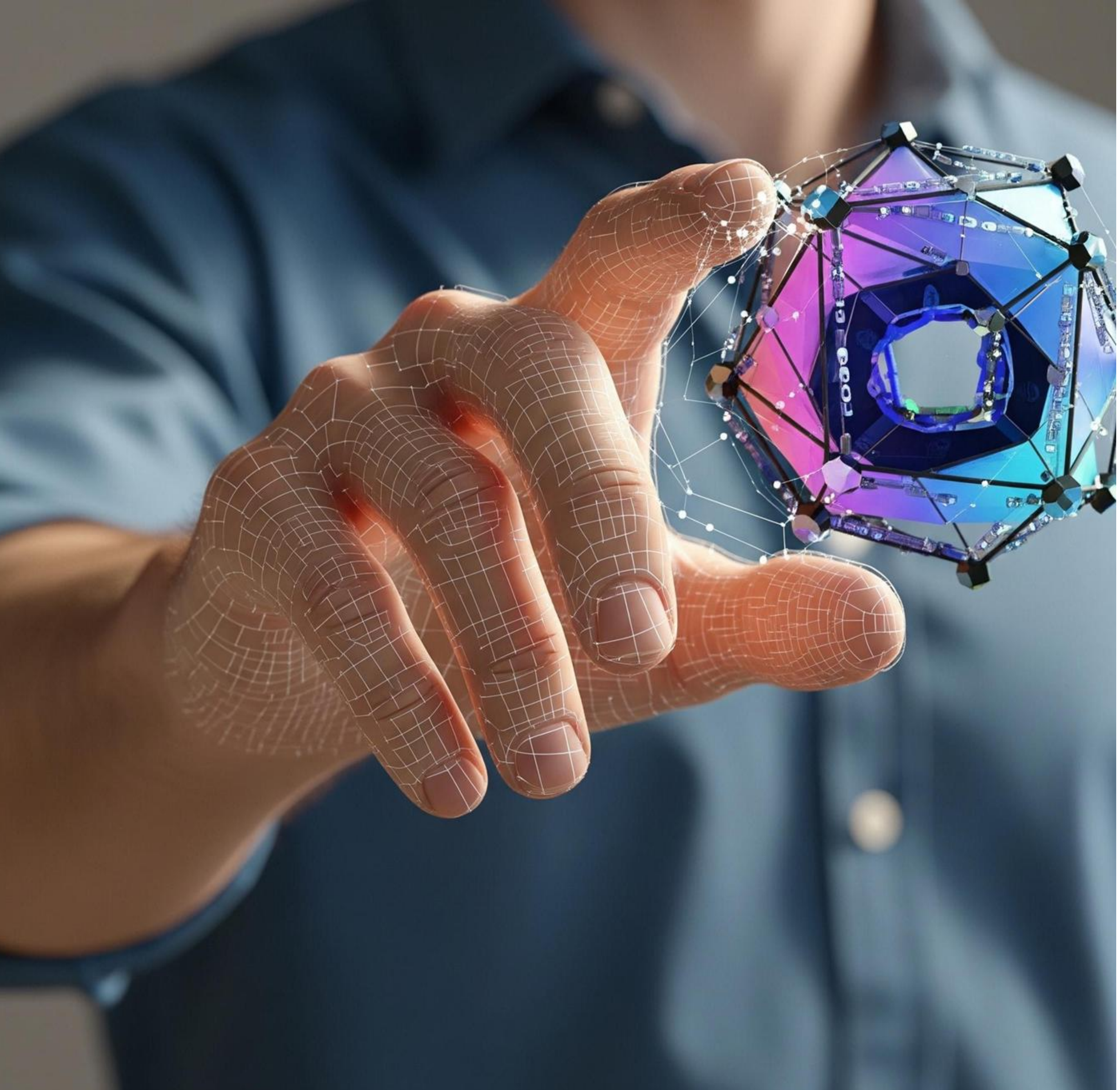
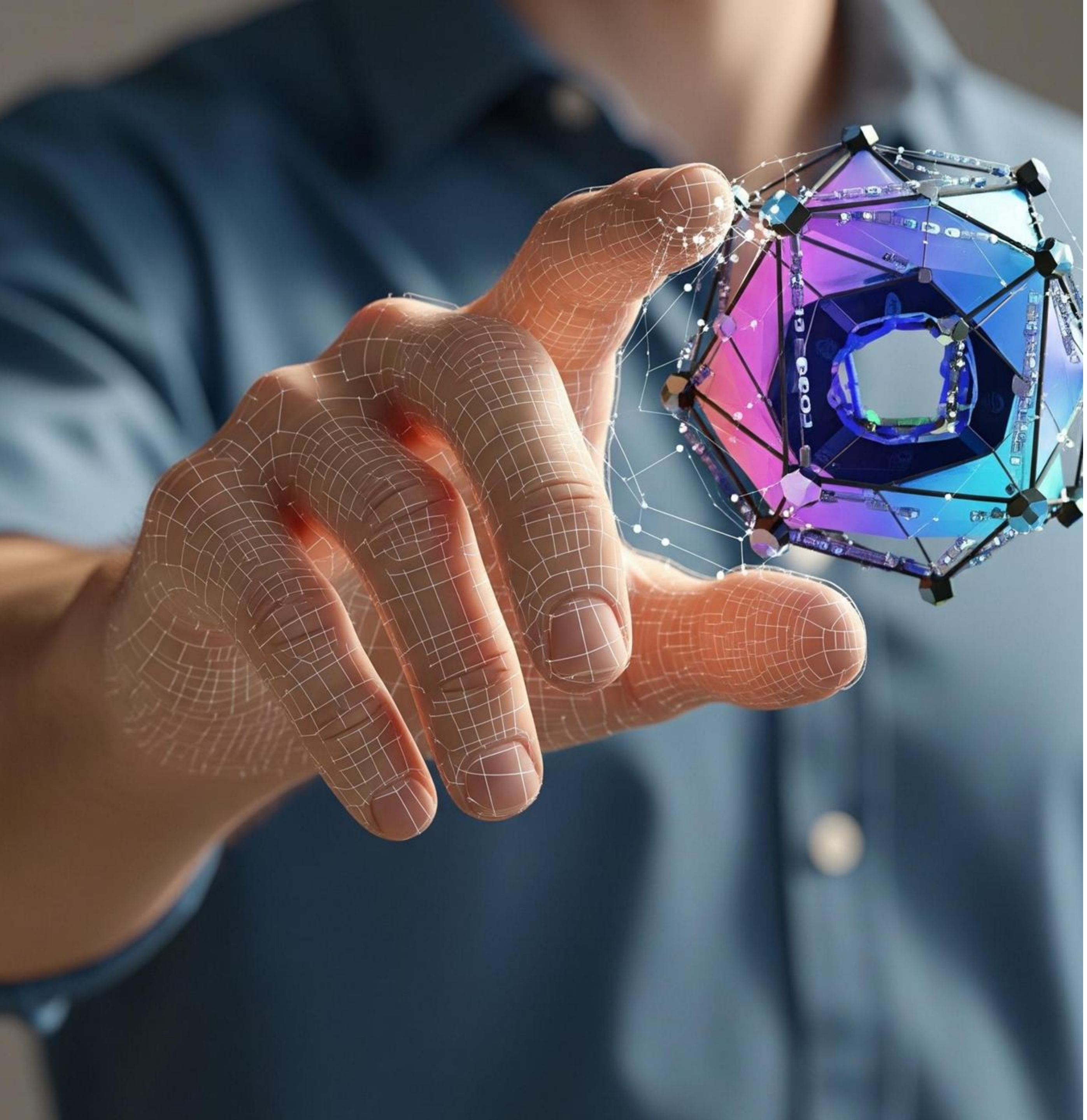


Geometric Deep Learning for Virtual Humans

Virtual Humans

Riccardo Marin

21st November 2025

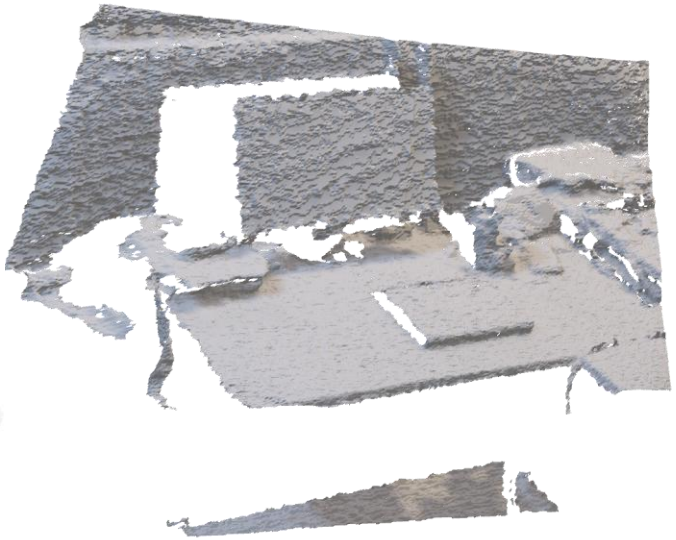


Recap

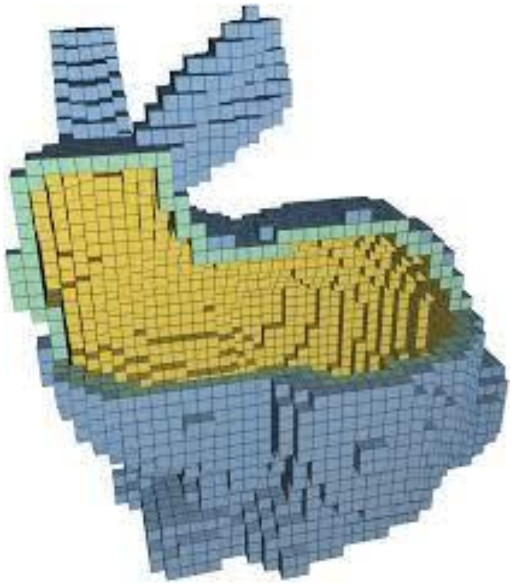

Ceci n'est pas une pipe.

Geometrical data have different structure

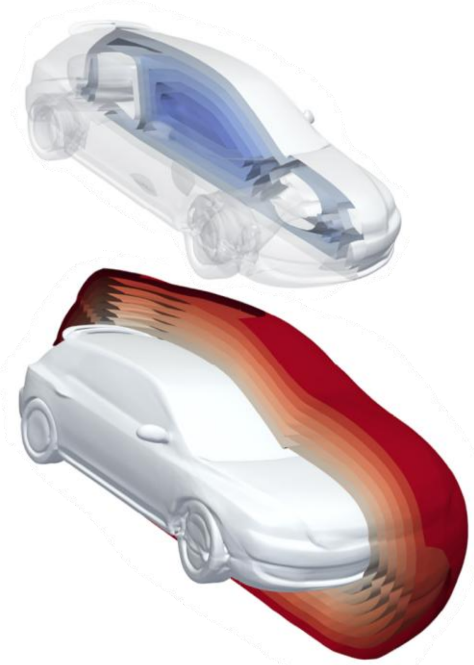
Point clouds



Range map



Volumetric



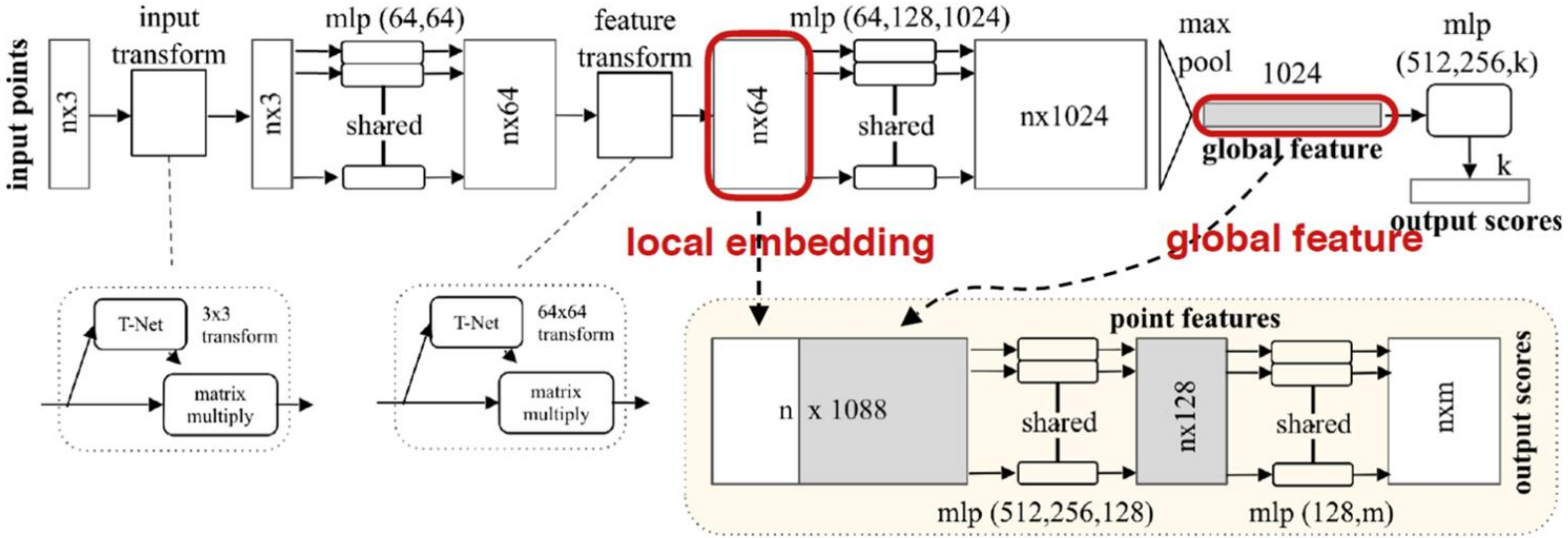
Implicit

“Fast Parallel Surface and Solid Voxelization on GPUs”, M. Schwarz et al., 2010

“Implicit Geometric Regularization for Learning Shapes”, A. Gropp et al., 2020

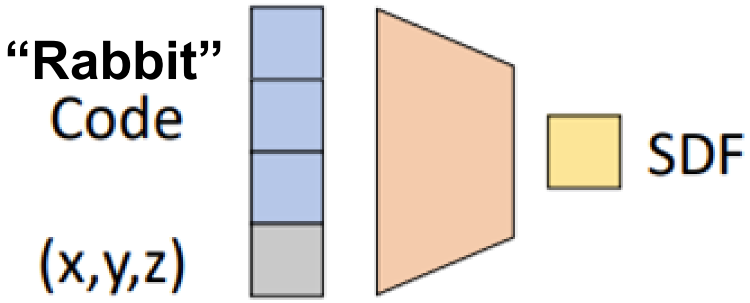
Slide credit E. *Rodolà*

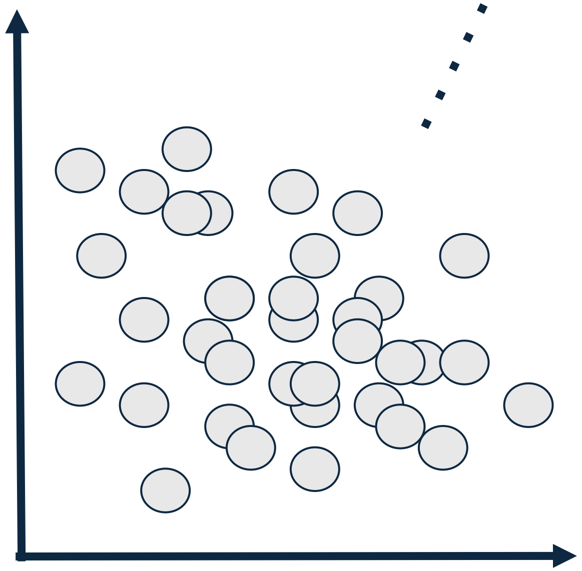
Point cloud classification and Segmentation



Point cloud Segmentation

Deep SDF



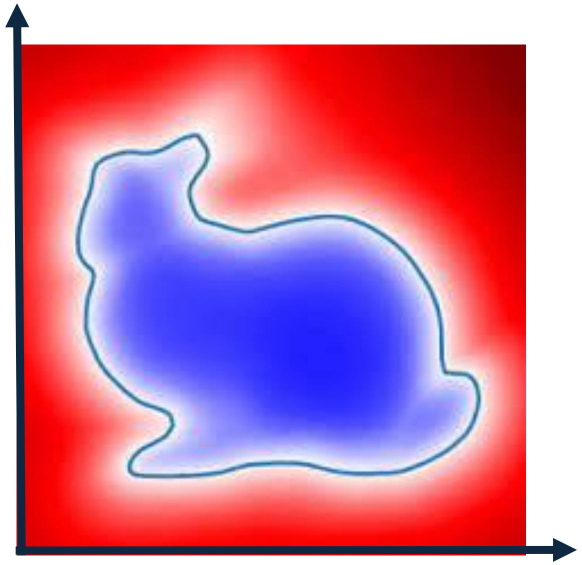


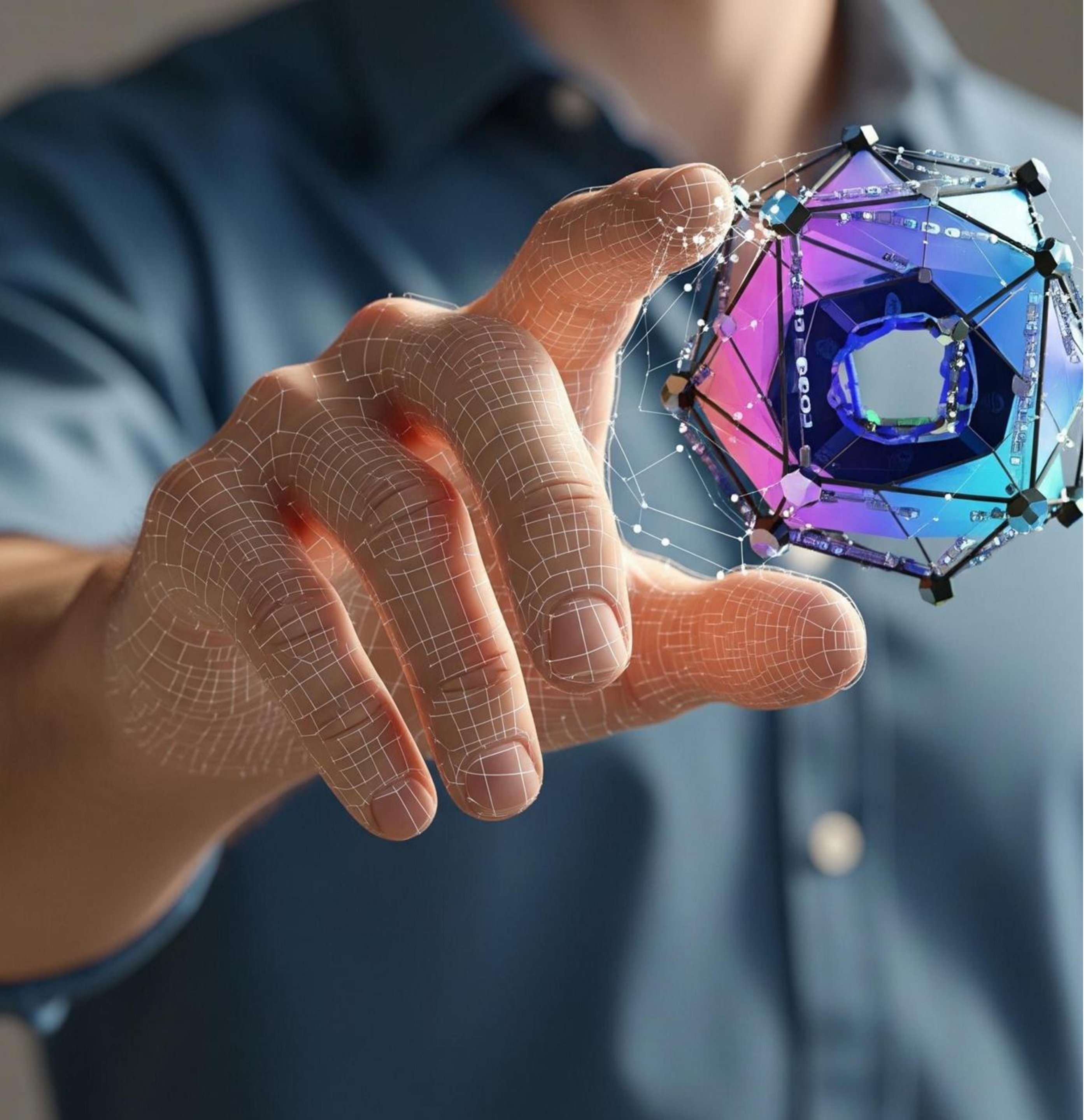
Problem

One network for a single shape

Solution

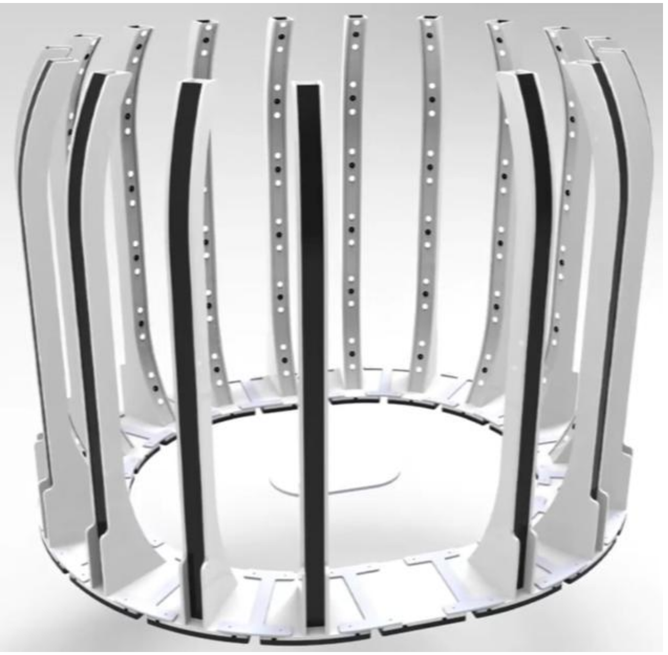
Condition the input

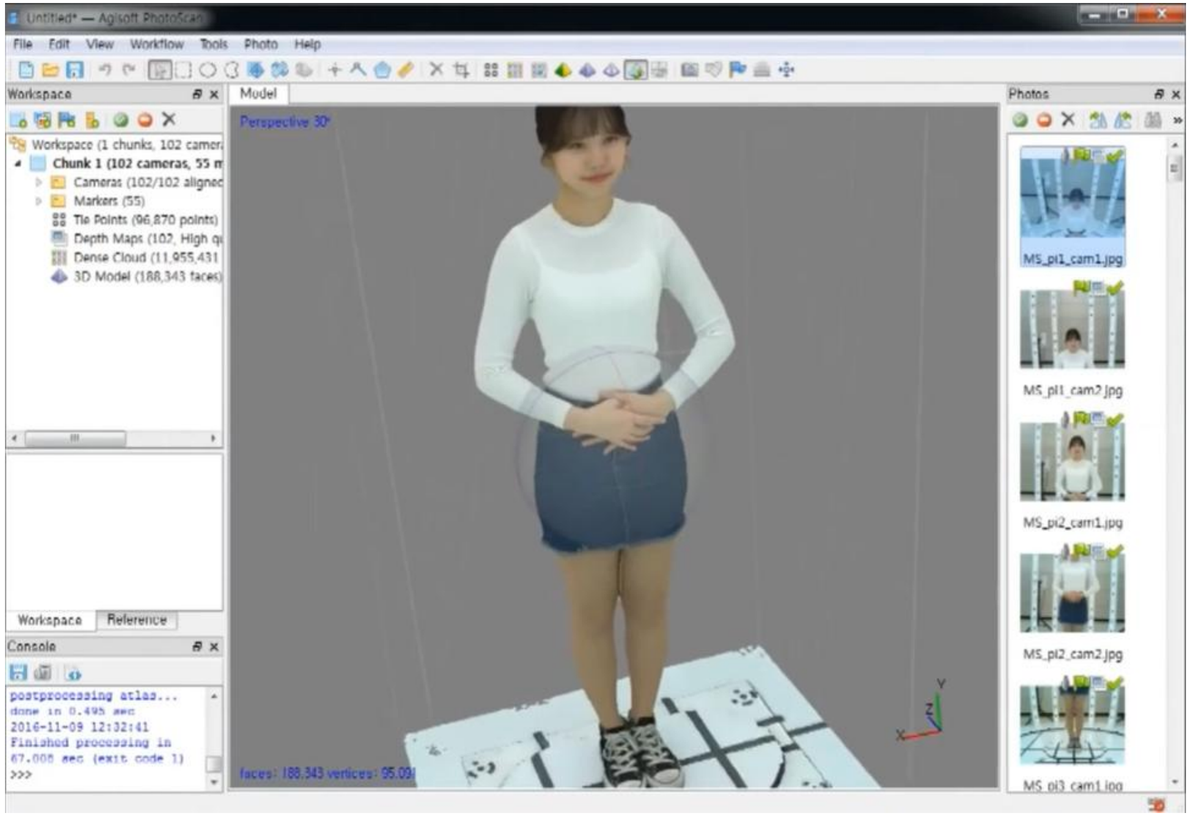




Virtual Humans

How do you obtain this?





https://www.youtube.com/watch?v=bxOtIgJ_5OA


Installation
(~10 min)
16x

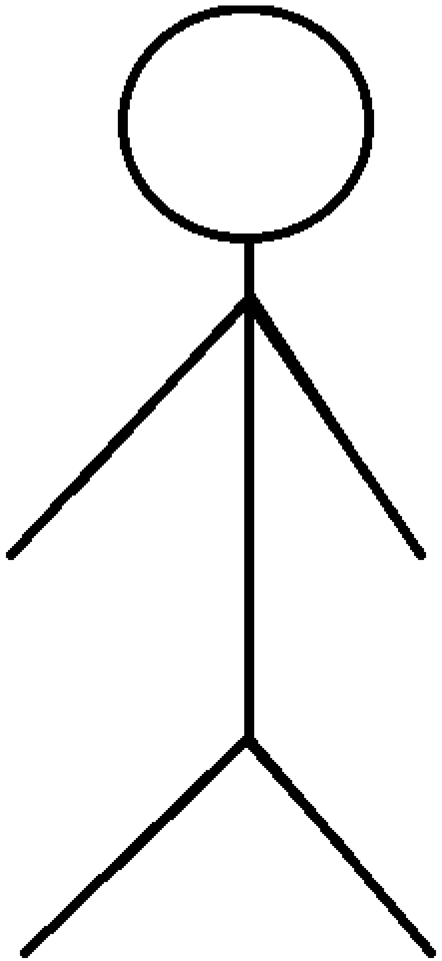
Mobile phone video

Mobile phone video

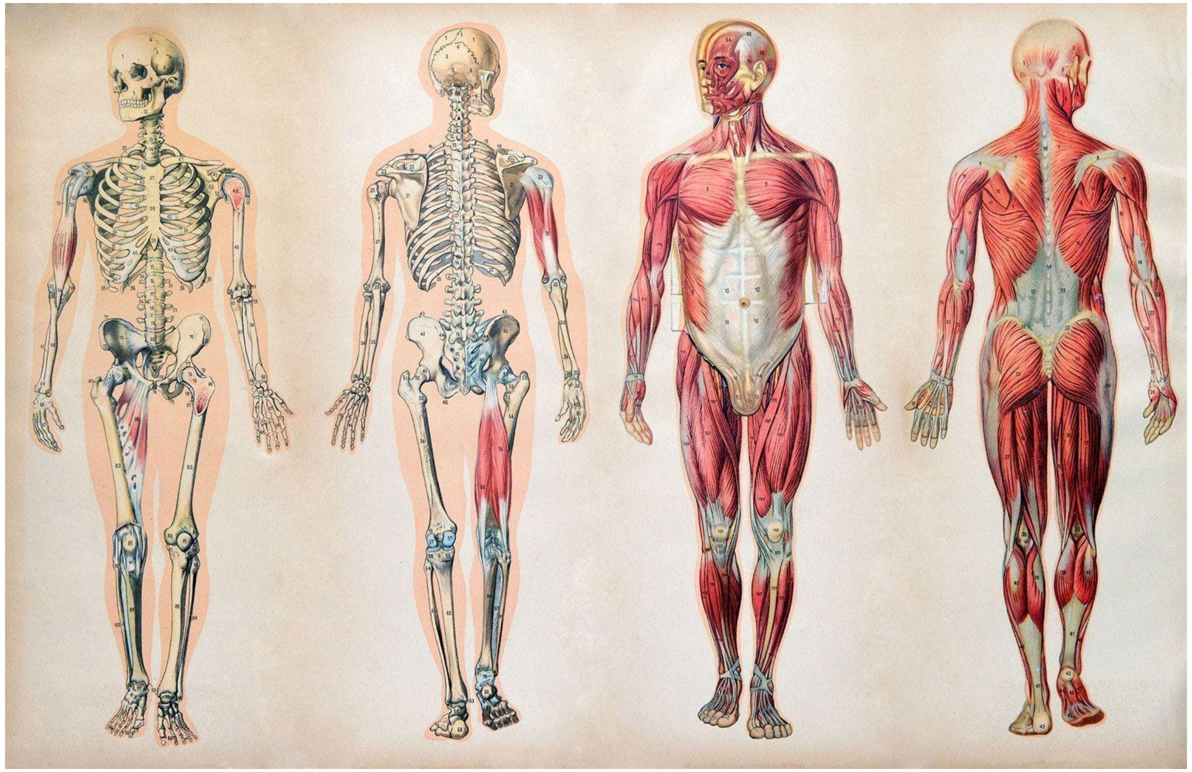
NeRF/GS Model (luma.ai)

What is a good representation for human data?

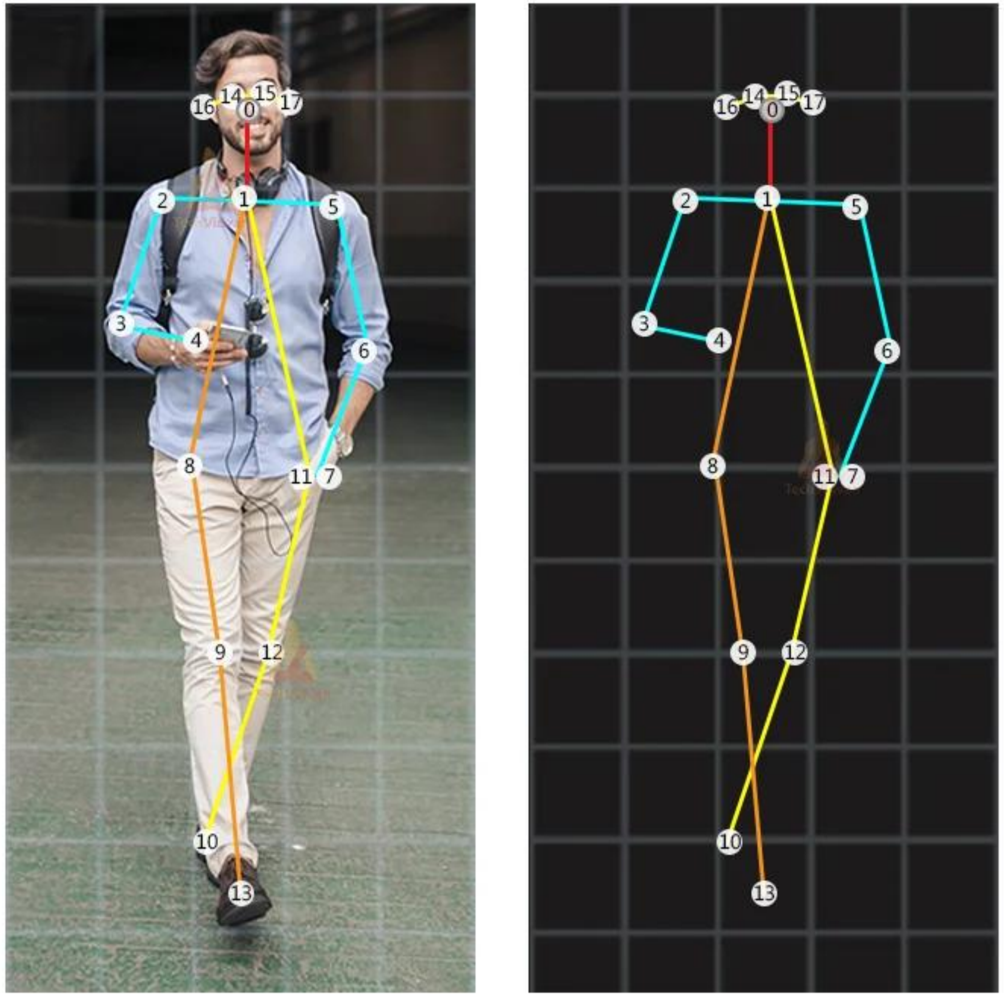
5 lines
1 circle



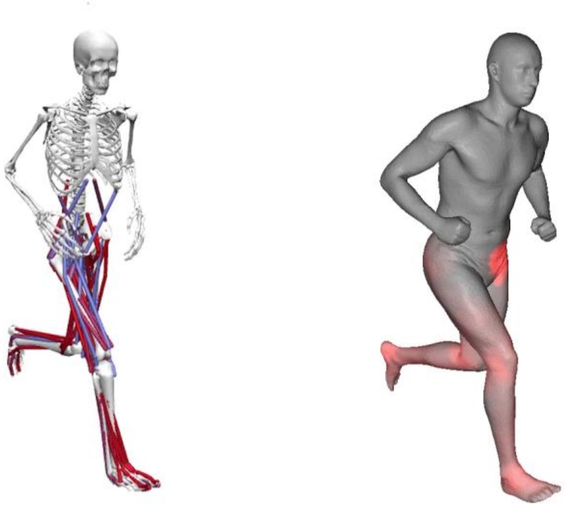
206 bones
600+ muscles
78 organs
…



What is a good representation for human data?



https://github.com/nitingour1203/human_pose_detection



BASH, Schleicher et al., 2021

The roles of a representation

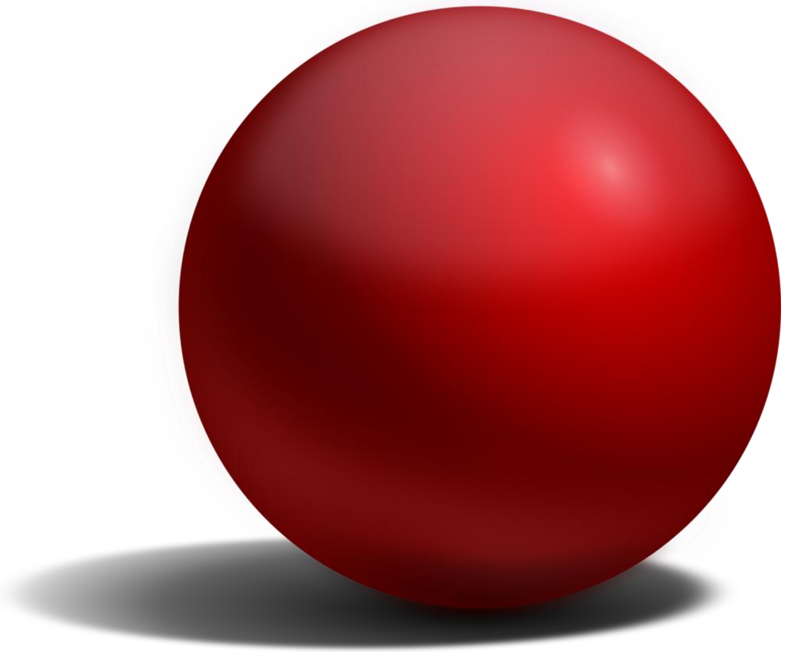
Convey the geometry

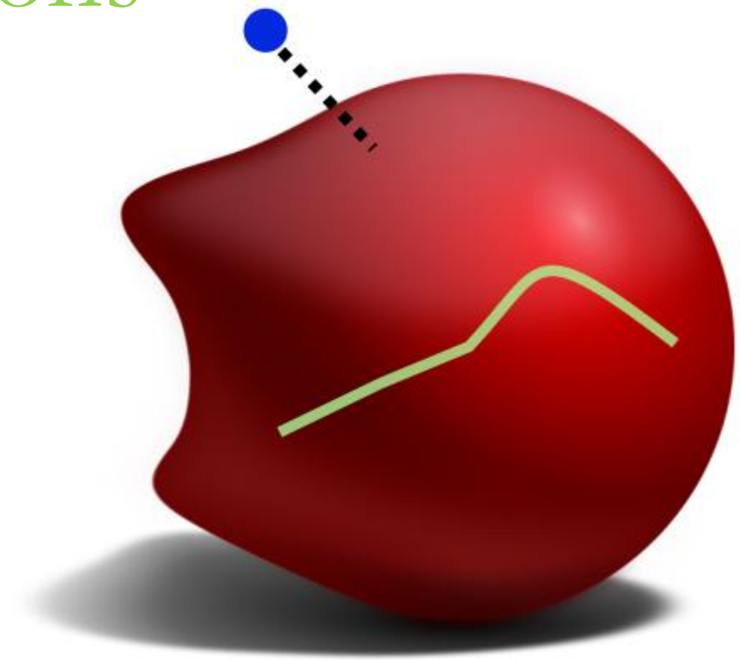
Support computations

Geometry Evaluations

Spatial Queries

Modifications

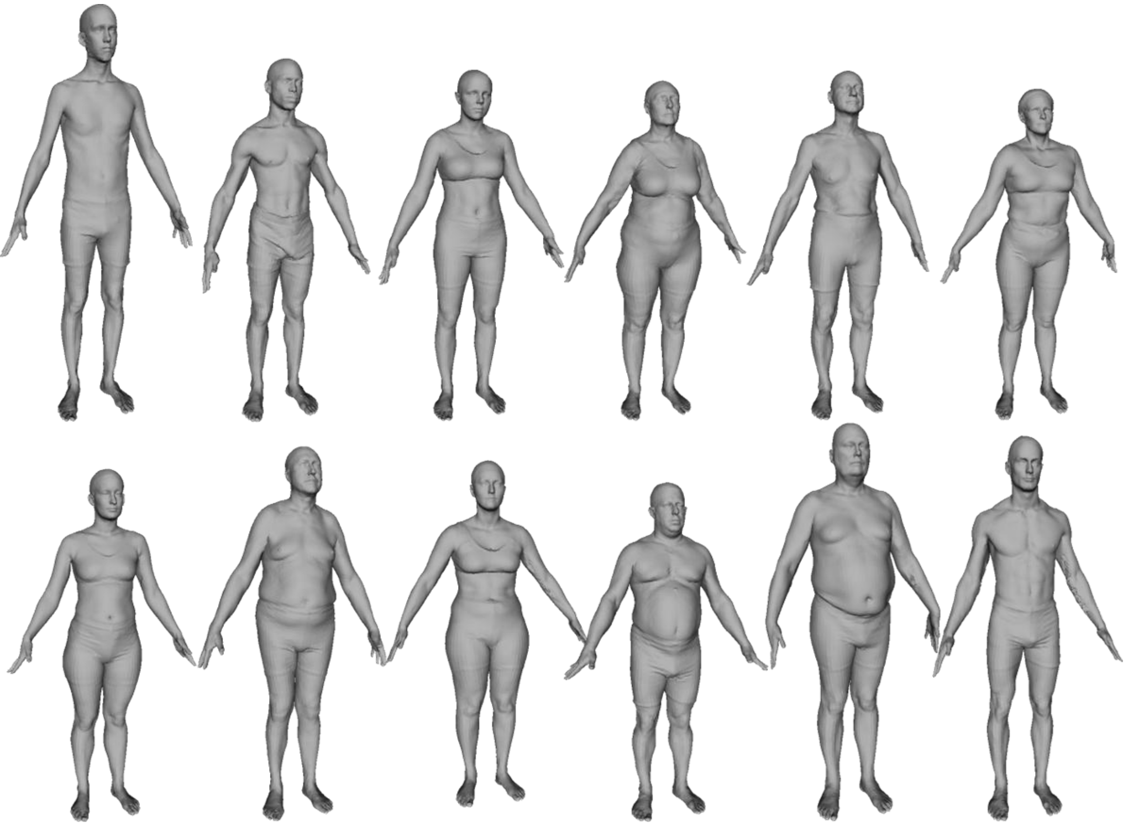




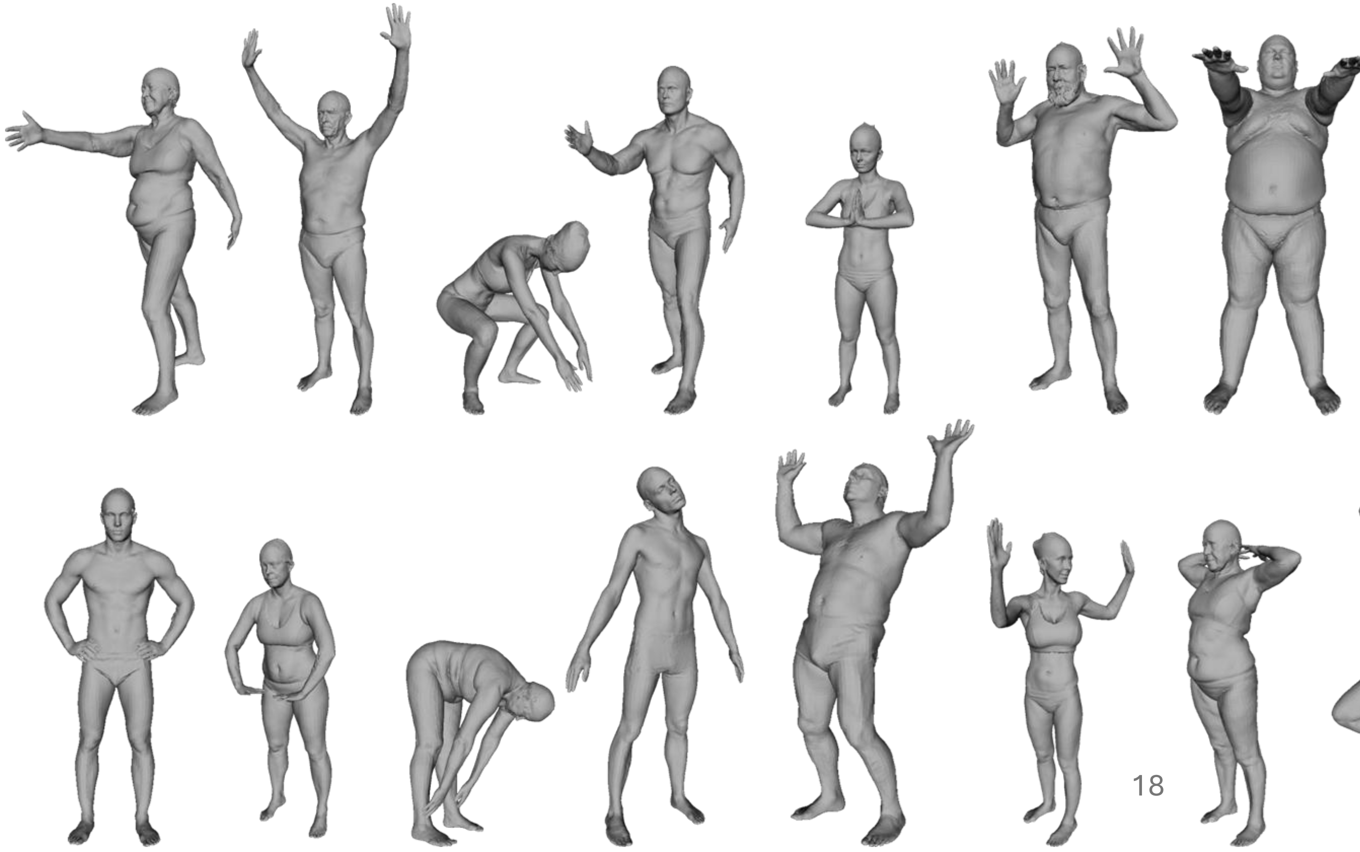
What geometry should a body representation convey?

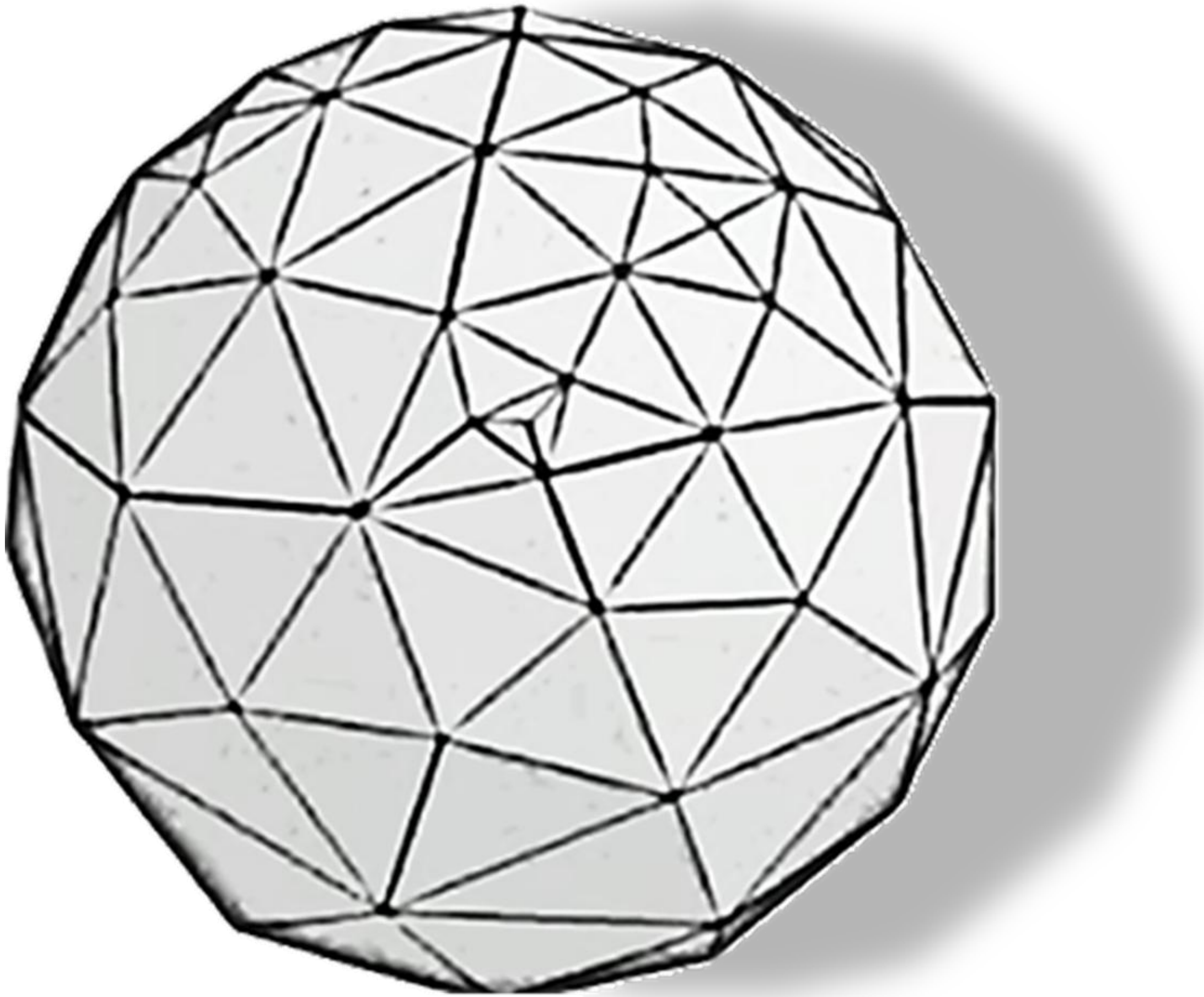
A good body model should **look and move** like real people.

Identities



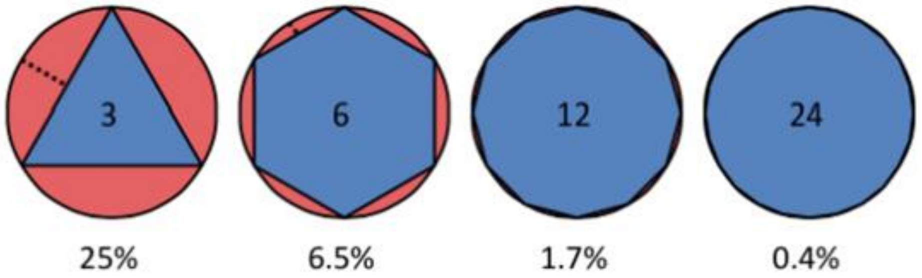
Poses





Piecewise **linear** approximation

Doubling the number of vertices reduces the error by 4



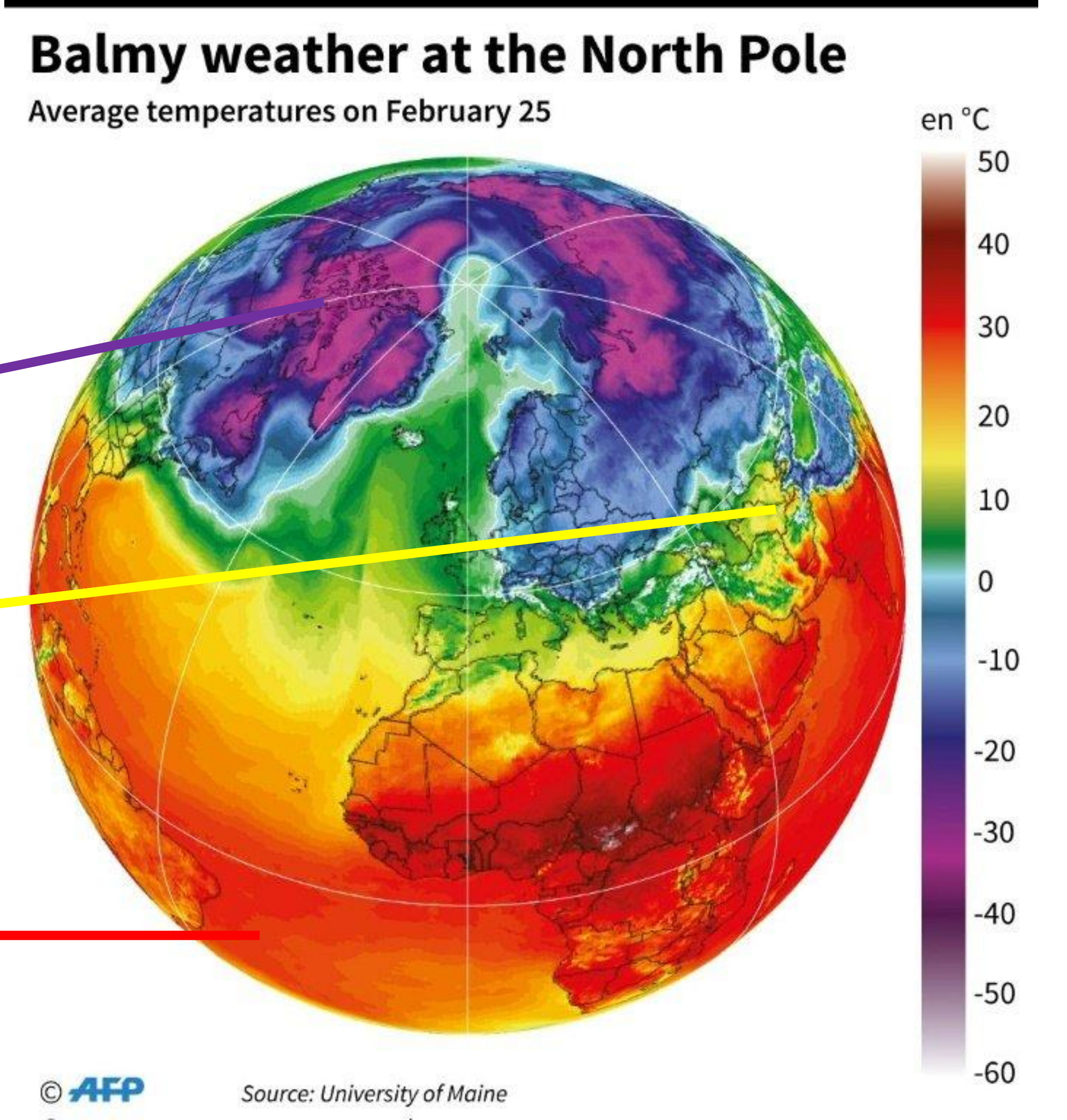
Covering the same geometry with a **point cloud would require a square number of points**

Geometry supports information

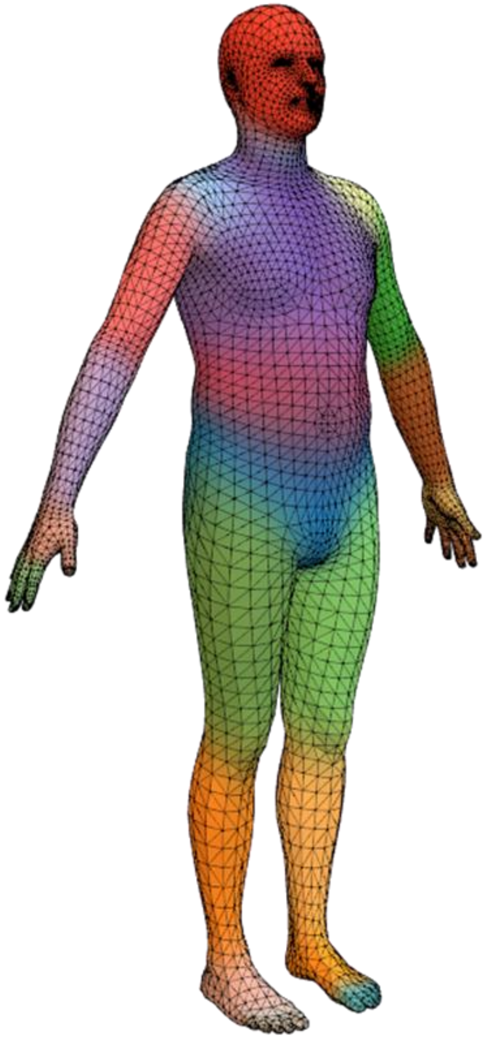
-34

16

30



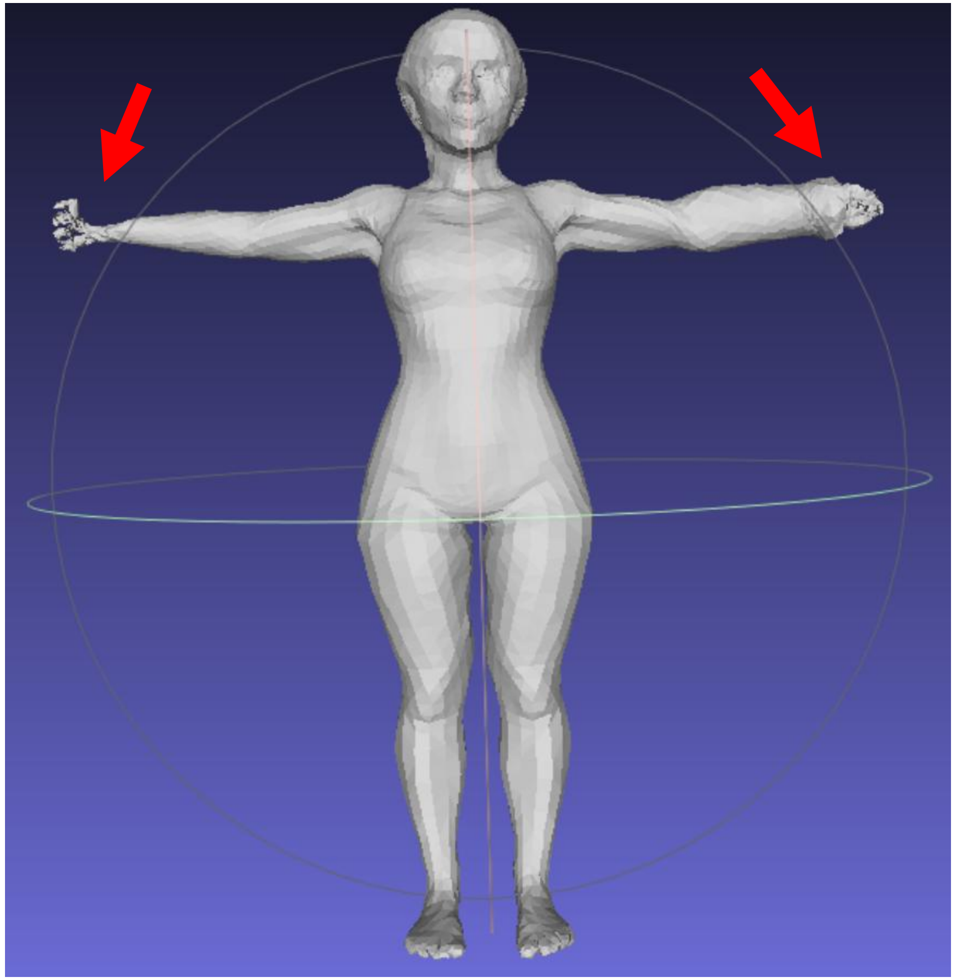
Bodies as meshes?



We represent the body as its **surface**

A discretized set of thousands of vertices and faces

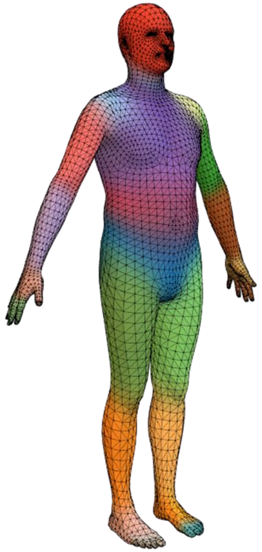
Problem: editing the identity and pose requires to specify tens of thousands of values. And not all of their combinations are realistic.



Morphing one body in different ones

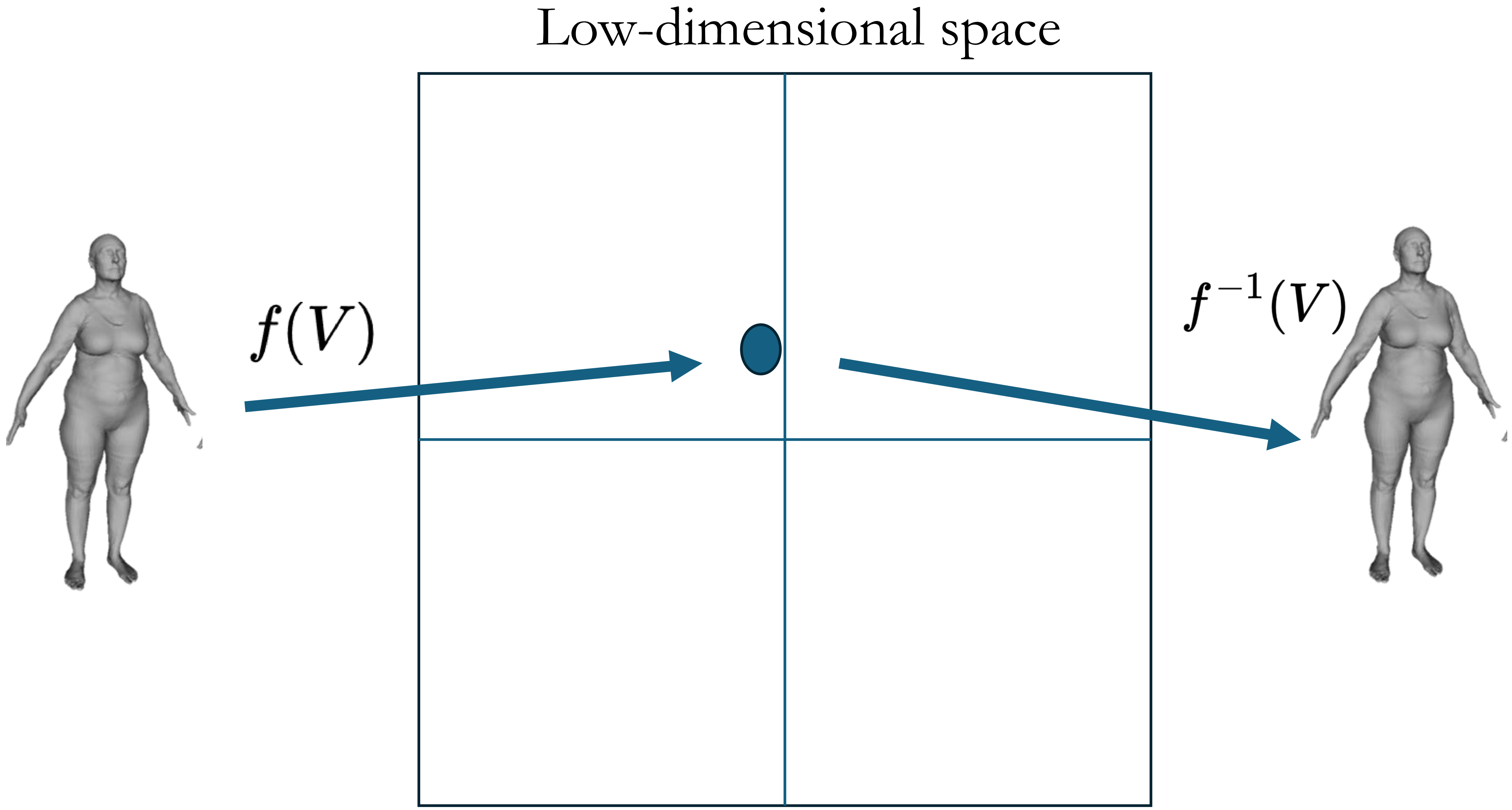
V

$=$

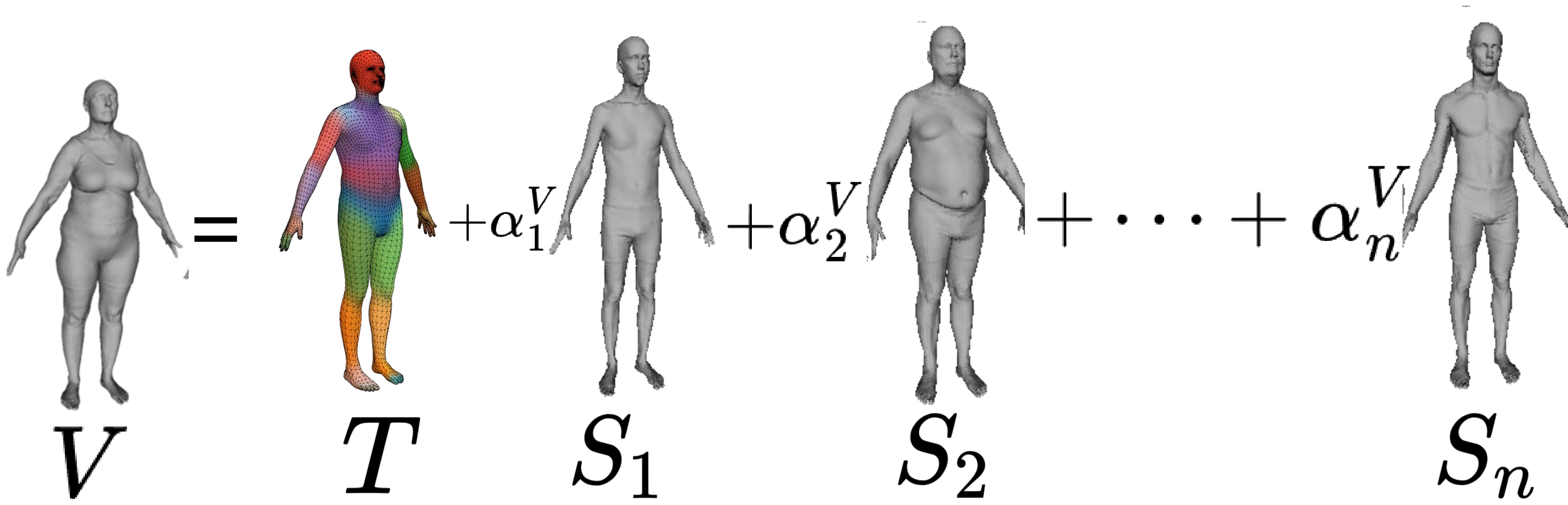


T

$+ f(V)$


Low-dimensional space
$f(V)$
$f^{-1}(V)$

Morphing one body in different ones - Linear Combination

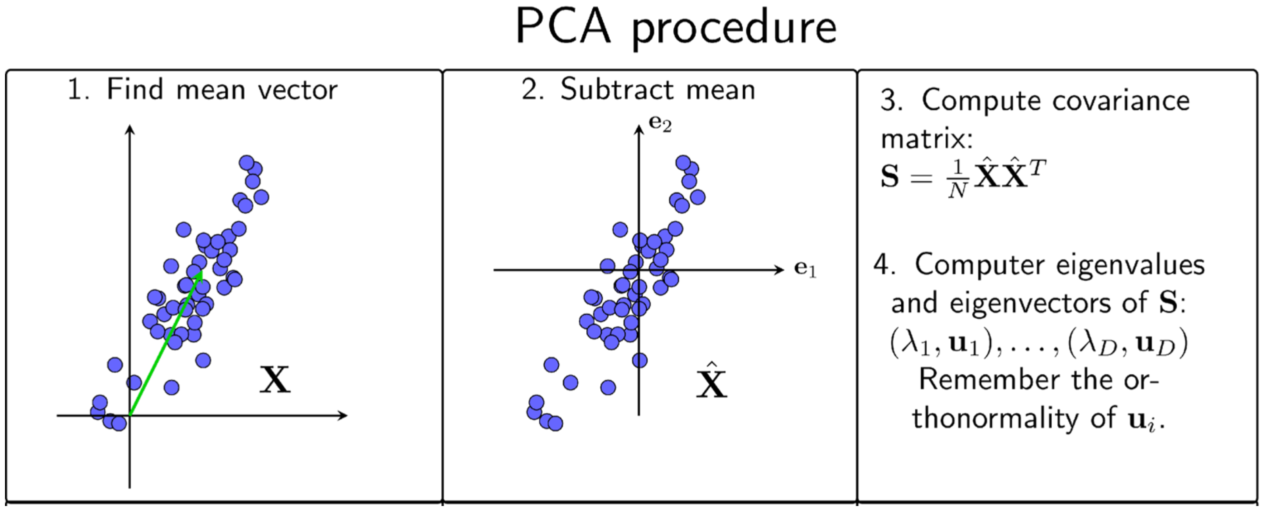


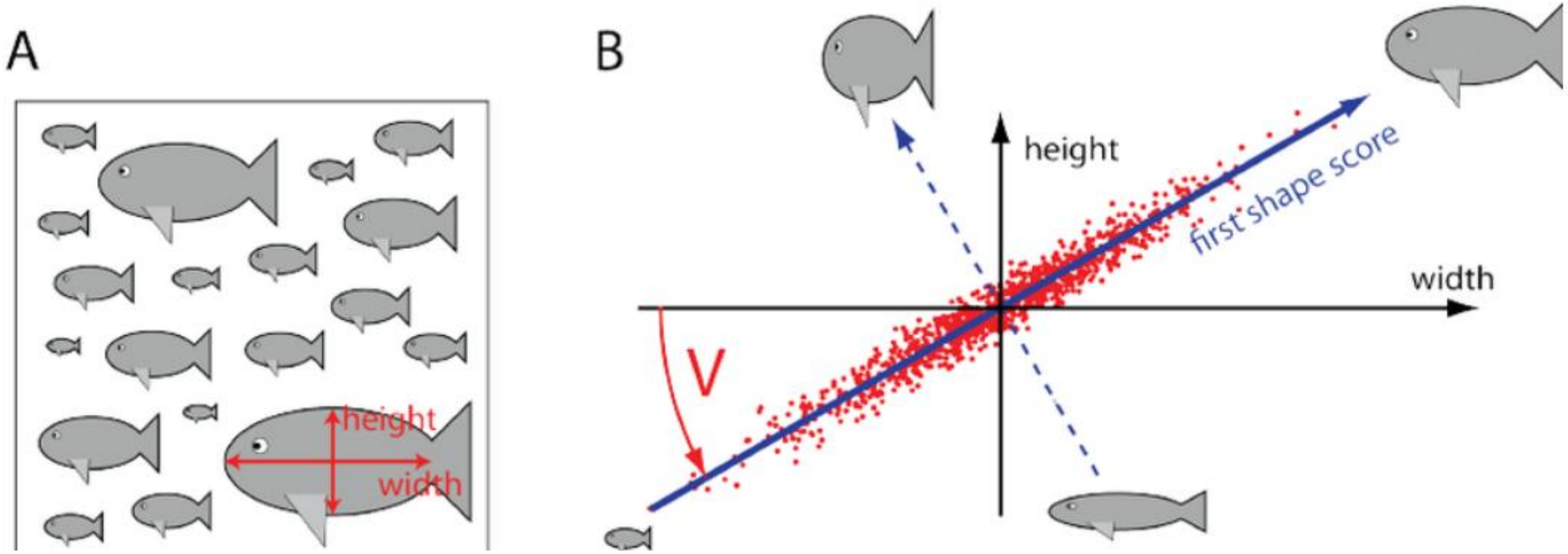
$$V = T + f(V) = T + \alpha_1^V(S_1) + \alpha_2^V(S_2) + \cdots + \alpha_n^V(S_n)$$

PCA in a nutshell

The Principle Component Analysis is a **dimensionality reduction** technique.

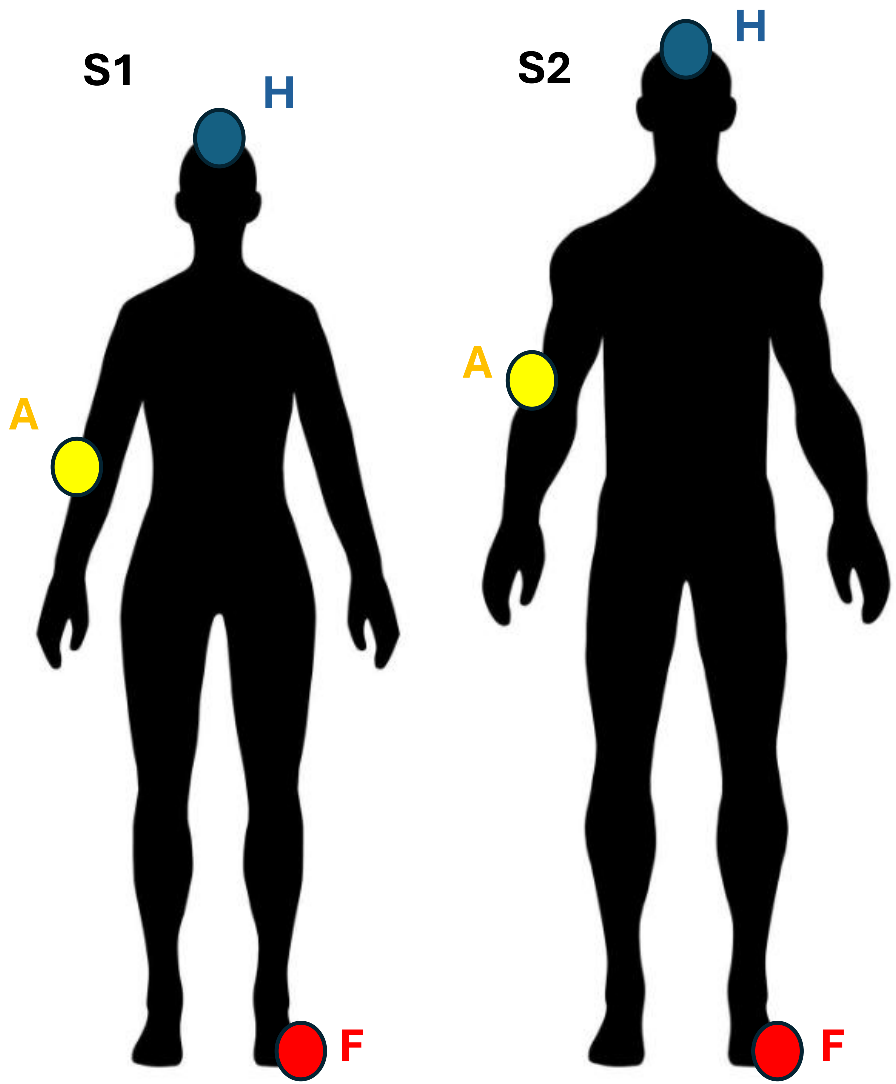
Given a dataset of samples with N correlated features, it returns K uncorrelated orthonormal vectors **ordered by explained variance**





https://www.researchgate.net/figure/llustration-of-principal-component-analysis-A-As-a-minimal-example-we-consider-a_fig1_263968032

How can we use PCA for 3D data?



S1 vertices $N \times 3$

1.1	- 2	0.1
1.5	- 1	0.4
0.1	1.1	0.7
…	…	…

S2 vertices $N \times 3$

0.5	1.2	-3.0
1.4	0.1	0.3
3.3	0	2.0
…	…	…

Vectorize →

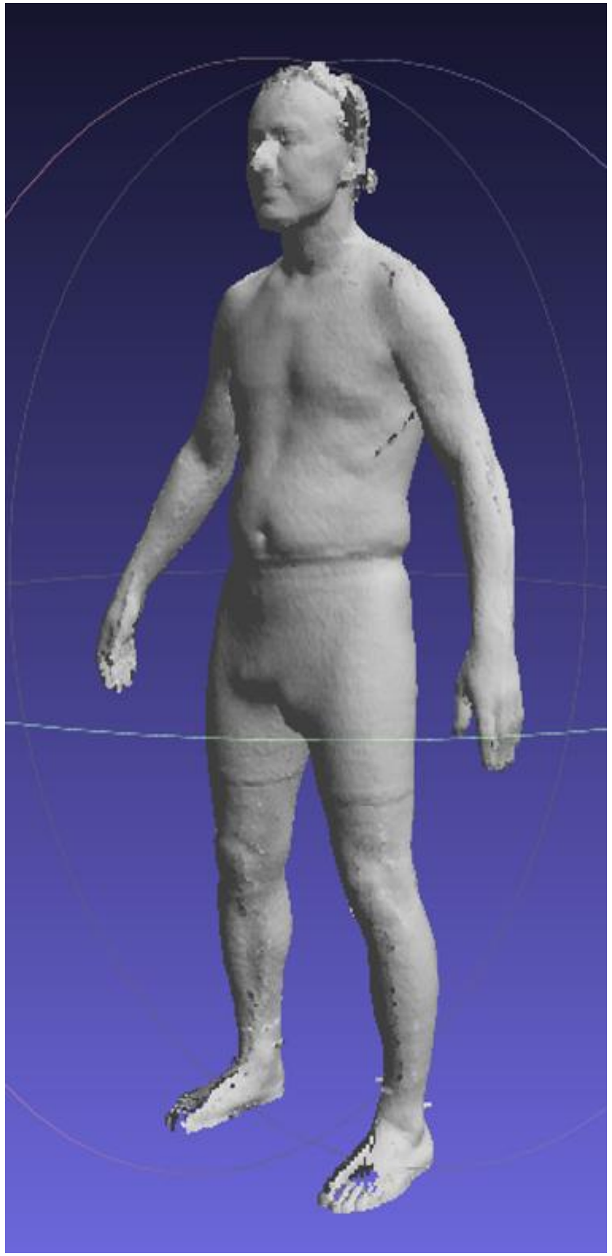
1.1	3.3	
-2	0	
0.1	2.0	
1.5	0.5	
-1	1.2	
0.4	-3.0	…
0.1	1.4	
1.1	0.1	
0.7	0.3	
…	…	

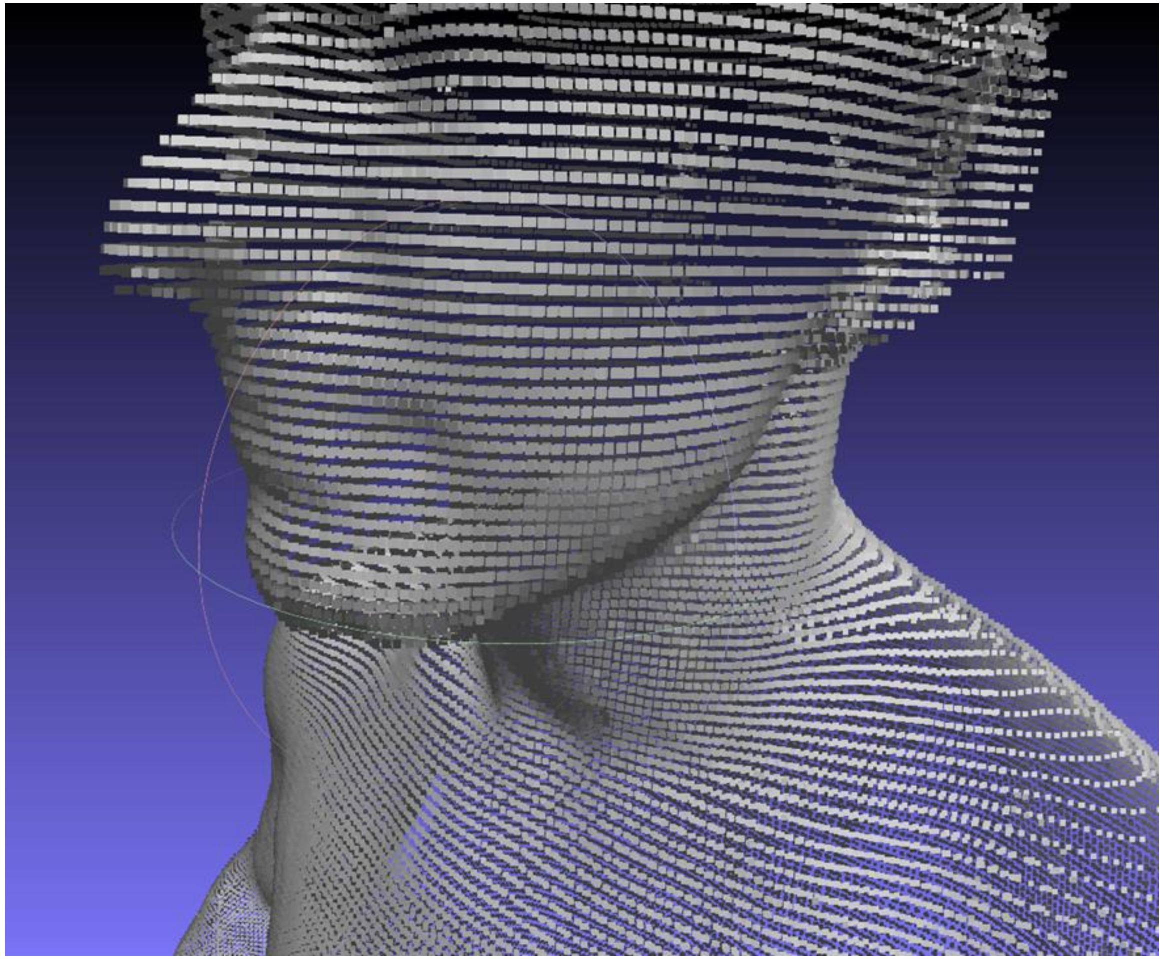
PCA →

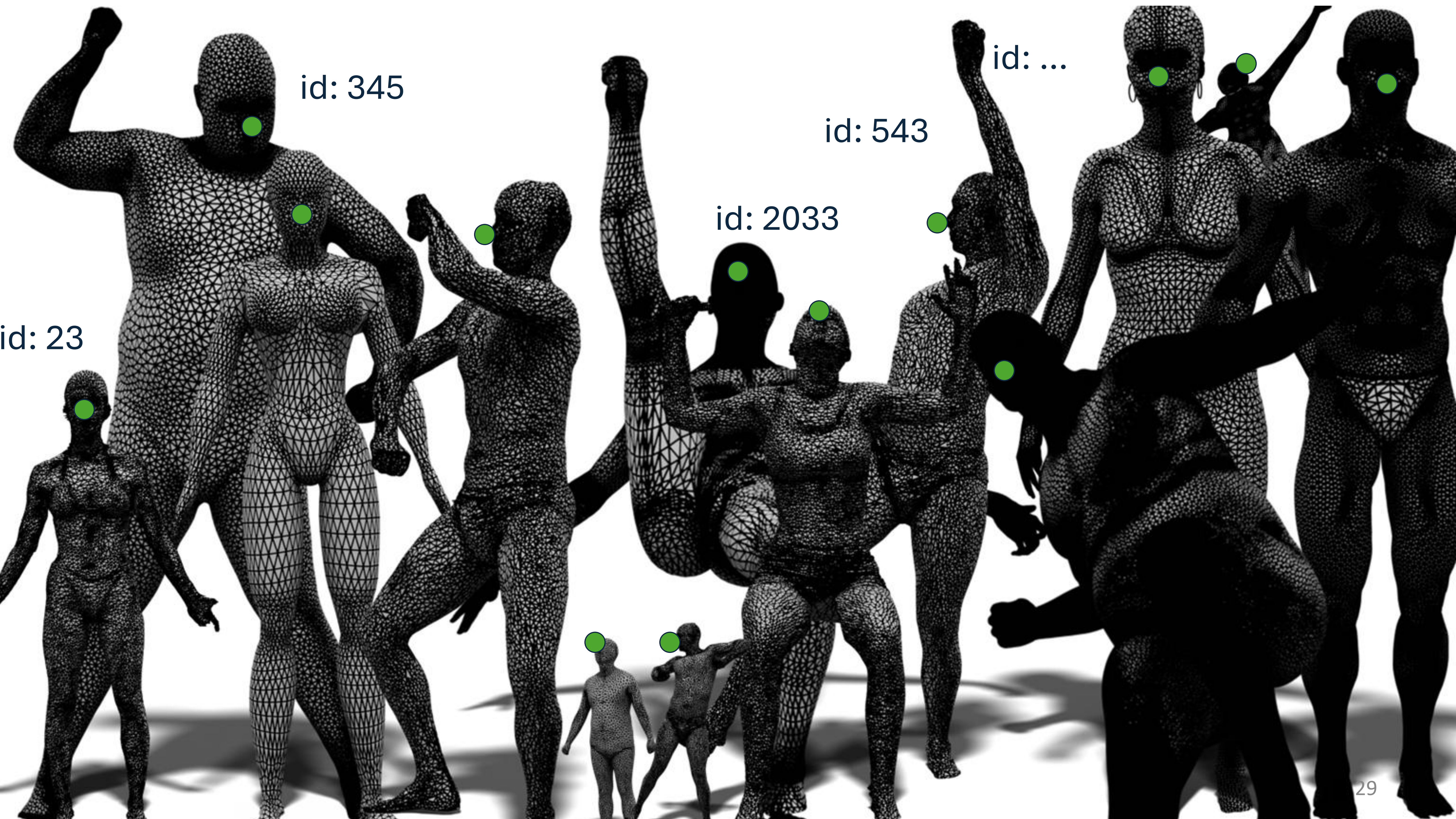
1st PCA	2nd PCA	…
P1_Hx	P2_Hx	
P1_Hy	P2_Hy	
P1_Hz	P2_Hz	
P1_Ax	P2_Ax	
P2_Ay	P2_Ay	
P3_Az	P2_Az	…
P1_Fx	P3_Fx	
P1_Fy	P3_Fy	
P1_Fz	P3_Fz	
…	…	

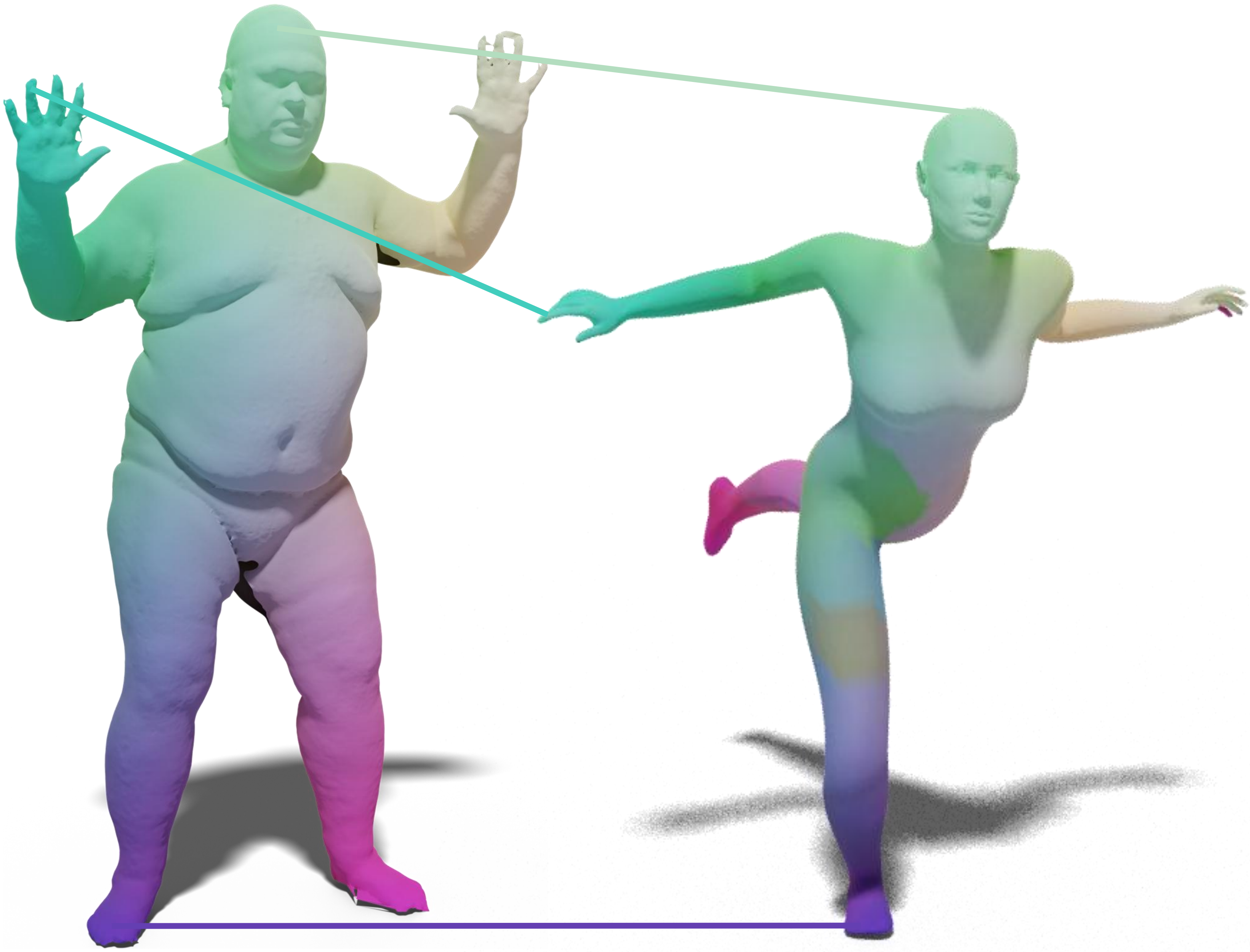
QUIZ: There are two big assumptions

Problem: Scans are just points, no semantic meaning






id: 345
id: 543
id: ...
id: 2033
id: 23



Same color
=
Same semantic meaning

How points evolve across different identities, poses, …

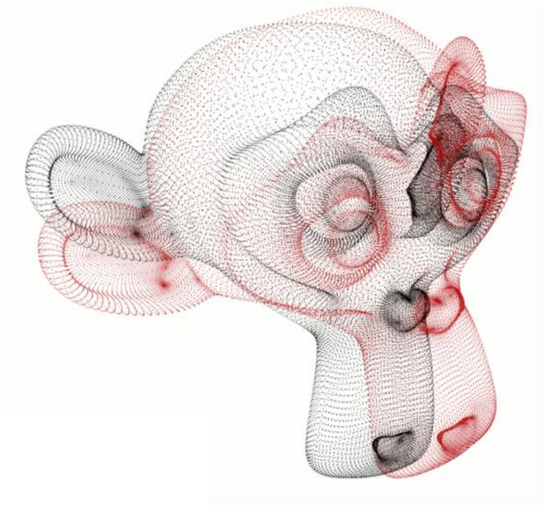
Point Clouds Registration

Template

$\mathbf{X} \in \mathbb{R}^{m \times 3}$

Target

$\mathbf{Y} \in \mathbb{R}^{n \times 3}$



Goal:

$$\|F(\mathbf{X}) - \Pi(\mathbf{Y})\|_F = 0$$

Deformation
(Registration)

$$F : \mathbf{X} \rightarrow \hat{\mathbf{X}} \in \mathbb{R}^{m \times 3}$$

Permutation
(Correspondence)

$$\Pi : \mathbf{Y} \rightarrow \mathbf{Y} \in \mathbf{Y}^m$$

Problems:

1) **Different number of points**
2) Generally, F and Π are both **unknown**
3) **Deformation cannot be free** (trivial solutions)
4) **Exact solutions do not exist** in practice

General Technique - ICP

Solving for the correspondence
(Nearest Neighbor)

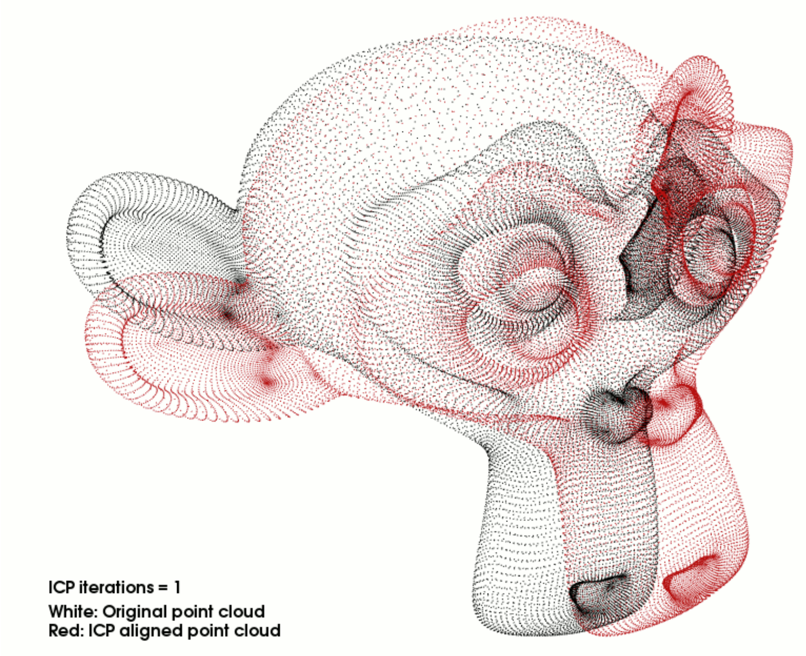
$$\widetilde{\Pi} = \arg\min_{\Pi} \|\widetilde{F}(\mathbf{X}) - \Pi(\mathbf{Y})\|_2^2$$

Iterate

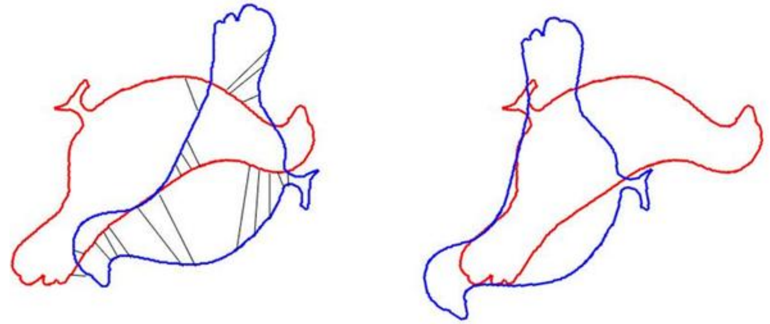
Solving for the deformation
(Optimization or closed form)

$$\widetilde{F} = \arg\min_{F} \sum_{i=1}^{m} \|F(\mathbf{x}_i) - \widetilde{\Pi}(\mathbf{Y})_i\|_2^2.$$

For the rigid case:
Rotation + Translation
(Procustes analysis)

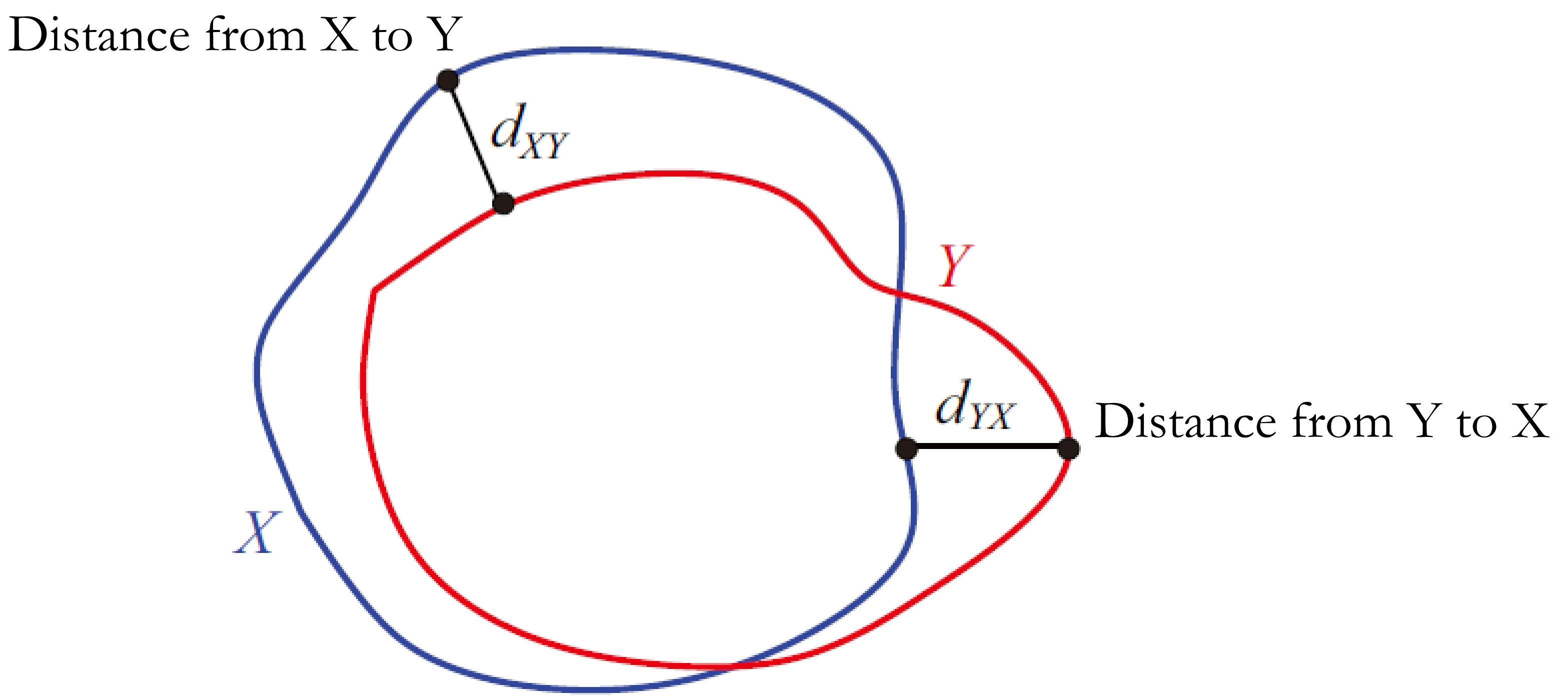


Prone to local minima



https://pcl.readthedocs.io/projects/tutorials/en/latest/interactive_icp.html

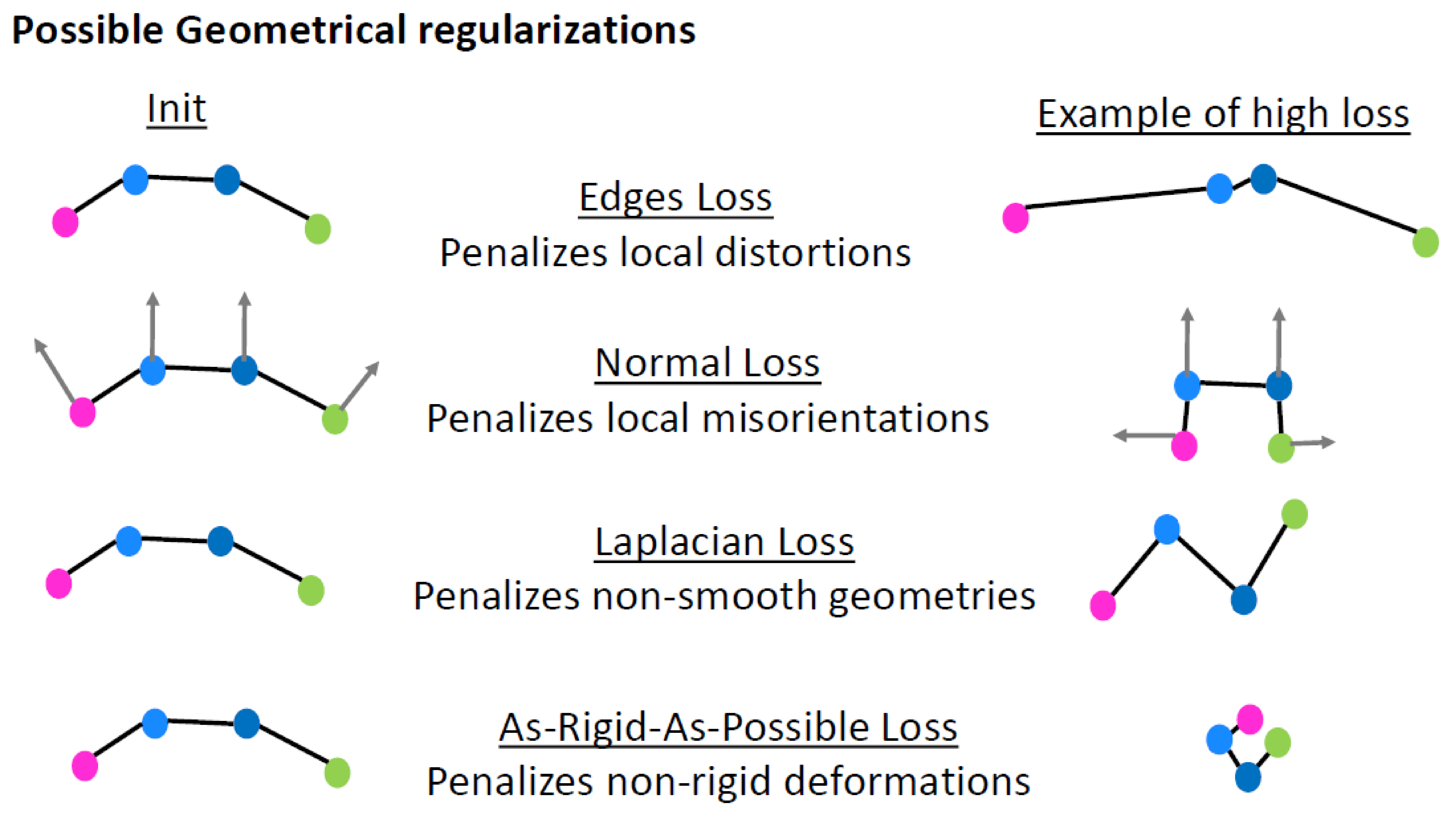
Optimize the deformation – Loss between sets



$$d_{CD}(\mathbf{X}, \mathbf{Y}) = \sum_{x \in \mathbf{X}} \min_{y \in \mathbf{Y}} \|x - y\|_2^2 + \sum_{y \in \mathbf{Y}} \min_{x \in \mathbf{X}} \|x - y\|_2^2$$

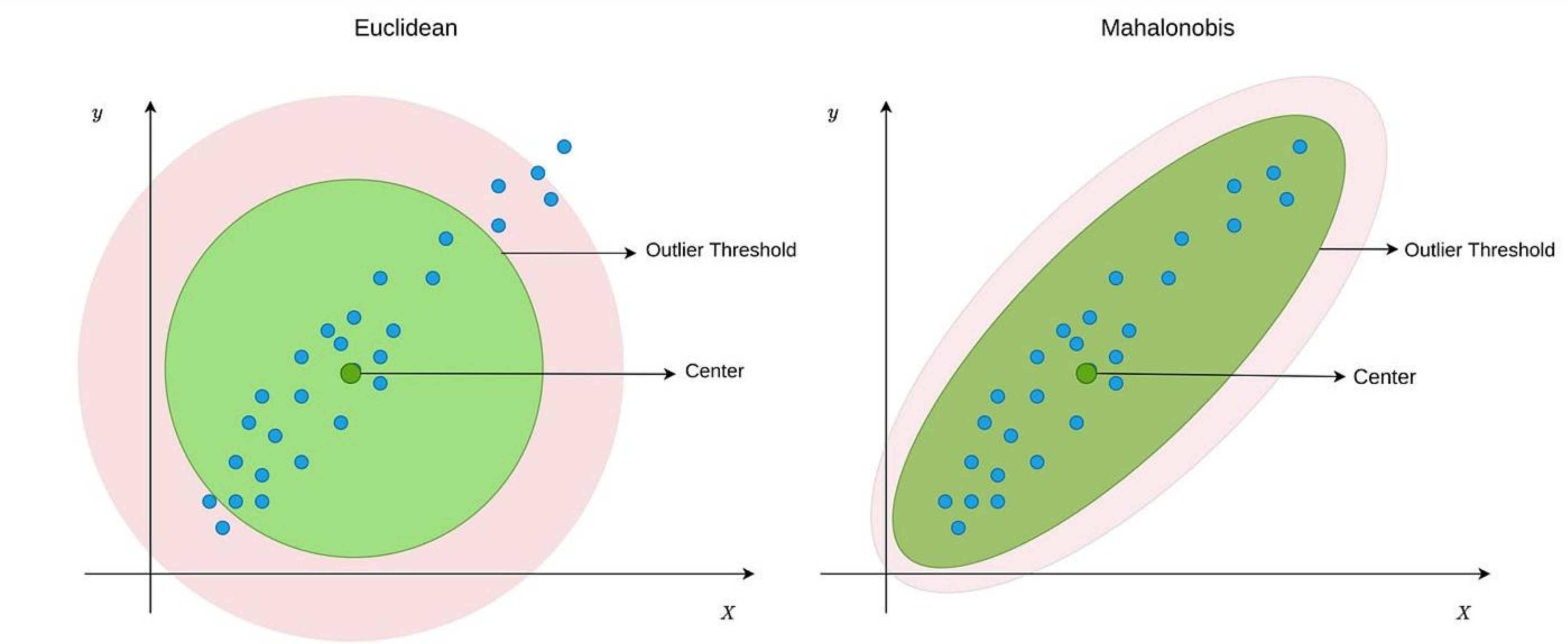
Basically, we minimize distances of a bidirectional nearest-neighbor

https://structseg2019.grand-challenge.org/Evaluation/

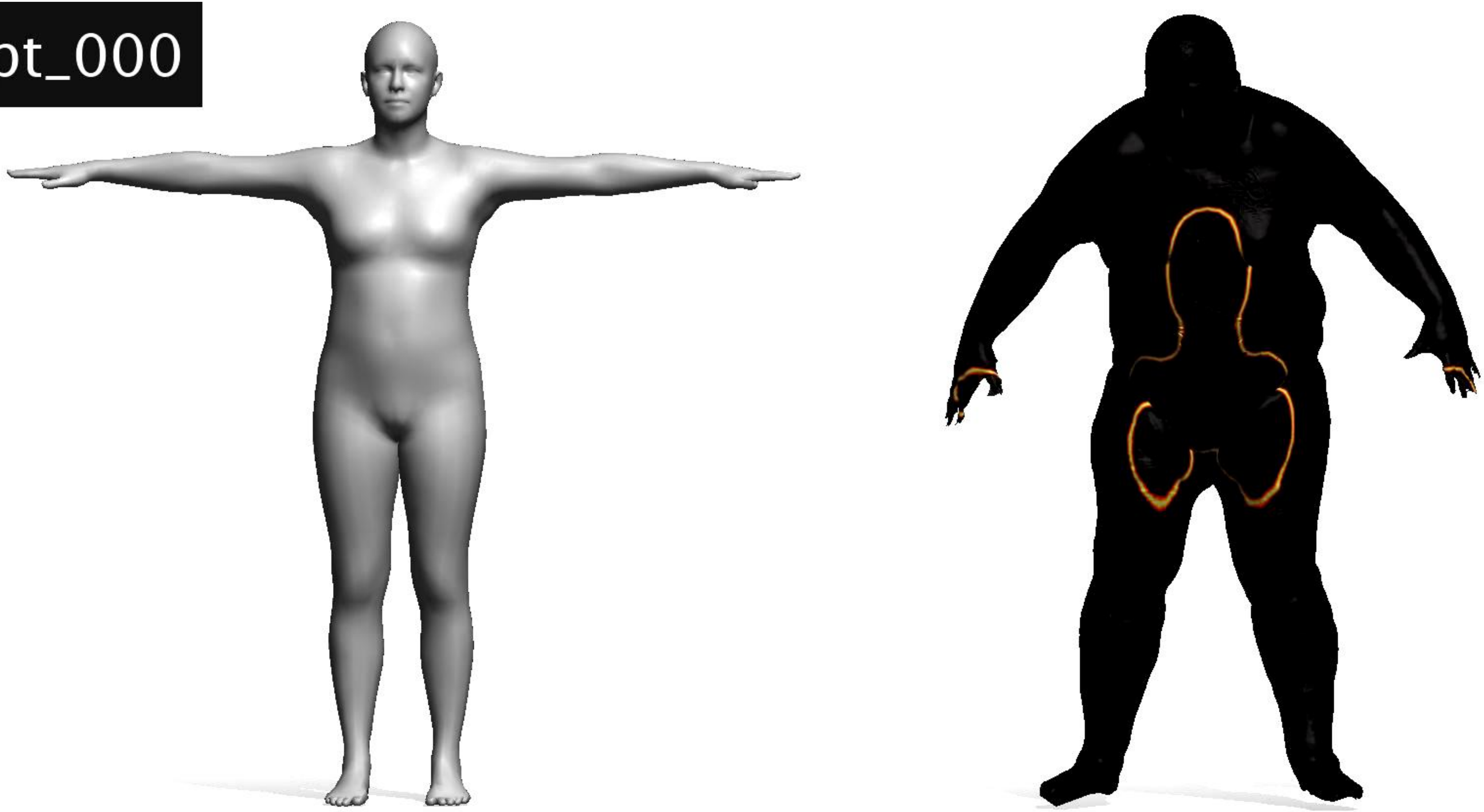

Possible Geometrical regularizations
Init
Example of high loss
Edges Loss
Penalizes local distortions
Normal Loss
Penalizes local misorientations
Laplacian Loss
Penalizes non-smooth geometries
As-Rigid-As-Possible Loss
Penalizes non-rigid deformations

Human Registration - Regularizations

Learning a Pose prior (penalizing out-of-distribution poses



https://github.com/nghorbani/human_body_prior opt_000



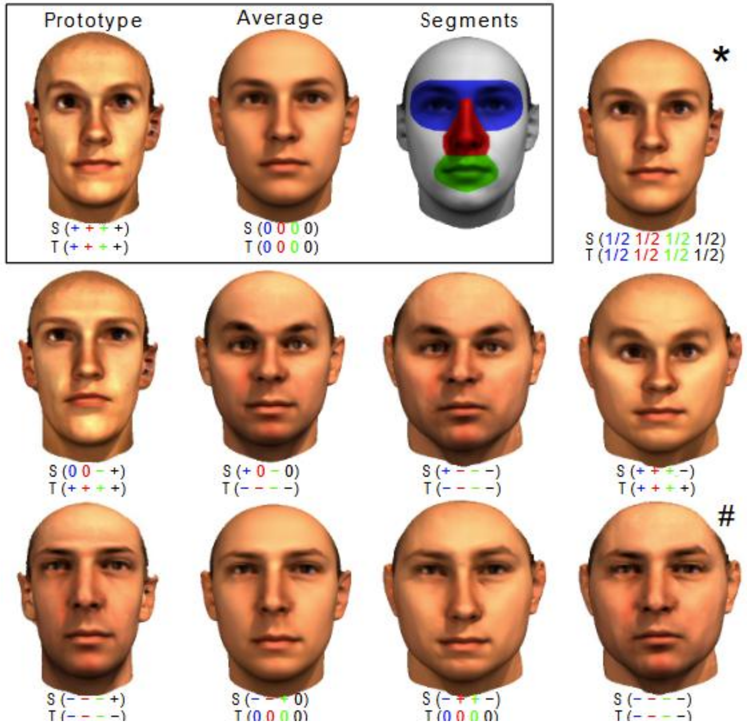
First application of PCA to humans

A Morphable Model For The Synthesis Of 3D Faces

Volker Blanz   Thomas Vetter

1999

Max-Planck-Institut für biologische Kybernetik,
Tübingen, Germany*

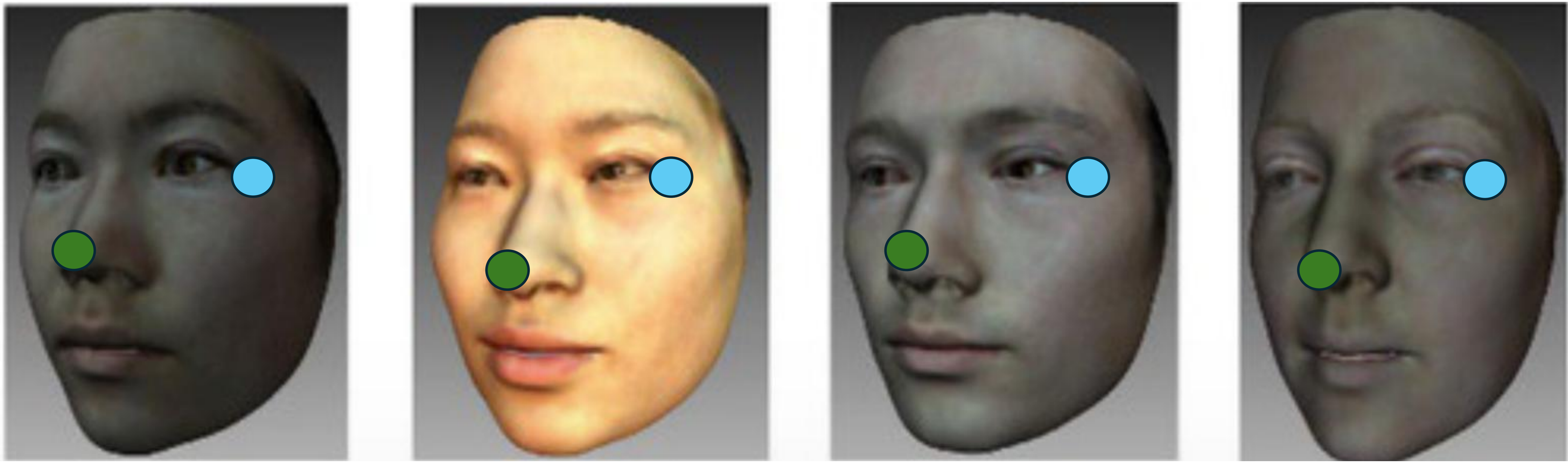


https://www.yumpu.com/en/document/read/37026456/head-face-3d-scanner-model-px-cyberware

First you collect a set of data

https://www.mdpi.com/2304-6732/9/12/924

Step 1: Obtain semantic Correspondence



Step 2: consider vertices' coordinates as variables for PCA

(i.e., 1000 vertices = 3000 features)

<u>Important: Data need to be spatially aligned</u>

<u>QUIZ: Why?</u>

https://www.mdpi.com/2304-6732/9/12/924

A (linear) morphable\parametric model for faces

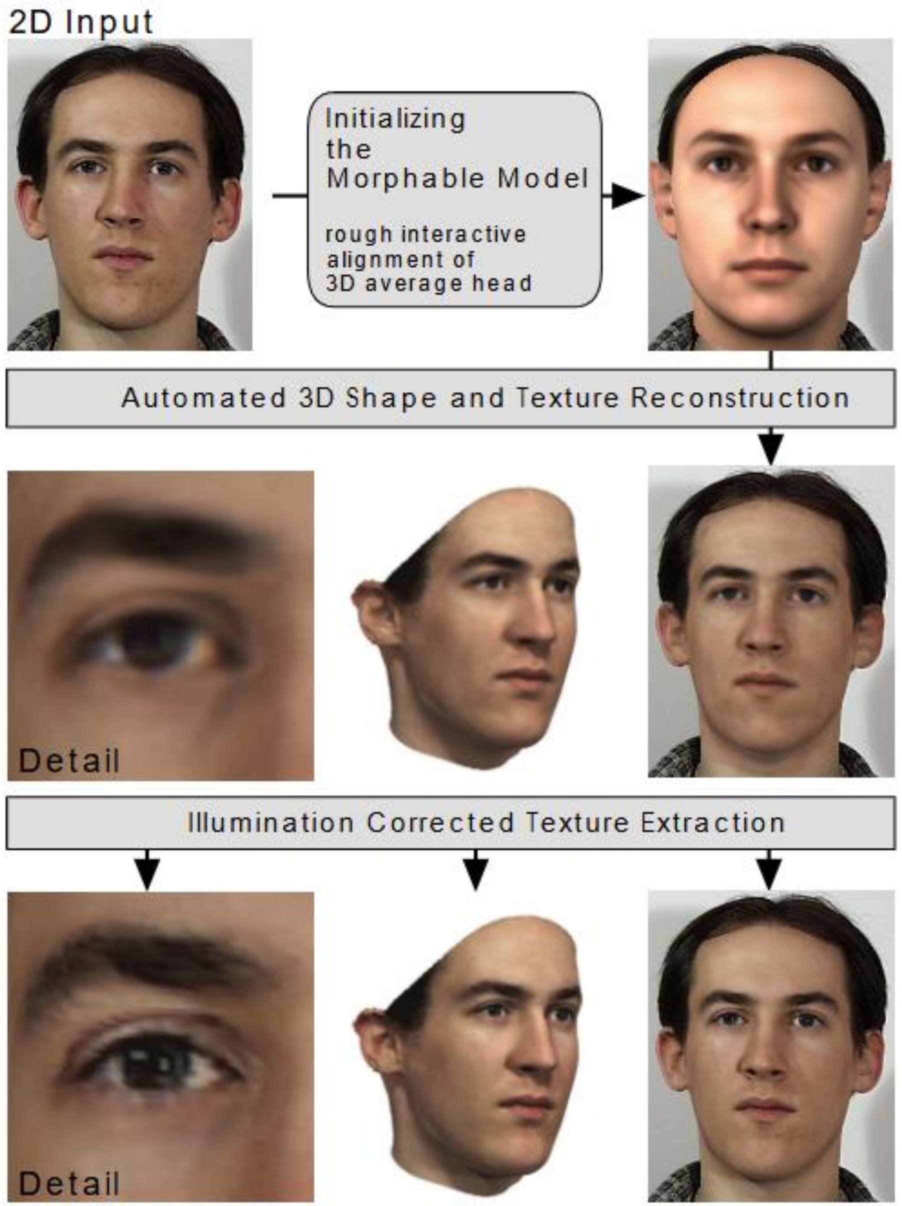
Any face can be expressed by manipulating the coefficients:

$$\mathbf{S}_{mod} = \sum_{i=1}^{m} a_i \mathbf{S}_i \,, \quad \mathbf{T}_{mod} = \sum_{i=1}^{m} b_i \mathbf{T}_i \,, \quad \sum_{i=1}^{m} a_i = \sum_{i=1}^{m} b_i = 1.$$

Trained on 200 samples,
around 70k vertices

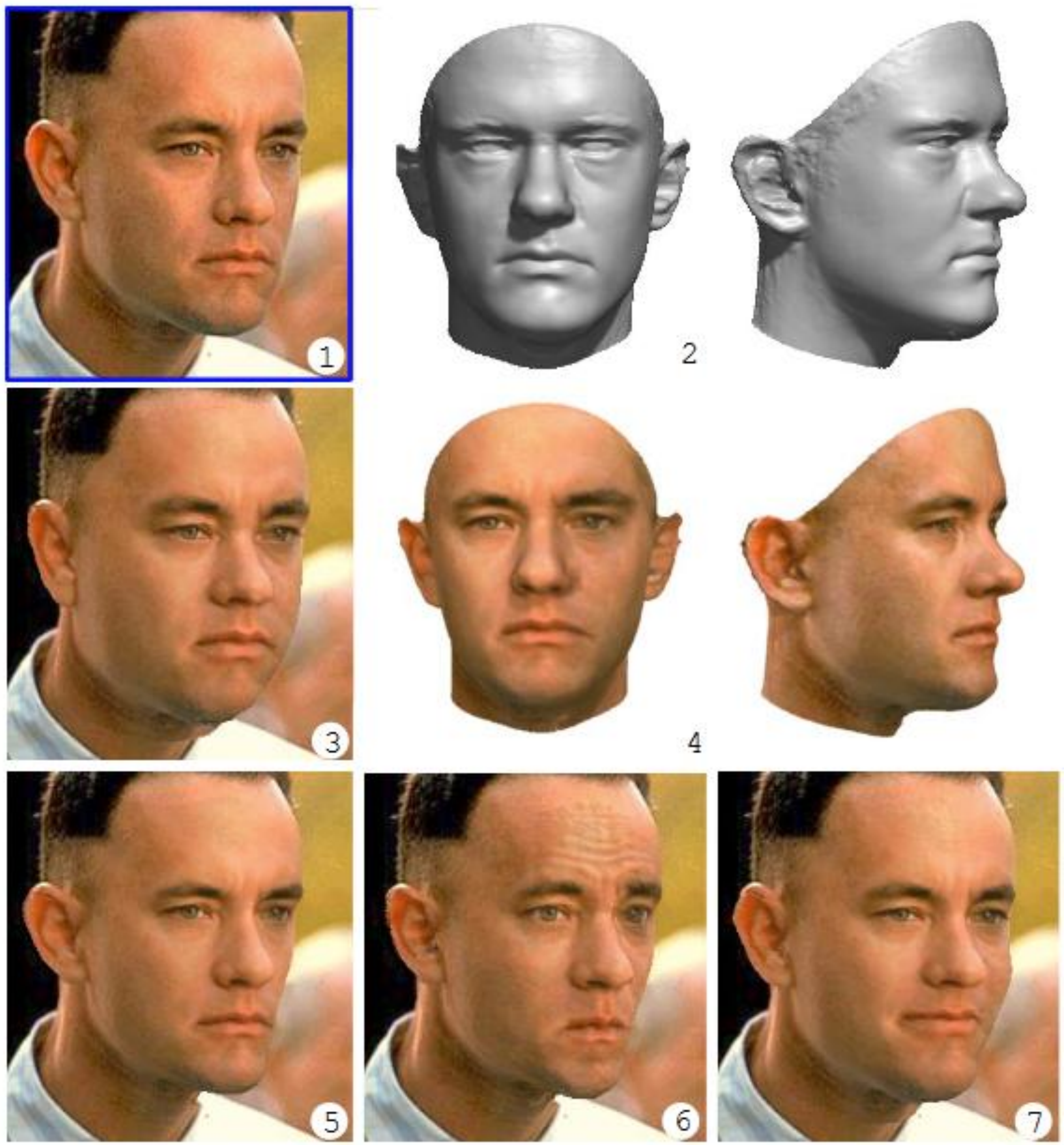
Easy to optimize

$$E_I = \sum_{x,y} \|\mathbf{I}_{input}(x,y) - \mathbf{I}_{model}(x,y)\|^2.$$



PCA Space for faces

deepfake before it became famous



CAESAR Dataset

The Caesar Project: A 3-D Surface Anthropometry Survey

K. M. Robinette, WPAFB, USA, H. Daanen, TNO, The Netherlands
E. Paquet NRC, Canada

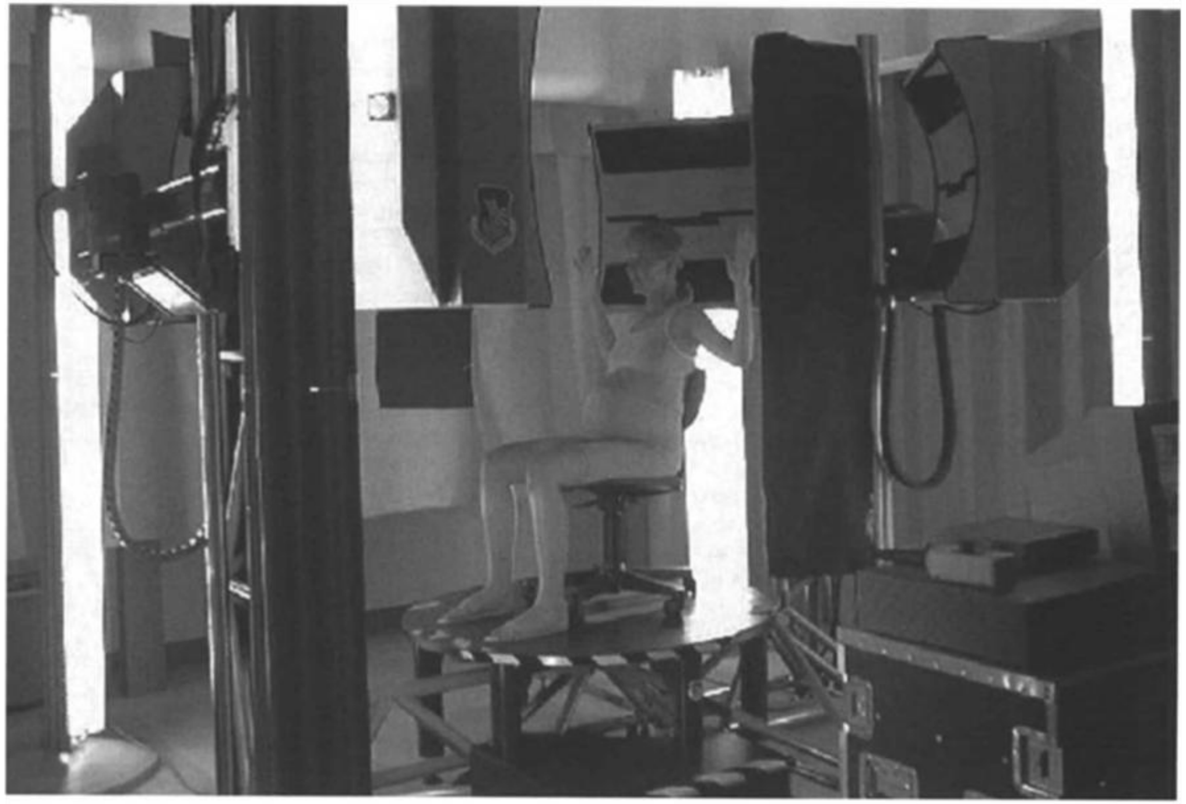


Figure 2.
WB4 Whole Body Scanner (Cyberware) in use during study of pregnant women.

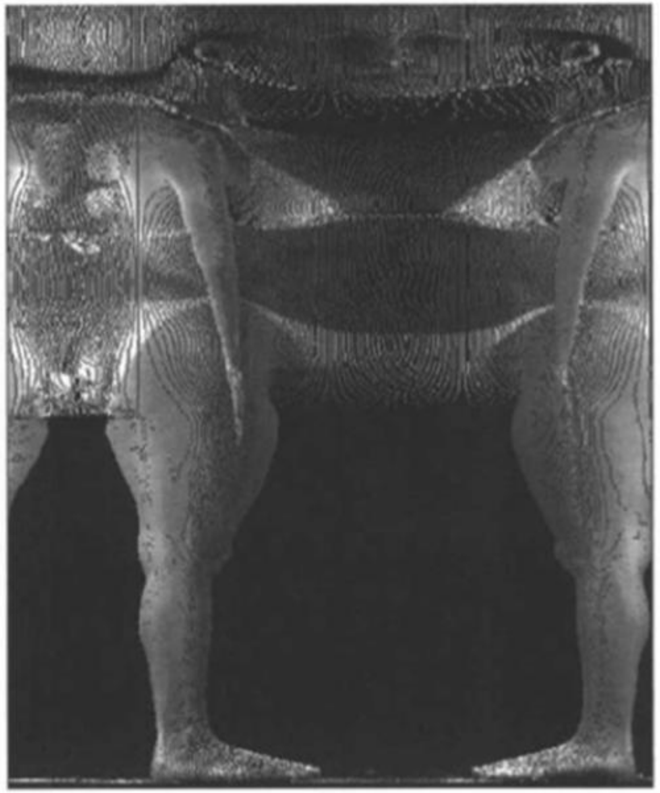
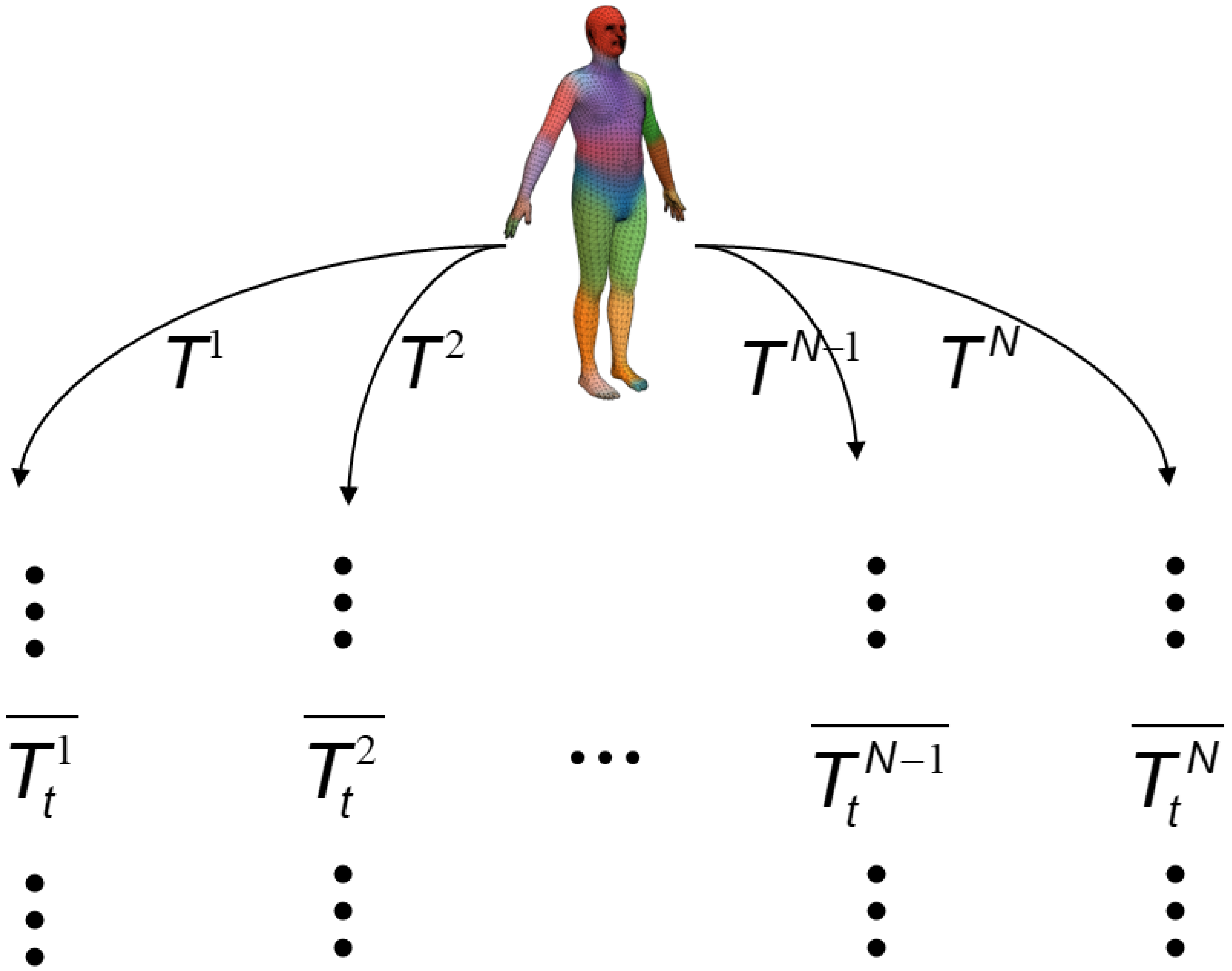


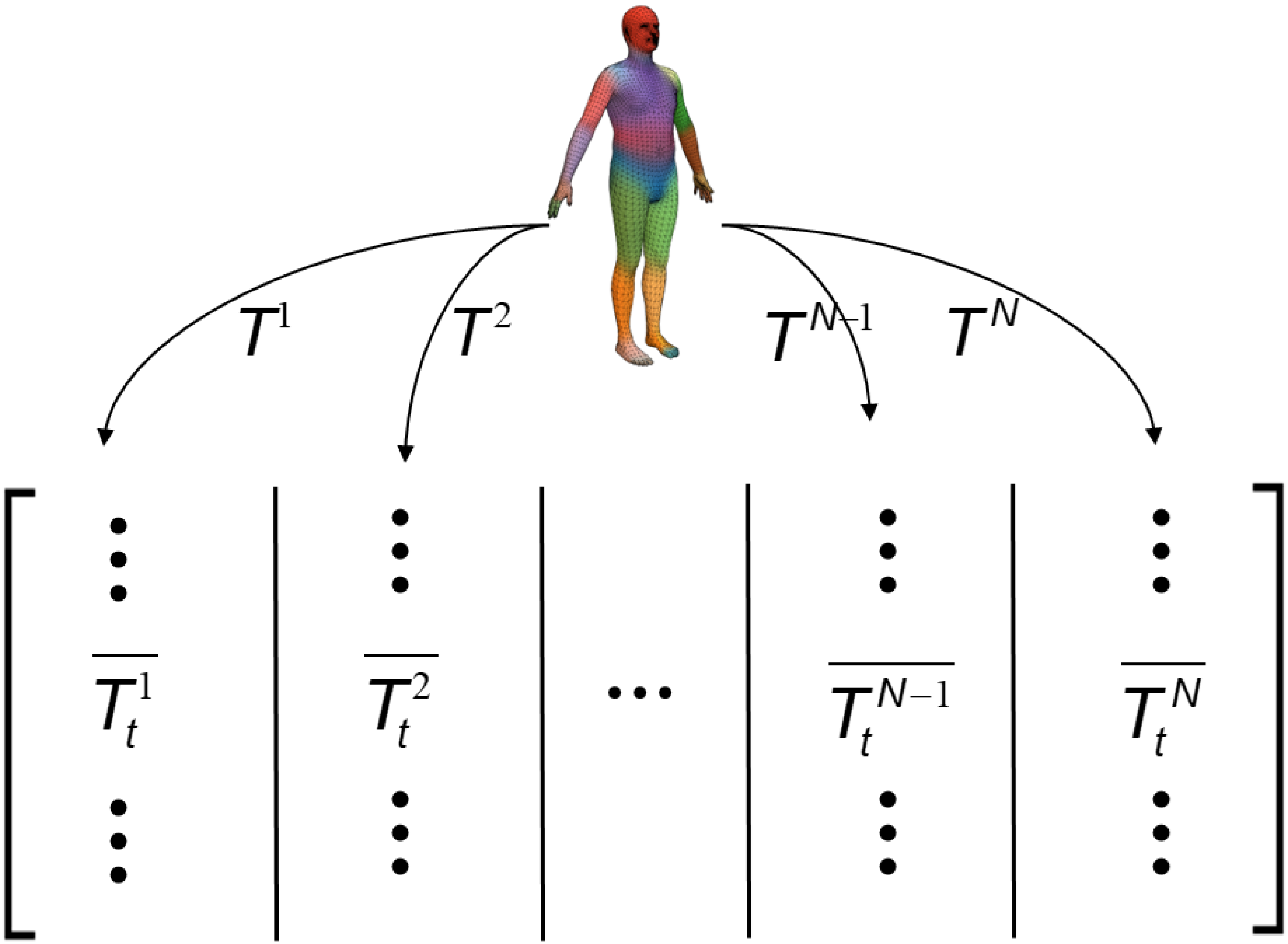
Figure 1.
Flat map of a 3D woman.

Why 3D?

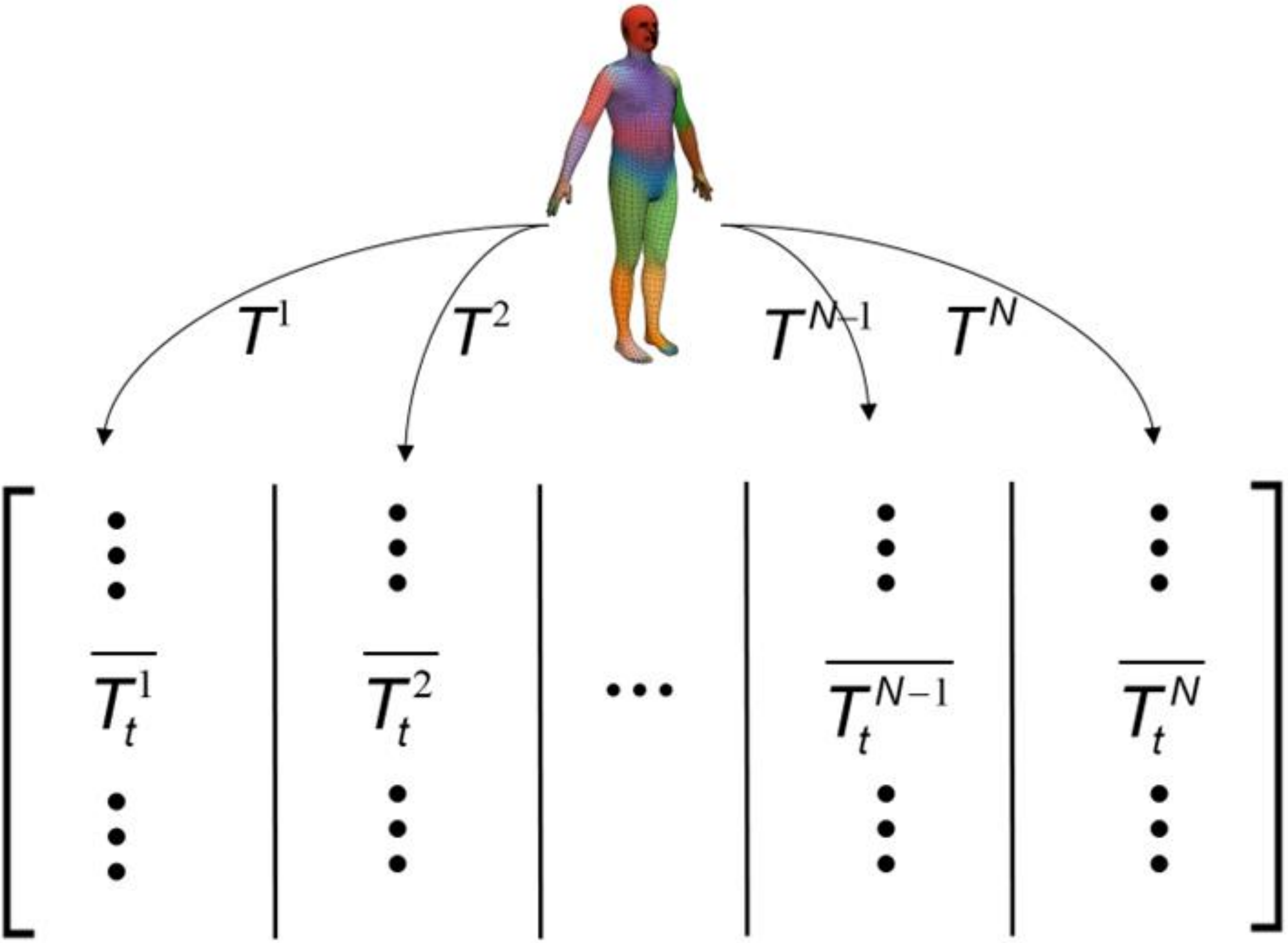
The world is not flat, and the flat-map representation of the world distorts the 3D reality. The amount of distortion inherent in the 2D view of 3D objects is readily apparent when viewing a flat map of the human body, as seen in Figure 1. Traditional body-size measuring tools are generally limited to one-dimensional information, and virtually all human models to date were built using at most 2D information. This limita-



- Vectorize the mesh vertices.
- Subtract the mean mesh.



- Make each body a column in a matrix.
- Perform PCA.



Low dimensional shape sub-space

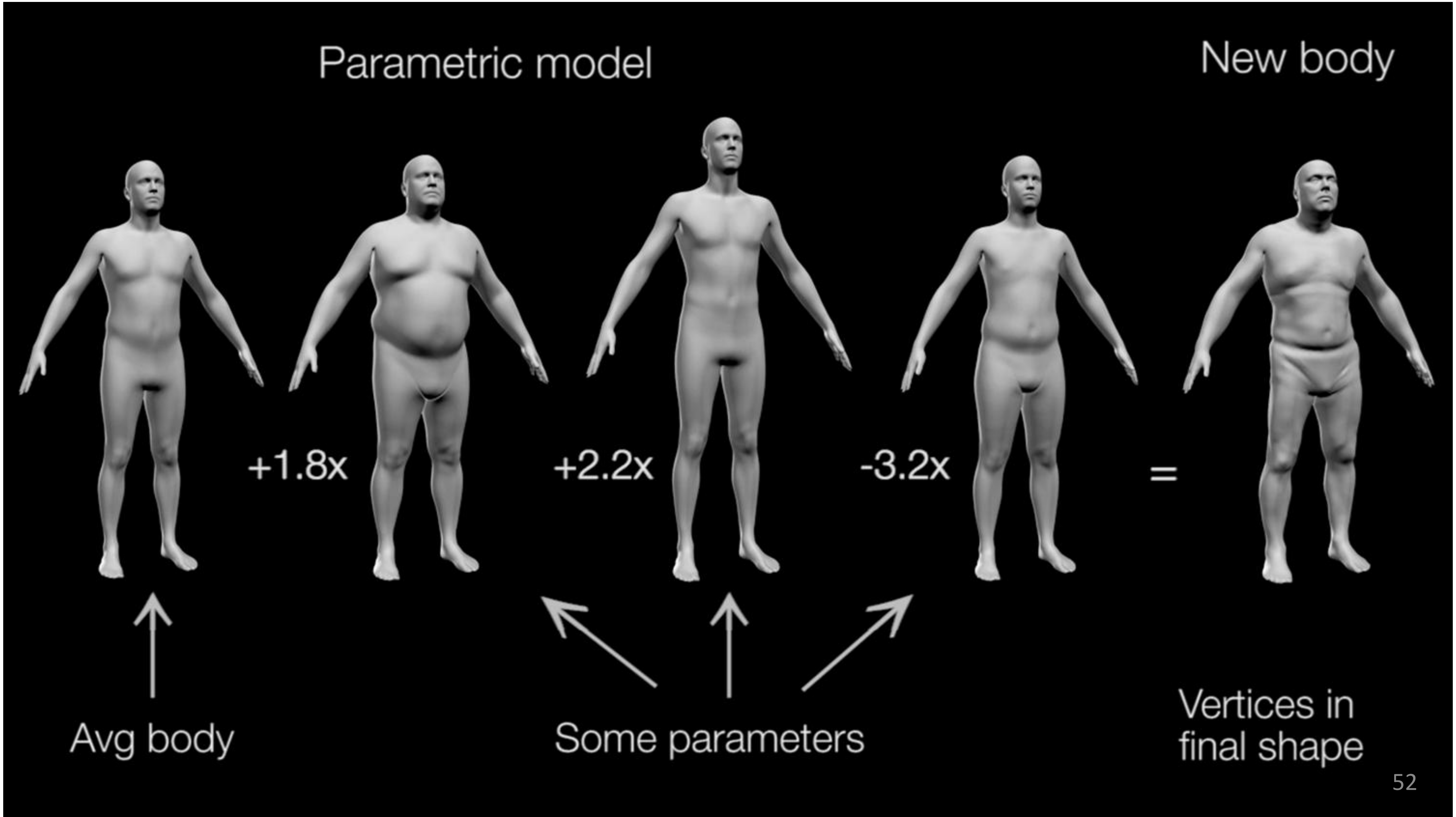
$$\begin{bmatrix} \vdots \\ \overline{T_t^*} \\ \vdots \end{bmatrix} = U\beta^* + \mu = \mathbf{S}(\boldsymbol{\beta})$$

β – shape parameters (e.g. 10-300D)

Mean male shape

Mean female shape

Identities by composition



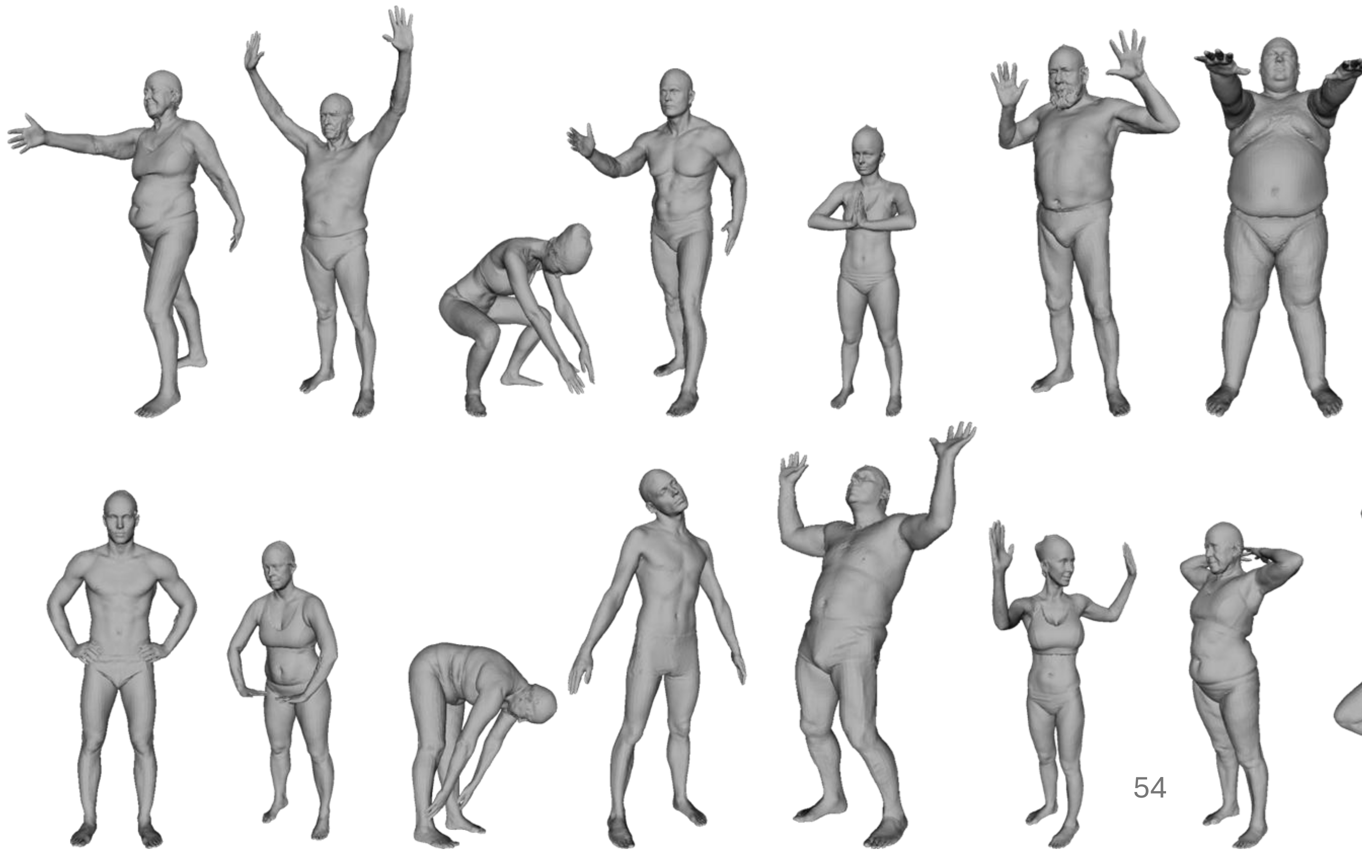
Learning a Model:

2) Shape Training

What geometry should a body representation convey?

A good body model should **look and move** like real people.

Poses



Skinning

Deforming a surface mesh according to skeleton transforms

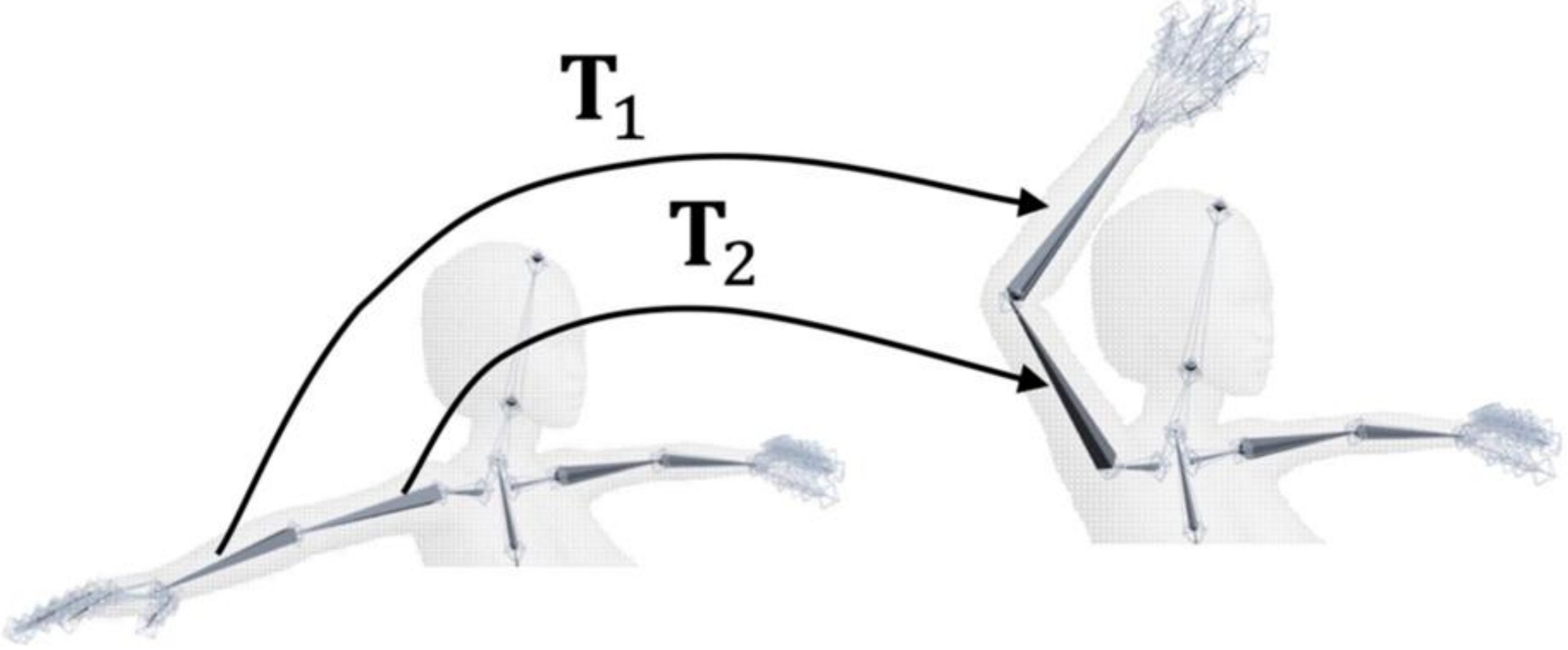
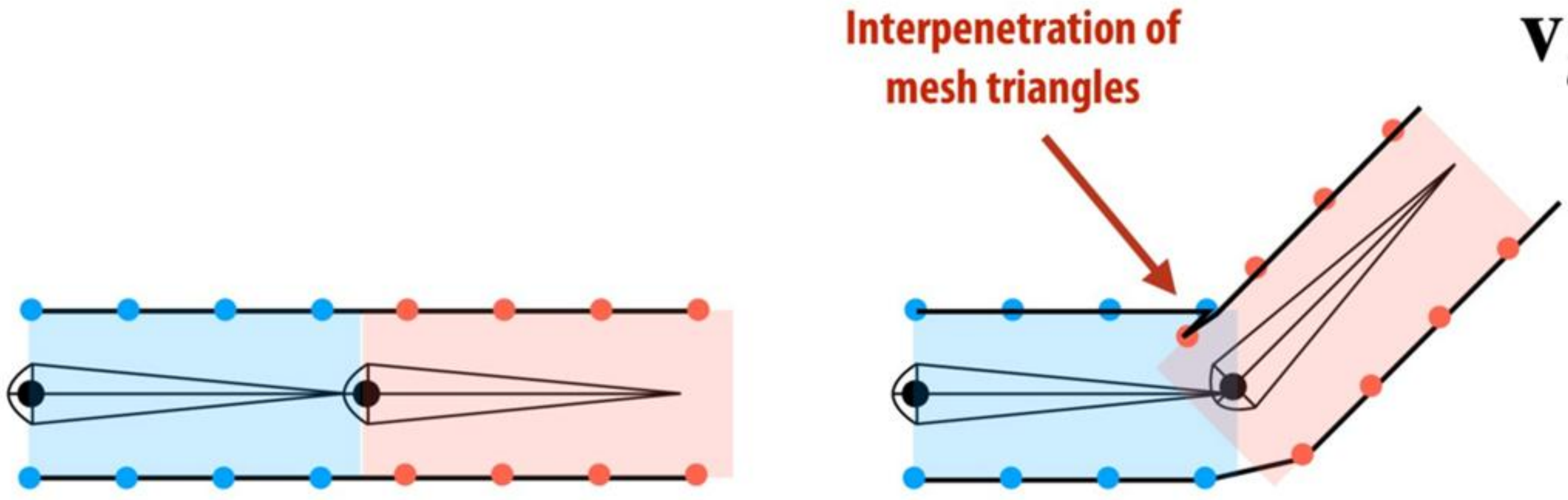


Image credit: Ladislav Kavan

Example: Rigid body Skinning

Vertices are rigidly attached to joints

$$\mathbf{v}_i' = \mathbf{T}_{b_i}\mathbf{v}_i$$



Original pose

Vertices transforms according to corresponding joint transform (notice surface interpenetration)

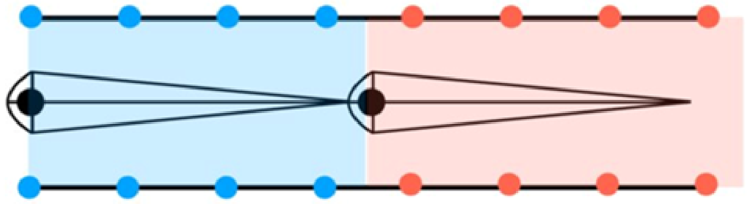
Blue verts = associated with first joint
Red verts = associated with second joint

Maya rigid bind example

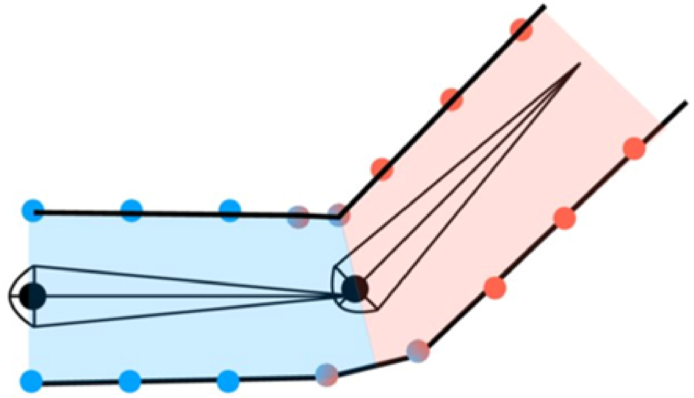
Linear Blend Skinning

Transform mesh vertices according to linear combination of transforms for nearby skeleton joint

$$\mathbf{v}_i' = \sum_{b=1}^{N} w_{ib} \mathbf{T}_b \mathbf{v}_i$$



Original pose

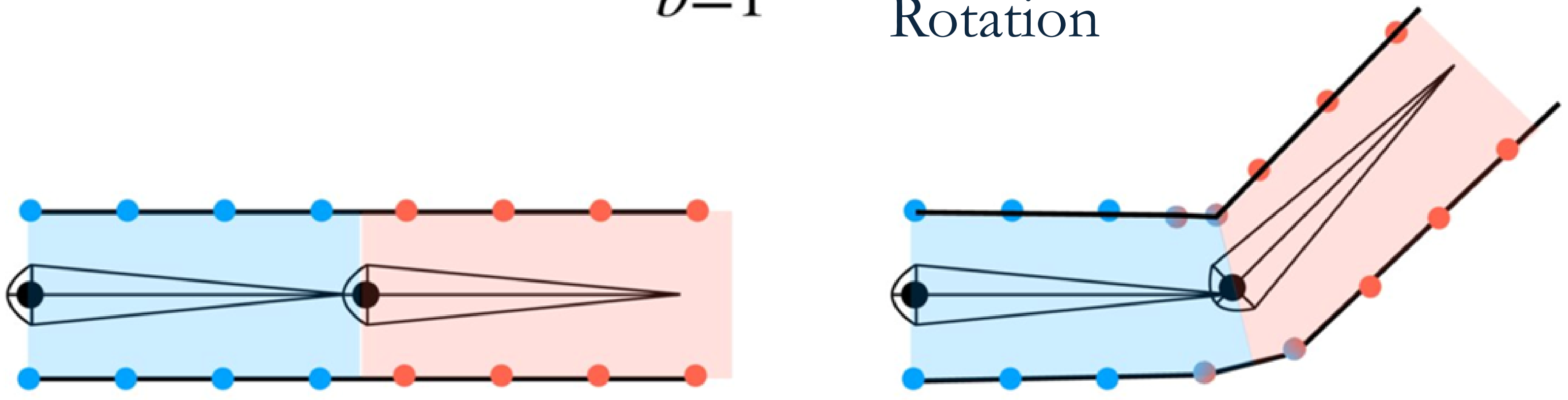


After transform

Linear Blend Skinning

Transform mesh vertices according to linear combination of transforms for nearby skeleton joint

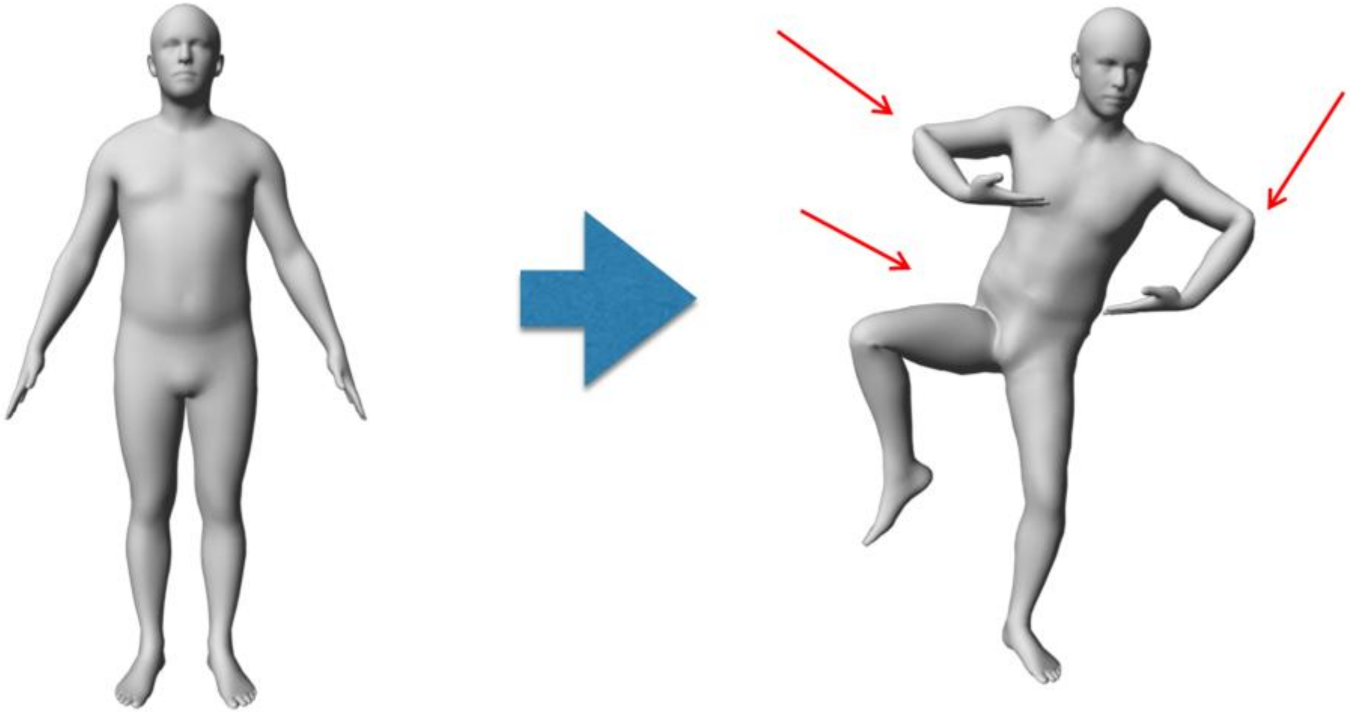
$$\mathbf{v}'_i = \sum_{b=1}^{N} \overset{\text{Blend weights}}{w_{ib}}\, \underset{\text{Rotation}}{\mathbf{T}_b}\, \mathbf{v}_i \longrightarrow \text{Template}$$



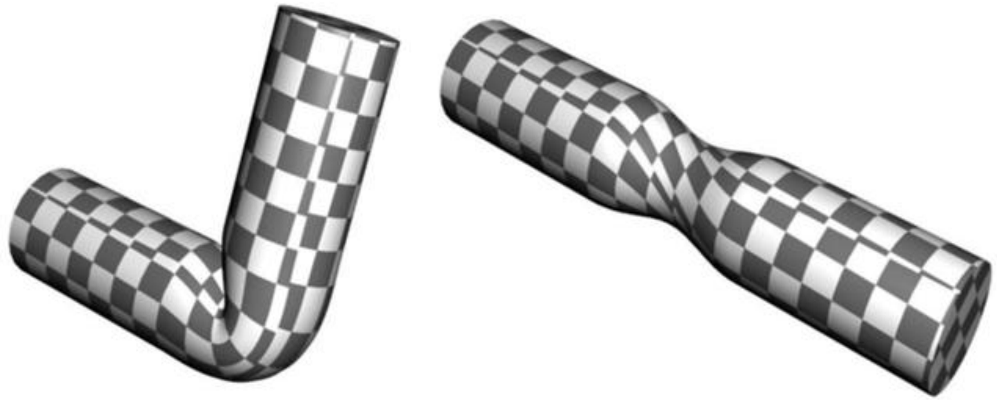
Original pose

After transform

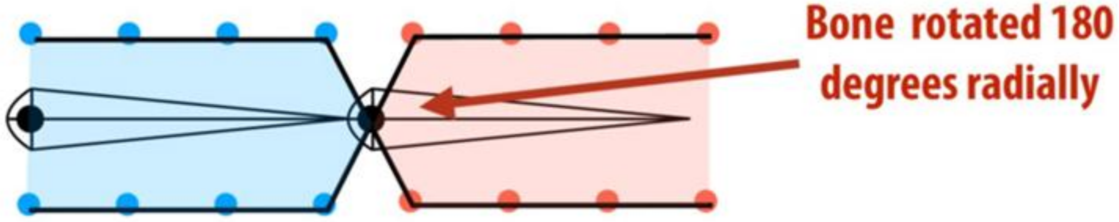
Problem: Large rotations cause artifacts



LBS loses volume



"candy wrapper effect"



Many more advanced solutions in literature: dual-quaternion skinning, joint-based or cage-based deformers, etc.

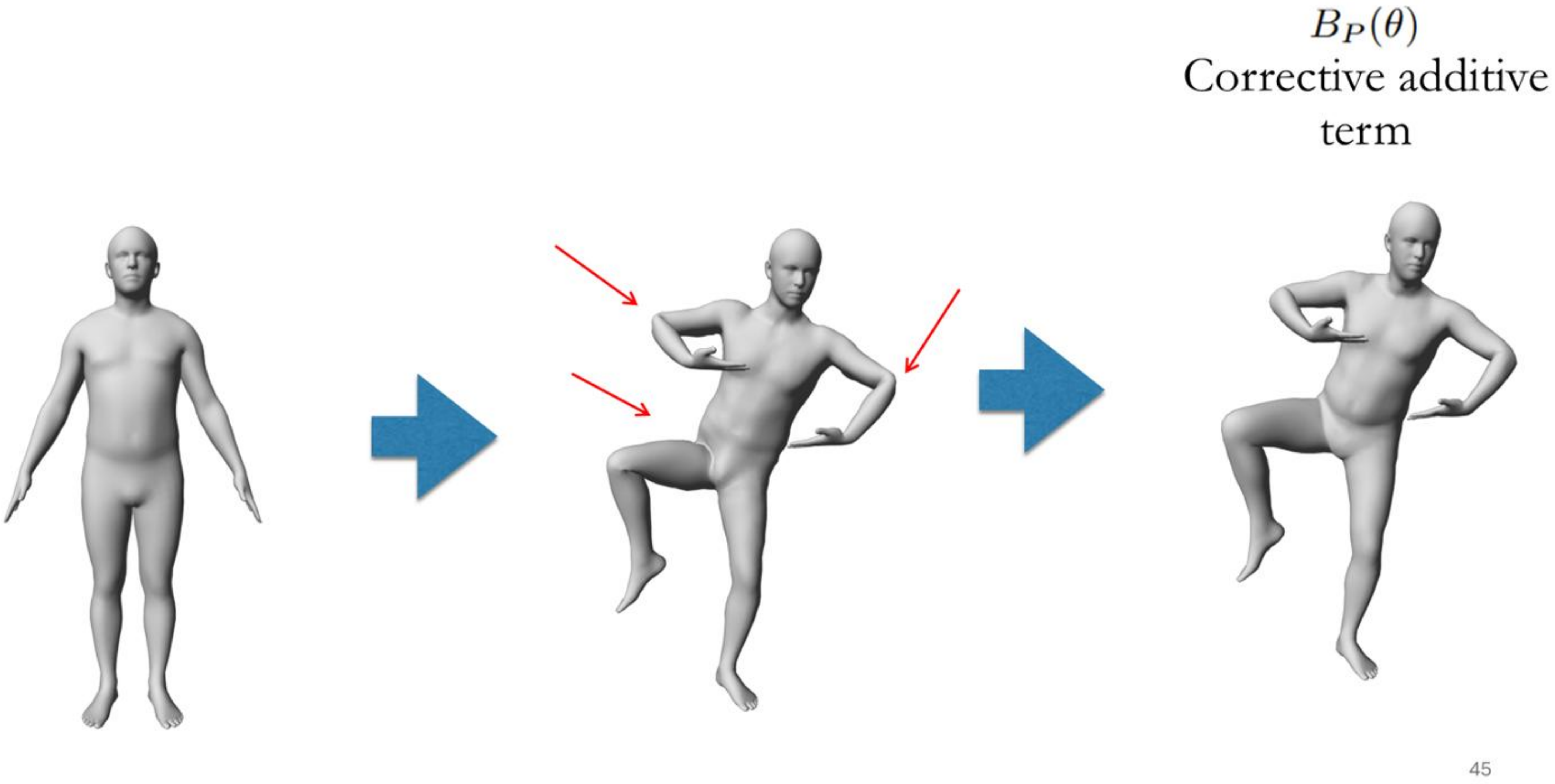
Image credit: Jacka et al.

39

Stanford CS248, Winter 2022

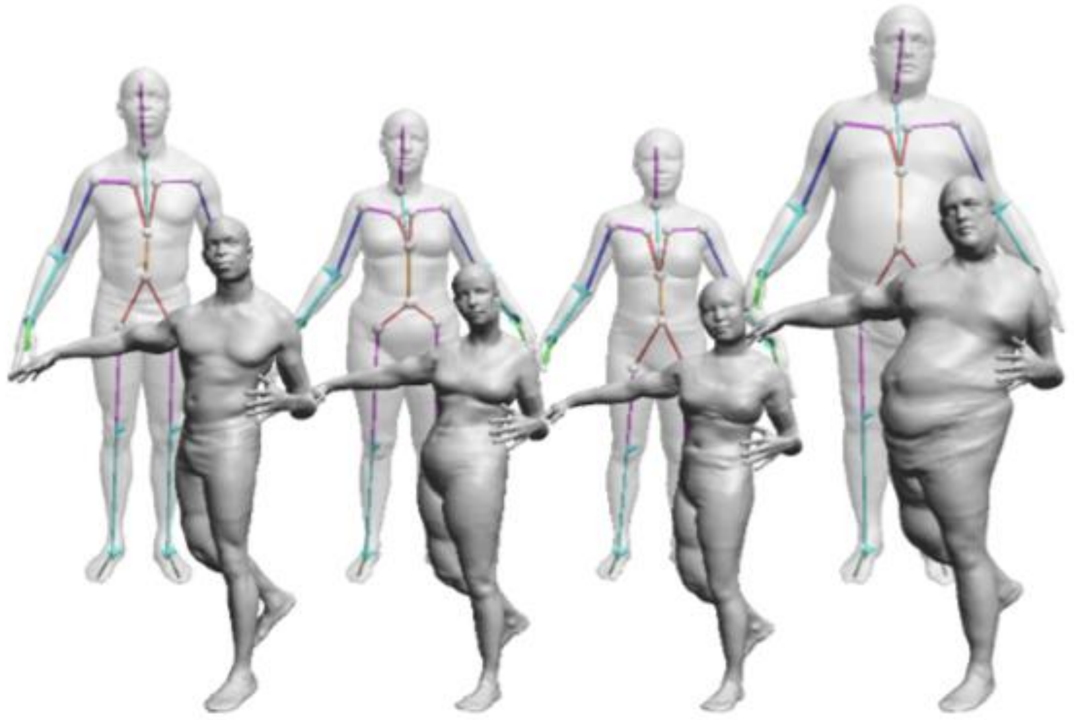
Quiz: LBS has been used by the first 3D animation movie. Do you know which one was?

Solution: pose-dependent blend shapes

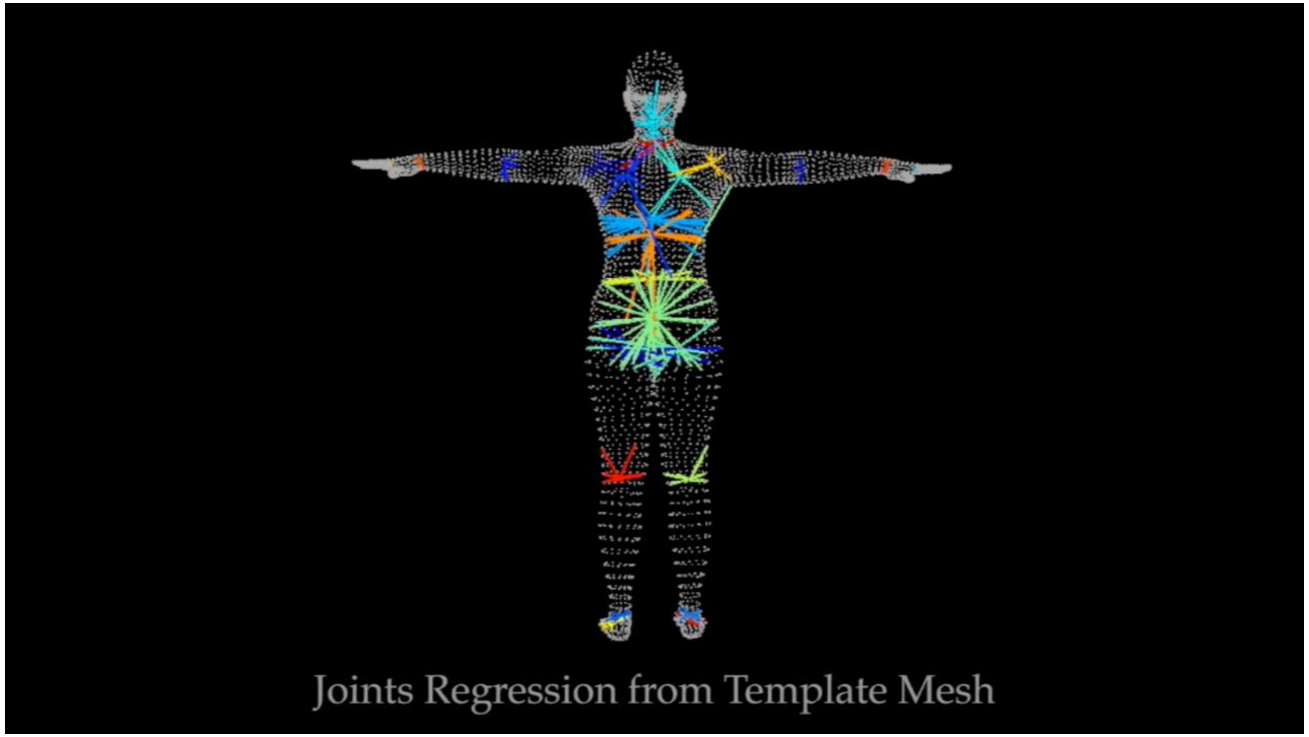


Problem: Skeleton changes for different identities

How to obtain the skeleton for different identities?



Solution:
Joints as a function of the vertices



$$\mathcal{J}(\bar{\mathbf{T}} + B_S(\vec{\beta}; \mathcal{S}))$$

Summary of SMPL model

$$M(\underbrace{\overset{\text{pose}}{\theta}, \overset{\text{shape}}{\beta}}_{\text{Input}}; \underbrace{\mathbf{T}, \mathcal{S}, \mathcal{P}, \mathcal{W}, \mathcal{J}}_{\text{Model parameters to be learned from data}})$$

Around 6M parameters

- $\mathbf{T}$ Template (average shape)
- $\mathcal{S}$ Shape blend shape matrix
- $\mathcal{P}$ Pose blend shape matrix
- $\mathcal{W}$ Blend weights matrix
- $\mathcal{J}$ Joint regressor matrix

In summary:

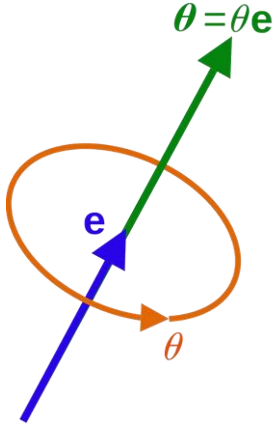
$$M(\underbrace{\overset{\text{pose}}{\vec{\theta}}, \overset{\text{shape}}{\vec{\beta}}}_{\text{Input}}; \underbrace{\mathbf{T}, \mathcal{S}, \mathcal{P}, \mathcal{W}, \mathcal{J}}_{\text{Model parameters to be learned from data}})$$

- $\mathbf{T}$ Template (average shape)
- $\mathcal{S}$ Shape blend shape matrix
- $\mathcal{P}$ Pose blend shape matrix
- $\mathcal{W}$ Blend weights matrix
- $\mathcal{J}$ Joint regressor matrix

SMPL has 6890 vertices and 24 joints. Its input are:

β **Identity**: coefficients for PCA
generally 10, can be up to 300

θ **Pose**: rotations for the joints
72 values: axis-angle rotation for every of the 24 joints



Angle: θ (in radians)

Axis: $\vec{e} = [e_x, e_y, e_z]$ (unit vector)

Angle-Axis: $\vec{\boldsymbol{\theta}} = \theta\vec{e} = [\theta e_x, \theta e_y, \theta e_z]$, *where* $|\vec{\boldsymbol{\theta}}| = \theta$

https://medium.com/data-science/3d-rotations-intuitions-and-limitations-f3ae2122fe23

What is a good representation model for bodies?

SMPL Principles:

- **Low-dimensional** (e.g., editable acting on a few parameters),

- **Differentiable** (e.g., optimizable for data fitting)

- **Compatible with standard graphics tools** (e.g. animation)

Given the identity and pose parameters, vertices are update by this equation:

Skinning Weights

T-pose Coordinates

Posed vertex

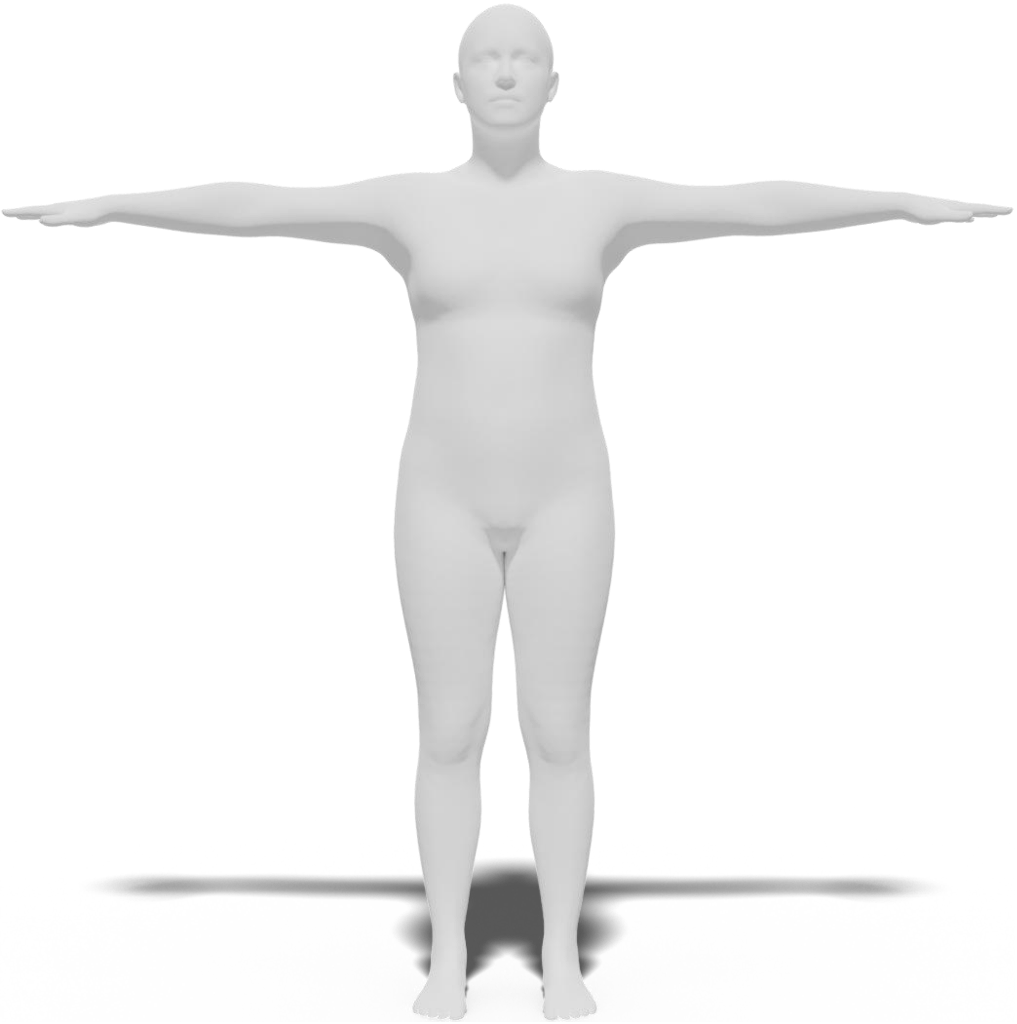
Pose correction

$$\bar{\mathbf{t}}'_i = \sum_{k=1}^{K} w_{k,i} G'_k(\vec{\theta}, J(\vec{\beta}))(\bar{\mathbf{t}}_i + \mathbf{b}_{S,i}(\vec{\beta}) + \mathbf{b}_{P,i}(\vec{\theta}))$$

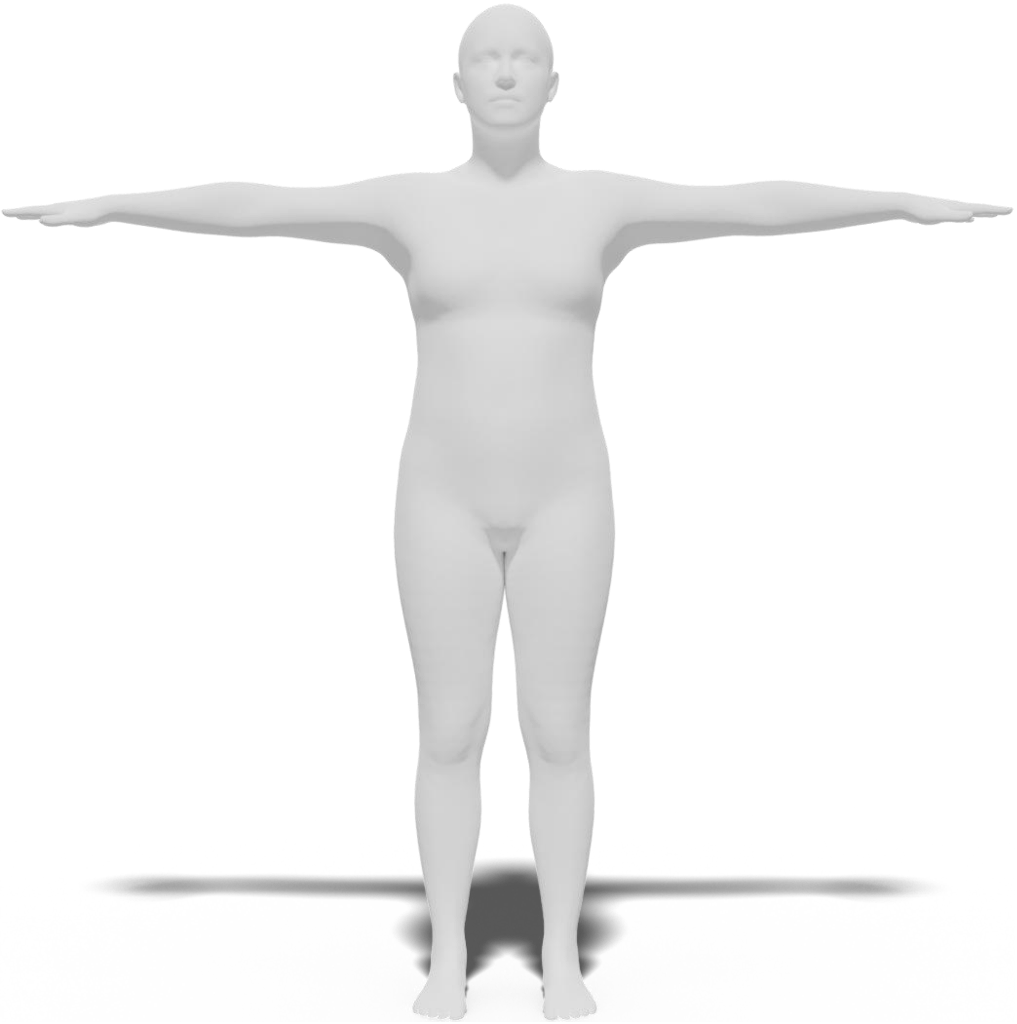
Joint Rotation

Subject Identity

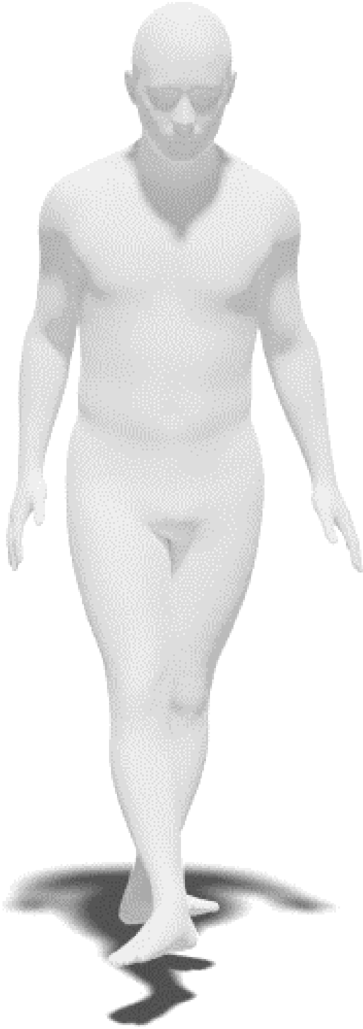
$$M(\mathbf{0}, \mathbf{I}) =$$



$M(\mathbf{0}, \mathbf{I}) =$



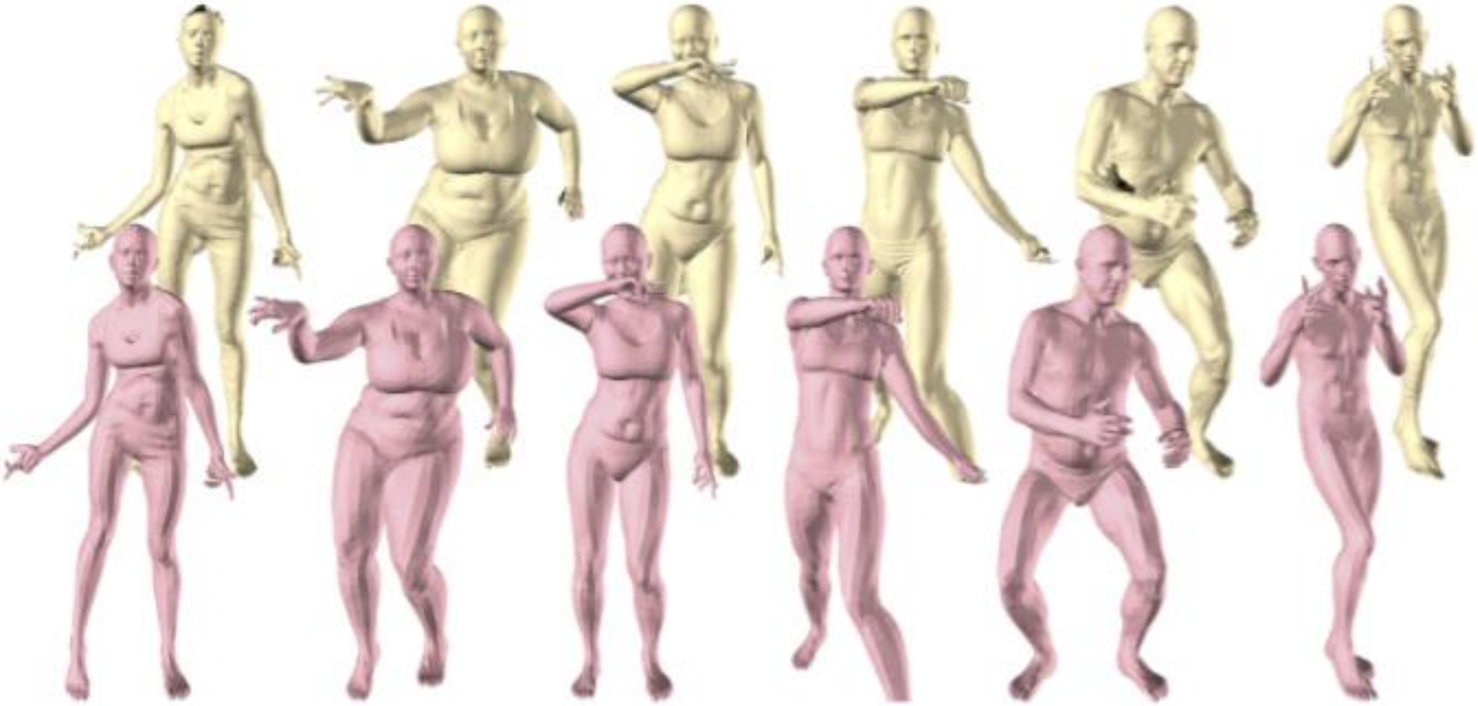
$M(\beta, \theta) =$



Follow-up: SMPL+H(ands)

Embodied Hands: Modeling and Capturing Hands and Bodies Together

JAVIER ROMERO[*,†], Body Labs Inc.
DIMITRIOS TZIONAS[*], Max Planck Institute for Intelligent Systems
MICHAEL J. BLACK, Max Planck Institute for Intelligent Systems



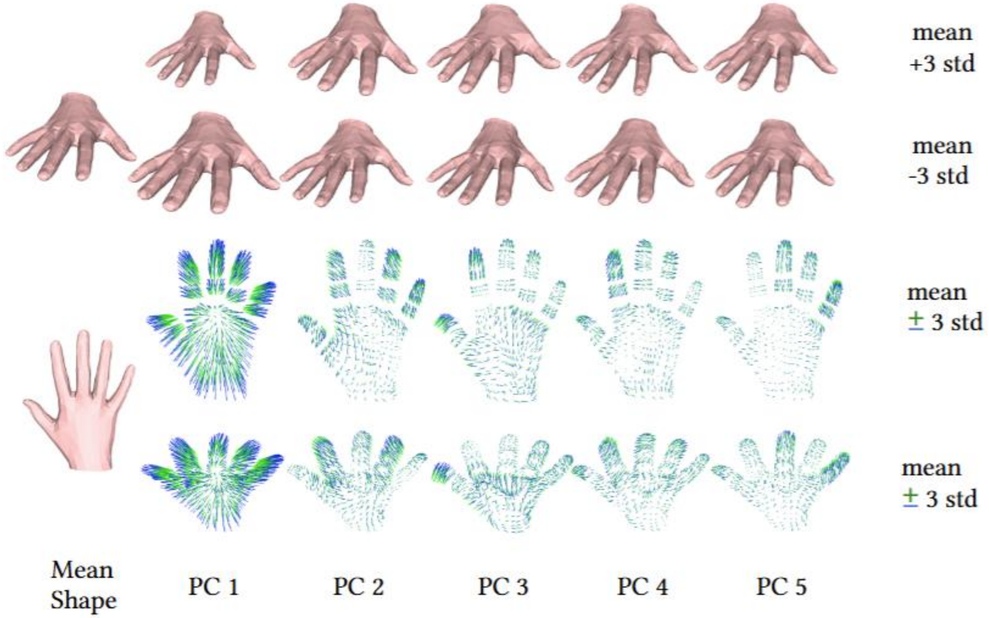


Fig. 4. PCA shape space. Each column depicts the effect of one of the first five *principal components* (PCs) of the learned hand shape space. The effect of each PC is shown by adding ±3 standard deviations (std) to the mean shape (left-most image), as indicated. See the Supplemental Video.

MANO hand model

- Each hand is defined by: $\theta^h, \beta^h, \mathcal{T}, \mathcal{W}$
  - $\theta^h \in \mathbb{R}^{48}$: 15 joints + one global rotation.
  - $\beta^h \in \mathbb{R}^{10}$: shape PCA components.
  - $\mathcal{T}$: hand template, part of SMPL template.
- Hand articulations are highly restricted: the pose is overparametrized.
  - Learn a hand pose PCA space: $f: \mathbb{R}^{45} \rightarrow \mathbb{R}^6$

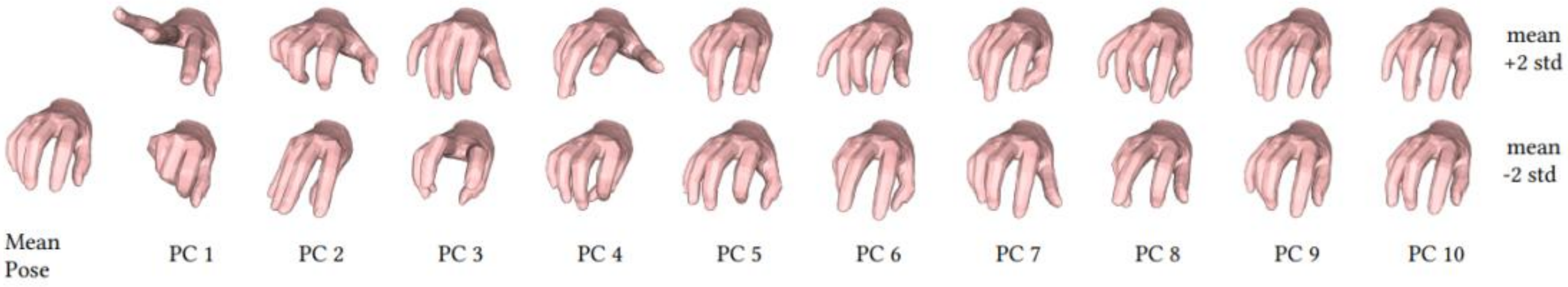
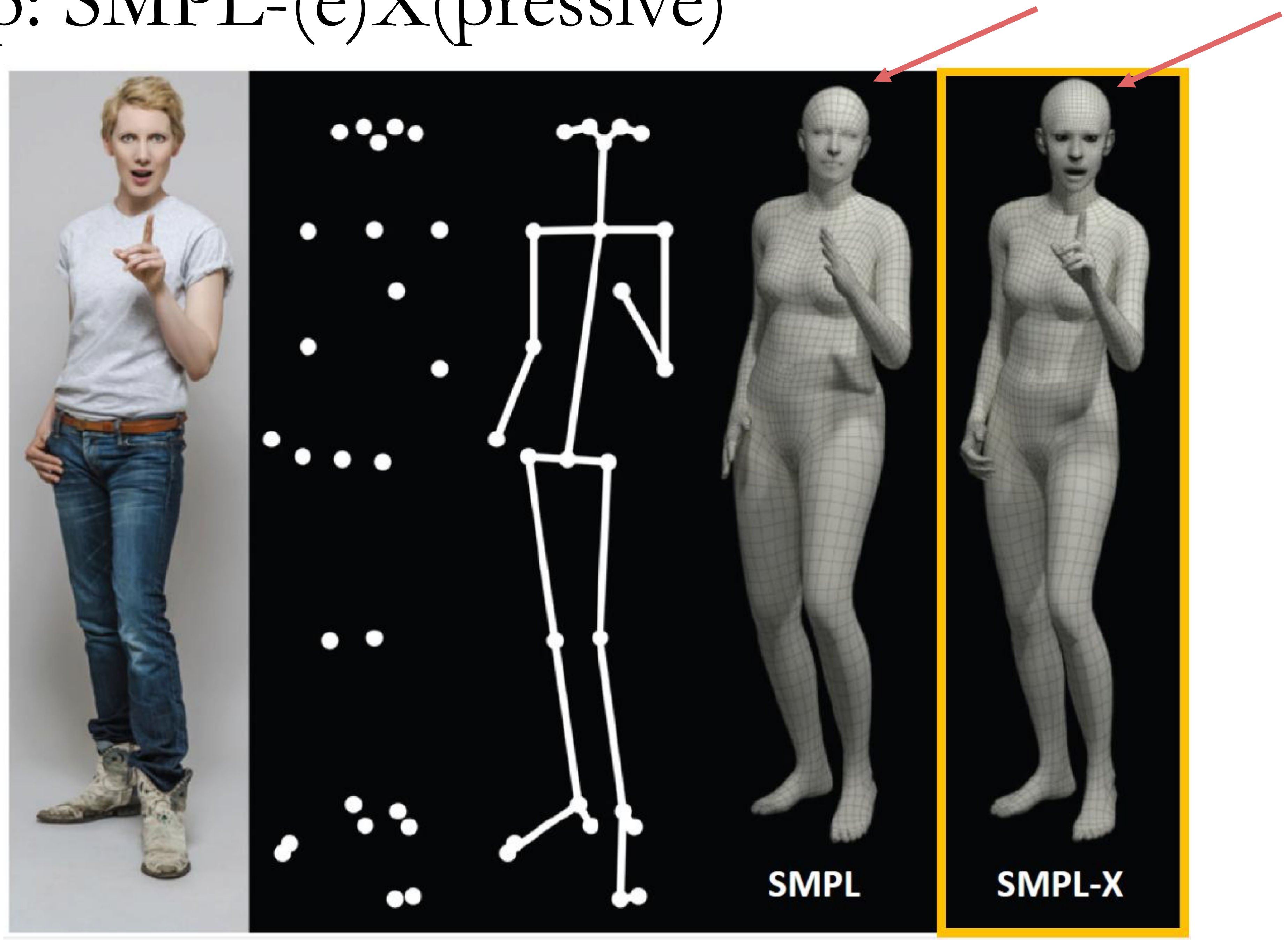


Fig. 10. **PCA pose space.** The left-most image depicts the *mean pose*, while the rest of the columns depict the effect of the first ten *principal components* (PCs) of the pose space. The effect of each PC is shown by adding ±2 standard deviations (std) to the mean pose, as indicated.

SMPL-H model

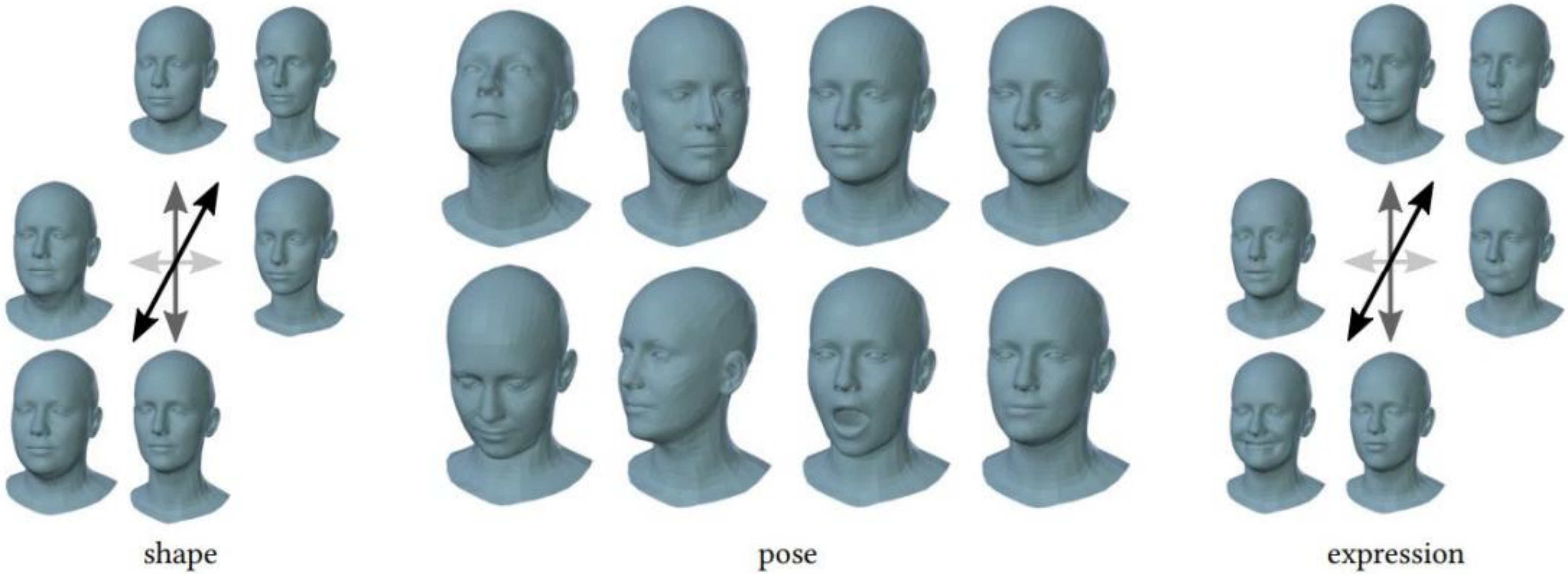
- Integrate MANO hand model to SMPL body model.
  - Total poses: $\theta \in \mathbb{R}^{156}$, 3 global orientation, 63 body pose, 45+45 hand poses.
  - Pose blend weights $\mathcal{W}$: concatenate hand and body blend weights.
  - Shape: $\beta \in \mathbb{R}^{10}(\text{or } \mathbb{R}^{300})$, shape parameters from SMPL, same shape blend weights.

Follow-up: SMPL-(e)X(pressive)



FLAME head model

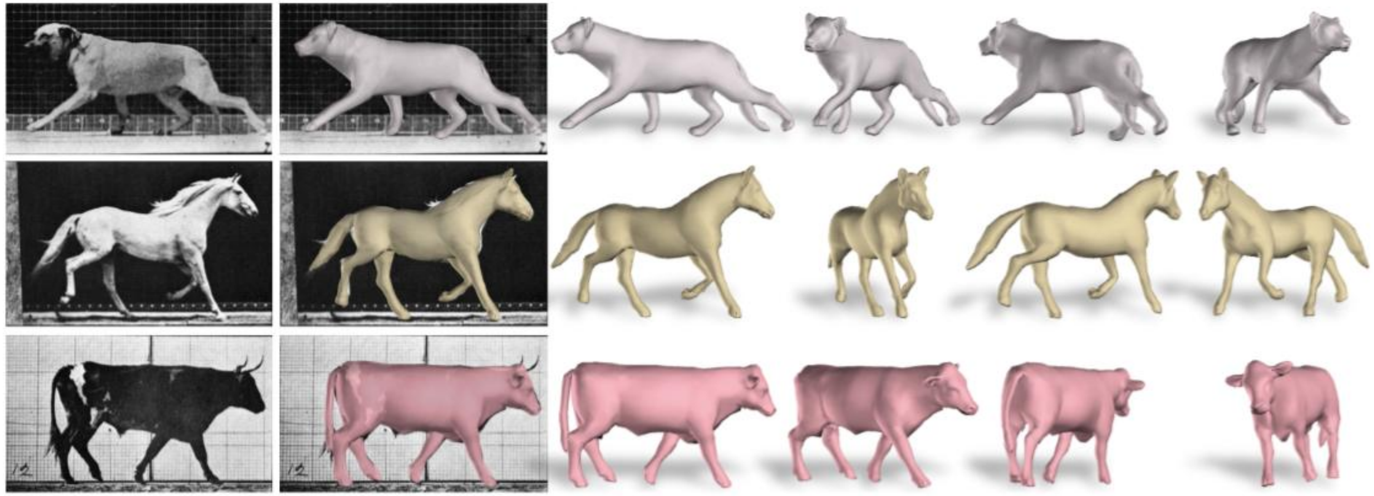
- Parametrize the head by shape β, pose θ, expression ψ
  - Pose $\theta \in \mathbb{R}^{3K+3}$: $K = 4$ joints (neck, jaw, eyeballs).
  - Shape and expression: PCA components.



Follow-up: SMAL

3D Menagerie: Modeling the 3D Shape and Pose of Animals

Silvia Zuffi[1] Angjoo Kanazawa[2] David Jacobs[2] Michael J. Black[3]
[1]IMATI-CNR, Milan, Italy, [2]University of Maryland, College Park, MD
[3]Max Planck Institute for Intelligent Systems, Tübingen, Germany
silvia@mi.imati.cnr.it, {kanazawa, djacobs}@umiacs.umd.edu, black@tuebingen.mpg.de



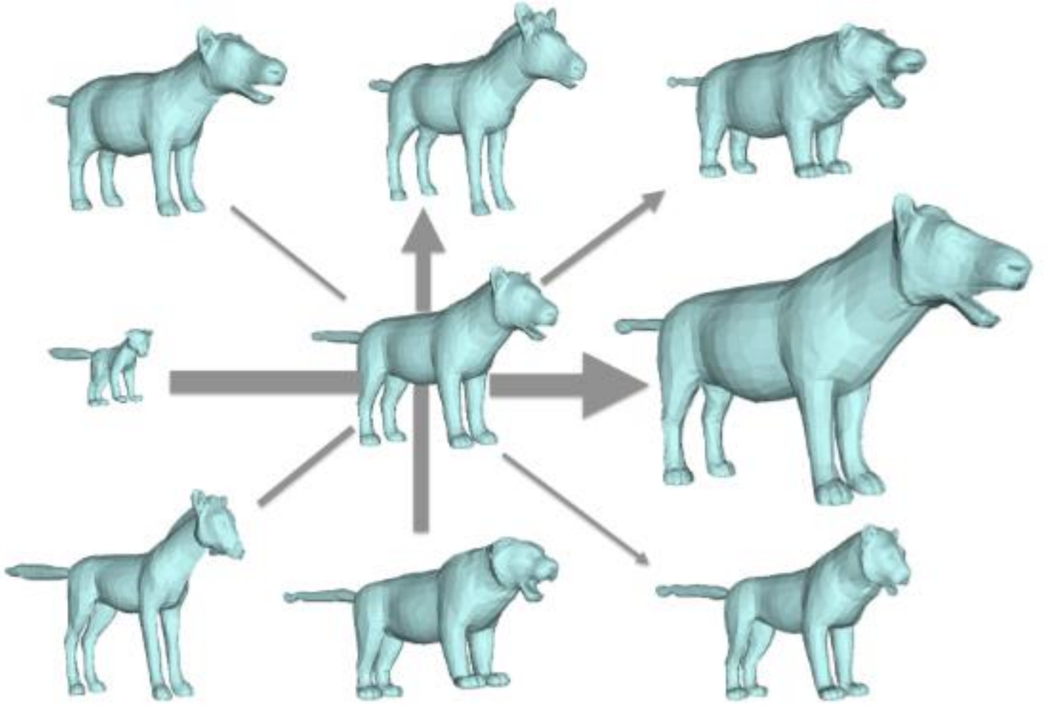


Figure 7: **PCA space.** First 4 principal components. Mean shape is in the center. The width of the arrow represents the order of the components. We visualise deviations of ±2std.

Learned from animals figurines

Follow-up: SMIL

Learning an Infant Body Model from RGB-D Data for Accurate Full Body Motion Analysis

Nikolas Hesse[1*], Sergi Pujades[2], Javier Romero[3], Michael J. Black[2], Christoph Bodensteiner[1], Michael Arens[1], Ulrich G. Hofmann[4], Uta Tacke[5], Mijna Hadders-Algra[6], Raphael Weinberger[7], Wolfgang Müller-Felber[7], and A. Sebastian Schroeder[7]

Skinned Multi-Linear Infant Model (SMIL)

Registration Results

Follow-up: Biomechanical skeleton

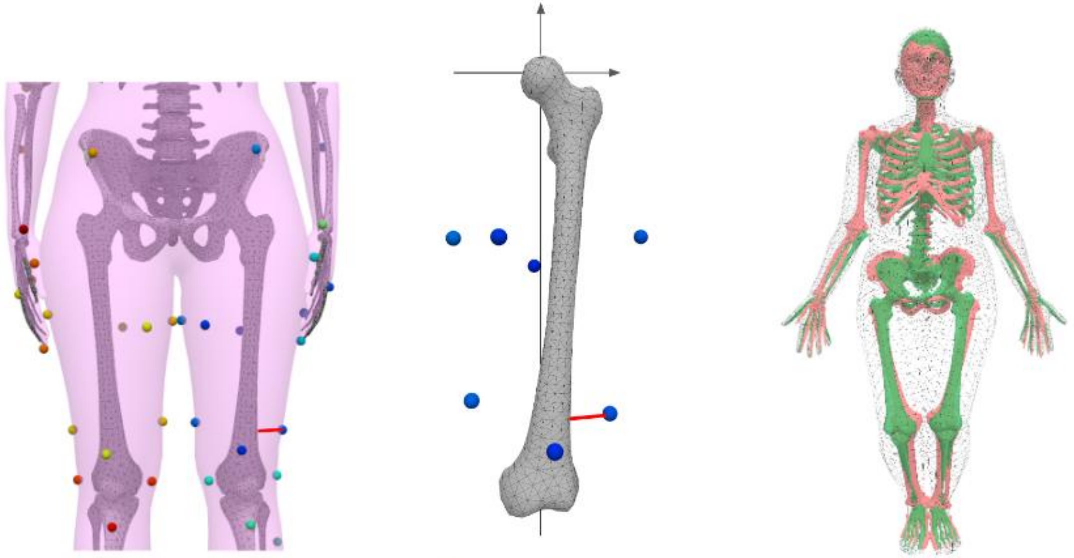
OSSO: Obtaining Skeletal Shape from Outside

Marilyn Keller[1] Silvia Zuffi[2] Michael J. Black[1] Sergi Pujades[3]

[1]Max Planck Institute for Intelligent Systems, Tübingen, Germany
[2]IMATI-CNR, Milan, Italy
[3]Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, France

From Skin to Skeleton:
Towards Biomechanically Accurate 3D Digital Humans

MARILYN KELLER, Max Planck Institute for Intelligent Systems, Germany
KEENON WERLING, Stanford University, USA
SOYONG SHIN, Carnegie Mellon University, USA
SCOTT DELP, Stanford University, USA
SERGI PUJADES, Inria centre at the University Grenoble Alpes, France
C. KAREN LIU, Stanford University, USA
MICHAEL J. BLACK, Max Planck Institute for Intelligent Systems, Germany

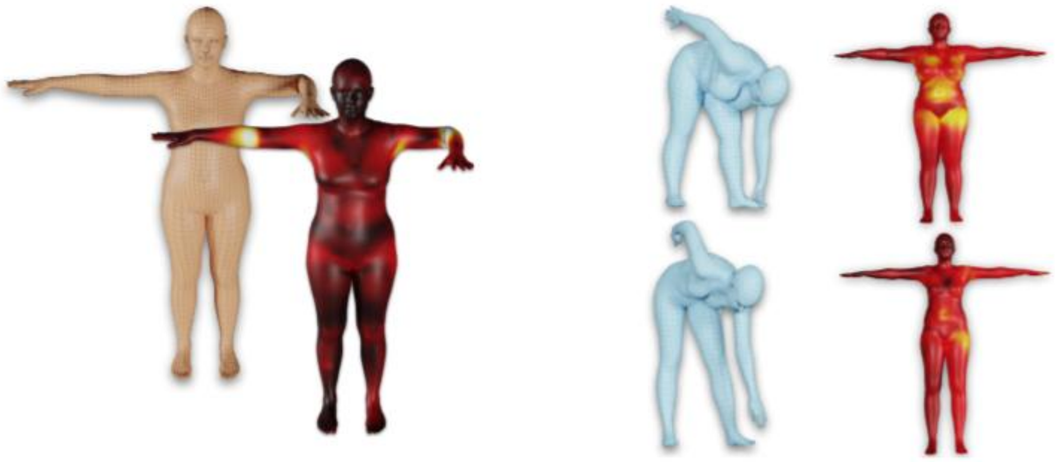


(a) Markers wrt OSSO

(b) Markers in the BSM bone frame

(c) AddBiomechanics fit result

SMPL: Limitations

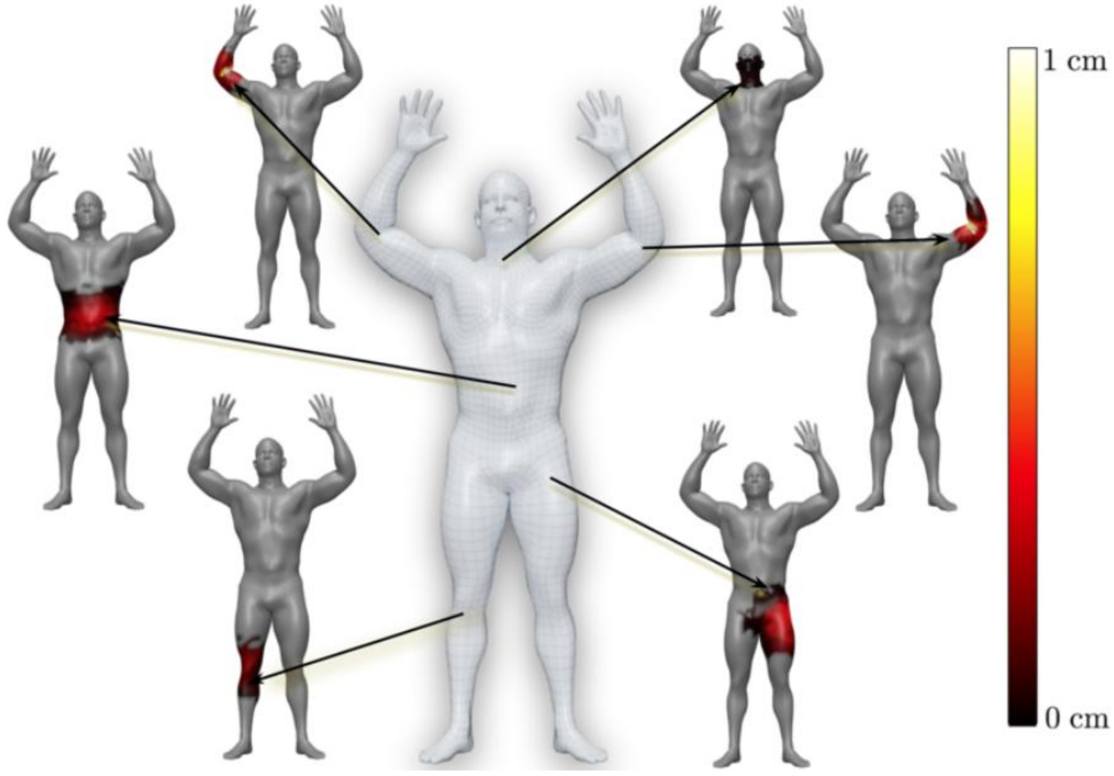


(a) (b)

SMPL has Spurious correlations
It factors out pose correction from identity
Training data fails to represent several body types

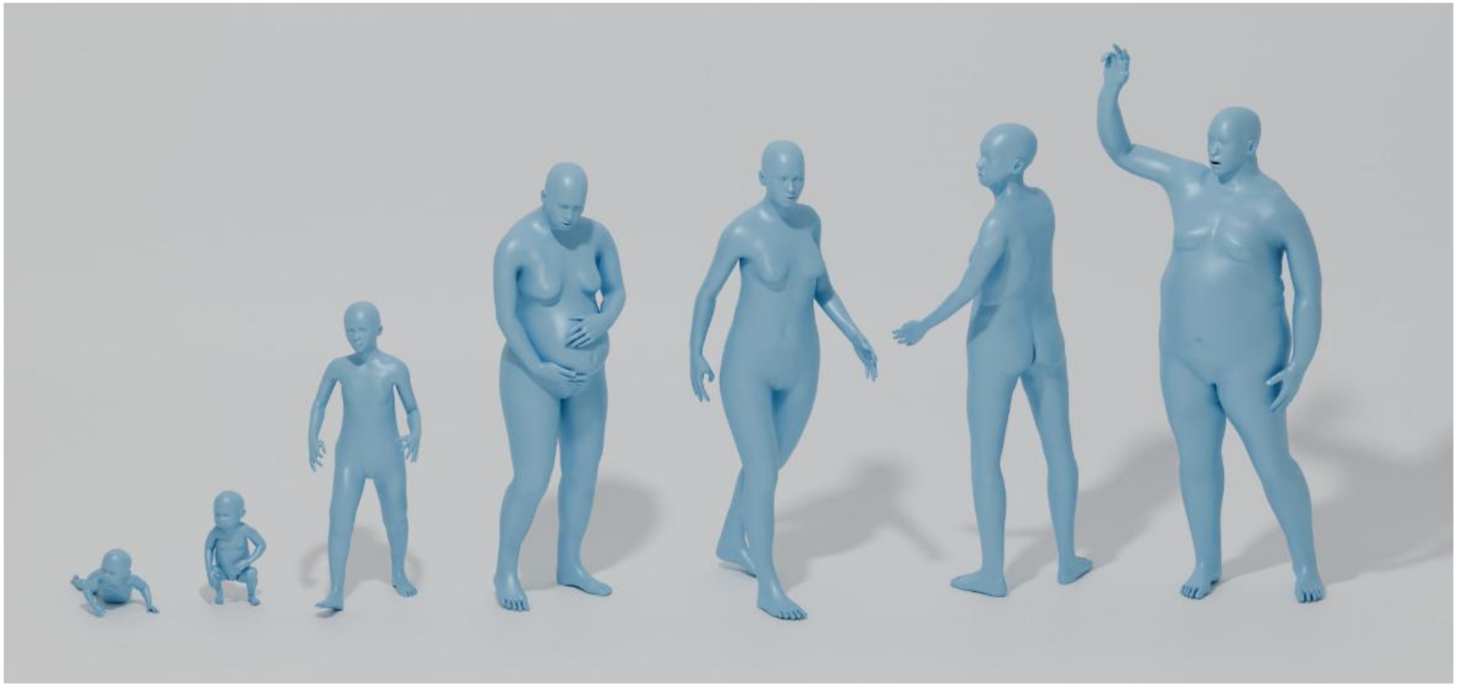
STAR: A Sparse Trained Articulated Human Body Regressor

Ahmed A. A. Osman, Timo Bolkart, Michael J. Black

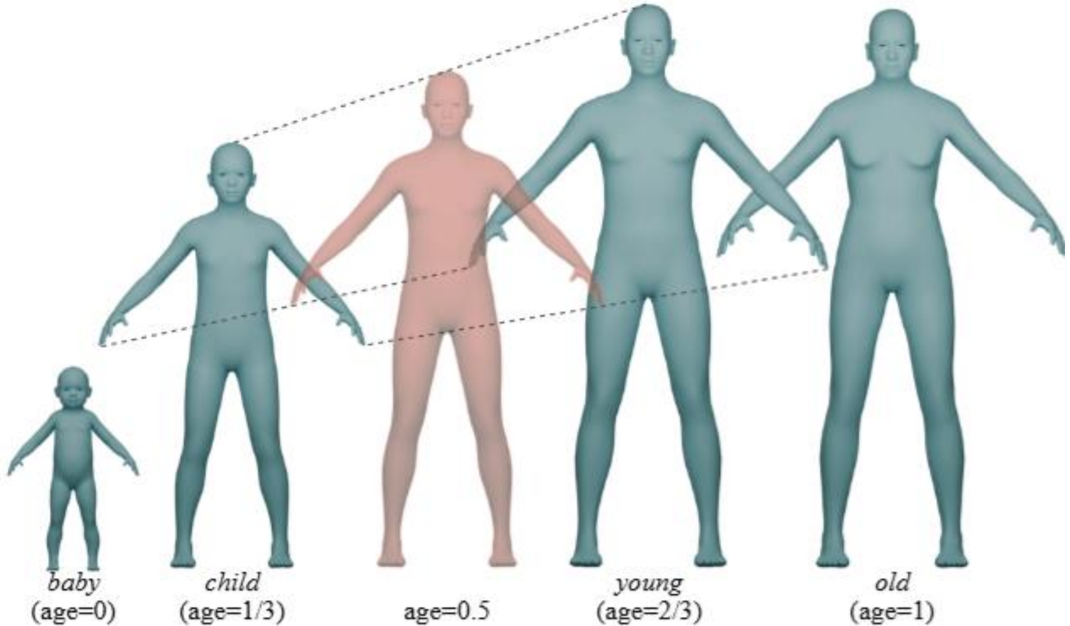


Human Mesh Modeling for Anny Body

Romain Brégier Guénolé Fiche Laura Bravo-Sánchez Thomas Lucas
Matthieu Armando Philippe Weinzaepfel Grégory Rogez Fabien Baradel
NAVER LABS Europe
https://github.com/naver/anny



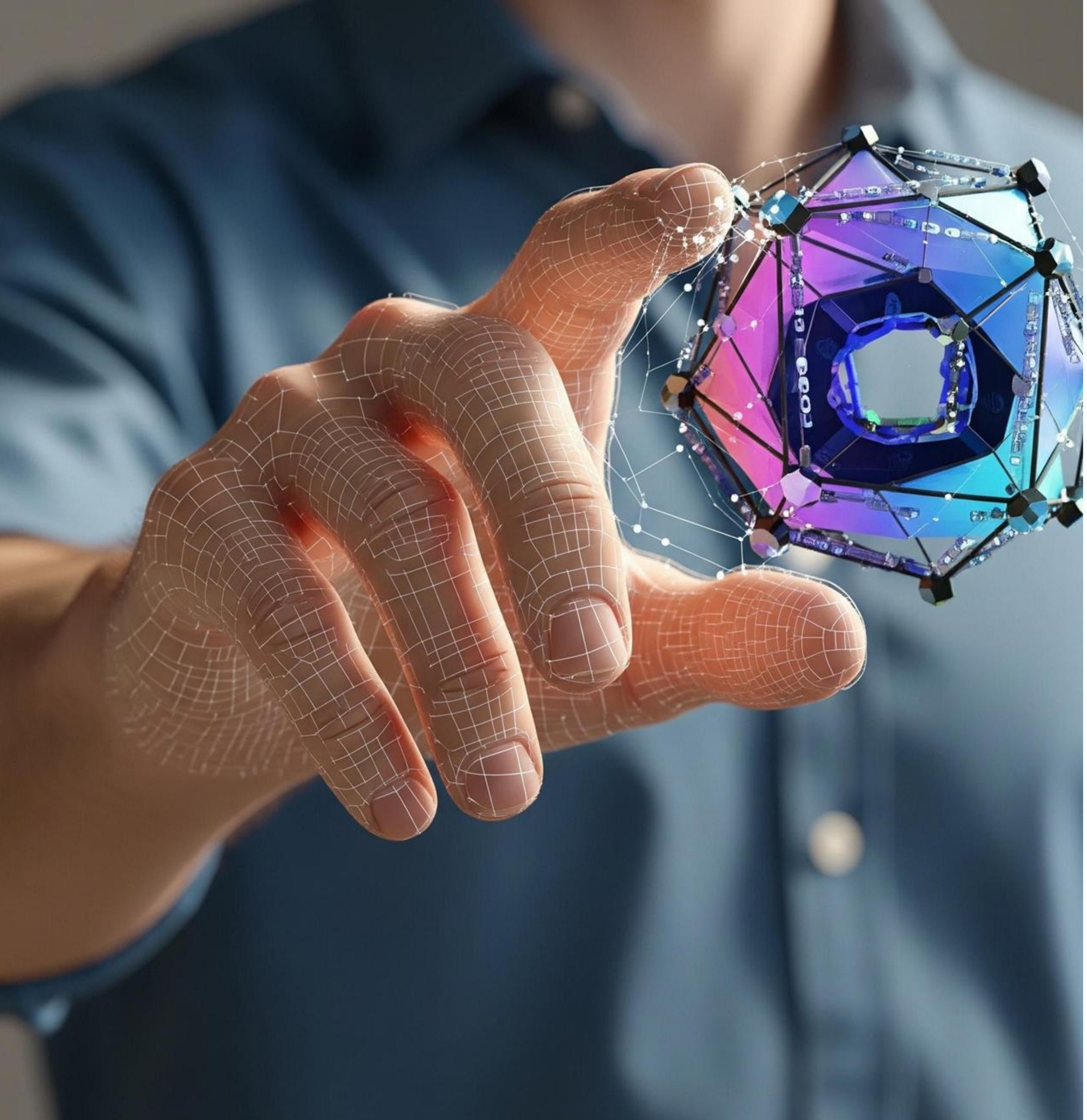
Similar to SMPL but:
1) Learned on syhtetic data
 + WHO anthropometric
2) 13K vertices, and 163 bones



MHR (Momentum Human Rig)

Meta

- **Identity Parameterization:** 45 shape parameters controlling body identity
- **Pose Parameterization:** 204 model parameters for full-body articulation
- **Facial Expression:** 72 expression parameters for detailed face animation
- **Multiple LOD Levels:** 7 levels of detail (LOD 0-6) for different performance requirements
- **Non-linear Pose Correctives:** Neural network-based pose-dependent deformations
- **PyTorch Integration:** GPU-accelerated inference for real-time applications
- **PyMomentum Integration:** Compatible to fast CPU solver



Beyond humans

Humans do not live in isolation

Nadia Magnenat-Thalmann
Daniel Thalmann

Virtual humans: thirty years of research, what next?

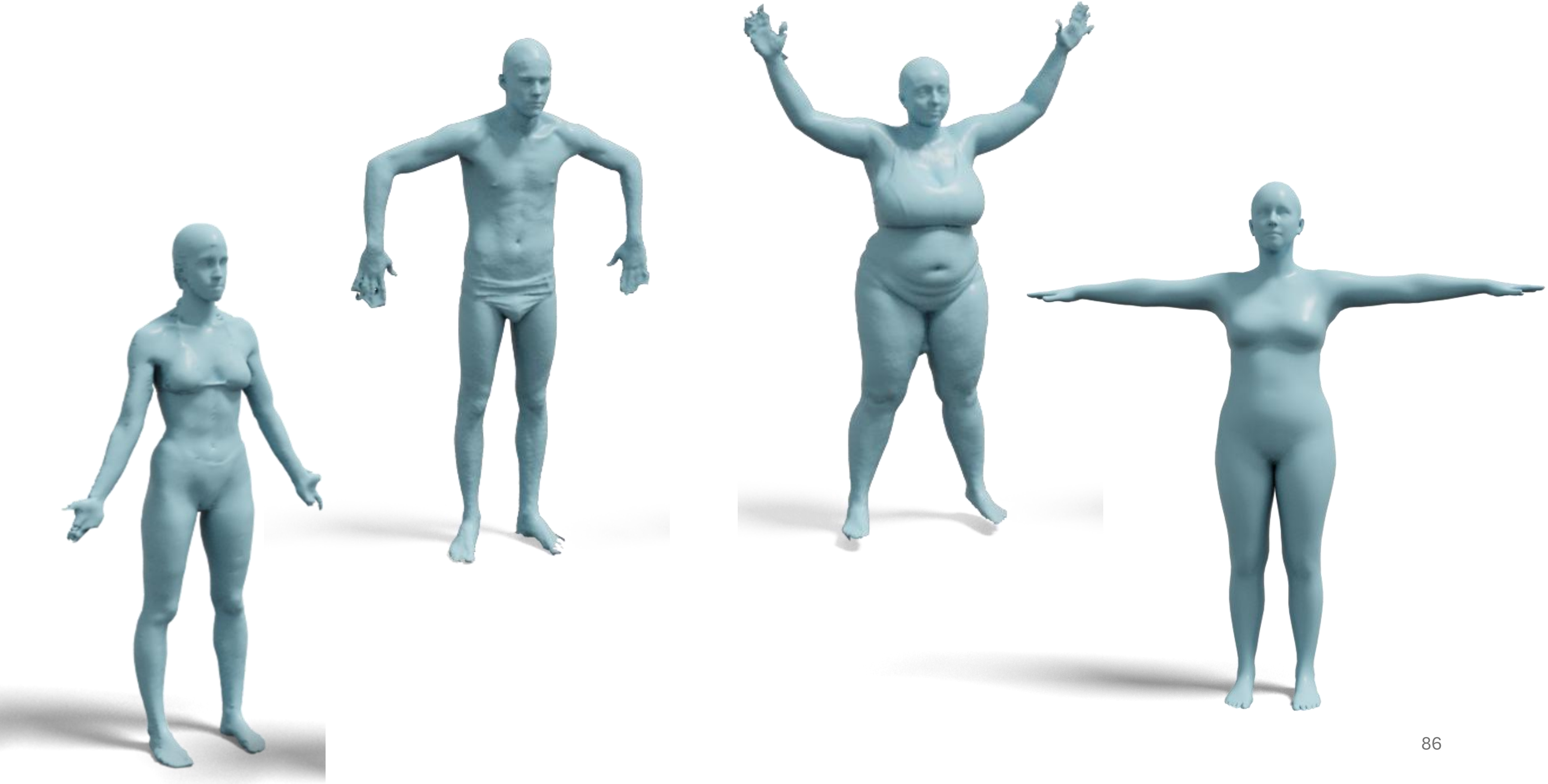
2005

The conception of such virtual humans is an immense challenge as it requires solving many problems in various areas. In this section, we will study six major challenges.

1. Creating virtual humans with a better skeleton and realistic deformable bodies.
2. Developing methods to generate on-the-fly flexible motion.
3. Generating complex behaviours of virtual humans inside their environments using a realistic perception of the environment.
4. Creating believable relationships between real and virtual humans based on emotion and personality.
5. Simulating realistic and believable behaviors of groups and crowds.
6. Generating realistic virtual clothed and haired people.

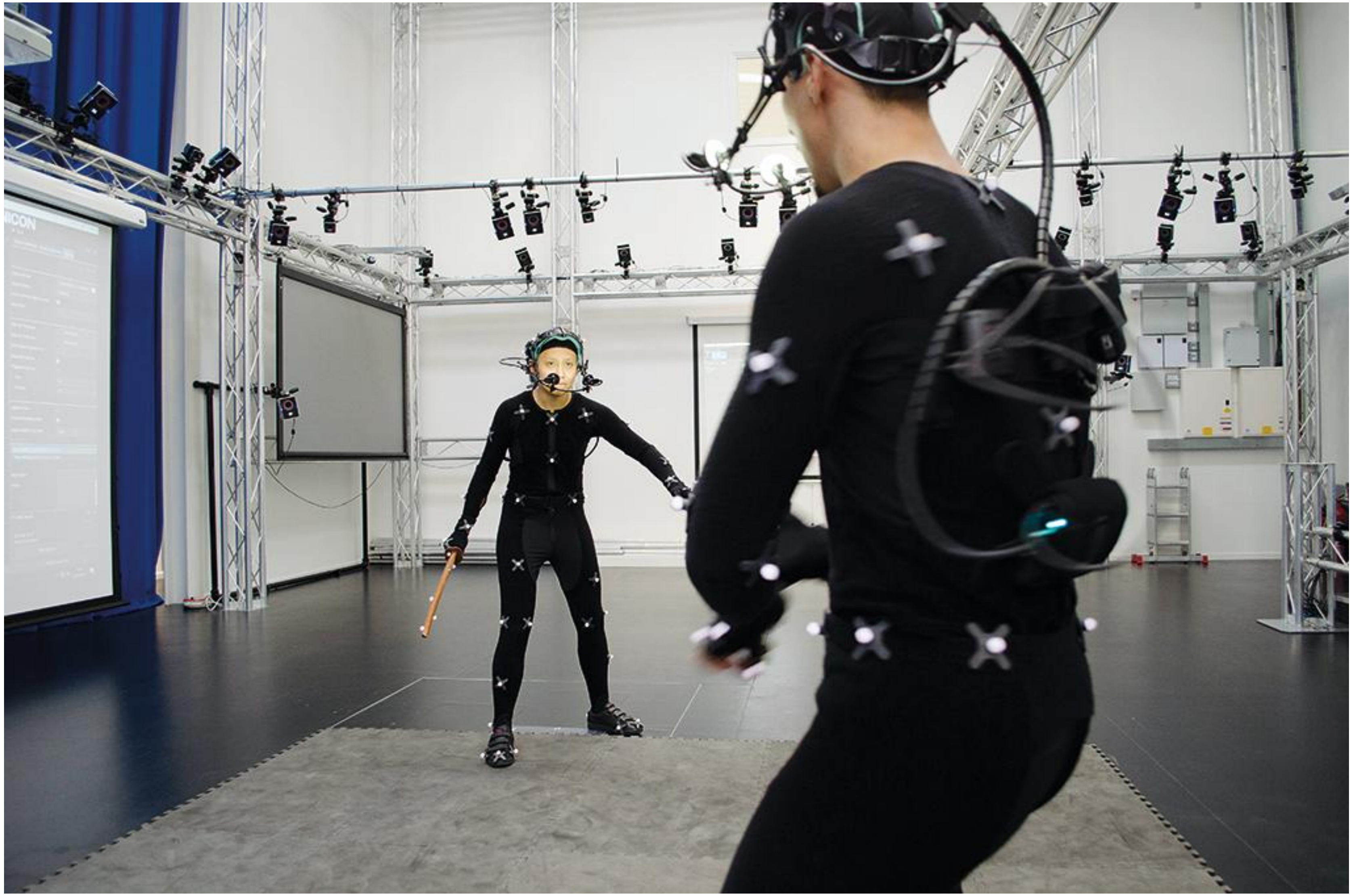
Fig. 5. Virtual humans waiting at a tram stop in Geneva

To be believable, an actor has to be affected by what takes place around it and needs to engage in social behaviors with other actors.






Motion capturing is limited and resource demanding



Controlled setting

Costly

Professionists required

Limited capturing volume

Wearables

Wearables

Interactions in large scenes

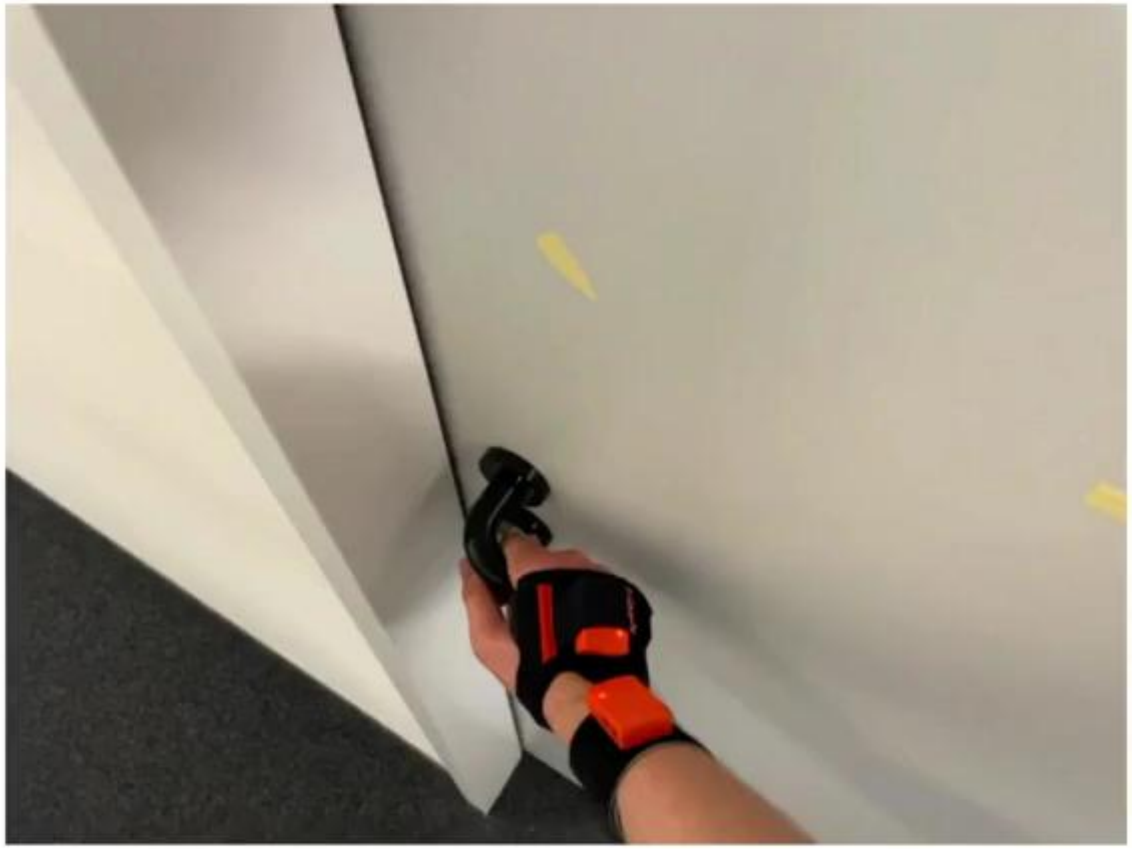
Human POSEitioning System (HPS):
3D Human Pose Estimation and Self-localization in Large Scenes from Body-Mounted Sensors

Vladimir Guzov* [1,2], Aymen Mir* [1,2], Torsten Sattler [3], Gerard Pons-Moll [1,2]

[1] University of Tübingen, Germany
[2] Max Planck Institute for Informatics, Saarland Informatics Campus, Germany
[3] CIIRC, Czech Technical University in Prague, Czech Republic

* Equal contribution

The object can be barely visible or not visible at all



Object can obstruct the whole view

Interaction Replica: Tracking human-object interaction and scene changes from human motion; Guzov, Chibane, Marin, He, Sattler, Pons-Moll, 3DV, 2024

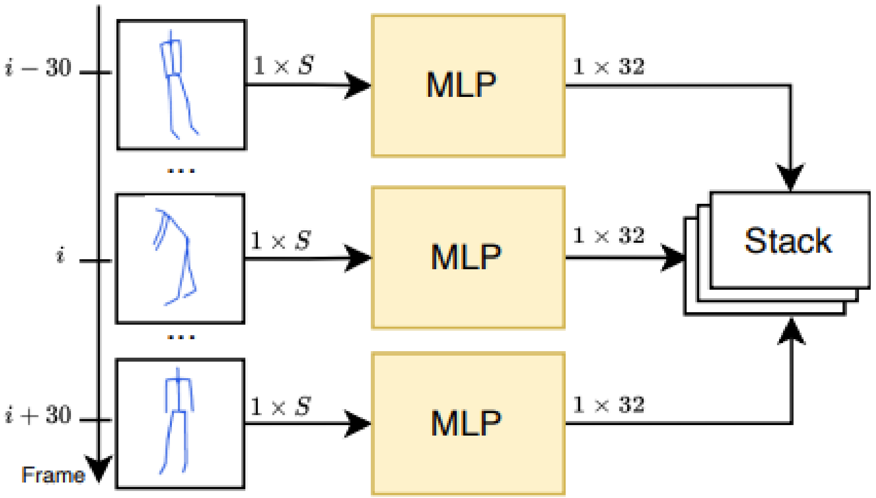
Interaction Replica: Tracking human-object interaction and scene changes from human motion

3DV submission #68

Includes Audio

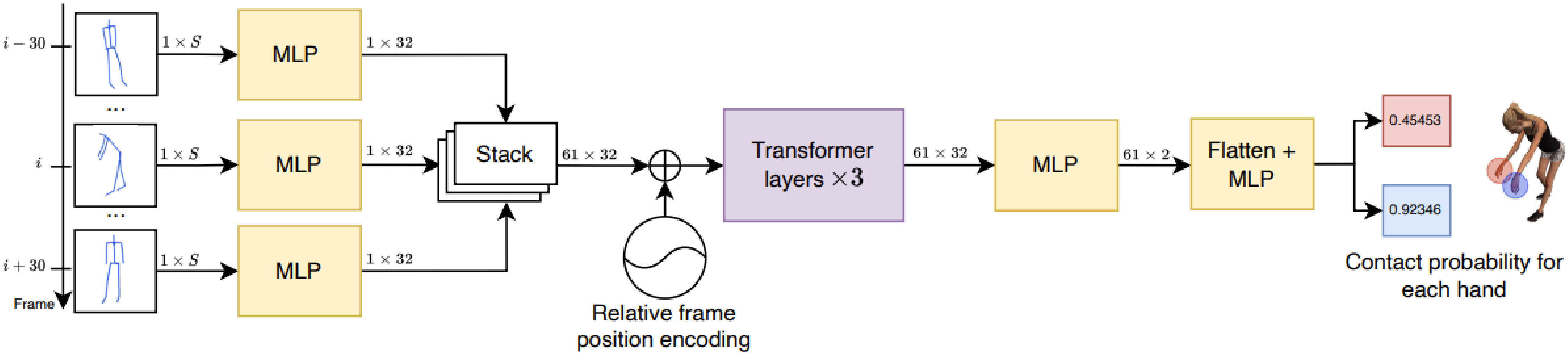
Interaction Replica: Tracking human-object interaction and scene changes from human motion; Guzov, Chibane, Marin, He, Sattler, Pons-Moll, 3DV, 2024

Contact predictor



Input:
61 sequential poses

Contact predictor



Input:
61 sequential poses

Output:
Contact probability for the central frame

Interaction Replica:
Tracking human-object interaction and scene changes from human motion

3DV submission #68

Includes Audio

Interaction Replica: Tracking human-object interaction and scene changes from human motion; Guzov, Chibane, Marin, He, Sattler, Pons-Moll, 3DV, 2024

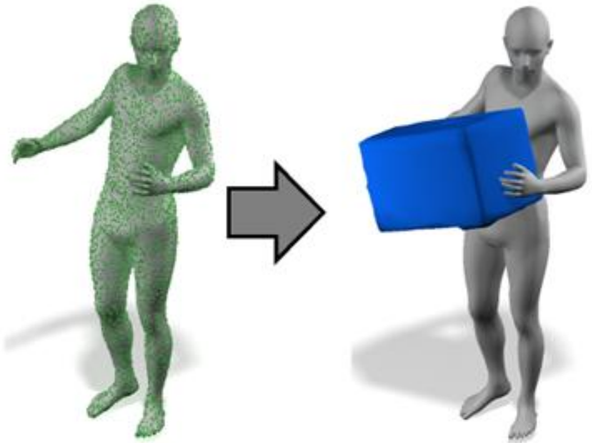
Interaction Replica: Tracking human-object interaction and scene changes from human motion

3DV submission #68

Includes Audio

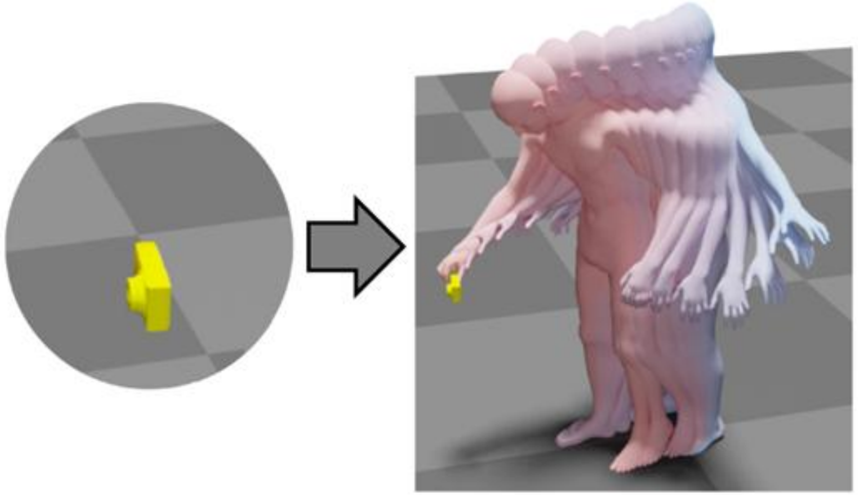
Interaction Replica: Tracking human-object interaction and scene changes from human motion; Guzov, Chibane, Marin, He, Sattler, Pons-Moll, 3DV, 2024

$$p(\mathcal{O}, \mathcal{I} | \mathcal{H})$$

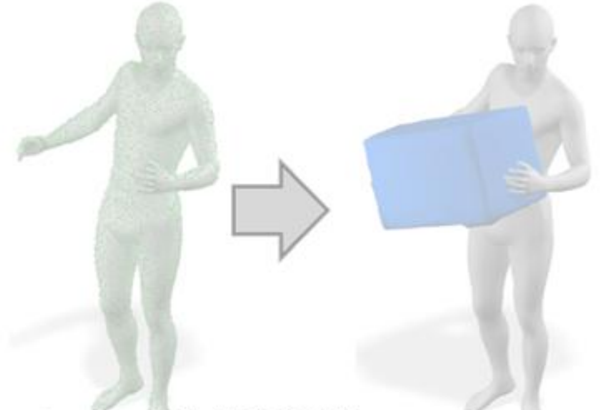


Object Pop-up I.A. Petrov, et al., CVPR'23

$$p(\mathcal{H}, \mathcal{I} | \mathcal{O})$$

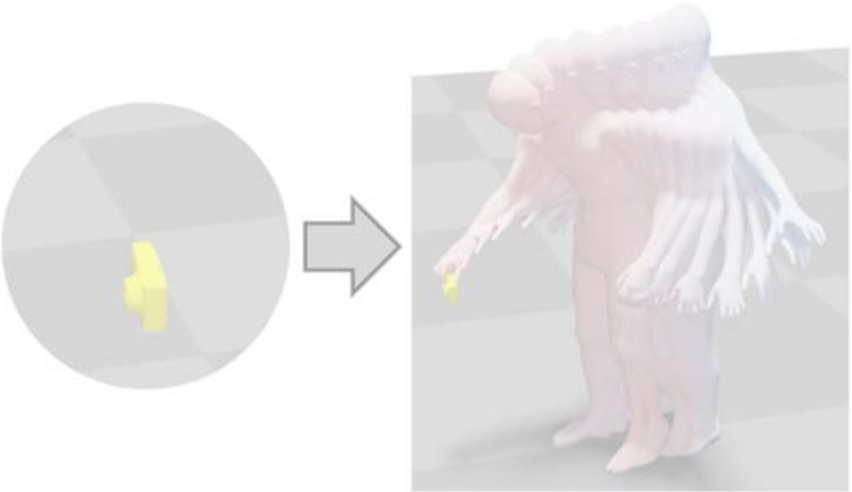


$$p(\mathcal{O}, \mathcal{I} | \mathcal{H})$$

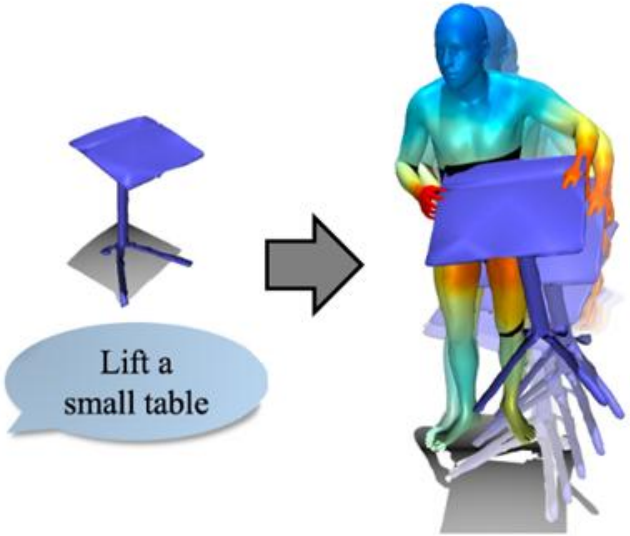


GOAL O. Taheri, et al., CVPR’22
SAGA Y. Wu, et al., ECCV’22

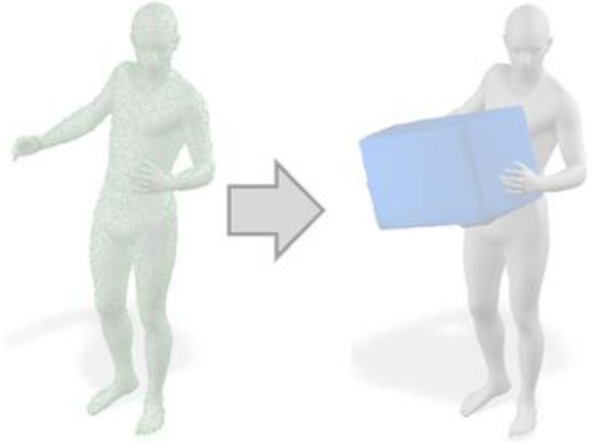
$p(\mathcal{H},\mathcal{I}|\mathcal{O})$



$p(\mathcal{H},\mathcal{O}|\mathcal{I})$

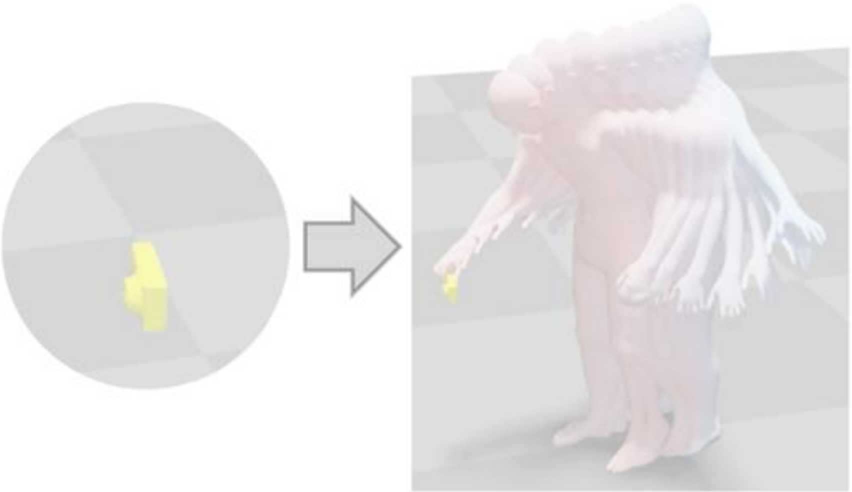


$p(\mathcal{O},\mathcal{I}|\mathcal{H})$

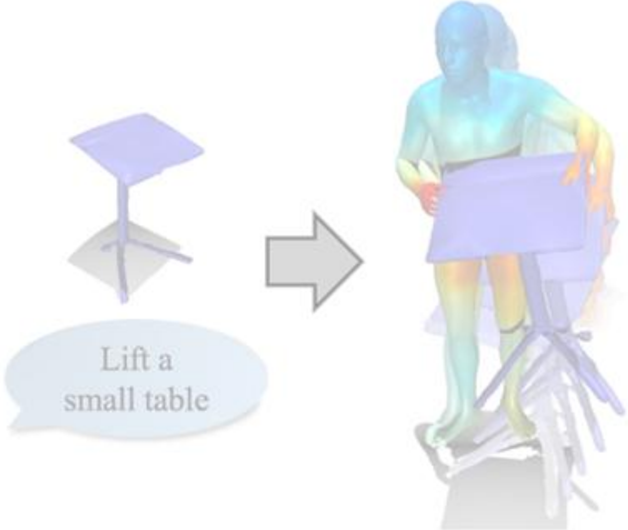


CG-HOI C. Diller, A. Dai, CVPR'24

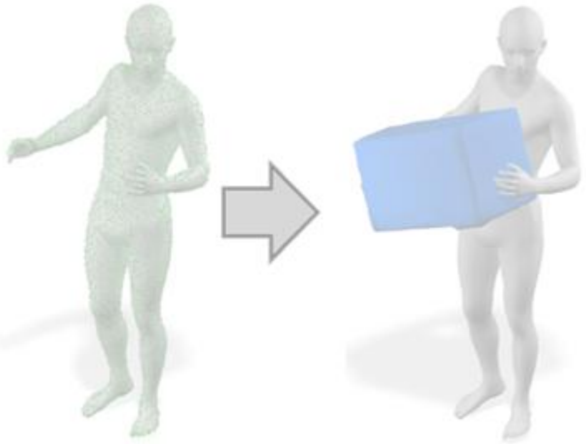
$p(\mathcal{H}, \mathcal{I} | \mathcal{O})$



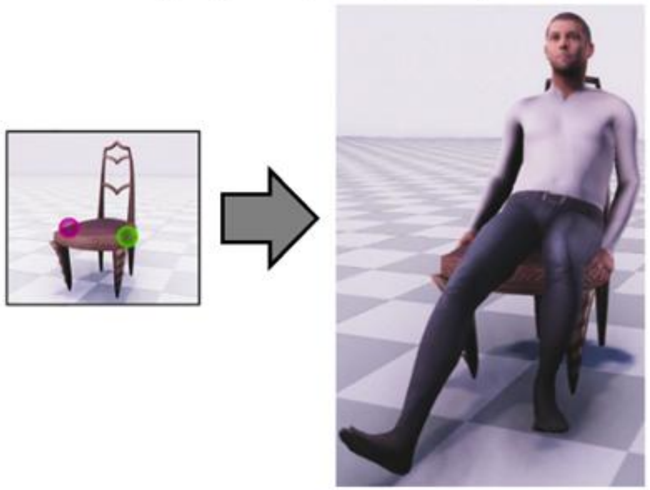
$p(\mathcal{H}, \mathcal{O} | \mathcal{I})$



$p(\mathcal{O}, \mathcal{I} | \mathcal{H})$



$p(\mathcal{H} | \mathcal{O}, \mathcal{I})$



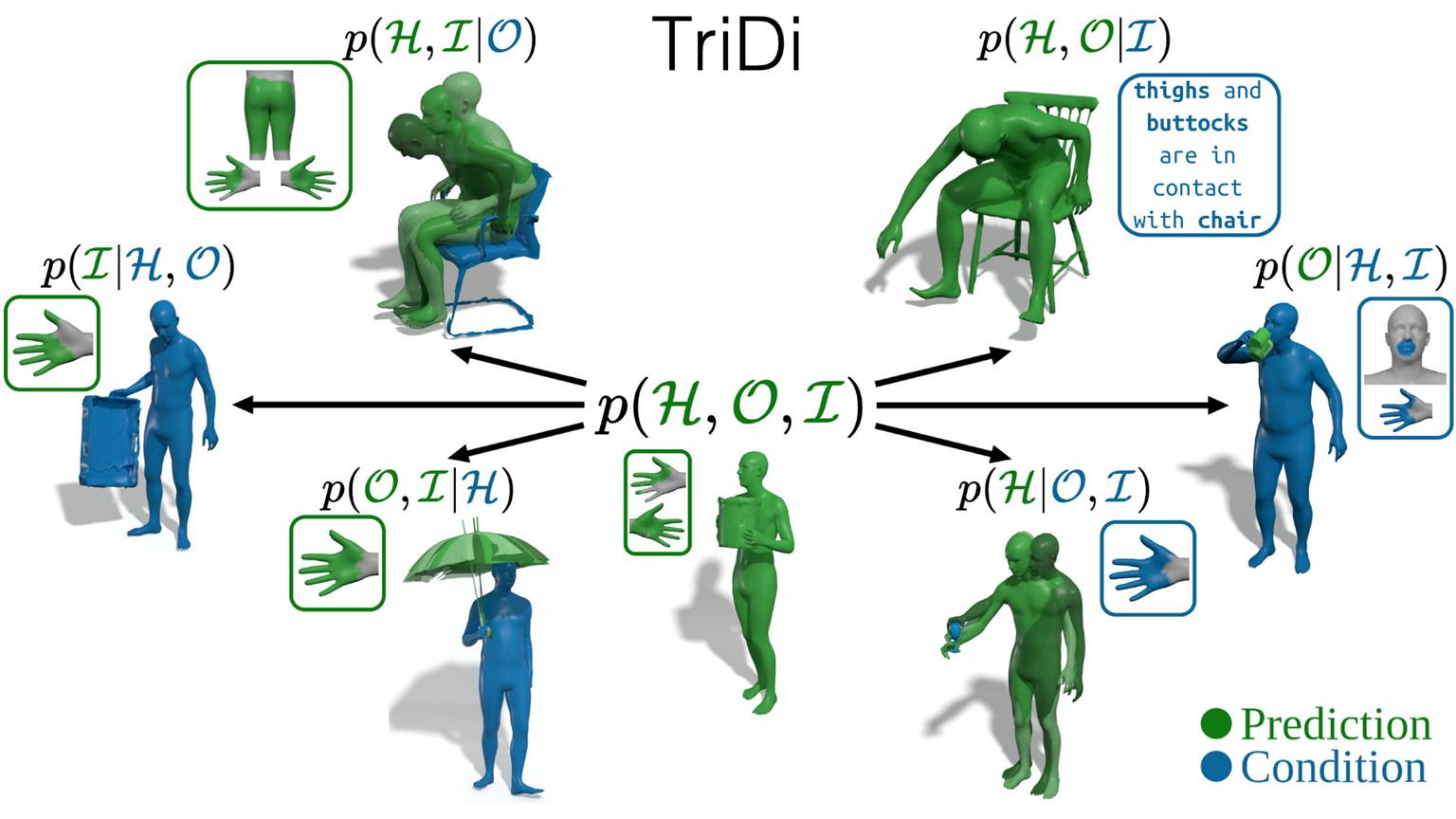
COUCH X. Zhang, et al., ECCV'22

$$p(\mathcal{H}, \mathcal{O}, \mathcal{I})$$

TriDi

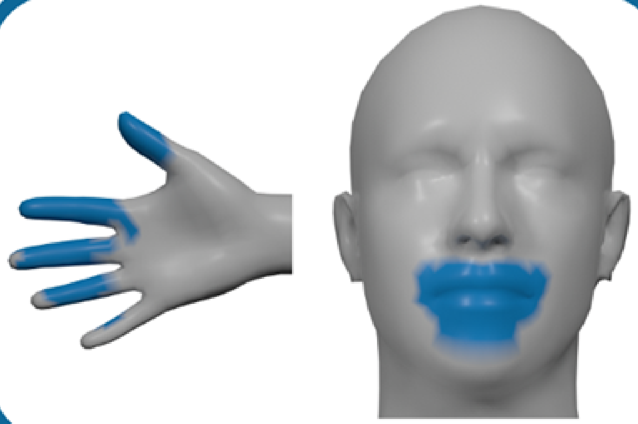
$$p(\mathcal{H}, \mathcal{O}, \mathcal{I})$$

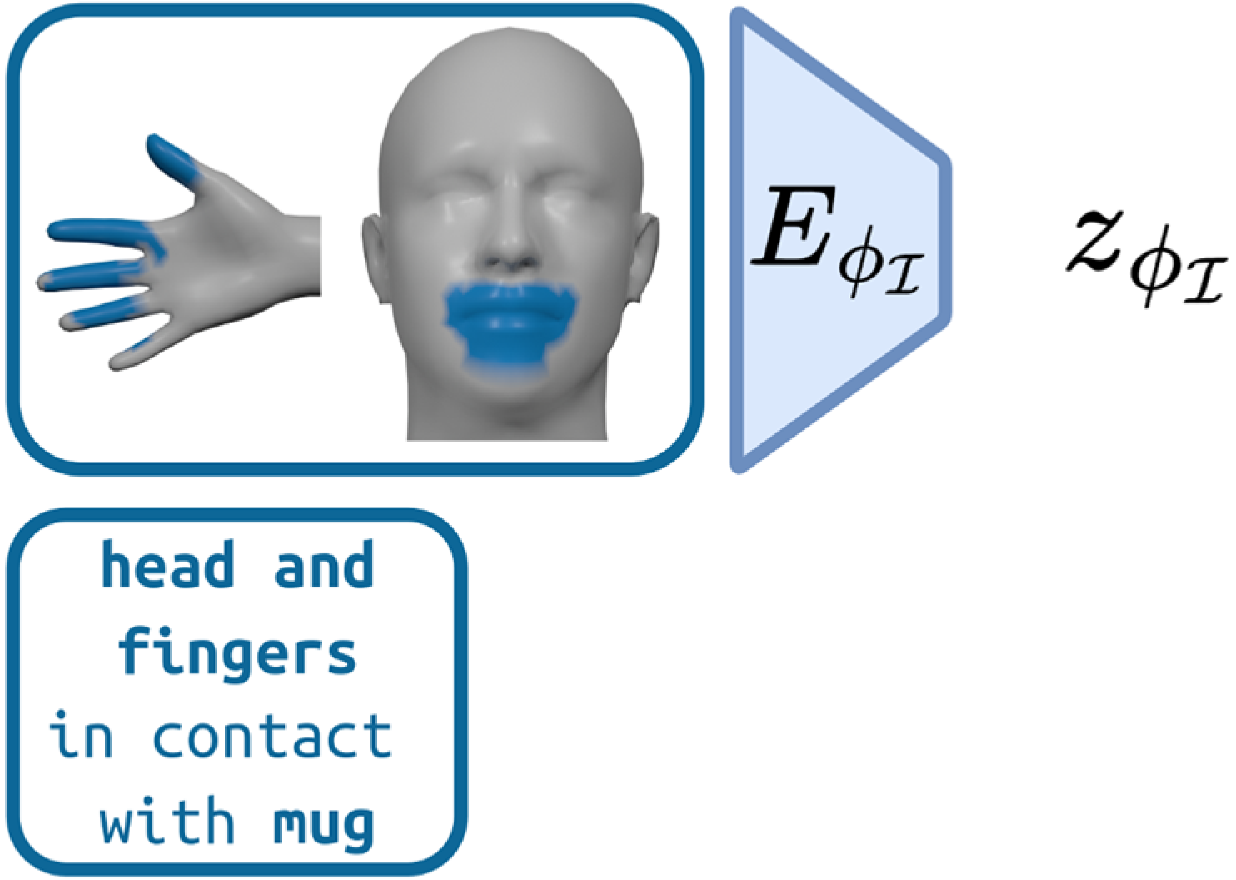
● Prediction

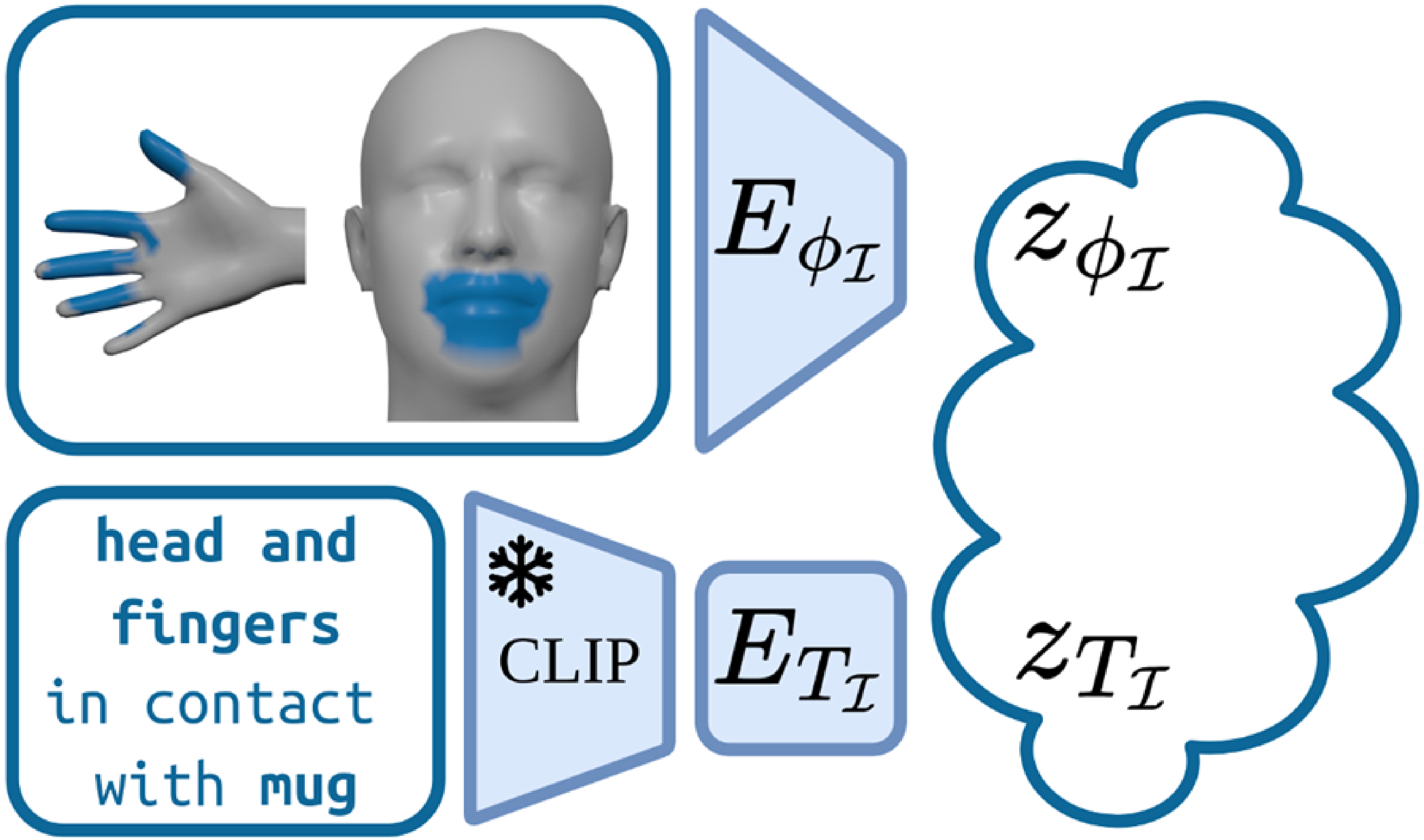


TriDi
$p(\mathcal{H}, \mathcal{I} | \mathcal{O})$
$p(\mathcal{H}, \mathcal{O} | \mathcal{I})$
thighs and buttocks are in contact with chair
$p(\mathcal{I} | \mathcal{H}, \mathcal{O})$
$p(\mathcal{O} | \mathcal{H}, \mathcal{I})$
$p(\mathcal{H}, \mathcal{O}, \mathcal{I})$
$p(\mathcal{O}, \mathcal{I} | \mathcal{H})$
$p(\mathcal{H} | \mathcal{O}, \mathcal{I})$
Prediction
Condition


Generation 1

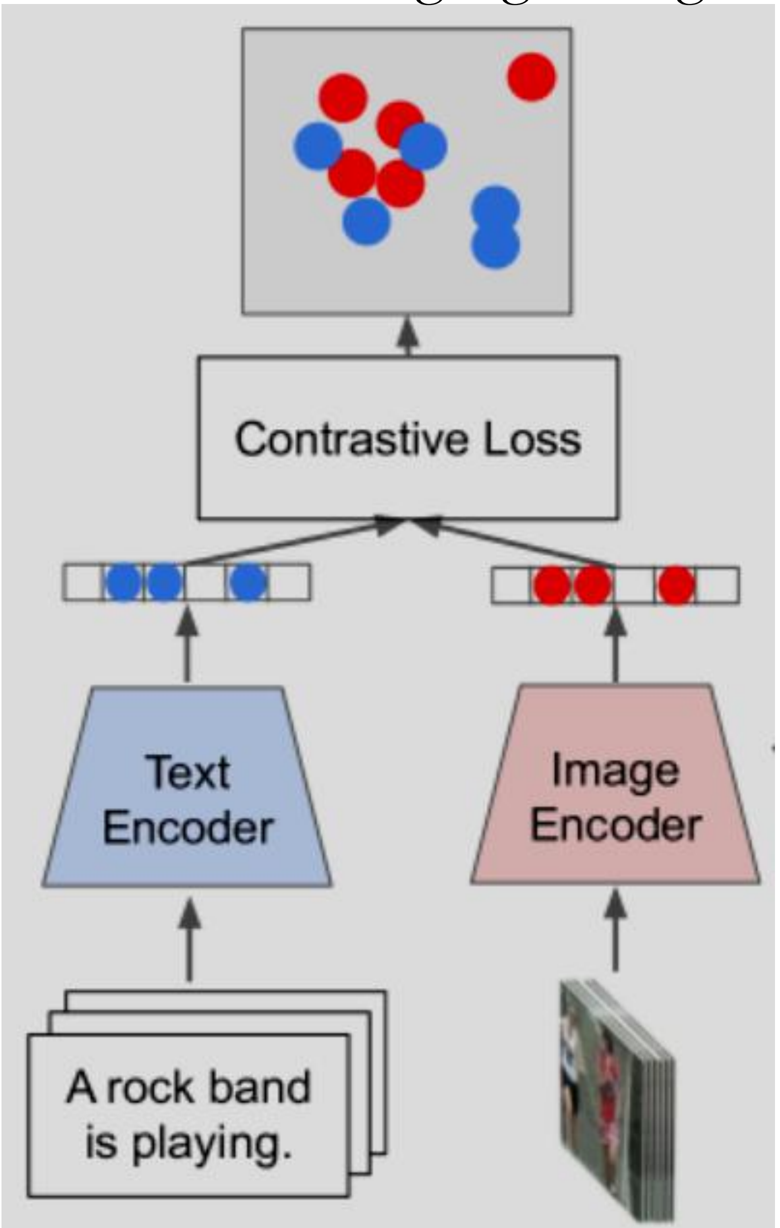
	Finegrained	Easy to control
	✓	✗
head and fingers in contact with **mug**	✗	✓


$E_{\phi_{\mathcal{I}}}$
$z_{\phi_{\mathcal{I}}}$
head and fingers in contact with mug

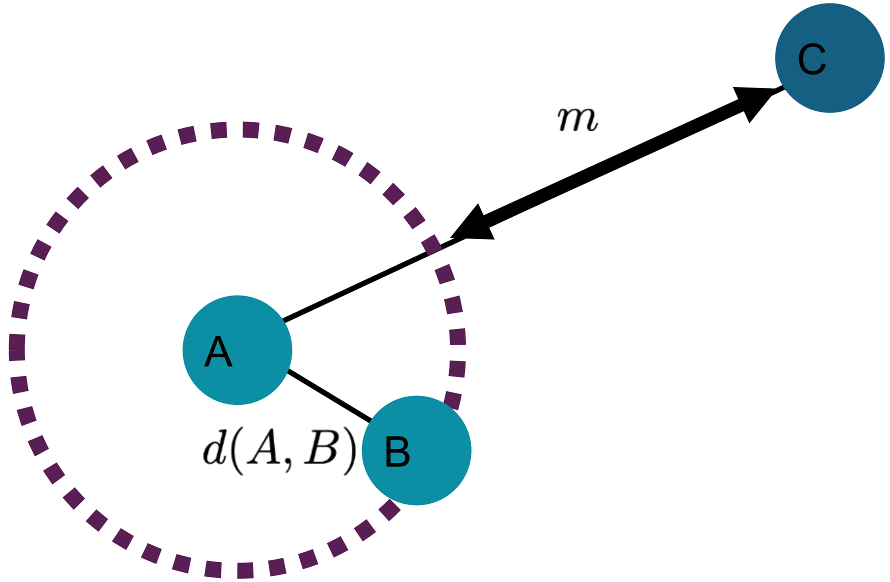


A. Radford, et al.: Learning transferable visual models from natural language supervision

Contrastive Language-Image Pre-Training (CLIP)

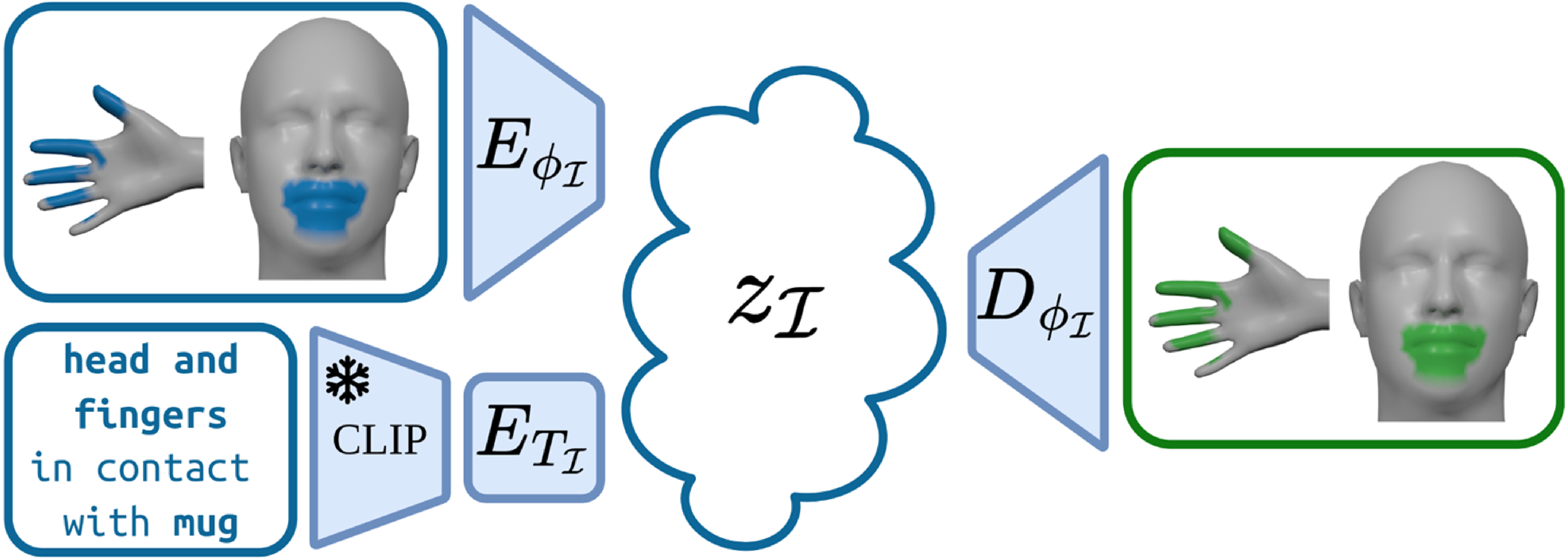


$$\mathcal{L}(A,B,C) = \max(0, \|f(A)-f(B)\|^2 - \|f(A)-f(C)\|^2 + m)$$



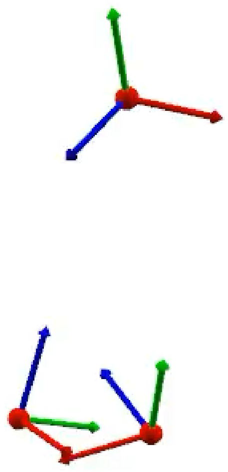
Bring closer similar input, and push away different ones
(up to a threshold)

https://ar5iv.labs.arxiv.org/html/2110.11499


$E_{\phi_{\mathcal{I}}}$
head and fingers in contact with mug
CLIP
$E_{T_{\mathcal{I}}}$
$z_{\mathcal{I}}$
$D_{\phi_{\mathcal{I}}}$

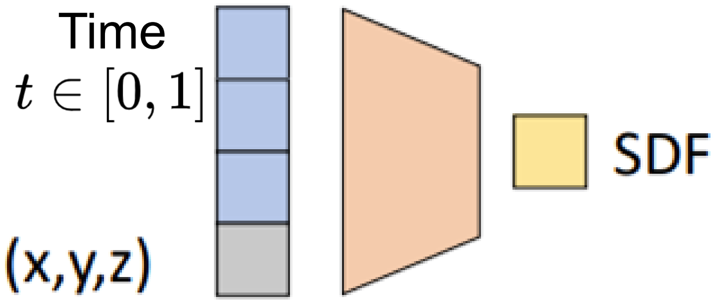
wearable devices
enable ego-centric vision

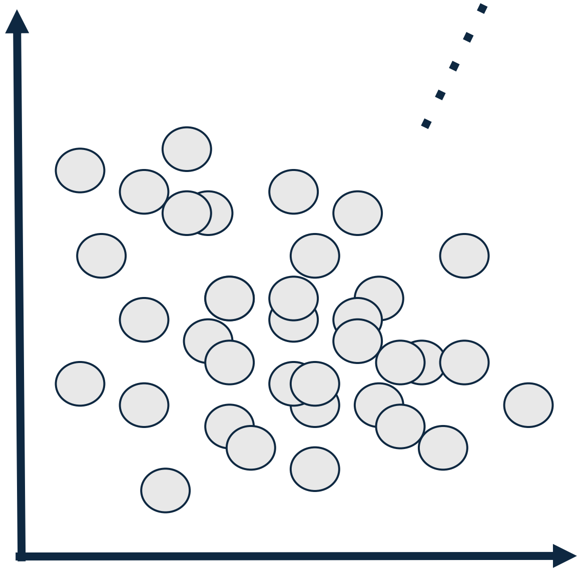
Input: 3-point tracking



ECHO prediction: HOI

Deep SDF



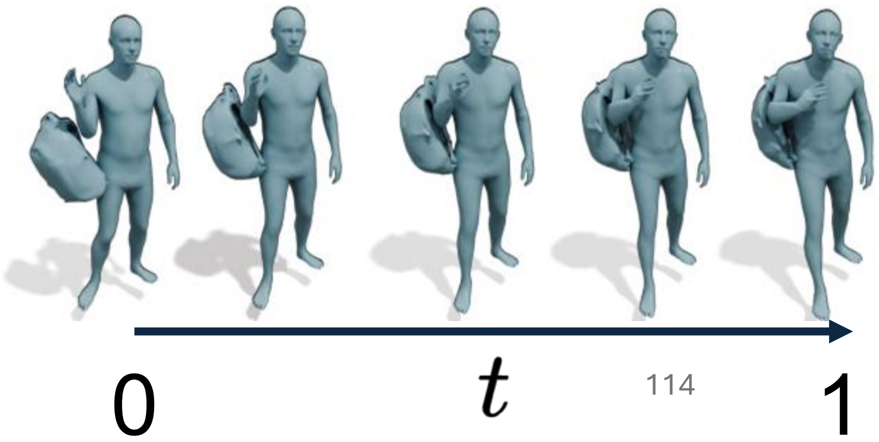


Problem

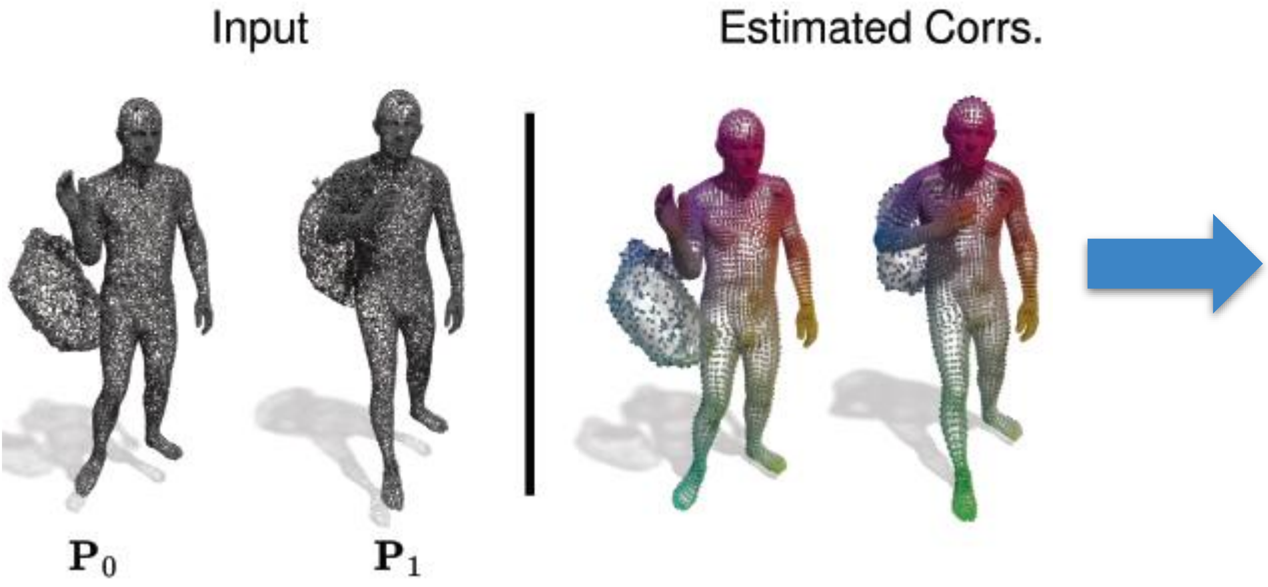
One network for a single shape

Solution

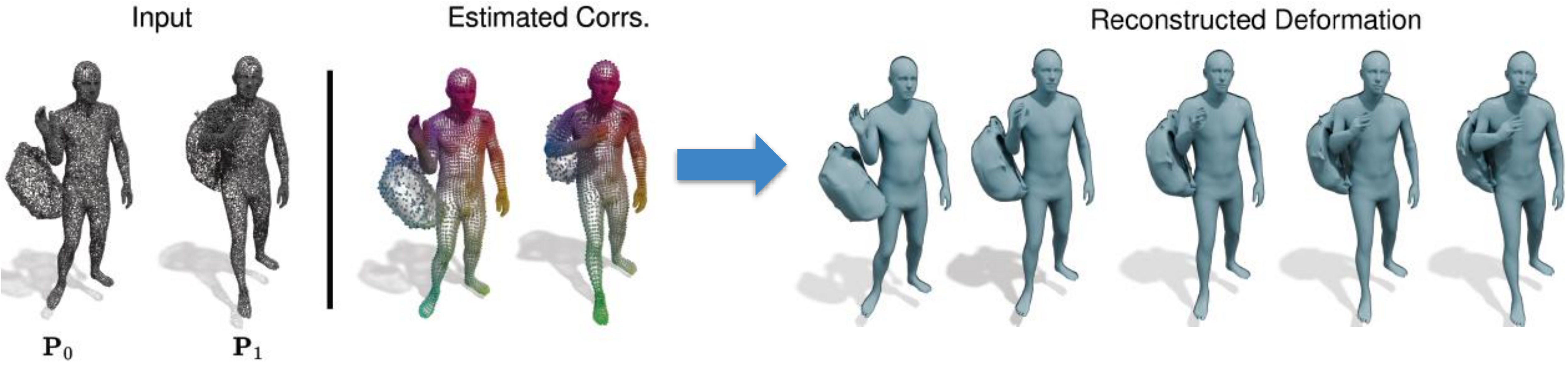
Condition the input



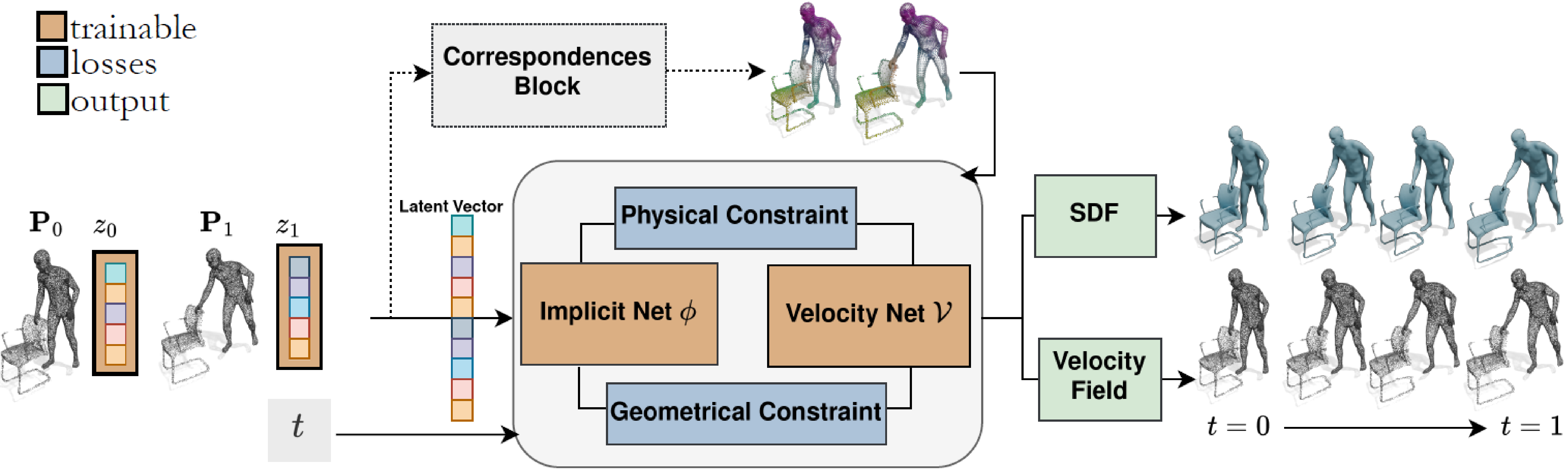
Goal: 3D to 4D reconstruction



Goal: 3D to 4D reconstruction

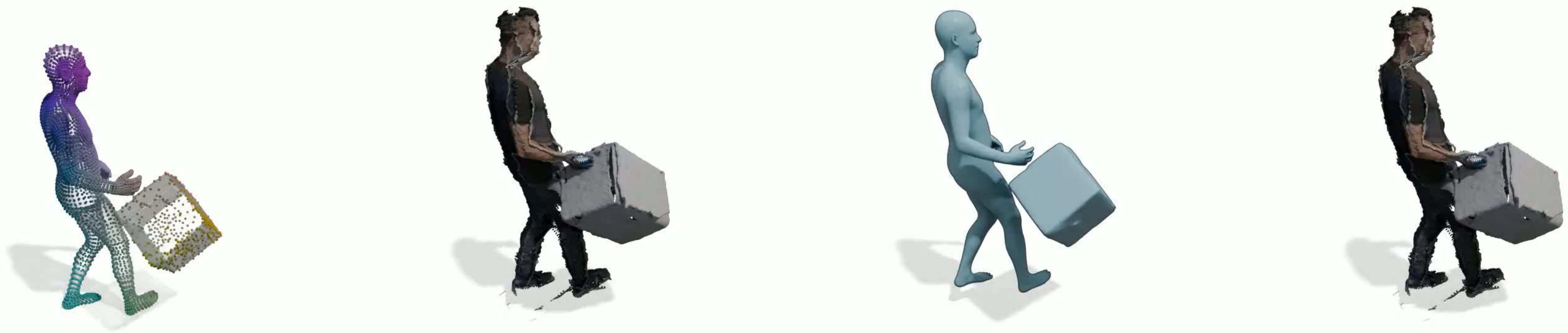


4Deform: Neural Surface Deformation for Robust Shape Interpolation



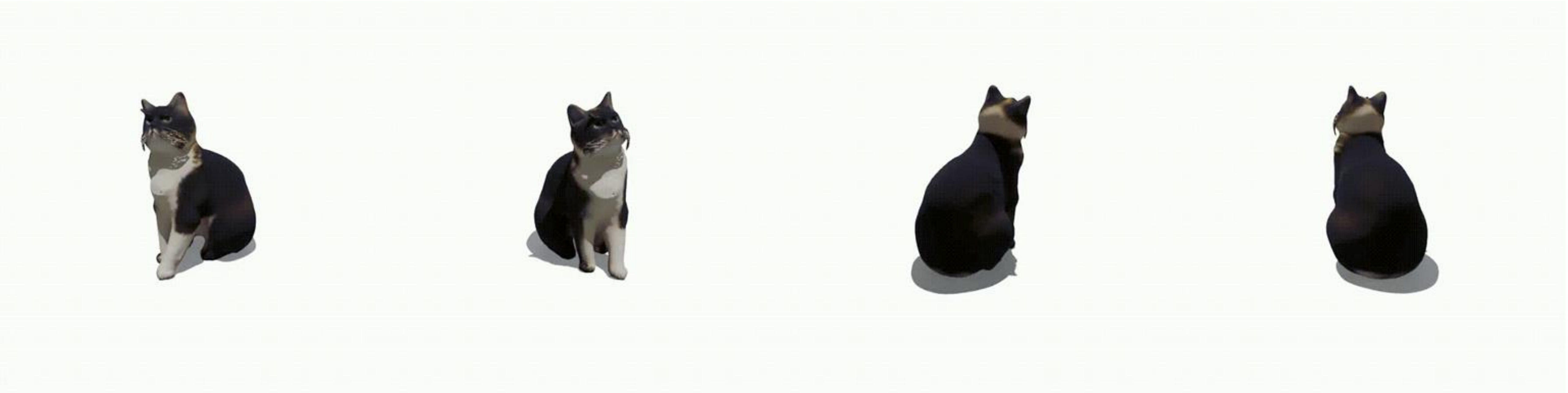
4Deform: Neural Surface Deformation for Robust Shape Interpolation

4Deform: Neural Surface Deformation for Robust Shape Interpolation



Input images

Input images

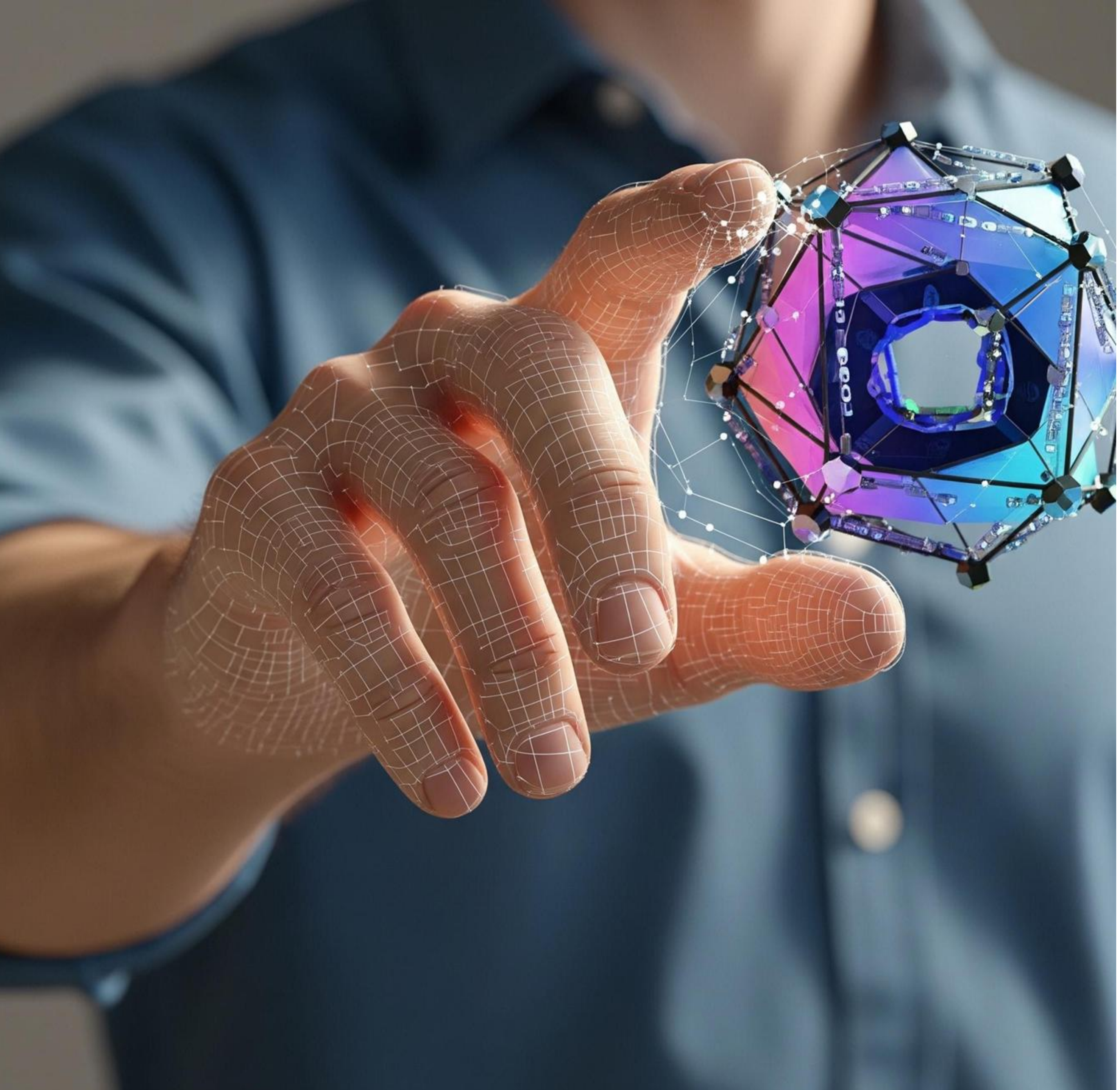


Input images

Input images

Sources and references:

- Learning Human Bodies in Motion
  https://bodymodelling.is.tuebingen.mpg.de/
- FAUST Dataset and Challenge:
  https://faust-leaderboard.is.tuebingen.mpg.de/
- Virtual Humans Tuebingen Course:
  https://www.youtube.com/watch?v=DFHuV7nOgsI&list=PL05umP7R6ij13it8Rptqo7lycHozvzCJn
- Human + cloths:
  - CAPE: https://cape.is.tue.mpg.de/
  - ETCH: https://boqian-li.github.io/ETCH/
  - 4D-Dress: https://eth-ait.github.io/4d-dress/
- Project page of SMPL: https://smpl.is.tue.mpg.de/
- SMPL made simple: https://smpl-made-simple.is.tue.mpg.de/
- Meshcapade Wiki: https://meshcapade.wiki/



Geometric Deep Learning for Virtual Humans

Virtual Humans

Riccardo Marin

21st November 2025