BOSS: Bones, Organs and Skin Shape Model

Karthik Shetty[1,2]   Annette Birkhold[2]   Srikrishna Jaganathan[1,2] Norbert Strobel[2] Bernhard Egger[1]
Markus Kowarschik[2]   Andreas Maier[1]

[1]FAU Erlangen-Nürnberg, Erlangen, Germany   [2]Siemens Healthineers AG, Forchheim, Germany

karthik.shetty@fau.de

Abstract

Objective: A digital twin of a patient can be a valuable tool for enhancing clinical tasks such as workflow automation, patient-specific X-ray dose optimization, markerless tracking, positioning, and navigation assistance in image-guided interventions. However, it is crucial that the patient's surface and internal organs are of high quality for any pose and shape estimates. At present, the majority of statistical shape models (SSMs) are restricted to a small number of organs or bones or do not adequately represent the general population. Method: To address this, we propose a deformable human shape and pose model that combines skin, internal organs, and bones, learned from CT images. By modeling the statistical variations in a pose-normalized space using probabilistic PCA while also preserving joint kinematics, our approach offers a holistic representation of the body that can benefit various medical applications. Results: We assessed our model's performance on a registered dataset, utilizing the unified shape space, and noted an average error of 3.6 mm for bones and 8.8 mm for organs. To further verify our findings, we conducted additional tests on publicly available datasets with multi-part segmentations, which confirmed the effectiveness of our model. Conclusion: This works shows that anatomically parameterized statistical shape models can be created accurately and in a computationally efficient manner. Significance: The proposed approach enables the construction of shape models that can be directly applied to various medical applications, including biomechanics and reconstruction.

1. Introduction

A virtual patient model that incorporates anatomy and physiology has the potential to enhance numerous medical diagnosis and therapy tasks. In particular, in the field of minimally invasive image-guided procedures, a detailed understanding of the patient being treated could facilitate (semi-) automated treatments. Given that not all necessary information about a patient is always available, a statistical model of body shape can serve as a foundation for incorporating patient-specific information. Such a Statistical Shape Model (SSM) of the human anatomy typically represents the average shape of multiple subjects and their variation in shape using a low-dimensional parameter space [1, 2]. Further, by incorporating kinematics, thereby the pose, the model can realistically represent a human body during a medical intervention.

The availability of such patient models opens up possibilities for automating clinical workflow steps entirely based on simulations. For example, in interventional environments such a model can be used to estimate and reduce the X-ray dose distribution within the patient as well as the dose the staff is exposed to due to scattered radiation [3,4]. Other applications are the automated positioning of the imaging system relative to a target region, such as an organ, or the virtual display of X-ray images that would be acquired based on the current system settings and position and patient pose prior to any X-ray exposure. It was also demonstrated such a generation of virtual X-ray images for bony structures based on pre-existing 3D hip models which are later iteratively fit to a patient's X-ray [5,6].

A virtual twin of the patient could be further employed for enhanced segmentation of anatomical structures [7], if (preoperative) images of a patient are available or are generated during the intervention. This would further facilitate the (pre- or intraoperative) generation of a patient-specific hierarchical model. It has been shown, that coupling a statistical knee model with a segmentation neural network a more precise segmentation of the knee based on magnetic resonance images can be achieved [8]. Shape models can be beneficial for the segmentation of 3D volumetric images such as Computed Tomography (CT) or Magnetic Resonance Imaging (MRI) [9–11], as well as for 2D projection images such as X-rays [12]. This is because the prior knowledge of the body shape can be used as a regularizer. Employed on a full body patient model, this technique could be applied online during image-guided (MR or X-ray) medical interventions and might therefore facilitate navigation

during the procedure. Additionally, shape models can be used to generate synthetic data to train deep neural networks, allowing them to better generalize [13, 14] to new subjects and overcome the limited anatomical variation often observed during domain translation.

Anatomic body models are similar to statistical shape models in that they are both computer-generated models of the human anatomy. However, anatomic body models are designed to provide a highly detailed representation of the anatomy of a single individual, while statistical shape models aim to capture the variation in shape that occurs across the population. Anatomic body models typically include different body parts, such as bones, muscles, organs, and vessels [15, 16]. They can be used for a variety of purposes, including visualization of anatomy for educational purposes, understanding relationships between different structures, and simulations of medical procedures. Despite their level of detail, the focus on a single individual's anatomy may limit their practicality in real-world applications. We hold the view that by leveraging the abundance of data and the progress in deep learning, the gap between statistical models and anatomical models can be bridged through a data-driven approach.

While prior studies in shape modeling focused primarily on specific anatomical regions (few bones or organs), this research aims to develop a comprehensive model representing the entire human body. Numerous works have proposed various solutions to create personalized anatomical models, such as Meng et.al.'s realistic spine model, which was learned from partial scans [17]. Kadlecek et.al. proposed personalized anatomical models for physics-based animation using only the surface 3D scans of skin [18]. However, the bone structures in these models are based on uniform scaling, which may not accurately represent the true structure. While the individual components of the BASH model provide accurate representations of kinematics and skin surface, combining them using linear interpolation may not fully capture the true anatomy. The most similar model to our work is likely the recent OSSO model, which combines a skin and bone model [19]. OSSO is a data-driven model that infers the shape of the skeleton from a given human surface scan. However, it is important to note that the skeleton model in OSSO is learned solely from 2D data, even though the method is data-driven.

In the field of patient modelling, the most prominent approach describing whole body anatomies is based on 4D extended cardiac-torso (XCAT) [20] employing nonuniform rational B-spline (NURBS). The underlying model is based on male and female CT scan from the Visible Human dataset [21] which were manually segmented on a fixed pose. The model was extended to incorporate cardiac and respiratory motions as a parametric function, based on cardiac- and respiratory-gated multislice CT data. To en-

hance the model's representation, it was further extended to handle CT scans from individuals of varying age, height, and body mass [22, 23]. However, this process is computationally expensive since each segmented CT scan must be registered to a template, followed by large deformation diffeomorphic metric mapping between the initial XCAT model and the registered scan. Furthermore, the robustness of this method decreases when patients are in different poses.

Various techniques have been introduced to create realistic models of human skin for individuals using captured 3D range scans that modify a generic 3D skin template to fit the particular person. For instance, Allen et al. employed nonrigid template fitting to compute correspondences between human body shapes in similar poses [24]. This method has been extended to work for varying poses [25, 26]. A dataset of varying shapes and poses can be utilized to create a data-driven model for building a human skin surface shape model that incorporates these variations [1, 2, 27]. These models represent the variations in human shape identity using a PCA space, and variations due to pose using an articulated skeleton-based skinning approach.

As these methods have shown promising results for the human skin, in this paper we extend the existing method based on SMPL [1] to combine skin, skeleton and internal organs as one parameterized full body model. We start by creating an articulated template mesh for skeleton and organs. Followed by registration of the individual structures, which is then pose-normalized to a rest pose. The shape space is then learnt on the rest pose. This is the first joint model for bones, organs and skin covering the full body. This demonstrates the applicability of the concepts of statistical shape models to a full-body scale, encompassing multiple layers. Such a comprehensive model enables holistic image processing and we expect it to be highly valuable to improve image acquisition and understanding.

2. Methodology

Our statistical model is based on a set publicly available whole body as well as partial CT scans (total ≈ 300). We performed automatic segmentation on them to isolate skeletons, internal organs and skin surfaces to model those components individually. Unlike SMPL or other well known statistical parametric models, the availability of a pose-only dataset is not feasible as we cannot subject a patient with unnecessary radiation. Usually while modelling the human skin, the pose-dataset is primarily used to learn the pose related shape variations, such as the bulging of the muscles.

2.1. SMPL Introduction

We provide necessary background on the SMPL framework, which forms the basis of our work [1]. The SMPL model is a statistical parametric function $M(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{t}; \boldsymbol{\Phi})$,

where $\boldsymbol{\beta}$ are the shape parameters, $\boldsymbol{\theta}$ are the pose parameters, $\boldsymbol{t}$ represents the global translation and, $\boldsymbol{\Phi}$ represents the learned model parameters. The output of this function is a triangulated surface mesh with $N = 6890$ vertices. The shape parameters $\boldsymbol{\beta}$ are represented as low-dimensional PCA coefficients, learned from a standard shape dataset [28] with $B_S(\boldsymbol{\beta}) : \mathbb{R}^{|\boldsymbol{\beta}|} \mapsto \mathbb{R}^{3N}$ representing offsets to a mean template mesh $\bar{\mathbf{T}}$ in the rest pose.

The pose of the model is defined by a kinematics chain with a set of relative rotation matrices $\mathcal{R} = [\mathbf{R}_1, \ldots, \mathbf{R}_K] \in \mathbb{R}^{3 \times 3}$ made up of $K = 24$ joints. Each set of rotation matrices is a function of the pose parameters $\boldsymbol{\theta}_i \in \mathbb{R}^3$ which represents the axis-angle rotations relative to the joints. Deforming the template model from its rest pose to a desired pose based on the relative rotations is known as forward kinematics. Let $\boldsymbol{t}_k \in \mathbb{R}^3$ represent the joint locations of the rest pose template. The new joint locations $\boldsymbol{q}_k \in \mathbb{R}^3$ of the deformed model are determined by

$$\boldsymbol{q}_k = \mathbf{G}_k(\boldsymbol{t}_k - \boldsymbol{t}_{p(k)}) + \boldsymbol{q}_{p(k)},$$

where $\mathbf{G}_k = \mathbf{R}_k \mathbf{G}_{p(k)} \in \mathbb{R}^{3 \times 3}$ are the global rotation of each joint k calculated recursively based on the kinematic tree. Here, $p(k)$ represents the parent joint for joint k. For more in depth description and implementations refer to [1, 29].

The deformation process requires additional parameters such as the joint regressor $\mathcal{J}(.)$, blend weights $\mathcal{W}$ and pose-dependent shape variations $B_P(.)$. The joint regressor $\mathcal{J}(\boldsymbol{\beta})$ computes the new joint location $\boldsymbol{t}_k$ for different body shapes in the rest pose. The blend weights and pose-dependent shape variations are learned from a pose dataset consisting of multiple subjects in various poses. The smoothing function of rotating vertices around a given joint is determined by the blend weights $\mathcal{W}$, using a given blend skinning function $W(.)$. The pose-dependent shape variations $B_P(\boldsymbol{\theta}) : \mathbb{R}^{|\boldsymbol{\theta}|} \mapsto \mathbb{R}^{3N}$ represents the offsets to a template mesh in the rest pose. Provided a template mesh $\bar{\mathbf{T}}$, a morphed model can be represented as

$$\mathcal{M} = W(\bar{\mathbf{T}} + B_S(\boldsymbol{\beta}) + B_P(\boldsymbol{\theta}), \boldsymbol{\theta}, \mathcal{J}(\boldsymbol{\beta}), \mathcal{W}). \quad (1)$$

In order to achieve smooth deformations, the Linear Blend Skinning (LBS) $W(.)$ method is used. This works by assigning blend weights $\mathcal{W}$ to each vertex of the mesh, which determine how much each segment of the skeleton affects the vertex's rotation. The resulting transformed vertex position $\mathbf{v}i'$ is given by the sum of the contributions from each segment

$$\mathbf{v}_i' = \sum_{k=1}^{K} w_{k,i} \mathcal{G}_k \mathbf{v}_i$$

Here, $w_{k,i}$ is an element of the blend weights $\mathcal{W}$ that corresponds to segment k and vertex index i, and $\mathcal{G}_k = [\mathbf{G}_k | \boldsymbol{q}_k]$ is the global transformation of joint k.

2.2. Data

In this study, 3D CT images of nearly 300 patients were used for the statistical modeling. Due to the limited public availability of whole body CT scans, we relied on two types of datasets, whole-body scans and partial-body scans. Whole-body scans typically consists of the entire human body from head to toe with a exceptions missing parts around the arm regions. In total, 42 valid scans from the Visceral dataset [30] make up the whole-body scans with an average voxel resolution of $0.87 \times 0.87 \times 2.0$ mm. The partial-body scans, on the other hand, predominantly cover scan areas ranging from the neck to the femur region. This set comprises 58 valid scans from the Visceral dataset [30], and 206 valid scans from the QIN-HeadNeck dataset [31] with an average voxel resolution of $0.91 \times 0.91 \times 1.5$ mm and $0.97 \times 0.97 \times 2.0$ mm, respectively. Additional metadata such as the height and weight of the patient during the CT acquisition was available for the QIN-HeadNeck dataset. On both datasets we have only included patients who have 12 thoracic vertebra and 5 lumbar vertebra as we are interested in the general population who can modeled based off the template skeleton model. Hence, we cannot account for varying amount of bones. In addition we excluded all patients with significant metallic implants as they cause imaging artifacts which degrade the quality of the segmented CT images. We further rejected patients with missing organs such as the kidneys. In total, we make use of 306 CT scans for building the model.

We anatomically segmented the 3D scans into 3 distinct sets, namely skin, bones and organs. This was automatically performed using the AI-Rad Companion Organs RT1 software (Siemens Healthcare GmbH, Erlangen, Germany). Note that we use the term organs loosely in this context as it refers to a combined representation of lungs, liver, heart, kidneys, bladder, rectum, esophagus, aorta, and the bowel region. From the segmented volumes we extracted skin $\mathbf{S}_s^i$, bone $\mathbf{S}_b^i$, and organ $\mathbf{S}_o^{i,k}$ surface meshes using the Marching Cubes [32] algorithm. Here $i \in [1, |\mathbf{S}|]$ represents the volume index and k represents the organ index. Fig. 1 shows an example surface mesh of a patient. Further, we manually annotated around 60 landmarks L_j^i both on the skin and on the bones with the bone's structure as reference guides. This also acts as a guide in determining the availability of segments such as the arms and legs.

2.3. Skin Model

In order to create a consolidated human model, it is fundamental that we establish dense correspondence between all scans in the dataset. The typical approach of achiev-

(a)   (b)

Figure 1. Example surface mesh with landmarks (depicted in the image with a blue cross) from the Visceral dataset [30] of (a) skin and (b) bone-organ obtained from a segmented CT scan.

ing this is to start from a common template mesh which ensures identical mesh topology for all the scans [26, 33]. This is accompanied by a skeleton-based deformation of the template to estimate the rough pose and shape of the scan. Subsequently a non-rigid deformation technique is tasked to fit the template to the surface scan. The two stage process of registering reduces the convergence speed and improves accuracy drastically. To avoid falling onto a local minima we make use of previously described manual landmarks to achieve the initial pose deformation of the template.

2.3.1 Skin Registration

With the availability of a trained SMPL model, the registration process is simplified due to the preexistence of a SSM. Hence, the pose and shape can be optimized simultaneously by minimizing $E_{lm} + E_{data}$, where E_{lm} is the landmark-loss (Eq. 4) and E_{data} the data-loss (Eq. 2).

$$E_{data} = \lambda_{d1} E_d(\mathbf{S}_s^i, M_s) + \lambda_{d2} E_d(M_s, \mathbf{S}_s^i) \qquad (2)$$

$$E_d(\mathbf{S}, M(\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{t}; \boldsymbol{\Phi})) = \sum_{m_j \in (M)} p_j \rho \left(||m_j - \mathcal{N}(\mathbf{S})||_2^2 \right)$$
$$\qquad (3)$$

$$E_{lm} = \sum_{l_j \in L(M)} ||l_j - L_j||_1 \qquad (4)$$

The data loss accounts for the distance between the skin SMPL model M and the surface $\mathbf{S}_s^i$. As correspondence is not present implicitly, we select the nearest neighbour $\mathcal{N}$ in

$\mathbf{S}_s^i$ for any given vertex in M for $E_d(\mathbf{S}_s, M_s)$ and inversely for $E_d(M_s, \mathbf{S}_s)$. For robustness, we discard matches where the angle between the corresponding normals are above a threshold of $30°$ and when the distance between the points are greater than 30 mm by setting p_j in (3) to either 1 or 0. We make use of the robust Geman-McClure function [34], represented by ρ to handle noisy data. The data-loss alone could fit the two surfaces if they are close and and on comparable poses, else the optimization process may get stuck in a local minima. The landmark-loss penalizes misalignment between the set of manually annotated landmarks L_j^i of each scan with the corresponding vertices $L(M)$ of the skin model. In addition to E_{lm} and E_{data} we further add regularization in the form of pose E_θ, shape E_β, weight E_w and height prior E_h. The pose prior is the same as Eq. 5 from [35]. Un-natural poses especially for the arms could lead to lower data error, hence these terms tries to keep the poses in a realistic range. We make use of the Gaussian mixture model provided by [35] for the pose prior. As the patient are lying on a table in a supine position, whereas the SMPL model was trained on standing pose we further add an exponential pose loss along the sagital plane for the joints near the thorax and abdomen region. The shape prior $E_\beta = \sum ||\beta||$ forces the shapes to be close to the mean shape. For the partial-body scans, we add additional prior loss E_h and E_w, when the patients height and weight are known. The height and weight of the skin model are measured in the rest pose for a given shape parameter $\boldsymbol{\beta}$. The height is measured from the head to toe, whereas the weight is measured as function of the mesh volume $V(M)$ as described in [36].

$$E_w = ||(V(M_s^i) + 4.937)/1.015 - w^i||_2^2. \qquad (5)$$

We optimize the following energy term

$$\underset{\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{t}}{\arg\min}(E_{lm} + E_{data} + E_\theta + E_\beta + E_w + E_h), \qquad (6)$$

to obtain an initial fit $\tilde{\mathbf{V}}_s^i$. The weight terms λ_x associated with the energy term E_x is omitted for easier readability. The initial fit is obtained by first by minimizing the global translation $\boldsymbol{t}$, followed by the pose $\boldsymbol{\theta}$, shape $\boldsymbol{\beta}$ and finally all three parameters simultaneously.

If the scans $\mathbf{S}_s^i$ were in a standing pose, registration would be achieved under the assumption that the SMPL model represents an adequate space of human shape variations. However, the model we would like to represent are the ones in a supine position, taken during a CT procedure. This usually causes the backs to become flat, stomach to be depressed and the chest bulged out. Hence, we perform non-rigid registration on the initial fit $\tilde{\mathbf{V}}_s^i$. Similar to the works from [24, 37], we represent a set of 3×4 affine transformation matrices $\mathbf{A}_j^i$ associated with each vertex of the

initial fit $\tilde{\mathbf{V}}_{s,k}^i$, with the aim to align the vertices to the scan $\mathbf{S}_s^i$. This is achieved by minimizing $E_{lm} + E_{data} + E_s + E_o$, where E_s and E_o are smoothing and orthogonality constraints [37] respectively. To achieve local rigidity, the affine transformations applied on the vertices $\tilde{\mathbf{V}}_{s,k}^i$ need to be close to the transformations on the neighbouring vertices $\tilde{\mathbf{V}}_{s,k}^i \in \mathcal{N}(\tilde{\mathbf{V}}_{s,j}^i)$. Therefore, the smoothness term $E_s(\tilde{\mathbf{V}}_s^i)$ can be defined as

$$E_s(\mathbf{p}) = \sum_{\{j,k|\{\mathbf{p}_j,\mathbf{p}_k\} \in \text{edges}(\mathbf{p})\}} c_{ij} \left\| \mathbf{A}_j \mathbf{p}_j - \mathbf{A}_k \mathbf{p}_k \right\|_2^2. \tag{7}$$

Here, c_{ij} represents the Laplacian cotangent edge weights, which tries to make changes on the transformation matrices $\mathbf{A}_j^i$ over the mesh as smooth as possible [38]. The orthogonality constraint additionally preserves local rigidity during registration by enforcing the affine transformation to be close to a rigid transformation by

$$E_o = \sum_j \left\| \mathbf{A}_j^i - \mathbf{R}_j^i \right\|_F^2. \tag{8}$$

Here $\mathbf{R}_j$ is the closest projection of $\mathbf{A}_j$ onto the rotation matrix group. This can be extracted by performing Singular Value Decomposition on the transformation matrix. All energies are minimized to obtain a final non-rigid fit $\mathbf{V}_s^i$ using a gradient-based LBFGS [39] minimization method and make use of automatic differentiation packages.

The main advantage of the two step process of registration is that it can handle scans with missing data or holes. Missing data here refers to the non-availability of scans sections such as the arms or legs from the partial-body dataset. Missing data is identified by the non-availability of landmarks for a given scan. As the SMPL model is divided into 24 sections, we can prevent pose deformation and data loss minimization on those sections. Using a SSM reduces the search space, and can provide shape in the realm of probable human shapes.

2.4. Bone-Organ Model

Unlike the skin model, publically available SSMs for bones and organs do not exist. For this purpose we create a deformable model from scratch. A template mesh is derived from an existing polygon data BodyParts3D, which was extracted from full body MRI images [40]. The bone model is made up of 70 segments, including skull, femur, humerus, forearm, lower leg, scapula, clavicle, sternum, hands, feet, vertebra, ribs, and pelvis. The organ model includes lungs, liver, kidneys, spleen, heart, bladder, rectum, esophagus, and aorta. We also incorporate the bowel region containing the stomach and intestines. However, segmentation for the individual bowel components are not available, rather a hull enclosing the stomach and intestines. The entire bone-organ template is made up of $104,546$ vertices and $209,418$ faces,

Figure 2. Template bone-organ model, consisting of lungs, liver, kidneys, spleen, heart along with aorta and bladder

of which $65,617$ vertices are made up of the bone section and the remaining for the internal organs.

On top of the template, we define a kinematics chain made up of $N_b = 63$ joints comprising 63 segments. Though we start with 70 individual segments, we consider femur-patella and all cervical vertebrae as combined segments. Linear blend skinning is adopted on femur-patella-tibia and cervical section to achieve a smooth deformable bone model. The blend weights are set to 1 for the rigid entity with respect to their own segments, whereas for the composite structure it is evenly distributed between the parent and child segment. The initial blend weights for the organs are set only with respect to the vertebral section. We make use of Blender [41] to automatically generate these weights.

We rigidly deform the vertebra of the bone-organ mesh model to one of the segmented CT volumes of comparable shape and size, such that the mesh represents a person laying in supine pose. Similarly, we define the skin template $\tilde{\mathbf{T}}_s$ in supine pose by re-posing the non-rigid skin model to a T-pose of the same CT volume. Additionally, we also rotate the arms and legs of the bone-organ model, such that they lay inside and follow the same T-pose as the skin from the SMPL model. The final template bone-organ template mesh $\tilde{\mathbf{T}}_{bo}$ is shown in Fig. 2 .

2.4.1 Bone Registration

The registration process in general follows the methodology as described in Sec. 2.3. However, estimating the rough pose followed by non-rigid registration is not feasible by virtue of the complex thin structure of bones. This problem

is particularly evident on the scapula and clavicle, which leads to incorrect poses for the rest of the template. To address this, we simultaneously estimate a rough shape and pose. The shape variations are achieved by applying a scale transform along a segment in world coordinates. Consequently, the joint locations along the kinematic chain are also scaled by the same amount.

Hereby a simplistic deformable bone model can be expressed as $M^b(\hat{\boldsymbol{\beta}}_b, \boldsymbol{\theta}_b, \boldsymbol{t}_b; \boldsymbol{\Phi}_b)$, where $\hat{\boldsymbol{\beta}}_b \in \mathbb{R}^{63 \times 3}$ represents the scaling parameters, $\boldsymbol{\theta}_b \in \mathbb{R}^{63 \times 3}$ represents the pose parameters, $\boldsymbol{t}_b \in \mathbb{R}^{63 \times 3}$ represents the individual segment translation parameters, and $\boldsymbol{\Phi}_b$ represent the model parameters comprising of the kinematic chain and the initial blend weights.

Figure 3. Example of the virtual edges between two vertebral segments to prevent mesh overlap during non-rigid registration. Blue and Orange lines represent inter-segment and intra-segment virtual edges respectively.

Similar to the skin registration from Sec. 2.3, registration of the bones can be performed with a few minor changes as described

$$\underset{\hat{\boldsymbol{\beta}}_b, \boldsymbol{\theta}_b, \boldsymbol{t}_b}{\arg\min}(E_{lm} + E_{data} + E_{\theta_b} + E_{\hat{\beta}_b} + E_{t_b} + E_{lm}^s). \quad (9)$$

The bone surface scans, represented by $\mathbf{S}_b^i$, are made up of 26 separate segments including 24 ribs, the sternum, and the rest of the skeletal structure. The set of vertices corresponding to a respective segment on the surface scans and template mesh are represented by $\mathbf{ss}$ and $\mathbf{sm}$ respectively. The data loss E_{data} term from Eq. 4 is replaced with Eq. 10. The pose $E_{\theta_b} = \sum ||\boldsymbol{\theta}_b^j - \boldsymbol{\theta}_b^{p(j)}||$ and shape $E_{\hat{\beta}_b} = \sum ||\hat{\boldsymbol{\beta}}_b^j - \hat{\boldsymbol{\beta}}_b^{p(j)}||$ prior loss forces the registered model maintain the original template shape and prevent unnatural poses. Here, $p(j)$ denotes the parent joint of joint j. The pose prior is defined only for the vertebral section. As the shape here represents a scaling function, the shape prior forces the scaling to be similar to that of its parent segment. The incorporation of individual translation $\boldsymbol{t}$ allows free floating segments such as the sternum, scapula, etc. to move freely. However, we add a translation prior $E_{t_b} = \sum ||\boldsymbol{t}_b^j||$ that prevents large translation movement far away from its initial position. In addition to the landmark error E_{lm}, we include a skin-based vicinity landmarks E_{lm}^s. This helps to

generate an estimated fit of the bones in the case of missing data around the leg and arm regions. Here, we use a learnt mapping between the skin vertices and joints of skeletons $\mathcal{J}(M_s)$, in particular for the arms, legs, hand and feet. The mapping is learnt on the set of registered scans where aforementioned sections where present in the CT data. Using the registered skin as a reference, we minimize the loss between the predicted joint locations $\mathcal{J}(M_s)$ and the joint locations obtained from the skeleton model $M^b(\boldsymbol{\beta}^b, \boldsymbol{\theta}^b)$.

$$E_{data} = \sum_{\{\mathbf{sm}_k, \mathbf{ss}_k \in \text{segments}(M_b, \mathbf{S}_b)\}} (\lambda_{d1} E_d(\mathbf{sm}_k, \mathbf{ss}_k) + \lambda_{d2} E_d(\mathbf{ss}_k, \mathbf{sm}_k))$$
$$(10)$$

Similar to the skin, non-rigid registration is performed for the bones by minimizing $E_{lm} + E_{data} + E_s + E_o$. Unlike skin, the bones are narrow structures, resulting in severe mesh overlap when using the same formulation. To prevent this, we create inter-segment and intra-segment virtual edges on the template mesh as shown in Fig.3. These virtual edges are defined based on a distance threshold and direction of the vertex normals. For inter-segment edges the normals between two set of vertices need to be facing each other whereas for the intra-segment edges the normals need to be facing away from each other.

Organ registration is a straightforward process that is similar to skin registration. The initial fit is obtained during the rigid registration of the bones. The non-rigid registration is accomplished through affine transformations applied directly to individual translations. To avoid mesh overlap, we include virtual edges between organs. However, the bowel region presents a challenge because there is no distinct segmentation, only a boundary. To address this, we use a higher regularizer loss and a penetration loss. The penetration loss calculates if any vertices of the bowel are inside another body region and penalizes the distance to the nearest boundary. Examples of registered scans are presented in Figure 4.

2.5. Model Formulation

With the availability of registered scans $\mathbf{V}^s$ and $\mathbf{V}^{bo}$ in the common mesh topology, we can now formulate the deformable shape model for the skin and internal body. It is necessary that all scans are normalized to the rest pose as defined by their respective templates $\tilde{\mathbf{T}}^s$ and $\bar{\mathbf{T}}^{bo}$ in order to remove variance related to body articulation. We first perform the unposing (transforming to rest pose) operation on the skin, followed by the internal body. Unposing the skin first, allows us to use the skin as a guide for unposing the internal body later. From the pose-normalized models, it is then possible to learn a joint shape-space of the skin-bone-organ model.

Figure 4. Examples of registered patients.

2.5.1 Skin Unposing

We unpose the registered skin mesh $\mathbf{V}_s^i$ to the rest pose $\mathbf{U}_s^i$, defined by the template skin mesh $\tilde{\mathbf{T}}_s$. As we prefer the patient skin model to operate similarly to that of the original SMPL model, we optimize for a new joint regressor $\mathcal{J}_s$, blend weights $\mathcal{W}_s$ and pose related shape variations B_{P_s}. This is done by minimizing the distance between the skin model vertices and the registration vertices as described in the following equation

$$\underset{\mathbf{U}_s^i,\tilde{\boldsymbol{\theta}}_s^i,B_{P_s},\mathcal{J}_s,}{\arg\min} \sum_i^{S_n} \left\| W(\mathbf{U}_s^i+B_{P_s}(\tilde{\boldsymbol{\theta}}_s^i),\tilde{\boldsymbol{\theta}}_s^i;\mathcal{J}_s,\mathcal{W}_s)-\mathbf{V}_s^i \right\|^2 . \quad (11)$$

During optimization, we initialize $\mathcal{J}_s$, $\mathcal{W}_s$ and B_{P_s} with the original SMPL parameters. The rest pose vertices $\mathbf{U}_s^i$ are initialized $M(\boldsymbol{\beta}^i)$ with only the shape parameters $\boldsymbol{\beta}^i$ obtained from the initial rigid skin fit. Similarly, the poses $\tilde{\boldsymbol{\theta}}_s^i$ are initialized with the poses $\boldsymbol{\theta}^i$ obtained from the initial rigid skin fit. To stabilize the optimization, we make use of multiple regularizers based on various assumptions.

As, the pose normalized subjects $\mathbf{U}_s^i$ needs to be aligned to the template mesh $\tilde{\mathbf{T}}_s$, we add an edge loss of the form $\sum_i^{S_n} ||\mathbf{U}_{s,e}^i - \tilde{\mathbf{T}}_{s,e}||$, where $\mathbf{U}_{s,e}^i, \tilde{\mathbf{T}}_{s,e} \in \text{edges}(\tilde{\mathbf{T}}_s)$ represents the normalized direction vector for a pair of neighbouring vertices. We further add constraints on $\tilde{\boldsymbol{\theta}}_s^i$, $\mathcal{J}_s$, $\mathcal{W}_s$ and B_{P_s} in the form of $L2$ loss, to not deviate too much from its initial values. In the original SMPL model, the joint regressor was computed using non-negative least squares, with a constraint that the weights add up to 1. We maintain similar setting during the joint optimization by normalizing

the joint regressor as $|\mathcal{J}_s|/|||\mathcal{J}_s|||_1$. During the optimization, we further add a regularizing term defined as

$$\sum_i^{S_n} ||\mathcal{J}_s(W(\mathbf{U}_s^i,\tilde{\boldsymbol{\theta}}_s^i,B_{P_s},\mathcal{J}_s)) - \mathcal{J}(\tilde{\mathbf{V}}_s^i)||^2,$$

on the joint locations, such that it is close to the joint locations from the original rigid fit.

Symmetry regularizer is applied on the pose-normalized mesh vertices $\mathbf{U}_s^i$ and on its joint locations $\mathcal{J}_s(\mathbf{U}_s^i)$ to enforce symmetry along the sagittal plane defined as

$$\sum_i^{S_n} ||\mathbf{U}_s^i - \mathrm{m}(\mathbf{U}_s^i)|| + ||\mathcal{J}_s(\mathbf{U}_s^i) - \mathrm{m}(\mathcal{J}_s(\mathbf{U}_s^i))||^2, \quad (12)$$

where m denotes the mirror vertices or mirror joints.

2.5.2 Bone-Organ Unposing

Similar to the skin model, we perform unposing of the bone-organ registered mesh $\mathbf{V}_{bo}^i$ to the rest pose $\mathbf{U}_{bo}^i$. However, performing exactly the same would lead to incorrect skeletal localization and mesh overlap of the internal body and the skin.

The objective function defined in Eq. 13 is to unpose the body-organ model by minimizing the distance between the model vertices $\bar{\mathbf{V}}_{bo}^i = W(\mathbf{U}_{bo}^i, \tilde{\boldsymbol{\theta}}_{bo}^i; \mathcal{J}_{bo}, \mathcal{W}_{bo})$ and the registration vertices. Additionally, the objective function defined in Eq. 13 maintains the distance $\mathcal{P}$ between the skin and bone-organ model among both posed and unposed states as shown in Eq. 14. We do this with pose-only $\bar{\boldsymbol{\theta}}_{bo}^i$ deformation of the unposed model $\bar{\mathbf{U}}_{bo}^i = W(\mathbf{U}_{bo}^i, \bar{\boldsymbol{\theta}}_{bo}^i; \mathcal{J}_{bo}, \mathcal{W}_{bo})$. For each skin vertex of the registered skin mesh $\mathbf{V}_s^i$, we find at most 6 bone vertices from the registered bone-organ mesh $\mathbf{V}_{bo}^i$ based on empirically chosen distance (based on human size and the particular bone segment) and vertex normal (skin and bone normal within $30 \deg$) thresholds. However, we ignore skin vertices around the complex shoulders, elbows and knees joints, along with bone vertices for scapula and clavicle. This avoids any pose related influences during optimization around these regions.

$$\underset{\mathbf{U}_{bo}^i,\tilde{\boldsymbol{\theta}}_{bo}^i,\mathcal{J}_{bo},\mathcal{W}_{bo}}{\arg\min} \sum_i^{S_n} \left\| \bar{\mathbf{V}}_{bo}^i - \mathbf{V}_{bo}^i \right\|^2 ; \quad (13)$$

$$\underset{\tilde{\boldsymbol{\theta}}_{bo}^i,\bar{\boldsymbol{\theta}}_{bo}^i,\mathcal{J}_{bo},\mathcal{W}_{bo}}{\arg\min} \sum_i^{S_n} ||\mathcal{P}(\bar{\mathbf{U}}_{bo}^i, \mathbf{U}_s^i) - \mathcal{P}(\bar{\mathbf{V}}_{bo}^i, \mathbf{V}_s^i)|| \quad (14)$$

We start off by initialising the joint regressor $\mathcal{J}_{bo}$ on the template bone-organ model $\mathbf{T}_{bo}$ based on the joint locations from its initial kinematic chain. The joints are always located between 2 bone segments. We randomly sample 50

Figure 5. The first three principal components of body shape are shown, varying about 2 standard deviations after normalizing the variance. One could infer that the height and weight of the patient are mostly explained by the first two components

closest vertices to the joint from both segments where the vertex normals approximately faces the vertex-joint direction. The joint regressor $\mathcal{J}_{bo}$ is learnt using a least square fit for the sampled vertices. The rest pose vertices $\mathbf{U}_{bo}^i$ are initialized $M^{bo}(\hat{\boldsymbol{\beta}}_b^i)$ with only the scale parameters $\hat{\boldsymbol{\beta}}_b^i$ obtained from the initial rigid bone fit. Similar to the skin model, the blend weights $\mathcal{W}_{bo}$ and poses $\tilde{\boldsymbol{\theta}}_{bo}^i$ are initialized. During unposing of the bone-organ model, we define that the motion of ribs, sternum and pelvis are a function of shape rather than a function of the pose. Hence, we initialize the pose to zero for these particular segments. Note that, only the ribs and sternum are leaf nodes in our kinematic chain, i.e. there are no child segments. However, for the femur, we additionally include the pelvis rotation, as it is its parent node.

To stabilize the optimization we make use of similar regularizes defined in the skin model for both the objectives functions. For the edge loss, we additionally incorporate the virtual edges from the registration process with lower

Figure 6. Cumulative variance of the skin-bone-organ model shape space.

weightage to reduce mesh interpenetration. Symmetry regularizer is applied only to the bone structures. Regularizing the joint locations of bone and skin along the arms and legs are done by minimizing the following equation:

$$||\mathcal{J}_{bo}(\mathbf{U}_{bo}^i) - \mathcal{J}_s(\mathbf{U}_s^i)||^2.$$

We alternate between optimizing Eq. 13 and Eq. 14, while carrying the optimized parameters between them. For optimization of Eq. 14, we initialize the pose $\bar{\boldsymbol{\theta}}_{bo}^i$ with zero, and regularize them towards zero. While alternating to optimize Eq. 13, we initialize unposed vertices $\mathbf{U}_{bo}^i$ with the obtained vertices $\bar{\mathbf{U}}_{bo}^i$ after optimizing Eq. 14.

2.5.3 Shape Space

From the unposed skin $\mathbf{U}_s^i$ and bone-organ $\mathbf{U}_{bo}^i$ volume, we learn the shape components with the aid of mean and principal shape components. We do not have complete registrations around skulls, arms and legs for some of the volume. Hence, we use a publicly available implementation [1] of Probabilistic Principal Component Analysis (PPCA) [42], which can handle missing data. By performing PPCA, we obtain a mean skin $\mathbf{T}_\mu^s$, bone-organ $\mathbf{T}_\mu^{bo}$ and vertex offsets to the mean in the form of shape space for skin $\mathbf{B}_\mu^s$ and $\mathbf{B}_\mu^{bo}$ bone-organ.

3. Evaluation

In Fig. 5, we visualize the first three shape components, while Fig. 6 displays the cumulative variance of the full model. The first 10 shape components captures 88% of the variance, and the first 20 components capture 92% of the variance. Although the shape space of the skin and bone-organ are coupled, the kinematic model is separated to account for differences in skeletal posture. We use a neutral

---

[1]https://github.com/allentran/pca-magic

Figure 7. (a) Box-plot of model generalization on the registered ACRIN dataset using 10 and 50 shape components. (b) Box-plot of model generalization on the registered ACRIN dataset by only using the metadata and skin surface as a guide to determine the internal body shape.

Figure 8. Residual error on ACRIN dataset with respect to the different number of shape components. The centroid represents, the mean error of the centre of mass for individual segments between reconstructed and registered models.

model rather than separating into two genders on basis of the number of CT volumes available. We map the gender, height, and weight of the registered patients where available to the shape space coefficients using a linear regressor. We provide qualitative results of random samples generated in Fig. 11.

To demonstrate the effectiveness of the skin-bone-organ model we perform evaluations on multiple datasets. First we assess the model generalizability on a registered test set. We then demonstrate model completion from metadata such as height, weight, or surface scans. We finally evaluate the model performance on various public datasets, where segmentations of organs or bones are provided.

3.1. Model Generalization

We register the skin, bones and organs on separate held out test set. For the test set we use CT volumes from ACRIN-NSCLC-FDG-PET [43]. We ignore all volumes where the z-spacing is greater then 3.27 mm, while retaining volumes greater than 2.5 mm only if clear separation of the vertebra is visible. In total we test on 78 scans, which were registered to its segmentations as described in the methodology section.

We minimize the vertex-to-vertex error between the reconstructed model and the registration while regularizing the process with shape and pose priors. The shape prior keeps the reconstructed model close to its mean, while the pose prior restricts individual vertebrae from deviating too

Figure 9. Model generalization on VerSe dataset using 10 and 50 shape components.

Figure 10. Model generalization on various datasets using 10 and 50 shape components.

much from their parent and child nodes. It's important to note that we do not translate individual segments, and all the shape variations are accounted for only by the shape space of the model. In the case of the ACRIN dataset, regions such as the legs, arms, and part of the skull are not available, so we exclude these regions while measuring metrics and during reconstruction.

In Figure 7a, we present the box plot illustrating the error in individual segments for 10 and 50 shape components. The vertebra includes all vertebral sections, the pelvis includes the pelvis and sacral, and the skeleton comprises the remaining bone structures. In Figure 8, we display the generalization error across different numbers of shape components. The results indicate that the bones can be generalized well, with an error of 3.66 mm at 25 components. However, the errors for individual organs are significantly higher, as we optimize the model for the entire body rather than individual organs. This is expected since the shape space accounts for variations in organ shapes and locations. Using only 25 components, we achieve an average organ error of 8.83 mm, which can provide a starting point for downstream tasks involving organ shape and placement. We also measure the mean centroid error for each segment, which represents the center of mass error. With 25 components, we observe a mean centroid error of 2.52 mm.

Fig. 7b displays the ability of our model to estimate the shape of inner organs and skeleton using patient metadata or skin surface. For estimating with metadata, we use the linear regressor with patient's gender, weight, and height to determine the model shape coefficients. On the other hand, when using the skin surface, we only use the shape coefficients of our skin model that fit the target skin. In both cases, we optimize only for the pose while using the previously determined shape coefficients. We observe an average overall error of 8.11 mm and 8.68 mm using metadata and skin surface, respectively.

3.2. Model Generalization on Public Organ Segmentations

We evaluate the generalization of our model on public datasets containing multiple organs or bones. Specifically, we test our model on Verse [44] for vertebra, CTPEL [45] for pelvis and sacral, Abdomen1K [46] for liver, kidney, and spleen, and StructSeg [47] for lungs and heart. For all methods, we follow a similar registration process. To register to each dataset, we obtain a surface mesh of the segmented volume using marching cubes [32], and then minimize the chamfer loss between the vertices of our model and the target surface mesh. For regularization we use the same procedure as defined in Sec. 3.1 by using pose and shape prior with no individual translations, i.e. the combined shape model has to best represent all target surfaces. The final errors presented are the bidirectional point-to-surface distance [17].

The model's generalization performance for the vertebrae region on the Verse dataset is presented in Fig.9. The full Verse dataset, including the train, test and validation sets, was used for testing. However, volumes containing L6 and T13 vertebrae, which are not represented by our model, were excluded. The generalization errors for the pelvis and sacral region in CTPEL, for the liver, kidney, and spleen in StructSeg, and for the lungs and heart of the Abdomen1k dataset are presented in Fig. 10. In the Abdomen1k dataset, volumes containing incomplete liver were excluded from the analysis. The results obtained on these datasets are consistent with those observed on the ACRIN dataset.

4. Conclusion

A digital twin of a patient can be a valuable tool to improve multiple clinical tasks, such as automating clinical workflows, estimating patient-specific X-ray dose exposure to patient and medical staff, positioning, markerless tracking and navigation assistance in image-guided interventions. For the first time, we provide a joint model with an accurate estimation of shapes and poses of the patient

Figure 11. Random samples drawn from the model.

surface, skeleton and internal organs. We propose a deformable shape and pose human model devised out of individual segments such as the skin, internal organs and bones, learnt from surfaces extracted from segmented whole-body CT images. With surface errors of 3.66 mm for bones and an average organ error of 8.83 mm, we believe that the statistical model will allow many automation tasks in clinical workflows and lays the basis for fast intraoperative model personalisation given patient-specific information.

Disclaimer

The concepts and information presented in this article are based on research and are not commercially available.

References

[1] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black, "SMPL: A skinned multi-person linear model," *ACM Trans. Graphics (Proc. SIGGRAPH Asia)*, vol. 34, no. 6, pp. 248:1–248:16, Oct. 2015.

[2] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis, "Scape: shape completion and animation of people," in *ACM SIGGRAPH 2005 Papers*, 2005, pp. 408–416.

[3] P. Roser, X. Zhong, A. Birkhold, N. Strobel, M. Kowarschik, R. Fahrig, and A. Maier, "Physics-driven learning of x-ray skin dose distribution in interventional procedures," *Medical Physics*, vol. 46, no. 10, pp. 4654–4665, 2019.

[4] P. B. Johnson, D. Borrego, S. Balter, K. Johnson, D. Siragusa, and W. E. Bolch, "Skin dose mapping for fluoroscopically guided interventions," *Medical Physics*, vol. 38, no. 10, pp. 5490–5499, 2011.

[5] M. Ehlke, H. Ramm, H. Lamecker, H.-C. Hege, and S. Zachow, "Fast generation of virtual x-ray images for reconstruction of 3d anatomy," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 12, pp. 2673–2682, 2013.

[6] A. Maier, H. G. Hofmann, C. Schwemmer, J. Hornegger, A. Keil, and R. Fahrig, "Fast simulation of x-ray projections of spline-based surfaces using an append buffer," *Physics in Medicine & Biology*, vol. 57, no. 19, p. 6193, 2012.

[7] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3d medical image segmentation: A review," *Medical Image Analysis*, vol. 13, no. 4, pp. 543 – 563, 2009.

[8] F. Ambellan, A. Tack, M. Ehlke, and S. Zachow, "Automated segmentation of knee bone and cartilage combining statistical shape knowledge and convolutional neural networks: Data from the osteoarthritis initiative," *Medical Image Analysis*, vol. 52, no. 2, pp. 109 – 118, 2019.

[9] M. Esfandiarkhani and A. H. Foruzan, "A generalized active shape model for segmentation of liver in low-contrast ct volumes," *Computers in Biology and Medicine*, vol. 82, pp. 59–70, 2017.

[10] H. Seim, D. Kainmueller, H. Lamecker, M. Bindernagel, J. Malinowski, and S. Zachow, "Model-based auto-segmentation of knee bones and cartilage in mri data," in *Proc. MICCAI Workshop Medical Image Analysis for the Clinic*, 2010, pp. 215–223.

[11] X. Zhong, M. Amrehn, N. Ravikumar, S. Chen, N. Strobel, A. Birkhold, M. Kowarschik, R. Fahrig, and A. Maier, "Deep action learning enables robust 3d segmentation of body organs in various ct and mri images," *Scientific Reports*, vol. 11, no. 1, pp. 1–14, 2021.

[12] N. Baka, B. L. Kaptein, M. de Bruijne, T. van Walsum, J. Giphart, W. J. Niessen, and B. P. Lelieveldt, "2d–3d shape reconstruction of the distal femur from stereo x-ray imaging using statistical shape models," *Medical image analysis*, vol. 15, no. 6, pp. 840–850, 2011.

[13] A. Gilbert, M. Marciniak, C. Rodero, P. Lamata, E. Samset, and K. Mcleod, "Generating synthetic labeled data from existing anatomical models: an example with echocardiography segmentation," *IEEE Transactions on Medical Imaging*, vol. 40, no. 10, pp. 2783–2794, 2021.

[14] S. Amirrajab, S. Abbasi-Sureshjani, Y. Al Khalil, C. Lorenz, J. Weese, J. Pluim, and M. Breeuwer, "Xcat-gan for synthesizing 3d consistent labeled cardiac mr images on anatomically variable xcat phantoms," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2020, pp. 128–137.

[15] S. L. Delp, F. C. Anderson, A. S. Arnold, P. Loan, A. Habib, C. T. John, E. Guendelman, and D. G. Thelen, "Opensim: open-source software to create and analyze dynamic simulations of movement," *IEEE transactions on biomedical engineering*, vol. 54, no. 11, pp. 1940–1950, 2007.

[16] R. Kelc, "Zygote body: a new interactive 3-dimensional didactical tool for teaching anatomy," 2012.

[17] D. Meng, M. Keller, E. Boyer, M. Black, and S. Pujades, "Learning a statistical full spine model from partial observations," in *Shape in Medical Imaging*. Springer International Publishing, Oct. 2020, pp. 122,133.

[18] P. Kadlecek, A.-E. Ichim, T. Liu, J. Krivanek, and L. Kavan, "Reconstructing personalized anatomical models for physics-based body animation," *ACM Trans. Graph.*, vol. 35, no. 6, 2016.

[19] M. Keller, S. Zuffi, M. J. Black, and S. Pujades, "OSSO: Obtaining skeletal shape from outside," in *Proceedings IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2022, pp. 20 492–20 501.

[20] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4d xcat phantom for multimodality imaging research," *Medical Physics*, vol. 37, no. 9, pp. 4902–4915, 2010.

[21] M. J. Ackerman, "The visible human project: a resource for education." *Academic medicine: journal of the Association of American Medical Colleges*, vol. 74, no. 6, pp. 667–670, 1999.

[22] W. P. Segars, H. Norris, G. M. Sturgeon, Y. Zhang, J. Bond, A. Minhas, D. J. Tward, J. T. Ratnanather, M. I. Miller, D. Frush, and E. Samei, "The development of a population of 4d pediatric xcat phantoms for imaging research and optimization," *Medical Physics*, vol. 42, no. 8, pp. 4719–4726, 2015.

[23] W. P. Segars, J. Bond, J. Frush, S. Hon, C. Eckersley, C. H. Williams, J. Feng, D. J. Tward, J. T. Ratnanather, M. I. Miller, D. Frush, and E. Samei, "Population of anatomically variable 4d xcat adult phantoms for imaging research and optimization," *Medical Physics*, vol. 40, no. 4, p. 043701, 2013.

[24] B. Allen, B. Curless, and Z. Popović, "Articulated body deformation from range scan data," *ACM Transactions on Graphics (TOG)*, vol. 21, no. 3, pp. 612–619, 2002.

[25] D. Hirshberg, M. Loper, E. Rachlin, and M. Black, "Coregistration: Simultaneous alignment and modeling of articulated 3D shape," in *European Conf. on Computer Vision (ECCV)*, ser. LNCS 7577, Part IV. Springer-Verlag, Oct. 2012, pp. 242–255.

[26] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel, "A statistical model of human pose and body shape," in *Computer graphics forum*, vol. 28, no. 2. Wiley Online Library, 2009, pp. 337–346.

[27] L. Pishchulin, S. Wuhrer, T. Helten, C. Theobalt, and B. Schiele, "Building statistical shape spaces for 3d human modeling," *Pattern Recognition*, vol. 67, pp. 276–286, 2017.

[28] K. M. Robinette, S. Blackwell, H. Daanen, M. Boehmer, and S. Fleming, "Civilian american and european surface anthropometry resource (caesar), final report. volume 1. summary," Sytronics Inc Dayton Oh, Tech. Rep., 2002.

[29] R. M. Murray, Z. Li, S. S. Sastry, and S. S. Sastry, *A mathematical introduction to robotic manipulation*. CRC press, 1994.

[30] M. Krenn, K. Grünberg, O. Jimenez-del Toro, A. Jakab, T. Salas Fernandez, M. Winterstein, M.-A. Weber, and G. Langs, *Datasets Created in VISCERAL*. Cham: Springer International Publishing, 2017, pp. 69–84.

[31] R. R. Beichel, E. J. Ulrich, C. Bauer, A. Wahle, B. Brown, T. Chang, K. Plichta, B. Smith, J. Sunderland, T. Braun, A. Fedorov, D. Clunie, M. Onken, V. A. Magnotta, Y. Menda, J. Riesmeier, S. Pieper, R. Kikinis, M. Graham, T. Casavant, M. Sonka, and J. Buatti, "Data from qin-headneck," 2015. [Online]. Available: https://wiki.cancerimagingarchive.net/x/qQ9p

[32] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," *ACM siggraph computer graphics*, vol. 21, no. 4, pp. 163–169, 1987.

[33] B. Allen, B. Curless, and Z. Popović, "The space of human body shapes: reconstruction and parameterization from range scans," *ACM transactions on graphics (TOG)*, vol. 22, no. 3, pp. 587–594, 2003.

[34] S. Geman and D. E. McClure, "Statistical methods for tomographic image reconstruction," 1987.

[35] F. Bogo, A. Kanazawa, C. Lassner, P. V. Gehler, J. Romero, and M. J. Black, "Keep it smpl: Automatic estimation of 3d human pose and shape from a single image," in *ECCV*, 2016.

[36] S. Ice, "Obesity in perspective (volume 2, parts 1 and 2 of fogarty international center series on preventive medicine)," *Psychosomatic Medicine*, vol. 39, no. 3, 1977.

[37] K. Li, J. Yang, Y.-K. Lai, and D. Guo, "Robust non-rigid registration with reweighted position and transformation sparsity," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, pp. 2255–2269, 2019.

[38] O. Sorkine, "Laplacian mesh processing."

[39] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," vol. 45, no. 1-3, pp. 503–528, Aug. 1989. [Online]. Available: https://doi.org/10.1007/bf01589116

[40] N. Mitsuhashi, K. Fujieda, T. Tamura, S. Kawamoto, T. Takagi, and K. Okubo, "Bodyparts3d: 3d structure database for anatomical concepts," *Nucleic acids research*, vol. 37, no. suppl_1, pp. D782–D785, 2009.

[41] B. O. Community, *Blender - a 3D modelling and rendering package*, Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. [Online]. Available: http://www.blender.org

[42] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, 1999.

[43] P. Kinahan, M. Muzi, B. Bialecki, B. Herman, and L. Coombs, "Data from the acrin 6668 trial nsclc-fdg-pet," 2019.

[44] A. Sekuboyina, M. E. Husseini, A. Bayat, M. Löffler, H. Liebl, H. Li, G. Tetteh, J. Kukačka, C. Payer, D. Štern, M. Urschler, M. Chen, D. Cheng, N. Lessmann, Y. Hu, T. Wang, D. Yang, D. Xu, F. Ambellan, T. Amiranashvili, M. Ehlke, H. Lamecker, S. Lehnert, M. Lirio, N. P. de Olaguer, H. Ramm, M. Sahu, A. Tack, S. Zachow, T. Jiang, X. Ma, C. Angerman, X. Wang, K. Brown, A. Kirszenberg, É. Puybareau, D. Chen, Y. Bai, B. H. Rapazzo, T. Yeah, A. Zhang, S. Xu, F. Hou, Z. He, C. Zeng, Z. Xiangshang, X. Liming, T. J. Netherton, R. P. Mumme, L. E. Court, Z. Huang, C. He, L.-W. Wang, S. H. Ling, L. D. Huynh, N. Boutry, R. Jakubicek, J. Chmelik, S. Mulay, M. Sivaprakasam, J. C. Paetzold, S. Shit, I. Ezhov, B. Wiestler, B. Glocker, A. Valentinitsch, M. Rempfler, B. H. Menze, and J. S. Kirschke, "VerSe: A vertebrae labelling and segmentation benchmark for multi-detector CT images," *Medical Image Analysis*, vol. 73, p. 102166, Oct. 2021. [Online]. Available: https://doi.org/10.1016/j.media.2021.102166

[45] B. Connolly and C. Wang, "Segmented ct pelvis scans with annotated anatomical landmarks," 2019. [Online]. Available: https://datahub.aida.scilifelab.se/10.23698/aida/ctpel

[46] J. Ma, Y. Zhang, S. Gu, C. Zhu, C. Ge, Y. Zhang, X. An, C. Wang, Q. Wang, X. Liu, S. Cao, Q. Zhang, S. Liu, Y. Wang, Y. Li, J. He, and X. Yang, "Abdomenct-1k: Is abdominal organ segmentation a solved problem?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.

[47] Hongsheng Li and M. Chen, "Automatic structure segmentation for radiotherapy planning challenge 2020," 2020. [Online]. Available: https://zenodo.org/record/3718884