

A Multilinear Model for Bidirectional Craniofacial Reconstruction

Jascha Achenbach¹, Robert Brylka², Thomas Gietzen², Katja zum Hebel^{3a},
Elmar Schömer⁴, Ralf Schulze^{3b}, Mario Botsch¹, Ulrich Schwannecke²

¹Bielefeld University, Computer Graphics & Geometry Processing Group

²RheinMain University of Applied Sciences, Wiesbaden, Rüsselsheim, Computer Vision and Mixed Reality Group

³University Medical Center of the Johannes Gutenberg University Mainz, ^aDepartment of Prosthetic Dentistry, ^bSection of Oral Radiology

⁴Johannes Gutenberg University Mainz, Institute of Computer Science

Abstract

We present a bidirectional facial reconstruction method for estimating the skull given a scan of the skin surface and vice versa estimating the skin surface given the skull. Our approach is based on a multilinear model that describes the correlation between the skull and the facial soft tissue thickness (FSTT) on the one hand and the head/face surface geometry on the other hand. Training this model requires to densely sample the Cartesian product space of skull shape times FSTT variation, which cannot be obtained by measurements alone. We generate this data by enriching measured data—volumetric computed tomography scans and 3D surface scans of the head—by simulating statistically plausible FSTT variations. We demonstrate the versatility of our novel multilinear model by estimating faces from given skulls as well as skulls from given faces within just a couple of seconds. To foster further research in this direction, we will make our multilinear model publicly available.

1. Introduction

The face constitutes a rather unique characteristic of our visual appearance and our identity. Its shape is mainly determined by the geometry of the underlying skull and the distribution of facial soft tissue on top of the bony structure. A better understanding—and an efficient computation model—of the relation of facial skin (*head*), the underlying bony structures (*skull*), and the facial soft tissue thickness (*FSTT*) will bring forward a wide range of applications.

For instance, an important topic in forensic medicine and anthropology is facial reconstruction from skeleton remains: By providing a human skull and several FSTT options, the goal is to reconstruct plausible facial appearances to enable recognition of the unknown subject. The other way around, i.e., deriving the skull from the face, also has high-potential applications. In a medical context, this technology can estimate the skull of a person based on a 3D face-scan only—without the need for X-radiation or other expensive medical imaging methods. A reasonably accurate radiation-free alternative will be beneficial, e.g., for patients with craniofacial malformations, where Computed Tomography (CT) is the standard imaging procedure [CHP03]. Another application is radiation-free bony cephalometric skull assessment in orthodontics, where often both the skull and face shape are of interest and a high radiation dose is prohibitive due to the typically young age of the patients [Eur04].

While there are several approaches for facial reconstruction based on skull remains, we are not aware of any work that reconstructs accurate skull geometry from 3D face scans. Both prob-

lems are challenging and have to be regularized by statistical priors from medical imaging data. However, building a dense and accurate model of the correlation between skull, FSTT, and facial skin requires training data that sufficiently samples the *Cartesian product space* of skull shape times FSTT variation. Even with a large number of CT scans this is intractable, since it would require measurements of the same individual at several tissue thickness states.

In this paper, we present such a combined statistical model and make it publicly available for research purposes. We employ a multilinear model that maps from skull shape and FSTT—both represented in low-dimensional parameter spaces—to high-resolution triangle meshes of the skull and the head/facial skin. Varying just the skull parameters generates geometries of different individuals all sharing the same FSTT. Varying the FSTT parameters allows simulating weight changes of a particular individual. Thanks to its multilinear nature our model can be evaluated as well as fitted in just a couple of seconds, allowing us to produce skull and skin variations from given skull shape and FSTT parameters, or to determine these parameters by fitting the multilinear model to a given skull or skin, measured, e.g., by medical imaging or a face scanner.

In order to train the multilinear model we build on our previous work [GBA*18], where from a set of volumetric CT scans and 3D surface scans of heads/faces we constructed three individual parametric models of skull shapes, head shapes, and FSTTs, respectively, and thereby de-coupled these three models (Section 3). In this paper, we first improve the FSTT representation and fitting from [GBA*18] by representing the FSTT as a sphere-mesh offset

from the skull (Section 4.1). This allows us to generate high-quality training data by computing face/head meshes from the Cartesian product of variations of skull geometries and variations of FSTTs, thereby effectively re-coupling the previously decoupled parametric models (Section 5.1). The resulting dense sampling of the product space of skull and FSTT variations enables the construction of a multilinear model (Section 5.2), which we can then fit to either given skull scans or face scans in a unified manner (Section 5.3). We show the versatility of our novel multilinear model and evaluate its reconstruction accuracy by estimating faces from given skulls as well as skulls from given faces (Section 6).

2. Related Work

Skull-based facial reconstruction traditionally relies on artists, who manually model the unknown face using moldable substances and FSTT statistics defined by a few predefined landmarks. Such methods have the inherent drawback that they are very time-consuming, highly subjective and considered to be unreliable [Wil10]. To get more reliable and statistically plausible facial estimates, various computer-aided facial reconstruction methods have been proposed [KHS03, TBK*05, TBL*07, RME*14, SZD*16, SZM*17, GBA*18]. All of these methods are based on FSTT statistics generated with the help of modern 3D imaging techniques such as CT, Magnetic Resonance Imaging (MRI), or Cone Beam Computed Tomography (CBCT). All these technologies enable high accuracy of data acquisition [HCH*15] and provide a solid basis for the collection of FSTT statistics.

By using modern 3D imaging techniques, FSTT statistics are no longer limited to a few predefined landmarks, but can be generated as a dense soft tissue map covering the entire facial area. Nevertheless, there are many approaches to facial reconstruction, such as [KHS03] or [RME*14], that use the FSTT in a few predefined landmarks only. The missing data are either interpolated between the landmarks or anatomical rules such as the location of the anatomical planes and anatomical regressions regarding the shape of the ears, nose, or mouth are used. Both methods result in rather poor reconstruction results and need a lot of manual work. Turner et al. [TBK*05] proposed a template fitting approach for automated skull registration and Tu et al. [TBL*07] applied it to a large CT database consisting of 280 individuals. Instead of creating and using FSTT statistics, they rely on the geometry of skulls and skin surfaces from the CT itself. Using principal component analysis (PCA) they construct a *face-space*, that allows to create a set of plausible reconstructions of a sought face. Shui et al. [SZD*16, SZM*17] use a dense soft tissue map derived from CT data. Given this dense soft tissue map, they fit a statistical head model to the data and thereby provide an estimate of the visual appearance of the person to be reconstructed. Their method requires some manual steps, such as pre-selecting landmarks on the skull, in order to guide the coarse registration of the template, or the reconstruction of mouth, eyes, and nose, which must be performed by a forensic expert.

Recently, in [GBA*18] we proposed a method for forensic facial reconstruction using a parametric skull model, a parametric FSTT model based on dense FSTT measurements, and a parametric head model. This approach is fully automatic and allows the generation

of different plausible head variants utilizing the parametric FSTT model. The current paper improves upon this previous work by reconstructing not only faces from skulls, but also skulls from faces, and by being computationally much more efficient thanks to the proposed multilinear model of skull, FSTT, and head.

The other way around, i.e., reconstructing a skull from skin surface data has a wide range of applications, especially in medicine, but is still relatively unexplored. The common techniques for reconstructing skulls with high precision are CT and MRI. To the best of our knowledge there is currently no method that allows the skull structure to be accurately estimated from a face scan alone. A method for reconstructing a coarse approximation of the skull based on the correlation between skin surface, FSTT at few landmarks, and skull was presented in [BB14]. The authors estimate the rigid head transformation in a facial performance by fitting a simplified skull model to the animated face model. However, their skull model is too simplified to be utilized for medical purposes.

Ali-Hamadi et al. [AHLG*13] presented a semi-automatic method for transferring a volumetric anatomical template model (consisting of bones, muscles and viscera) to any target character. To map the internal anatomy into the target character they manually estimate the fat distribution and warp the template by Laplacian deformation, while satisfying additional constraints, e.g., that bones must stay straight and symmetric across the sagittal plane. Even if the reconstructed interior follows anatomical rules and gives visual pleasing results, the focus of this approach is to transfer the model to all kinds of targets, like animals or cartoon characters, and not on precisely reconstructing the inner of a human body.

In [KIL*16] a fully automated approach for reconstructing physics-based anatomical models based on a tetrahedral template mesh representing an average male was presented. To fit the target as closely as possible, the template model is warped through symmetric as-rigid-as-possible deformation. The work focuses on the reconstruction of large and medium anatomical details, leaving out parts like hands, toes, and the face, which is the main component of our current work. Another approach, presented by Ichim et al. [IKNDP16] builds a volumetric face rig based on thickness measurements from forensic studies and employs it for physics-based animation. In [IKKP17] this approach was extended to include a novel muscle activation model that separates active and passive soft tissue layers. Again, all the above approaches are based on skull models that are too simplified to be used for medical purposes.

3. Fundamental Models and Data

In order to generate training data for our multilinear model, i.e., pairs of skin surfaces and corresponding skulls, we follow the approach presented in [GBA*18]. This section gives a brief overview of the models and methods used to generate the training data (see [GBA*18] for more details).

Our aim is to generate three parametric models from CT data describing skull, head, and FSTT variants. To build the parametric models, existing head CT scans (from four different devices) were collected from the DICOM database of the PACS system of the University Medical Center Mainz. The local ethical approval

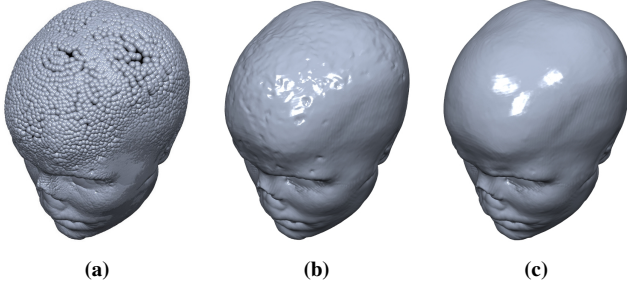


Figure 1: Adding facial tissue, specified by FSTT distribution, onto a given skull: (a) Union-of-Spheres from [GBA*18], (b) sphere-mesh based on unoptimized radii with dent-like artifacts, (c) smooth sphere-mesh based on optimized radii.

board [Eth] has approved the processing of the pseudonymized existing CTs to generate the statistical models under the approval number No 837.244.15 (10012) (date: 05.08.2015).

To generate the different parametric models, we only used CT scans that meet the criteria that a) the facial skull of the patient is *completely imaged*, b) the *slice thickness* is less than or equal to 1 mm and c) the subject has neither significant oral or maxillofacial deformations nor missing parts. Three different types of data were used to generate the proposed multilinear model. 1) 56 skulls extracted from CT scans to generate a parametric skull model, 2) 78 heads (39 extracted skin surfaces from CT scans and 39 heads from 3D surface scans) to generate a parametric head model, and 3) 39 corresponding skin surfaces and skulls extracted from CT scans to generate a parametric FSTT model. In addition, 5 pairs of skin surfaces and skulls from CT scans are used for evaluation and are therefore not included in building the parametric models.

From the collected CT data, we automatically extracted cleaned polygonal meshes of the skin surfaces and skulls. This automatic step includes cropping each CT to a volume of interest, extracting the skin surface and skull using Marching Cubes [LC87] and removing unwanted parts such as soft tissue structure (respiratory tract, ear canal) in case of the skin surface and the spine or internal bone structures in case of the skull. The 39 skin surfaces extracted from CT were finally supplemented by 39 high-quality optical 3D head scans. The resulting 56 skulls and 78 skin surfaces are aligned in a common reference frame by fitting a template skull or template head to each mesh individually. This fitting process is based on a similarity transformation followed by a fine-scale deformation step. By applying PCA to the fitted skulls and the fitted skin surfaces, we generate parametric models for the skull consisting of a tetrahedral mesh with $M = 69122$ vertices and the head consisting of a skin surface mesh with $N = 24574$ vertices.

In the next section, we present a method to obtain an accurate estimate of dense FSTT statistics. We also describe how to create a parametric FSTT model that can be used together with a suitable parametric head model to generate plausible head variants.

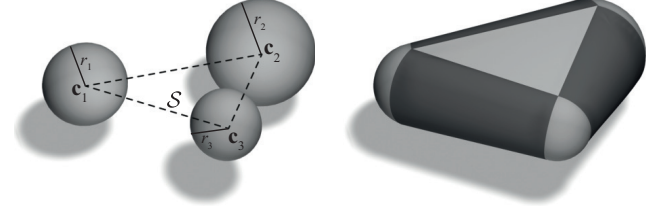


Figure 2: Left: Skeleton triangle S with centers $\mathbf{c}_i$ and radii r_i . Right: The resulting sphere-mesh wedge (image from [TPT16]).

4. Facial Soft Tissue Thickness

The FSTT is defined by a scalar thickness radius r_i for each vertex on the outside of the skull model, i.e., where a meaningful tissue thickness between the skull bone and the skin surface can be determined. The set of these radii is denoted by $\mathcal{R} = (r_1, \dots, r_m)$, where $m \approx 16.5k$ is the number of outer skull vertices (from the overall $M \approx 69k$ skull vertices).

In our previous work [GBA*18] the geometric representation of an FSTT-specified offset from a given skull was constructed as the union of spheres centered at each outer skull vertex $\mathbf{c}_i$ with its corresponding FSTT radius r_i , as shown in Figure 1a. In the following, we replace the discontinuous non-smooth union of spheres by a *sphere-mesh* [TGB13], which leads to a continuous surface envelope around the skull representing the FSTT offset (Figure 1b).

4.1. Sphere-Mesh Representation

Sphere-meshes are a variant of convolution surfaces [BS91] and were originally used for shape approximation [TGB13]. Recently, they have also been employed for hand modeling and tracking [TPT16, TTR*17]. For representing the FSTT-offset from a skull through sphere meshes, we consider all triangles on the outer skull surface where each vertex $\mathbf{c}_i$ has an associated FSTT thickness radius r_i . Each such triangle $(\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3)$ is convolved with a sphere whose spatially varying radius is determined by barycentric interpolation of the FSTT thicknesses r_1, r_2, r_3 , leading to a *triangle wedge* as shown in Figure 2. With $B(\mathbf{x}, \mathbf{c}, r) = \|\mathbf{x} - \mathbf{c}\| - r$ denoting the signed distance from a sphere of radius r centered at $\mathbf{c}$, the triangle wedge is implicitly defined as the zero-set of

$$\min_{\substack{\alpha, \beta, \gamma \geq 0 \\ \alpha + \beta + \gamma = 1}} B(\mathbf{x}, \alpha \mathbf{c}_1 + \beta \mathbf{c}_2 + \gamma \mathbf{c}_3, \alpha r_1 + \beta r_2 + \gamma r_3),$$

where α , β , and γ are the barycentric coordinates.

If we denote the set of all wedges as $\mathcal{T}$, the FSTT, as the union of all wedges, is defined as the zero-set of its signed distance

$$\text{dist}(\mathbf{x}) = \min_{t \in \mathcal{T}} \min_{\substack{\alpha, \beta, \gamma \geq 0 \\ \alpha + \beta + \gamma = 1}} B(\mathbf{x}, \alpha \mathbf{c}_1^t + \beta \mathbf{c}_2^t + \gamma \mathbf{c}_3^t, \alpha r_1^t + \beta r_2^t + \gamma r_3^t). \quad (1)$$

From the above *implicit* representation we can extract an *explicit* triangle mesh through the Marching Cubes algorithm. In our experiments a voxel size of 2 mm turned out to provide a good trade-off between precision and computing time, such that we use this voxel size for all reconstructed sphere-meshes.

The diagram shows a horizontal line with points c_1 , c , and c_2 in order from left to right. A vertical line segment connects point p at the top to point c at the bottom. A point s is located on the segment pc , and a right-angle symbol is shown at s , indicating that ps is perpendicular to the horizontal line. Two circular arcs are centered at c_1 and c_2 with radii r_1 and r_2 respectively. The region between these two arcs is shaded gray and labeled S . The distance between c and c_2 is also labeled S .

© 2018 The Author(s)

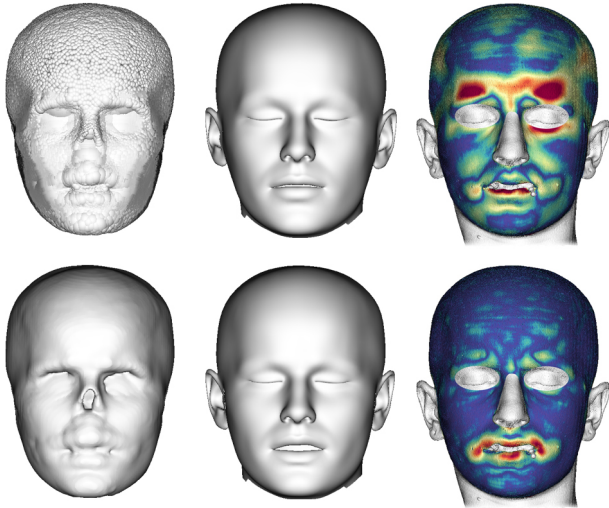


Figure 5: Fitting the head model to an FSTT-offset of a given skull. Representing the offset as union of spheres (top) leads to a higher geometric error as our proposed sphere-mesh representation (bottom). From left to right: FSTT-offset, fitted head model, color-coded distance to true skin (blue: = 0mm, red: ≥ 3 mm).

In comparison, Tkach et al. [TPT16, TTR*17] decompose wedges into triangles, spheres, and cones and solve a nonlinear optimization for fitting a sphere-mesh of 30 nodes to their hand model. In contrast, our approach is fully implicit and seamlessly handles even special cases when radii are larger than skeleton triangles. Our fitting requires simple linear least squares systems only and efficiently and robustly optimizes our $m \approx 16.5$ k FSTT radii.

As shown in Figure 1c, our optimization successfully removes the artifacts due to CT noise, leading to a smooth FSTT geometry. Based on the techniques presented in this section, we improve all 39 FSTT distributions from our database, from which we compute an improved parametric FSTT PCA-model analogous to [GBA*18].

4.3. Fitting a Head Model

In order to reconstruct a 3D face from both a given skull and a given FSTT distribution, we fit our parametric head model to the FSTT-specified geometric offset from the outer skull surface. To this end, the template head model is first coarsely aligned through a similarity transform and by optimizing parameters of the head-PCA, followed by a fine-scale non-rigid deformation. To guarantee plausible reconstructions, the fitting process is regularized by penalizing large PCA weights as well as strong bending [GBA*18].

In [GBA*18] the FSTT-offset was represented as a union of spheres, leading to a discontinuous non-smooth surface. Fitting the head model to this type of target geometry requires a rather strong bending regularization and even a dedicated regularization term for down-weighting correspondences in the interior part of the FSTT (see Figure 5, top row for an exemplary fitting result). In contrast, our proposed sphere-mesh representation with optimized FSTT radii provides a smoother and more accurate FSTT offset. We

convert the implicit sphere-mesh to an explicit triangle mesh using Marching Cubes (grid spacing of 2 mm) and point-sample the triangle mesh to compute correspondences. The higher surface smoothness and FSTT accuracy allow for less regularization and therefore result in more precise fits (see Figure 5, bottom row). A quantitative evaluation yields an average RMS fitting error of 0.51 mm for the optimized sphere-meshes, compared to an average RMS fitting error of 0.82 mm for the union of spheres, which is an improvement of 37% in fitting accuracy.

5. Multilinear Model

Our goal is to develop a model that (i) maps from skull shape and FSTT distribution—both controlled by low-dimensional parameter vectors $\mathbf{w}_{\text{skull}}$ and $\mathbf{w}_{\text{fstt}}$ —to a 3D head/skin surface, and (ii) can also invert this map to infer skull and FSTT from a given face scan.

Our parametric PCA-models for skull shape (Section 3) and FSTT (Sections 3 and 4.2) can map skull and FSTT parameters to specific skull and FSTT instances. Adding the FSTT onto the skull through the sphere-mesh representation (Section 4.1) and fitting the head model to it (Section 4.3) eventually implements the forward mapping. However, this multi-step approach requires about 90s, which is prohibitive for interactive applications, and it cannot easily be inverted. Inspired by previous approaches that have successfully applied multilinear models in the context of faces, using separate parameter sets for person identity and facial expressions [VBPP05, BW13, CWZ*14], we in the following generate a multilinear model that can efficiently and robustly compute the head surface from skull and FSTT parameters and vice versa.

5.1. Generating Training Data

Multilinear models have to be trained on the full Cartesian product of their independent parameter sets. In our context, this means generating a set of skull shapes and a set of FSTT distributions, and providing as training data each skull shape equipped with each FSTT variation (input) and the respective head surface (output).

It obviously is not possible to collect such data from measurements alone, as it would require multiple CT scans of the same person under different, controlled body weight variations. Our CT scans include different skulls with different FSTT distributions, but not their dense Cartesian product. By building independent PCA models for skull shape and FSTT distributions, we effectively decoupled these two components, allowing us to re-couple them by generating synthetic head models for the Cartesian product of skull shape variation times FSTT variation as outlined above, thus obtaining statistically and anatomically plausible training data.

As a trade-off between computational effort and coverage of input data, we sample skull variations along six PCA-dimensions and FSTT variations along five PCA-dimensions, which covers more than 70% of the variation included in our CT data. For each principal component we sample two offsets at ± 2 standard deviations along that component. Furthermore, we include the mean skull with mean FSTT from our parametric models, which in total yields $2^{5+6} + 1 = 2049$ pairs of skulls and FSTT distributions. For each of these pairs we compute the sphere-mesh offset and fit the head

model as described before, which takes about 90 s for each model on a desktop PC with 4×3.6 GHz Intel Xeon. We also experimented with sampling more PCA dimensions to cover more than 75 % of the variation in our CT data, but this did not lead to significant improvements in fitting accuracy and did not justify the increased computation effort for a more complex multilinear model.

5.2. Generating the Multilinear Model

Our training data consist of 2048 pairs of skull mesh ($M = 69122$ vertices) and skin mesh ($N = 24574$ vertices), which we each stack into a column vector $\mathbf{X}_i \in \mathbb{R}^{d_{\text{vert}}}$ with $d_{\text{vert}} = 3N + 3M$. These pairs are obtained as $d_{\text{skull}} = 64$ skull variants, each containing $d_{\text{fstd}} = 32$ FSTD distributions. Following [BW13], we center each $\mathbf{X}_i$ by subtracting the model constructed from the mean skull with the mean FSTD from our parametric models, denoted by $\bar{\mathbf{X}}$.

To construct the multilinear model (MLM in the following), we arrange the 2048 mean-centered geometry vectors $\mathbf{X}_i$ into a three-dimensional array $\mathcal{D} \in \mathbb{R}^{d_{\text{vert}} \times d_{\text{skull}} \times d_{\text{fstd}}}$, which is formally called a third order (3-mode) *data tensor*. This way, the three *mode spaces* of $\mathcal{D}$ are associated with skin/skull vertex geometry, skull variations, and FSTD variations. This data tensor $\mathcal{D}$ is then decomposed by *higher-order singular value decomposition* [DL97] as

$$\mathcal{D} = \mathcal{M} \times_{\text{skull}} \mathbf{U}_{\text{skull}} \times_{\text{fstd}} \mathbf{U}_{\text{fstd}}, \quad (3)$$

where

$$\mathcal{M} = \mathcal{D} \times_{\text{skull}} \mathbf{U}_{\text{skull}}^T \times_{\text{fstd}} \mathbf{U}_{\text{fstd}}^T \quad (4)$$

is a *multilinear model tensor* $\mathcal{M} \in \mathbb{R}^{d_{\text{vert}} \times d_{\text{skull}} \times d_{\text{fstd}}}$. $\mathbf{U}_{\text{skull}} \in \mathbb{R}^{d_{\text{skull}} \times d_{\text{skull}}}$ and $\mathbf{U}_{\text{fstd}} \in \mathbb{R}^{d_{\text{fstd}} \times d_{\text{fstd}}}$ are orthogonal matrices containing the left singular vectors of the corresponding mode spaces. If we choose n to be either ‘skull’ or ‘fstd’, the matrix $\mathbf{U}_n$ is constructed as follows (please see [BW13, VBPP05] for more details): We first unfold $\mathcal{D}$ along the n -th mode to a matrix $\mathbf{T}_n$ by stacking as columns all vectors of $\mathcal{D}$ aligned with the n -th mode. Then the matrix $\mathbf{U}_n \in \mathbb{R}^{d_n \times d_n}$ can be computed via standard matrix SVD as $\mathbf{T}_n = \mathbf{U}_n \mathbf{S}_n \mathbf{V}_n^T$. For instance, unfolding $\mathcal{D}$ along the skull-mode leads to a matrix $\mathbf{T}_{\text{skull}} \in \mathbb{R}^{d_{\text{skull}} \times (d_{\text{vert}} \cdot d_{\text{fstd}})}$. Given $\mathcal{D}$ and $\mathbf{U}_n$, the n -th *mode product* $\mathcal{D} \times_n \mathbf{U}_n^T$ acts on each vector $\mathbf{v} \in \mathbb{R}^{d_n}$ in $\mathcal{D}$ ’s mode- n space via the linear transformation $\mathbf{v} \mapsto \mathbf{U}_n^T \mathbf{v}$.

Finally, given skull parameters $\mathbf{w}_{\text{skull}} \in \mathbb{R}^{d_{\text{skull}}}$ and FSTD parameters $\mathbf{w}_{\text{fstd}} \in \mathbb{R}^{d_{\text{fstd}}}$, the MLM computes the corresponding combined skin/skull mesh $\mathbf{X} \in \mathbb{R}^{d_{\text{vert}}}$ by tensor contraction as

$$\mathbf{X}(\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}}) = \bar{\mathbf{X}} + \mathcal{M} \times_{\text{skull}} \mathbf{w}_{\text{skull}}^T \times_{\text{fstd}} \mathbf{w}_{\text{fstd}}^T. \quad (5)$$

This evaluation takes less than a second, making the MLM well suited for interactive applications, for instance in order to explore FSTD variations for a given skull in a forensic context.

5.3. Multilinear Model Fitting

The MLM maps skull parameters $\mathbf{w}_{\text{skull}}$ and FSTD parameters $\mathbf{w}_{\text{fstd}}$ to a geometry $\mathbf{X}(\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}})$, which includes both the N head vertices and the M skull vertices. Inverting this process means determining the parameters $\mathbf{w}_{\text{skull}}$ and $\mathbf{w}_{\text{fstd}}$ such that the corresponding

model $\mathbf{X}(\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}})$ closely matches a given geometry observation, which could for instance be a face scan or a skull scan extracted from CT. The inverse process therefore amounts to nonrigid registration (or fitting) of the MLM to a given point cloud $\mathcal{P}$.

This fitting procedure requires a coarse initial alignment, which can be performed manually (by selecting landmarks) or computed automatically, depending on the type of scanner data available [GBA*18, AZB15]. Following [BW13] we initialize the MLM as the mean shape $\bar{\mathbf{X}} = \mathbf{X}(\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}})$. To speed up the fitting process, we uniformly sub-sample the scanner data $\mathcal{P}$ to approximately 100k points, without noticeably sacrificing geometric fidelity.

After this initialization, we alternately compute closest point correspondences $\mathcal{C}$ between the given point cloud $\mathcal{P}$ and the current state $\mathbf{X}(\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}})$ and optimize the model parameters. We again prune correspondences if their distance is too high or their normal vectors deviate too much. Furthermore, we prune correspondences for error-prone areas like the teeth, the inner part of the skull, hair, ears, or eye regions, which we have pre-selected on the template.

Given a set of correspondences $(\mathbf{p}_i, \mathbf{x}_i) \in \mathcal{C}$, we minimize their squared distances by optimizing for similarity transform (scaling s , rotation $\mathbf{R}$, translation $\mathbf{t}$) and model parameters $\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}}$:

$$E_{\text{fit}}(\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}}, s, \mathbf{R}, \mathbf{t}) = \frac{1}{|\mathcal{C}|} \sum_{i \in \mathcal{C}} \|s \mathbf{R} \mathbf{x}_i(\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}}) + \mathbf{t} - \mathbf{p}_i\|^2, \quad (6)$$

where $\mathbf{p}_i \in \mathcal{P}$ is a scanner point and $\mathbf{x}_i(\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}})$ its closest point on the current state $\mathbf{X}(\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}})$, which is typically located within a triangle and expressed through barycentric coordinates. To prevent over-fitting, we add a Tikhonov regularization term

$$E_{\text{reg}}(\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}}) = \frac{1}{d_{\text{skull}}} \sum_{k=1}^{d_{\text{skull}}} \left(\frac{w_{id,k} - \bar{w}_{id,k}}{\sigma_{id,k}} \right)^2 + \frac{1}{d_{\text{fstd}}} \sum_{l=1}^{d_{\text{fstd}}} \left(\frac{w_{fstt,l} - \bar{w}_{fstt,l}}{\sigma_{fstt,l}} \right)^2, \quad (7)$$

with $\sigma_{\text{skull},k}^2$ and $\sigma_{\text{fstd},l}^2$ being the variance of the principal components computed from the covariance matrices after unfolding $\mathcal{D}$ along the respective modes [BW16]. Similar to [VBPP05], we then minimize the combined objective function

$$E(\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}}, s, \mathbf{R}, \mathbf{t}) = E_{\text{fit}}(\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}}, s, \mathbf{R}, \mathbf{t}) + E_{\text{reg}}(\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}}) \quad (8)$$

using block-coordinate descent, i.e., we alternately solve for either MLM parameters $\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstd}}$ or pose parameters $s, \mathbf{R}, \mathbf{t}$, while fixing the respective other parameters. This energy minimization is alternated with the computation of new correspondences and iterated until convergence, which typically takes 3–5 iterations and requires about 30 s on average.

The result of the fitting process are model parameters $\mathbf{w}_{\text{skull}}$ and $\mathbf{w}_{\text{fstd}}$, which through Eq. (5) can be evaluated to a skin mesh and a skull mesh that closely matches the scanner point cloud $\mathcal{P}$.

6. Evaluation

We evaluate our method on 5 different pairs of skulls and corresponding skin surfaces extracted from CT scans that are not included in our training data introduced in Section 3. We present results for fitting our MLM to scanner data in order to either infer

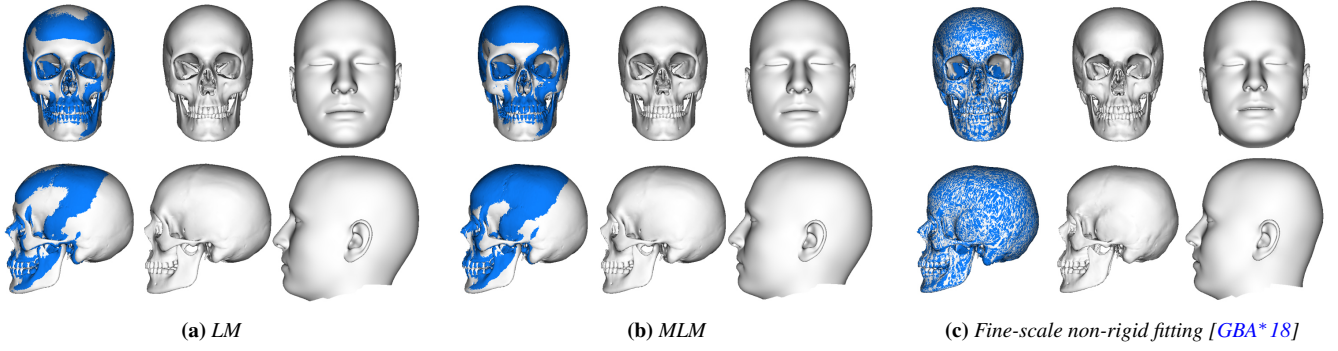


Figure 6: Comparison of (a) LM and (b) MLM fitting and reconstruction results, (c) Automatic forensic facial reconstruction approach presented in [GBA*18] with fine-scale non-rigid fitting result and facial reconstruction based on mean FSTT for one subject of our evaluation data set. Each from left to right: Skull fit (white) with skull extracted from CT (blue) as overlay, skull fit and skin reconstruction.

skin surface from skull or vice versa. We compare our MLM with two different approaches: (1) a linear model (LM) created through PCA of the 2049 combined skin/skull pairs $\mathbf{X}_i$ and (2) the forensic facial reconstruction approach presented in [GBA*18]. Note that due to privacy reasons the extracted or reconstructed skin surface can only be shown for one single subject.

Generating and Fitting a Linear Model. Analogous to the generation of the MLM, we use the 2049 pairs of skin/skull mesh from our synthetic training data to generate the LM. The vertices of each skin mesh and corresponding skull mesh are again stacked into a column vector $\mathbf{X}_i \in \mathbb{R}^{d_{\text{vert}}}$. After subtracting the mean $\bar{\mathbf{X}} \in \mathbb{R}^{d_{\text{vert}}}$ over all training data from each of the $\mathbf{X}_i$ we arranged the resulting mean-centered geometry vectors into a $d_{\text{vert}} \times 2049$ -dimensional matrix. PCA of this matrix gives $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_d]$ consisting of the first d principal components. To obtain the same number of degrees of freedom as for the MLM, we chose $d = d_{\text{skull}} + d_{\text{fstt}} = 96$. Given a weight vector $\mathbf{w} \in \mathbb{R}^d$, the LM allows to generate a combined skin surface and skull mesh as

$$\mathbf{X}(\mathbf{w}) = \bar{\mathbf{X}} + \mathbf{U}\mathbf{w}. \quad (9)$$

Fitting the LM to a given face/skull geometry is very similar to fitting the MLM. Again, we distinguish between fitting to scanner data of skin/head or skull, which basically differs in the way of how correspondences are computed. Given an initial alignment, we perform a non-rigid registration to estimate the weights $\mathbf{w}$ by minimizing a Tikhonov regularized linear least-squares problem.

6.1. Inferring Skin Surface from Skull

To analyze our skin reconstruction process, we fit both the LM and the MLM to the skulls extracted from our evaluation data set. Figure 6 shows skull fitting and skin surface reconstruction results for one specific subject based on the LM (Figure 6a) or the MLM (Figure 6b), respectively. The resulting skin reconstruction of the LM is an arbitrary skin surface, related to the underlying PCA space and by no means a reconstruction based on the mean FSTT distribution. It is comparable to the MLM if $\mathbf{w}_{\text{fstt}}$ is not adjusted. Because both models are built on the same training data, both reconstructions are

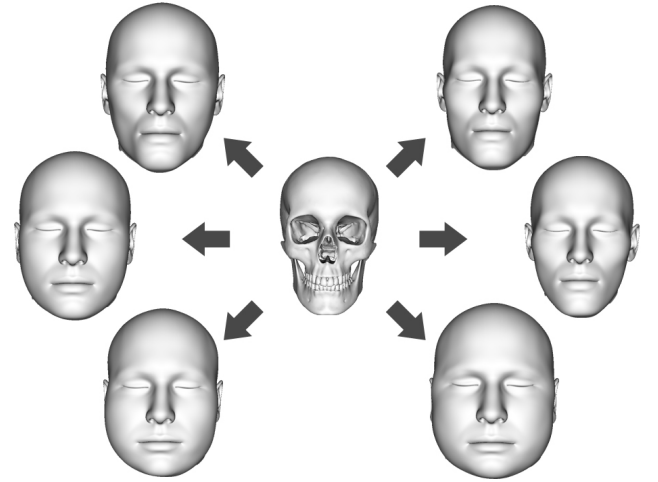


Figure 7: Multilinear model fitting: Skin surface variants given a skull. Skin variants can be simply generated by fixing skull parameters $\mathbf{w}_{\text{skull}}$ and varying FSTT parameters $\mathbf{w}_{\text{fstt}}$.

visually very similar. Moreover, fitting the MLM takes about 28 s, while fitting the LM takes 10 s on average.

Figure 6c shows a fitted skull and a skin reconstruction based on the approach presented in [GBA*18]. Since our skull fittings in Figures 6a and 6b are constrained by the LM and the MLM, respectively, the result is less accurate compared the fine-scale non-rigid registration of [GBA*18]. The RMS error based on the skull evaluation mask (see Figure 9) results in 0.34 mm [GBA*18] vs. 1.13 mm for the MLM. However, the resulting skin estimations of both approaches are visually very similar.

Fitting based on the LM has the inherent drawback that there is no control over the FSTT distribution, which results in a single non-changeable skin surface reconstruction. The benefits of the MLM come into play when reconstructing skin surface variants for a specific skull, since the MLM allows to generate different head variants by varying $\mathbf{w}_{\text{fstt}}$. Figure 7 shows different head surface variants generated by manipulating the FSTT for a given, fixed skull.

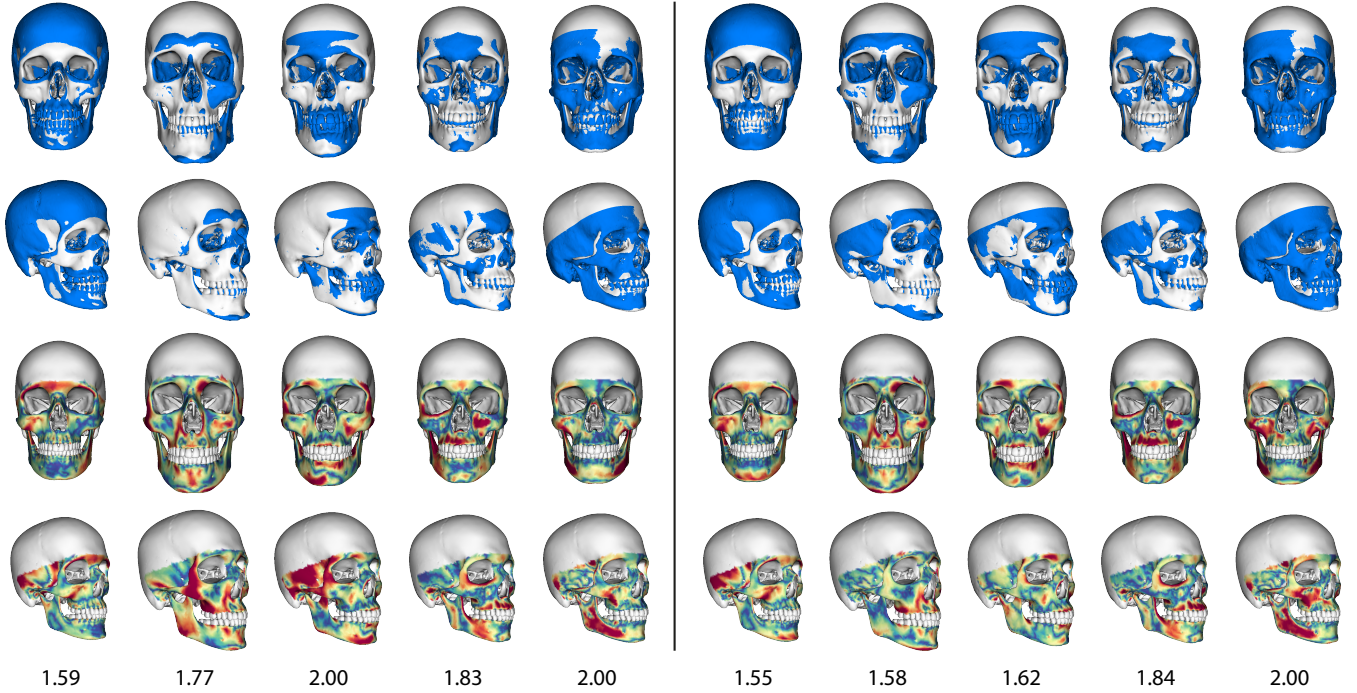


Figure 8: Skull reconstructions given skin surface scans from our 5 evaluation data sets. Reconstruction results for the linear model (left) and for the multilinear model (right). For each model from top to bottom: Skulls extracted from CT (blue) and our skull reconstructions (white) as overlay, the minimal distance to the actual skull (blue: = 0mm, red: ≥ 3 mm), and the RMS errors in mm.

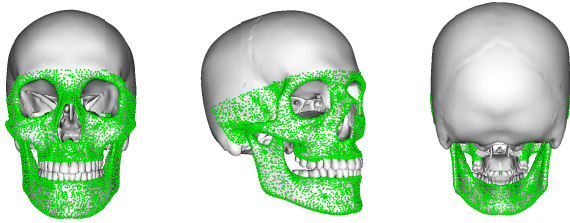


Figure 9: Evaluation mask for skull reconstructions (green).

The presented MLM allows to generate skin variants nearly in real-time, only at the cost of evaluating Eq. (5). In contrast, the skin reconstruction process presented in [GBA*18] is based on several time consuming steps resulting in a computing time of about 2 min.

6.2. Inferring Skull Shape from Face Scan

To analyze the accuracy of our skull reconstruction process, we fitted the MLM and the LM to the extracted skin surfaces from the evaluation data sets. For privacy reasons, we can only show skull reconstructions but not the skin surface fittings. For the evaluation we create a point mask, which is limited to the facial area of the skull. Since our CT data set for creating the FSTT statistics is partially incomplete for the upper part of the skull, we additionally restrict the evaluation mask to the smallest available calvaria part and also exclude teeth. The final evaluation mask used is shown in Figure 9, with points of interest colored green.

Distance is measured from each point of interest on a reconstructed skull to the surface of the corresponding extracted skull. The average RMS fitting error over all 5 reconstructed skulls is 1.72 mm using the MLM and 1.85 mm using the LM. As can be seen clearly in Figure 8, both models not only allow to reconstruct the correct size of the skull, but also correctly reproduce the shape of the skull, in particular the emplacement of the mandibular. For both models the fitting error is below 2 mm. While reconstructing skull from given skin surfaces using the MLM gives slightly better results, it takes about 30 s compared to 8 s using the LM.

6.3. Simulating Weight Changes for Face Scans

Fitting the MLM to a face scan does not only reveal the skull parameters $\mathbf{w}_{\text{skull}}$, but also the FSTT parameters $\mathbf{w}_{\text{fstt}}$ of the scanned individual. Given the skull shape and FSTT distribution of the person, we can simulate weight changes by varying $\mathbf{w}_{\text{fstt}}$.

Since the MLM does not reconstruct hairs or eyes, we start to obtain a realistic head reconstruction by fitting a head template to a photogrammetric face scan using the method of [AWLB17]. This head template has the same triangulation as the template used in Section 4.3, but has open eyes, eyeballs, and teeth. Furthermore, the nonlinear fine-scale deformation of [AWLB17] allows a reasonable reconstruction of hair geometry (Figure 10, top row).

Since both models (the realistic face model and the MLM) were fitted to the same scanner data, they are well aligned to each other.



Figure 10: The multilinear model makes it possible to vary FSTT for a specific individual. Three scanned persons reconstructed by [AWLB17] (first row) with varied FSTT (second and third row).

When changing the FSTT of the scanned person from $\mathbf{w}_{\text{fstt}}$ to $\tilde{\mathbf{w}}_{\text{fstt}}$, we can therefore simply transfer the per-vertex displacement

$$\mathbf{X}(\mathbf{w}_{\text{skull}}, \tilde{\mathbf{w}}_{\text{fstt}}) - \mathbf{X}(\mathbf{w}_{\text{skull}}, \mathbf{w}_{\text{fstt}})$$

computed by the MLM onto the realistic face model, thereby obtaining the thinner or thicker models shown in Figure 10. Finally, the positions of the eyeballs are adjusted to accommodate the slight displacements in the eye lids due to the simulated weight changes.

7. Conclusions

We presented a multilinear model that maps a set of low-dimensional parameters for skull shape and FSTT distribution to an accurate and high-quality mesh of both the skin and the skull geometry. To foster further research in this direction, we make our multilinear model publicly available for research purposes at doi:10.4119/unibi/2930619.

The required training data, a dense Cartesian product of skull and FSTT variations, cannot be obtained by measurements alone. Based on individual parametric models of skull shape, FSTT distributions, and head shapes, we generate synthetic—but statistically plausible—training data by computing head models for given skull shapes and FSTT distributions. Representing the FSTT-offset by a sphere-mesh improved the quality of the resulting head models.

We demonstrated our model to have several interesting and high-potential applications in the medical context. First, it allows to simulate plausible head shapes for given skull and FSTT variations at interactive rates, which is an important component in explorative computer-aided forensics. Second, by fitting the multilinear model to face scans, we can infer both the skull shape and the FSTT of the scanned individual and successively simulate weight changes. More importantly, the estimation of a reasonably accurate skull shape from a photogrammetric face scan helps to reduce or even avoid X-radiation for bony cephalometric skull assessments in orthodontics or for surgical planning for patients with craniofacial malformations. To the best of our knowledge, there is no approach to infer the skull shape from a skin surface scan at our precision.

One limitation of our model is that there is no guarantee that eyes and mouth stay at their anatomically correct positions relative to the skull when varying the FSTT parameters. The slight movement of eyes and mouth can be avoided in the future by incorporating corresponding constraints into the training data generation. Furthermore, since real faces and FSTTs are asymmetric to a certain extent, our learned FSTT component inherits this asymmetry, which can result in asymmetric head shapes when varying FSTT parameters.

Besides improving the theoretical properties of our model, collecting more CT scans of a larger variation of real people is needed to increase the variability and expressiveness of our model. Access to more training data would also allow for exploring different learning algorithms, or to investigate other approaches for filling up the sparse scanned data to achieve a dense set of training data.

8. Acknowledgments

The work was supported by the research project “Kephalos” funded by the German Federal Ministry of Education and Research (BMBF) and the Cluster of Excellence Cognitive Interaction Technology “CITEC” (EXC 277) at Bielefeld University, funded by the German Research Foundation (DFG). We gratefully acknowledge the Department of Diagnostic and Interventional Radiology, University Medical Center of the Johannes Gutenberg University Mainz for providing us with the DICOM data of the human heads.

References

- [AHLG*13] ALI-HAMADI D., LIU T., GILLES B., KAVAN L., FAURE F., PALOMBI O., CANI M.-P.: Anatomy Transfer. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 32, 6 (2013), 188:1–188:8. 2
- [AWLB17] ACHENBACH J., WALTEMAT T., LATOSCHIK M., BOTSCH M.: Fast Generation of Realistic Virtual Humans. In *Proceedings of ACM Symposium on Virtual Reality Software and Technology* (2017), ACM, pp. 1–10. 8, 9
- [AZB15] ACHENBACH J., ZELL E., BOTSCH M.: Accurate Face Reconstruction through Anisotropic Fitting and Eye Correction. In *Proceedings of Vision, Modeling & Visualization* (2015), pp. 1–8. 6
- [BB14] BEELER T., BRADLEY D.: Rigid Stabilization of Facial Expressions. *ACM Trans. Graph.* 33, 4 (2014), 44:1–44:9. 2
- [BKP*10] BOTSCH M., KOBELT L., PAULY M., ALLIEZ P., LEVY B.: “Polygon Mesh Processing”. AK Peters, 2010. 4
- [BS91] BLOOMENTHAL J., SHOEMAKE K.: Convolution Surfaces. *SIGGRAPH Comput. Graph.* 25, 4 (1991), 251–256. 3

- [BW13] BOLKART T., WUHRER S.: Statistical Analysis of 3D Faces in Motion. In *International Conference on 3D Vision - 3DV* (2013), pp. 103–110. 5, 6
- [BW16] BOLKART T., WUHRER S.: A Robust Multilinear Model Learning Framework for 3D Faces. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), pp. 4911–4919. 6
- [CHP03] CARUSO P., HARRIS G., PADWA B.: CT imaging of craniofacial malformations. *Neuroimaging Clinics of North America* 13, 3 (2003), 541–572. 1
- [CWZ*14] CAO C., WENG Y., ZHOU S., TONG Y., ZHOU K.: Face-Warehouse: A 3D Facial Expression Database for Visual Computing. *IEEE Transactions on Visualization and Computer Graphics* 20, 3 (2014), 413–425. 5
- [DL97] DE LATHAUWER L.: *Signal Processing Based on Multilinear Algebra*. PhD thesis, Katholieke Universiteit Leuven, 1997. 6
- [Eth] ETHIK-KOMMISSION: Landesärztekammer Rheinland-Pfalz, Deutschhausplatz 2, 55116 Mainz. www.laek-rlp.de/ethik/. Accessed 2018-08-21. 3
- [Eur04] EUROPEAN COMMISSION, DIRECTORATE-GENERAL FOR ENERGY AND TRANSPORT ; DIRECTORATE H - NUCLEAR SAFETY AND SAFEGUARDS: Radiation protection 136 : European guidelines on radiation protection in dental radiology; the safe use of radiographs in dental practice. https://ec.europa.eu/energy/sites/ener/files/documents/136_0.pdf, 2004. Accessed 2018-08-21. 1
- [GBA*18] GIETZEN T., BRYLKA R., ACHENBACH J., ZUM HEBEL K., SCHÖMER E., BOTSCH M., SCHWANECKE U., SCHULZE R.: A method for automatic forensic facial reconstruction based on dense statistics of soft tissue thickness, 2018. [arXiv:1808.07334](https://arxiv.org/abs/1808.07334). 1, 2, 3, 4, 5, 6, 7, 8
- [HCH*15] HWANG H.-S., CHOE S.-Y., HWANG J.-S., MOON D.-N., HOU Y., LEE W.-J., WILKINSON C.: Reproducibility of Facial Soft Tissue Thickness Measurements Using Cone-Beam CT Images According to the Measurement Methods. *Journal of Forensic Sciences* 60, 4 (2015), 957–965. 2
- [IKKP17] ICHIM A.-E., KADLEČEK P., KAVAN L., PAULY M.: Phace: Physics-based Face Modeling and Animation. *ACM Trans. Graph.* 36, 4 (2017), 153:1–153:14. 2
- [IKNDP16] ICHIM A.-E., KAVAN L., NIMIER-DAVID M., PAULY M.: Building and Animating User-Specific Volumetric Face Rigs. In *Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation* (2016), pp. 107–117. 2
- [KHS03] KÄHLER K., HABER J., SEIDEL H.-P.: Reanimating the Dead: Reconstruction of Expressive Faces from Skull Data. *ACM Trans. Graph.* 22, 3 (2003), 554–561. 2
- [KIL*16] KADLEČEK P., ICHIM A.-E., LIU T., KŘIVÁNEK J., KAVAN L.: Reconstructing Personalized Anatomical Models for Physics-based Body Animation. *ACM Trans. Graph.* 35, 6 (2016), 213:1–213:13. 2
- [LC87] LORENSEN W. E., CLINE H. E.: Marching Cubes: A High Resolution 3D Surface Construction Algorithm. *SIGGRAPH Comput. Graph.* 21, 4 (1987), 163–169. 3
- [RME*14] ROMEIRO R., MARROQUIM R., ESPERANÇA C., BRED A., FIGUEREDO C. M.: Forensic Facial Reconstruction Using Mesh Template Deformation with Detail Transfer over HRBF. In *27th SIB-GRAPI Conference on Graphics, Patterns and Images* (2014), pp. 266–273. 2
- [SZD*16] SHUI W., ZHOU M., DENG Q., WU Z., JI Y., LI K., HE T., JIANG H.: Densely Calculated Facial Soft Tissue Thickness for Craniofacial Reconstruction in Chinese Adults. *Forensic Science International* 266 (2016), 573.e1 – 573.e12. 2
- [SZM*17] SHUI W., ZHOU M., MADDOCK S., HE T., WANG X., DENG Q.: A PCA-Based Method for Determining Craniofacial Relationship and Sexual Dimorphism of Facial Shapes. *Computers in Biology and Medicine* 90 (2017), 33–49. 2
- [TBK*05] TURNER W., BROWN R., KELLIHER T., TU P., TAISTER M., MILLER K.: A novel method of automated skull registration for forensic facial approximation. *Forensic Science International* 154 (2005), 149–158. 2
- [TBL*07] TU P., BOOK R., LIU X., KRAHNSTOEVEER N., ADRIAN C., WILLIAMS P.: Automatic Face Recognition from Skeletal Remains. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2007), pp. 1–7. 2
- [TGB13] THIERY J.-M., GUY E., BOUBEKEUR T.: Sphere-Meshes: Shape Approximation using Spherical Quadric Error Metrics. *ACM Trans. Graph. (Proc. SIGGRAPH Asia)* 32, 6 (2013), 178:1–178:12. 3
- [TPT16] TKACH A., PAULY M., TAGLIASACCHI A.: Sphere-meshes for Real-time Hand Modeling and Tracking. *ACM Trans. Graph.* 35, 6 (2016), 222:1–222:11. 3, 5
- [TTR*17] TKACH A., TAGLIASACCHI A., REMELLI E., PAULY M., FITZGIBBON A.: Online Generative Model Personalization for Hand Tracking. *ACM Trans. Graph.* 36, 6 (2017), 243:1–243:11. 3, 5
- [VBPP05] VLASIC D., BRAND M., PEISTER H., POPOVIĆ J.: Face Transfer with Multilinear Models. *ACM Trans. Graph.* 24, 3 (2005), 426–433. 5, 6
- [Wil10] WILKINSON C.: Facial reconstruction—anatomical art or artistic anatomy? *Journal of Anatomy* 216 2 (2010), 235–50. 2
- [ZPK16] ZHOU Q.-Y., PARK J., KOLTUN V.: Fast Global Registration. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II* (2016), pp. 766–782. 4