Virtual Humans – Winter 23/24

Lecture 8_1 – Vertex based clothing

Prof. Dr.-Ing. Gerard Pons-Moll
University of Tübingen / MPI-Informatics

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Topics today

- Clothing representation as vertex displacements and how to do registration
- Predicting people in 3D clothing from images
- Learning a model of clothing as a function of pose, shape and style

Clothing Representation

SMPL + Clothing

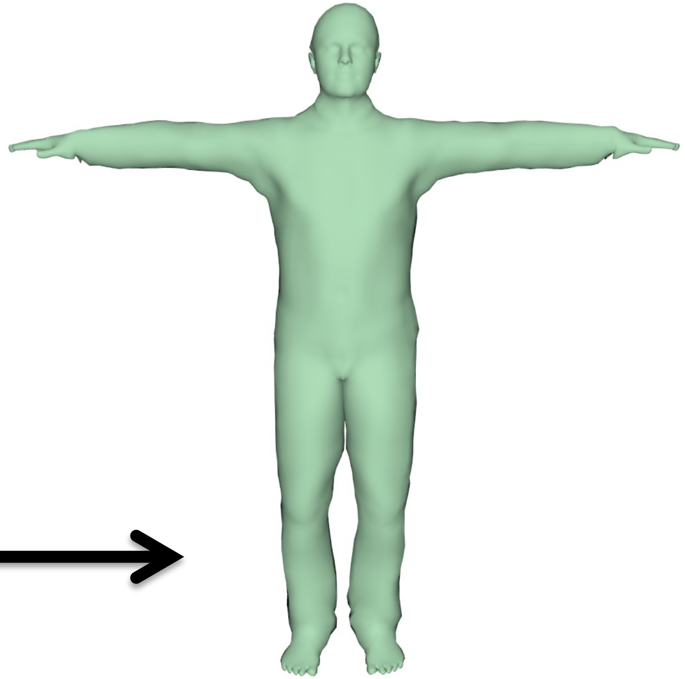
Vertices in a 0-pose

$$T(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{D}) = \mathbf{T}_{\mu} + B_{\boldsymbol{s}}(\boldsymbol{\beta}) + B_{\boldsymbol{p}}(\boldsymbol{\theta}) + \mathbf{D}$$

$\boldsymbol{\theta}$ Pose parameters

$\boldsymbol{\beta}$ Shape parameters

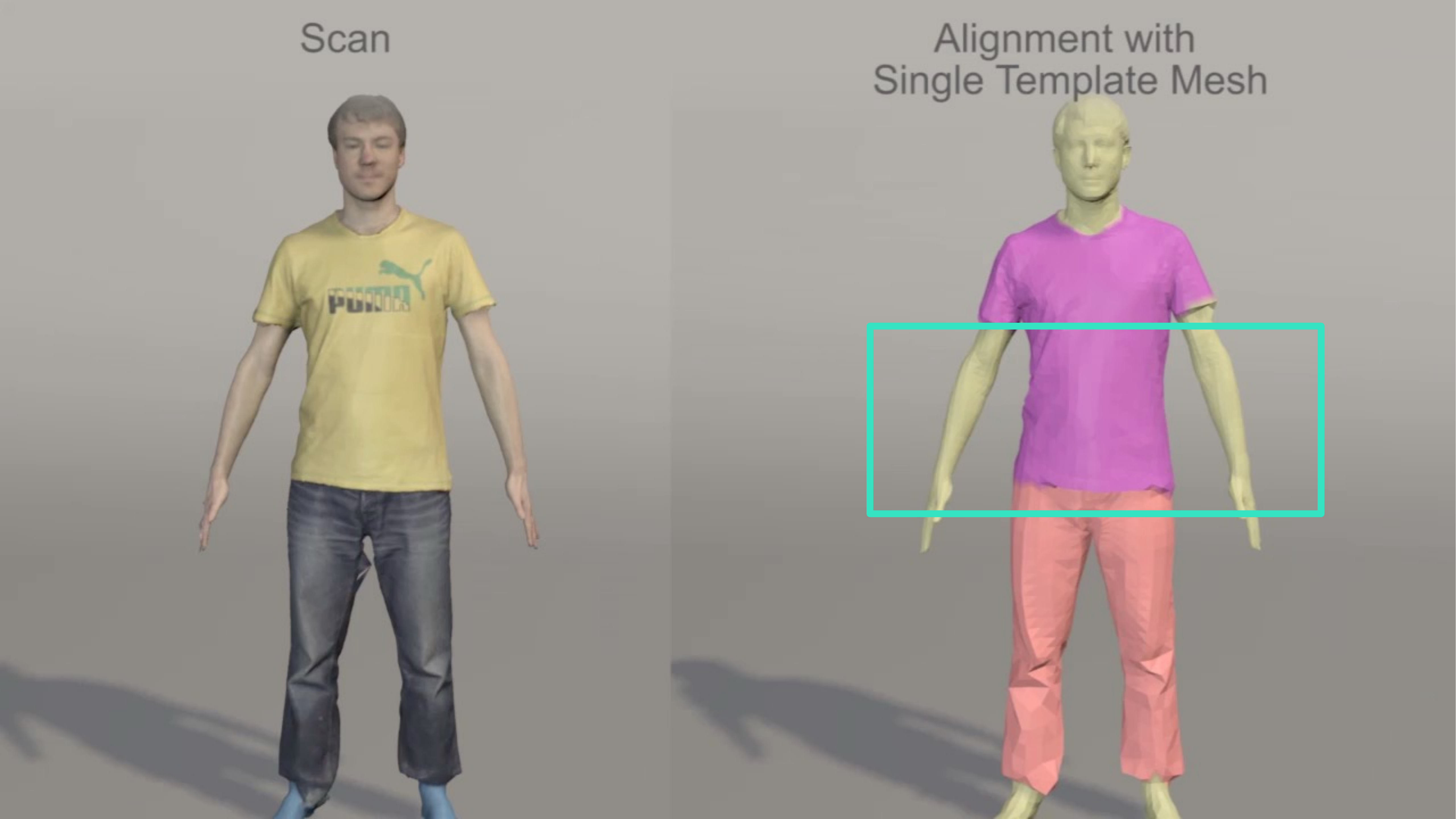
$\mathbf{D}$ Personal details + clothing



Registration with Clothing

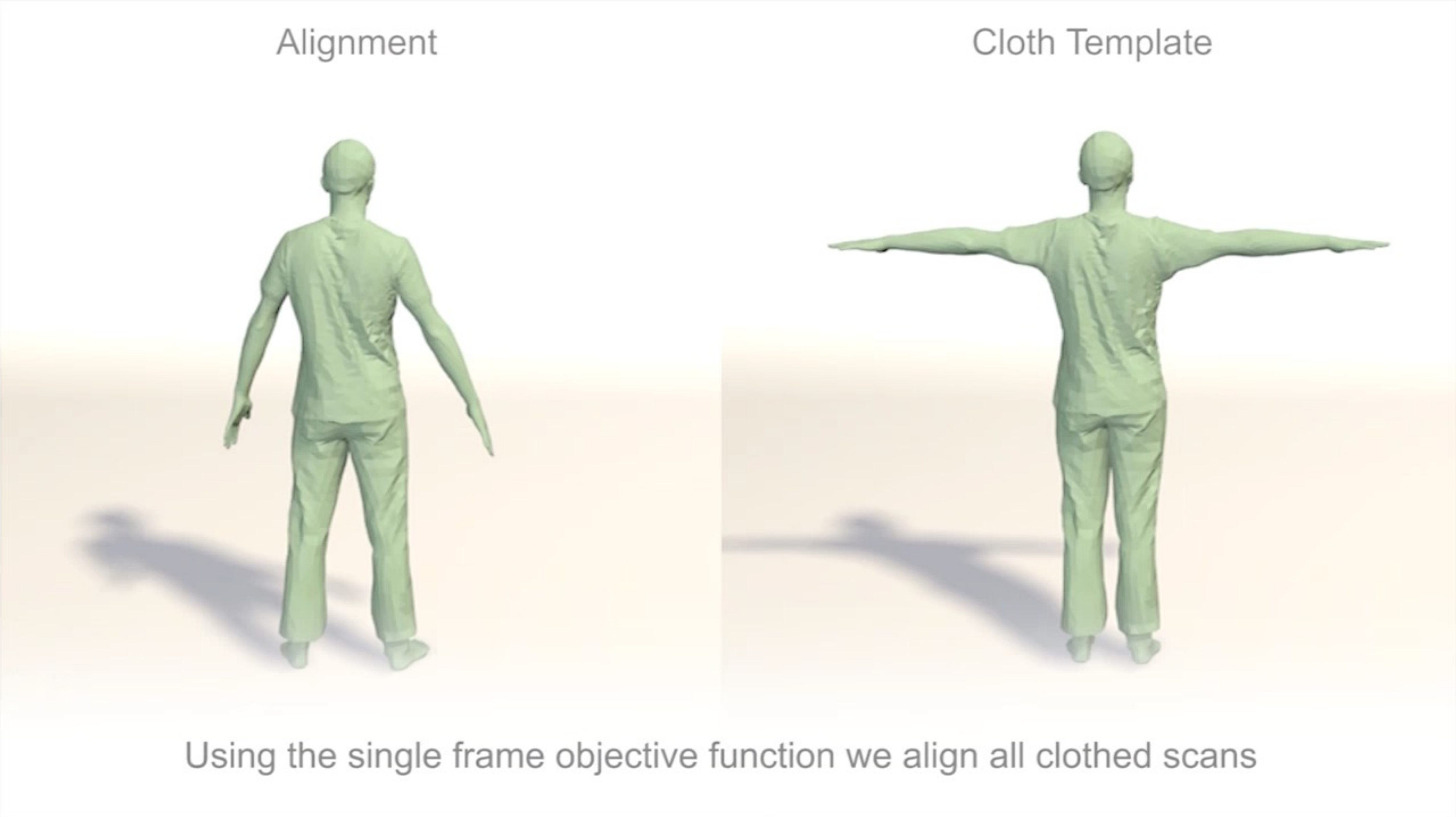
Pons-Moll et al. (ClothCap) SiggraphAsia'17

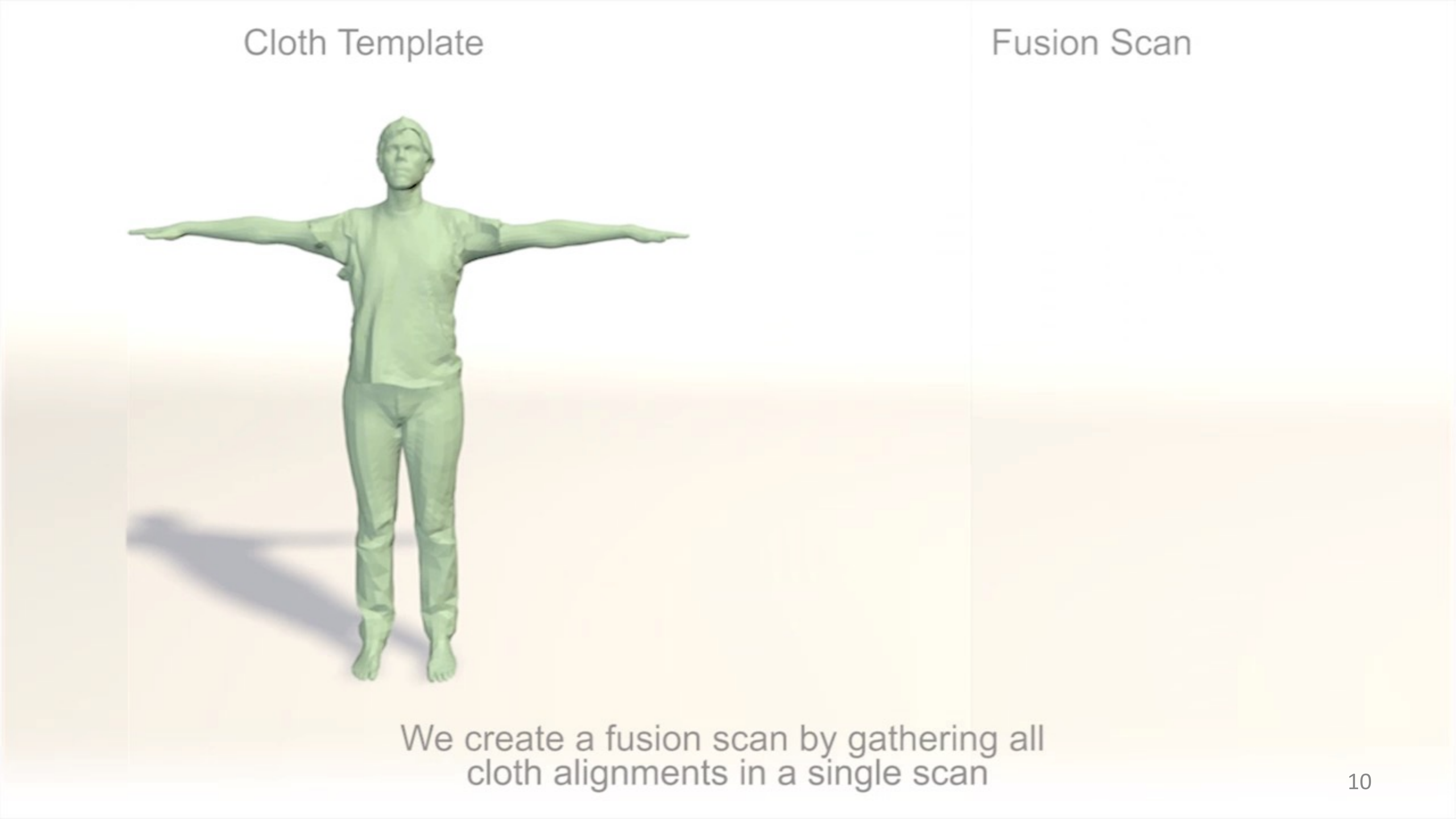
Scan


Scan
Alignment with
Single Template Mesh

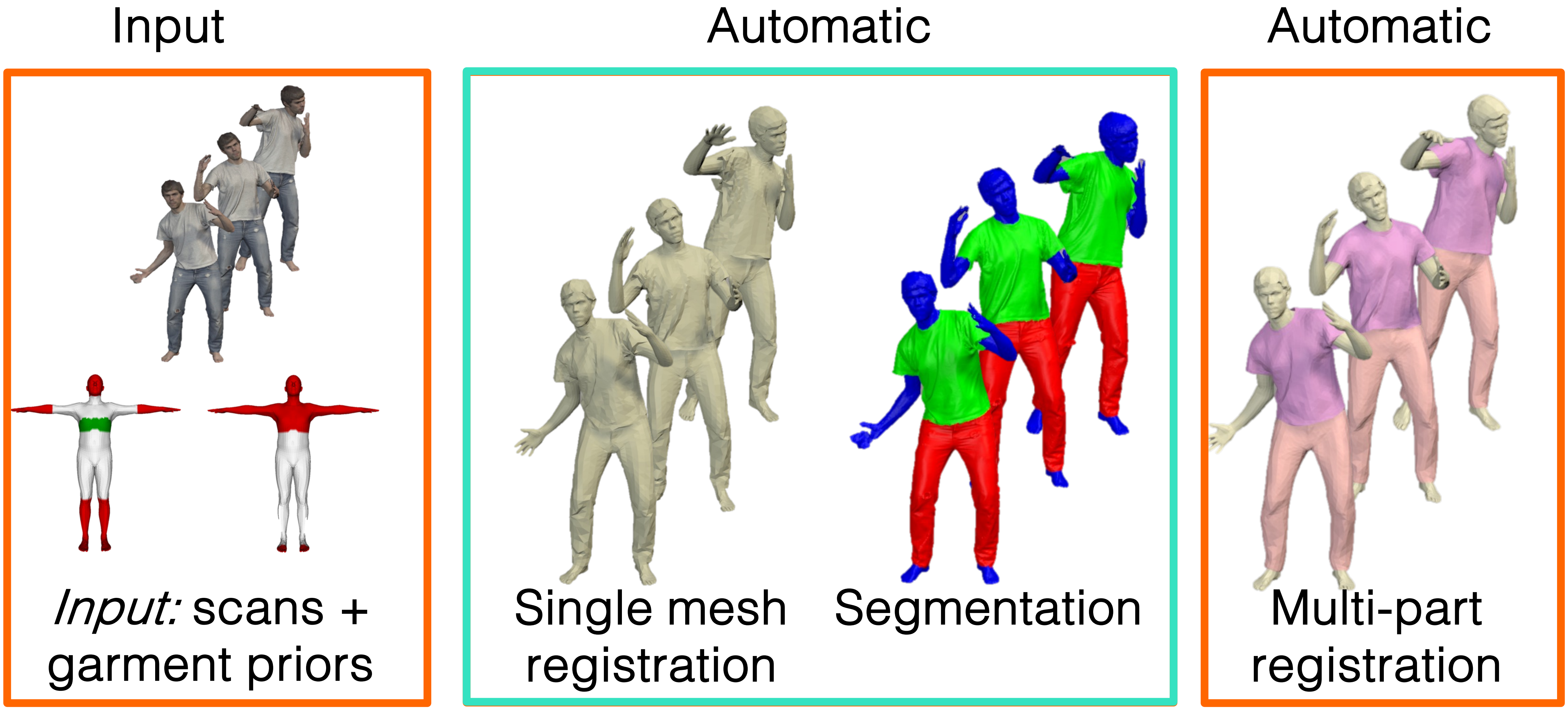

Alignment with
Single Template
Alignment with
Segmented Templates
Full ClothCap
Alignment

First: Shape under Clothing



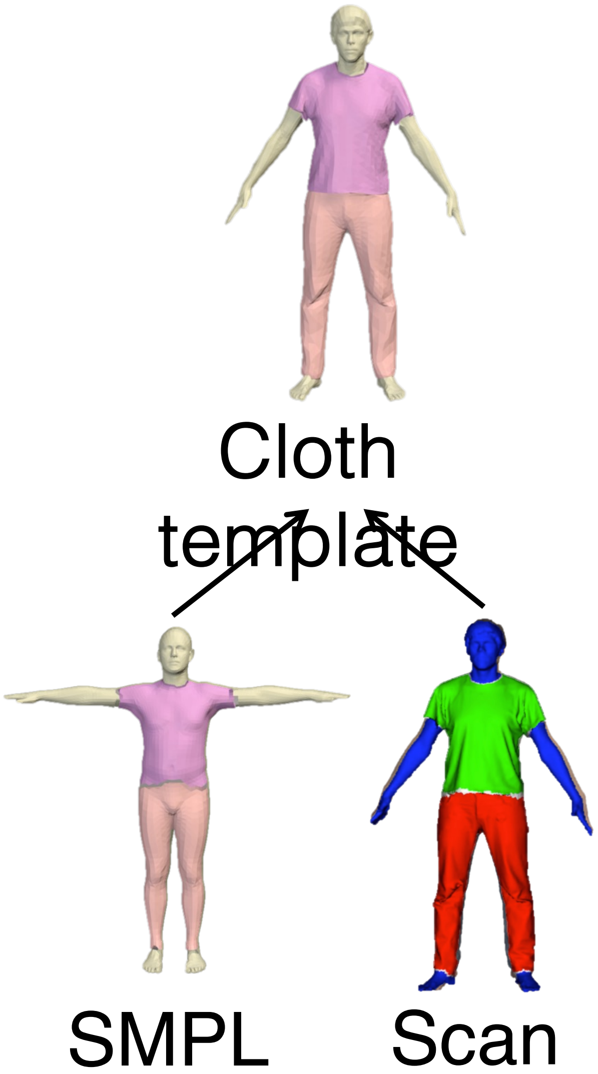

Cloth Template
Fusion Scan
We create a fusion scan by gathering all
cloth alignments in a single scan

ClothCap Overview



Pons-Moll et al. (ClothCap) SiggraphAsia’17

Multi-part Mesh Registration



Pons-Moll et al. (ClothCap) SiggraphAsia’17

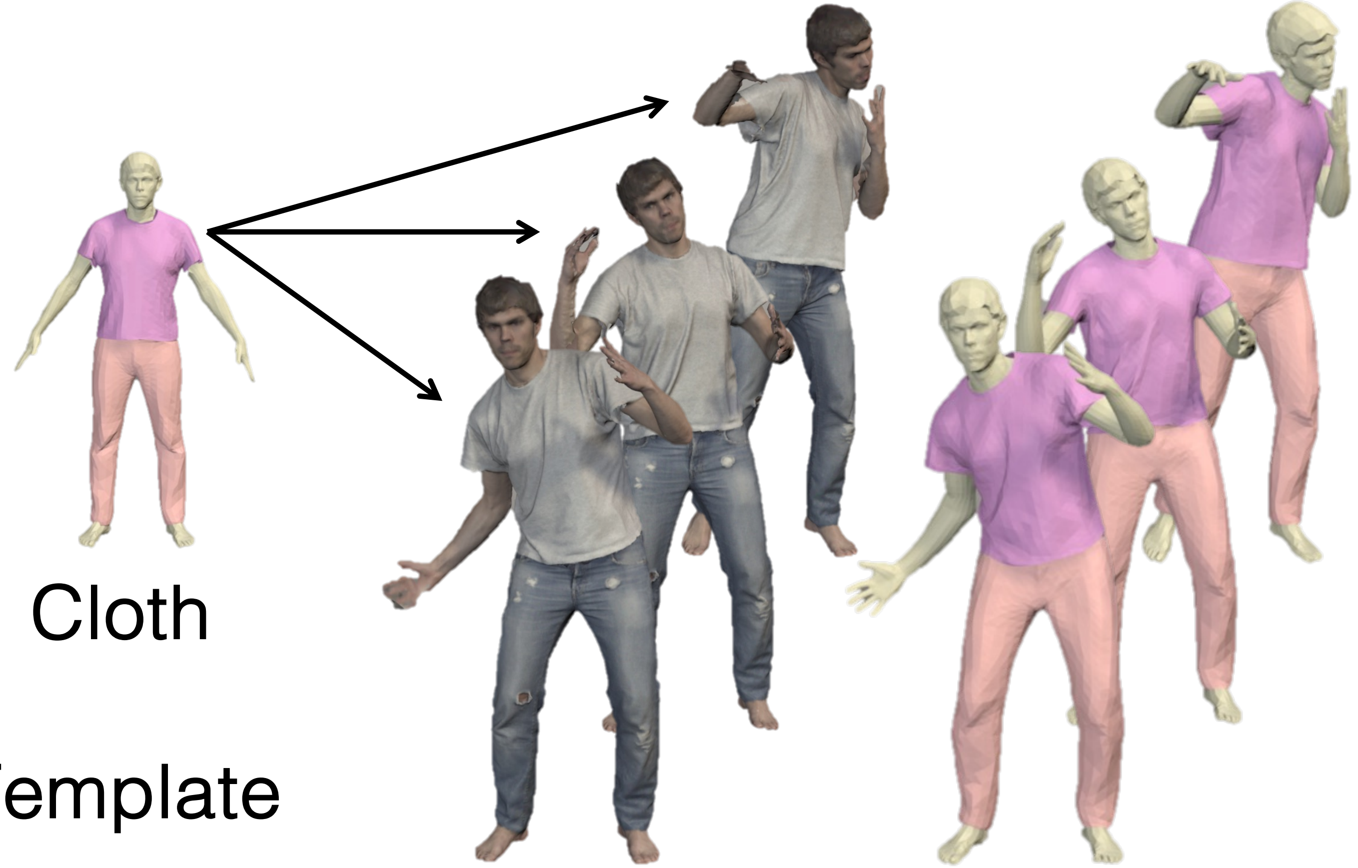
Multi-part Mesh Registration

Cloth

Template

Pons-Moll et al. (ClothCap) SiggraphAsia’17

Multi-part Mesh Registration

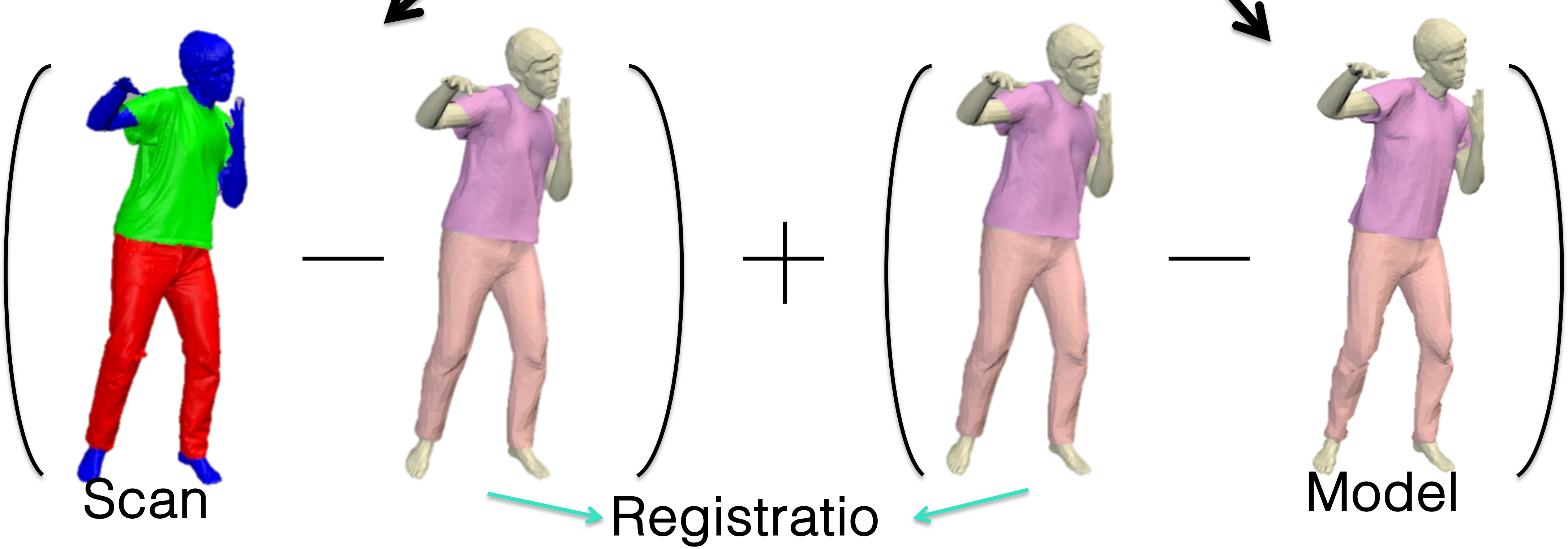


Cloth

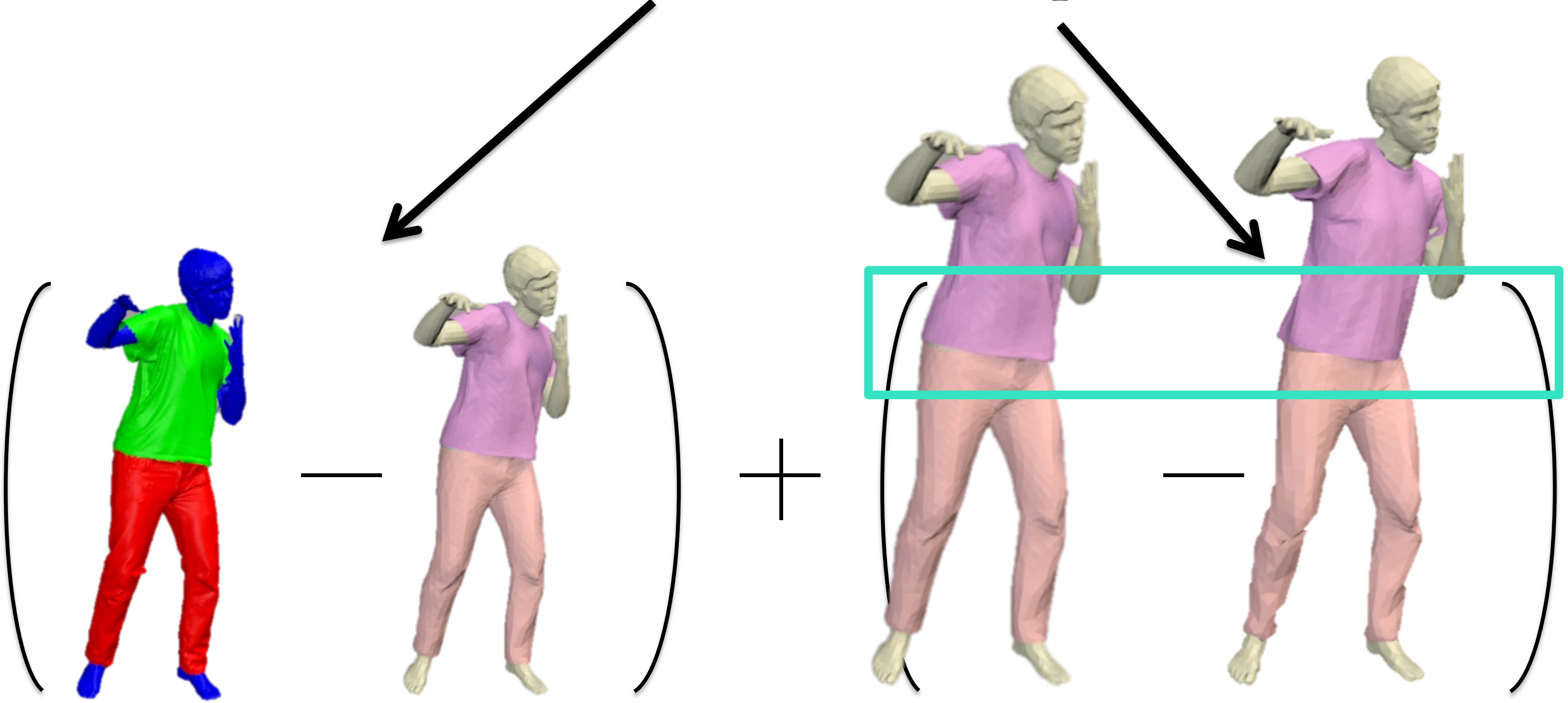
Template

Pons-Moll et al. (ClothCap) SiggraphAsia’17

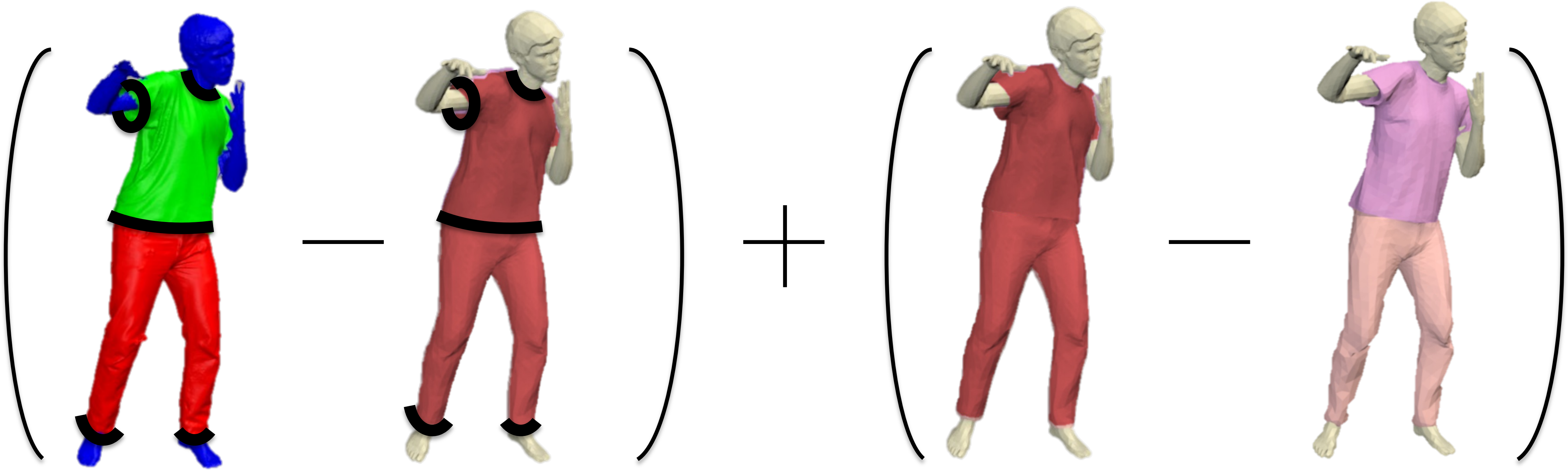
$$E(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}) = E_{\mathrm{data}}(\mathbf{v}) + E_{\mathrm{cpl}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}) +$$



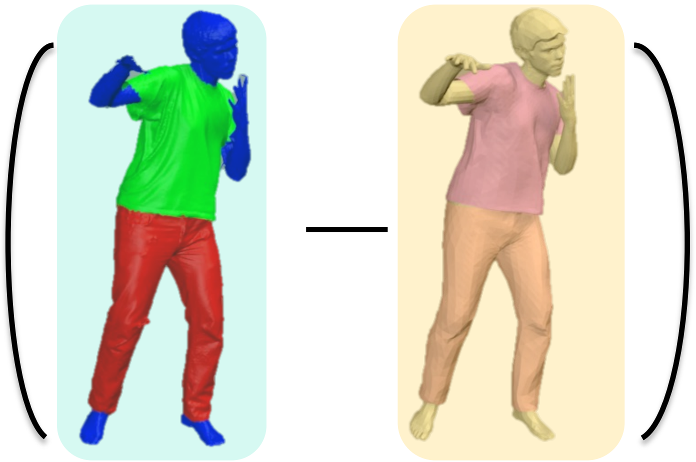
$$E(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}) = E_{\mathrm{data}}(\mathbf{v}) + E_{\mathrm{cpl}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}) +$$



$$E(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}) = E_{\text{data}}(\mathbf{v}) + E_{\text{cpl}}(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{v}) + \\ + \underline{E_{\text{boundary}}}(\mathbf{v}) + \underline{E_{\text{lap}}}(\mathbf{v})$$



Objective function terms: data term



$$E_{\mathrm{data}}(\mathbf{v}) = \sum_{g=1}^{N} E_{\mathrm{g}}(\mathbf{v}_g; S_g)$$

Per garment scan-to-mesh distance

Vertices of garment g

Segmented scan garment g

Pons-Moll et al. (ClothCap) SiggraphAsia'17

Objective function terms: data term

$$E_{\text{boundary}}(\mathbf{v}) = \sum_{g=1}^{N} E_{\text{g}}(\mathbf{v}_g; \mathcal{S}_g)$$

Per garment scan-to-curve distance

Vertices of ring r

Scan ring r

Pons-Moll et al. (ClothCap) SiggraphAsia'17

Objective function terms: Boundary Smoothness

Curve parameterized by arclength $\gamma(s) = (x(s), y(s), z(s))$

s

Curvature squared $k(s)^2 = x''(s)^2 + y''(s)^2 + z''(s)^2$

To make boundaries smooth, minimize curvature for each ring r:

$$E_{\text{smth}}(\mathbf{v}) = \underbrace{\sum_{r=1}^{R_l}}_{\text{Rings}} \sum_{n} \|\underbrace{\mathbf{v}_{r,n-1} - 2\mathbf{v}_{r,n} + \mathbf{v}_{r,n+1}}_{\text{Ordered vertices along the ring}}\|^2$$

Rings

Ordered vertices along the ring

Objective function: Laplacian Term

Given a mesh, the adjacency matrix Z is defined as:

$$\mathbf{Z}_{ij} = \begin{cases} 1, & \text{if } \mathbf{v}_i \text{ and } \mathbf{v}_j \text{ are connected} \\ 0, & \text{otherwise.} \end{cases}$$

Let $\mathbf{H}$ be a diagonal matrix where $\mathbf{H}_{ii}$ equals the number of neighbors of vertex i.

The Graph Laplacian is defined as

$$\mathbf{G}_{\text{lap}} = \mathbf{I} - \mathbf{H}^{-1}\mathbf{Z}$$

Pons-Moll et al. (ClothCap) SiggraphAsia'17

Objective function: Laplacian Term

To make a mesh smooth, we minimize a Laplacian term

$$E_{\text{lap}}(\mathbf{v}) = \sum_{g=1}^{N_{\text{garm}}} \|\mathbf{G}_{\text{lap}}^{g} \mathbf{v}_g\|_F^2$$

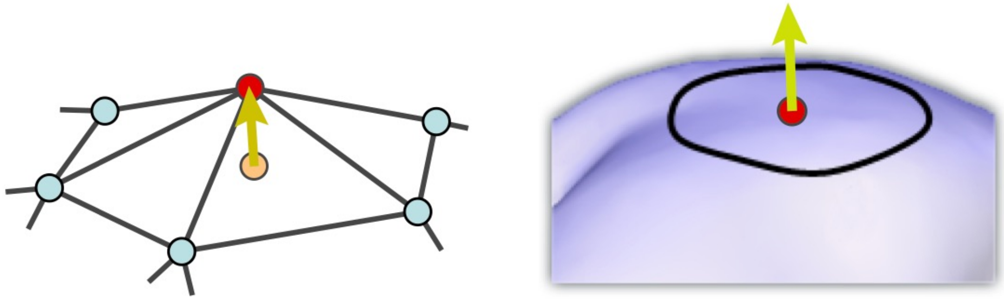
Graph Laplacian matrix for garment g

Vertices of garment g

Pons-Moll et al. (ClothCap) SiggraphAsia'17

Objective function: Laplacian Term

$$\mathbf{G}_{\text{lap}}^{g}\mathbf{v}_{g}$$



$\delta_i = \frac{1}{d_i}\sum_{j\in N(i)}(\mathbf{v}_i - \mathbf{v}_j)$ $\qquad$ $\frac{1}{|\gamma|}\int_{\mathbf{v}\in\gamma}(\mathbf{v}_i - \mathbf{v})dl(\mathbf{v})$

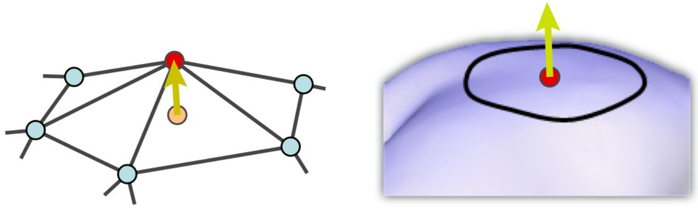
Figure 1: *The vector of the differential coordinates at a vertex approximates the local shape characteristics of the surface: the normal direction and the mean curvature.*

The Laplacian matrix times the matrix of vertices computes the difference from vertex v_i and the average of its neighbors v_j

See Sorkine et al. Laplacian Mesh Processing. EG'05

Objective function: Laplacian Term

What are we minimzing? $$E_{\text{lap}}(\mathbf{v}) = \sum_{g=1}^{N_{\text{garm}}} \|\mathbf{G}^{g}_{\text{lap}} \mathbf{v}_g\|_F^2$$



$\delta_i = \frac{1}{d_i}\sum_{j\in N(i)}(\mathbf{v}_i - \mathbf{v}_j)$ $\frac{1}{|\gamma|}\int_{\mathbf{v}\in\gamma}(\mathbf{v}_i - \mathbf{v})dl(\mathbf{v})$

Figure 1: *The vector of the differential coordinates at a vertex approximates the local shape characteristics of the surface: the normal direction and the mean curvature.*

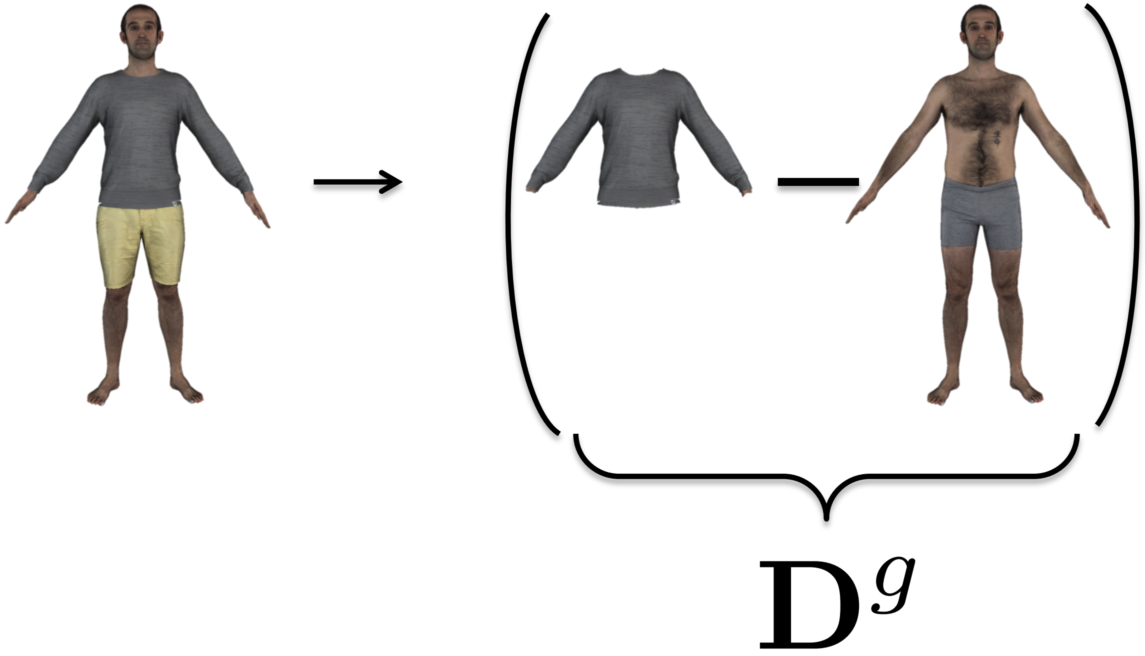
$$E(\mathbf{v}_i) = \left\| \mathbf{v}_i - \frac{1}{\mathbf{H}_{ii}} \sum_{j\in\mathcal{N}_j} \mathbf{v}_j \right\|^2$$

We minimize the norm of differential coordinates

Pons-Moll et al. (ClothCap) SiggraphAsia'17

SMPL + Garments

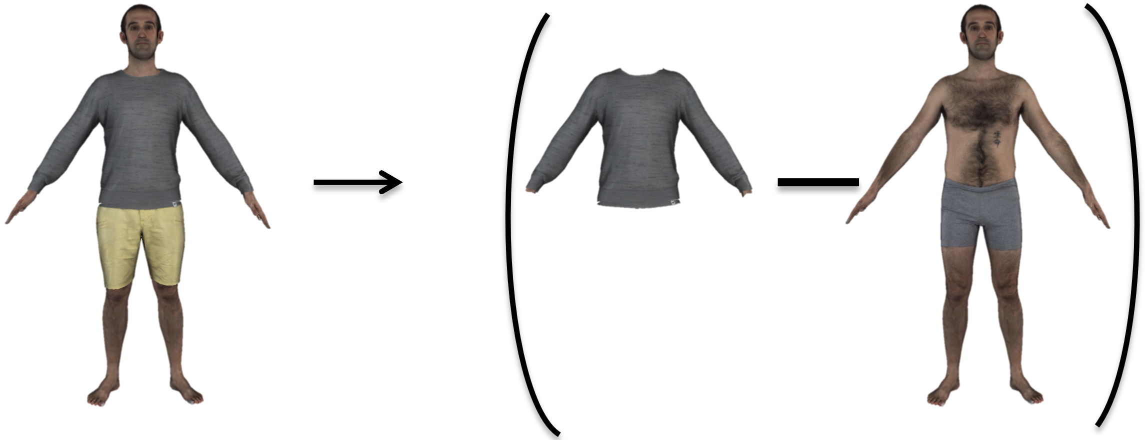
$$\mathbf{D}^g = \mathbf{G}^g - \mathbf{I}^g T(\boldsymbol{\beta}, \mathbf{0}_\theta, \mathbf{0}_\mathbf{D})$$



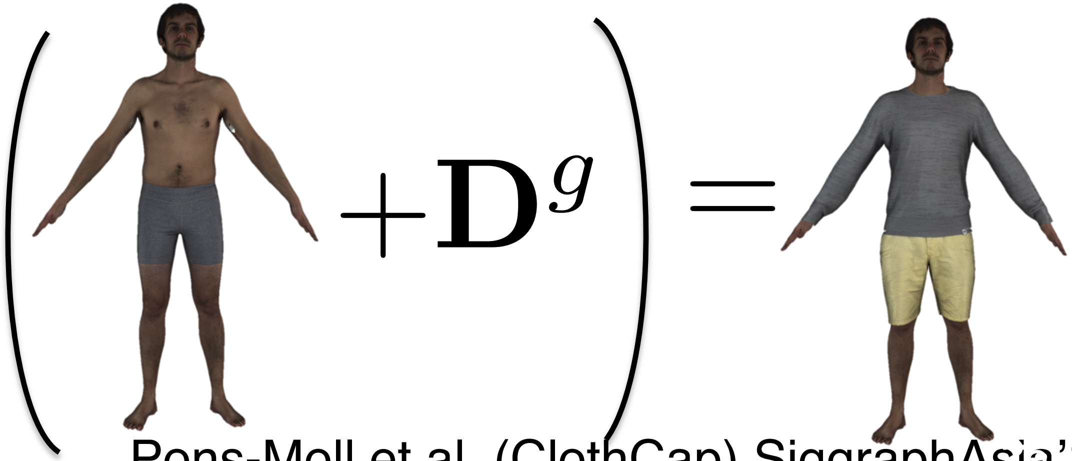
Pons-Moll et al. (ClothCap) SiggraphAsia'17

SMPL + Garments

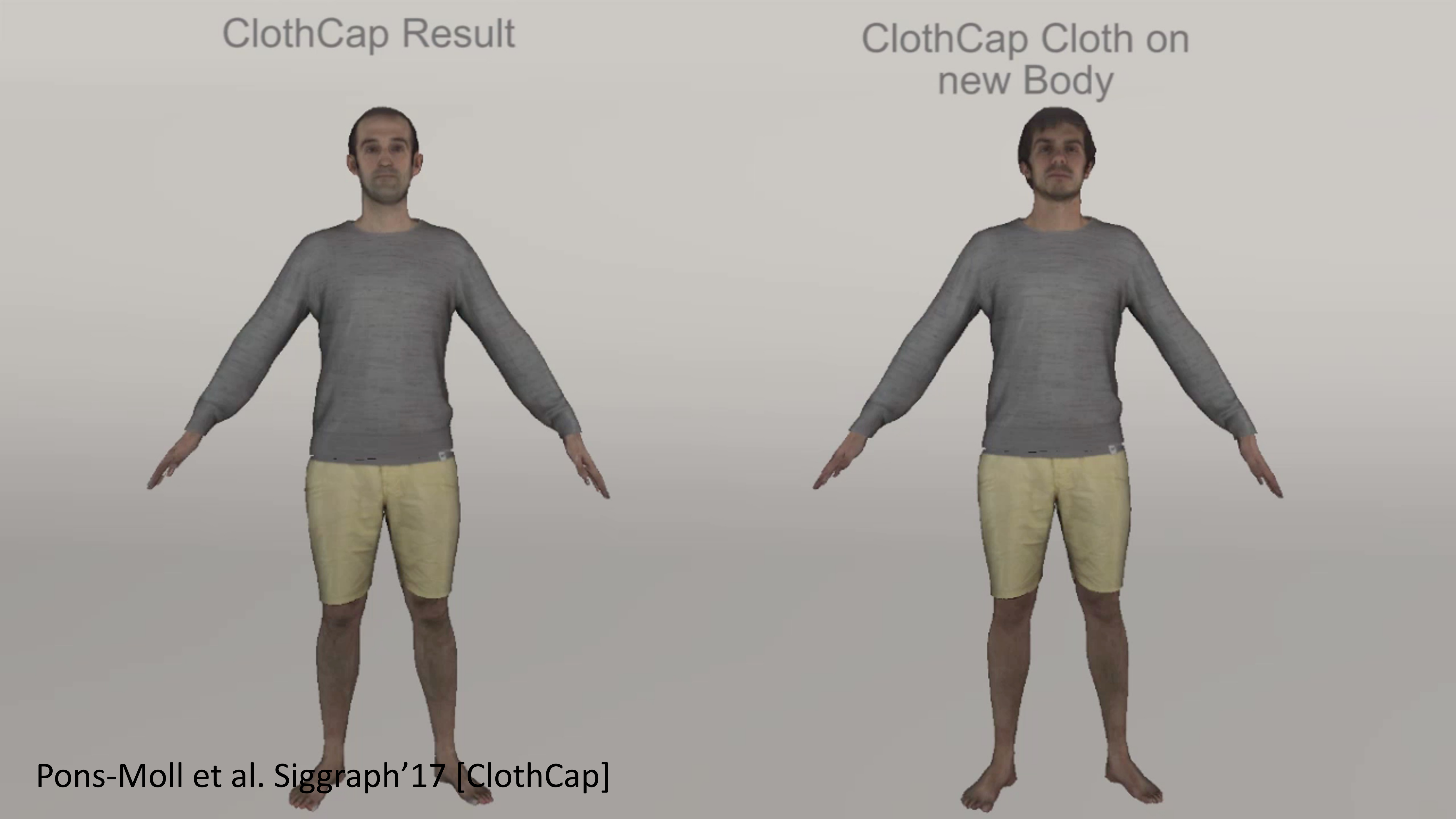
$$\mathbf{D}^g = \mathbf{G}^g - \mathbf{I}^g T(\boldsymbol{\beta}, \mathbf{0}_\theta, \mathbf{0}_\mathbf{D})$$



$$T^g(\boldsymbol{\beta}, \boldsymbol{\theta}, \mathbf{D}^g) = \mathbf{I}^g T(\boldsymbol{\beta}, \mathbf{0}_\theta, \mathbf{0}_\mathbf{D}) + \mathbf{D}^g$$



Pons-Moll et al. (ClothCap) SiggraphAsia'17



Pons-Moll et al. Siggraph'17 [ClothCap]


4D Scan
ClothCap Result
ClothCap Cloth on new Body

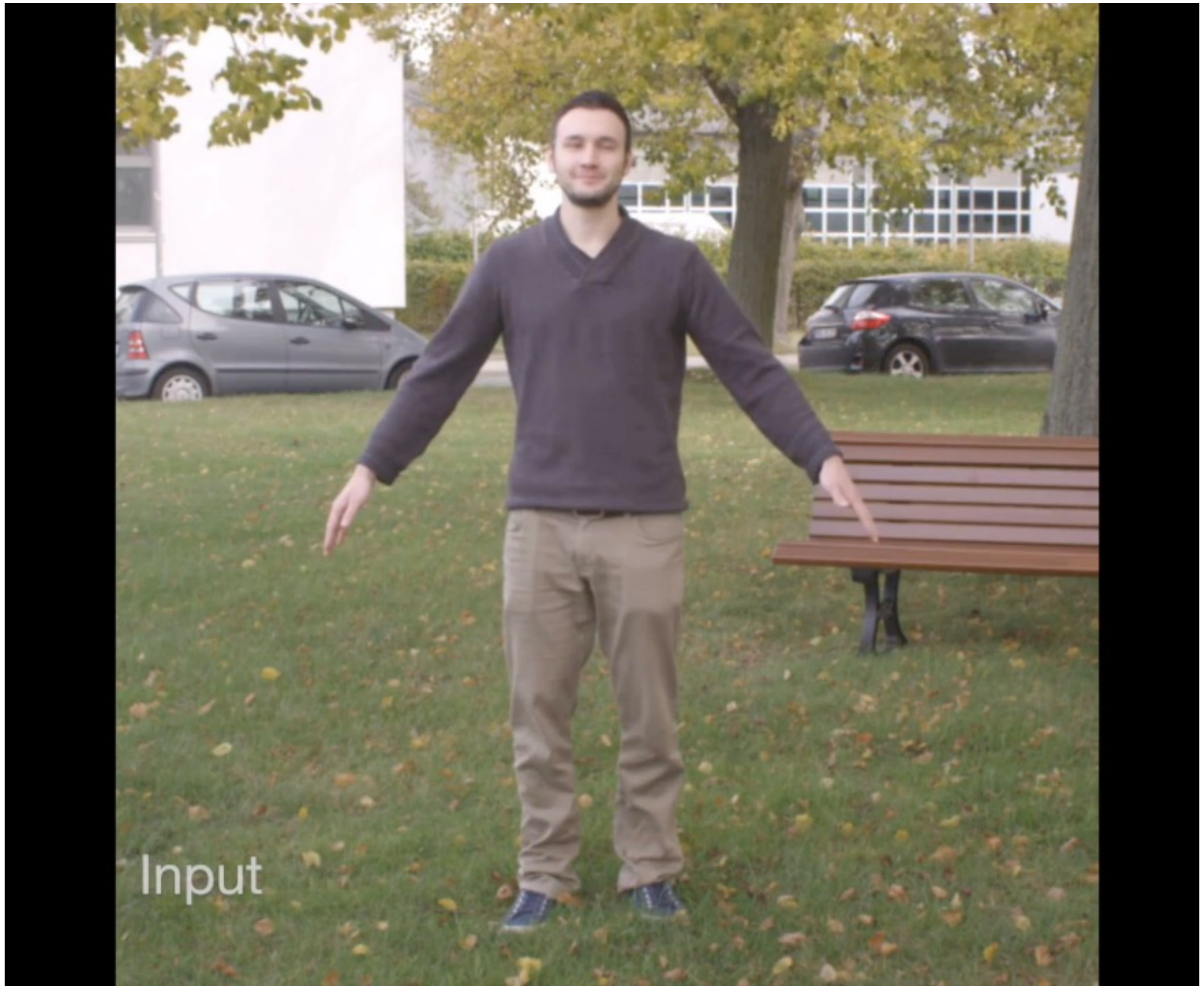

ClothCap Result
ClothCap Cloth on new Body


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

Results: Garments with Different Topology

People in Clothing from Images

Goal: 3D Reconstruction of People from a Single Video



Alldieck et al. CVPR'18

Optimization

Output codes

pose $\boldsymbol{\theta}$

shape $\boldsymbol{\beta}$

clothing $\mathbf{D}$

SMPL + D

$M(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{D})$

3D world

$R(\cdot)$

Video

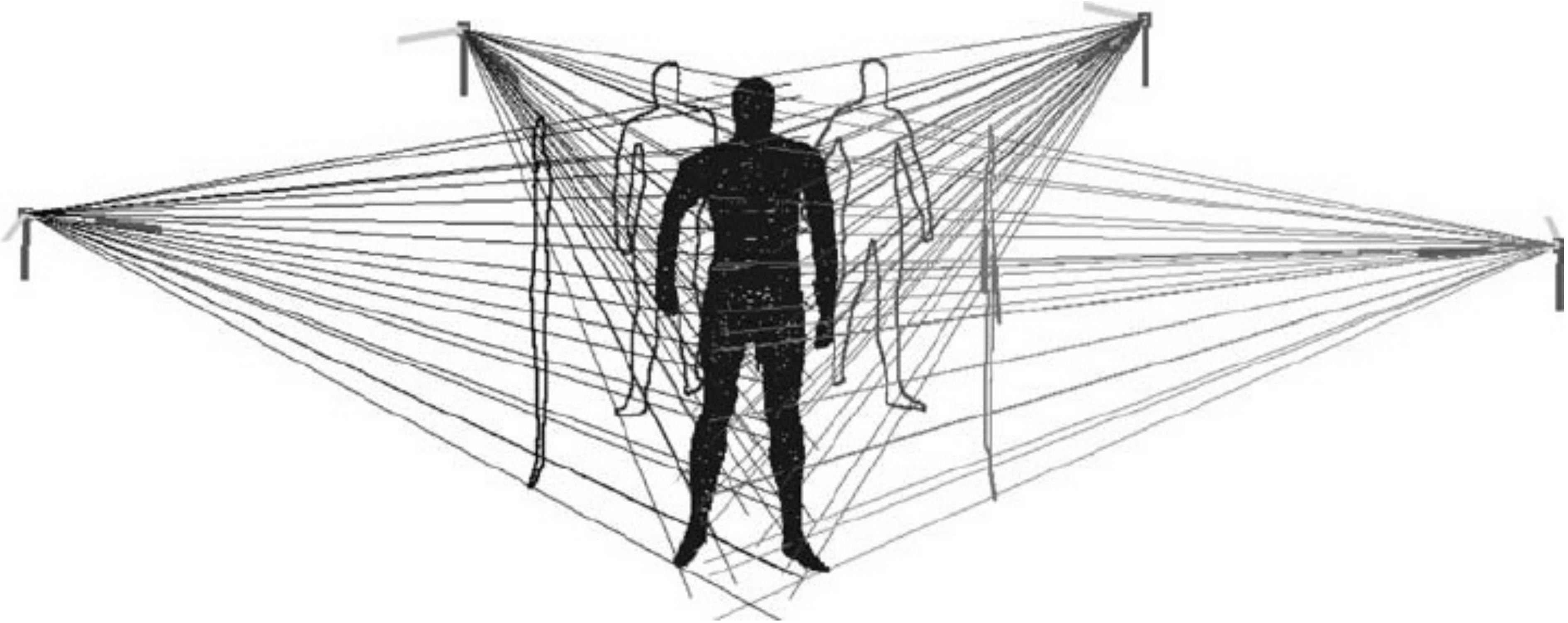
$$\arg\min_{\boxed{\theta,\beta,\mathbf{D}}} \sum_i \mathrm{dist}(R(M(\theta_i, \beta, \mathbf{D})), \mathbf{I}_i)$$

Optimize all poses at once is slow

Alldieck et al. CVPR'18

Key Idea: Extend Visual Hulls to Dynamic Human Motion

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



Alldieck et al. CVPR'18

How Can We Generalize It to Dynamic Human Motion ?

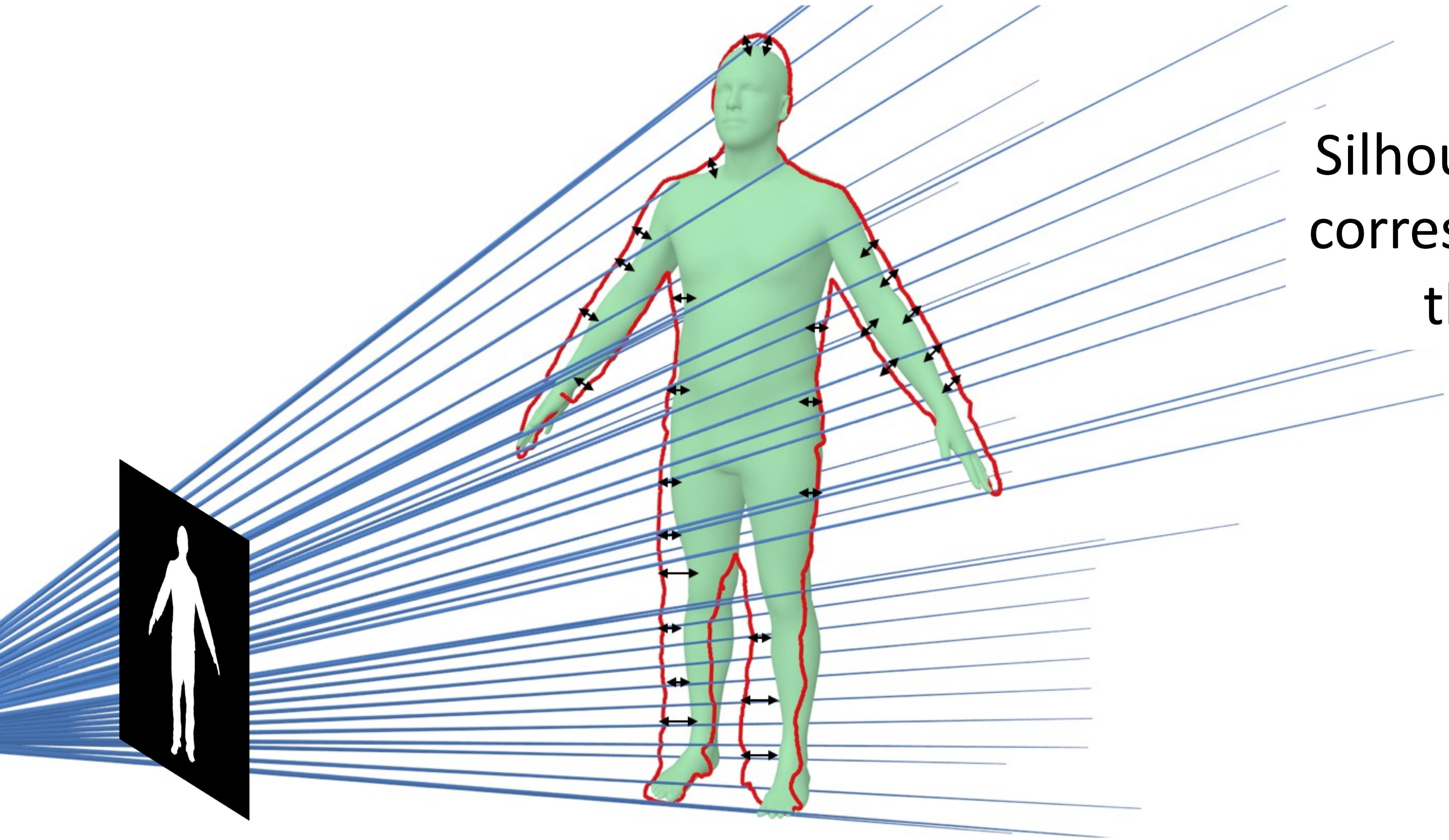
Person is moving!

Alldieck et al. CVPR'18

How Can We Generalize It to Dynamic Human Motion ?

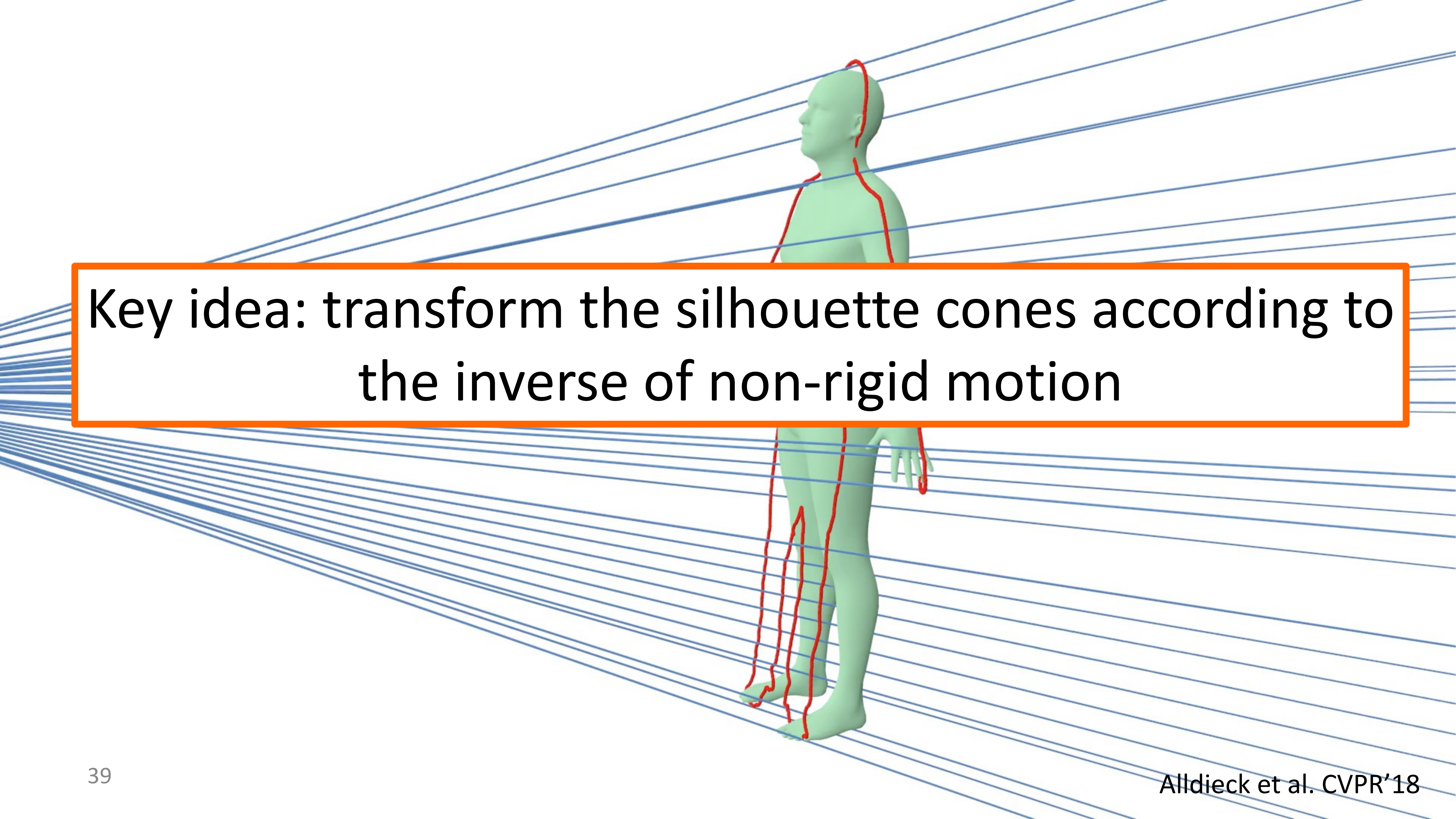
Estimate the 3D human pose and shape per frame

Alldieck et al. CVPR'18



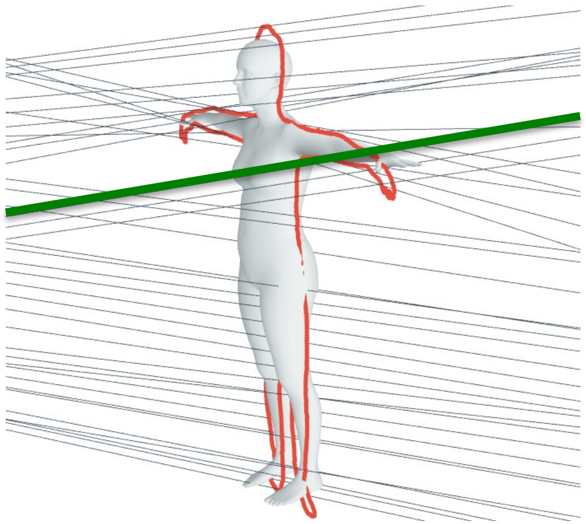
Silhouette rays with correspondences on the surface

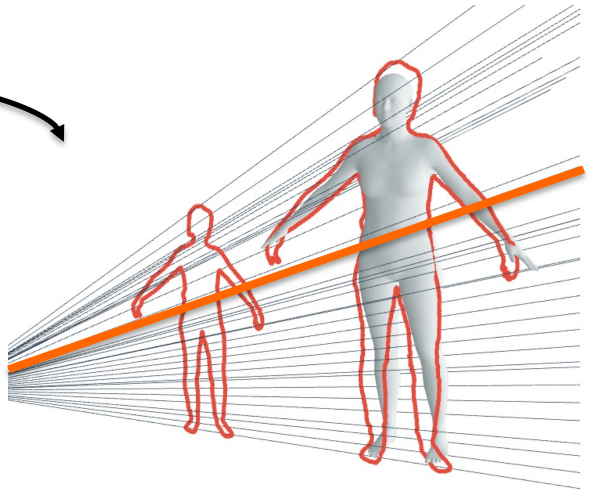
Alldieck et al. CVPR’18



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

Alldieck et al. CVPR’18





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

Alldieck et al. CVPR’18

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

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

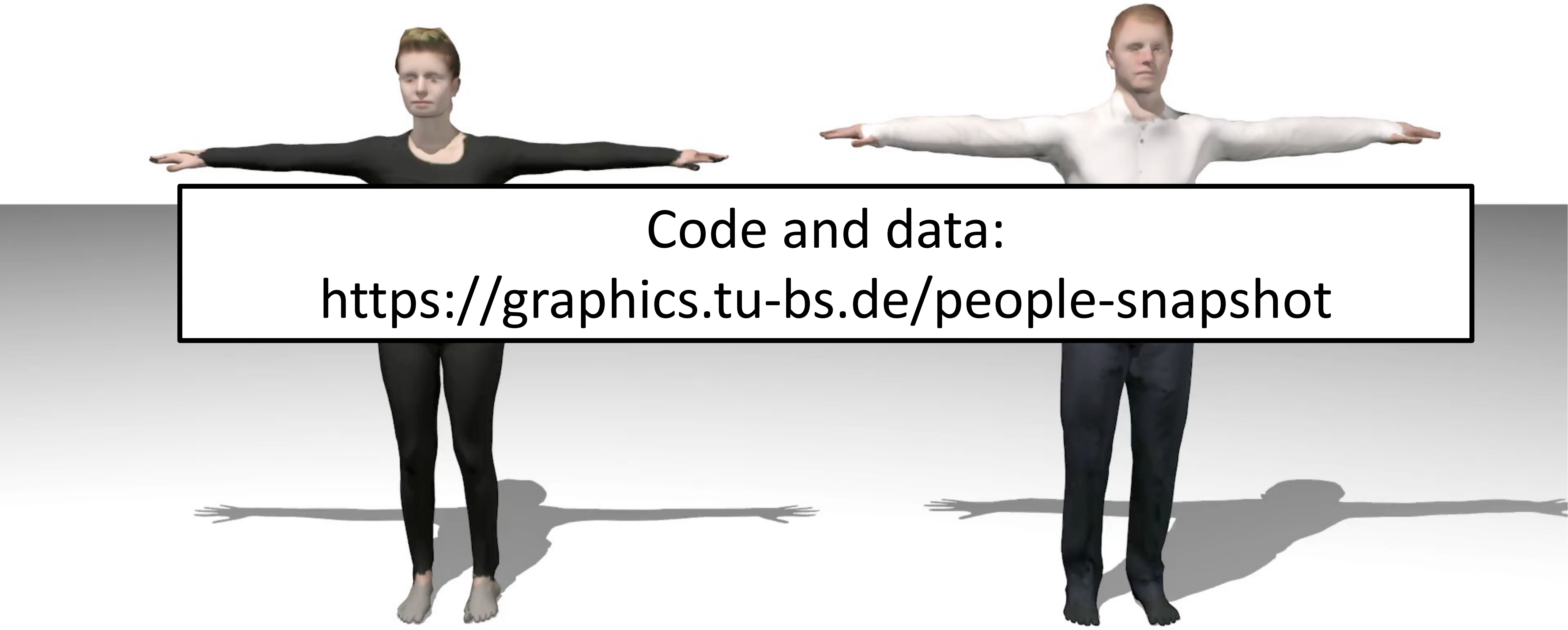
$$E_{\text{data}} = \sum_{(\mathbf{v},\mathbf{r})\in\mathcal{M}} \rho(\mathbf{v}\times\mathbf{r}_n - \mathbf{r}_m)$$

Sum of **point to line** distances

Prior Terms:
- Symmetry
- Prior on Shape
- Surface Smoothness

Alldieck et al. CVPR’18

Alldieck et al. CVPR’18

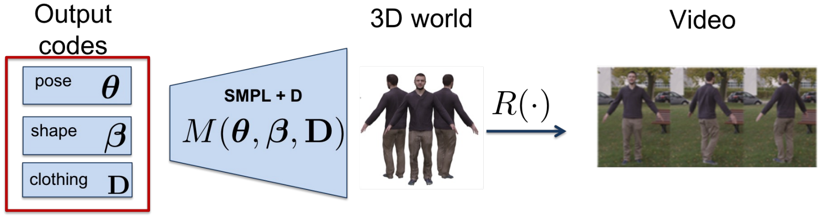


Code and data:
https://graphics.tu-bs.de/people-snapshot

Alldieck et al. CVPR’18

Alldieck et al. 3DV ‘18

Limitations



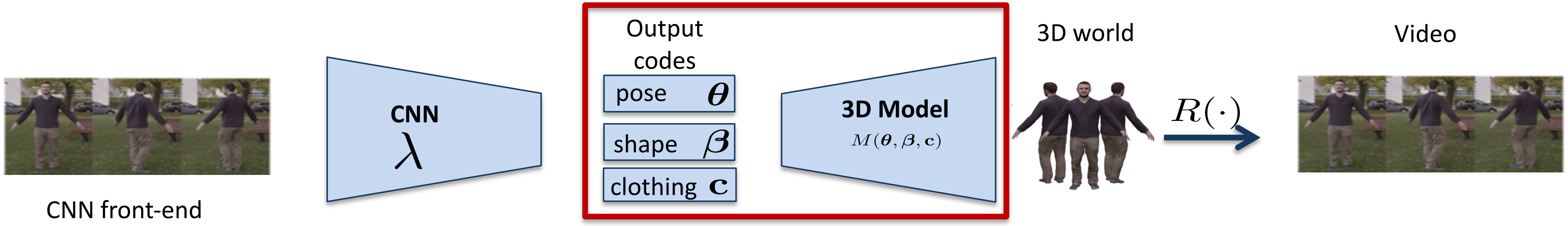
$$\arg\min_{\boxed{\theta,\beta,\mathbf{D}}} \sum_i \mathrm{dist}(R(M(\theta_i, \beta, \mathbf{c})), \mathbf{I}_i)$$

- Optimization: Local minima and slow

- Clothing as a single offset field is limiting:
  – Can not separate body from clothing

Alldieck et al. CVPR’18

Self-supervised Full Surface Reconstruction



$$\boldsymbol{\theta} \mapsto \boldsymbol{\theta}(\mathbf{I}, \lambda), \quad \boldsymbol{\beta} \mapsto \boldsymbol{\beta}(\mathbf{I}, \lambda), \quad \mathbf{c} \mapsto \mathbf{c}(\mathbf{I}, \lambda)$$

Thiemo Alldieck, Marcus Magnor, Bharat Lal Bhatnagar, Christian Theobalt, Gerard Pons-Moll
Learning to Reconstruct People in Clothing from a Single RGB Camera
in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
BibTeX PDF Video Code/Data

CVPR’19

Bharat Lal Bhatnagar, Garvita Tiwari, Christian Theobalt, Gerard Pons-Moll
Multi-Garment Net: Learning to Dress 3D People from Images
in *IEEE International Conference on Computer Vision (ICCV)*, 2019.

ICCV’19

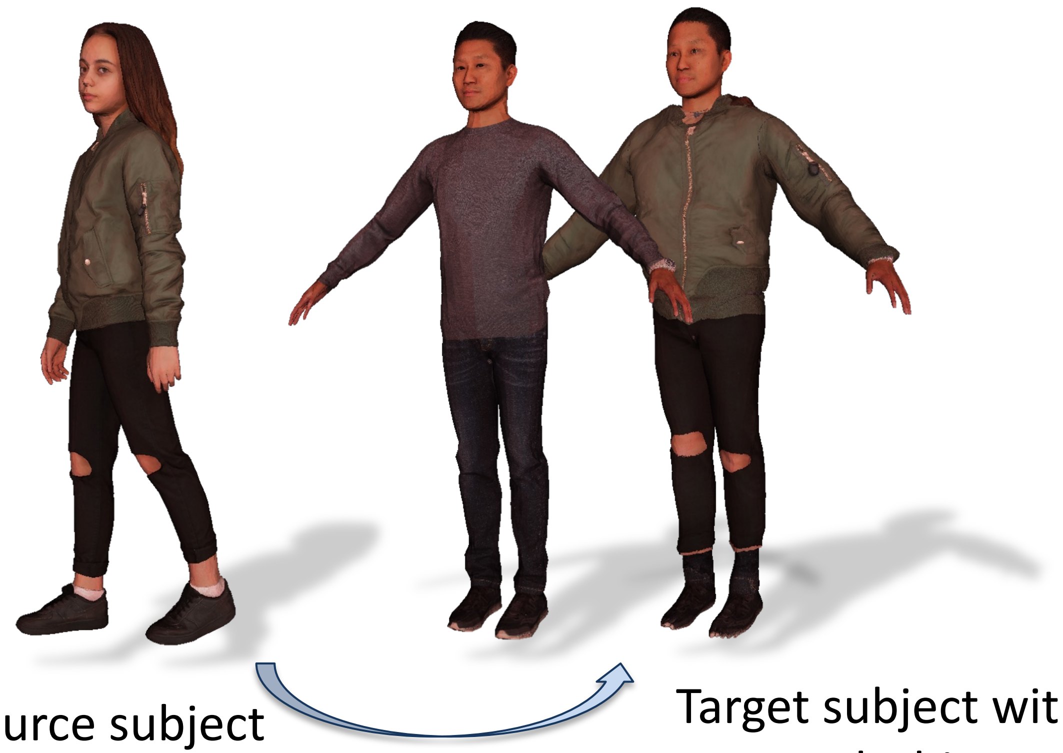
Marc Habermann, Weipeng Xu, Michael and Zollhoefer, Gerard Pons-Moll, Christian Theobalt
DeepCap: Monocular Human Performance Capture Using Weak Supervision
in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

CVPR’20
Best Student Paper
Honorable Mention

Multi-Garment Net: Learning to Dress People from Images

Source subject

Target subject with source clothing



Source subject

Target subject with source clothing

Bhatnagar et al. ICCV'19

SMPL + Clothing

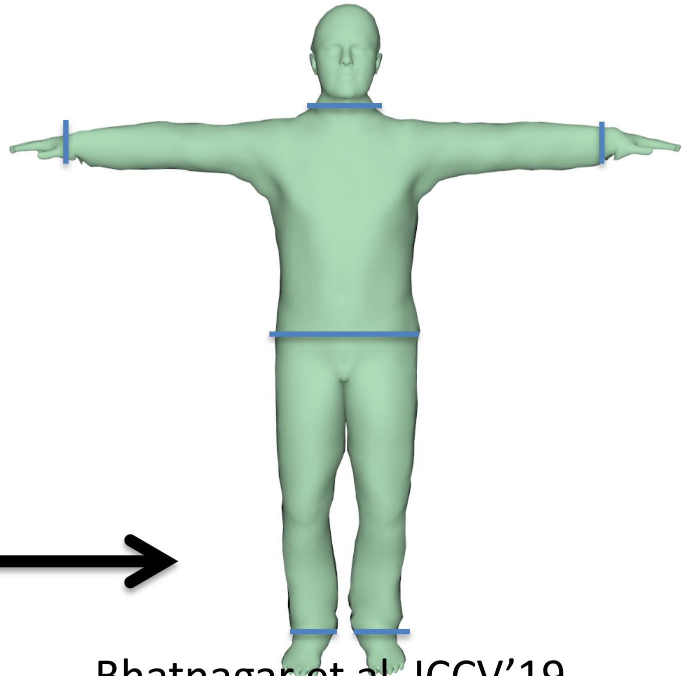
Vertices in a 0-pose

$$T(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{D}) = \mathbf{T}_\mu + B_s(\boldsymbol{\beta}) + B_p(\boldsymbol{\theta}) + \mathbf{D}$$

$\boldsymbol{\theta}$ Pose parameters

$\boldsymbol{\beta}$ Shape parameters

$\mathbf{D}$ Personal details + clothing



Bhatnagar et al. ICCV'19

Registration

1) Segment the scans into garments
2) Estimate body shape under clothing
3) Non rigidly register each garment template to each scan → joint optimization

Bhatnagar et al. ICCV'19

Digital Wardrobe

–Bhatnagar et al. ICCV'19

Dressing Shapes from Images

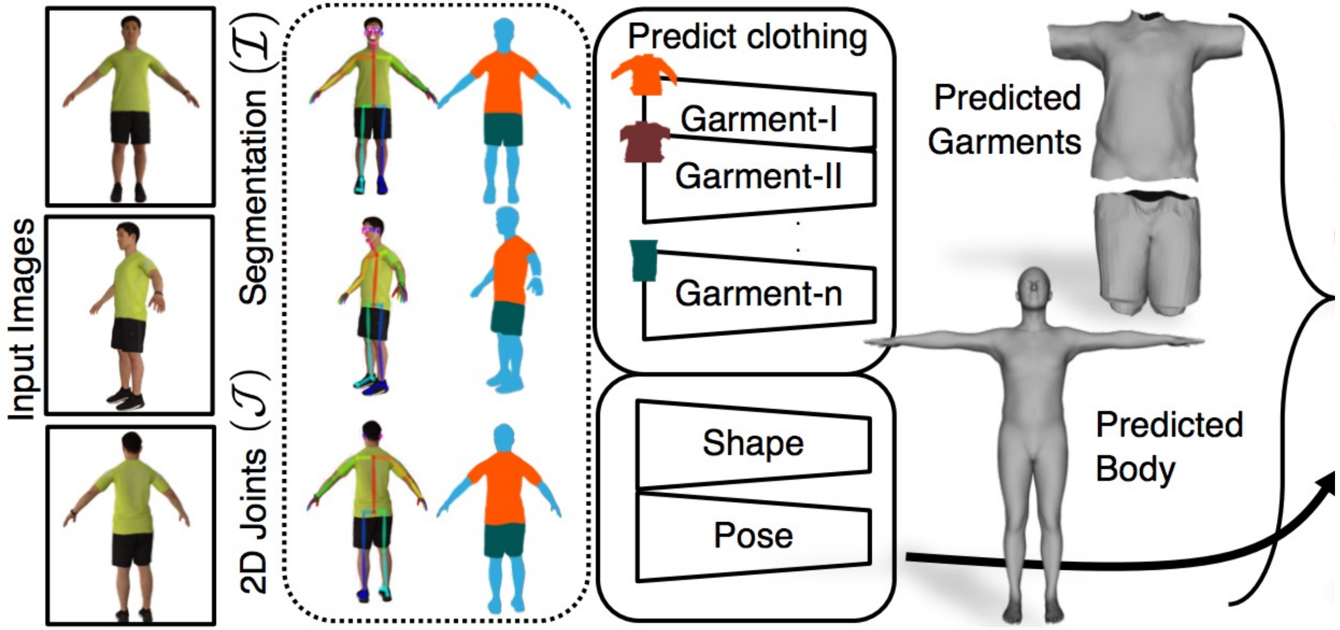
Source:8 images of a person turning

Target: scan

Result: 3Dmesh

Bhatnagar et al. ICCV'19

Multi-Garment Net: MGN



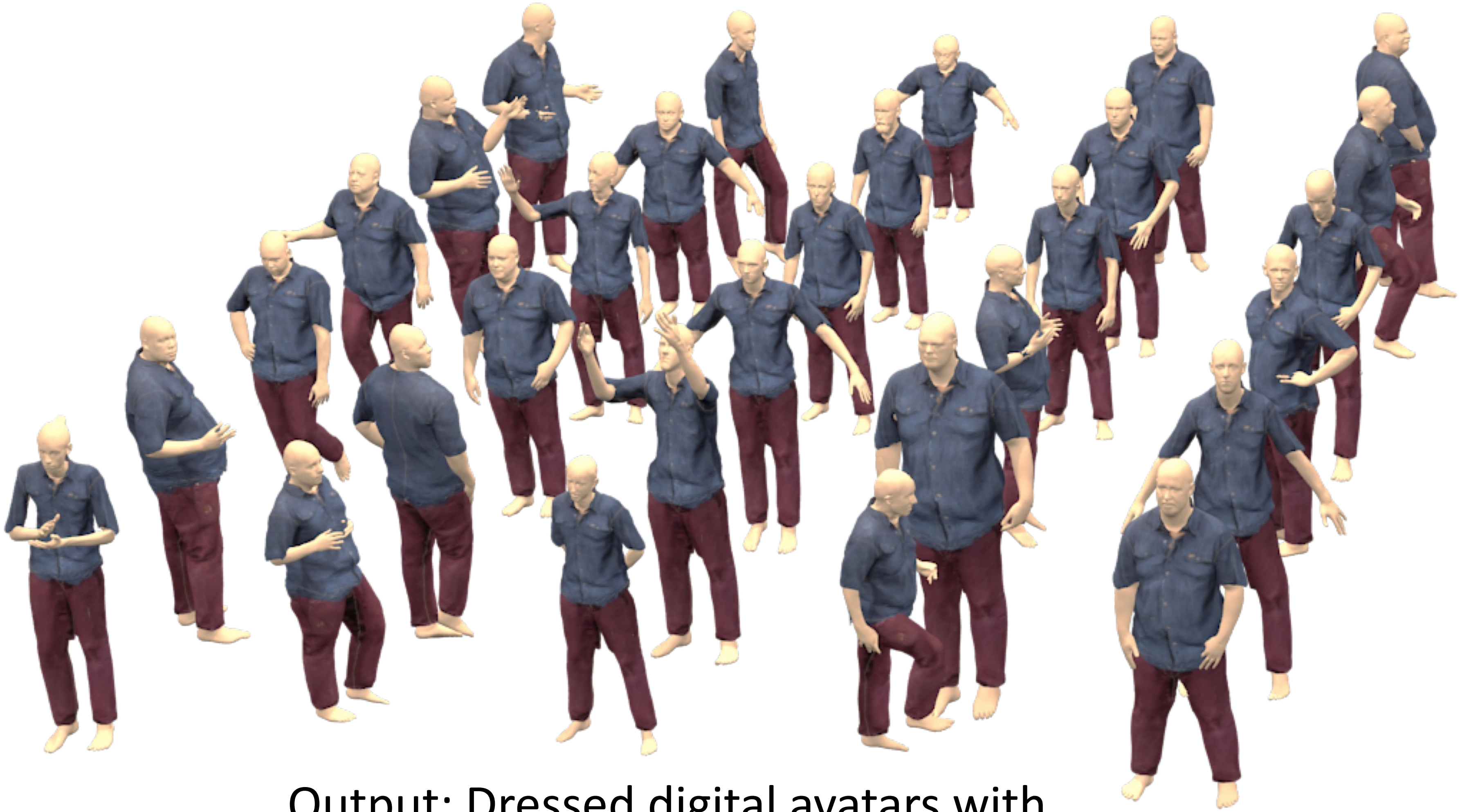
$$\mathbf{G}^g = \mathbf{B}^g \boldsymbol{z}^g + \mathbf{D}^{\mathrm{hf},g}$$

Codes for clothing, pose and shape

Bhatnagar et al. ICCV'19

Dressing in different shapes and poses

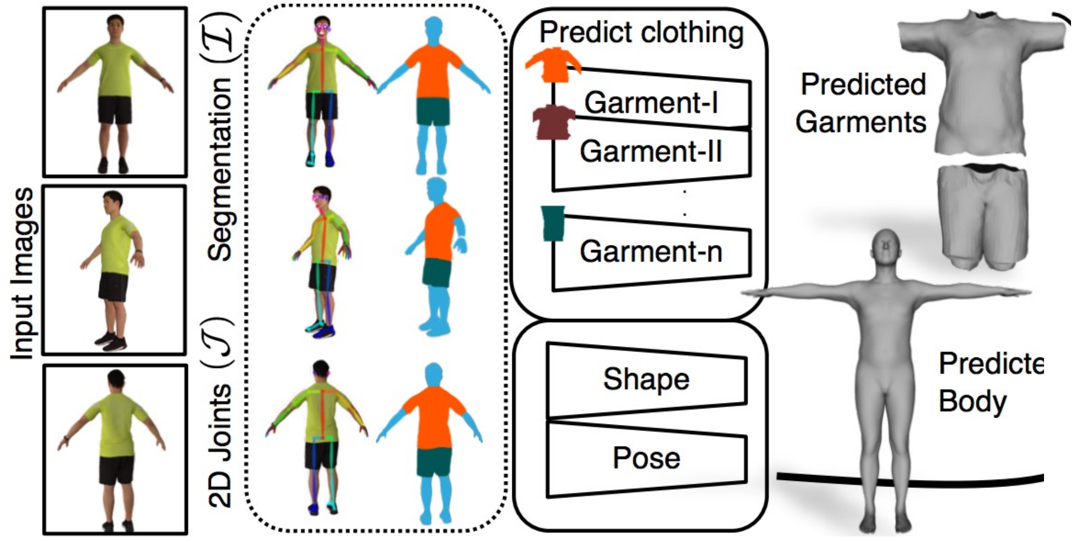
Input: 8 images



Output: Dressed digital avatars with input clothing

Bhatnagar et al. ICCV'19

Remaining Problem: Details

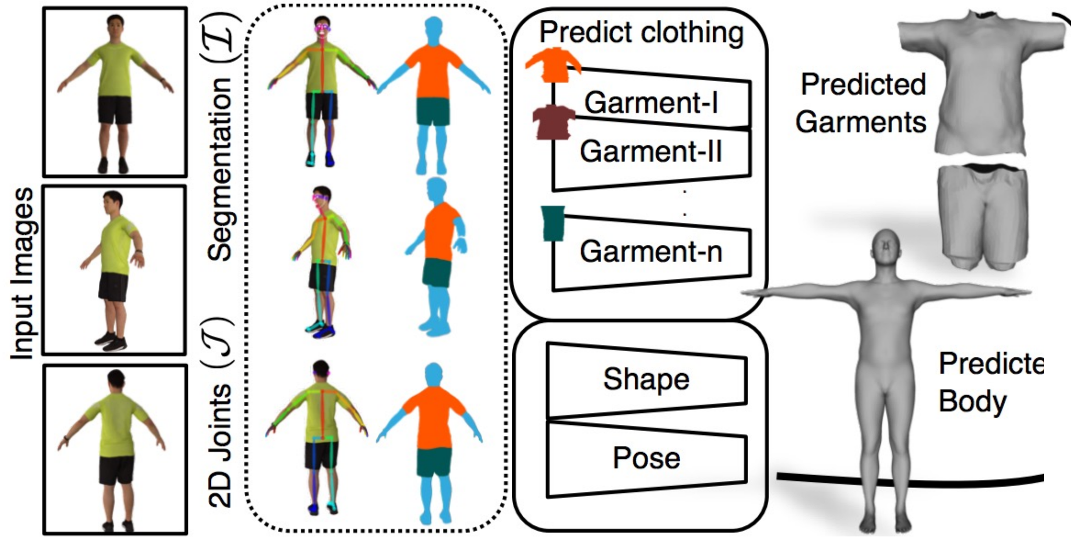


$$f : \mathcal{I} \mapsto \mathbf{D} \in \mathbb{R}^{3N}$$

Predicting Displacements directly as a function of the image is hard

Bhatnagar et al. ICCV'19

Remaining Problem: Details

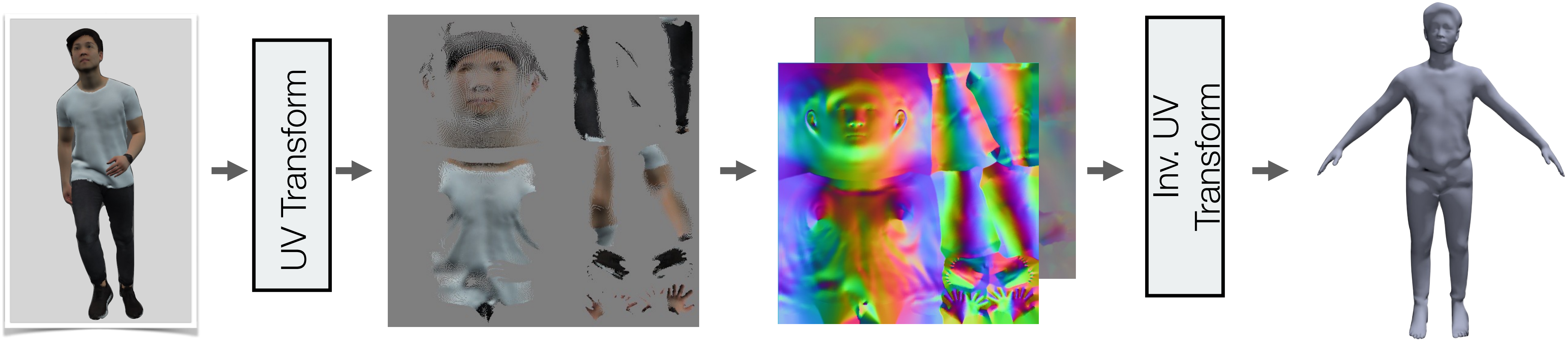


$$f : \mathcal{I} \mapsto \mathbf{D} \in \mathbb{R}^{3N}$$

Predicting Displacements directly as a function of the image is hard

Bhatnagar et al. ICCV'19

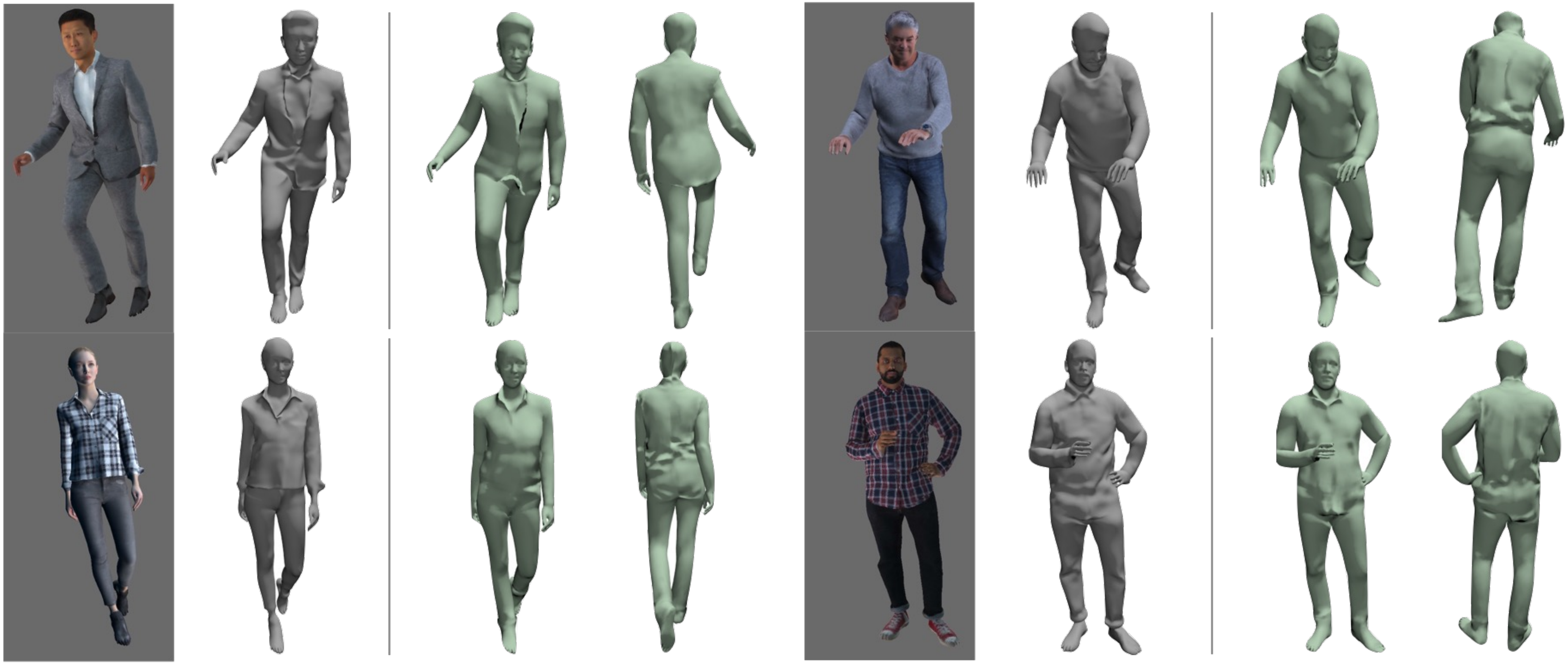
Tex2Shape: Detailed Full Human Body Geometry from a Single Image Exploiting UV-maps



Input UV Unwrap Normals and 3D displacements Output

Alldieck et al. ICCV’19
Lazova et al. 3DV’19 (for texture completion)
Mir et al. CVPR’20 (transfer texture from shopping websites)

Results



Alldieck et al. ICCV’19

Results

Alldieck et al. ICCV'19

Learning to Transfer Texture from Clothing Images to 3D Humans

Aymen Mir, Thiemo Alldieck, Gerard Pons-Moll

Mir et al. CVPR’20

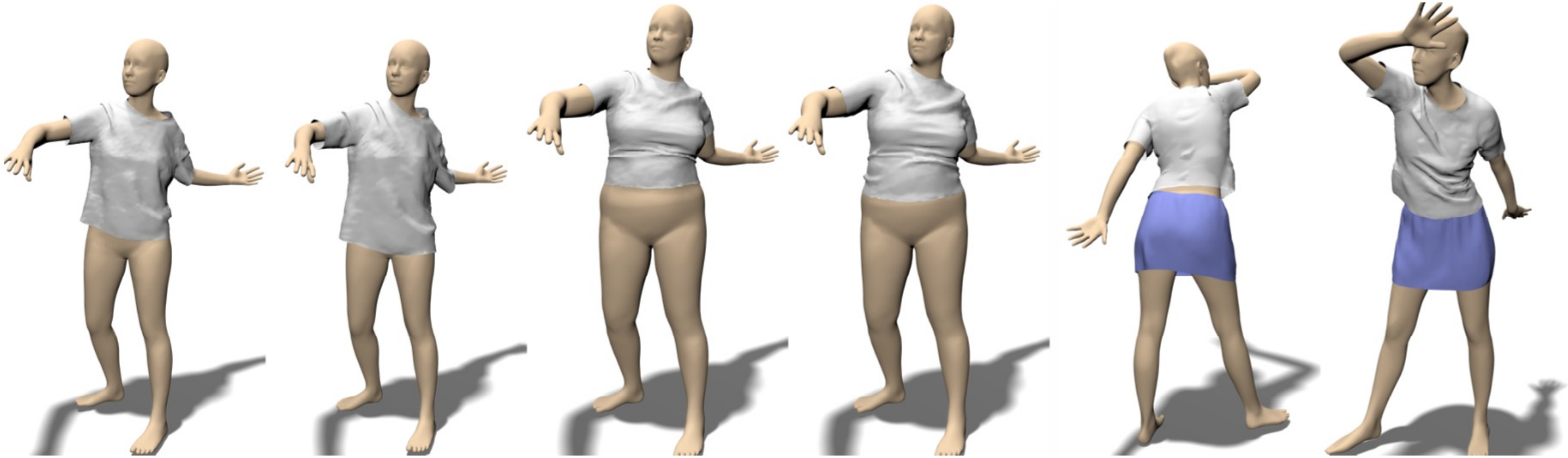
Take home messages

- Displacements are the simplest representation for clothing
- Video Avatars demonstrated 3D **reconstruction** of people in clothing is possible from a **single video**
- Exploit temporal information: shape barely changes over time
- Encoding body separately from clothing allows more **control**
- **Codes** carry **meaning** and allow **control**
- Pixel-aligned predictions in **UV-space** yield detailed reconstruction

Learning a Model of Clothing

TailorNet: Predicting 3D Clothing as a Function of Human Pose, Shape and Garment Style

Chaitanya Patel, Zhou Liao and Gerard Pons-Moll



Patel et al. CVPR’20 [Oral]

SMPL + Clothing

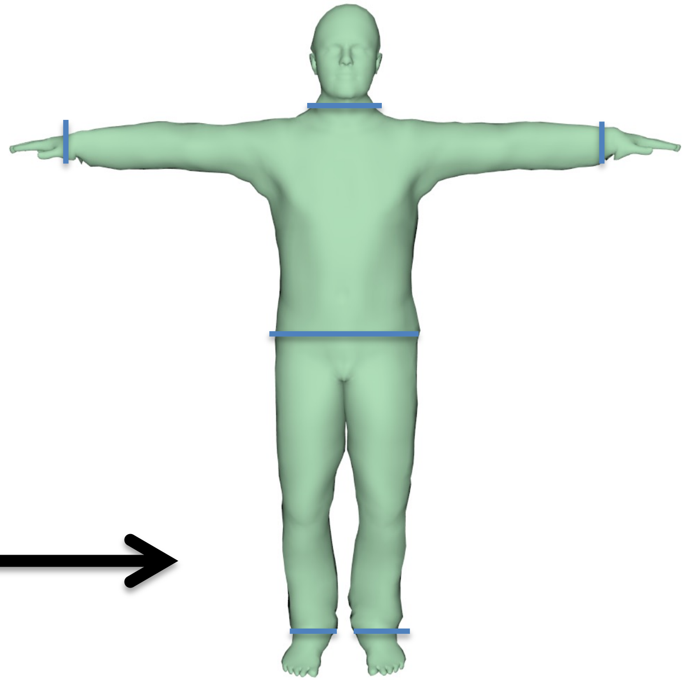
Vertices in a 0-pose

$$T(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{D}) = \mathbf{T}_\mu + B_s(\boldsymbol{\beta}) + B_p(\boldsymbol{\theta}) + \mathbf{D}$$

$\boldsymbol{\theta}$ Pose parameters

$\boldsymbol{\beta}$ Shape parameters

$\mathbf{D}$ Personal details + clothing

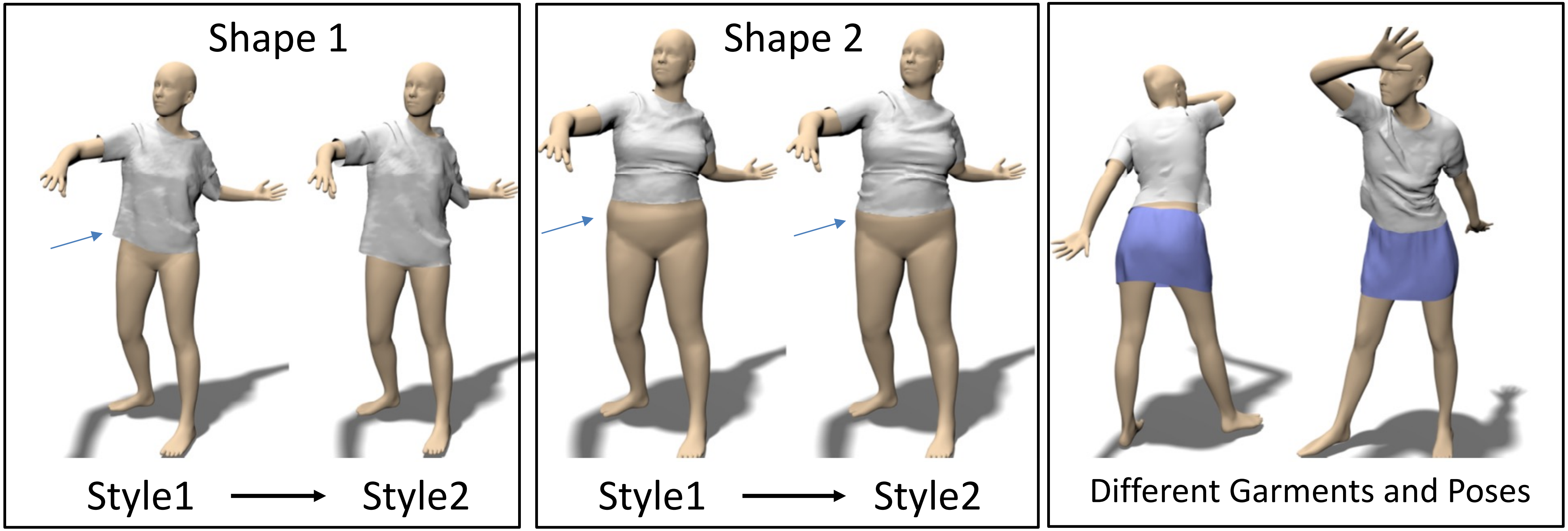


Goal: Clothing as a function of Pose, Shape and *Style*

$$D(\theta, \beta, \gamma) : \mathbb{R}^{|\theta|} \times \mathbb{R}^{|\beta|} \times \mathbb{R}^{|\gamma|} \mapsto \mathbb{R}^{m \times 3}$$

Pose Shape Style

Vertices of garment



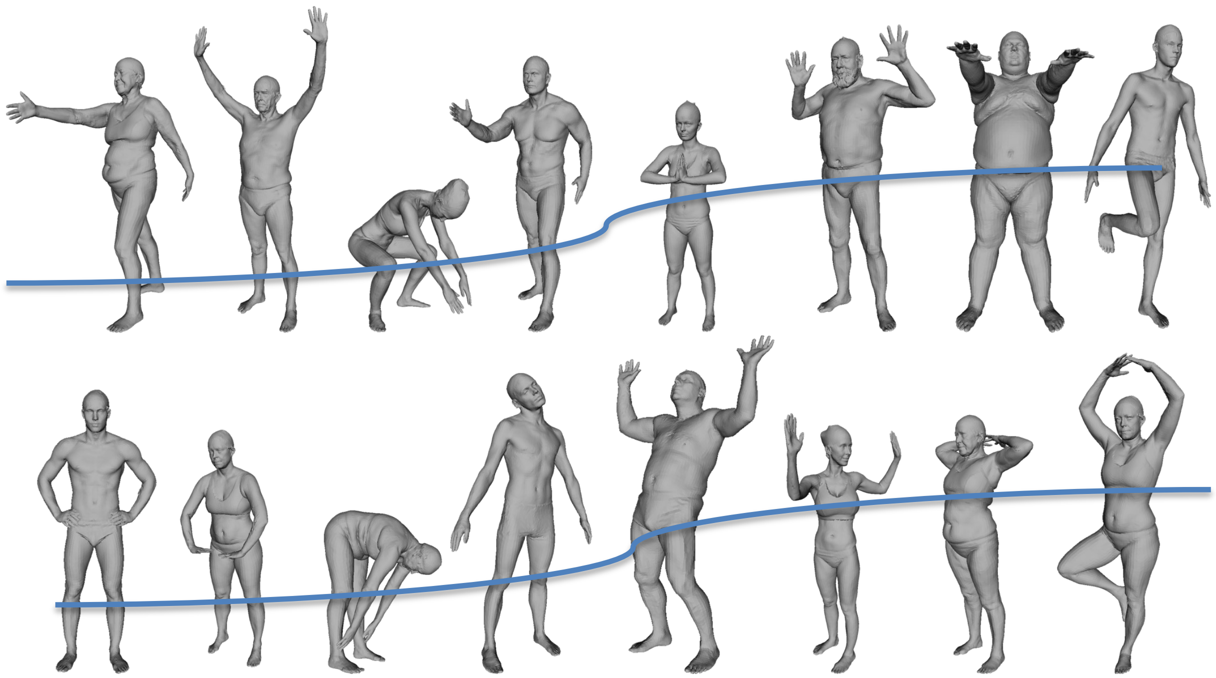
Patel et al. CVPR'20

TailorNet Style-Space

- Step 1: Unpose all publicly available garments of MGN ICCV'17
- Step2: Run physics based simulation on garments
- Step3: do PCA
- Alternate steps 2 and 3
- Style parameters are controlled with PCA coefficients γ

Patel et al. CVPR'20

TailorNet training data: pose and shape variation



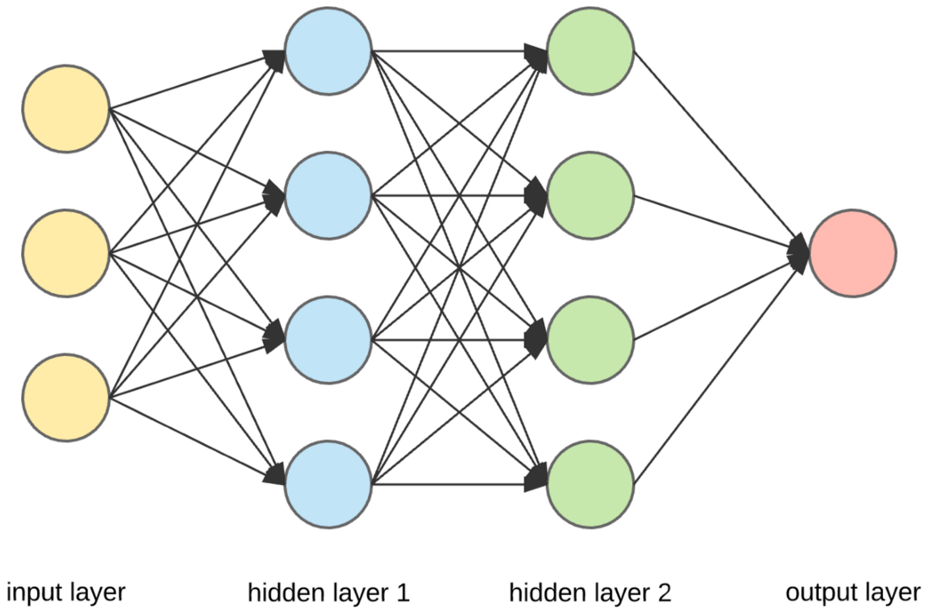
- Interpolate poses and animate the sequence with physics simulation
- Vary the shape smoothly along path
- Do this for multiple styles (garment types)
- This creates variation over pose, shape and style
- Unpose all data

Patel et al. CVPR'20

TailorNet, first idea

$$D(\theta, \beta, \gamma) : \mathbb{R}^{|\theta|} \times \mathbb{R}^{|\beta|} \times \mathbb{R}^{|\gamma|} \mapsto \mathbb{R}^{m \times 3}$$

Pose
Shape
Style



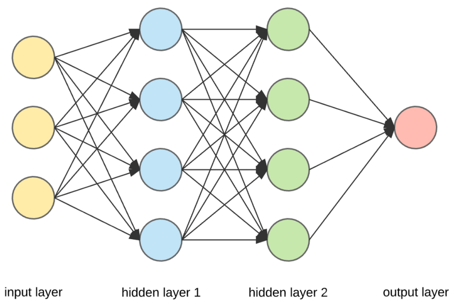
Unposed vertices

Patel et al. CVPR’20

TailorNet, first idea

$$D(\theta, \beta, \gamma) : \mathbb{R}^{|\theta|} \times \mathbb{R}^{|\beta|} \times \mathbb{R}^{|\gamma|} \mapsto \mathbb{R}^{m \times 3}$$

Pose
Shape
Style



Unposed vertices

Empirical observation: MLP generalizes well to new poses, but produces <u>smooth results</u> when trained over multiple shapes and styles

Hypothesis: High-frencency wrinkle patterns vary a lot depending on the shape and style for the same pose. When jointly learned the model averages

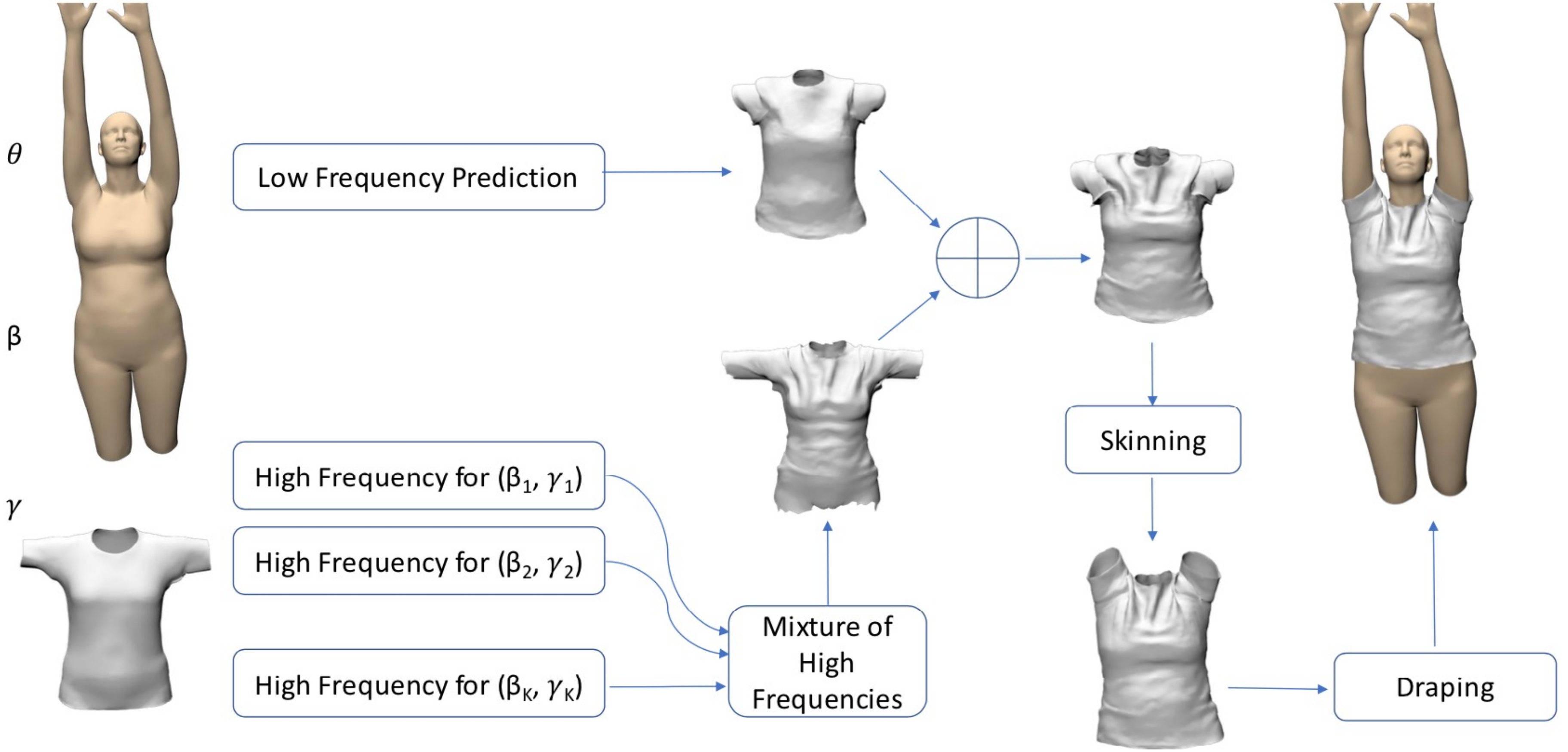
Patel et al. CVPR'20

$$D(\boldsymbol{\theta}, \phi) = D^{LF}(\boldsymbol{\theta}, \phi) + \sum_{k=1}^{K} \Psi(\phi, \phi_k) D^{HF}_{\phi,k}(\boldsymbol{\theta})$$

$$\phi = (\beta, \gamma)$$

Low-frequency

High-frequency



Prototypes

Patel et al. CVPR'20

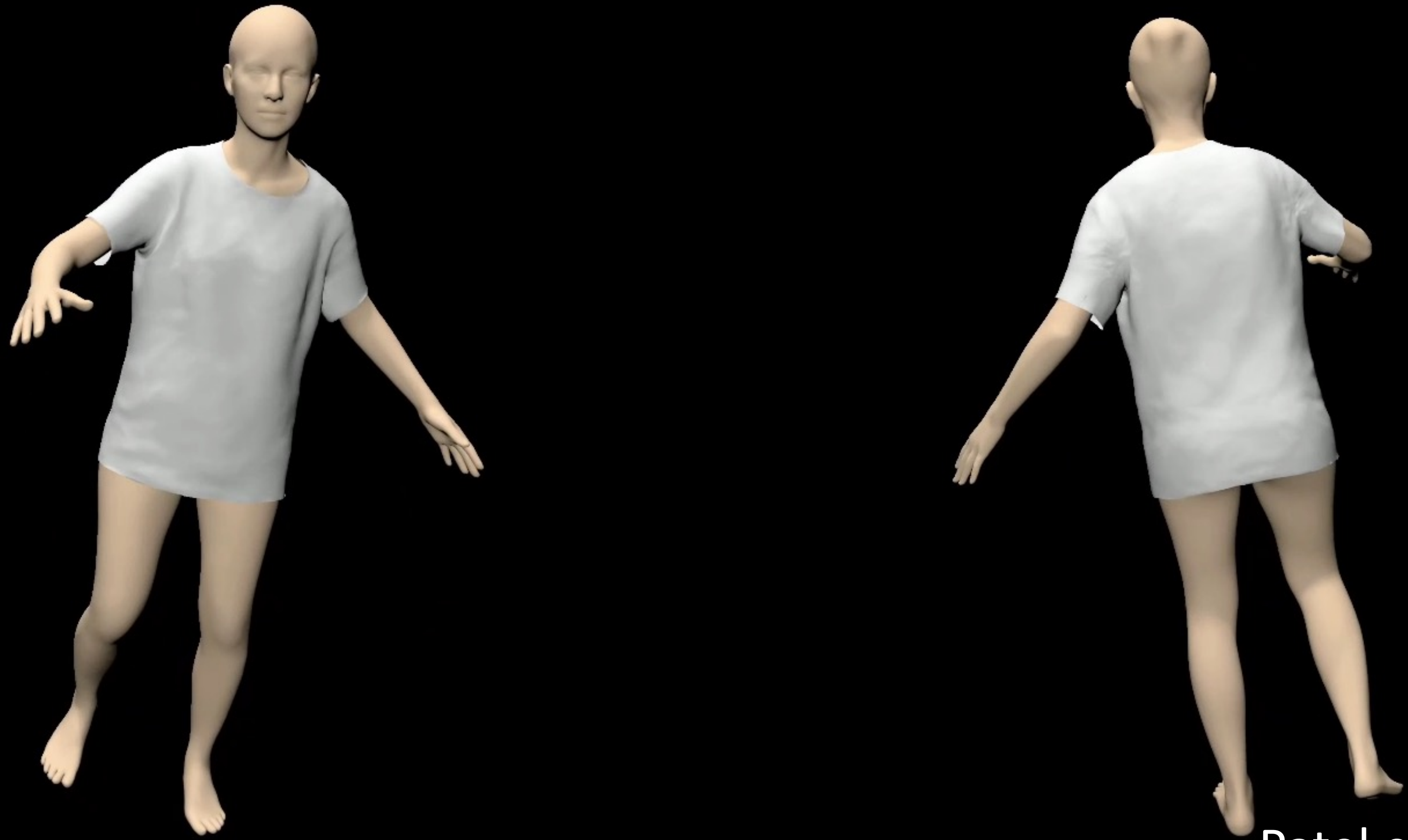
Results: Generalization to completely new poses

The Virtual Tailor: Predicting Clothing in 3D as a Function of Human Pose, Shape and Garment Style

Paper ID 6098

Patel et al. CVPR’20

Change style – keep shape fixed

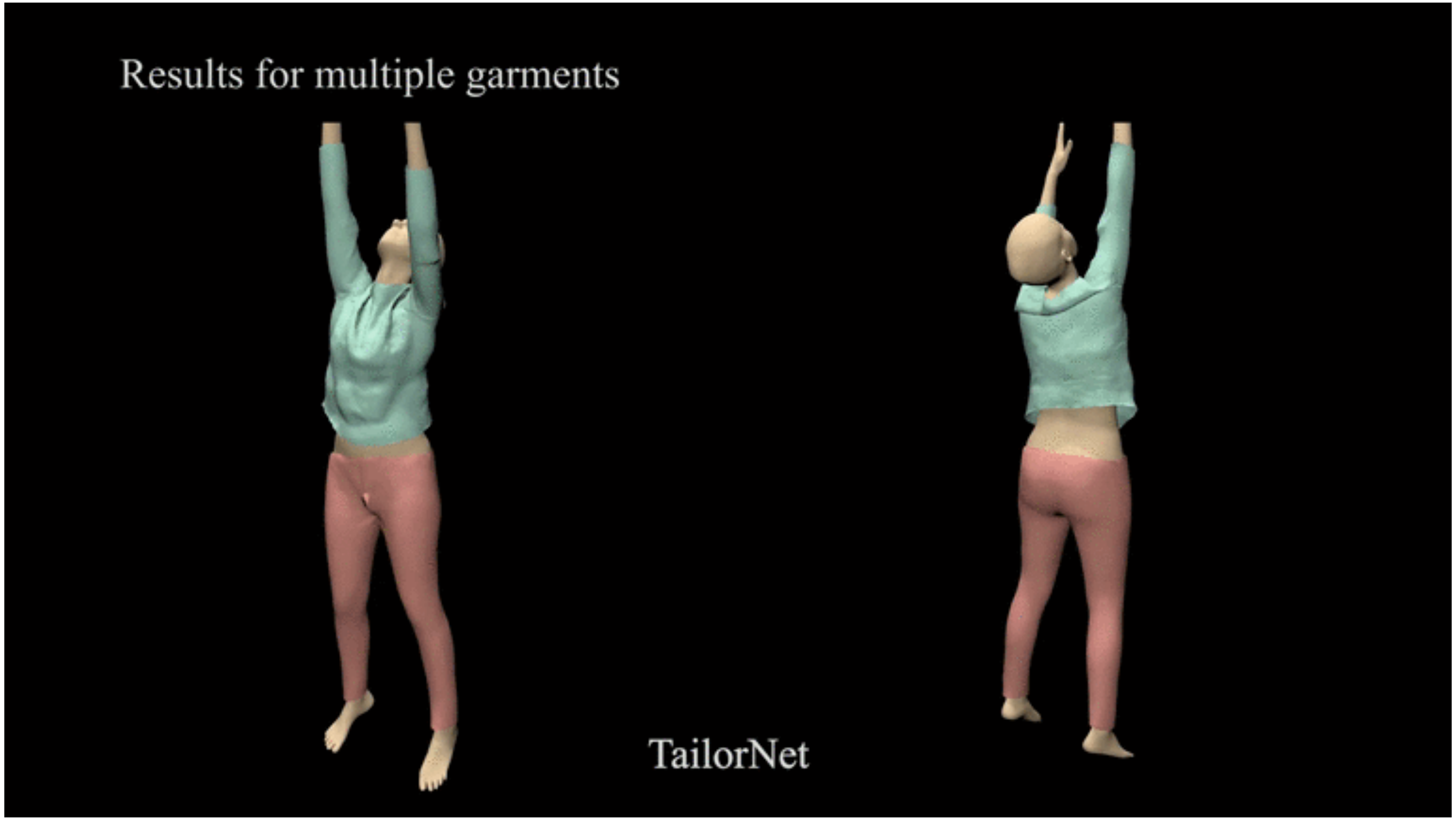


Patel et al. CVPR'20 Oral

Keep style – Change shape

Patel et al. CVPR'20

TailorNet for different garments

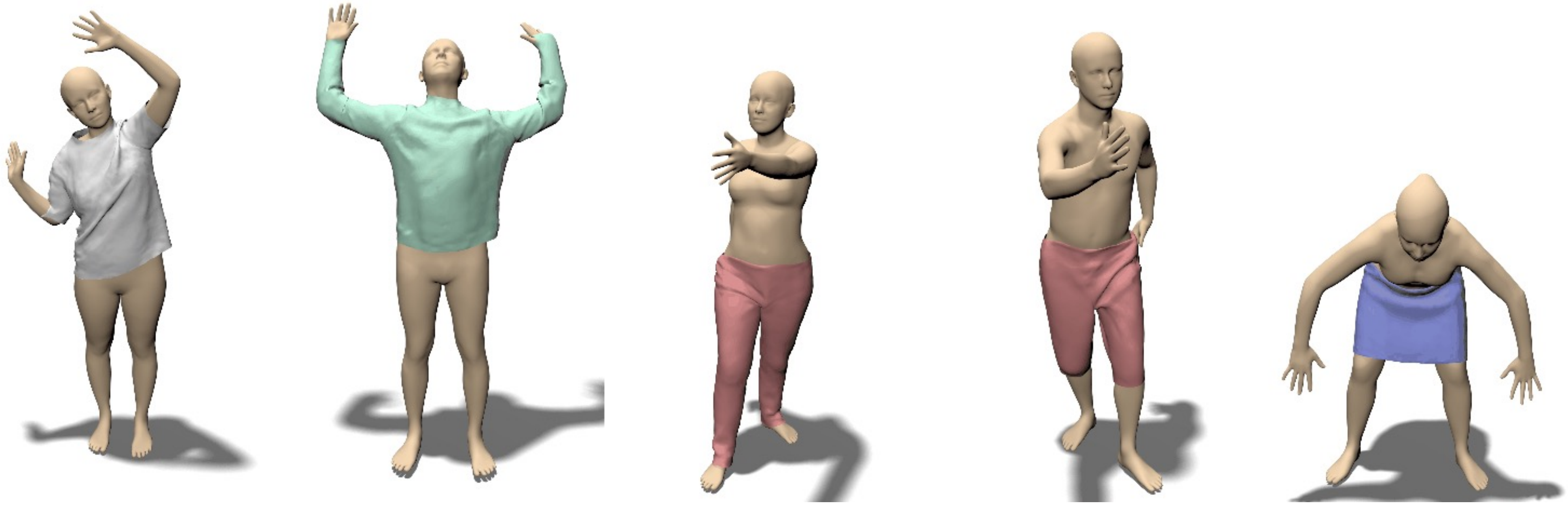


Patel et al. CVPR'20

TailorNet with Texture

Patel et al. CVPR'20

Shape variation for different garments



Patel et al. CVPR’20

Different Garments

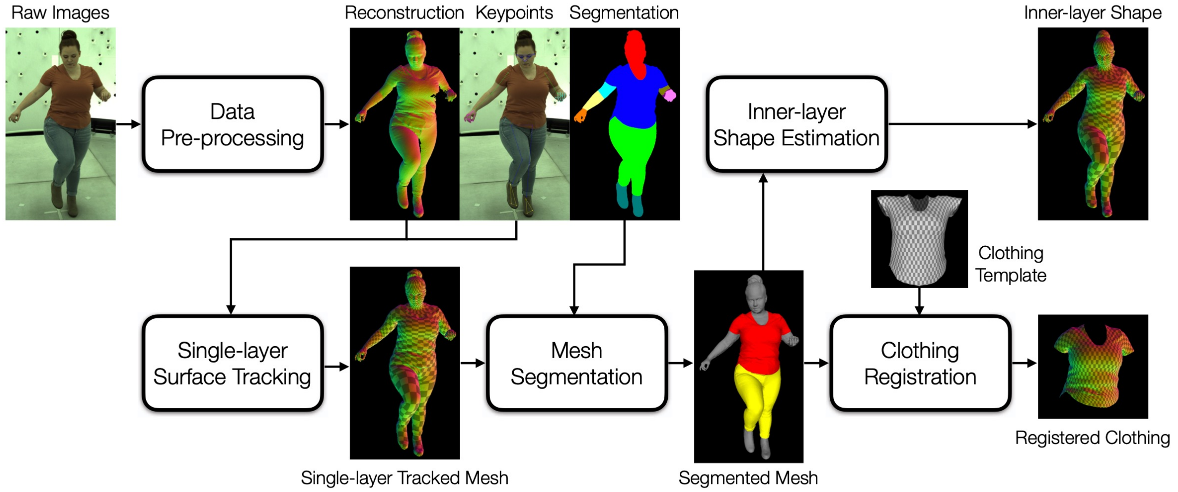
Patel et al. CVPR’20

SNUG: Self-supervised

TailorNet: Generate data with physics simulation. Train.
SNUG: Physics simulation is used within training

Santesteban et al. CVPR’22

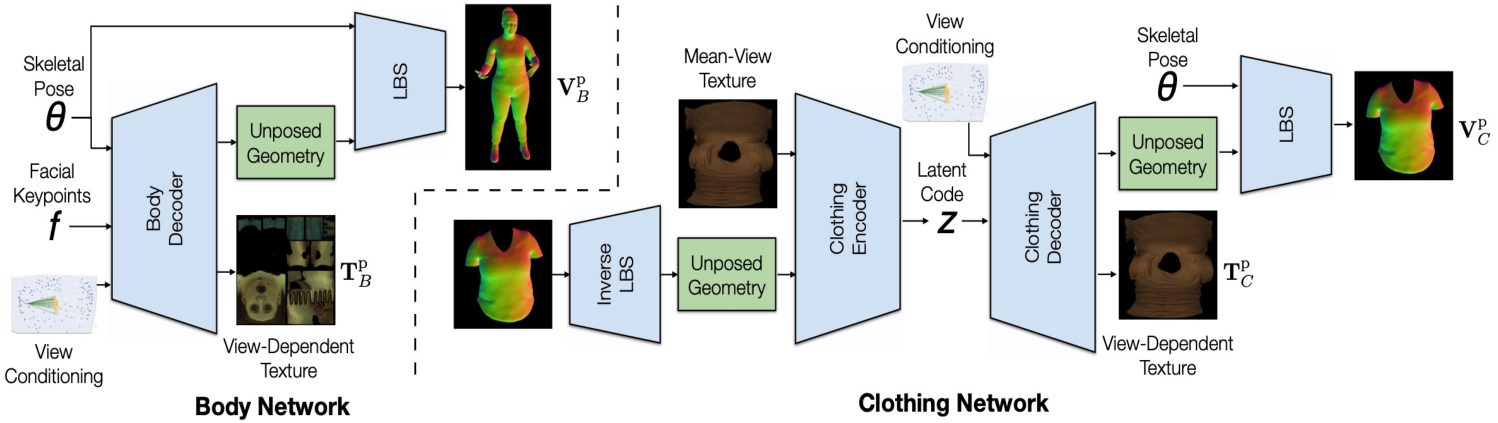
2 layer Codec Avatars (from Meta/Facebook)



<u>Training data:</u>

Registration of clothing.

Conceptually similar to ClothCap



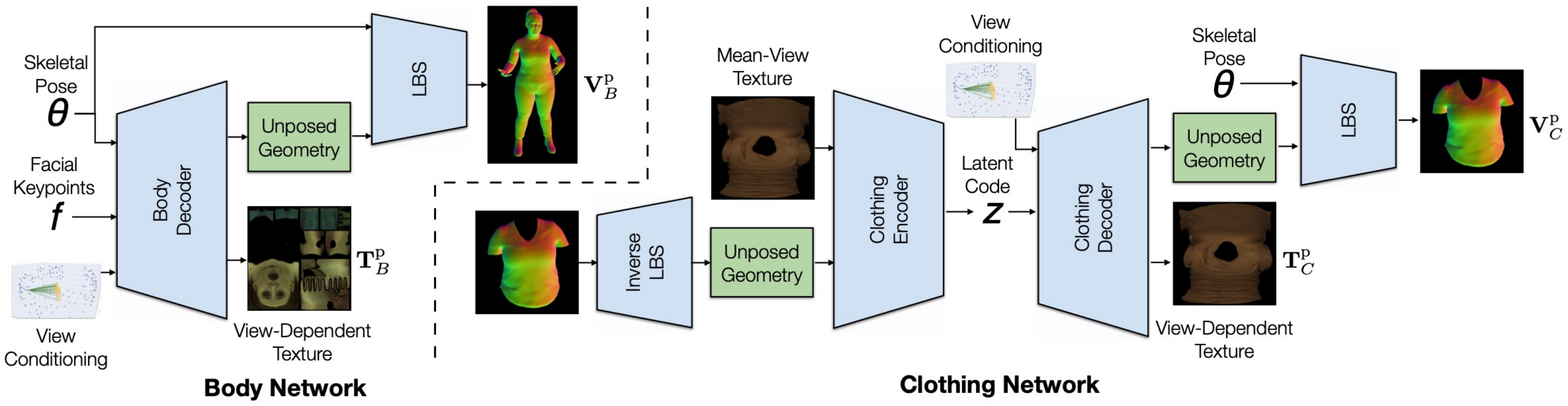
<u>Learning:</u>

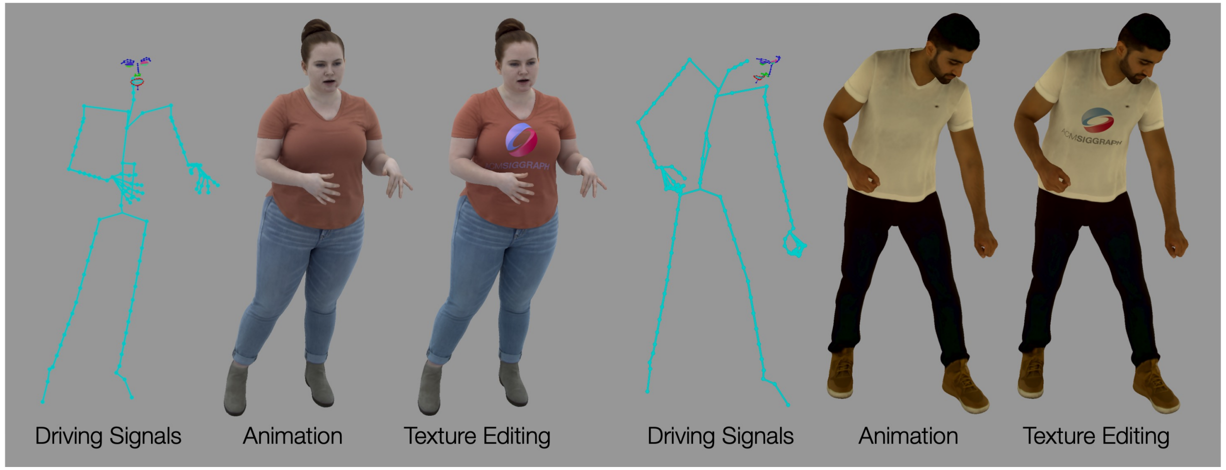
VAE to regress

- Cloth deformation
- Pose dependent texture

Xiang et al. Siggraph Asia'21

2 layer Codec Avatars (from Meta/Facebook)





Xiang et al. Siggraph Asia’21

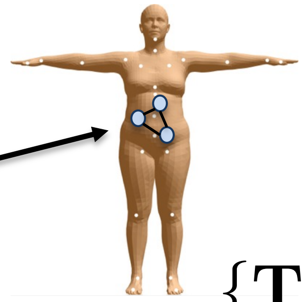
Remaining Problem

Mesh based representations are limited to surfaces with 1 “**topology**”

$\mathbf{T} \in \mathbb{R}^{3N}$

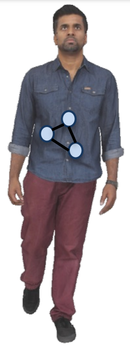
$\mathbf{F} \in \mathbb{Z}^{3N}$

Connectivity



Deform

$$\{\mathbf{T}, \mathbf{F}\} \mapsto \{\mathbf{T}', \mathbf{F}\}$$



✓ tight clothing

https://www.shopclues.com/

✗ Complex topologies

✗ General objects

CONCLUSIONS

- Clothing is much harder than undressed bodies (representation, registration, image fitting is harder)
- Vertex based models require registration of training data
- Mesh based **parametric** models like SMPL are **powerful** and **easy to use and control and compatible** with graphics pipelines
- But they are **limited to 1 topology per model**, and it is hard to produce detail
- In the next lecture: implicit surface models of clothing

Main papers in this lecture

- **Cloth registration and shape under cloth**

  Gerard Pons-Moll, Sergi Pujades, Sonny Hu, Michael Black
  ClothCap: Seamless 4D Clothing Capture and Retargeting
  in ACM Transactions on Graphics (SIGGRAPH), vol. 36, no. 4, 2017.

  Chao Zhang, Sergi Pujades, Michael Black, Gerard Pons-Moll
  Detailed, accurate, human shape estimation from clothed 3D scan sequences
  in IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2017.

- **Peope in clothing from images**

  Thiemo Alldieck, Marcus Magnor, Weipeng Xu, Christian Theobalt, Gerard Pons-Moll
  Video Based Reconstruction of 3D People Models
  in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

  Aymen Mir, Thiemo Alldieck, Gerard Pons-Moll
  Learning to Transfer Texture from Clothing Images to 3D Humans
  in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020

  Bharat Lal Bhatnagar, Garvita Tiwari, Christian Theobalt, Gerard Pons-Moll
  Multi-Garment Net: Learning to Dress 3D People from Images
  in IEEE International Conference on Computer Vision (ICCV), 2019

- **Learning models of clothing**

  Chaitanya Patel, Zhouyingcheng Liao, Gerard Pons-Moll
  TailorNet: Predicting Clothing in 3D as a Function of Human Pose, Shape and Garment Style
  in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

DATA & CODE:

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

max planck institut
informatik

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN