



HAL
open science

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

Marilyn Keller, Keenon Werling, Soyong Shin, Scott Delp, Sergi Pujades, C.
Karen Liu, Michael J Black

► **To cite this version:**

Marilyn Keller, Keenon Werling, Soyong Shin, Scott Delp, Sergi Pujades, et al.. From Skin to Skeleton: Towards Biomechanically Accurate 3D Digital Humans. ACM Transactions on Graphics, 2023, 42 (6), pp.1-12. 10.1145/3618381 . hal-04396607

HAL Id: hal-04396607

<https://inria.hal.science/hal-04396607>

Submitted on 16 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

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

MARILYN KELLER, Max Planck Institute for Intelligent Systems, Germany

KEENON WERLING, Stanford University, USA

SOYONG SHIN, Carnegie Mellon University, USA

SCOTT DELP, Stanford University, USA

SERGI PUJADES, Inria centre at the University Grenoble Alpes, France

C. KAREN LIU, Stanford University, USA

MICHAEL J. BLACK, Max Planck Institute for Intelligent Systems, Germany



Fig. 1. (a) We fit our new Biomechanical Skeleton Model, BSM, to SMPL [Loper et al. 2015] mesh sequences from AMASS [Mahmood et al. 2019]. This gives paired data enabling us to learn the mapping from skin to skeleton. (b) We use this to create SKEL, a parametric body model with skin and skeleton meshes, driven by biomechanical pose parameters and incorporating the shape space of SMPL. SKEL is like SMPL but with more realistic degrees of freedom. Fitting SKEL to DFAUST scans [Bogo et al. 2017] results in SKEL’s scapula sliding (c) and the forearms twisting appropriately (d).

Great progress has been made in estimating 3D human pose and shape from images and video by training neural networks to directly regress the parameters of parametric human models like SMPL. However, existing body models have simplified kinematic structures that do not correspond to the true joint locations and articulations in the human skeletal system, limiting their potential use in biomechanics. On the other hand, methods for estimating biomechanically accurate skeletal motion typically rely on complex motion capture systems and expensive optimization methods. What is needed is a parametric 3D human model with a biomechanically accurate skeletal structure that can be easily posed. To that end, we develop SKEL, which re-rigs the SMPL body model with a biomechanics skeleton. To enable this, we need training data of skeletons inside SMPL meshes in diverse poses.

Authors’ addresses: Marilyn Keller, Max Planck Institute for Intelligent Systems, Max Planck Ring 4, Tübingen, 72076, Germany, marilyn.keller@tuebingen.mpg.de; Keenon Werling, Stanford University, Stanford, CA, USA; Soyong Shin, Carnegie Mellon University, Pittsburgh, PA, USA; Scott Delp, Stanford University, Stanford, CA, USA; Sergi Pujades, Inria centre at the University Grenoble Alpes, Grenoble, France; C. Karen Liu, Stanford University, Stanford, CA, USA; Michael J. Black, Max Planck Institute for Intelligent Systems, Tübingen, Germany.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

© 2023 Copyright held by the owner/author(s).

0730-0301/2023/12-ART553

<https://doi.org/10.1145/3618381>

We build such a dataset by optimizing biomechanically accurate skeletons inside SMPL meshes from AMASS sequences. We then learn a regressor from SMPL mesh vertices to the optimized joint locations and bone rotations. Finally, we re-parametrize the SMPL mesh with the new kinematic parameters. The resulting SKEL model is animatable like SMPL but with fewer, and biomechanically-realistic, degrees of freedom. We show that SKEL has more biomechanically accurate joint locations than SMPL, and the bones fit inside the body surface better than previous methods. By fitting SKEL to SMPL meshes we are able to “upgrade” existing human pose and shape datasets to include biomechanical parameters. SKEL provides a new tool to enable biomechanics in the wild, while also providing vision and graphics researchers with a better constrained and more realistic model of human articulation. The model, code, and data are available for research at <https://skel.is.tue.mpg.de>.

ACM Reference Format:

Marilyn Keller, Keenon Werling, Soyong Shin, Scott Delp, Sergi Pujades, C. Karen Liu, and Michael J. Black. 2023. From Skin to Skeleton: Towards Biomechanically Accurate 3D Digital Humans. *ACM Trans. Graph.* 42, 6, Article 553 (December 2023), 12 pages. <https://doi.org/10.1145/3618381>

1 INTRODUCTION

Human motion is captured, modeled and studied in diverse fields, including computer vision, graphics, gaming, biomechanics, medicine, ergonomics and more. The tools and representations used, however, vary significantly. Vision and graphics methods often represent the

articulated body pose using an approximate 3D skeleton, whereas, in biomechanics and sports medicine, an accurate kinematic skeleton is of paramount importance for disease diagnosis. The capture methods also vary significantly. Computer vision focuses on estimating 3D humans from images and videos while the biomechanics community focuses on highly accurate marker-based motion capture (mocap) systems. This paper takes a step towards combining the best of these disciplines, providing new and improved tools to each; see Fig. 1.

Specifically, we focus on advances in computer vision that infer the 3D pose and shape of the human body in the form of parametric body models like SMPL [Loper et al. 2015]. The field has advanced rapidly and the accuracy of markerless video-based 3D motion capture is catching up with marker-based techniques. Unfortunately, the kinematic structure of models such as SMPL is not physically accurate, limiting applicability in biomechanics. On the other hand, the biomechanics field has developed detailed skeletal models to represent the anatomic motion of the knee, spine, shoulder, etc. The vision and graphics communities are currently not benefiting from these more accurate models of the body and its joints.

To address these issues, we unify the SMPL body model with BSM, a new Biomechanical Skeleton Model. While previous work has addressed the problem of putting skeletons inside 3D body models [Ali-Hamadi et al. 2013; Kelc 2012; Keller et al. 2022; Shetty et al. 2023], such approaches have not addressed the problem of precisely locating the skeleton within a *moving* body. The key challenge is the lack of training data that pairs the posed 3D human body shape with the ground-truth skeleton. We address this by creating a novel dataset called BioAMASS. To create BioAMASS, we take sequences of 3D bodies from the AMASS dataset [Mahmood et al. 2019] that cover a wide range of body shapes and challenging poses. To obtain pseudo-ground-truth skeletons, we place virtual motion capture markers on the body surface. We then use the recent method, AddBiomechanics [Werling et al. 2022], to solve for the BSM skeleton given the virtual markers.

With this paired dataset, we can now solve several problems that were previously impossible. First, we train a regressor to estimate the 3D anatomical BSM joint locations of the body given a posed SMPL mesh. Note that these locations significantly differ from the joints in SMPL. This is useful for generating more relevant training data for 2D or 3D joint detectors, as today, such methods are typically trained from manually labeled joints or projected SMPL joints.

Next, we re-rig the SMPL body model with the BSM biomechanical skeleton, i.e. we use the BSM parameters to drive a SMPL mesh, and we call the resulting model SKEL, which is short for “Skeletal Kinematics Enveloped by a Learned body model”. To do so, it is critical that the skeleton is properly scaled, located and oriented inside the SMPL body mesh. To that end, we propose a data-driven strategy that places the bones inside the body while ensuring that their orientations are compatible with the anatomic constraints of the limbs. Like SMPL, SKEL provides a body surface but with a skeleton inside that has biomechanical degrees of freedom. For example, the spine in SKEL is modeled by a spline derived from biomechanics. Additionally, shoulders are a complicated structure that is typically crudely approximated in vision and graphics models. SKEL replaces the approximate shoulder of SMPL with a biomechanical shoulder

blade [Seth et al. 2016] that slides along an ellipsoid defined around the thorax. The forearm rotation is another place where standard graphics models like SMPL differ from biomechanics. Instead of a simple rotation around the elbow, SKEL models the motion of the radius and ulna bones to drive forearm pronation and supination.

SKEL has several uses. Specifically, we consider the problem of taking a SMPL body model and computing the correct skeleton inside. To do so, we simply fit SKEL to the posed SMPL mesh by optimizing the SKEL pose to minimize the vertex-to-vertex distance between the meshes. We apply this process to archival datasets such as 3DPW [Von Marcard et al. 2018] and BEDLAM [Black et al. 2023]. This effectively *upgrades* existing computer vision datasets to contain biomechanical ground truth, extending their use to biomechanics. For example, one could evaluate, or learn to directly regress biomechanical parameters from video.

We evaluate two methods for estimating the skeleton from SMPL: direct regression of BSM joints from SMPL and fitting SKEL to SMPL. Accuracy is defined in terms of 3D joint location error. Since there is no ground-truth for this task, we take the joint locations estimated by AddBiomechanics as pseudo-ground-truth. We find that both of our methods produce significantly more accurate joint predictions than SMPL. We also provide extensive qualitative experiments that show the articulated structure of SKEL and its use in upgrading existing human motion datasets to support biomechanics.

SKEL can also be used in the other direction. Given an input skeleton mesh obtained after fitting a biomechanical model to mocap data, SKEL can be used to add a plausible skin surface; this is useful for visualization of mocap data. Since there are an infinite number of body shapes that are consistent with a given skeleton, the predicted shape can be constrained, e.g. with the subject’s weight.

To the best of our knowledge, SKEL is the first model where the body surface and anatomical skeleton are directly controlled by the same set of shape and pose parameters (β, q). The BioAMASS dataset, the code to create it from the AMASS dataset, as well as the SKEL model, are available for research purposes at <https://skel.is.tue.mpg.de>.

2 RELATED WORK

The accurate representation and animation of human bodies play an important role in computer graphics, vision, and biomechanics. There have been significant recent advances in the creation of statistical body surface models, biomechanical anatomical models, and techniques for extracting these models from motion capture data.

Body models. In vision and graphics, statistical body shape models are widely used [Allen et al. 2003; Angelov et al. 2005; Loper et al. 2015; Osman et al. 2020, 2022; Pavlakos et al. 2019; Wang et al. 2020; Xu et al. 2020]. These models are trained using 3D scans of people with many body shapes in many poses and provide an accurate representation of the human body surface. However, their skeletal structure and their joint locations are not designed to correspond to the anatomical functional joints of the body. For example, their kinematic tree does not match the degrees of freedom of the human anatomic skeleton. The knee and elbow flexion, the spine, the elbow and the arm supination are typically modeled by ball joints, while those functional joints have only one major degree

of freedom or are more complex than a pure rotation, such as the knee, spine, or the shoulder.

Predicting the location of joints, such as the femur head, from 2D images of clothed individuals is inherently ill-posed because the joint location is not directly observed. Instead of directly estimating joints from video, one can fit, or regress, SMPL or SMPL-X body model parameters [Kanazawa et al. 2018; Kocabas et al. 2020; Pavlakos et al. 2019]. From SMPL one can then extract the 3D joint locations, but unfortunately, SMPL joints are not anatomically correct. In this paper, we quantify the error of the SMPL joint locations w.r.t. a biomechanical model and learn a regressor that better predicts the functional joint locations inside SMPL.

Biomechanical skeleton models. In contrast to body models, skeletal models used in biomechanics, e.g. Rajagopal et al. [2016], Seth et al. [2016], Nitschke et al. [2020], define the degrees of freedom of the human skeleton with a focus on anatomic realism. This is critical for kinematic and kinetic analysis. The size and motion of these skeleton models are computed from optical motion capture data using optimization frameworks like OpenSim [Delp et al. 2007] or AddBiomechanics [Werling et al. 2022]. This is the classical approach in biomechanics for measuring the precise location of the functional joints.

Motion capture. While marker-based motion capture (mocap) is the preferred method for analyzing movement, it is expensive, invasive and time consuming. It is also hard to reproduce the exact marker placement on different subjects and most methods assume that the markers are rigidly attached to the body, which is not true due to soft tissue motion. MoSh [Loper et al. 2014; Mahmood et al. 2019] unifies mocap and statistical body models by fitting the parameters of the model to match the marker data. This approach can even mitigate the issues of soft tissue motion.

Traditional mocap, however, typically prevents subjects from wearing normal clothing, complicating capture and limiting its applications. Consequently, many research and commercial solutions for markerless motion capture exist [Bittner et al. 2022; Peng et al. 2023; Uhlrich et al. 2022]. For example, OpenCap [Uhlrich et al. 2022] enables biomechanics from smartphone videos. They use OpenPose [Cao et al. 2017] to detect the subject’s 2D joint locations in several camera views and reconstruct their 3D locations. A biomechanical skeleton model is then fit to these 3D joints. However, existing 2D joint detectors [Cao et al. 2017; Fang et al. 2023; Mathis et al. 2018] have limited biomechanical accuracy, since they are typically trained using manually annotated 2D images. The “ground truth” joint locations do not correspond to the actual functional 3D joint locations. When the joints predicted from images are compared to joint locations computed from motion capture systems [Needham et al. 2021], the differences are as high as 30 to 50 mm for joints such as the knee.

Bones inside bodies. Our goal is to properly place the skeleton inside a parametric body model, providing the best of both worlds. A common approach in previous work uses an anatomic skeleton model and deforms it to register it to a target body mesh [Ali-Hamadi et al. 2013; Gilles et al. 2010; Kadlecěk et al. 2016; Saito et al. 2015; Zhu et al. 2015]. This registration is challenging as these skeleton models do not, in contrast to SMPL, have a shape space

of deformations. Thus the applied deformations may create non-plausible anatomies. In contrast, OSSO [Keller et al. 2022] learns to predict the geometry of the bones from a SMPL body mesh. They learn this 3D geometry from 2D medical images, where both the surface of the person and the skeleton can be observed. Although this approach gives a plausible skeleton shape that fits inside the subject, the resulting skeleton model can not be easily animated as it does not have a kinematic tree. For a lying pose, OSSO yields precise skeletal geometry that is close to the ground truth scans, but the reposing of the skeleton requires an optimization process that can lead to biomechanically impossible poses. The recent BOSS model [Shetty et al. 2023] improves on OSSO by learning a skin-bone-organs model from segmented 3D medical data. While the skin and skeleton model share the same shape space, their kinematic trees used for rigging are different. This does not allow the synchronous posing of both skin and skeleton and an expensive optimization step is required.

Bodies from bones. Going in the other direction, one can infer the body shape given a skeleton. For example, BASH [Schleicher et al. 2021] uses the SCAPE body model [Angelov et al. 2005] to envelop a biomechanical skeletal and muscle model [Nitschke et al. 2020]. However, the SCAPE model is only scaled to match the limb lengths of the skeleton. Shape accuracy is not critical because their goal is to better visualize muscle activation by displaying it on the human surface.

In contrast to prior work, SKEL provides a properly scaled skeleton inside any SMPL body model. Any optimization or regression method that estimates SMPL parameters can now be used to produce biomechanical skeletal parameters. SKEL effectively connects parametric shape models with biomechanical skeletons for the first time to enable the integration of these technologies and fields.

3 METHOD OVERVIEW

Our driving goal is to create SKEL, a model that combines skin and skeleton meshes in which both are synchronously rigged with the same pose parameters \mathbf{q} , and can be reshaped by inheriting the SMPL shape space. To create this model, we must know the location of the anatomic joints and bone rotations inside the human body. There is no large-scale medical dataset of subjects in motion where one can extract both the body and skeleton meshes, and static medical scans do not fully constrain the skeleton in motion. For this, we need bodies in motion and leverage the AMASS dataset [Mahmood et al. 2019] to address this challenge. In Sec. 4 we first present our new custom Biomechanical Skeleton Model, BSM, and describe how to align it inside AMASS sequences of SMPL bodies in motion to obtain the new BioAMASS dataset. Leveraging BioAMASS, Sec. 5 shows how we learn the $SKEL(\boldsymbol{\beta}, \mathbf{q})$ model, which inherits the shape space $\boldsymbol{\beta}$ from SMPL and the pose vector \mathbf{q} from the new BSM biomechanical model. It enables direct animation of the skin and the skeleton meshes using shape and pose parameters, $\boldsymbol{\beta}$ and \mathbf{q} , respectively. Creating SKEL involves two important steps: learning the bone locations and orientations (Sec. 5.1) inside the body, and rigging the skin and bone motions to a common kinematic tree parameterized by \mathbf{q} (Sec. 5.2).

4 THE BIOAMASS DATASET

The goal of the BioAMASS dataset is to enable the learning of the location and orientation of the 3D bones inside a body surface in motion. To create BioAMASS, we use the SMPL [Loper et al. 2015] model for the body surface and a new biomechanical skeleton model, BSM, for the bones. We first introduce these two models, and then describe how we fit BSM to SMPL and create BioAMASS.

4.1 The SMPL body surface model

We model the 3D body surface using the SMPL function, which takes as input shape parameters β and pose parameters $\theta \in \mathbb{R}^{72}$, and outputs a 3D mesh with vertices $\mathbf{v} \in \mathbb{R}^{6890 \times 3}$. The SMPL model includes a joint regressor defined in Eq. 10 of [Loper et al. 2015]. It computes the 3D joint locations of the kinematic tree for the shape parameters β . Each joint is parameterized by three degrees of freedom in an axis-angle representation. The SMPL kinematic tree is artist-defined and only approximately corresponds to the human anatomy. The SMPL equation, summarized in Eq. 5 and 6 of [Loper et al. 2015], first deforms a template mesh \mathbf{T} using learned deformations driven by the shape and pose parameters. Then linear blend skinning (LBS) is used to pose the vertices and produce the body mesh vertices.

4.2 The BSM skeletal model

To model the human skeleton, we create BSM, a custom skeleton model using the OpenSim framework [Delp et al. 2007]; BSM is described by a file in “.osim” format. BSM consists of 24 rigid groups of bones with joints defined between them as well as a mesh representing the geometry of each bone group. On top of each bone, a set of virtual markers is defined; these markers are used to fit BSM to motion capture sequences.

The BSM is represented by three functions that take scaling and pose parameters as input. Using forward kinematics, these functions output the skeleton joint locations, $BSM^J(\mathbf{s}, \mathbf{q})$, the bone meshes vertices, $BSM^v(\mathbf{s}, \mathbf{q})$, and the posed marker locations, $BSM^m(\mathbf{s}, \mathbf{q}, \mathbf{m}_0)$. The scale parameter $\mathbf{s} \in \mathbb{R}^{24 \times 3}$ scales each of the 24 unposed bones along the axis (x,y,z), while the pose parameters $\mathbf{q} \in \mathbb{R}^{46}$ represent the 46 degrees of freedom of the articulated model. The model markers are defined by designating their 3D coordinates $\mathbf{m}_0 \in \mathbb{R}^{N_m \times 3}$ in the corresponding bone reference frame. Each marker is rigidly attached to one bone and, when the bone is scaled with \mathbf{s} , the marker location is scaled accordingly. In contrast to SMPL, BSM has a more realistic kinematic tree but lacks a shape space.

Body models like SMPL typically treat every joint as a ball joint with three angular degrees of freedom. In reality, the joints of the body differ significantly from this assumption. Consequently, for BSM we use more realistic models of the spine, shoulder, and forearm.

Lower body. For BSM’s lower body model, we use the model from Rajagopal et al. [2016], which implements the knee flexion model from Walker et al. [1988].

Spine. We extend the original OpenSim framework with a new custom joint that we call “constant curvature”, to model the spine bending with a constant length. Our BSM model’s spine is made of 3 such joints, enabling lumbar, thoracic, and cervical bending, as

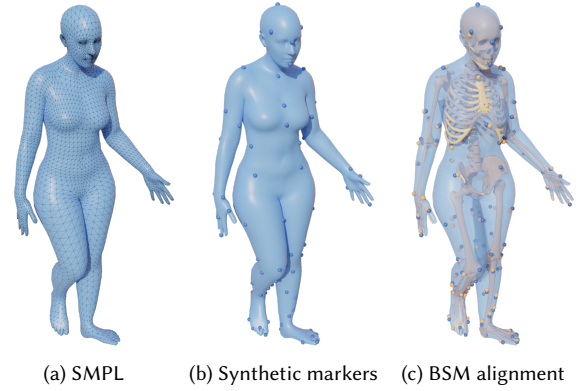


Fig. 2. Creation of the paired skeleton and body dataset. Given a SMPL motion sequence (a), we generate synthetic markers (b), and fit a biomechanical model to the markers using AddBiomechanics [Werling et al. 2022] (c).

illustrated in Fig. 9a. Given the parent joint location J_{i-1} and a spine curve of length l , the child joint J_i will move on a curve of constant arc length and curvature, parameterized by one termination angle $\mathbf{q}_i = [q_x, q_z, q_y] \in \mathbb{R}^3$, represented as Euler-angles in XZY. The child joint location is $J_i = R(\mathbf{q}_i) \cdot (J_{i-1} - J_i) + \mathbf{t}^{\text{spine}}(\mathbf{q}_i)$, where

$$\mathbf{t}^{\text{spine}}(\alpha) = \begin{cases} x = (r - \cos(\alpha)) * \frac{-\sin(q_z)}{\sin(\alpha)} \\ y = r * \sin(\alpha) \\ z = (r - \cos(\alpha)) * \frac{\cos(q_z) * \sin(q_x)}{\sin(\alpha)} \end{cases} \quad (1)$$

with $\alpha = \arcsin \sqrt{\sin(q_z)^2 + (\cos(q_z) * \sin(q_x))^2}$, and $r = \frac{l}{\alpha}$.

Shoulder blades. In BSM we follow the Seth et al. [2016] model and parameterize the shoulder blade joint such that it slides along an ellipsoid defined around the thorax, making the scapula slide along the ribs. The three degrees of freedom are linked to scapula abduction, elevation, and upward rotation as illustrated in Fig. 9b.

Forearm. The forearm pronation and supination are modeled by a single degree of freedom; which is distinct from the elbow flexion, wrist flexion, and wrist deviation [Rajagopal et al. 2016]. The forearm is made of two bones: the radius and the ulna. The ulna is linked to the humerus through a hinge joint, enabling the elbow flexion. During the forearm pronation and supination, the hand rotates while the ulna stays fixed. We model this by rotating the radius along the axis defined by the ulna’s parent joint location and the radius extremity as illustrated in Fig. 9c.

4.3 Fitting BSM to SMPL

To leverage the SMPL body meshes in AMASS, we define a mocap marker set on the SMPL mesh and obtain synthetic sequences of markers. We use these as input to fit our BSM skeleton using AddBiomechanics [Werling et al. 2022], a recent biomechanical optimization framework. Fig. 2 illustrates this pipeline.

4.3.1 Establishing marker correspondences with BSM. To fit BSM to SMPL, we define the same markers on both models. Theoretically, we could define each skin vertex of SMPL to be a marker attached

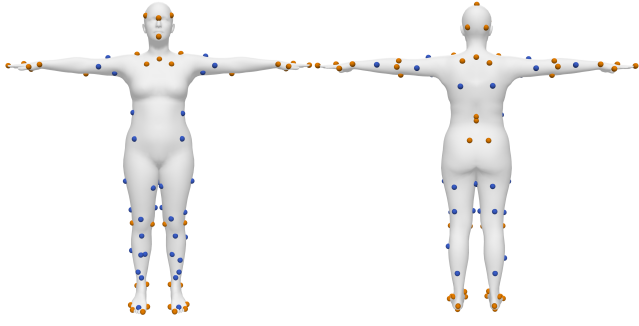


Fig. 3. The markers defined on SMPL: bony in orange, soft in blue.

to BSM. But OpenSim rigidly attaches markers to the bones, hence we define a set of markers that are mostly influenced by one bone and not subject to significant soft tissue deformation.

Specifically, we define 57 *bony markers* that are close to the bones, as typically done in motion capture. Each marker is defined on BSM and SMPL by examining tight SMPL fits to 3D scans and identifying specific SMPL vertices. Figure 3 shows all the *bony markers* on SMPL in orange.

Although this marker set follows the rigidity assumption, is too sparse in some areas to properly constrain the location of the bones. So we introduce an additional 48 *soft markers*, located on soft body parts (blue in Fig. 3). To define a new BSM marker, it needs to be positioned on the BSM skeleton template. While this can be achieved quite precisely for *bony markers*, it is harder to estimate at what distance to the bones *soft markers* should lie. Moreover, this distance varies significantly for different body shapes (e.g. due to adipose tissue). Initializing markers close to the bones for subjects with more adipose tissue can lead to AddBiomechanics over-stretching the bones to fit the SMPL markers.

To address this marker offset issue, we propose a method to automatically define markers on BSM with personalized offsets depending on the body shape. We leverage the OSSO model [Keller et al. 2022], which predicts the location and shape of the skeleton inside SMPL. In contrast to BSM, OSSO models the geometry of the skeleton with respect to the body shape and, as it was trained on medical scans, it learned the offset between the bones and the skin. We can thus use it to compute where skin markers should be located with respect to the bone surface, given a body shape. We first compute the relationship between the OSSO and BSM bones. Precisely, we register each OSSO bone mesh to the corresponding BSM bone mesh and effectively obtain all OSSO bones in the reference frames of the BSM bones. This relationship only needs to be computed once. Then, for each AMASS subject, we use OSSO to obtain their skeleton mesh. We use the lying down pose in which OSSO is trained to obtain the best possible OSSO prediction. Now, given a marker location on the SMPL mesh and the computed OSSO bone mesh inside the body, we parameterize the marker location using the closest triangle on the OSSO bone mesh (Fig. 4a). This allows us to transfer the marker location onto the OSSO bone mesh and, consequently to the corresponding template BSM bone (Fig. 4b).

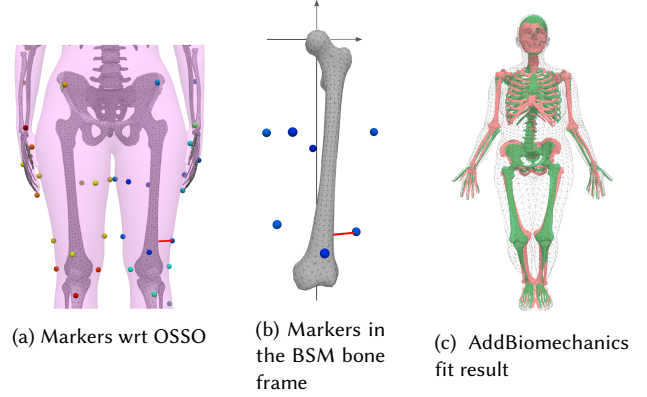


Fig. 4. (a) The OSSO skeleton is aligned to the subject's SMPL mesh. (b) We deduce the personalized markers location $\mathbf{m}_0(\beta)$ on the BSM bone template. (c) On high BMI subjects, a shape-agnostic marker definition for all subjects yields over-stretched bones (red). Using personalized marker locations $\mathbf{m}_0(\beta)$ defined using OSSO prevents this over-stretching (green).

We use this method to generate a BSM model for each subject, with personalized markers $\mathbf{m}_0(\beta)$, thus avoiding over-stretching the bones during the AddBiomechanics optimization, as shown Fig. 4c. We experimented with different marker sets, adjusting their number and placement, to obtain the best possible fits from AddBiomechanics; i.e. minimizing the marker errors and yielding a satisfactory fit visually.

4.3.2 Fitting BSM to motion data. With corresponding markers defined on both SMPL and BSM, we can fit the BSM skeleton to any SMPL mesh. Given a sequence of N_f frames and N_m target 3D marker locations per frame, \mathbf{m}_k^T ($k \in \{1, \dots, N_m\}$), extracted from the sequence of SMPL meshes, we use AddBiomechanics [Werling et al. 2022] to obtain the BSM scale parameters \mathbf{s} and the N_f poses $\{\mathbf{q}_f\}$. We optimize a bi-level objective, to find the best \mathbf{s} such that inverse-kinematics with these scales yields poses $\{\mathbf{q}_f\}$ with minimal distance to the N_m target markers:

$$\arg \min_{\mathbf{s}, \delta} \left(\sum_{f=1}^{N_f} \arg \min_{\mathbf{q}_f} \sum_{k=1}^{N_m} \lambda_k (\text{BSM}^m(\mathbf{s}, \mathbf{q}_f, \mathbf{m}_0(\beta) + \delta)_k - \mathbf{m}_k^T) \right) + \lambda_p P(\mathbf{s}, \beta), \quad (2)$$

where $\delta \in \mathbb{R}^{N_m \times 3}$ is a 3D per marker offset. The weighting factor $\lambda_k \in \mathbb{R}$ is set to a low value for *soft markers* and a high value for *bony markers* to allow larger fitting errors due to secondary soft tissue motions.

The prior P regularizes the scale of the bones, given the subject's height, weight, and biological sex as in [Werling et al. 2022]. We automatically estimate the height and weight of each subject from their SMPL shape parameters β , by assuming that the body has a uniform density [Choutas et al. 2022; Pujades et al. 2019] and thus re-parameterize this prior term as $P(\mathbf{s}, \beta)$.

Despite the scale prior, using a generic marker set can lead to AddBiomechanics over-stretching the bones for heavy subjects. Defining personalized marker locations $\mathbf{m}_0(\beta)$ on the skeleton template as described in the previous section helps further regularize the bone scales (Fig. 4c).

We apply this optimization process to a subset of AMASS consisting of 113 subjects and 2198 motion sequences, amounting to over 9 hours of motion data. The paired SMPL meshes and BSM skeletons form the BioAMASS dataset. For each subject p there is a SMPL body shape β_p and the scaled personalized BSM model s_p . Further, for each motion frame f it includes the bone angles \mathbf{q}_f as well as the bone joint locations \mathbf{J}_f^B . Figure 5 shows examples of the BioAMASS dataset.

5 THE SKEL MODEL

Now we have BSM skeletons inside SMPL but we want to go further and parameterize the 3D body model with the biomechanical skeleton. To that end, we develop SKEL, which is designed to be compatible with SMPL and posed like BSM. This allows us to leverage SMPL’s learned shape space as well as all the existing datasets where SMPL bodies are estimated from different modalities. To create SKEL, we must put SMPL vertices and the BSM skeleton together in the same reference frame. We pose the BSM skeleton mesh inside a SMPL body in T-pose, the zero-pose of the SMPL body model. To that end, in Sec. 5.1, we learn to regress the anatomical joint locations from SMPL using the BioAMASS dataset. Then, in Sec. 5.2 we describe how we rig the SMPL model using the BSM joint rotations.

Note that, in SMPL, all joint orientations are defined in a global T-pose space with an axis-aligned frame of reference for each joint as illustrated in Fig. 6 right. This means that SMPL assumes, for example, that the elbow rotation axis is aligned with the world y-axis, independent of the orientation of the humerus. The over-parametrized nature of SMPL allows plausible arm articulation by combining several axis rotations. But BSM, with its reduced degrees of freedom for the rotations, requires the local frame on which the rotation is applied to be precisely aligned with the anatomy in order to obtain a proper anatomic rigging. In addition, the location and orientation of the humerus and ulna bones have to be coherent with the rotation axis. In SMPL this coherence does not exist: the joint reference frames are not aligned with the articulation axis. As shown in Fig. 6, the elbow frame is not aligned with the segment defining the humerus position. Hence, we first learn to predict the location of the joints inside SMPL and, with these, we learn to properly orient the bones inside a SMPL body mesh.

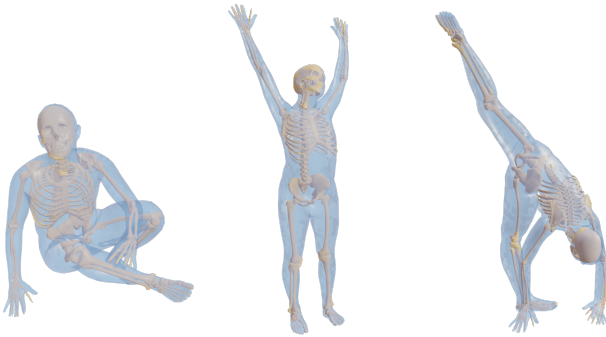


Fig. 5. BioAMASS: examples of BSM fits to AMASS poses.

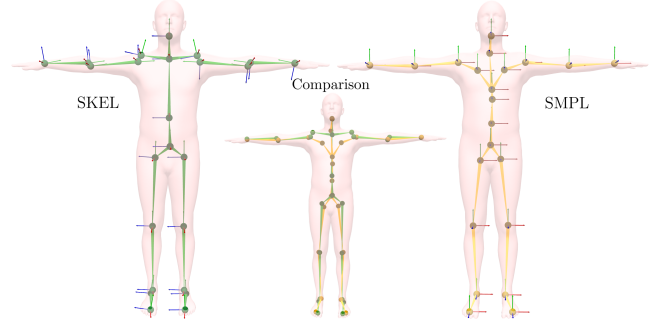


Fig. 6. **Left:** SKEL kinematic tree with learned anatomical joint locations. **Right:** SMPL’s kinematic tree. **Middle:** the superposition of both. In contrast to SMPL, which has axis-aligned rotation axes, SKEL’s rotation axes are bone-aligned.

5.1 Establishing the bone locations and orientations

Anatomical joint locations. Given paired SMPL body meshes and their corresponding BSM anatomic joint locations, we learn a function that predicts the joints from the body surface. We proceed similarly to Loper et al. [2015] by learning a joint regressor \mathcal{J} that takes as input the SMPL mesh vertices $\mathbf{v}^{\text{smpl}} \in \mathbb{R}^{6890 \times 3}$ and predicts the new anatomic joints $\mathbf{J}^B \in \mathbb{R}^{24 \times 3}$. We follow the Keller et al. [2022] OSSO methodology, by formulating a non-negative least squares problem for each joint i , and solving it with an active set method [Lawson and Hanson 1995]. We train these regressors from the posed vertices and joints of the BioAMASS dataset.

Figure 6 shows the new regressed kinematic tree in green. Notice that the hip joint locations, corresponding to the femur heads, are more anatomically correct than the ones in SMPL. The comparison also shows significant differences at the shoulders, as well as more subtle, but important, differences for the other joints.

Bone orientations. We aim to find the orientation of the bones inside the SMPL T-pose mesh, i.e. find the rotation R_i to apply to the i -th BSM bone template mesh, to position it inside the SMPL T-pose mesh. In BSM, the rest position of each individual bone template is centered at the origin and oriented along the canonical axis x, y, z . In the following, we refer to the “bone axis” as the axis passing through the bone’s proximal and distal ends.

In contrast to BSM, in SMPL T-pose, the bones should be positioned and oriented between pairs of regressed anatomical joints. This brings two challenges: (i) the rotation of the bone around its bone axis is not known, and (ii) as the regressed joint location depends on β , the orientation of the bones also varies with β .

To solve those two issues, we split the bone rotation $R_i(\beta)$ into a learned base rotation R_i^{base} and a shape-dependant rotation $R_i^{\beta}(\beta)$:

$$R_i(\beta) = R_i^{\beta}(\beta) \cdot R_i^{\text{base}}, \quad (3)$$

where R_i^{base} is learned to define the bone’s orientation around its bone axis, ensuring that bones are properly orientated wrt their parent bone. $R_i^{\beta}(\beta)$ is computed dynamically to align the bone to the segment defined by its parent and child joints, so that the bone stays in its socket regardless of the shape of the subject.

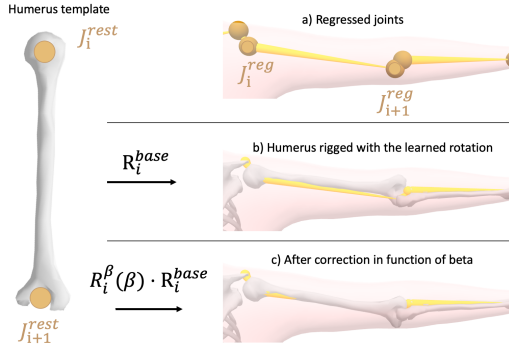


Fig. 7. Left: Humerus template in the rest pose. We want to find its transformation to position it inside SMPL’s arm. On the right we show, in order: a) the anatomical bone joints J_i^{reg} regressed from SMPL skin vertices (pink). b) We center the bone on J_i^{reg} and orient it with R_i^{base} . This provides a rough alignment, rotating the bone properly around its bone axis. c) We then compute and apply the personalized rotation $R_i^\beta(\beta)$ to perfectly align the bone with the limb segment. Notice how the ulnar head now properly fits in the humerus distal end.

First, we learn R_i^{base} from BioAMASS. For each bone i , we can define a corresponding SMPL joint and limb. For example, the right humerus bone corresponds to the 17th joint and the right upper arm of SMPL. Thus, for each frame f of our dataset, we obtain the bone BSM rotation $R_{i,f}^B$ and its SMPL rotation $R_{i,f}^S$.

R_i^{base} is the rotation that the bone needs to undergo so that when chained to the SMPL rotation, the corresponding BSM rotation is obtained. For each bone i we learn its base rotation R_i^{base} by minimizing

$$\sum_{f=1}^{N_f} (R_{i,f}^B - R_{i,f}^S R_i^{base})^2 \quad (4)$$

over the N_f frames of the dataset.

This rotation properly orients the bone around its bone axis. But, as shown in Fig. 7, this rotation alone does not guarantee that the bones are aligned between their T-pose parent and child joints.

Thus we explicitly compute $R_i^\beta(\beta)$, a shape-dependent corrective rotation that aligns the bone segment ($R_i^{base}(J_{i+1}^{rest} - J_i^{rest})$) with ($J_{i+1}^{reg}(\beta) - J_i^{reg}(\beta)$), where J_i^{rest} is the location of joint i in the bone rest pose and $J_i^{reg}(\beta)$ is the shape-dependent regressed joint. The rotation axis of $R_i^\beta(\beta)$ is computed from the cross-product of the segments. As shown in Fig. 7, this effectively ensures a proper fit of the bone geometry into the regressed joint location.

It is worth noting that computing a direct rotation between the rest bone and the regressed segment ($J_{i+1}^{reg}(\beta) - J_i^{reg}(\beta)$) leaves a degree of rotation open: the rotation around the bone axis. With the proposed approach, we obtain an anatomically coherent placement of the skeleton. Thanks to BioAMASS, a consensus orientation R_i^{base} is found, which is then specialized per subject with $R_i^\beta(\beta)$. Effectively, the per-joint R_i^{base} is learned from the dataset once and

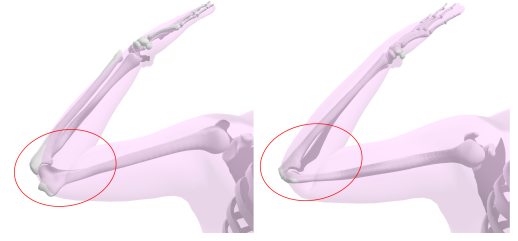


Fig. 8. Left: Rigging the skeleton to the regressed joints and posing them using SMPL parameters θ can yield unrealistic articulations. We see that the humerus posed with the SMPL upper arm transformation does not yield the correct humerus orientation. Right: BSM fit for the same frame.

kept fixed, and each per-joint $R_i^\beta(\beta)$ is recomputed when the shape parameters change.

5.2 Building SKEL: A single rig for skin and bones

As we saw in Fig. 6, the SMPL kinematic tree is not suited to rig the skeletal structure, as its joints do not match the anatomic ones. Moreover, because of its over-parameterization, applying SMPL’s transformation to the bones can yield unrealistic bone orientations, as shown in Fig. 8.

Consequently, we re-rig SMPL with new anatomic degrees of freedom using the learned bone locations and orientations (Sec. 5.1).

The SKEL function. The SKEL function takes as input a vector of SMPL shape parameters, β , and the $\mathbf{q} \in \mathbb{R}^{46}$ pose parameters of BSM. SKEL outputs $(\mathbf{v}^{skin}, \mathbf{v}^{skel}, \mathbf{J})$ where \mathbf{v}^{skin} are the body surface vertices, \mathbf{v}^{skel} the skeleton mesh vertices, and \mathbf{J} the learned anatomic joint locations.

Skin. SKEL builds on the additive approach of SMPL, starting with a mean template mesh $\mathbf{T} \in \mathbb{R}^{6890 \times 3}$ and adding the learned displacements $\beta \cdot \mathbf{B}_S + B_P(\mathbf{q})$, where \mathbf{B}_S is the PCA shape basis learned in Loper et al. [2015] and $B_P(\mathbf{q})$ are pose dependent displacements. The posed SKEL body vertices \mathbf{v}^{skin} are then computed with the following linear blend skinning equation:

$$\mathbf{v}^{skin}(\beta, \mathbf{q}) = \left[\sum_{i=1}^{N_j} W_i^{skin} G_i^{skin}(\mathbf{q}, \beta) \right] (\mathbf{T} + \beta \cdot \mathbf{B}_S + B_P(\mathbf{q})) \quad (5)$$

where $G_i^{skin}(\mathbf{q}, \beta)$ is a rigid transformation that will be defined in Eq. (6). It translates and rotates the vertices associated with the i -th limb depending on the pose parameter \mathbf{q} . W_i^{skin} is a 6890×24 matrix of skinning weights indicating how the vertices of the SMPL mesh are affected by each rigid transformation i . Those weights are inherited from SMPL, by defining a corresponding SMPL joint for each of the $N_j = 24$ joints of SKEL.

To define the transformations G_i^{skin} we use the composition of rigid transformations $T(\mathbf{R}, \mathbf{t})$ defined by a rotation matrix \mathbf{R} and a translation \mathbf{t} , as well as per-joint local transformations $G_k^B(\mathbf{q}_i, \beta)$, which are pure rotations for most joints, and a combination of rotation and translation for the spine and shoulder blades as explained in Sec. 4.2. The global transformation to apply to the skin vertices

is computed as $G_i^{\text{skin}}(\mathbf{q}, \boldsymbol{\beta}) =$

$$\prod_{k=0}^i T(R_k(\boldsymbol{\beta}), \mathbf{J}_k(\boldsymbol{\beta})) G_k^B(\mathbf{q}_k, \boldsymbol{\beta}) T(R_k(\boldsymbol{\beta}), 0)^{-1} T(0, \mathbf{J}_k(\boldsymbol{\beta}))^{-1} \quad (6)$$

The green term transforms the i -th limb vertices back to the unposed bone space, by centering it on its joint location $T(0, \mathbf{J}_k(\boldsymbol{\beta}))^{-1}$, and undoing the T-pose bone rotation $T(R_k(\boldsymbol{\beta}), 0)^{-1}$. Then, the joint-specific transformation $G_k^B(\mathbf{q}_k, \boldsymbol{\beta})$ is applied. Finally, the bone vertices are posed back to SMPL's posed space by applying the rotation $R_k(\boldsymbol{\beta})$ and the translation $\mathbf{J}_k(\boldsymbol{\beta})$. $\mathbf{J}_k(\boldsymbol{\beta})$ is the k -th joint location in T-pose ($\mathbf{q} = 0$) as defined in Eq. (7). The leading product enforces the kinematic tree structure.

The pose-dependent deformations of SKEL are inherited from SMPL. For each degree of freedom of SKEL, we define a corresponding degree of freedom of SMPL and transfer the pose-dependent deformations $B_{p_i}(\mathbf{q}_i)$. For SKEL's joints that do not have an equivalent joint in SMPL, we default to linear blend skinning with no pose correctives. While this transfer is not optimal and creates artifacts in extreme poses (see Sup. Video), SKEL can match SMPL meshes with an average vertex-to-vertex error below 3 cm; see Fig. 12. We leave the learning of SKEL-specific pose-dependent deformations using BioAMASS for future work.

Joints. SKEL's unposed joints are regressed from the unposed skin vertices $\mathbf{v}^{\text{skin}}(\boldsymbol{\beta}, \mathbf{q} = 0)$ with the learned anatomical joint regressor \mathcal{J} , to get the unposed joints $\mathbf{J}^B(\boldsymbol{\beta})$. Those joints are then posed with the parameter \mathbf{q} , like the skin vertices, by applying the rigid transformations G_i^{skin} :

$$\mathbf{J}(\boldsymbol{\beta}, \mathbf{q}) = \left[\sum_{i=1}^{N_J} W_i^J G_i^{\text{skin}}(\mathbf{q}, \boldsymbol{\beta}) \right] \mathbf{J}^B(\boldsymbol{\beta}) \quad (7)$$

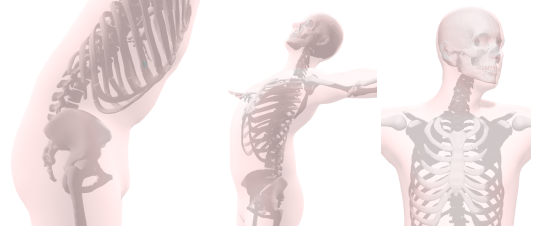
only with different weights W_i^J that ensure that the proper joint is affected by the transformation. Note that for SKEL we use a simplified hinge joint at the knee.

Skeleton. To obtain the shaped and posed skeleton mesh, a similar equation is used. We name the initial skeleton template mesh \mathbf{T}^B in which every bone mesh is axis-aligned and has its parent joint at the world's origin (Fig. 4b right shows the unposed template femur). This mesh is scaled using $s(\mathbf{J}^B(\boldsymbol{\beta}))$, a per-bone scaling factor defined by the regressed joint locations, namely the limb lengths they define. Then the scaled vertices are posed to obtain the posed skeleton vertices

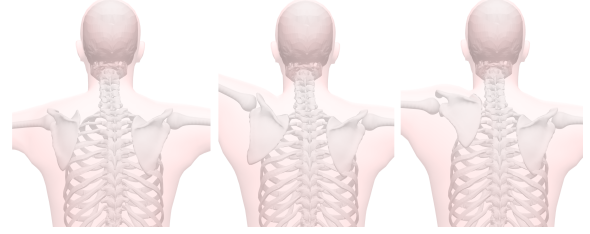
$$\mathbf{v}^{\text{skel}}(\mathbf{q}, \boldsymbol{\beta}) = \left[\sum_{i=1}^{N_J} W_i^{\text{skel}} G_i^{\text{skel}}(\mathbf{q}, \boldsymbol{\beta}) \right] (s(\mathbf{J}^B(\boldsymbol{\beta})) \cdot \mathbf{T}^B) \quad (8)$$

where W_i^{skel} are boolean per-bone weights, except for the spine and rib cage where the weights are interpolated to be 0 at the bottom of the spine section and 1 at the top. The skeleton vertex transformations are computed with

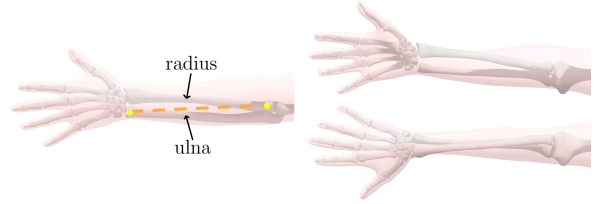
$$G_i^{\text{skel}}(\mathbf{q}, \boldsymbol{\beta}) = \prod_{k=0}^i T(0, \mathbf{J}_k(\boldsymbol{\beta})) T(R_k(\boldsymbol{\beta}), 0) G_k^B(\mathbf{q}_k, \boldsymbol{\beta}) \quad (9)$$



(a) Lumbar flexion, thorax extension, and head twist.



(b) Scapula's abduction, elevation, and upward rotation.



(c) Left: axis of rotation of the radius (lateral view). Right: forearm supination and pronation (top view).

Fig. 9. Illustration of SKEL's degrees of freedom. The bone and body surface meshes are controlled by the same kinematic tree.

in which the unposed bone mesh is transformed by the joint transformation $G_k^B(\mathbf{q}_k, \boldsymbol{\beta})$, then oriented with $R_k(\boldsymbol{\beta})$ to be aligned with the limb's skin and translated to its T-pose joint $\mathbf{J}_k(\boldsymbol{\beta})$.

Finally, we define the range of possible angles for specific degrees of freedom like the shoulder blades, knee, arms, and spine motions. Figure 9 illustrates SKEL's degrees of freedom for the spine, shoulder blades, and arm pronation. Note that the deformation of the body surface (pink) is driven by the BSM pose, thus combining the SMPL surface model with an anatomical skeleton.

6 EVALUATION

In this section, we evaluate the fit accuracy of the BioAMASS dataset, the learned anatomical joint regressors, and the skeleton meshes obtained by fitting SKEL to SMPL meshes.

6.1 Evaluating BioAMASS fits

In Sec. 4 we simulate optical motion capture markers on SMPL sequences and fit the BSM biomechanical skeleton to them. We evaluate these fits by computing the Mean Absolute Error (MAE) between the target and the fitted markers. In Tab. 1, for each subset of AMASS, we report the average error of bony and soft markers across all frames. For comparison, these distances are similar to the body shape reconstruction error from markers reported in [Loper

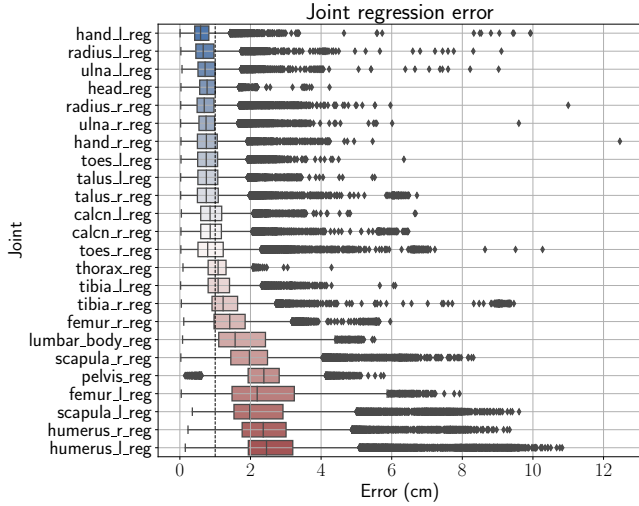


Fig. 10. Anatomical joints regression error over the female DFAUST dataset.

et al. 2014] and significantly more accurate than the held-out marker error [Loper et al. 2014].

6.2 Joint regressors

We evaluate the regressors learned in Sec. 5.1 on unseen body meshes by comparing the regressed values with the reference BSM alignment. We train our anatomical joint regressors on the CMU [CMU Graphics Lab 2000] and MPI_Limits [Akhter and Black 2015] datasets, which are part of AMASS [Mahmood et al. 2019]. CMU contains good variation in body shape, while MPI_Limits contains extreme poses. Once trained, we evaluate our regressor on the DFAUST dataset [Bogo et al. 2017], with various motion sequences for 10 subjects with diverse BMIs; DFAUST contains precise SMPL fits to 3D scan sequences.

For each frame of the DFAUST dataset, BioAMASS provides the anatomical joint locations \mathbf{J}^B that we consider ground truth. Then, from the frame’s SMPL mesh, we use our learned joint regressor to regress the anatomical joint location \mathbf{J}^{reg} . In Fig. 10 we report the per joint regression errors $|\mathbf{J}_i^{reg} - \mathbf{J}_i^B|$, which are below a centimeter for most joints. Some joints, such as the humerus, have higher errors. We inspected the outlying frames and observed some failure cases of the AddBiomechanics fits for the shoulder joints, which explains the higher values. The regressed anatomical joints are, in these cases, more plausible than those obtained with BSM, as we show in the supplementary video.

Table 1. Marker fitting error of the BSM model on the AMASS dataset.

MAE in (cm)	Bony markers	Soft markers
DFAUST	1.54	2.00
CMU	1.70	2.37
MPI_Limits	1.70	2.37

Further, we evaluate the femur and tibia joint location given by different methods as shown in Fig. 11. We consider \mathbf{J}^B as the ground truth joint locations and compute the 3D Euclidean distance error of the joints given by SMPL, \mathbf{J}^{smpl} , the anatomical joints we regress from SMPL, \mathbf{J}^{reg} , and the anatomical joints, \mathbf{J}^{skel} , obtained by fitting fit SKEL to the SMPL mesh. As expected, the SMPL joints have higher error compared to the learned anatomical ones.

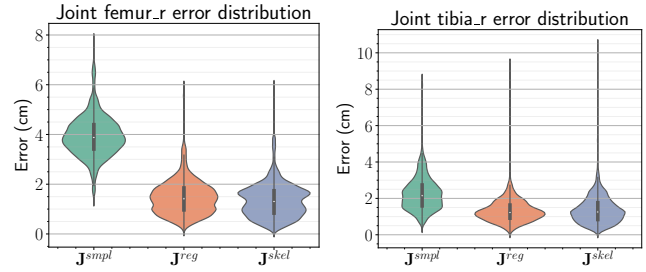


Fig. 11. On DFAUST female subjects, we predict the joint locations and show the Euclidean distance errors wrt the “ground truth” BSM joint location for the right femur (left) and right tibia (right). We compare 3 methods: \mathbf{J}^{smpl} : using the joints directly from the SMPL fit to the DFAUST bodies. \mathbf{J}^{reg} : joint regressed from the SMPL mesh using our learned anatomical joint regressor. \mathbf{J}^{skel} : anatomical joints obtained by fitting SKEL to the SMPL mesh.

6.3 SKEL fits to SMPL

Since SKEL has the same surface mesh topology and shape parameters β as SMPL, it can be directly fit to existing SMPL meshes by optimizing its pose parameters to minimize the vertex-to-vertex error.

To quantitatively evaluate how similar SKEL shapes are to SMPL, we consider motion sequences from the DFAUST dataset and their SMPL fits with 10 shape parameters. We fit SKEL to each of these SMPL meshes by optimizing its pose parameters \mathbf{q} . To evaluate the mesh fits, we compute the mean absolute difference (MAD) between SKEL skin vertices and the target SMPL vertices, and then average over all the frames. For males, we find an average difference of 1.1 cm and an average max difference of 2.5 cm, while for females we obtain an average mean difference of 0.9 cm and max of 1.9 cm. A visualization of these differences on the SMPL body mesh is shown in Fig. 12. The larger differences can be explained by the approximate pose-dependant blend shapes inherited from SMPL, which could be retrained in future work.

Fitting SKEL to SMPL provides joint locations with similar accuracy as the regressed ones, as reported on Fig. 11. Let us note that direct joint regression is faster than estimating the SKEL model fit. Applications that require the joint locations but not the skeleton pose parameters, and for which time is critical, should prefer the direct regression approach.

Upgrading SMPL datasets with SKEL. Since SKEL is compatible with SMPL, we can fit SKEL to SMPL meshes from the 3DPW dataset [Von Marcard et al. 2018] and the synthetic BEDLAM [Black et al. 2023] dataset (Fig. 13). The full sequences are shown in the supplementary video. This effectively upgrades these datasets to include biomechanical pose parameters.

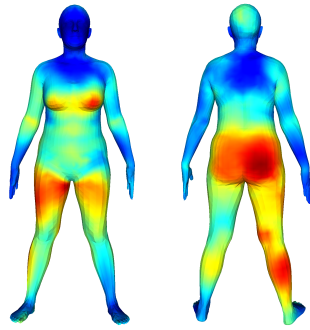


Fig. 12. Average per vertex distance between SKEL and SMPL fit to the females of the DFAUST dataset. Blue: 0 cm, Red 2cm.

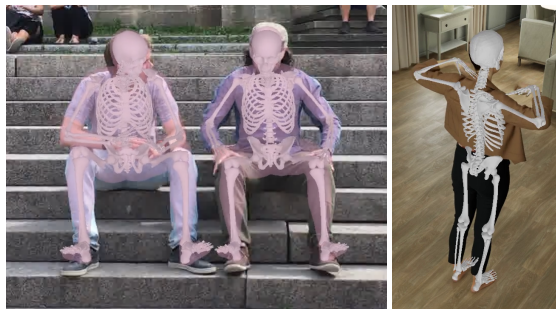


Fig. 13. SKEL can be fit to existing SMPL datasets to upgrade them with biomechanical pose parameters. Left: SKEL skeleton mesh on a frame of 3DPW [Von Marcard et al. 2018]. Right: SKEL skeleton mesh on a frame of BEDLAM [Black et al. 2023].

6.4 Qualitative comparisons with OSSO

SKEL fits to SMPL also yield anatomically correct orientations of the bones. To illustrate this we compare the SKEL predictions to OSSO skeletons [Keller et al. 2022] on the MOYO dataset [Tripathi et al. 2023]. The SKEL skeletons yield more anatomically correct joint location and biomechanically relevant bone angles, as visible in Fig. 14; see, for example, the knee orientation as well as arm supination. See Sup. Mat. and Sup. Video for more examples.

6.5 Disentangling body shape and bone lengths

Since our skeleton mesh is fully defined by the joint segment lengths, we can modify the body shape of a person, while maintaining their skeletal identity. This can be helpful for generating a plausible skin mesh from a given biomechanical skeleton. As illustrated in Fig. 15, we optimize the SKEL shape parameters β to fit a subject’s limb lengths with different target weights. This results in body meshes with different body shapes but the same bone lengths.

7 DISCUSSION AND CONCLUSION

In this paper we describe SKEL, a new parametric 3D human body shape model driven by anatomically sound parameters, providing consistent skin and bone geometries. SKEL is learned from BioAMASS, a new dataset of skeletons inside SMPL meshes in diverse

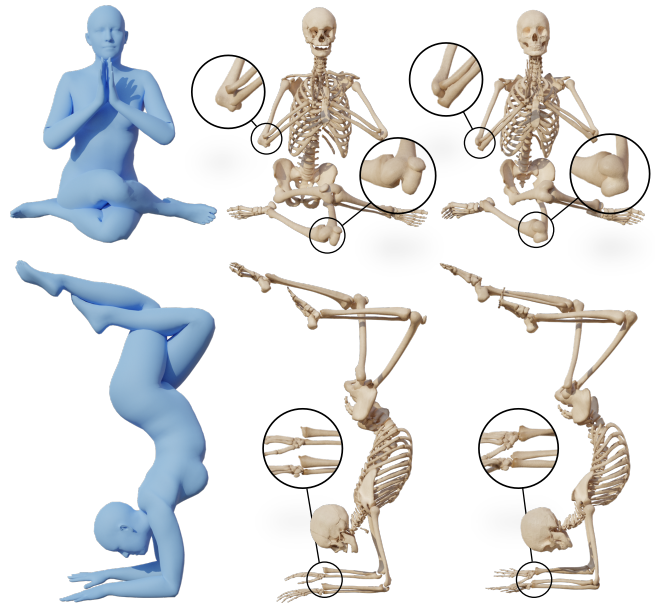


Fig. 14. Qualitative comparison between the OSSO and SKEL skeletons fitted to MOYO SMPL meshes [Tripathi et al. 2023]. From left to right: Input SMPL mesh, OSSO skeleton, SKEL skeleton. First row: Due to the anatomic degrees of freedom of SKEL, the humerus and femur orientation are properly recovered, while OSSO fails. Second row: OSSO does not model the forearm supination: the radius is not properly rotated with respect to the ulna. The forearm bones have an anatomically correct orientation inside SKEL.



Fig. 15. Given an input skeleton, and a target weight, SKEL can generate plausible skins while preserving the skeletal structure. From left to right, we set the weight to be 70, 100, and 130 kg.

AMASS poses. We build BioAMASS by optimizing BSM, a new biomechanically accurate skeleton model, to fit inside SMPL mesh sequences. Using this paired internal and external data we then learn a regressor from SMPL mesh vertices to the anatomic joint locations and bone orientations. SKEL inherits the shape space from

SMPL and the new anatomic kinematic parameters from BSM. From the point of view of vision and graphics, the new model can be used in place of SMPL and it has fewer and more anatomically sound pose parameters (46 for SKEL vs 72 for SMPL). This is advantageous, for instance, to regress more accurate anatomic joints from video compared to current approaches solely based on SMPL joints. From the biomechanics point of view, SKEL provides a shape space, which is advantageous to adapt the model to varied body shapes without overstretching certain bones. In addition, it provides an animatable model that can take BSM poses and add a SMPL skin for visualization.

BioAMASS accuracy limitation. Although the skeletal structures and joint locations computed by AddBiomechanics are anatomically plausible, they should not be considered as *actual* ground truth, but rather a pseudo-ground truth. Obtaining actual ground-truth bone measurements of people in motion is not technically feasible. Thus we rely on marker-based motion capture to obtain estimates of bone motion; this is the current “gold standard” in biomechanics. Thus we inherit the accuracy limits of this method, especially for the humerus head prediction, as shown in Fig. 10. A key next step is to use SKEL in the diagnosis of disease and injury and to compare this with traditional motion capture methods. This is necessary to validate the clinical relevance of the model and methods. It is worth noting that the learning and rigging pipeline described in Sec. 5 are, in fact, independent of the biomechanical model. If a new biomechanical model is clinically validated, one can rerun our approach with it to obtain an improved dataset and model.

SKEL extensions and future work. There are several directions for extending and improving SKEL. For instance, in the current model the hands of SKEL are rigid. Using a biomechanical model with more expressive hands, our approach could be used to put it in correspondence inside SMPL-X [Pavlakos et al. 2019]. The bone locations could also be supervised with static medical data observations, such as the ones provided by [Wang et al. 2019].

Additionally, the current SKEL model inherits the skinning weights and pose-dependent blend shapes from SMPL. These could be retrained from the BioAMASS dataset to make the skin surface deformation more accurate. Ideally, the pose-correctives should be retrained from 3D scan data with BSM as the native parameterization. This would allow the pose-corrective offsets to be directly learned as a function of BSM parameters.

Finally, SKEL is a step towards a more complete model of the body in motion. A next step is to add muscle geometry and muscle activation. For example, we can exploit the estimated BSM skeleton to infer muscle activation using standard biomechanics techniques. This would allow us to upgrade BioAMASS with estimated muscle activity.

Conclusion. In summary, SKEL effectively connects data-driven parametric body shape models with biomechanical skeletons for the first time to enable the integration of these technologies and fields, paving the way towards a new generation of body models and methods that combine the best of both worlds.

Acknowledgments and Disclosures. Michael J. Black (MJB) has received research gift funds from Adobe, Intel, Nvidia, Meta/Facebook,

and Amazon. MJB has financial interests in Amazon, Datagen Technologies, and Meshcapade GmbH. While MJB is a consultant for Meshcapade, his research was performed solely at MPI.

Soyong Shin performed this work while an intern at the Max Planck Institute for Intelligent Systems.

Marilyn Keller was supported by the International Max Planck Research School for Intelligent Systems. Sergi Pujades’ work was funded by the ANR SEMBA project.

This work uses *aitviewer* for visualization [Kaufmann et al. 2022].

REFERENCES

- Ijaz Akhter and Michael J. Black. 2015. Pose-Conditioned Joint Angle Limits for 3D Human Pose Reconstruction. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 1446–1455. <https://doi.org/10.1109/CVPR.2015.7298751>
- Dicko Ali-Hamadi, Tiantian Liu, Benjamin Gilles, Ladislav Kavan, François Faure, Olivier Palombi, and Marie-Paule Cani. 2013. Anatomy transfer. *ACM Transactions on Graphics* 32, 6, Article 188 (Nov. 2013), 8 pages. <https://doi.org/10.1145/2508363.2508415>
- Brett Allen, Brian Curless, and Zoran Popović. 2003. The Space of Human Body Shapes: Reconstruction and Parameterization from Range Scans. *ACM Transactions on Graphics* 22, 3 (July 2003), 587–594. <https://doi.org/10.1145/882262.882311>
- Dragomir Anguelov, Praveen Srinivasan, Daphne Koller, Sebastian Thrun, Jim Rodgers, and James Davis. 2005. SCAPE: Shape Completion and Animation of People. *ACM Transactions on Graphics* 24, 3 (July 2005), 408–416. <https://doi.org/10.1145/1073204.1073207>
- Marian Bittner, Wei-Tse Yang, Xucong Zhang, Ajay Seth, Jan van Gemert, and Frans CT van der Helm. 2022. Towards Single Camera Human 3D-Kinematics. *Sensors* 23, 1 (2022), 341.
- Michael J. Black, Priyanka Patel, Joachim Tesch, and Jinlong Yang. 2023. BEDLAM: A Synthetic Dataset of Bodies Exhibiting Detailed Lifelike Animated Motion. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 8726–8737.
- Federica Bogo, Javier Romero, Gerard Pons-Moll, and Michael J. Black. 2017. Dynamic FAUST: Registering Human Bodies in Motion. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 6233–6242.
- Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2017. Realtime multi-person 2d pose estimation using part affinity fields. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 7291–7299.
- Vasileios Choutas, Lea Müller, Chun-Hao P. Huang, Siyu Tang, Dimitrios Tzionas, and Michael J. Black. 2022. Accurate 3D Body Shape Regression using Metric and Semantic Attributes. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2718–2728.
- CMU Graphics Lab 2000. CMU Graphics Lab Motion Capture Database. <http://mocap.cs.cmu.edu>. Accessed: 2012-12-11.
- Scott L. Delp, Frank C. Anderson, Allison S. Arnold, Peter Loan, Ayman Habib, Chand T. John, Eran Guendelman, and Darryl G. Thelen. 2007. OpenSim: Open-source software to create and analyze dynamic simulations of movement. *IEEE Transactions on Biomedical Engineering* 54, 11 (2007), 1940–1950.
- Hao-Shu Fang, Jiefeng Li, Hongyang Tang, Chao Xu, Haoyi Zhu, Yuliang Xiu, Yong-Lu Li, and Cewu Lu. 2023. AlphaPose: Whole-Body Regional Multi-Person Pose Estimation and Tracking in Real-Time. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 6 (2023), 7157–7173. <https://doi.org/10.1109/TPAMI.2022.3222784>
- Benjamin Gilles, Lionel Reveret, and Dinesh Pai. 2010. Creating and animating subject-specific anatomical models. *Comput. Graph. Forum* 29, 8 (Dec. 2010), 2340–2351. <https://doi.org/10.1111/j.1467-8659.2010.01718.x>
- Petr Kadlecěk, Alexandru-Eugen Ichim, Tiantian Liu, Jaroslav Krivánek, and Ladislav Kavan. 2016. Reconstructing Personalized Anatomical Models for Physics-Based Body Animation. *ACM Transactions on Graphics* 35, 6, Article 213 (Nov. 2016), 13 pages. <https://doi.org/10.1145/2980179.2982438>
- Angjoo Kanazawa, Michael J. Black, David W. Jacobs, and Jitendra Malik. 2018. End-to-end Recovery of Human Shape and Pose. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 7122–7131.
- Manuel Kaufmann, Velko Vechev, and Dario Mylonopoulos. 2022. *aitviewer*. <https://doi.org/10.5281/zenodo.1234>
- Robi Kelc. 2012. Zygote body: A new interactive 3-dimensional didactical tool for teaching anatomy. *Webmedcentral Anatomy* (2012).
- Marilyn Keller, Silvia Zuffi, Michael J. Black, and Sergi Pujades. 2022. OSSO: Obtaining Skeletal Shape from Outside. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 20492–20501.
- Muhammed Kocabas, Nikos Athanasiou, and Michael J. Black. 2020. VIBE: Video Inference for Human Body Pose and Shape Estimation. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Piscataway, NJ, 5252–5262. <https://doi.org/10.1109/CVPR45622.2020.00525>

- [//doi.org/10.1109/CVPR42600.2020.00530](https://doi.org/10.1109/CVPR42600.2020.00530)
- Charles L Lawson and Richard J Hanson. 1995. *Solving least squares problems*. Society for Industrial and Applied Mathematics.
- Matthew Loper, Naureen Mahmood, Javier Romero, Gerard Pons-Moll, and Michael J Black. 2015. SMLP: A skinned multi-person linear model. *ACM Transactions on Graphics* 34, 6 (2015), 1–16.
- Matthew M. Loper, Naureen Mahmood, and Michael J. Black. 2014. MoSh: Motion and Shape Capture from Sparse Markers. *ACM Transactions on Graphics* 33, 6 (Nov. 2014), 220:1–220:13. <https://doi.org/10.1145/2661229.2661273>
- Naureen Mahmood, Nima Ghorbani, Nikolaus F. Troje, Gerard Pons-Moll, and Michael J. Black. 2019. AMASS: Archive of Motion Capture as Surface Shapes. In *International Conf. on Computer Vision (ICCV)*. IEEE, 5442–5451.
- Alexander Mathis, Pranav Mamidanna, Kevin M Cury, Taiga Abe, Venkatesh N Murthy, Mackenzie Weygandt Mathis, and Matthias Bethge. 2018. DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature neuroscience* 21, 9 (2018), 1281–1289.
- Laurie Needham, Murray Evans, Darren P Cosker, Logan Wade, Polly M McGuigan, James L Bilzon, and Steffi L Colyer. 2021. The accuracy of several pose estimation methods for 3D joint centre localisation. *Scientific reports* 11, 1 (2021), 20673.
- Marlies Nitschke, Eva Dorschky, Dieter Heinrich, Heiko Schlarb, Bjoern M Eskofier, Anne D Koelewijn, and Antonie J van den Bogert. 2020. Efficient trajectory optimization for curved running using a 3D musculoskeletal model with implicit dynamics. *Scientific reports* 10, 1 (2020), 1–12.
- Ahmed A. A. Osman, Timo Bolkart, and Michael J. Black. 2020. STAR: Sparse Trained Articulated Human Body Regressor. In *European Conf. on Computer Vision (ECCV)*, Vol. LNCS 12355. Springer, 598–613. https://doi.org/10.1007/978-3-030-58539-6_36
- Ahmed A A Osman, Timo Bolkart, Dimitrios Tzionas, and Michael J. Black. 2022. SUPR: A Sparse Unified Part-Based Human Body Model. In *European Conf. on Computer Vision (ECCV)*. Springer, Springer, 568–585. <https://supr.is.tue.mpg.de>
- Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. 2019. Expressive Body Capture: 3D Hands, Face, and Body from a Single Image. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 10975–10985.
- Zhiheng Peng, Kai Zhao, Xiaoran Chen, Yingfeng Chen, Changjie Fan, Bowei Tang, Siyu Xia, and Weijian Shang. 2023. See Through the Inside and Outside: Human Body and Anatomical Skeleton Prediction Network. In *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 915–916.
- Sergi Pujades, Betty Mohler, Anne Thaler, Joachim Tesch, Naureen Mahmood, Nikolas Hesse, Heinrich H Bühlhoff, and Michael J Black. 2019. The Virtual Caliper: Rapid Creation of Metrically Accurate Avatars from 3D Measurements. *IEEE Trans. Vis. Comput. Graph.* 25, 5 (2019), 1887–1897. <https://doi.org/10.1109/TVCG.2019.2898748>
- Apoorva Rajagopal, Christopher L Dembia, Matthew S DeMers, Denny D Delp, Jennifer L Hicks, and Scott L Delp. 2016. Full-body musculoskeletal model for muscle-driven simulation of human gait. *IEEE Transactions on Biomedical Engineering* 63, 10 (2016), 2068–2079.
- Shunsuke Saito, Zi-Ye Zhou, and Ladislav Kavan. 2015. Computational bodybuilding: Anatomically-based modeling of human bodies. *ACM Transactions on Graphics* 34, 4 (2015), 1–12.
- Robert Schleicher, Marlies Nitschke, Jana Martschinke, Marc Stamminger, Björn Eskofier, Jochen Klucken, and Anne Koelewijn. 2021. BASH: Biomechanical Animated Skinned Human for Visualization of Kinematics and Muscle Activity. In *Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP) (2021-02-08/2021-02-10)*. Springer, 25–36. <https://doi.org/10.5220/0010210600250036>
- Ajay Seth, Ricardo Matias, António P. Veloso, and Scott L. Delp. 2016. A Biomechanical Model of the Scapulothoracic Joint to Accurately Capture Scapular Kinematics during Shoulder Movements. *PLoS One* 11, 1 (01 2016), 1–18. <https://doi.org/10.1371/journal.pone.0141028>
- Karthik Shetty, Annette Birkhold, Srikrishna Jaganathan, Norbert Strobel, Bernhard Egger, Markus Kowarschik, and Andreas Maier. 2023. BOSS: Bones, organs and skin shape model. *Computers in Biology and Medicine* 165 (2023), 107383. <https://doi.org/10.1016/j.compbiomed.2023.107383>
- Shashank Tripathi, Lea Müller, Chun-Hao P. Huang, Omid Taheri, Michael Black, and Dimitrios Tzionas. 2023. 3D Human Pose Estimation via Intuitive Physics. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 4713–4725. <https://ipman.is.tue.mpg.de>
- Scott D. Uhlich, Antoine Falisse, Łukasz Kidziński, Julie Muccini, Michael Ko, Akshay S. Chaudhari, Jennifer L. Hicks, and Scott L. Delp. 2022. OpenCap: 3D human movement dynamics from smartphone videos. *bioRxiv* (2022). <https://doi.org/10.1101/2022.07.07.499061>
- Timo Von Marcard, Roberto Henschel, Michael J Black, Bodo Rosenhahn, and Gerard Pons-Moll. 2018. Recovering accurate 3d human pose in the wild using imus and a moving camera. In *European Conf. on Computer Vision (ECCV)*. 601–617.
- PS Walker, JS Rovick, and DD Robertson. 1988. The effects of knee brace hinge design and placement on joint mechanics. *Journal of Biomechanics* 21, 11 (1988), 965–974.
- Bohan Wang, George Matcuk, and Jernej Barbič. 2019. Hand Modeling and Simulation Using Stabilized Magnetic Resonance Imaging. *ACM Transactions on Graphics* 38, 4, Article 115 (July 2019), 14 pages. <https://doi.org/10.1145/3306346.3322983>
- Haoyang Wang, Riza Alp Güler, Iasonas Kokkinos, George Papandreou, and Stefanos Zafeiriou. 2020. BLSM: A bone-level skinned model of the human mesh. In *European Conf. on Computer Vision (ECCV)*. Springer, 1–17.
- Keenon Werling, Michael Raitor, Jon Stingel, Jennifer L. Hicks, Steve Collins, Scott L. Delp, and C. Karen Liu. 2022. Rapid bilevel optimization to concurrently solve musculoskeletal scaling, marker registration, and inverse kinematic problems for human motion reconstruction. *bioRxiv* (2022). <https://doi.org/10.1101/2022.08.22.504896> arXiv:<https://www.biorxiv.org/content/early/2022/08/23/2022.08.22.504896.full.pdf>
- Hongyi Xu, Eduard Gabriel Bazavan, Andrei Zanfir, William T. Freeman, Rahul Sukthankar, and Cristian Sminchisescu. 2020. GHUM & GHUML: Generative 3D human shape and articulated pose models. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 6184–6193.
- Lifeng Zhu, Xiaoyan Hu, and Ladislav Kavan. 2015. Adaptable Anatomical Models for Realistic Bone Motion Reconstruction. *Comput. Graph. Forum* 34, 2 (2015), 459–471. <https://doi.org/10.1111/cgf.12575>