

# Inside Humans: Creating a Simple Layered Anatomical Model from Human Surface Scans

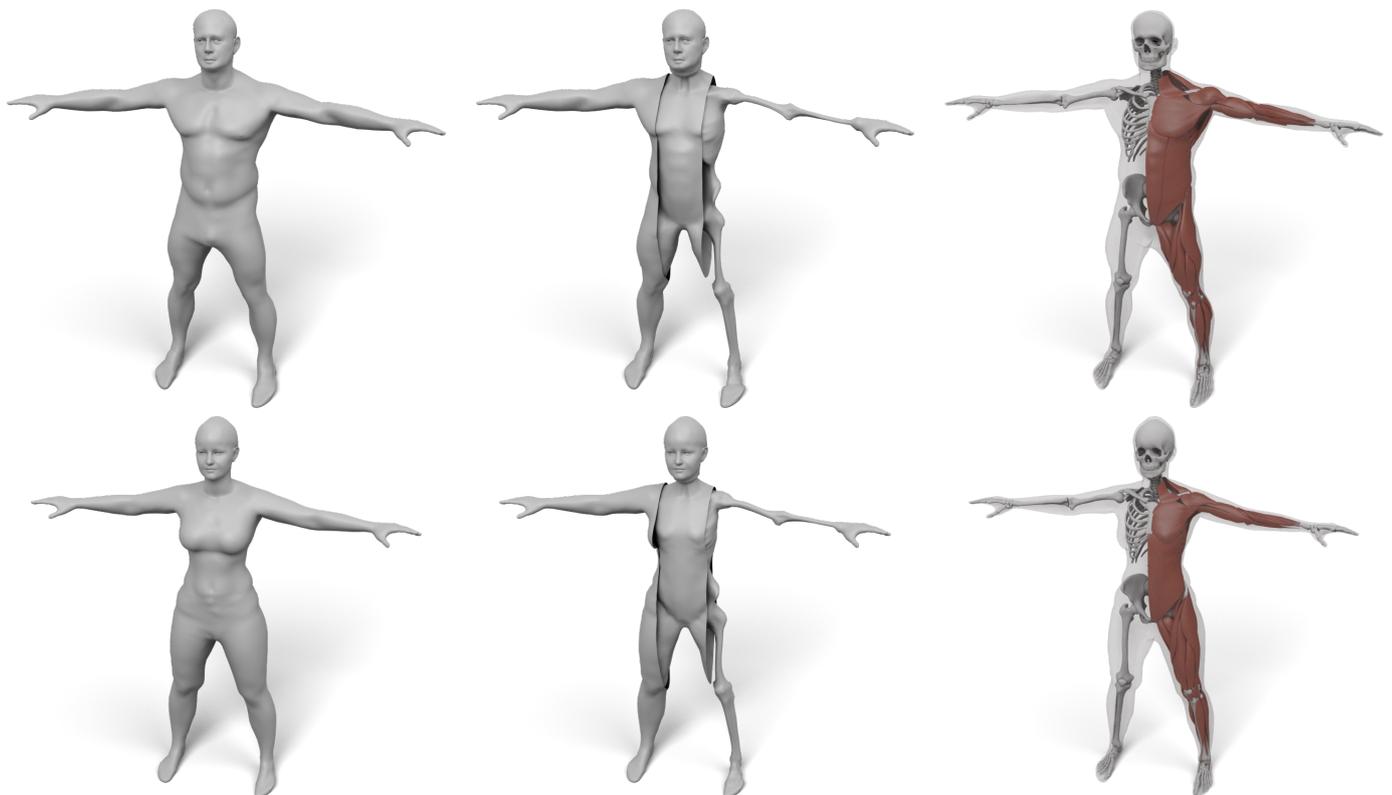
Martin Komaritzan<sup>1,\*</sup>, Stephan Wenninger<sup>1</sup> and Mario Botsch<sup>1</sup>

<sup>1</sup>Computer Graphics Group, TU Dortmund University, Dortmund, Germany

Correspondence\*:

Martin Komaritzan

[martin.komaritzan@tu-dortmund.de](mailto:martin.komaritzan@tu-dortmund.de)



**Figure 1.** Starting from the surface of a human (left), we fit a three-layered model consisting of a skin, muscle, and skeleton layer (middle), which enables physical simulations in a simple and intuitive way. Interior structures, such as individual models of muscles and bones, can also be transferred using our layered model (right).

## ABSTRACT

3D morphable models are widely used to describe the variation of human body shapes. However, these models typically focus on the *surface* of the human body, since the acquisition of the *volumetric interior* would require prohibitive medical imaging. In this paper we present a novel approach for creating a volumetric body template and for fitting this template to the surface scan of a person in a just a few seconds. The body model is composed of three surface layers for bones, muscles, and skin, which enclose the volumetric muscle and fat tissue in between them. Our approach includes a data-driven method for estimating the amount of muscle mass and fat

mass from a surface scan, which provides more accurate fits to the variety of human body shapes compared to previous approaches. We also show how to efficiently embed fine-scale anatomical details, such as high resolution skeleton and muscle models, into the layered fit of a person. Our model can be used for physical simulation, statistical analysis, and anatomical visualization in computer animation and medical applications, which we demonstrate on several examples.

**Keywords:** virtual humans, anatomical details, template fitting, character animation, shape deformation

## 1 INTRODUCTION

Virtual humans are present in our everyday lives. They can be found in movies, computer games, and commercials. In addition, they are employed in a rapidly growing number of applications in virtual reality (VR) and augmented reality (AR), even ranging to computational medicine. All these applications benefit from realistic virtual representations of human .

If we look at a human, its appearance is mostly determined by everything we can directly see (skin, hair, cloth, etc.). Hence, it is not surprising that research has focused on capturing, analyzing, and animating *surface models* of humans. Consequently, there is a vast amount of surface-based capturing approaches, suitable for almost every level of detail and budget: From complex multi-camera photogrammetry setups that capture finest-scale wrinkles of the human face (Riviere et al., 2020) over approaches that compute ready-to-animate models from simple smart-phone videos (Wenninger et al., 2020) to machine learning approaches that reconstruct a virtual model from a single image (Weng et al., 2019). For the purpose of creating convincing animations of and interactions with those models, large amounts of 3D captured data have been collected to build sophisticated surface-based models (Loper et al., 2015; Anguelov et al., 2005; Bogo et al., 2017). Those models compensate for the fact that humans are not empty hulls or homogeneous solids by capturing and analyzing more and more data of that surface hull. Another way to approach this is to model *volumetric* virtual humans by incorporating (discrete approximations of) their interior anatomical structures. While surface-based models might be sufficient for many applications, for others (e.g., surgery simulation) a volumetric model is an essential prerequisite.

While detailed volumetric models of the human body exist (Ackerman, 1998; Christ et al., 2009; Zygote, 2020), they can be very tedious to work with. Since they usually consist of hundreds of different bones and muscles, merely creating a volumetric tetrahedral mesh for simulation purposes can be frustratingly difficult. Moreover, those models represent average humans and transferring their volumetric structure and anatomical details to a specific human model (e.g., a scanned person) is not straightforward. Although there are a couple of approaches for transferring the interior anatomy from a volumetric template model into a surface-based virtual human (Dicko et al., 2013; Kadleček et al., 2016), these methods either deform bone structures in a non-plausible manner (Dicko et al., 2013) or require a complex numerical optimization (Kadleček et al., 2016).

In this paper we present a robust and efficient method for transferring an interior anatomy template into a surface mesh in just a couple of seconds. A key component is a simple decomposition of the human body into three layers that are bounded by surfaces sharing the same triangulation: the *skin surface* defines the outer shape of the human, the *muscle surface* envelopes its individual muscles, and the *skeleton surface* wraps the subject's skeleton (see Figure 1 middle). The muscle layer is hence enclosed in between the skeleton and muscle surface, and the subcutaneous fat tissue by the muscle surface and skin surface. This layered template model is derived from the Zygote body model (Zygote, 2020), which provides an accurate representation of both the male and female anatomy. We propose simple and fast methods for fitting the

layered template to surface scans of humans and for transferring the high-resolution anatomical details (Zygote, 2020) into these fitted layers (see Figure 1 right). Our method is robust, efficient, and fully automatic, which we demonstrate on about 1700 scans from the European CAESAR dataset (Robinette et al., 2002).

Our approach enriches simple surface scans by plausible anatomical details, which are suitable for educational visualizations and volumetric simulations. We note, however, that due to the lack of true volumetric information, it is not a replacement of volumetric imaging techniques in a medical context. Our main contributions are:

- A novel approach for creating a layered volumetric template defined by skin, muscle, and bone surfaces, which all have the same triangulation, thereby making volumetric tessellation straightforward.
- A robust and efficient method for transferring the layered volumetric template model into a given surface scan of a human in just a couple of seconds.
- A regressor that extracts the amount of muscle and fat mass of a subject from the skin surface only, thereby making manual specification of muscle and fat distribution unnecessary.
- Our approach takes differences between male and female anatomy into account by deriving individual volumetric templates and individual muscle/fat regressors for men and women.

## 2 RELATED WORK

Using a layered volumetric model of a virtual character has been shown beneficial compared to a surface-only model in multiple previous works. Deul and Bender (2013) compute a simple layered model representing a bone, muscle, and fat layer, which they use for a multi-layered skinning approach. Simplistic layered models have also been used to extend the SMPL surface model (Loper et al., 2015) in order to support elastic effects in skinning animations (Kim et al., 2017; Romero et al., 2020). Compared to these works, our three layers yield an anatomically more accurate representation of the human body, while still being simpler and more efficient than complex irregular tetrahedralizations. Saito et al. (2015) show that a layer that envelopes muscles yields more convincing muscle growth simulations and reduces the number of tetrahedral elements required in their computational model. They also show how to simulate different variations of bone sizes, muscle mass and fat mass for a virtual character.

When it comes to the generation of realistic personalized anatomical structures from a given skin surface, most previous works focus on the human head: Ichim et al. (2016) register a template skull model to a surface-scan of the head in order to build a combined animation model using both physics-based and blendshape-based face animation. Ichim et al. (2017) also incorporate facial muscles and a muscle activation model to allow more advanced face animation effects. Gietzen et al. (2019) and Achenbach et al. (2018) use volumetric CT head scans and surface-based head scans in order to learn a combined statistical model of the head surface, the skull surface, and the enclosed soft tissue, which allows them to estimate the head surface from the skull shape and vice versa. Regarding the other parts of the body, Zhu et al. (2015) propose an anatomical model of the upper and lower limbs that can be fit to surface scans and is able to reconstruct motions of the limbs.

To our knowledge, there are just two former approaches for generating an anatomical model of the *complete core* human body (torso, arms, legs) from a given skin surface. In their pioneering work, Dicko et al. (2013) transfer the anatomic details from a template model to various humanoid target models, ranging from realistic body shapes to stylized non-human characters. They transfer the template's anatomy through a harmonic space warp and per-bone affine transformations, which, however, might distort muscles

and bones in an implausible way. Different distributions of subcutaneous fat can be (and have to be) painted manually into a special fat texture. The work of Kadleček et al. (2016) is most closely related to our approach. They build an anatomically plausible volumetric model from a set of 3D scans of a person in different poses. An inverse physics simulation is used to fit a volumetric anatomical template model to the set of surface scans, where custom constraints prevent muscles and bones from deforming in an unnatural manner. We discuss the main differences of our approach and Dicko et al. (2013) and Kadleček et al. (2016) in Section 4.

Estimating the body composition from surface measures or 3D surface scans (like we do in Section 3.3) has been tackled before. There are numerous formulas for computing body fat percentage (BF), or body composition in general, from certain circumferences, skinfold thicknesses, age, gender, height, weight, and density measurements. Prominent examples are the skinfold equations, or the Siri- and Brozek formulas (Jackson and Pollock, 1985; Siri, 1956; Brožek et al., 1963). These formulas, however, either rely on anthropometric measurements that have to be taken by skilled personnel or on measuring the precise body density via expensive devices, such as BOD PODs (Fields et al., 2002). Ng et al. (2016) compute BF based on a 3D body scan of the subject, but their formula is tailored towards body scans and measurements taken with the Fit3D Scanner (Fit3D, 2021). Even with the help of the authors we could not successfully apply their formulas to scans taken with different systems, since we could always find examples resulting in obviously wrong (or even negative) BF. Recently, Maalin et al. (2020) showed that modelling body composition through body fat alone is an inferior measure for defining the shape of a person compared to a combined model of fat mass and muscle mass. We therefore adapt their data to estimate fat mass and muscle mass from surface scans alone (Section 3.3). Incorporating these estimations into the volumetric fitting process allows us to determine how much of the soft tissue layer is described by muscle tissue more plausibly than Kadleček et al. (2016).

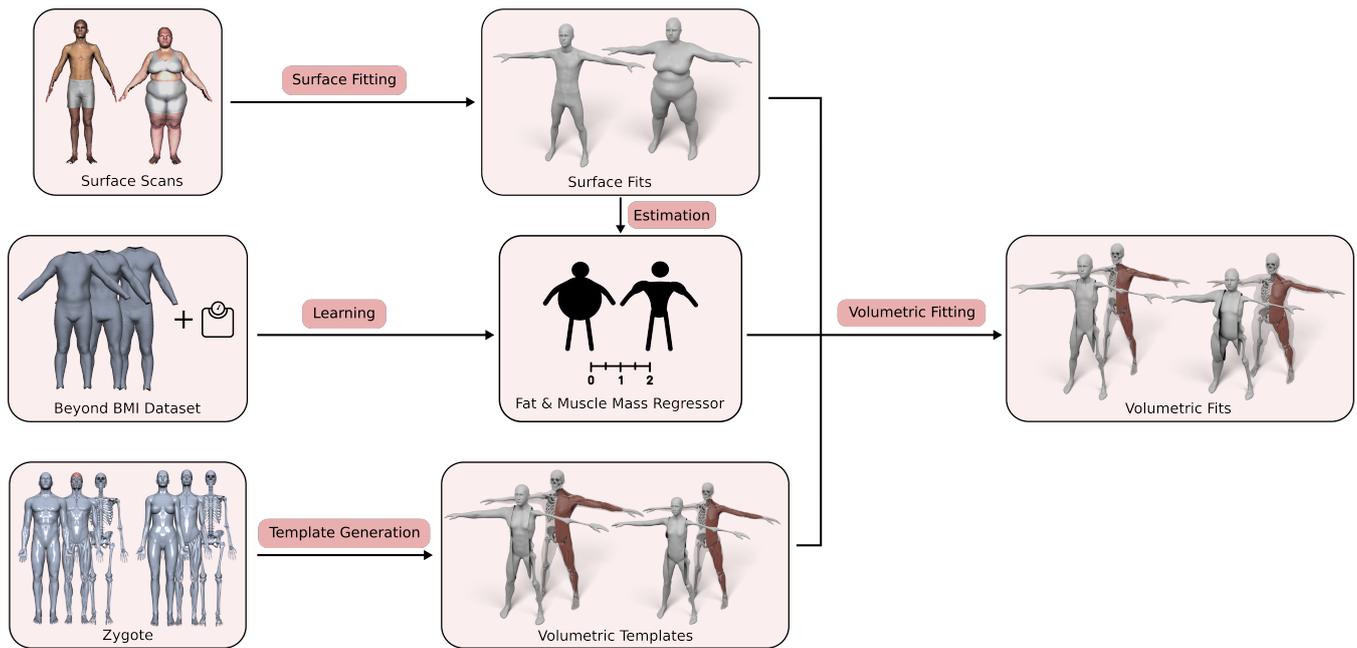
### 3 METHOD

Our approach consists of three main contributions: First, the generation of the volumetric three-layer template, described in Section 3.2, where we derive the skin, muscle, and skeleton layers from the male and female Zygote model (Zygote, 2020). Second, an efficient method for fitting this layered model (including all contained anatomical details) (in)to a given human surface scan (Section 3.4). Third, the estimation of a person’s body composition, i.e., how much of the person’s soft tissue is described by muscles and fat (Section 3.3). By adapting the BeyondBMI dataset (Maalin et al., 2020) to our template, we derive this information from the surface scan alone and use it to inform the volumetric template fitting. Figure 2 shows an overview of the whole process, starting from the different input data sets, the template model and the muscle/fat regressor, to the final personalized anatomical fit.

#### 3.1 Data Preparation

In our approach we make use of several publicly or commercially available datasets for model generation, model learning, and evaluation:

- *Zygote*: The Zygote model (Zygote, 2020) provides high-resolution models for the male and female anatomy. We use their skin, muscle, and skeleton models for building our layered template.
- *BeyondBMI*: Maalin et al. (2020) scanned about 400 people and additionally measured their fat mass (FM), muscle mass (MM), and body mass index (BMI) using a medical-grade eight-electrode bioelectrical impedance analysis. They provide annotated (synthetic) scans of 100 men and 100 women, each computed by averaging shape and annotations of two randomly chosen subjects. From this data we learn a regressor that estimates fat and muscle mass from the skin surface.

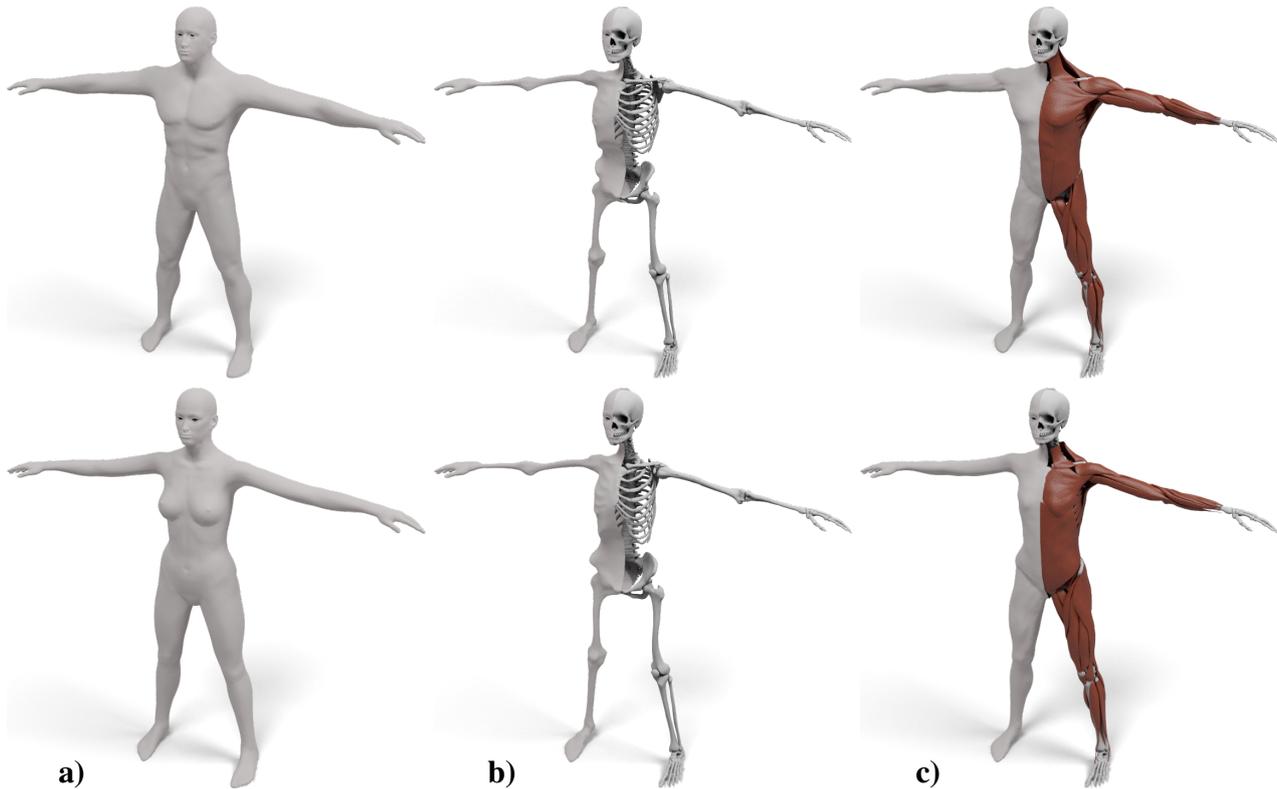


**Figure 2.** Overview of our volumetric template fitting approach. From the Zygote model (Zygote, 2020), we build layered volumetric templates for the male and female anatomy. By adapting the BeyondBMI dataset (Maalin et al., 2020) we learn a model for estimating fat and muscle mass from a surface model. Given a person's surface scan, we then estimate its fat/muscle mass and use this information to fit the volumetric template (in) to the surface scan, which yields the personalized anatomical model.

- *Hasler*: The dataset of Hasler et al. (2009) contains scans of 114 subjects in 35 different poses, captured by a 3D laser scanner. The scans are annotated with fat and muscle mass percentage as measured by a consumer-grade impedance spectroscopy body fat scale. We use this dataset to evaluate the regressor learned from the BeyondBMI data.
- *CAESAR*: The European subset of the CAESAR scan database (Robinette et al., 2002) consists of 3D scans (with about 70 selected landmarks) equipped with annotations (e.g., weight, height, BMI) of about 1700 subjects in a standing pose. We use this data to evaluate our overall fitting procedure.

All these data sources use different model representations, i.e., either different mesh tessellations or even just point clouds. In a preprocessing step we therefore re-topologize the skin surfaces of these datasets to a common triangulation by fitting a surface template using the non-rigid surface-based registration of Achenbach et al. (2017).

This approach is based on an animation-ready, fully rigged, statistical template model. Its mesh tessellation (about 21k vertices), animation skeleton, and skinning weights come from the Autodesk Character Generator (Autodesk, 2014). It uses a 10-dimensional PCA model representing the human body shape variation and we will call it the *surface template* in the following. In a preprocessing step we fit the surface template to all input surface scans to achieve a common triangulation and thereby establish dense correspondence. This fitting process is guided by a set of landmarks, which are either specified manually or provided by the dataset. A nonlinear optimization then determines alignment (scaling, rotation, translation), body shape (PCA parameters), and pose (inverse kinematics on joint angles) in order to minimize squared distances of user-selected landmarks and automatically determined closest point correspondences in a non-rigid ICP manner (Bouaziz et al., 2014b). Once the model parameters are optimized, a fine-scale



**Figure 3.** Our layered template for both male (top) and female (bottom): the skin surface (a), the skeleton surface enveloping the skeleton (b), and the muscle surface enveloping both muscles and skeleton (c). For (b) and (c) the left half shows the enveloping surface, the right half the enveloped anatomical details.

out-of-model deformation improves the matching accuracy and results in the final template fit. For more details we refer to (Achenbach et al., 2017).

### 3.2 Generating the Volumetric Template

We use the male and female Zygote body model (Zygote, 2020) as a starting point for our volumetric model. Our volumetric template is defined by the *skeleton surface*  $\mathcal{B}$  (for bones), the *muscle surface*  $\mathcal{M}$ , and the *skin surface*  $\mathcal{S}$ . The skeleton is enveloped by the skeleton surface, the muscle layer is enclosed between the skeleton surface and the muscle surface, and the (subcutaneous) fat layer is enclosed by the muscle surface and the skin surface. The soft-tissue layer is the union of the fat and muscle layers. In our layered model we exclude the head, hands, and toes. These regions will be identical to the skin surface in all layers. See Figure 3 for a visualization of the layered template.

The three surfaces  $\mathcal{B}$ ,  $\mathcal{M}$ , and  $\mathcal{S}$  will be constructed to share the same triangulation, providing a straightforward one-to-one correspondence between the  $i^{\text{th}}$  vertices on each surface, which we denote by  $\mathbf{x}_i^{\mathcal{B}}$ ,  $\mathbf{x}_i^{\mathcal{M}}$ , and  $\mathbf{x}_i^{\mathcal{S}}$ , respectively. Each two corresponding triangles  $(\mathbf{x}_i^{\mathcal{S}}, \mathbf{x}_j^{\mathcal{S}}, \mathbf{x}_k^{\mathcal{S}})$  on  $\mathcal{S}$  and  $(\mathbf{x}_i^{\mathcal{M}}, \mathbf{x}_j^{\mathcal{M}}, \mathbf{x}_k^{\mathcal{M}})$  on  $\mathcal{M}$  span a volumetric element of the fat layer. Similarly, the volumetric elements of the muscle layer are spanned by pairs of triangles  $(\mathbf{x}_i^{\mathcal{M}}, \mathbf{x}_j^{\mathcal{M}}, \mathbf{x}_k^{\mathcal{M}})$  on  $\mathcal{M}$  and  $(\mathbf{x}_i^{\mathcal{B}}, \mathbf{x}_j^{\mathcal{B}}, \mathbf{x}_k^{\mathcal{B}})$  on  $\mathcal{B}$ . We call these elements, built from six vertices of two triangles, *prisms*, and will either use them directly in a simulation or (trivially) split them into three tetrahedra each, resulting in a simple conforming volumetric tessellation.

The following two sections describe how to generate the skeleton surface  $\mathcal{B}$  (Section 3.2.1) and the muscle surface  $\mathcal{M}$  (Section 3.2.2). The skin surface  $\mathcal{S}$  is generated by fitting the surface-based template of Achenbach et al. (2017) to the skin of the anatomical model (Zygote, 2020), as described in Section 3.1.

### 3.2.1 The Skeleton Surface

The skeleton surface  $\mathcal{B}$  should enclose all the bones of the detailed skeleton model, as shown in Figure 3, center. We achieve this by shrink-wrapping the skin surface  $\mathcal{S}$  onto the skeletal bones. To avoid problems caused by gaps between bones (e.g. rib-cage, tibia/fibula), we first generate a skeleton wrap  $\mathcal{W}$ , a watertight genus-0 surface that encapsulates the bones, and then shrink-wrap the skin surface to  $\mathcal{W}$  instead. The wrap surface  $\mathcal{W}$  can easily be generated by a few iterations of shrink-wrapping, remeshing, and smoothing of a bounding sphere in a 3D modeling software like Blender or Maya. This results in a smooth, watertight, and two-manifold surface  $\mathcal{W}$  that excludes regions like the interior of the rib-cage and small holes like in the pelvis or between ulna and radius.

We generate the skeleton surface  $\mathcal{B}$  by starting from the skin surface  $\mathcal{S}$ , i.e., setting  $\mathcal{X} = \mathcal{S}$ , and then minimizing a nonlinear least squares energy that is composed of a fitting term, which attracts the surface  $\mathcal{X}$  to the bone wrap  $\mathcal{W}$ , and a regularization term, which prevents  $\mathcal{X}$  from deforming in a physically implausible manner from its initial state  $\bar{\mathcal{X}} = \mathcal{S}$ :

$$\mathcal{B} = \arg \min_{\mathcal{X}} w_{\text{fit}} E_{\text{fit}}(\mathcal{X}, \mathcal{W}) + w_{\text{reg}} E_{\text{reg}}(\mathcal{X}, \bar{\mathcal{X}}). \quad (1)$$

The regularization is formulated as a discrete bending energy that penalizes the change of mean curvature, measured as the change of length of the Laplacian:

$$E_{\text{reg}}(\mathcal{X}, \bar{\mathcal{X}}) = \sum_{\mathbf{x}_i \in \mathcal{X}} A_i \|\Delta \mathbf{x}_i - \mathbf{R}_i \Delta \bar{\mathbf{x}}_i\|^2, \quad (2)$$

where  $\mathbf{x}_i$  and  $\bar{\mathbf{x}}_i$  denote the vertex positions of the deformed surface  $\mathcal{X}$  and the initial surface  $\bar{\mathcal{X}}$ , respectively. The matrix  $\mathbf{R}_i \in SO(3)$  denotes the optimal rotation aligning the vertex Laplacians  $\Delta \mathbf{x}_i$  and  $\Delta \bar{\mathbf{x}}_i$ , which are discretized using the cotangent weights and the Voronoi areas  $A_i$  (Botsch et al., 2010).

The fitting term penalizes the squared distance of vertices  $\mathbf{x}_i \in \mathcal{X}$  from their target positions  $\mathbf{t}_i \in \mathcal{W}$ :

$$E_{\text{fit}}(\mathcal{X}, \mathcal{W}) = \sum_{\mathbf{x}_i \in \mathcal{X}} w_i A_i \|\mathbf{x}_i - \mathbf{t}_i\|^2. \quad (3)$$

The target positions  $\mathbf{t}_i$  are points (not necessarily vertices) on the skeleton wrap  $\mathcal{W}$  of either one of three types: closest point correspondences, fixed correspondences, or collision targets. The weight  $w_i$  is determined solely by the type of target position  $\mathbf{t}_i$  (0.1 for closest point correspondences, 1 for fixed correspondences, 100 for collision targets). We define just one target  $\mathbf{t}_i$  for each vertex  $\mathbf{x}_i$ . The default is a closest point correspondence per vertex, which can be overridden by a fixed correspondence, and both of them will be overridden by the collision target in case of a detected collision. Below we explain the three target types.

*Closest point correspondences* are updated in each iteration of the minimization to the closest position on  $\mathcal{W}$  to the vertex  $\mathbf{x}_i \in \mathcal{X}$ , i.e.,  $\mathbf{t}_i = \arg \min_{\mathbf{y} \in \mathcal{W}} \|\mathbf{x}_i - \mathbf{y}\|$ .

Near complicated regions, like the armpit or the rib-cage, the skin has to stretch considerably to deform toward the skeleton wrap. As a consequence, corresponding triangles  $(\mathbf{x}_i^{\mathcal{S}}, \mathbf{x}_j^{\mathcal{S}}, \mathbf{x}_k^{\mathcal{S}})$  on the skin surface  $\mathcal{S}$  and  $(\mathbf{x}_i^{\mathcal{B}}, \mathbf{x}_j^{\mathcal{B}}, \mathbf{x}_k^{\mathcal{B}})$  on the eventual skeleton surface  $\mathcal{B}$  will not be approximately on top of each other, but instead be tangentially shifted. These two triangles span a volumetric element that we call a *prism*. Misaligned triangles will lead to heavily sheared prisms, which can cause artifacts in physical simulations.

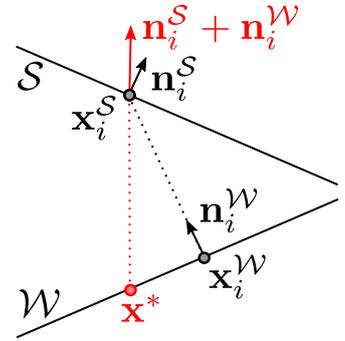


**Figure 4.** Standard nonrigid registration from skin to skeleton (left) results in a bad tangential alignment of corresponding triangles, causing sheared prisms, which we visualize by color-coding the alignment error (4). Using fixed correspondences reduces this error (center). Also shifting closest point correspondences with bad alignment reduces the error even further (right).

We define a per-vertex score penalizing misalignment of corresponding vertices  $\mathbf{x}_i^{\mathcal{S}} \in \mathcal{S}$  and  $\mathbf{x}_i^{\mathcal{W}} \in \mathcal{W}$  w.r.t. their common averaged normal  $\mathbf{n}_i^{\mathcal{S}} + \mathbf{n}_i^{\mathcal{W}}$ :

$$E_{\text{align}}(\mathbf{x}_i^{\mathcal{S}}, \mathbf{x}_i^{\mathcal{W}}) = \left| \frac{(\mathbf{n}_i^{\mathcal{S}} + \mathbf{n}_i^{\mathcal{W}}) \cdot (\mathbf{x}_i^{\mathcal{S}} - \mathbf{x}_i^{\mathcal{W}})}{\|\mathbf{n}_i^{\mathcal{S}} + \mathbf{n}_i^{\mathcal{W}}\| \cdot \|\mathbf{x}_i^{\mathcal{S}} - \mathbf{x}_i^{\mathcal{W}}\|} - 1 \right|. \quad (4)$$

A 2D example of this is shown on the right, where the closest correspondence of  $\mathbf{x}_i^{\mathcal{S}}$  is  $\mathbf{x}_i^{\mathcal{W}}$ . The position that maximizes the minimal angle at both vertices is  $\mathbf{x}^*$ , where the connecting line (dotted red) aligns with the average normal.



*Fixed correspondences* are responsible for reducing these tangential shifts and thereby improving the prism shapes. We determine them for some vertices at the beginning of the fit as explained in the following and keep them fixed throughout the optimization. Since the alignment error increases faster if the distance between skin surface and skeleton wrap is small, we specify fixed correspondences for vertices on  $\mathcal{S}$  that have a distance less than 3 cm to  $\mathcal{W}$ . For each such vertex we randomly sample points in the geodesic neighborhood of  $\mathbf{x}_i^{\mathcal{W}}$  and select the one that minimizes (4) as fixed alignment constraint, where we generate normal vectors of sample points using barycentric Phong interpolation. To avoid interference of spatially close fixed correspondences, we add them in order of increasing distance to the skeleton, but only if their distance to all previously selected points is larger than 5 cm. In that way, we get a well distributed set of fixed correspondences, favoring those with a small skin-to-skeleton distance. Figure 4, center, shows that this already reduces the alignment error by a large amount.

Closest point correspondences can also drag vertices to locations with high alignment error. In each iteration of the nonrigid ICP, we compute  $E_{\text{align}}(\mathbf{x}_i^{\mathcal{S}}, \mathbf{x}_i)$  for each vertex on  $\mathcal{S}$  and its counterpart on the current state of  $\mathcal{X}$ . If this error exceeds a limit of 0.01, which corresponds to an angle deviation of  $8^\circ$  from the optimal angle, we sample the one-ring neighborhood of vertex  $\mathbf{x}_i$  on  $\mathcal{X}$  and set  $\mathbf{x}_i$  to the sample with minimal alignment error and update its closest point correspondence on  $\mathcal{W}$ . This strategy reduces the alignment error even further, as shown in Figure 4, right.

In the process of moving the surface  $\mathcal{X}$  toward  $\mathcal{W}$ , these two meshes might intersect each other, violating our goal that in the converged state the surface  $\mathcal{X}$  (i.e.,  $\mathcal{B}$ , due to (1)) should fully enclose  $\mathcal{W}$ . We therefore

detect these collisions during the optimization and resolve them through *collision targets*. We use the exact continuous collision detection of Brochu et al. (2012) to detect collisions. In case of a collision, we back-track the triangles' linear path from the current  $\mathcal{X}$  to the initial  $\mathcal{S}$  to find the non-colliding state closest to  $\mathcal{X}$ . This state defines *collision targets*  $t_i$  for colliding vertices  $x_i$ , which override the other types of target positions. In case of multiple collision targets  $t_i$  for the same vertex  $x_i$ , we determine all non-colliding states separately and choose the one that is closest to the initial skin surface  $\mathcal{S}$ . Minimizing (1) leads to the final skeleton surface  $\mathcal{B}$  (Figure 3b). See Appendix 1 for more details about the optimization strategy.

### 3.2.2 The Muscle Surface

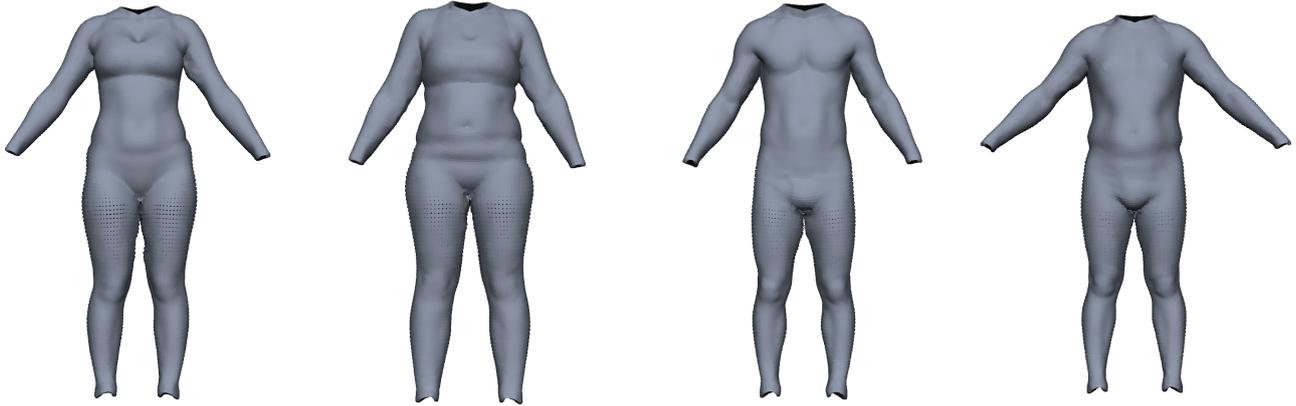
We generate the muscle surface  $\mathcal{M}$  by minimizing the same energy as in Equation (1), but using a different method for finding the correspondences  $t_i$  in Equation (3), which exploits that we already established correspondence between skin surface  $\mathcal{S}$  and skeleton surface  $\mathcal{B}$ . We do not employ closest point correspondences, but instead set for each vertex  $x_i$  a *fixed* correspondence  $t_i$  to the first intersection of the line from skin vertex  $x_i^{\mathcal{S}}$  to skeleton vertex  $x_i^{\mathcal{B}}$  with the high-resolution muscle model (Zygotte, 2020). If there is no intersection (e.g., at the knee), we set  $t_i = x_i^{\mathcal{B}}$  and assign a lower weight  $w_i$ . When the minimization converges and we decrease  $w_{\text{reg}}$ , we project the vertices of the current muscle surface  $x_i^{\mathcal{M}}$  to their corresponding skin-to-skeleton line from  $x_i^{\mathcal{S}}$  to  $x_i^{\mathcal{B}}$ . Due to the collision handling, the resulting muscle surface  $\mathcal{M}$  will enclose the high-resolution muscle model. To ensure that our volumetric elements always have a non-zero volume, even in regions where there is no muscle between skin and bone, we ensure a minimal offset of 1 mm from to the skeleton mesh. The resulting muscle surface  $\mathcal{M}$  is visualized in Figure 3c. Note that the muscle layer does not exclusively contain muscles: Especially in the abdominal region, a large amount of the muscle layer is filled by organs. We therefore define a *muscle thickness map* that for each vertex  $i$  stores the accumulated length of the segments of the line  $(x_i^{\mathcal{S}}, x_i^{\mathcal{B}})$  that are covered by muscles. This map will be used later in Section 3.4.3.

## 3.3 Estimating Fat Mass and Muscle Mass

Having generated the volumetric layered template, we want to be able to fit it to a given surface scan of a person. To regularize this under-determined problem, we first have to estimate how much of the person's soft tissue is explained by fat mass (*FM*) and muscle mass (*MM*), respectively. This is a challenging problem since we want to capture a single surface scan of the person only and therefore cannot rely on information provided by additional hardware, such as a DXA scanner or a body fat scale. Kadleček et al. (2016) handle this problem by describing the person's shape primarily through muscles, i.e., by growing muscles as much as possible and defining the remaining soft tissue volume as fat. This strategy results in adipose persons having considerably more muscle mass than leaner people. Although there is a certain correlation between total body mass (and also BMI) and muscle mass – because the higher weight has a training effect especially on the muscles of the lower limbs (Tomlinson et al., 2016) – this general trend is not sufficient to define the body composition of people.

Maalin et al. (2020) measured both FM and MM using a medical-grade eight-electrode bio-electrical impedance analysis and acquired a 3D surface scan. From this data, they built a model that can vary the shape of a person based on specified muscle or fat variation, similar to Piryankova et al. (2014). Our model should perform the inverse operation, i.e., estimate FM and MM from a given surface scan. We train our model on their BeyondBMI dataset (Section 3.1), which consists of scans of 100 men and 100 women captured in an approximate A-pose (see Figure 5), each annotated with FM, MM, and BMI.

By applying the surface fitting described in Section 3.1 to the BeyondBMI dataset, we make their scans compatible to our template and un-pose their scans to a common T-pose, thereby making any subsequent statistical analysis pose-invariant. After re-excluding the head, hands, and feet of our surface template, we



**Figure 5.** Examples for the BeyondBMI dataset provided by Maalin et al. (2020) consisting of scans of 100 men and 100 women, annotated with fat mass, muscle mass, and BMI. The scans lack geometric data for head, hands, and feet and are captured in approximate A-pose (with noticeable variation in pose).

are left with  $N = 100$  meshes per sex that consist of  $V = 7665$  vertices  $\mathbf{x}_i$ . We denote the  $j^{\text{th}}$  training mesh by a  $3V$ -dimensional vector of stacked vertex coordinates

$$\mathbf{X}_j = \left( \left( \mathbf{x}_1^j \right)^\top, \dots, \left( \mathbf{x}_V^j \right)^\top \right)^\top \in \mathbb{R}^{3V}$$

and perform PCA on the data matrix  $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_N) \in \mathbb{R}^{3V \times N}$ . Let  $\mathbf{P} \in \mathbb{R}^{3V \times k}$  be the basis of the subspace spanned by the first  $k$  principal components and  $\boldsymbol{\mu}$  the mean of the training data. Since the data is now pose-normalized, the dimensionality reduction can focus solely on differences in human body shape. As a result, our model only needs  $k = 12$  PCA components to explain 99.5% of the data variance, while the original BeyondBMI dataset needs  $k = 24$  components to cover the same percentage due to noticeable variations in pose during the scanning process (see Figure 5). We then perform linear regression to estimate FM and MM from PCA weights, as proposed by Hasler et al. (2009).

For a first evaluation of this model, we perform a leave-one-out test on the BeyondBMI dataset, i.e., excluding each scan once, building the regressors as described above from the remaining  $N - 1$  scans, and measuring the mean absolute error of the predictions. We again use  $k = 12$  PCA components, as this covers almost all the variance present in the dataset and gives the linear regression enough degrees of freedom. The leave-one-out evaluation yields a mean absolute error (MAE) of  $\text{MAE}_{\text{FM}} = 1.20 \text{ kg} (\pm 0.93)$  and  $\text{MAE}_{\text{MM}} = 1.01 \text{ kg} (\pm 0.79)$  for the female dataset, where the fat mass lies in the range 6.27–34.71 kg and the muscle mass in the range 21.59–31.63 kg. The linear regression shows an average  $R^2$  score of 0.84, confirming that there is indeed a linear relationship between PCA coordinates and the FM/MM measurements. Performing the leave-one-out test on the male dataset shows similar values:  $\text{MAE}_{\text{FM}} = 1.37 \text{ kg} (\pm 1.00)$  and  $\text{MAE}_{\text{MM}} = 1.46 \text{ kg} (\pm 1.11)$ , fat mass in the range 3.91–27.83 kg, muscle mass in the range 31.51–51.20 kg, and an average  $R^2$  score of 0.88.

We compared the linear model to a support vector regression (using scikit-learn (Pedregosa et al., 2011) with default parameters and RBF kernels), but in contrast to Hasler et al. (2009) we found that for the BeyondBMI dataset this approach performs considerably worse:  $\text{MAE}_{\text{FM}} = 2.98 \text{ kg} (\pm 2.85)$  and  $\text{MAE}_{\text{MM}} = 1.24 \text{ kg} (\pm 1.02)$  with an average  $R^2$  score of 0.64 for the female dataset, and  $\text{MAE}_{\text{FM}} =$

2.63 kg ( $\pm 2.60$ ) and  $MAE_{MM} = 2.48$  kg ( $\pm 1.82$ ) with an average  $R^2$  score of 0.58 for the male dataset. We therefore keep the simpler and better-performing linear regression model.

Whenever we fit the volumetric model to a given body scan, as explained in the next section, we first use the proposed linear regressors to estimate the person's fat mass and muscle mass and use this information to generate the muscle and fat layers in Section 3.4.3.

### 3.4 Fitting the Volumetric Template to Surface Scans

Given a surface scan, we transfer the template anatomy into it through the following steps: First, we fit our *surface* template to the scan, which establishes one-to-one correspondence with the volumetric template and puts the scan into the same T-pose as the template (Section 3.1). After this pre-processing, we deform the *volumetric* template to match the scanned subject. To this end, we adjust global scaling and per-bone local scaling, such that body height and limb lengths of template and scan match (Section 3.4.1). This is followed by a quasi-static deformation of the volumetric template that considers the skin surface  $\mathcal{S}$  as hard constraint and yields the skeleton surface  $\mathcal{B}$  through energy minimization (Section 3.4.2). Given the skin surface  $\mathcal{S}$ , the bone surface  $\mathcal{B}$ , and the estimated fat mass and muscle mass from Section 3.3, the muscle surface  $\mathcal{M}$  is determined (Section 3.4.3). Having transferred all three layer surfaces to the scan we finally warp the detailed anatomical model to the target (Section 3.4.4).

#### 3.4.1 Global and Local Scaling

Fitting the surface template to the scanner data puts the latter into the same alignment (rotation, translation) and the same pose as the volumetric template. The next step is to correct the mismatch in scale by adjusting body height and limb lengths of the volumetric template.

This scaling does influence all three of the template's surfaces. Since the shape of the skeleton surface  $\mathcal{B}$  will be constrained to the result after scaling, we have to scale in a way that keeps bone lengths and bone diameters within a plausible range. The *length* of prominent bones, like the upper arm or the upper leg (humerus and femur), can be well approximated by measures on the surface of the model. But finding the correct bone *diameters* is impossible without measurements of the subject's interior. In particular for corpulent or adipose subjects, the subcutaneous fat layer dominates the appearance of the skin surface, preventing us from precisely determining the bone diameters from the surface scan. It has been shown that there is a moderate correlation of bone length and bone diameter (Aydin Kabakci et al., 2017; Ziylan and Murshid, 2002) and (obviously) a strong correlation of body height and bone length (Dayal et al., 2008). We therefore perform a *global isotropic* scaling depending on body height (affecting bone lengths and diameters) as well as *local anisotropic* scaling depending on limb lengths (affecting bone lengths only).

The global scaling is determined from the height difference of scan and template and is applied to all vertices of the template model. It therefore scales all bone lengths and bone diameters uniformly. Directly scaling with the height ratio of scan and template, however, can result in bones too thin or too thick for extreme target heights. Thus, we damp the height ratio  $r = h_{\text{scan}}/h_{\text{template}}$  by  $r \leftarrow 0.5(r - 1) + 1$ , which means that a person that is 20% taller than the template will have 10% thicker bones than the template. This heuristic results in visually plausible bone diameters for all our scanned subjects.

After the global scaling, the local scaling further adjusts the limb lengths of the template to match those of the scan. The (fully rigged) surface-based template has been fit to both the scan (Section 3.1) and the template (Section 3.2). This fit provides a simple skeleton graph (used for skinning animation) for both models. We use the length mismatch of the respective skeleton graph segments to determine the required scaling for upper and lower arms, upper and lower legs, feet, and torso. We scale these limbs in their



**Figure 6.** Scaling the template (opaque) to match the scan (semi-transparent): The pre-processing aligns the scan with the template and puts it into the same pose (left). Body height and limb lengths of the template are then adjusted by a global uniform scaling (center), followed by local scaling for limbs and spine (right).

corresponding bone directions (or the spine direction for the torso) using the bone stretching of Kadleček et al. (2016). As mentioned before, this changes the limb lengths but not the bone diameters.

This two-step scaling process is visualized in Figure 6. As a result, the scaled template matches the scan with respect to alignment, pose, body height, and limb lengths. Its layer surfaces, which we denote by  $\bar{\mathcal{S}}$ ,  $\bar{\mathcal{M}}$ , and  $\bar{\mathcal{B}}$ , provide a good initialization for the optimization-based fitting described in the following.

### 3.4.2 Skeleton Fitting

Given the coarse registration of the previous step, we now fit the skin surface  $\mathcal{S}$  and skeleton surface  $\mathcal{B}$  by minimizing a quasi-static deformation energy. Since the template’s skin surface  $\mathcal{S}$  should match the (skin) surface of the scan and since both meshes have the same triangulation, we can simply copy the skin vertex positions from the scan to the template and consider them as hard Dirichlet constraints. It therefore remains to determine the vertex positions of the skeleton surface  $\mathcal{B}$ , such that the soft tissue enclosed between skin surface  $\mathcal{S}$  and skeleton surface  $\mathcal{B}$  (fat + muscles, which we call *flesh*) deforms in a physically plausible manner. This is achieved by minimizing a quasi-static energy consisting of three terms:

$$E(\mathcal{B}) = w_{\text{reg}} E_{\text{reg}}(\mathcal{B}, \bar{\mathcal{B}}) + w_{\text{flesh}} E_{\text{flesh}}(\mathcal{B}, \mathcal{S}) + w_{\text{coll}} E_{\text{coll}}(\mathcal{B}, \mathcal{S}). \quad (5)$$

The first term is responsible for keeping the skeleton surface (approximately) rigid and uses the same formulation as Equation (2), with  $\bar{\mathcal{B}}$  and  $\mathcal{B}$  denoting the skeleton surface before and after the deformation, respectively. We employ a soft constraint with high weight  $w_{\text{reg}}$  instead of deforming bones in a strictly rigid manner (Kadleček et al., 2016), since we noticed that for very thin subjects the skeleton surface might otherwise protrude the skin surface and therefore a certain amount of bone deformation is required. We also do not penalize deviation from rigid or affine transformations as Dicko et al. (2013) since this penalizes smooth shape deformation in the same way as locally flipped triangles, which we observed to cause artifacts in the skeleton surface. The discrete bending energy of Equation (2), with a suitably high regularization weight  $w_{\text{reg}}$ , allows for moderate *smooth* deformations and gave better results in our experiment.

The second term prevents strong deformations of the prism elements  $p \in \mathcal{P}$ , spanned by corresponding triangles  $(\mathbf{x}_i^{\mathcal{S}}, \mathbf{x}_j^{\mathcal{S}}, \mathbf{x}_k^{\mathcal{S}})$  on the skin surface and  $(\mathbf{x}_i^{\mathcal{B}}, \mathbf{x}_j^{\mathcal{B}}, \mathbf{x}_k^{\mathcal{B}})$  on the skeleton surface. While we penalize deformation of the top/bottom triangles, we allow changes of prism heights, i.e., anisotropic scaling in

the direction from surface to bone, since otherwise the fat layer cannot grow to bridge the gap from the skeleton surface to the skin surface. This behavior is modelled by the anisotropic strain limiting energy

$$E_{\text{flesh}}(\mathcal{B}, \mathcal{S}) = \frac{1}{2} \sum_{p \in \mathcal{P}} \left\| \mathbf{F}_p - \mathbf{R}_p \mathbf{B}_p \tilde{\mathbf{S}}_p \mathbf{B}_p^\top \right\|_F^2, \quad (6)$$

where  $\mathbf{F}_p \in \mathbb{R}^{3 \times 3}$  is the deformation gradient of the element  $p$ , i.e., the linear part of the best affine transformation that maps the un-deformed prism  $\bar{p}$  to the deformed prim  $p$  in the least squares sense. If  $\mathbf{D}_p \in \mathbb{R}^{3 \times 5}$  denotes the edge direction matrix of the prism  $p$  and  $\bar{\mathbf{D}}_p$  the respective matrix of  $\bar{p}$ , then  $\mathbf{F}_p = \arg \min_{\mathbf{F}} \|\mathbf{D}_p - \mathbf{F} \bar{\mathbf{D}}_p\|_F^2$ . Polar decomposition (Shoemake and Duff, 1992)  $\mathbf{F}_p = \mathbf{R}_p \mathbf{S}_p$  decomposes  $\mathbf{F}_p$  into a rotation  $\mathbf{R}_p$  and scale/shear  $\mathbf{S}_p$ .  $\mathbf{B}_p$  is a rotation matrix that aligns the z-axis with the surface normal of the prism's corresponding skin triangle, i.e., the direction in which we allow stretching. The matrix  $\tilde{\mathbf{S}}_p$  represents the anisotropic scaling  $\text{diag}(1, 1, \alpha)$ , where  $\alpha \in [\alpha_{\min}, \alpha_{\max}]$  allows to tune the amount of stretching in normal direction that should be allowed. We use  $\alpha_{\min} = 0.2$  and  $\alpha_{\max} = 5.0$  to allow stretching and compression of the element by a factor of five before the energy of this element increases.

Third, we detect all collisions  $\mathcal{C}$ , defined as vertices of the skeleton surface  $\mathcal{B}$  that are outside of the skin surface  $\mathcal{S}$ . For these colliding vertices we add a collision penalty term

$$E_{\text{coll}}(\mathcal{B}, \mathcal{S}) = \frac{1}{2} \sum_{\mathbf{x}_i \in \mathcal{C}} w_i \|\mathbf{x}_i - \pi_{\mathcal{S}}(\mathbf{x}_i)\|^2, \quad (7)$$

where  $\pi_{\mathcal{S}}(\mathbf{x}_i)$  is the projection of the colliding vertex  $\mathbf{x}_i$  to a position 2 mm beneath the closest triangle on the skin surface  $\mathcal{S}$ . The weight  $w_i$  is defined per vertex, is set to 1 the first time a vertex is colliding, and is increased by 1 each time the minimization was not able to resolve the collision. The iterative minimization of (5) as well as the computation of the individual elements of (6) is further detailed in Appendix 1.

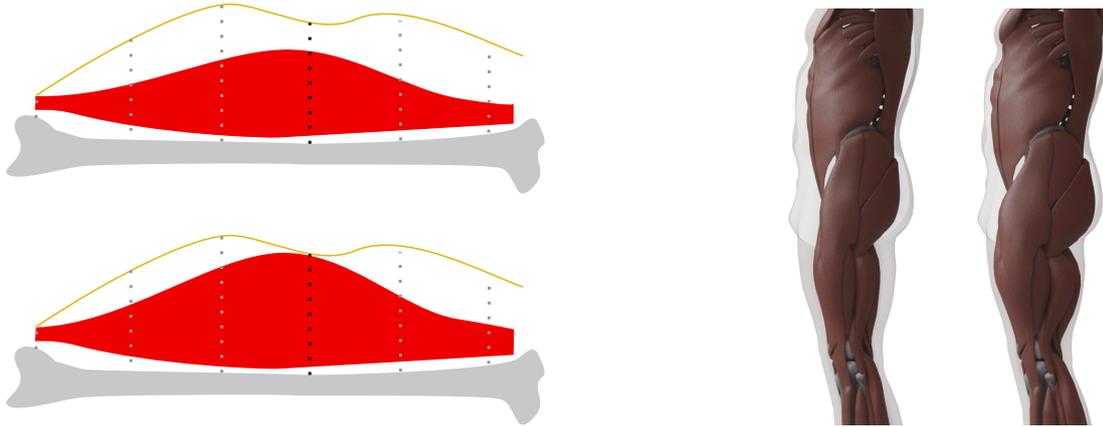
### 3.4.3 Muscle Fitting

Having determined the skin surface  $\mathcal{S}$  and skeleton surface  $\mathcal{B}$ , we now fit the muscle surface  $\mathcal{M}$  in between  $\mathcal{S}$  and  $\mathcal{B}$ , such that the ratio of fat mass (FM) and muscle mass (MM) resembles the values estimated by our regressors (Section 3.3). We proceed in three steps: First, we transfer the template's muscle distribution to the fitted skin and skeleton surfaces, which we call *average muscle layer* in the following. Second, we grow and shrink the muscles as much as anatomically and physically plausible, yielding the *minimum* and *maximum muscle layers*. Third, we find a linear interpolation between these two extremes that matches the predicted fat mass and muscle mass as good as possible.

The average muscle surface is transferred from the *scaled* template  $\bar{\mathcal{M}}$  (Section 3.4.1, Figure 6) by minimizing an energy consisting of two objectives:

$$E(\mathcal{M}) = w_{\text{reg}} E_{\text{reg}}(\mathcal{M}, \bar{\mathcal{M}}) + w_{\text{line}} E_{\text{line}}(\mathcal{M}, \mathcal{B}, \mathcal{S}). \quad (8)$$

The first term tries to preserve the shape of the scaled template's muscle surface  $\bar{\mathcal{M}}$  and is modelled using the regularization energy of Equation (2). The second term preserves the template's property that each



**Figure 7.** Left: When computing the maximum muscle surface, we move muscle vertices toward the skin by an amount proportional to their *muscle potential*, which for each vertex is the length of the dotted line intersected with the muscle. The vertex with the black dotted line defines the maximum allowed stretch in this example. Right: An example of our minimum and maximum muscle layers for the same target. These two surfaces define the lower and upper limit for the muscle mass and vice versa for the fat mass.

muscle vertex  $\mathbf{x}_i^{\mathcal{M}}$  resides on the line segment from its corresponding skeleton vertex  $\mathbf{x}_i^{\mathcal{B}}$  to its skin vertex  $\mathbf{x}_i^{\mathcal{S}}$ , by penalizing the squared distance from that line:

$$E_{\text{line}}(\mathcal{M}, \mathcal{B}, \mathcal{S}) = \frac{1}{2} \sum_{\mathbf{x}_i \in \mathcal{M}} \left\| \mathbf{x}_i - \pi(\mathbf{x}_i, \mathbf{x}_i^{\mathcal{B}}, \mathbf{x}_i^{\mathcal{S}}) \right\|^2, \quad (9)$$

where  $\pi(\mathbf{x}_i, \mathbf{x}_i^{\mathcal{B}}, \mathbf{x}_i^{\mathcal{S}})$  is the projection of  $\mathbf{x}_i$  onto the line  $(1 - \alpha)\mathbf{x}_i^{\mathcal{B}} + \alpha\mathbf{x}_i^{\mathcal{S}}$ ,  $\alpha \in [0, 1]$ . Minimizing (8) leads to flat abdominal muscles like in the template model, which is unrealistic for corpulent or adipose subjects, because the majority of body fat resides in two different fat tissues: the *subcutaneous fat*, which resides between skin and muscle surface, and the *visceral fat*, which accumulates in the abdominal cavity, i.e., under the muscle layer. Since the bulging of the abdomen due to visceral fat causes a bulging of the belly, we inversely want the abdominal muscles in  $\mathcal{M}$  to slightly bulge out in case of a belly bulge in the skin surface  $\mathcal{S}$ . The latter is a combined effect of visceral and subcutaneous fat in the abdominal region. We model this effect by adjusting  $E_{\text{line}}$  for each vertex  $\mathbf{x}_i$  in the abdominal region. Instead of using the full interval  $\alpha \in [0, 1]$ , we adjust the lower boundary to  $\alpha_{\min} = \frac{\|\bar{\mathbf{x}}_i^{\mathcal{M}} - \bar{\mathbf{x}}_i^{\mathcal{B}}\|}{\|\bar{\mathbf{x}}_i^{\mathcal{S}} - \bar{\mathbf{x}}_i^{\mathcal{B}}\|}$ , i.e., the parameter  $\alpha$  where for the (scaled) template the muscle surface intersects the line. The iterative minimization of (8) is further detailed in Appendix 1.

Having transferred the average muscle surface, we next grow/shrink muscles as much as possible in order to define the maximum/minimum muscle surfaces. Since certain muscle groups might be better developed than others, we perform the muscle growth/shrinkage separately for the major muscle groups, namely upper legs (including buttocks), lower legs, upper arms, lower arms, chest, abdominal muscles, shoulders, and back. Muscles are built from fibres and grow perpendicular to the fibre direction. In all cases relevant for us, the fibres are approximately perpendicular to the direction from  $\mathcal{M}$  to  $\mathcal{S}$ , thus muscle growth/shrinkage will move vertices  $\mathbf{x}_i^{\mathcal{M}}$  along the line from  $\mathbf{x}_i^{\mathcal{B}}$  to  $\mathbf{x}_i^{\mathcal{S}}$ . The amount of vertex movement along these directions is proportional to the muscle thickness map of the template (computed in Section 3.2.2). We determine how much we can grow a muscle before it collides with the skin surface in the thicker parts of the muscle (instead of close to its endpoints where it connects to the bone). Figure 7 shows an example, where the leftmost muscle vertex is already close to the skin and would prevent any growth if we took endpoint

regions into account. For each muscle group, we also define an upper limit for muscle growth that prevents the muscles from increasing further even if the skin distance is large (e.g., for adipose subjects). For determining the minimal muscle surface, we repeat the process in the opposite direction (towards the skeleton surface). To prevent distortions of the muscle surface, we do not set the new vertex positions directly, but instead use them as target positions  $t_i$  (using Equation (3)) and regularize with Equation (8). Figure 7 (right) shows an example of minimum/maximum muscle surfaces computed by this procedure.

We determine the final muscle surface  $\mathcal{M}$  by linear interpolation between the minimum and maximum muscle surfaces, such that the resulting fat mass FM and muscle mass MM match the values predicted by the regressors (denoted by  $FM^*$  and  $MM^*$ ) as good as possible. To this end we have to compute FM and MM from an interpolated muscle surface  $\mathcal{M}$ . We can compute the volume  $V_{FL}$  of the fat layer (between  $\mathcal{S}$  and  $\mathcal{M}$ ) and the volume  $V_{ML}$  of the muscle layer (between  $\mathcal{M}$  and  $\mathcal{B}$ ) and convert these to masses  $m_{FL}$  and  $m_{ML}$  by multiplying with the (approximate) fat and muscle densities  $\rho_F = 0.9 \text{ kg/l}$  and  $\rho_M = 1.1 \text{ kg/l}$ , respectively.

The resulting masses require some corrections though: First, we have to add the visceral fat (VAT), which is not part of our fat layer but resides in the abdominal cavity. We estimate the VAT mass  $m_{VAT}$  by computing the difference of the cavity volumes of the scaled template and of the final fit, thereby assuming a negligible amount of VAT in the template. Second, we subtract the skin mass  $m_{skin}$  from the fat layer mass. We assume an average skin thickness of 2 mm, multiply this by the skin's surface area and the density  $\rho_F$ . Third, our fat layer includes the complete reproductive apparatus in the crotch region. This volume is even larger due to the underwear that was worn during scanning and incorrectly increases the fat layer mass by  $m_{crotch}$ . Our corrected fat mass is then

$$FM = m_{FL} + m_{VAT} - m_{skin} - m_{crotch}. \quad (10)$$

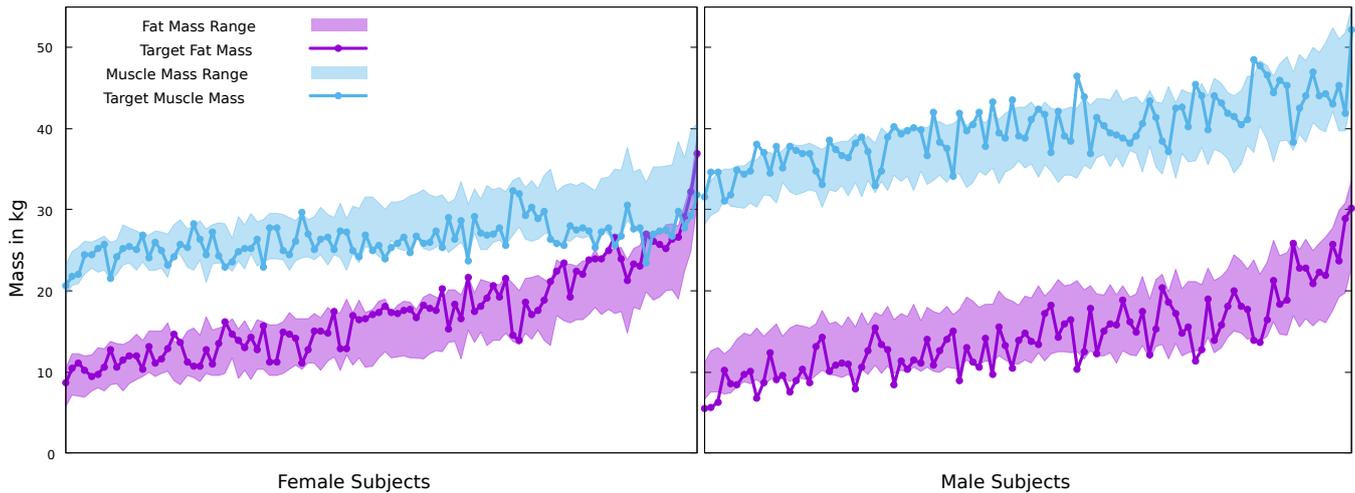
We correct the muscle mass by subtracting the mass  $m_{abd}$  of the abdominal cavity, which is incorrectly included in the muscle layer. The remaining muscle mass is always too small even when using the maximum muscle surface, due to all muscles not considered in the muscle layer, such as heart, face, and hand muscles or the diaphragm. It is known that the lean body mass roughly scales with the squared body height (Heymsfield et al., 2011), which is the basis of the well known body and muscle mass indices. We analogously assume the missing muscle mass to be proportional to the squared height  $h$  of the subject, i.e.,  $m_h = kh^2$ , with a constant  $k$  to be determined later. The corrected muscle mass is therefore

$$MM = m_{ML} - m_{abd} + m_h. \quad (11)$$

There are other terms like the fat of head, hands, and toes, which could be added, or the volume of blood vessels and tendons, which could be subtracted. We assume those terms to be negligible.

Since the total volume of the soft tissue layer  $V_{ST} = V_{ML} + V_{FL}$  is constant, the muscle layer mass  $m_{ML}$  is coupled to the fat layer mass  $m_{FL}$  via  $m_{ML} = (V_{ST} - V_{FL})\rho_M$ . We want to compute the fat layer mass such that the resulting FM and MM minimize the least squares error to the values predicted by the regressor:  $E = (FM - FM^*)^2 + (MM - MM^*)^2$ . Inserting (10) and (11) into  $E$ , rewriting  $m_{ML}$  in terms of  $m_{FL}$ , and setting the derivative  $dE/dm_{FL} = 0$  yields the optimal fat layer mass

$$m_{FL} = \frac{FM^* - m_{VAT} + m_{skin} + m_{crotch} + \rho(V_{ST}\rho_M - m_{abd} + m_h - MM^*)}{1 + \rho^2}, \quad (12)$$



**Figure 8.** True muscle and fat masses for the female and male subjects of the BeyondBMI dataset, plotted on top of the possible ranges defined by our minimum and maximum muscle surfaces. Note that our minimal fat mass is coupled to the maximal muscle mass and vice versa.

with the density ratio  $\rho = \rho_M / \rho_F$ .

The minimum/maximum muscle surface yields a maximum/minimum fat layer mass. The optimized fat layer mass is clamped to meet this range, thereby defining the final fat layer mass. We then choose the linear interpolant between the minimum and maximum muscle surface that matches this fat mass, which we find through bisection search.

We did this for the scans of 100 men and 100 women from the BeyondBMI dataset (Maalin et al., 2020), where we know the true values for FM and MM from measurements, and optimized the value of  $k$  for this dataset, yielding  $k_{\text{male}} = 1.5$  and  $k_{\text{female}} = 1.0$ . This is plausible since women in general have a lower muscle mass. For instance, the average muscle mass of the male subjects in the dataset is indeed 50% higher than the average MM for the female subjects. The mean absolute errors (MAE) for the BeyondBMI dataset are  $\text{MAE}_{\text{MM}} = 0.37 \text{ kg} (\pm 0.31)$ ,  $\text{MAE}_{\text{FM}} = 0.46 \text{ kg} (\pm 0.38)$  for the female subjects and  $\text{MAE}_{\text{MM}} = 0.46 \text{ kg} (\pm 0.39)$ ,  $\text{MAE}_{\text{FM}} = 0.57 \text{ kg} (\pm 0.48)$  for the male subjects. Figure 8 shows how well our model can adjust to the target values of muscle and fat mass. All values are inside or at least close to the predicted possible range of minima and maxima. Moreover, in most cases the muscle/fat mass values for the same person split the two ranges at about an inverse point (e.g., close to maximum muscle and close to minimum fat), which leads to the low errors stated above.

### 3.4.4 Transferring Original Anatomical Data

After fitting the skin surface  $\mathcal{S}$  to the scan and transferring the skeleton surface  $\mathcal{B}$  and the muscle surface  $\mathcal{M}$  into the scan, the final step is to transform the high-resolution anatomical details (Zygote’s bone and muscle models in our case) from the volumetric template to the scanned subject. We implement this in an efficient and robust manner as a mesh-independent space warp  $\mathbf{d}: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  that maps the original template’s skin surface  $\hat{\mathcal{S}}$ , muscle surface  $\hat{\mathcal{M}}$ , and skeleton surface  $\hat{\mathcal{B}}$  (all marked with a hat) to the scanned subject’s layer surfaces  $\mathcal{S}$ ,  $\mathcal{M}$ , and  $\mathcal{B}$ , respectively. All geometry that is embedded in between these surfaces will smoothly be warped from template to scan.

Dicko et al. (2013) also employ a space warp for their anatomy transfer, which they discretized by interpolating values  $\mathbf{d}_{ijk}$  on a regular 3D grid constructed around the object. Their space warp is computed

by interpolating the skin deformation  $\hat{\mathcal{S}} \mapsto \mathcal{S}$  on the boundary and being harmonic in the interior (i.e.,  $\Delta \mathbf{d} = 0$ ), which requires the solution of a large sparse Poisson system for the coefficients  $\mathbf{d}_{ijk}$ .

We follow the same idea, but use a space warp based on triharmonic radial basis functions (RBFs) (Botsch and Kobbelt, 2005), which have been shown to yield higher quality deformations with lower geometric distortion than many other warps (including FEM-based harmonic warps) (Sieger et al., 2013). The RBF warp is defined as a sum of  $n$  RBF kernels and a linear polynomial:

$$\mathbf{d}(\mathbf{x}) = \sum_{j=1}^n \mathbf{w}_j \varphi_j(\mathbf{x}) + \mathbf{a}^\top \mathbf{x} + b, \quad (13)$$

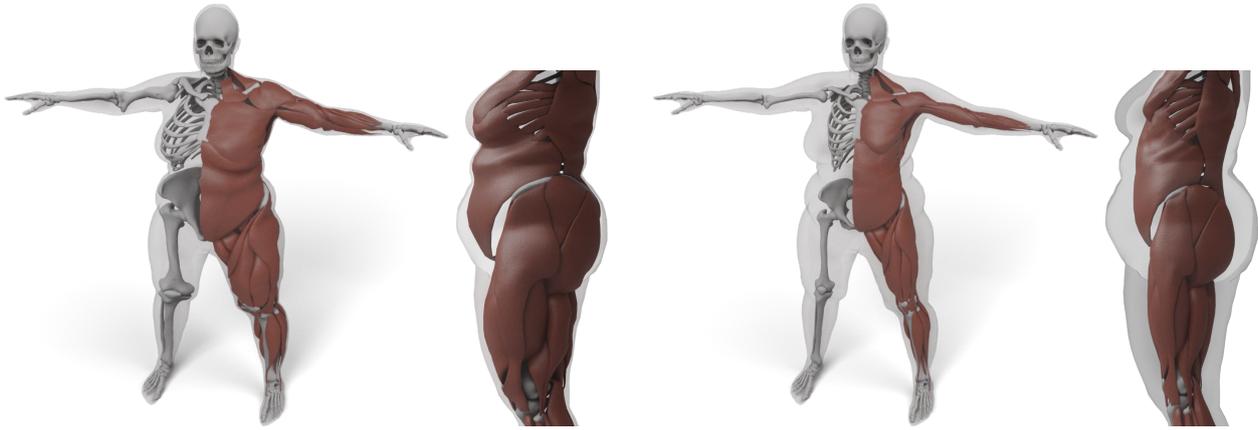
where  $\mathbf{w}_j \in \mathbb{R}^3$  is the coefficient of the  $j^{\text{th}}$  radial basis function  $\varphi_j(\mathbf{x}) = \varphi(\|\mathbf{x} - \mathbf{c}_j\|)$ , which is centered at  $\mathbf{c}_j \in \mathbb{R}^3$ . As kernel function we use  $\varphi(r) = r^3$ , leading to highly smooth triharmonic warps ( $\Delta^3 \mathbf{d} = 0$ ). The term  $\mathbf{a}^\top \mathbf{x} + b$  is a linear polynomial ensuring linear precision of the warp.

In order to warp the high-resolution *bone model* from the template to the scan, we setup the RBF warp to reproduce the deformation  $\hat{\mathcal{B}} \mapsto \mathcal{B}$ . To this end, we select 5000 vertices  $\hat{\mathbf{x}}_i \in \hat{\mathcal{B}}$  from the template's skeleton surface by farthest point sampling. The corresponding vertices on the scan's skeleton surface are denoted by  $\mathbf{x}_i \in \mathcal{B}$ . At these vertices  $\hat{\mathbf{x}}_i$  the deformation function  $\mathbf{d}(\hat{\mathbf{x}}_i)$  should interpolate the displacements  $\mathbf{d}_i = \mathbf{x}_i - \hat{\mathbf{x}}_i$ . These constraints lead to a dense, symmetric, but indefinite  $(n+4) \times (n+4)$  linear system, which we solve for the coefficients  $\mathbf{w}_1, \dots, \mathbf{w}_n, \mathbf{a}, b$  using the LU factorization of Eigen (Guennebaud et al., 2018); see (Sieger et al., 2013) for details. The resulting RBF warp  $\mathbf{d}$  then transforms each vertex  $\mathbf{x}$  of the high-resolution bone model as  $\mathbf{x} \leftarrow \mathbf{x} + \mathbf{d}(\mathbf{x})$ . Note that this process can trivially be parallelized over all model vertices, which we implement using OpenMP. For warping the high-resolution *muscle model* we follow the same procedure, but collect 7000 constraints from the vertices  $\hat{\mathbf{x}}_i \in \mathcal{S} \cup \mathcal{M}$  of the skeleton and muscle surfaces, since these enclose the muscle layer.

## 4 RESULTS AND APPLICATIONS

Generating a personalized anatomical model for a given surface scan of a person consists of the following steps: First, the surface template is registered to the scanner data (triangle mesh or point cloud) as described in Section 3.1 and Achenbach et al. (2017). After manually selecting 10–20 landmarks, this process takes about 50 sec. Fitting the surface template establishes dense correspondence with the surface of the volumetric template and puts the scan into the same T-pose as the volumetric template. Fitting the volumetric template by transferring the three layer surfaces (Sections 3.4.1, 3.4.2, 3.4.3) takes about 15 sec. Transferring the high-resolution anatomical models of bones and muscles (145k vertices) takes about 4.5 sec for solving the linear system (which is an offline pre-processing) and 0.5 sec for transforming the vertices (Section 3.4.4). Timings were measured on a desktop workstation, equipped with an Intel Core i9 10850K CPU and a Nvidia RTX 3070 GPU.

Dicko et al. (2013) and Kadleček et al. (2016) are the two approaches most closely related to ours. Dicko et al. (2013) also use a space warp for transferring anatomical details, but since they only use the skin surface as constraint, the interior geometry can be strongly distorted. To prevent this, they restrict bones to affine transformations, which, however, might still contain unnatural shearing modes and implausible scaling. Our space warp yields a higher smoothness due to the use of  $C^\infty$  RBF kernels instead of  $C^0$  trilinear interpolation and reduces unnatural distortion of bones and muscles by using three layer surfaces as constraints instead of the skin surface only and by optimizing these layers w.r.t. anatomical distortion. In Figure 9 we compare the result of warping the anatomical structures using a harmonic basis and 7000



**Figure 9.** Result of transferring the anatomy by using just the skin layer and a harmonic basis (left). Here, both muscles and bones deform too much to fit overweight targets. We use the additional muscle and skeleton layer and a triharmonic basis (right) to prevent unnatural deformations.

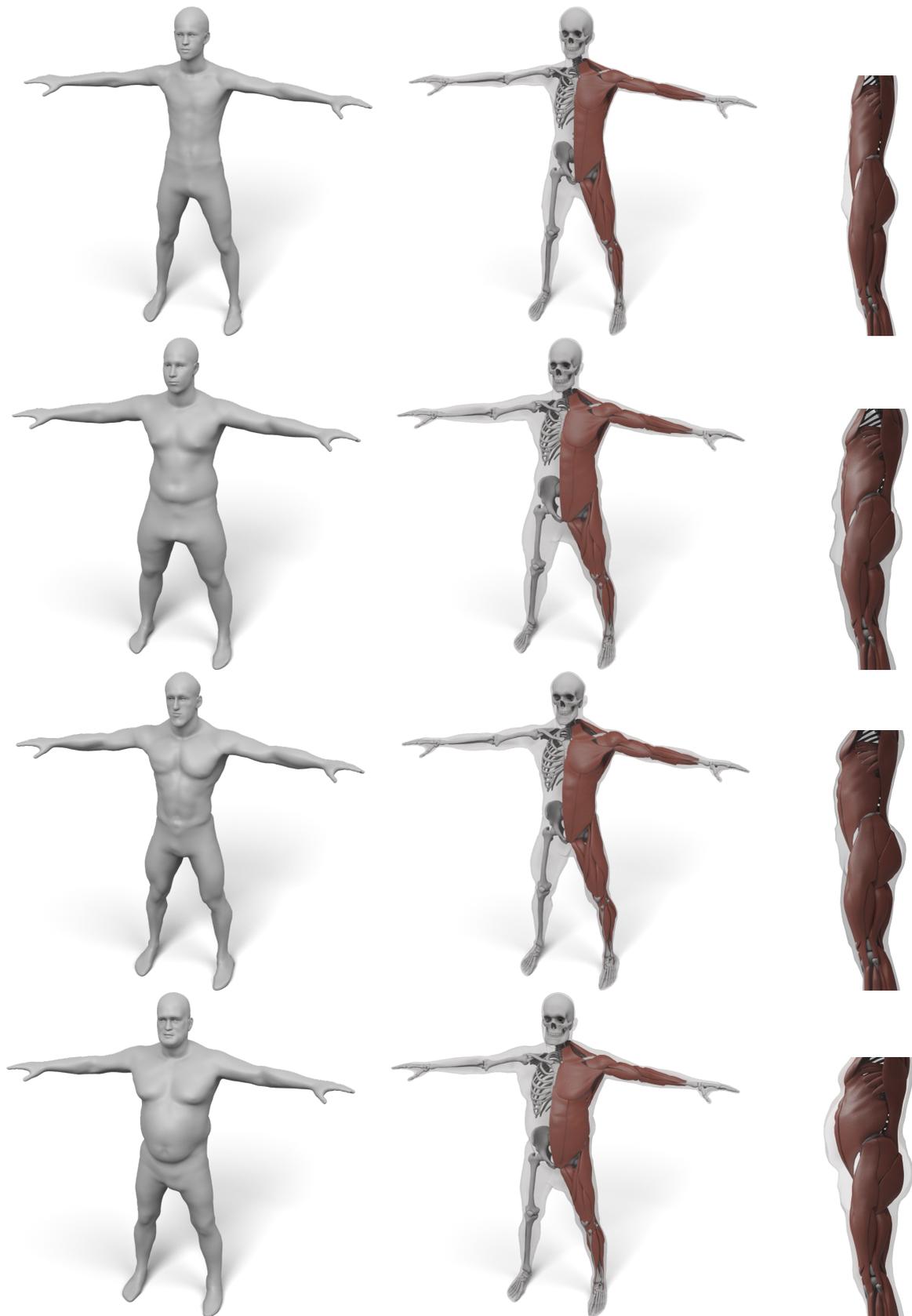
centers from only the skin surface to our three-layered, triharmonic warp result. The former does show drastic and unrealistic deformations of both muscles and bones while our approach solves those issues. Note that additionally restricting the bones to affine transformations like Dicko et al. (2013) would still produce unnaturally thick bones (e.g. the upper leg bone) and muscles.

Compared to Kadleček et al. (2016), we require a single input scan only, since we infer (initial guesses for) joint positions and limb lengths from the full-body PCA of Achenbach et al. (2017). Putting the scan into T-pose prevents us from having to solve bone geometry and joint angles simultaneously, which makes our approach much faster than theirs (15 sec vs. 30 min). Moreover, our layered model yields a conforming volumetric tessellation with constant and homogeneous per-layer materials, which more effectively prevents bones from penetrating skin or muscles. In their approach the rib cage often intersects the muscle layer for thin subjects as mentioned by Kadleček et al. (2016) in the limitations and shown in Figure 12 (bottom row) of their work. Furthermore, we automatically derive the muscle/fat body composition from the surface scan, which yields more plausible results than growing muscles as much as possible (Kadleček et al., 2016), since the latter leads to more corpulent people always having more muscles. Our model extracts the amount of muscle and fat using data of real humans and can therefore adopt to the variety of human shapes (low FM and high MM, high FM and low MM, and everything in between). Finally, we support both male and female subjects by employing individual anatomical templates and muscle/fat regressors for men and women.

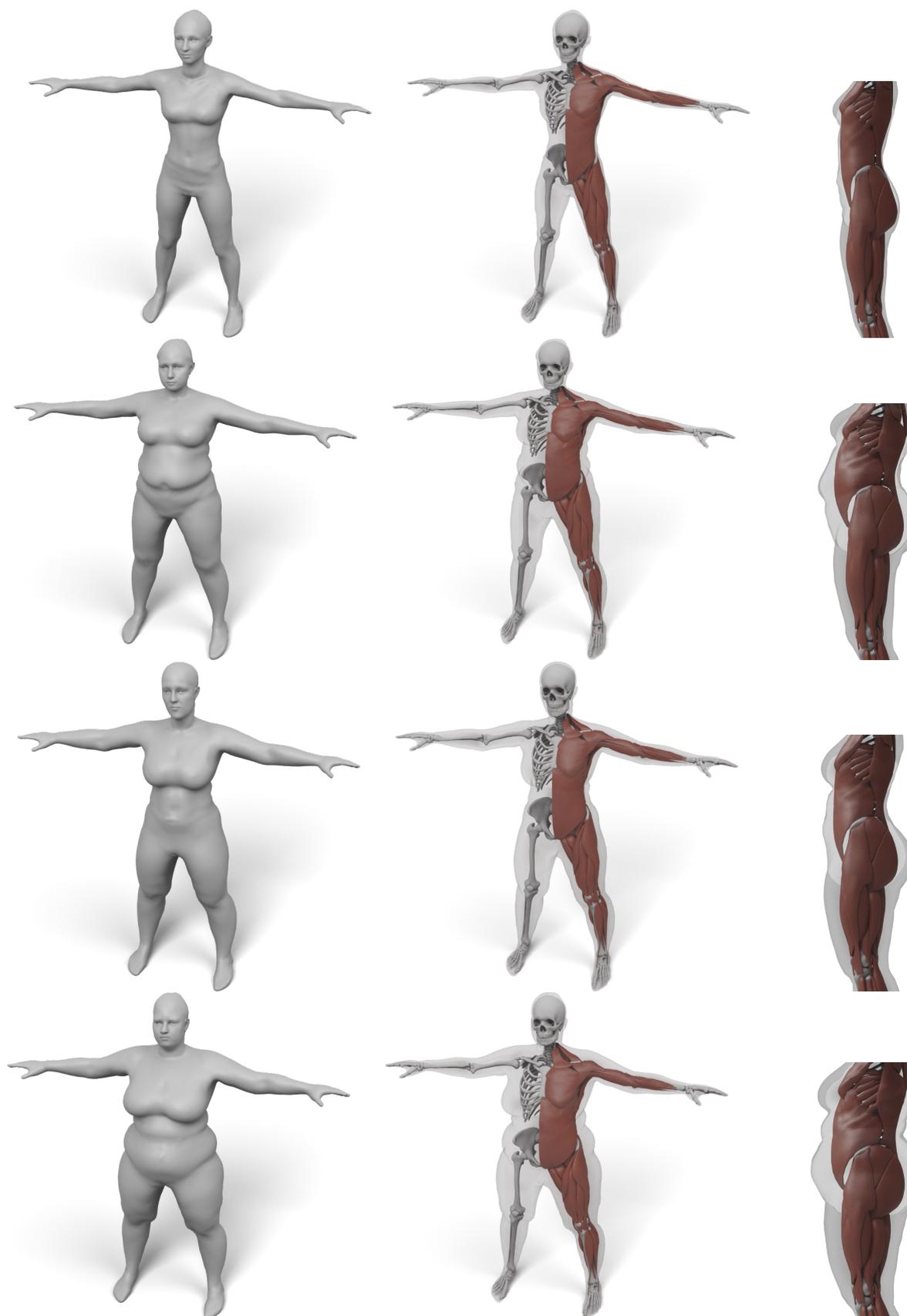
#### 4.1 Evaluation on Hasler Dataset

In order to further evaluate the generalization abilities of the linear FM/MM models (Section 3.3) to other data sources, we estimate FM and MM for a subset of registered scans from the *Hasler* dataset (Hasler et al., 2009) and measure the prediction error. We selected scans of 10 men and 10 women, making sure to cover the extremes of the weight, height, fat, and muscle percentage distribution present in the data.

For the female sample, the predictions show a mean absolute error of  $MAE_{FM} = 0.65 \text{ kg} (\pm 0.44)$  and  $MAE_{MM} = 4.39 \text{ kg} (\pm 1.71)$ . For the male sample, the model shows a similar error for the MM prediction, but performs worse at predicting FM:  $MAE_{FM} = 3.32 \text{ kg} (\pm 1.98)$  and  $MAE_{MM} = 4.14 \text{ kg} (\pm 2.74)$ . Compared to the leave-one-out tests on the *BeyondBMI* data, the average error increases noticeably, which can partly be explained by differences in the measurement procedure between the two datasets: While Hasler et al. (2009) used a consumer-grade body fat scale, Maalin et al. (2020) used a medical-grade scale,



**Figure 10.** Some examples for various male body shape types. For each input surface the transferred muscles and skeleton are shown in front and side view.



**Figure 11.** Some examples for various female body shape types. For each input surface the transferred muscles and skeleton are shown in front and side view.

which should lead to more accurate measurements. Nevertheless, these results show that our regressor generalizes well to other data sources, providing a simple and sufficiently accurate method for estimating FM and MM from body scans.

Given the FM and MM values of a target from our regressor, we choose the optimal muscle surface between the minimal and maximal muscle surface as explained in Section 3.4.3. Comparing the final FM and MM of the volumetric model to the ground truth measurements of the *Hasler* dataset we get end-to-end errors of  $MAE_{FM} = 0.70 \text{ kg } (\pm 0.52)$ ,  $MAE_{MM} = 4.19 \text{ kg } (\pm 1.39)$  (female) and  $MAE_{FM} = 3.49 \text{ kg } (\pm 2.02)$ ,  $MAE_{MM} = 3.81 \text{ kg } (\pm 2.56)$  (male). This evaluation shows that the additional error induced by fitting the muscle layer is very low.

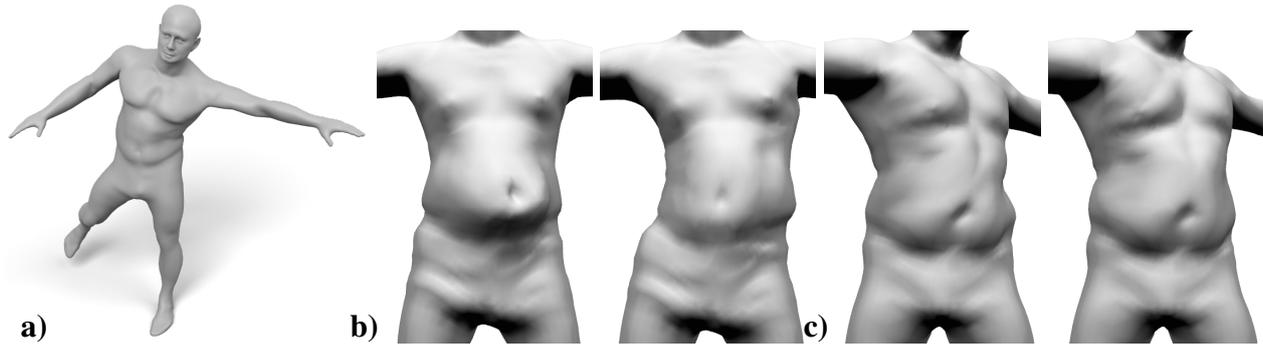
## 4.2 Evaluation on CAESAR Dataset

In order to demonstrate the flexibility and robustness of our method, we evaluate it by generating anatomical models for all scans of the European Caesar data set (Robinette et al., 2002), consisting of 919 scans of women and 777 scan of men, with height range 131–218 cm for men and from 144–195 cm for women (we restricted to scans with complete annotation and taken in standing pose). A few examples for men and women can be seen in Figure 10 and Figure 11.

For the about 1700 CAESAR scans, our muscle and fat mass regressors yield just one slightly negative value for the fat mass of the thinnest male (body weight 48 kg, height 1.72 m, BMI 16.14 kg/m<sup>2</sup>). For all other subjects, we get values ranging from 3.5–38.9% body fat (mean 20.3%) for male subjects and 8–45.3% (mean 28.9%) for female subjects. The range of predicted muscle masses is 24.9–57.8 kg (men) and 20.1–37.7 kg (women). When determining the optimal interpolation between the minimum and maximum muscle layer (Section 3.4.3) we meet the estimated target values up to mean errors  $MAE_{FM} = 1.08 \text{ kg } (\pm 0.9)$  (male),  $MAE_{FM} = 1.41 \text{ kg } (\pm 1.35)$  (female) and  $MAE_{MM} = 0.88 \text{ kg } (\pm 0.74)$  (male),  $MAE_{MM} = 1.15 \text{ kg } (\pm 1.11)$  (female). Note that even the scan with predicted negative FM can be reconstructed robustly. In this case the muscle surface will be the maximum muscle surface, which in general is a suitable estimate for very skinny subjects.

The CAESAR dataset does not include ground truth data for fat and muscle mass of the scanned individuals. Thus, in order to further evaluate the plausibility of our estimated body composition, we compare it to known body fat percentiles. Percentiles are used as guidelines in medicine and provide statistical reference values one can compare individual measurements to. For instance, a 10<sup>th</sup> percentile of 20.8% body fat means that 10% of the examined population have a body fat percentage less than 20.8%. Assuming that the European CAESAR dataset is a representative sample of the population, the percentiles we get from our reconstructions of the CAESAR scans should match the percentiles of the European population. We compared the values produced by our fat and muscle mass regressors (Section 3.3) to Kyle et al. (2001), who measured body fat using 4-electrode bio-electrical impedance analysis from 2735 male and 2490 female western European adults. Our body fat percentiles on the CAESAR dataset are very well in agreement with their results, as shown in the following table:

Percentile		5 <sup>th</sup>	10 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	90 <sup>th</sup>	95 <sup>th</sup>
Male	Our estimate	10.2	12.3	16.0	20.3	24.6	28.1	30.7
	Kyle et al. (2001)	10.9	12.6	15.7	19.2	23.5	27.0	29.2
Female	Our estimate	18.6	21.1	24.7	28.5	33.7	37.4	39.3
	Kyle et al. (2001)	18.5	20.8	23.8	28.1	32.6	37.5	40.5



**Figure 12.** Our layered anatomical model can be animated using an extension of Fast Projective Skinning (FPS), as shown in (a). When the character performs a jump to the left (b), our realistic skeleton correctly restricts the dynamic jiggling to the belly region (b-left), while the original FPS deforms the complete torso (b-right). For a static twist of the torso (c), the rib-cage of our layered model keeps the chest region rather rigid and concentrates the deformation to the belly (c-left). Without a proper anatomical model, the deformation of FPS is distributed over the complete torso (c-right).

### 4.3 Physics-Based Character Animation

One application of our model is simulation-based character animation (Deul and Bender, 2013; Komaritzan and Botsch, 2018, 2019), where the transferred volumetric layers can improve the anatomical plausibility. We demonstrate the potential by extending the Fast Projective Skinning (FPS) of Komaritzan and Botsch (2019). FPS already uses a simplified volumetric skeleton built from spheres and cylinders, a skeleton surface wrapping this simple skeleton, and one layer of volumetric prism elements spanned between skin and skeleton surface. Whenever the skeleton is posed, the vertices of the skeleton surface are moved, and a projective dynamics simulation of the soft tissue layer updates the skin surface.

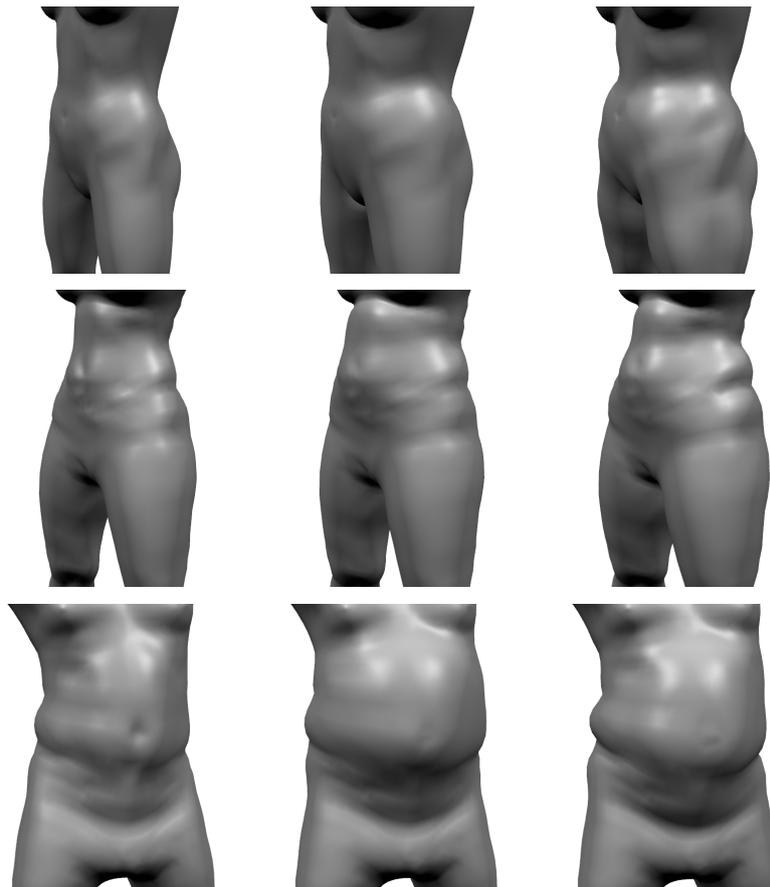
We replace their synthetic skeleton by our more realistic version and split their soft tissue layer into our separate muscle and fat layers. This enables us to use different stiffness values for the fat and muscle layers (the latter being three times larger). Moreover, our skeleton features a realistic rib-cage, whereas FPS only uses a simplified spine in the torso region. As a result, our extended version of FPS yields more realistic results in particular in the torso and belly region, as shown in Figure 12.

### 4.4 Simulation of Fat Growth

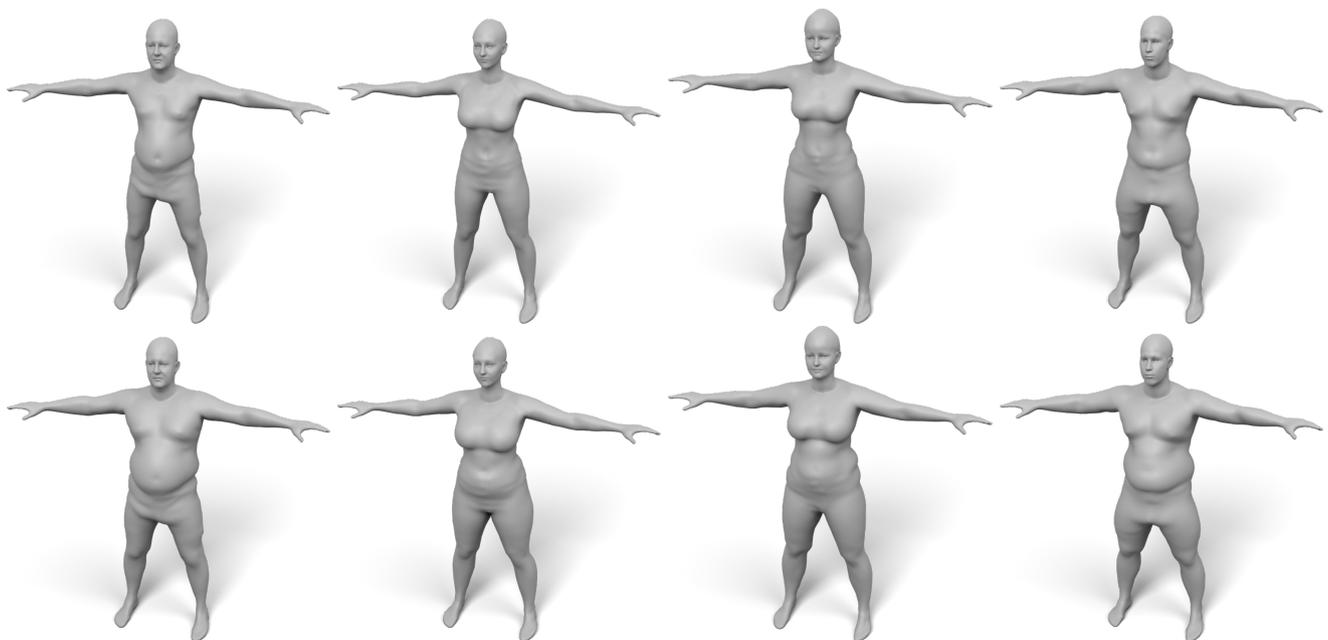
Our anatomical model can also be used to simulate an increase of body fat, where its volumetric nature provides advantages over existing surface-based methods.

In their computational bodybuilding approach, Saito et al. (2015) also propose a method for growing fat. They, however, employ a purely *surface-based* approach that conceptually mimics blowing up a rubber balloon. This is modelled by a pressure potential that drives skin vertices outwards in normal direction, regularized by a co-rotated triangle strain energy. The user can (and should) specify a scalar field that defines where and how strong the skin surface should be “blown up”, which is used to modulate the per-vertex pressure forces. Despite the regularization we sometimes noticed artifacts at the boundary of the fat growing region and therefore add another regularization through Equation (2). This approach allows the user to tune the amount of subcutaneous fat, but unless a carefully designed growth field is specified, the fat growth looks rather uniform and balloon-like (see Figures 10, 11 in Saito et al. (2015)).

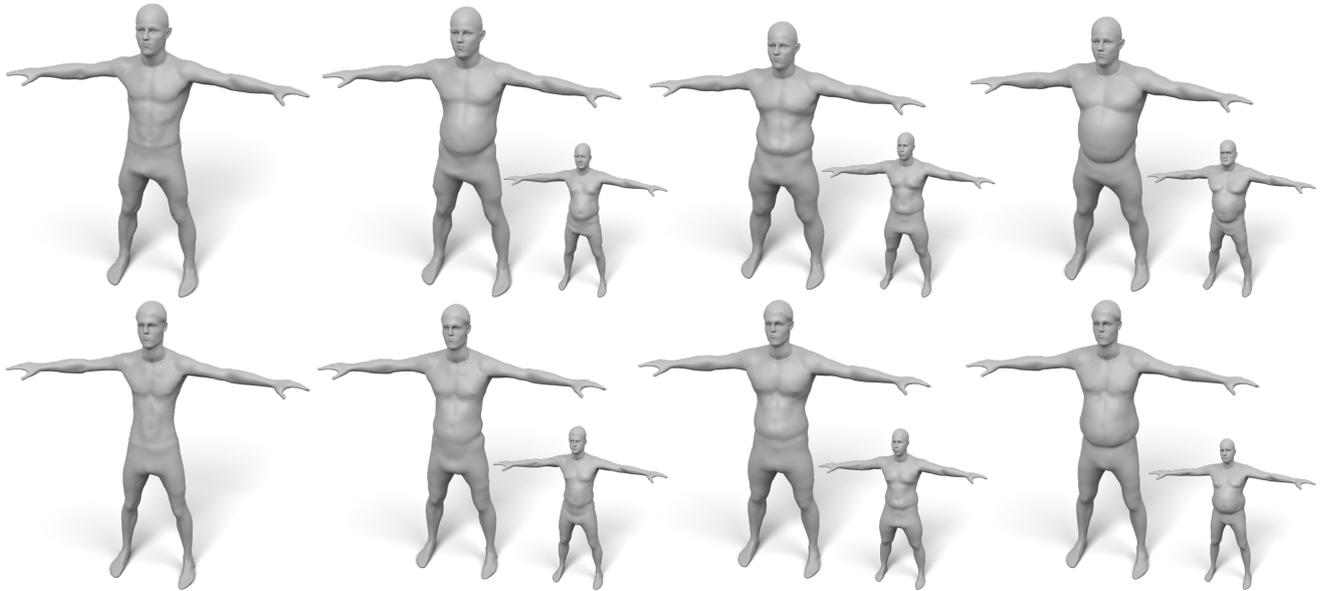
Every person has an individual fat distribution and gaining weight typically intensifies these initial fat depots. We model this behavior by scaling up the local prism volumes of our fat layer. Each fat prism can be split into three tetrahedra, which define volumetric elements  $t_j \in \mathcal{T}$  with initial volumes  $\bar{V}_j$ . A simple



**Figure 13.** Given a reconstructed model (left), the pressure-based fat growth of Saito et al. (2015) leads to a more uniform increases in fat volume (center), while our volume-based fat growth increases the initial fat distribution.



**Figure 14.** Examples of our fat growth simulation, with input models shown in the top row and their weight-gained version in the bottom row.



**Figure 15.** Examples of “fat transfer”. The two subjects (top/bottom) on the left have a very low amount of body fat. Therefore, scaling their fat volumes is not suitable for fat growth. Instead, we copy the fat distributions of other subjects (shown as small insets) and transfer them to the skinny targets.

uniform scaling  $s \cdot \bar{V}_j$  achieves the desired effect that fat increases more in fat-intense regions. The growth simulation is implemented by minimizing the energy

$$E_{\text{grow}}(\mathcal{S}) = w_{\text{vol}}E_{\text{vol}}(\mathcal{S}) + w_{\text{reg}}E_{\text{reg}}(\mathcal{S}, \bar{\mathcal{S}}) + w_{\text{rest}}E_{\text{rest}}(\mathcal{S}, \bar{\mathcal{S}}) \quad (14)$$

with the Laplacian regularization of Equation (2), the displacement regularization

$$E_{\text{rest}}(\mathcal{S}, \bar{\mathcal{S}}) = \sum_{\mathbf{x}_i \in \mathcal{S}} A_i \|\mathbf{x}_i - \bar{\mathbf{x}}_i\|^2 \quad (15)$$

and the volume fitting term

$$E_{\text{vol}}(\mathcal{S}) = \sum_{t_j \in \mathcal{T}} \bar{V}_j (\text{vol}(t_j) - s \cdot \bar{V}_j)^2, \quad (16)$$

where  $\bar{\mathcal{S}}$  and  $\mathcal{S}$  denote the skin surface before/after the fat growth and  $s$  is the global fat scaling factor. Saito et al. (2015) argued that *anisotropically* scaling fat tetrahedra in *one* direction does not produce plausible results. However, *isotropically* scaling the volume leaves the minimization more freedom and yields convincing results. Figure 13 compares the pressure-based and volumetric fat growth simulations. Figure 14 shows some more examples produced by combining both methods.

Our volume-based fat growth has another advantage: If we want to grow fat on a very skinny person, the initial (negligible) fat distribution does not provide enough information on where to grow fat, such that both approaches would do a poor job. But since we can easily fit the volumetric template to several subjects, we can “copy” the distribution of fat prism volumes from another person and “paste” it onto the skinny target, which simply replaces the target volumes in (16). This enables to fat transfer between different subjects, which is shown in Figure 15.

## 5 CONCLUSION

We created a simple layered volumetric template of the human anatomy and presented an approach for fitting it to surface scans of men and women of various body shapes and sizes. Our method generates plausible muscle and fat layers by estimating realistic muscle and fat masses from the surface scan alone. In addition to the layered template, we also showed how to transfer internal anatomical structures, such as bones and muscles, using a high-quality space warp. Compared to previous work, our method is fully automatic and considerably faster, enabling the simple generation of personalized anatomical models from surface body scans. Besides educational visualization, we demonstrated the potential of our model for physics-based character animation and anatomically plausible fat growth simulation.

Our approach has some limitations: First, we do not generate individual layers for head, hands and toes, where in particular the head would require special treatment. Combining our layered body model with the multi-linear head model of Achenbach et al. (2018) is therefore a promising direction for future work. Second, our regressors for fat and muscle mass could be further optimized by training on more body scans with known body composition. Given more and more accurate training data, as for instance provided by DXA scans, we could extend the fat/muscle estimations to individual body parts. Third, we do not model tendons and veins. Those would have to be included in all layers and could be transferred in the same way as high-resolution muscle and bone models. Fourth, the fact that the three layers of our model share the same topology/connectivity can also be considered a limitation, since we cannot use different, adaptive mesh resolutions in different layers. A promising direction for future work is the use of our anatomical model for generating synthetic training data for statistical analysis and machine learning applications, where the simple structure of our layered model can be beneficial.

## ACKNOWLEDGMENTS

The authors are grateful to Jascha Achenbach for valuable discussion and implementation hints and to Hendrik Meyer for his help with the renderings of our models. This research was supported by the German Federal Ministry of Education and Research (BMBF) through the project ViTraS (ID 16SV8225). The scale and ruler emojis in Figure 2 are designed by OpenMoji (<https://openmoji.org>) and provided through CC BY-SA 4.0 License.

## REFERENCES

- Achenbach, J., Brylka, R., Gietzen, T., Zum Hebel, K., Schömer, E., Schulze, R., et al. (2018). A multilinear model for bidirectional craniofacial reconstruction. In *Proc. of Eurographics Workshop on Visual Computing for Biology and Medicine*. 67–76
- Achenbach, J., Waltemate, T., Latoschik, M. E., and Botsch, M. (2017). Fast generation of realistic virtual humans. In *Proc. of ACM Symposium on Virtual Reality Software and Technology*. 12:1–12:10
- Ackerman, M. J. (1998). The visible human project. *Proceedings of the IEEE* 86, 504–511
- Angelov, D., Srinivasan, P., Koller, D., Thrun, S., Rodgers, J., and Davis, J. (2005). SCAPE: Shape completion and animation of people. *ACM Transactions on Graphics* 24, 408–416
- [Dataset] Autodesk (2014). Character generator. <https://charactergenerator.autodesk.com/>
- Aydin Kabakci, A., Buyukmumcu, M., Yilmaz, M., Cicekcibasi, A., Akin, D., and Cihan, E. (2017). An osteometric study on humerus. *International Journal of Morphology* 35, 219–226
- Bogo, F., Romero, J., Pons-Moll, G., and Black, M. J. (2017). Dynamic FAUST: Registering human bodies in motion. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*. 5573–5582

- Botsch, M. and Kobbelt, L. (2005). Real-time shape editing using radial basis functions. *Computer Graphics Forum* 24, 611–621
- Botsch, M., Kobbelt, L., Pauly, M., Alliez, P., and Lévy, B. (2010). *Polygon Mesh Processing* (AK Peters, CRC press)
- Bouaziz, S., Deuss, M., Schwartzburg, Y., Weise, T., and Pauly, M. (2012). Shape-up: Shaping discrete geometry with projections. *Comput. Graph. Forum* 31, 1657–1667
- Bouaziz, S., Martin, S., Liu, T., Kavan, L., and Pauly, M. (2014a). Projective dynamics: Fusing constraint projections for fast simulation. *ACM Transactions on Graphics* 33, 115:1–115:11
- Bouaziz, S., Tagliasacchi, A., and Pauly, M. (2014b). Dynamic 2D/3D registration. In *Eurographics Tutorials*. 1–17
- Brochu, T., Edwards, E., and Bridson, R. (2012). Efficient geometrically exact continuous collision detection. *ACM Transactions on Graphics* 31, 96:1–96:7
- Brožek, J., Grande, F., Anderson, J. T., and Keys, A. (1963). Densitometric analysis of body composition: Revision of some quantitative assumptions. *Annals of the New York Academy of Sciences* 110, 113–140
- Christ, A., Kainz, W., Hahn, E. G., Honegger, K., Zefferer, M., Neufeld, E., et al. (2009). The virtual family—development of surface-based anatomical models of two adults and two children for dosimetric simulations. *Physics in Medicine & Biology* 55, N23–N38
- Dayal, M. R., Steyn, M., and Kuykendall, K. L. (2008). Stature estimation from bones of south african whites. *South African Journal of Science* 104, 124–128
- Deul, C. and Bender, J. (2013). Physically-based character skinning. In *Proc. of Virtual Reality Interactions and Physical Simulations*
- Deuss, M., Deleuran, A. H., Bouaziz, S., Deng, B., Piker, D., and Pauly, M. (2015). ShapeOp – a robust and extensible geometric modelling paradigm. In *Proc. of Design Modelling Symposium*. 505–515
- Dicko, A.-H., Liu, T., Gilles, B., Kavan, L., Faure, F., Palombi, O., et al. (2013). Anatomy transfer. *ACM Transactions on Graphics* 32, 188:1–188:8
- Fields, D. A., Goran, M. I., and McCrory, M. A. (2002). Body-composition assessment via air-displacement plethysmography in adults and children: a review. *The American Journal of Clinical Nutrition* 75, 453–467
- [Dataset] Fit3D (2021). Fit3d scanner systems. <https://fit3d.com/>
- Gietzen, T., Brylka, R., Achenbach, J., Zum Hebel, K., Schömer, E., Botsch, M., et al. (2019). A method for automatic forensic facial reconstruction based on dense statistics of soft tissue thickness. *PloS one* 14, e0210257
- [Dataset] Guennebaud, G., Jacob, B., et al. (2018). Eigen v3. <http://eigen.tuxfamily.org>
- Hasler, N., Stoll, C., Sunkel, M., Rosenhahn, B., and Seidel, H.-P. (2009). A statistical model of human pose and body shape. *Computer Graphics Forum* 28, 337–346
- Heysmsfield, S. B., Heo, M., Thomas, D., and Pietrobelli, A. (2011). Scaling of body composition to height: relevance to height-normalized indexes. *The American journal of clinical nutrition* 93, 736–740
- Ichim, A.-E., Kadleček, P., Kavan, L., and Pauly, M. (2017). Phace: Physics-based face modeling and animation. *ACM Transactions on Graphics* 36, 153:1–153:14
- Ichim, A.-E., Kavan, L., Nimier-David, M., and Pauly, M. (2016). Building and animating user-specific volumetric face rigs. In *Symposium on Computer Animation*. 107–117
- Jackson, A. S. and Pollock, M. L. (1985). Practical assessment of body composition. *The Physician and Sportsmedicine* 13, 76–90
- Kadleček, P., Ichim, A.-E., Liu, T., Křivánek, J., and Kavan, L. (2016). Reconstructing personalized anatomical models for physics-based body animation. *ACM Transactions on Graphics* 35, 213:1–213:13

- Kim, M., Pons-Moll, G., Pujades, S., Bang, S., Kim, J., Black, M. J., et al. (2017). Data-driven physics for human soft tissue animation. *ACM Transactions on Graphics* 36, 54:1–54:12
- Komaritzan, M. and Botsch, M. (2018). Projective skinning. *Proceedings of the ACM on Computer Graphics and Interactive Techniques* 1
- Komaritzan, M. and Botsch, M. (2019). Fast projective skinning. In *Proc. of ACM Motion, Interaction and Games*. 22:1–22:10
- Kyle, U. G., Genton, L., Slosman, D. O., and Pichard, C. (2001). Fat-free and fat mass percentiles in 5225 healthy subjects aged 15 to 98 years. *Nutrition* 17, 534–541
- Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., and Black, M. J. (2015). SMPL: A skinned multi-person linear model. *ACM Transactions on Graphics* 34, 248:1–248:16
- Maalin, N., Mohamed, S., Kramer, R. S., Cornelissen, P. L., Martin, D., and Tovée, M. J. (2020). Beyond BMI for self-estimates of body size and shape: A new method for developing stimuli correctly calibrated for body composition. *Behavior Research Methods*
- Ng, B. K., Hinton, B. J., Fan, B., Kanaya, A. M., and Shepherd, J. A. (2016). Clinical anthropometrics and body composition from 3D whole-body surface scans. *European Journal of Clinical Nutrition* 70, 1265–1270
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830
- Piryankova, I., Stefanucci, J., Romero, J., de la Rosa, S., Black, M., and Mohler, B. (2014). Can I recognize my body's weight? The influence of shape and texture on the perception of self. *ACM Transactions on Applied Perception* 11, 13:1–13:18
- Riviere, J., Gotardo, P., Bradley, D., Ghosh, A., and Beeler, T. (2020). Single-shot high-quality facial geometry and skin appearance capture. *ACM Transactions on Graphics* 39, 81:1–81:12
- Robinette, K. M., Blackwell, S., Daanen, H., Boehmer, M., and Fleming, S. (2002). *Civilian american and european surface anthropometry resource (CEASAR), final report. volume 1: Summary*. Tech. rep., Sytronics Inc
- Romero, C., Otaduy, M. A., Casas, D., and Perez, J. (2020). Modeling and estimation of nonlinear skin mechanics for animated avatars. *Computer Graphics Forum* 39, 77–88
- Saito, S., Zhou, Z.-Y., and Kavan, L. (2015). Computational bodybuilding: Anatomically-based modeling of human bodies. *ACM Transactions on Graphics* 34, 41:1–41:12
- Shoemake, K. and Duff, T. (1992). Matrix animation and polar decomposition. In *Proceedings of the Conference on Graphics Interface*. 258–264
- Sieger, D., Menzel, S., and Botsch, M. (2013). High quality mesh morphing using triharmonic radial basis functions. In *Proceedings of the 21st International Meshing Roundtable*. 1–15
- Siri, W. E. (1956). *Body composition from fluid spaces and density: analysis of methods*. Tech. Rep. UCRL-3349, Lawrence Berkeley National Laboratory
- Tomlinson, D., Erskine, R., Morse, C., Winwood, K., and Onambélé-Pearson, G. (2016). The impact of obesity on skeletal muscle strength and structure through adolescence to old age. *Biogerontology* 17, 467–483
- Weng, C.-Y., Curless, B., and Kemelmacher-Shlizerman, I. (2019). Photo Wake-Up: 3D character animation from a single photo. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*. 5908–5917
- Wenninger, S., Achenbach, J., Bartl, A., Latoschik, M. E., and Botsch, M. (2020). Realistic virtual humans from smartphone videos. In *Proc. of ACM Symposium on Virtual Reality Software and Technology*. 29:1–29:11

Zhu, L., Hu, X., and Kavan, L. (2015). Adaptable anatomical models for realistic bone motion reconstruction. *Computer Graphics Forum* 34, 459–471

Ziylan, T. and Murshid, K. A. (2002). An analysis of anatolian human femur anthropometry. *Turkish Journal of Medical Sciences* 32, 231–235

[Dataset] Zygote (2020). <https://www.zygote.com>

## 1 APPENDIX: IMPLEMENTATION DETAILS

For minimization of the energies (1), (5), and (8), we use the projective framework of (Bouaziz et al., 2012, 2014a), implemented through an adapted local/global solver from the ShapeOp library (Deuss et al., 2015). It has the advantages of being unconditionally stable, easy-to-use and flexible enough to handle a wide range of energies. Here we give the weights for the different energy terms and give implementation details.

We fit the skin surface  $\mathcal{S}$  to the skeleton wrap  $\mathcal{W}$  by minimizing (1), restated here:

$$\mathcal{B} = \arg \min_{\mathcal{X}} w_{\text{fit}} E_{\text{fit}}(\mathcal{X}, \mathcal{W}) + w_{\text{reg}} E_{\text{reg}}(\mathcal{X}, \bar{\mathcal{X}}).$$

We first initialize  $\mathcal{X}$  with  $\mathcal{S}$  and set  $w_{\text{reg}} = w_{\text{fit}} = 1$ . When the minimization converges, we update the initial Laplacians  $\Delta \bar{\mathbf{x}}_i$  in (2) to the Laplacians  $\Delta \mathbf{x}_i$  of the current solution  $\mathcal{X}$  and decrease  $w_{\text{reg}}$  by a factor of 0.1. This is repeated until  $w_{\text{reg}}$  reaches  $10^{-7}$ . In order to speed up the fitting process, we first remove high frequency details of the skin surface (e.g., nipples and navels) by Laplacian smoothing (Botsch et al., 2010) before computing the initial Laplacians  $\Delta \bar{\mathbf{x}}_i$ . Since we exclude head, hands, and toes from the layered template, those regions are fixed throughout the whole process.

For fitting the skeleton surface of the template to a surface scan, we minimize Equation (5)

$$E(\mathcal{B}) = w_{\text{reg}} E_{\text{reg}}(\mathcal{B}, \bar{\mathcal{B}}) + w_{\text{flesh}} E_{\text{flesh}}(\mathcal{B}, \mathcal{S}) + w_{\text{coll}} E_{\text{coll}}(\mathcal{B}, \mathcal{S}),$$

where  $E_{\text{flesh}}$  penalizes deformations of individual prisms but allows some stretching in the direction from skeleton to skin. In order to use this energy in the projective framework, we have to determine the amount of stretching  $\alpha$  for each prism in  $\tilde{\mathbf{S}}_p = \text{diag}(1, 1, \alpha)$ . Given the polar decomposition  $\mathbf{F}_p = \mathbf{R}_p \mathbf{S}_p$  of a prism’s deformation gradient, the stretching is given by  $\alpha = (\mathbf{B}_p^T \mathbf{S}_p \mathbf{B}_p)_{3,3}$ . This is clamped to the range  $[\alpha_{\text{min}}, \alpha_{\text{max}}]$ . We use  $\alpha_{\text{min}} = 0.2$  and  $\alpha_{\text{max}} = 5.0$  to allow stretching and compression of the element by a factor of five before the energy of this element increases. We set the weights  $w_{\text{reg}} = 0.1$ ,  $w_{\text{flesh}} = 0.01$ , and  $w_{\text{coll}} = 50$ . The minimization is iterated until convergence, meaning that for a fixed set of iterations the decrease of the energy falls below some threshold. In the converged state, we detect collisions and start the minimization again until convergence. This is repeated until no collisions are found in a converged solution. For all of our subjects, the minimization always converged within less than 20 iterations.

In order to fit the templates muscle surface to the target, we perform the minimization of (8)

$$E(\mathcal{M}) = w_{\text{reg}} E_{\text{reg}}(\mathcal{M}, \bar{\mathcal{M}}) + w_{\text{line}} E_{\text{line}}(\mathcal{M}, \mathcal{B}, \mathcal{S}).$$

We initialize  $\mathcal{M}$  with  $\bar{\mathcal{M}}$  and set  $w_{\text{reg}} = 0.01$ ,  $w_{\text{line}} = 1.0$ . When the minimization converges, we update the Laplacians in  $E_{\text{reg}}$  to those of the current solution and decrease  $w_{\text{reg}}$  by a factor of 0.5. This is iterated until the maximal distance of a vertex to its bone-to-skin line (see (9)) is less than 0.2 mm. Lastly, we project each vertex onto its corresponding bone-to-skin line to get a perfect alignment.