Virtual Humans – Winter 23/24

Lecture 1_2 – Introduction to Human Models - Intro

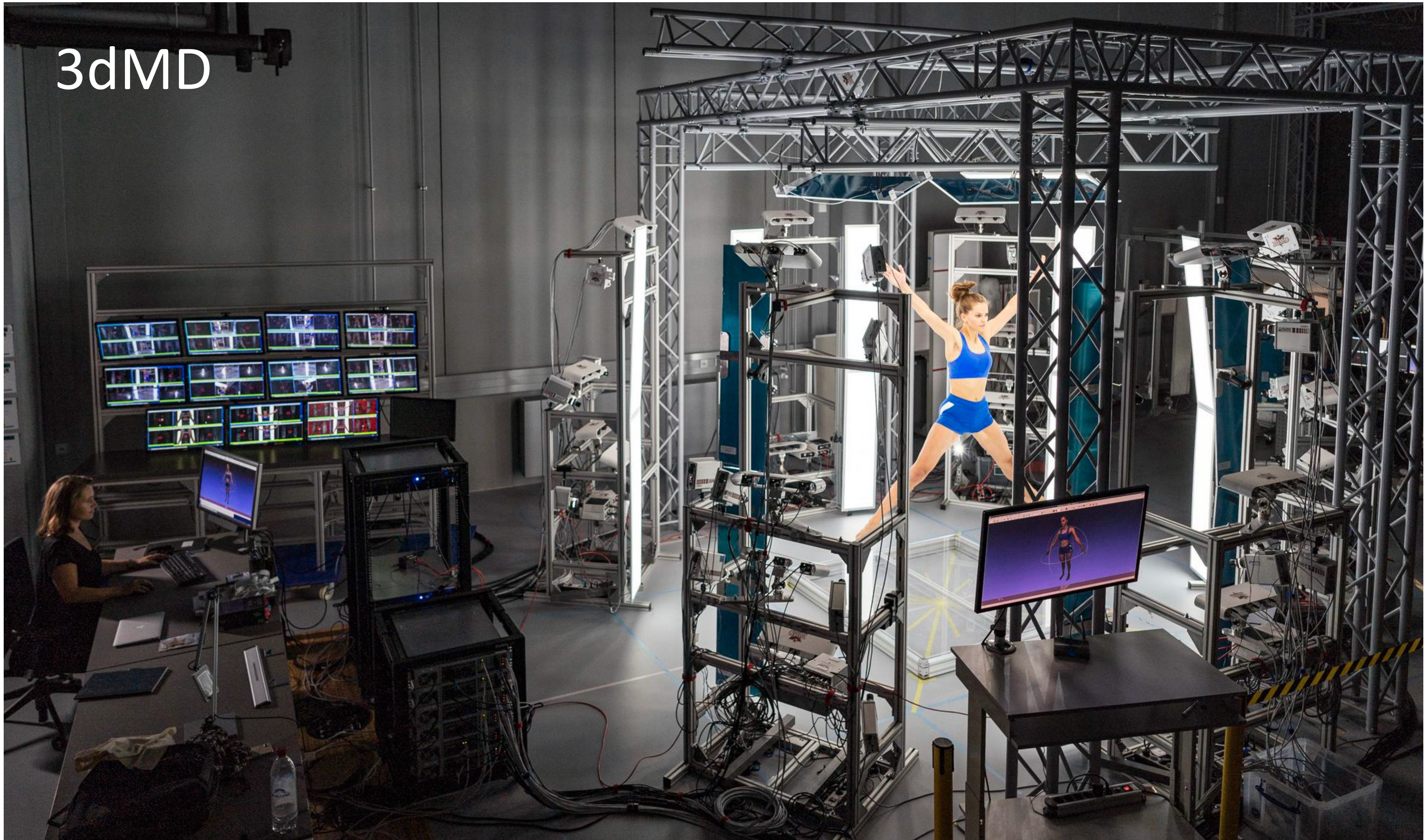
Prof. Dr.-Ing. Gerard Pons-Moll
University of Tübingen / MPI-Informatics

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Goal: Virtual humans

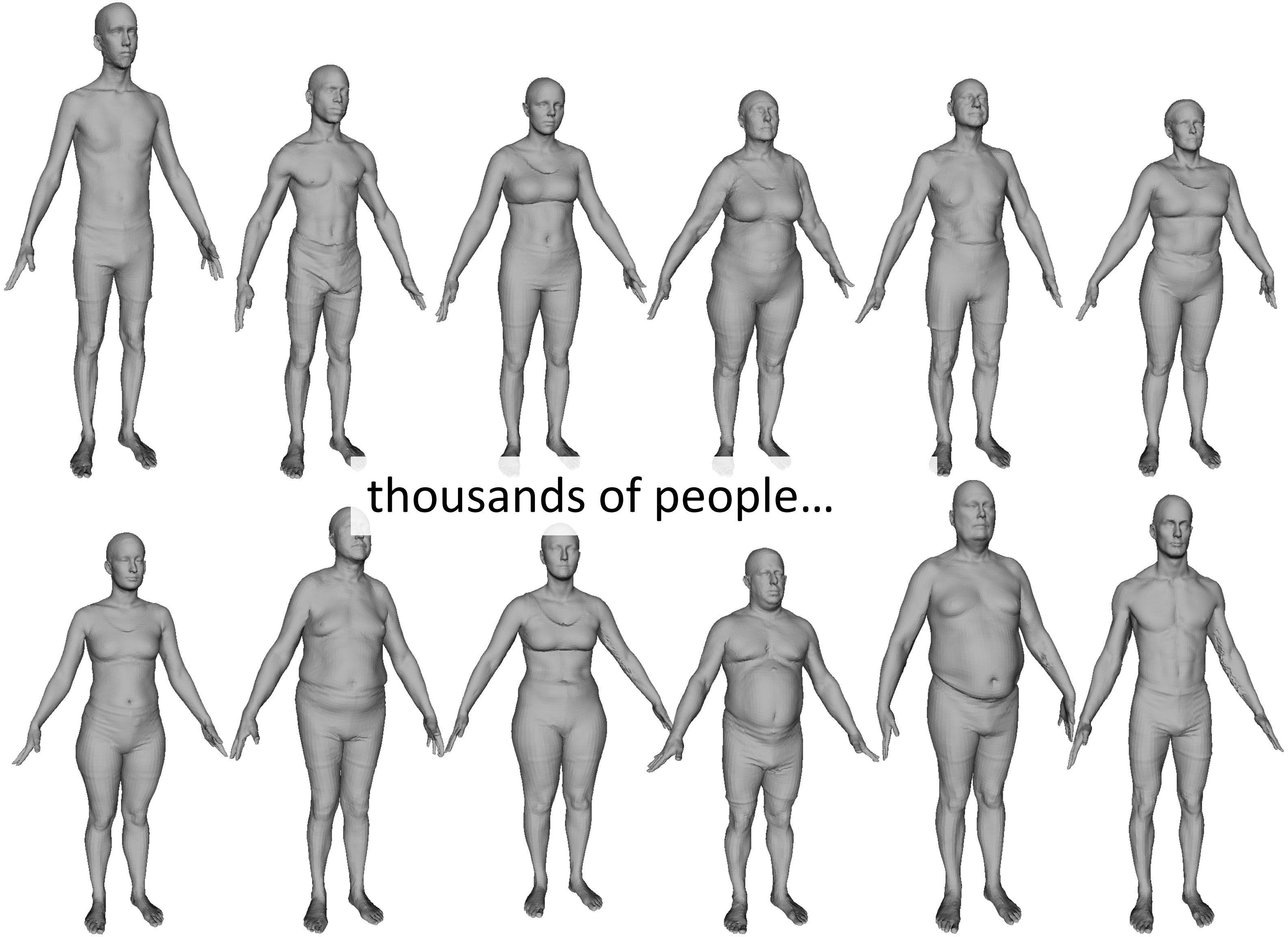
Define a simple **mathematical model** of body shape.
It should **look** like real people.
It should **move** like real people.
It should be low-D, differentiable, have joints, and be easy to animate and fit to data.
It should be **compatible** with standard graphics tools.

Capturing the body surface in 4D



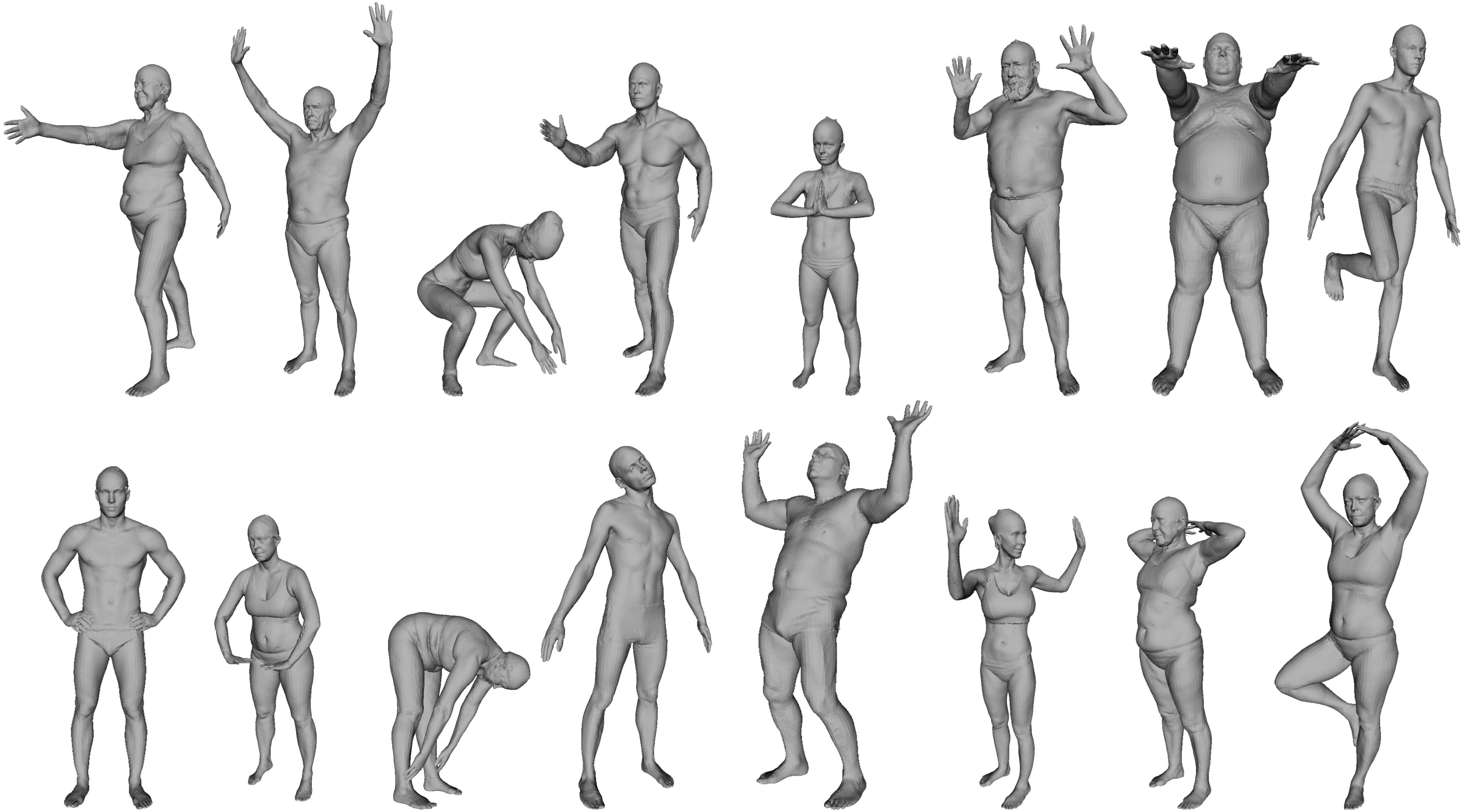
4D scanner: 3D at 60 fps

Collect 3D scans from

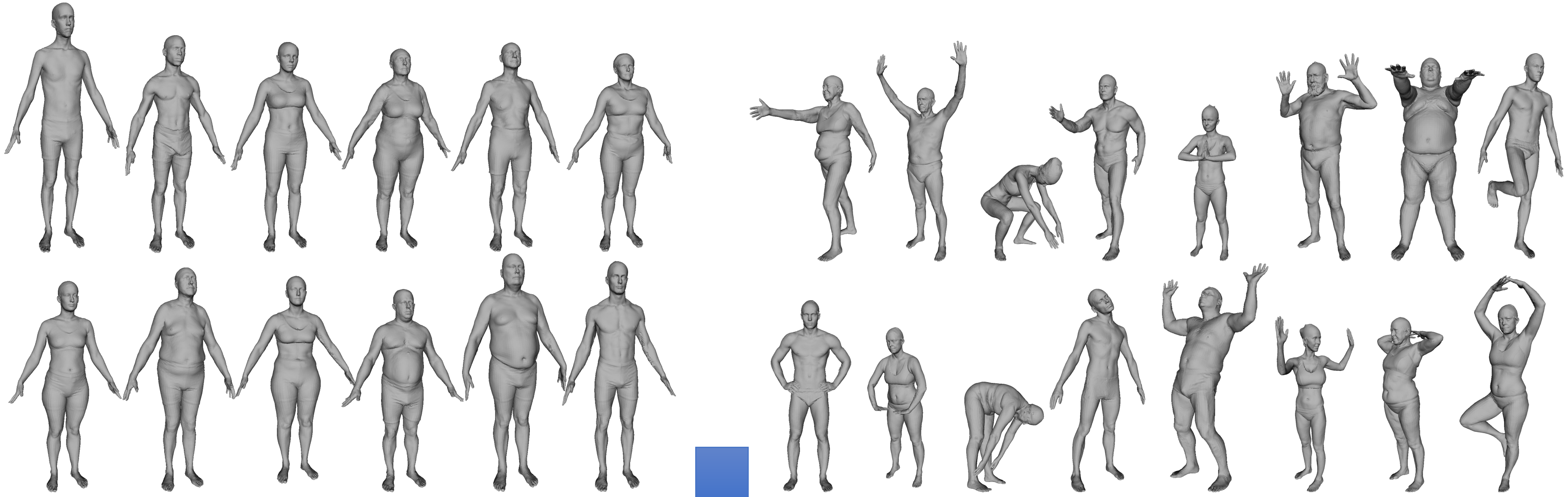


thousands of people…

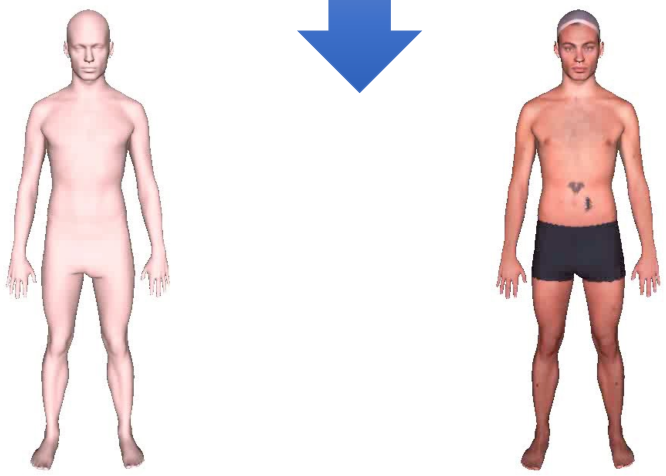
and thousands of poses



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



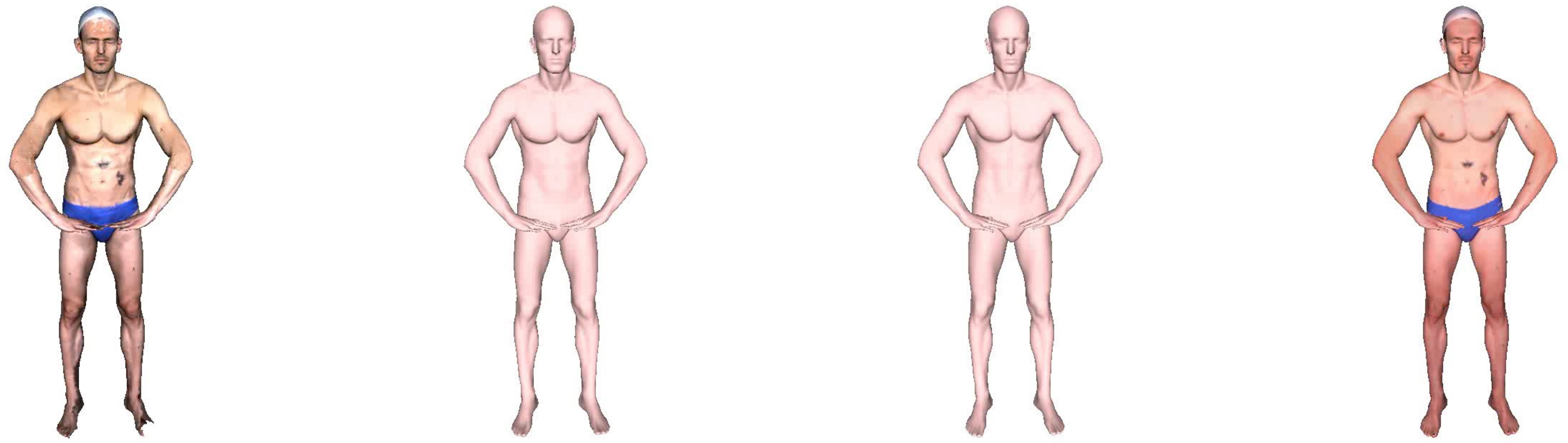
$$M(\theta, \beta, \delta, A)$$



A body model M takes a small number of pose, shape, and other parameters and returns a 3D mesh.

What is a body model?

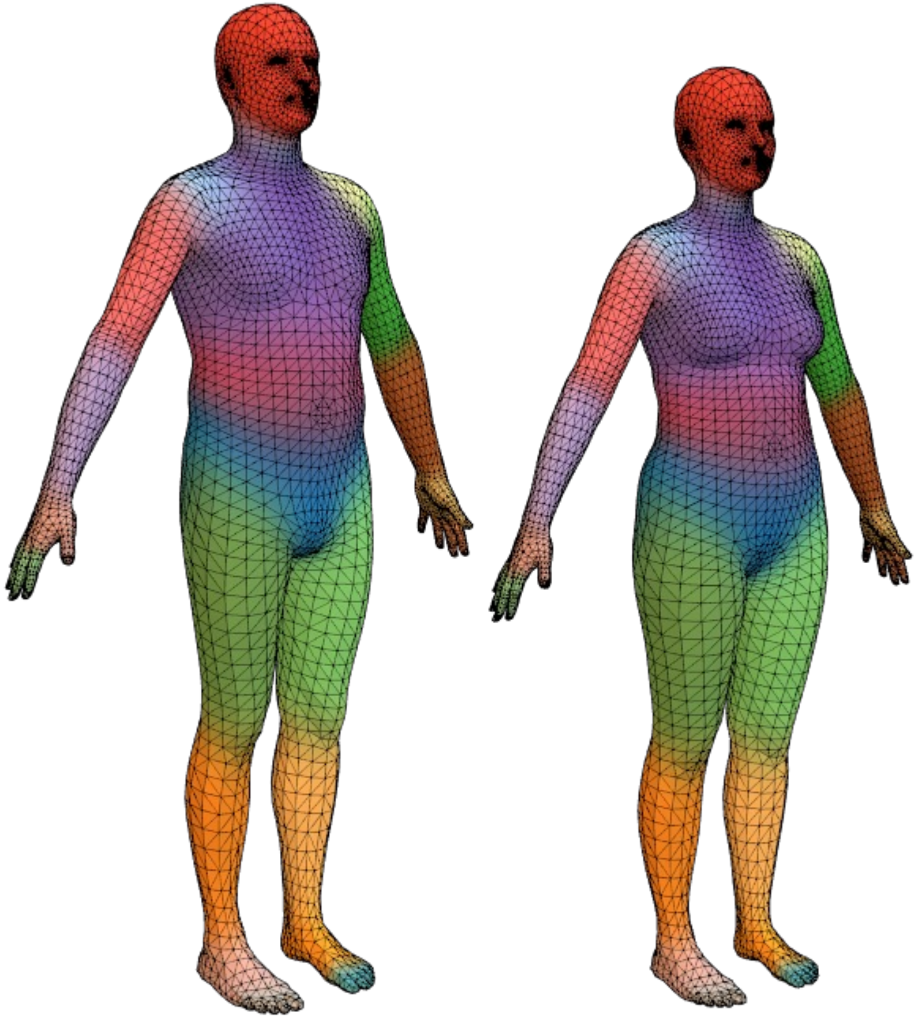
3D scan with texture | **Ground truth shape** | **Model** | **Model** with texture



$$\underset{\text{model}}{M}(\overset{\text{pose}}{\theta}, \overset{\text{shape}}{\beta}, \overset{\text{dynamics}}{\delta}, \overset{\text{texture}}{A})$$

A model takes a small number of pose, shape, and texture parameters and returns a 3D mesh.

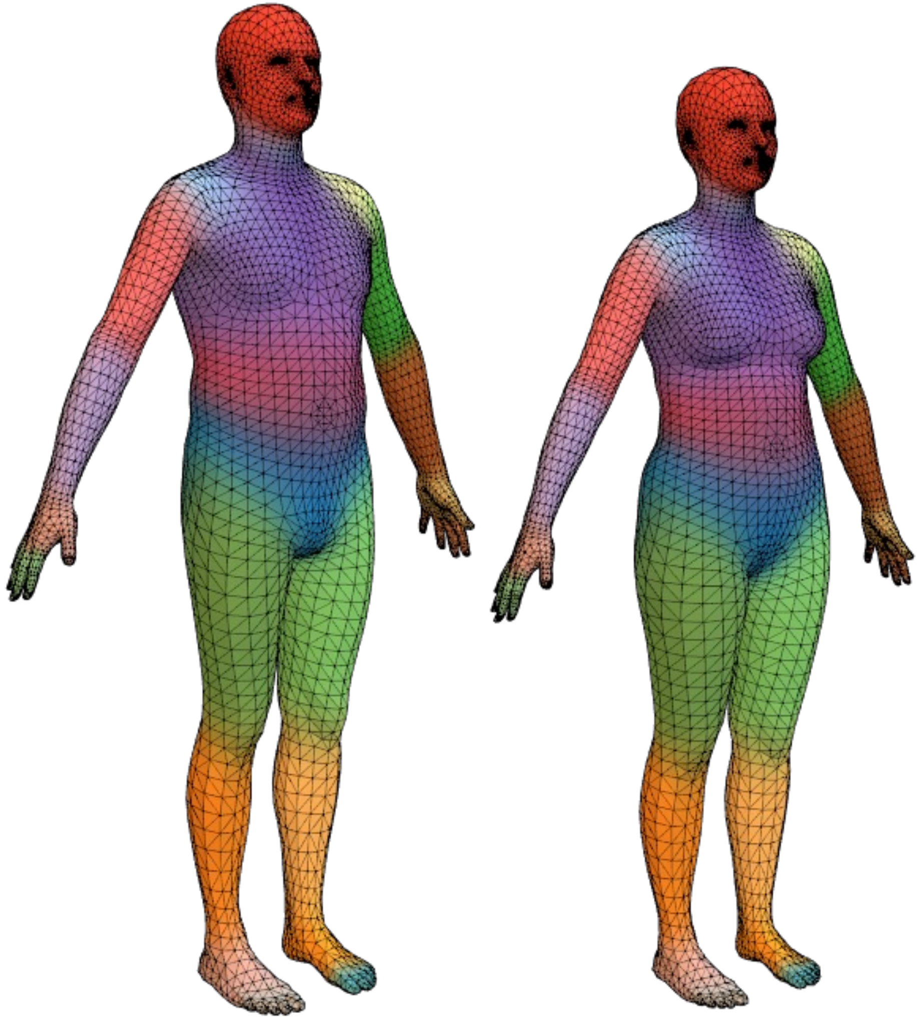
Modeling bodies



We represent the whole body as a 3D mesh.

Mesh is segmented into parts that can rotate in a kinematic tree.

Modeling bodies

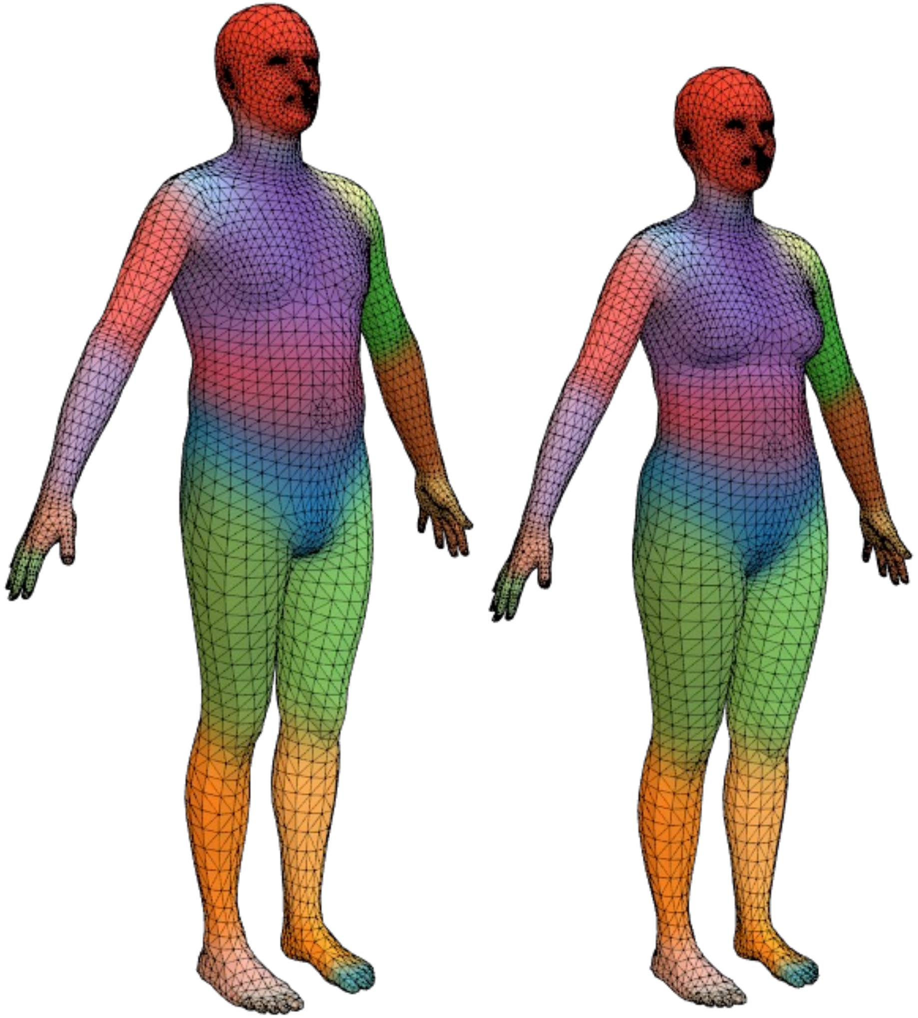


SMPL represents the body with about 7,000 points (vertices) in 3D.

So body is then just 21,000 numbers.

Problem: Most settings of these numbers don't correspond to people or anything else.

Modeling bodies



Goal:

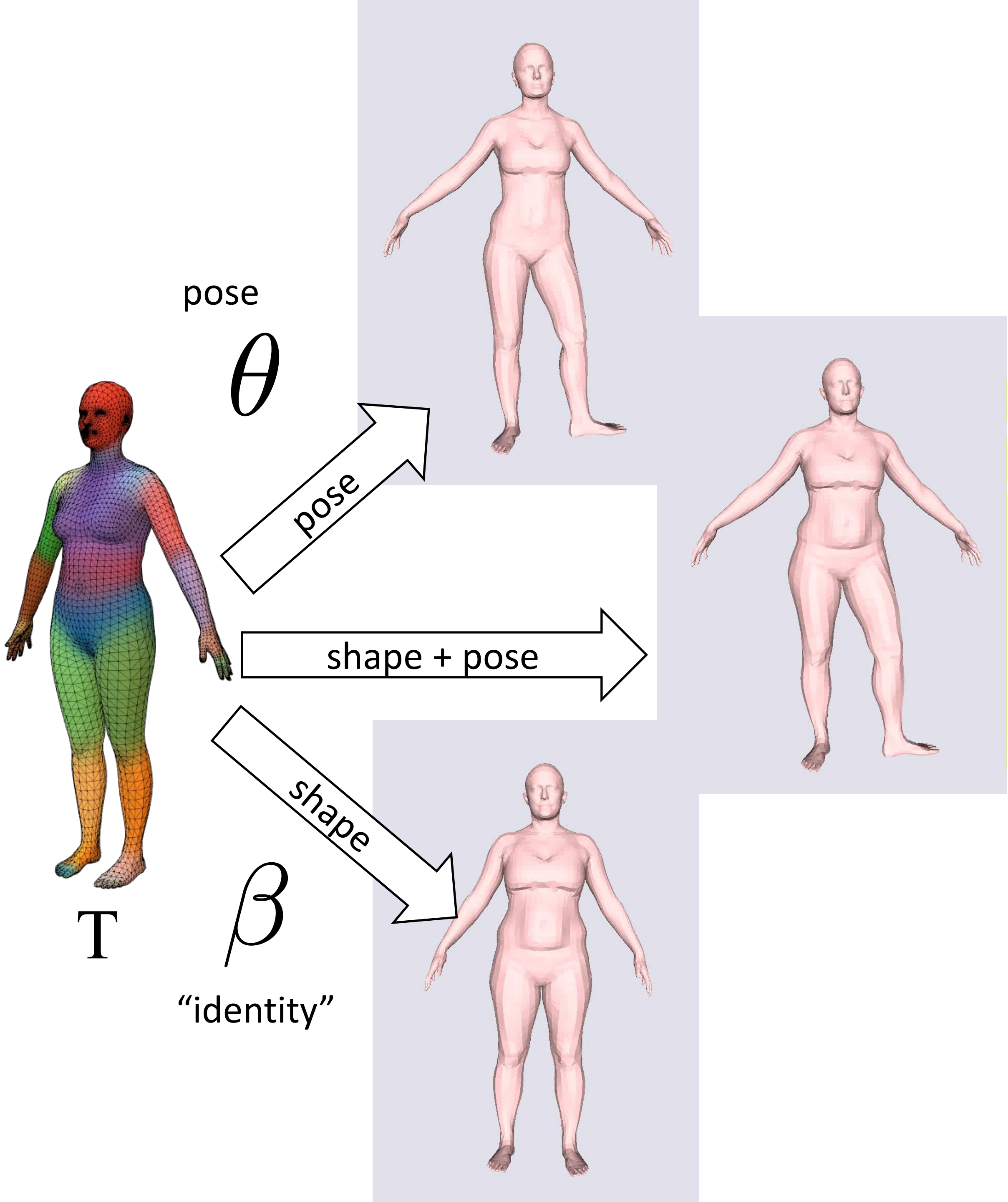
Characterize which settings of 21,000 numbers correspond to real people.

For this we define a mathematical “function” that outputs bodies.

Factored model

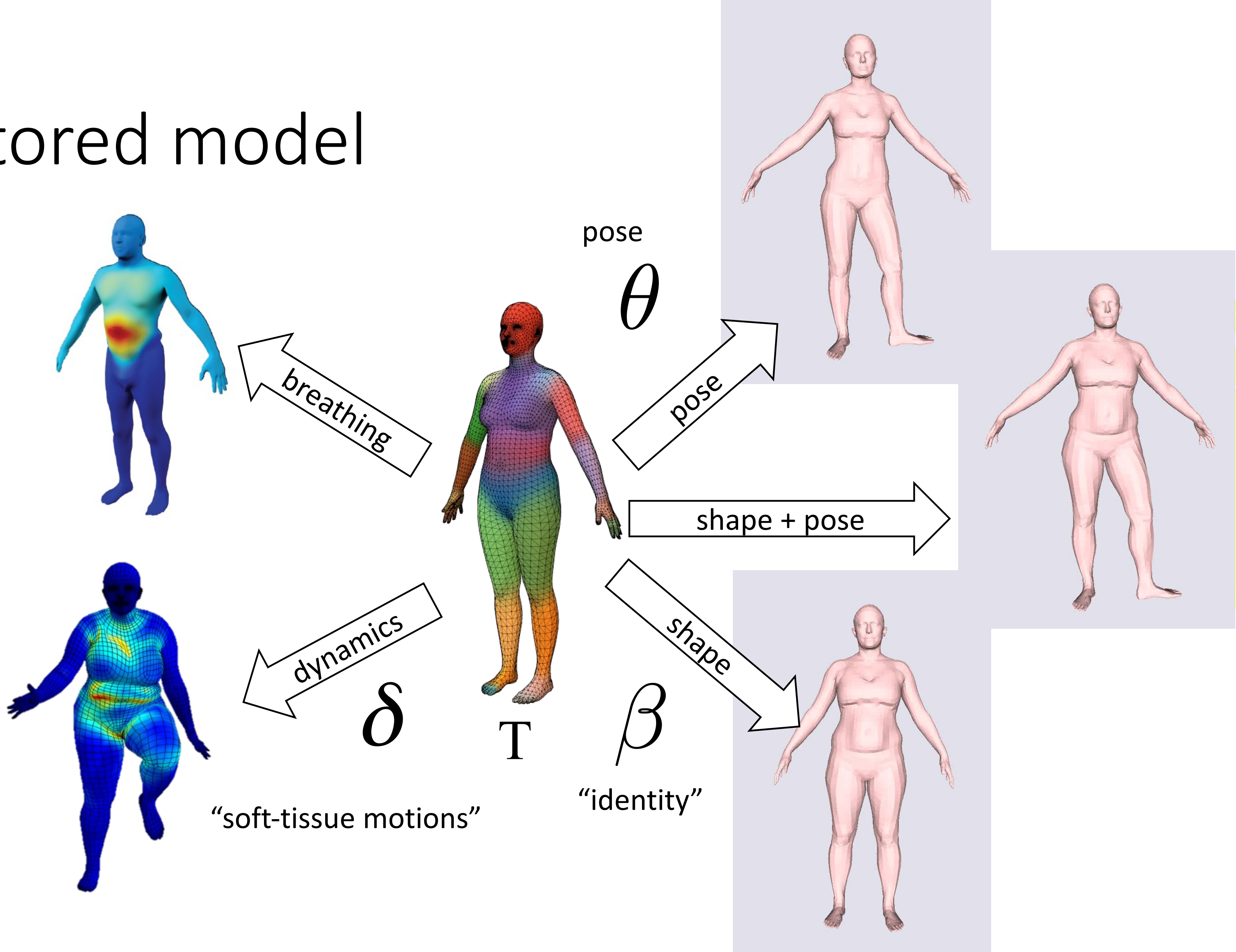
A model parameterizes "deviations" from a template mesh.

Simplifies learning and inference.

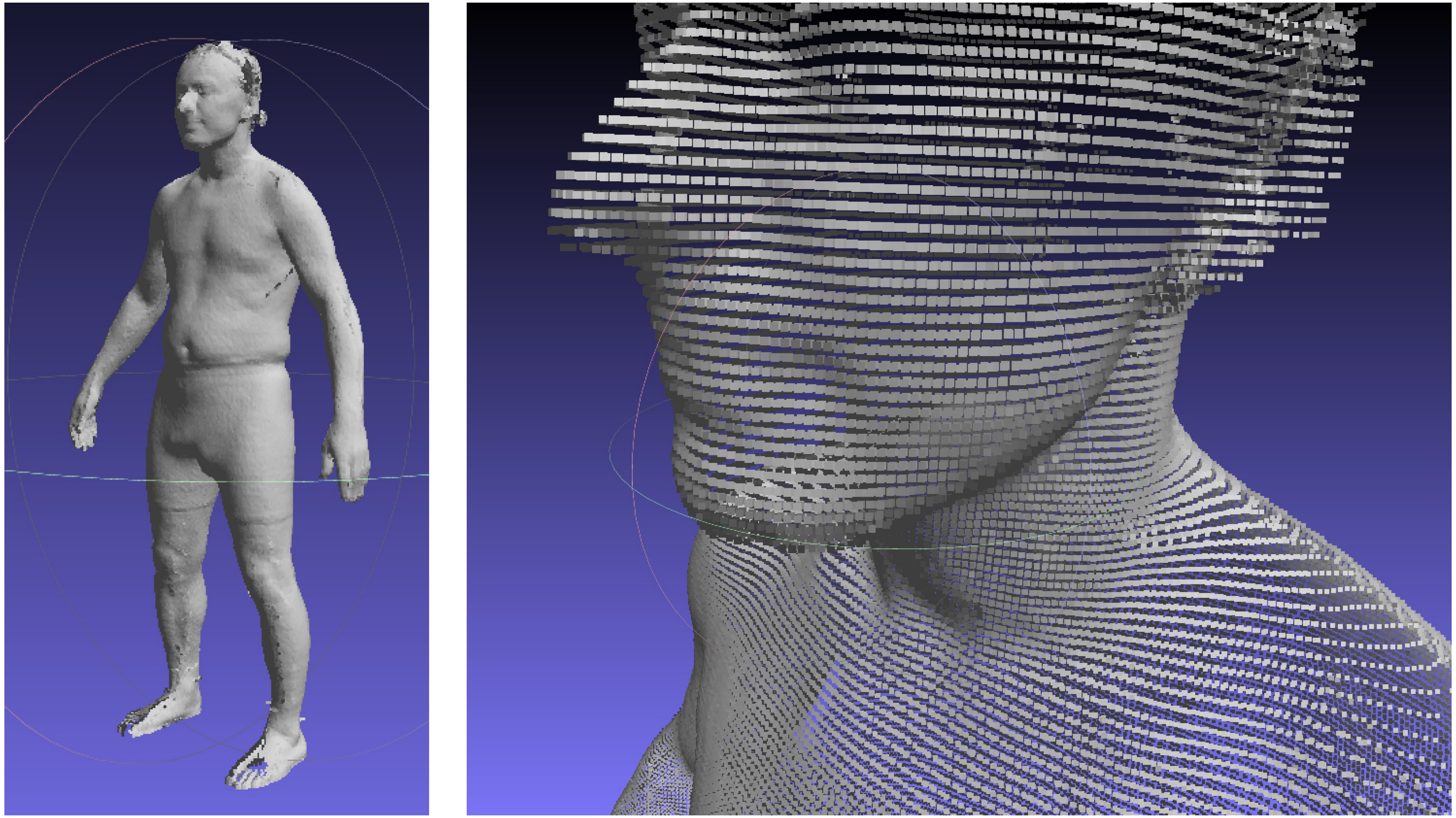


SCAPE: Anguelov et al., SIGGRAPH 2005

Factored model



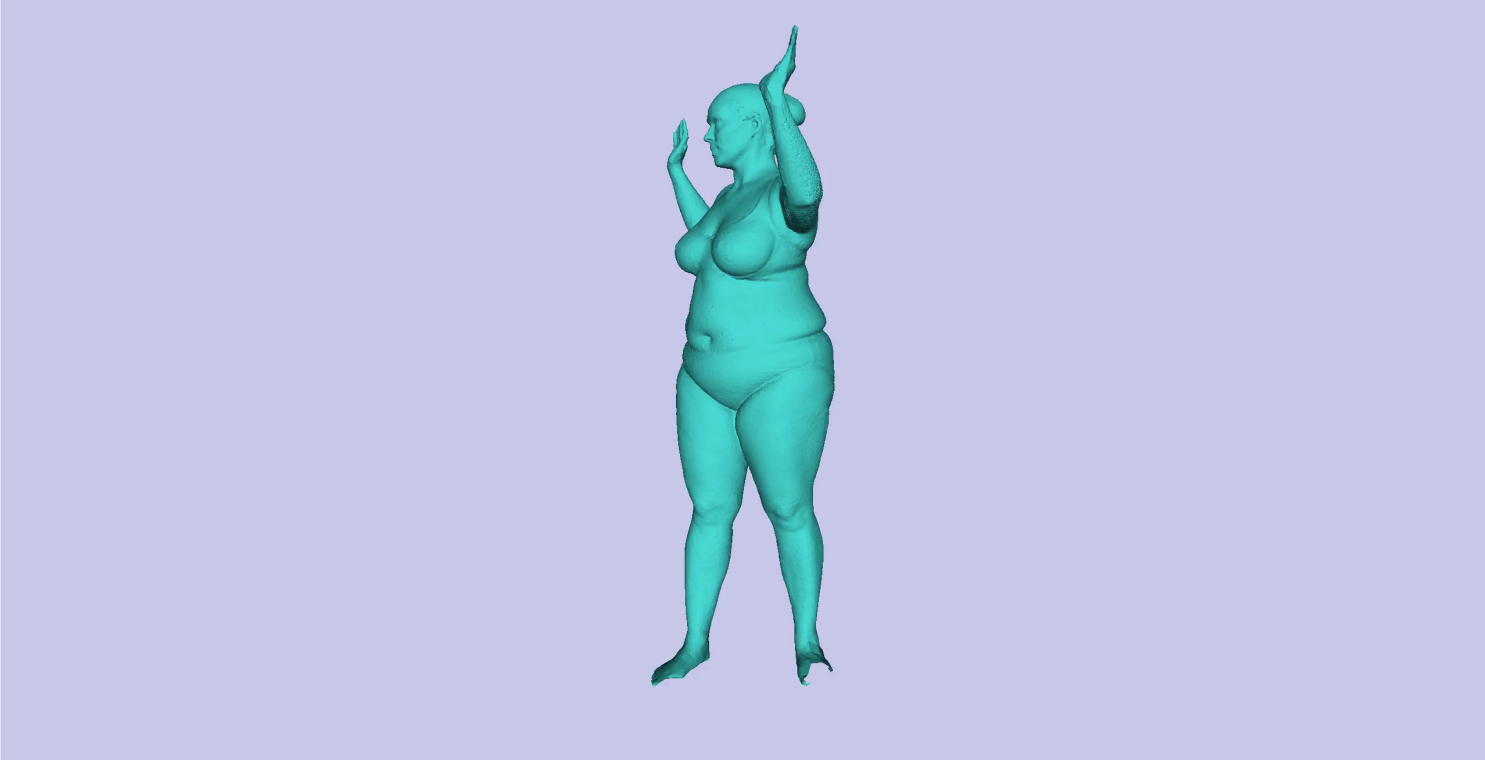
Problem: Scans are not bodies



Point clouds – collections of unordered 3D points.

Co-registration

A common template T is brought into alignment with each scan S_i



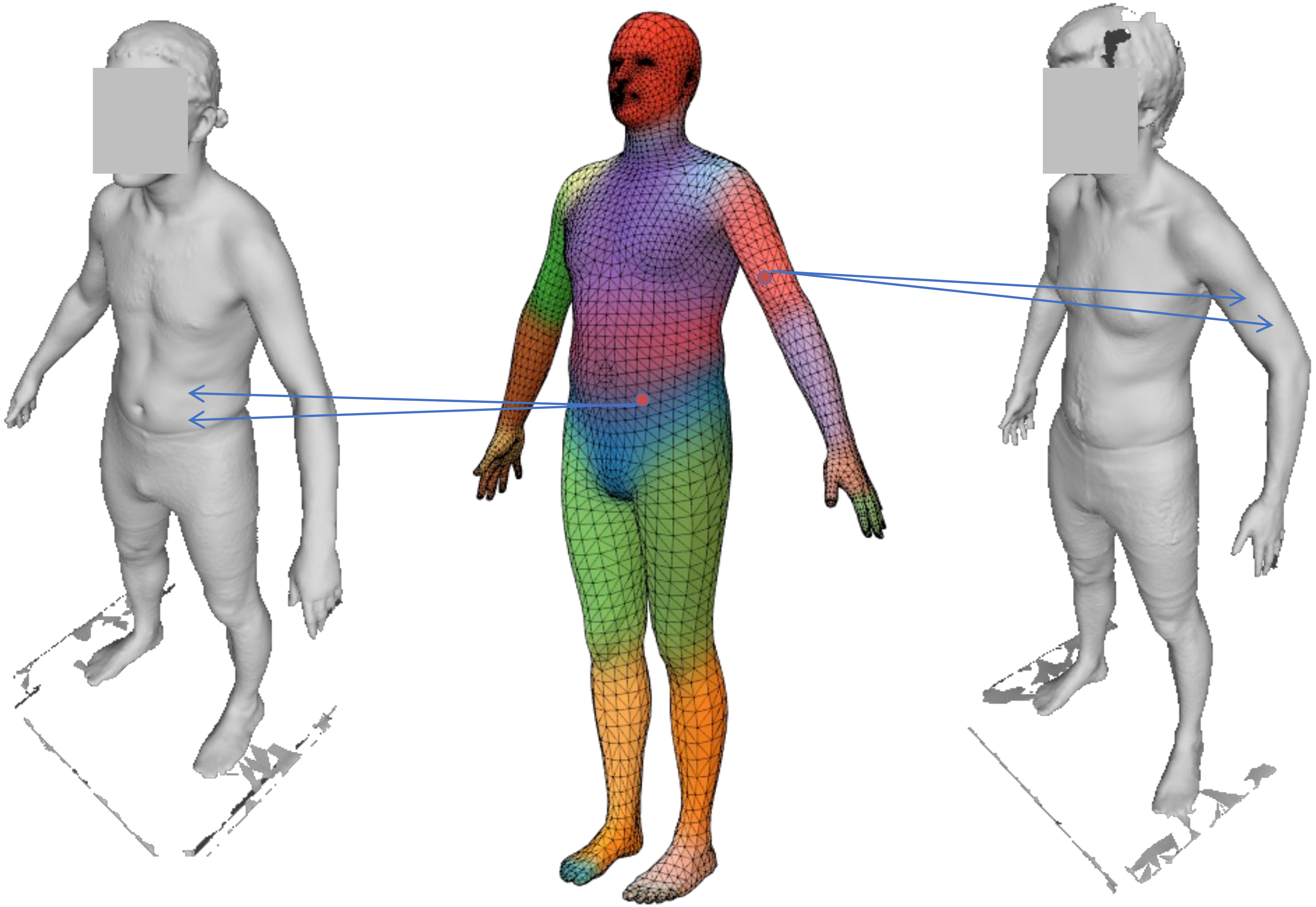
Register **template** to **scan**, while *regularizing to the model*.

Build the model at the same time.

The better the model, the better the registration, and vice versa.

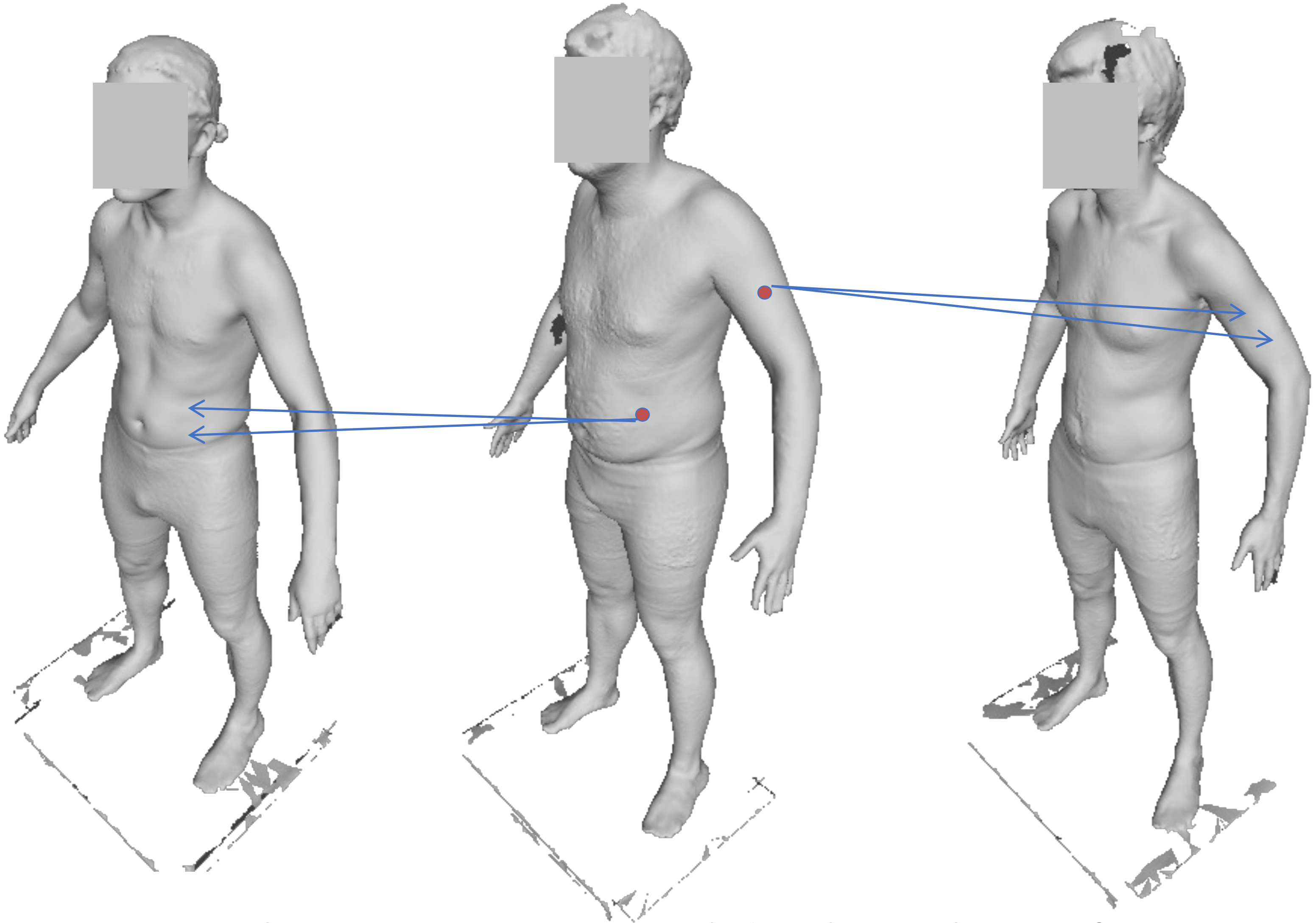
Hirshberg, et al., Coregistration: Simultaneous alignment and modeling of articulated 3D shape, ECCV, 2012.

Why is it hard? Resolution.



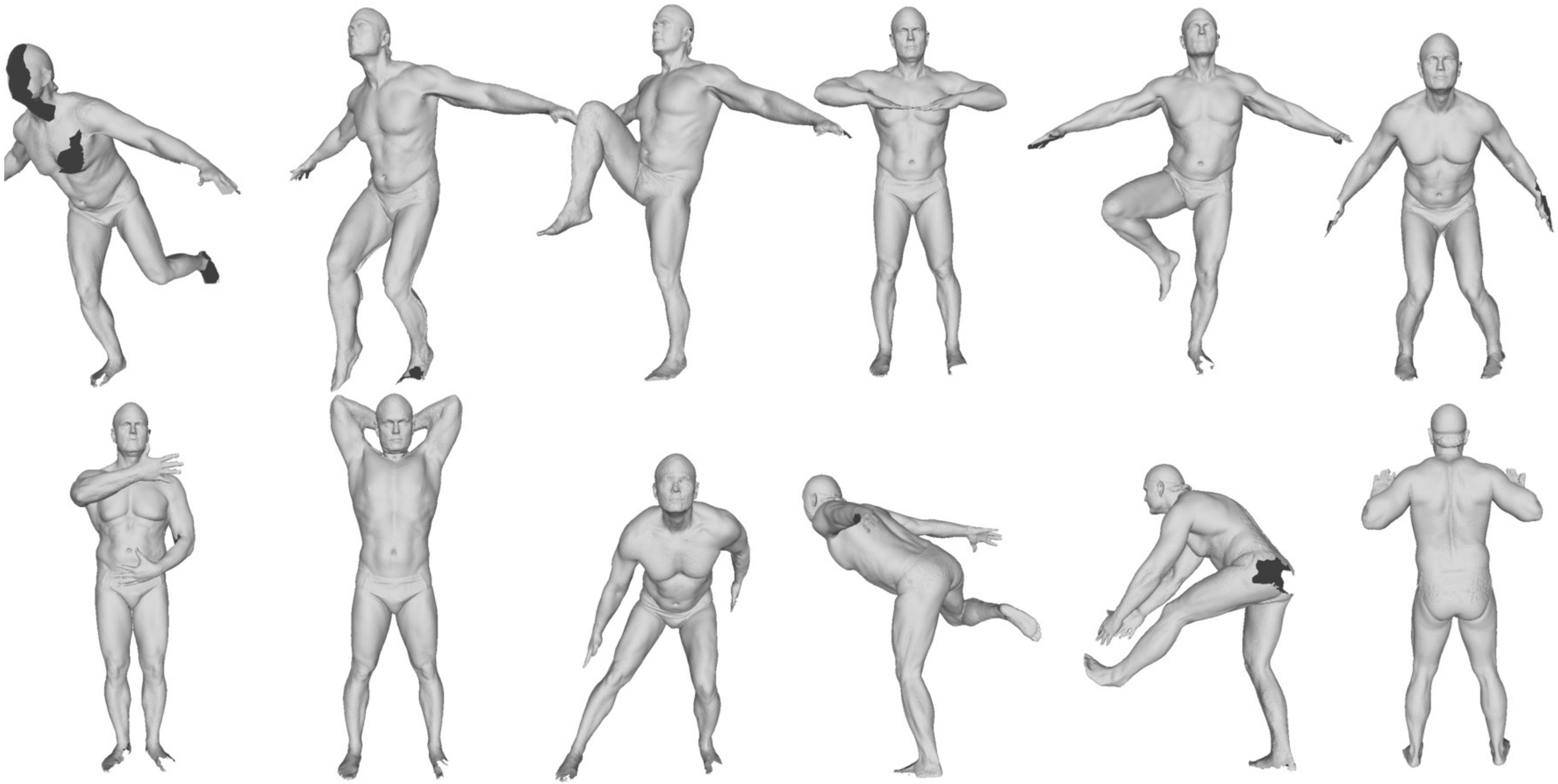
Low-resolution template but high-resolution scans.

Why is it hard? Ambiguous.



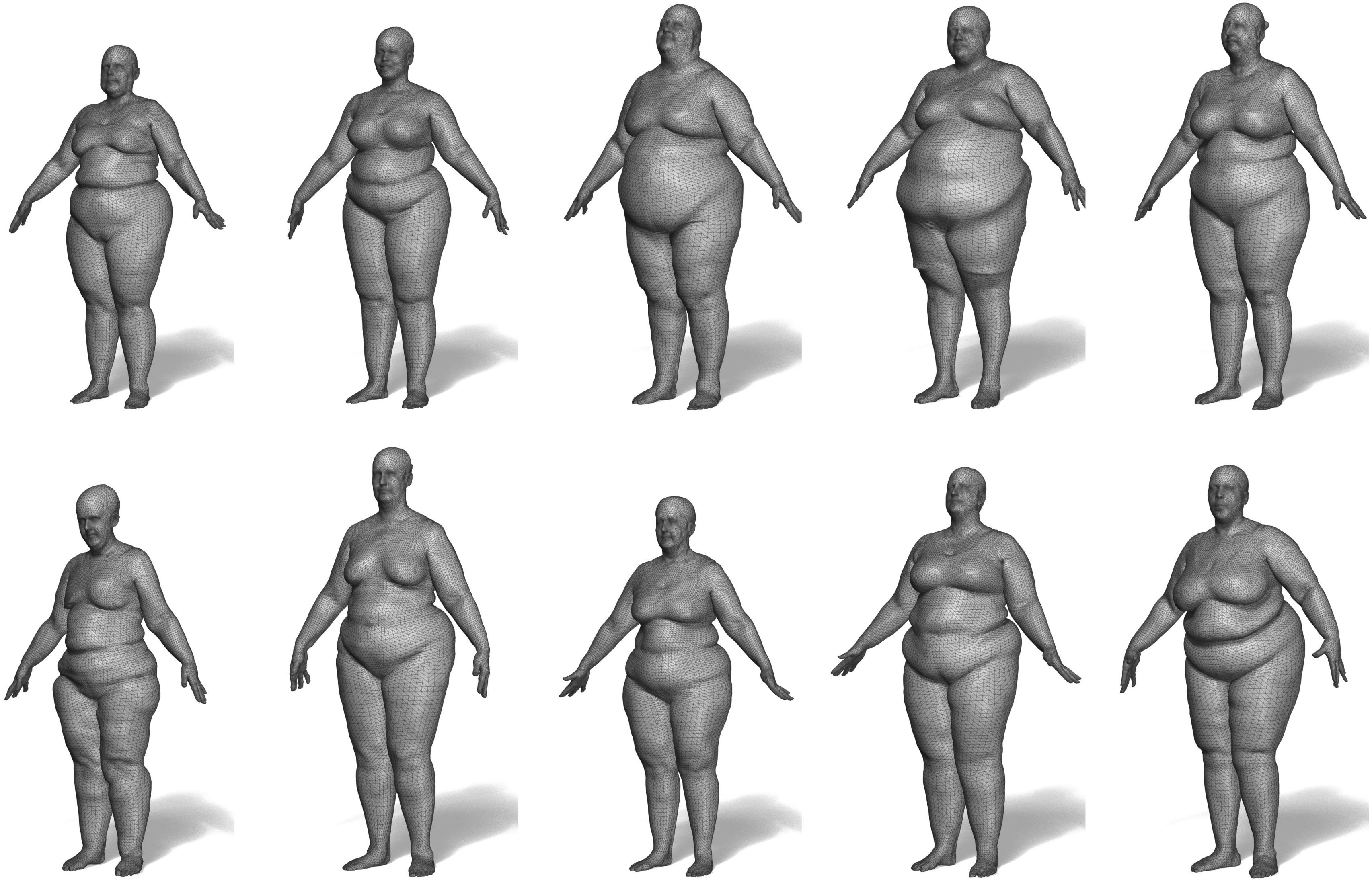
In smooth areas, points can slide along the surface.

Why is it hard? Pose, contacts, missing data

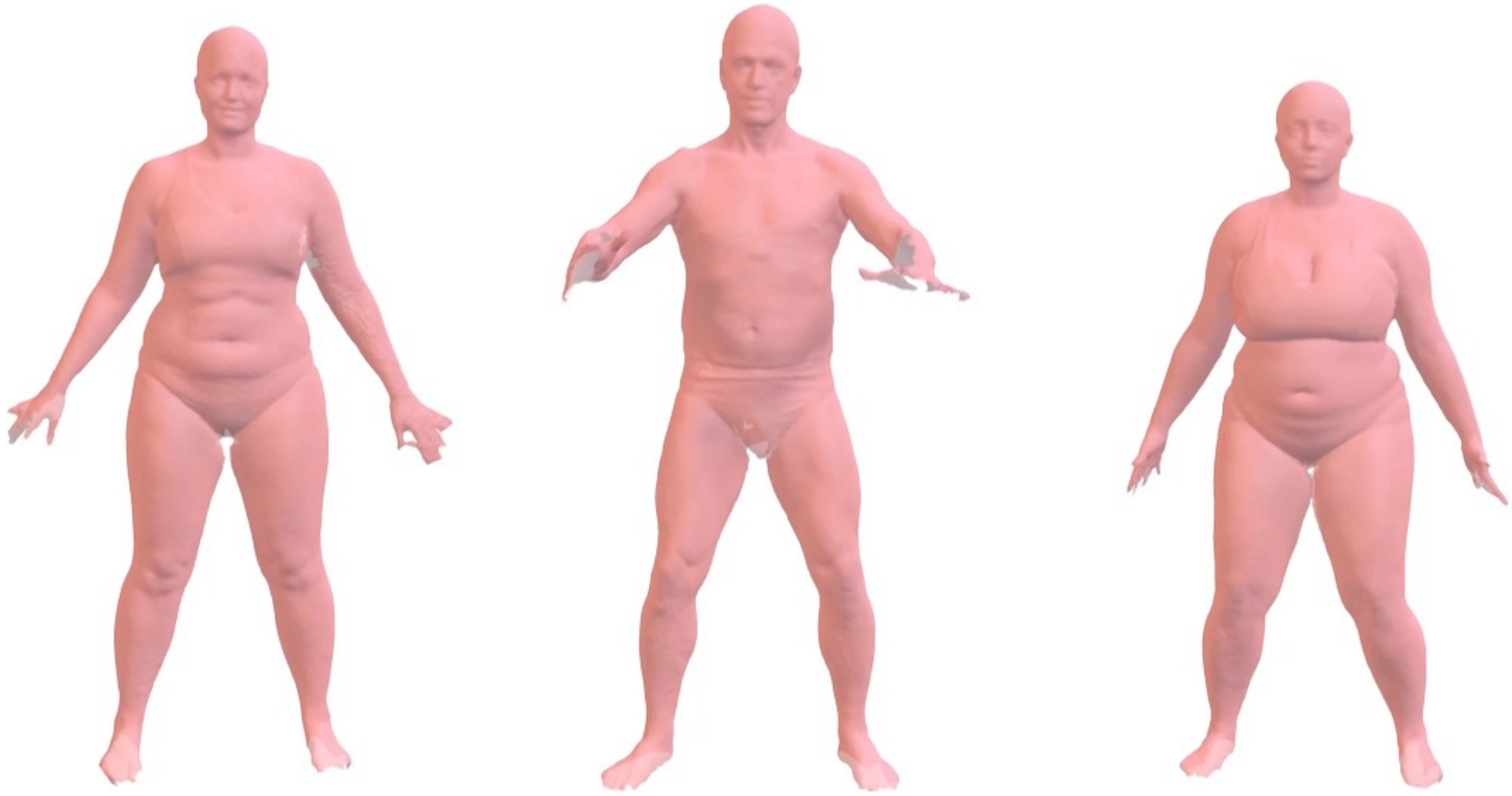


Shape change, self contact, missing data.

Example CAESAR registrations

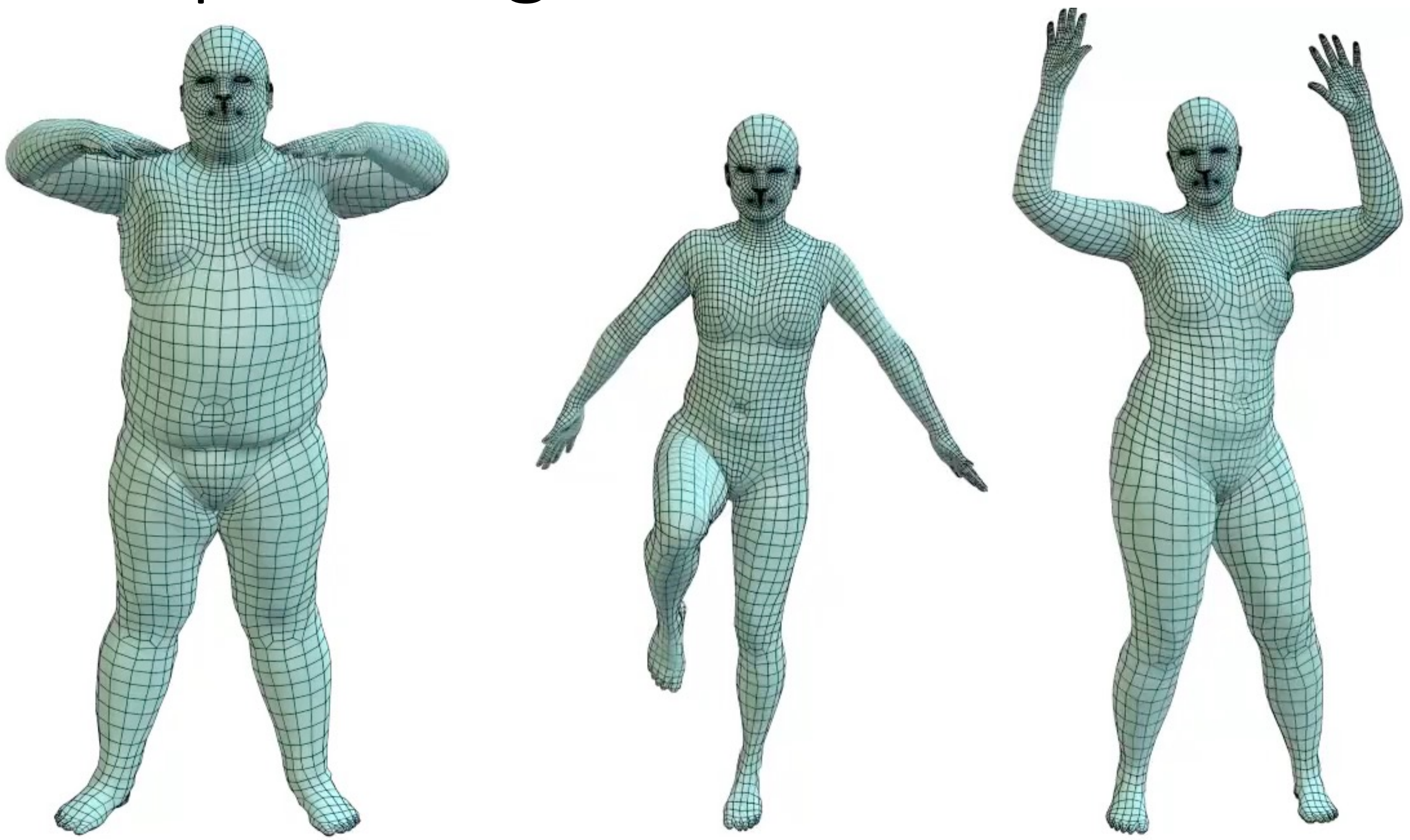


Raw 4D scans (60 fps)



Just point clouds

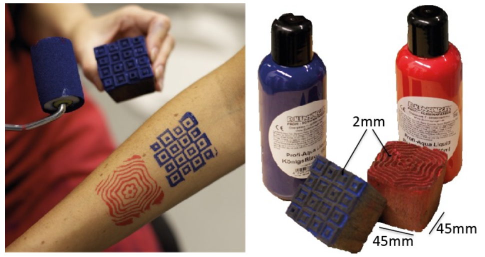
4Cap: Temporal registrations



Dyna dataset. Pons-Moll et al. SIGGRAPH 2015

Is our registration any good?

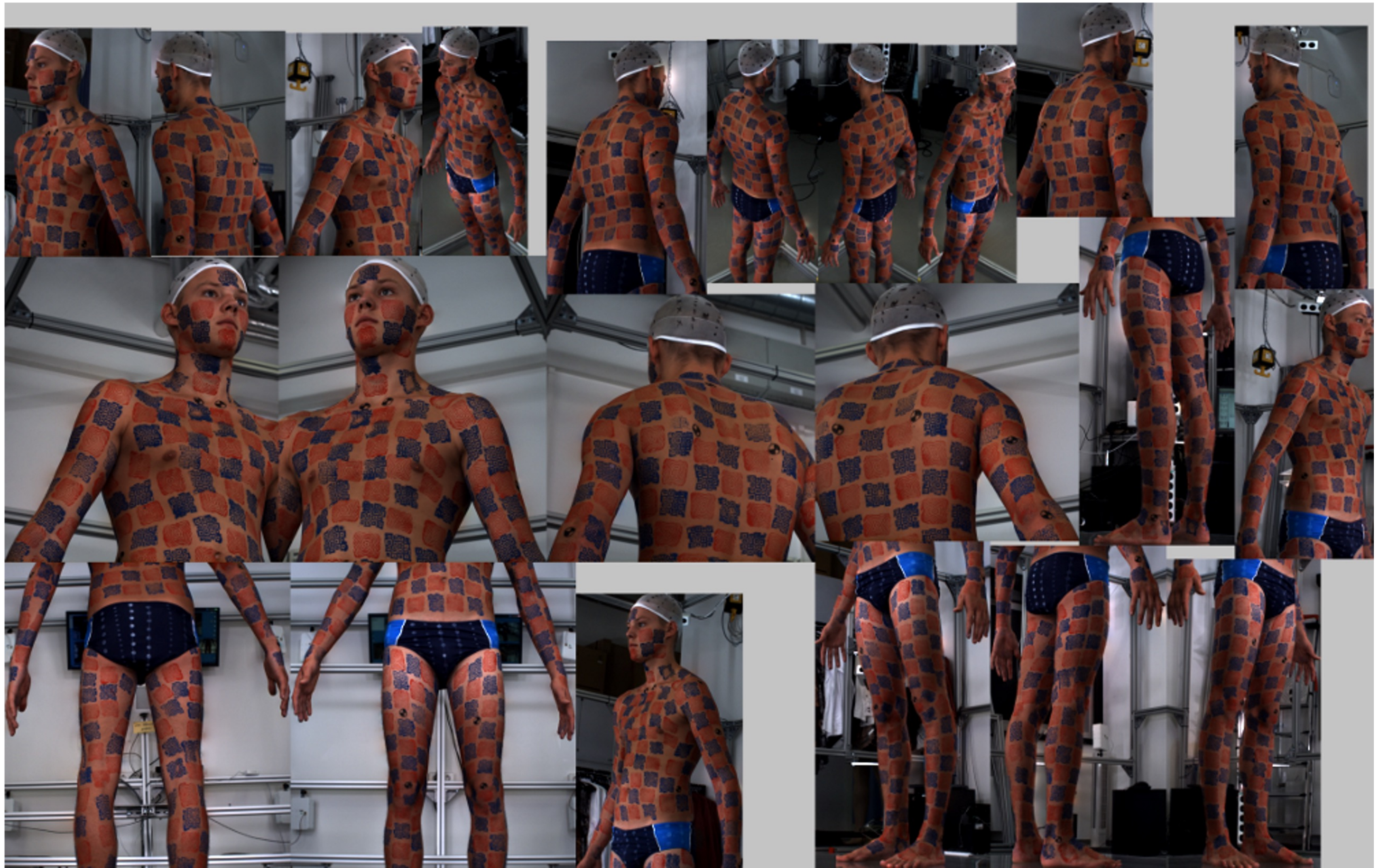
- *dense* intra-subject correspondences



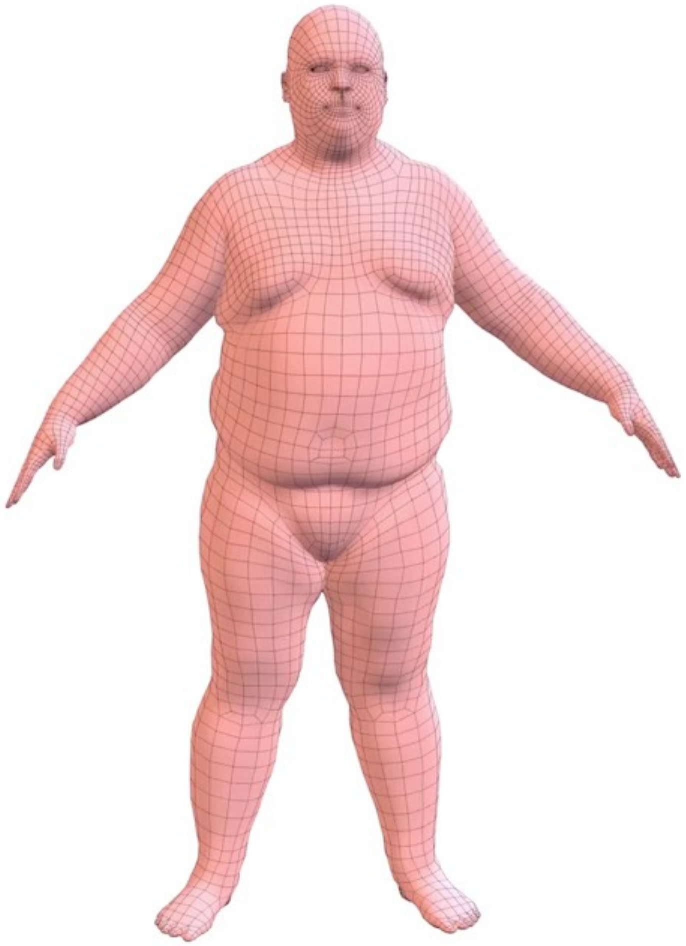
- *sparse* inter-subject correspondences

High-frequency body paint

Scan texture captured by 22 RGB cameras

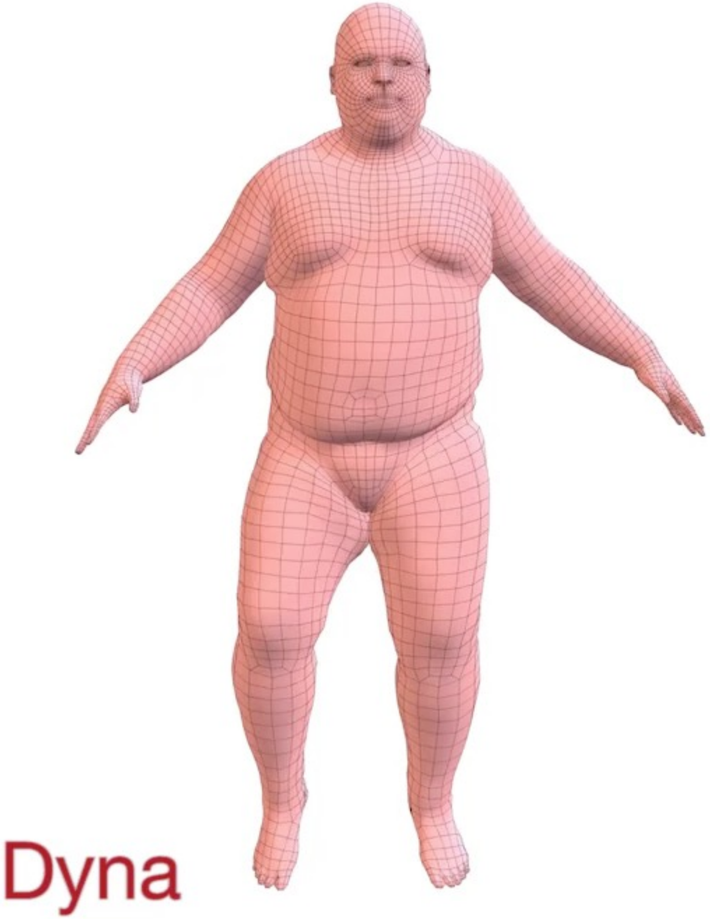


Problem: Geometry is not sufficient



Bogo et al, CVPR 2017

Solution: Geometry + Texture



Bogo et al, CVPR 2017

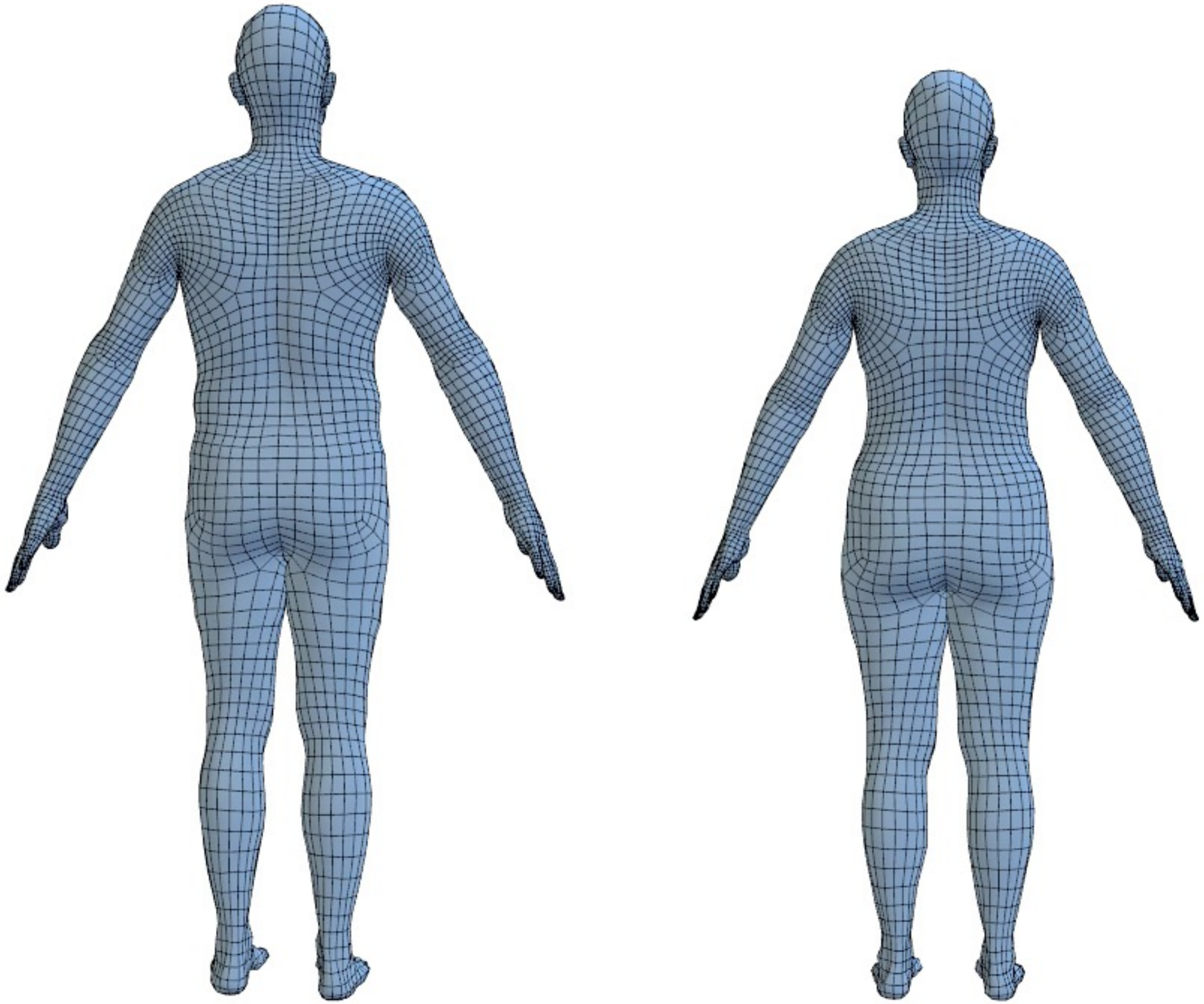
Shape training set

~2000 meshes per gender (CAESAR dataset)

In correspondence with a template mesh

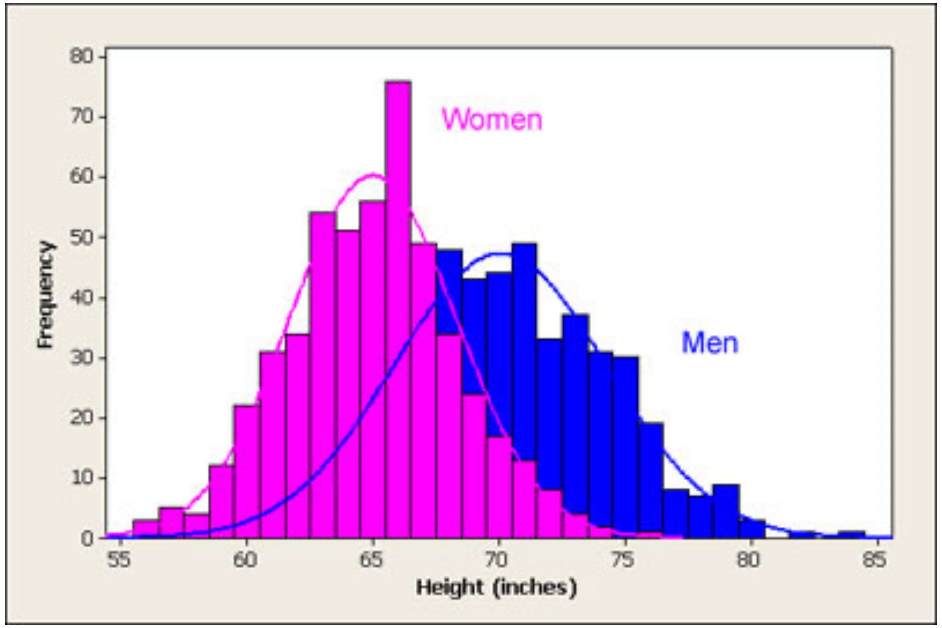
Normalized to the same pose (*important*, later)

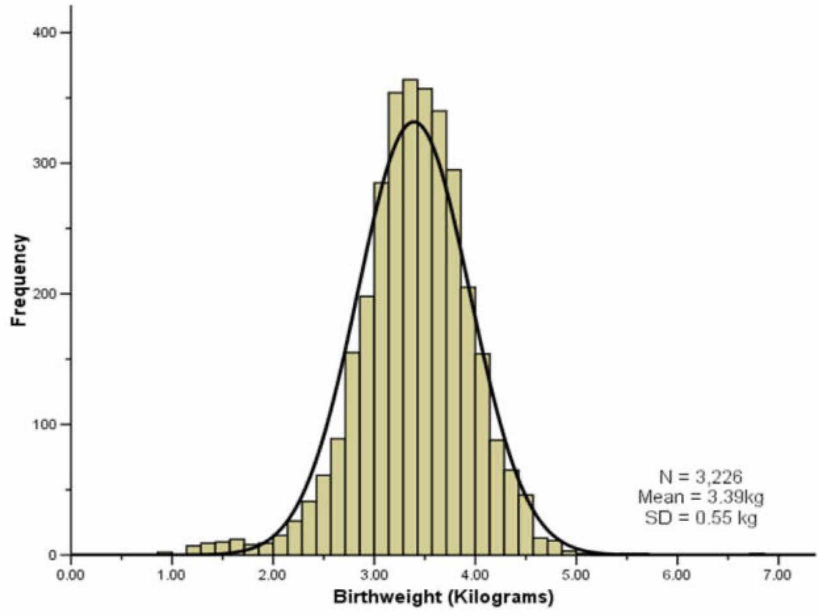
Average man and woman



Average shape of ~2000 men and women in the US and Europe.

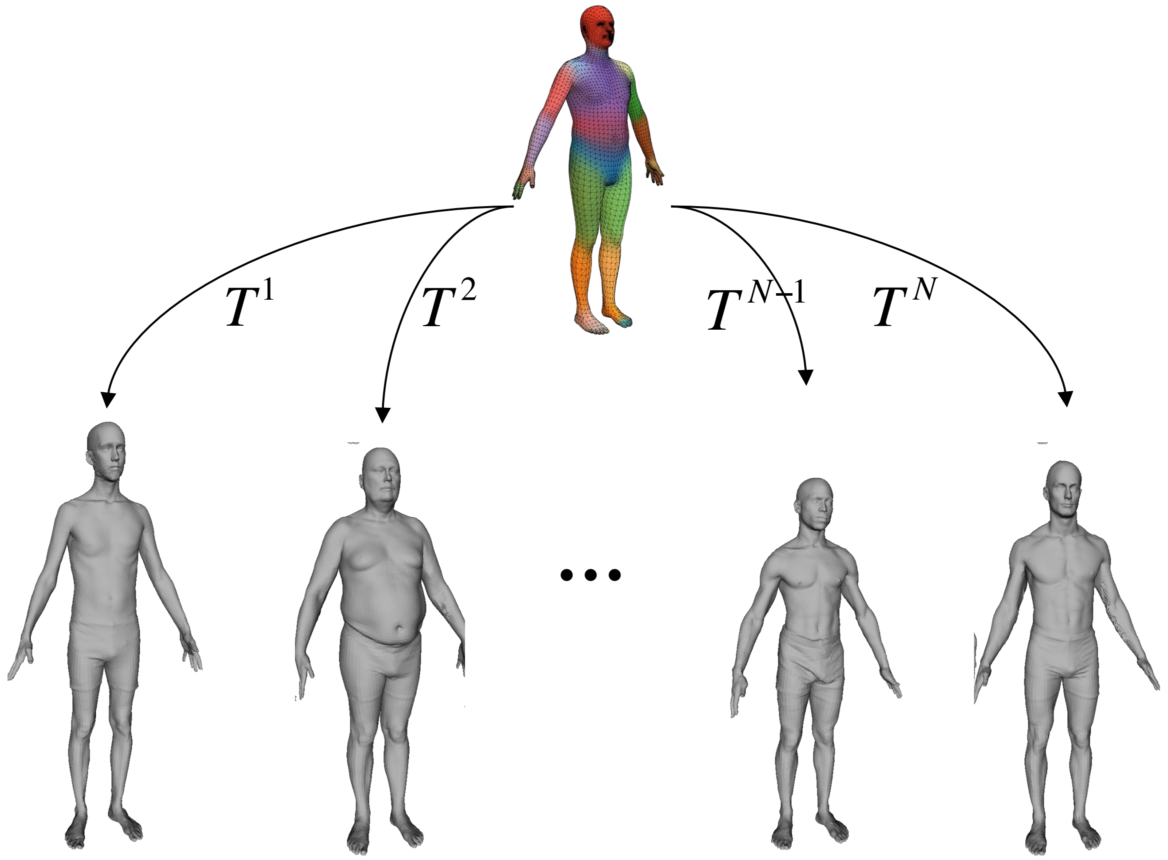
Human shape statistics





Human shapes are well represented by a linear Gaussian model.

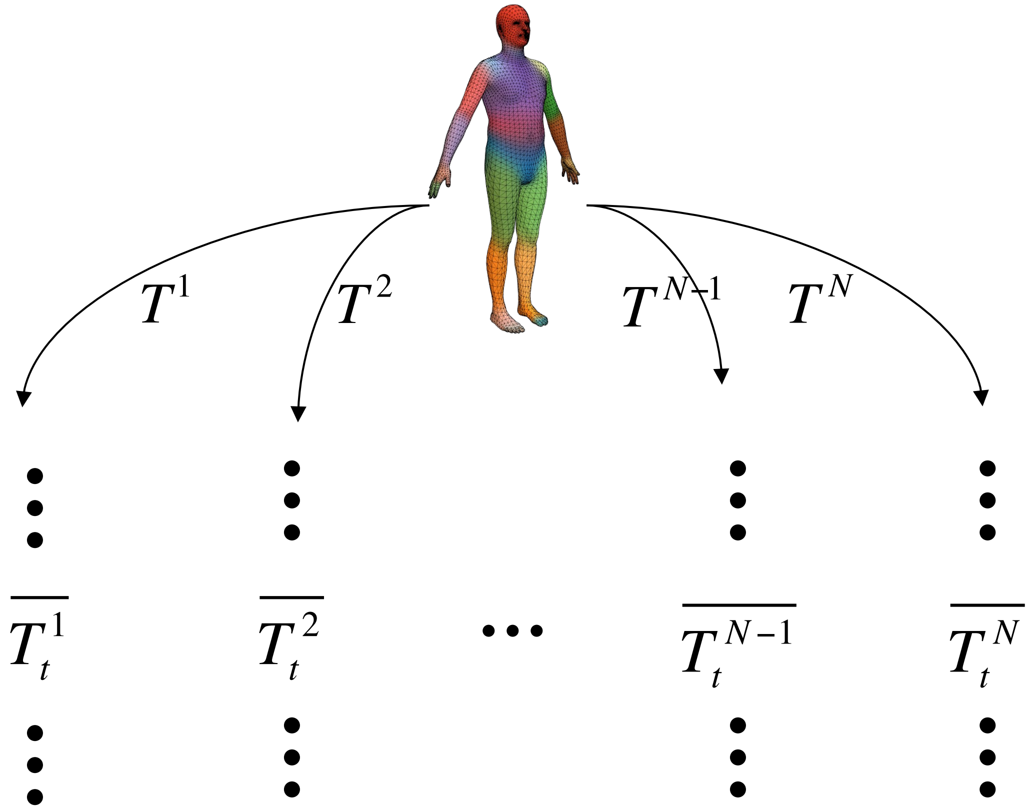
Shape blend shapes



Recall: All in the same pose.

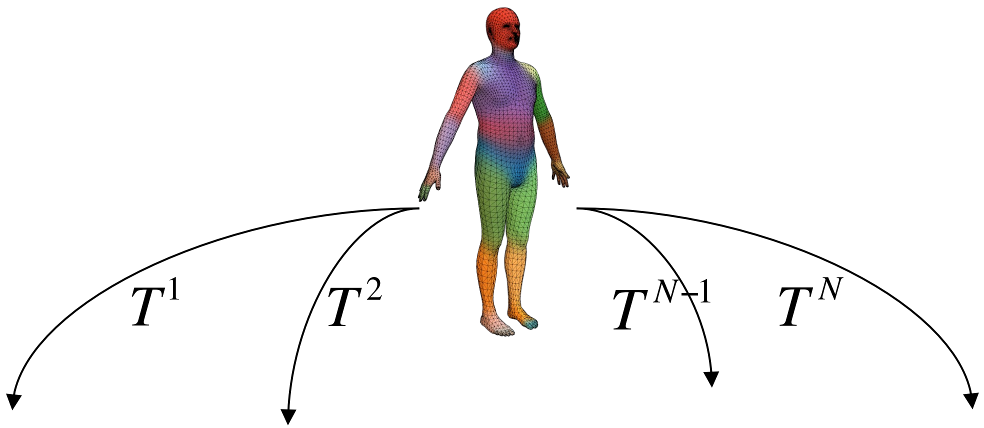
- Each body is a point in a 21,000 dimensional space.
- This is too big to deal with.
- Need to reduce the dimensionality.
- Not enough data (2000 bodies) to use a non-linear deep autoencoder and we don't need to.

Shape blend shapes



- Vectorize the mesh vertices.
- Subtract the mean mesh.
- Because the pose is removed, everything lives in Euclidean space.

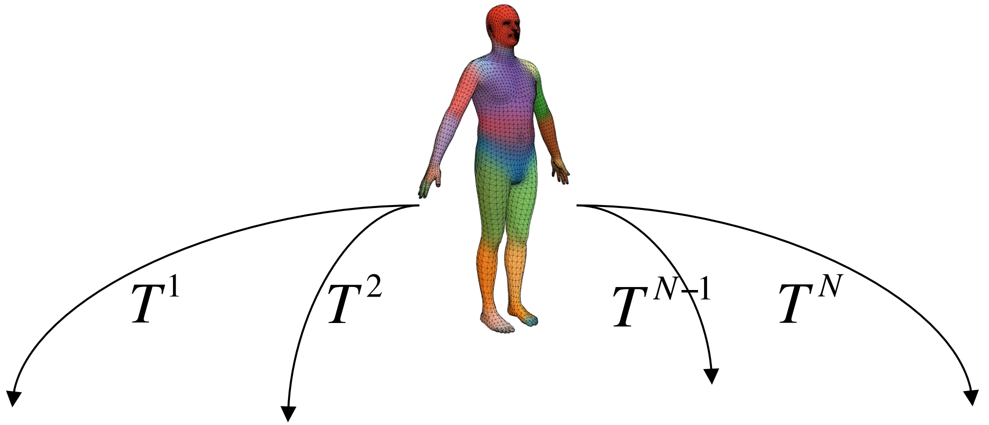
Shape blend shapes



$$\left[\begin{array}{c|c|c|c|c} \vdots & \vdots & & \vdots & \vdots \\ \overline{T_t^1} & \overline{T_t^2} & \cdots & \overline{T_t^{N-1}} & \overline{T_t^N} \\ \vdots & \vdots & & \vdots & \vdots \end{array}\right]$$

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

Shape blend shapes



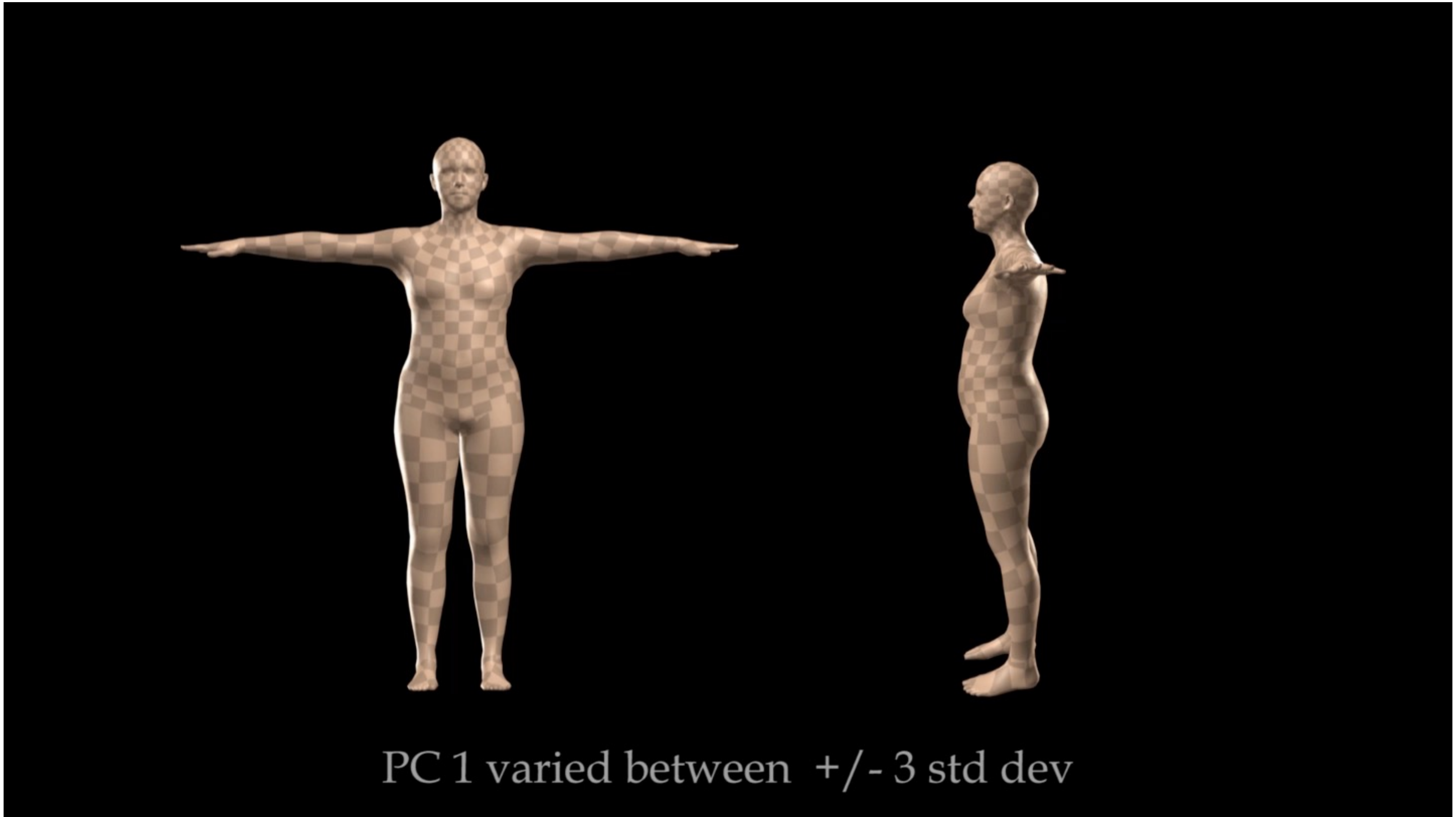
$$\left[\begin{array}{c|c|c|c|c} \vdots & \vdots & & \vdots & \vdots \\ \overline{T_t^1} & \overline{T_t^2} & \cdots & \overline{T_t^{N-1}} & \overline{T_t^N} \\ \vdots & \vdots & & \vdots & \vdots \end{array}\right]$$

Low dimensional shape sub-space

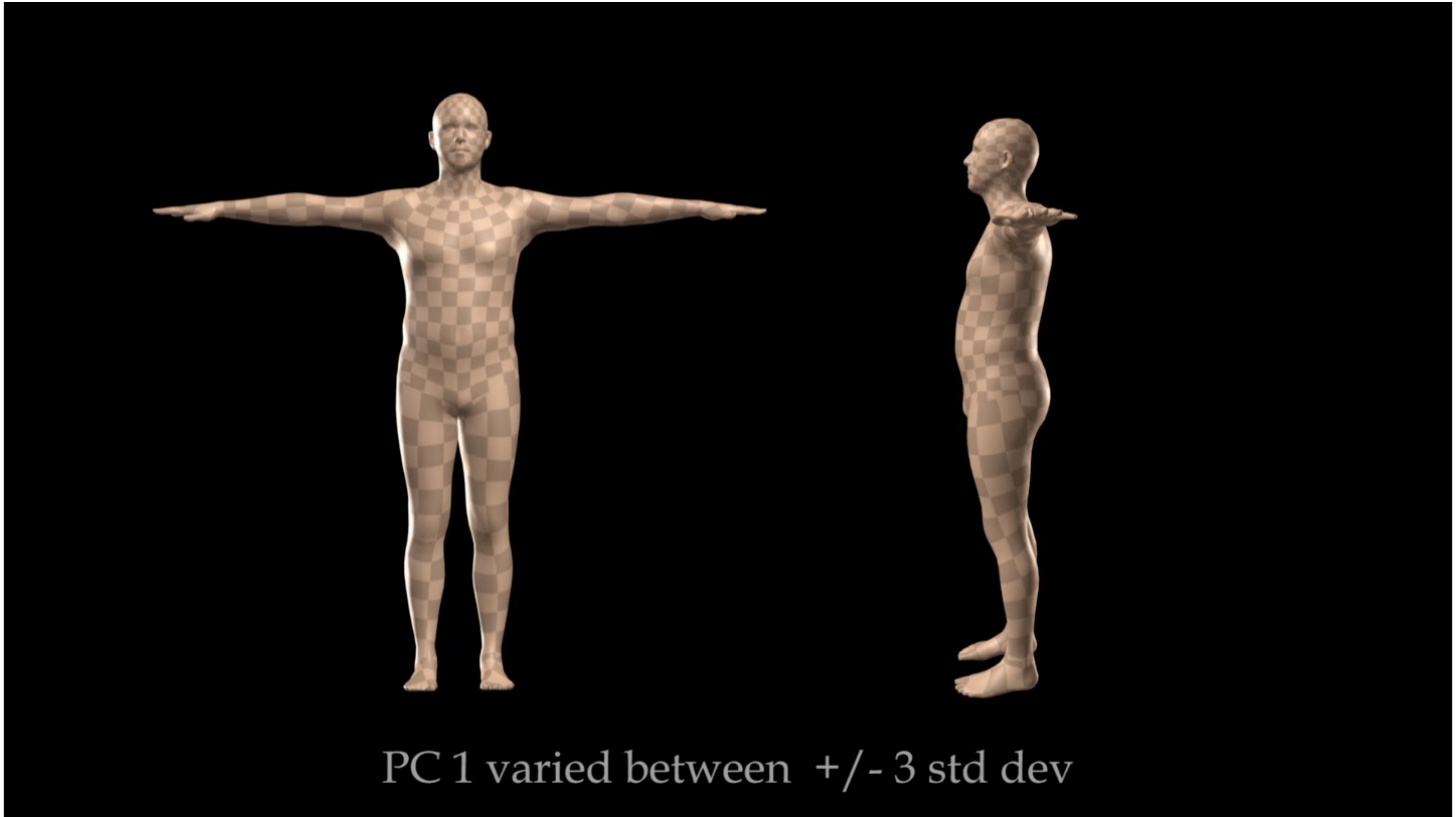
$$\begin{bmatrix} \vdots \\ \overline{T_t^*} \\ \vdots \end{bmatrix} = U\beta^* + \mu = \mathbf{S}(\boldsymbol{\beta})$$

β – shape parameters (e.g. 10-300D)

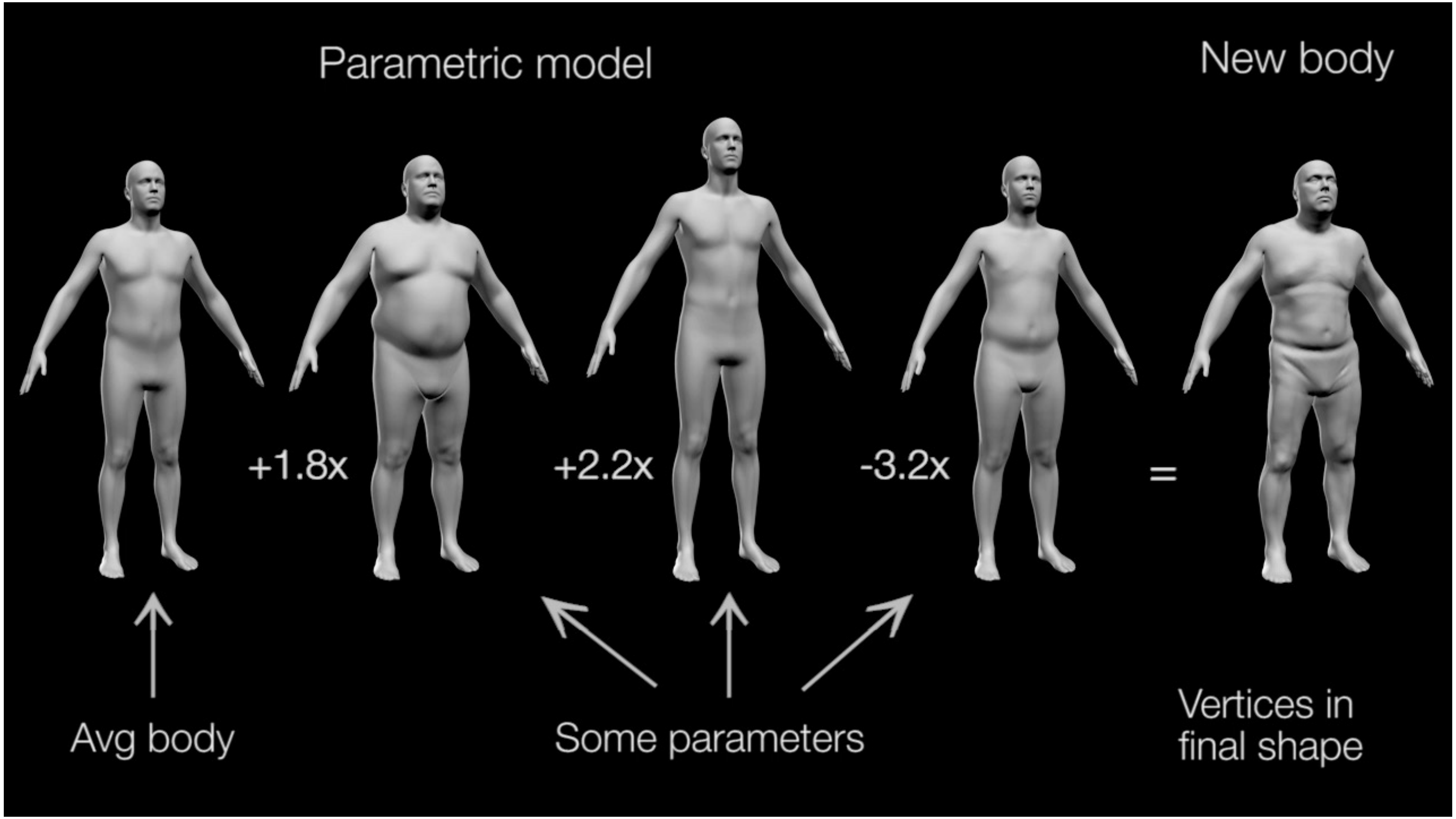
PCA shape space



PCA shape space

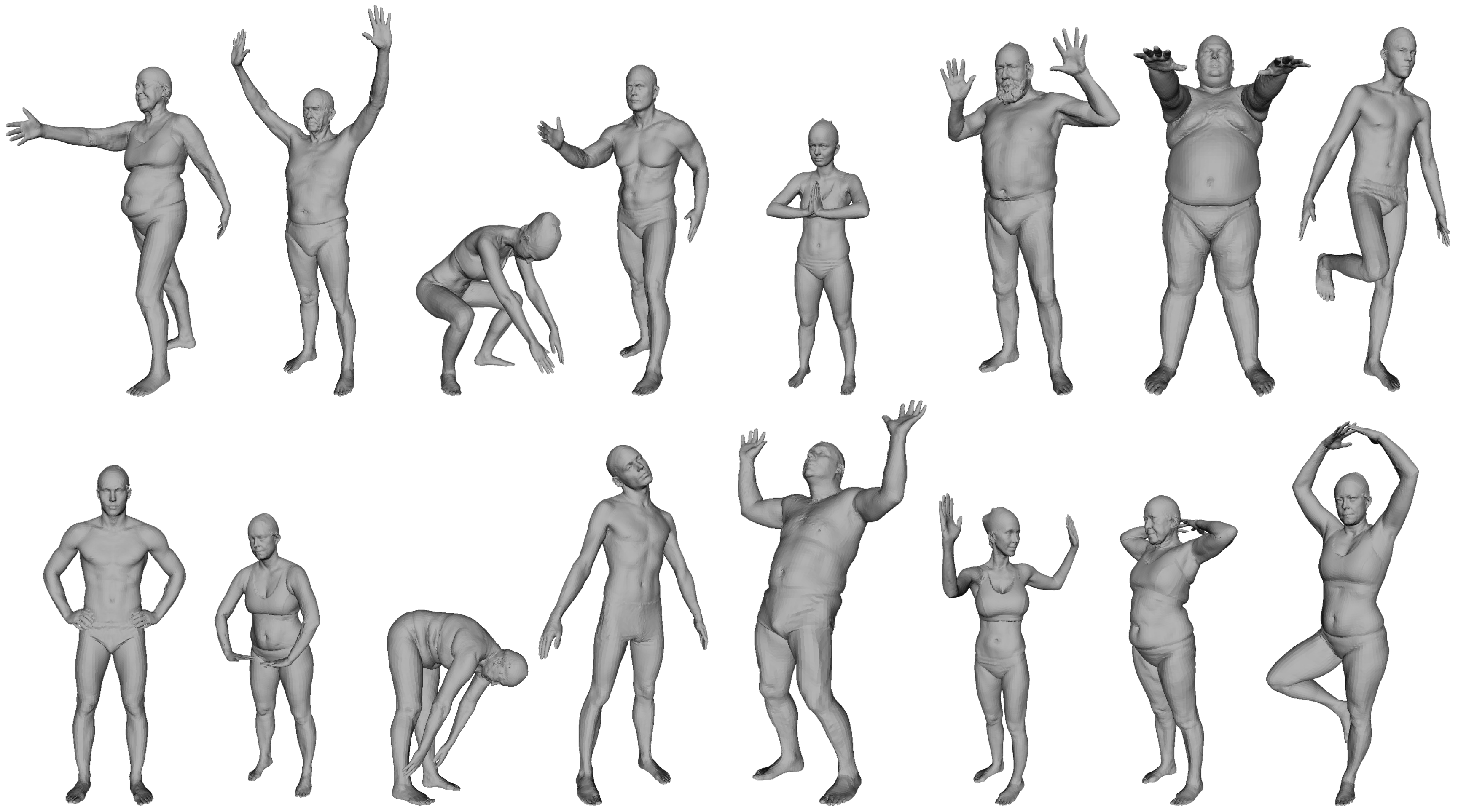


Body math



Parametric model
New body
+1.8x
+2.2x
-3.2x
=
Avg body
Some parameters
Vertices in
final shape

Pose training set

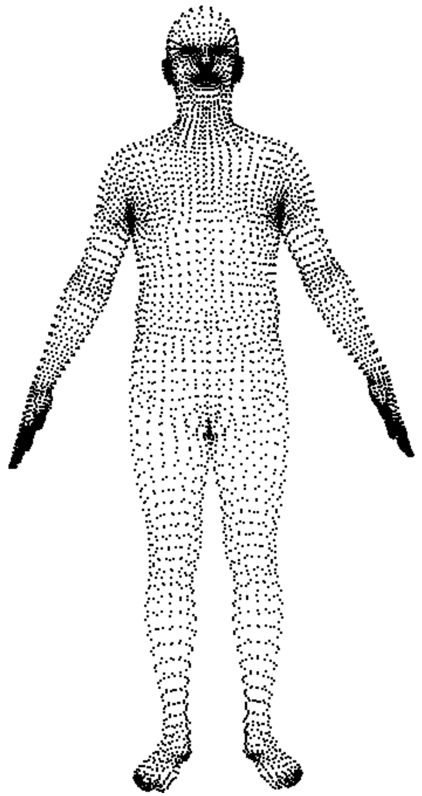


20 males, 24 females, 1800 registrations

Standard Skinning

Standard skinning produces vertices from…

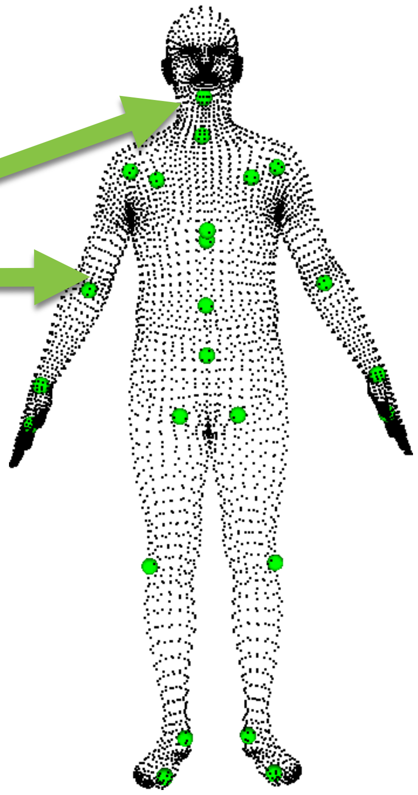
- Rest pose vertices: $\mathbf{T} \in \mathbb{R}^{3N}$
- Joint locations: $\mathbf{J} \in \mathbb{R}^{3K}$
- Weights: $\mathcal{W} \in \mathbb{R}^{N \times K}$
- Pose parameters: $\theta \in \mathbb{R}^{3K}$



Standard Skinning

Standard skinning produces vertices from…

- Rest pose vertices: $\mathbf{T} \in \mathbb{R}^{3N}$
- Joint locations: $\mathbf{J} \in \mathbb{R}^{3K}$
- Weights: $\mathcal{W} \in \mathbb{R}^{N \times K}$
- Pose parameters: $\theta \in \mathbb{R}^{3K}$



Standard Skinning

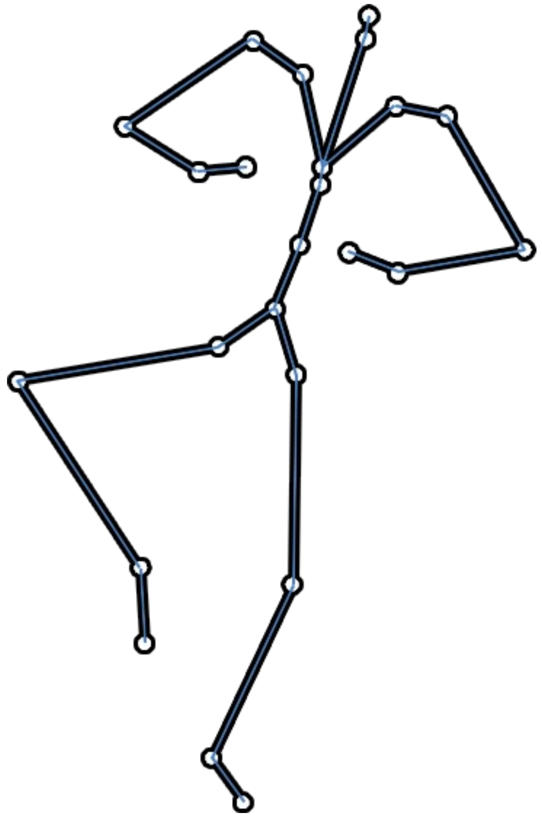
Standard skinning produces vertices from…

- Rest pose vertices: $\mathbf{T} \in \mathbb{R}^{3N}$
- Joint locations: $\mathbf{J} \in \mathbb{R}^{3K}$
- Weights: $\mathcal{W} \in \mathbb{R}^{N \times K}$
- Pose parameters: $\theta \in \mathbb{R}^{3K}$

Standard Skinning

Standard skinning produces vertices from…

- Rest pose vertices: $\mathbf{T} \in \mathbb{R}^{3N}$
- Joint locations: $\mathbf{J} \in \mathbb{R}^{3K}$
- Weights: $\mathcal{W} \in \mathbb{R}^{N \times K}$
- Pose parameters: $\theta \in \mathbb{R}^{3K}$



Skinning function

$$\boxed{W(\mathbf{T}, \mathbf{J}, \mathcal{W}, \theta) \mapsto \text{vertices}}$$

- Rest pose vertices: $\mathbf{T} \in \mathbb{R}^{3N}$
- Joint locations: $\mathbf{J} \in \mathbb{R}^{3K}$
- Weights: $\mathcal{W} \in \mathbb{R}^{N \times K}$
- Pose parameters: $\theta \in \mathbb{R}^{3K}$

SMPL: Skinned Multi-Person Linear Model

Standard skinning $W(\boxed{\mathbf{T}}, \boxed{\mathbf{J}}, \mathcal{W}, \theta) \mapsto \text{vertices}$

SMPL model

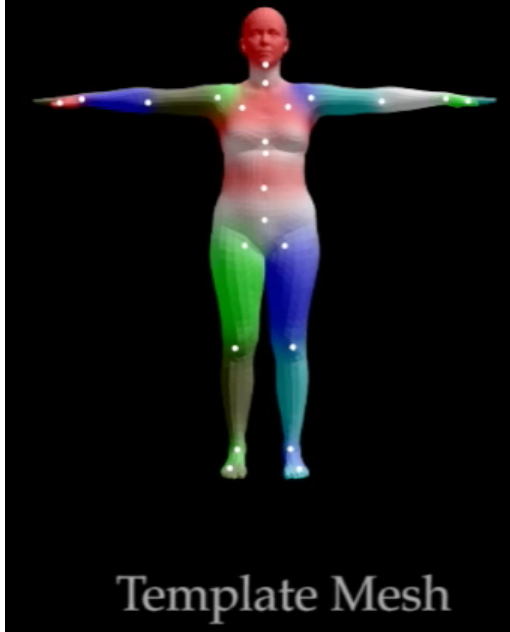
$$M(\theta, \beta) = W(\boxed{\mathbf{T}_F(\beta, \theta)}, \boxed{\mathbf{J}(\beta)}, \mathcal{W}, \theta) \mapsto \text{vertices}$$

SMPL is skinning parameterized by pose θ, and shape β

Loper et al., SIGGRAPH 2015

SMPL Summary

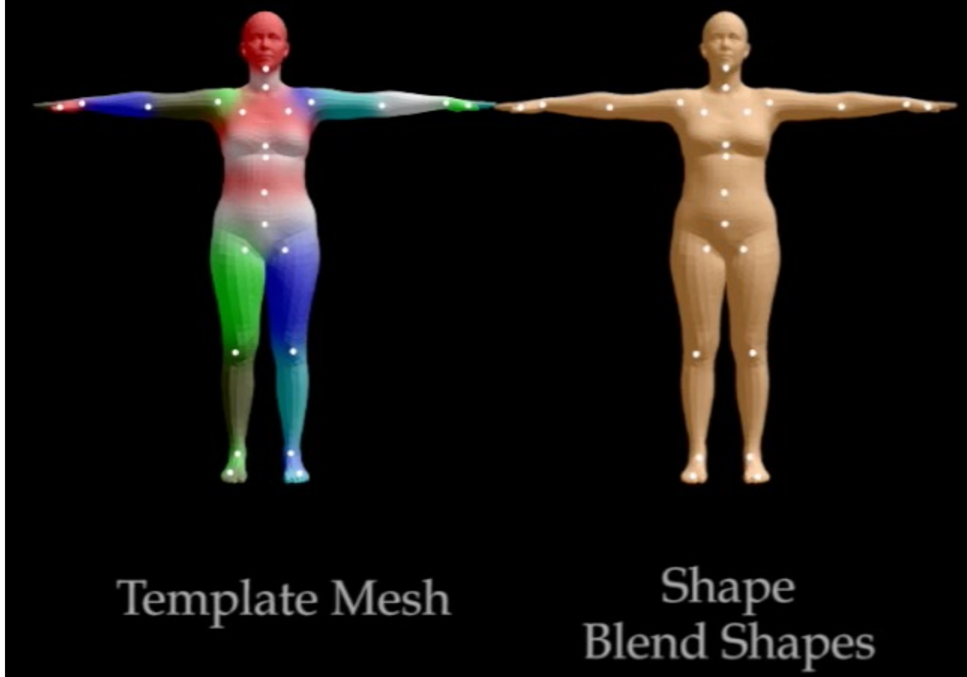
Template mesh in the rest pose. Joints locations.
Blend weights for linear blend skinning.



SMPL Summary

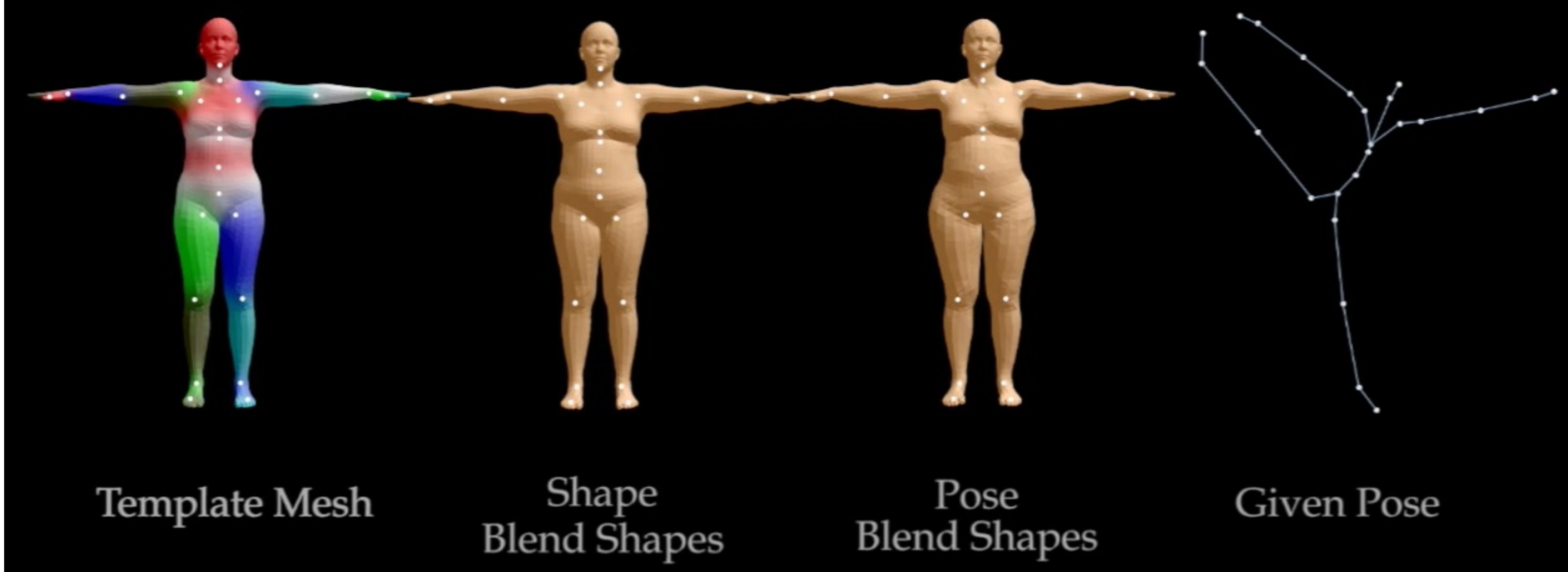
Shape blend shapes represent different body shapes.

Joint locations are a function of body shape.



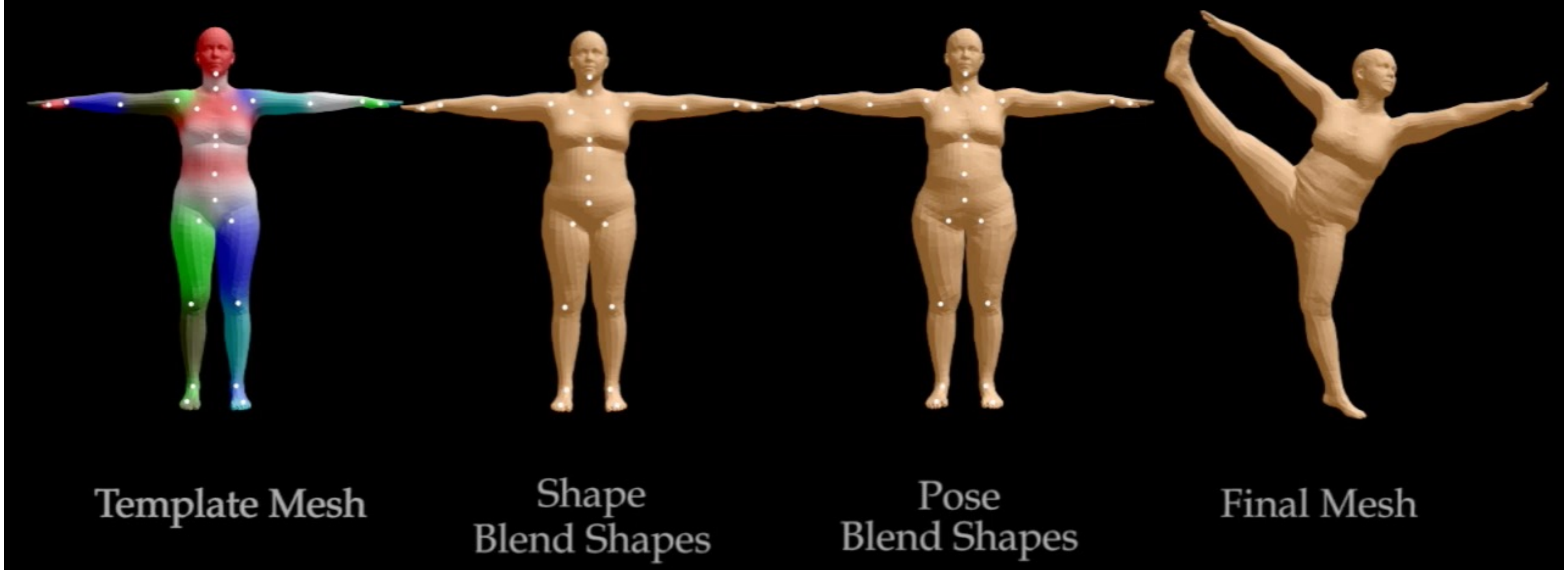
SMPL Summary

Pose blend shapes depend on body pose and correct skinning artifacts and model deformations.

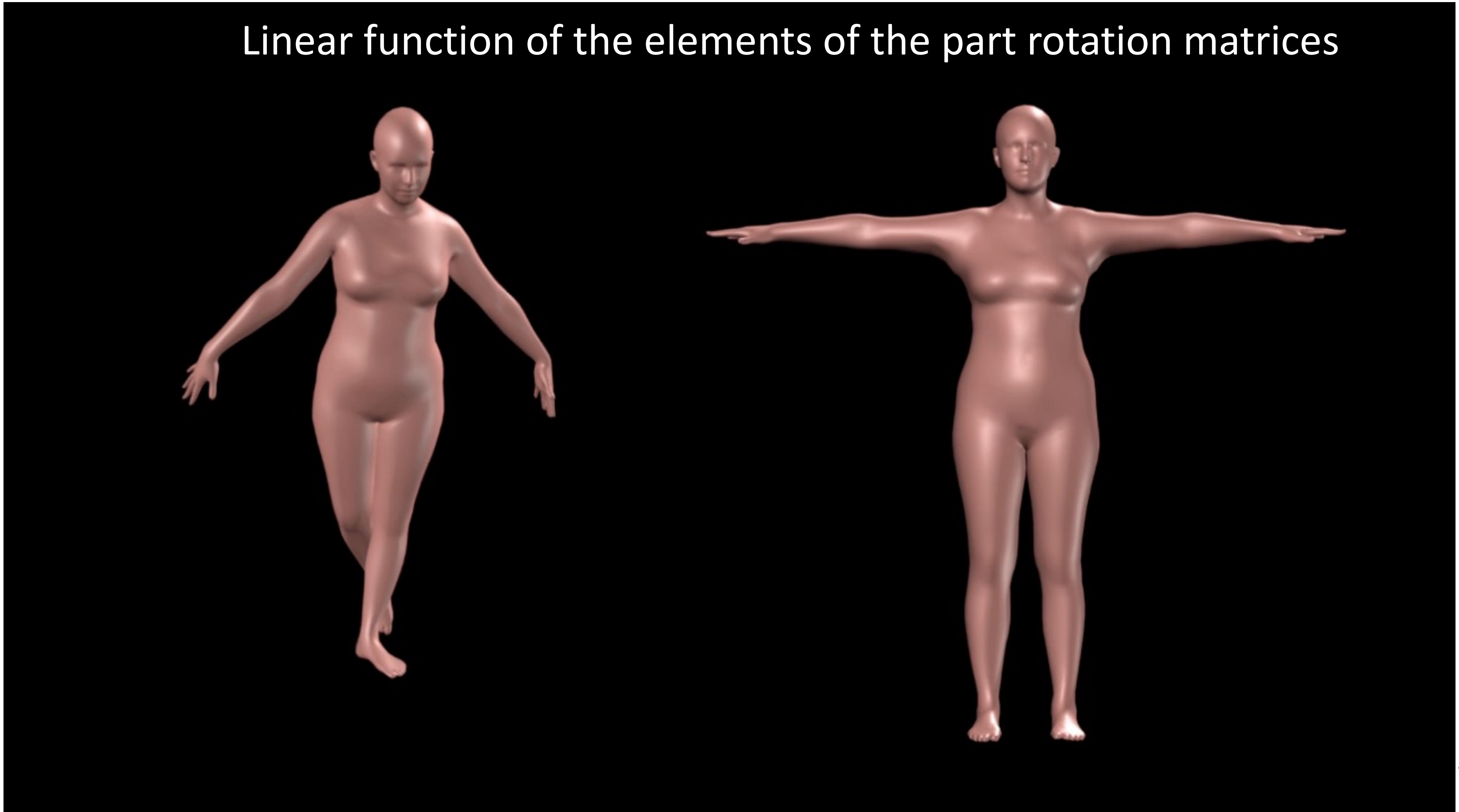


SMPL Summary

Pose blend shapes are a function of the pose (linear in the elements of the part rotation matrices).



Learned Pose Blendshapes



SMPL

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

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

$\mathcal{S}$ Shape blend shape matrix

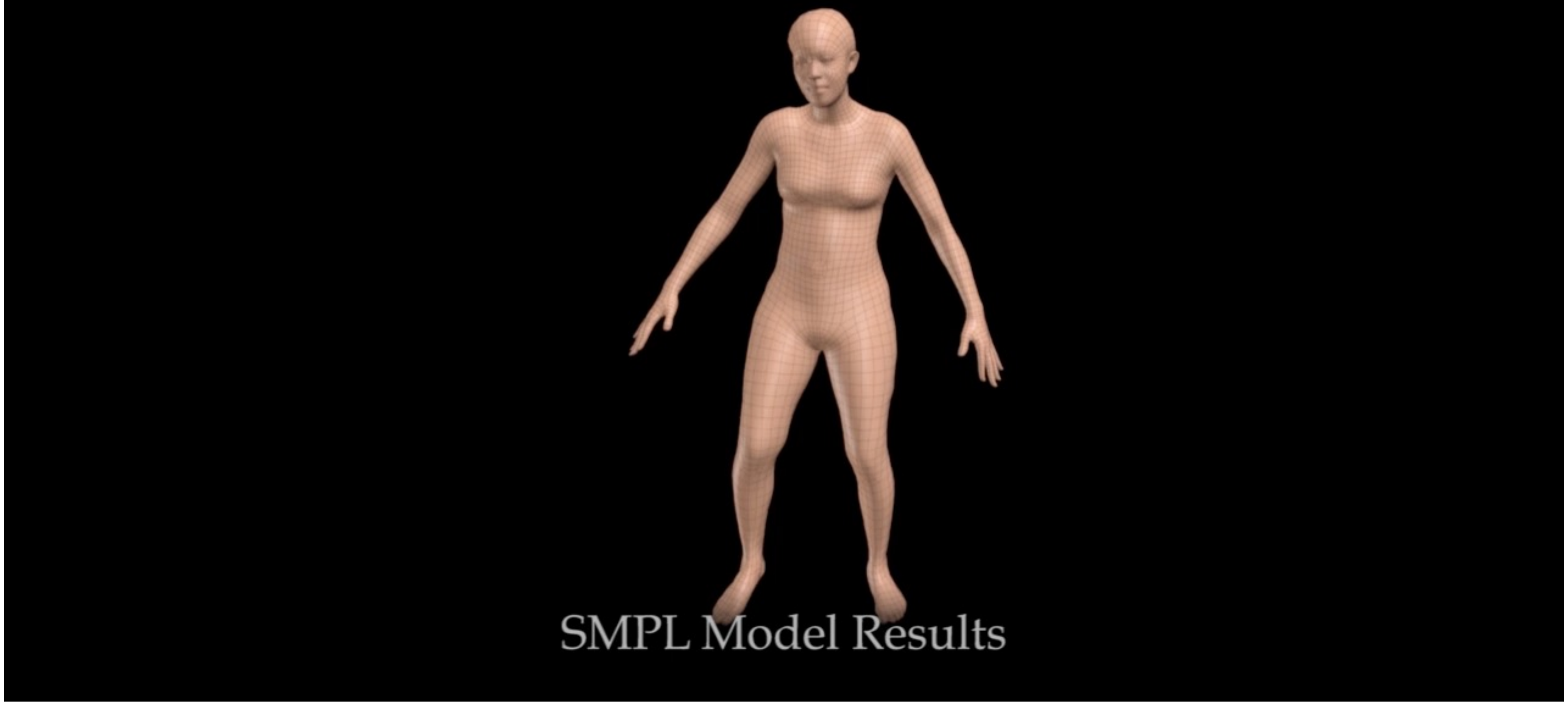
$\mathcal{P}$ Pose blend shape matrix

$\mathcal{W}$ Blend weights matrix

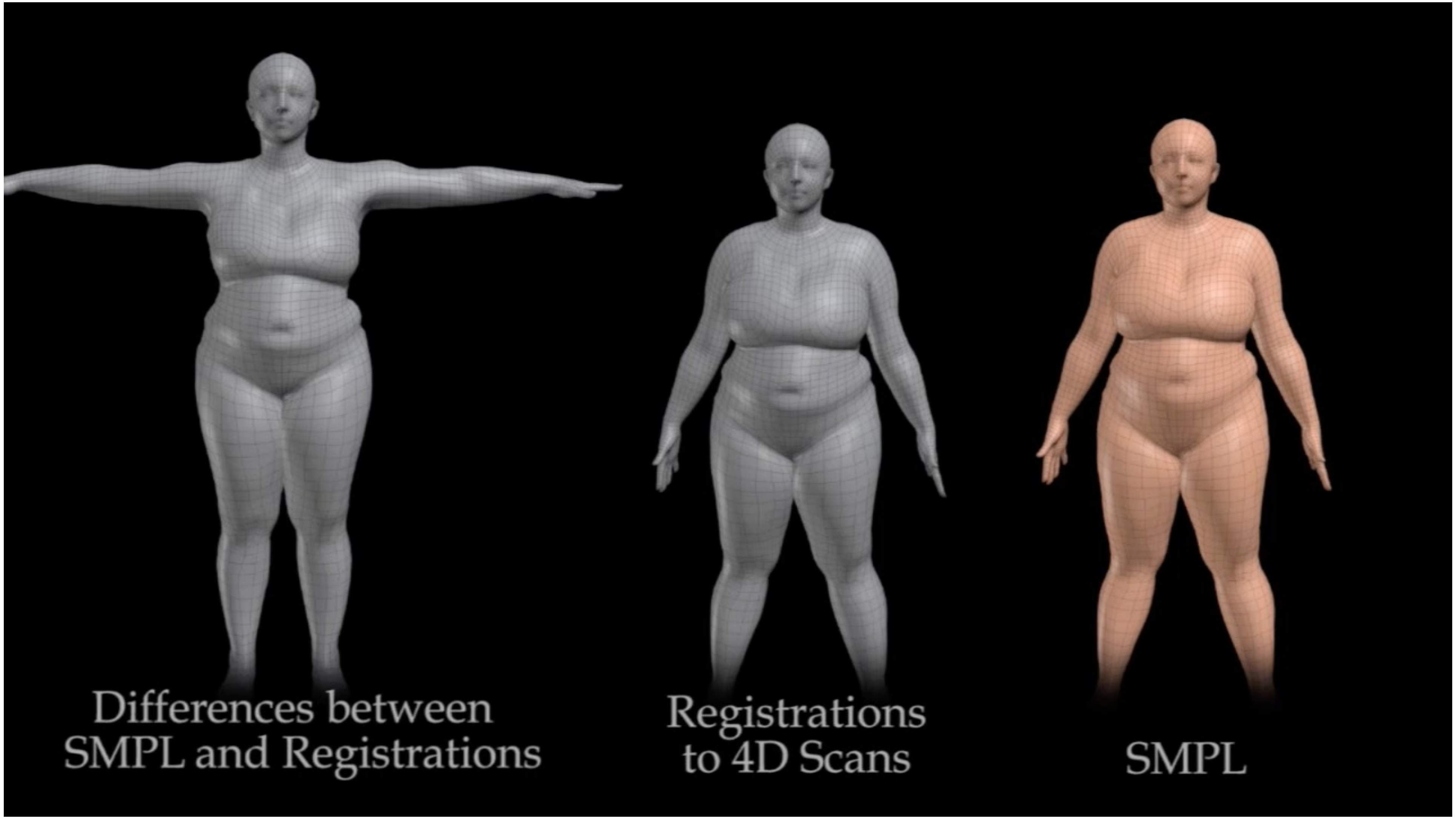
$\mathcal{J}$ Joint regressor matrix

SMPL: A Skinned Multi-Person Linear Model, Loper et al., SIGGRAPH 2015

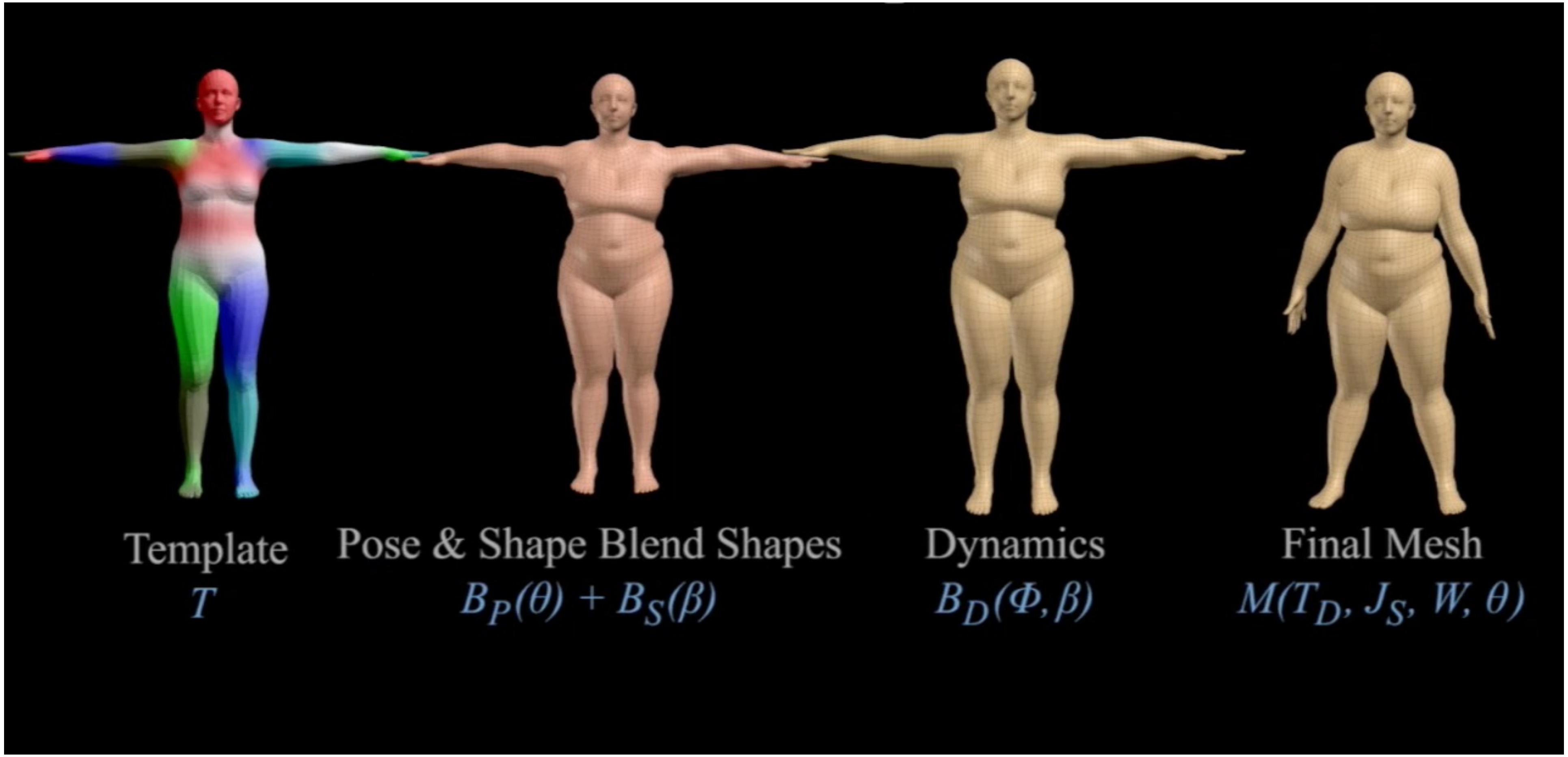
Key idea: Everything is learned from registered data to minimize surface-to-surface error.



Soft-tissue motions

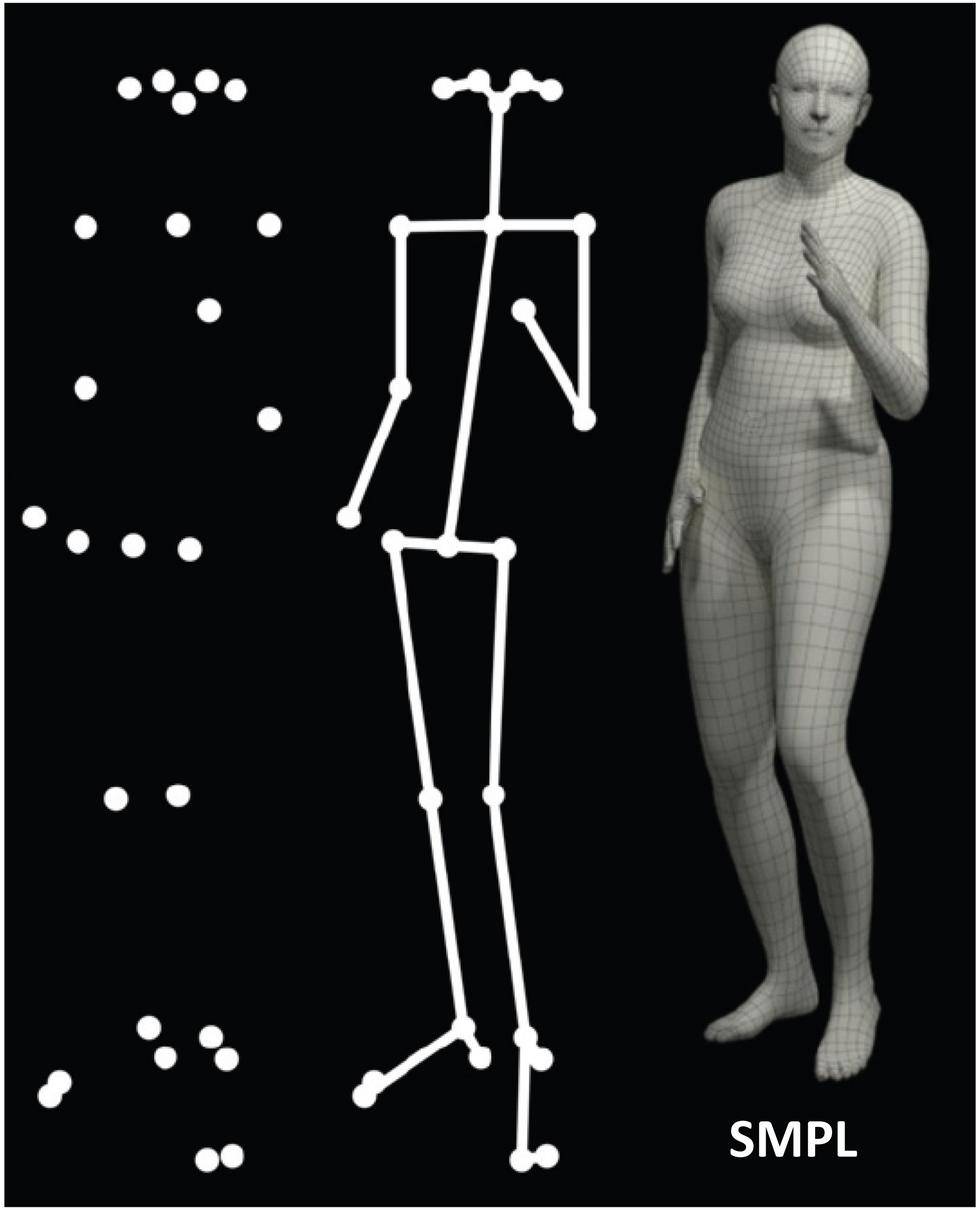


DMPL: Factored and additive



Just another linear, additive, shape term.

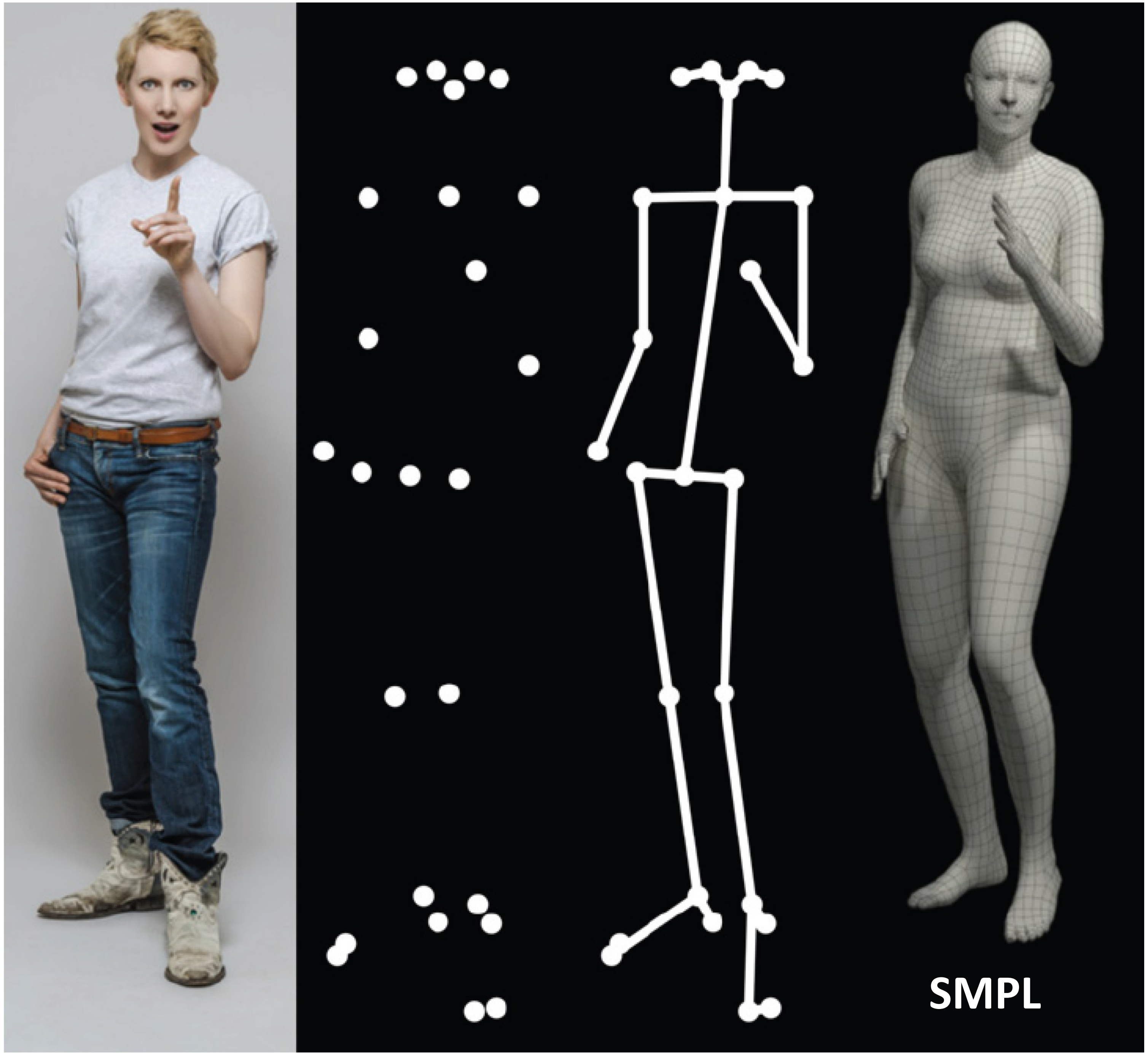
DMPL X 2



SMPL

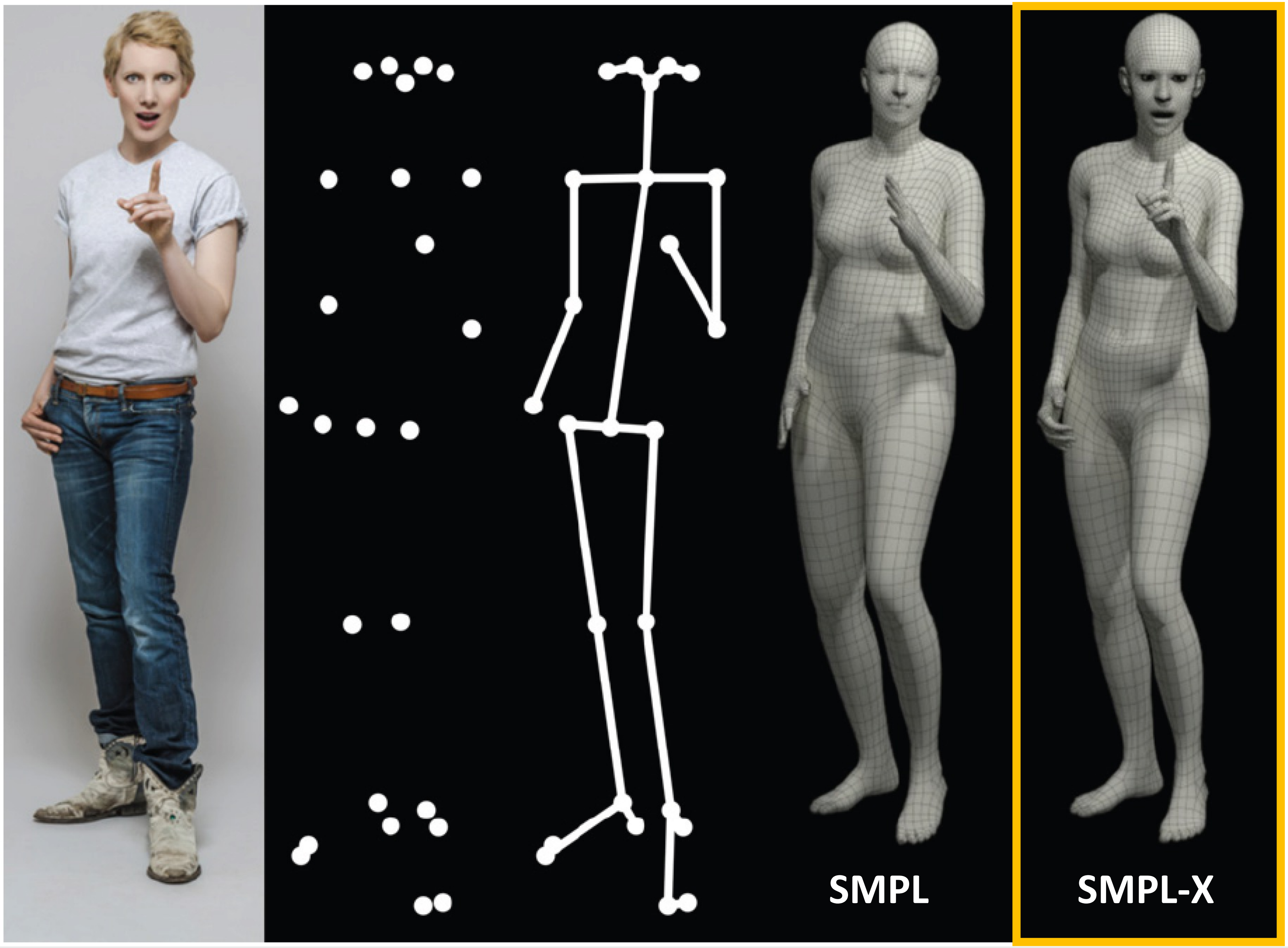
Popularized 3D body pose and shape.

Widely used in academia and industry.



It’s not enough.

We interact with the world and each other using our hands and faces.


SMPL
SMPL-X

FLAME face model

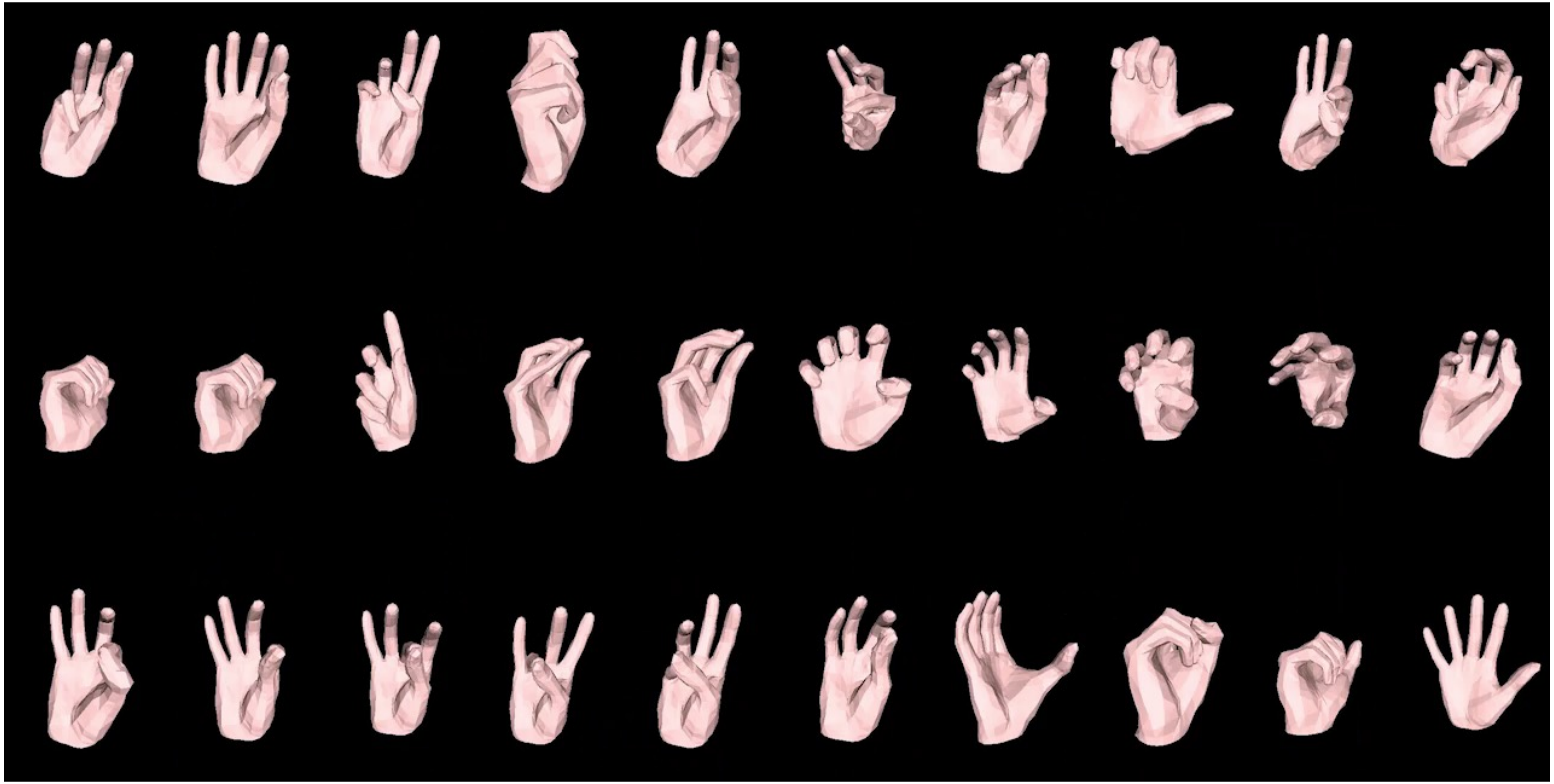
Shape Pose Expression

Based on SMPL: Linear blend skinning with learned pose-dependent deformations, PCA shape model, and a linear space of expressions (see non-linear CoMA model)

Li et al., “Learning a model of facial shape and expression from 4D scans,” SIGGRAPH Asia 2017

MANO hand model

2018 scans: left and right hands of 31 subjects doing 31+ poses



Romero et al., “Embodied Hands: Modeling and Capturing Hands and Bodies Together,” *SIG Asia*, 2017

MANO hand model

Romero et al., "Embodied Hands: Modeling and Capturing Hands and Bodies Together," *SIG Asia*, 2017

Up until here

- Hopefully by now you have a rough idea of what it entails to build a human model, but we will see this in more technical detail
- In the lecture, we will learn many more things!