Virtual Humans – Winter 23/24

Lecture 8_1 – Vertex based clothing

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

\[ T(\Theta(\beta, \delta)) = T_{\mu} + B_{s}(\beta) + B_{p}(\theta) + D \]

\( \Theta \) Pose parameters

\( \beta \) Shape parameters

\( D \) Personal details + clothing
Registration with Clothing
First: Shape under Clothing

Alignment

Cloth Template

Using the single frame objective function we align all clothed scans
We create a fusion scan by gathering all cloth alignments in a single scan.
ClothCap Overview

Input: scans + garment priors

Automatic

Single mesh registration

Segmentation

Multi-part registration

Pons-Moll et al. (ClothCap) SiggraphAsia'17
Multi-part Mesh Registration

SMPL

Cloth template

Scan

Pons-Moll et al. (ClothCap) SiggraphAsia’17
Multi-part Mesh Registration

Pons-Moll et al. (ClothCap) SiggraphAsia’17
Multi-part Mesh Registration

Pons-Moll et al. (ClothCap) SiggraphAsia’17
$$E(\theta, \beta, v) = E_{\text{data}}(v) + E_{\text{cpl}}(\theta, \beta, v) +$$

Scan

Registratio

Model
\[ E(\theta, \beta, v) = E_{\text{data}}(v) + E_{\text{cpl}}(\theta, \beta, v) + \]
\[ E(\theta, \beta, \mathbf{v}) = E_{\text{data}}(\mathbf{v}) + E_{\text{cpl}}(\theta, \beta, \mathbf{v}) + E_{\text{boundary}}(\mathbf{v}) + E_{\text{lap}}(\mathbf{v}) \]
Objective function terms: data term

\[ E_{\text{data}}(\mathbf{v}) = \sum_{g=1}^{N} E_{g}(\mathbf{v}_g; S_g) \]

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

\[ E_{\text{boundary}}(v) = \sum_{g=1}^{N} E_g(v_g; S_g) \]

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

\[ \gamma(s) = (x(s), y(s), z(s)) \]

Curvature squared

\[ k(s)^2 = x''(s)^2 + y''(s)^2 + z''(s)^2 \]

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

\[ E_{\text{smth}}(\mathbf{v}) = \sum_{r=1}^{R_l} \sum_{n} \| \mathbf{v}_{r,n-1} - 2\mathbf{v}_{r,n} + \mathbf{v}_{r,n+1} \|^2 \]

Rings

Ordered vertices along the ring
Objective function: Laplacian Term

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

$$Z_{ij} = \begin{cases} 
1, & \text{if } v_i \text{ and } v_j \text{ are connected} \\
0, & \text{otherwise.}
\end{cases}$$

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

The Graph Laplacian is defined as

$$G_{\text{lap}} = I - H^{-1}Z$$

Pons-Moll et al. (ClothCap) SiggraphAsia’17
Objective function: Laplacian Term

To make a mesh smooth, we minimize a Laplacian term

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

Graph Laplacian matrix for garment $g$
Vertices of garment $g$

Pons-Moll et al. (ClothCap) SiggraphAsia’17
Objective function: Laplacian Term

$$E_{\text{lap}}(\mathbf{v}) = \sum_{g=1}^{N} \mathbf{g}_{\text{arm}} \mathbf{x}_{g}^{k} \mathbf{G}_{\text{lap}}^{g} \mathbf{v}_{g}$$

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

$$\delta_{i} = \frac{1}{d_{i}} \sum_{j \in N(i)} (v_i - v_j)$$

$$\frac{1}{|\gamma|} \int_{\gamma} (v_i - v) dl(v)$$

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.

See Sorkine et al. Laplacian Mesh Processing. EG’05
Objective function: Laplacian Term

What are we minimizing?

$$E_{\text{lap}}(v) = \sum_{g=1}^{N_{\text{garm}}} \| G_{\text{lap}}^g v_g \|_F^2$$

We minimize the norm of differential coordinates

$$E(v_i) = \left\| v_i - \frac{1}{H_{ii}} \sum_{j \in \mathcal{N}_j} v_j \right\|^2$$

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.

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

\[ D^g = G^g - I^g T(\beta, 0_\theta, 0_D) \]

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

\[ D^g = G^g - I^g T(\beta, 0_\theta, 0_D) \]

\[ T^g(\beta, \theta, D^g) = I^g T(\beta, 0_\theta, 0_D) + D^g \]

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

ClothCap Cloth on new Body

Pons-Moll et al. Siggraph’17 [ClothCap]
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
Optimization

Optimize all poses at once is slow

Output codes
- pose $\theta$
- shape $\beta$
- clothing $D$

3D world

$M(\theta, \beta, D)$

$R(\cdot)$

Video

$\arg \min_{\theta, \beta, D} \sum_i \text{dist}(R(M(\theta_i, \beta, D)), I_i)$

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
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
Key idea: transform the silhouette cones according to the inverse of non-rigid motion
\[ \mathbf{r} = \left( \sum_{k=1}^{K} w_{k,i} G_k(\theta, J_\beta) \right)^{-1} \mathbf{r}' - b_{P,i}(\theta). \]
Optimize a Single Shape to Fit all Unposed Silhouette Cones

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

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

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

Sum of point to line distances
Code and data:
https://graphics.tu-bs.de/people-snapshot
Limitations

• Optimization: Local minima and slow

• Clothing as a single offset field is limiting:
  – Can not separate body from clothing
Self-supervised Full Surface Reconstruction

\[
\begin{align*}
\theta &\mapsto \theta(I, \lambda), & \beta &\mapsto \beta(I, \lambda), & c &\mapsto c(I, \lambda)
\end{align*}
\]

CNN front-end

3D Model

\[M(\theta, \beta, c)\]

Output codes

\[\theta, \beta, c\]

3D world

Video

CVPR’19

Thiemo Alldieck, Marcus Magnor, Bharat Lal Bhatnagar, Christian Theobalt, Gerard Pons-Moll
Learning to Reconstruct People in Clothing from a Single RGB Camera

ICCV’19

Bharat Lal Bhatnagar, Garvita Tiwari, Christian Theobalt, Gerard Pons-Moll
Multi-Garment Net: Learning to Dress 3D People from Images

CVPR’20

Marc Habermann, Weipeng Xu, Michael and Zollhoefer, Gerard Pons-Moll, Christian Theobalt
DeepCap: Monocular Human Performance Capture Using Weak Supervision

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

\[ T(\theta, \beta, D) = T_\mu + B_s(\beta) + B_p(\theta) + D \]

\( \theta \) Pose parameters

\( \beta \) Shape parameters

\( D \) 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

\[ G^g = B^g z^g + D^{hf,g} \]

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

Predicting Displacements directly as a function of the image is hard
Remaining Problem: Details

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

Predicting Displacements directly as a function of the image is hard
Tex2Shape: Detailed Full Human Body Geometry from a Single Image Exploiting UV-maps

Input → UV Transform → UV Unwrap → Normals and 3D displacements → Inv. UV Transform → 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
Learning to Transfer Texture from Clothing Images to 3D Humans
Aymen Mir, Thiemo Alldieck, Gerard Pons-Moll
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

\[ T(\theta, \beta, D) = T_\mu + B_s(\beta) + B_p(\theta) + D \]

\( \theta \) Pose parameters

\( \beta \) Shape parameters

\( D \) Personal details + clothing
Goal: Clothing as a function of Pose, Shape and Style

\[ D(\theta, \beta, \gamma) : \mathbb{R}^{|\theta|} \times \mathbb{R}^{|\beta|} \times \mathbb{R}^{|\gamma|} \rightarrow \mathbb{R}^{m \times 3} \]

Vertices of garment

Patel et al. CVPR’20
TailorNet Style-Space

- Step 1: Unpose all publicly available garments of MGN ICCV’17
- Step 2: Run physics based simulation on garments
- Step 3: do PCA
- Alternate steps 2 and 3
- Style parameters are controlled with PCA coefficients $\gamma$
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
TailorNet, first idea

\[ D(\theta, \beta, \gamma) : \mathbb{R}^{|\theta|} \times \mathbb{R}^{|\beta|} \times \mathbb{R}^{|\gamma|} \rightarrow \mathbb{R}^{m \times 3} \]
TailorNet, first idea

\[ D(\theta, \beta, \gamma) : \mathbb{R}^{|\theta|} \times \mathbb{R}^{|\beta|} \times \mathbb{R}^{|\gamma|} \rightarrow \mathbb{R}^{m \times 3} \]

Empirical observation: MLP generalizes well to new poses, but produces smooth results 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.
\[ D(\theta, \phi) = D^{LF}(\theta, \phi) + \sum_{k=1}^{K} \Psi(\phi, \phi_k)D^{HF}_{\phi,k}(\theta) \]

\[ \phi = (\beta, \gamma) \]

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
Change style – keep shape fixed
Keep style – Change shape
TailorNet for different garments

Results for multiple garments
TailorNet with Texture

Patel et al. CVPR’20
Shape variation for different garments
Different Garments

Patel et al. CVPR’20
SNUG: Self-supervised

TailorNet: Generate data with physics simulation. Train.
SNUG: Physics simulation is used within training
2 layer Codec Avatars (from Meta/Facebook)

Training data:
Registration of clothing. Conceptually similar to ClothCap

Learning:
- 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”

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

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

Connectivity

Deform

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

- **Complex topologies**

- **General objects**

https://www.shopclues.com/
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

  Chao Zhang, Sergi Pujades, Michael Black, Gerard Pons-Moll
  Detailed, accurate, human shape estimation from clothed 3D scan sequences

- **People in clothing from images**
  Thiemo Alldieck, Marcus Magnor, Weipeng Xu, Christian Theobalt, Gerard Pons-Moll
  Video Based Reconstruction of 3D People Models

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