Prediction of 3D Body Parts from Face Shape and Anthropometric Measurements

Alessio Gallucci[1], Dmitry Znamenskiy[2], and Milan Petkovic[1,2]
[1] Eindhoven University of Technology, Eindhoven, Netherlands
[2] Philips Research, Eindhoven, Netherlands
Email: {a.gallucci, m.petkovic}@tue.nl, {dmitry.znamenskiy, milan.petkovic}@philips.com

***Abstract*—While 3D body models have been vastly studied in the last decade, acquiring accurate models from the sparse information about the subject and few computational resources is still a main open challenge. In this paper, we propose a methodology for finding the most relevant anthropometric measurements and facial shape features for the prediction of the shape of an arbitrary segmented body part. For the evaluation, we selected 12 features that are easy to obtain or measure including age, gender, weight and height; and augmented them with shape parameters extracted from 3D facial scans. For each subset of features, with and without facial parameters, we predicted the shape of 5 segmented body parts using linear and non-linear regression models. The results show that the modeling approach is effective and giving sub cm reconstruction accuracy. Moreover, adding face shape features always significantly improves the prediction.**

***Index Terms*—human body modeling, 3D face shape, 3D body shape, human body shape prediction, anthropometry, sizing parameters, 3D scan data, 3D scans**

I. Introduction

The use of 3D human body shape has the potential of changing the way we interact with the world in a wide variety of ways. Applications of this technology have been proved helpful in several fields such as healthcare, cognitive science [1], [2], online shopping [3], [4], clothing [5] and virtual reality [6], [7]. For example, in the healthcare domain, the knowledge of 3D body shape can help in the assessment of the Psoriasis Area and Severity Index (PASI) [8], dosing chemotherapy according to the Body Surface Area (BSA) [9] or estimating a burned body part [10]. All this application lack precision in estimation [11]-[13] and accurate body shape prediction would help the dosage of a particular drug. The prediction of (a less accurate) 3D models from available metadata and body measurements can be considered as a lower cost alternative to full body 3D scanning and processing involving the recognition and processing of different body parts. Note that different practical applications can require 3D shapes and measurements with different precisions. Moreover, the obtaining of less accurate 3D body models computed from the available measurements can be used as a pre-processing tool for accurate registration of 3D models to raw scans.

One of the obstacles in the efficient processing of full body 3D models is the high volume of data. Cost and volume of the data required can be drastically reduced by learning a statistical representation of the human shape space, as described in [14], [15]. Only sparse data, combined with the learned space, are needed to reconstruct a full body scan instead of a dense representation. In the next section, we give an overview considering the prior art which relates and predicts the representation of the 3D body in the shape applications can require the prediction of 3D body shape using the least possible amount of metadata and low-cost body measurements. Thus in this paper, we address the following points we consider novel: first, we evaluate the predictive power of different combinations of features and study how the error drops when their number increases. Second, we consider facial 3D scan as a lower-cost and less-obtrusive alternative to a full body 3D scan. We analyze the improvement of the body shape prediction when the metadata and the measurements are augmented with features extracted from the facial 3D scans. Third, we apply the above analysis to body parts, which can be arbitrarily segmented on the body.

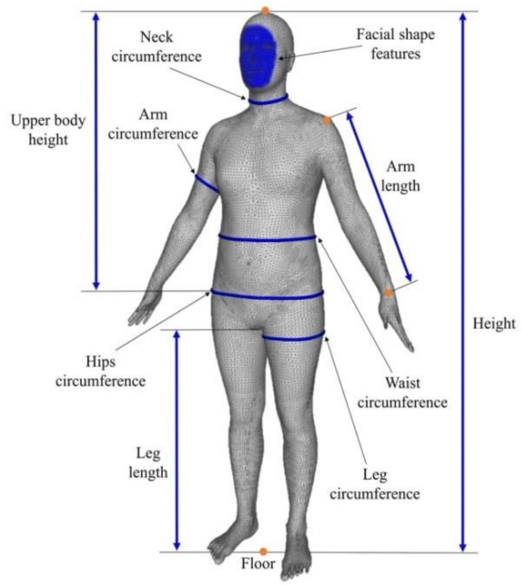


Figure 1. Facial shape features and measurements computed from the registered body meshes: Height, upper body height, leg and arm length, the perimeters for waist, hips, arm, leg quadriceps and neck.

Manuscript received January 17, 2020; revised July 27, 2020.

doi: 10.18178/joig.8.3.67-74

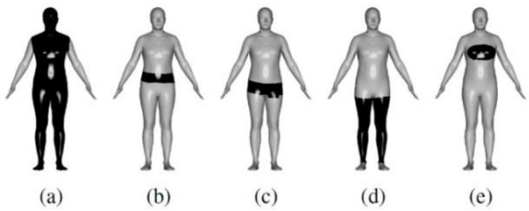


Figure 2. The body parts, also called segmentation masks, with segmented area highlighted in black. We refer to them as a) full body mask without arms, b) waistband, c) hips band, d) legs mask and e) breasts mask. The fraction of the segmented vertices with respect to the whole body are 0.78, 0.05, 0.08, 0.26, 0.02 respectively.

In our analysis, we considered 12 features that can be relatively easily and reliably collected from a subject: gender, age, weight, and nine measurements shown in Fig. 1 (most of the figures were generated using MeshLab [16]). Note that we have considered these features as an example, and others can be taken into account depending on the application and the data available. We apply our methodology for the prediction of five example body shapes, shown as segmentation masks in Fig. 2. We selected those five due to possible applications in healthcare and personal care. For each body part, we assessed how well it can be predicted given each possible subset of the measurements. For each subset of features, we considered how much the prediction accuracy improves when adding to the feature set the features describing the facial geometry, i.e. coefficients in the facial shape space introduced in [17].

The rest of the paper is organized as follows. Section II introduces the literature related to body shape analysis and modeling. Subsequently, in Section III, we describe our approach: the registration of the database's population adopting a common template model, the encoding into a parametric shape space using Principal Component Analysis (PCA) and the prediction model used to link face and body shapes. In the last part of the section, we introduce the error measure used for the evaluation of the prediction. The results, presented in Section IV, demonstrate that face shape has a positive correlation with different body parts including hips, waist, breast and legs. In the conclusion section, we summarize our view on the 3D body shape prediction from sparse meta-data and the face, and we elaborate on future applications and directions of our work.

II. Related Work

Many works have created models that correlate a body statistical shape space to other features, descriptors or meta-information but none of them define a strategy to find the optimal features and none include in the prediction another shape space (in our case the face).

Blanz and Vetter [17] defined how to learn a statistical shape space of the face and then used measurements and semantic descriptors to modify the face appearance. In [14], [15], Allen *et al.* were the first to employ the paradigm explained by [17] on 3D body scans. The authors paved the way for the application of this new method in exploring and studying human shape space. They first registered 250 scans, from dataset [18], solving an optimization problem that minimizes sparse markers' distance, vertices' distance and smoothness of the transformation. Then they learned a linear function mapping anthropometric measurements to the shape coefficient. In [19], [20], Seo *et al.* defined a model that can be modified or generated using only anthropometric measurements. They used radial basis interpolation to reconstruct the relationship between sizing parameters to shape space. Hasler *et al.* [21] includes in the registration phase the high level semantic parameters allowing the generation of realistic body meshes. Wuhrer and Shu [22] generated realist body shape fitting anthropometric measurements using non-linear optimization. Tsoli *et al.* [23] built a model to predict measurements from 3D scans. More recently in [24], [25], Hill *et al.* defined a linguistic space using common body words like fat, rounded or skinny. They first used Amazon Mechanical Turk to link descriptors and body shape by rating photographs. Streuber *et al.* [26] similarly used crowdsourcing to define verbal descriptors and to demonstrate that they are sufficient for retrieving a realistic 3D scans. While previous work finds a relationship between body measurements/characteristics and body shapes, they do not define a strategy to find the optimal subset of them for a specific body part. Moreover, they do not use facial features and/or another shape space as predictor.

Other works explored the correlation between face shape and textures to body parameters: Windhager *et al.* [27] linked facial features of young Caucasian females to body fat proportion using geometric morphometrics. Similarly, Mayer *et al.* [28] retrieved high-resolution face images and registered them using geometric morphometric. However, their experiments do not use the parametric modeling of the human body shape but they predicted a positive correlation between body mass index and waist-to-hip ratio with facial shape and texture. A similar approach to our work is presented in [29] where the authors model the difference between real and virtual measurements and fit a more advanced model with kinematic skeleton. However, they use a linear model for the mapping between features and body shape relying on very specific and not very accurate body measurements. They used VR controllers for collection adding the weight, probably because is a very strong predictor. Moreover, they selected the features based on their acquisition accuracy rather than their predictive power as presented in our work.

Multiple techniques are available to retrieve a 3d representation of a person from different sources (images, depth cameras, sparse markers, silhouettes, etc…). For example, Balan *et al.* [30] reconstructed the parametric shape model [31] using multiple images while more recent works [32]-[35] leverages only a single image and convolutional neural networks.

III. Method

We developed two parametric models, following the method explained in [14], [17], one for the body and one for the face shape. The face model was derived using more than 3000 3D scans, including the Size China Dataset [36][1]. The parametric body model was derived using more than 4000 full body scans standing in a frontal tree position as shown in Fig. 3a. The scans were taken mainly from the CAESAR dataset [18].

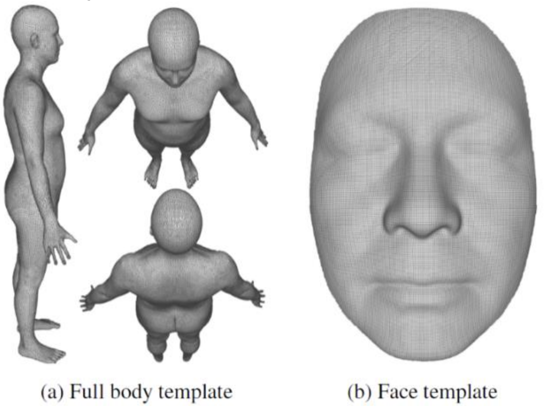


Figure 3. (a) Body template mesh standing in tree position and containing N_P≈53000 vertices. (b) Face template mesh containing N_Q≈23000 vertices.

In the following subsections, we describe the registration of the template meshes into 3D scans, the encoding of the registered models into the selected parameters. Then, we introduce the non-linear prediction model used to find the best subset of features for each segmented body part and, finally, the error measure used to evaluate the experiments.

A. Registration

In order to register every face and full body mesh, we employed state of the art non-rigid registration techniques [37]-[39]. We used a template mesh with about $N_P \approx$ 53000 vertices for the body, see Fig. 3a, and another template mesh with about $N_Q \approx$ 23000 vertices for the face, see Fig. 3b. Both template models were then used to register the full body scans dataset. We assessed the quality of the registration via visual inspection and other measures outlined in the survey [37]. For about $N \approx 3750$ full body scans both registrations have shown low fit error (below 0.5mm Root Mean Squared Error (RMSE), as surfaces distance, for the registration of the facial mesh and below 1.0mm RMSE for the full body).

Registration led to the following representation of each participant as the two morphed template meshes. Let $v_{i,j} \in \mathbb{R}^3$ be the full body morphed coordinates of vertex $j \in N_P$ at participant $i \in N$. Furthermore, we can write the morphed coordinates of all vertices of scan $i \in N$ as a single flattened vector, stacking all vertices' coordinates together, as

$$\boldsymbol{p}_i^r = \left(\boldsymbol{v}_{i,1}^r, \boldsymbol{v}_{i,2}^r, \dots, \boldsymbol{v}_{i,N_P}^r\right) \in \mathbb{R}^{3N_P} \tag{1}$$

and collecting all participants into a rectangular matrix we have

$$P_r = (\boldsymbol{p}_1^r; \boldsymbol{p}_2^r; \dots; \boldsymbol{p}_N^r)' \in \mathbb{R}^{N\times 3N_P} \tag{2}$$

In the same manner the definition of the face representation is $Q_r = (\boldsymbol{q}_1^r; \boldsymbol{q}_2^r; \dots; \boldsymbol{q}_N^r)' \in \mathbb{R}^{N\times 3N_Q}$.

[1]All other scans were collected at Philips

B. Parametric Spaces

The registered meshes were parametrized with Principal Component Analysis (PCA) transformation, using 200 eigenvectors for the body and 180 eigenvectors for the face. The PCA transformation can be written in matrix form as

$$P_r = \bar{P}_r + YD' + E_r \tag{3}$$

where $\bar{P}_r \in \mathbb{R}^{N\times 3N_P}$ is the matrix of N times repeated average mesh coordinates

$$\begin{gathered}\bar{\boldsymbol{p}} = \left(\bar{p}_{1_x}, \bar{p}_{1_y}, \dots, \bar{p}_{N_{P_z}}\right) \in \mathbb{R}^{3N} \\ \bar{p}_{j_x} = \frac{\sum_i P_r(i, j_x)}{N_P}\end{gathered} \tag{4}$$

$D \in \mathbb{R}^{3N_P\times 200}$ is the reduced eigenvectors matrix, composed of the 200 \`principal\` eigenvectors (*i.e.* eigenvectors with highest eigenvalues) of the covariance matrix $(P_r - \bar{P}_r)'(P_r - \bar{P}_r)$, $Y \in \mathbb{R}^{N\times 200}$ is the reduced matrix of PCA coefficients, and $E_r \in \mathbb{R}^{3N_P}$ is the residual error, *i.e.*

$$P_r \approx P = \bar{P}_r + YD' \tag{5}$$

The transformation (5) gives a compact representation of 53000×3-dimensional vectors of vertex coordinates P_r with the 200-dimensional PCA coefficient vectors Y. In the same way, we apply the PCA transformation to the registered facial meshes:

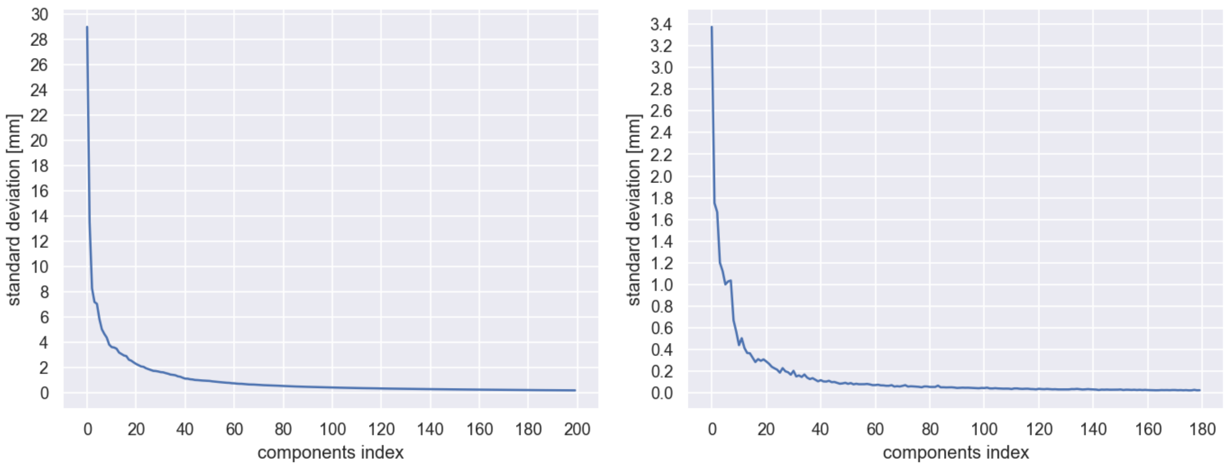
$$Q_r \approx Q = \overline{Q_r} + X_Q D_Q' \tag{6}$$

where $\bar{Q}_r \in \mathbb{R}^{N\times 3N_Q}$ is the matrix of N times repeated average mesh coordinates, D_Q consists of the 180 \`principal\` eigenvectors of the covariance matrix $(Q_r - \bar{Q}_r)'(Q_r - \bar{Q}_r)$, and $X_Q \in \mathbb{R}^{N\times 200}$ are the facial PCA coefficients. The results of the encoding for both models is shown in Fig. 4. The residual error between P_r and P, computed using equation (14) and explained in Section D, is less than 2.5 mm. Similarly, the residual error for the face is less than 0.3 mm.

C. Prediction Model

In this section, we describe how the body shape coefficients Y are predicted using the subject's features, denoted as $X_F \in \mathbb{R}^{N\times(N_F+1)}$, (where \`+1\` corresponds to free term in the regression model) and the face shape space X_Q. As subject features, we have considered *reported weight*, *age*, *gender*, and body measurements extracted from the registered meshes such as *body height*, *arm length*, *waist circumference*. This set was augmented by including their interactions up to $d = 3$-rd degree. Thus, considering in total N_F personal features, the expanded set corresponds to the terms of the polynomial with degree d build from them. This holds for all features except the ones with lower interactions allowed, like *gender*. In the following, we denote the augmented set of features by $X_G \in \mathbb{R}^{N\times(N_G+1)}$, where the reader can derive the general formula for N_G using basic combinatorial techniques [40] as

$$N_G = \binom{N_F + d}{d} - 1 \tag{7}$$



(a) Body shape coefficients

(b) Face shape coefficients

Figure 4. The significance of the encoding, i.e. the standard deviation of the PCA coefficients for the body (a) and for the face (b). The decision to use 200 principal components for the body and 180 for the face was a heuristic decision seeking a compromise between the requirements to represent all shape spaces adequately and to not encode noise. The standard deviation of the last PCA body shape component is 0.18mm and for the face it is 0.025mm.

which, in the case when the (binary) gender feature is included, becomes

$$N_G = \binom{N_F + d}{d} - 1 - (N_F + 1) \tag{8}$$

Equations (7), (8) are given for completeness but are not needed to understand the rest of the paper or run algorithms which can simply count the combinations. To facilitate the notation, we include the constant term in both X_F and X_G, but it is not counted in N_F and N_G.

Then, we performed multi-linear regression for the body coefficients Y

$$Y = XB + \varepsilon \tag{9}$$

with four settings of the independent variable X, with and without interactions and with and without face coefficients:

$$\begin{aligned} &(a) \quad X = X_F \in \mathbb{R}^{N\times(N_F+1)} \\ &(b) \quad X = X_G \in \mathbb{R}^{N\times(N_G+1)} \\ &(c) \quad X = [X_F, X_Q] \in \mathbb{R}^{N\times(N_F+1+N_Q)} \\ &(d) \quad X = [X_G, X_Q] \in \mathbb{R}^{N\times(N_G+1+N_Q)} \end{aligned} \tag{10}$$

Next, we evaluated the predictions of specific body parts, using the segmentation masks shown in Fig. 2. The arms were excluded from the segmentation masks deliberately since subjects had visible variability in the arm positions and the for lack of a pose model. To improve the prediction for each body part, instead of solving the basic regression (9), we solved the weighted versions as shown below. Let $I_m \in \mathbb{R}^{3N_P\times 3N_P}$ be the diagonal matrix of mask m, where $I_m(j,j) = 1$ if and only if the vertex is part of the segmentation mask. Recall $P = \bar{P}_r + YD'$ (equation 5) and note that for each body part m we want to have $I_m P$ accurately predicted. Then, assuming the regression model $Y = XB$, we get

$$\begin{aligned} \bar{P}_r I_m + YD'I_m &= \bar{P}_r I_m + XBD'I_m + \varepsilon D'I_m \\ YD'I_m &= XBD'I_m + \varepsilon D'I_m \\ YD'I_m D &= XBD'I_m D + \varepsilon D'I_m D \\ Y\Sigma_m &= XB\Sigma_m + \varepsilon\Sigma_m \end{aligned} \tag{11}$$

where $\Sigma_m = D'I_m D \in \mathbb{R}^{200\times 200}$. The least mean square estimate of B in the above equation is

$$\hat{B}_m = ((X'X)^{-1}X'Y\Sigma_m)\Sigma_m^{-1} \tag{12}$$

for each mask m.

D. Fitness Measures

For each model and mask, we performed a leave-one-out cross validation on the N participants. In other words, the estimation of $\hat{B}$ has been carried out every time, leaving out the participant to predict. Once computed the predicted body coefficients $\hat{Y}$ we need to convert back, decode, using the PCA transformation (5) to reach the predicted vertices $\hat{P}$ as

$$\hat{P} = \bar{P}_r + \hat{Y}D' = \bar{P}_r + X\hat{B}D' \tag{13}$$

To evaluate the prediction, we first aligned the predicted $\hat{P}(i,:)$ to the original coordinates $\forall i \in [1, N]$ with weighted Procrustes [41], and then we computed the vertex-wise RMSE over all participants for each vertex v_{ij} versus its predicted position $\hat{v}_{ij}$

$$E_j = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\|\hat{\boldsymbol{v}}_{i,j} - \boldsymbol{v}_{i,j}\right\|_2^2} \tag{14}$$

Since comparing the distribution of the vertices errors on the surface is beyond the scope of the research, as a final measure of fitness for the masks, we used the mean absolute error for all vertices:

$$E = \frac{1}{N_P}\sum_{j=1}^{N_P}|E_j| \tag{15}$$

Unlike other works, which used mainly point to surface distance, the above error measure also penalizes misplacement of the body part points on the surface and therefore can be considered more accurate.

IV. RESULTS

We evaluated 2 groups of features, listed in Table I, with 12 features in total. The first group is composed of reported gender, age and weight (without clothes), all acquired in [11]. The second group includes parametric measurements that were computed from the registered body meshes: the height computed as head to floor; upper body height as head to the highest touchable point of the pelvis; arm length as the distance between acromion (shoulder) to the distal end of the middle finger; leg length from crotch to floor; the perimeters for waist as the midpoint between the lower margin of the last palpable rib and the top of the iliac crest; hips circumference it is performed at the most prominent point, on the major trochanters, and at the level of the maximum relief of the gluteal muscles; arm circumference taken from the midpoint of the total length of the arm, between acromion and olecranon; leg quadriceps circumference taken from the midpoint of the total length of the thigh; neck circumference taken from the midpoint of the total length of the neck. The covariance matrix of all the features is presented in Table II.

TABLE I. FEATURES DEFINITION WHERE CQ STANDS FOR CAESAR QUESTIONNAIRE AND PM FOR PARAMETRIC MEASUREMENT

Name	Type	Mean + Std	Source
Gender	Male[1] or Female[2]	1.53 ± 0.5	CQ
Age	Yrs	38.00 ± 12.59	CQ
Weight	Kg	74.56 ± 18.09	CQ
Height	Y-length [mm]	1701.38 ± 100.78	PM
Waist	Circumference [mm]	889.83 ± 150.45	PM
Arm	Circumference[mm]	306.00 ± 43.45	PM
Hip	Circumference[mm]	1037.99 ± 106.07	PM
Leg	Circumference[mm]	614.94 ± 67.91	PM
Neck	Circumference[mm]	364.09 ± 43.33	PM
Leg	Y-distance[mm]	764.41 ± 55.67	PM
Arm	distance[mm]	557.64 ± 40.59	PM
Upper Body	Y-distance[mm]	750.72 ± 42.90	PM

TABLE II. FEATURES CORRELATION MATRIX

	Age	Gender	Weight	Height	WaistC	ArmC	HipC	LegC	NeckC	LegL	ArmL	UBodyH
Age	1.00	0.01	0.22	-0.08	0.40	0.26	0.26	0.12	0.26	-0.17	-0.03	-0.06
Gender		1.00	-0.42	-0.64	-0.34	-0.31	0.03	0.01	-0.68	-0.39	-0.65	-0.57
Weight			1.00	0.55	0.90	0.88	0.82	0.78	0.84	0.26	0.55	0.55
Height				1.00	0.32	0.31	0.19	0.14	0.52	0.86	0.91	0.85
WaistC					1.00	0.85	0.82	0.74	0.83	0.02	0.36	0.36
ArmC						1.00	0.82	0.81	0.76	0.01	0.31	0.36
HipC							1.00	0.92	0.56	-0.03	0.19	0.24
LegC								1.00	0.52	-0.09	0.14	0.20
NeckC									1.00	0.20	0.55	0.53
LegL										1.00	0.83	0.57
ArmL											1.00	0.68
UBodyH												1.00

We assessed the importance of each feature by performing a search over all possible combinations of the set X_F resulting in 2^{12}=4096 possible subset of features. We consider the empty subset as the error compared to the average of the population. For each subset, we compared four different feature designs $X_F, X_G, [X_F, X_Q], [X_G, X_Q]$. The maximum number of features reached by models without the face is N_G minus all combinations of the gender. In our example, the maximum number of features is N_F=12 and all combinations of gender from second order are N_F+1 hence using equation (8) we have that the maximum number of regressors, when using interactions is N_G=441. Considering instead the example with age, gender, weight and height, where N_F=4 we have N_G=29. In the following, we presents the errors for the full body mask without arms and next the errors for all the remaining four body parts.

A. Full Body Mask without Arms

Table III shows for each cardinality of X_F the best model for $X = X_G$ and the error using the other three input matrices. The most accurate single feature is the height with $E = 20.89$mm, because it is a major indicator of body size. The only feature that actually outperforms height is the body volume. However, we excluded it from the analysis since, it is a very uncommon measurement to be taken in a real life scenario. Note that height alone is giving a smaller error than using weight and face shape combined. The best combination of two features is the height and the weight, resulting in 17.93mm residual error. The minimal error is achieved using all 12 features, and it is 13.92mm for X_G and 13.00mm for $[X_G, X_Q]$. The average error reduction for those 12 best models is 1.33mm, and the average effect of adding the face coefficients X_Q is the drop of the error with 8.12%. We think that adding new measurements will only affect the error minimally because of the small variability of different subjects' pose. In fact, the pose cannot be predicted from the measurements, and we believe that the addition of pose normalization methods would result in lower errors.

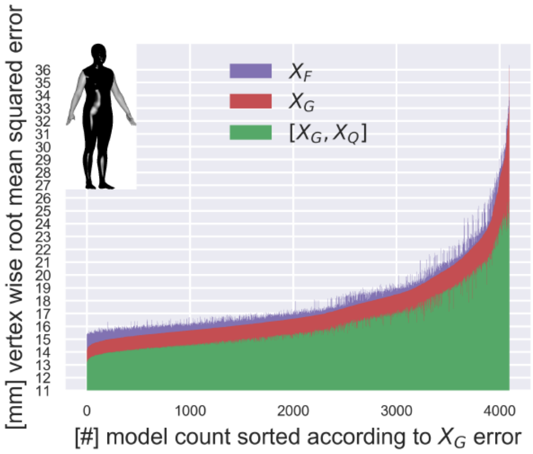


Figure 5. Full body mask with all 4095 models sorted according to the error of $X = XG$. The minimum error of 13mm is achieved using all 12 features plus the face coefficients.

In order to evaluate the significance of adding the face shape, we considered the model with $X = [X_Q, X_G]$ where X_G is augmented from $X_F =$ [age, gender, weight, height]. This model has an error of 15.91mm, which is better than the error of the model with $N_F = 4$ best predictors without face. Hence the face shape can replace detailed parametric measurements. Thus, for example, the face coefficients combined with age, gender and weight features give lower error than the prediction using waist, hip circumference and leg length features. In Fig. 5 all possible subsets, excluding the empty one, visually demonstrate that the face has a significant positive contribution to the prediction. In fact, the average error drop, when extending X_G to $[X_G, X_Q]$, is 0.98mm or 9.72%. On opposite, the more features are considered the bigger the effect of adding interactions between them adding of interactions, as one can see in the Table IV to Table VII, when comparing columns X_F to X_G.

TABLE III. ERROR E FOR FULL BODY WITHOUT ARMS USING X_G BEST FEATURES

N_F	N_G	X_F	$[X_F, X_Q]$	X_G	$[X_G, X_Q]$	Features
0	0	36.70 ± 10.69	36.70 ± 10.69	36.70 ± 10.69	36.70 ± 10.69	Avg distance
1	3	20.92 ± 5.36	17.46 ± 3.52	20.89 ± 5.35	17.45 ± 3.52	Height
2	9	18.11 ± 3.20	16.29 ± 2.64	17.93 ± 3.14	16.21 ± 2.60	Weight, Height
3	19	17.00 ± 3.20	15.28 ± 2.70	16.75 ± 3.13	15.15 ± 2.64	Weight, Height, LegL
4	34	16.31 ± 2.80	14.95 ± 2.47	16.00 ± 2.69	14.73 ± 2.38	Height, WaistC, HipC, LegL
5	55	15.91 ± 2.66	14.65 ± 2.35	15.56 ± 2.57	14.37 ± 2.28	Height, WaistC, HipC, LegL, UBodyH
6	83	15.69 ± 2.67	14.51 ± 2.37	15.22 ± 2.55	14.13 ± 2.29	Height, WaistC, HipC, LegC, LegL, UBodyH
7	119	15.56 ± 2.61	14.46 ± 2.34	15.02 ± 2.51	13.98 ± 2.25	Age, Height, WaistC, HipC, LegC, LegL, UBodyH
8	164	15.50 ± 2.60	14.42 ± 2.34	14.80 ± 2.47	13.81 ± 2.23	Age, Weight, Height, WaistC, HipC, LegC, LegL, UBodyH
12	441	$\mathbf{15.35 \pm 2.54}$	$\mathbf{14.32 \pm 2.30}$	$\mathbf{13.92 \pm 2.28}$	$\mathbf{13.00 \pm 2.07}$	All 12 features
4	29	17.41 ± 2.91	16.15 ± 2.65	17.11 ± 2.79	15.91 ± 2.54	Age, Gender, Weight, Height

TABLE IV. ERROR E FOR WAISTBAND USING X_G BEST FEATURES

N_F	N_G	X_F	$[X_F, X_Q]$	X_G	$[X_G, X_Q]$	Features
0	0	22.64 ± 3.66	22.64 ± 3.66	22.64 ± 3.66	22.64 ± 3.66	Avg distance
1	3	12.68 ± 2.34	10.51 ± 1.53	12.59 ± 2.32	10.44 ± 1.51	HipC
2	9	11.21 ± 1.08	9.59 ± 0.96	11.00 ± 1.05	9.47 ± 0.94	WaistC, HipC
3	15	10.63 ± 1.20	9.48 ± 0.92	10.37 ± 1.19	9.29 ± 0.97	Gender, WaistC, HipC
4	29	10.41 ± 1.11	9.38 ± 0.95	10.04 ± 1.08	9.11 ± 0.93	Gender, WaistC, HipC, LegC
5	49	10.29 ± 1.08	9.29 ± 0.92	9.87 ± 1.06	8.96 ± 0.90	Gender, WaistC, HipC, LegC, UBodyH
6	83	10.21 ± 1.08	9.20 ± 0.87	9.65 ± 0.98	8.78 ± 0.86	Height, WaistC, HipC, LegC, LegL, UBodyH
7	119	10.06 ± 1.05	9.13 ± 0.89	9.44 ± 0.98	8.65 ± 0.85	Age, Height, WaistC, HipC, LegC, LegL, UBodyH
8	155	10.00 ± 1.05	9.11 ± 0.88	9.25 ± 1.01	8.50 ± 0.87	Age, Gender, Height, WaistC, HipC, LegC, LegL, UBodyH
12	441	$\mathbf{9.92 \pm 1.02}$	$\mathbf{9.05 \pm 0.87}$	$\mathbf{8.59 \pm 0.94}$	$\mathbf{7.93 \pm 0.82}$	All 12 features
4	29	12.06 ± 1.30	10.96 ± 1.13	11.54 ± 1.32	10.47 ± 1.14	Age, Gender, Weight, Height

TABLE V. ERROR E FOR HIPS BAND USING X_G BEST FEATURES

N_F	N_G	X_F	$[X_F, X_Q]$	X_G	$[X_G, X_Q]$	Features
0	0	20.42 ± 3.01	20.42 ± 3.01	20.42 ± 3.01	20.42 ± 3.01	Avg distance
1	3	11.76 ± 2.67	9.59 ± 1.63	11.70 ± 2.66	9.51 ± 1.61	HipC
2	9	10.76 ± 1.88	9.22 ± 1.52	10.58 ± 1.79	9.09 ± 1.47	HipC, NeckC
3	19	10.29 ± 1.80	8.92 ± 1.46	9.94 ± 1.78	8.69 ± 1.40	WaistC, HipC, LegC
4	29	9.88 ± 1.62	8.80 ± 1.42	9.48 ± 1.52	8.50 ± 1.34	Gender, WaistC, HipC, LegC
5	55	9.70 ± 1.56	8.69 ± 1.37	9.25 ± 1.47	8.34 ± 1.29	Height, WaistC, HipC, LegC, LegL
6	83	9.44 ± 1.47	8.50 ± 1.27	8.88 ± 1.36	8.06 ± 1.20	Height, WaistC, HipC, LegC, LegL, UBodyH
7	119	9.32 ± 1.44	8.44 ± 1.26	8.73 ± 1.34	7.95 ± 1.18	Age, Height, WaistC, HipC, LegC, LegL, UBodyH
8	164	9.28 ± 1.44	8.42 ± 1.25	8.59 ± 1.31	7.85 ± 1.16	Age, Weight, Height, WaistC, HipC, LegC, LegL, UBodyH
12	441	$\mathbf{9.19 \pm 1.41}$	$\mathbf{8.37 \pm 1.24}$	$\mathbf{8.00 \pm 1.21}$	$\mathbf{7.33 \pm 1.09}$	All 12 features
4	29	11.66 ± 1.55	10.59 ± 1.36	11.17 ± 1.52	10.12 ± 1.33	Age, Gender, Weight, Height

TABLE VI. ERROR E FOR BREASTS USING X_G BEST FEATURES

N_F	N_G	X_F	$[X_F, X_Q]$	X_G	$[X_G, X_Q]$	Features
0	0	13.77 ± 4.24	13.77 ± 4.24	13.77 ± 4.24	13.77 ± 4.24	Avg distance
1	3	9.38 ± 2.54	7.45 ± 1.70	9.30 ± 2.52	7.44 ± 1.69	WaistC
2	6	7.73 ± 1.72	7.15 ± 1.58	7.64 ± 1.66	7.07 ± 1.53	Gender, WaistC
3	15	7.55 ± 1.66	7.03 ± 1.55	7.42 ± 1.60	6.93 ± 1.49	Gender, Weight, WaistC
4	29	7.47 ± 1.64	6.96 ± 1.53	7.27 ± 1.55	6.81 ± 1.46	Gender, Weight, WaistC, UBodyH
5	49	7.42 ± 1.64	6.94 ± 1.52	7.18 ± 1.53	6.73 ± 1.45	Age, Gender, Weight, WaistC, UBodyH
6	76	7.39 ± 1.64	6.91 ± 1.52	7.10 ± 1.52	6.68 ± 1.43	Gender, Weight, Height, WaistC, LegC, LegL
7	111	7.36 ± 1.63	6.90 ± 1.51	7.01 ± 1.50	6.61 ± 1.42	Age, Gender, Weight, Height, WaistC, LegC, LegL
8	155	7.34 ± 1.63	6.89 ± 1.51	6.92 ± 1.48	6.54 ± 1.40	Age, Gender, Weight, Height, WaistC, LegC, LegL, UBodyH
12	441	$\mathbf{7.29 \pm 1.60}$	$\mathbf{6.85 \pm 1.50}$	$\mathbf{6.50 \pm 1.38}$	$\mathbf{6.15 \pm 1.30}$	All 12 features
4	29	7.83 ± 1.81	7.27 ± 1.67	7.57 ± 1.69	7.07 ± 1.58	Age, Gender, Weight, Height

TABLE VII. ERROR E FOR LEGS USING X_G BEST FEATURES

N_F	N_G	X_F	$[X_F, X_Q]$	X_G	$[X_G, X_Q]$	Features
0	0	19.52 ± 4.96	19.52 ± 4.96	19.52 ± 4.96	19.52 ± 4.96	Avg distance
1	3	13.99 ± 2.27	11.91 ± 1.78	13.94 ± 2.25	11.89 ± 1.78	LegL
2	9	12.14 ± 1.46	10.89 ± 1.26	12.07 ± 1.45	10.83 ± 1.25	LegC, LegL
3	19	11.65 ± 1.39	10.78 ± 1.25	11.54 ± 1.37	10.68 ± 1.22	Height, LegC, LegL
4	34	11.51 ± 1.35	10.69 ± 1.22	11.37 ± 1.31	10.55 ± 1.20	Height, HipC, LegC, LegL
5	55	11.40 ± 1.29	10.61 ± 1.20	11.15 ± 1.26	10.42 ± 1.18	Height, WaistC, HipC, LegC, LegL
6	83	11.31 ± 1.28	10.57 ± 1.19	11.00 ± 1.24	10.31 ± 1.16	Height, WaistC, HipC, LegC, LegL, UBodyH
7	119	11.25 ± 1.27	10.54 ± 1.19	10.86 ± 1.24	10.20 ± 1.16	Weight, Height, WaistC, HipC, LegC, LegL, UBodyH
8	164	11.21 ± 1.27	10.52 ± 1.20	10.72 ± 1.22	10.08 ± 1.15	Age, Height, WaistC, ArmC, HipC, LegC, LegL, UBodyH
12	441	$\mathbf{11.13 \pm 1.27}$	$\mathbf{10.47 \pm 1.19}$	$\mathbf{10.12 \pm 1.15}$	$\mathbf{9.53 \pm 1.09}$	All 12 features
4	29	12.88 ± 1.60	11.99 ± 1.44	12.70 ± 1.55	11.84 ± 1.40	Age, Gender, Weight, Height

B. Additional Four Masks

An interesting observation is that, while height is coming first for the body, it is not the case for hips and waist band prediction, where, weight gives a better accuracy among the single feature predictors. As expected the circumferences are now playing a much more significant role in the specific masks compared to the full body mask. This is shown in Table IV to Table VII.

For both the waist and hips masks, the best performing feature is the hip circumference, registering an error of 12.59mm and 11.70mm respectively. The lowest error reached using all features for the waist mask is 8.59mm whereas the hip mask achieved a minimum error of 8.00mm. For the breast mask, the best single feature is the waist circumference that reaches an error of 9.30mm, and as foreseen, gender plays an important role as well. For this mask the lowest error, achieved using all features, is 6.50mm. Finally, analyzing the error registered in the leg mask, it can be noticed that the leg length plays the most crucial role, reaching an error of 13.94mm. It is followed by the leg circumference and the height. The minimum error achieved in this mask, using all the features, is 10.12mm.

Overall, the face improves the most the hips band where the reduction for the best 12 models is 10.45% (0.99mm). For the waist mask the average reduction is 9.71% (0.98mm) and for the full body, described in Section A, the drop is 8.12% (1.33mm). Finally the reduction for the legs is 7.32% (0.84mm) and the face achieves the least reduction in the breasts area with 7.14% (0.54mm). Thus, we can deduce that the face is more relevant in predicting hips and waists compared to the legs and breasts.

V. Discussion and Conclusion

With this work, we showed how to couple two high-resolution parametric spaces of body and face with metadata and low-cost measurements. Initially, we predicted the body shape parameters using anthropometric measurements. In addition, we included the face shape parameters to our predictive model leading to the conclusion that they always improve the prediction.

As far as our analysis is concerned, additional research could lead to the increase of the accuracy of our predictions. In the future, it would be helpful to include a skeleton model to factor out the pose, as shown in [42]. This, in turn, will prevent information loss in the PCA encoding due to factors affecting the pose (e.g. the position of arms and legs). The regression model can be enhanced using regularizing techniques. We believe that Lasso [43] is the best to set the tail components of the face to zero when needed. We avoided presenting those regularizations since it is beyond the scope of this paper. A further direction of research is the prediction of the face components out of images of the face. This way, one could predict the body shape coefficients using pictures instead of the 3D shape. A possible path to follow is extracting landmarks after aligning and then extrapolating 3D shapes. Suitable techniques to follow this approach are explained in [44], [45].

An interesting application of the procedure described in this paper could lie in correlating other body parts to one another. In principle, any body part could be registered and encoded via PCA. As an example, the investigation of the relationship of foot features on the back has been studied via Geometric Morphometric in [46], [47]. They correlated foot shape with anthropometric measurements like height, Body Mass Index (BMI) and gender. Their work can be enriched by registering belly, hips or back areas and then by studying the effect on the back with our approach. Although, several studies have been conducted using BMI as a base factor, body shape coefficients have the potential of conveying more information and thus improving the prediction.

Conflict of Interest

All the authors do not have any conflict of interests.

Author Contributions

All the authors contributed equally to the paper.

Acknowledgment

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 779780.

References

[1] K. K. Cornelissen, K. McCarty, P. L. Cornelissen, and M. J. Tovée, "Body size estimation in women with anorexia nervosa and healthy controls using 3D avatars," *Sci. Rep.*, vol. 7, no. 1, p. 15773, 2017.

[2] S. C. Mölbert, *et al.*, "Assessing body image in anorexia nervosa using biometric self-avatars in virtual reality: Attitudinal components rather than visual body size estimation are distorted," *Psychol. Med.*, vol. 48, no. 4, pp. 642-653, 2018.

[3] K. Kristensen, *et al.*, "Towards a next generation universally accessible 'online shopping-for-apparel' system," *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8006, part 3, pp. 418-427, 2013.

[4] L. G. Christiansen, A. L. Brooks, E. P. Brooks, and T. Rosenørn, "A virtual dressing room for people with Asperger's syndrome," in *Proc. International Conference on Universal Access in Human-Computer Interaction*, 2014, pp. 163-170.

[5] N. Magnenat-Thalmann, H. Seo, and F. Cordier, "Automatic modeling of virtual humans and body clothing," *J. Comput. Sci. Technol.*, vol. 19, no. 5, pp. 575-584, 2004.

[6] C. Kuster, *et al.*, "Towards next generation 3D teleconferencing systems," in *Proc. 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video*, 2012, pp. 1-4.

[7] S. Orts-Escolano, *et al.*, "Holoportation: Virtual 3d teleportation in real-time," in *Proc. the 29th Annual Symposium on User Interface Software and Technology*, 2016, pp. 741-754.

[8] T. Fredriksson and U. Pettersson, "Severe psoriasis — Oral therapy with a new retinoid," *Dermatology*, vol. 157, no. 4, pp. 238-244, 1978.

[9] D. Pinkel, "The use of body surface area as a criterion of drug dosage in cancer chemotherapy," *Cancer Res.*, 1958.

[10] G. A. Knaysi, G. F. Crikelair, and B. Cosman, "The rule of nines: Its history and accuracy," *Plast. Reconstr. Surg.*, 1968.

[11] E. D. Dommasch, D. B. Shin, A. B. Troxel, D. J. Margolis, and J. M. Gelfand, "Reliability, validity and responsiveness to change of the Patient Report of Extent of Psoriasis Involvement (PREPI) for measuring body surface area affected by psoriasis," *Br. J. Dermatol.*, 2010.

[12] R. J. Hunter, M. A. Navo, P. H. Thaker, D. C. Bodurka, J. K. Wolf, and J. A. Smith, "Dosing chemotherapy in obese patients: Actual versus assigned Body Surface Area (BSA)," *Cancer Treat. Rev.*, 2009.
[13] D. Parvizi, *et al.*, "The potential impact of wrong TBSA estimations on fluid resuscitation in patients suffering from burns: Things to keep in mind," *Burns*, vol. 40, no. 2, pp. 241-245, 2014.
[14] B. Allen, B. Curless, and Z. Popović, "The space of human body shapes: Reconstruction and parameterization from range scans," *ACM Trans. Graph.*, vol. 22, no. 3, p. 587, 2003.
[15] B. Allen, B. Curless, and Z. Popovic, "Exploring the space of human body shapes: Data-Driven synthesis under anthropometric control," in *Proc. Digit. Hum. Model. Des. Eng. Symp.*, 2004.
[16] P. Cignoni, M. Callieri, M. Corsini, M. Dellepiane, F. Ganovelli, and G. Ranzuglia, "MeshLab: An open-source mesh processing tool," in *Proc. Eurographics Italian Chapter Conference*, 2008.
[17] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proc. 26th Annu. Conf. Comput. Graph. Interact. Tech.*, 1999, pp. 187-194.
[18] K. M. Robinette, H. Daanen, and E. Paquet, "The CAESAR project: A 3-D surface anthropometry survey," in *Proc. Second International Conference on 3-D Digital Imaging and Modeling (Cat. No.PR00062)*, 1999, pp. 380-386.
[19] H. Seo and N. Magnenat-Thalmann, "An automatic modeling of human bodies from sizing parameters," in *Proc. Symp. Interact. 3D Graph.*, 2003.
[20] H. Seo, F. Cordier, and N. Magnenat-Thalmann, "Synthesizing animatable body models with parameterized shape modifications," in *Proc. Eurographics/SIGGRAPH Symposium on Computer Animation*, 2003, pp. 120-125.
[21] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H. P. Seidel, "A statistical model of human pose and body shape," *Comput. Graph. Forum*, vol. 28, no. 2, pp. 337-346, 2009.
[22] S. Wuhrer and C. Shu, "Estimating 3D human shapes from measurements," *Mach. Vis. Appl.*, vol. 24, no. 6, pp. 1133-1147, 2013.
[23] A. Tsoli, M. Loper, and M. J. Black, "Model-based anthropometry: Predicting measurements from 3D human scans in multiple poses," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, 2014.
[24] M. Hill, S. Streuber, C. Hahn, M. Black, and A. O'Toole, "Exploring the relationship between body shapes and descriptions by linking similarity spaces," *J. Vis.*, vol. 15, no. 12, p. 931, 2015.
[25] M. Q. Hill, S. Streuber, C. A. Hahn, M. J. Black, and A. J. O'Toole, "Creating body shapes from verbal descriptions by linking similarity spaces," *Psychol. Sci.*, vol. 27, no. 11, pp. 1486-1497, 2016.
[26] S. Streuber, *et al.*, "Body talk: Crowdshaping realistic 3D avatars with words," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1-14, 2016.
[27] S. Windhager, K. Patocka, and K. Schaefer, "Body fat and facial shape are correlated in female adolescents," *Am. J. Hum. Biol.*, 2013.
[28] C. Mayer, S. Windhager, K. Schaefer, and P. Mitteroecker, "BMI and WHR are reflected in female facial shape and texture: A geometric morphometric image analysis," *PLoS One*, 2017.
[29] S. Pujades, *et al.*, "The virtual caliper: Rapid creation of metrically accurate avatars from 3D measurements," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 5, pp. 1887-1897, 2019.
[30] A. O. Balan, L. Sigal, M. J. Black, J. E. Davis, and H. W. Haussecker, "Detailed human shape and pose from images," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.
[31] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis, "SCAPE: Shape completion and animation of people," *ACM Transactions on Graphics (TOG)*, vol. 24, no. 3, pp. 408-416, 2005.
[32] J. J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," *Advances in Neural Information Processing Systems*, pp. 1799-1807, 2014.
[33] F. Bogo, A. Kanazawa, C. Lassner, P. Gehler, J. Romero, and M. J. Black, "Keep it SMPL: Automatic estimation of 3D human pose and shape from a single image," in *Proc. European Conference on Computer Vision*, 2016, pp. 561-578.
[34] G. Pavlakos, X. Zhou, K. G. Derpanis, and K. Daniilidis, "Coarse-to-fine volumetric prediction for single-image 3D human pose," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7025-7034.
[35] A. Kanazawa, M. J. Black, D. W. Jacobs, and J. Malik, "End-to-End recovery of human shape and pose," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7122-7131.
[36] R. Ball and J. F. M. Molenbroek, "Measuring Chinese heads and faces," in *Proc. 9th Int. Congr. Physiol. Anthropol. (Hum. Divers. Des. Life)*, January 2008, pp. 150-155.
[37] G. K. L. Tam, *et al.*, "Registration of 3d point clouds and meshes: A survey from rigid to Nonrigid," *IEEE Trans. Vis. Comput. Graph.*, 2013.
[38] O. V. Kaick, H. Zhang, G. Hamarneh, and D. Cohen-Or, "A survey on shape correspondence," *Eurographics Symp. Geom. Process.*, vol. 30, no. 6, pp. 1681-1707, 2011.
[39] X. Li and S. S. Iyengar, "On computing mapping of 3D objects," *ACM Comput. Surv.*, vol. 47, no. 2, pp. 1-45, 2014.
[40] W. Feller, *An Introduction to Probability Theory and Its Applications: Volume One*, John Wiley & Sons, 1950.
[41] F. L. Bookstein, "Principal warps: Thin-Plate splines and the decomposition of deformations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 6, pp. 567-585, 1989.
[42] L. Pishchulin, S. Wuhrer, T. Helten, C. Theobalt, and B. Schiele, "Building statistical shape spaces for 3D human modeling," *Pattern Recognit.*, vol. 67, pp. 276-286, 2017.
[43] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Stat. Soc. Ser. B*, 1996.
[44] X. Xiong and F. L. Torre, "Supervised descent method and its applications to face alignment," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, 2013.
[45] L. A. Jeni, J. F. Cohn, and T. Kanade, "Dense 3D face alignment from 2D videos in dense 3D face alignment from 2D videos in real-time," in *Proc. IEEE International Conference on Automatic Face & Gesture Recognition and Workshops*, May 2015.
[46] J. Domjanic, M. Fieder, H. Seidler, and P. Mitteroecker, "Geometric morphometric footprint analysis of young women," *J. Foot Ankle Res.*, 2013.
[47] J. Domjanic, H. Seidler, and P. Mitteroecker, "A combined morphometric analysis of foot form and its association with sex, stature, and body mass," *Am. J. Phys. Anthropol.*, vol. 157, no. 4, pp. 582-591, 2015.

Alessio Gallucci was in born August 1991, in Moncalieri, Turin, Italy. He obtained his Bachelor and Master in Applied Mathematics at Politecnico di Torino, Italy. He is currently pursuing his PhD at Eindhoven University of Technology, The Netherlands. His current research interests include machine learning, digital image processing and pattern recognition.

Dmitry N. Znamenskiy was born in 1976 in Moscow, USSR. He obtained Master in Mathematics at Moscow State University and PhD at Vrije Universiteit in Amsterdam, studying interacting stochastic processes on large (random) graphs. At Philips Research since 2005 working on Image and Signal processing, Machine Learning, 3D Registration, analysis and design of 3D information systems.

Milan Petković is the head of the Data Science department in Philips, which creates AI innovation for Philips businesses. He is also a part-time full professor at the Eindhoven University of Technology and a Vice President of the Big Data Value Association. He has published more than 50 journal and conference papers as well as several books including a recent Springer book on Data Science in Healthcare.