Virtual Humans
Under a Shape Analysis Spotlight

https://github.com/riccardomarin/HumanAnalysis

Riccardo Marin

26th November 2025
STAG26

About me

https://riccardomarin.github.io/

Verona
PhD Computer Science

Paris
École Polytechnique
(Visiting)

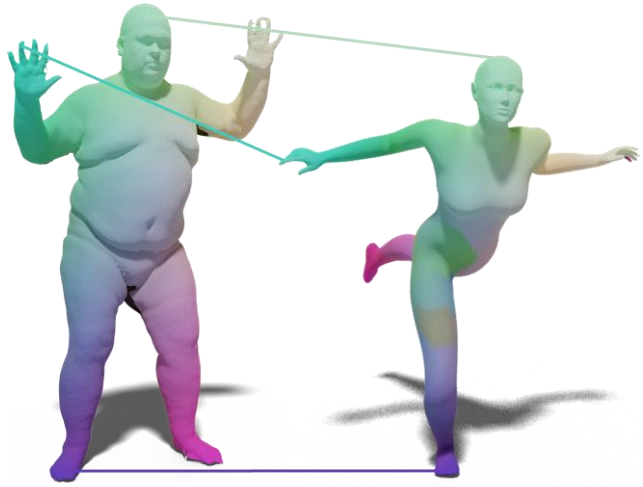
Rome
Sapienza
(Postdoc)

Tuebingen
AI Center
(Postdoc)

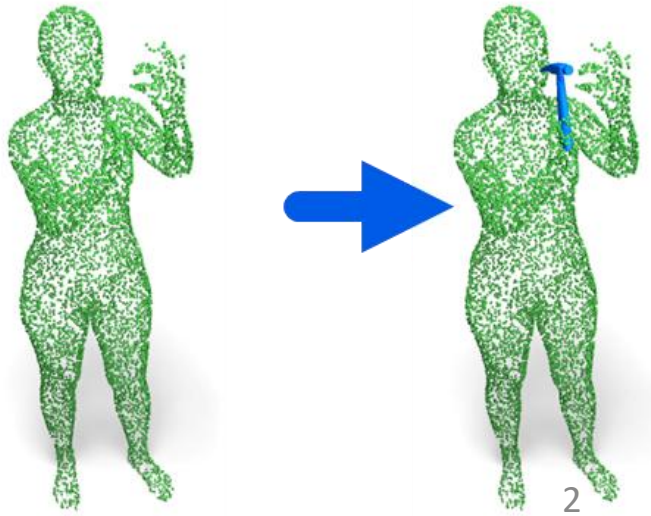
Munich
TUM/MCML
(Postdoc/Professor)

Research Topics:

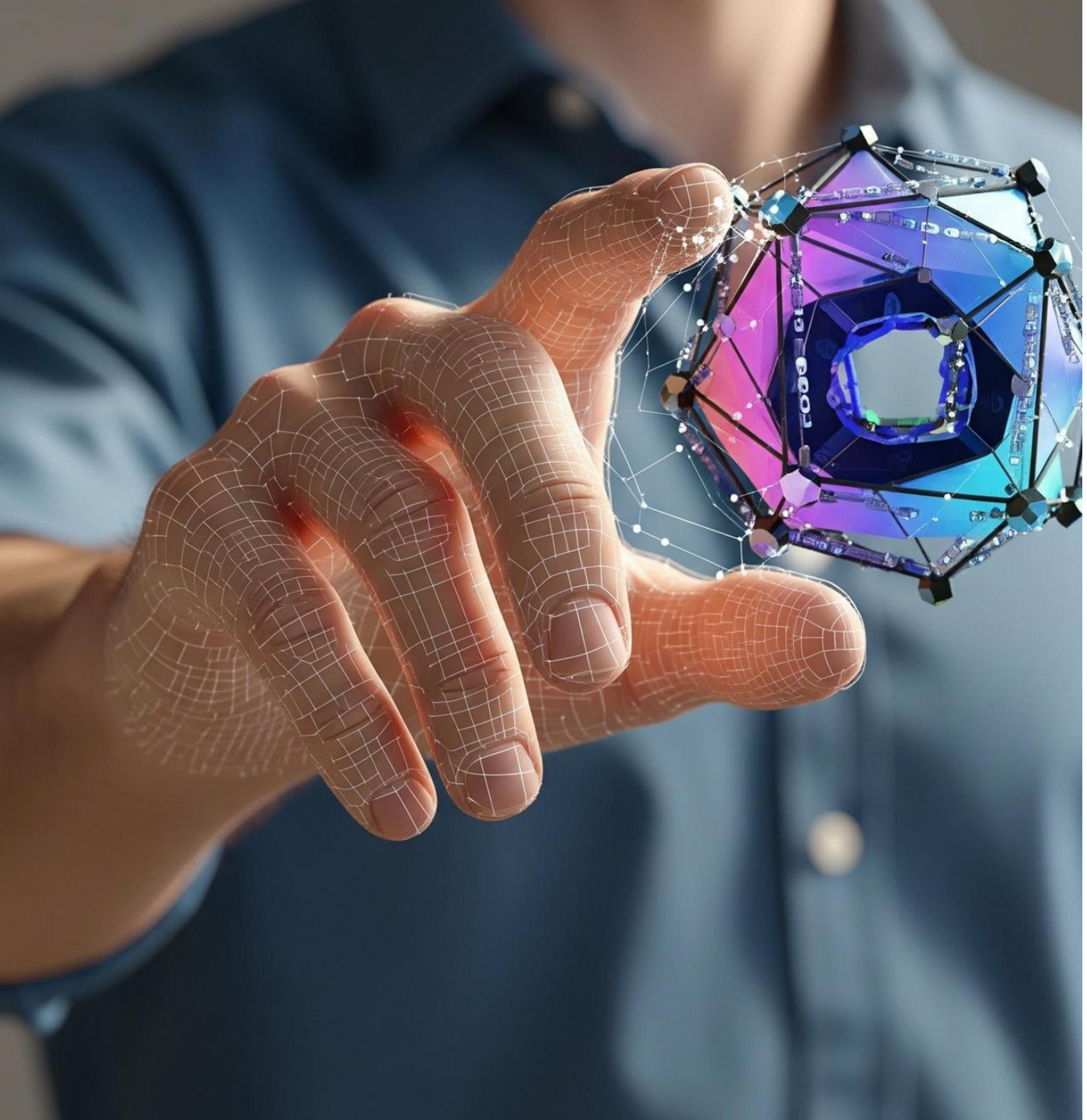
3D Matching/Shape Analysis



Virtual Humans




STAG 2018
Brescia, October 18-19, 2018



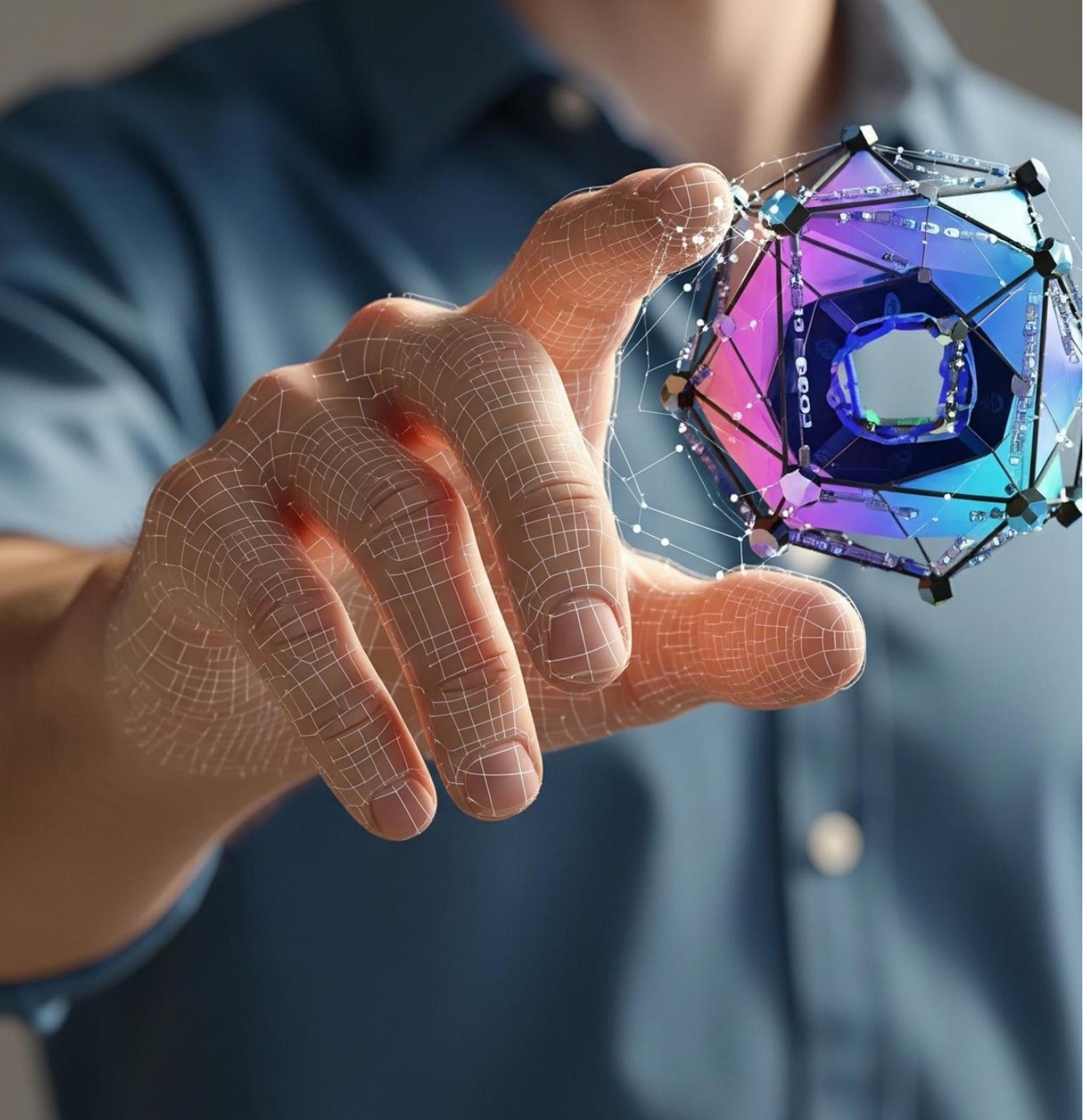
Schedule

Body Models

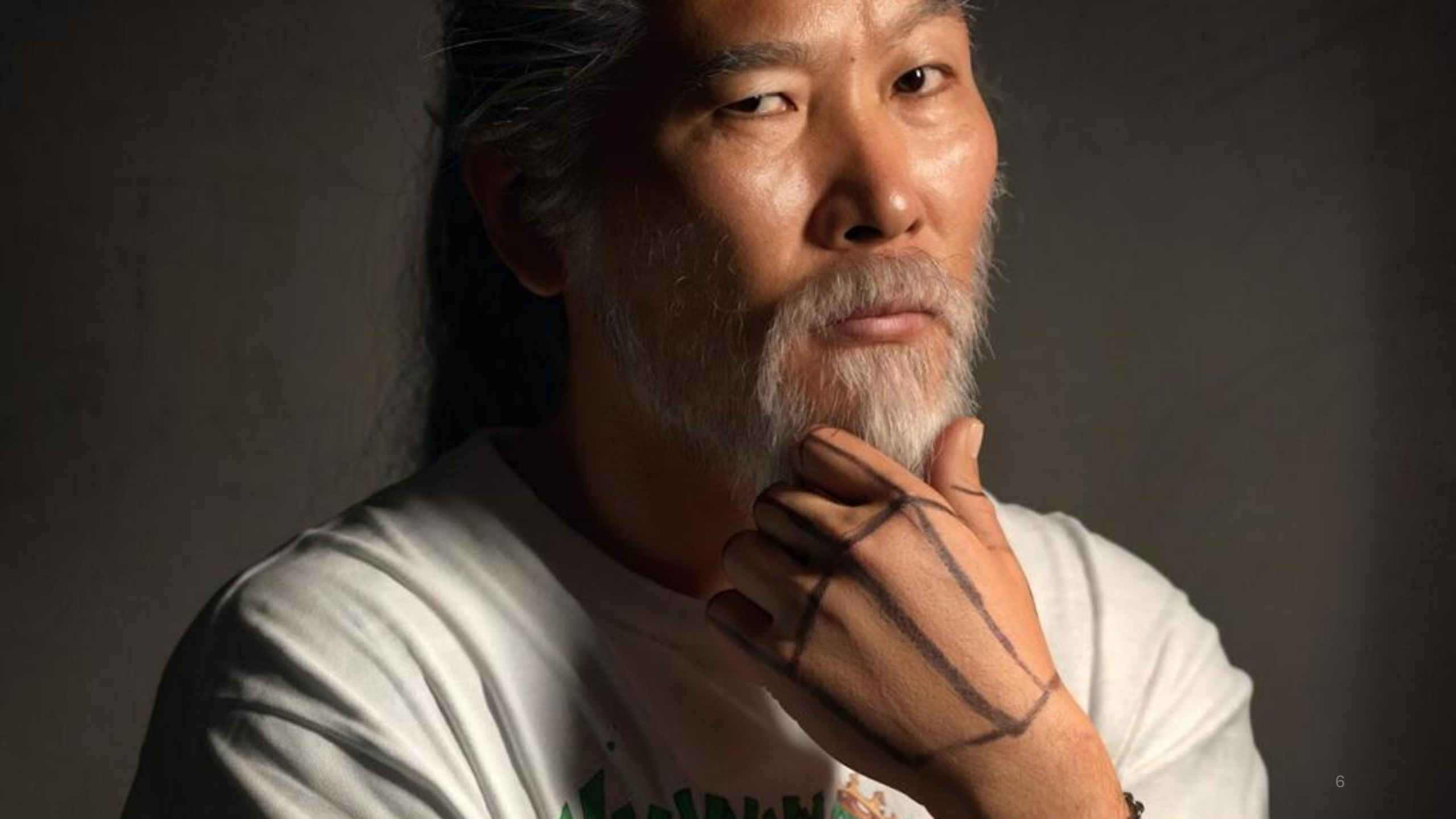
Spectral Shape Analysis

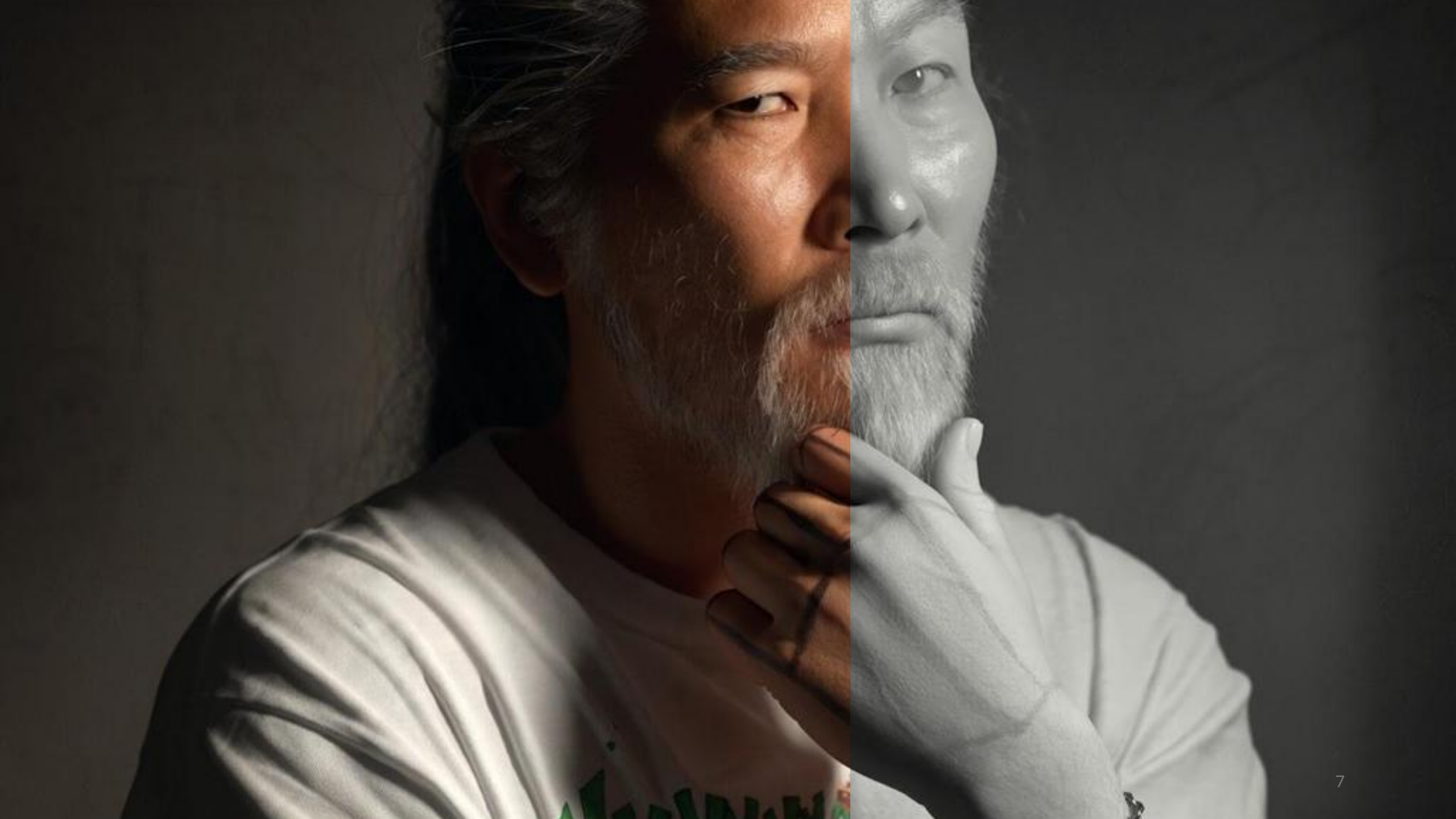
Humans Registration

Beyond Isolated Humans



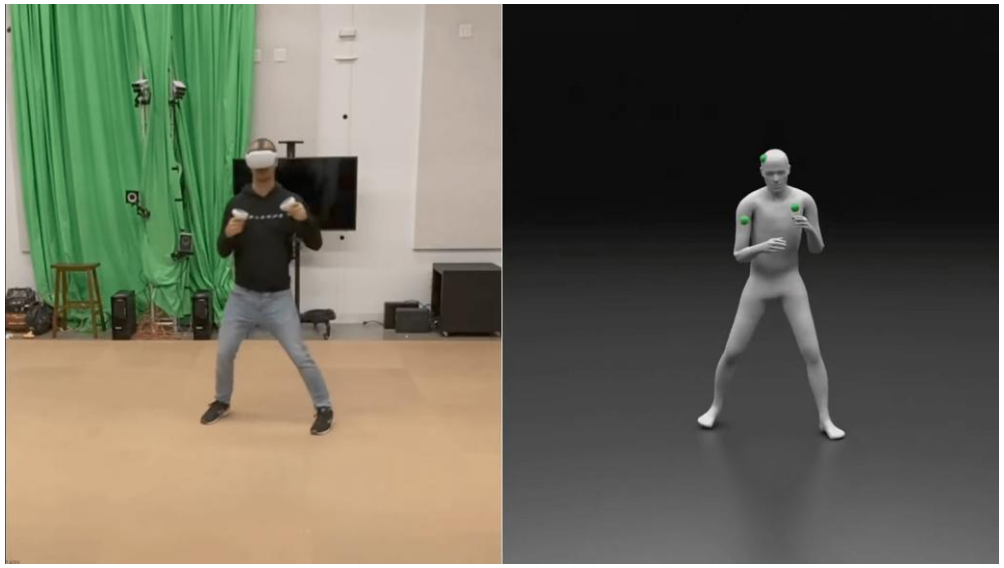
Body models



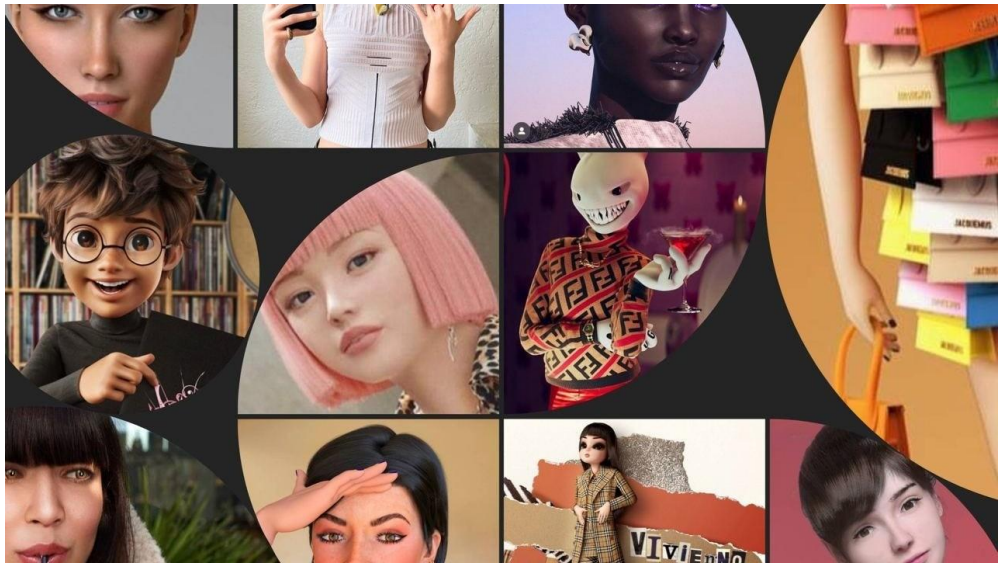


https://www.artstation.com/ianspriggs

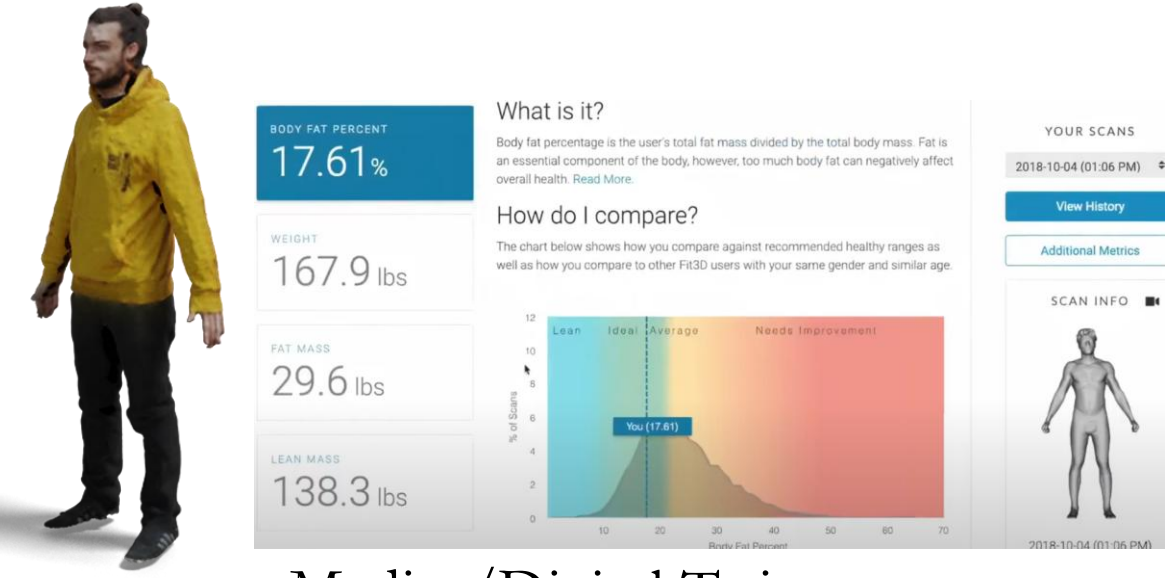
What are Virtual Humans?



VR/XR



Virtual Characters

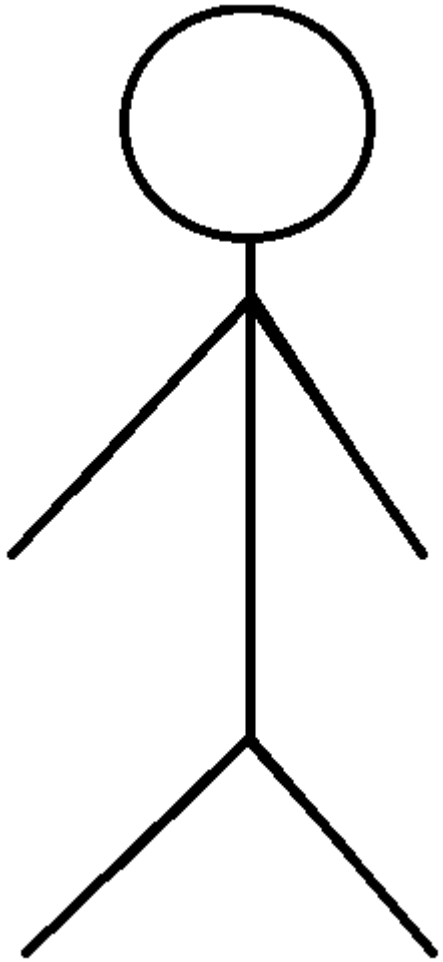


Medics/Digital Twin

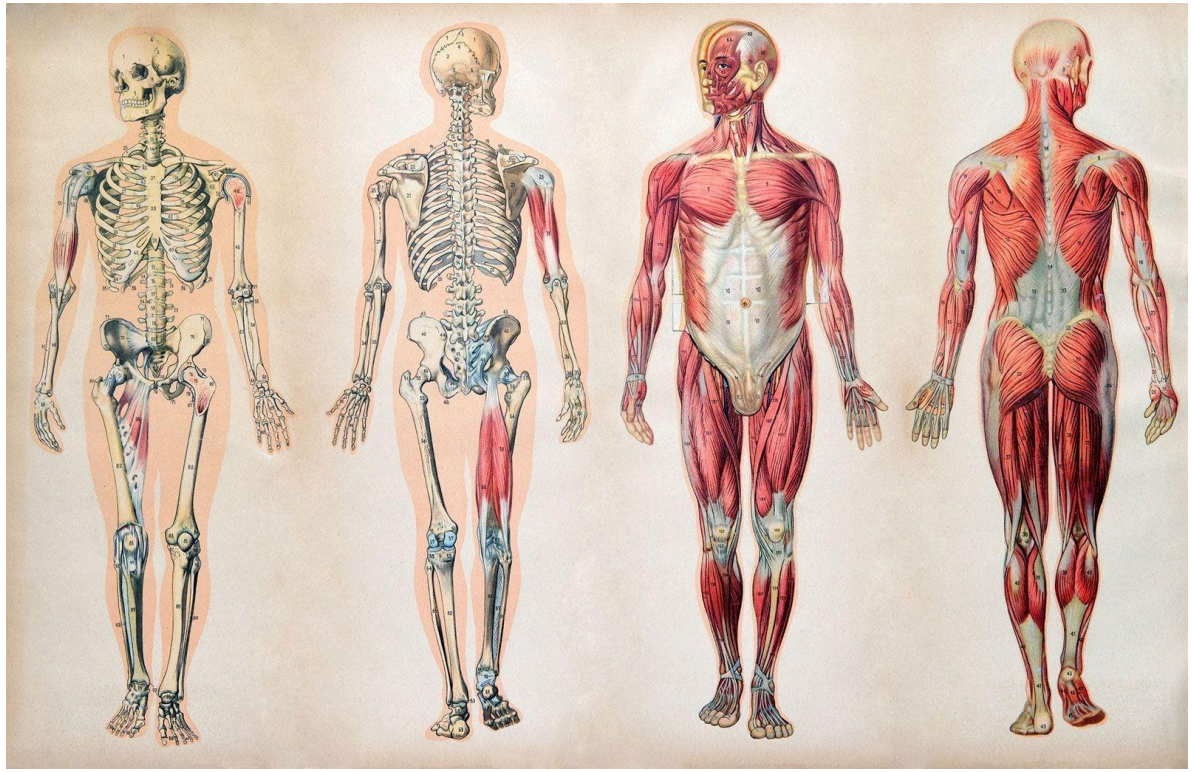
Assistance and Therapy

What is a good representation for human data?

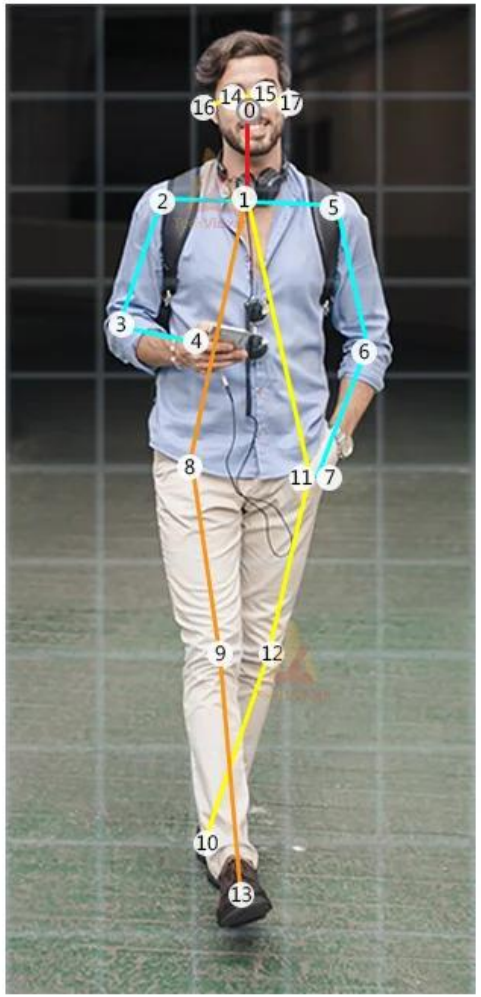
5 lines
1 circle

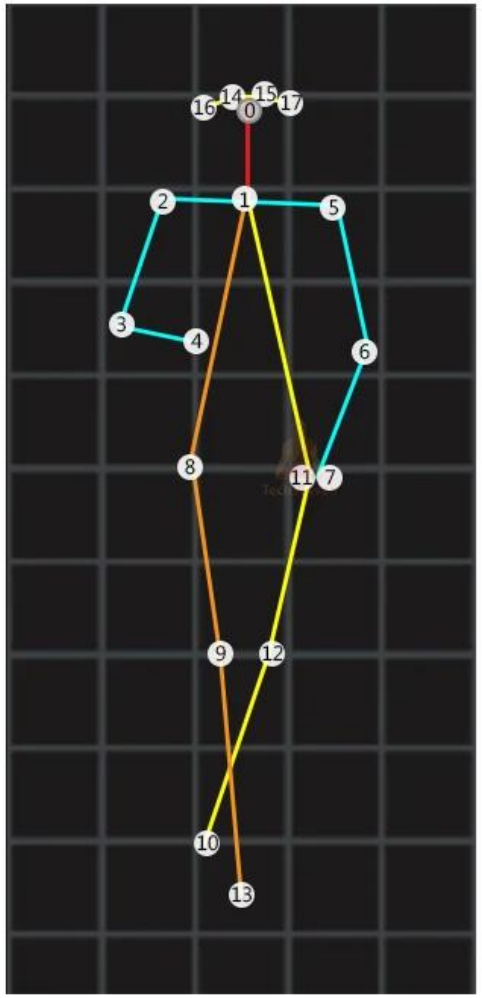


206 bones
600+ muscles
78 organs
…

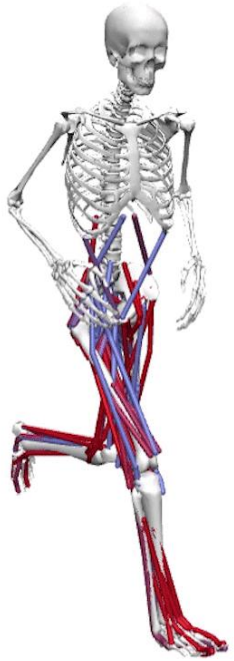


What is a good representation for human data?





https://github.com/nitingour1203/human_pose_detection



BASH, Schleicher et al., 2021

The roles of a representation

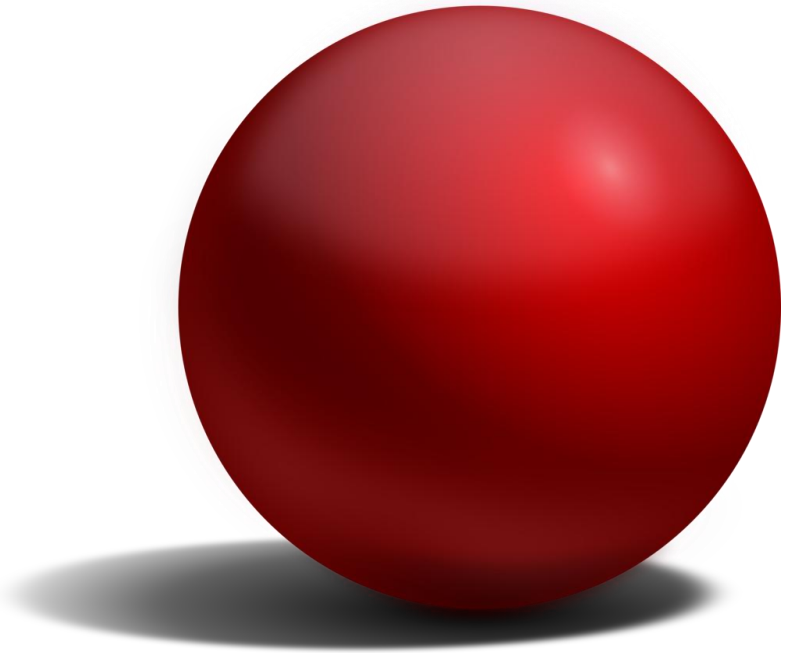
Convey the geometry

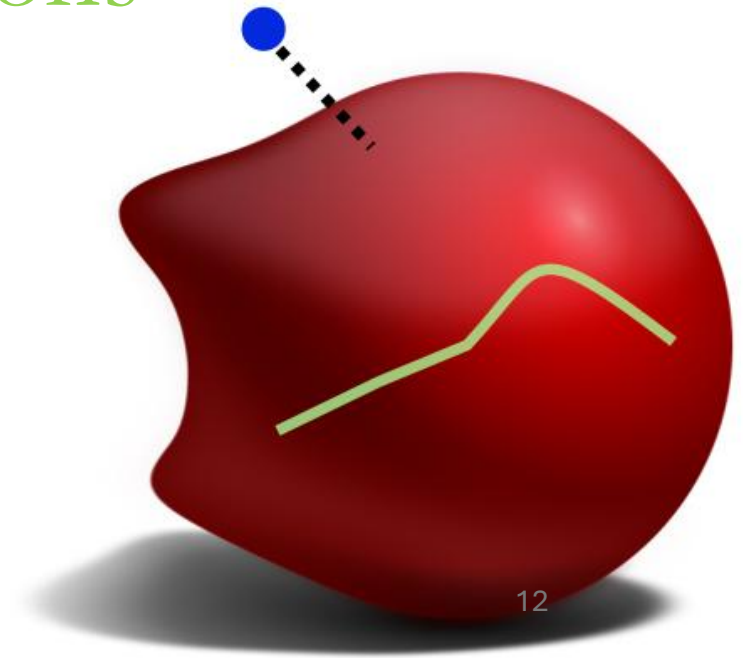
Support computations

Geometry Evaluations

Spatial Queries

Modifications

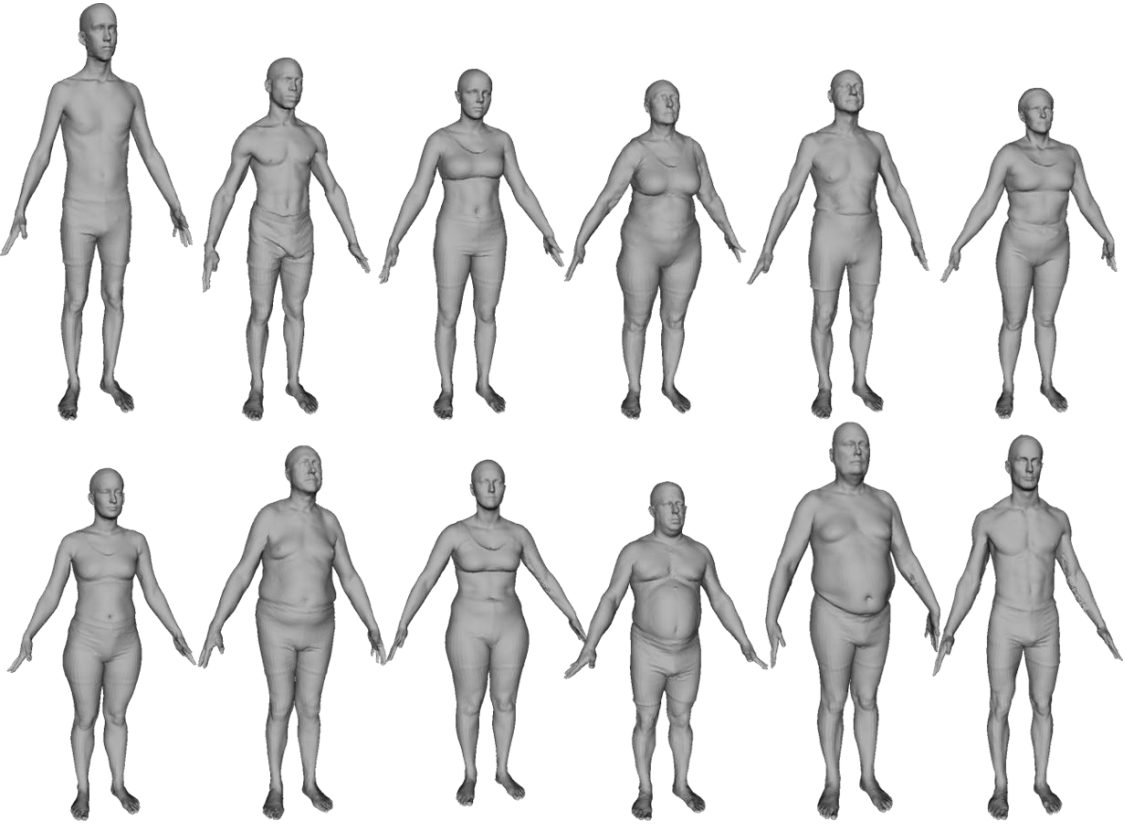




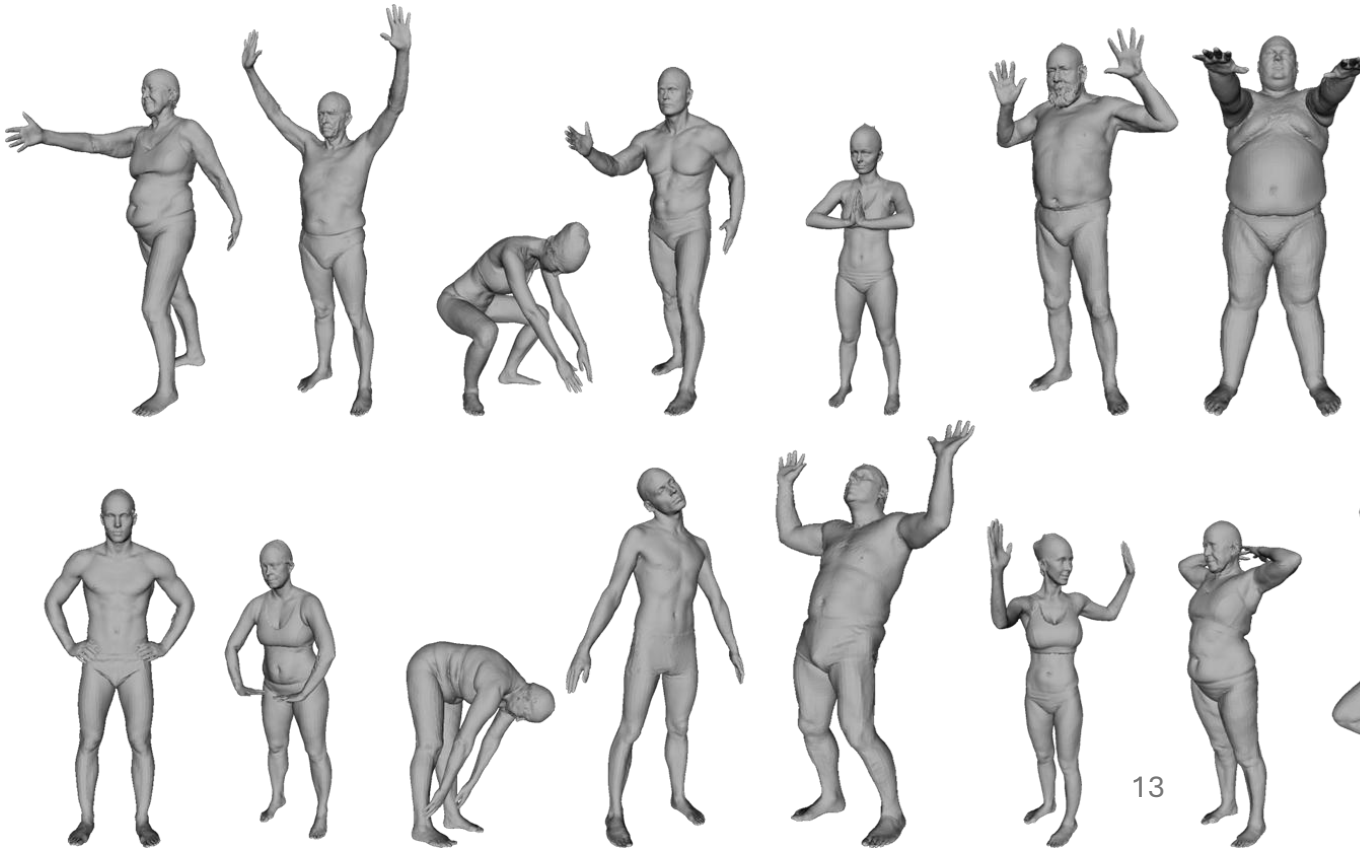
What geometry should a body representation convey?

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

Identities



Poses



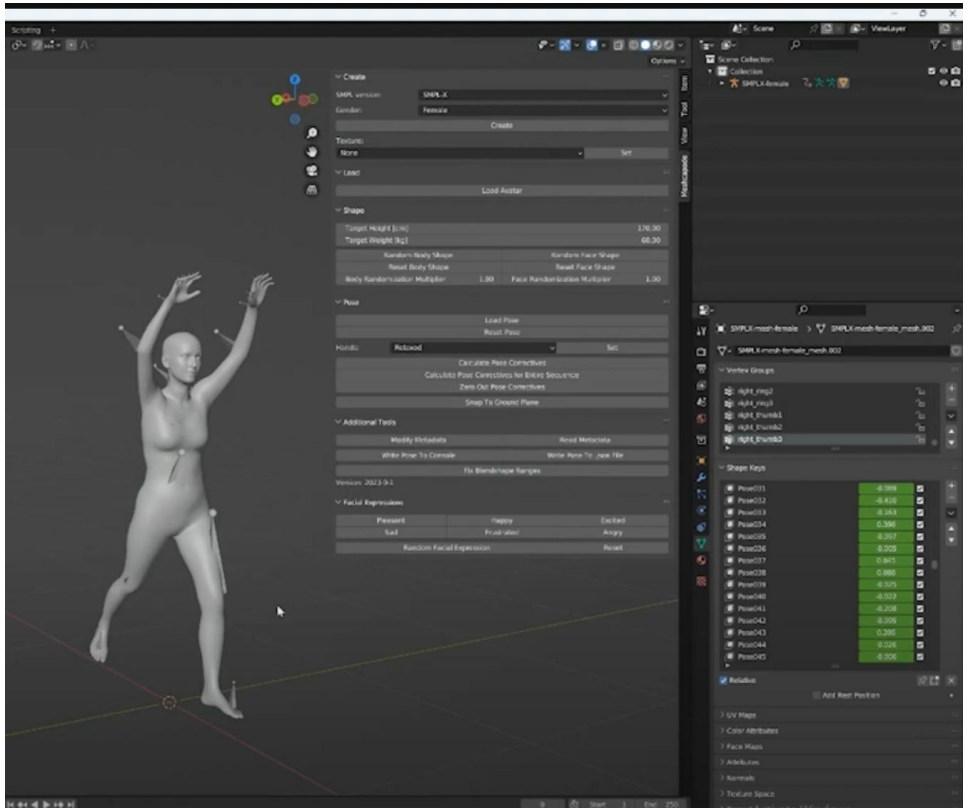
What operations should a body representation support?

Application oriented

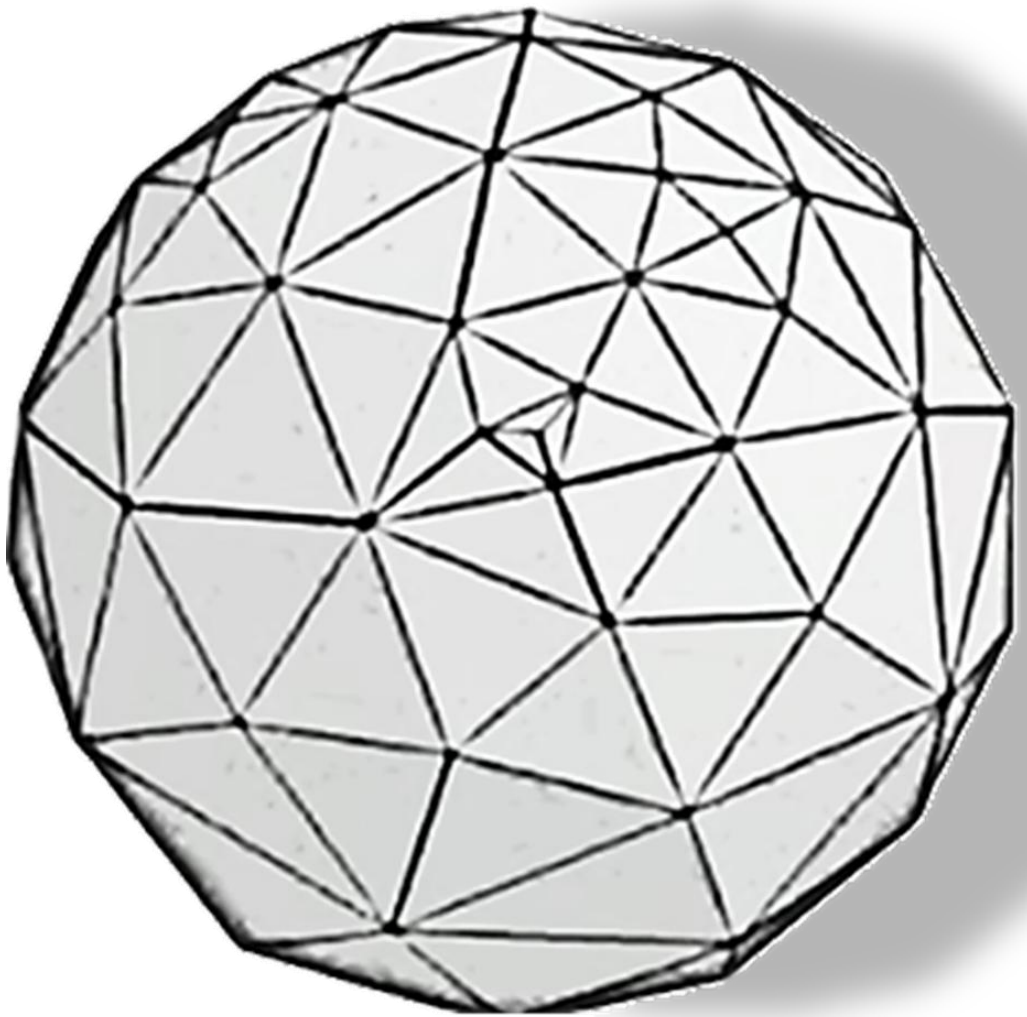
Optimization

https://smplify.is.tuebingen.mpg.de/

Software integration

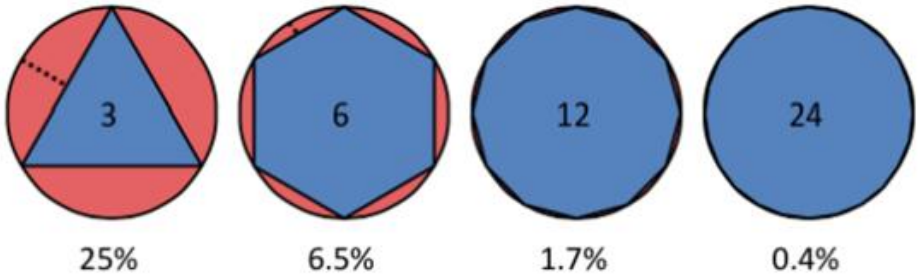


https://medium.com/meshcapade/building-human-foundation-models-with-smpl-a5251bb294fd



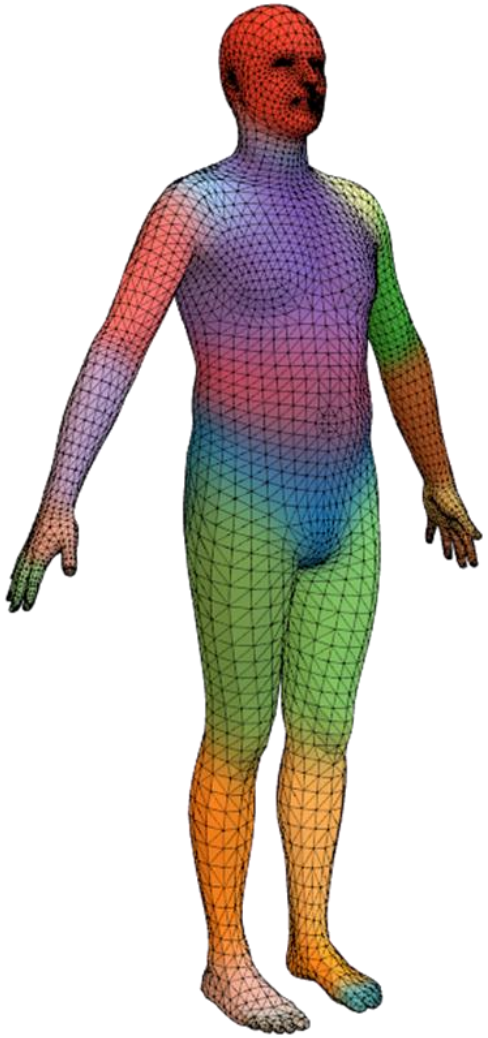
Piecewise **linear** approximation

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**

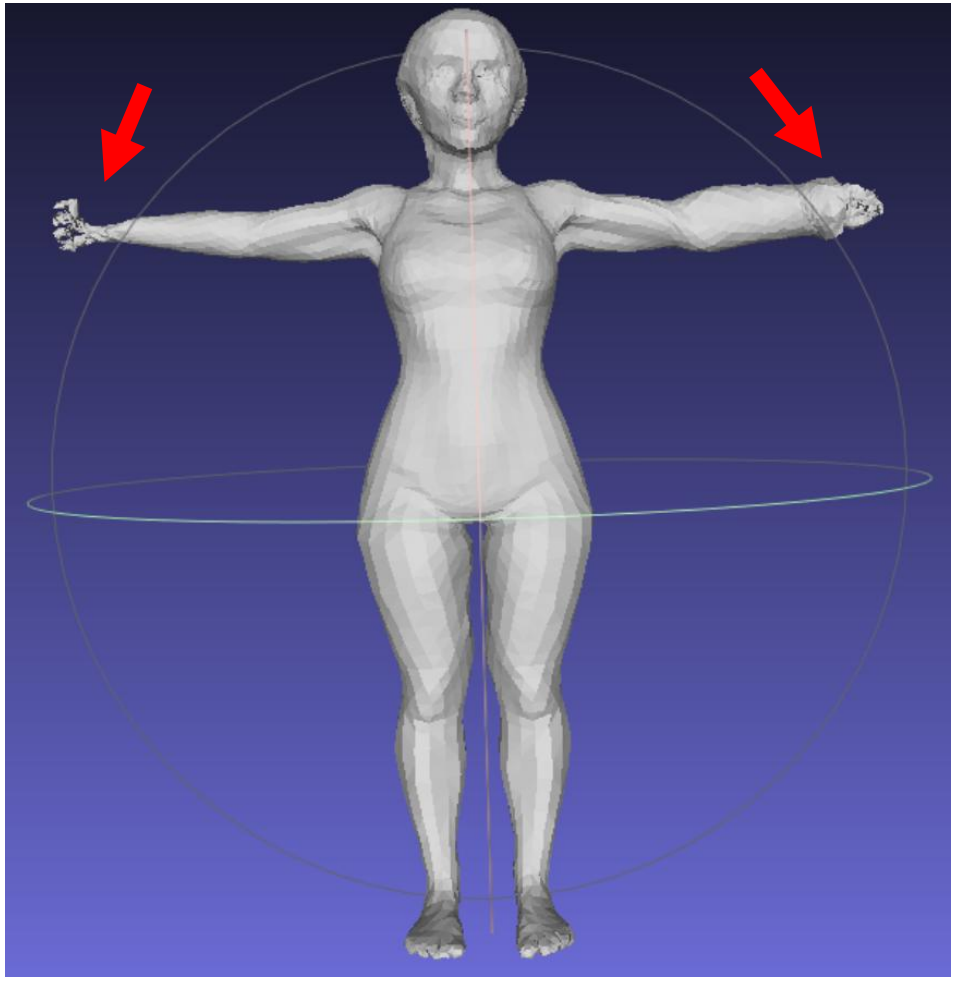
Bodies as meshes?



We represent the body as its **surface**

A discretized set of thousands of vertices and faces

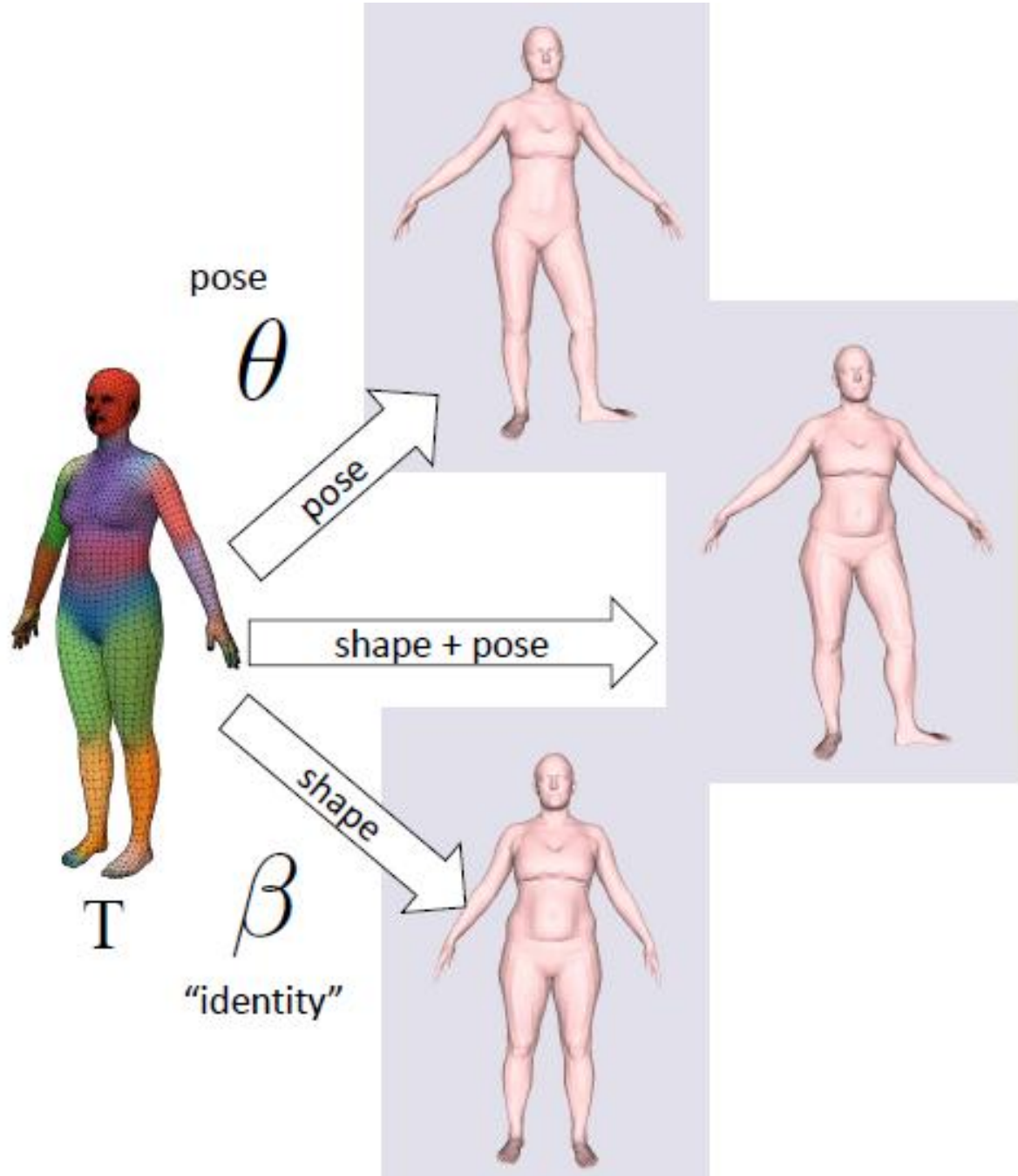
Problem: editing the identity and pose requires to specify tens of thousands of values. And not all of their combinations are realistic.



Factored model

A model parameterizes "deviations" from a template mesh.

Simplifies learning and inference.



SCAPE: Anguelov et al., SIGGRAPH 2005

SCAPE: Shape Completion and Animation of People

Dragomir Anguelov* Praveen Srinivasan* Daphne Koller* Sebastian Thrun* Jim Rodgers*
Stanford University

James Davis†
University of California, Santa Cruz

Figure 1: Animation of a motion capture sequence taken for a subject, of whom we have a single body scan. The muscle deformations are synthesized automatically from the space of pose and body shape deformations.

Using the process above, we obtained two data sets: a *pose data set*, which contains scans of 70 poses of a particular person in a wide variety of poses, and a *body shape data set*, which contains scans of 37 different people in a similar (but not identical) pose. We also added eight publicly available models from the CAESAR data set [Allen et al. 2003] to our data set of individuals.

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

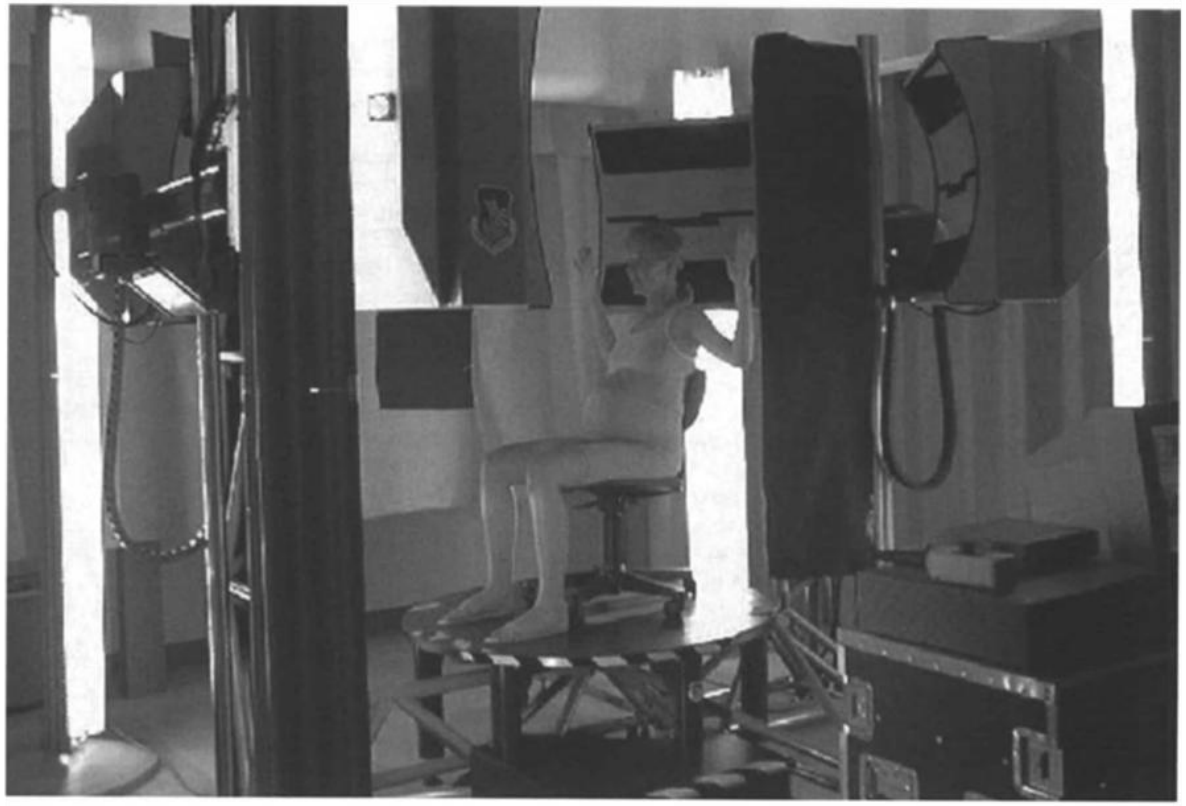


Figure 2.
WB4 Whole Body Scanner (Cyberware) in use during study of pregnant women.

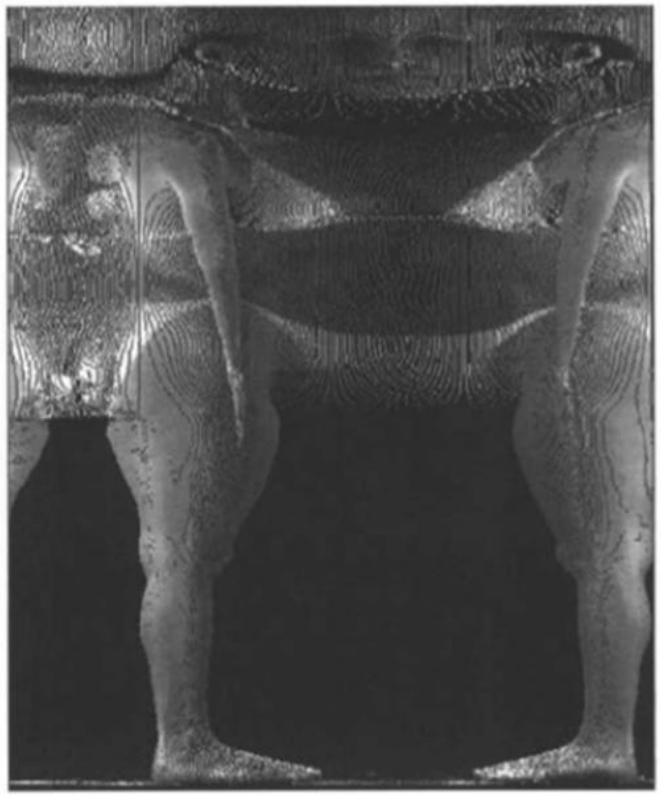
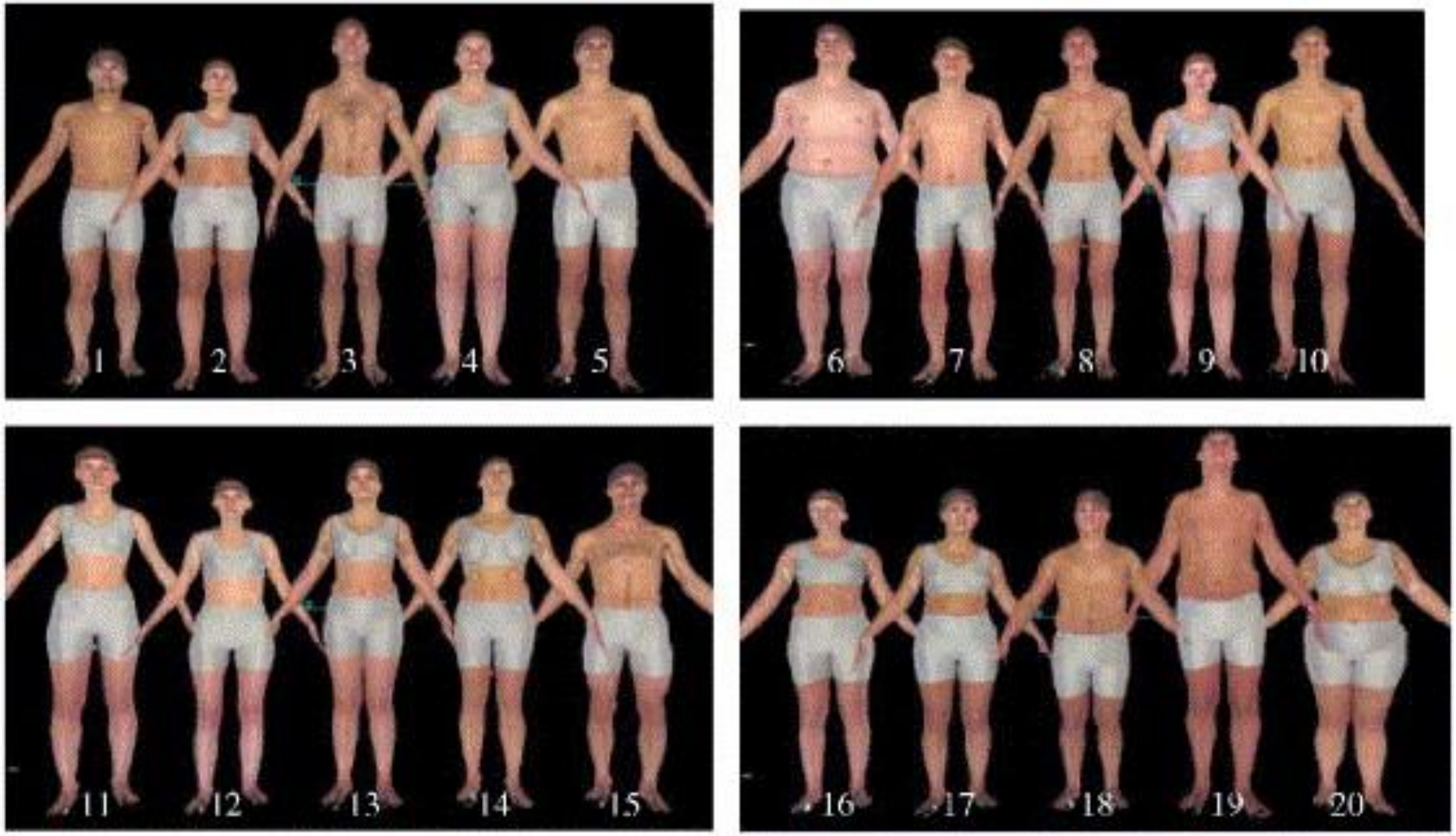


Figure 1.
Flat map of a 3D woman.

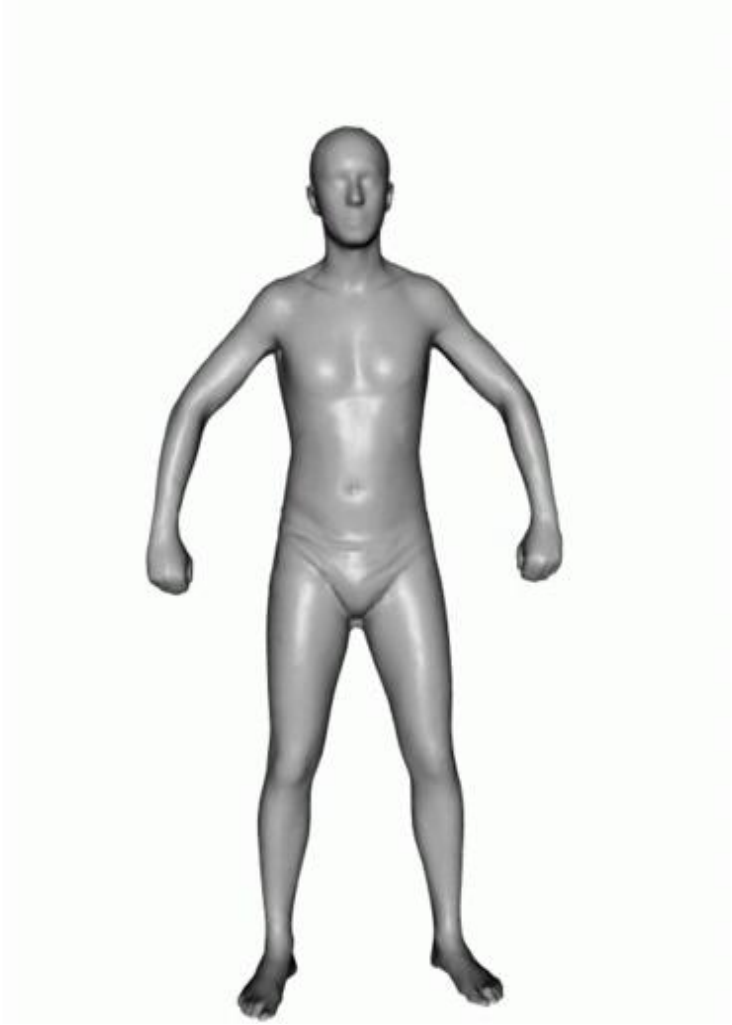
Why 3D?

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-

CAESAR Dataset



SCAPE: Shape Completion and Animation of People



Issues:

1) Learned from a **few data**, not much realistic
2) Based on **triangles deformation**
3) Combination of **part-based and local rotations**
4) Requires **stitching terms**
5) **Difficult to optimize**

Given a set of transformation matrices Q and R associated with a pose instance, our method's predictions can be used to synthesize a mesh for that pose. For each individual triangle, our method makes a prediction for the edges of p_k as $R_k Q_k \hat{v}_{k,j}$. However, the predictions for the edges in different triangles are rarely consistent. Thus, to construct a single coherent mesh, we solve for the location of the points $y_1, \ldots, y_M$ that minimize the overall least squares error:

$$\underset{y_1,\ldots y_M}{\operatorname{argmin}} \sum_k \sum_{j=2,3} \left\| R^i_{\ell[k]} Q^i_k \hat{v}_{j,k} - (y_{j,k} - y_{1,k}) \right\|^2 \qquad (2)$$

The SMPL (simple) body model

2015

SMPL: A Skinned Multi-Person Linear Model

Matthew Loper [*12] Naureen Mahmood[†1] Javier Romero[†1] Gerard Pons-Moll[†1] Michael J. Black[†1]

[1]Max Planck Institute for Intelligent Systems, Tübingen, Germany
[2]Industrial Light and Magic, San Francisco, CA

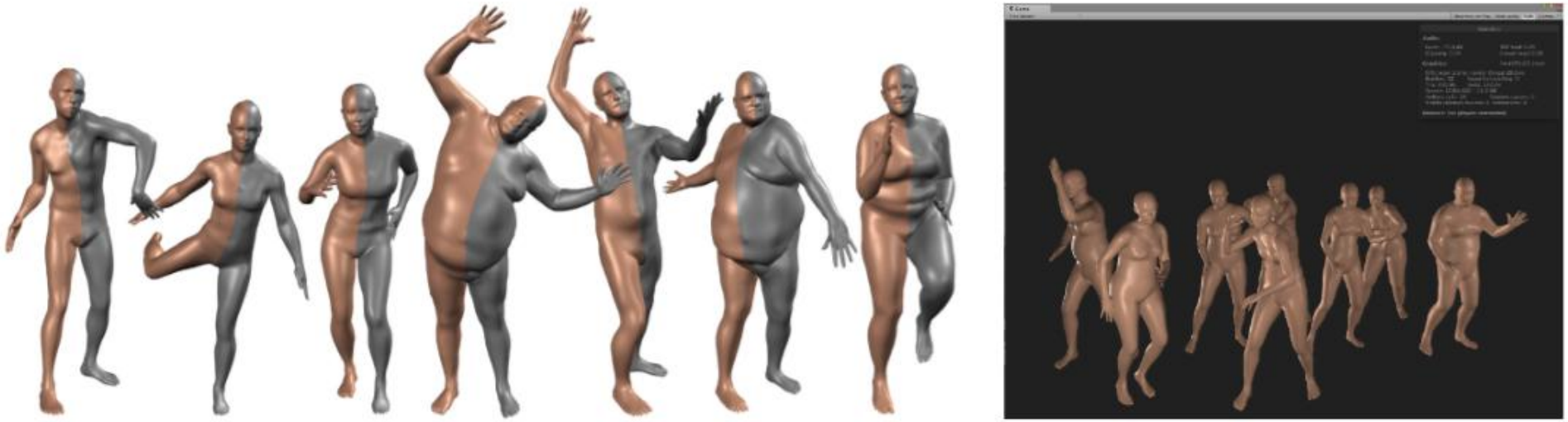
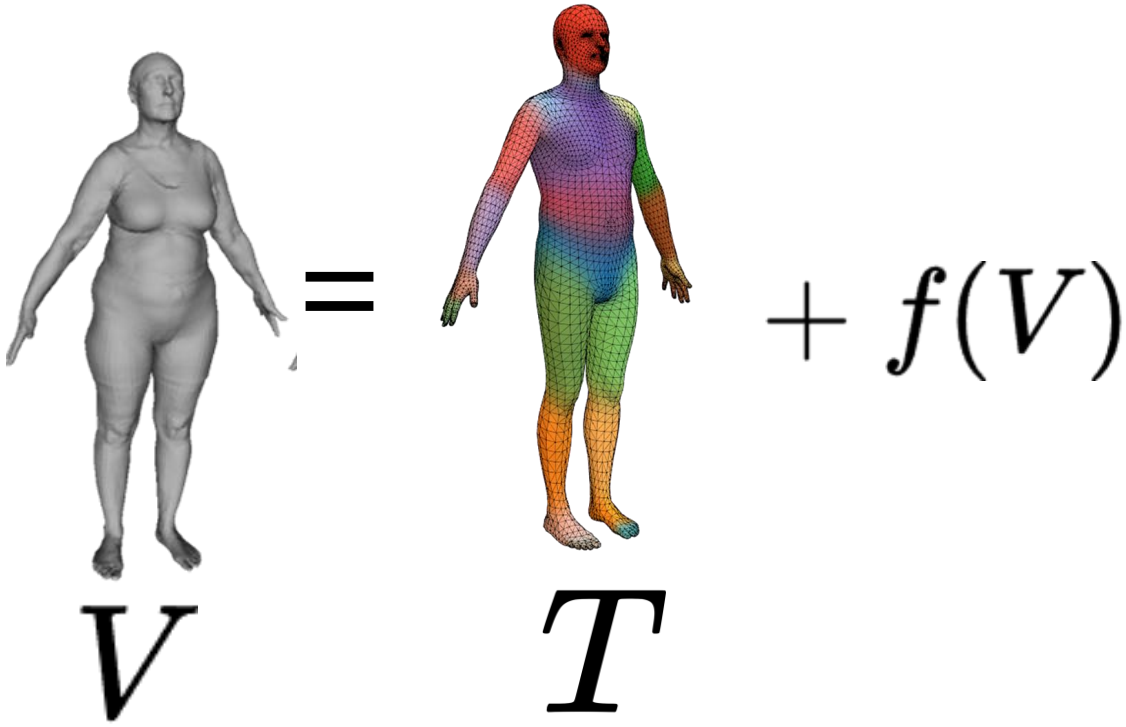


Figure 1: *SMPL is a realistic learned model of human body shape and pose that is compatible with existing rendering engines, allows animator control, and is available for research purposes. (left) SMPL model (orange) fit to ground truth 3D meshes (gray). (right) Unity 5.0 game engine screenshot showing bodies from the CAESAR dataset animated in real time.*

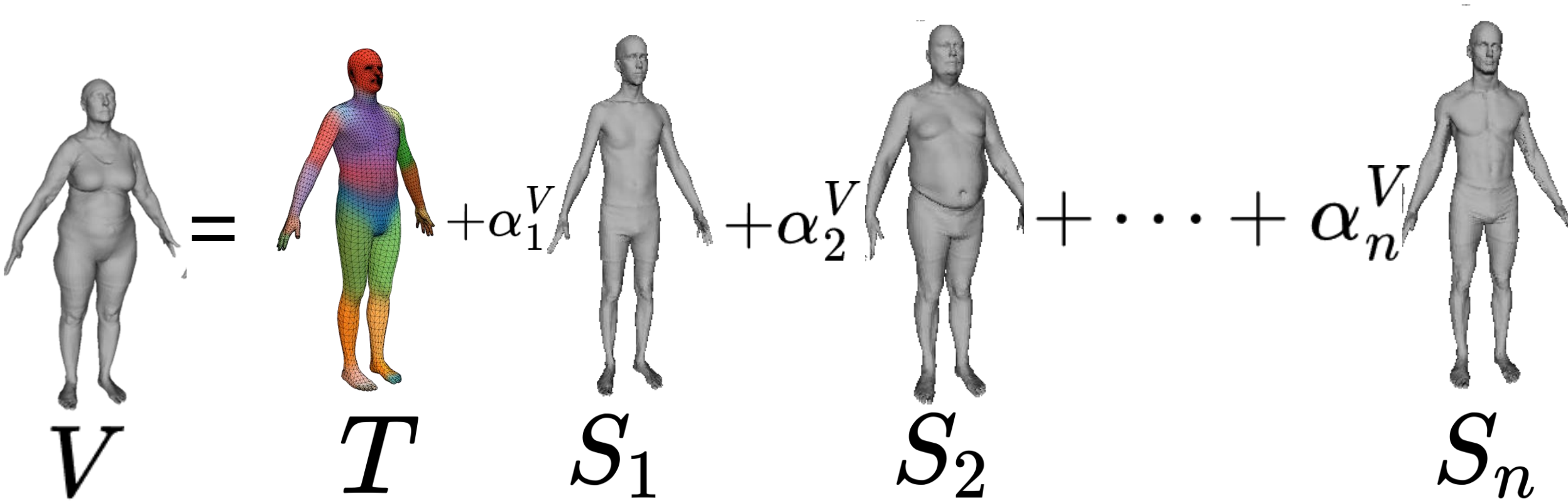
Desired properties

- **Simple**: controllable by only a few parameters (low-dimensional)
- **Realistic**: look and move as a real human
- **Differentiable**: usable in optimization problems
- **Applications-ready**: supported by rendering\graphics software

Morphing one body in different ones



Morphing one body in different ones - Linear Combination

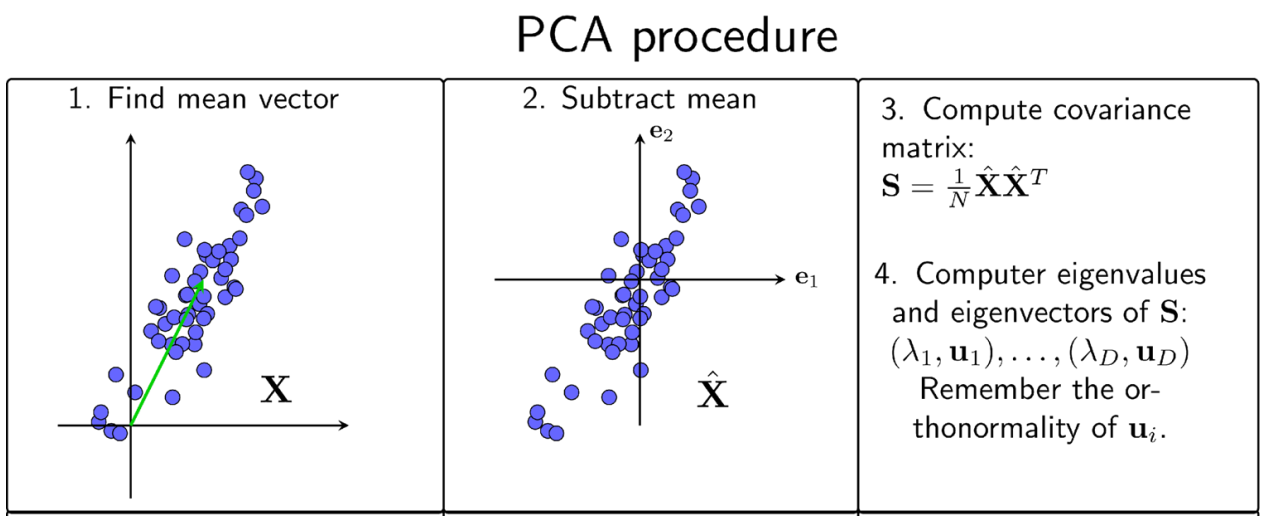


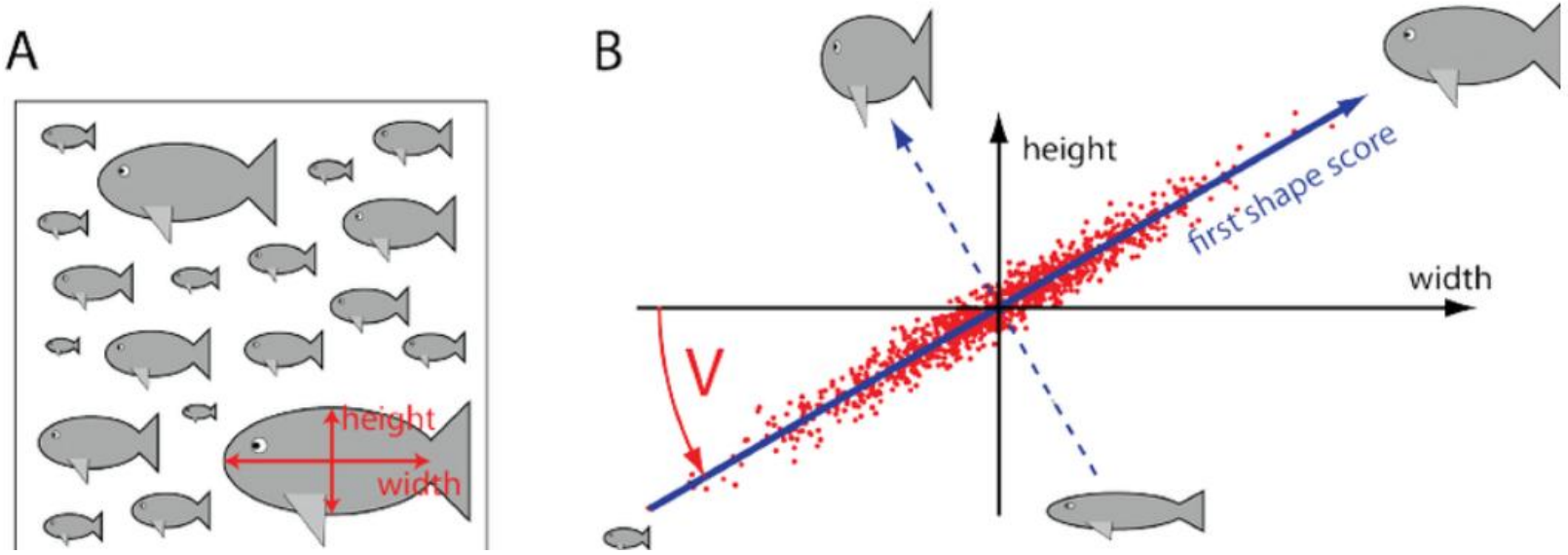
$$V = T + f(V) = T + \alpha_1^V(S_1) + \alpha_2^V(S_2) + \cdots + \alpha_n^V(S_n)$$

PCA in a nutshell

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

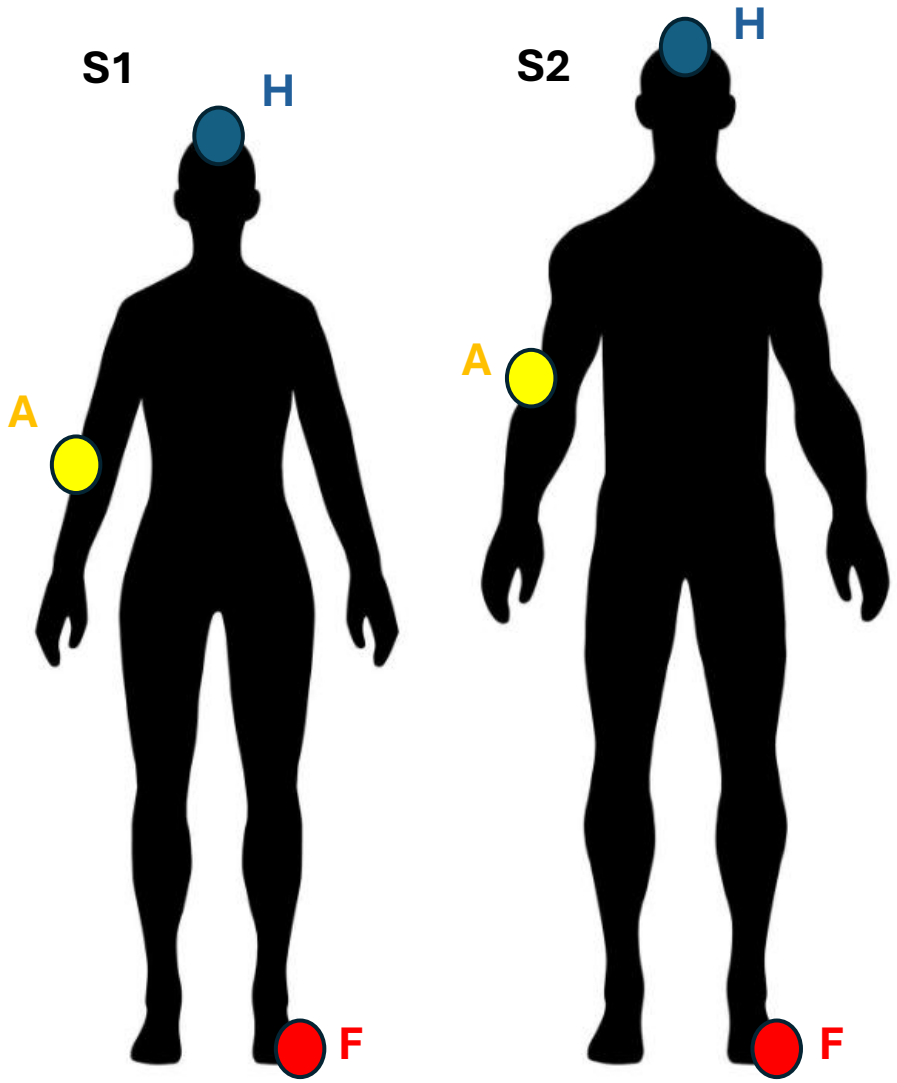
Given a dataset of samples with N correlated features, it returns K uncorrelated orthonormal vectors **ordered by explained variance**





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	
…	…	

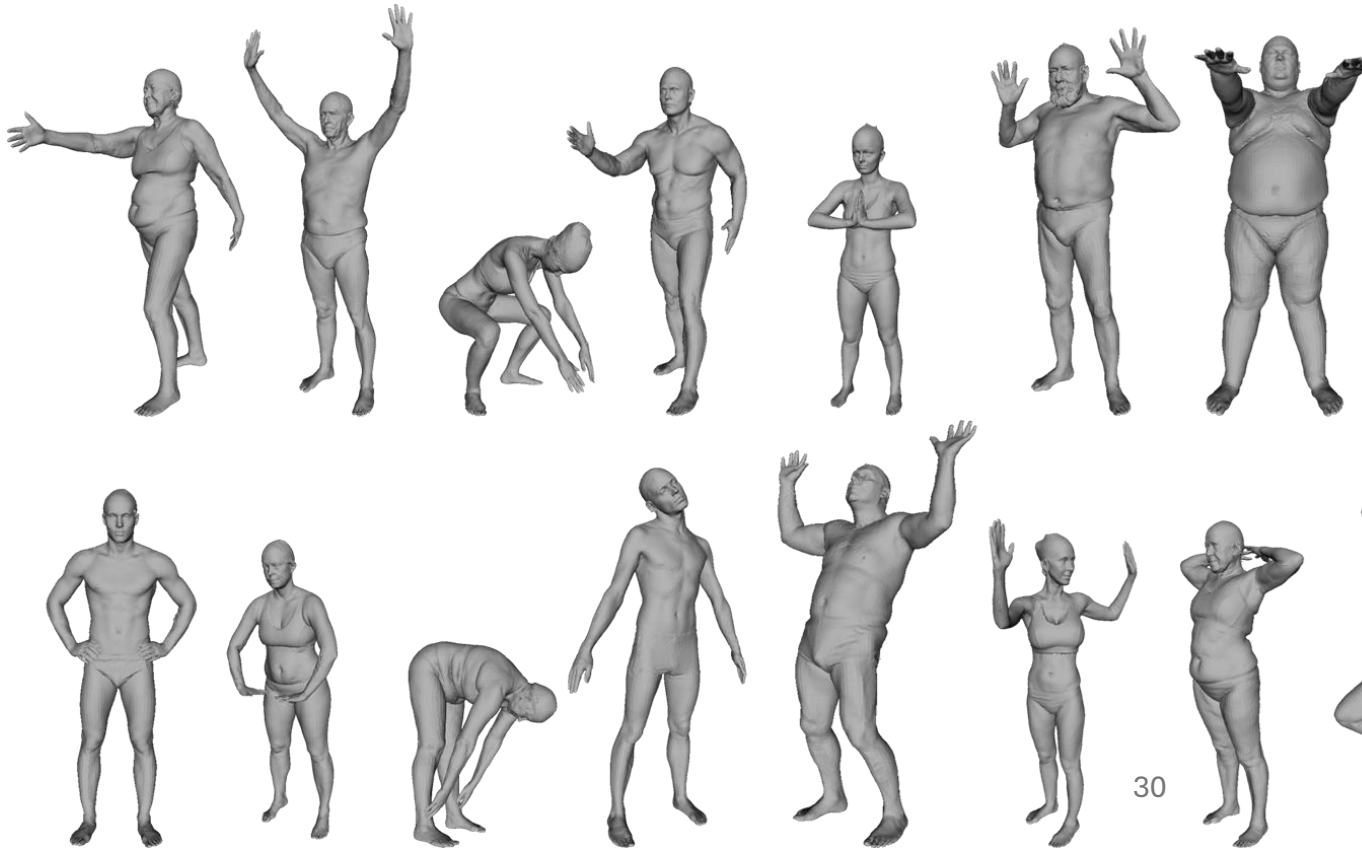
Learning a Model:

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

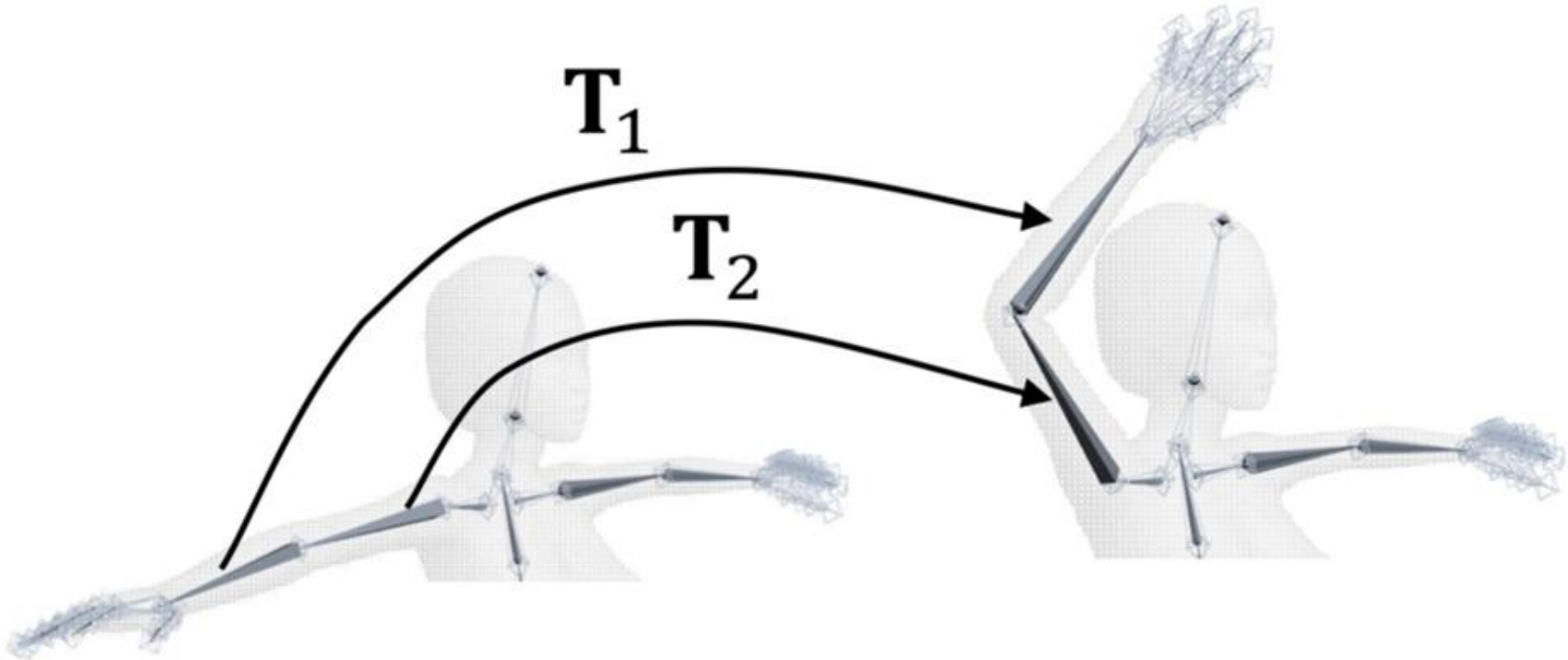
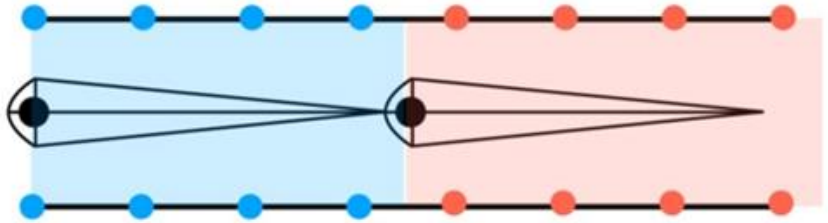


Image credit: Ladislav Kavan

Example: Rigid body Skinning

Vertices are rigidly attached to joints

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

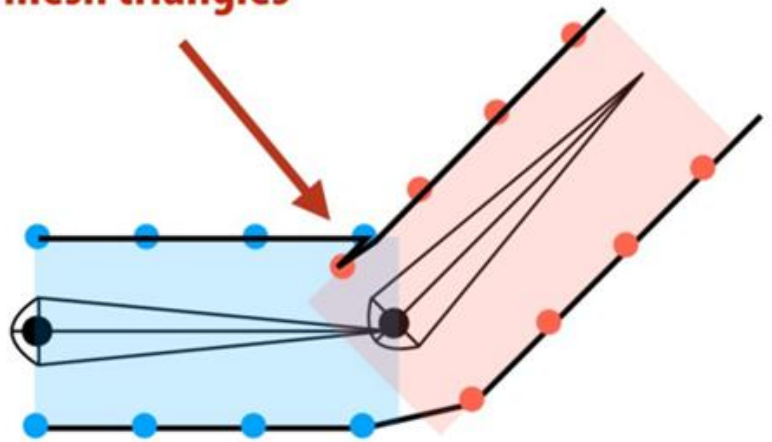


Original pose

Blue verts = associated with first joint

Red verts = associated with second joint

Interpenetration of mesh triangles



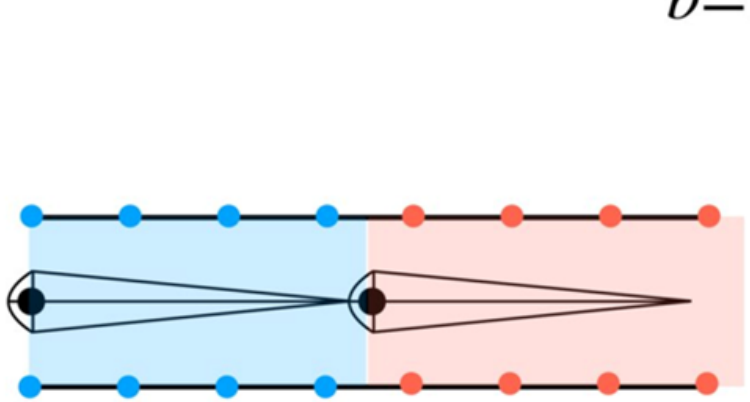
Vertices transforms according to corresponding joint transform (notice surface interpenetration)

Maya rigid bind example

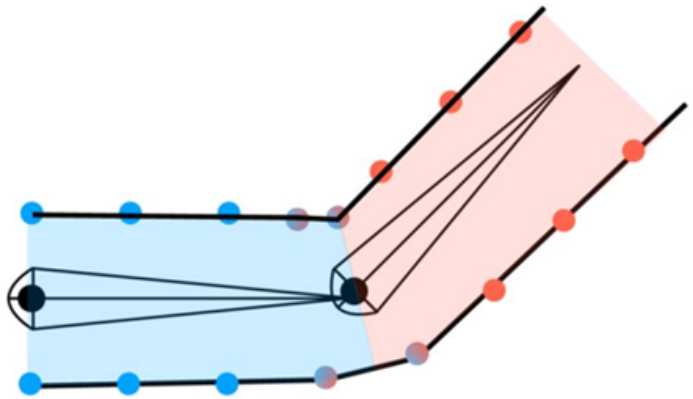
Linear Blend Skinning

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

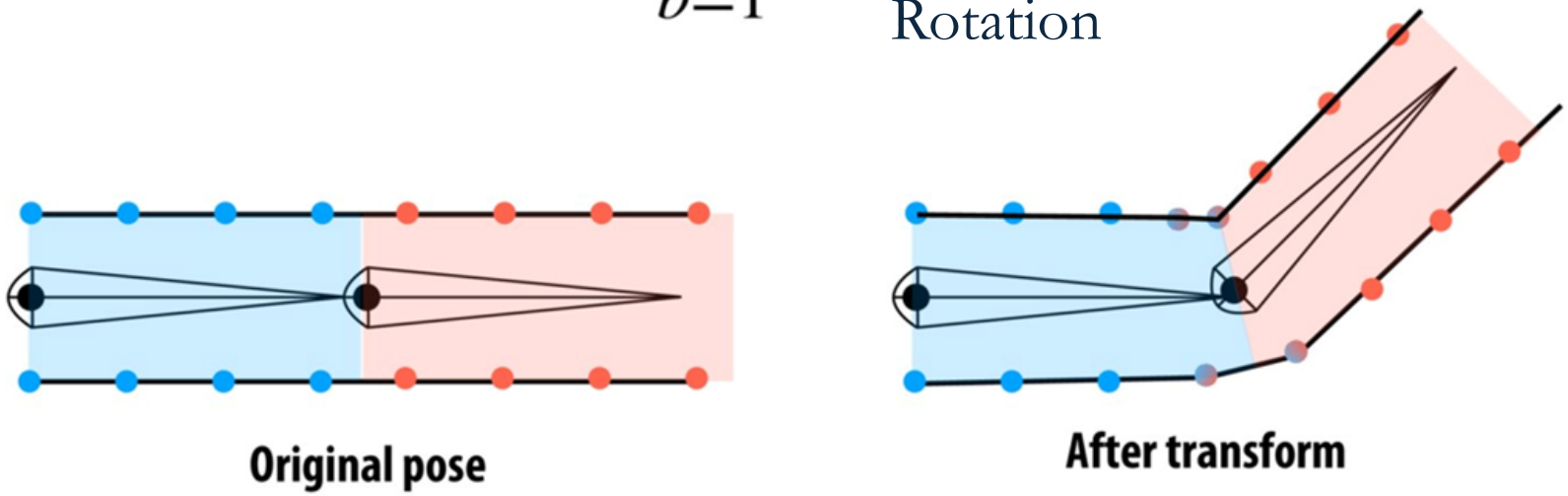
Transform mesh vertices according to linear combination of transforms for nearby skeleton joint

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

Blend weights

Template

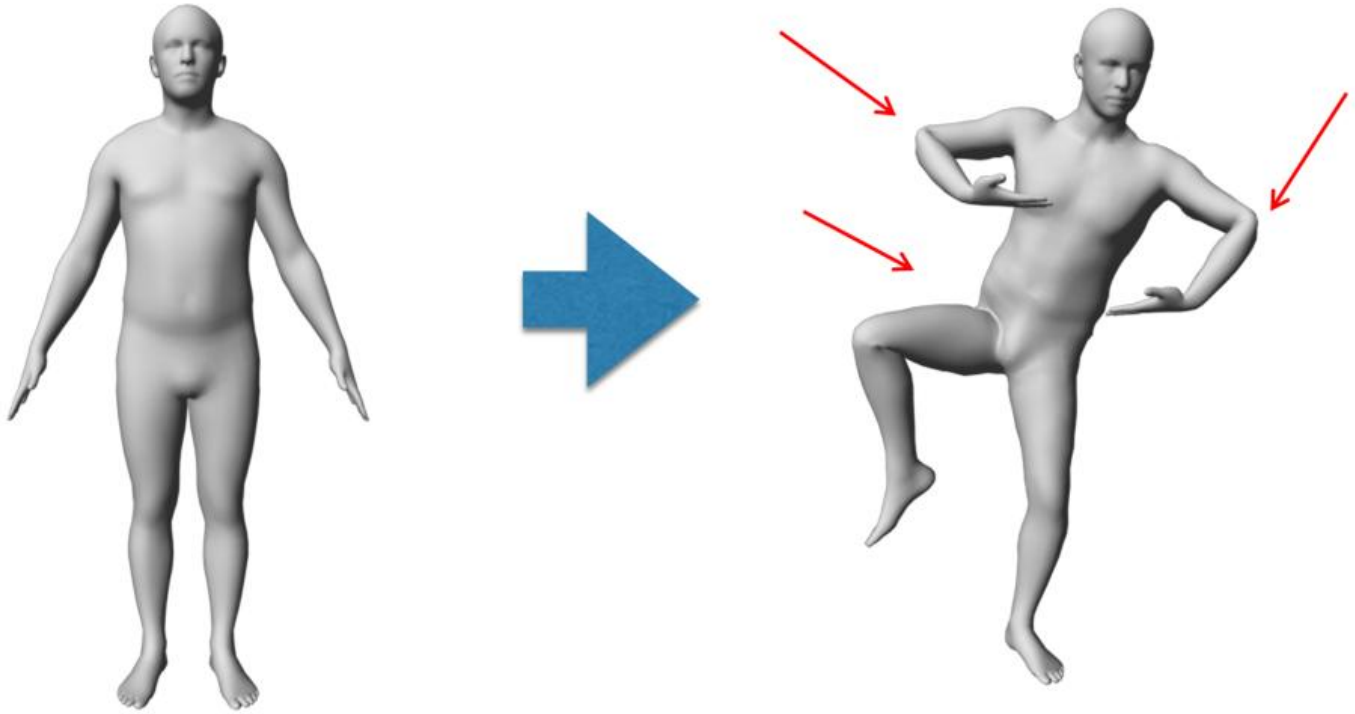
Rotation



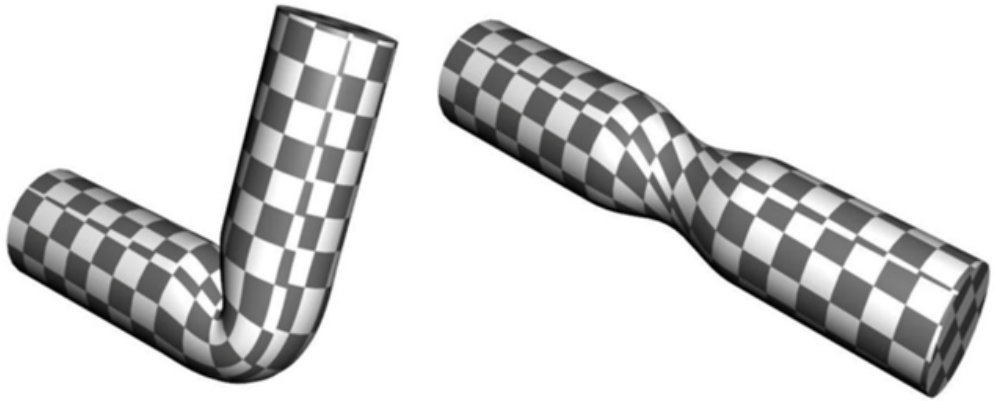
Original pose

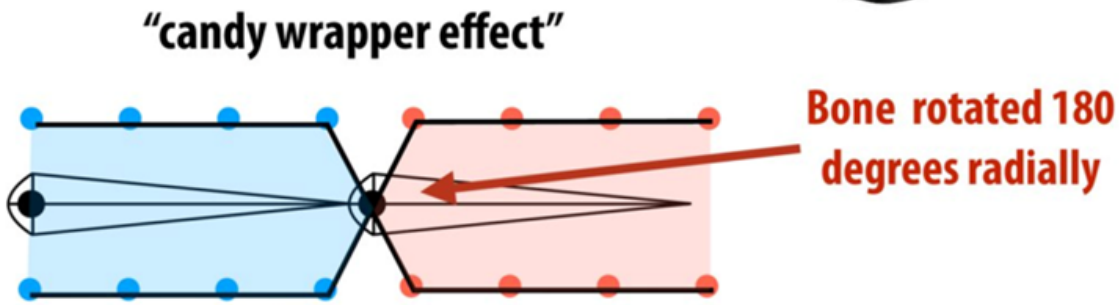
After transform

Problem: Large rotations cause artifacts



LBS loses volume





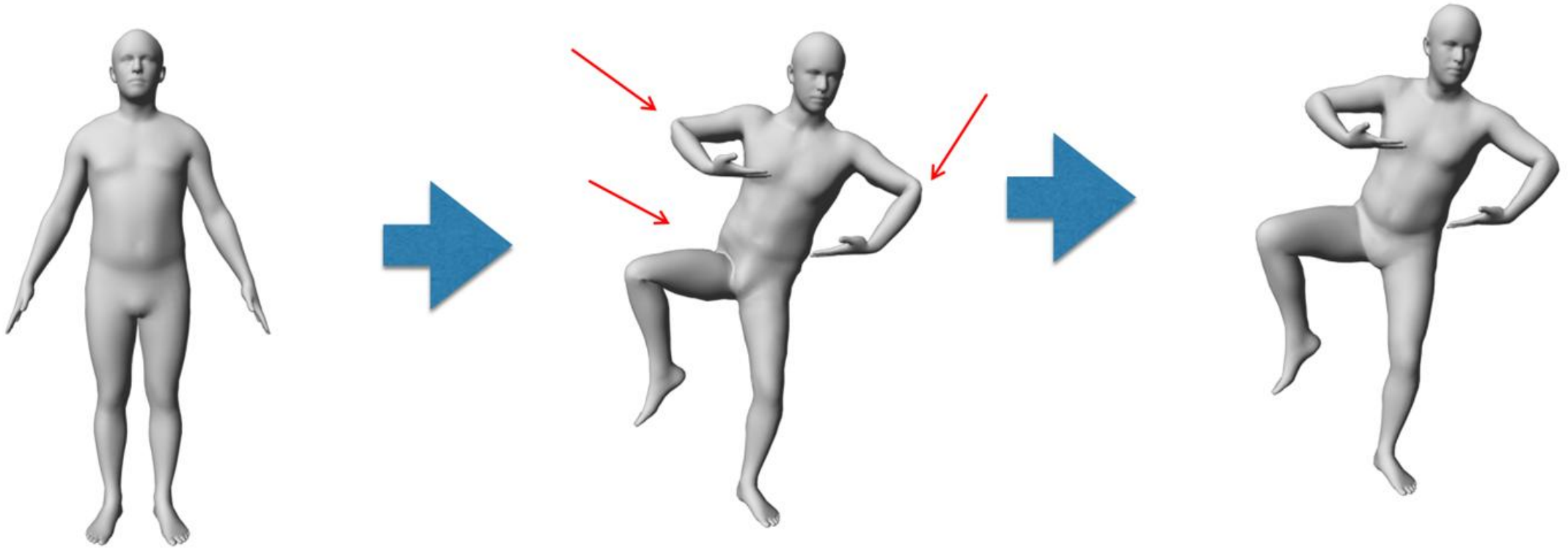
Many more advanced solutions in literature: dual-quaternion skinning, joint-based or cage-based deformers, etc.

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

LBS artifacts

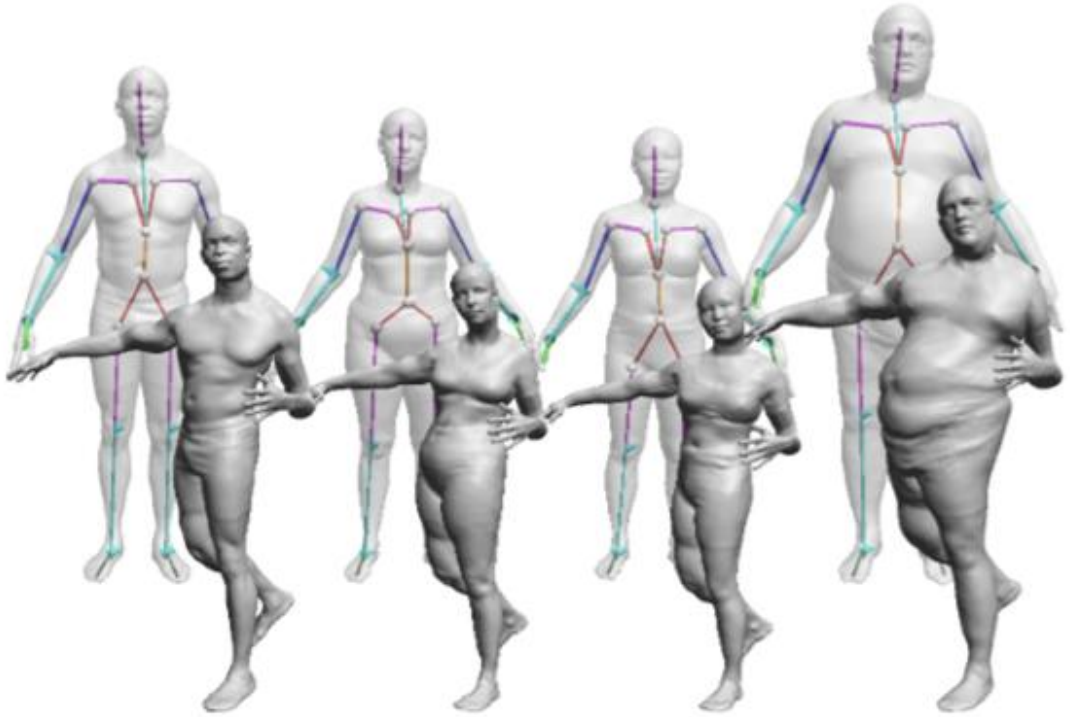
Solution: pose-dependent blend shapes

$B_P(\theta)$
Corrective additive term

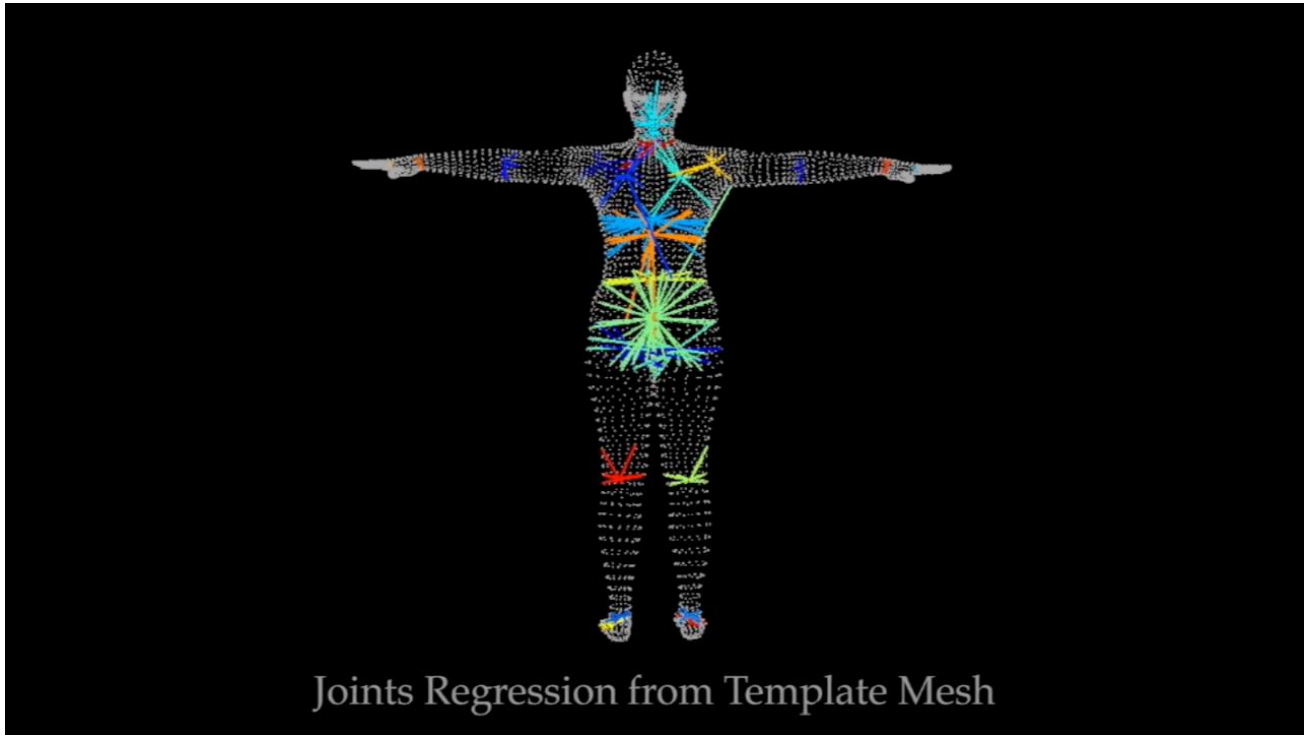


Problem: Skeleton changes for different identities

How to obtain the skeleton for different identities?



Solution:
Joints as a function of the vertices



$$\mathcal{J}(\bar{\mathbf{T}} + B_S(\vec{\beta}; \mathcal{S})) \rightarrow J_{\mathcal{M}} = RX_{\mathcal{M}}$$

In summary:

pose shape

$$M(\vec{\theta}, \vec{\beta}; \mathbf{T}, \mathcal{S}, \mathcal{P}, \mathcal{W}, \mathcal{J})$$

Input

Model parameters to be learned from data

$\mathbf{T}$ Template (average shape)

$\mathcal{S}$ Shape blend shape matrix

$\mathcal{P}$ Pose blend shape matrix

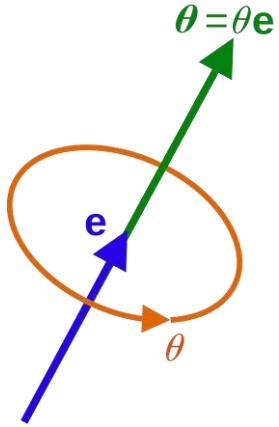
$\mathcal{W}$ Blend weights matrix

$\mathcal{J}$ 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



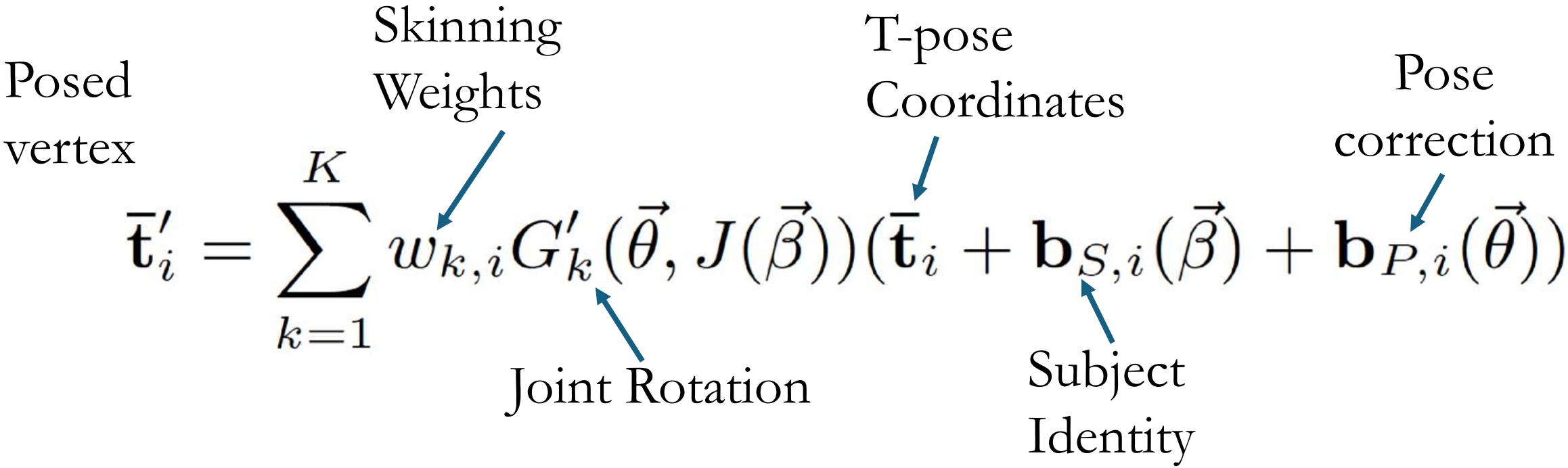
Angle: θ (in radians)

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

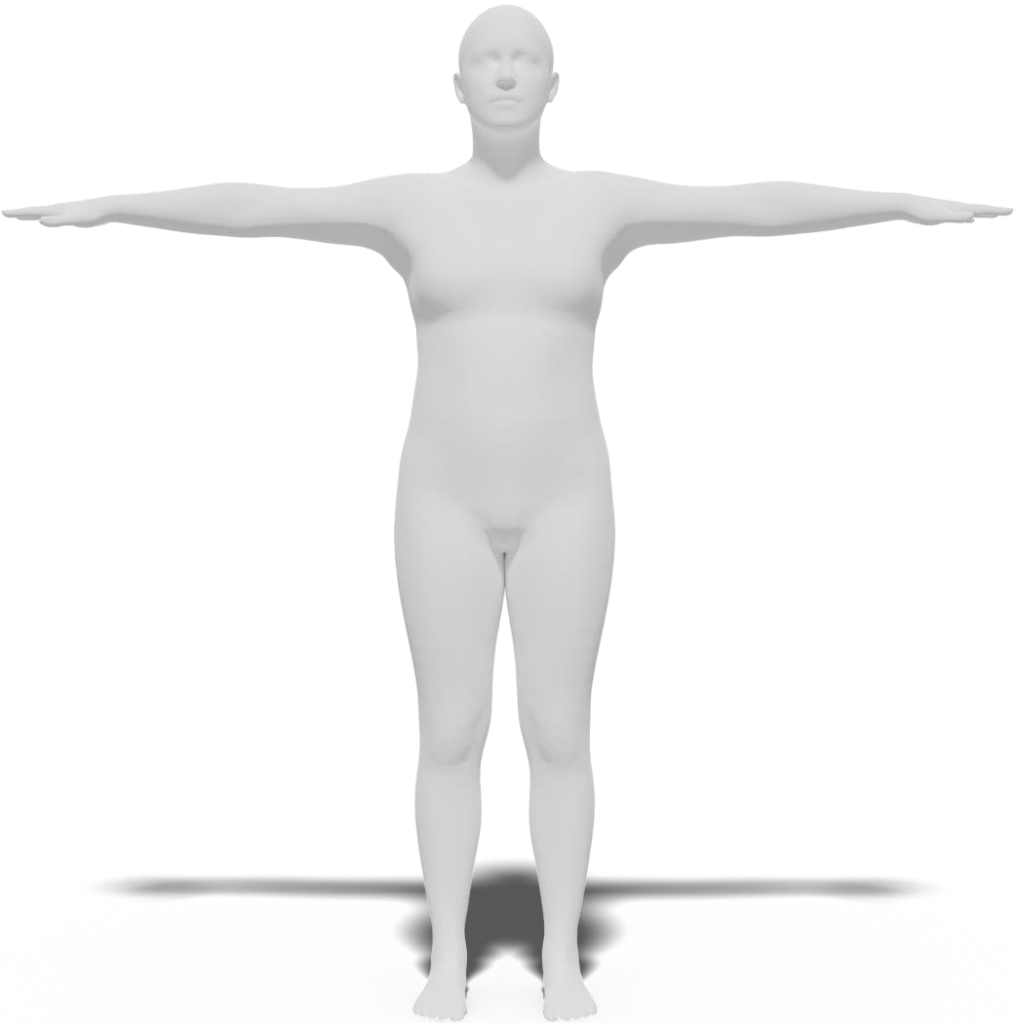
Angle-Axis: $\vec{\theta} = \theta\vec{e} = [\theta e_x, \theta e_y, \theta e_z]$, *where* $|\vec{\theta}| = \theta$

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

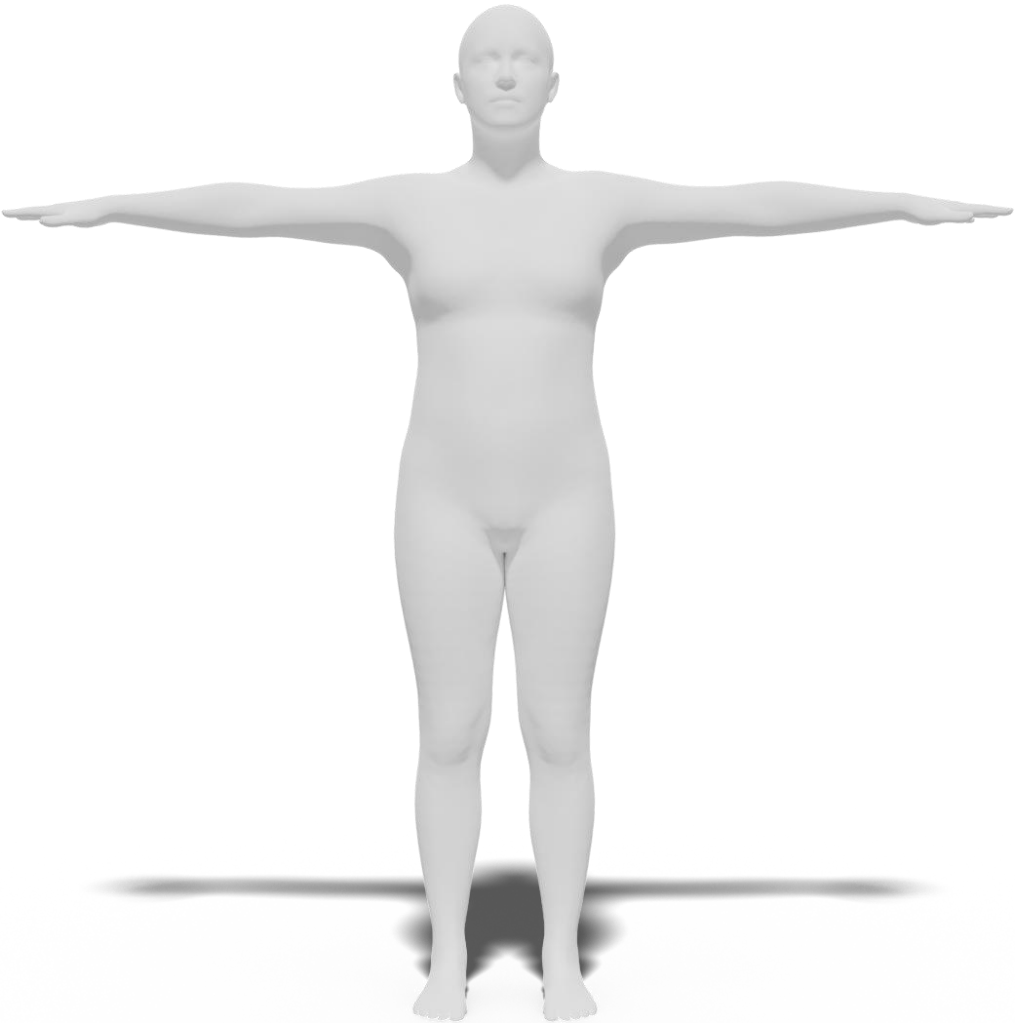
Given the identity and pose parameters, vertices are update by this equation:



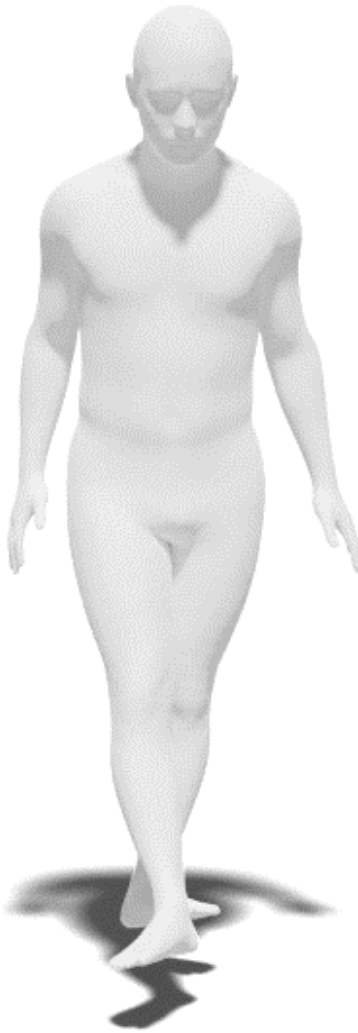
$$M(\mathbf{0}, \mathbf{I}) =$$



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



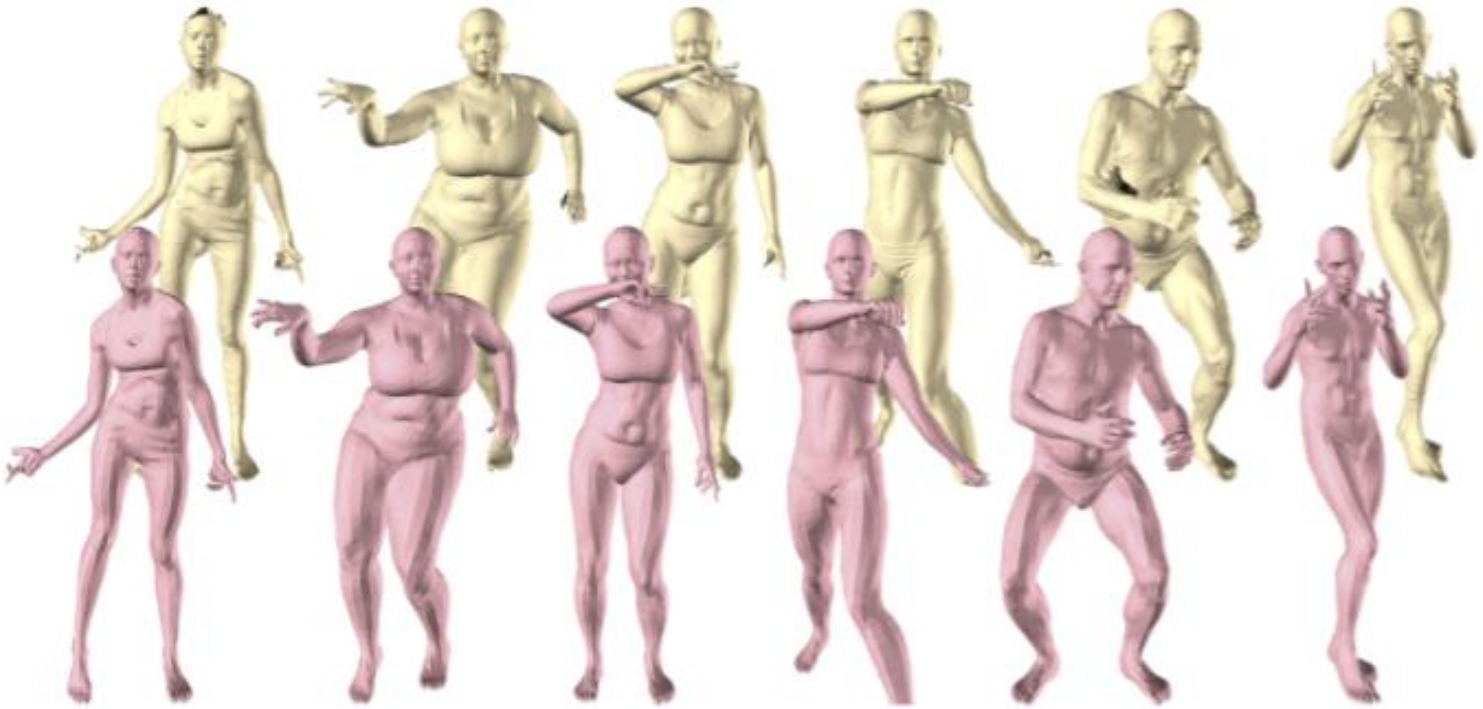
$$M(\beta, \theta) =$$



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



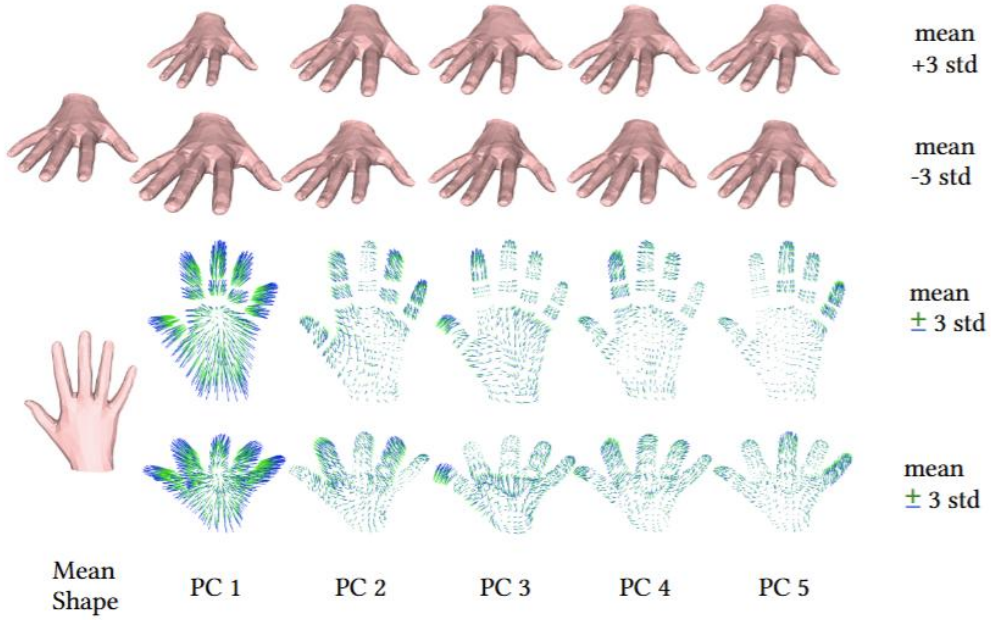
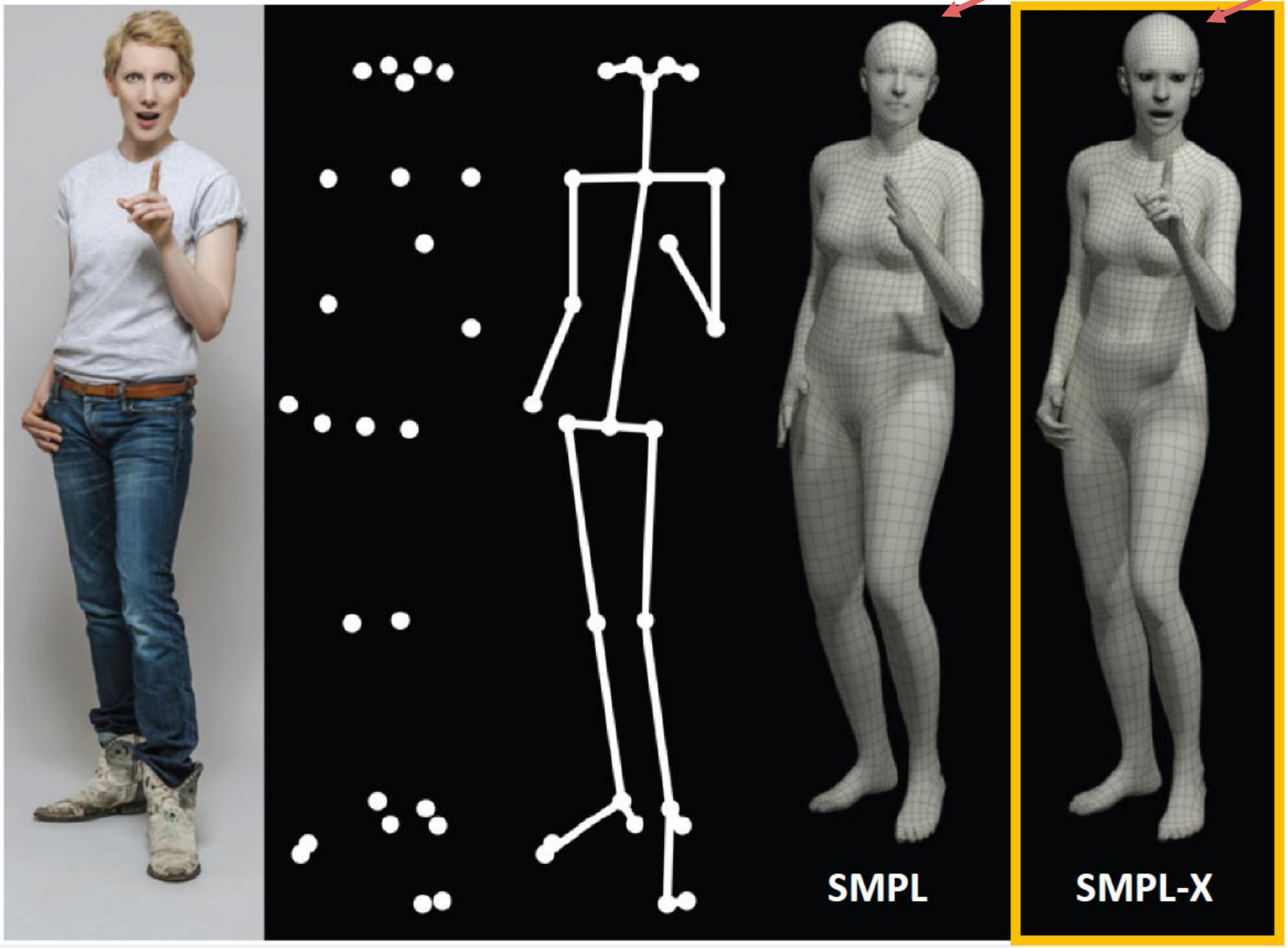


Fig. 4. PCA shape space. Each column depicts the effect of one of the first five *principal components* (PCs) of the learned hand shape space. The effect of each PC is shown by adding ±3 standard deviations (std) to the mean shape (left-most image), as indicated. See the Supplemental Video.

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



Follow-up: SMIL

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

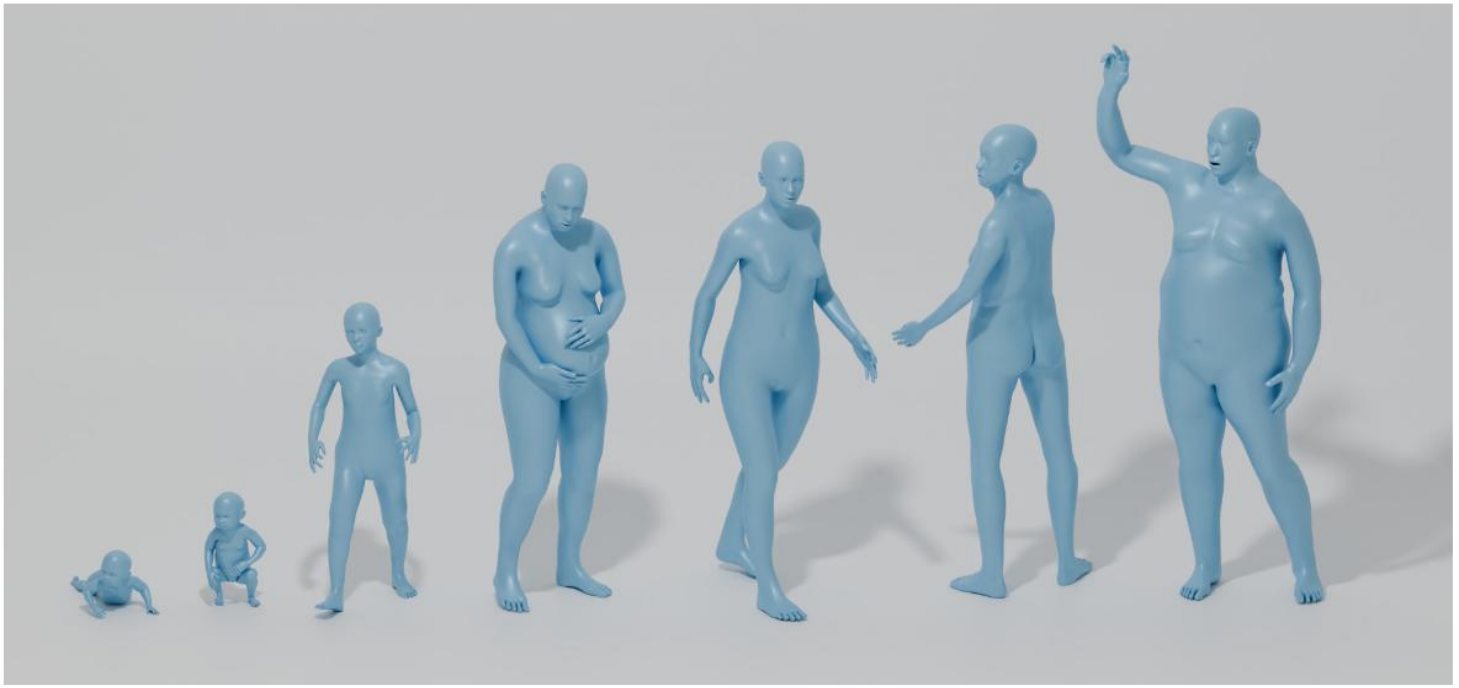
Nikolas Hesse[1]*, Sergi Pujades[2], Javier Romero[3], Michael J. Black[2], Christoph Bodensteiner[1], Michael Arens[1], Ulrich G. Hofmann[4], Uta Tacke[5], Mijna Hadders-Algra[6], Raphael Weinberger[7], Wolfgang Müller-Felber[7], and A. Sebastian Schroeder[7]

Skinned Multi-Linear Infant Model (SMIL)

Registration Results

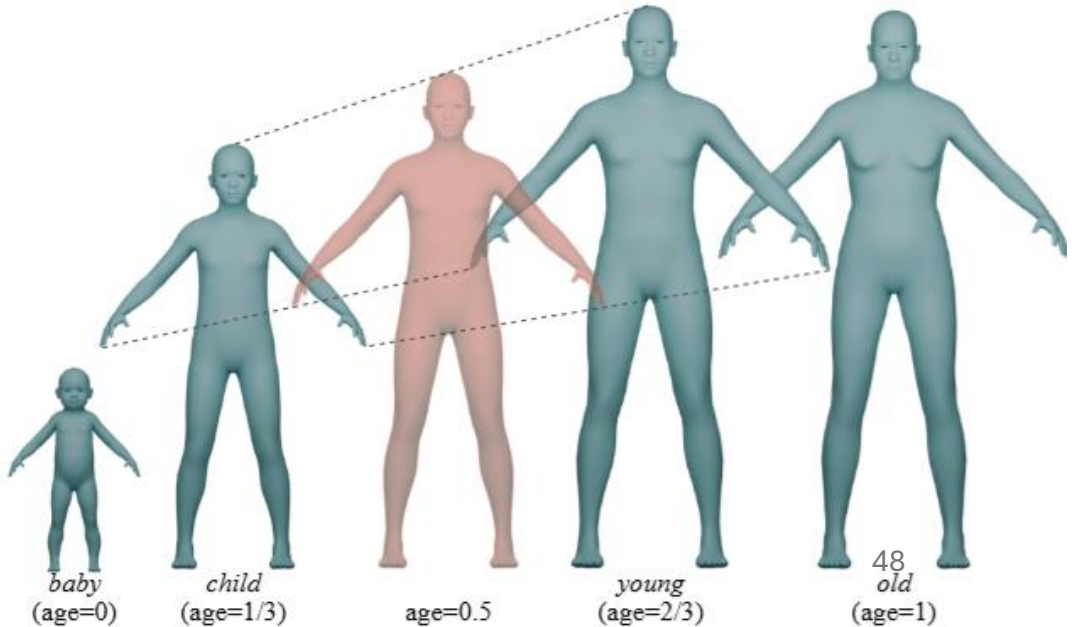
Human Mesh Modeling for Anny Body

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

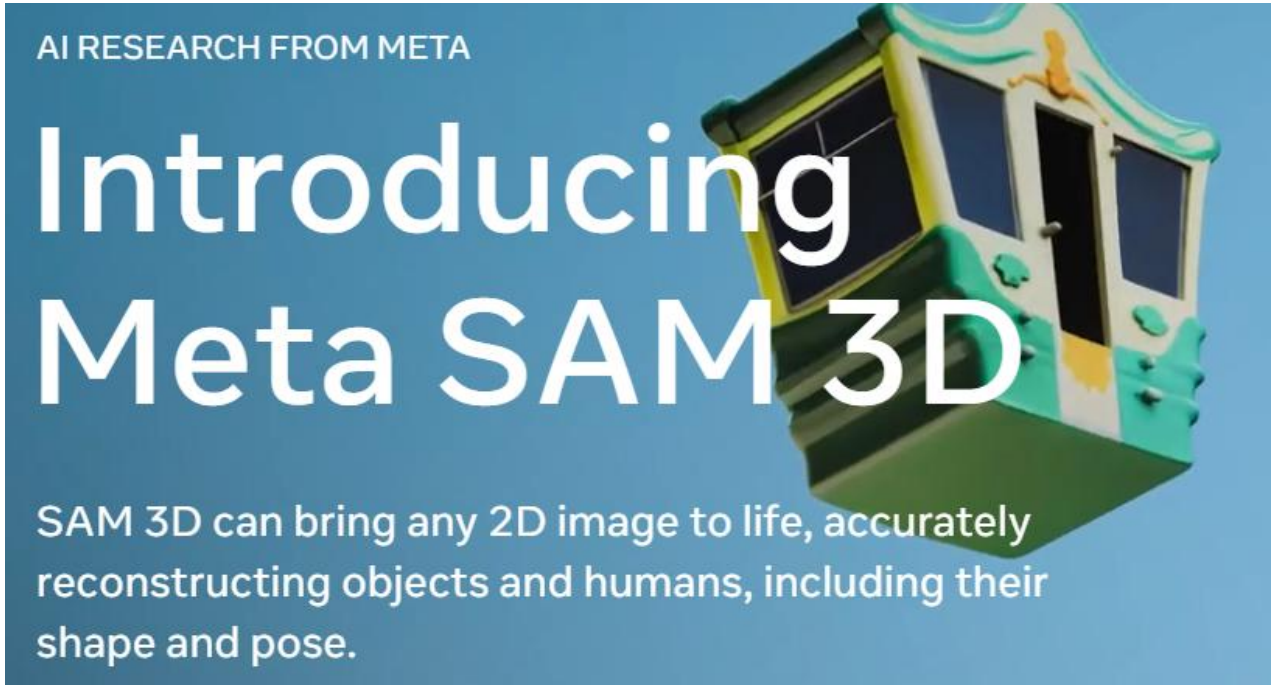


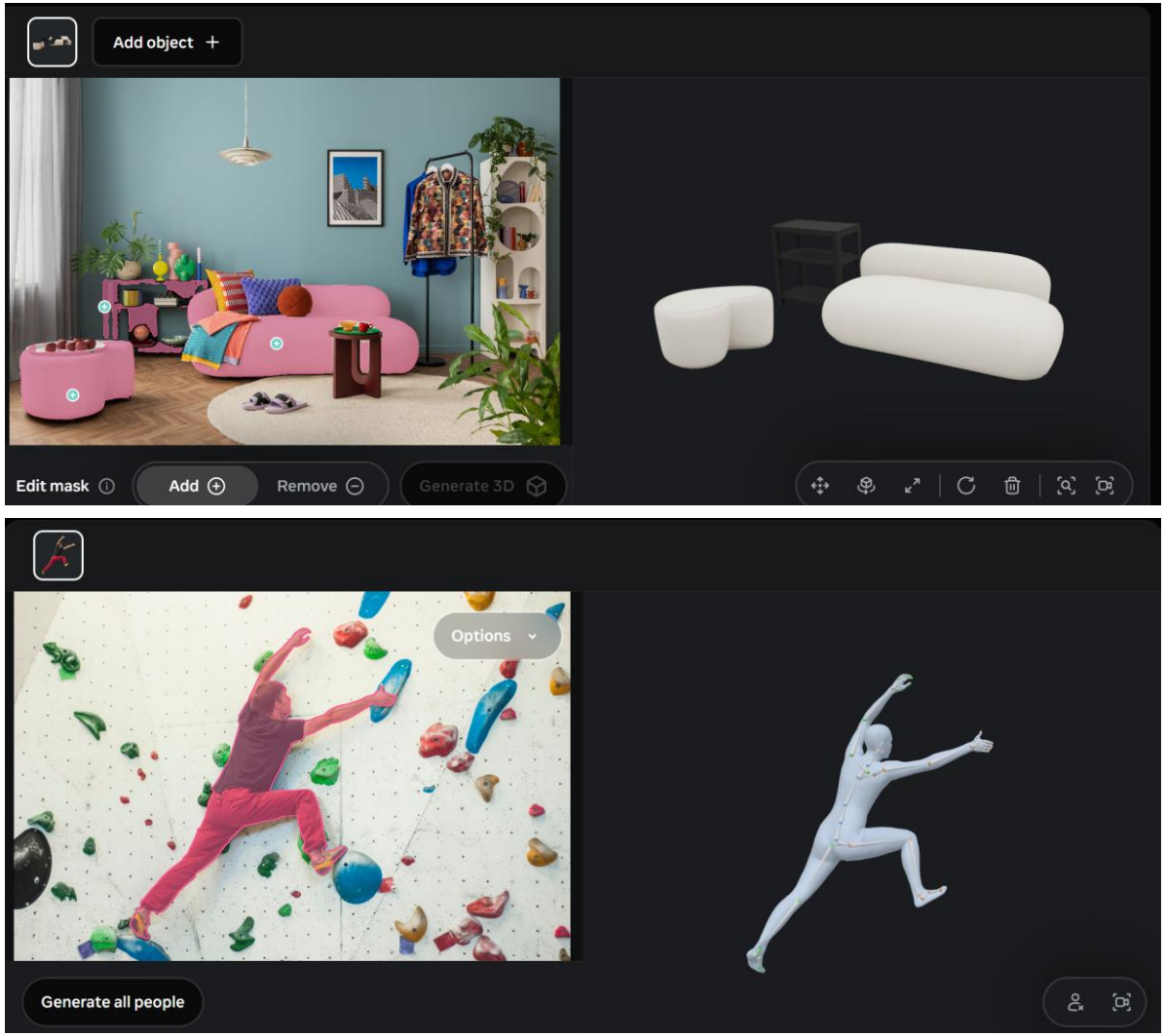
Similar to SMPL but:

1) Learned on syhtetic data (MetaHuman)
 + WHO anthropometric
2) 13K vertices, and 163 bones



Meta released SAM 3D Models



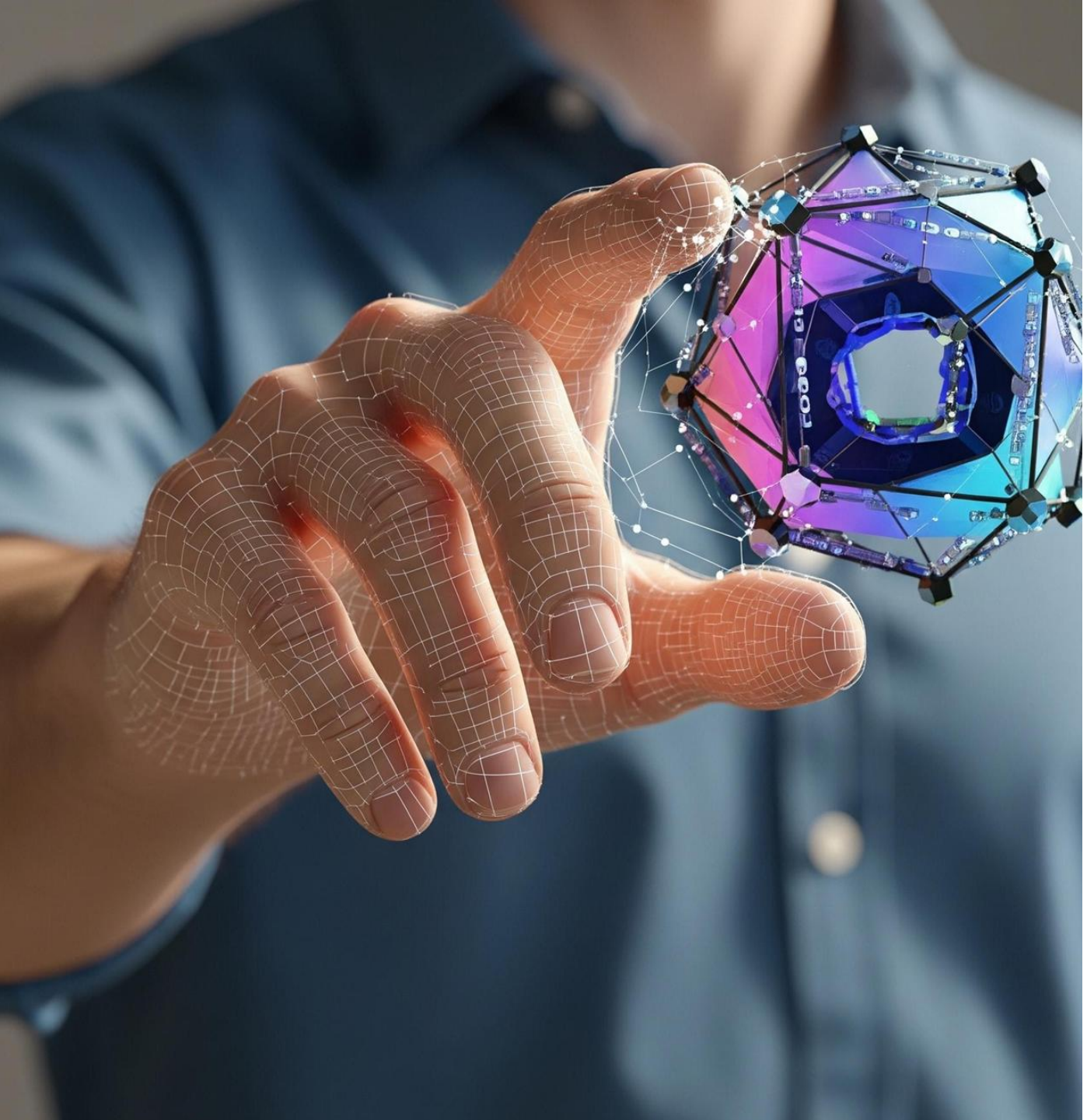


https://aidemos.meta.com/segment-anything/gallery

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



Shape Analysis

How to learn these?

pose $\vec{\theta}$ shape $\vec{\beta}$

$$M(\vec{\theta}, \vec{\beta}; \mathbf{T}, \mathcal{S}, \mathcal{P}, \mathcal{W}, \mathcal{J})$$

Input

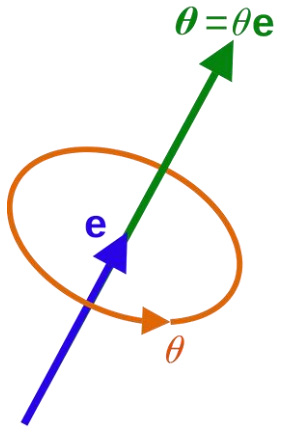
Model parameters to be learned from data

- $\mathbf{T}$ 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



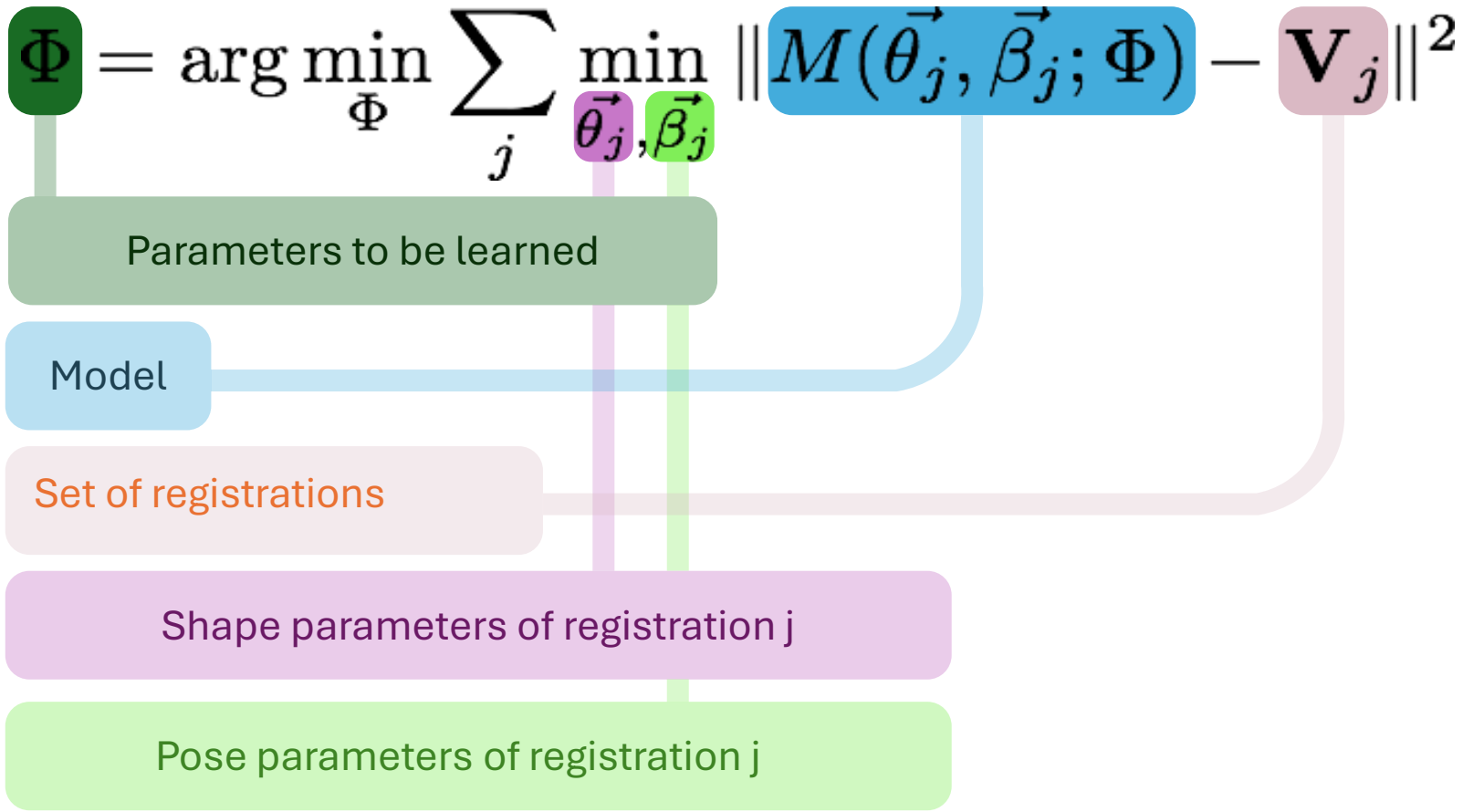
Angle: θ (in radians)

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

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

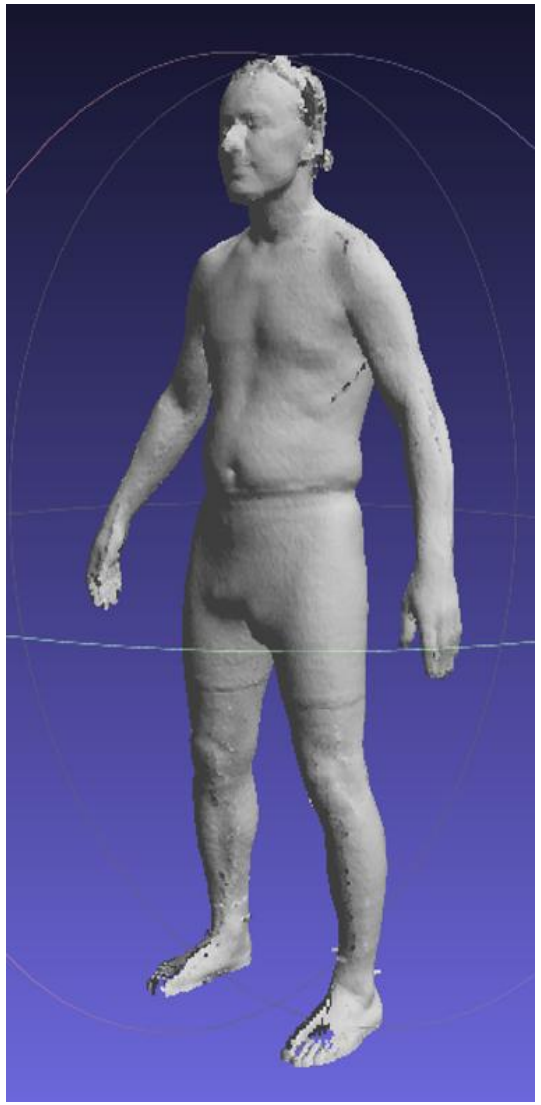
https://medium.com/data-science/3d-rotations-intuitions-and-limitations-f3ae2122fe23

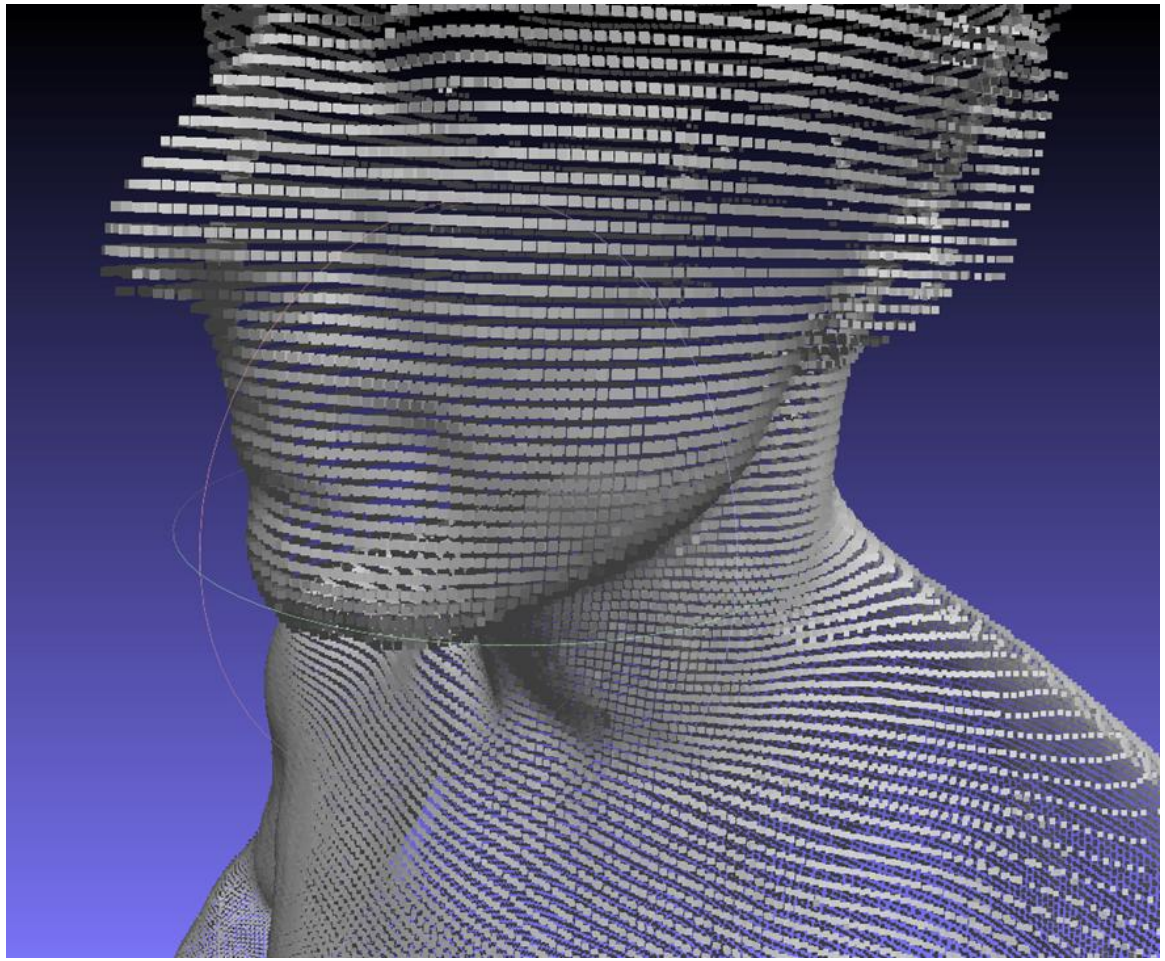
Target function



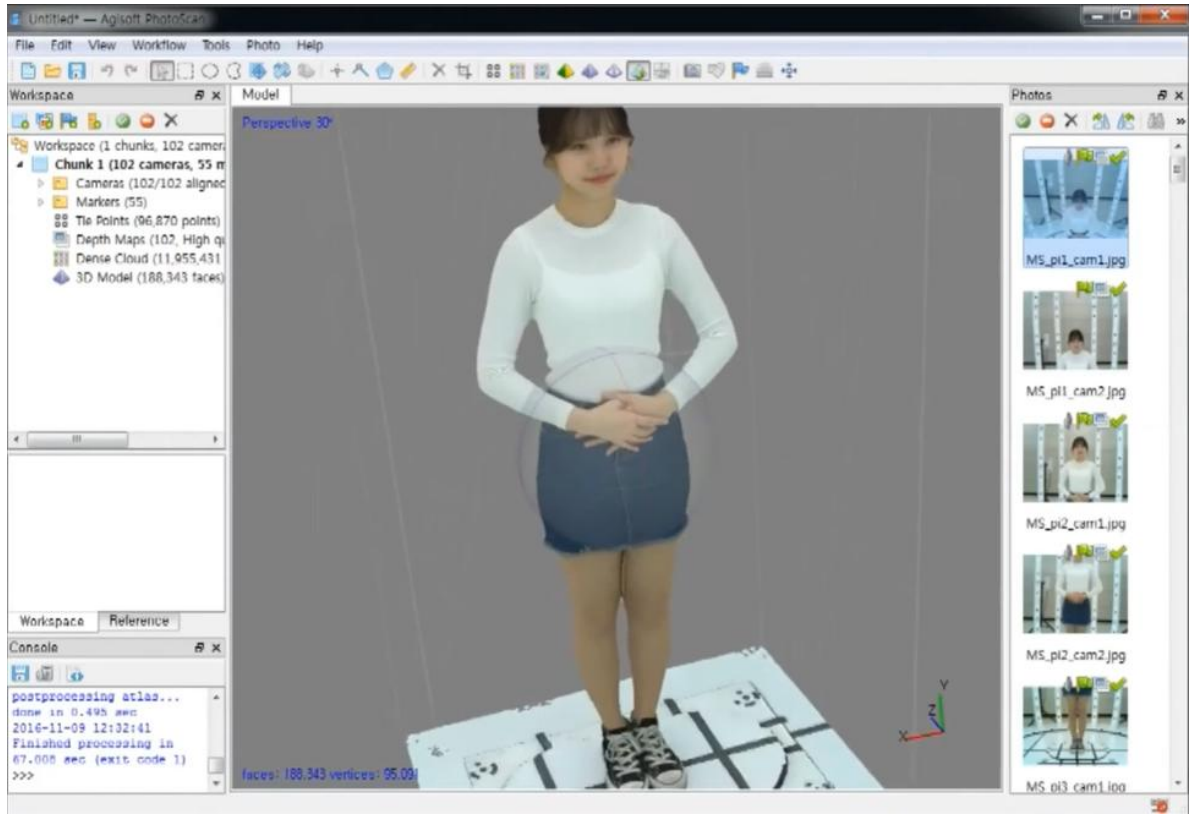
The parameters to be learned are those that best express a training set.

Problem: Scans are just points, no semanticity





How do you obtain this?



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

3D Scanning


Augmented Reality
Detection and Pose Estimation
3D Scene Understanding in Point Clouds

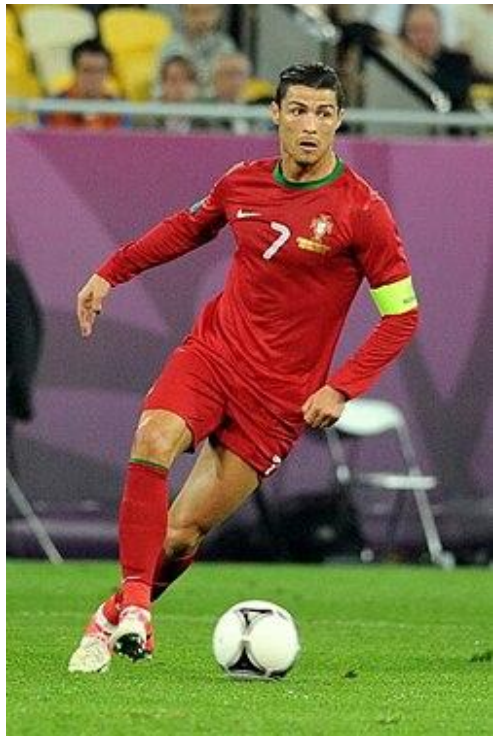
Mobile phone video

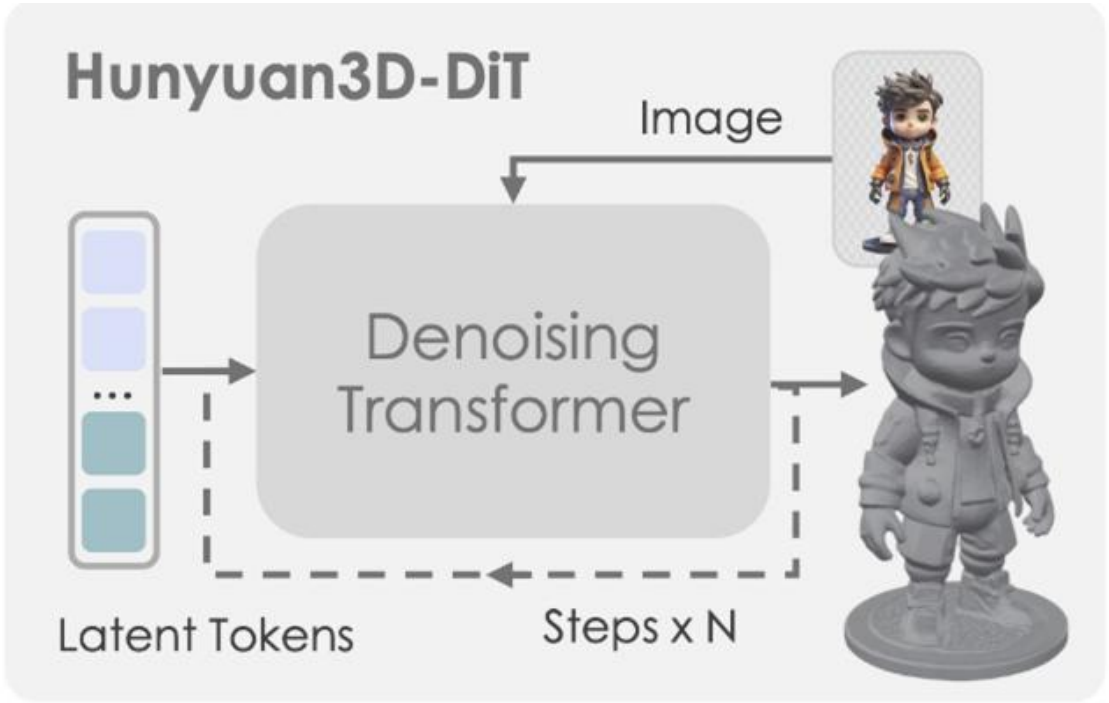
Mobile phone video

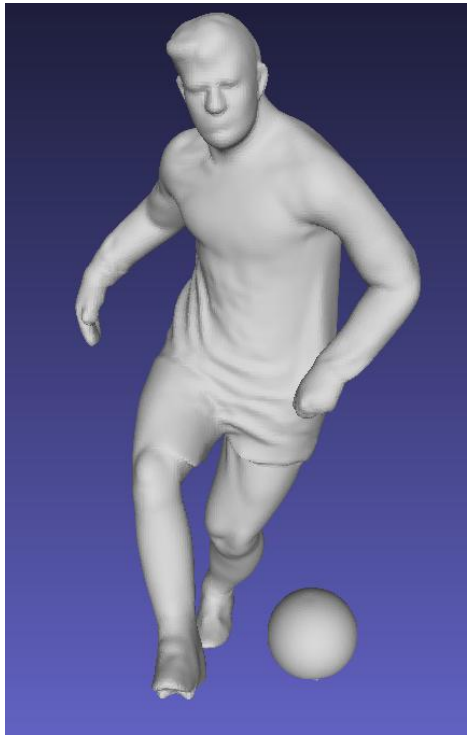
NeRF/GS Model (luma.ai)

3D/4D Generative AI as a new data source

https://huggingface.co/spaces/tencent/Hunyuan3D-2





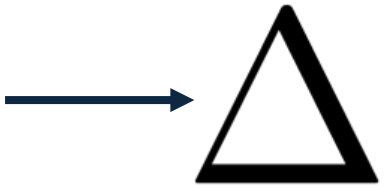


Disparate sources of data call for tools to relate them!

Spectral shape analysis

N

Vertices



LBO

Given a mesh

Spectral shape analysis

N

Vertices

Given a mesh

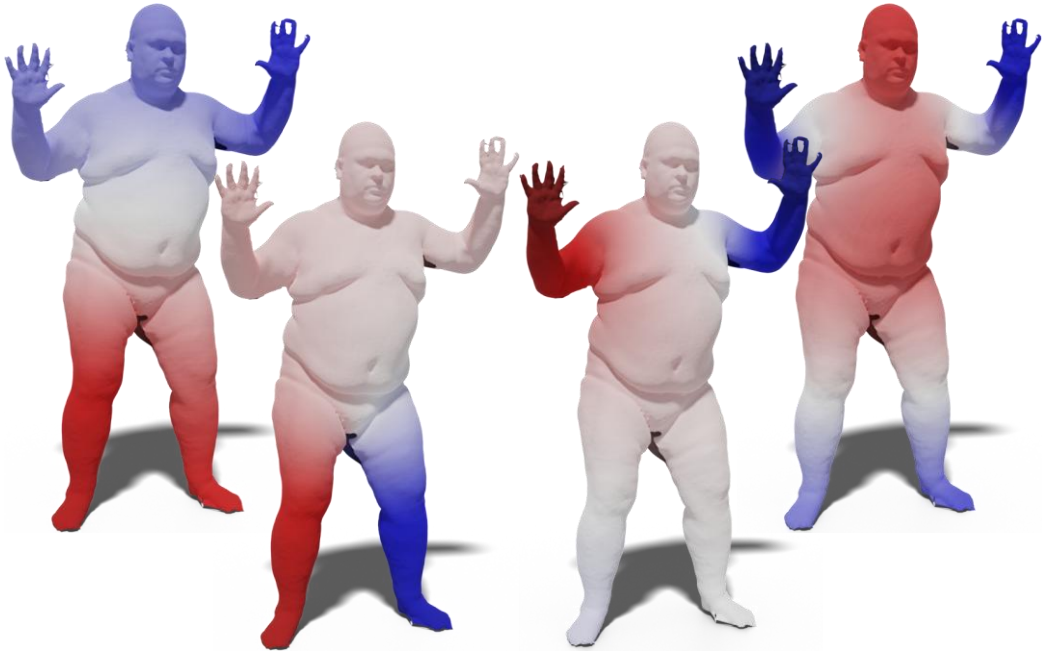
Δ

LBO

Eigenvectors

(Pointwise features)

N x k



Spectral shape analysis

N

Vertices

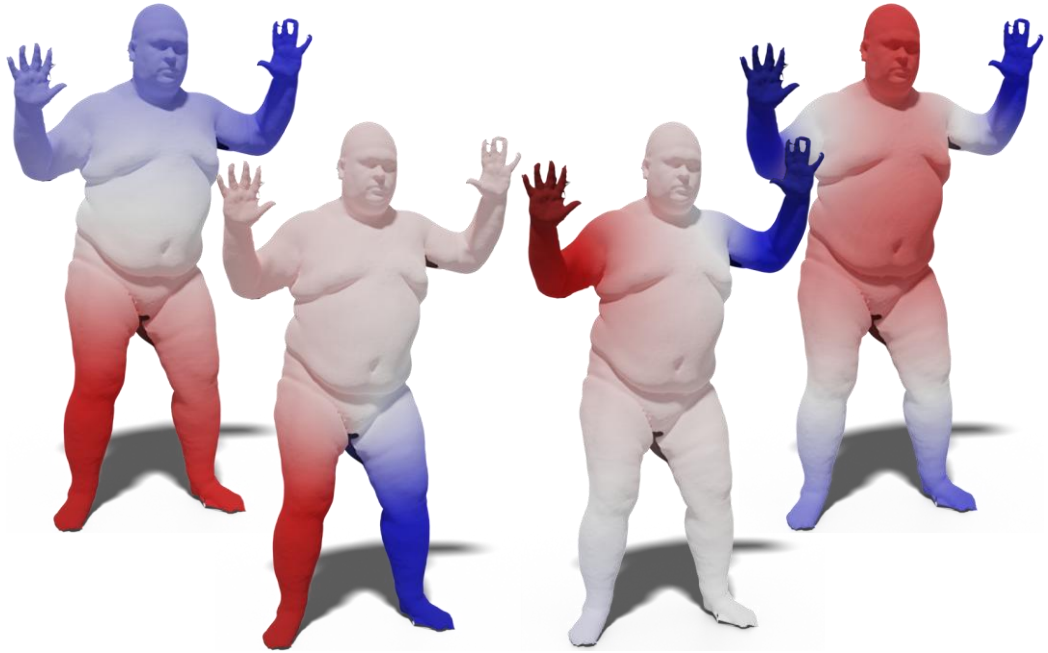
Given a mesh

LBO

Eigenvectors

(Pointwise features)

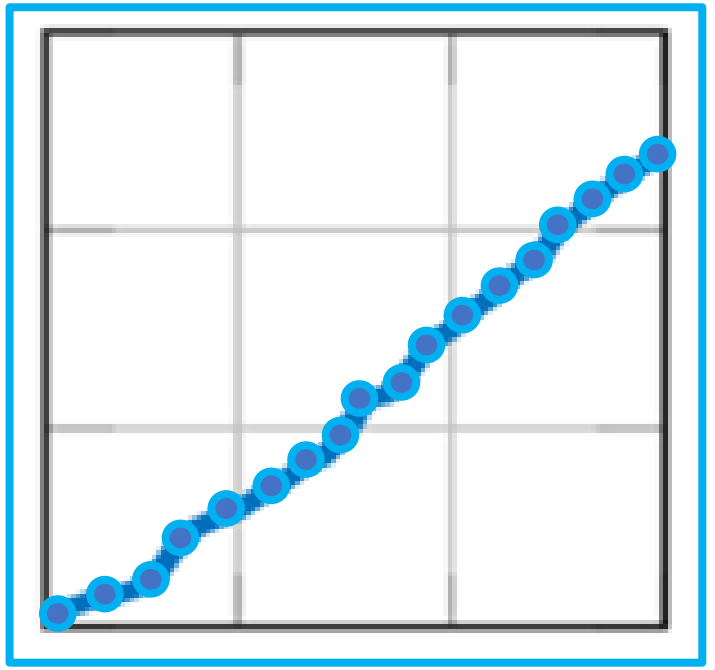
N x k



k

Eigenvalues

(Latent code)



Spectral shape analysis

N

Vertices

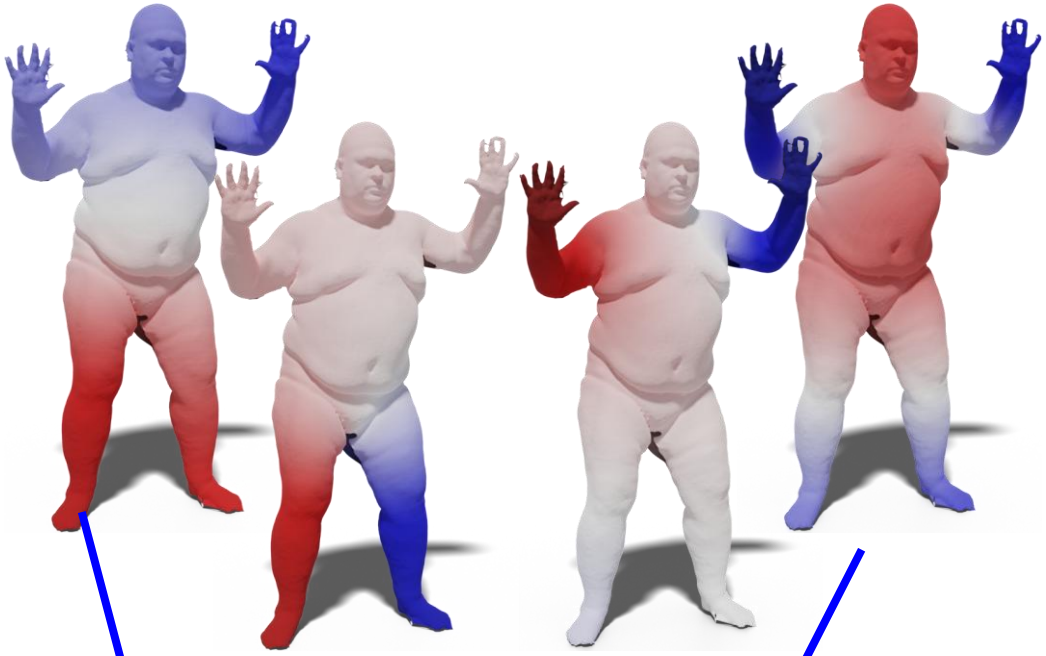
Given a mesh

LBO

Eigenvectors

(Pointwise features)

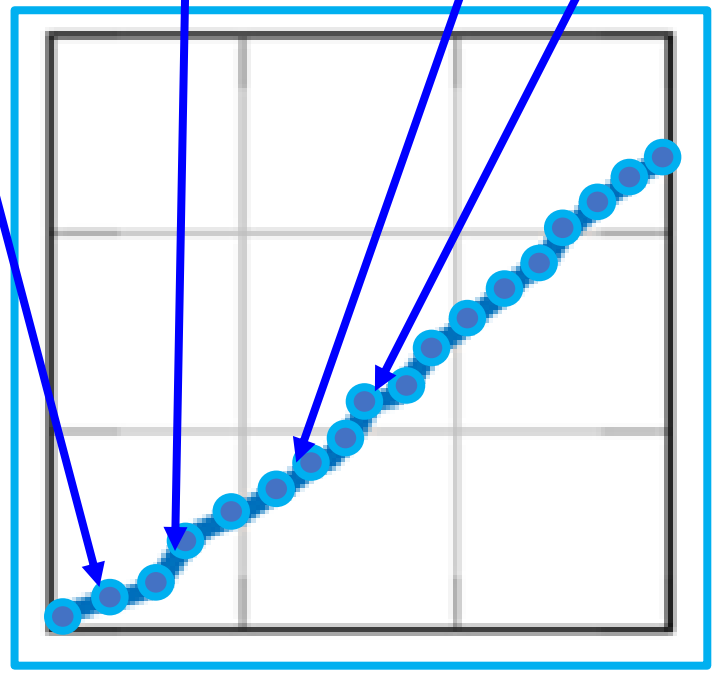
N x k



k

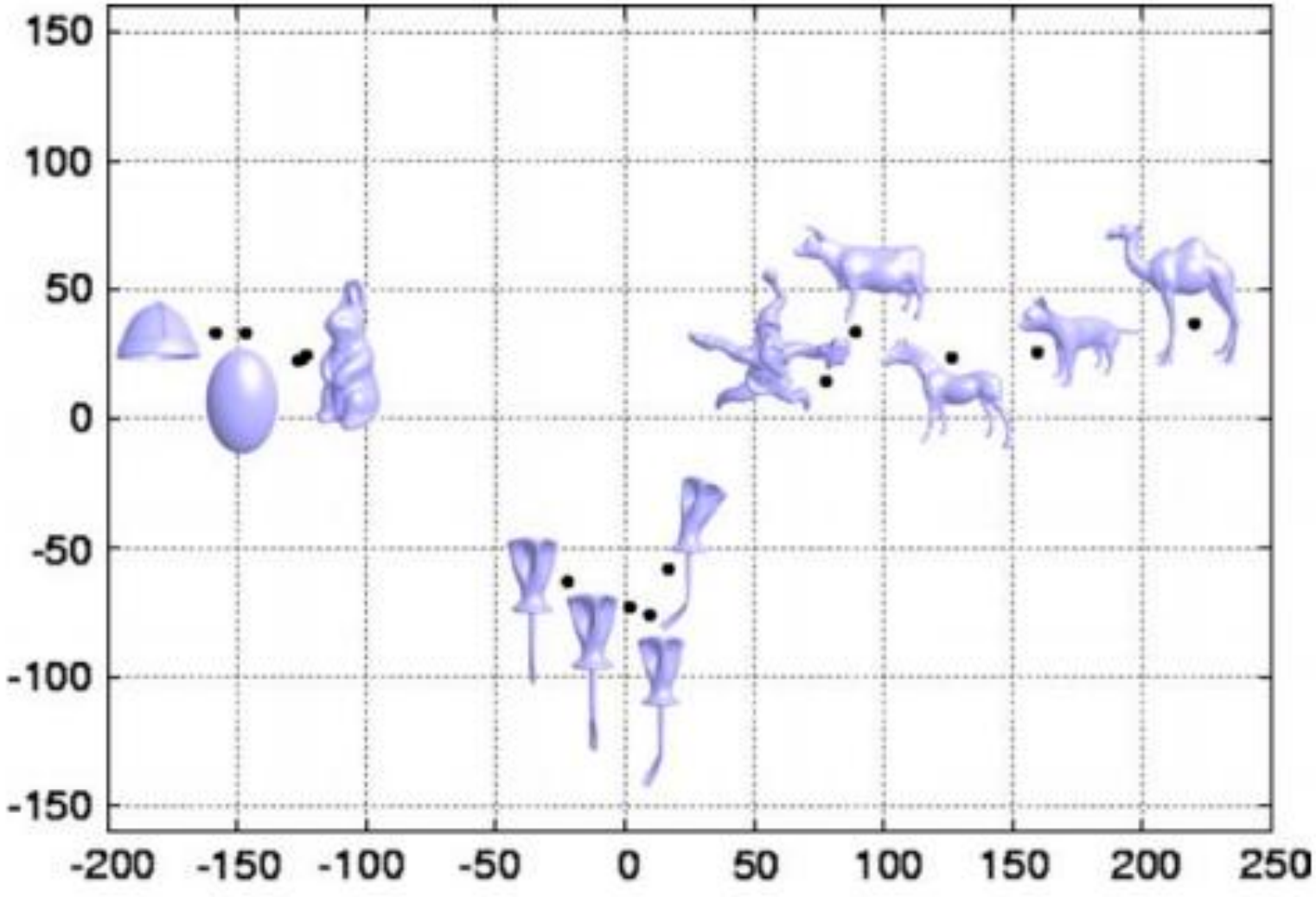
Eigenvalues

(Latent code)

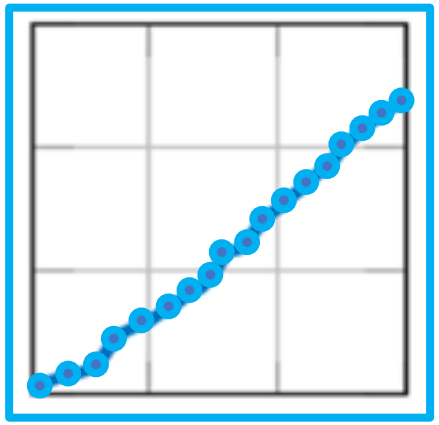


Application (Latent Code): Shape classification

Eigenvalues as a latent space

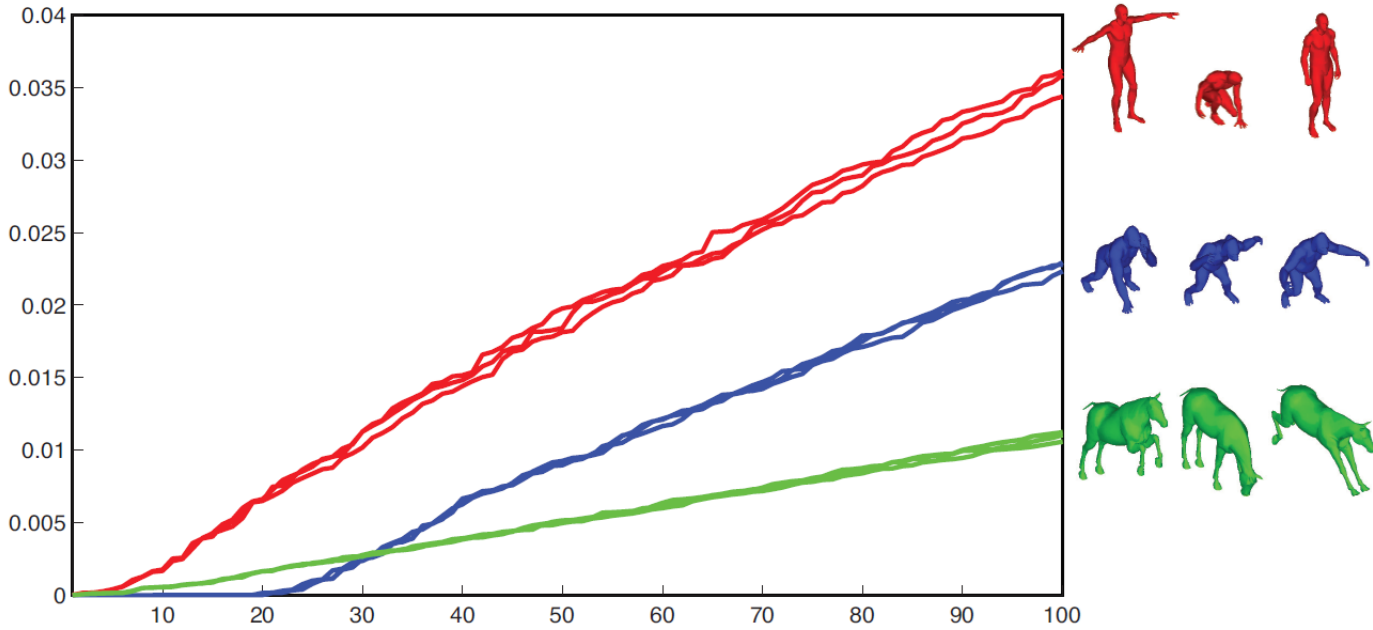


Application (Latent Code): Shape classification



Eigenvalues
sorted by increasing value

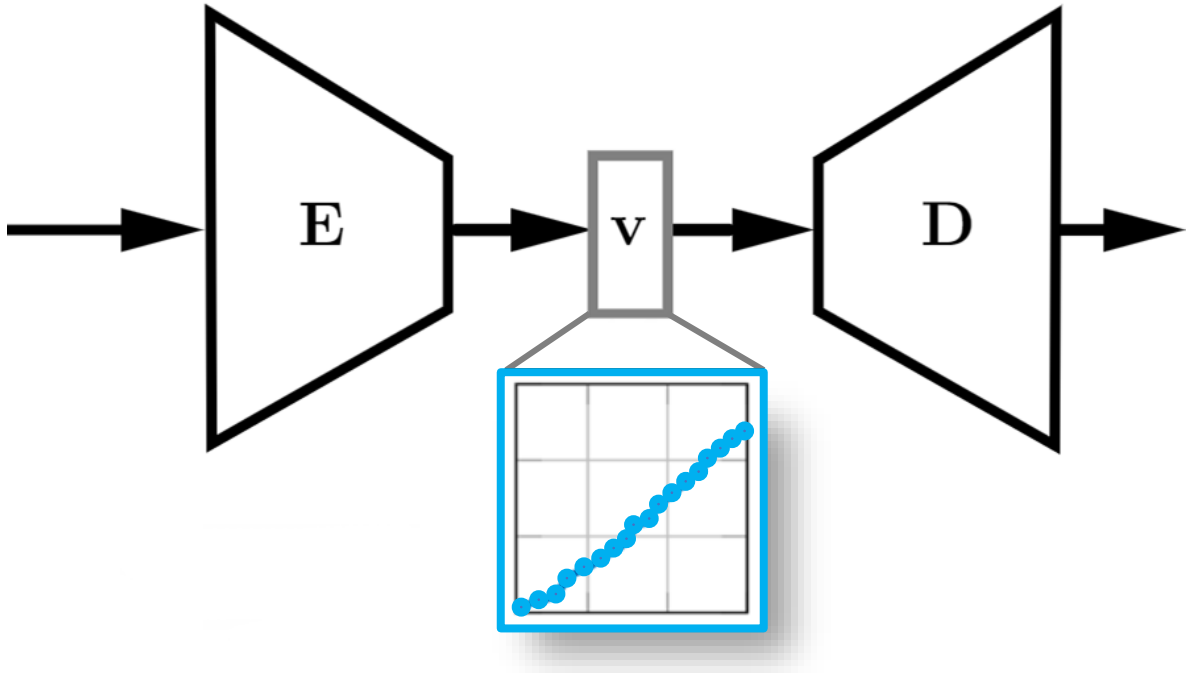
Shape-DNA



$$\text{cspec}_n(M,g) = \{\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n\} \in \mathbb{R}^n_{\geq 0}$$

Eigenvalues represent the global shape signature
(Generally normalized by the first non-zero)

Application (Latent Code): Shape retrival?

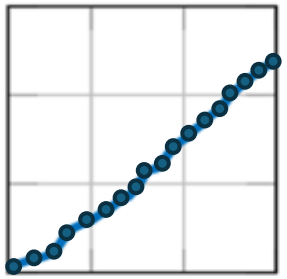


Why not use the eigenvalues as a latent representation?

Mesh

~4000 vertices

Laplacian eigenvalue λ_i



30 Scalars

however...

RESEARCH ANNOUNCEMENTS

1992

BULLETIN (New Series) OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 27, Number 1, July 1992

ONE CANNOT HEAR THE SHAPE OF A DRUM

CAROLYN GORDON, DAVID L. WEBB, AND SCOTT WOLPERT

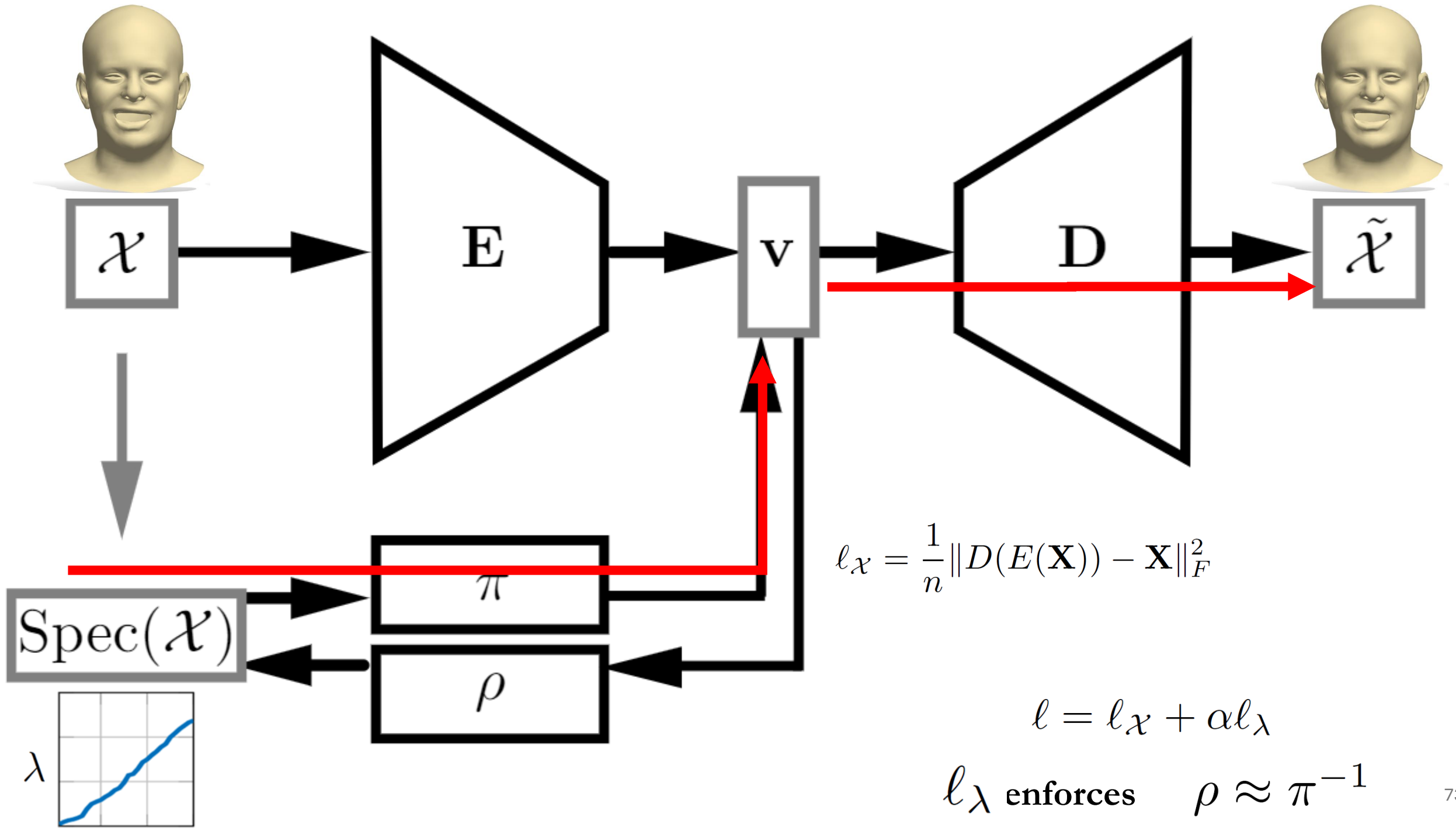
ABSTRACT. We use an extension of Sunada's theorem to construct a nonisometric pair of isospectral simply connected domains in the Euclidean plane, thus answering negatively Kac's question, "can one hear the shape of a drum?" In order to construct simply connected examples, we exploit the observation that an orbifold whose underlying space is a simply connected manifold with boundary need not be simply connected as an orbifold.

1. KAC'S QUESTION

Let (M, g) be a compact Riemannian manifold with boundary. Then M has a Laplace operator Δ, defined by $\Delta(f) = -\operatorname{div}(\operatorname{grad} f)$, that acts on smooth functions on M. The *spectrum* of M is the sequence of eigenvalues of Δ. Two Riemannian manifolds are *isospectral* if their spectra coincide (counting multiplicities). A natural question concerning the interplay of analysis and geometry is: must two isospectral Riemannian manifolds actually be isometric? (When M has nonempty boundary, one can consider the *Dirichlet spectrum*, i.e., the spectrum of Δ acting on smooth functions that vanish on the boundary, or the *Neumann spectrum*, that of Δ acting on functions with vanishing normal derivative at the boundary.) If M is a domain in the Euclidean plane then the Dirichlet eigenvalues of Δ are essentially the frequencies produced by a drumhead shaped like M, so the question has been phrased by Bers and Kac [16] (the latter attributes the problem to Bochner) as "can one hear the shape of a drum?" We answer this question negatively by constructing a pair of nonisometric simply connected plane domains that have both the same Dirichlet spectra and the same Neumann spectra. The domains are depicted in Figure 1.

The simple idea exploited here also permits us to construct the following: (1) a pair of isospectral flat surfaces (with boundary) one of which has a unit-length closed geodesic while the other has only a unit-length closed billiard trajectory; (2) a pair of isospectral potentials for the Schrödinger operator on

Received by the editors July 11, 1991 and, in revised form, November 5, 1991.
1991 *Mathematics Subject Classification.* Primary 58G25, 35P05, 53C20.
The authors gratefully acknowledge partial support from NSF grants.

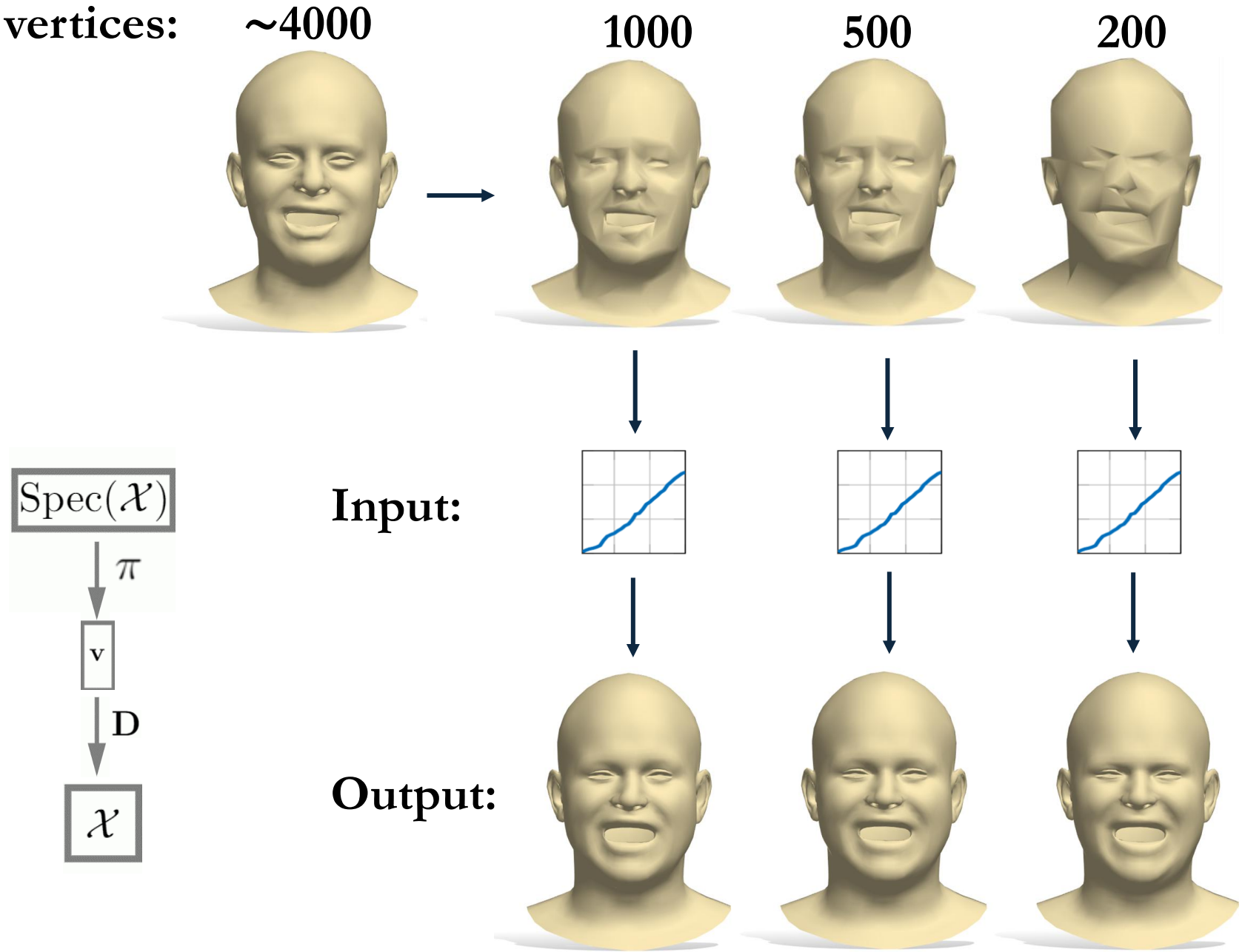


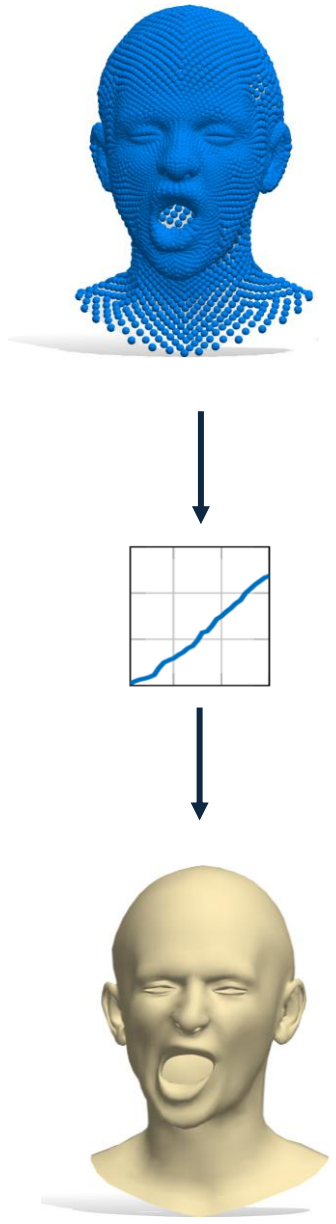
$$\ell_{\mathcal{X}} = \frac{1}{n}\|D(E(\mathbf{X})) - \mathbf{X}\|_F^2$$

$$\ell = \ell_{\mathcal{X}} + \alpha\ell_{\lambda}$$

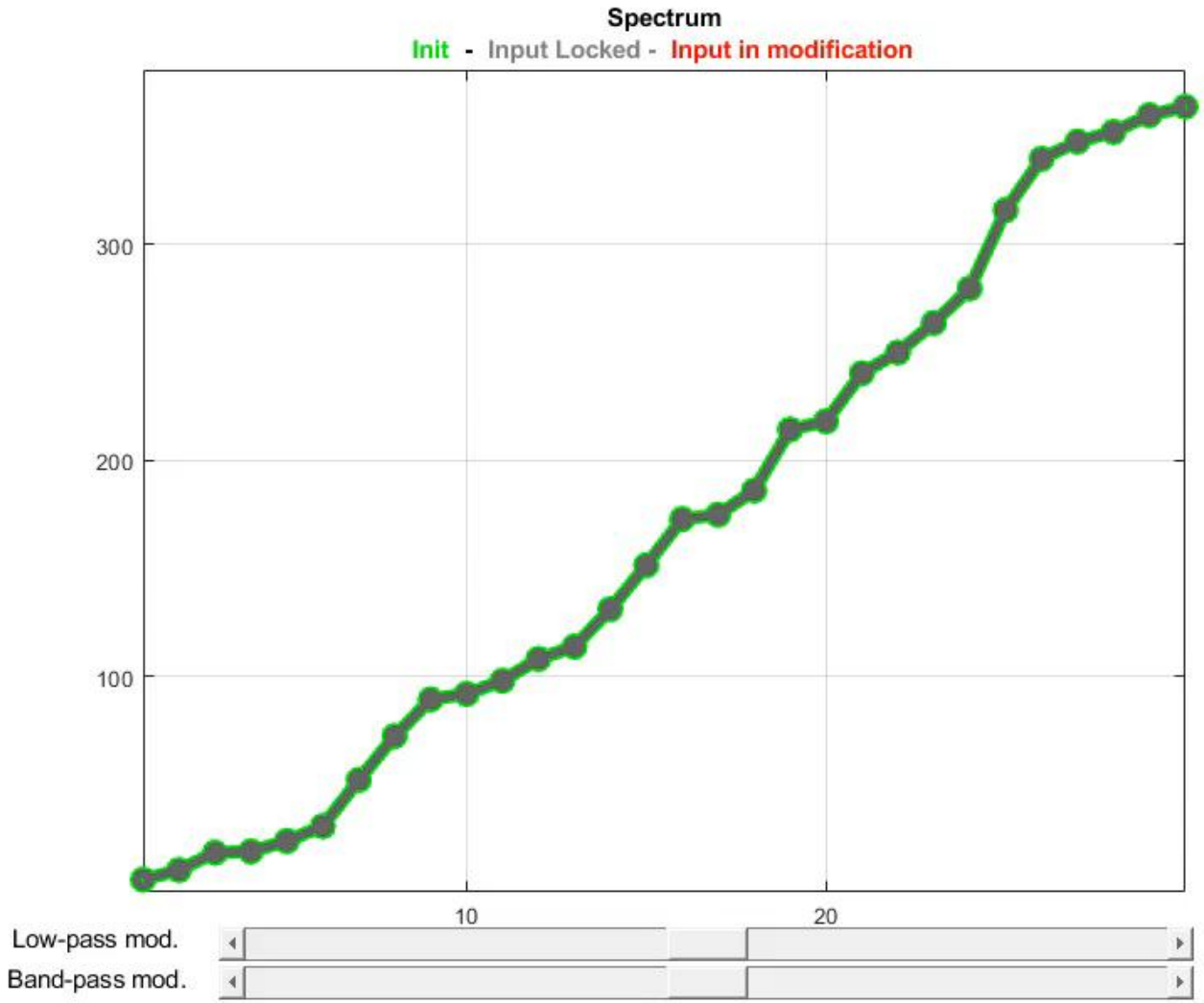
ℓ_{λ} **enforces** $\rho \approx \pi^{-1}$

Different discretizations and representations

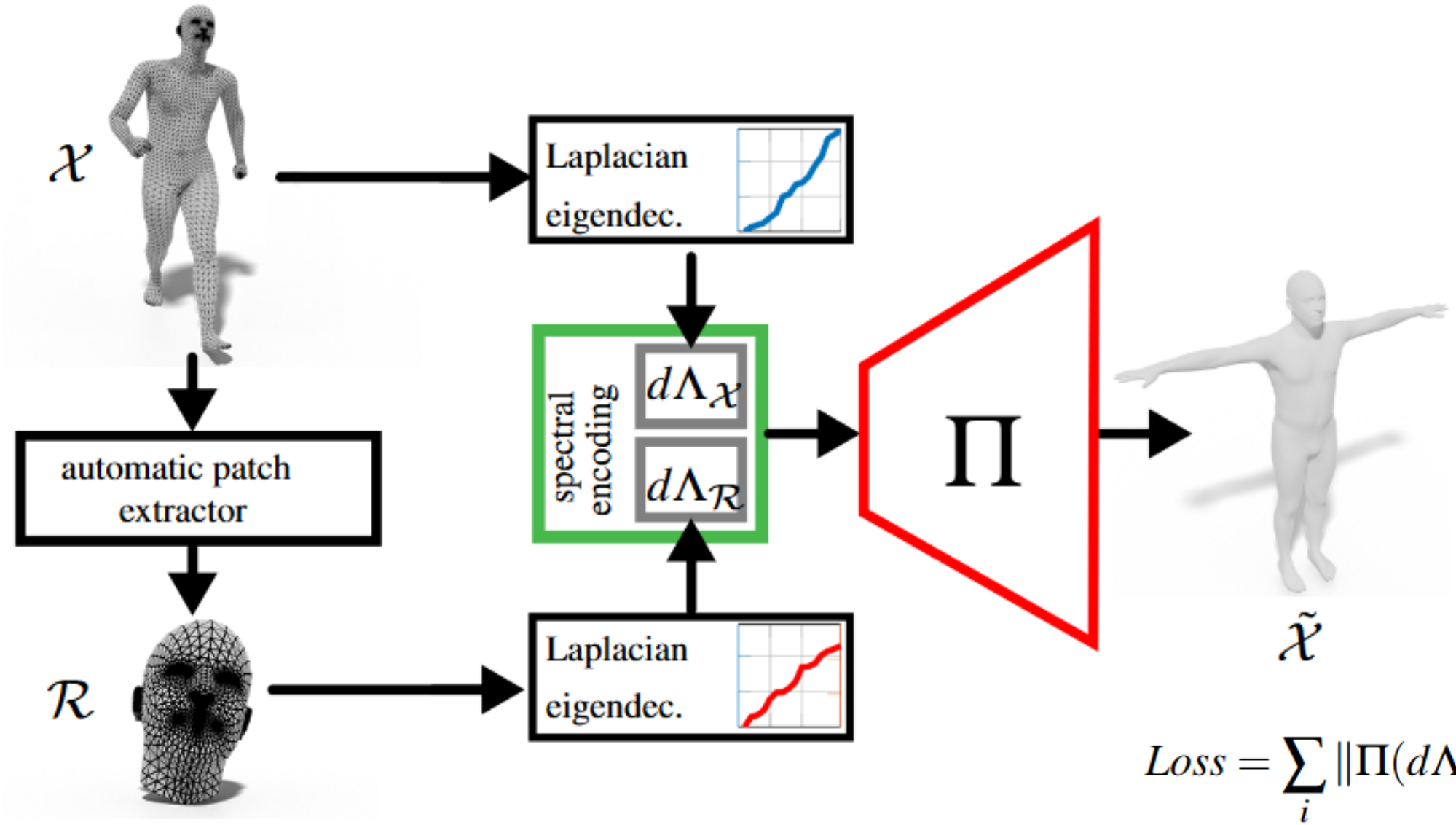




Latent space exploration



Decoding of spectral composition

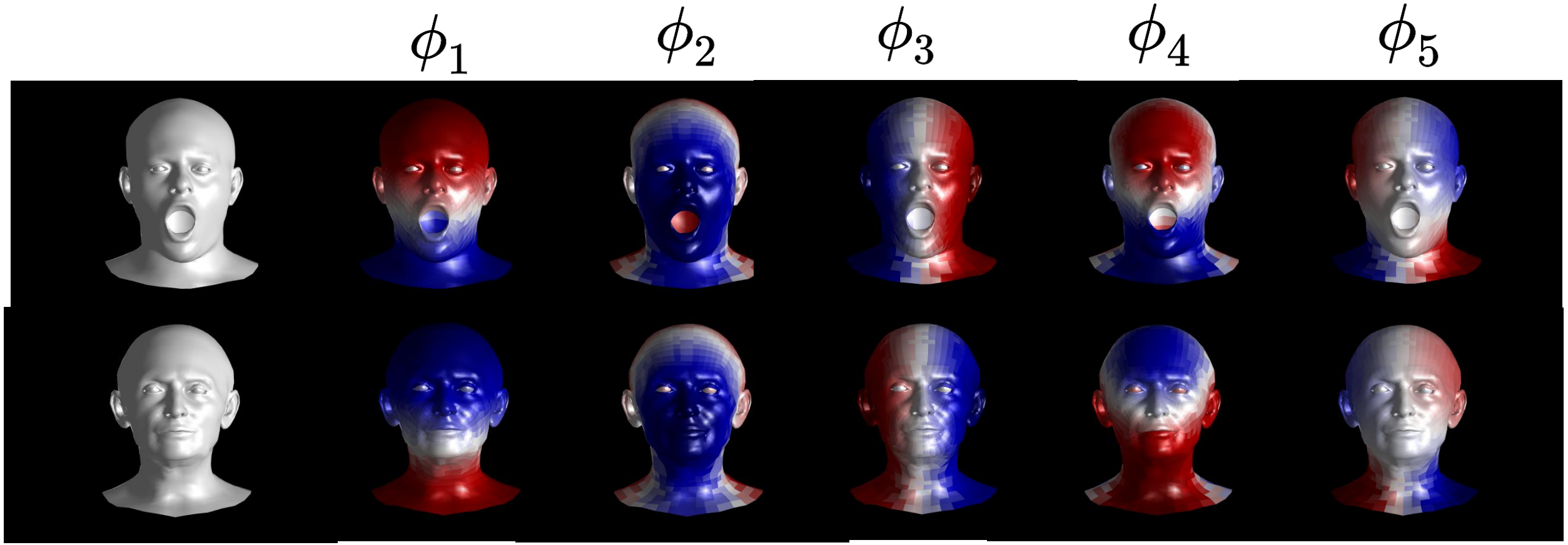


$$Loss = \sum_i \|\Pi(d\Lambda_i) - X_i\|_F^2$$

Localized shape modelling with global coherence: an inverse spectral approach, Pegoraro, Melzi, Castellani, Marin, Rodolà, SGP2022

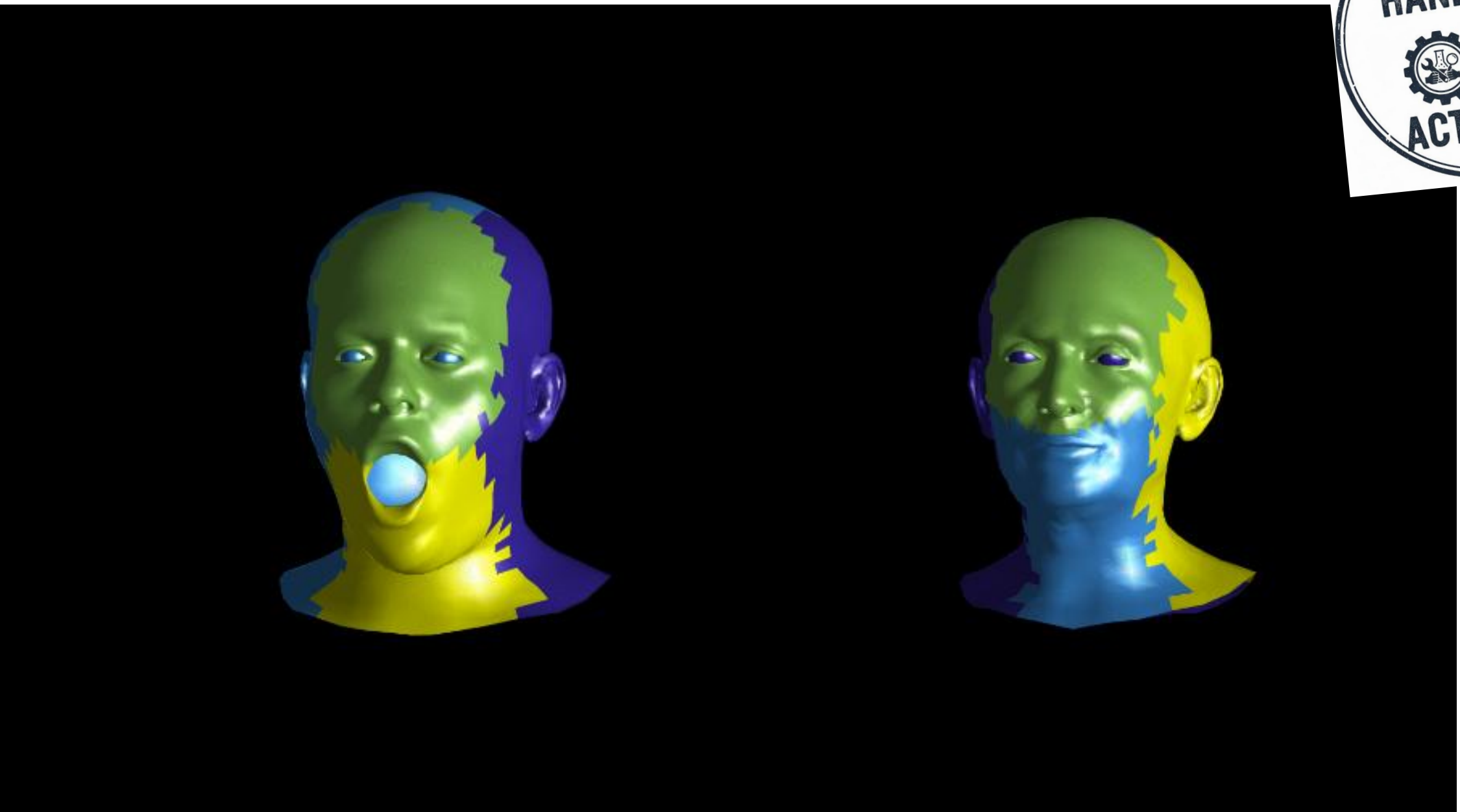
Localized shape modelling with global coherence: an inverse spectral approach, Pegoraro, Melzi, Castellani, Marin, Rodolà, SGP2022

(Pointwise features) Application: Shape Segmentation

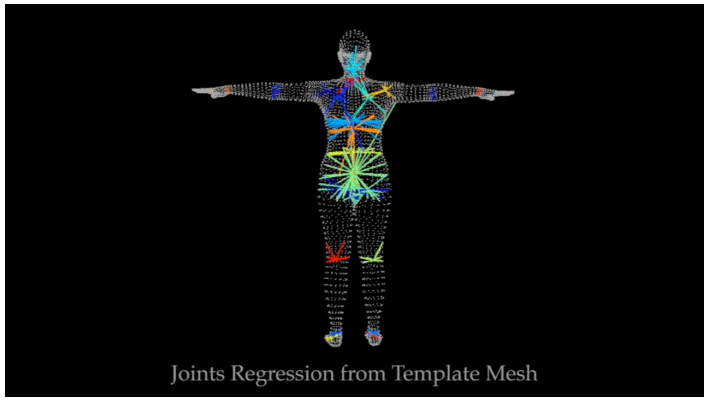


KMeans in feature space

(Pointwise features) Application: Shape Segmentation

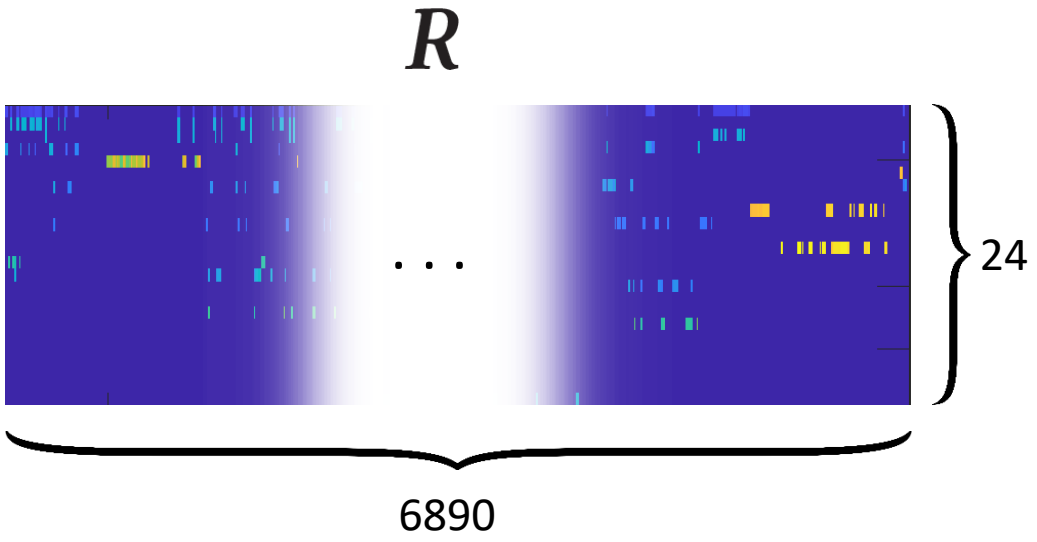


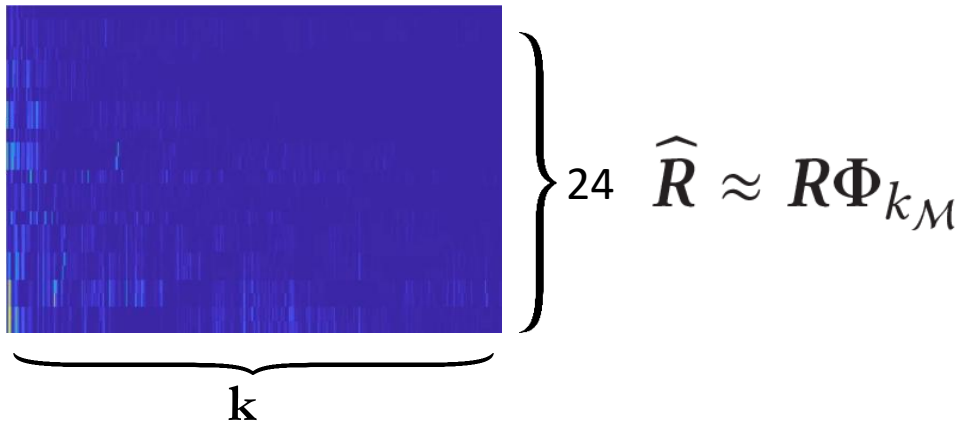
(Pointwise features) Application: Spectral regressor



$$J_{\mathcal{M}} = RX_{\mathcal{M}}$$

$$\approx \widehat{R}\widehat{X}_{\mathcal{M}} = \widehat{R}\Phi^{\dagger}_{k_{\mathcal{M}}} X_{\mathcal{M}}$$

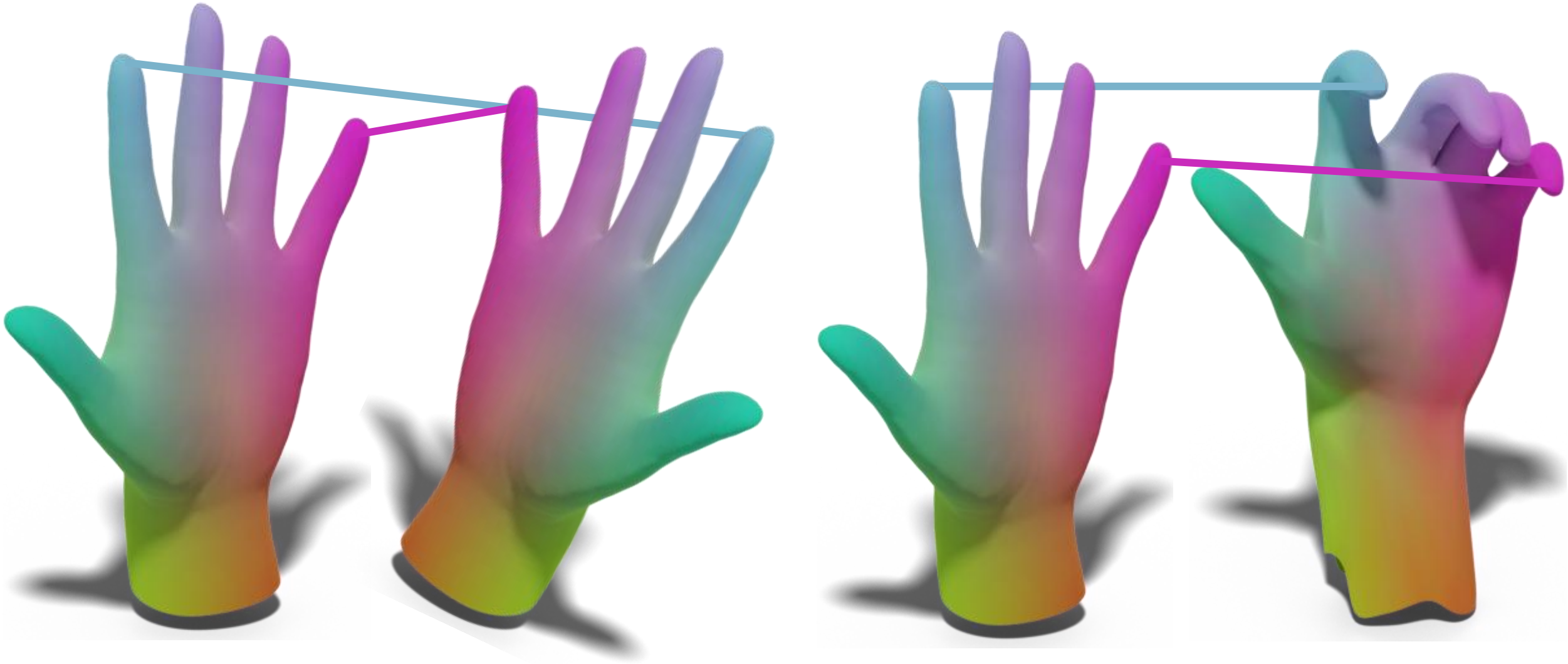




$$\widehat{R} \approx R\Phi_{k_{\mathcal{M}}}$$

A functional skeleton transfer, P Musoni, R Marin, S Melzi, U Castellani, ACM CGIT, 2021

Shape matching



Functional Map (2012)

Maks Ovsjanikov

Functional Maps: A Flexible Representation of Maps Between Shapes

Maks Ovsjanikov† Mirela Ben-Chen‡ Justin Solomon‡ Adrian Butscher‡ Leonidas Guibas‡
† LIX, École Polytechnique ‡ Stanford University

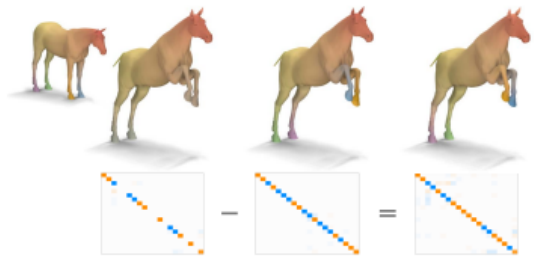


Figure 1: *Horse algebra: the functional representation and map inference algorithm allow us to go beyond point-to-point maps. The source shape (top left corner) was mapped to the target shape (left) by posing descriptor-based functional constraints which do not disambiguate symmetries (i.e. without landmark constraints). By further adding correspondence constraints, we obtain a near isometric map which reverses orientation, mapping left to right (center). The representation allows for algebraic operations on shape maps, so we can subtract this map from the ambivalent map, to retrieve the orientation preserving near-isometry (right). Each column shows the first 20x20 block of the functional map representation (bottom), and the action of the map by transferring colors from the source shape to the target shape (top).*

Abstract

We present a novel representation of maps between pairs of shapes that allows for efficient inference and manipulation. Key to our approach is a generalization of the notion of map that puts in correspondence real-valued functions rather than points on the shapes. By choosing a multi-scale basis for the function space on each shape, such as the eigenfunctions of its Laplace-Beltrami operator, we obtain a representation of a map that is very compact, yet fully suitable for global inference. Perhaps more remarkably, most natural constraints on a map, such as descriptor preservation, landmark correspondences, part preservation and operator commutativity become linear in this formulation. Moreover, the representation naturally supports certain algebraic operations such as map sum, difference and composition, and enables a number of applications, such as function or annotation transfer without establishing point-to-point correspondences. We exploit these properties to devise an efficient shape matching method, at the core of which is a single linear solve. The new method achieves state-of-the-art results on an isometric shape matching benchmark. We also show how this representation can be used to improve the quality of maps produced by existing shape matching methods, and illustrate its usefulness in segmentation transfer and in the joint analysis of shape collections.

Keywords: shape matching, representation, correspondence, shape maps, functional space

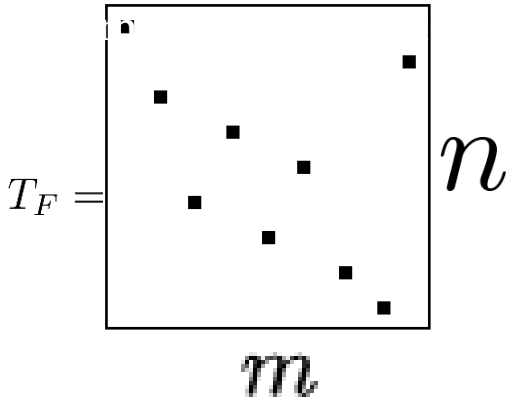
1 Introduction

Shape matching lies at the core of many operations in geometry processing. While several solutions to rigid matching are well established, non-rigid shape matching remains difficult even when the space of deformations is limited to e.g. approximate isometries. Part of the difficulty in devising a robust and efficient non-rigid shape matching method is that unlike the rigid case, where the deformation can be represented compactly as a rotation and translation, non-rigid shape matchings are most frequently represented as pairings (correspondences) of points or regions on the two shapes. This representation makes direct map estimation and inference intractable, since the space of possible point correspondences is exponential in size. For example, isometric matching techniques try to find correspondences that preserve geodesic distances as well as possible, but such optimization problems can be shown to be an NP-hard subclass of the quadratic assignment problem [Çela 1998]. Perhaps more importantly, this representation does not naturally support constraints such as map continuity or global consistency.

Additionally, in many practical situations, it is neither possible nor necessary to establish point-to-point correspondences between a pair of shapes, because of inherent shape ambiguities or because the user may only be interested in approximate alignment. Such ambiguous or approximate map inference is difficult to phrase in terms of point-to-point correspondences.

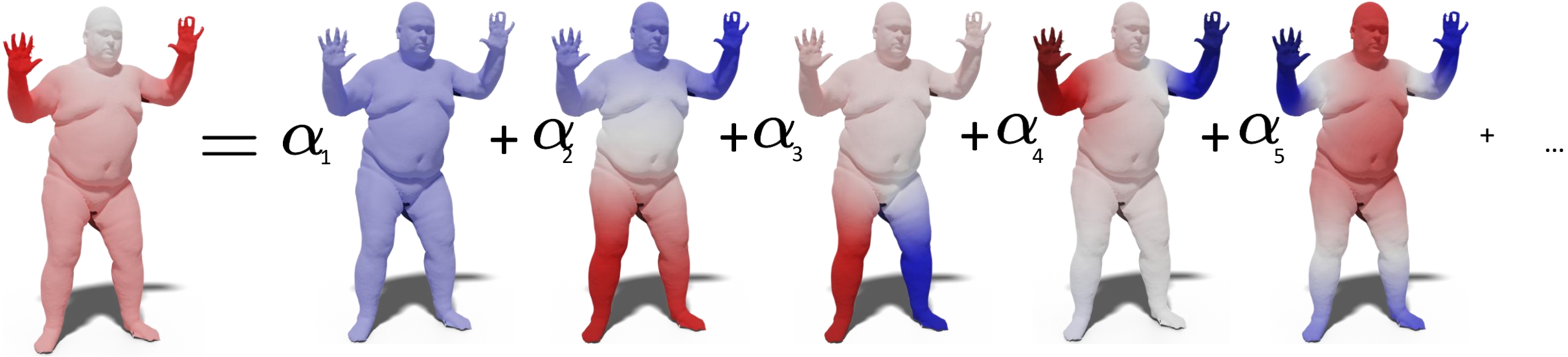
The majority of existing methods try to tackle these challenges by limiting their search for correspondences between a small set of landmark points and extending those to a dense set of correspondences on entire shapes during final post-processing ([Bronstein et al. 2006; Huang et al. 2008; Lipman and Funkhouser 2009; Kin-Chung Au et al. 2010; Ovsjanikov et al. 2010; Kim et al. 2011; Tevs et al. 2011; Sahillioğlu and Yemez 2011] among many others). This strategy has also been justified theoretically, since under general conditions a small set of landmark correspondences is known to be sufficient to obtain a unique dense mapping between isometric surfaces ([Lipman and Funkhouser 2009; Ovsjanikov et al. 2010]).

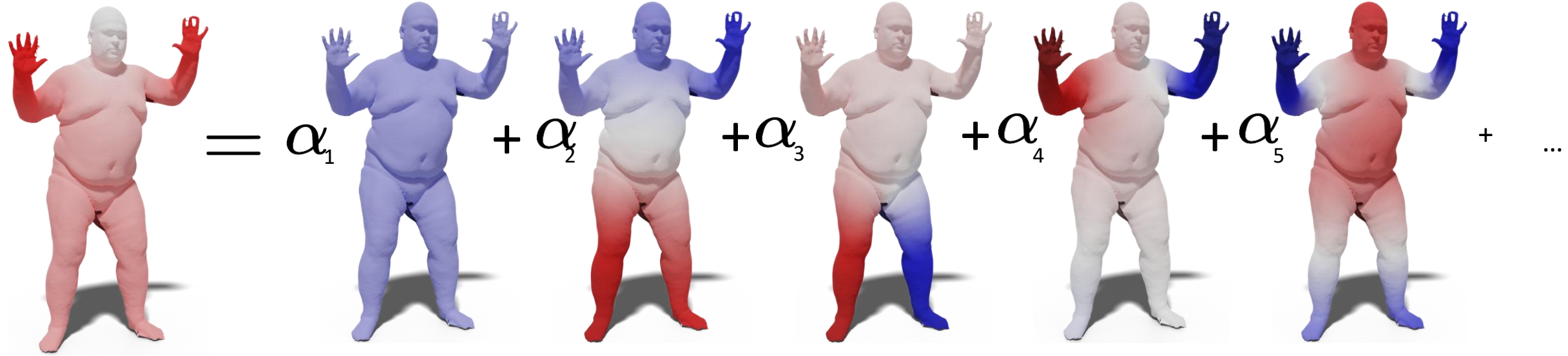
Matching is generally represented as a permutation matrix

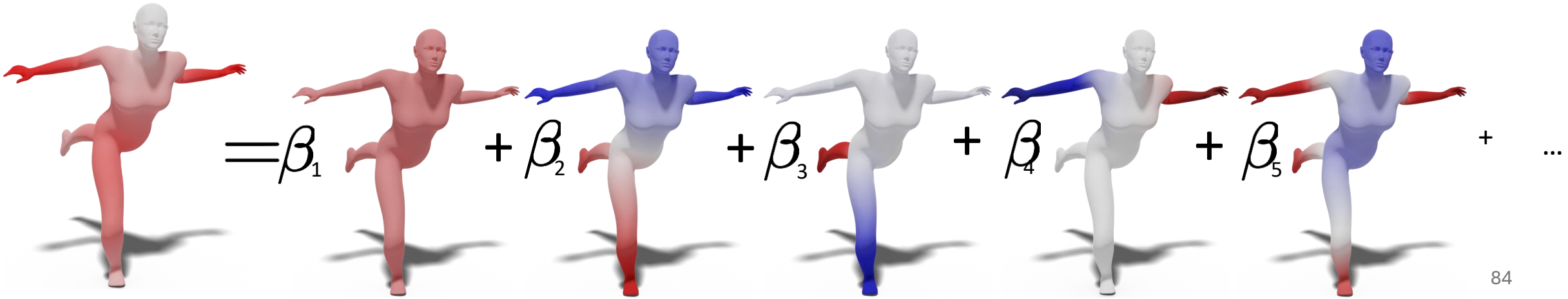


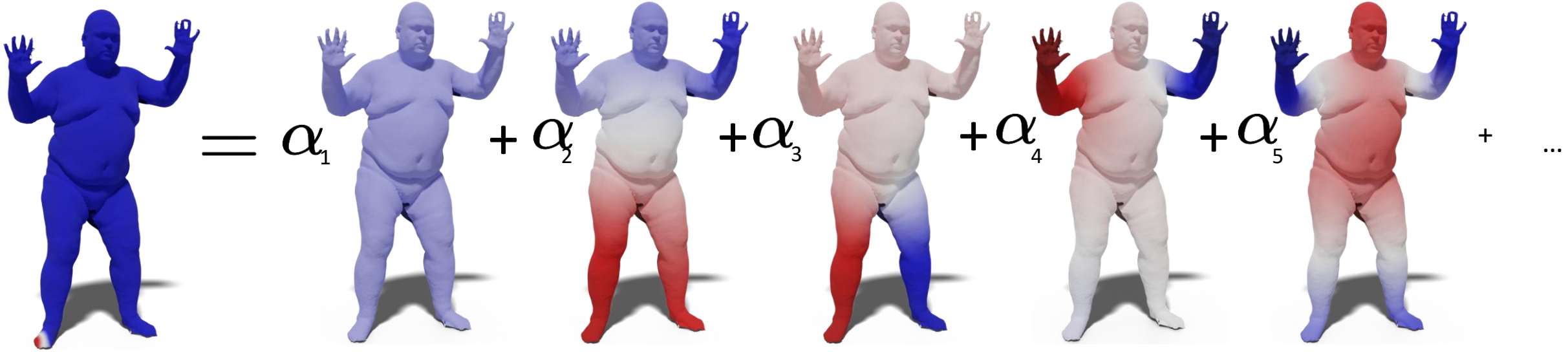
Big, sparse, discrete values

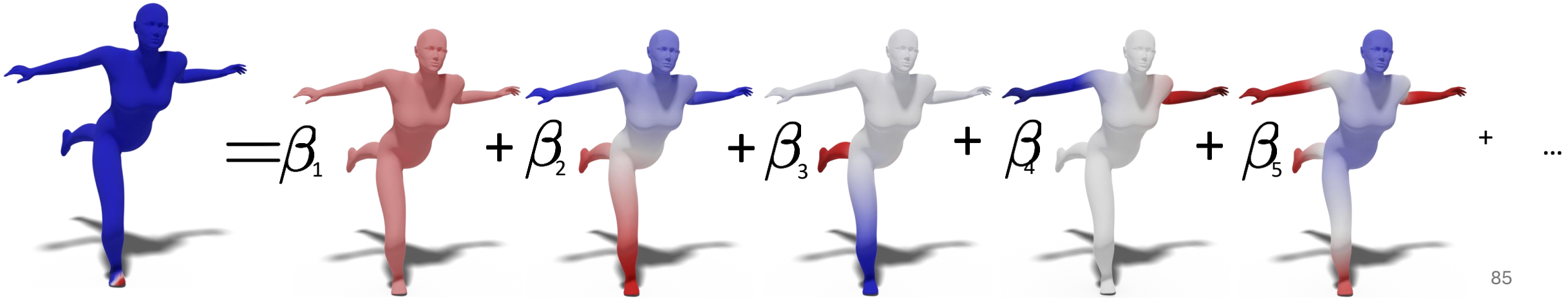
Motivation:
Can we express a correspondence in a more compact way?

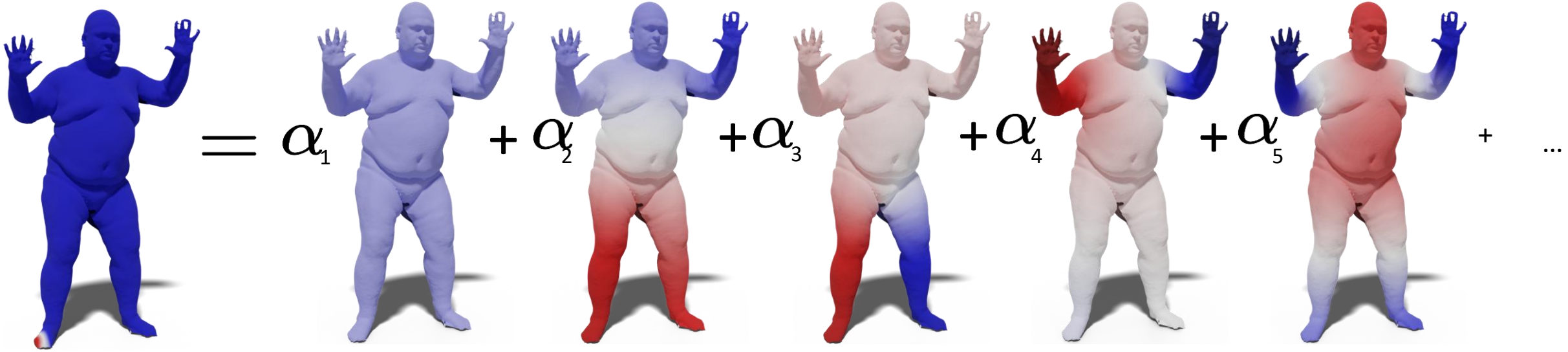

= α1 + α2 + α3 + α4 + α5 + …


= α1 + α2 + α3 + α4 + α5 + …

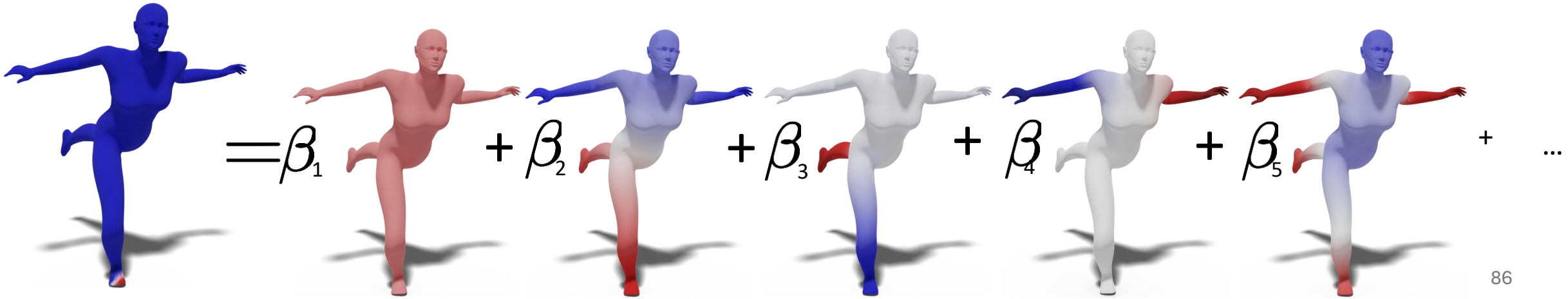

= β1 + β2 + β3 + β4 + β5 + …

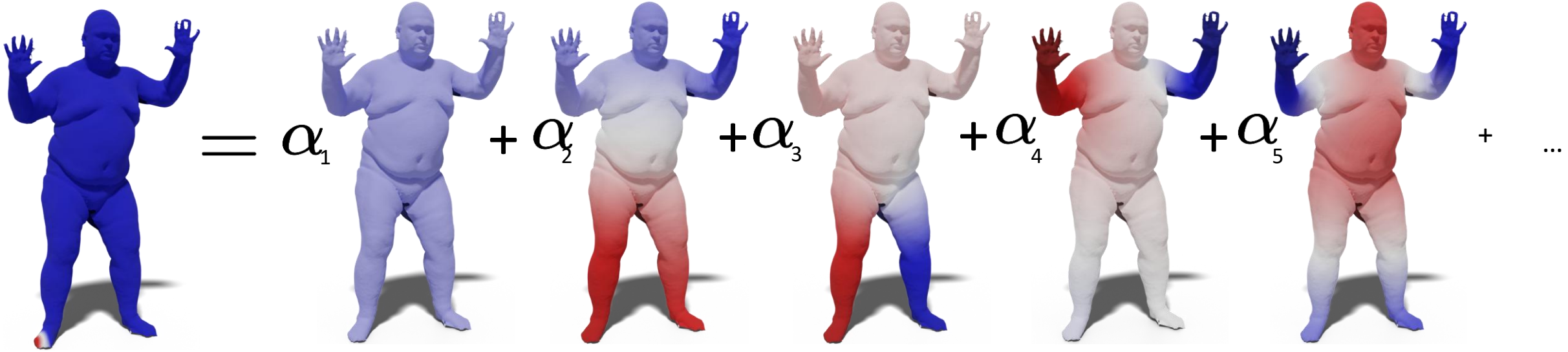

= α1 + α2 + α3 + α4 + α5 + …


= β1 + β2 + β3 + β4 + β5 + …



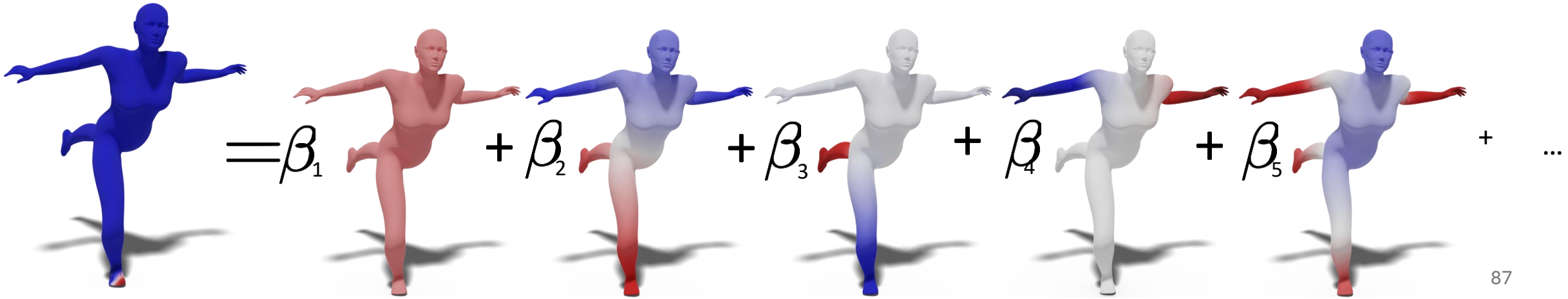
$$B = CA$$

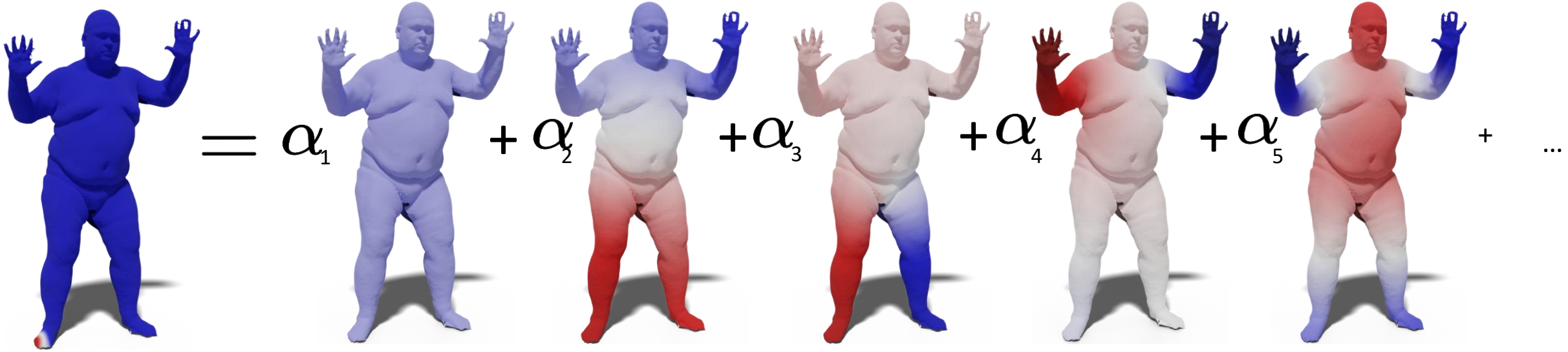




$$B = CA$$

$$\arg\min_{C} ||B - CA|| + reg(C)$$

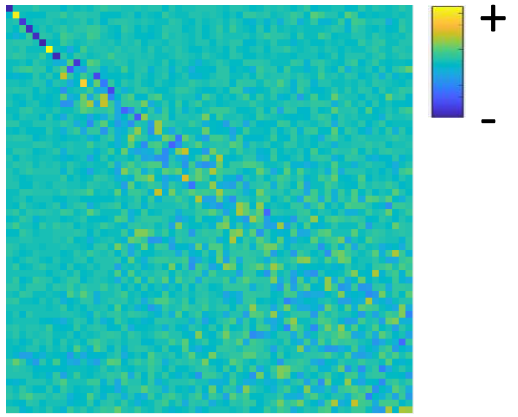


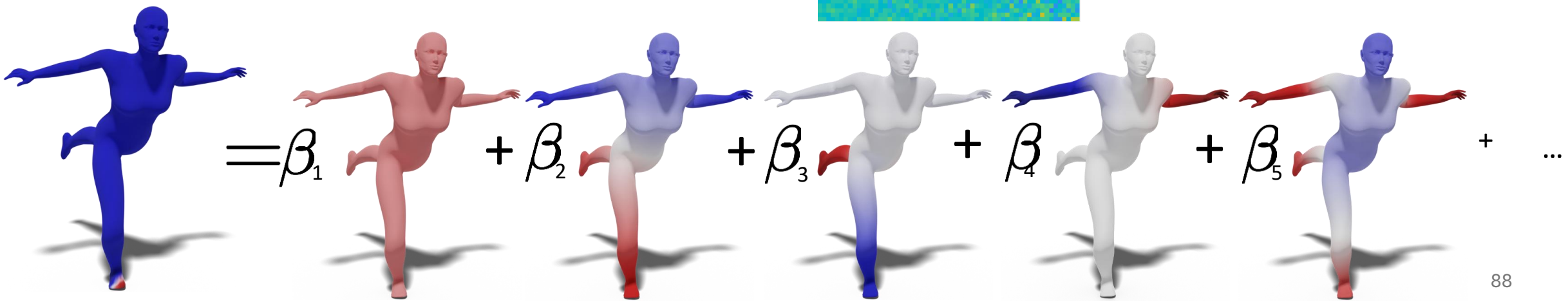


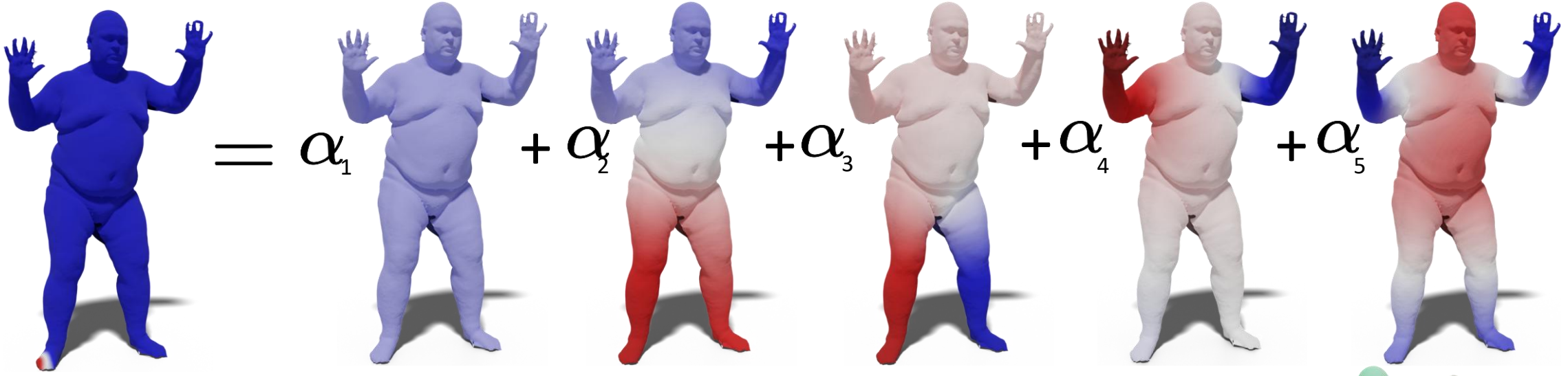
$$B = CA$$

$$\arg\min_{C} ||B - CA|| + reg(C)$$

$C =$



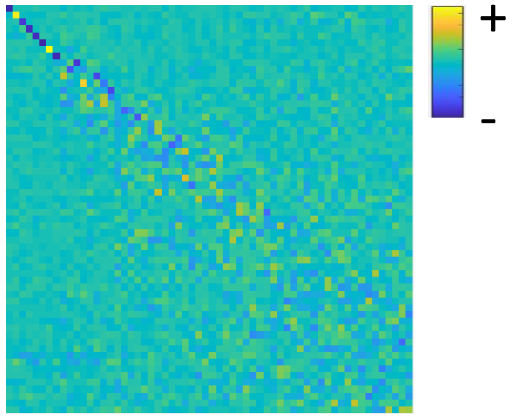


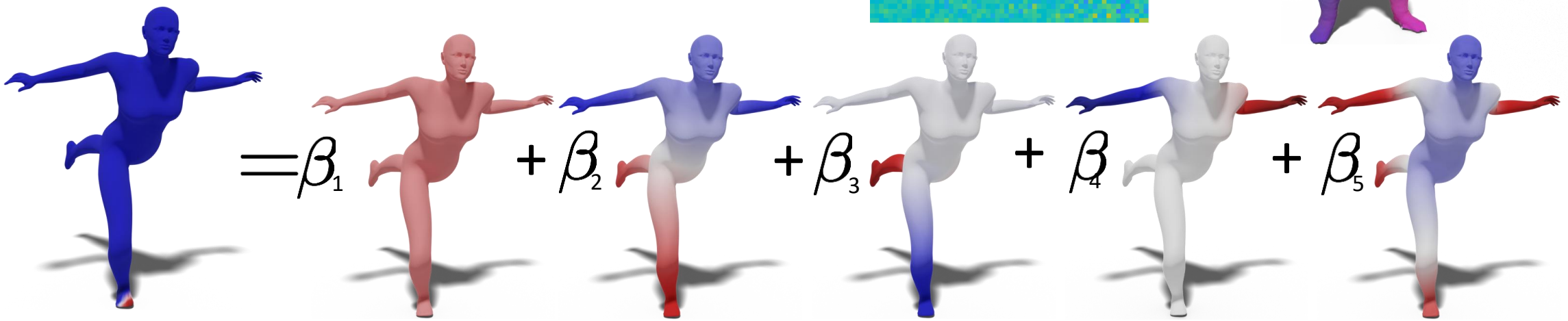


$$B = CA$$

$$\arg\min_C ||B - CA|| + reg(C)$$

$C =$



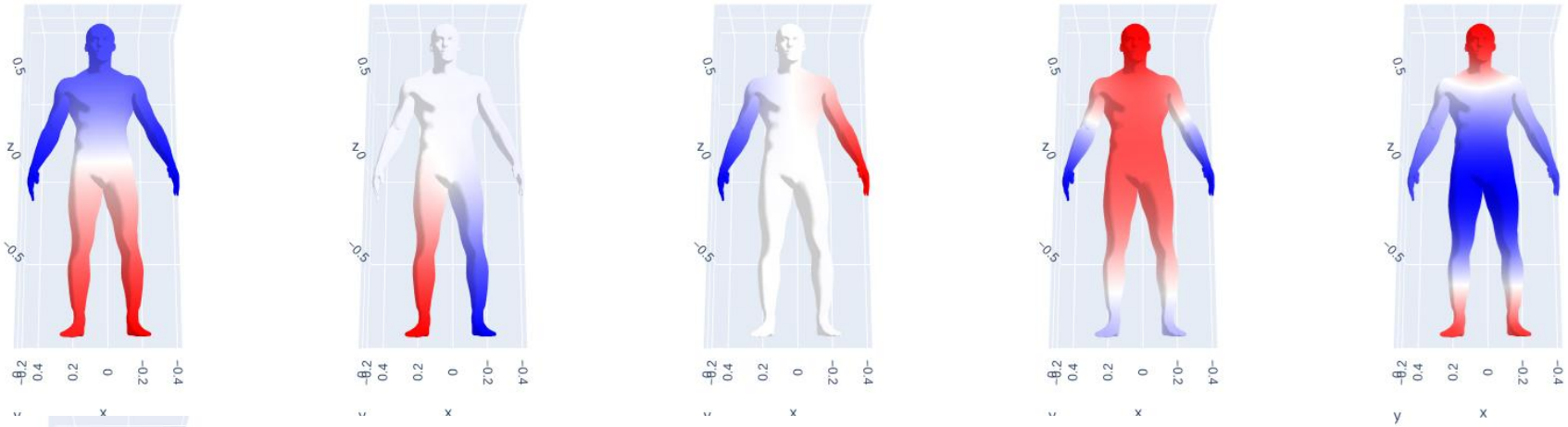


Convert a functional map into a pointwise correspondence

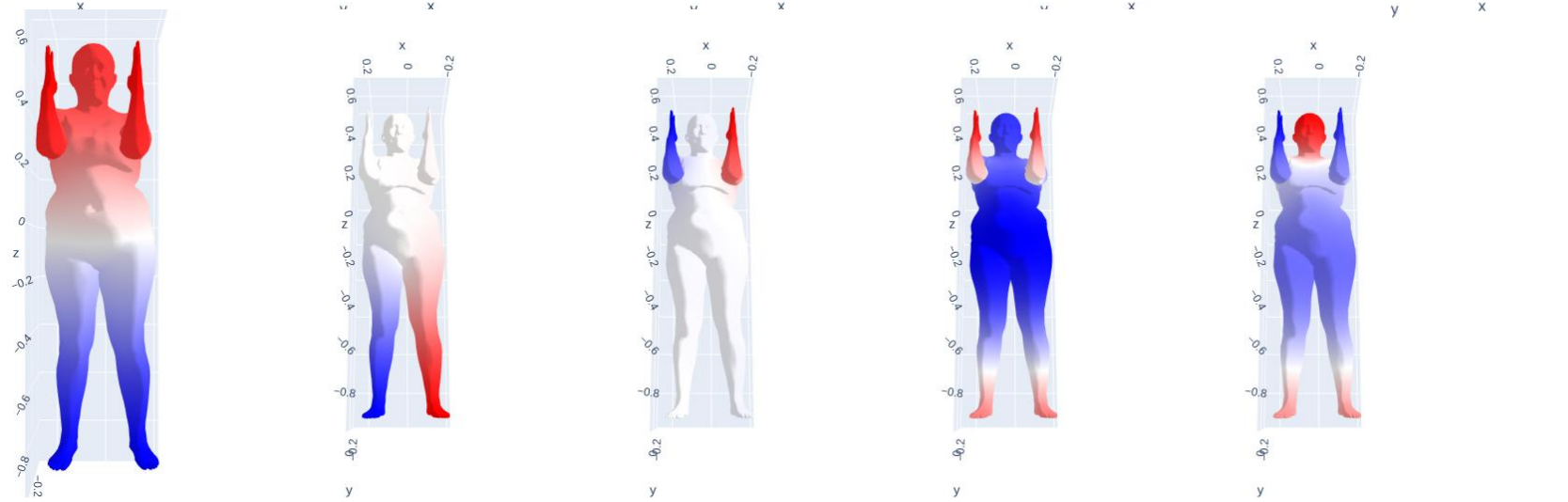
We use the functional map to align functions and we do nearest neighbor in the feature space

$$\Pi = KNN(\Phi_y, \Phi_x C)$$

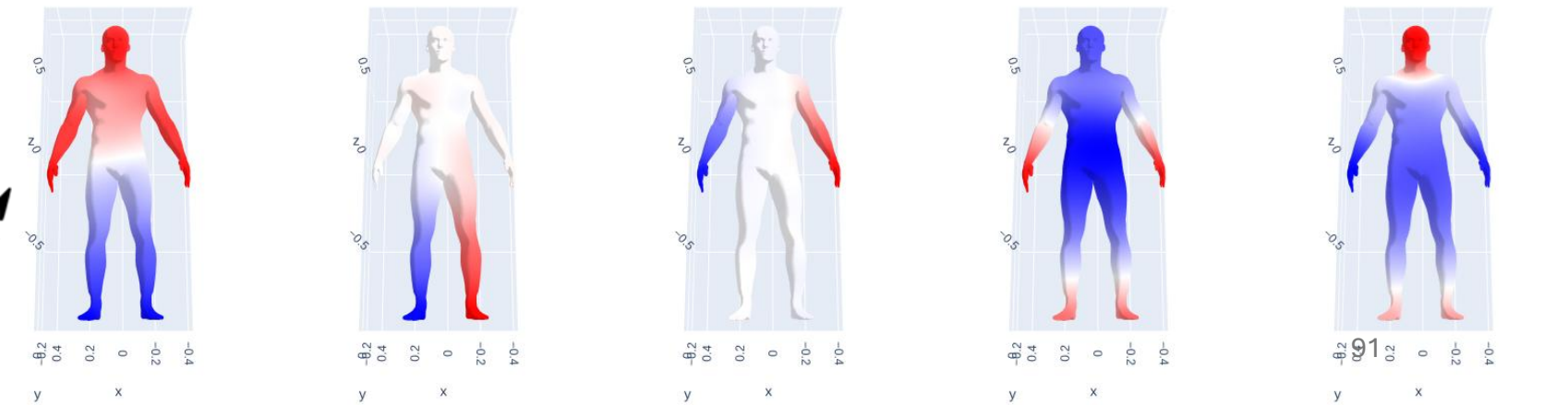
Φ_x



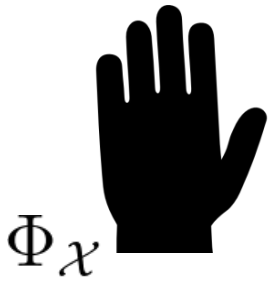
Φ_y



$\Phi_x C$



Deep Learning Intuition



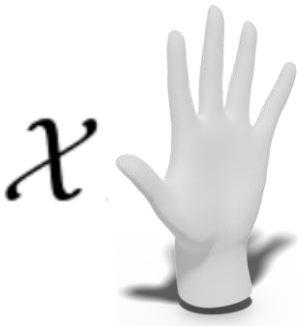
$\Phi_{\mathcal{X}}$

Learned
Features
Embedding

$\Phi_{\mathcal{Y}}$

$$\Phi_{\mathcal{X}} = \Phi_{\mathcal{Y}}$$

- **Learning Pointwise structure**
+ Natural Alignement



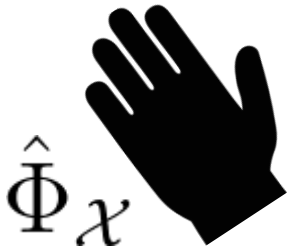
$\mathcal{X}$

3D Embedding

$\mathcal{Y}$

$$f(\mathcal{X}) = \mathcal{Y}$$

+ Natural Embedding
- **Registration is complex**



$\hat{\Phi}_{\mathcal{X}}$

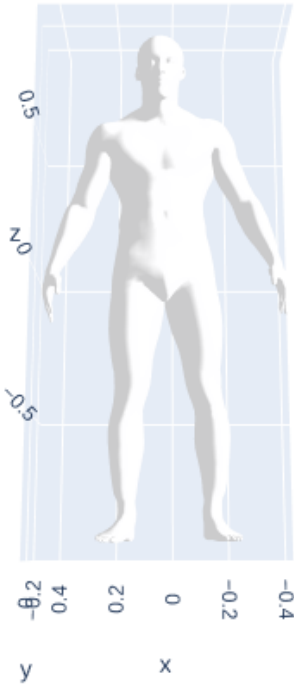
FMAP

$\hat{\Phi}_{\mathcal{Y}}$

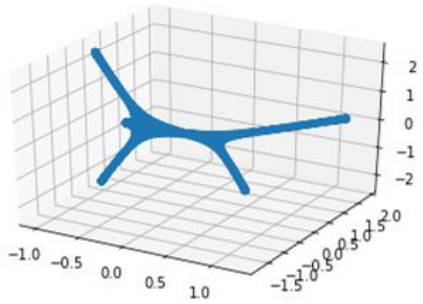
$$\hat{\Phi}_{\mathcal{X}} A = \hat{\Phi}_{\mathcal{Y}}$$

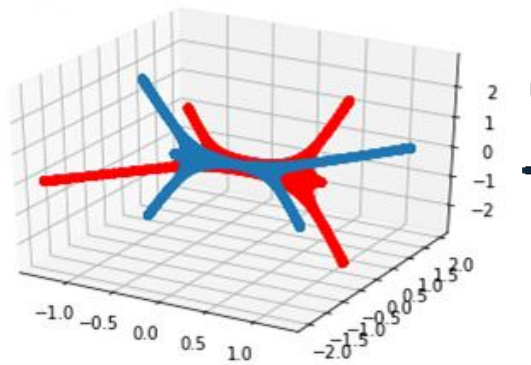
+ **Retain global structure**
+ **Alignement is simple**

Correspondence Learning via Linearly-invariant Embedding; R Marin, MJ Rakotosaona, S Melzi, M Ovsjanikov; NeurIPS, 2020

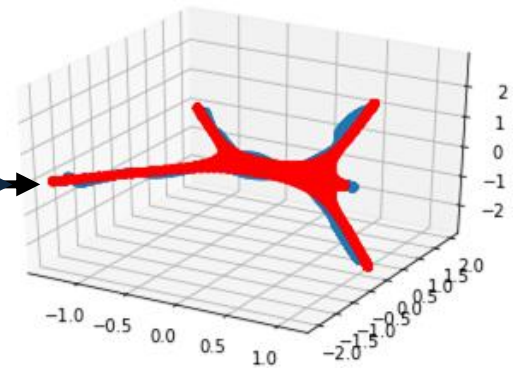


Φ_x

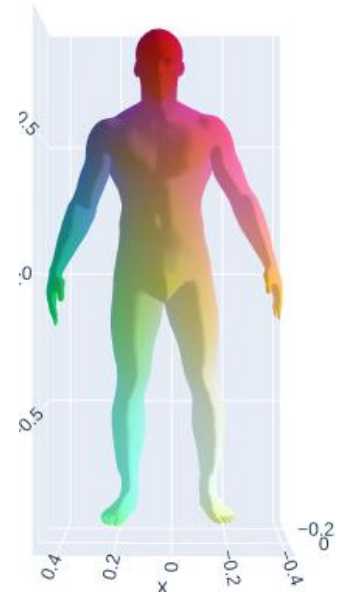


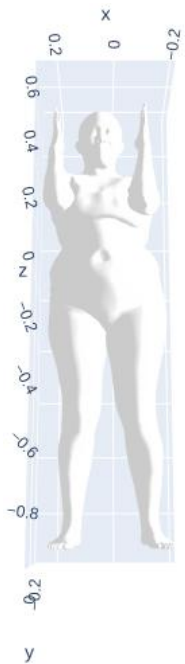


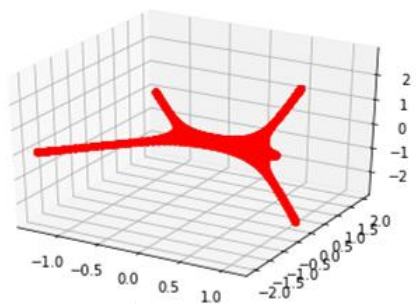
C



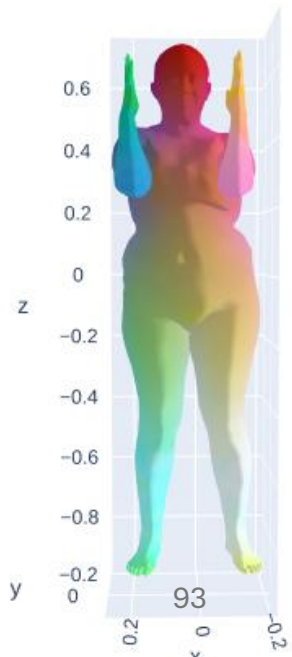
$\Pi = KNN(\Phi_y, \Phi_x C)$

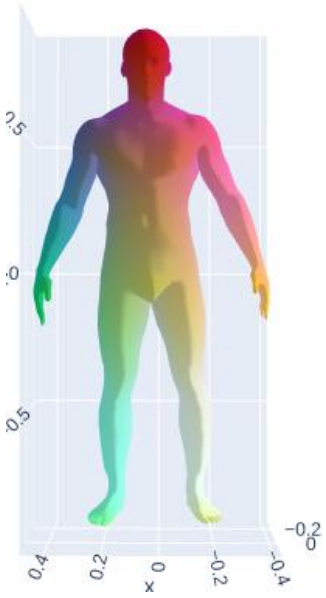






Φ_y

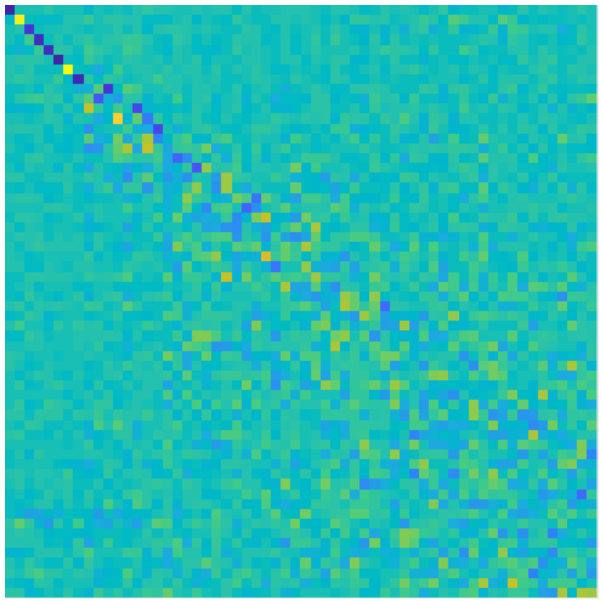


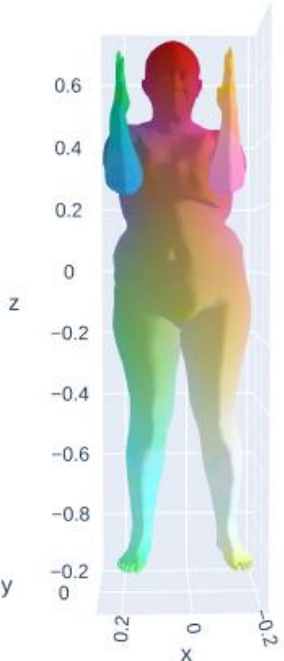


Π

$$\Phi_x C = \Pi \Phi_y$$

$$C = \Phi_x^{\dagger} \Pi \Phi_y$$





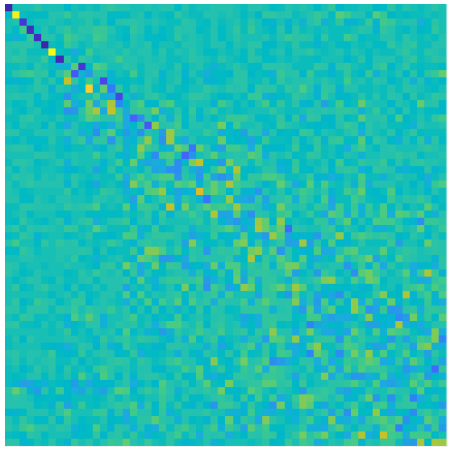
What happen if we iterate the two?

Starting from a correspondence

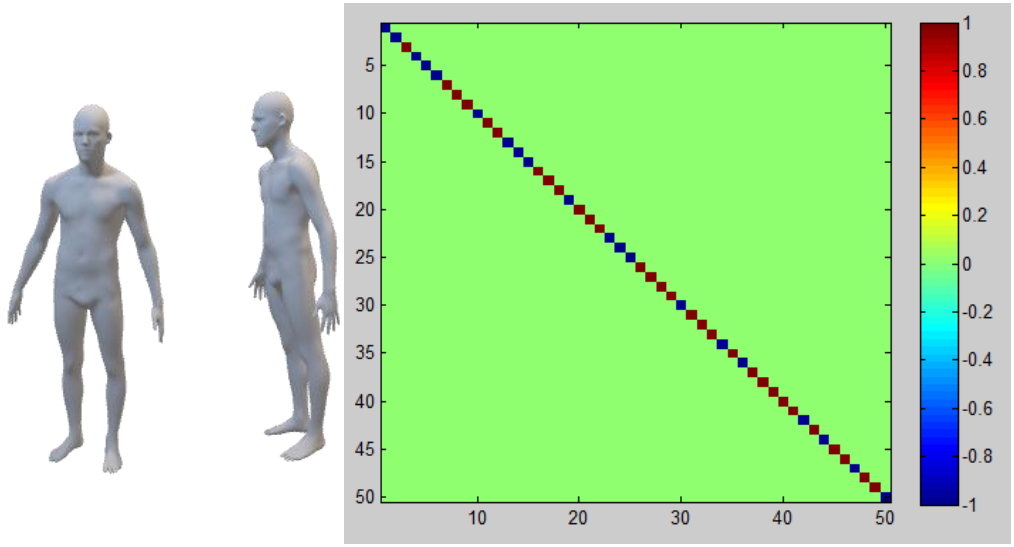
We convert it into a functional map C

We convert it back into a correspondence

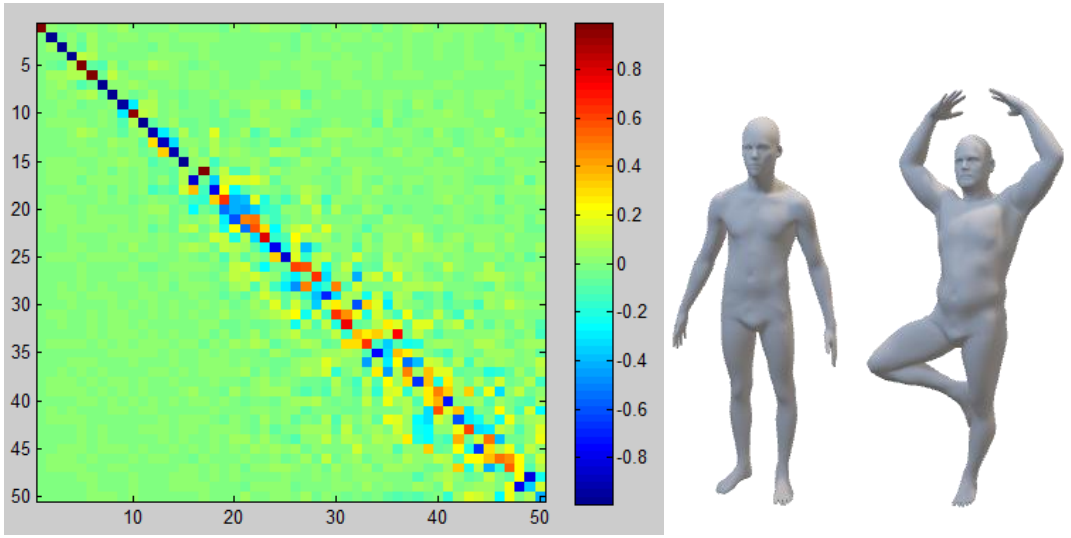
$$\Pi \longrightarrow C = \Phi_{\mathcal{X}}^{\dagger} \Pi \Phi_{\mathcal{Y}} \longrightarrow \hat{\Pi} = KNN(\Phi_{\mathcal{Y}}, \Phi_{\mathcal{X}} C) \longrightarrow$$



What happen if we iterate the two?



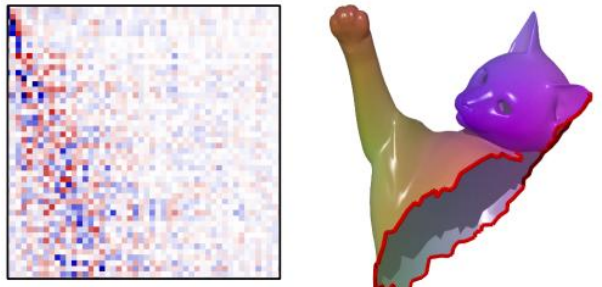
near-isometry

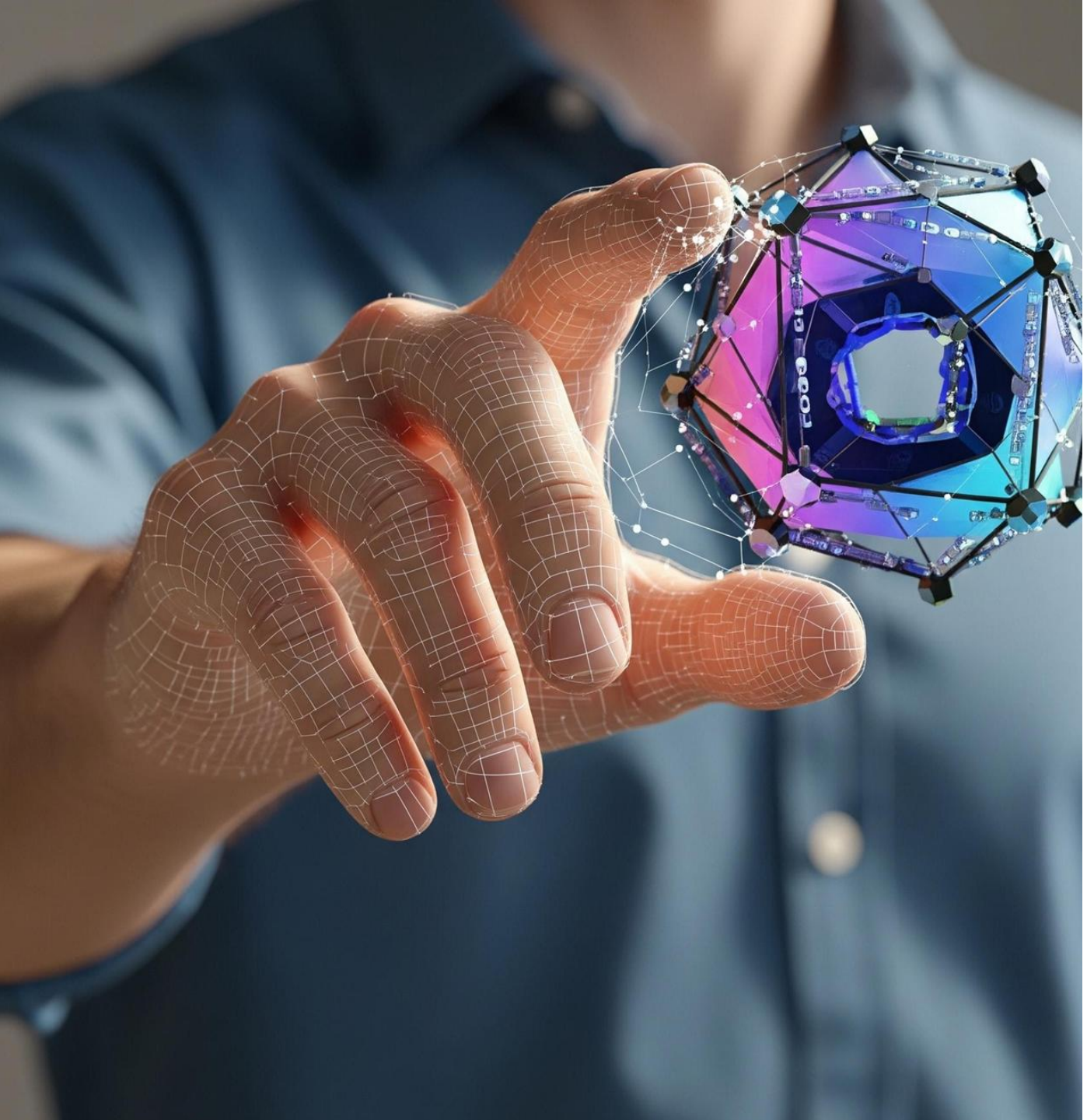


We can **enforce properties**
(isometries, volume preservation, commutativity with operators, …) and it is **differentiable** (i.e., you can learn it)

It is **general**
(e.g., graphs, images, partiality, any kind of basis, …)

Applications beyond matching
(symmetry detection, segmentation transfer, …)

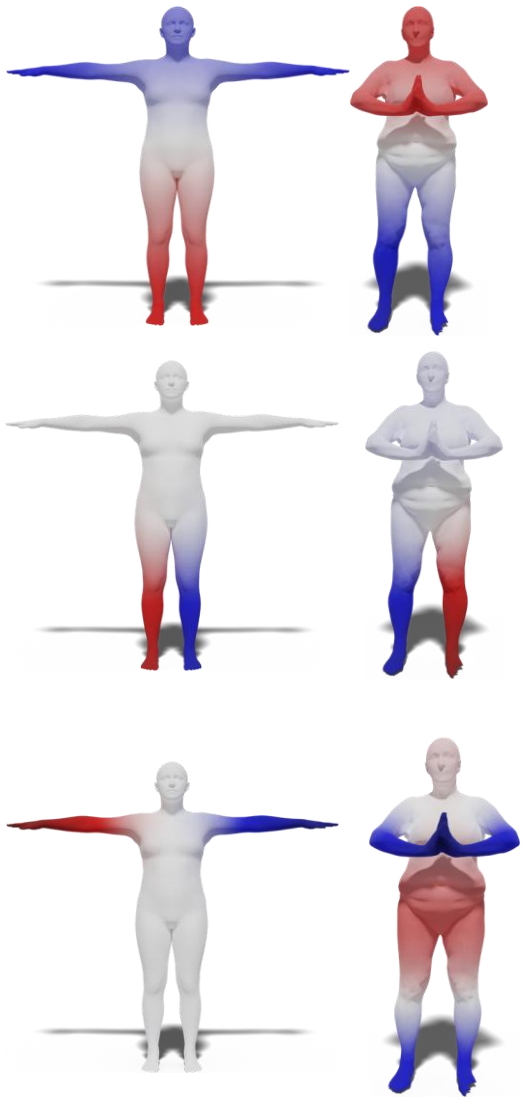




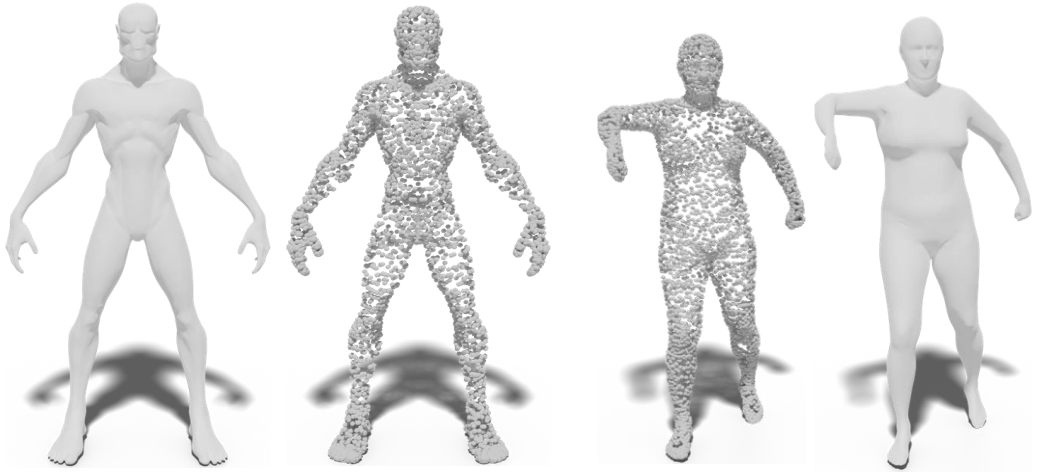
Registration

Main issues

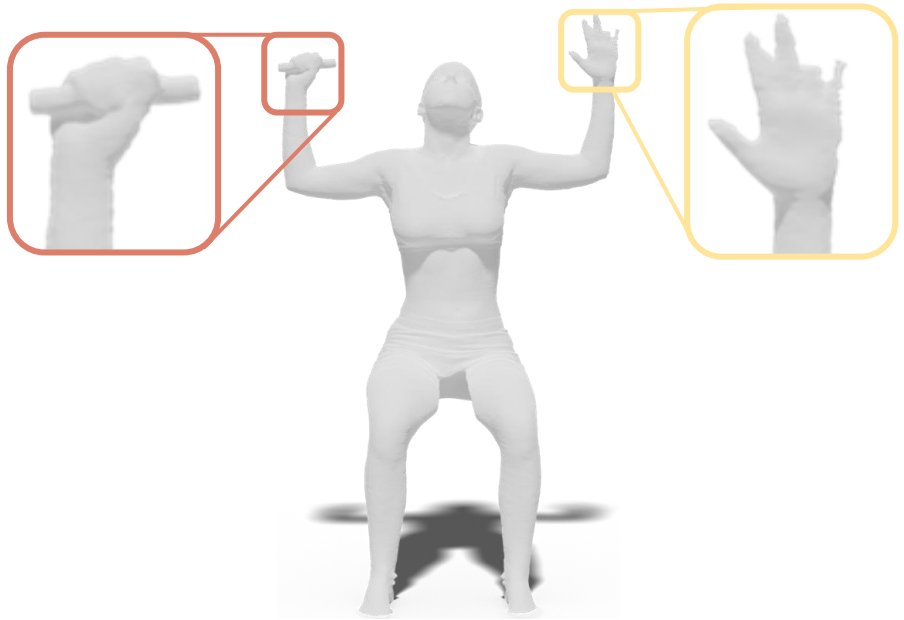
a) Topological noise



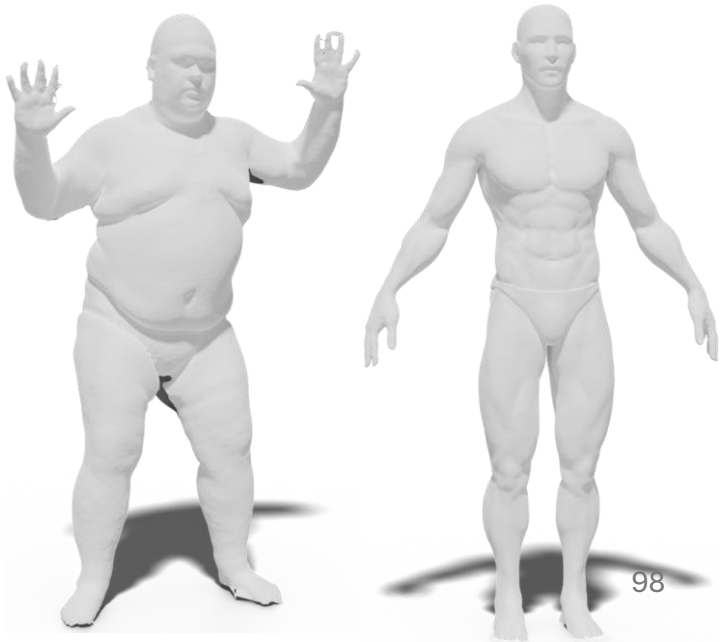
b) Pointclouds



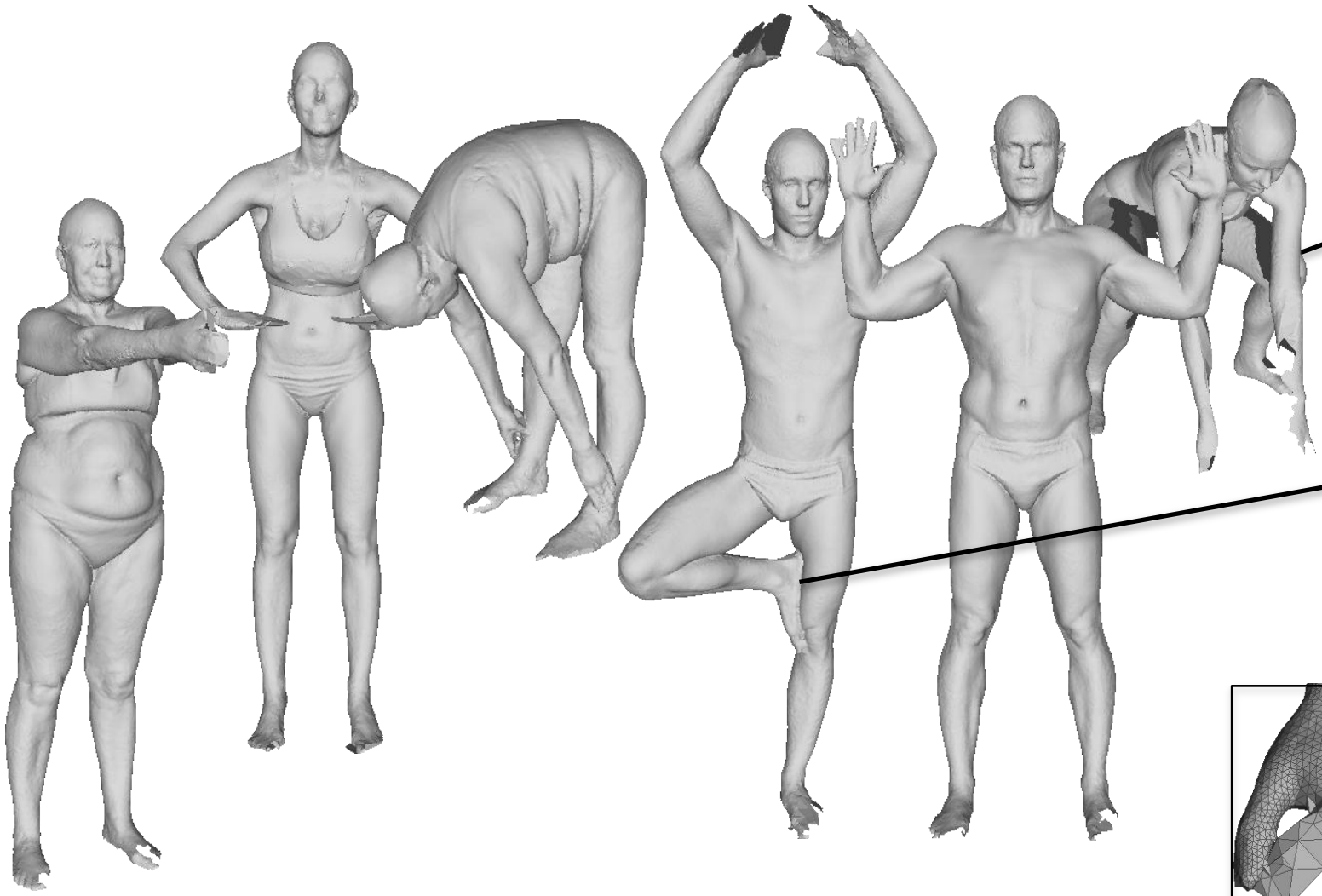
c) Clutter\Partiality

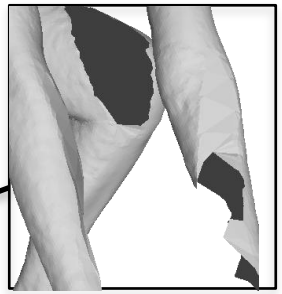


c) Heavy non-isometries



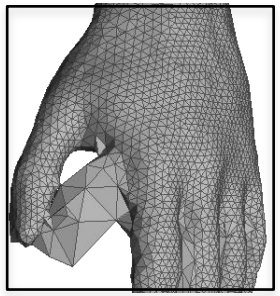
Human Registration has all of these issues

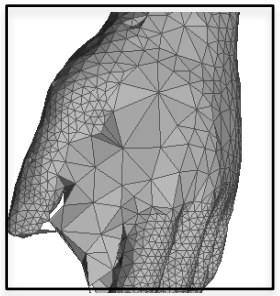




Missing Data

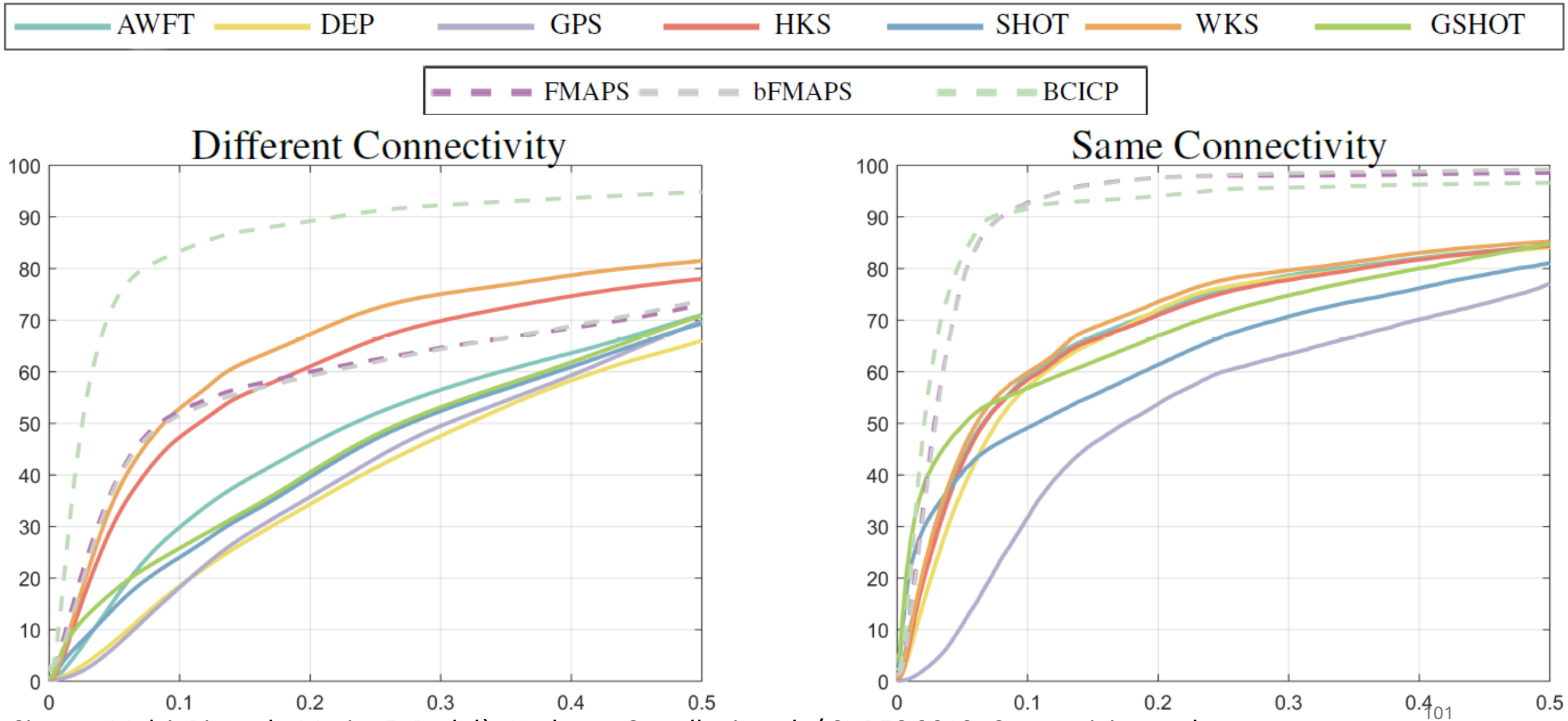
Self-Contact





Varying topology

Connectivity matters for matching methods



Simone Melzi, Riccardo Marin, E. Rodolà, Umberto Castellani et al. / SHREC 2019: Connectivity track

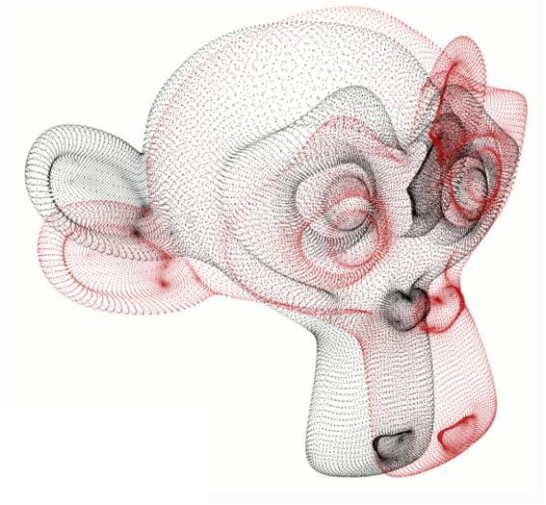
Point Clouds Registration

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)

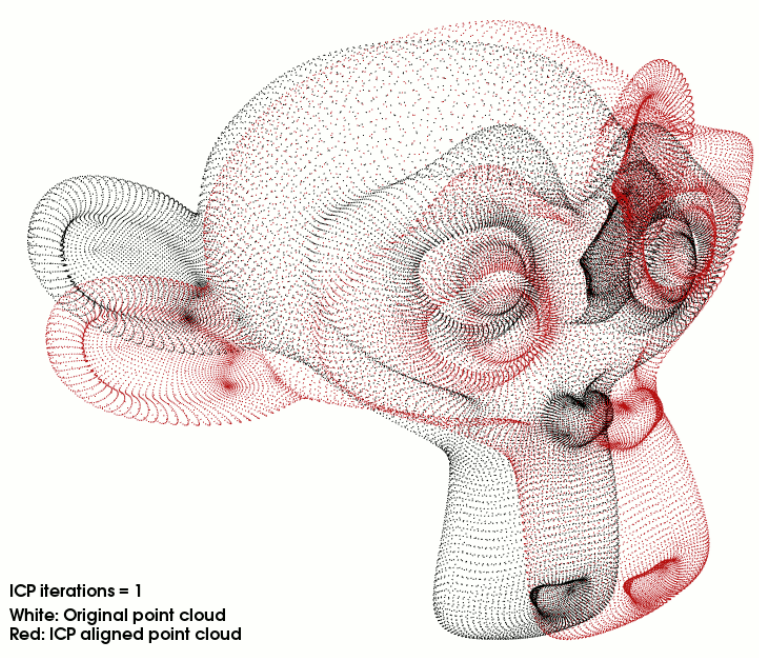
$$\widetilde{\Pi} = \arg\min_{\Pi} \|\widetilde{F}(\mathbf{X}) - \Pi(\mathbf{Y})\|_2^2$$

Iterate

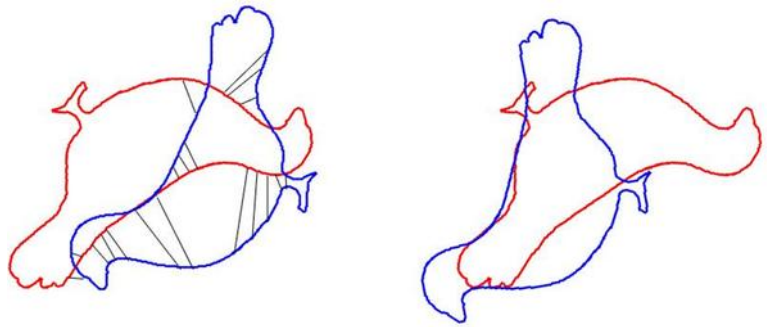
Solving for the deformation
(Optimization or closed form)

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

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

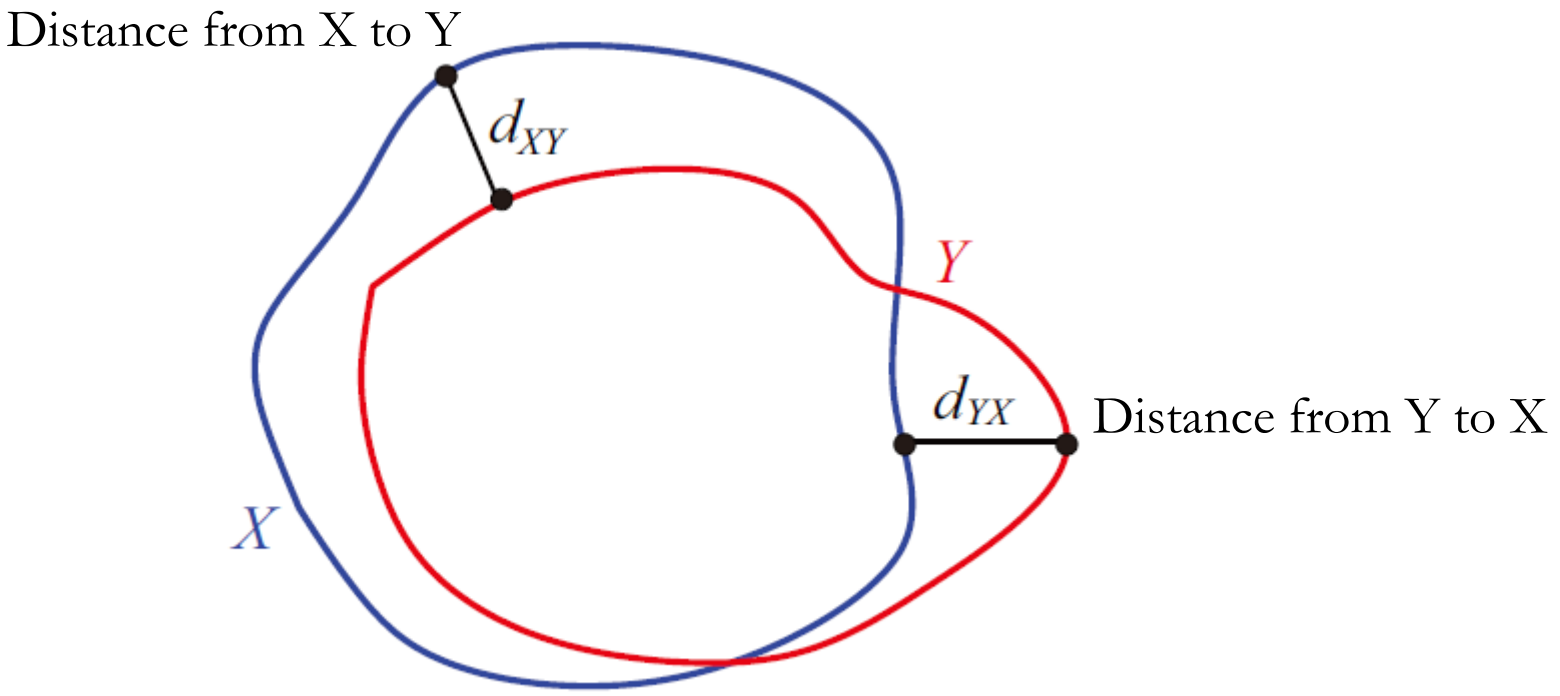


Prone to local minima



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

Optimize the deformation – Loss between sets



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

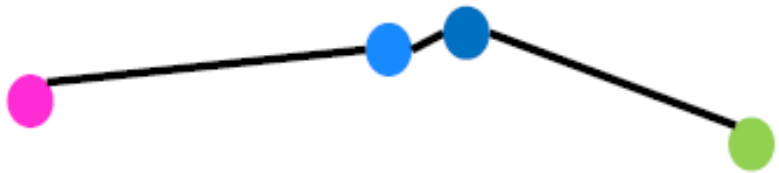
Basically, we minimize the distances of a bidirectional nearest-neighbor

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

Possible Geometrical regularizations

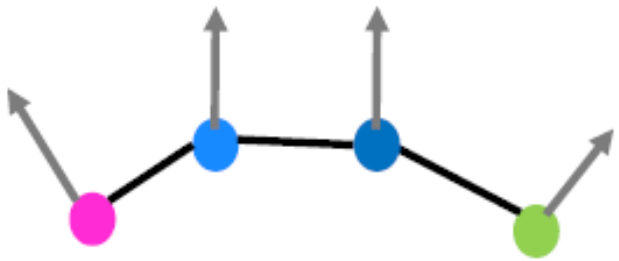
Init

Example of high loss



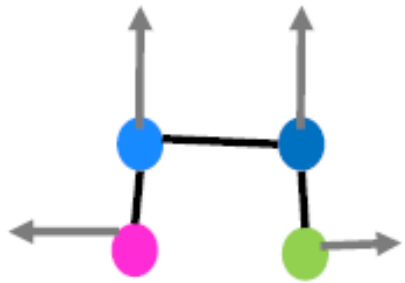
Edges Loss

Penalizes local distortions



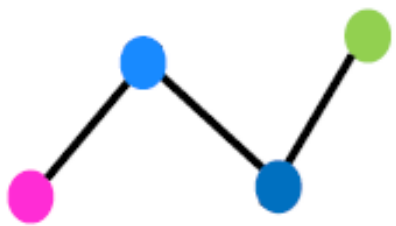
Normal Loss

Penalizes local misorientations



Laplacian Loss

Penalizes non-smooth geometries



As-Rigid-As-Possible Loss

Penalizes non-rigid deformations

Human Registration - Regularizations

- Optimizing SMPL is a **regularization**
  (optimizing 82 values instead of 6890x3)
- We can further **regularize the parameters:**

  For the identity:

  1) generally, L2\L1 regularization for the beta parameters

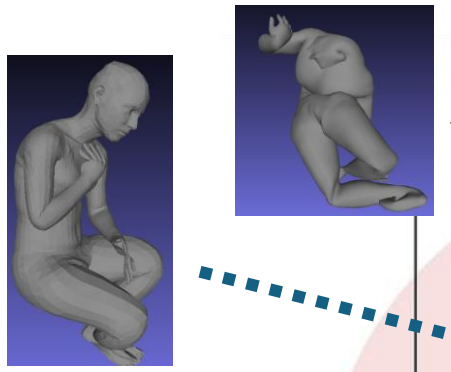
  For the pose:

  1) L2 loss on the angles (avoiding unnecessary twists)
  2) Hand-crafted limits for the joints' rotations
  3) Learning a Pose prior (penalizing out-of-distribution poses)

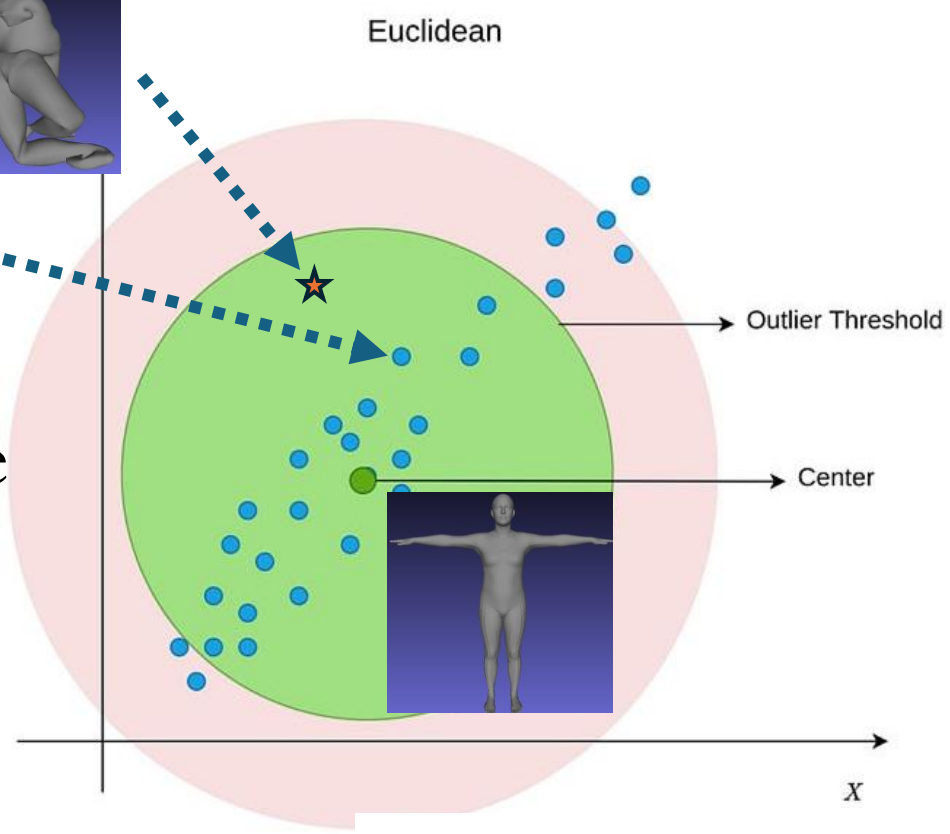
https://github.com/nghorbani/human_body_prior

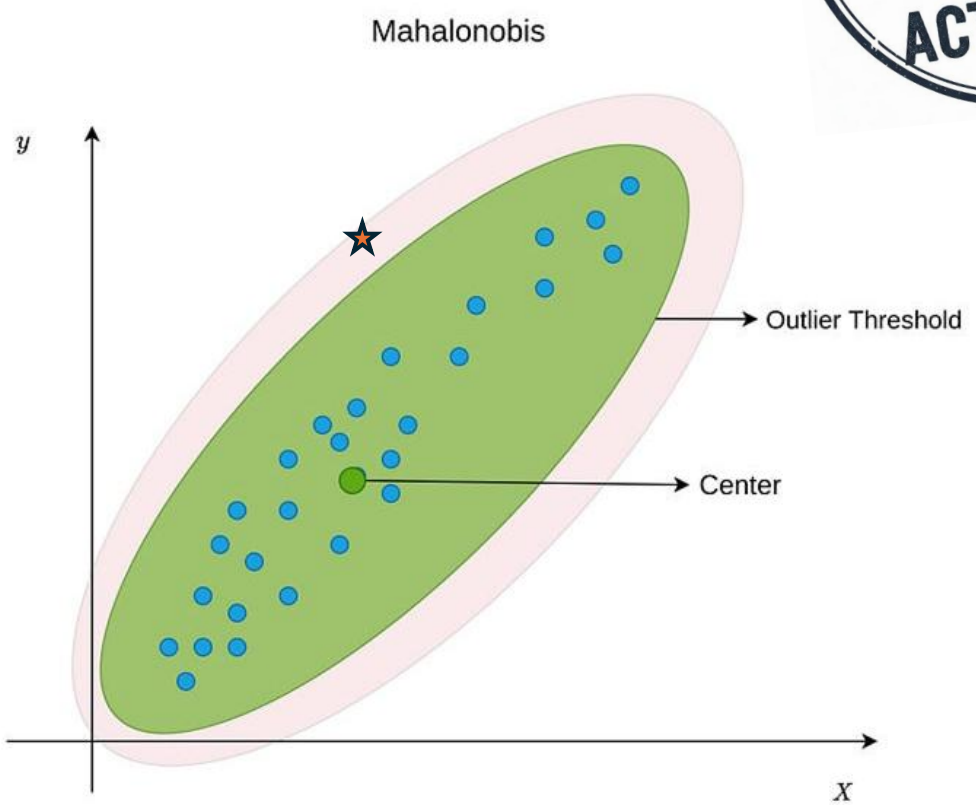
Human Registration - Regularizations

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



Pose space



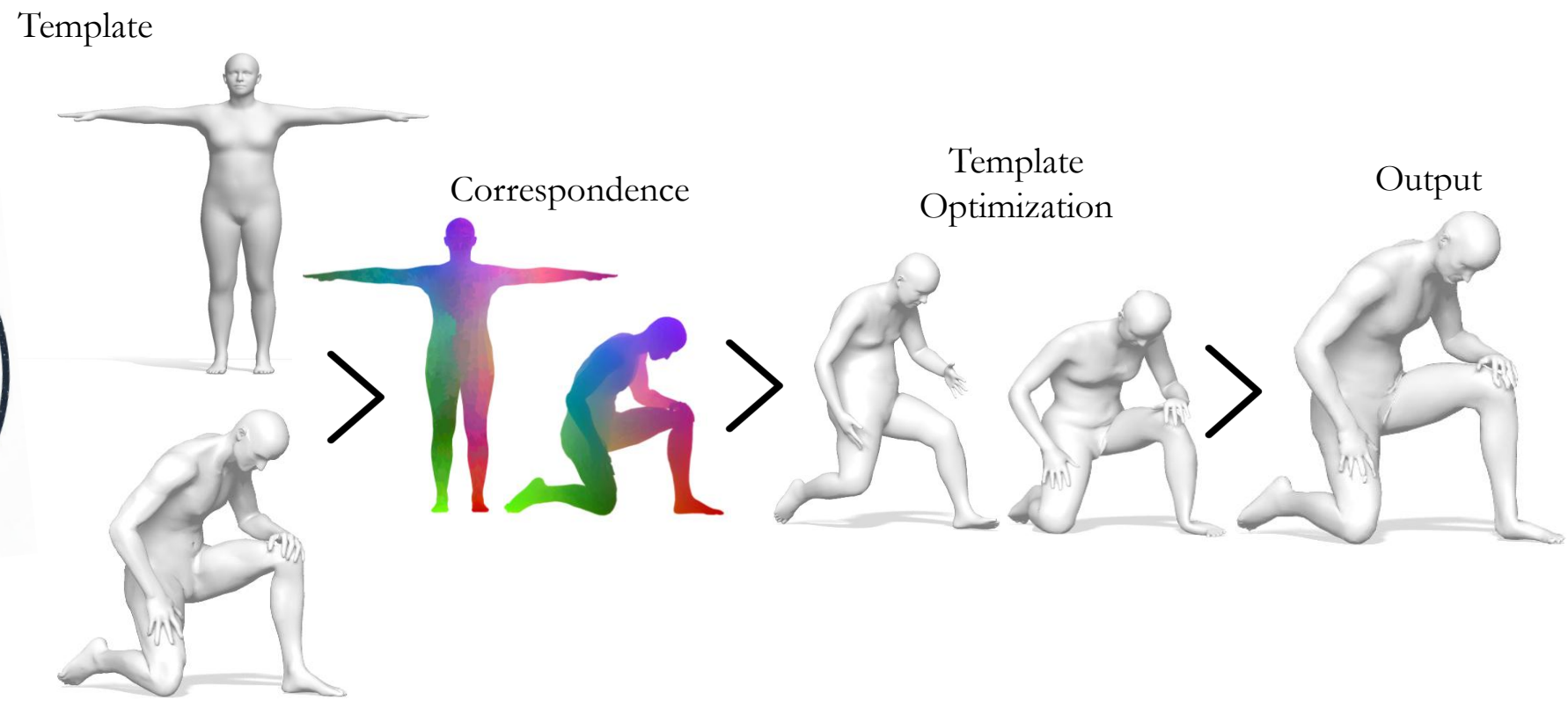


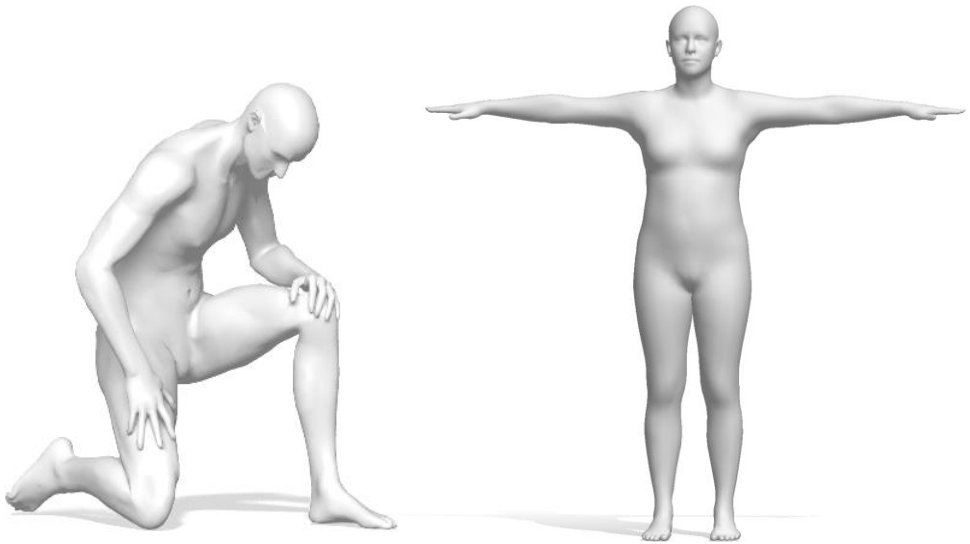
$$d_M(\vec{x}, Q) = \sqrt{(\vec{x} - \vec{\mu})^{\mathsf{T}} \boldsymbol{\Sigma}^{-1} (\vec{x} - \vec{\mu})}.$$

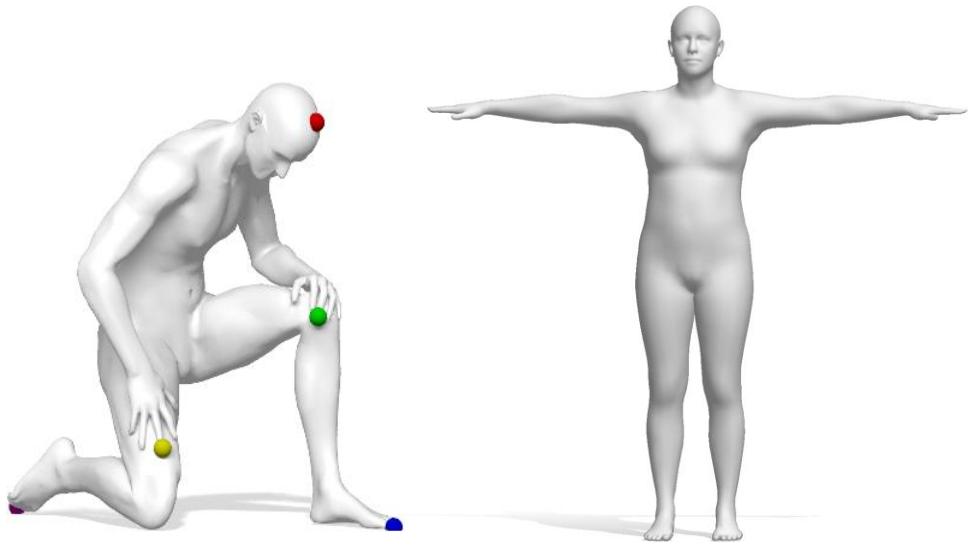
https://github.com/nghorbani/human_body_prior

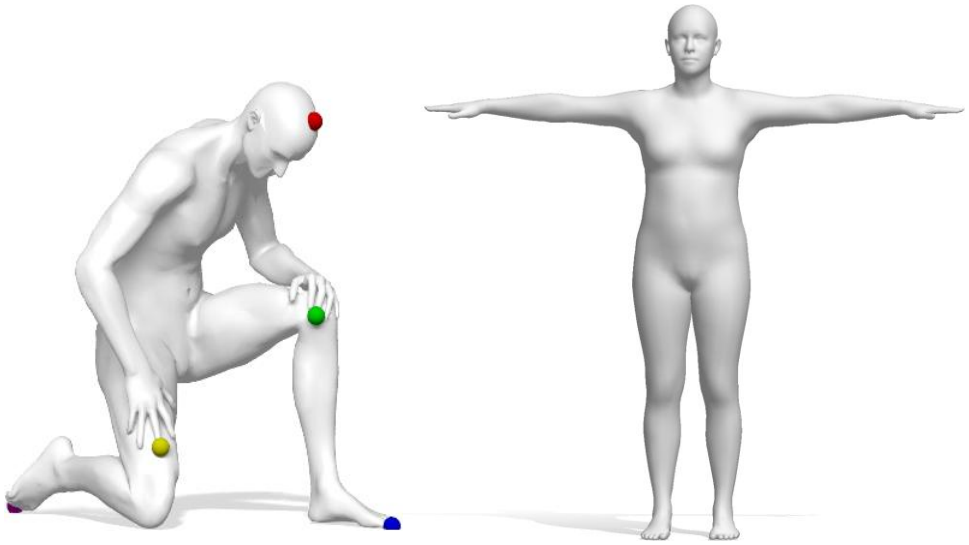
FARM:
Functional Automatic Registration Method for 3D Human Bodies

R. Marin[1†], S. Melzi[1†], E. Rodolà[2] and U. Castellani[1]



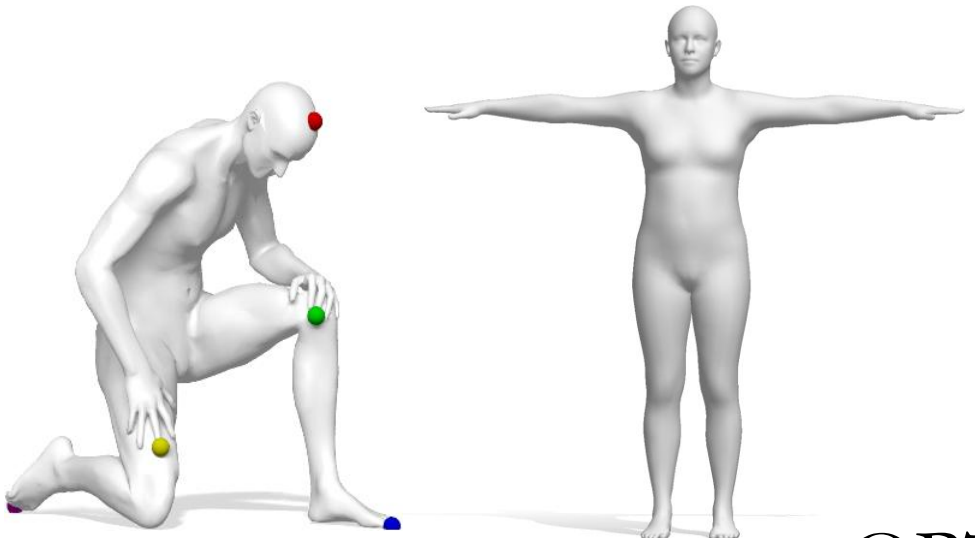






FMAP

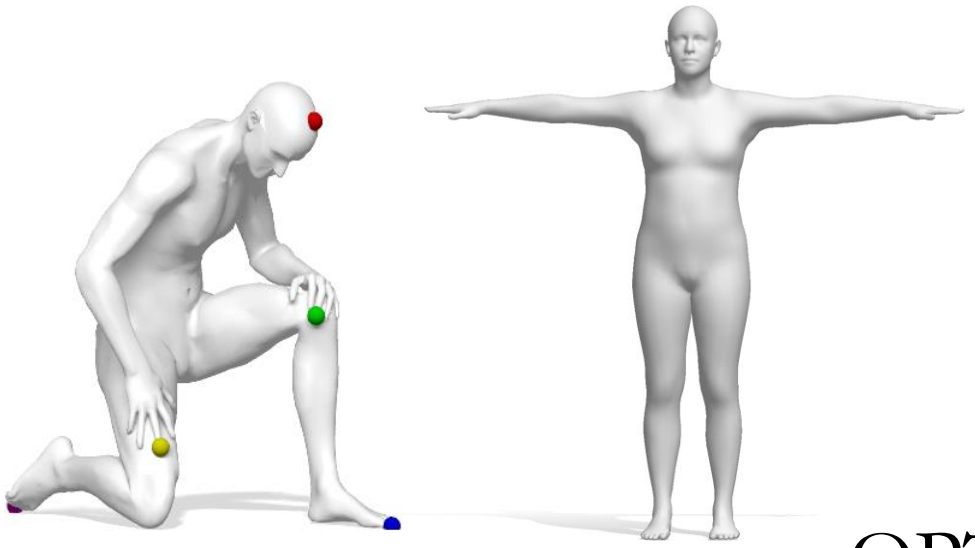
$$\mathcal{E} = w_S E_S + w_L E_L + w_V E_V + w_\beta E_\beta + w_\theta E_\theta$$



FMAP

OPT1

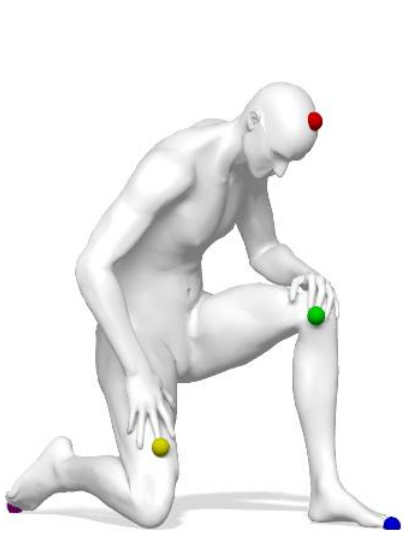
$$\mathcal{E} = w_S E_S + w_L E_L + w_V E_V + w_\beta E_\beta + w_\theta E_\theta$$

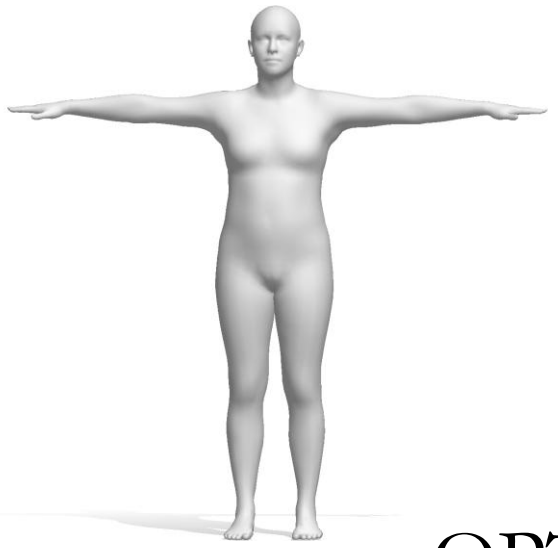


FMAP

OPT1

$$\mathcal{E} = w_S E_S + w_L E_L + w_V E_V + w_\beta E_\beta + w_\theta E_\theta$$



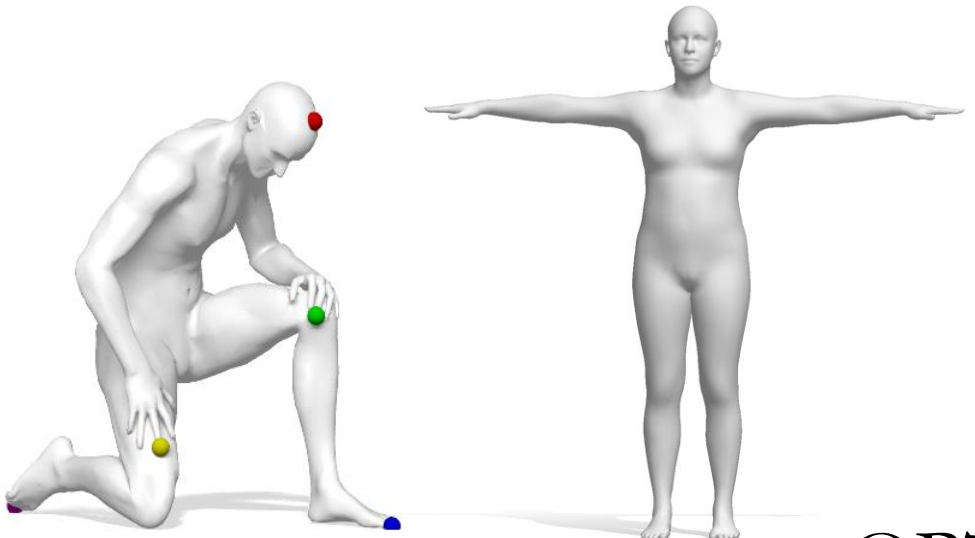


FMAP

OPT1

OPT2

$$\mathcal{E} = w_S E_S + w_L E_L + w_V E_V + w_\beta E_\beta + w_\theta E_\theta$$



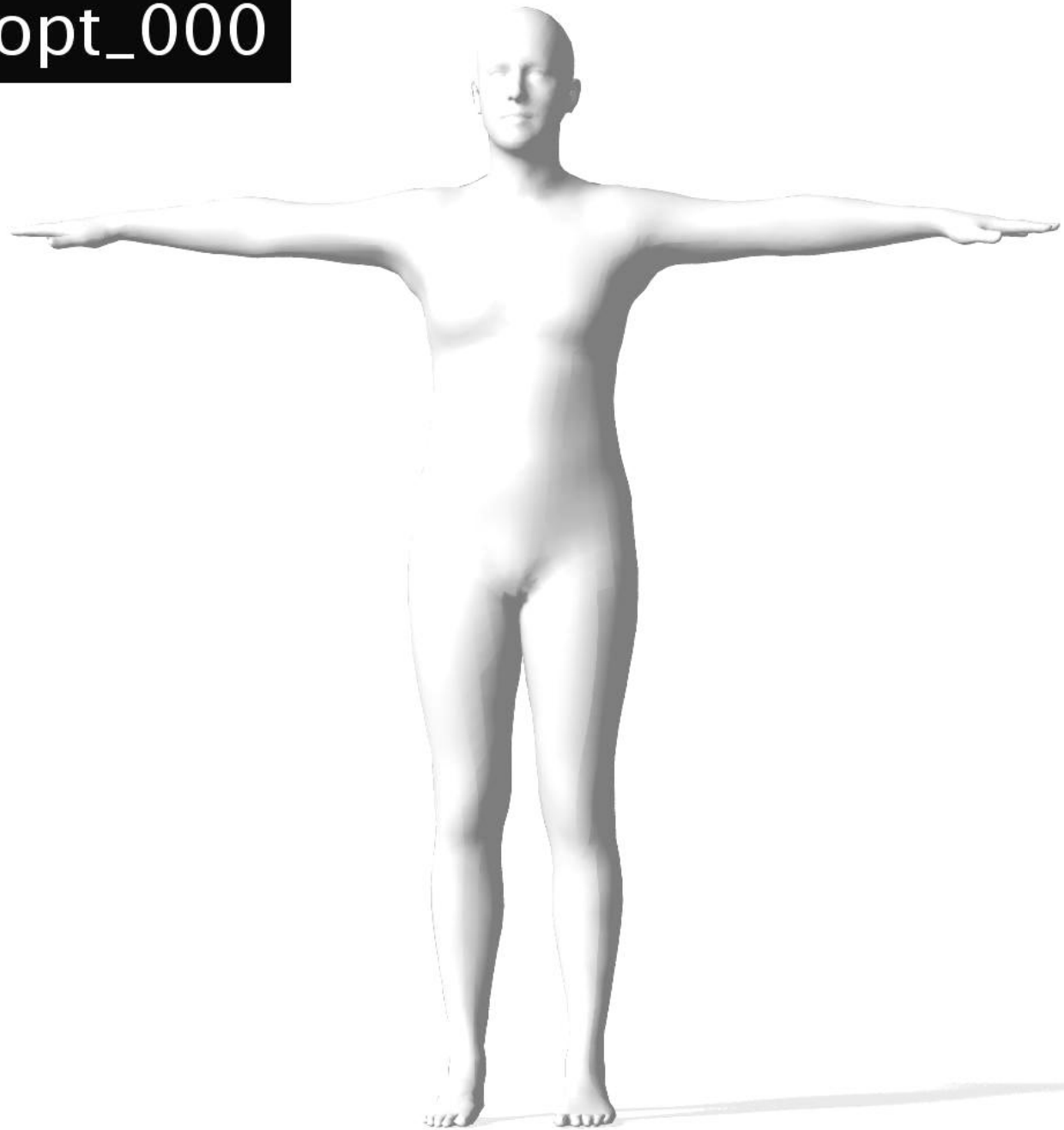
FMAP

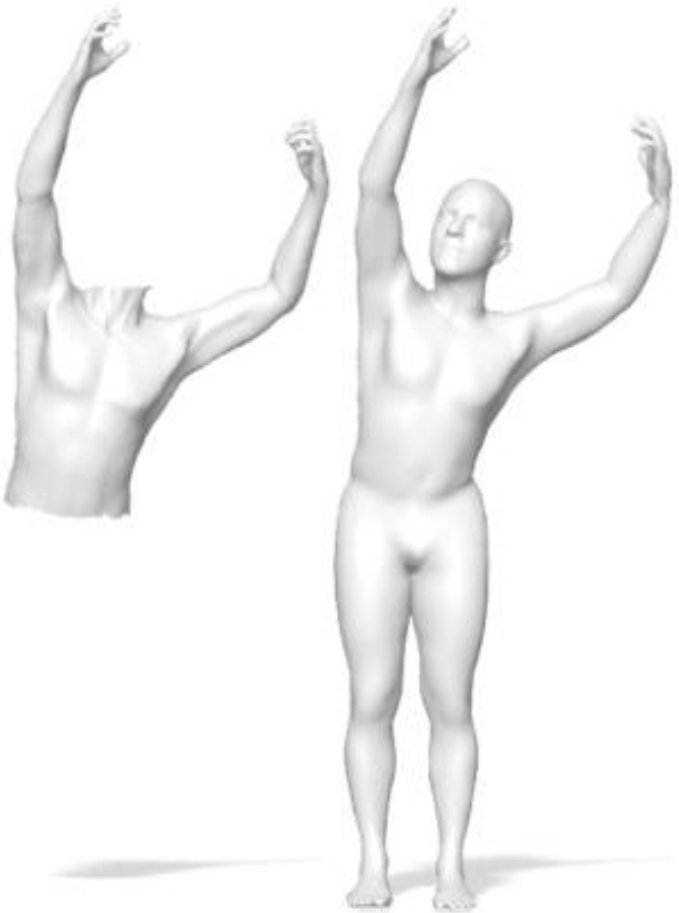
OPT1

OPT2

ARAP

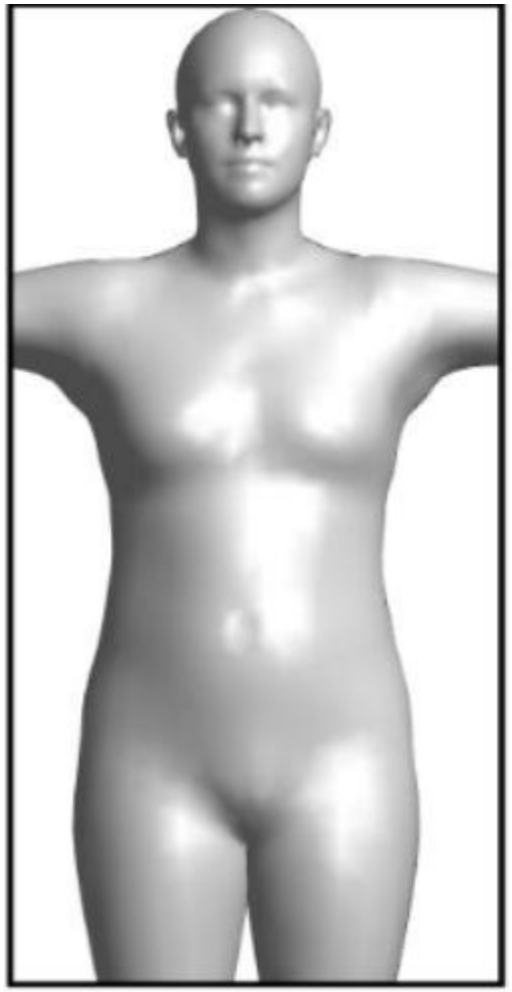
opt_000



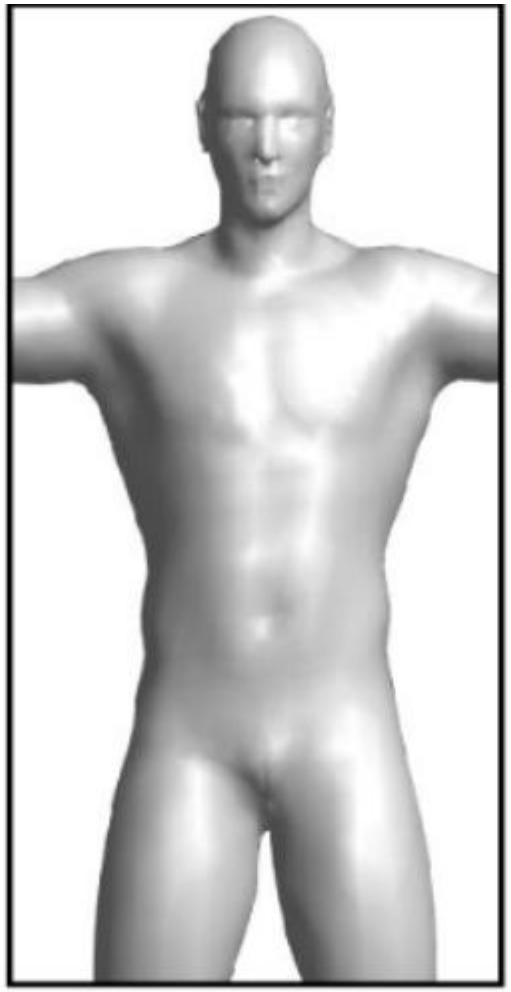


Problem: Is the registration bound by the template resolution?

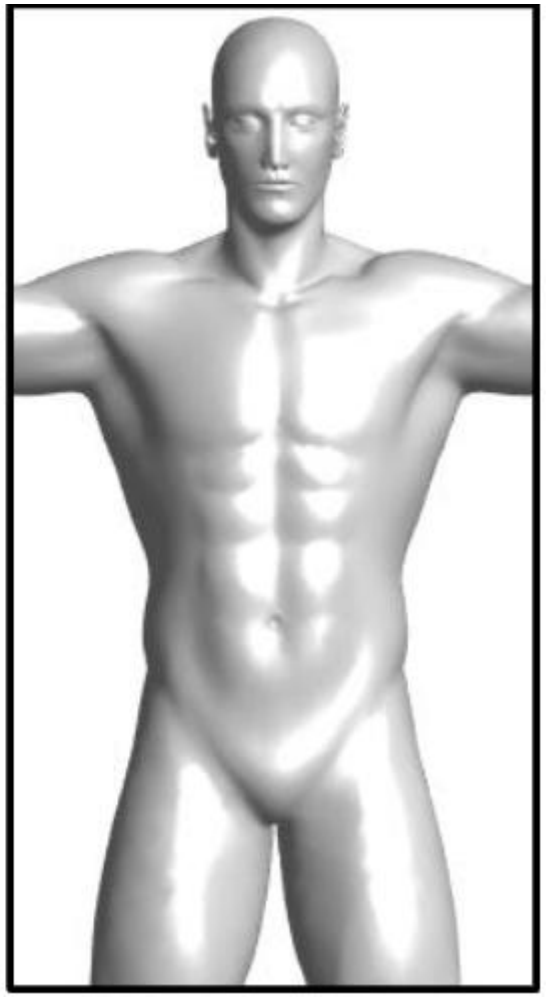
SMPL



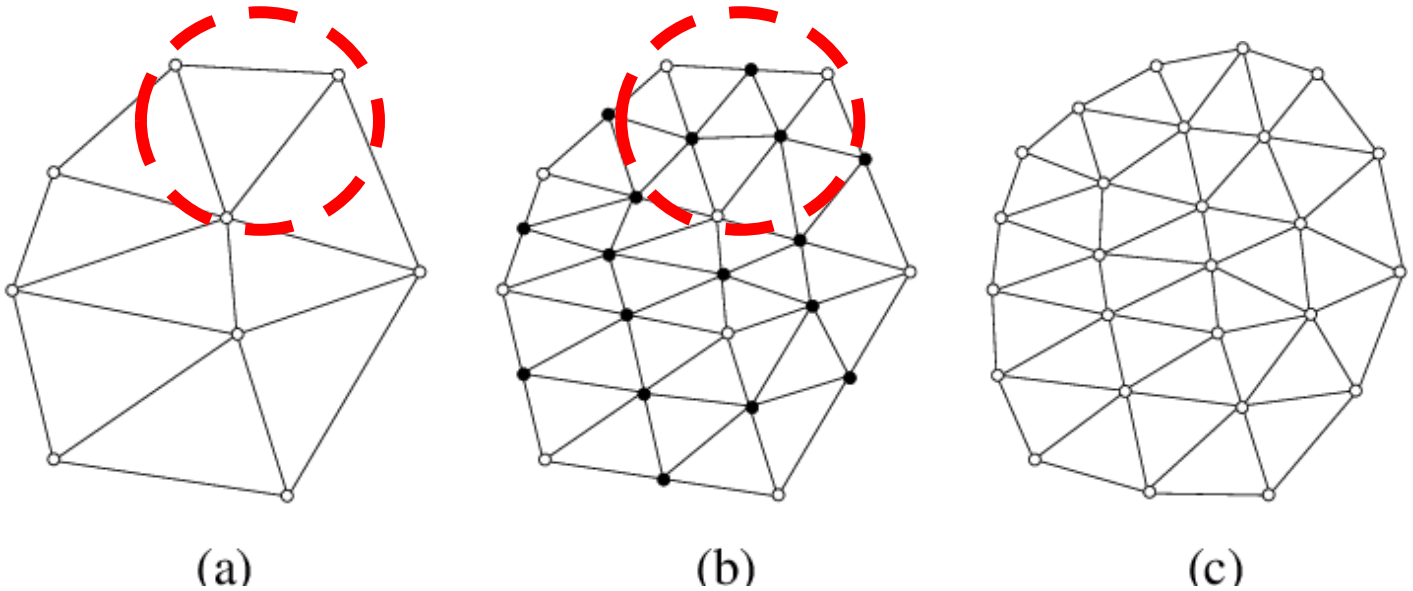
FARM

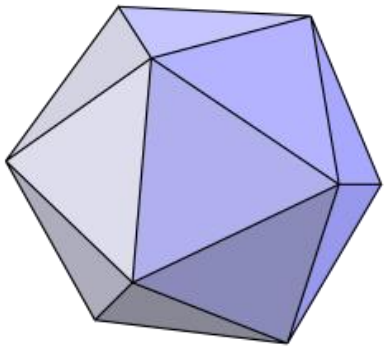


Target

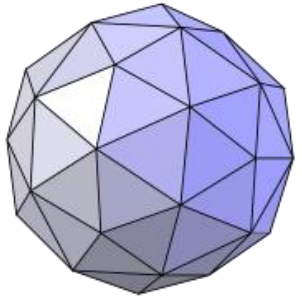


Loop Subdivision

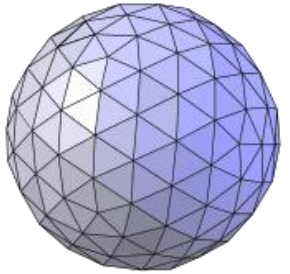




Source



1st
Subdivision



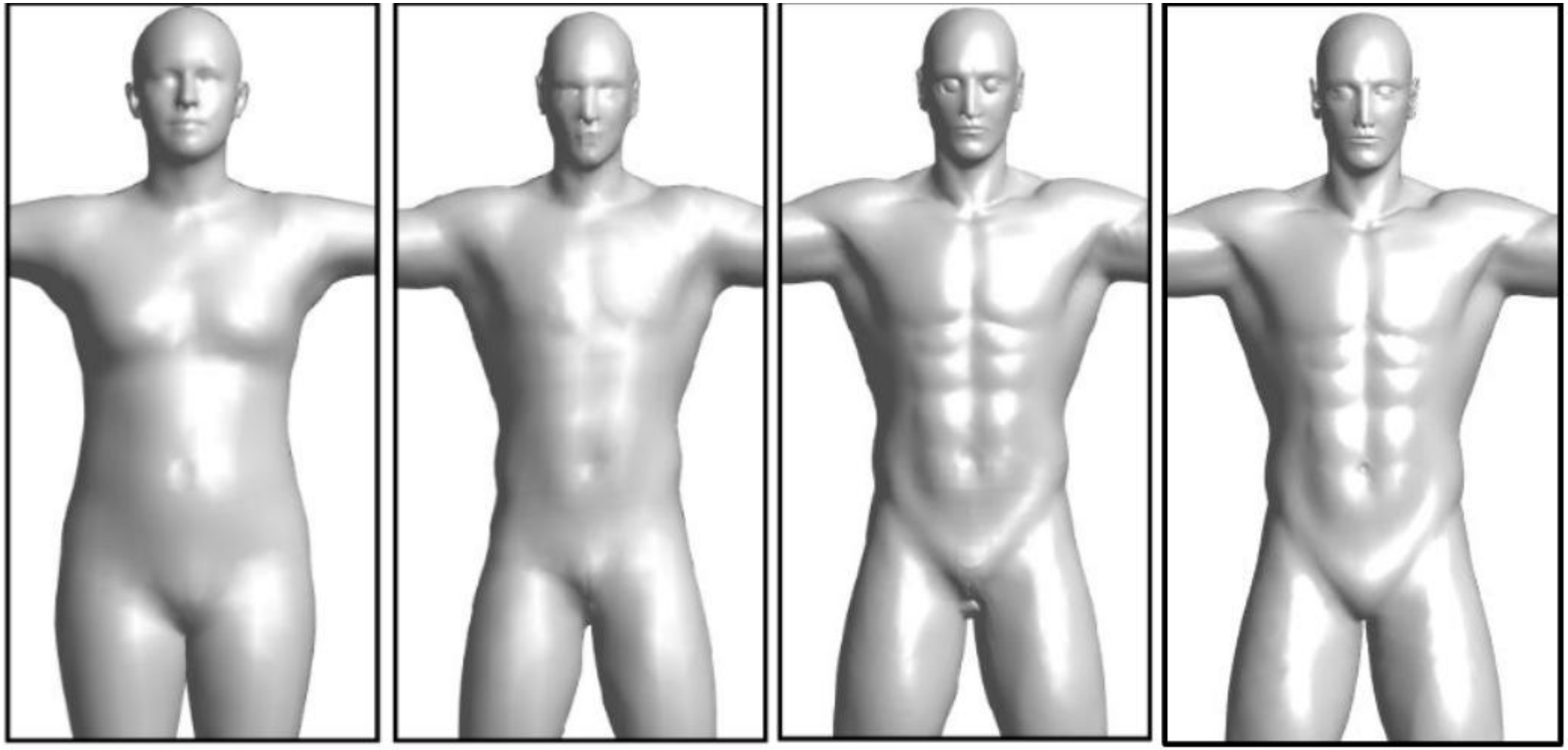
2nd
Subdivision

SMPL

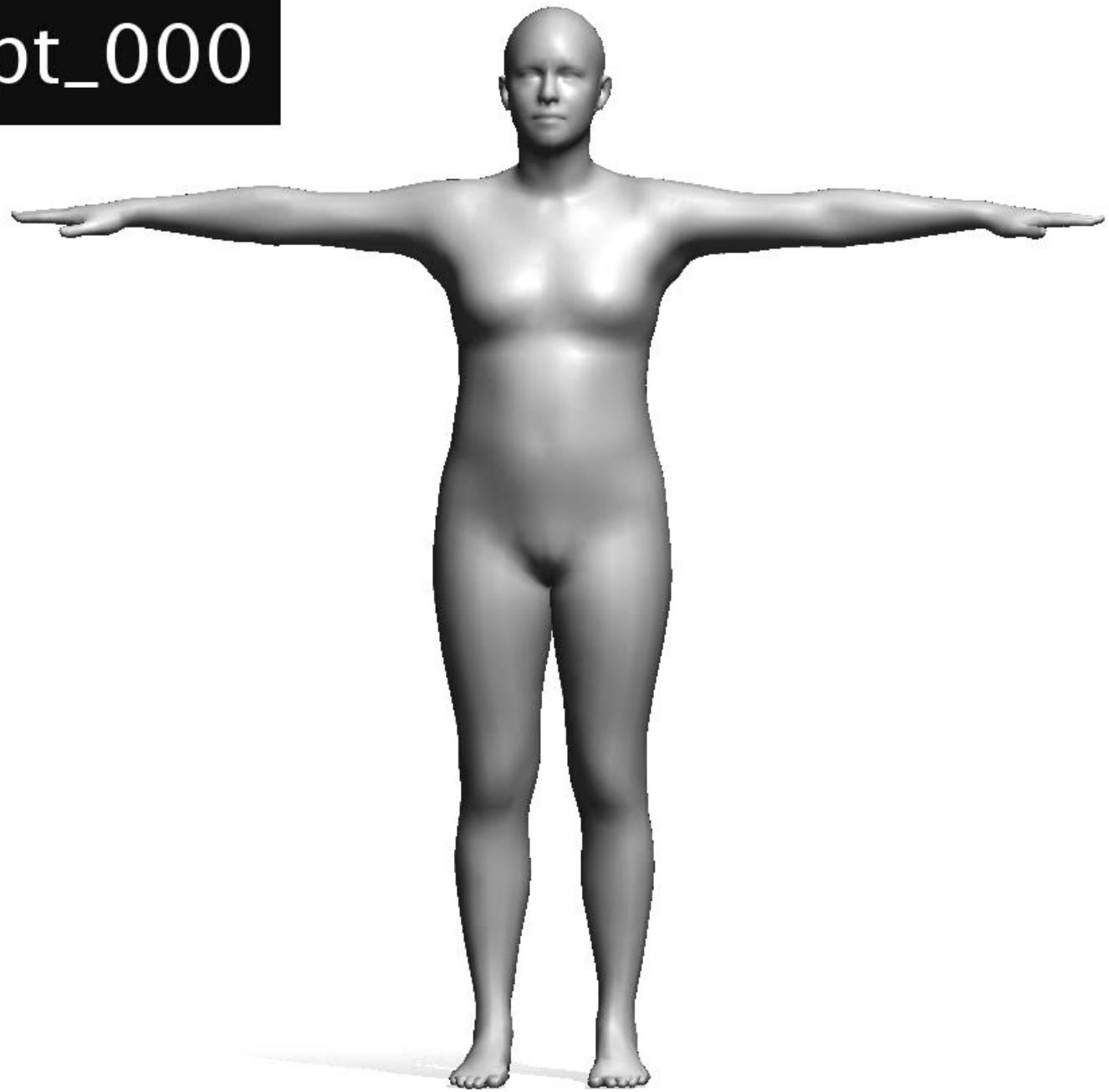
FARM

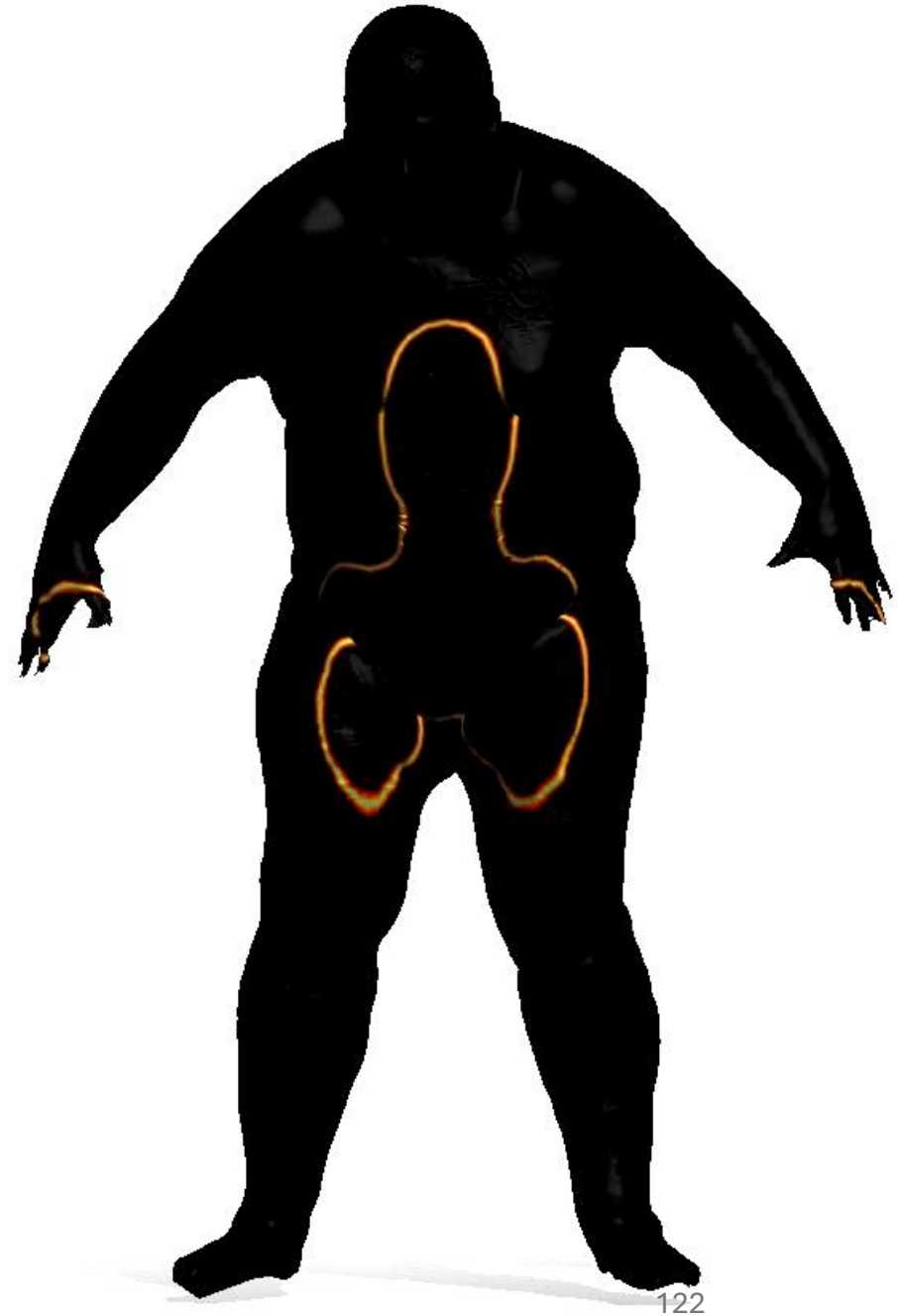
FARM+
Subdivision

Target



opt_000



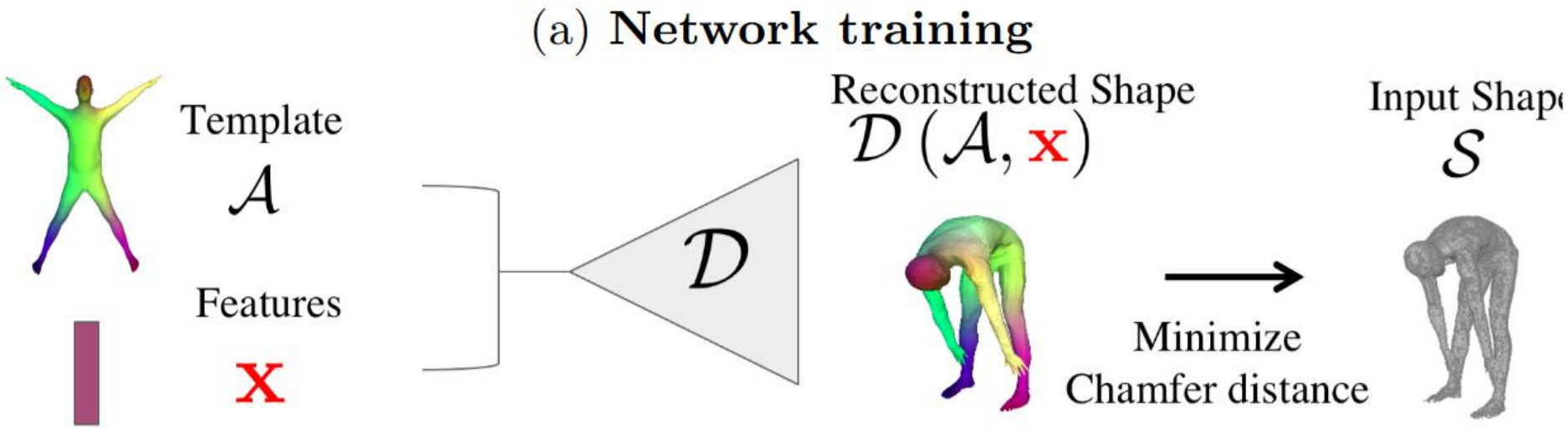


Learning based registration

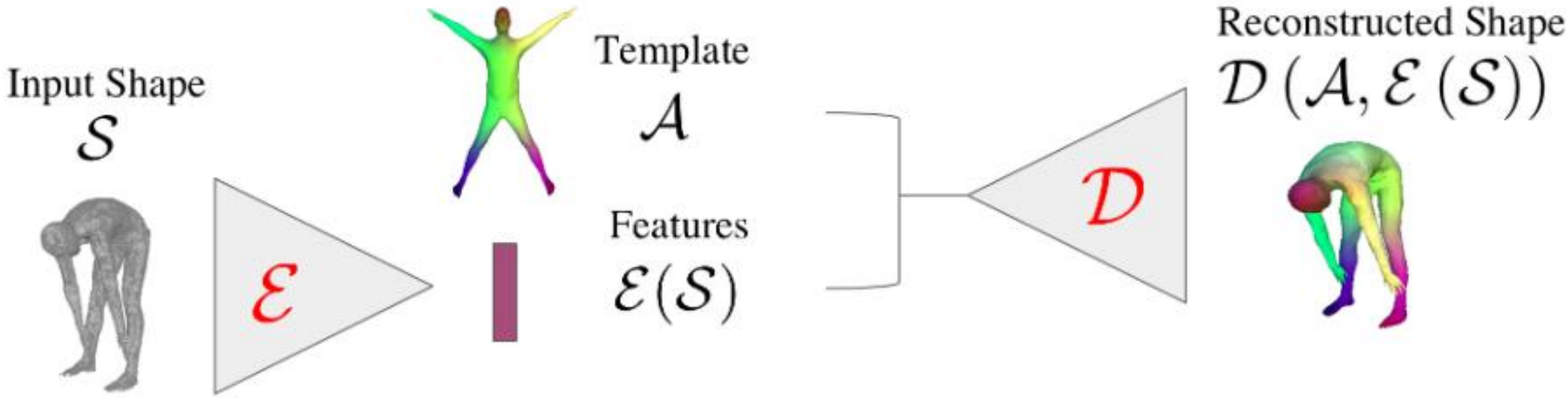
3D-CODED : 3D Correspondences by Deep Deformation

Thibault Groueix[1], Matthew Fisher[2], Vladimir G. Kim[2], Bryan C. Russell[2], and Mathieu Aubry[1]

[1] LIGM (UMR 8049), École des Ponts, UPE
[2] Adobe Research
http://imagine.enpc.fr/~groueixt/3D-CODED/



3DCoded



Supervised Training

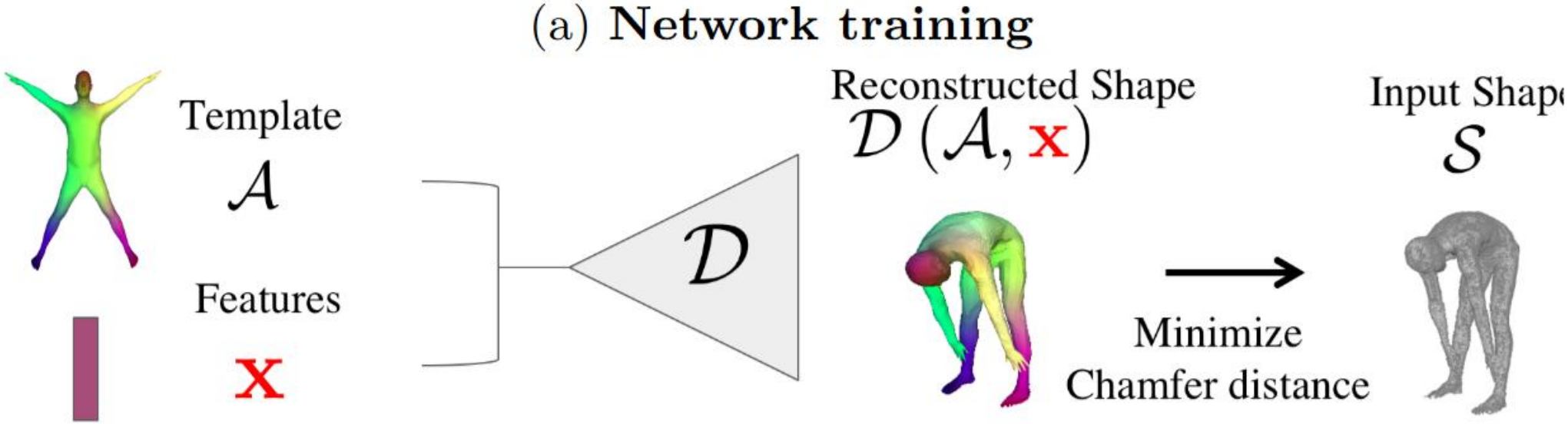
$$\mathcal{L}^{\text{sup}}(\theta, \phi) = \sum_{i=1}^{N} \sum_{j=1}^{P} |\mathcal{D}_\theta \left(\mathbf{p}_j; \mathcal{E}_\phi \left(\mathcal{S}^{(i)} \right) \right) - \mathbf{q}_j^{(i)}|^2$$

where the sums are over all P vertices of all N example shapes.

Unsupervised Training

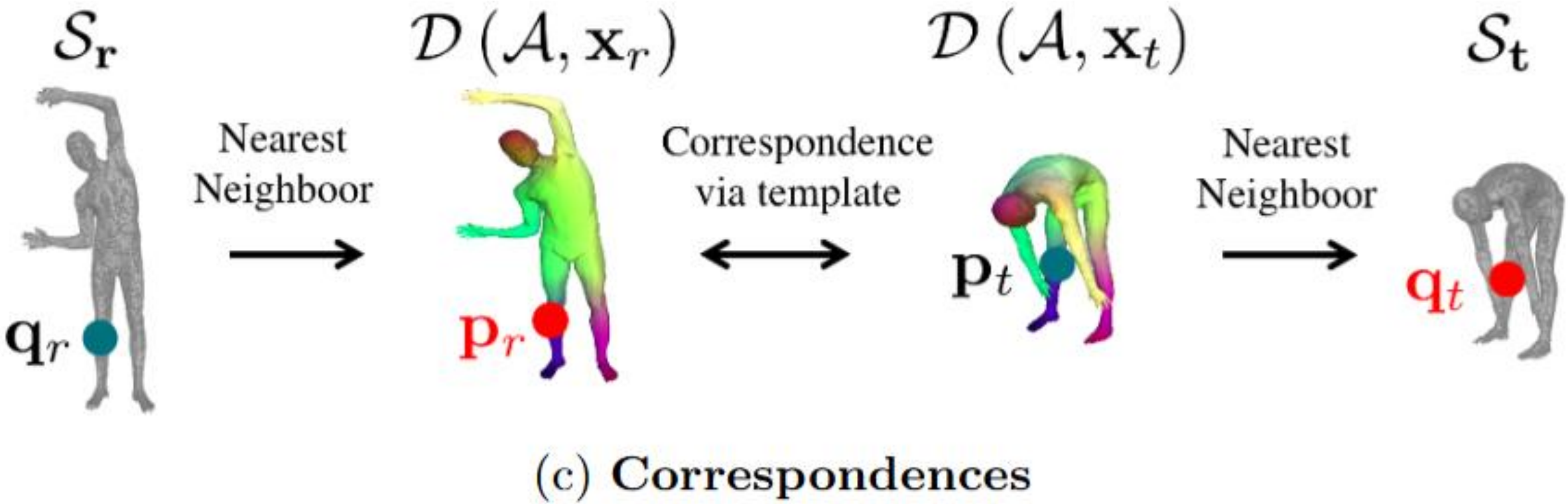
$$\mathcal{L}^{\text{unsup}} = \mathcal{L}^{\text{CD}} + \lambda_{Lap} \mathcal{L}^{\text{Lap}} + \lambda_{edges} \mathcal{L}^{\text{edges}}$$

Test time optimization



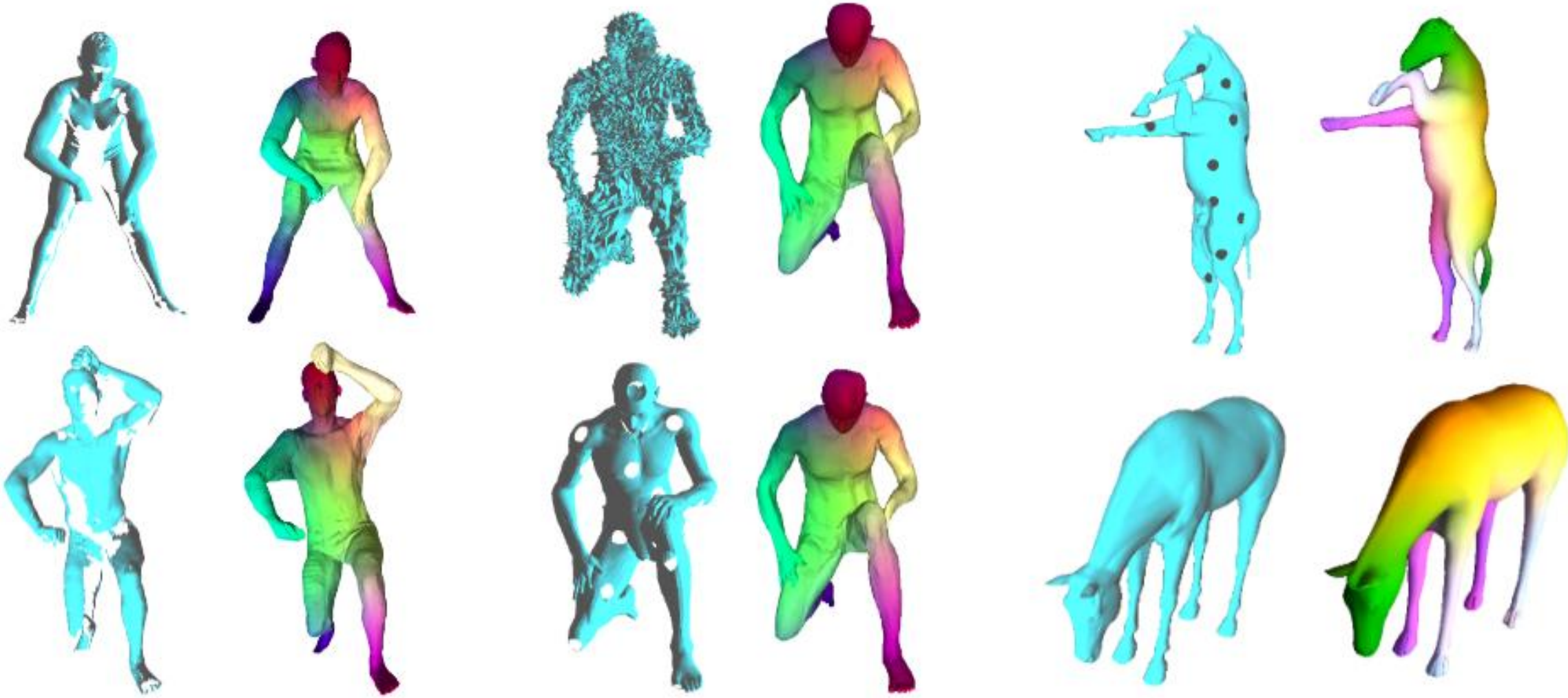
$$\mathcal{L}^{\mathrm{CD}}(\mathbf{x}; \mathcal{S}) = \sum_{\mathbf{p} \in \mathcal{A}} \min_{\mathbf{q} \in \mathcal{S}} |\mathcal{D}_{\theta}(\mathbf{p}; \mathbf{x}) - \mathbf{q}|^2 + \sum_{\mathbf{q} \in \mathcal{S}} \min_{\mathbf{p} \in \mathcal{A}} |\mathcal{D}_{\theta}(\mathbf{p}; \mathbf{x}) - \mathbf{q}|^2$$

Correspondence retrieval



(c) **Correspondences**

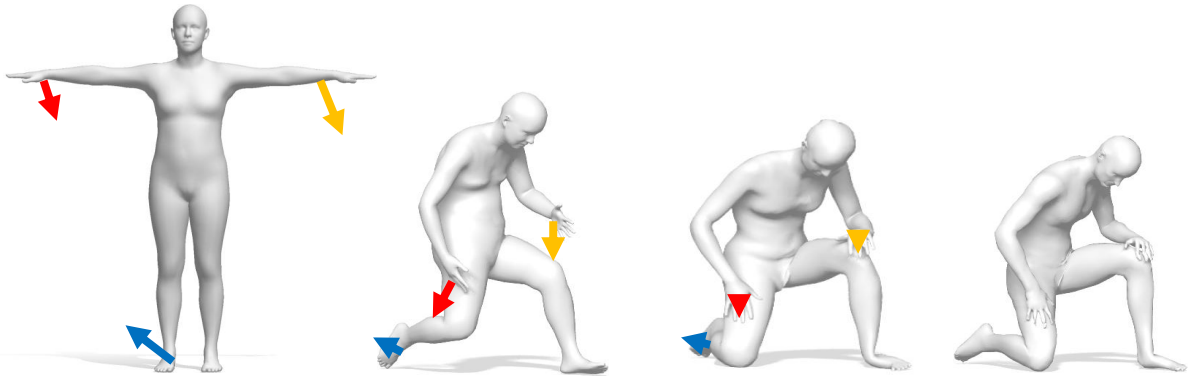
Results



(a) SCAPE [4] (b) TOSCA [12] (c) TOSCA animals [12]

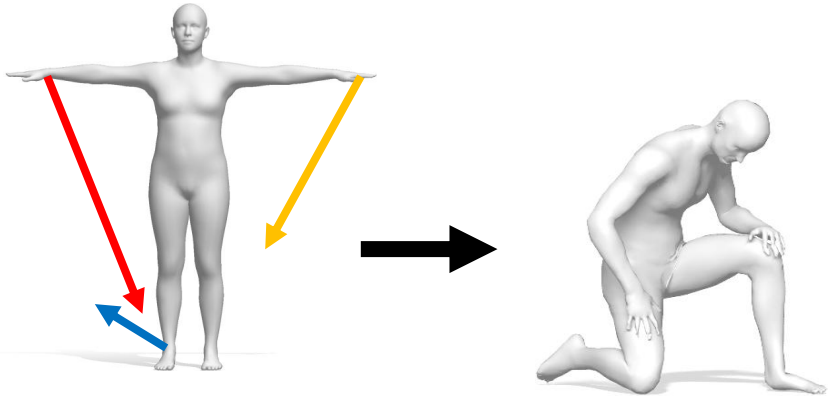
Precision vs speed

Optimization
(FARM)



- Iterative process
- Little steps till convergence
- Regularized
- Slow
- Hand-crafted features

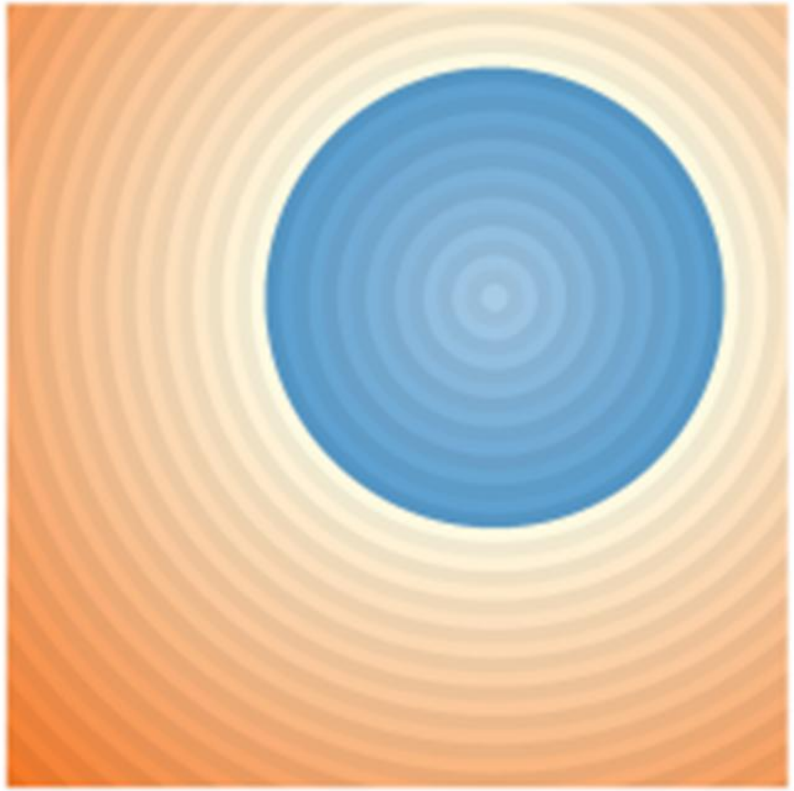
Learning regression
(3DCoded)



- Single forward-pass
- Difficult prediction
- Black-box (No control)
- Fast
- Learning from data

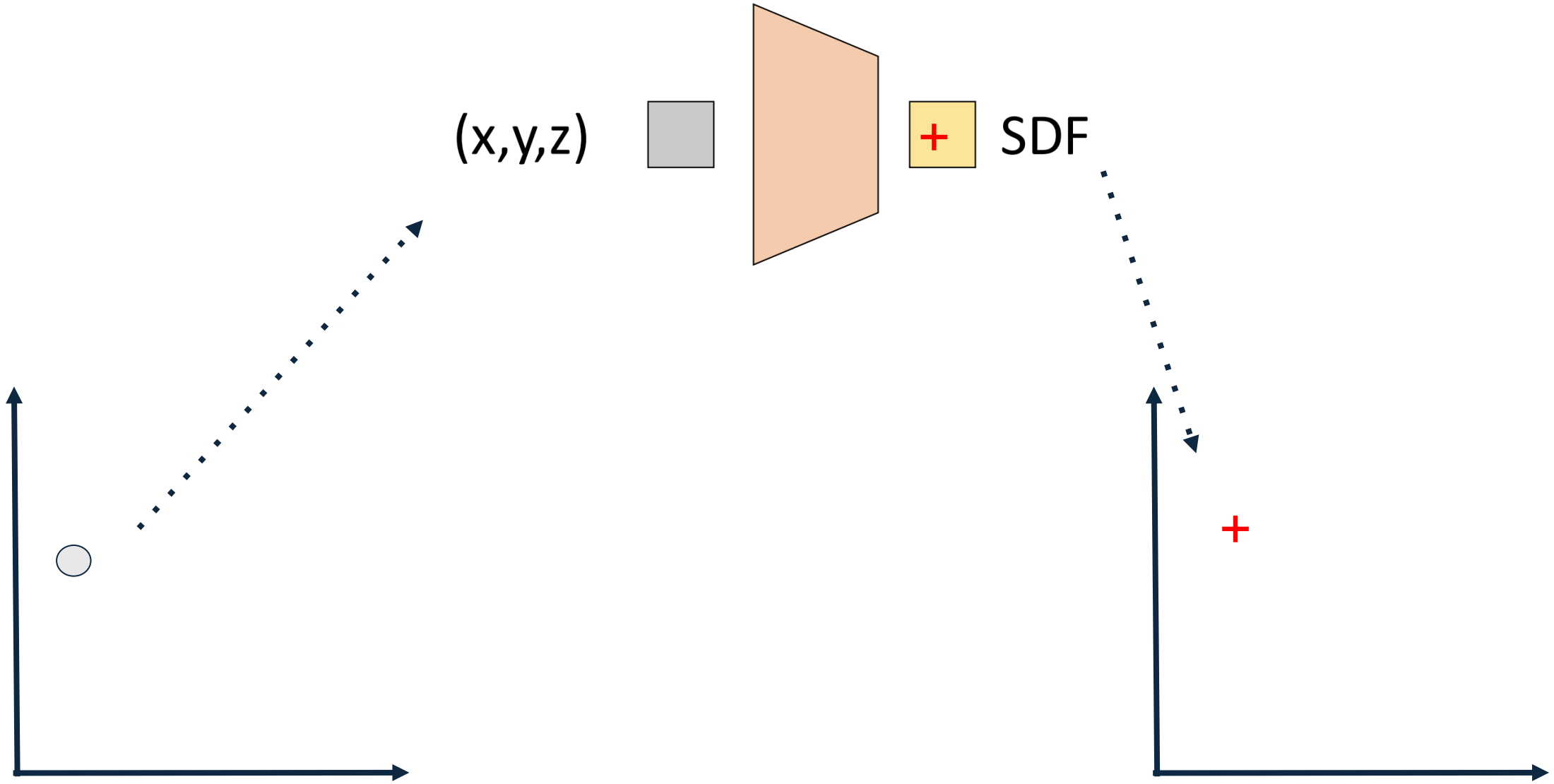
Can we obtain a tradeoff?

Neural field - Signed distance field/function

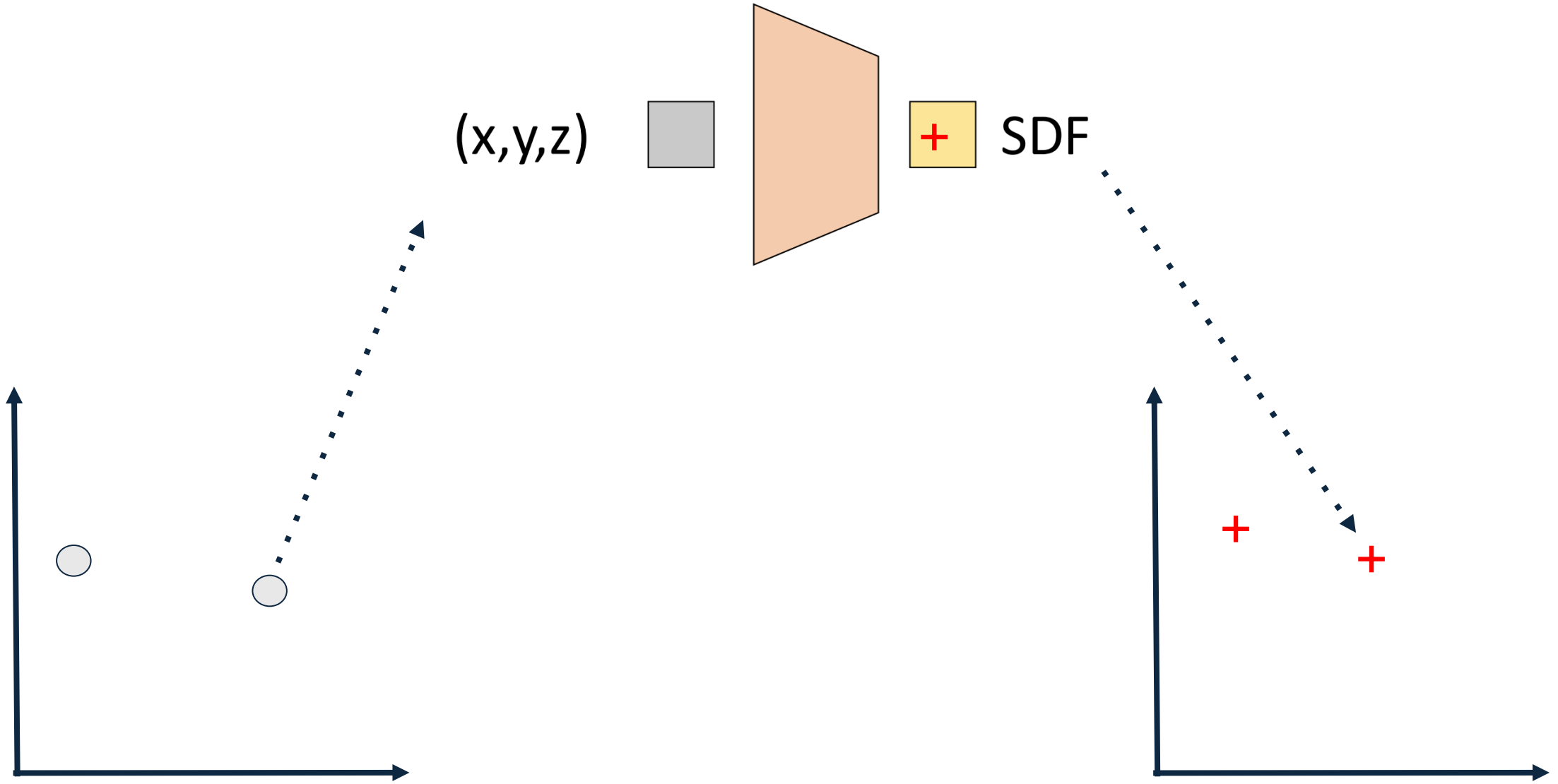


$$f : \mathbb{R}^3 \rightarrow \mathbb{R}, f(x) = \begin{cases} d, \text{ if } x \text{ is outside} \\ -d, \text{otherwise} \end{cases}$$

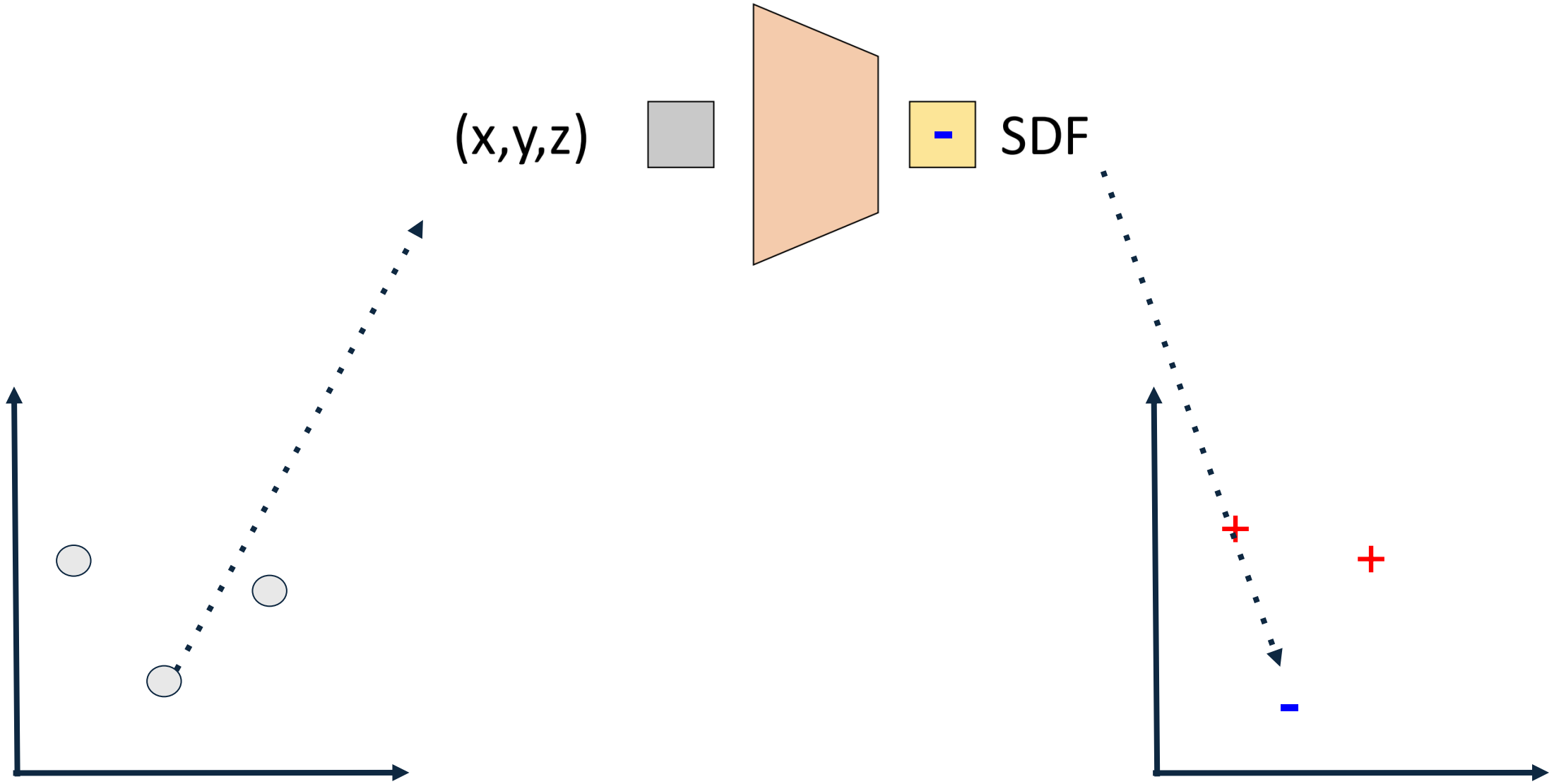
Deep SDF



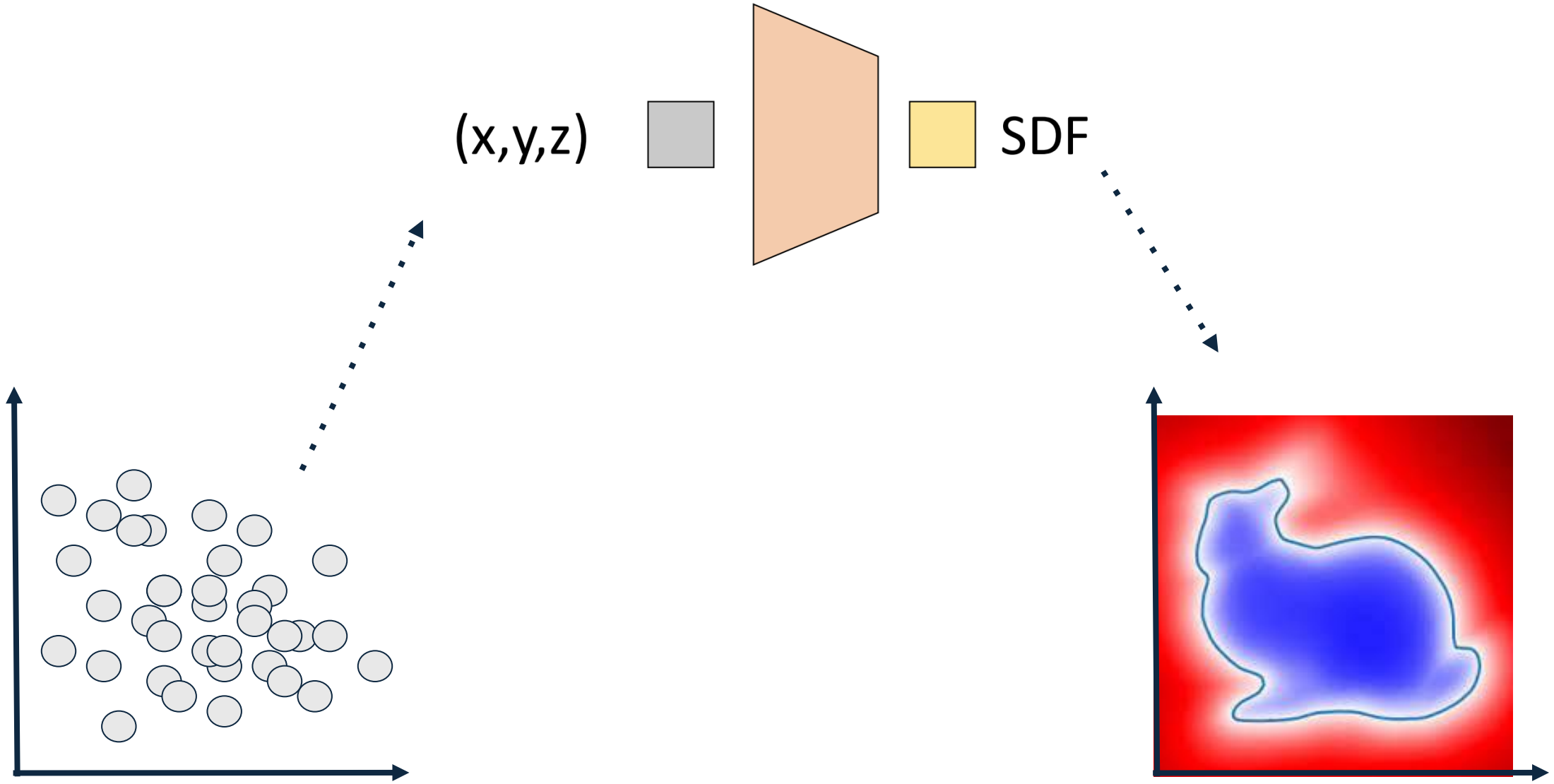
Deep SDF



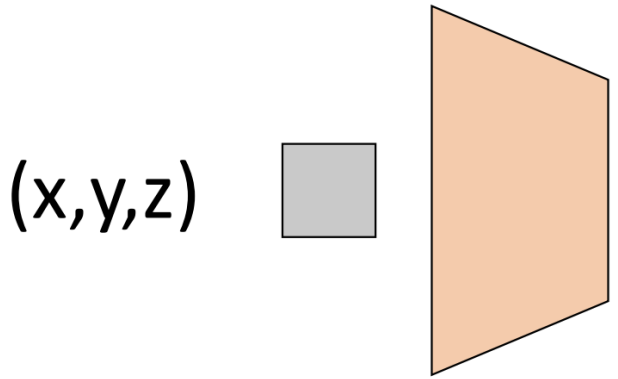
Deep SDF

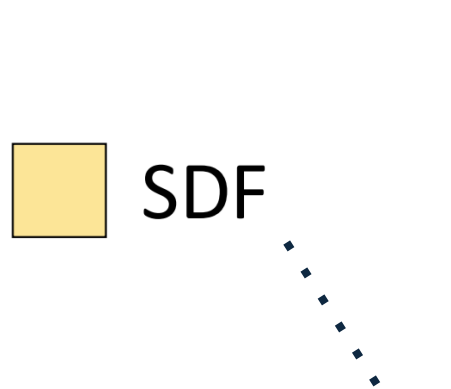


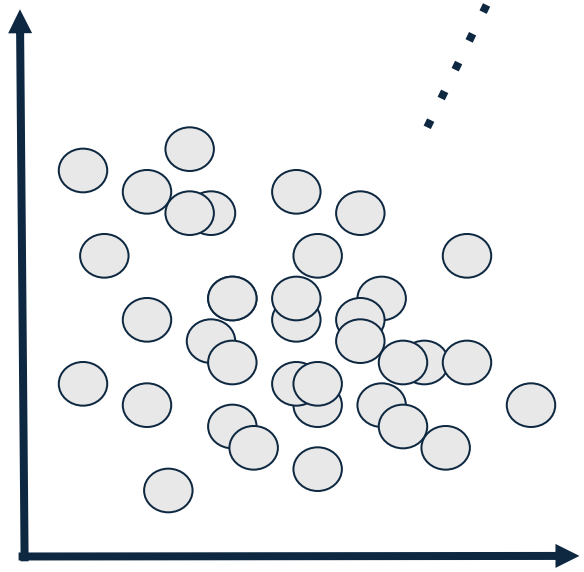
Deep SDF



Deep SDF

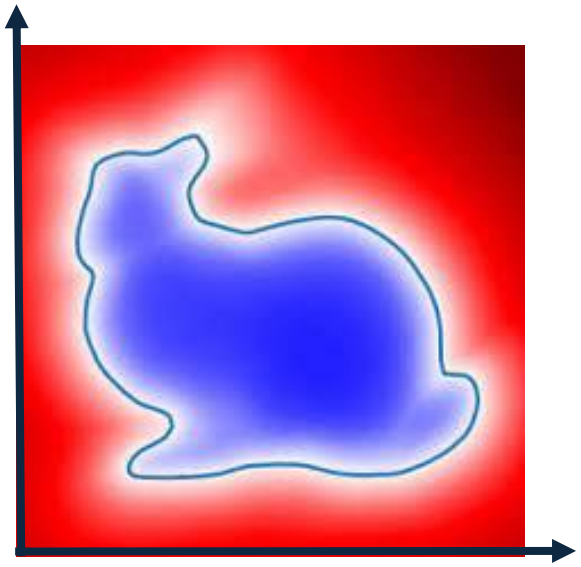




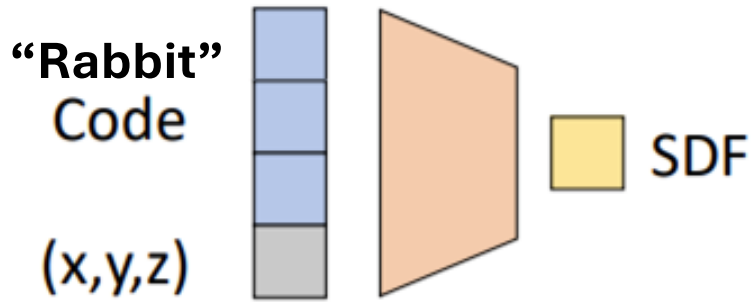


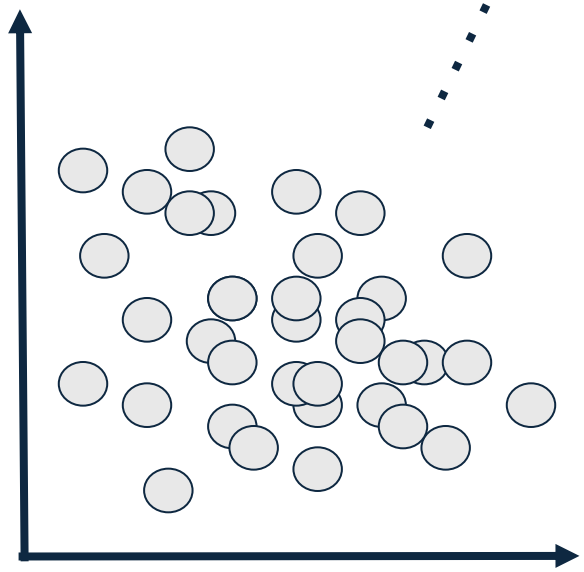
Problem

One network for a single shape



Deep SDF



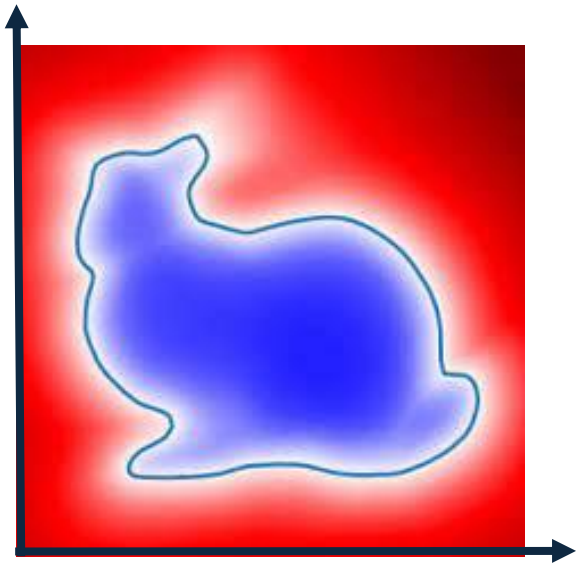


Problem

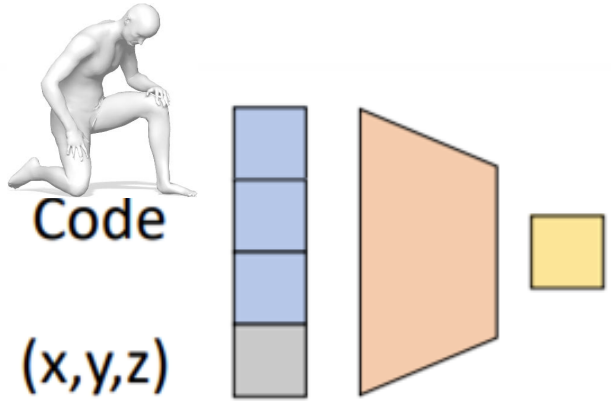
One network for a single shape

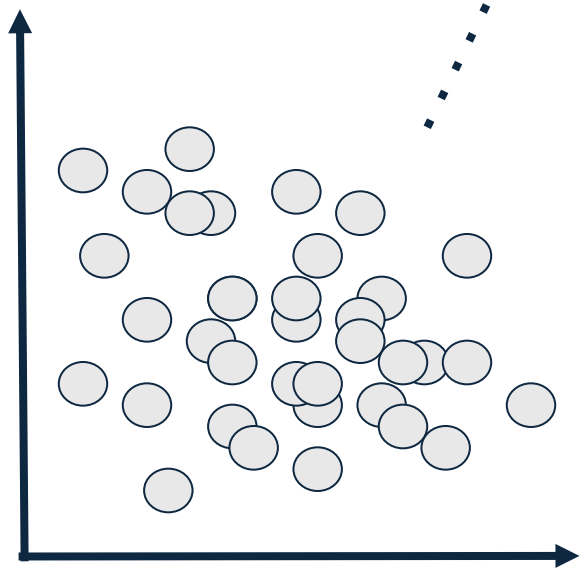
Solution

Condition the input



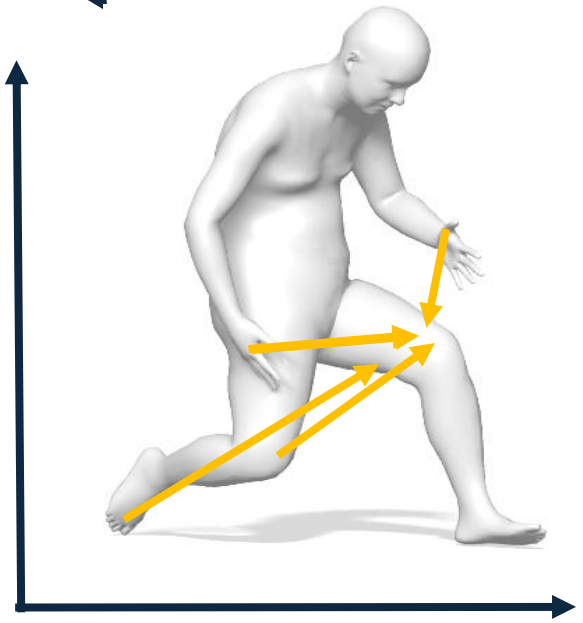
Neural Deformation Field



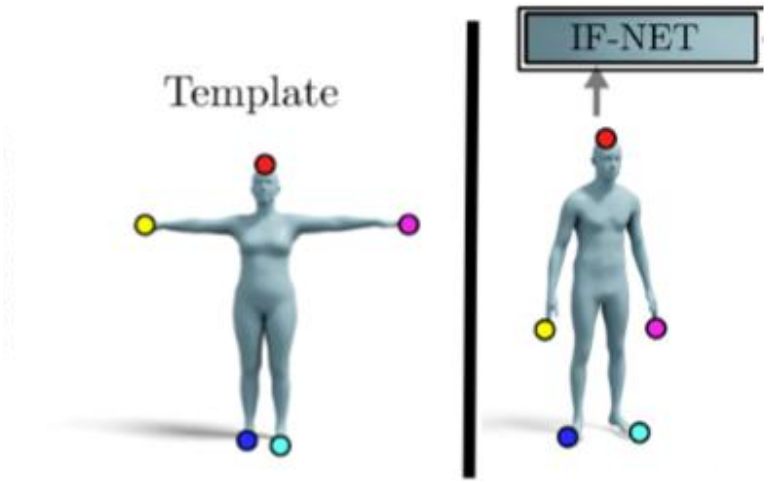


Idea

Train a network that
for any point in space
predicts small offsets toward
the targets

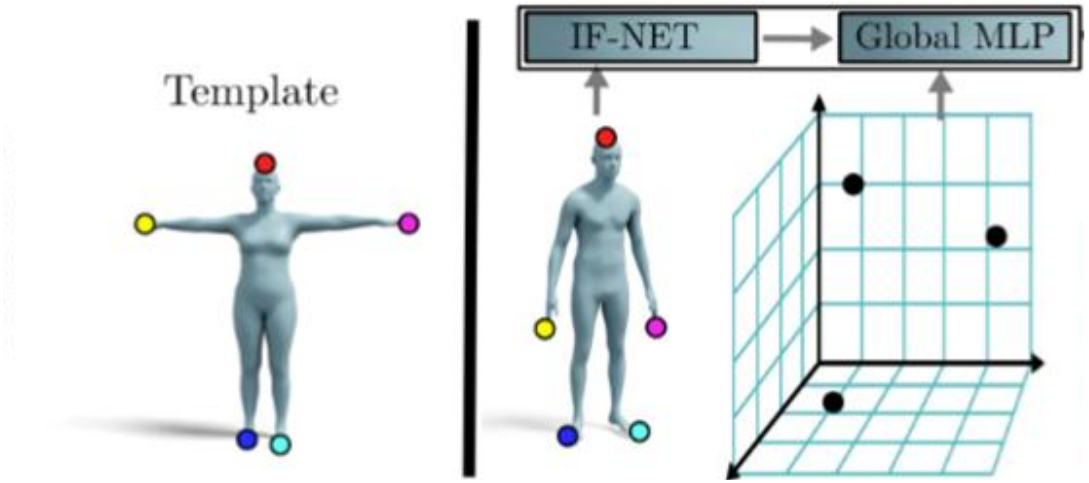


Neural Fields Deformation



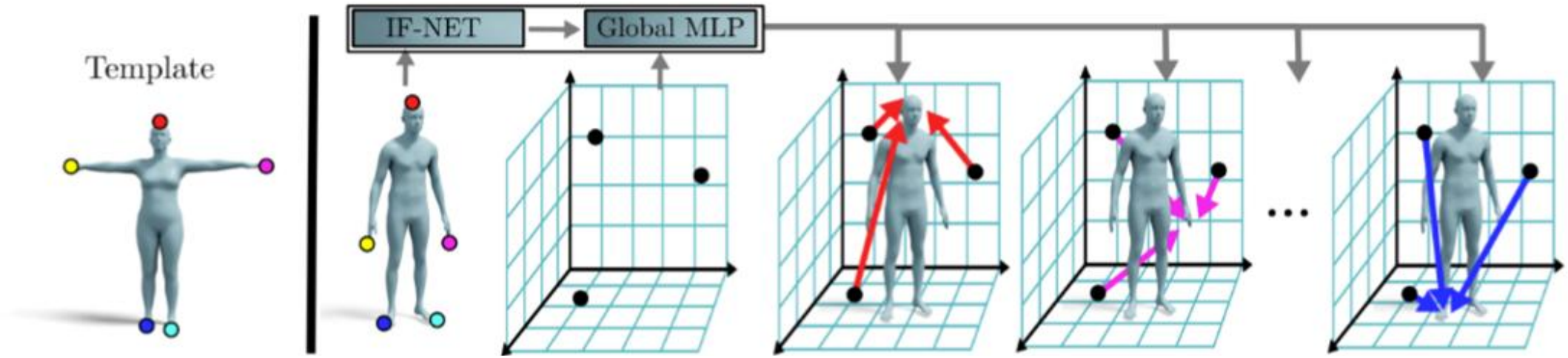
Learned Vertex Descent: A New Direction for 3D Human Model Fitting, Corona et al., 2022

Neural Fields Deformation



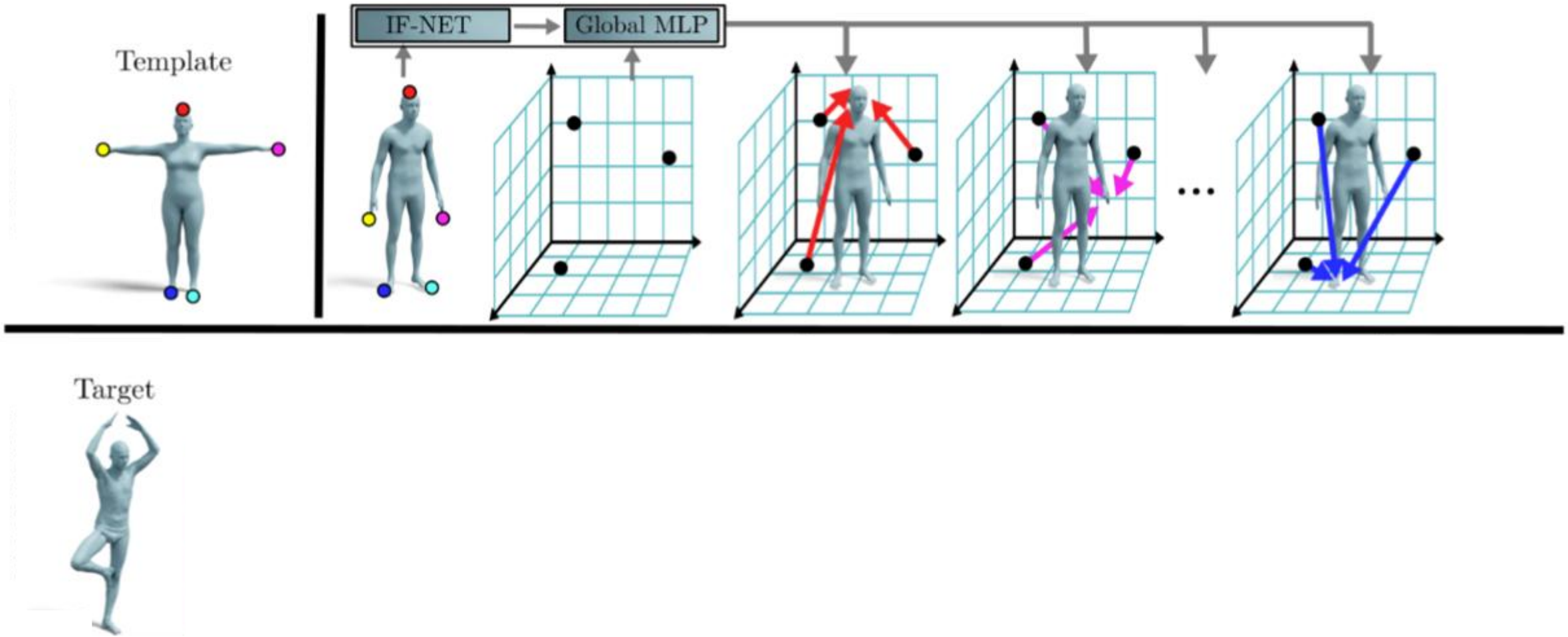
Learned Vertex Descent: A New Direction for 3D Human Model Fitting, Corona et al., 2022

Neural Fields Deformation



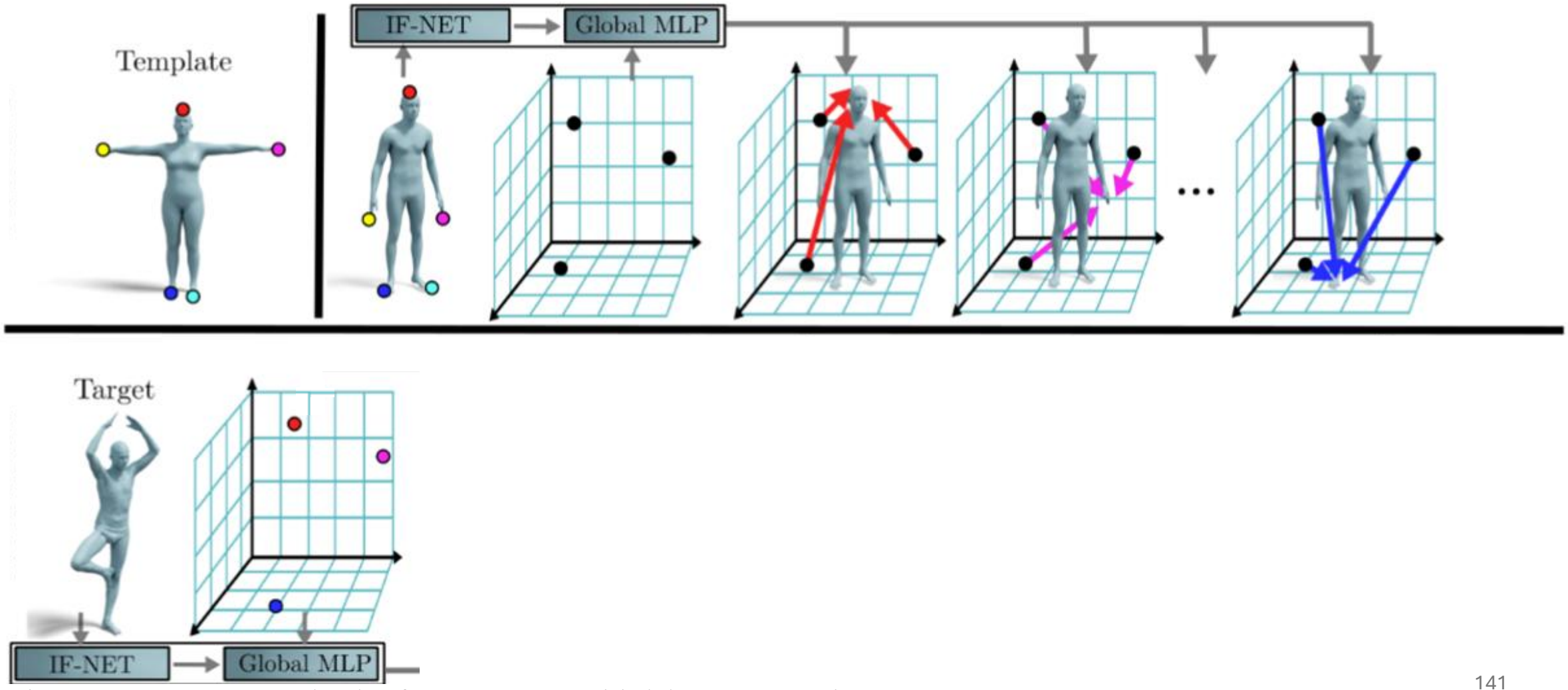
Learned Vertex Descent: A New Direction for 3D Human Model Fitting, Corona et al., 2022

Neural Fields Deformation



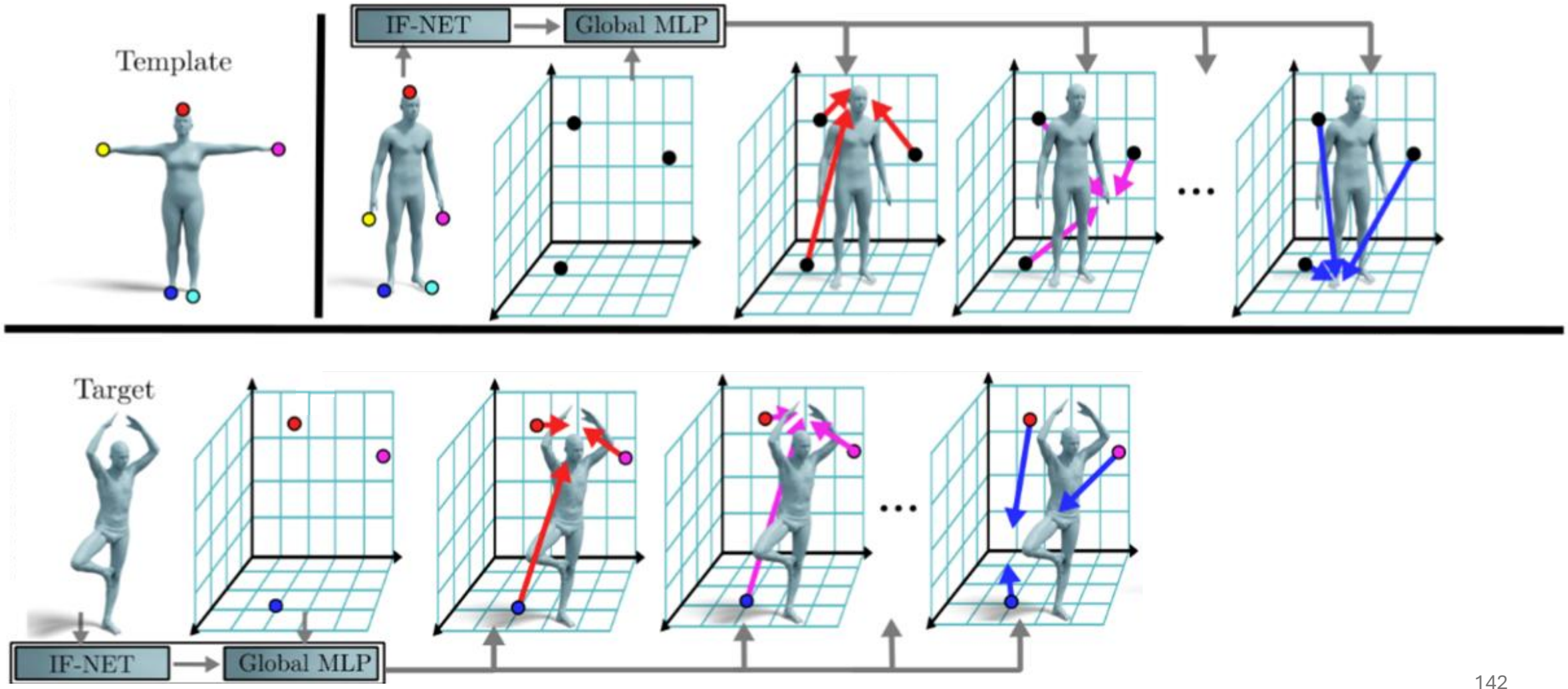
Learned Vertex Descent: A New Direction for 3D Human Model Fitting, Corona et al., 2022

Neural Fields Deformation



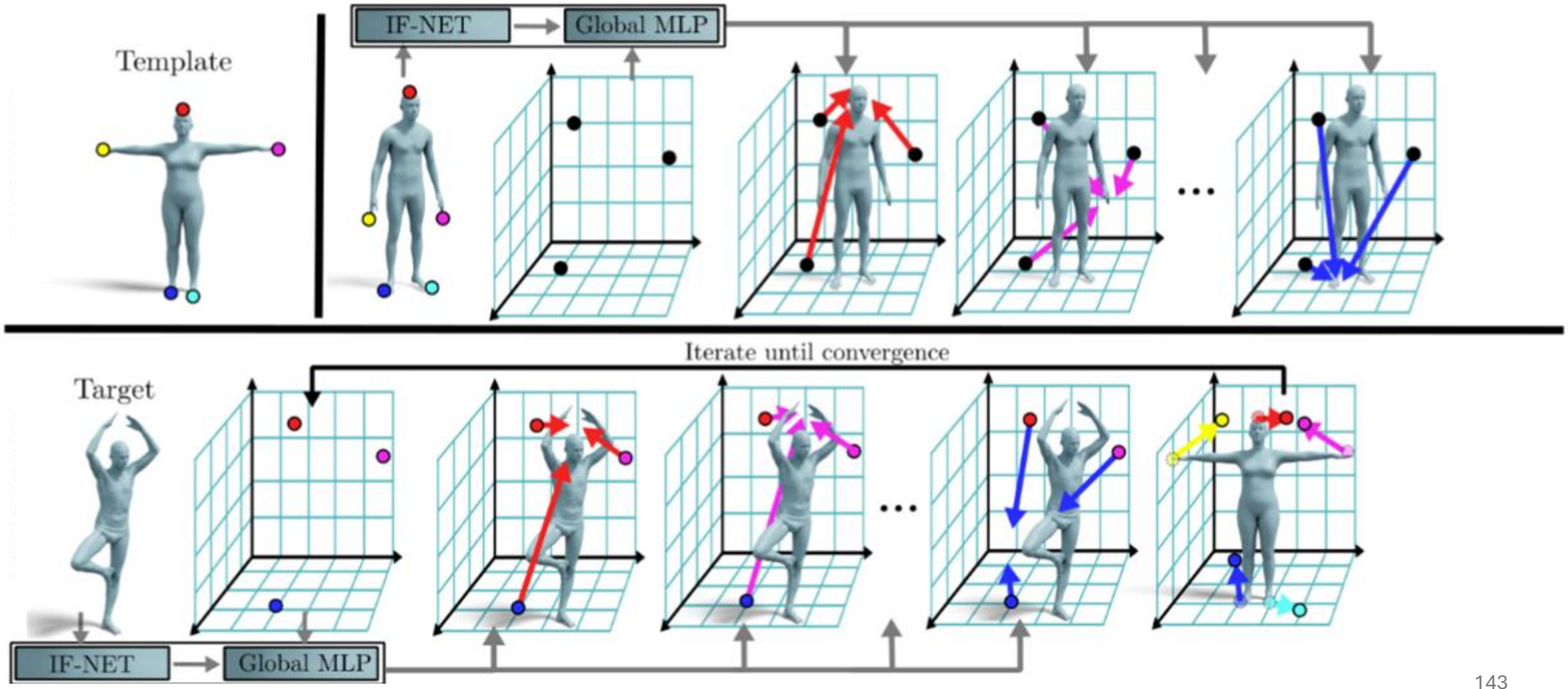
Learned Vertex Descent: A New Direction for 3D Human Model Fitting, Corona et al., 2022

Neural Fields Deformation



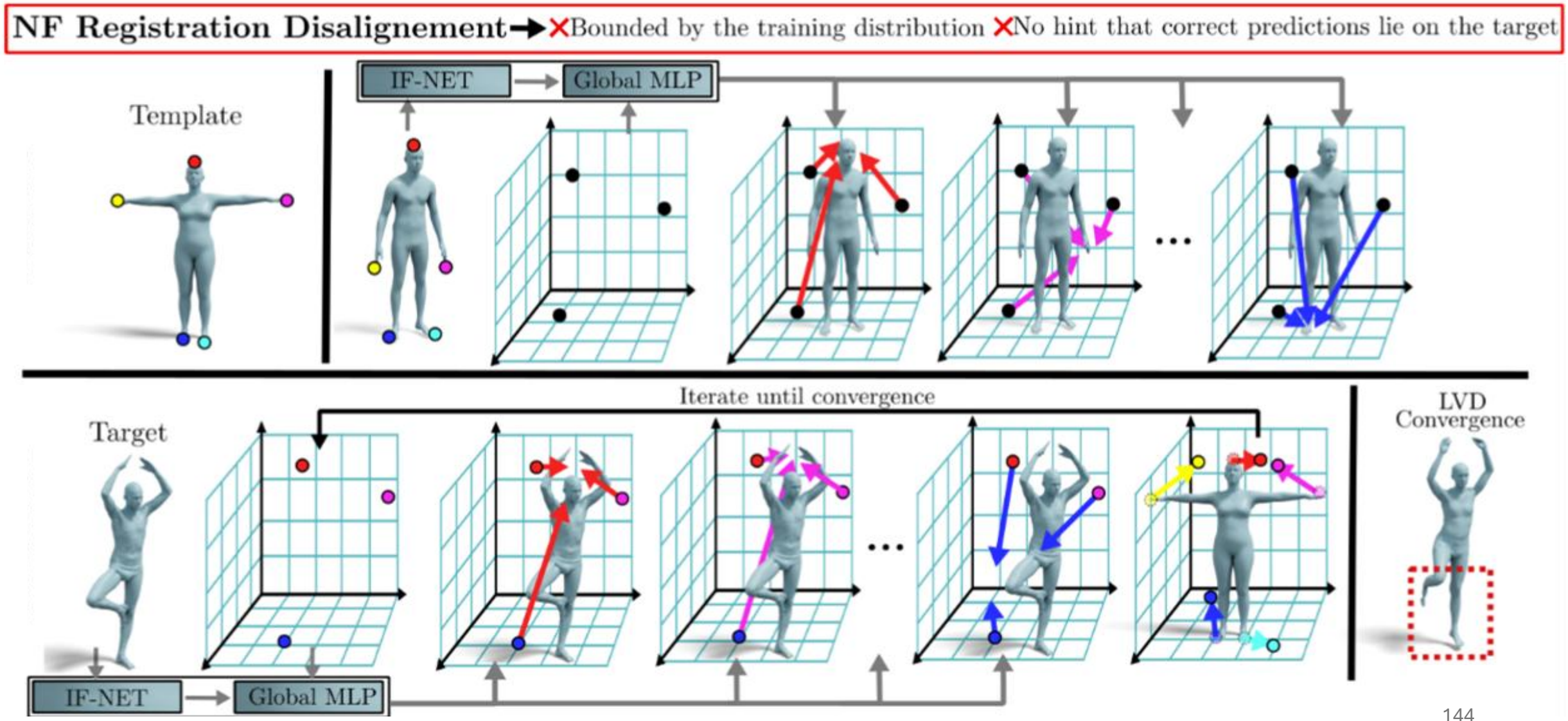
Learned Vertex Descent: A New Direction for 3D Human Model Fitting, Corona et al., 2022

Neural Fields Deformation



Learned Vertex Descent: A New Direction for 3D Human Model Fitting, Corona et al., 2022

Neural Fields Deformation



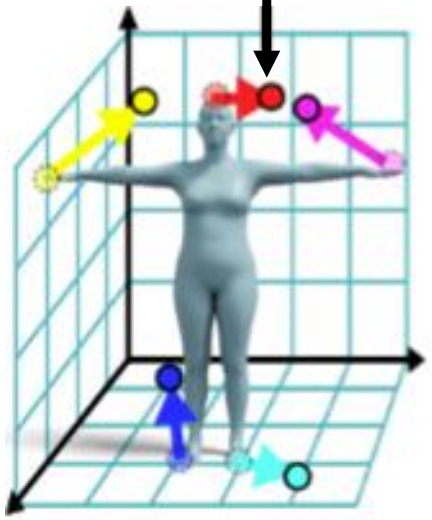
Learned Vertex Descent: A New Direction for 3D Human Model Fitting, Corona et al., 2022

Solution

Neural ICP (NICP)

Self-Supervised tuning on Target Geometry (inference)

LVD



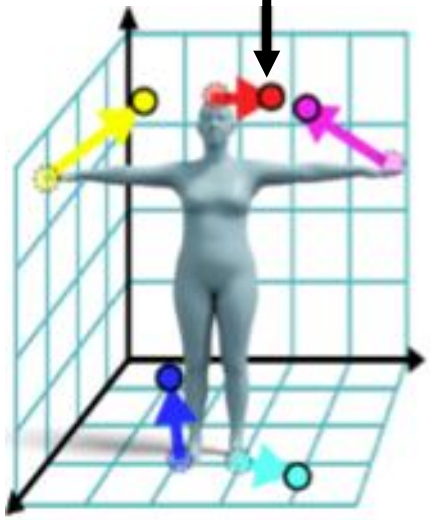
Convergence

Solution

Neural ICP (NICP)

Self-Supervised tuning on Target Geometry (inference)

LVD



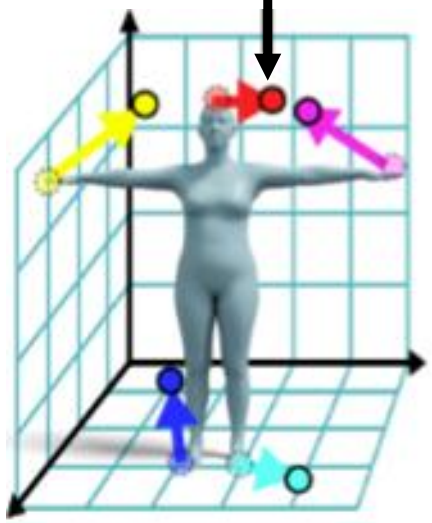
Convergence

Solution

Neural ICP (NICP)

Self-Supervised tuning on Target Geometry (inference)

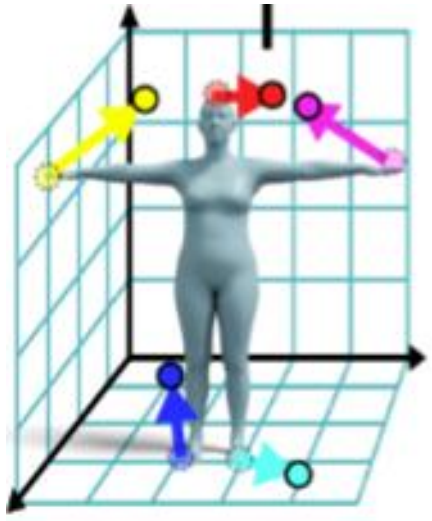
LVD



Convergence

NICP

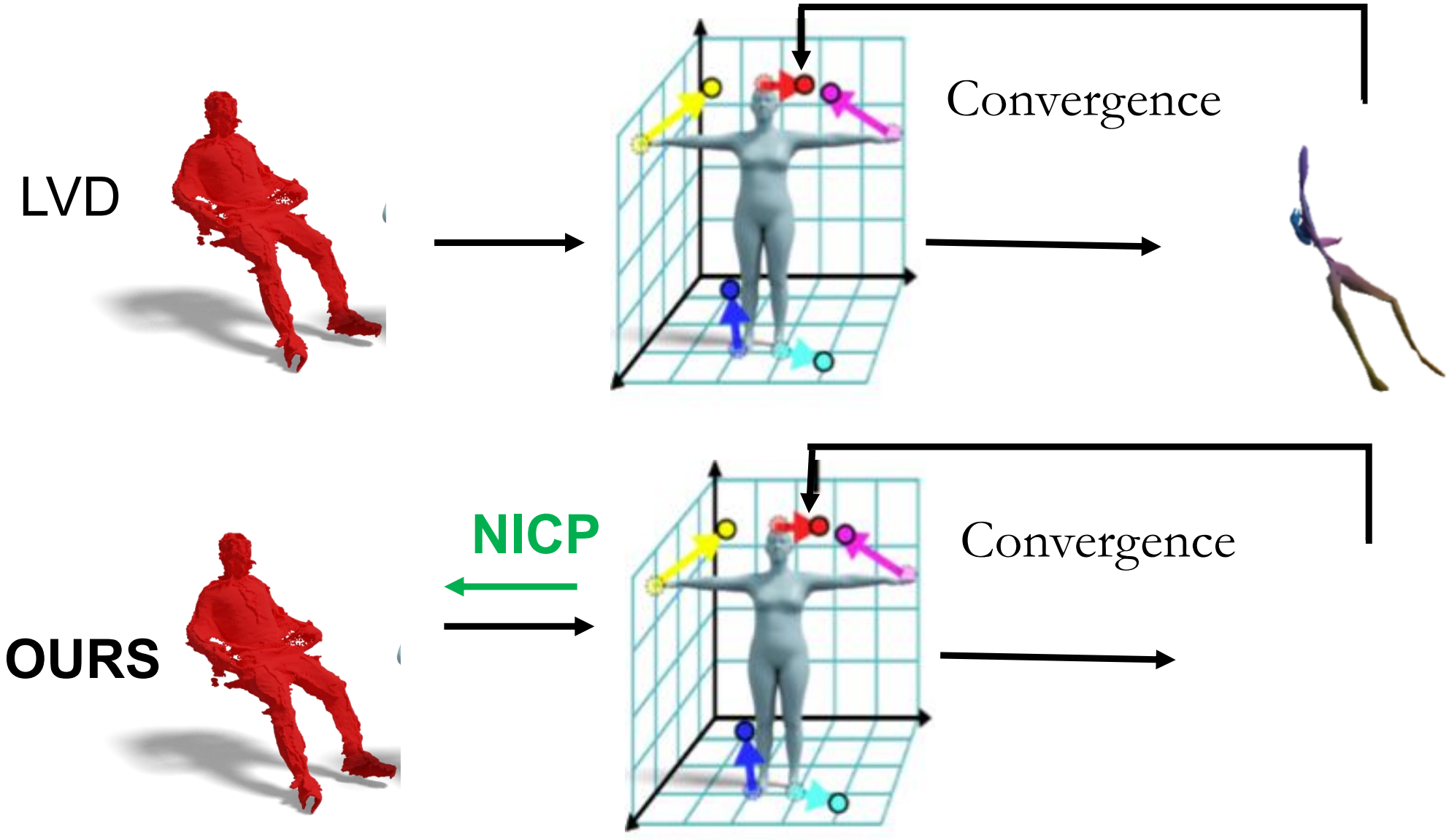
OURS



Solution

Neural ICP (NICP)

Self-Supervised tuning on Target Geometry (inference)

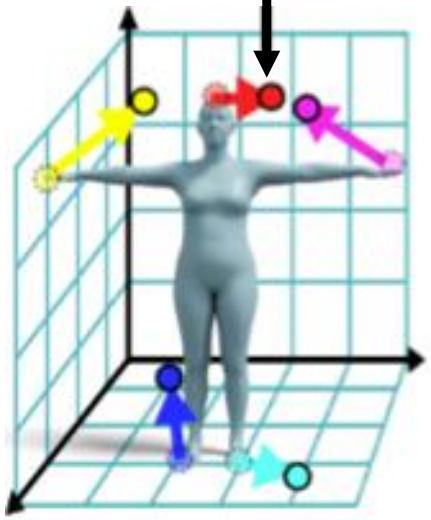


Solution

Neural ICP (NICP)

Self-Supervised tuning on Target Geometry (inference)

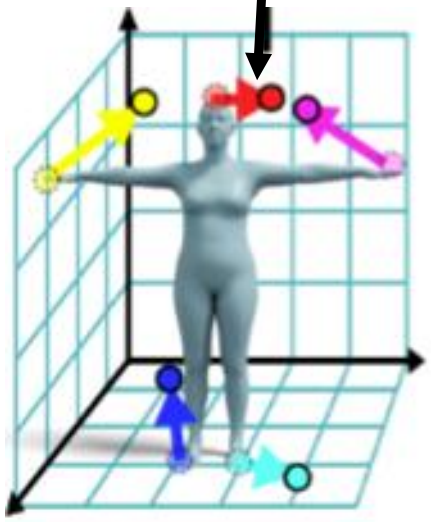
LVD



Convergence

NICP

OURS

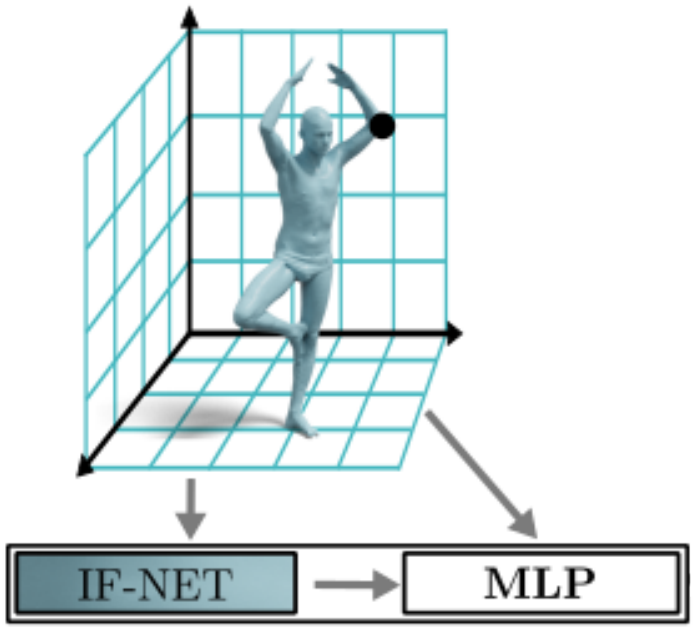


Convergence

Solution

Neural ICP (NICP)

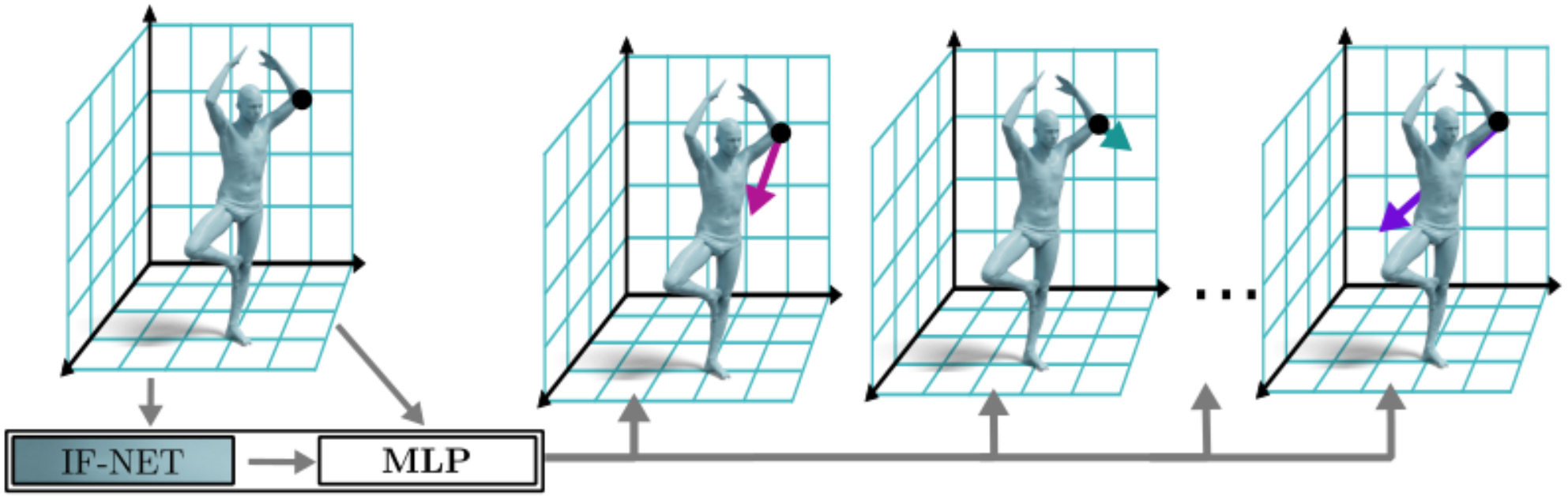
Self-Supervised tuning on Target Geometry (inference)



Solution

Neural ICP (NICP)

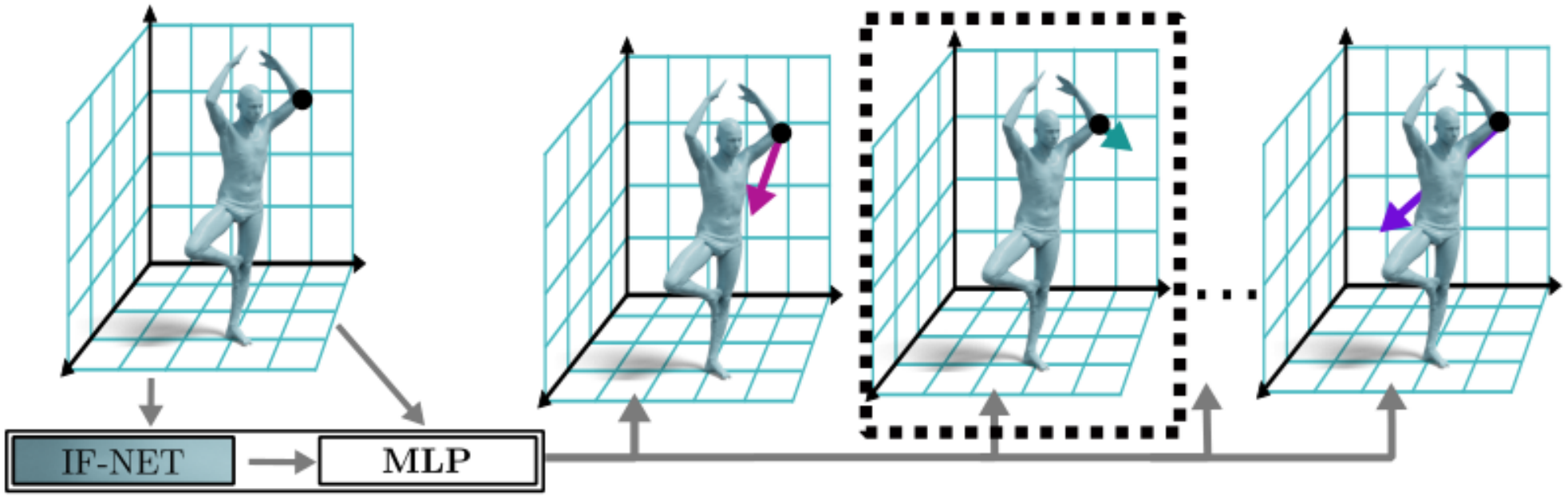
Self-Supervised tuning on Target Geometry (inference)



Solution

Neural ICP (NICP)

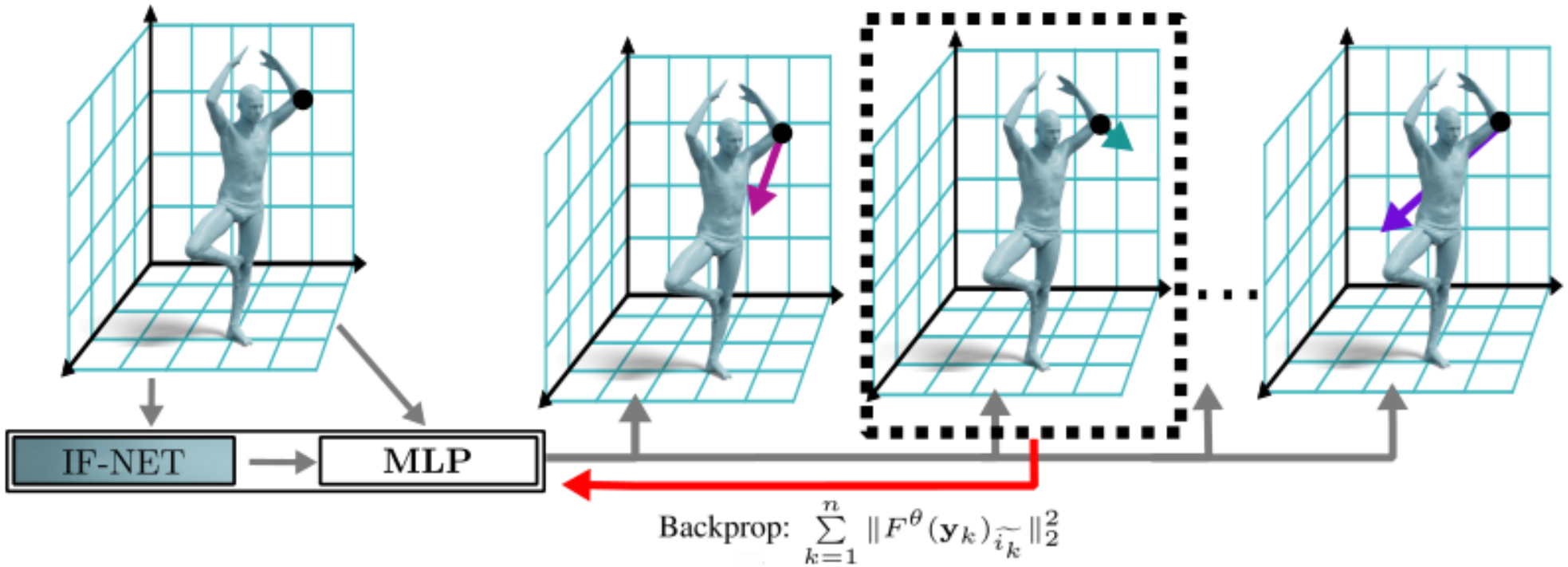
Self-Supervised tuning on Target Geometry (inference)



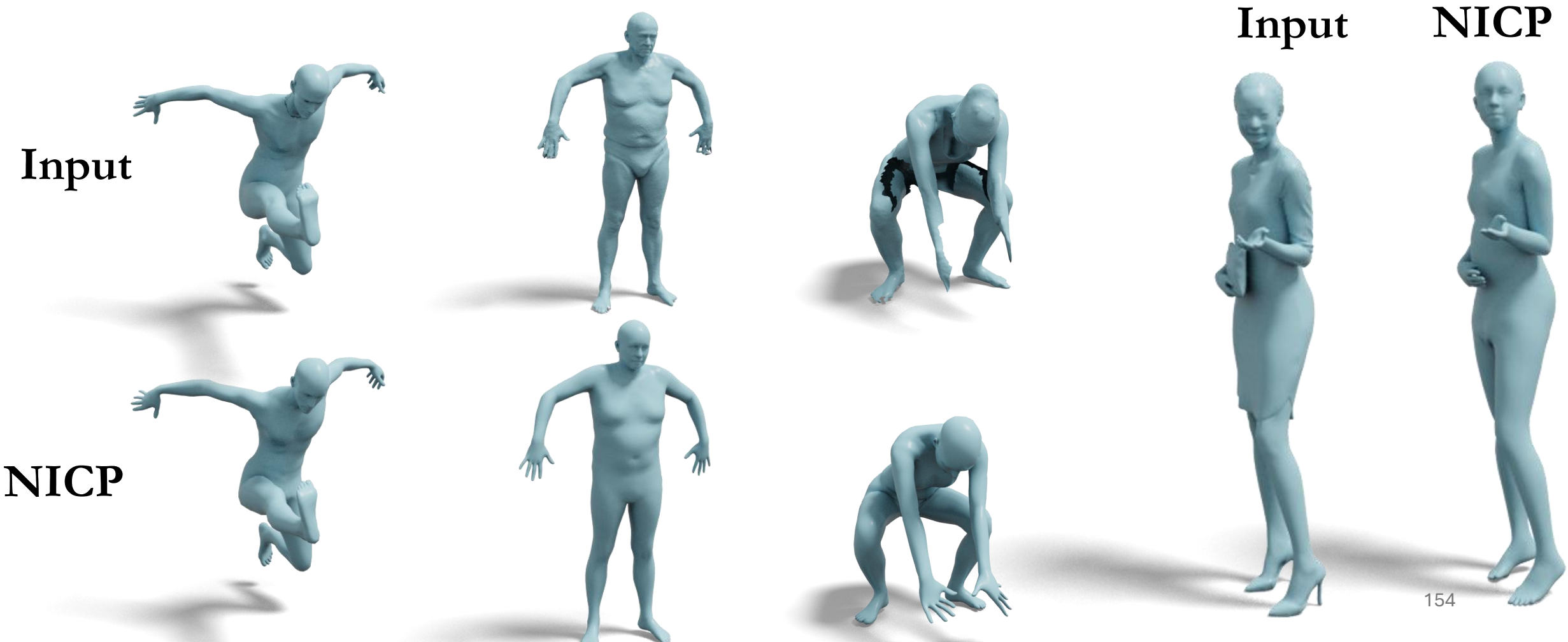
Solution

Neural ICP (NICP)

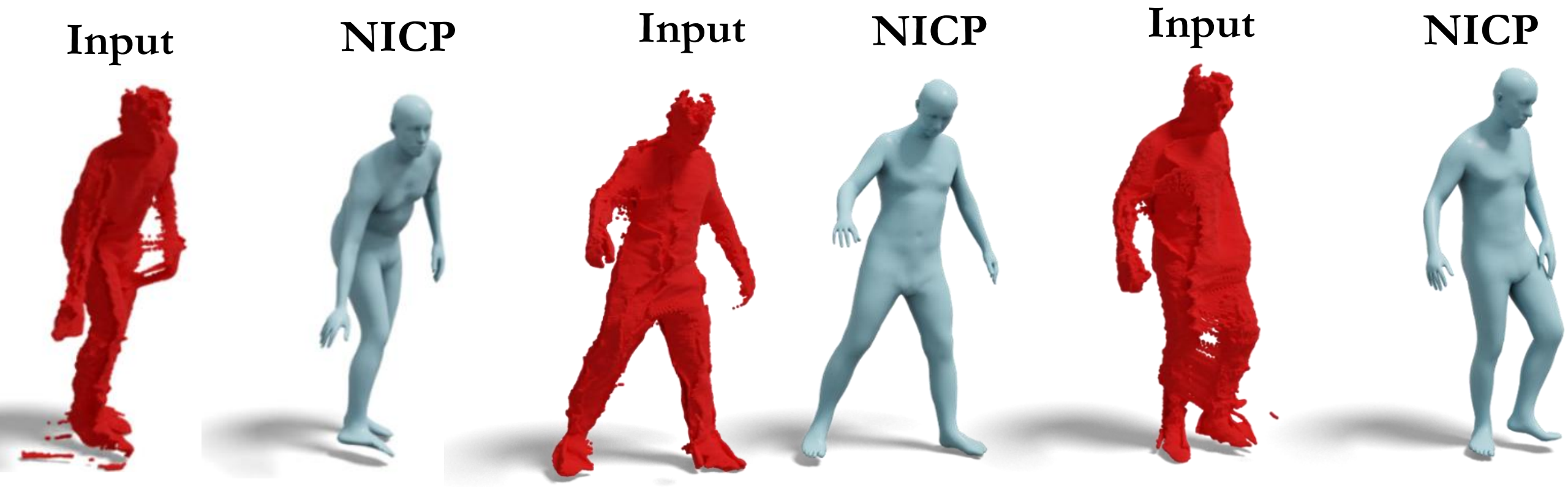
Self-Supervised tuning on Target Geometry (inference)



Results – Real Challenges



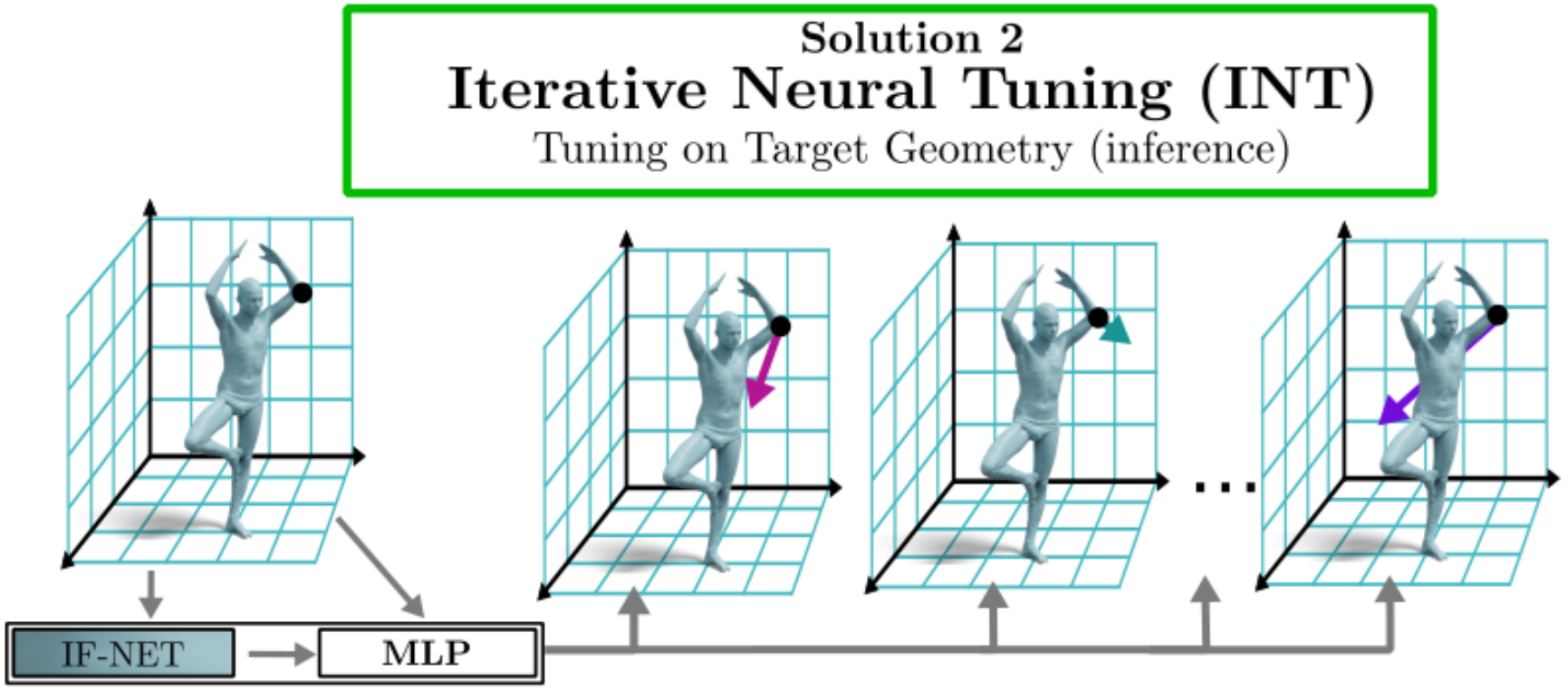
Results – Real Noise Scans



NICP: Neural ICP for 3D Human Registration at Scale

Riccardo Marin[1,2], Enric Corona[3], and Gerard Pons-Moll[1,2,4]

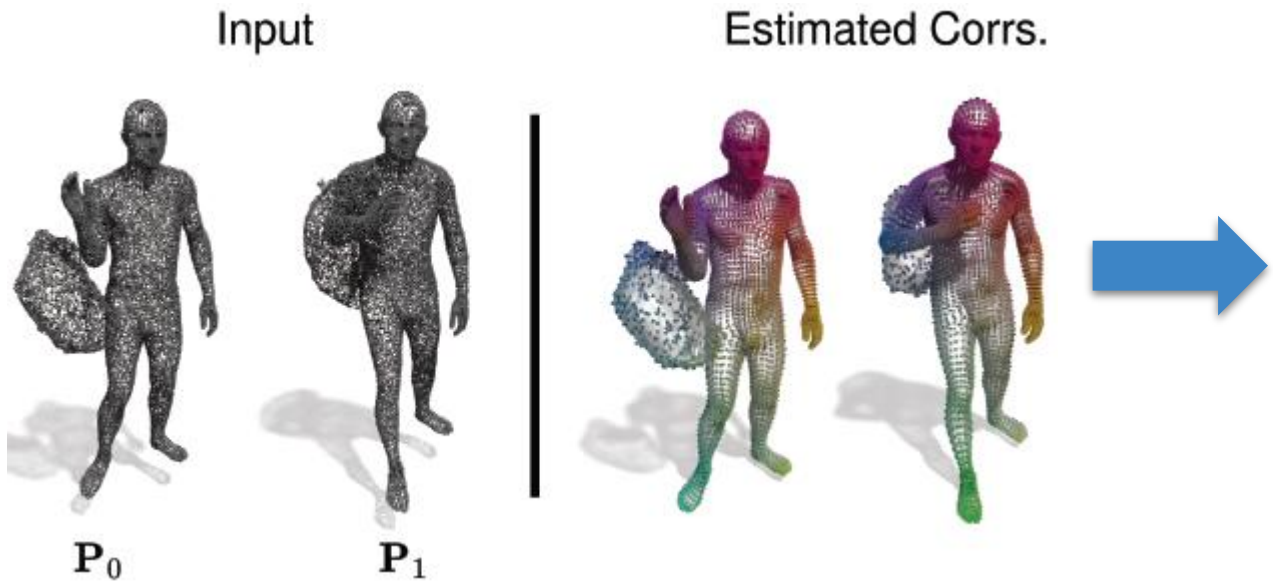
NICP



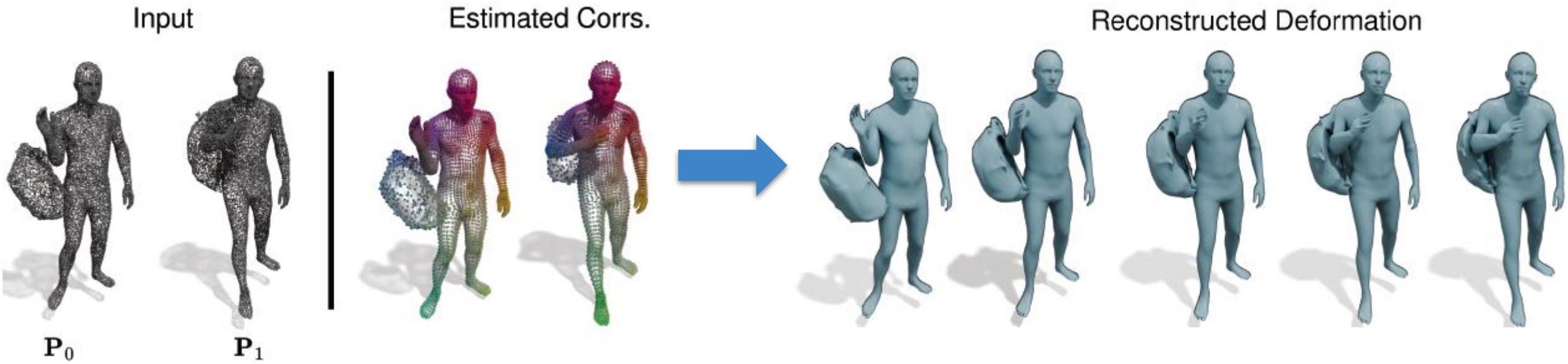
Deformation should converge on the target surface.

Can we enforce further constraints on the deformation field?

Goal: 3D to 4D reconstruction

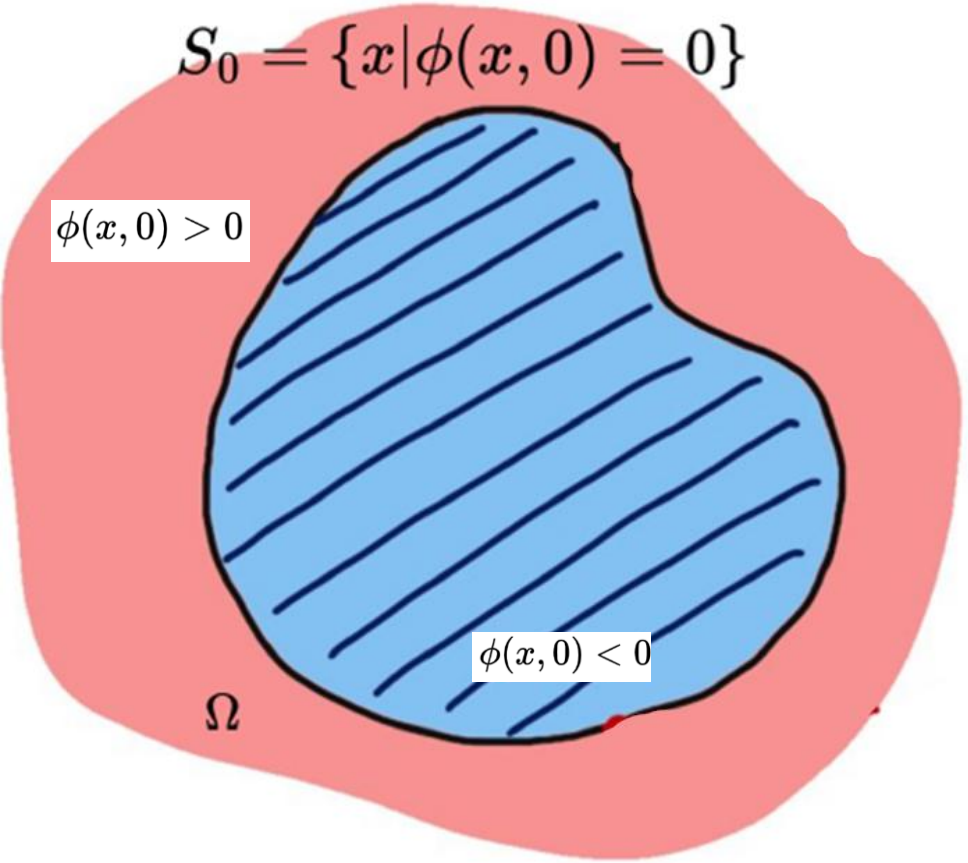


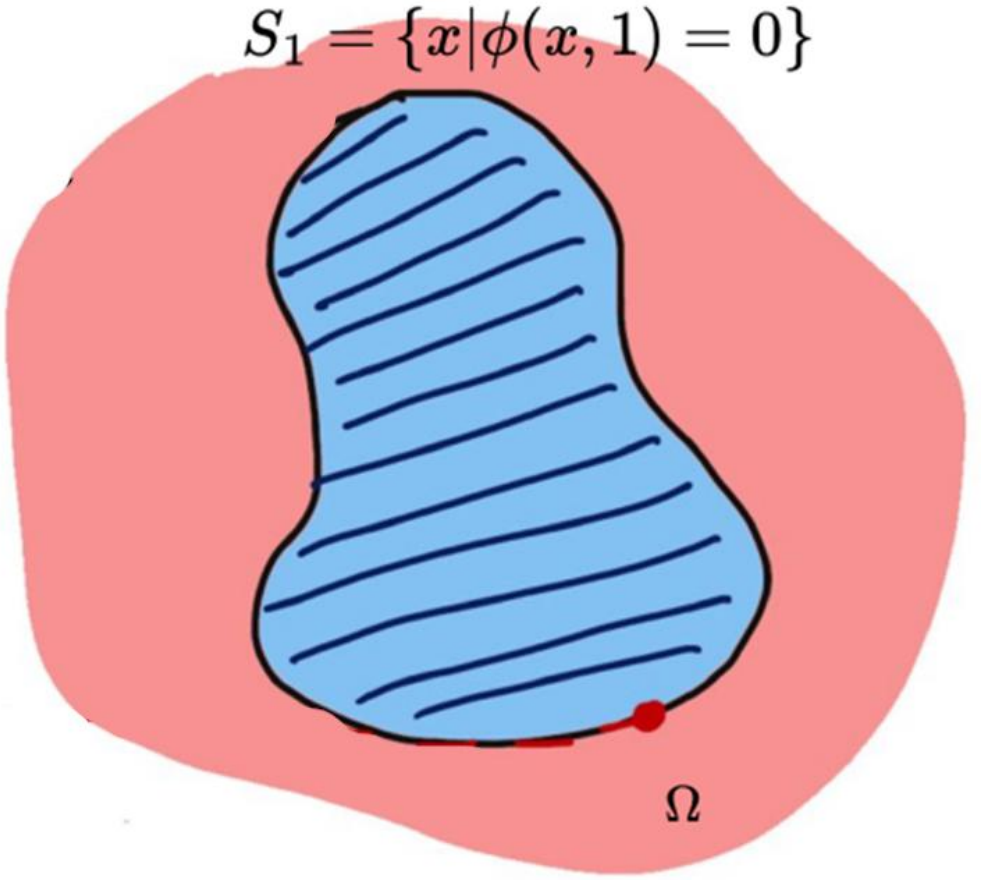
Goal: 3D to 4D reconstruction



The level-set method

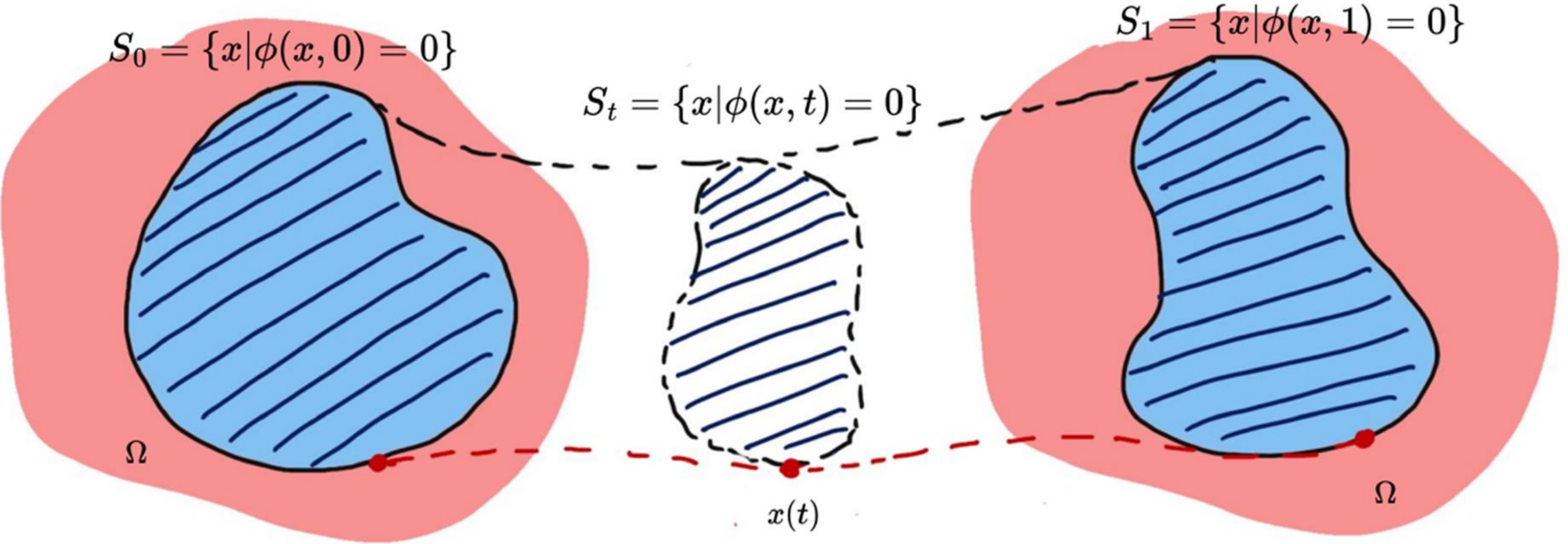
$\phi(\mathbf{x}(t), t)$ Implicit Representation





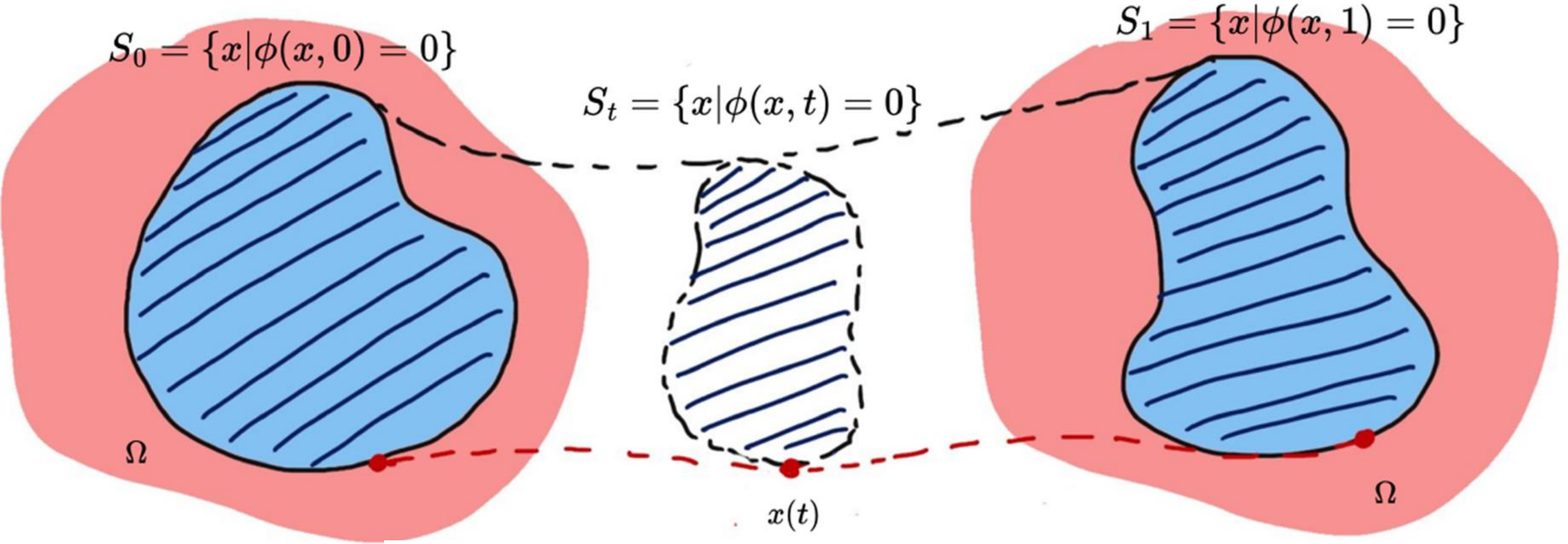
The level-set method

$\phi(\mathbf{x}(t), t)$ Implicit Representation



The level-set method

$\phi(\mathbf{x}(t), t)$ Implicit Representation



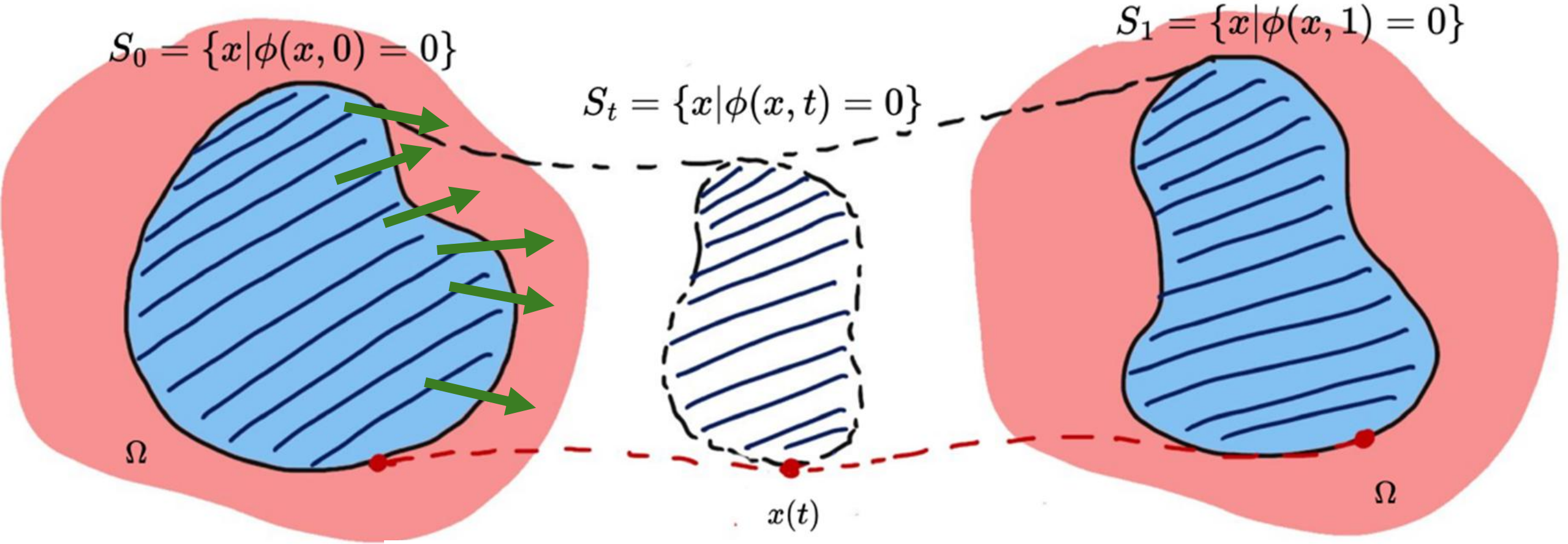
$$\frac{d}{dt}\phi(\mathbf{x}(t), t) = \nabla\phi(\mathbf{x}(t), t) \cdot \frac{d\mathbf{x}}{dt} + \frac{\partial\phi}{\partial t}(\mathbf{x}(t), t) = 0$$

The change over time of the implicit rep.…

… is function of its steepness times velocity

The level-set method

$\phi(\mathbf{x}(t), t)$ Implicit Representation



$$\frac{d}{dt}\phi(\mathbf{x}(t), t) = \nabla\phi(\mathbf{x}(t), t) \cdot \frac{d\mathbf{x}}{dt} + \frac{\partial\phi}{\partial t}(\mathbf{x}(t), t) = 0$$

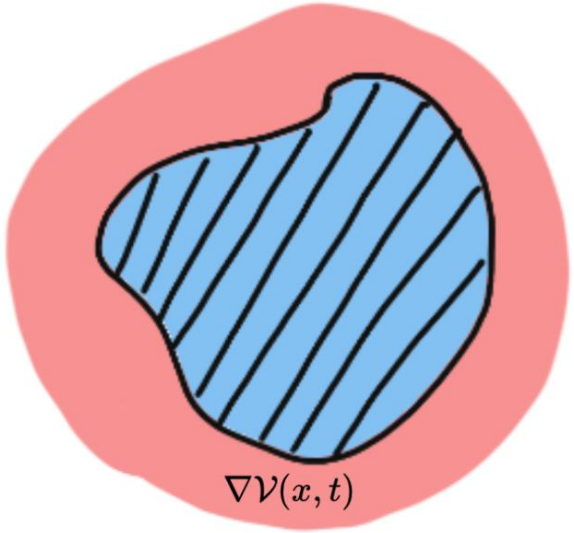
The change over time of the implicit rep.…

… is function of its steepness times velocity

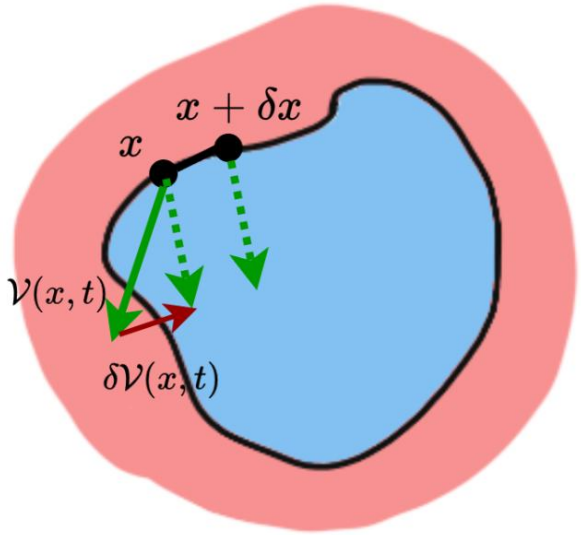
Neural Surface Deformation Via Velocity Fields

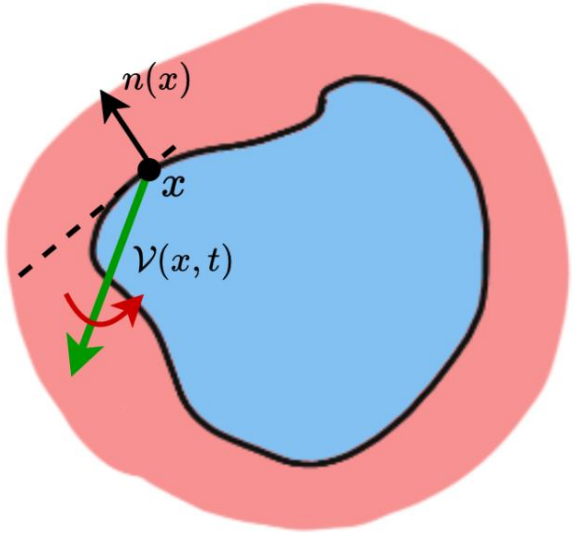
Explicit velocity lets us enforce a number of physical and geometrical constraints!

Volume preserving

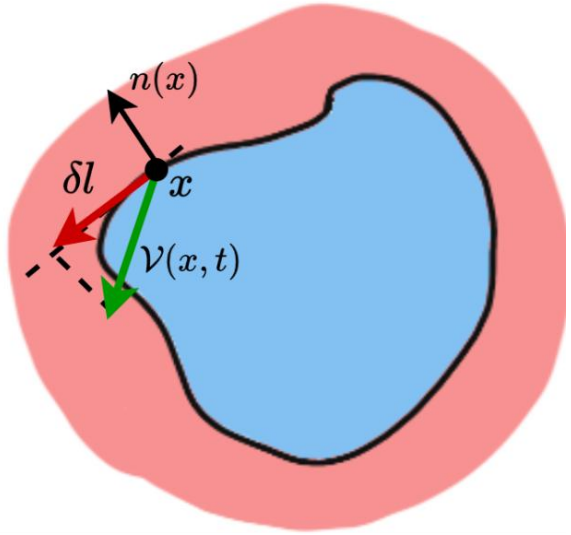


Spatial smoothness





Distortion constraint



Stretching constraint

Physically Plausible Deformation

Level-set Loss

$$\mathcal{L}_i = \int_\Omega \|\partial_t \phi + \mathcal{V} \cdot \nabla \phi + \lambda_l \phi \mathcal{R}(x,t)\|_{l^2} dx \qquad \mathcal{R}(\mathbf{x},t) = -\langle (\nabla \mathcal{V}) \frac{\nabla \phi}{\|\nabla \phi\|}, \frac{\nabla \phi}{\|\nabla \phi\|} \rangle$$

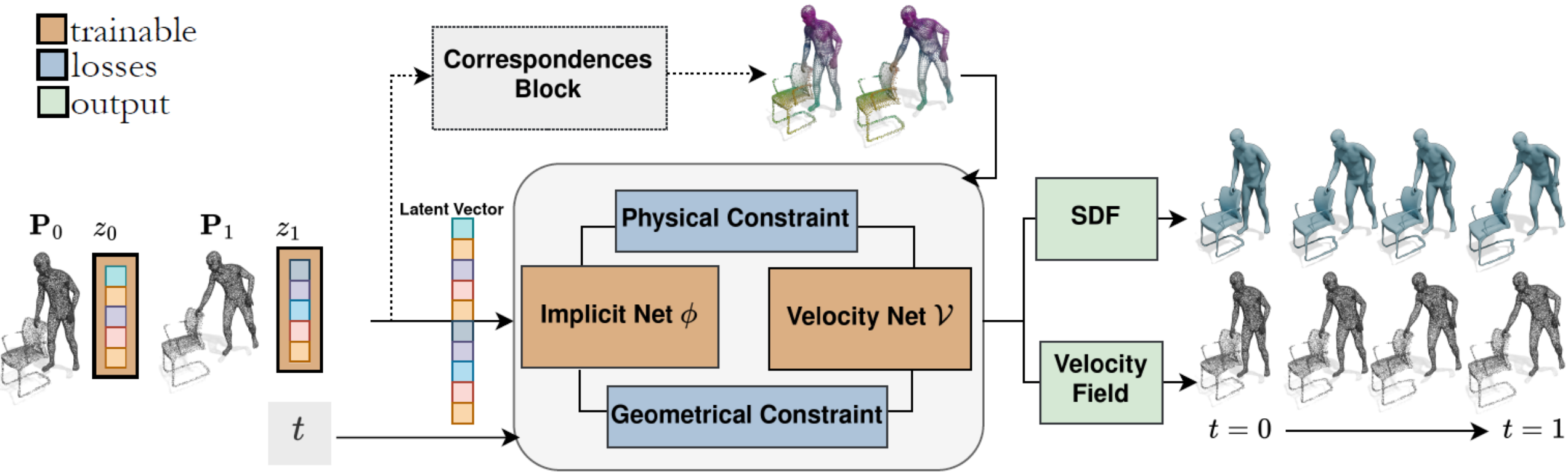
Implicit Net Variable

Velocity Net Variable

- Spatial smoothness loss $\mathcal{L}_s = \int_\Omega ||(-\alpha\Delta + \gamma \mathbf{I})\mathcal{V}(\mathbf{x},t)||_{l^2} d\mathbf{x}$
- Volume preservation loss $\mathcal{L}_v = \int_\Omega |\nabla \cdot \mathcal{V}(\mathbf{x},t)| d\mathbf{x}$
- Stretching loss $\mathcal{L}_{st} = \int_\Omega ||\mathbf{P}^\top (\nabla \mathcal{V}^\top \nabla \mathcal{V} + \nabla \mathcal{V} + \nabla \mathcal{V}^\top)\mathbf{P}||_F d\mathbf{x} \qquad \mathbf{P} = \mathbf{I} - \nabla\phi\nabla\phi^\top$
- Distortion loss $\mathcal{L}_d = \int_\Omega ||\frac{1}{6} Tr(\mathbf{D})^2 - \frac{1}{2} Tr(\mathbf{D}\cdot\mathbf{D})^2||_F d\mathbf{x} \qquad \mathbf{D} = \frac{1}{2}(\nabla\mathcal{V} + (\nabla\mathcal{V})^\top)$

Sang, L et. al, 4Deform: Neural Surface Deformation for Robust Shape Interpolation (CVPR 2025)

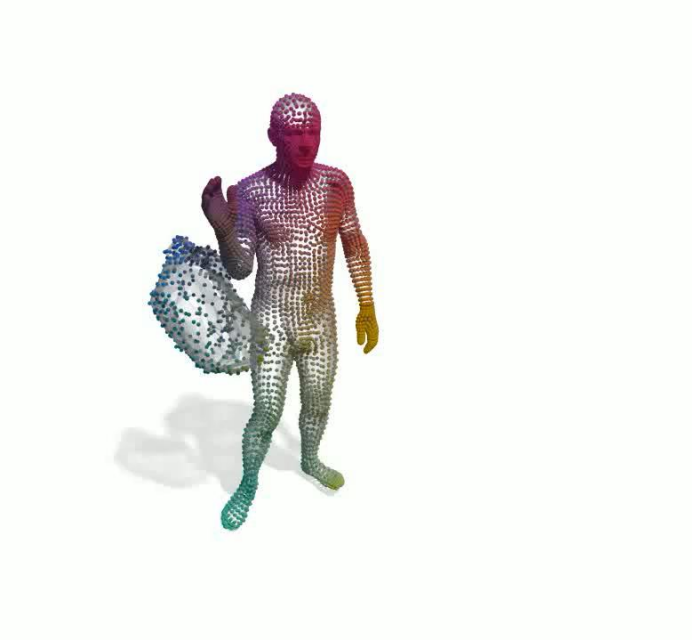
4Deform: Neural Surface Deformation for Robust Shape Interpolation



Upsampling with change of topology

Training:

1FPS



Upsampling with change of topology

Training:

1FPS

Inference Input:

1FPS

Upsampling with change of topology

Training:

1FPS

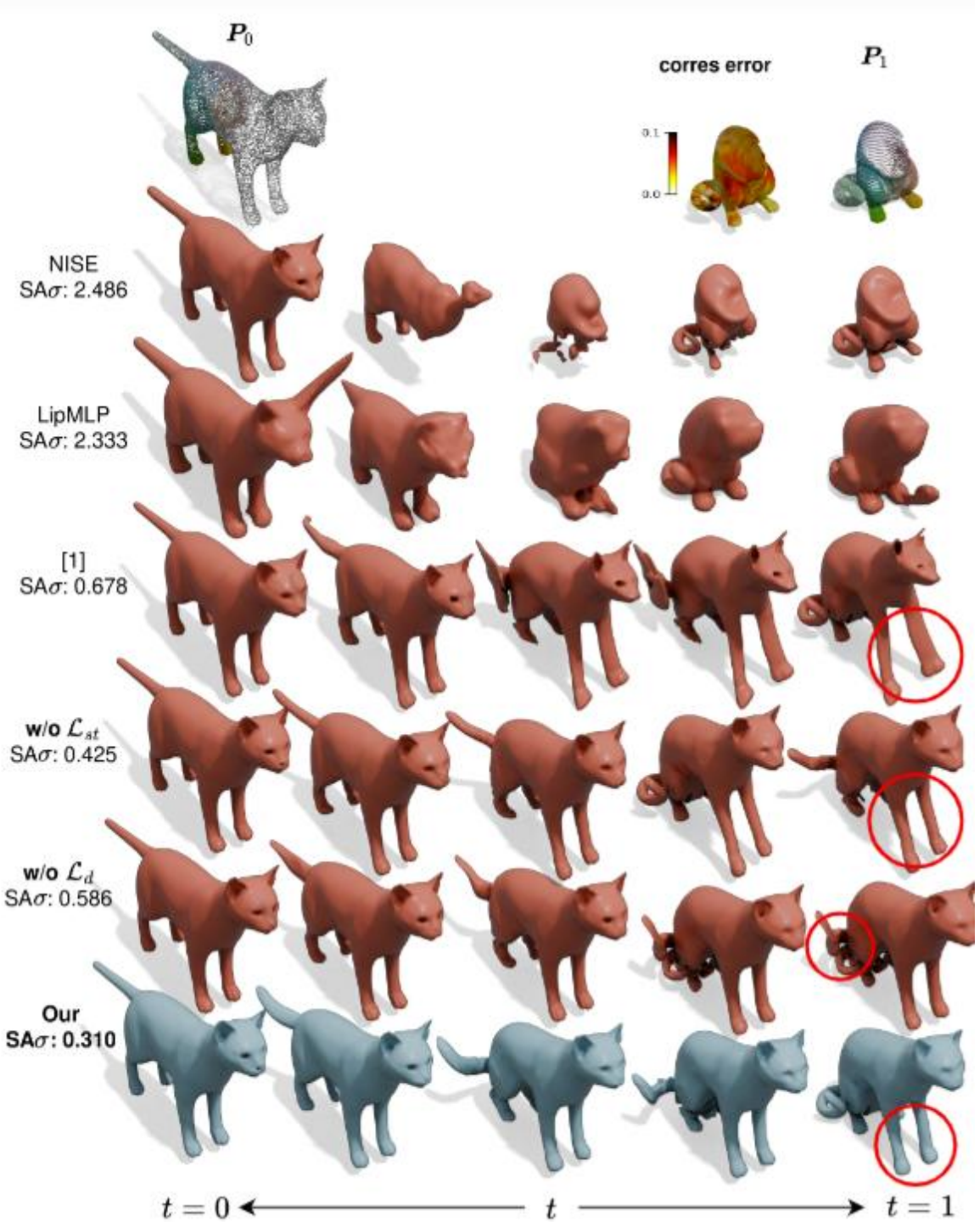
Inference Input:

1FPS

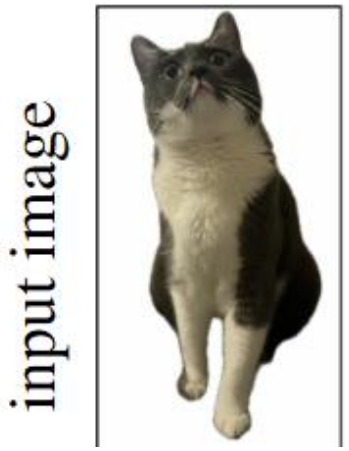
Output:

Continous

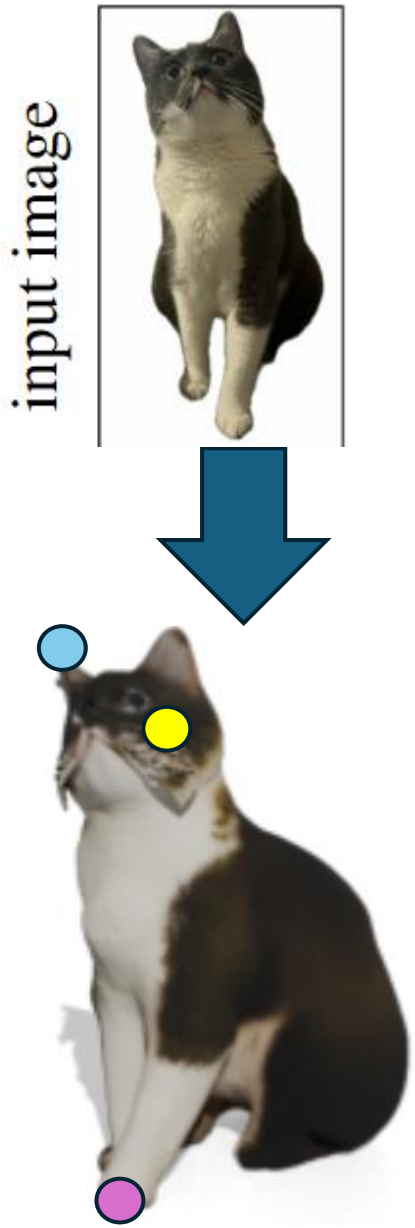
Partiality

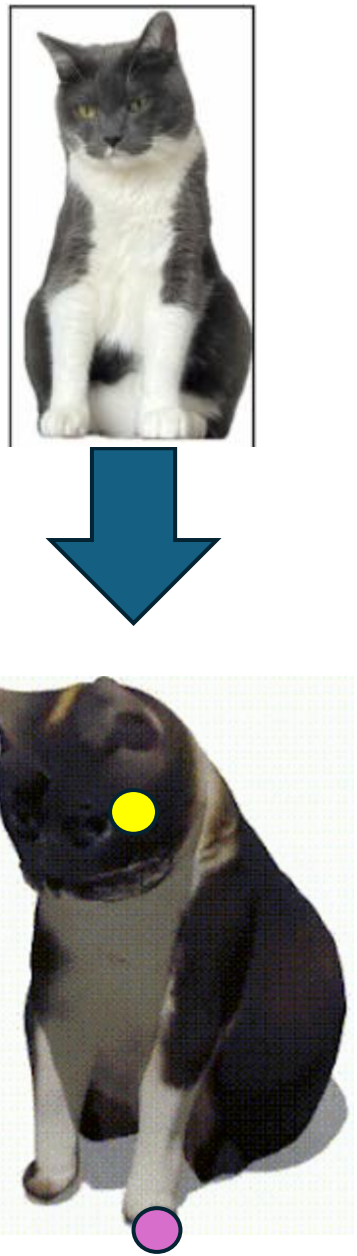

P_0
corres error
P_1
0.1
0.0
NISE
SAσ: 2.486
LipMLP
SAσ: 2.333
[1]
SAσ: 0.678
w/o $\mathcal{L}_{st}$
SAσ: 0.425
w/o $\mathcal{L}_d$
SAσ: 0.586
Our
SAσ: 0.310
$t = 0$
t
$t = 1$

We can also start from 2D images

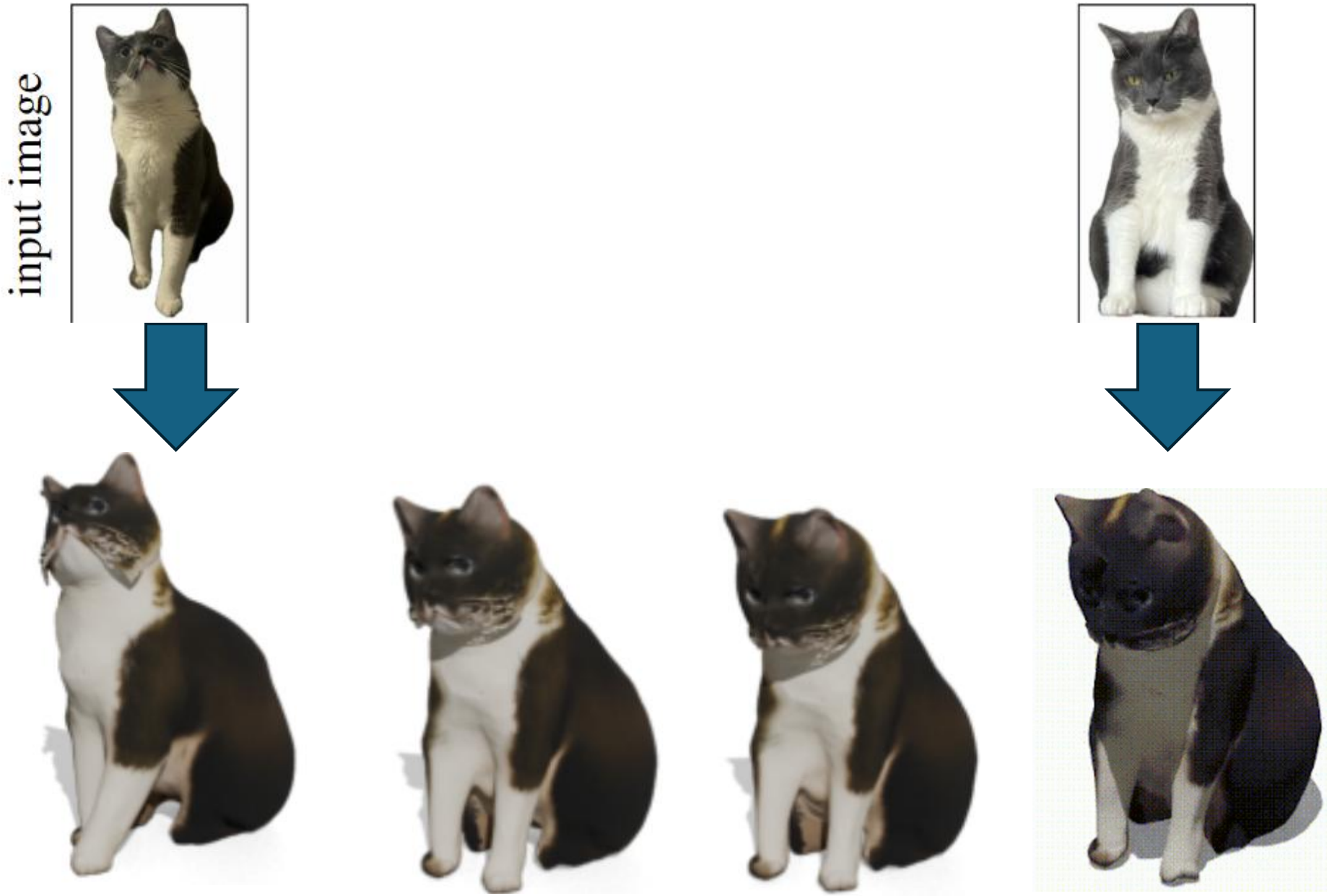


We can also start from 2D images





We can also start from 2D images



Input images

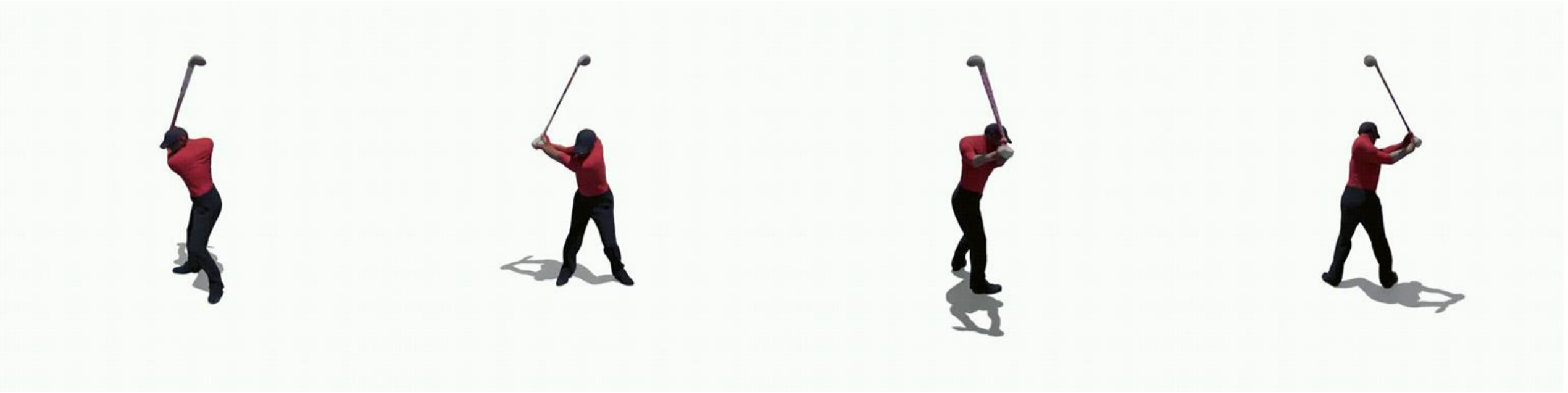
Input images

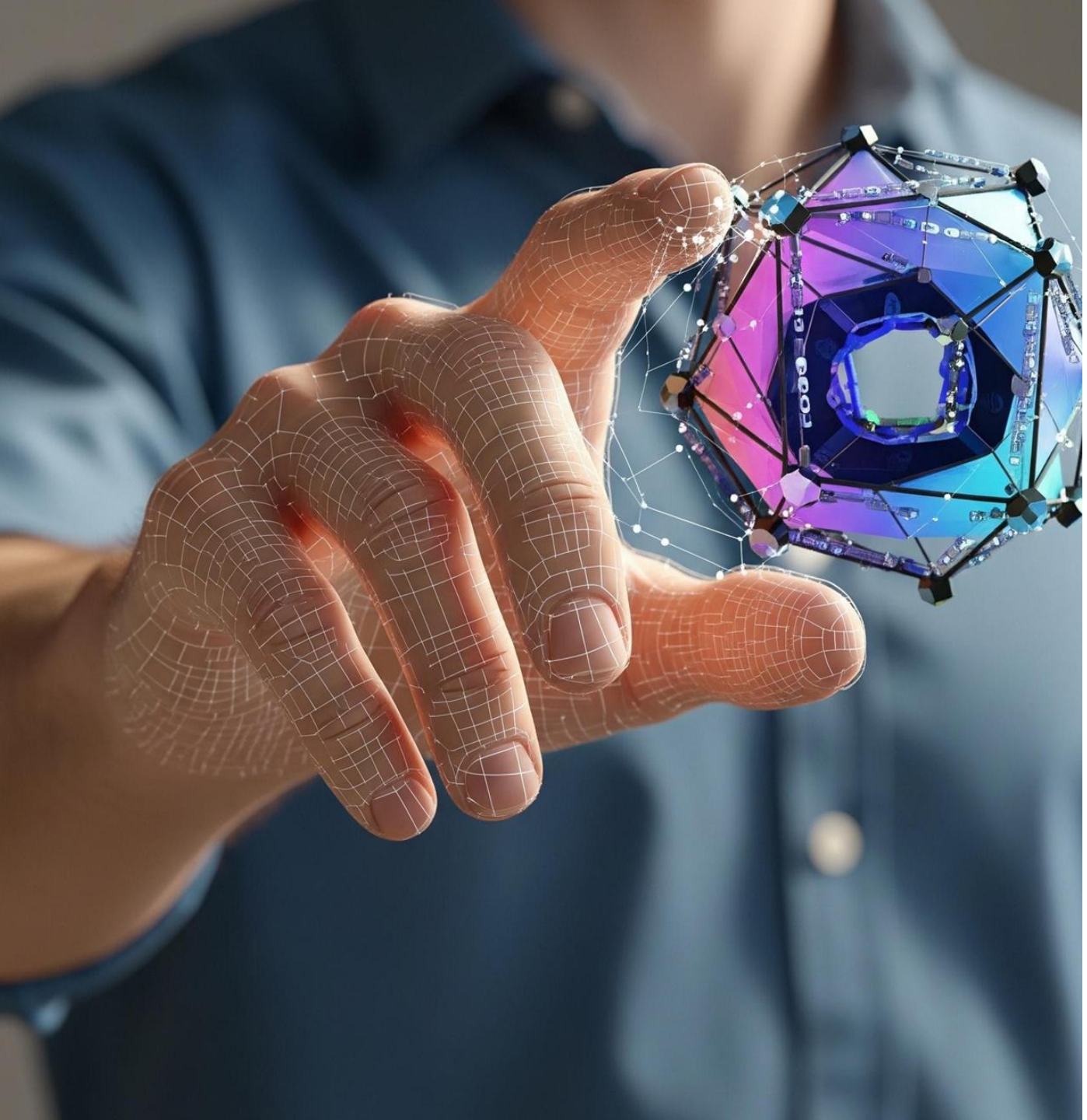




Input images

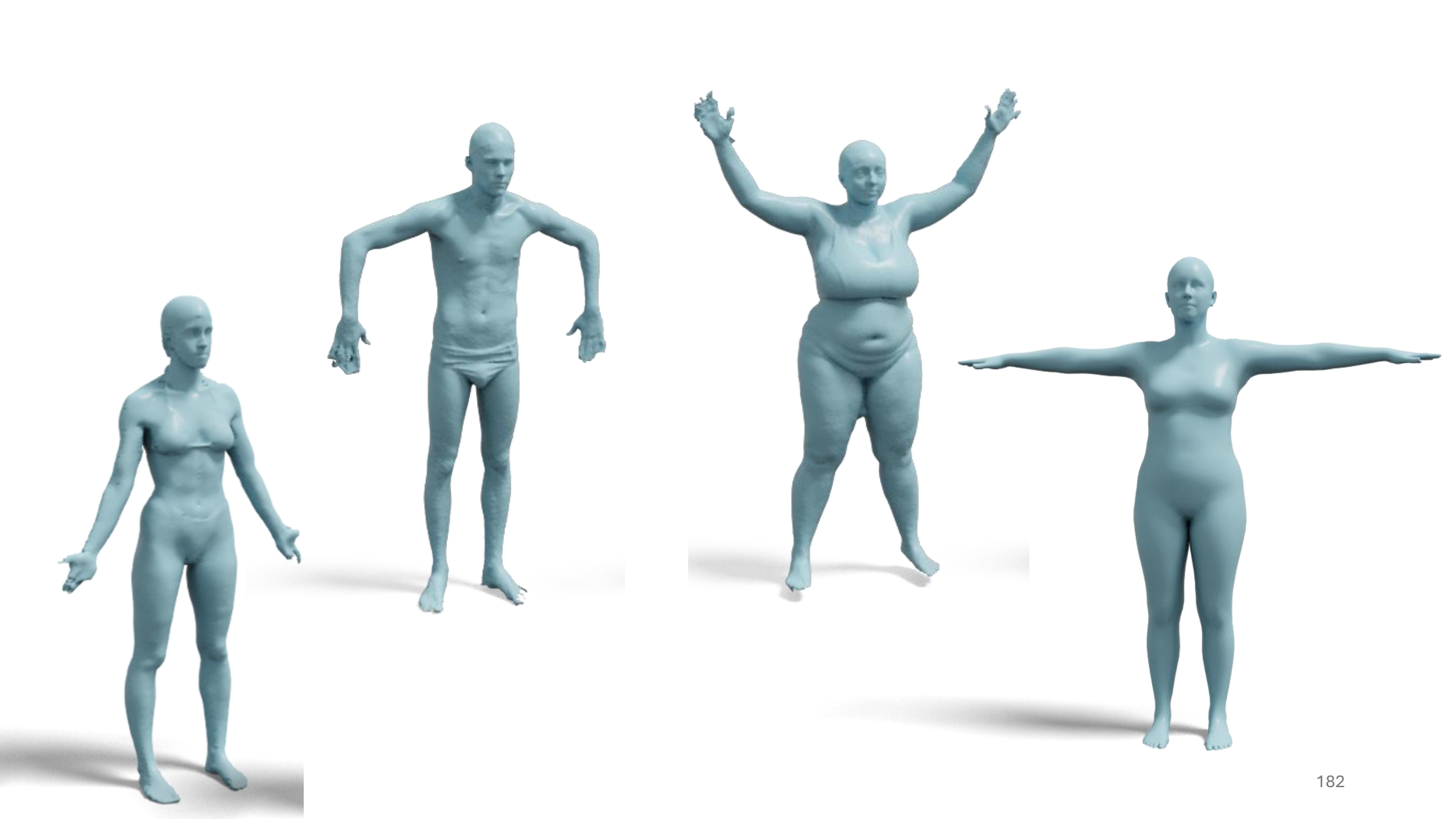
Input images





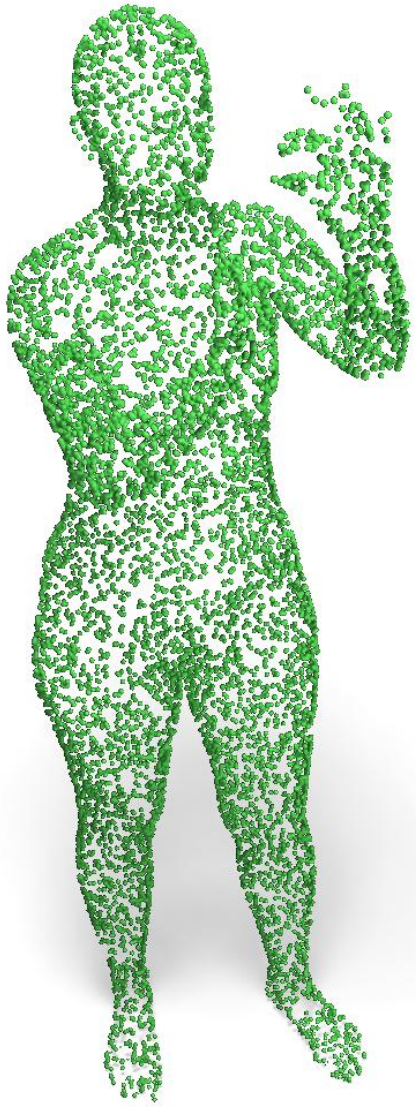
Beyond isolated humans

Humans do not live in isolation



Input

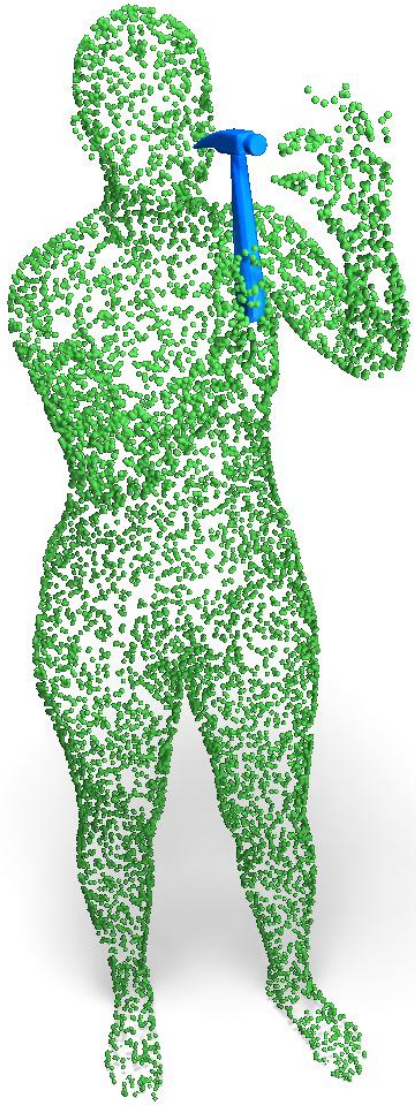
Point Cloud



Object pop-up

Output

Posed object



Object pop-up: Can we infer 3D objects and their poses from human interaction alone?




Ilya A. Petrov[1,2] Riccardo Marin[1,2] Julian Chibane[1,3] Gerard Pons-Moll[1,2,3]

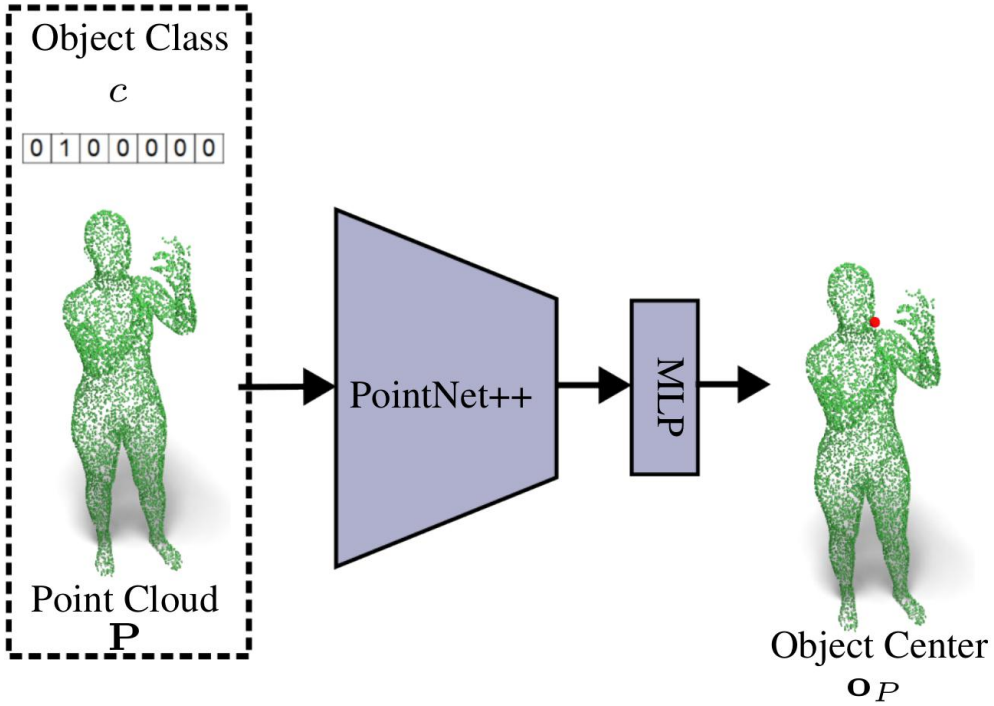
[1]University of Tübingen; [2]Tübingen AI Center;
[3]Max Planck Institute for Informatics, Saarland Informatics Campus

https://virtualhumans.mpi-inf.mpg.de/object_popup/

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

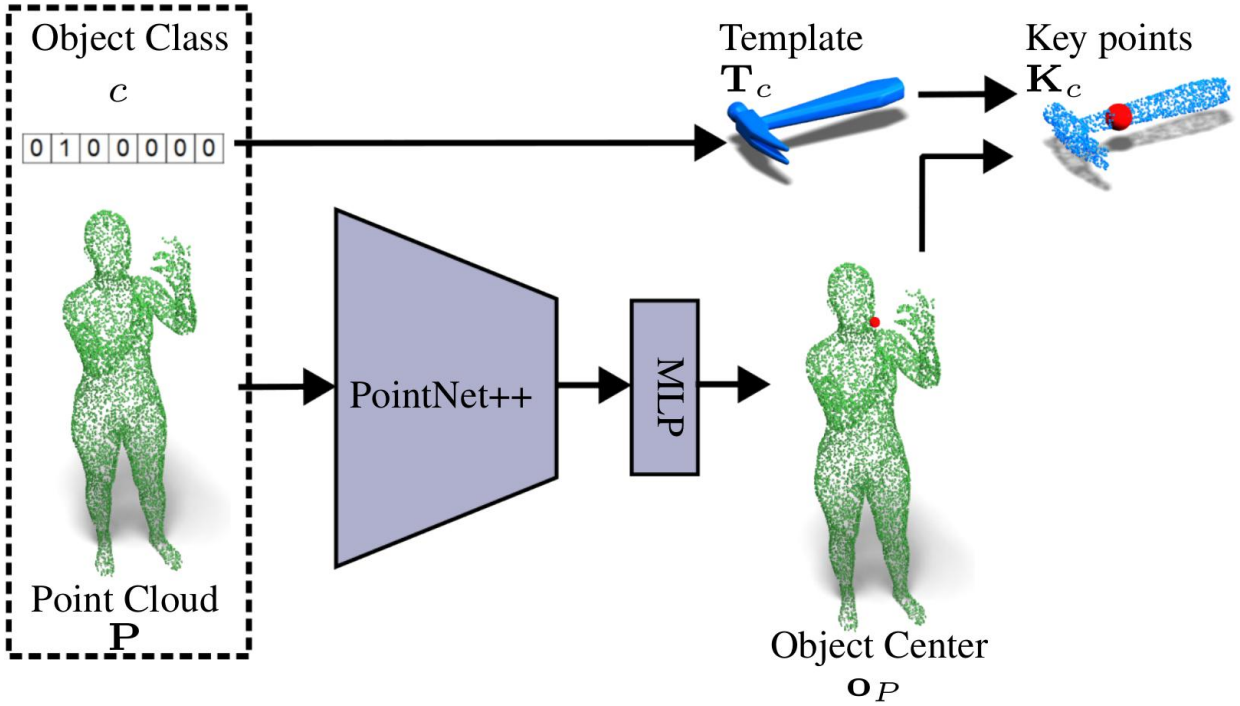
Includes Audio

Object pop-up: whole body features and center prediction



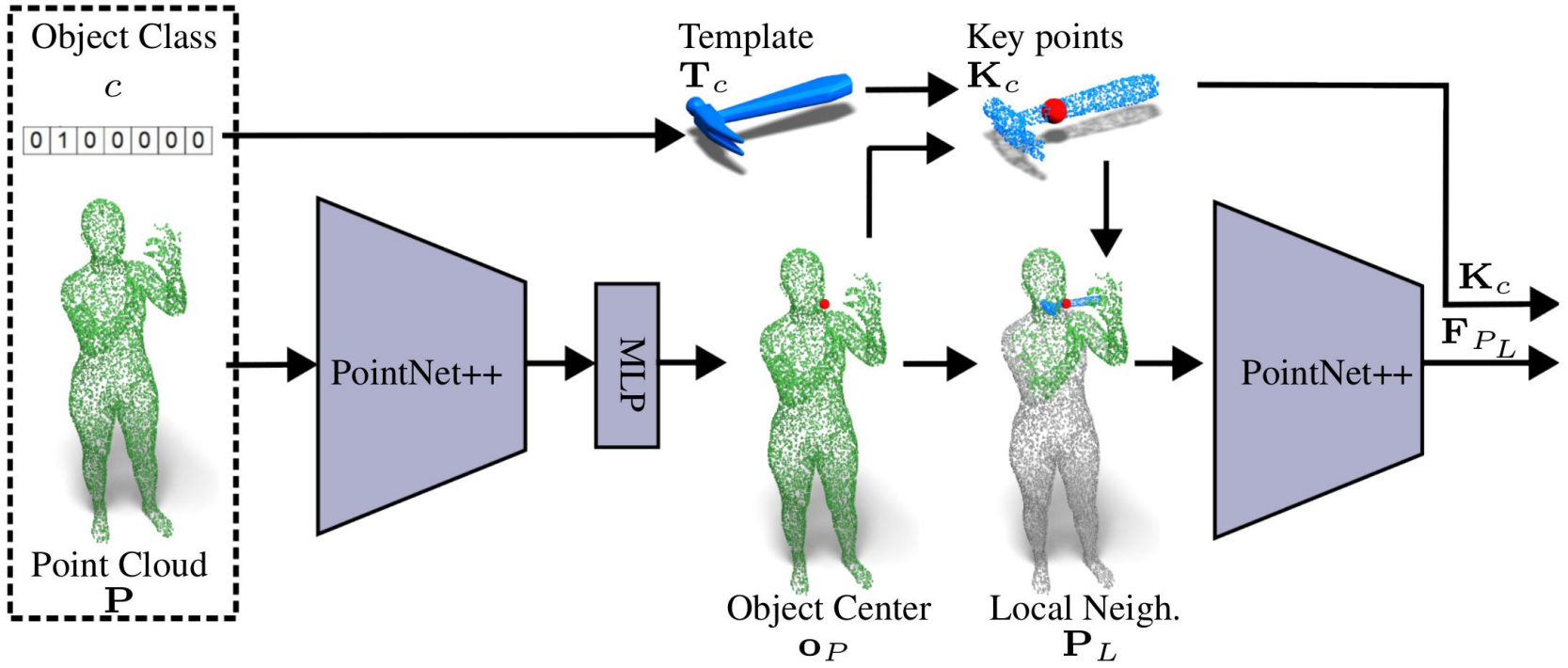
Petrov, I. A., Marin, R., Chibane, J., & Pons-Moll, G. (2023). Object pop-up: Can we infer 3d objects and their poses from human interactions alone?, CVPR 2023

Object pop-up: object keypoints



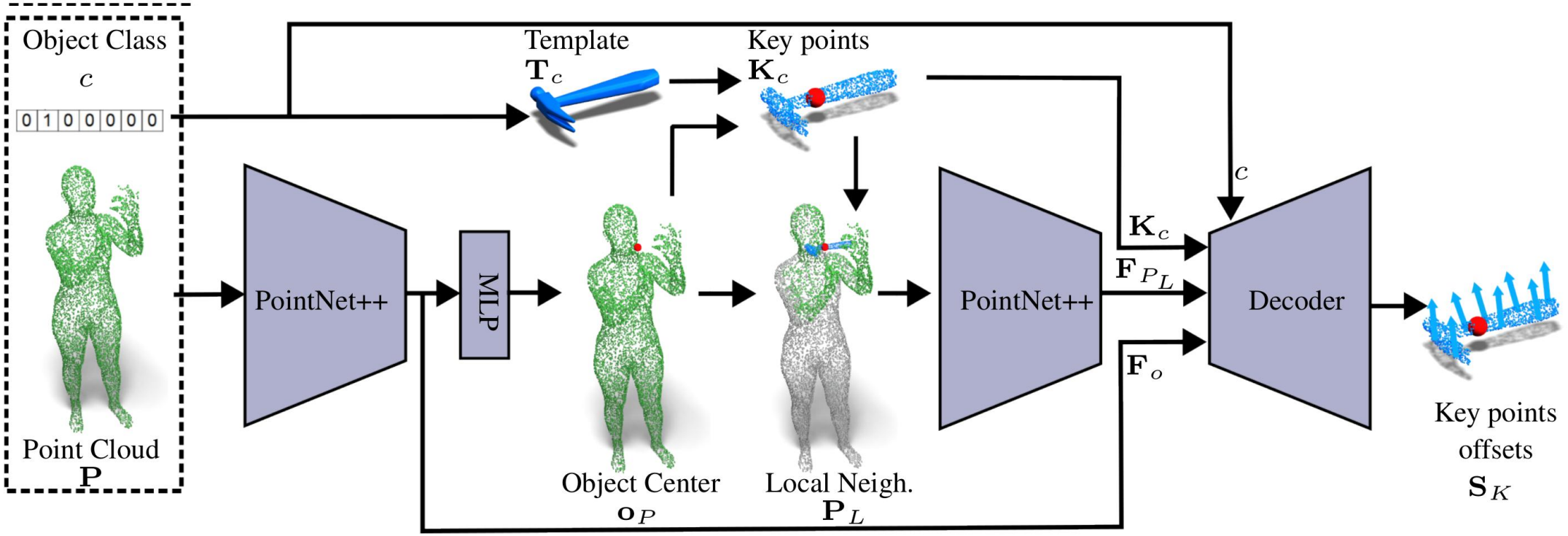
Petrov, I. A., Marin, R., Chibane, J., & Pons-Moll, G. (2023). Object pop-up: Can we infer 3d objects and their poses from human interactions alone?, CVPR 2023

Object pop-up: local features



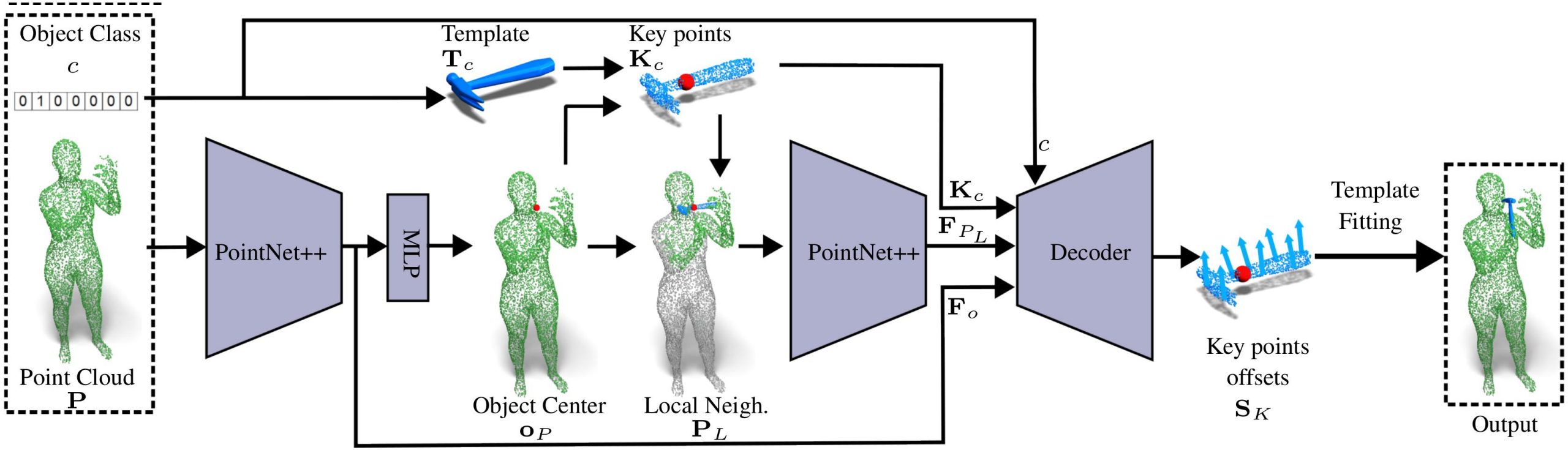
Petrov, I. A., Marin, R., Chibane, J., & Pons-Moll, G. (2023). Object pop-up: Can we infer 3d objects and their poses from human interactions alone?, CVPR 2023

Object pop-up: per-point offset prediction



Petrov, I. A., Marin, R., Chibane, J., & Pons-Moll, G. (2023). Object pop-up: Can we infer 3d objects and their poses from human interactions alone?, CVPR 2023

Object pop-up: overview

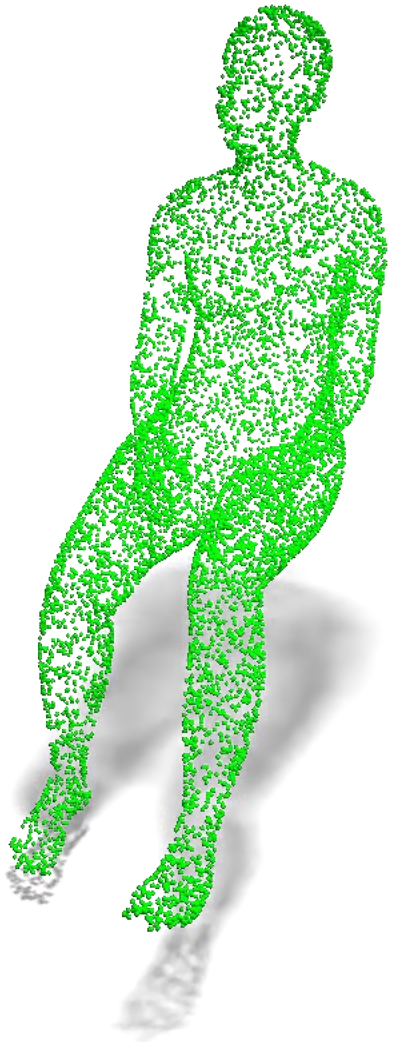


Petrov, I. A., Marin, R., Chibane, J., & Pons-Moll, G. (2023). Object pop-up: Can we infer 3d objects and their poses from human interactions alone?, CVPR 2023


Input

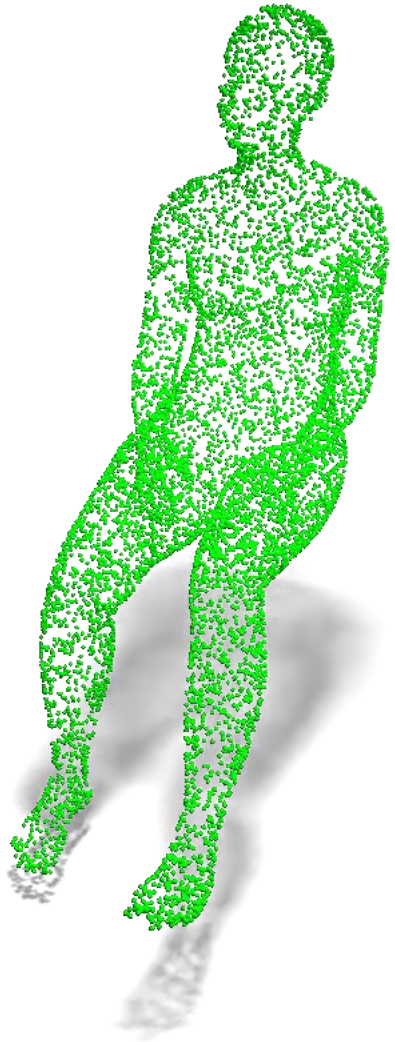

Input

Can we classify an object of interaction?

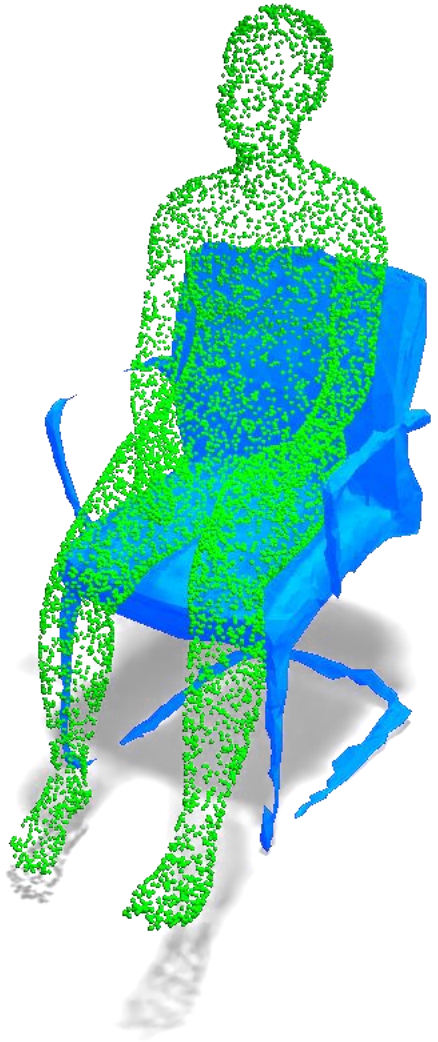


Yes

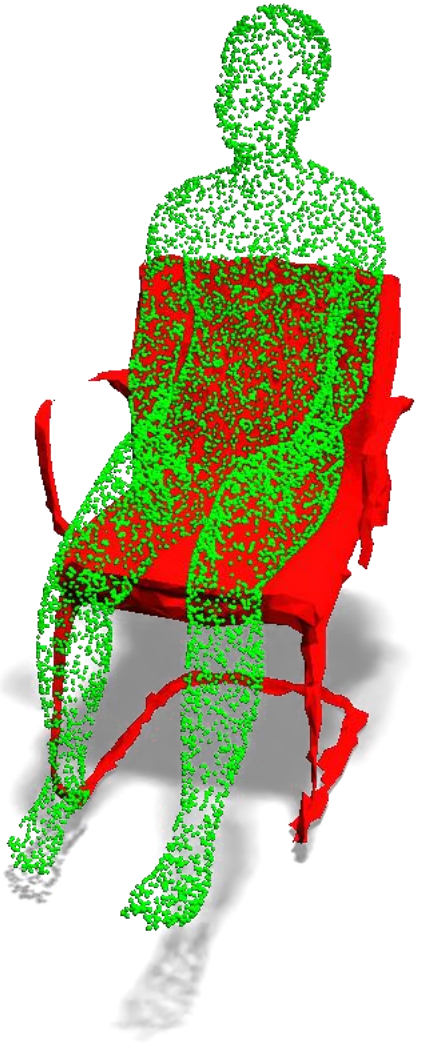
Input



Prediction



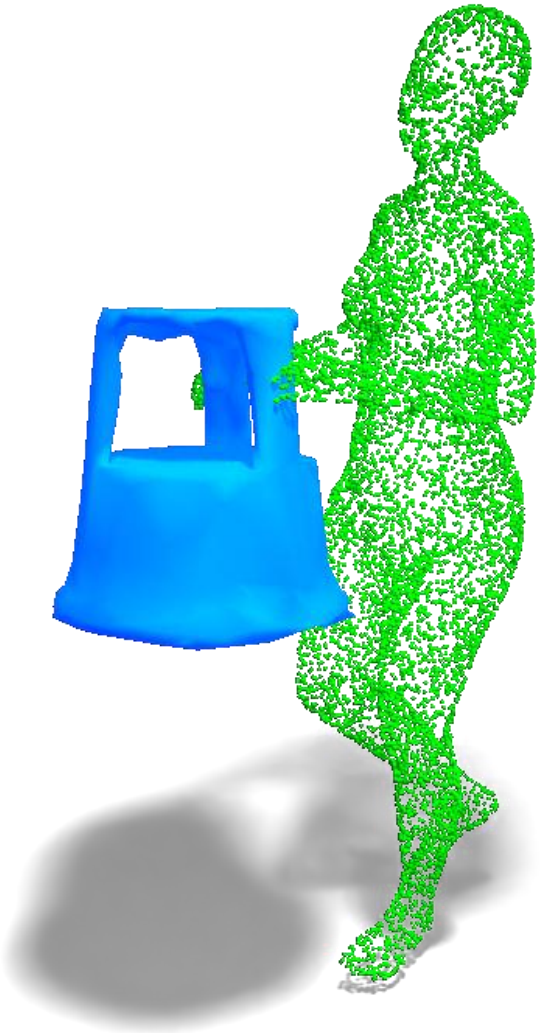
Ground-truth



But the task is inherently ambiguous

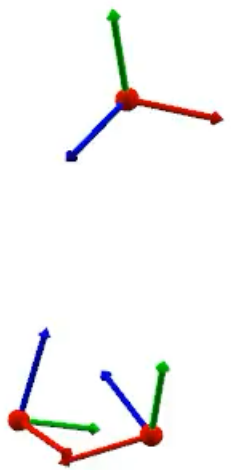
Input

Prediction



Ground-truth

Input: 3-point tracking

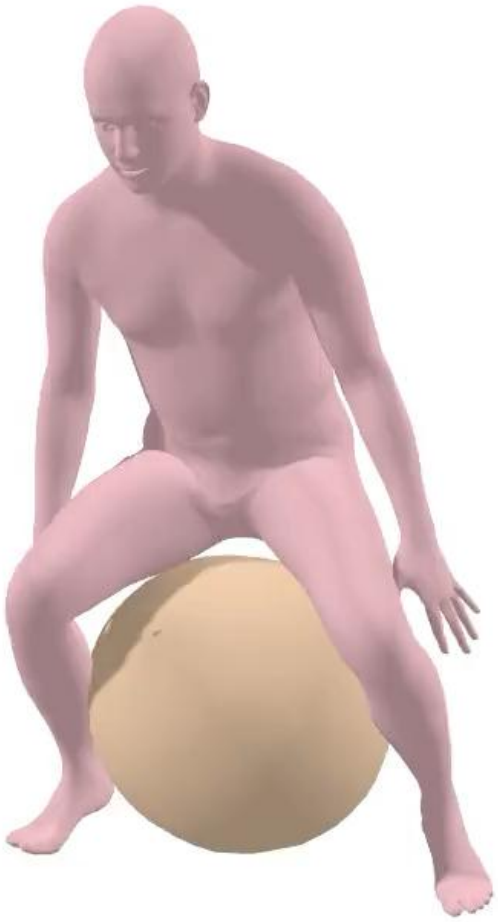


ECHO prediction: HOI

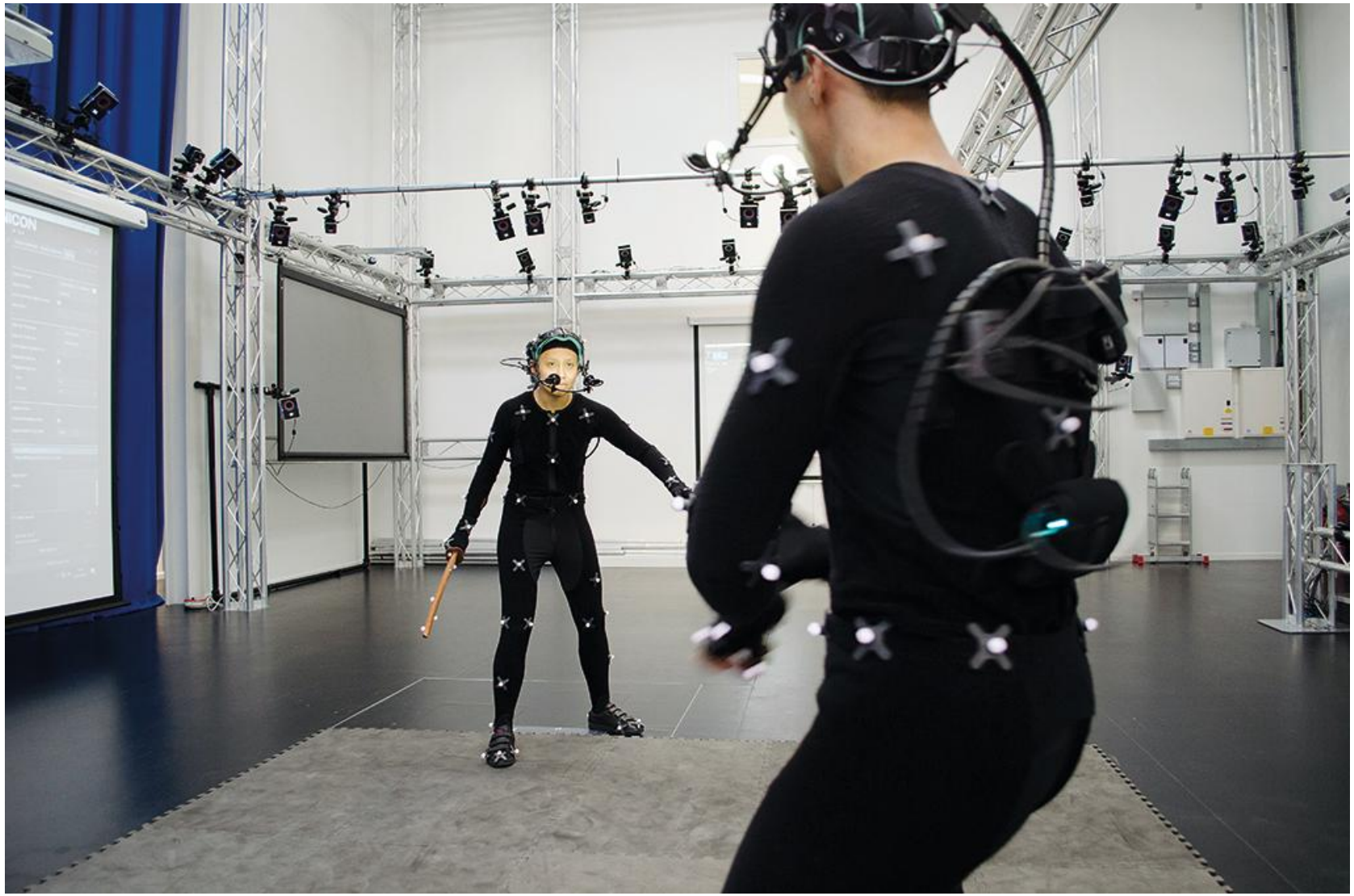
GT data

ECHO (ours)

BoDiffusion + Obj.



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

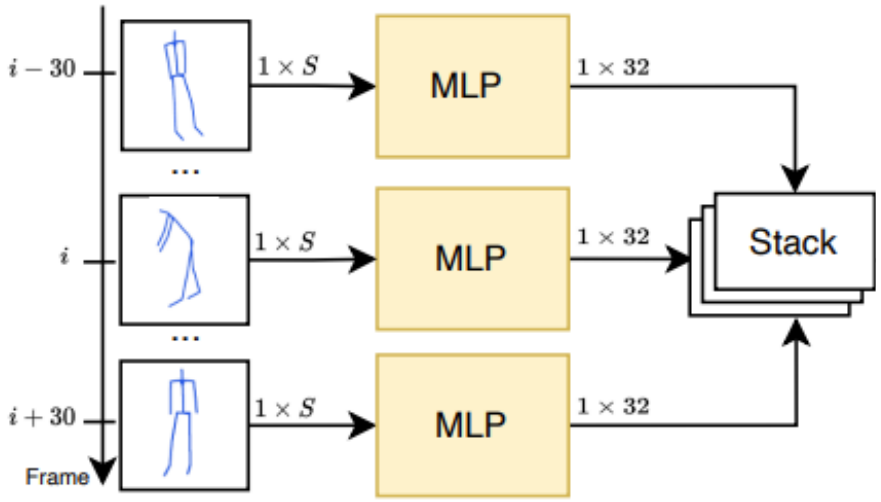
Interaction Replica: Tracking human-object interaction and scene changes from human motion; Guzov, Chibane, Marin, He, Sattler, Pons-Moll, 3DV, 2024

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

3DV submission #68

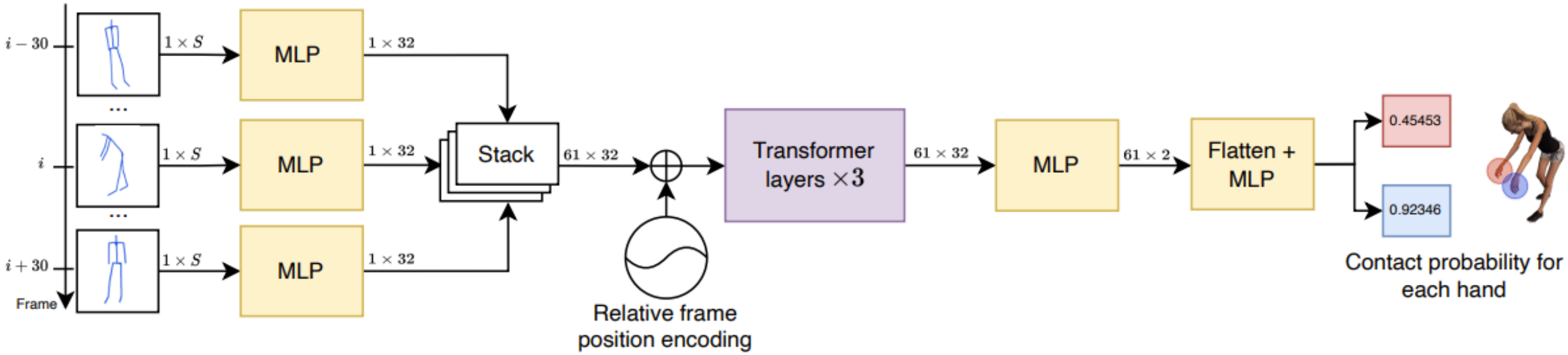
Includes Audio

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

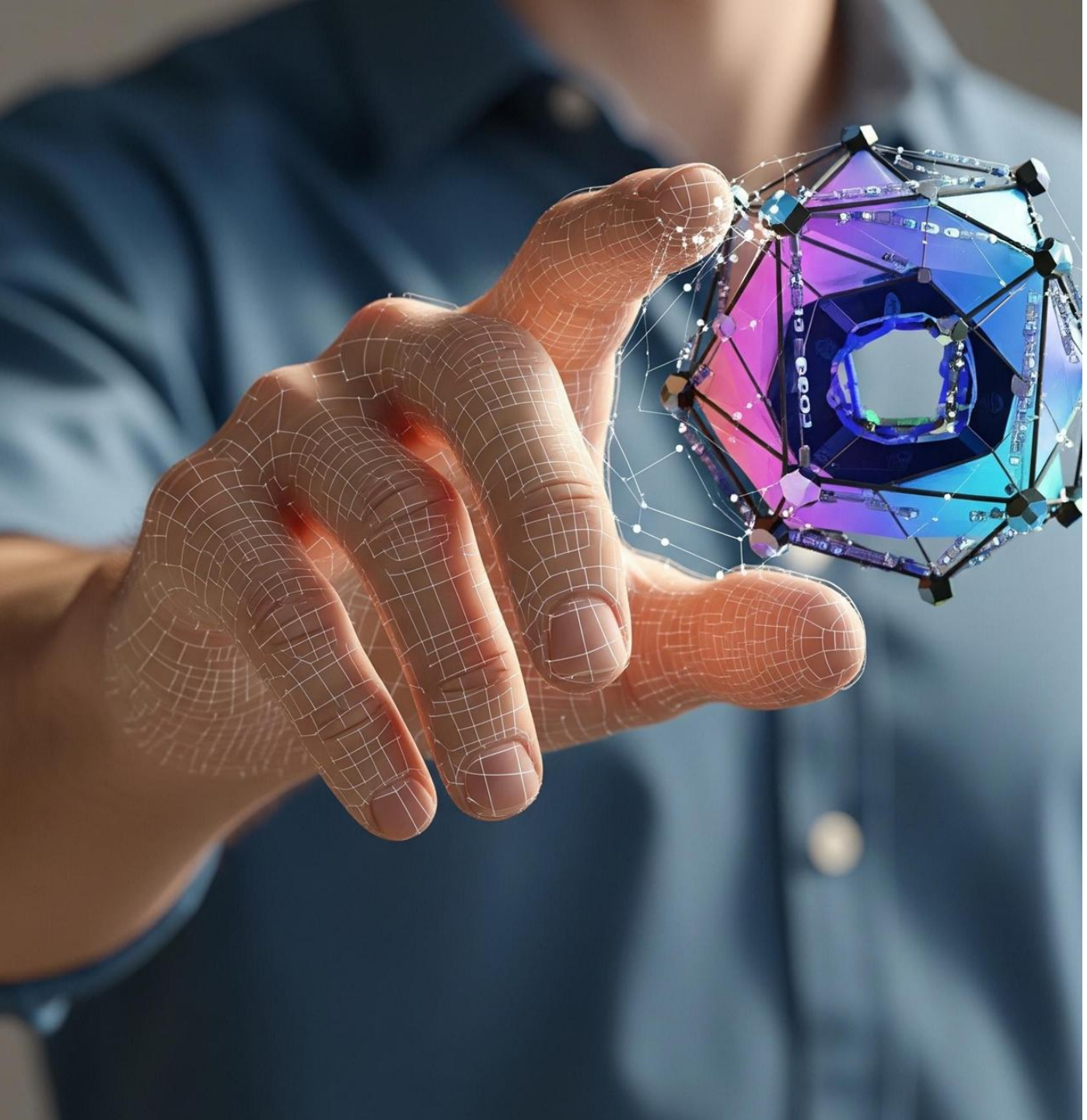
3DV submission #68

Includes Audio

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

3DV submission #68

Includes Audio

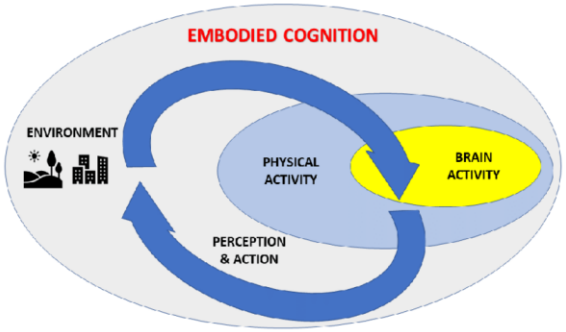


Conclusions

Does intelligence need a body?

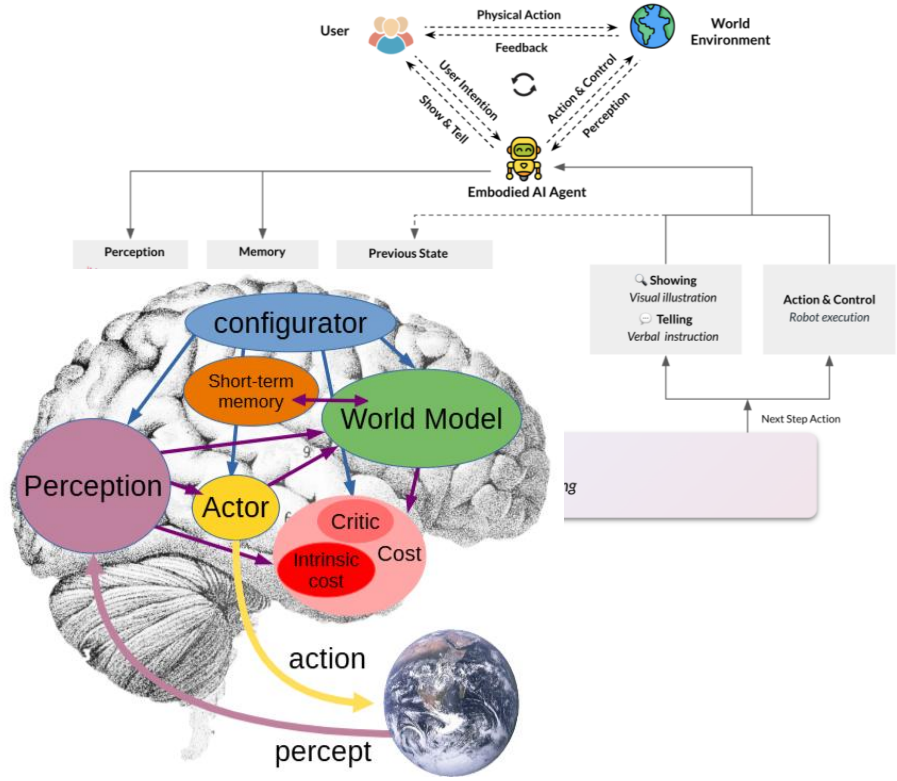
Descartes: No

https://pmc.ncbi.nlm.nih.gov/articles/PMC3512413/



Psycologists:
embodied cognition
(e.g., how movements relate to language and memory)

https://pubmed.ncbi.nlm.nih.gov/20739194/



CV\AI Researchers (LeCunn, Malik):
Embodied AI

https://openreview.net/pdf?id=BZ5a1r-kVsf
https://arxiv.org/pdf/2506.22355

Dr. Fei-Fei Li

From Words to Worlds: Spatial Intelligence is AI's Next Frontier

FEI-FEI LI

NOV 10, 2025

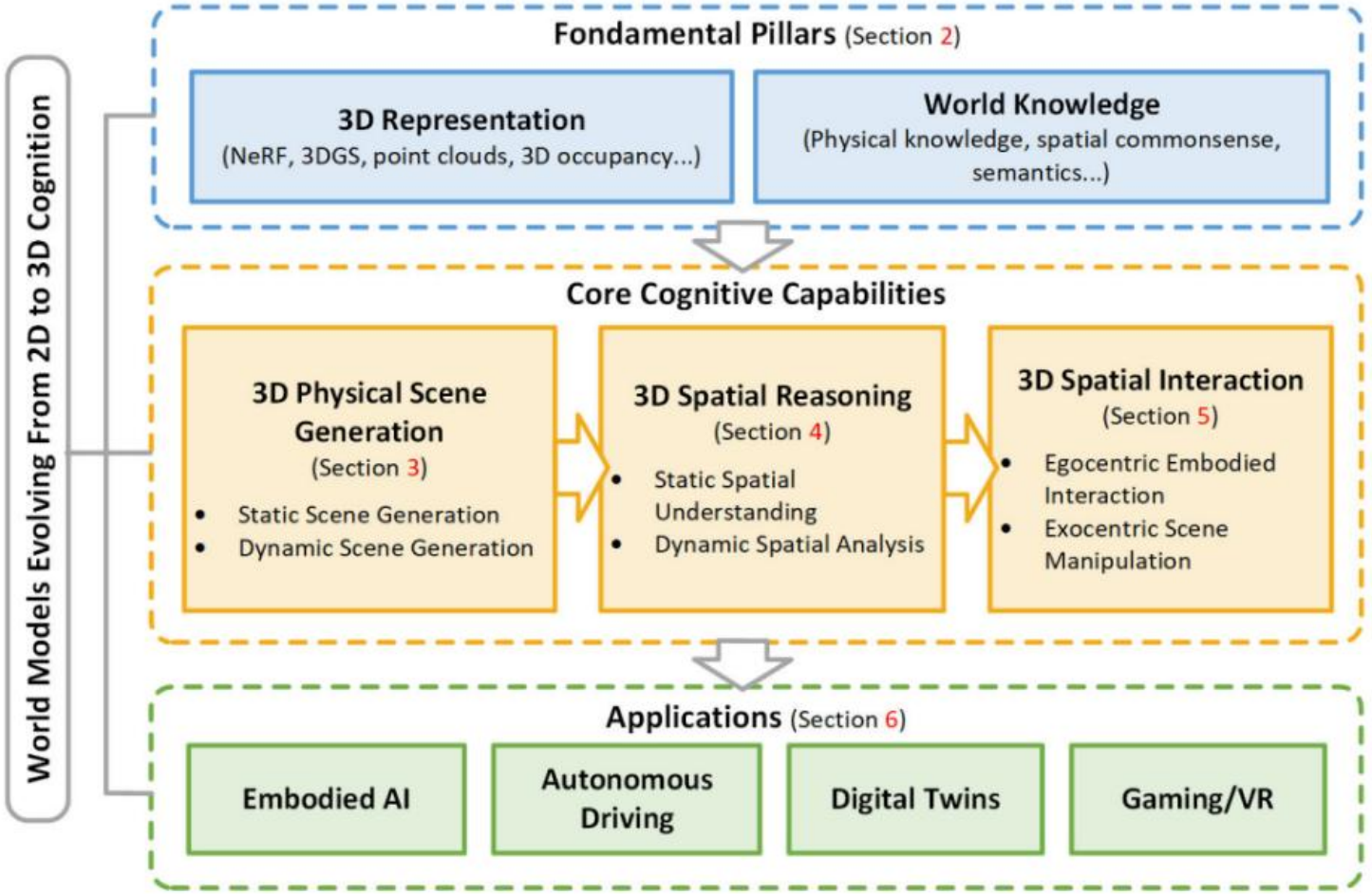
Tremendous progress has indeed been made in the past few years. Multimodal LLMs (MLLMs), trained with voluminous multimedia data in addition to textual data, have introduced some basics of spatial awareness, and today's AI can analyze pictures, answer questions about them, and generate hyperrealistic images and short videos. And through breakthroughs in sensors and haptics, our most advanced robots can begin to manipulate objects and tools in highly constrained environments.

Yet the candid truth is that AI's spatial capabilities remain far from human level. And the limits reveal themselves quickly. State-of-the-art MLLM models rarely perform better than chance on estimating distance, orientation, and size—or "mentally" rotating objects by regenerating them from new angles. They can't navigate mazes, recognize shortcuts, or predict basic physics. AI-generated videos—nascent and yes, very cool—often lose coherence after a few seconds.

Almost a half billion years after nature unleashed the first glimmers of spatial intelligence in the ancestral animals, we're lucky enough to find ourselves among the generation of technologists who may soon endow machines with the same capability—and privileged enough to harness those capabilities for the benefits of people everywhere. Our dreams of truly intelligent machines will not be complete without spatial intelligence.

SpatialAI

From 2D to 3D Cognition: A Brief Survey of General World Models, Xie et al., 2025



Sources and references:

- Learning Human Bodies in Motion
  https://bodymodelling.is.tuebingen.mpg.de/

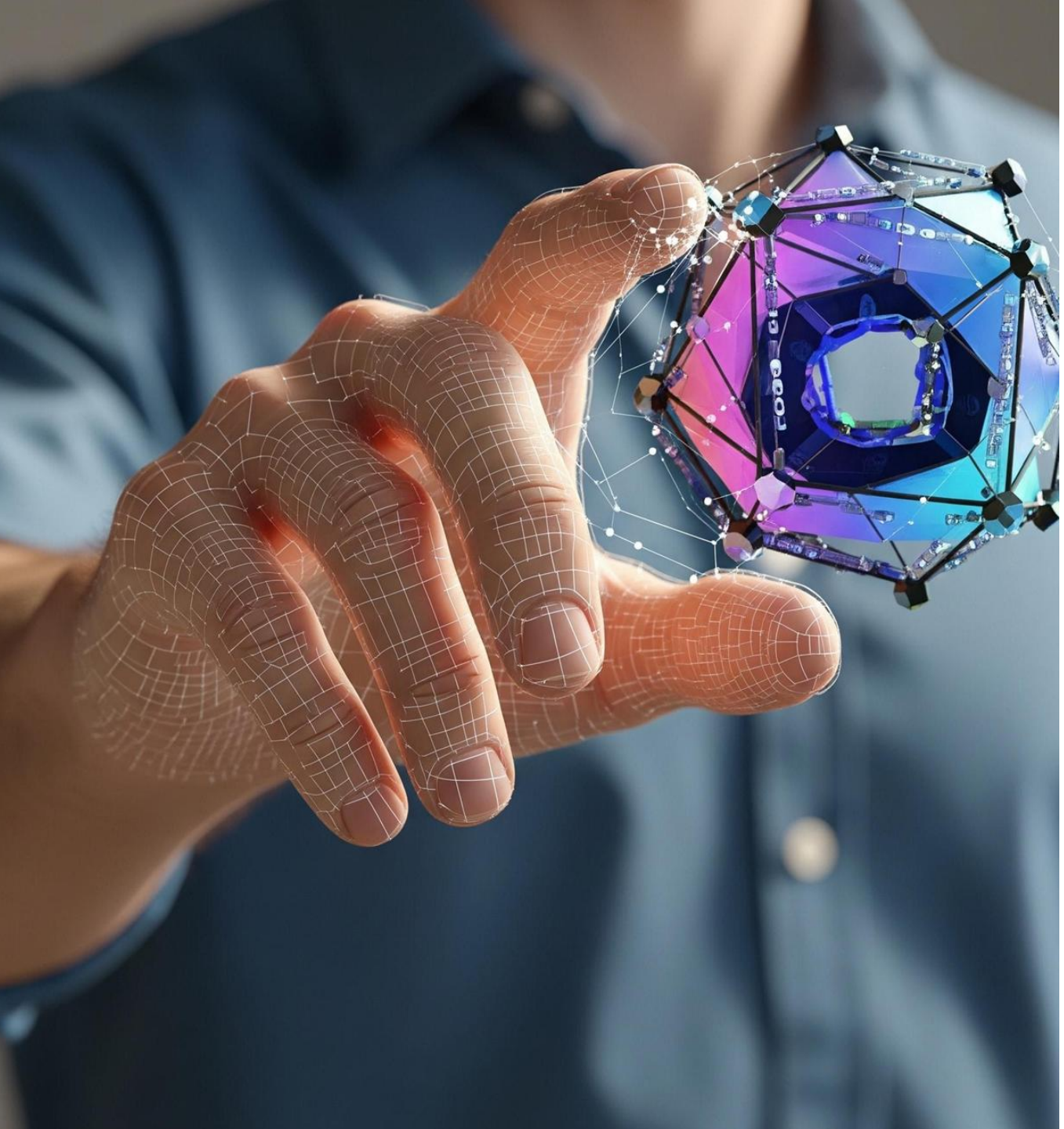
- FAUST Dataset and Challenge:
  https://faust-leaderboard.is.tuebingen.mpg.de/

- 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/



Virtual Humans
Under a Shape Analysis Spotlight

https://github.com/riccardomarin/HumanAnalysis

Riccardo Marin

26th November 2025
STAG26