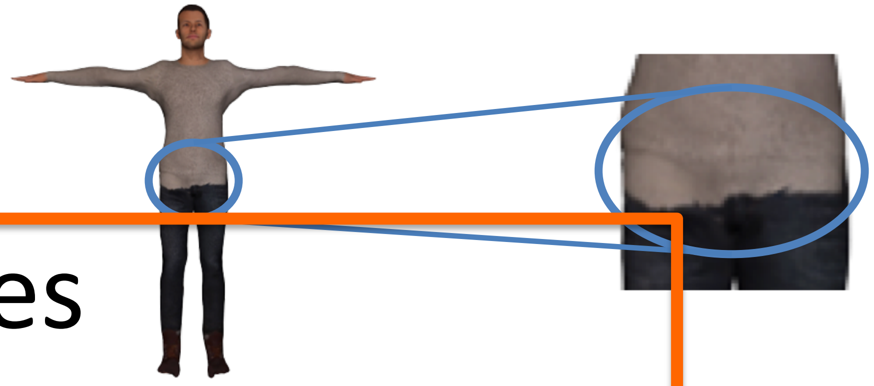
HMR
Kanazawa et al.
CVPR'18

Our CVPR18
(120 frames)

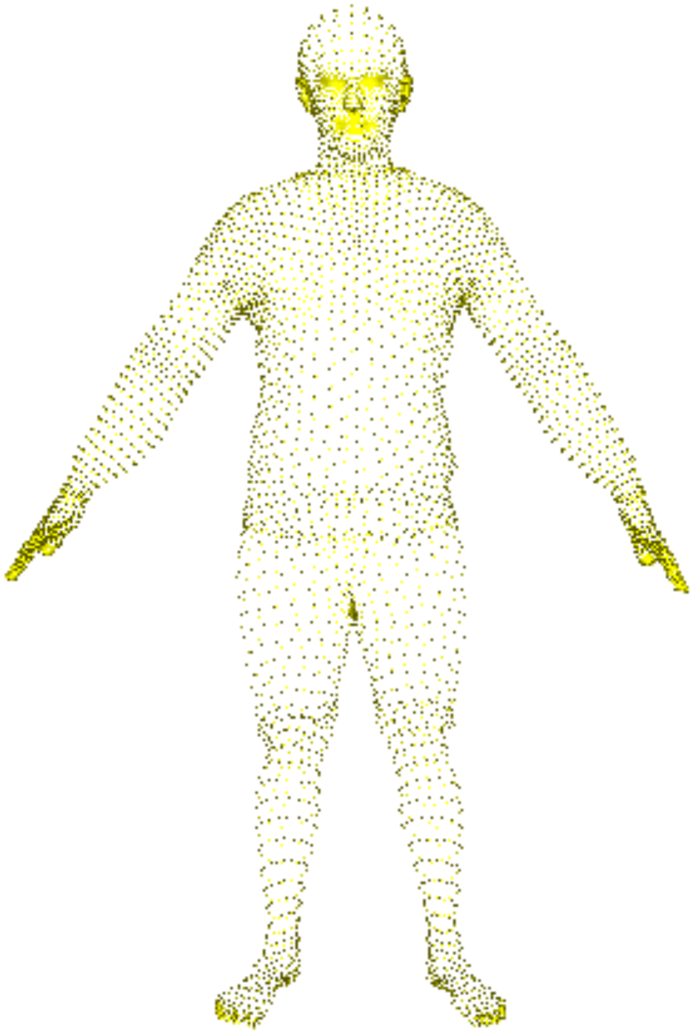
Tex2Shape

Remaining Problems (with the Representation)

- Lack of geometric detail
- Un-sharp clothing boundaries
- Can not retarget to new bodies
- Limited to clothing with “body topology”



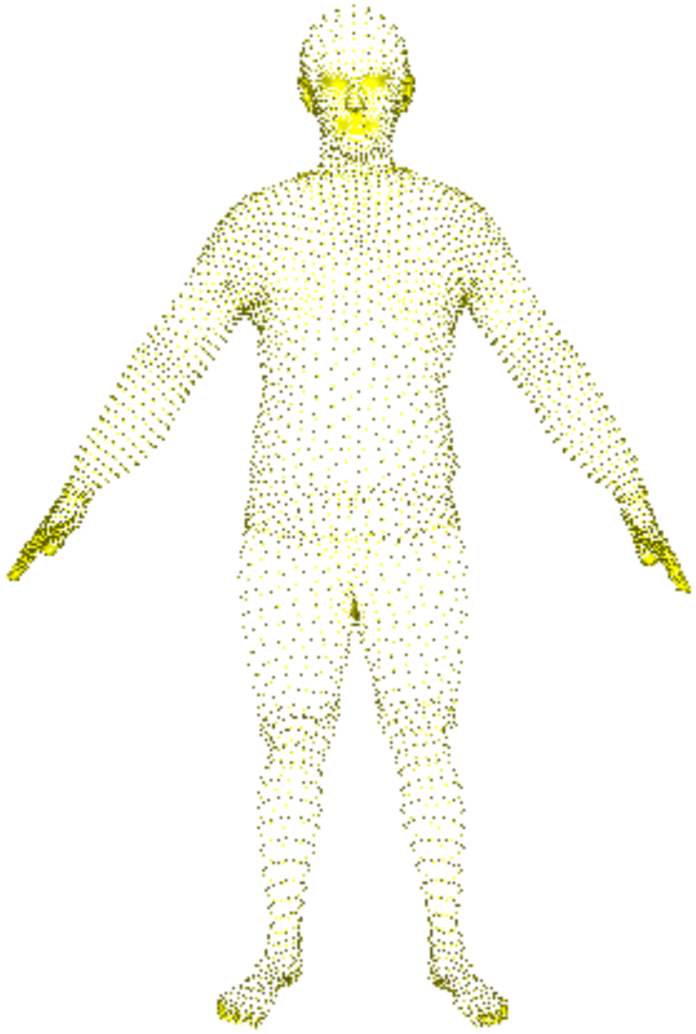
Capture and Dress People Digitally



Just a scan!
- Un-ordered point cloud
- **No control**: can not change shape, motion, clothing
- Useless without further processing

Pons-Moll et al.
SIGGRAPH 2017

Capture and Dress People Digitally



Pons-Moll et al.
SIGGRAPH 2017

Capture and Dress People Digitally

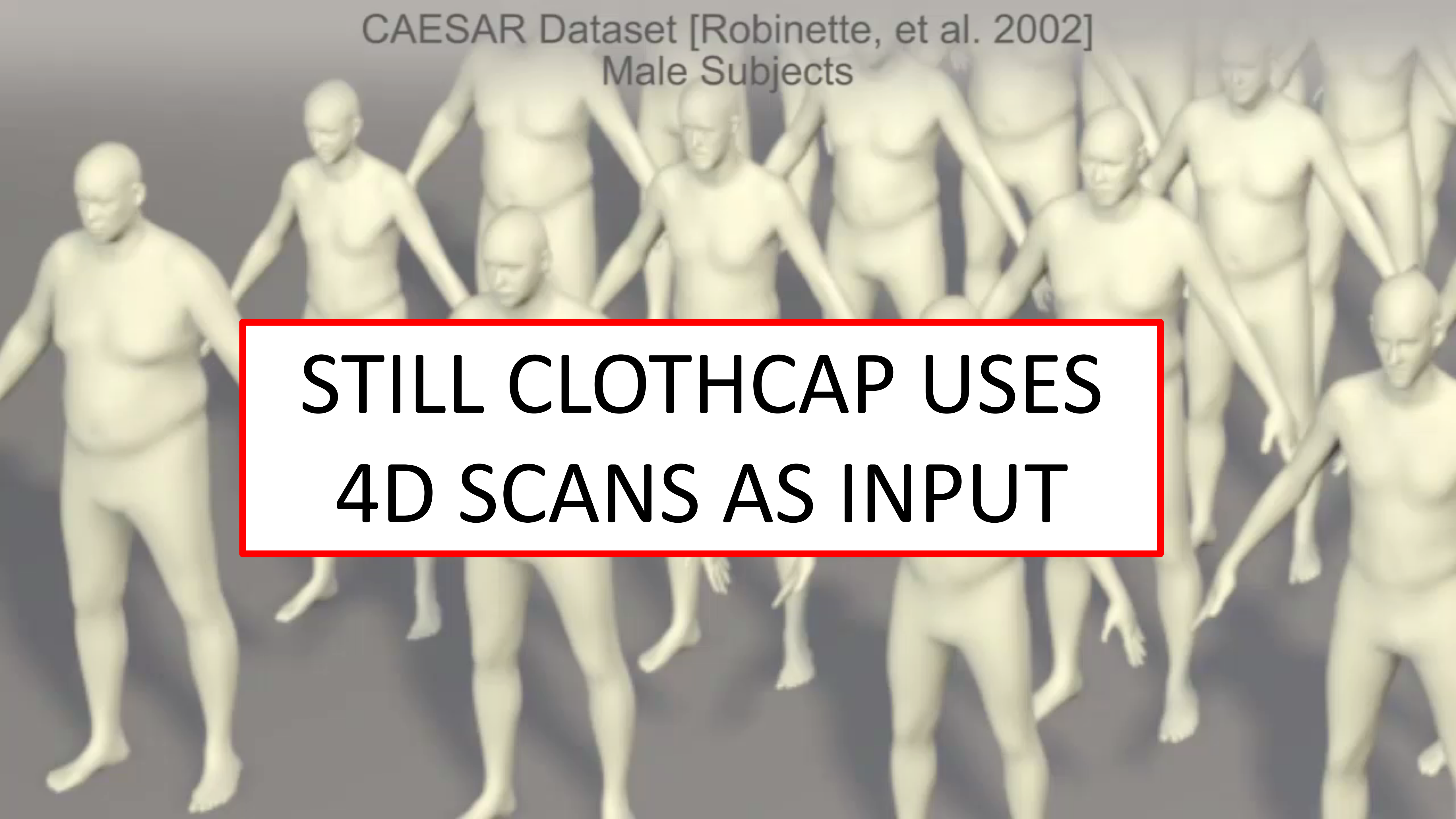
Pons-Moll et al.
SIGGRAPH 2017

Capture and Dress People Digitally

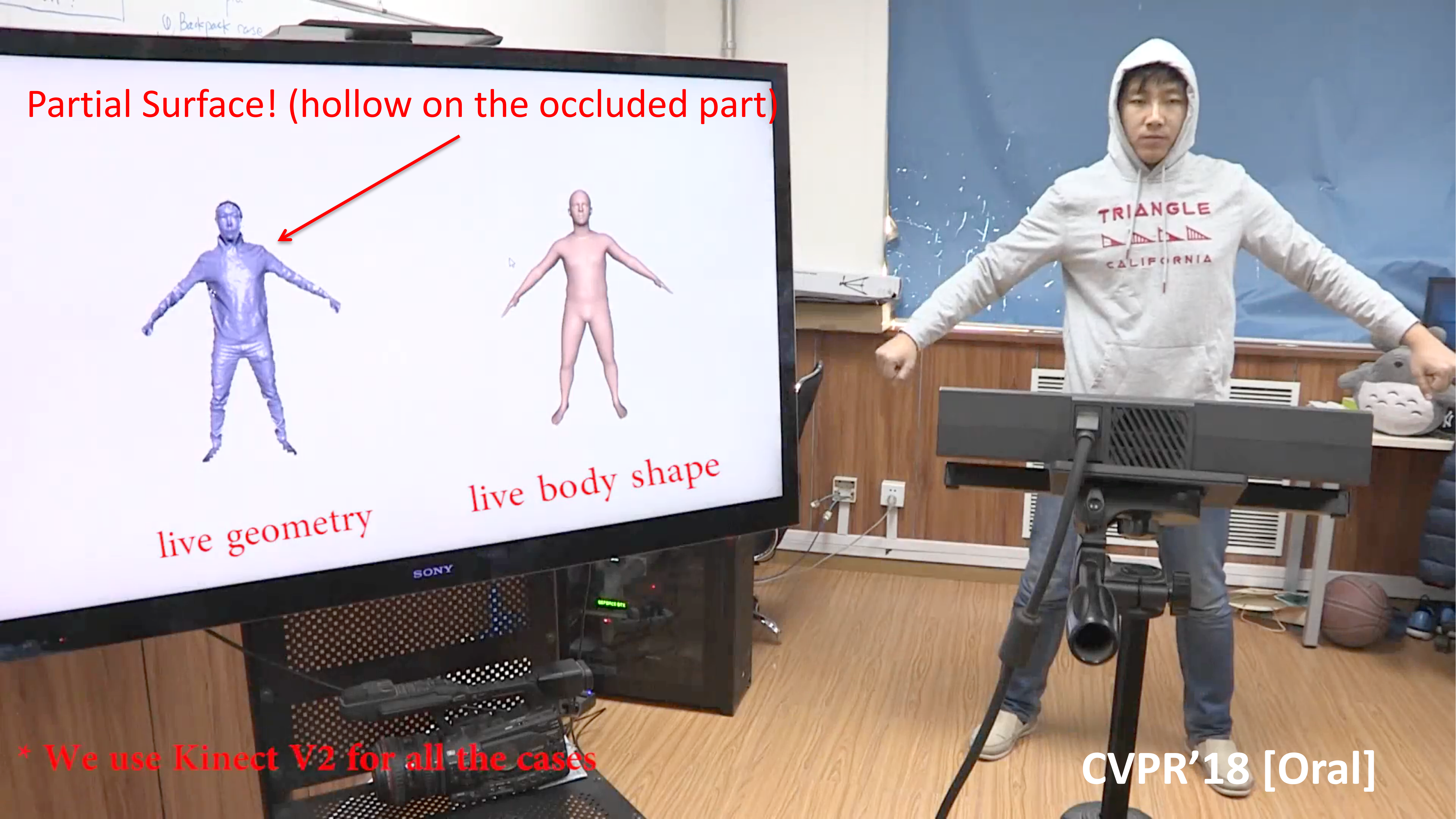
Pons-Moll et al.
SIGGRAPH 2017

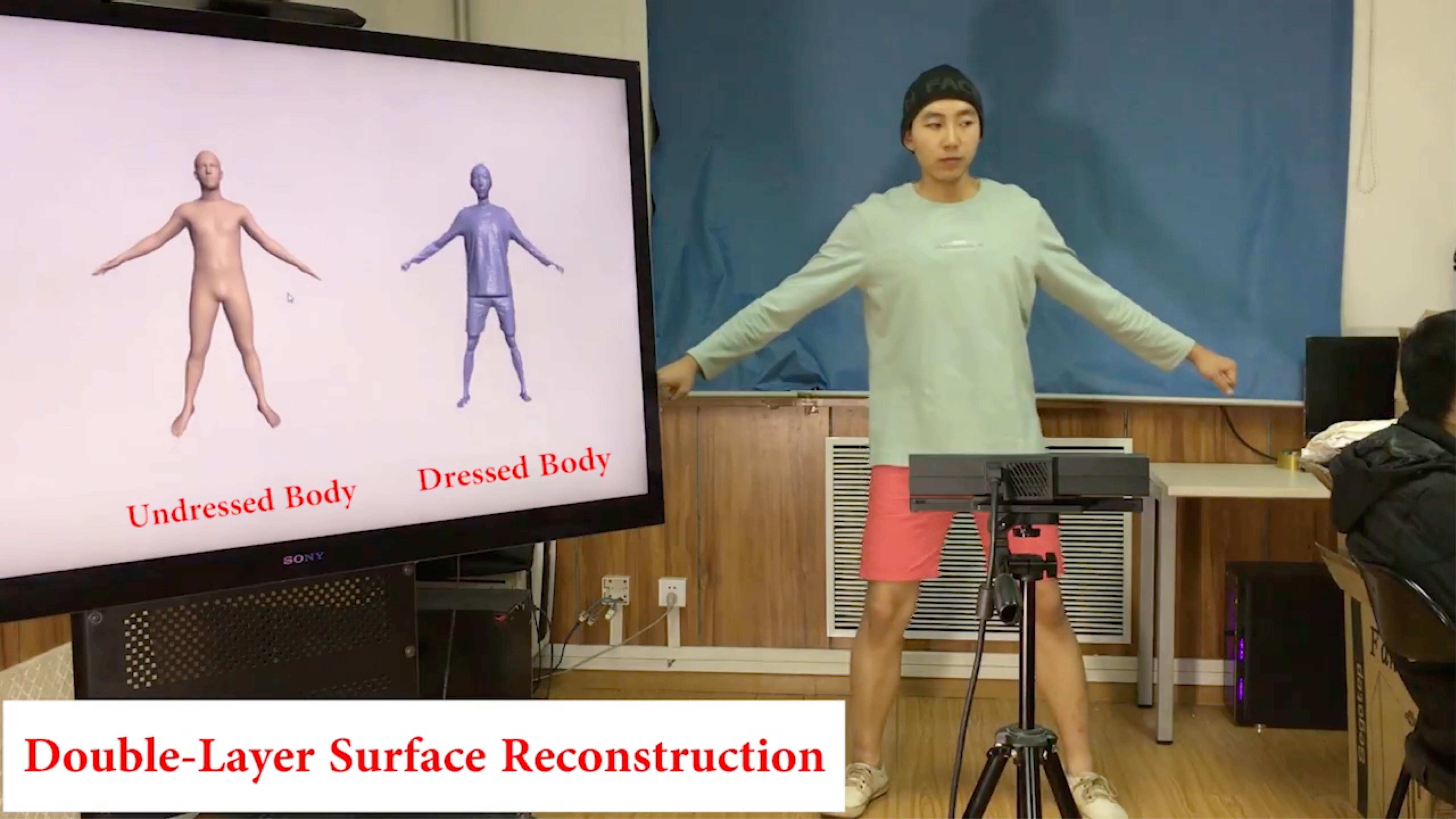
Capture and Dress People Digitally

Pons-Moll et al.
SIGGRAPH 2017

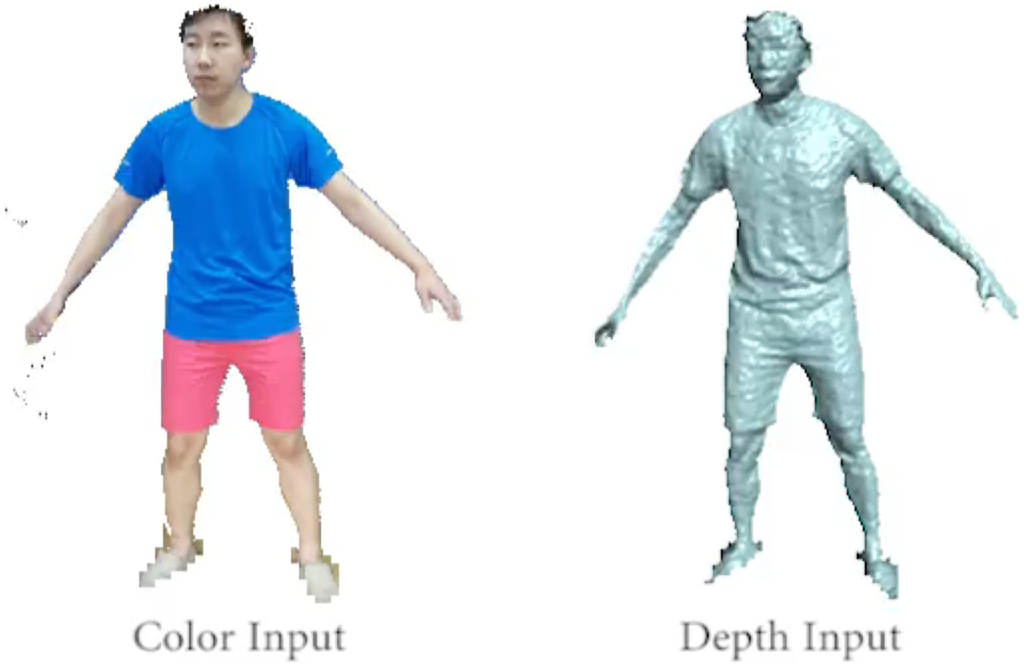


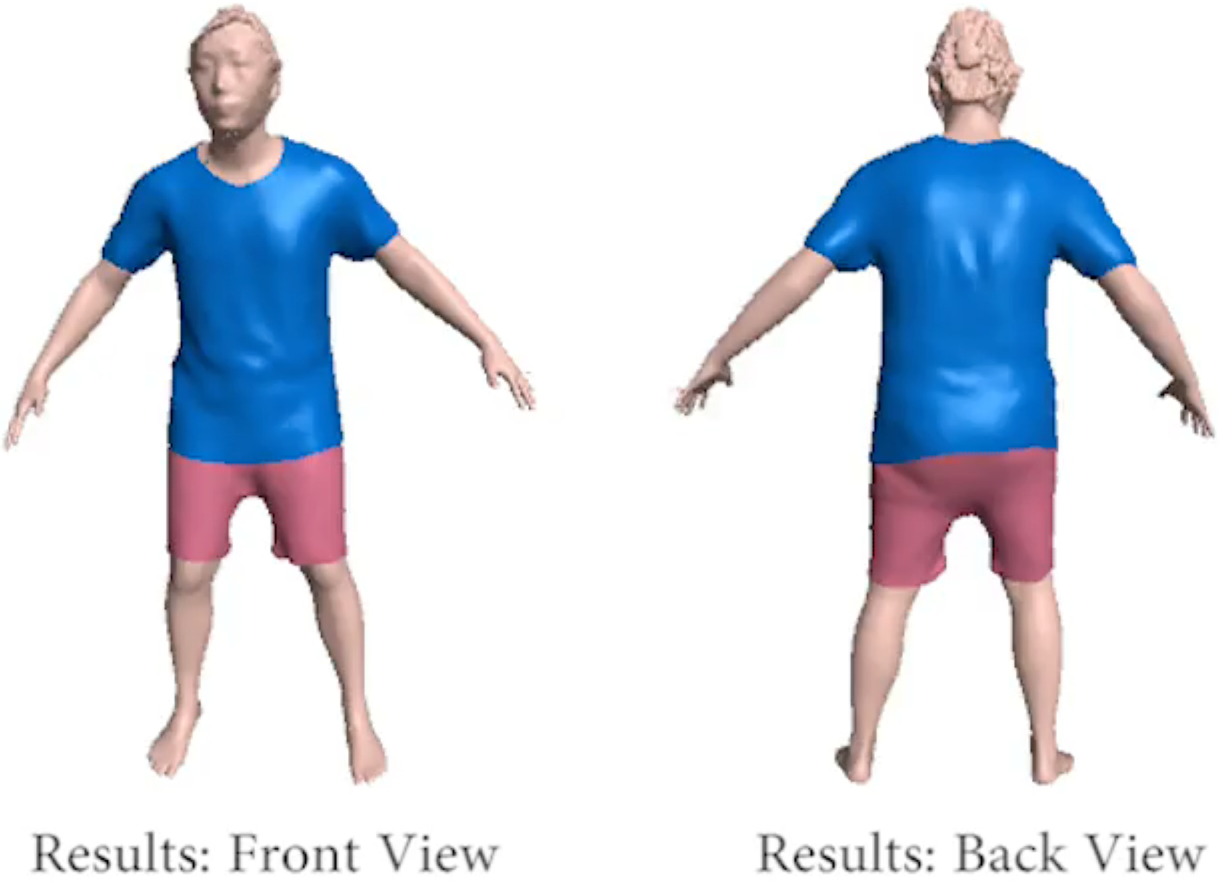
STILL CLOTHCAP USES 4D SCANS AS INPUT


Partial Surface! (hollow on the occluded part)
live geometry
live body shape
SONY
TRIANGLE
CALIFORNIA
* We use Kinect V2 for all the cases
CVPR'18 [Oral]


Undressed Body
Dressed Body
SONY
Double-Layer Surface Reconstruction

Performance capture with multi-layer and physics simulation





Tao et al. CVPR’19

Different topologies

Color Input

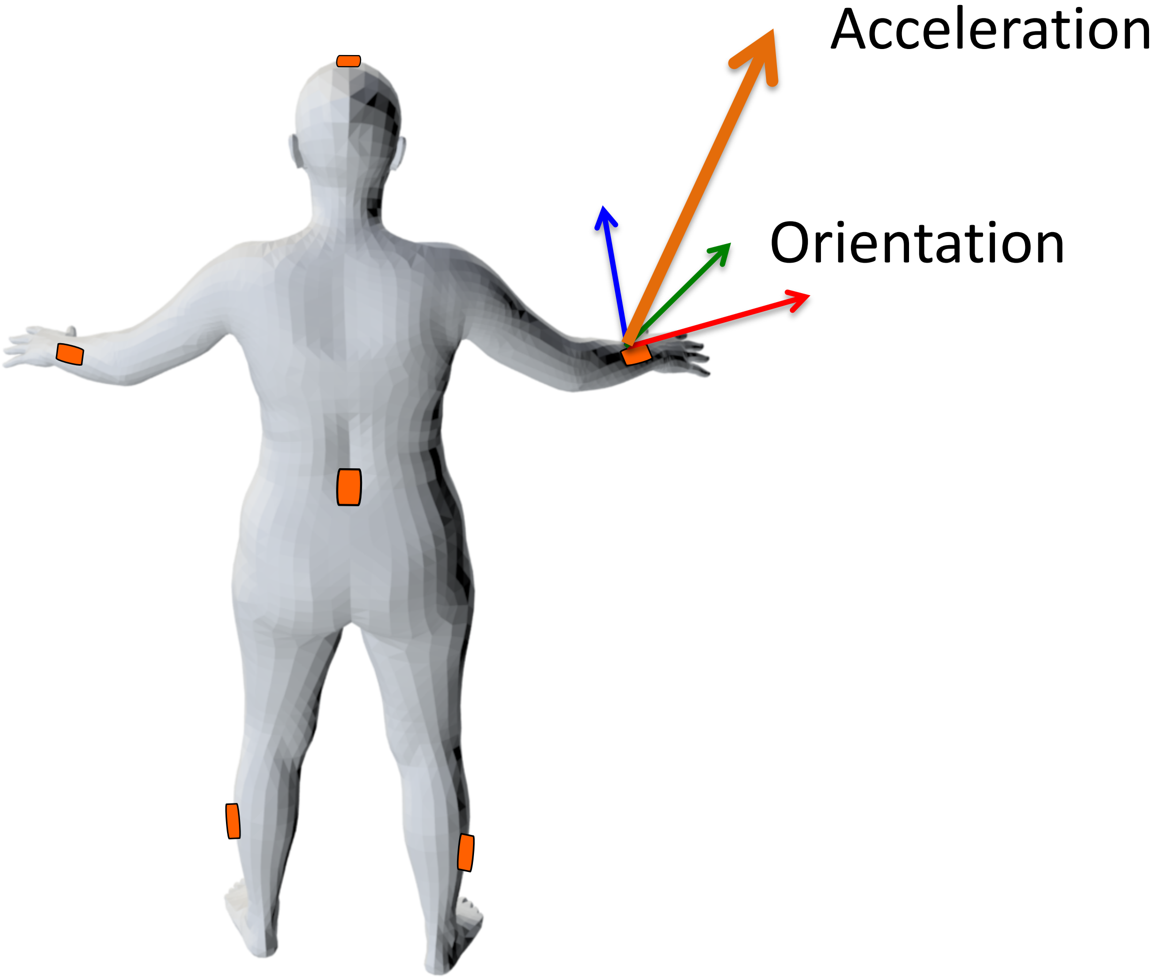
Depth Input

Results: Front View

Results: Back View

Motion Capture from Sparse IMUs

IMU = Inertial Measurement Unit
(Xsens)



Sparse Inertial Poser

Automatic 3D Human Pose Estimation from Sparse IMUs

Supplementary material

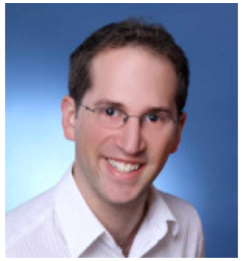
Eurographics'17
Paper ID 1112

T. Marcard, B. Rosenhahn, M. Black, G. Pons-Moll. **Eurographics ’17** **Best Paper Award**

Climbing

Recovering Accurate 3D Human Pose in the Wild Using IMUs and a Moving Camera

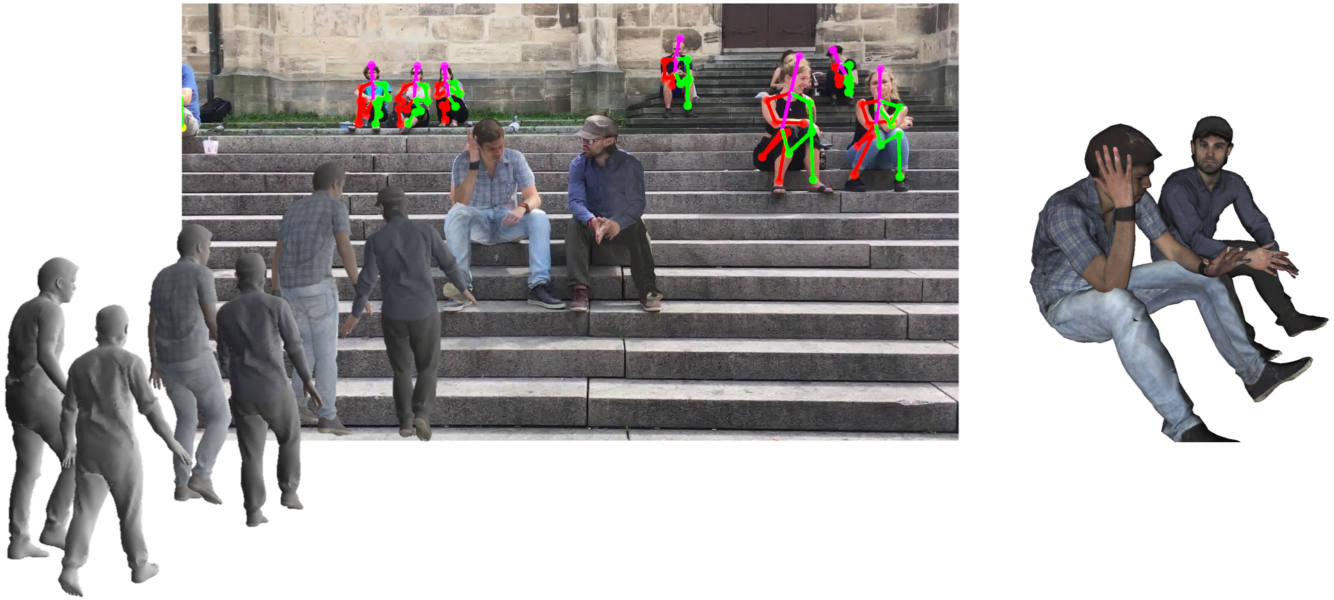

T. von Marcard


R. Henschel


M. Black

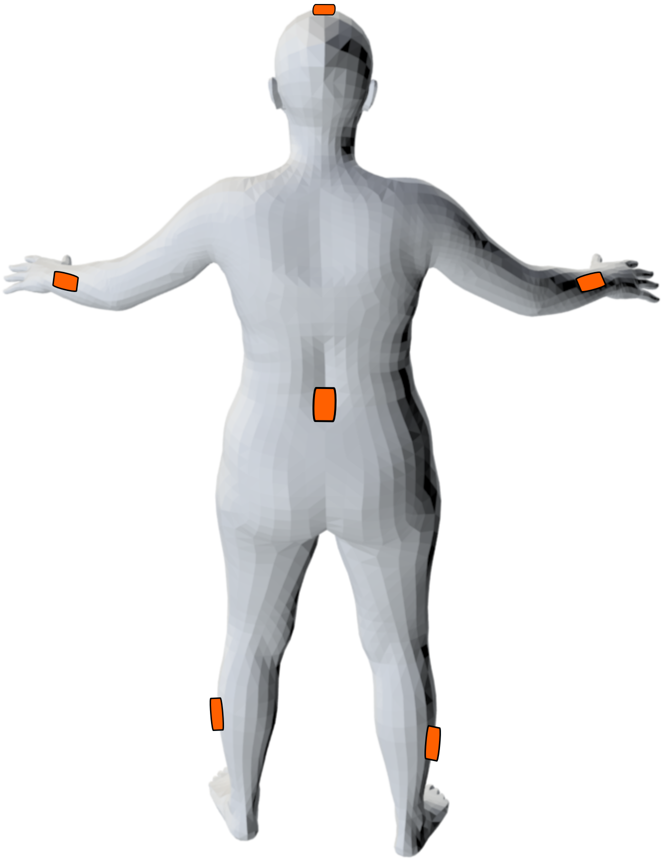

B. Rosenhahn


G. Pons-Moll

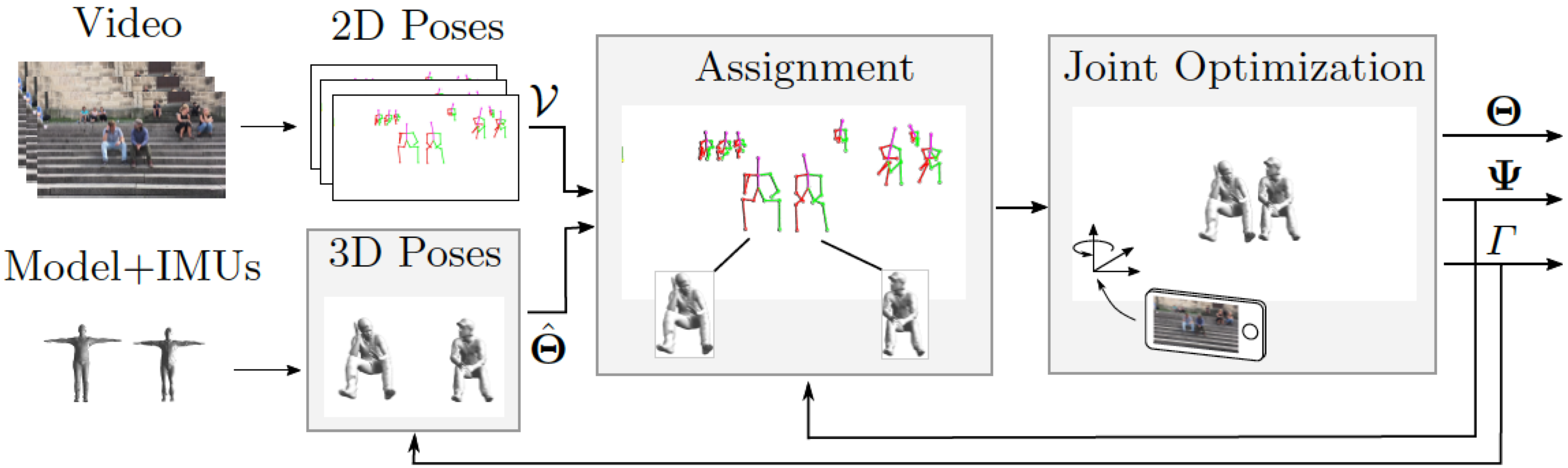


ECCV'18

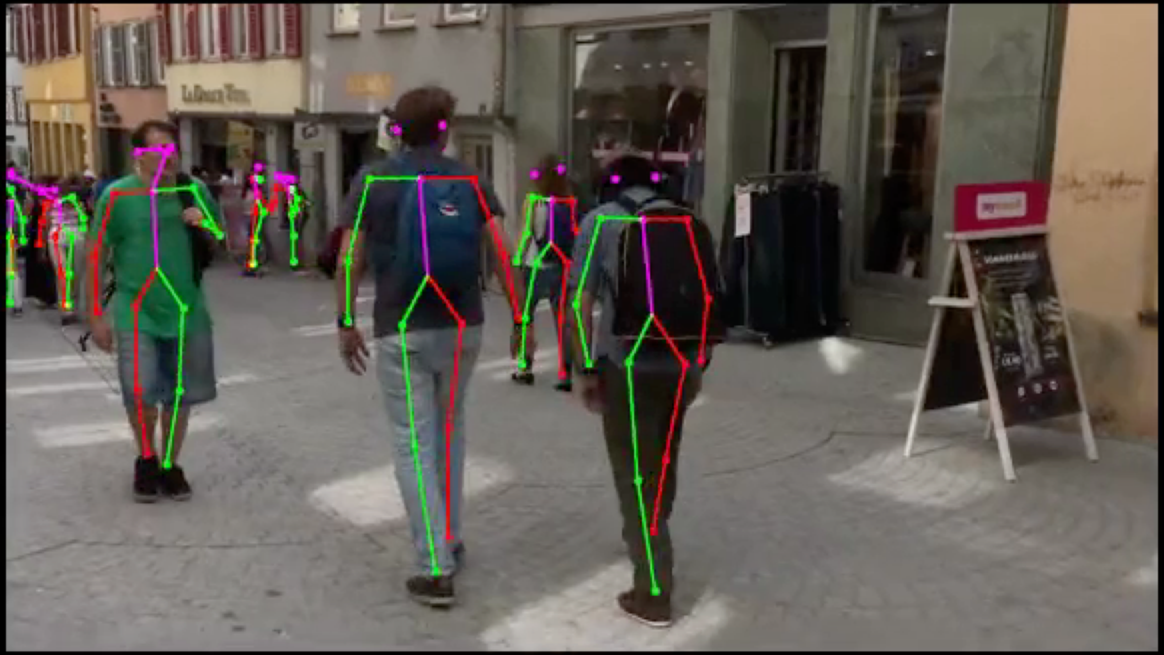
3DPW: 3D Poses in the Wild



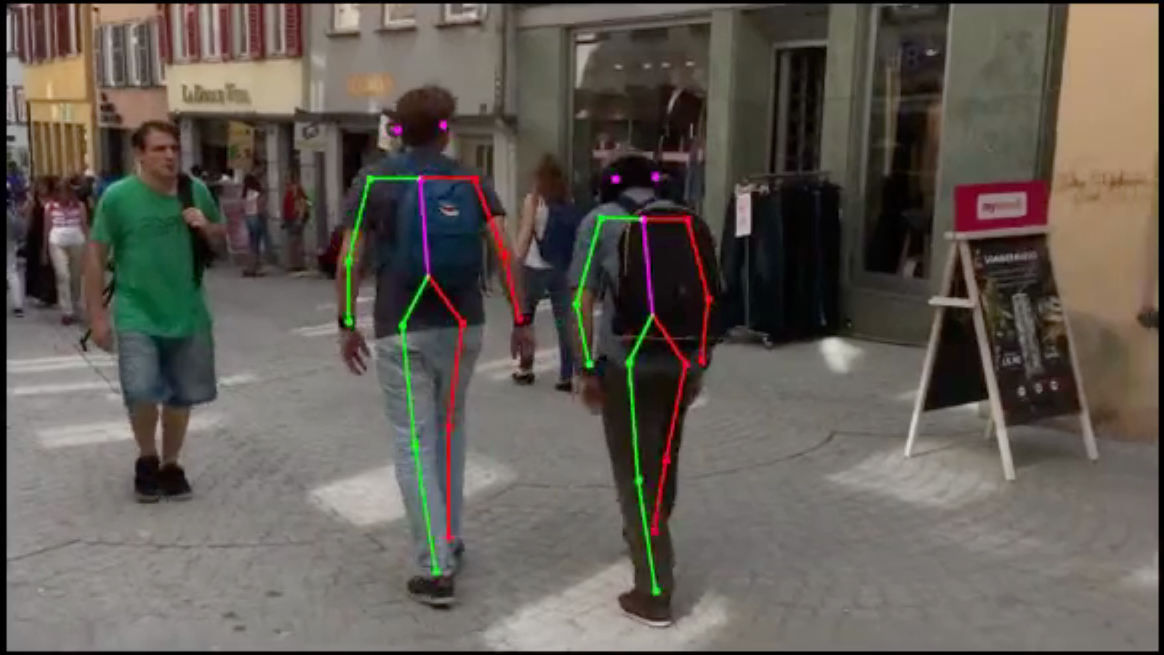
A single moving camera and IMUs on the person



Person Identification

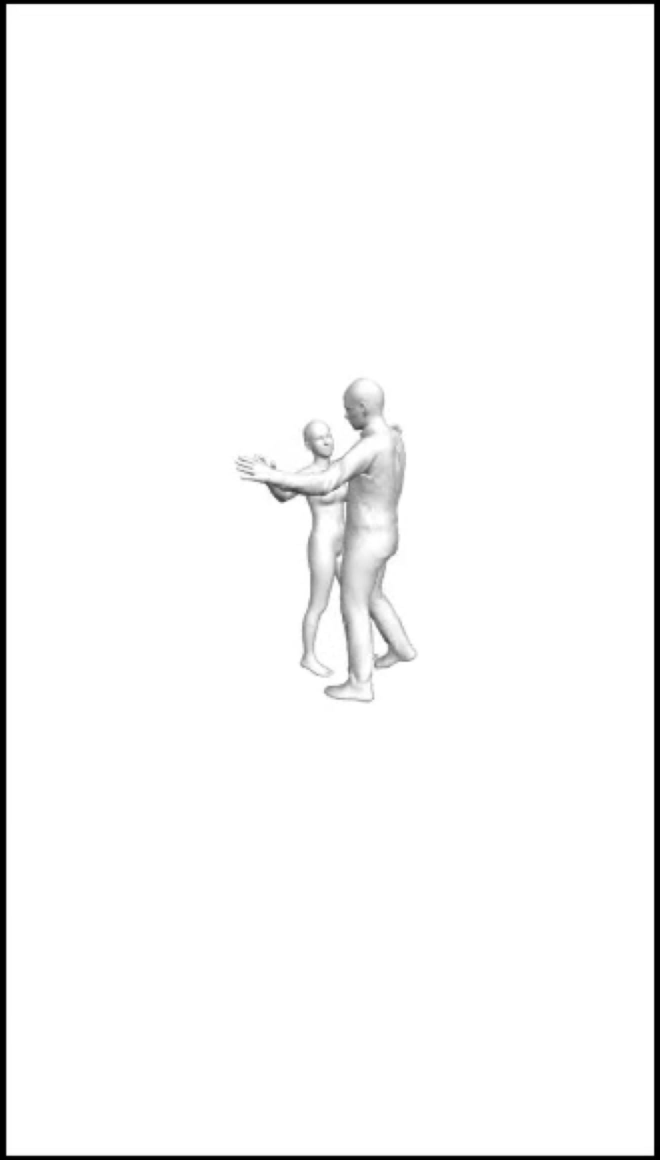


All 2D Poses

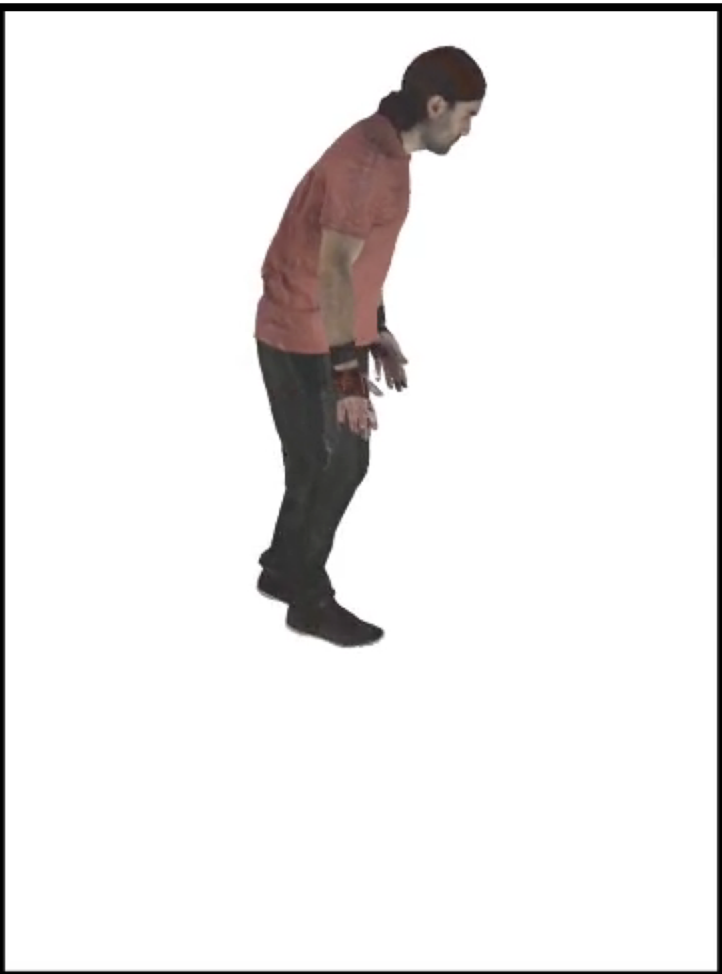


Assigned 2D Poses

3D Pose Estimation

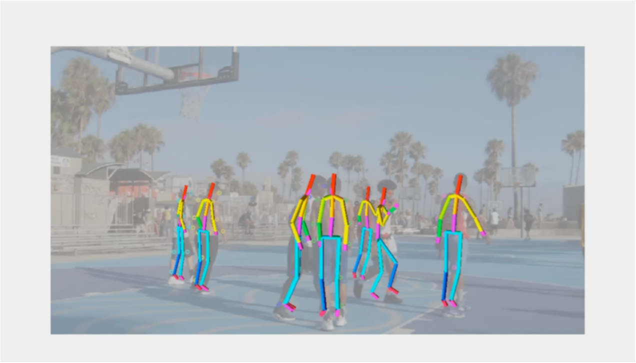


Full dataset available:
http://virtualhumans.mpi-inf.mpg.de/3DPW/

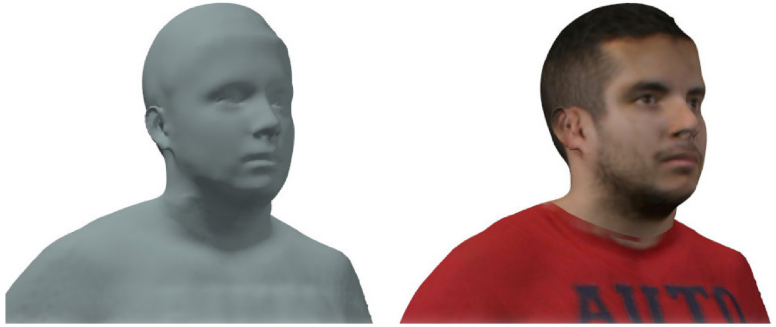


3DPW

- 60 video sequences.
- 2D pose annotations.
- 3D reference poses.
- Camera poses for every frame in the sequences.
- 3D body scans and 3D people models (re-poseable and re-shapeable). Each sequence contains its corresponding models.
- 18 3D models in different clothing variations.



Multiple People (3DV’18)
D. Mehta, O. Sotnychenko, F. Mueller, Weipeng Xu, S.Sridha, G.Pons-Moll, C. Theobalt



Detailed Human Avatars from Monocular Video
T. Alldieck, M. Magnor, W. Xu,C. Theobalt, G.Pons-Moll

Real-Time Monocular Performance Capture
M. Habermann, W. Xu, M. Zollhoefer, G.Pons-Moll, C. Theobalt

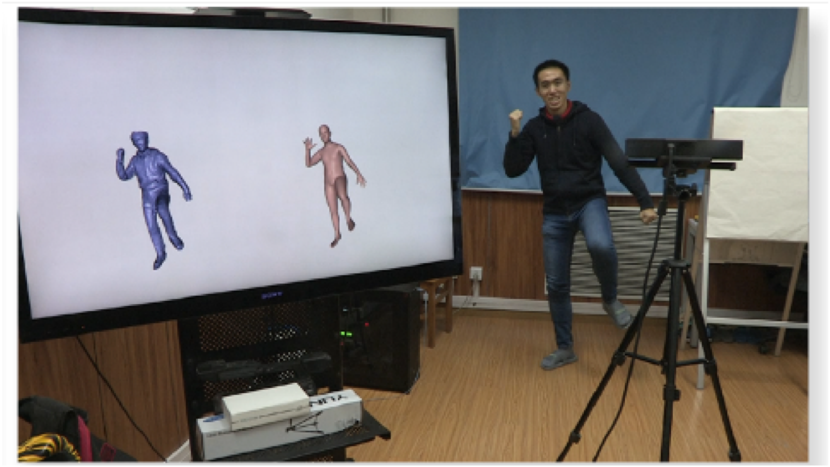
Shape and Motion from Markers
N. Mahmood, G. Pons-Moll, N. Riza, N. Troje, M. Black

Recovering accurate 3D human pose in the wild using IMUs and a single camera (ECCV’18)
Marcard, Henschel, Black, Rosenhahn, Pons-Moll

Generating People with GANs
C.Lassner, G. Pons-Moll, P. Gehler ICCV’17

CONCLUSIONS

- To achieve **realism** we need to **learn** digital humans by capturing **real** ones
- **Clothing** is one of the main **missing** components in current statistical body models
- We need **perception** algorithms that **reason** about the **3D world**, not about pixels



Real Virtual Humans

DATA & CODE:

https://virtualhumans.mpi-inf.mpg.de/software.html

Open positions: Postdoc, PhD, Master

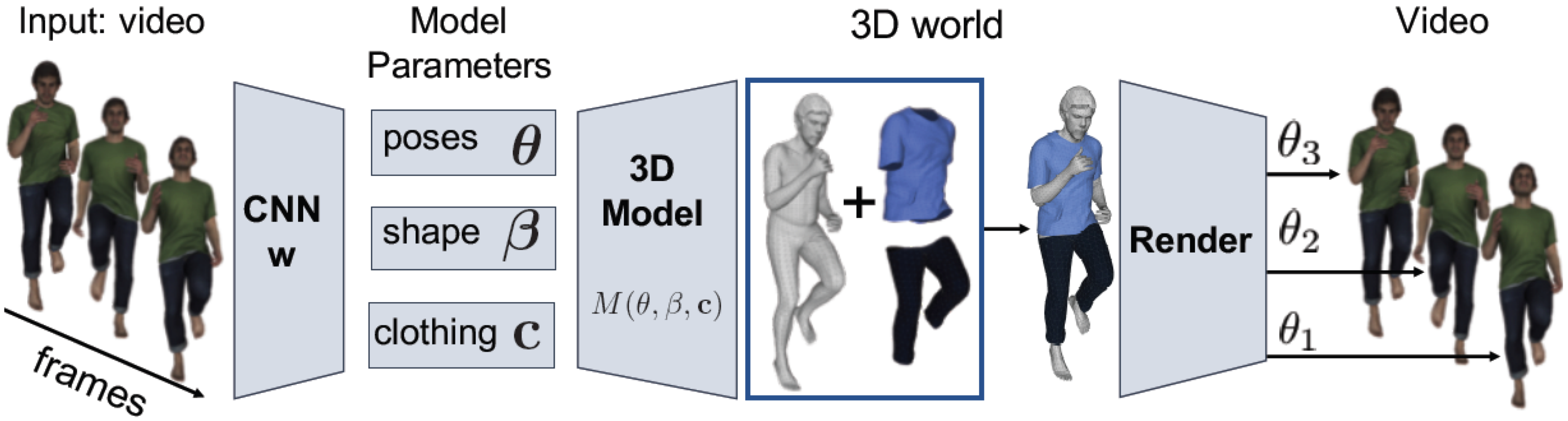
Emmy Noether-Programm
DFG Deutsche Forschungsgemeinschaft

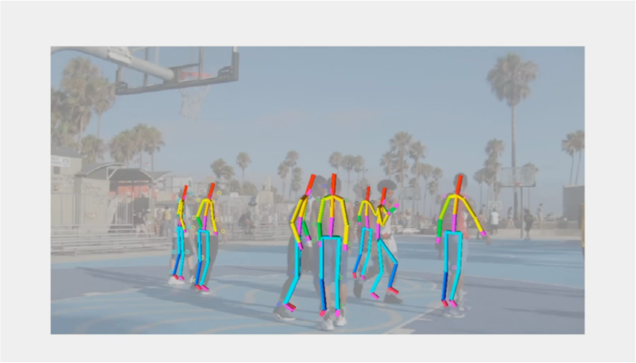
max planck institut informatik

facebook

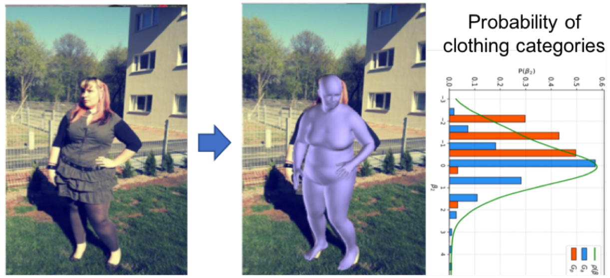
Google
Google Faculty Research Awards

Body and Garments Separately?





Multiple People (3DV'18)
D. Mehta, O. Sotnychenko, F. Mueller, Weipeng Xu, S.Sridha, G.Pons-Moll, C. Theobalt



Clothing Preferences
Shape Evasion
H. Sattar, G.Pons-Moll, M. Fritz

Real-Time Monocular Performance Capture
M. Habermann, W. Xu, M. Zollhoefer, G.Pons-Moll, C. Theobalt

Shape and Motion from Markers
N. Mahmood, G. Pons-Moll, N. Riza, N. Troje, M. Black

Generating People with GANs
C.Lassner, G. Pons-Moll, P. Gehler ICCV'17

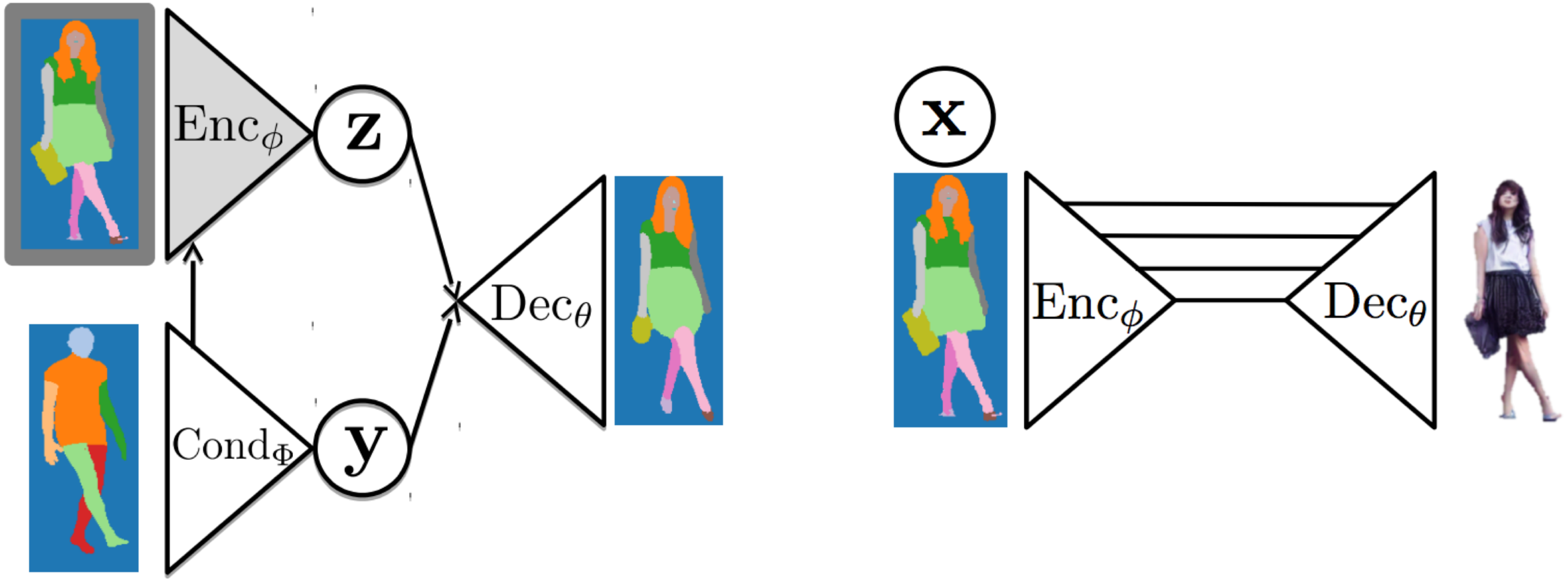
CAESAR Dataset [Robinette, et al. 2002]
Male Subjects

A generative model of people in clothing

Christoph Lassner, Gerard Pons-Moll and Peter Gehler

ICCV 2017 - Spotlight

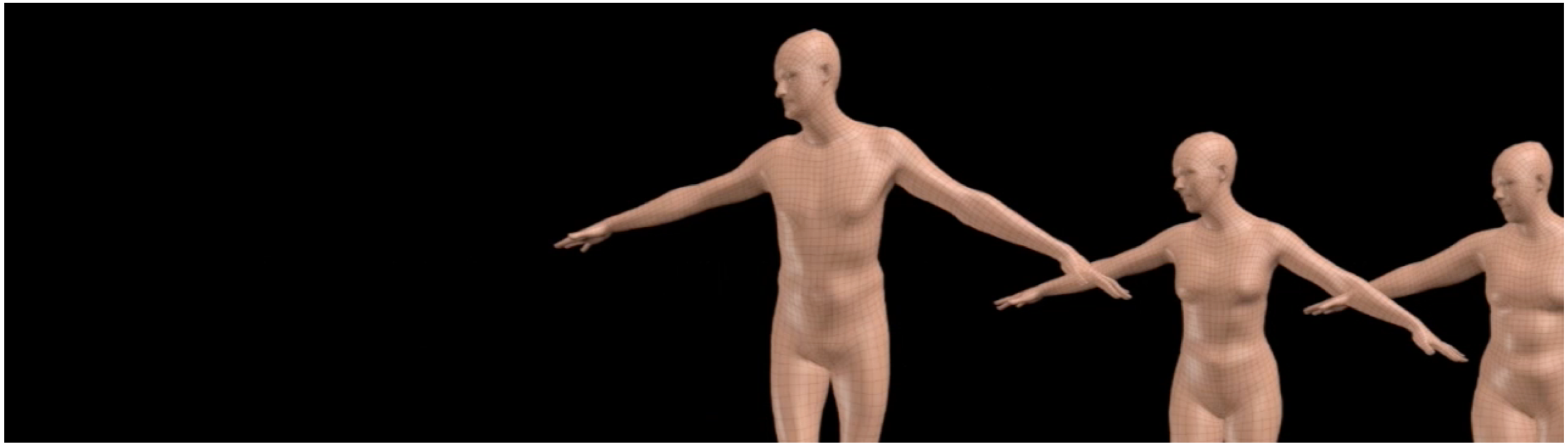
Variational auto-encoder and image translation network



Generate people in random fashion styles

Condition on color

SMPL: A model of pose and shape



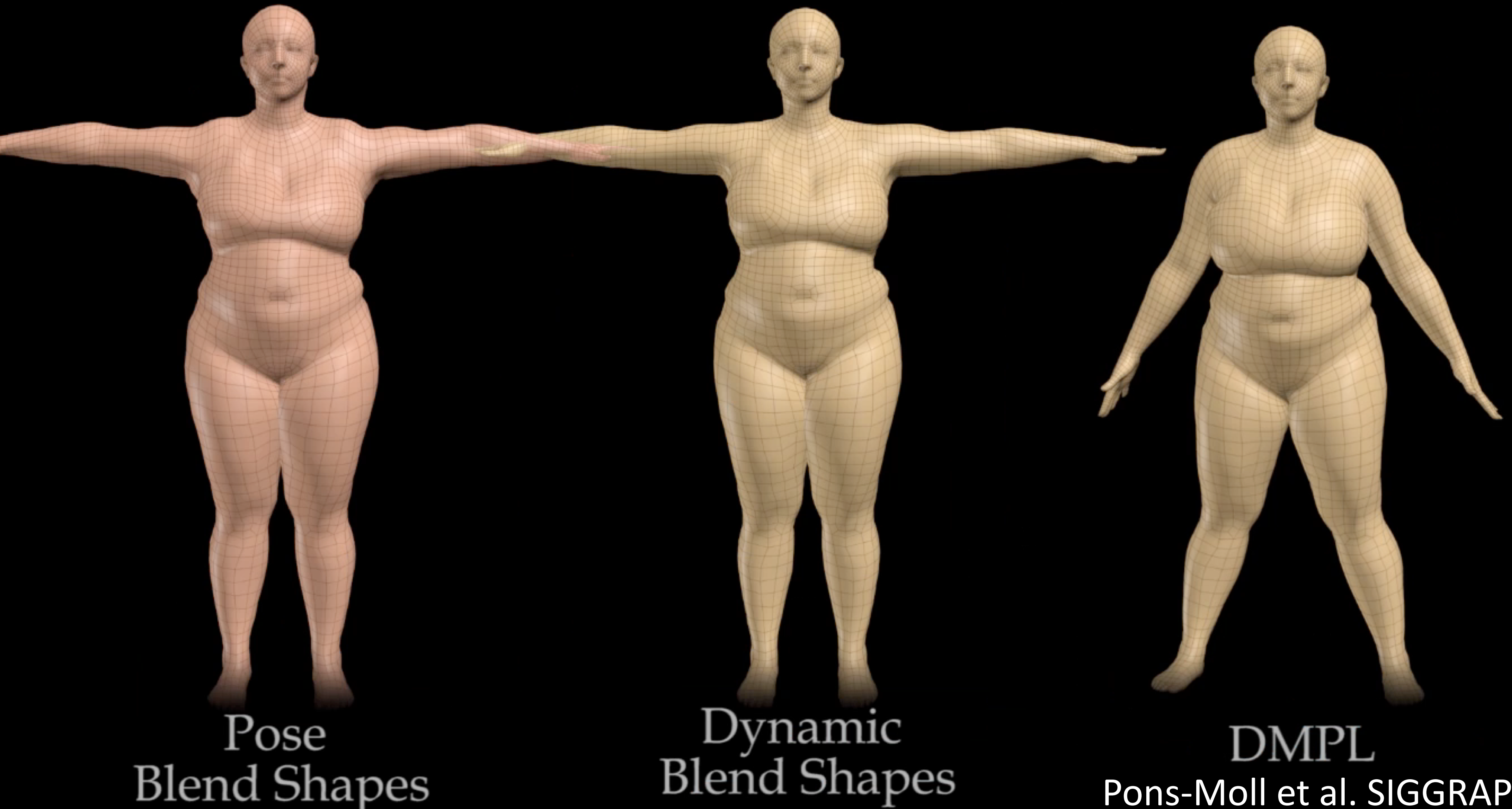
$$M(\theta, \beta; \mathbf{w}) : \mathbb{R}^{|\theta|+|\beta|} \mapsto \mathbb{R}^{3N}$$

$$\text{Latent parameters} \mapsto \text{vertices}$$

SMPL Model

M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, M. Black **SIGGRAPH Asia'15**

DYNA: A model soft-tissue



Pons-Moll et al. SIGGRAPH’15