SOMA: From Surface Observations to Muscle Anatomy

Eduardo Alvarado[1], Emily Kim[1], Gerrit Nolte[2], Friedemann Runte[2], Mario Botsch[2], Marc Habermann[1], and Christian Theobalt[1]

[1] Max Planck Institute for Informatics, Saarland Informatics Campus
[2] TU Dortmund University

https://vcai.mpi-inf.mpg.de/projects/SOMA/

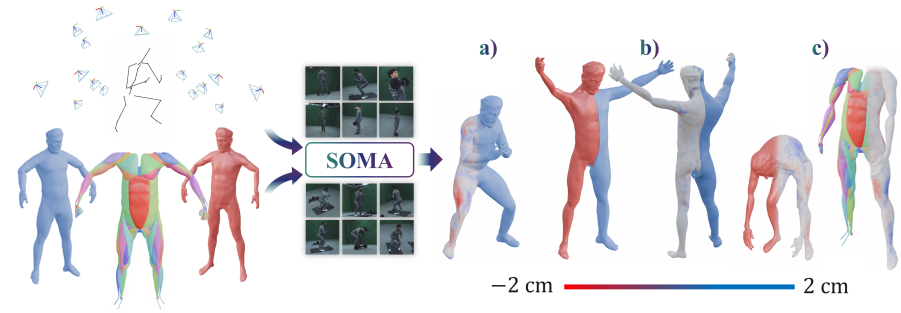

Fig. 1: Our model SOMA allows us to convert arbitrary motion into anatomically plausible a) skin and b) muscle deformations, also targeting c) individual muscles, all grounded with real-data.

Abstract. With the growing demand for realistic virtual humans, parametric body models have become a cornerstone of modern medicine, sports or entertainment applications. However, most of these models are inherently limited: they only capture the 3D surface of the skin, offering no insight into the complex bio-mechanical structures that generate motion. As more applications expand towards biomechanics, the need for virtual human models that go beyond the skin has become increasingly evident. Traditional soft-tissue simulations, such as FEM, are accurate but non-scalable and too computationally expensive for most common applications. Alternatively, existing biomechanical tools can simulate muscular forces and activations, but do not model changes in external shape, restricting how activations correlate with actual observable anatomy. This motivates a novel inverse research problem: recovering muscle deformations directly from visible surface observations - i.e., from the skin, and thus the pose. In this work, we present SOMA (from **S**urface **O**bservations to **M**uscle **A**natomy), a person-specific model that infers spatio-temporal muscle behavior from surface signals obtained using RGB cameras, and SKIM, a subject-specific soft-tissue deformation dataset. To the best of our knowledge, this is the first method that attempts to recover muscle deformations from multi-view RGB data. We show how our method provides anatomically grounded

animations without the complexity of traditional simulations, leading to a scalable and cost-effective solution. Data and code are available.

Keywords: Parametric Human Models · Muscle deformation

1 Introduction

The creation of realistic and controllable digital humans is a long-standing challenge that increasingly demands biomechanical fidelity. Pioneering works in parametric models for human body surfaces [8, 18, 39, 41, 44, 46, 48, 50, 53] have standardized the representation of body shape and pose in a low-dimensional format, enabling scalable human representations. Building on these foundations, recent vision-based approaches aim to retrieve dynamic surface geometry directly from monocular [15, 16, 30] or multi-view RGB inputs [4, 11, 38, 57, 61, 69], bridging the gap between parametric abstraction and real-world visual data. However, these models are typically based solely on surface geometry, offering no insight into the underlying internal structures. As applications in VR, sports science, and medical simulation demand greater detail, bridging the gap between external appearance and internal physiology has become a crucial step for the next generation of digital humans.

Current approaches to modeling internal anatomy have significant limitations. Heuristic methods, often used to fit anatomy to surface scans, rely on simplified assumptions that yield visually plausible but anatomically inaccurate results [14, 21, 23]. Conversely, data-driven methods have been proven successful in solving such inverse problems and fitting anatomical structures using the human surface as input. These approaches leverage medical imaging (MRI, CT) to adapt a volumetric anatomical template, for example, for the entire body skeleton [28, 29, 63] or specific regions, like the skull [3, 24, 49]. However, they suffer from acquisition complexity, privacy concerns, and the difficulty of capturing dynamic, full-body motion [35, 66], especially for methods involving soft tissue parts [27, 32, 37, 40, 62]. The challenge becomes exponentially harder when we are faced with dynamic muscle behavior, due to its anisotropic nature and inability to obtain dynamic, full-body magnetic resonance images. Physics-based simulations can generate dynamic muscle behavior using Finite Element Methods (FEM) [42, 43, 59] or Projective Dynamics [10, 52], and provide useful contributions to learning-based methods, generating synthetic data to predict muscle shape from the pose and external interactions [17].

Yet, the emergence of biomechanics frameworks [56] has led many methods to focus solely on predicting their mechanical behavior, also from the surface skin [12, 47, 54, 55], thus neglecting the prediction of muscle shape. Consequently, there is a clear lack of scalable methods for capturing internal muscle dynamics from accessible, real-world visual data, making it difficult to build parametric models that accurately reflect internal shapes. Additionally, muscle appearance, in comparison to skeletal motion, is influenced by a broader set of factors, such as physiological cross-sectional area (PCSA) at rest, muscle activations or external interactions, all of which introduce significant complexity.

These challenges underscore the need for a fundamentally new approach, one that bypasses the reliance on inaccessible medical data and instead leverages visual observations. In this work, we introduce SOMA, a vision-based framework designed to recover dynamic muscle deformations directly from poses, e.g., retrieved from video (see Fig. 1). Our method combines surface signals with individualized anatomical templates (e.g., derived from 3D scans), enabling the recovery of rich internal representations. This allows for anatomically grounded modeling that is scalable, non-invasive, and well-suited for downstream applications.

Our approach operates in three main stages. First, we introduce a vision-based data acquisition framework using a custom marker-embedded suit to capture high-resolution spatio-temporal signals of skin deformation during motion. Second, we propose an end-to-end data-driven pipeline that learns a non-linear blend-shape representation of a canonical muscle and skin template, supervised by these surface observations. Third, we integrate individual 3D muscle meshes to refine the deformation output, propagating the learned surface changes to detailed anatomical geometry. In summary, the contribution of this work is threefold:

- SKIM, a multi-view RGB dataset bridging surface observations and internal anatomy via high-res, spatio-temporal marker tracking and ground-truth pose.
- SOMA, a parametric model recovering muscle deformation from surface signals, utilizing decoupled blendshapes supervised by physics-based geometric priors.
- A framework propagating learned volumetric deformations to individual muscle meshes, enabling interactive internal anatomy in standard graphics engines.

2 Related Work

We review the most relevant literature to our approach across three primary domains: surface-only human body models, inverse anatomical models, and muscle-based simulations (see Tab. 1).

Surface-only Human Body Models. Many parametric models of the body [6, 18, 39, 46, 48, 65], hands [51], and face [36] rely on standard Linear Blend Skinning (LBS) and blendshapes.They are recovered via monocular RGB [15, 16, 25, 31], multi-view setups [4, 38, 57, 61, 69], or IMUs [20, 64], achieving high surface fidelity. However, deformations remain purely geometric, implicitly learning soft-tissue changes without representing the underlying anatomy.

To enhance realism, some methods mimic biomechanical effects [32] or use EMG data to predict muscle activations as 2D surface textures [12,55] or animation drivers [47]. Relying solely on visual data, Neumann et al. [44] learn data-driven surface deformations from arm muscle bulges. Ultimately, these prior works reduce musculature to abstract signals, image-space textures, or statistical surface effects. Because they lack explicit 3D anatomical structure, they cannot leverage true volumetric deformation properties. Our work addresses this gap by capturing and modeling explicit, subject-specific 3D muscle deformations directly from multi-view RGB body poses.

Inverse Anatomical Models. Unlike surface-only representations, inverse models [23, 26, 27, 33] infer the underlying internal anatomy (e.g., bones, muscles)

from external observations. This introduces the inverse challenge: recovering internal structures from sparse, often noisy surface data rather than simulating them from known biomechanical parameters.

Heuristic and Template-Based Approaches. These methods non-rigidly register a generic anatomical template to a subject's surface scan. This strategy enables transferring anatomy to new body shapes [26] and creating animation-ready models [14, 21]. While visually plausible, their reliance on heuristics and generic templates fundamentally limits their ability to faithfully reconstruct true variations in the muscles' anatomical shape, their origins, and insertions.

Data-driven Anatomical Reconstruction. Recent data-driven methods predict subject-specific internal anatomy directly from surface shapes using paired 3D scan and medical datasets. Much focus has been on recovering skeletons [28, 29, 63], with extensions to the face/skull [3, 45, 49] and hands [37]. For full-body volumes, layered anatomical templates [33, 62] produce static predictions from real-world scans [40] but cannot recreate dynamic results. Similarly, methods like HIT [27] learn continuous implicit representations from MRI to infer internal tissues from the outer surface. Since these models are trained on single, static poses, they share a fundamental limitation: they lack real-world data on how muscles dynamically deform during motion. Our work aims to address this gap.

Table 1: Human Models using anatomical features. "✓" indicates methods that use real-world data but are limited to static poses or heuristics, often struggling to capture accurate dynamic motion.

Method	3D Muscle Representation	Real-world Data	Full-body
SMPL [39]	✗	✓	✓
Neumann et al. [44]	✗	✓	✗
Kavan et al. [26]	✓	✗	✓
Komaritzan et al. [33]	✓	✓	✓
Keller et al. [27]	✓	✓	✓
Han et al. [17]	✓	✗	✗
Kemper et al. [9]	✓	✗	✓
Ours	✓	✓	✓

Muscle-based Simulation. Finite Element Method (FEM) simulations offer high accuracy by modeling volumetric biomechanical meshes [42, 59], enabling realistic personalized anatomy [23, 52] but at expensive computational costs. To achieve real-time performance, other methods [17] approximate physics by training neural networks on offline FEM data or using simplified line-actuator models like OpenSim [54, 56], sacrificing volumetric detail. Besides, such simulations struggle with reality grounding. Because dynamic real-world data (e.g., MRI [60]) is scarce and mostly used for validation [7, 19], simulations rely on generic physical parameters. Consequently, learning volumetric muscle deformation directly from *in-vivo* motion data remains an open problem.

3 Skin-to-Internal Muscle Dataset (SKIM)

Inferring internal muscle deformation from RGB videos is inherently ill-posed, as it relies solely on indirect skin surface observations. Resolving this requires coupling high-precision, temporally consistent skin annotations with an individualized volumetric muscle template.

To provide these high-resolution, spatio-temporally coherent, and minimally occluded observations, we introduce the SKIM dataset (see Fig. 2). SKIM utilizes a high-fidelity data acquisition framework centered on a custom skin-tight suit embedded with ArUco markers, captured via multi-view RGB markerless motion capture.

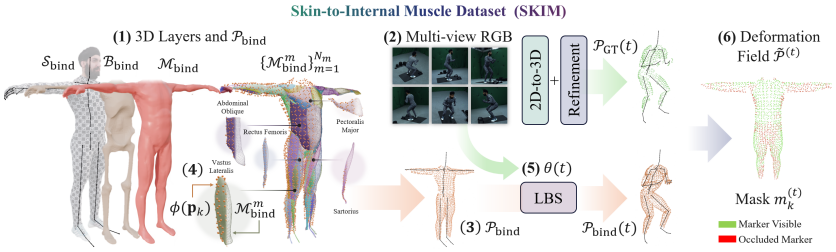

Fig. 2: SKIM Dataset Overview. For five subjects, we provide comprehensive multi-layer data: (1) high-resolution 3D skin, bone, and individual muscle meshes; (2) 45 minutes of multi-view RGB recordings; (3) canonical ArUco marker point clouds with (4) internal muscle bindings; (5) ground-truth skeletal poses; and (6) 3D marker trajectories, residual deformation fields, and visibility masks capturing soft-tissue dynamics.

3.1 Capture Setup

We designed a custom ArUco-embedded suit [13] to capture dynamic skin deformations in our multi-view studio. This suit enables robust spatio-temporal tracking of dense surface correspondences while ensuring consistent muscle-relative placement, minimizing occlusions via known topologies, and eliminating subject-specific calibration.

ArUco-based Suit Specifications. Constructed from stretchable, matte fabric for tight skin contact and minimal reflections, the suit's full set of landmarks is defined as $\mathcal{C} = \{\mathbf{n}_{m,c} \mid m \in \{0, \ldots, M-1\}, c \in \{0,1,2,3\}\}$, where each landmark $\mathbf{n}_{m,c}$ represents corner c of the ArUco marker with dictionary ID m.

Markerless Human Capture. A 120-camera mocap studio [1] capturing at 25 Hz provides both the suit marker detections and the ground-truth 3D skeletal poses $\boldsymbol{\theta}_t \in \mathbb{R}^{J\times 6}$, where J joints are parameterized in a continuous 6D representation [68] at time t. A separate 140-camera scanner [2] reconstructs the static body mesh templates (Sec. 3.2).

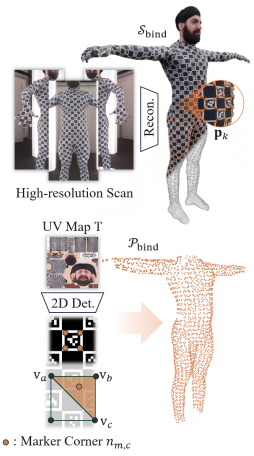

Fig. 3: Canonical Representation.

3.2 Canonical Point Correspondences

To enable anatomically grounded modeling from surface observations, we first construct a canonical anatomical template that encodes the marker layout of the suit in a subject-specific 3D configuration. This template serves as a static reference point cloud $\mathcal{P}_{\text{bind}} = \{\mathbf{p}_k \in \mathbb{R}^3\}_{k=1}^{\mathcal{C}}$, where each point $\mathbf{p}_k$ is the 3D position of a specific marker corner $\mathbf{n}_{m,i,c} \in \mathcal{C}$.

To construct this template, we use our body scanner to generate a high-resolution static skin mesh of the subject [5] wearing the suit, denoted as $\mathcal{S}_{\text{bind}}$. The mesh $\mathcal{S}_{\text{bind}}$ consists of a set of vertices $\mathcal{V}_{\text{bind}} = \{\mathbf{v}_j \in \mathbb{R}^3\}_{j=1}^{V}$ and faces $\mathcal{F}_{\text{bind}} = \{\mathbf{f}_k\}_{k=1}^{F}$, where each face $\mathbf{f}_k$ is a triangle defined by three vertex indices. To robustly detect marker IDs and their corner positions, we unwrap the texture of $\mathcal{S}_{\text{bind}}$ into a 2D atlas $\mathbf{T}$, perform ArUco detection on $\mathbf{T}$, and re-project the detected marker corners back onto the 3D mesh. During unwrapping, texture seams might occur, leading to markers being cut. The process to solve such artifacts is explained in detail in the Supplementary Material. For each detected corner, we identify the corresponding triangle $\mathbf{f}_k = (\mathbf{v}_a, \mathbf{v}_b, \mathbf{v}_c)$ and compute its 3D location using barycentric interpolation $\mathbf{p}_k = \lambda_1\mathbf{v}_a + \lambda_2\mathbf{v}_b + \lambda_3\mathbf{v}_c$, s.t. $\sum \lambda_i = 1, \lambda_i \geq 0$, being $\mathbf{p}_k \in \mathbb{R}^3$ a point in the canonical point cloud $\mathcal{P}_{\text{bind}}$ (see Fig. 3).

3.3 Internal Anatomical Modeling

The internal canonical anatomy is reconstructed from the outer mesh $\mathcal{S}_{\text{bind}}$ using a volumetric fitting [33], yielding three additional representations (see Fig. 4):

1. A global muscle mesh $\mathcal{M}_{\text{bind}}$ that approximates the aggregated muscle volume beneath the skin. This mesh shares the same topology as $\mathcal{S}_{\text{bind}}$.
2. A set of individualized muscle meshes $\{\mathcal{M}_{\text{bind}}^m\}_{m=1}^{N_m}$, where each mesh represents the geometry of a specific muscle m in its resting shape.
3. A skeleton mesh $\mathcal{B}_{\text{bind}}$ that wraps the skeleton layer beneath the muscle layer. This mesh also shares the same topology as $\mathcal{S}_{\text{bind}}$.

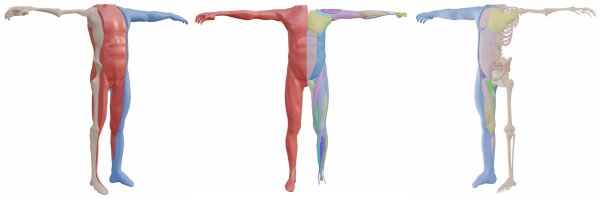

Fig. 4: Anatomical Representations. Left: Set of bone $\mathcal{B}_{\text{bind}}$, muscle $\mathcal{M}_{\text{bind}}$ and skin layer $\mathcal{S}_{\text{bind}}$. Center: High-res muscles $\{\mathcal{M}_{\text{bind}}^m\}_{m=1}^{N_m}$. Right: Full High-res Representations.

Finally, to associate each marker with its corresponding muscle, we cast a ray from the marker position $\mathbf{p}_k$ along the negative surface normal $-\alpha\mathbf{n}_k$ of its underlying triangle in $\mathcal{S}_{\text{bind}}$. The first intersection point of this ray with the individual muscle $\mathcal{M}_{\text{bind}}^m$ sets the anatomical correspondence. This defines a

mapping function $\phi(\mathbf{p}_k) = m$, s.t. $\mathbf{p}_k - \alpha \mathbf{n}_k \in \mathcal{M}^m_{\text{bind}}$ and $\alpha > 0$, ensuring precise marker-to-muscle correspondence. Our canonical representation of markers and muscle anatomy is shown in Fig. 2.

3.4 Marker Tracking

Having established the (1) high-resolution skin and muscle meshes, (2) a dense set of surface markers, and (3) their specific anatomical correspondences, we now turn to leveraging our multi-view capture system to track these markers over time. For each frame t, we detect visible marker corners in the 2D image plane of each camera stream j, yielding detections $\mathcal{D}^{(j)}_t \subset \mathbb{R}^2$. Corners identified across multiple views are triangulated and refined by minimizing the multi-view reprojection error to find the optimal 3D coordinate $\mathbf{p}^*_{k,t}$:

$$\mathbf{p}^*_{k,t} = \arg\min_{\mathbf{p}\in\mathbb{R}^3} \sum_{j\in\mathcal{V}_{k,t}} \left\| \Pi_j(\mathbf{p}) - \mathbf{d}^{(j)}_{k,t} \right\|^2, \tag{1}$$

where Π_j is the camera projection operator, $\mathbf{d}^{(j)}_{k,t} \in \mathcal{D}^{(j)}_t$, and $\mathcal{V}_{k,t}$ is the set of cameras observing marker k at time t. To address observation sparseness from self-occlusions, we apply post-processing strategies, such as temporal smoothing and consensus filtering, to yield refined positions $\tilde{\mathbf{p}}^*_{k,t}$. These steps are explained in detail in the Supplementary Material. The final temporally coherent ground-truth point cloud for the successfully tracked subset $\mathcal{K}_t \subset \{1, \ldots, \mathcal{C}\}$ is defined as:

$$\mathcal{P}_{\text{GT}}(t) = \{\tilde{\mathbf{p}}^*_{k,t} \mid k \in \mathcal{K}_t\}. \tag{2}$$

3.5 Residual Estimation

To compare the static canonical point cloud $\mathcal{P}_{\text{bind}}$ against the temporally accurate but sparse observations $\mathcal{P}_{\text{GT}}(t)$, we articulate $\mathcal{P}_{\text{bind}}$ using the ground-truth skeletal pose $\boldsymbol{\theta}_t$ via Linear Blend Skinning (LBS). By assigning LBS weights $\mathbf{w}_k \in \mathbb{R}^J$ to each marker $\mathbf{p}_k \in \mathcal{P}_{\text{bind}}$ via barycentric interpolation from the rigged skin mesh $\mathcal{S}_{\text{bind}}$, the driven marker position is:

$$\mathbf{p}_k(\boldsymbol{\theta}_t) = \sum_{j=1}^{J} w_{k,j} \left(\mathbf{R}_j(\boldsymbol{\theta}_t)\, \mathbf{p}_k + \mathbf{t}_j(\boldsymbol{\theta}_t) \right), \tag{3}$$

where $\mathbf{R}_j(\boldsymbol{\theta}_t)$ and $\mathbf{t}_j(\boldsymbol{\theta}_t)$ are the global rotation and translation of joint j, derived from the continuous 6D pose representation [68]. Soft-tissue dynamics (e.g., muscle bulging) cause the observed position $\mathbf{p}^*_{k,t}$ to deviate from the kinematic prediction $\mathbf{p}_k(\boldsymbol{\theta}_t)$. We isolate these non-rigid deformations as pose-corrective displacements in the canonical space by computing the world-space residual and applying the inverse blended transformation:

$$\boldsymbol{\delta}_{k,t} = \left(\sum_{j=1}^{J} w_{k,j} \mathbf{R}_j(\boldsymbol{\theta}_t) \right)^{-1} \left(\mathbf{p}^*_{k,t} - \mathbf{p}_k(\boldsymbol{\theta}_t) \right). \tag{4}$$

These vectors form a sparse, temporally evolving field of canonical displacements $\mathcal{D}^{(t)} = \{(\mathbf{p}_k, \boldsymbol{\delta}_{k,t}) \mid k \in \mathcal{K}_t\}$, serving as the ground truth for our muscle deformation model.

4 SOMA

Our approach builds on the concept of corrective pose-dependent blendshapes [34] but extends them to a volumetric, multi-layer anatomy representation. As we introduced in Sec. 3.2, let $\mathcal{M}_{\text{bind}} \in \mathbb{R}^{3N}$ denote the global muscle mesh, $\mathcal{B}_{\text{bind}} \in \mathbb{R}^{3N}$ the global skeleton mesh, and $\mathcal{S}_{\text{bind}} \in \mathbb{R}^{3N}$ the skin mesh in the rest T-pose $\boldsymbol{\theta}^*$. These layers share the same topology and skeletal rigging but represent distinct anatomical boundaries. Our pipeline is illustrated in Fig. 5.

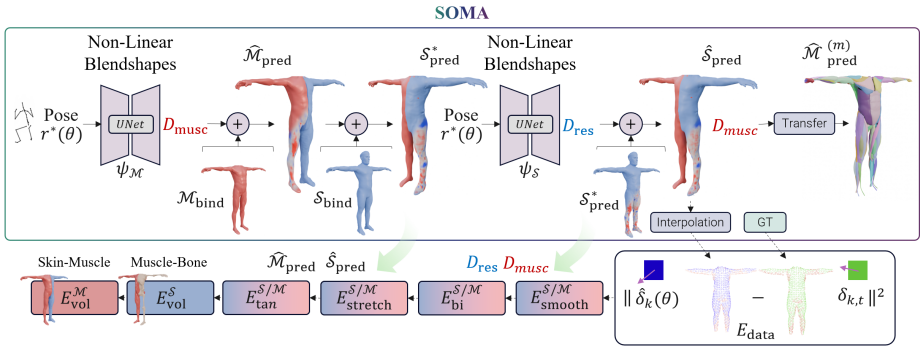

Fig. 5: Overview of SOMA. From pose θ, cascaded blendshapes predict muscle deformation ($\mathbf{D}_{musc}$) and skin residuals ($\mathbf{D}_{res}$) to reconstruct final meshes $\hat{\mathcal{M}}_{pred}$ and $\hat{\mathcal{S}}_{pred}$ from the bind geometry. To regularize marker supervision (E_{data}), we enforce biomechanical priors: (1) Vector constraints for smoothness (E_{smooth}) and sliding (E_{tan}); (2) Geometric constraints ($E_{stretch}$, E_{bi}) and soft-tissue incompressibility (E_{vol}).

4.1 Decoupled Non-Linear Blendshapes Model

To capture complex soft tissue dynamics, we decouple deformation into interdependent structural muscle and volumetric skin layers, rather than using standard single-layer blendshapes.

Muscle Layer Deformation The muscle layer captures primary structural changes (e.g., muscle bulging) driven by skeletal pose $\boldsymbol{\theta}$. We employ a non-linear U-Net architecture, $\Psi_{\mathcal{M}}$, to compute a displacement field $\mathbf{D}_{\text{musc}}$ from pose features $\mathbf{r}^*(\boldsymbol{\theta}) \in \mathbb{R}^{9(J-1)}$ (defined as relative rotation matrices minus identity [39]). The deformed canonical muscle configuration is:

$$\hat{\mathcal{M}}_{\text{pred}}(\boldsymbol{\theta}) = \mathcal{M}_{\text{bind}} + \mathbf{D}_{\text{musc}}(\boldsymbol{\theta}), \quad \text{where } \mathbf{D}_{\text{musc}} = \Psi_{\mathcal{M}}(\mathbf{r}^*). \tag{5}$$

Skin Layer and Residual Offset To allow for soft tissue compression (isochoric deformation) during articulation, the skin surface must not rigidly follow the muscle. We formulate skin deformation as a *residual* function of the muscle layer:

$$\mathbf{D}_{\text{skin}}(\boldsymbol{\theta}) = \mathbf{D}_{\text{musc}}(\boldsymbol{\theta}) + \mathbf{D}_{\text{res}}(\boldsymbol{\theta}), \tag{6}$$

where $\mathbf{D}_{\text{res}} = \Psi_{\mathcal{S}}(\mathbf{r}^*)$ is a residual offset predicted by a second U-Net. This residual provides the necessary degrees of freedom for sliding or inward compression to satisfy physical volume constraints. As this decomposition is inherently ambiguous, resolving the disentanglement is a challenge of our proposed regularization (Sec. 4.3). Finally, the canonical skin mesh $\hat{\mathcal{S}}_{\text{pred}} = \mathcal{S}_{\text{bind}} + \mathbf{D}_{\text{skin}}(\boldsymbol{\theta})$ is transformed into the final posed mesh $\mathcal{S}_{\text{pred}}$ via LBS using pose $\boldsymbol{\theta}$ and weights $\mathbf{W}$, consistent with Eq. 3.

4.2 Muscle and Skin Surface Supervision

We train the model end-to-end by minimizing a combined energy of data supervision and biomechanical priors: $E_{\text{total}} = \lambda_{\text{data}} E_{\text{data}} + E_{\text{reg}}$.

The data term E_{data} supervises the learned corrective field using the ground-truth canonical marker residuals $\boldsymbol{\delta}_{k,t}$ (Sec. 3.5). To predict the corresponding residual $\hat{\boldsymbol{\delta}}_k(\boldsymbol{\theta})$, we interpolate the canonical skin displacement $\mathbf{D}_{\text{skin}}$ using the marker's barycentric coordinates $\mathbf{b}_k$ on its corresponding triangle vertices $\mathbf{v}_{f_j}$:

$$\hat{\boldsymbol{\delta}}_k(\boldsymbol{\theta}) = \sum_{j=1}^{3} b_{k,j} \mathbf{D}_{\text{skin}}(\mathbf{v}_{f_j}, \boldsymbol{\theta}). \tag{7}$$

The data loss is the Mean Squared Error (MSE) between the predicted and ground-truth residuals over the C total markers, masked by their visibility $m_{k,t} \in \{0, 1\}$ at frame t:

$$E_{\text{data}} = \frac{1}{N_t} \sum_{k=1}^{C} m_{k,t} \|\hat{\boldsymbol{\delta}}_k(\boldsymbol{\theta}) - \boldsymbol{\delta}_{k,t}\|^2, \tag{8}$$

where N_t is the number of visible markers. By computing this loss on canonical residuals rather than absolute world positions, the optimization becomes strictly invariant to global pose, forcing the network to focus solely on local non-rigid soft-tissue deformation. By comparing residuals in the canonical frame rather than absolute world positions, the loss becomes invariant to the global pose, allowing the optimization to focus solely on the local non-rigid deformation.

4.3 Bio-mechanically Inspired Deformation Priors

Minimizing E_{data} alone is ill-posed due to the disentanglement ambiguity (Sec. 4). To resolve this, we introduce biomechanical priors enforcing that muscles bulge outwards while maintaining structural stability, and skin behaves as an elastic sheet sliding over the tissue. We formulate these constraints as a combined regularization term E_{reg} applied to both the muscle ($\mathcal{M}$) and skin ($\mathcal{S}$) layers:

$$E_{\text{reg}} = E_{\text{smooth}} + E_{\text{bi}} + E_{\text{stretch}} + E_{\text{tang}} + E_{\text{vol}}. \tag{9}$$

Spatial Smoothness. To ensure smooth deformations, we apply an area-normalized Laplacian regularization. For simplicity, we refer to $\mathbf{D}_{\text{musc}}$ as $\mathbf{D}_{\mathcal{M}}$ and $\mathbf{D}_{\text{res}}$ as $\mathbf{D}_{\mathcal{S}}$:

$$E_{\text{smooth}} = \sum_{l \in \{\mathcal{M}, \mathcal{S}\}} \lambda^{l}_{\text{smooth}} \| \sqrt{\mathbf{M}^{-1}} \mathbf{L}_l \mathbf{D}_l \|^2, \tag{10}$$

where $\mathbf{L}_l$ is the uniform discrete Laplacian operator and $\mathbf{M}$ is the diagonal mass matrix of Voronoi areas, ensuring discretization invariance. We set $\lambda^{\mathcal{S}}_{\text{smooth}} \gg \lambda^{\mathcal{M}}_{\text{smooth}}$ to prevent high-frequency artifacts in the skin profile while allowing the underlying muscle layer to bulge freely.

Bending Resistance. To prevent localized high-frequency artifacts (e.g., isolated vertex spikes) and enforce the wide, organic deformations characteristic of thick soft tissue [22], we apply a second-order Laplacian (biharmonic) regularization. Acting as a bending resistance, this is formulated as:

$$E_{\text{bi}} = \sum_{l \in \{\mathcal{M}, \mathcal{S}\}} \lambda^{l}_{\text{bi}} \| \mathbf{L}_l \mathbf{M}^{-1} (\mathbf{L}_l \mathbf{D}_l) \|^2_{\mathbf{M}^{-1}}. \tag{11}$$

Applying the discrete Laplacian $\mathbf{L}_l$ twice penalizes high-curvature deformations, while the inverse mass matrix $\mathbf{M}^{-1}$ maintains discretization invariance.

Surface Stretching. To preserve structural integrity and prevent unrealistic expansion, we penalize edge length deviations from their canonical rest state $L_{uv,0}$:

$$E_{\text{stretch}} = \sum_{l \in \{\mathcal{M}, \mathcal{S}\}} \lambda^{l}_{\text{stretch}} \sum_{(u,v) \in \mathcal{E}_l} \omega_{uv} (\| \mathbf{v}^{l}_{u} - \mathbf{v}^{l}_{v} \| - L_{uv,0})^2, \tag{12}$$

where $\mathcal{E}_l$ denotes the edges of layer l. The weight $\omega_{uv} = (\text{area}(\mathbf{f}_i) + \text{area}(\mathbf{f}_j))/3$ assigns the area of the incident triangles $\mathbf{f}_{i,j}$ to each edge, normalized such that $\sum \omega_{uv} = 1$ to ensure scale invariance.

Tangential Sliding. To discourage unrealistic sliding, we penalize the tangential component of the displacement relative to the canonical vertex normal $\mathbf{n}_v$:

$$E_{\text{tang}} = \sum_{l \in \{\mathcal{M}, \mathcal{S}\}} \lambda^{l}_{\text{tang}} \sum_{v \in \mathcal{V}_l} A_v \| \mathbf{D}_{l,v} - (\mathbf{D}_{l,v} \cdot \mathbf{n}_v) \mathbf{n}_v \|^2, \tag{13}$$

where A_v is the local vertex area. By enforcing $\lambda^{\mathcal{M}}_{\text{tang}} \gg \lambda^{\mathcal{S}}_{\text{tang}}$, we ensure muscles remain structurally anchored (deforming primarily outwards), while allowing the skin to slide more freely over the underlying tissue during articulation.

Soft Tissue Volume Preservation. To mimic the incompressibility of human soft tissue [58], we enforce volume preservation on the subcutaneous layer (skin-to-muscle, $\mathcal{S}$) and deep tissue (muscle-to-bone, $\mathcal{M}$). Modeling the space between

these boundaries as sets of volumetric prisms $\mathcal{P}_l$, we penalize deviations from their rest volume p_0:

$$E_{\text{vol}} = \sum_{l \in \{\mathcal{M}, \mathcal{S}\}} \lambda_{\text{vol}}^{l} \frac{1}{|\mathcal{P}_l^*|} \sum_{p \in \mathcal{P}_l^*} \frac{\big(\text{Vol}(p) - \text{Vol}(p_0)\big)^2}{|\text{Vol}(p_0)| + \epsilon}, \tag{14}$$

where we evaluate only a valid subset of non-degenerate prisms $\mathcal{P}_l^*$ for numerical stability, and ϵ prevents singularities in extremely thin regions. These constraints operate in different spaces: the volume ($\lambda_{\text{vol}}^{\mathcal{S}}$) is evaluated in the *canonical space* and the muscle volume ($\lambda_{\text{vol}}^{\mathcal{M}}$) in the *posed space*, forcing the network to learn canonical muscle bulges that physically counteract LBS compression. We compute exact signed volumes via 2-point Gauss-Legendre Quadrature. By evaluating signed rather than absolute volume, geometric inversions (layer intersections) naturally yield negative volumes and massive energy penalties, inherently preventing structural collapse. The mathematical derivation can be found in the Supplementary Material.

Finally, to animate the high-resolution individual muscles, we propagate the learned canonical displacements from the unified boundary layer $\mathcal{M}_{\text{bind}}$ to the underlying muscle meshes via a pre-computed barycentric binding mechanism, ensuring physical synchrony. The full binding and deformation mechanism is detailed in the Supplementary Material.

5 Experiments and Results

We evaluate our muscle-driven deformation model against kinematic baselines, volumetric estimators, and ablated variations. Our analysis assesses three aspects: (1) **Data Fidelity** (tracking accuracy of ground-truth sparse markers); (2) **Anatomical Plausibility** (soft tissue volume preservation and layer interpenetration); and (3) **Geometric Quality** (surface smoothness and structural integrity).

5.1 Data and Experimental Setup

We evaluate our model using time-synchronized skeletal poses and tracked markers from our dataset (Sec. 3), testing generalization on unseen poses. Because acquiring *in-vivo* volumetric ground truth during dynamic motion is infeasible, we validate physical plausibility using marker tracking and physics-inspired metrics. Specifically, we evaluate Global Surface Tracking (MPME, MedPME, P90), Dynamic Region Error (DRE) for isolating complex soft-tissue dynamics, Intersection Ratio for collision avoidance, and Volume Preservation for isochoric plausibility. We compare our explicit physical formulation against a geometric kinematic lower-bound (LBS applied to the static scan $\mathcal{S}_{\text{bind}}$) and a state-of-the-art uncoupled volumetric predictor (HIT [27] via SMPL [39]). Detailed metric definitions and baseline configurations are provided in the Supplementary Material.

Table 2: Quantitative evaluation on unseen motion sequences. Internal anatomical plausibility is evaluated for the subcutaneous fat volume ($\mathcal{S} \rightarrow \mathcal{M}$) and muscle volume ($\mathcal{M} \rightarrow \mathcal{B}$). LBS reports no volume skin change as it does not apply pose-correctives.

	Global Tracking (mm) ↓			**DRE (mm)** ↓			**Int. (%)** ↓		**Δ Vol (%)** ↓	
Method	MPME	MedPME	P90	> 15	> 20	> 25	$\mathcal{S} \rightarrow \mathcal{M}$	$\mathcal{M} \rightarrow \mathcal{B}$	$\mathcal{S} \rightarrow \mathcal{M}$	$\mathcal{M} \rightarrow \mathcal{B}$
LBS	**13.11**	15.64	26.03	40.46	88.25	134.28	0.85	0.93	N/A	18.81
HIT [27]	115.34	50.15	354.80	**24.09**	119.72	151.65	7.61	N/A	2.99	**1.95**
Ours	13.43	**11.51**	**19.20**	25.51	**46.99**	**64.52**	**0.85**	**0.93**	**1.37**	10.35

5.2 Quantitative Evaluation

Quantitative results on unseen validation sequences (see Tab. 2) demonstrate our dual-layer formulation balances high-fidelity tracking with anatomical constraints. While standard LBS yields a competitive global mean error (MPME) as rigid body parts heavily dilute the metric, it fundamentally fails to preserve internal structures, exhibiting volumetric joint collapse (*candy-wrapper* effect). Conversely, off-the-shelf estimators like HIT [27] rely on generic templates lacking subject-specific proportions, resulting in high tracking errors.

Our method, SOMA, resolves these issues. By capturing non-linear muscle bulging and preventing joint collapse, it improves global tracking (MedPME) and significantly reduces extreme errors (P90). Notably, in highly dynamic soft-tissue regions ($\tau > 25$ mm) where LBS degrades, our formulation cuts tracking error by more than half. Crucially, SOMA guarantees internal physical plausibility. Standard LBS exhibits a falsely low intersection ratio (0.86%) only because skin and deep tissue share similar skinning weights and collapse together. HIT attempts to restore volume but lacks coupled constraints, causing severe anatomical collisions (7.61%). Only our explicitly coupled formulation successfully restores isochoric volume while strictly avoiding interpenetration.

5.3 Qualitative Evaluation

We provide a real-time visualization tool built in Viser [67] (as shown in Fig. 7). By importing the trained model and subject bind meshes ($\mathcal{M}_{\text{bind}}$, $\mathcal{S}_{\text{bind}}$), the tool enables interactive full-body inference, per-muscle querying, and isolated single-muscle prediction using only locally associated markers, therefore capturing active muscle bulging (e.g., quadriceps, triceps) during motion. We also compare dynamic deformations against HIT [27] (see Fig. 6). While static baselines like these produce plausible resting volumes, they over-smooth high-frequency dynamics. For more example motions, please see the Supplementary Material.

5.4 Ablations

We ablate our architectural choices and loss components across three categories: **network architecture**, **vector regularization** (on displacement fields), and **physical energies** (on surface geometry).

Fig. 6: Qualitative Comparison against HIT [27]. Our explicit model (right) predicts finer-grained dynamic deformations. For cross-topology evaluation (scan vs. SMPL), we map our marker representation onto HIT's meshes via barycentric interpolation.

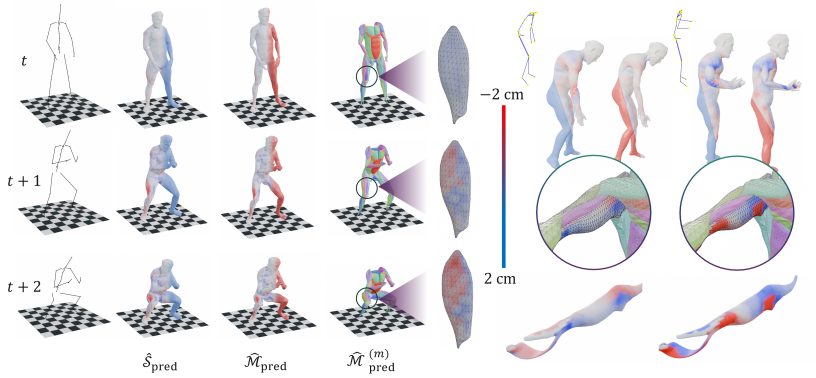

Fig. 7: From skin to individual muscle deformation. Given a target pose $\boldsymbol{\theta}$, we visualize the inferred non-linear displacements of the skin and muscle layers (blue: inward compression, red: outward expansion).

Network Architecture. We compare Linear Blendshapes, an MLP, and a U-Net for predicting canonical displacements in Tab. 3. Although similar, linear models struggle more with non-linear soft-tissue bulging and lack the spatial weight-sharing needed to enforce global volume constraints. Though MLPs add non-linearities, the U-Net best captures both local surface details and spatial deformations without overfitting the sparse marker data.

Vector Regularization (E_{smooth}, E_{tan}). Removing these terms (*w/o Vector*) surprisingly degrades subcutaneous volume preservation (see Tab. 4). Without tangential constraints, vertices drift laterally to minimize marker error, shearing the underlying prisms. This prevents E_{vol} from enforcing outward biological bulging and slightly degrades global tracking.

Physical Energies ($E_{bi}, E_{stretch}, E_{vol}$). Our *Full Model* achieves the lowest tracking error (13.43 mm), demonstrating that these geometric priors act as structural anchors guiding the network to better local minima. Removing all physical constraints (*w/o Physics*) yields the worst tracking and high volume distortion. Ablating structural tension (*w/o* $E_{bi}, E_{stretch}$) causes the highest volume errors,

Table 3: Ablation of network architectures. While the U-Net achieves the lowest error, Linear and MLPs are sufficient to effectively fit the markers. This demonstrates the significant benefit of training person-specific models on high-quality tracking data, while highlighting the inherent anatomical misalignment issues faced by generic templates.

	Global ↓	**DRE ↓**	**Int. (%)**		**Δ Vol (%)**	
Architecture	MPME	> 25	$\mathcal{S} \to \mathcal{M}$	$\mathcal{M} \to \mathcal{B}$	$\mathcal{S} \to \mathcal{M}$	$\mathcal{M} \to \mathcal{B}$
Linear	16.52	64.58	0.85	0.93	3.57	11.02
MLP	14.03	64.93	0.86	1.09	5.74	11.05
U-Net	**13.43**	**64.52**	**0.85**	**0.93**	**1.37**	**10.35**

Table 4: Ablation of Regularization Terms. While all priors significantly improve anatomical stability, the smoothness and stretch energies ($E_{\text{bi, str}}$) prove most critical. Removing them causes the most severe volumetric distortion, highlighting that membrane tension is essential to prevent mesh artifacts during dynamic motion.

	Global ↓	**DRE ↓**	**Int. (%)**		**Δ Vol (%)**	
Method	MPME	> 25	$\mathcal{S} \to \mathcal{M}$	$\mathcal{M} \to \mathcal{B}$	$\mathcal{S} \to \mathcal{M}$	$\mathcal{M} \to \mathcal{B}$
w/o Physics	13.71	64.93	0.87	0.95	9.37	13.54
w/o Vector	13.46	64.57	0.87	0.94	8.27	12.63
w/o $E_{\text{bi, str}}$	13.54	64.48	0.88	0.94	10.54	13.76
w/o E_{vol}	13.67	64.57	0.87	0.97	8.28	11.87
Full (Ours)	**13.43**	**64.52**	**0.85**	**0.93**	**1.37**	**10.35**

as the mesh artificially satisfies volume constraints via severe high-frequency wrinkling instead of smooth bulging. Ablating E_{vol} spikes subcutaneous volume error to 8.28%, confirming its necessity in correcting LBS artifacts. Finally, while the full formulation reduces outer fat volume error to 1.37%, the deep muscle layer ($\mathcal{M} \to \mathcal{B}$) retains a 10.35% deviation. This represents a realistic physical compromise: the optimizer prioritizes strict non-intersection over full deep-tissue inflation to prevent massive skeletal collisions.

6 Limitations and Future Work

While our proposed method is among the first to recover interior muscle deformations solely from visual data, it still has some limitations. First, we rely on volumetric fitting [33] for the initial template, which makes us vulnerable to initialization errors; future work could jointly optimize baseline anatomy and dynamic displacements. Second, requiring a marker suit restricts convenience. Leveraging our data to train robust, markerless skin trackers would be a promising alternative. Third, force- and co-contraction-driven variation at a fixed joint angle constitutes a remaining unmodeled component. Finally, our method is currently subject-specific. Expanding the dataset to build a more generalizable parametric model offers an exciting path toward in-the-wild anatomically accurate performance capture.

7 Conclusion

We introduced SOMA, a novel approach that models skin and muscle geometry driven by skeletal motion. To achieve this, we developed a capture setup for high-fidelity spatio-temporal surface tracking. As learning internal anatomy from visual data is inherently ill-posed, our blendshapes formulation leverages biomechanically-inspired priors. At inference, SOMA animates both layers using only user-defined skeletal poses. We believe this work is a crucial step toward accessible, data-driven biomechanical digital humans, inspiring future work in the directions of human surface correspondence estimation and person-agnostic anatomical human models.

8 Acknowledgements

This research has been funded by the Ministry of Culture and Science of the State of North Rhine-Westphalia through the project inVirtuo 4.0 (PB22-063-B) and by the Federal Ministry of Education and Research of Germany and the state of North Rhine-Westphalia as part of the Lamarr Institute for Machine Learning and Artificial Intelligence.

References

1. Captury – markerless motion capture technology. https://captury.com/, accessed: 2025-12-30
2. Treedys – body scanner. https://github.com/treedys/tps, accessed: 2020-12-30
3. Achenbach, J., Brylka, R., Gietzen, T., zum Hebel, K., Schömer, E., Schulze, R., Botsch, M., Schwanecke, U.: A Multilinear Model for Bidirectional Craniofacial Reconstruction. The Eurographics Association (2018)
4. Achenbach, J., Waltemate, T., Latoschik, M.E., Botsch, M.: Fast generation of realistic virtual humans. In: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology. pp. 1–10. VRST '17, Association for Computing Machinery, New York, NY, USA (Nov 2017). https://doi.org/10.1145/3139131.3139154
5. Achenbach, J., Waltemate, T., Latoschik, M.E., Botsch, M.: Fast generation of realistic virtual humans. In: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology. VRST '17, Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3139131.3139154, https://doi.org/10.1145/3139131.3139154
6. Anguelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., Davis, J.: SCAPE: Shape completion and animation of people. ACM Trans. Graph. **24**(3), 408–416 (Jul 2005). https://doi.org/10.1145/1073204.1073207
7. Blemker, S., Delp, S.: Three-Dimensional Representation of Complex Muscle Architectures and Geometries1. Annals of Biomedical Engineering **33** (Aug 2005). https://doi.org/10.1007/s10439-005-7385-0
8. Bogo, F., Romero, J., Pons-Moll, G., Black, M.J.: Dynamic FAUST: Registering human bodies in motion. In: IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (Jul 2017)
9. Botsch, M.: Skeletal-driven animation of anatomical humans via neural deformation gradients (2026)
10. Bouaziz, S., Martin, S., Liu, T., Kavan, L., Pauly, M.: Projective dynamics: Fusing constraint projections for fast simulation. ACM Trans. Graph. **33**(4), 154:1–154:11 (Jul 2014). https://doi.org/10.1145/2601097.2601116
11. Chen, H., Park, H., Macit, K., Kavan, L.: Capturing detailed deformations of moving human bodies. ACM Trans. Graph. **40**(4) (Jul 2021). https://doi.org/10.1145/3450626.3459792, https://doi.org/10.1145/3450626.3459792
12. Chiquier, M., Vondrick, C.: Muscles in Action (Mar 2023). https://doi.org/10.48550/arXiv.2212.02978
13. Garrido-Jurado, S., Muñoz-Salinas, R., Madrid-Cuevas, F.J., Marín-Jiménez, M.J.: Automatic generation and detection of highly reliable fiducial markers under occlusion. In: Pattern Recognition. Lecture Notes in Computer Science, vol. 8827, pp. 113–122. Springer (2014). https://doi.org/10.1007/978-3-319-16865-4_10
14. Gilles, B., Revéret, L., Pai, D.K.: Creating and Animating Subject-Specific Anatomical Models. Computer Graphics Forum **29**(8), 2340–2351 (2010). https://doi.org/10.1111/j.1467-8659.2010.01718.x
15. Habermann, M., Xu, W., Zollhoefer, M., Pons-Moll, G., Theobalt, C.: LiveCap: Real-time Human Performance Capture from Monocular Video (Jan 2019). https://doi.org/10.48550/arXiv.1810.02648
16. Habermann, M., Xu, W., Zollhoefer, M., Pons-Moll, G., Theobalt, C.: DeepCap: Monocular Human Performance Capture Using Weak Supervision (Mar 2020). https://doi.org/10.48550/arXiv.2003.08325

17. Han, Y., Chen, Y., Ong, C., Chen, J., Hicks, J., Teran, J.: A Neural Network Model for Efficient Musculoskeletal-Driven Skin Deformation. ACM Trans. Graph. **43**(4), 118:1–118:12 (Jul 2024). https://doi.org/10.1145/3658135
18. Hasler, N., Stoll, C., Sunkel, M., Rosenhahn, B., Seidel, H.P.: A Statistical Model of Human Pose and Body Shape. Computer Graphics Forum **28**(2), 337–346 (2009). https://doi.org/10.1111/j.1467-8659.2009.01373.x
19. Hicks, J.L., Uchida, T.K., Seth, A., Rajagopal, A., Delp, S.L.: Is my model good enough? Best practices for verification and validation of musculoskeletal models and simulations of movement. Journal of Biomechanical Engineering **137**(2), 020905 (Feb 2015). https://doi.org/10.1115/1.4029304
20. Hieu, N.Q., Thai Hoang, D., Nguyen, D.N., Abu Alsheikh, M.: Reconstructing Human Pose From Inertial Measurements: A Generative Model-Based Compressive Sensing Approach. IEEE Journal on Selected Areas in Communications **42**(10), 2674–2687 (Oct 2024). https://doi.org/10.1109/JSAC.2024.3414604
21. Ichim, A.E., Kavan, L., Nimier-David, M., Pauly, M.: Building and animating user-specific volumetric face rigs. In: Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation. pp. 107–117. SCA '16, Eurographics Association, Goslar, DEU (Jul 2016)
22. Jacobson, A., Baran, I., Popović, J., Sorkine-Hornung, O.: Bounded biharmonic weights for real-time deformation. Commun. ACM **57**(4), 99–106 (Apr 2014). https://doi.org/10.1145/2578850, https://doi.org/10.1145/2578850
23. Kadleček, P., Ichim, A.E., Liu, T., Křivánek, J., Kavan, L.: Reconstructing personalized anatomical models for physics-based body animation. ACM Transactions on Graphics **35**(6), 213:1–213:13 (Dec 2016). https://doi.org/10.1145/2980179.2982438
24. Kadleček, P., Kavan, L.: Building accurate physics-based face models from data. Proc. ACM Comput. Graph. Interact. Tech. **2**(2) (Jul 2019). https://doi.org/10.1145/3340256, https://doi.org/10.1145/3340256
25. Kanazawa, A., Black, M.J., Jacobs, D.W., Malik, J.: End-to-End Recovery of Human Shape and Pose. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7122–7131 (Jun 2018). https://doi.org/10.1109/CVPR.2018.00744
26. Kavan, L., Faure, F., Palombi, O., Cani, M.: Anatomy Transfer. ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH Asia **32**(6) (2013). https://doi.org/10.1145/2508363.2508415
27. Keller, M., Aora, V., Abdelmouttaleb, D.: HIT: Estimating Internal Human Implicit Tissues from the Body Surface (2024)
28. Keller, M., Werling, K., Shin, S., Delp, S., Pujades, S., C. Karen, L., Black, M.J.: From Skin to Skeleton: Towards Biomechanically Accurate 3D Digital Humans (2023)
29. Keller, M., Zuffi, S., Black, M.J., Pujades, S.: OSSO: Obtaining Skeletal Shape from Outside (Apr 2022). https://doi.org/10.48550/arXiv.2204.10129
30. Kocabas, M., Athanasiou, N., Black, M.J.: VIBE: Video Inference for Human Body Pose and Shape Estimation (Apr 2020). https://doi.org/10.48550/arXiv.1912.05656
31. Kolotouros, N., Pavlakos, G., Black, M.J., Daniilidis, K.: Learning to Reconstruct 3D Human Pose and Shape via Model-fitting in the Loop (Sep 2019). https://doi.org/10.48550/arXiv.1909.12828
32. Komaritzan, M., Botsch, M.: Fast Projective Skinning. In: Proceedings of the 12th ACM SIGGRAPH Conference on Motion, Interaction and Games. pp. 1–10. MIG '19, Association for Computing Machinery, New York, NY, USA (Oct 2019). https://doi.org/10.1145/3359566.3360073

33. Komaritzan, M., Wenninger, S., Botsch, M.: Inside Humans: Creating a Simple Layered Anatomical Model from Human Surface Scans. Frontiers in Virtual Reality **2** (2021)
34. Lewis, J.P., Cordner, M., Fong, N.: Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques. p. 165–172. SIGGRAPH '00, ACM Press/Addison-Wesley Publishing Co., USA (2000). https://doi.org/10.1145/344779.344862, https://doi.org/10.1145/344779.344862
35. Li, B., Yang, L., Solenthaler, B.: Efficient Incremental Potential Contact for Actuated Face Simulation. In: SIGGRAPH Asia 2023 Technical Communications. pp. 1–4. SA '23, Association for Computing Machinery, New York, NY, USA (Nov 2023). https://doi.org/10.1145/3610543.3626161
36. Li, T., Bolkart, T., Black, M.J., Li, H., Romero, J.: Learning a model of facial shape and expression from 4D scans. ACM Trans. Graph. **36**(6), 194:1–194:17 (Nov 2017). https://doi.org/10.1145/3130800.3130813
37. Li, Y., Zhang, L., Qiu, Z., Jiang, Y., Li, N., Ma, Y., Zhang, Y., Xu, L., Yu, J.: NIMBLE: A Non-rigid Hand Model with Bones and Muscles. pp. 1788–1796 (Oct 2022). https://doi.org/10.1145/3503161.3548148
38. Liao, Z., Golyanik, V., Habermann, M., Theobalt, C.: VINECS: Video-based Neural Character Skinning (Jul 2023). https://doi.org/10.48550/arXiv.2307.00842
39. Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., Black, M.J.: SMPL: A skinned multi-person linear model. ACM Trans. Graph. **34**(6), 248:1–248:16 (Oct 2015). https://doi.org/10.1145/2816795.2818013
40. Maalin, N., Mohamed, S., Kramer, R.S.S., Cornelissen, P.L., Martin, D., Tovée, M.J.: Beyond BMI for self-estimates of body size and shape: A new method for developing stimuli correctly calibrated for body composition. Behavior Research Methods **53**(3), 1308–1321 (Jun 2021). https://doi.org/10.3758/s13428-020-01494-1
41. Mahmood, N., Ghorbani, N., Troje, N.F., Pons-Moll, G., Black, M.J.: AMASS: Archive of motion capture as surface shapes. In: International Conference on Computer Vision. pp. 5442–5451 (Oct 2019)
42. Modi, V., Fulton, L., Jacobson, A., Sueda, S., Levin, D.I.W.: EMU: Efficient Muscle Simulation in Deformation Space. COMPUTER GRAPHICS forum **40**(1), 234–248 (2021). https://doi.org/10.1111/cgf.14185
43. Murai, H., Hong, Q.Y., Yamane, K., Hodgins, J.: Dynamic Skin Deformation Simulation Using Musculoskeletal Model and Soft Tissue Dynamics. In: Disney Research (2016)
44. Neumann, T., Varanasi, K., Hasler, N., Wacker, M., Magnor, M., Theobalt, C.: Capture and Statistical Modeling of Arm-Muscle Deformations. Computer Graphics Forum **32**(2pt3), 285–294 (2013). https://doi.org/10.1111/cgf.12048
45. Nicolas, W., Ulrich, S., Mario, B.: SparseSoftDECA — Efficient high-resolution physics-based facial animation from sparse landmarks. Computers & Graphics **119**, 103903 (Apr 2024). https://doi.org/10.1016/j.cag.2024.103903
46. Osman, A.A.A., Bolkart, T., Black, M.J.: STAR: Sparse Trained Articulated Human Body Regressor. vol. 12351, pp. 598–613 (2020). https://doi.org/10.1007/978-3-030-58539-6_36
47. Park, J., Jung, E., Lee, J., Won, J.: [SIGGRAPH 2025] MAGNET: Muscle Activation Generation Networks for Diverse Human Movement (Jun 2025)
48. Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A.A.A., Tzionas, D., Black, M.J.: Expressive body capture: 3d hands, face, and body from a single image. In: Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) (2019)

49. Qiu, Z., Li, Y., He, D., Zhang, Q., Zhang, L., Zhang, Y., Wang, J., Xu, L., Wang, X., Zhang, Y., Yu, J.: SCULPTOR: Skeleton-Consistent Face Creation Using a Learned Parametric Generator (Dec 2022). https://doi.org/10.48550/arXiv.2209.06423
50. Ramon, P., Romero, C., Tapia, J., Otaduy, M.A.: Sflsh: Shape-dependent soft-flesh avatars. In: SIGGRAPH Asia Conference Papers (SA Conference Papers '23). ACM (2023). https://doi.org/10.1145/3610548.3618242
51. Romero, J., Tzionas, D., Black, M.J.: Embodied hands: Modeling and capturing hands and bodies together. ACM Trans. Graph. **36**(6), 245:1–245:17 (Nov 2017). https://doi.org/10.1145/3130800.3130883
52. Saito, S., Zhou, Z.Y., Kavan, L.: Computational bodybuilding: Anatomically-based modeling of human bodies. ACM Transactions on Graphics **34**(4), 41:1–41:12 (Jul 2015). https://doi.org/10.1145/2766957
53. Santesteban, I., Garces, E., Otaduy, M.A., Casas, D.: Softsmpl: Data-driven modeling of nonlinear soft-tissue dynamics for parametric humans. Computer Graphics Forum **39**(2), 65–75 (2020). https://doi.org/https://doi.org/10.1111/cgf.13912, https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.13912
54. Schleicher, R., Nitschke, M., Martschinke, J., Stamminger, M., Eskofier, B., Klucken, J., Koelewijn, A.: BASH: Biomechanical Animated Skinned Human for Visualization of Kinematics and Muscle Activity. Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (2021)
55. Schneider, D., Reiß, S., Kugler, M., Jaus, A., Peng, K., Sutschet, S., Sarfraz, M.S., Matthiesen, S., Stiefelhagen, R.: Muscles in Time: Learning to Understand Human Motion by Simulating Muscle Activations (Oct 2024). https://doi.org/10.48550/arXiv.2411.00128
56. Seth, A., Sherman, M., Reinbolt, J.A., Delp, S.L.: OpenSim: A musculoskeletal modeling and simulation framework for in silico investigations and exchange. Procedia IUTAM **2**, 212–232 (Jan 2011). https://doi.org/10.1016/J.PIUTAM.2011.04.021
57. Shetty, A., Habermann, M., Sun, G., Luvizon, D., Golyanik, V., Theobalt, C.: Holoported Characters: Real-time Free-viewpoint Rendering of Humans from Sparse RGB Cameras (Jul 2024). https://doi.org/10.48550/arXiv.2312.07423
58. Sleboda, D.A., Roberts, T.J.: Incompressible fluid plays a mechanical role in the development of passive muscle tension. Biology Letters **13**(1), 20160630 (01 2017). https://doi.org/10.1098/rsbl.2016.0630, https://doi.org/10.1098/rsbl.2016.0630
59. Smith, B., Goes, F.D., Kim, T.: Stable Neo-Hookean Flesh Simulation. ACM Trans. Graph. **37**(2), 12:1–12:15 (Mar 2018). https://doi.org/10.1145/3180491
60. Wang, B., Matcuk, G., Barbič, J.: Hand MRI dataset (2020), http://www.jernejbarbic.com/hand-mri-dataset
61. Wang, Y., Han, Q., Habermann, M., Daniilidis, K., Theobalt, C., Liu, L.: NeuS2: Fast Learning of Neural Implicit Surfaces for Multi-view Reconstruction (Nov 2023). https://doi.org/10.48550/arXiv.2212.05231
62. Wenninger, S., Kemper, F., Schwanecke, U., Botsch, M.: TailorMe: Self-Supervised Learning of an Anatomically Constrained Volumetric Human Shape Model. Computer Graphics Forum **43**(2), e15046 (2024). https://doi.org/10.1111/cgf.15046
63. Xia, Y., Zhou, X., Vouga, E., Huang, Q., Pavlakos, G.: Reconstructing Humans with a Biomechanically Accurate Skeleton (Mar 2025). https://doi.org/10.48550/arXiv.2503.21751
64. Xiao, X., Wang, J., Feng, P., Gong, A., Zhang, X., Zhang, J.: Fast Human Motion reconstruction from sparse inertial measurement units considering the human shape.

Nature Communications **15**(1), 2423 (Mar 2024). https://doi.org/10.1038/s41467-024-46662-5

65. Xu, H., Bazavan, E.G., Zanfir, A., Freeman, W.T., Sukthankar, R., Sminchisescu, C.: GHUM & GHUML: Generative 3D Human Shape and Articulated Pose Models. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 6183–6192 (Jun 2020). https://doi.org/10.1109/CVPR42600.2020.00622
66. Yang, L., Zoss, G., Chandran, P., Gross, M., Solenthaler, B., Sifakis, E., Bradley, D.: Learning a Generalized Physical Face Model From Data. ACM Trans. Graph. **43**(4), 94:1–94:14 (Jul 2024). https://doi.org/10.1145/3658189
67. Yi, B., Kim, C.M., Kerr, J., Wu, G., Feng, R., Zhang, A., Kulhanek, J., Choi, H., Ma, Y., Tancik, M., Kanazawa, A.: Viser: Imperative, web-based 3d visualization in python. arXiv preprint arXiv:2507.22885 (2025)
68. Zhou, Y., Barnes, C., Lu, J., Yang, J., Li, H.: On the continuity of rotation representations in neural networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5745–5753 (2019)
69. Zhu, H., Zhan, F., Theobalt, C., Habermann, M.: TriHuman: A Real-time and Controllable Tri-plane Representation for Detailed Human Geometry and Appearance Synthesis. ACM Trans. Graph. **44**(1), 4:1–4:17 (Oct 2024). https://doi.org/10.1145/3697140