3DBodyTex: Textured 3D Body Dataset

Alexandre Saint[1], Eman Ahmed[1], Abd El Rahman Shabayek[1], Kseniya Cherenkova[1,2], Gleb Gusev[2], Djamila Aouada[1], and Björn Ottersten[1]

[1] SnT, University of Luxembourg <firstname.lastname@uni.lu>
[2] Artec Europe S.à.r.l. <gleb@artec-group.com>

Abstract

In this paper, a dataset, named 3DBodyTex, of static 3D body scans with high-quality texture information is presented along with a fully automatic method for body model fitting to a 3D scan. 3D shape modelling is a fundamental area of computer vision that has a wide range of applications in the industry. It is becoming even more important as 3D sensing technologies are entering consumer devices such as smartphones. As the main output of these sensors is the 3D shape, many methods rely on this information alone. The 3D shape information is, however, very high dimensional and leads to models that must handle many degrees of freedom from limited information. Coupling texture and 3D shape alleviates this burden, as the texture of 3D objects is complementary to their shape. Unfortunately, high-quality texture content is lacking from commonly available datasets, and in particular in datasets of 3D body scans. The proposed 3DBodyTex dataset aims to fill this gap with hundreds of high-quality 3D body scans with high-resolution texture. Moreover, a novel fully automatic pipeline to fit a body model to a 3D scan is proposed. It includes a robust 3D landmark estimator that takes advantage of the high-resolution texture of 3DBodyTex. The pipeline is applied to the scans, and the results are reported and discussed, showcasing the diversity of the features in the dataset.

1. Introduction

Modelling the 3D body shape has applications in fashion, ergonomics, online retail, entertainment and other industries. As 3D scanners have become mainstream, several datasets of 3D body scans have been collected and proposed for research or commercial purposes [7, 42, 20, 27, 9, 40, 11, 30]. The content of the datasets varies depending on the aim and the method of acquisition. Examples

Figure 1. Sample scans of the 3DBodyTex dataset.

of varying characteristics include the number of subjects and poses, the resolution of the data, and the accompanying attributes such as landmark annotations and anthropometric measurements. Although all these datasets contain the 3D shape information, in the form of triangle meshes or point clouds, none provide clean high-resolution texture information. However, the texture information and the 3D human shape are intimately related and complementary. The two pieces of information combined provide more dimensions to model and process, *e.g.* [9, 26, 16]. This richer information is desired in several scenarios. First, real world applications dealing with the 3D human body shape [6, 22, 7, 18, 25, 35, 33] use cheap and low-end scanning systems that provide noisy and low-quality data, usually from a single view. Applications of this type require accurate 3D models to interpret and denoise the data accordingly. To develop such accurate models, higher quality datasets containing both 3D geometric and texture information are required. Moreover, fundamental problems in computer vision involve modelling the human body in 2D, ranging from body pose and shape estimation from images [37, 22, 17, 8] to action recognition [29]. In this case, large amounts of high-quality 2D data with ground truth an-

notations, such as body landmarks, are crucial. Higher quality 3D body scan datasets with texture information allow to generate 2D content at will, with natural-looking results and in a controlled environment. Furthermore, compared to 2D, the computational complexity of analysing 3D data is high, as observed, for example, in recent geometric deep learning techniques [12, 4]. A way to alleviate the burden of processing 3D data directly is to substitute the analysis with 2D methods which are always a step ahead or to consider the fusion of 2D and 3D data to improve the overall performance [2, 15]. For example, coupling 3D shape and texture information allows to generate realistic 2D content that can be analysed through the lens of fast and proven 2D models [37]. To investigate the possibilities of this approach, a higher-quality dataset of 3D body scans with texture information is crucial.

This paper presents 3DBodyTex, a dataset of *high-quality static 3D body scans with high-resolution texture information.* 3DBodyTex complements existing datasets of static 3D body scans and adds another dimension to the data by incorporating high-quality texture information. The advantage of the texture, and its complementary role to the 3D geometric data, is demonstrated by proposing a fully automatic pipeline for fitting a body shape model to a 3D scan. The pipeline includes a novel robust 3D landmark detector and pose estimator exploiting the texture information. The pipeline works automatically even in challenging poses with only a single 3D scan, as opposed to methods working on scan sequences [43, 41]. Thus, the main contributions of this paper are to:

- provide a high-quality dataset of 3D body scans in close-fitting clothing with high-resolution texture information and high-resolution watertight meshes,
- show an example on how the texture information, absent from all other related datasets, enables novel analysis and modelling with 2D and 3D body shape data,
- present a novel robust and fully automatic pipeline for body model fitting to a 3D scan in challenging poses using the high-quality texture information of 3DBodyTex.

The remainder of this paper is organised as follows. The related datasets are presented and discussed in Section 2. In Section 3, the proposed 3DBodyTex dataset is described and its features highlighted. Section 4 presents a novel automatic pipeline for fitting a 3D body model. It illustrates how the unique texture feature of 3DBodyTex can be exploited in a 3D body modelling task to achieve full automation, even with challenging poses, as opposed to existent approaches. The experimental results of the fitting method are presented in Section 5, showcasing the diverse features of 3DBodyTex. Finally, promising usage of 3DBodyTex to further advance the field of computer vision is discussed in Section 6.

2. Related datasets

There are multiple datasets of static 3D body scans available to the research community. Some are freely available for research [7, 42, 20, 27, 9, 40, 11], others are commercial such as [30]. They are reviewed and discussed below with their advantages and disadvantages. Their purpose and common usage scenarios are highlighted. Table 1 compares the available datasets. The first column of Table 1 is the proposed 3DBodyTex dataset to be released freely to the research community. There also exist some related body scan datasets of scan sequences [28, 10] discussed briefly at the end of the section.

CAESAR [30] is a dataset of 3D laser scans of people of the "European and North American population". It is a commercial dataset primarily intended for the industry although also used in research. There are scans of about 2400 North American males and females and 2000 European subjects in three poses. The dataset comes with comprehensive body measurements taken manually. Those help improve the design of products adapted to today's bodies. Further assets include anthropometric landmarks detected automatically and demographic information on the participants. In the literature, CAESAR has been widely used for statistical body shape modelling [7, 25, 42]. The raw data comes in the form of meshes with texture information. The meshes have a fine resolution but contain holes. The texture is of low resolution, tainted by body markers, dark and unevenly illuminated.

SCAPE [7] provides a dataset of registered meshes of a single male person in 70 poses. The dataset is popular to learn the pose deformations of body shape models when the pose deformations are decoupled from the shape deformations, as in the seminal SCAPE [7] body model. It is also popular for evaluating 3D mesh registration and correspondence methods. As the mesh topology is irregular, with thin and elongated triangles, it is prone to numerical instabilities. Moreover, the registrations have artefacts such as flipped triangles and surface sliding in smooth areas. The accuracy of the vertex correspondence between poses is thus uncertain. The body shape is provided without texture.

SPRING [42] provides a dataset of about 3000 meshes of males and females in a single standard "A" pose registered on the CAESAR [30] laser scans. The meshes in dense correspondence are suitable to learn the space of human shape variation [5]. The mesh topology of SPRING is the same as for SCAPE [7], enabling both datasets to be used in conjunction to learn an articulated body shape model with decoupled pose and shape deformations [7, 22]. The authors of SPRING propose to learn semantic parameters of the human shape space. As in SCAPE, SPRING contains regis-

	3DBodyTex	SCAPE	SPRING	MPI	MPII	FAUST	K3D-hub	TOSCA	CAESAR
High-quality texture	✓	✗	✗	✗	✗	✗	✗	✗	✗
Landmarks	✓	✗	✗	✗	✗	✗	✓	✗	✓
Regular topology	✓	✗	✗	✓	✓	✗	✓	✓	✓
With raw scans	✓	✗	✗	✓	✓	✗	✓	n/a	✓
# poses	$\leq$35	70	1	$\leq$ 35	1	30	5	$\leq$ 20	3
# subjects	200	1	≈3000	114	≈4300	10	50	3	4400
Watertight raw scans	✓	n/a	n/a	✗	n/a	✗	✗	n/a	✗
# scans	400	n/a	n/a	520	n/a	300	250	n/a	≈12000
# vertices	≈300k	n/a	n/a	180–450k	n/a	≈180k	≈150k	n/a	≈150k
# triangles	≈600k	n/a	n/a	n/a	n/a	≈300k	≈300k	≈90k	≈300k
With registered template	✓	✓	✓	✓	✓	✓(*)	✗	✓	✗
# registrations	400	70	≈3000	520	4300	100	n/a	39	n/a
# vertices	6890	12500	12500	6449	6449	6890	n/a	≈45k	n/a
# triangles	13776	25000	25000	12894	12894	13776	n/a	≈90k	n/a
Real people	✓	✓	✓	✓	✓	✓	✓	✗	✓
Freely available	✓	✓	✓	✓	✓	✓	✓	✓	✗

Table 1. Comparison of static 3D body scan datasets.
(*) Only for a subset of the scans.

tration artefacts and the texture information is lacking.

MPI [20] provides a dataset of 520 scans along with the registration of a common template. There are 114 different persons, each in a subset of 35 poses. The dataset is commonly used to learn statistical body shape models with correlated pose and shape deformations. The raw scans are high-resolution point clouds with between 180k and 450k points. Large holes are present on horizontal areas of the bodies because of the single angle of view of the range scanner. This happens, for example, on the back and the chest when the upper body is bent at 90 degrees. Protruding hair is only approximately cropped on the registered meshes. As a side effect, part of the head is cropped unrealistically. Landmarks and texture information are not provided.

MPII Human Shape [27], similarly to SPRING [42], provides a few thousand meshes registered on the CAESAR [30] dataset. The dataset is the result of evaluating different approaches to build statistical body shape models. It comes in three variants with different posture normalisation strategies [27]. The mesh topology is the same as MPI [20] and can be used to complement the latter for learning the shape space. Again, the texture information is lacking.

FAUST [9] is a dataset of 300 scans of 10 people in 30 different poses. The dataset is meant as a benchmark for 3D mesh registration. The raw scans are of high resolution with about 180k vertices and over 300k triangles. Large holes are present in occluded regions and in the extremities of the body. As part of the training data, 100 scans, out of the 300, are provided with a reference registration of a common template mesh. The registration is quantified and more accurate than most other datasets because it was produced from dense texture patterns painted on the bodies. Unfortunately, these patterns taint the appearance of the bodies unrealistically, making the texture information only usable as an accurate spatial reference for tasks such as mesh registration. In any case, the texture information is not provided. As such, the dataset is challenging as a mesh registration benchmark but it is not optimal for applications such as 3D shape modelling and analysis or the generation of 2D data from 3D models. Nonetheless, FAUST does show that texture information is important to model the shape more accurately.

The Kinect-based 3D human body (K3D-hub) dataset [40] is made of 250 low-quality human body scans acquired with the Microsoft Kinect sensor. It contains 50 different subjects in 5 different poses. Due to the limitations of the acquisition hardware, the scans are noisy, the body shape features are smoothed out and the meshes contain holes. This dataset was proposed with the aim to advance the development of methods processing low-quality 3D body shape data. Some body landmark annotations are provided but the texture information is not part of the dataset.

TOSCA [11] contains 3D non-rigid shapes in a variety of poses. It is widely used in experiments of non-rigid shape retrieval, similarity and correspondence [23, 31]. One female and two male characters are available in 7, 12 and 20 poses, respectively. The meshes are clean but not closed.

They were generated synthetically. The shape deformations between poses are thus not representative of the human body shape. Some complex details are represented such as the cavity of the mouth, ears and nostrils, which are not needed in applications of shape modelling dealing with the outer appearance. The synthetic shapes do not come with any mapped texture information.

Some datasets provide 4D body scans (Dyna [28], D-FAUST [10]), *i.e*. sequences of scans recorded over time at high frame rates. For these datasets the focus is not on the quantity of subjects nor the natural body texture, but rather the body pose variation and the soft-tissue motion over time [28]. As a result, the number of subjects is limited, even though the number of poses can be quite high. Texture is not provided in these datasets but, as for FAUST [9], D-FAUST [10] shows that texture information, combined with the 3D shape, provides crucial complementary information for shape modelling.

All the datasets of static 3D scans mentioned above have been exploited for 3D data analysis such as body shape modelling, segmentation, shape similarity and shape correspondence. The 3DBodyTex dataset, presented in Section 3, proposes richer data with both high-resolution texture information and 3D shape. Such data offers new possibilities for 3D shape modeling using the complementary texture information, either directly in 3D or in combination with 2D methods.

3. Proposed 3DBodyTex dataset

The proposed 3DBodyTex dataset contains 3D scans combining unique high-quality and high-resolution texture information with detailed 3D body shape. Examples are given in Figure 1. The characteristics of 3DBodyTex are highlighted and discussed below, and compared in Table 1 to the state-of-the-art datasets discussed in Section 2.

3DBodyTex contains high-resolution textured static 3D scans of people in close-fitting clothing. There are in total 400 scans of 200 different subjects captured in at least two poses: a standard "U" pose and one other pose among a fixed set of 35 poses described below. Some subjects are represented in more than two poses. Two example poses are presented in Figure 2.

The raw scans were captured with the Artec Shapify Booth 3D full-body scanner [1]. The scans are high-resolution meshes with high-quality texture information. The texture is evenly illuminated and reflects the natural appearance of the body. No landmarks or markers taint the colour. The raw meshes are watertight. They contain 300k vertices and 600k triangles with a regular topology. They are provided unprocessed. Only the faces have been blurred for anonymisation. The highly-detailed meshes can be further adapted in resolution and topology to best fit target applications.

Figure 2. Example poses of the 3DBodyTex dataset for two different subjects. First and third: The standard "U" pose common to all subjects. Second and fourth: Examples of varied poses differing between subjects.

As far as could be found, the proposed dataset is the first to contain high-quality texture information. Texture of this quality is indeed lacking in datasets freely available to the research community, and is found only with limited quality in the commercial CAESAR [30] dataset (see Section 2). Texture information is crucial in many applications because it captures an essential dimension of real body data. Combined with the 3D geometric information, it allows to better analyse and represent the body shape. It can also be used to generate high quantities of varied synthetic content with a realistic appearance. This is the approach used in SURREAL [36], a dataset of 2D renderings of a lit and textured body model over real background scenes. However, the texture used for minimal and close-fitting clothing was generated from the limited quality CAESAR [30] dataset, resulting in images of limited quality and realism. Moreover, the body shape was generated from a textured 3D body shape model. This kind of model smooths out the fine details that are present in the raw scans of 3DBodyTex. Thus, 3DBodyTex can be used to create more realistic and higher quality virtual scenes.

The provided mesh resolution is very high with 300k vertices and 600k triangles. The mesh topology is watertight and regular. This is in contrast to most datasets. All the datasets providing only registered meshes [7, 42, 27] have a resolution one order of magnitude lower. The FAUST [9] and CAESAR [30] datasets attain high resolution with the raw meshes but those contain holes. In FAUST, the local resolution varies unpredictably from fine to coarse. The MPI [20] dataset also provides similar resolution but on point clouds only. Moreover, holes are present due to the fixed angle of view of the acquisition system. The high resolution and the absence of holes of 3DBodyTex provides fine details of the body shape globally without any missing information. This is currently not available in related data sources. This provides accurate and complete information for applications such as body shape modelling or 2D content generation.

Subjects were scanned in a standard "U" pose and in a subset of 35 other varied poses. In contrast to existing

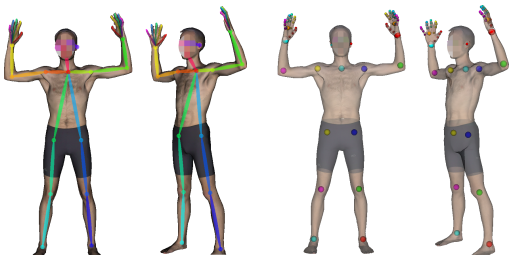

Figure 3. The body landmarks of the COCO dataset [24] and the hand landmarks of [34]. Left half: The 2D landmarks estimated by the 2D detector [37, 14, 34] in two different views. Right half: The 3D landmarks estimated robustly from all the 2D viewpoints by the proposed method (Section 4.1.2). The coloured spheres inside the translucent bodies indicate the landmark positions in 3D. As expected, the body joints lie all inside the body volume. The finger joints are finely estimated.

datasets, the standard pose is the "U" pose instead of the "A" pose. The advantage is that the whole body surface is exposed to the scanner. The shape of the underarms is not approximated as in datasets using the "A" pose. The other pose variations were assigned randomly to scanned participants. Those poses are varied as each subject performs them differently. There are no touching body parts; therefore, no hidden areas on the scans. This makes 3DBodyTex a rich and reliable dataset of human poses and shapes, as required by a large set of 3D modelling and analysis applications.

Furthermore, the number of poses of the datasets can be augmented by using shape deformation techniques such as deformation transfer [35, 33]. This is especially suited to applications that can balance the trade-off between realistic shape deformations and having a higher number of poses. The high-quality texture of 3DBodyTex is crucial in having the reshaped scans keep a natural-looking appearance. This is harder to obtain with the related textureless datasets discussed in Section 2.

As presented above, the poses are chosen to expose the body surface to the scanner as much as possible. Long hair is held back or covered with a cap to expose the head, the neck, the shoulders and the back. The standard "U" pose exposes the underarms. Most of the subjects are scanned in close-fitting sports clothing (tight shorts and sports bras). Some scans come in more challenging conditions with loose hair or additional clothing such as socks or close-fitting T-shirts.

Additionally, reference body landmarks are provided. These include the main body joints, the ears, the eyes, the nose (from the COCO dataset [24]) and some keypoints of the hands [34], see Figure 3. The landmarks were detected automatically by the method presented in Section 4.1 on the raw dataset without blurred faces. They contain some inevitable estimation errors. The knuckles of the hands have higher noise but all other landmarks are reliably and systematically detected. They allow, for example, the accurate recovery of the 3D pose of the scans, as presented in Section 4 for the shape fitting with the body model. They can also be used as reference points for applications such as body segmentation and shape correspondence.

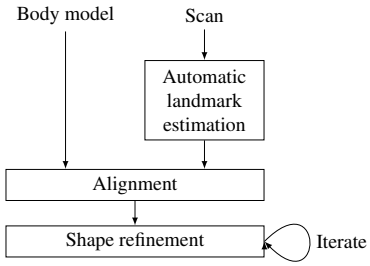

Figure 4. Proposed automatic body model fitting pipeline.

Finally, the registration of a common mesh template to all the scans is provided. This establishes a dense correspondence between the scans. The template mesh has a clean and regular quad-based topology with 6890 vertices and 13776 triangles. It is symmetric in the sagittal plane of the body, making it suitable, for example, for precise texturing with garments and for the definition of topological constraints.

4. Proposed automatic fitting to a 3D scan

The unique high-resolution texture information of 3DBodyTex allows the fitting of a body model to the scans, even in challenging poses, in a fully automatic pipeline (Figure 4). In contrast, existing methods [19, 38, 39, 28] assume manual landmark annotation. The pipeline presented in this section improves upon related works striving towards automation. It does not make any assumption on the input pose [21, 40, 28], *e.g.* the "A" pose. It also does the full shape fitting instead of fitting only the pose [32].

The pipeline relies on an existing method for 2D body landmark estimation. In this work, OpenPose [37] is used to exploit the realistic and textured body data provided by 3DBodyTex. This method combines several attractive attributes: runtime efficiency; systematic placement of landmarks learned from manual annotations; specific to the human body; detection of skeleton and facial/hands landmarks jointly; robustness to left/right ambiguity from front and back views. Similar 2D methods could be used but 3D methods gathering those attributes are, as far as could be found, not available [12, 4]. The usage of a fast and automatic landmark estimator allows to scale the proposed method to large datasets.

The proposed pipeline consists of three main steps. First, the 3D scan is projected from different viewpoints onto images. The 2D body shape stays realistic thanks to the high-

resolution texture information and the scans of real people. Second, the 3D positions of key body landmarks are estimated. The 2D landmarks are detected independently in each 2D view [37, 14, 34], then aggregated robustly into 3D positions. Finally, a parametric body shape model is fitted to the scan using the estimated 3D landmark positions and the scan surface. The method relies on a pretrained body model that can generate the 3D shape of a human body as a triangle mesh from some pose and shape parameters. SCAPE [7] is used in this work but any equivalent model would work. The proposed fitting scheme thus takes advantage of both texture and 3D shape information from the scan. Estimating the landmarks in 2D from multiple viewpoints enables the correct estimation of challenging poses containing both occlusion and ample flexing of the body parts.

The rest of this section details the pipeline with, first, the automatic estimation of the landmarks, and then, the formulation and solving of the body model fitting.

4.1. Body landmarks

The body landmarks serve as guides for fitting the pose of target scan. A correspondence is thus required between the landmarks on the body model and the scan. The N_l considered landmarks are the main body joints and some key anatomical points on the face and the hands (Figure 3).

The landmarks are defined manually on the body model. It is a preparation step done once and for all independently of the scans to be fitted. On a target scan, the landmarks are detected automatically. The definition of the landmarks on the body model and the automatic detection process on a scan are described below.

4.1.1 Landmarks on the body model

The 3D landmark positions of the body model, $l \in \mathbb{R}^{N_l \times 3}$, are obtained by regressing the positions of the N_v vertices of the mesh, $v \in \mathbb{R}^{N_v \times 3}$,

$$l = Lv. \tag{1}$$

Each row of the regression matrix $L \in \mathbb{R}^{N_l \times N_v}$ corresponds to a landmark. Each landmark is defined by one vertex or several vertices in a close neighbourhood, resulting in sparse rows in L. For example, a body joint is defined by the strip of vertices surrounding it such that the regression matrix outputs its centre of mass. A point landmark, such as the tip of the nose, is defined by a single vertex or by the set of vertices in a patch centred on the key point.

4.1.2 Automatic landmark detection

Because of the fine texture information of the 3D scans in 3DBodyTex, accurate landmark positions can be detected automatically even in challenging poses. First, the scan is projected from several views on images with a white background using a perspective camera model. The camera parameters and the depth of the projected points are known. The 2D landmarks are detected automatically in each view using an off-the-shelf 2D landmark detector [37] used as is without fine tuning. This is illustrated with two viewpoints in Figure 3. Using the depth information, the 3D rays joining the camera to the detected landmarks are identified independently in each view.

The 3D position of a landmark is estimated by robustly aggregating candidate positions from the different views. To define the candidate landmark locations, the intersection points of the rays with the scan are computed. For surface landmarks (ears, eyes, nose and base knuckles of the hands), the intersections are all retained as possible candidates. For body joints, the mean of each pair of subsequent intersections is computed and retained as a candidate position. The final landmark locations are estimated in a robust way by taking the median of the candidate locations. The estimation of the 3D position is thus resilient to bad detections in some of the views provided enough views are generated. Figure 3 shows the result of the 3D landmark estimation.

The views are generated systematically. The number of views and the viewpoints can be adapted to a specific application or to the sensitivity of the 2D detector. Overall, the combined texture information and the real appearance of the data allow for an automatic and efficient detection of the landmarks in difficult poses.

4.2. Formulation of the fitting problem

The fitting of the body model to a scan happens in two steps (Figure 4). First, the model is aligned to the scan by using the body landmarks as guides. Second, the shape of the model is refined to fit the scan tightly, assuming scans in minimal form-fitting clothing. To enforce a tight fitting constraint, a correspondence between the points of the model and the scan is computed.

The problem is formulated as the minimisation of the energy function $E(y, r, s)$. The variables are the vertex positions $y \in \mathbb{R}^{N_v \times 3}$, the pose $r \in \mathbb{R}^{N_p \times 3}$ and the shape parameters $s \in \mathbb{R}^{N_s}$. The pose is encoded with N_p axis-angle 3D vectors, expressing the global orientation of the model and the relative rotations of the $N_p - 1$ body parts. Since the body shape model is parametric, the vertex positions are themselves function of the pose and the shape parameters, $y = y(r, s)$. The energy is split in several terms

$$E(y, r, s) = w_l E_l(y) + w_v E_v(y) + w_s E_s(y) + E_m(y, r, s), \tag{2}$$

with relative scalar weights w_*. The landmark fitting is ensured by $E_l(y) = \|Ly - \hat{l}\|_2^2$, with $\hat{l}$ the set of detected landmarks and L the landmark position regressor of the body model, as in equation (1). The tight shape fitting is ensured by $E_v = \sum_{(y,z)\in C} \|y - z\|_2^2$, which encourages pairs

(y, z) of corresponding vertices from the body model and the scan to be close. The construction of the set C of correspondences is described in Section 4.3. The body shape is further constrained to stay within reasonable limits of the learned shape space. This is achieved by penalising shape parameters more than three standard deviations away from the mean $E_s = \sum_i \max(0, |s_i|/\sigma_i - 3)^2$. The last term E_m is the regularisation of the mesh with the learned body model. A body model deforms a template body mesh into a mesh with specific body shape and pose from the corresponding parameters, r and s. The particular model considered in this work, SCAPE [7], achieves body shape and pose deformation in two steps: deformation of individual triangles followed by their reconnection to get a watertight mesh. A triangle, with vertex indices (i, j, k) and vertex positions (x_i, x_j, x_k), is deformed by applying some linear transformations $R, Q, S \in \mathbb{R}^{3\times3}$ to the edges (i, j) and (i, k), *i.e.* $\Delta x'_{ij} = RQS\Delta x_{ij}$. The transformations are learned a priori on a database of body scans as part of the training of the body model [7]. The articulation of the skeleton is encoded by $R(r)$ and the pose-dependent deformations by $Q(r)$. Both depend on the pose parameters r. The shape deformations determining the identity of an individual are encoded by $S(s)$ and depend on the shape parameters s. All triangles of a body part share a common R, while Q and S are different for each triangle. After individual deformation of all the triangles, the mesh gets disconnected. To reconnect the mesh, new vertex positions y are sought that form a watertight mesh and that respect as much as possible the fixed individual triangle deformations R,Q and S. This leads to the linear optimisation problem

$$y^* = \arg\min_y \sum_{(i,j)\in E} \|RQS\Delta x_{ij} - \Delta y_{ij}\|_2^2, \qquad (3)$$

with E the set of triangle edges and x the vertices of the template mesh. This transformation and reconnection process forms the SCAPE [7] body model. So the energy term of equation (2) representing the body model regularisation is $E_m(y, r, s) = \sum_{(i,j)\in E} \|R(r)Q(r)S(s)\Delta x_{ij} - \Delta y_{ij}\|_2^2$. However, this term makes (2) nonlinear because all the variables, r, s and y, are optimised for.

4.3. Solving the fitting problem

The alignment and shape refinement phases (Figure 4) are solved sequentially by non-linear least squares minimisation.

During the alignment phase, the shape-focused energy terms in (2) are disabled by setting $w_v = w_s = 0$. The relative weight of the landmark fitting term is set to $w_l = 1$. On top of the pose parameters, the shape parameters s are also optimised for the model to adapt to the scan dimensions.

For the shape refinement phase, all energy terms in (2) are active with $w_l = w_v = w_s = 1$. Several iterations are performed to gradually update the shape parameters s. When the model surface is close to the scan surface, the parameters stall and the fitting is terminated. At each iteration, a vertex correspondence is established between the model and the scan to define the shape fitting term, E_v. The form the correspondence, for each vertex of the model, the closest neighbouring vertex on the scan is selected. The pair of vertices is retained if they are within a threshold distance d, and if their vertex normals are within an angular threshold α. A good initial alignment of the model is assumed to be provided by the alignment phase.

Figure 5. Result of the fitting on a challenging pose. Left: Target scan. Middle: Target scan and fitted model overlaid. Right: Fitted model.

5. Experimental results

To highlight the different features of 3DBodyTex, the automatic body model fitting pipeline of Section 4 is applied to the scans. The results are presented in light of the different attributes of the dataset such as pose, gender and texture.

Figure 5 shows the result of fitting a scan far from the rest "A" pose of the body model. As can be seen qualitatively, the correct posture is recovered correctly even though the method is fully automatic and necessitates no annotation of the body scan.

Figure 6 shows a quantitative evaluation of the fitting on the "U" pose and some further challenging poses of 3DBodyTex. The distance between a scan and the model is reported per vertex by colour mapping the values from 0 cm to 1 cm. The distance from scan to model is defined by the Euclidean distance between each vertex of the scan and the nearest vertex of the body model. As can be observed, the fitting is good on most parts of the body with a fitting within a few millimetres of the scan. The body poses are again well recovered, albeit the absence of any human intervention. Long hair is either held back or contained with a cap. When there is protruding hair, as in the second and third rows in Figure 6, it is well discarded by the body model. The hands are also discarded by the fitting because the body model has clenched fists whereas most of the scans have open hands. Other difficult regions to fit are the chin, the underarms, the lower part of the chest and the crotch. Those regions have sharp curvature where the body model tends to smooth out the transitions.

Overall, the body model is closely fitted to the scans.

When computing the root mean square error on the whole body of each fitted scan and aggregating across all scans, the mean error is 4 mm and the standard deviation 13 mm. The deviation is mostly explained by gaps between the body model and non-skin areas, such as the hair, as well as between the clenched fists of the body model and the open hands of the scans.

Despite the fully-automatic nature of the pipeline, challenging poses can be recovered correctly from a single static scan as can be seen in Figure 5 and 6.

Figure 6. Fitting error for some sample scans of 3DBodyTex in the "U" pose and some non-trivial poses. Left: Target scan. Middle and right: Fitting error as measured from scan to model for the front and back views, respectively. The colour scale varies from 0 cm to 1 cm. Values over 1 cm are coloured white. This is the case for regions discarded by the fitting because not part of the body shape, such as protruding hair or opened hands (since the body model has clenched fists).

6. Conclusion

In this paper, the 3DBodyTex dataset of static 3D body scans with high-quality texture and of high resolution was presented. It was shown how 3DBodyTex complements or supersedes other datasets. As far as could be found, 3DBodyTex is the only dataset of diverse body scans with high-quality texture information and watertight high-resolution meshes. 3DBodyTex will be released freely for academic use. The texture is detailed, evenly illuminated and free of tainting markers or other reference patterns. The raw scans are watertight meshes with a regular and high-resolution topology containing 300k vertices and 600k triangles. Registrations of a common template mesh to all the scans are provided. The template mesh has a clean, regular and symmetric structure, enabling, for example, precise texturing and the definition of physical constraints on the body.

The unique high-quality texture and the high resolution of the scans of 3DBodyTex is of advantage in shape modelling applications. 3DBodyTex is foreseen to be used in 2D, 2.5D and 3D settings. In 2D, it can be used to generate large quantities of natural-looking 2D content. This can be used to develop and evaluate 2D methods or also to bypass the burden of processing 3D data directly. In 3D, it provides realistic data with unique modalities, namely, the texture information and the high-resolution meshes in many combinations of body shapes and poses. Possible applications include segmentation, shape modelling, shape matching, and geometric deep learning, among others.

The richness of 3DBodyTex was illustrated through the task of fitting a 3D body shape model to 3D body scans in challenging poses in a fully automatic pipeline. The high-quality and realistic texture information is taken advantage of to recover difficult body poses. This is done by leveraging robust and efficient 2D methods for body landmark estimation [37]. Additionally, the proposed method scales to large datasets. Overall, 3DBodyTex brings new possibilities to develop novel methods of 3D and 2D shape analysis in computer vision and related fields.

Acknowledgement

This work was funded by the National Research Fund (FNR), Luxembourg, AFR PPP reference 11806282, and by Artec Europe SARL. The authors are grateful to the volunteers for the scanning, to the whole Computer Vision Lab at SnT for collecting the data, and to the contributors of the open source libraries used in this work [13, 3, 37, 14, 34].

References

[1] Artec shapify booth. https://www.artec3d.com/portable-3d-scanners/shapifybooth. Retrieved: 2018-05-06. 4

[2] H. Afzal, D. Aouada, D. Fofi, B. Mirbach, and B. Ottersten. Rgb-d multi-view system calibration for full 3d scene reconstruction. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 2459–2464. IEEE, 2014. 2

[3] S. Agarwal, K. Mierle, and Others. Ceres solver. http://ceres-solver.org. 8

[4] E. Ahmed, A. Saint, A. E. R. Shabayek, K. Cherenkova, G. Gusev, D. Aouada, and B. Ottersten. Deep learning advances on different 3d data representations: A survey, 2018. arXiv preprint. 2, 5

[5] B. Allen, B. Curless, and Z. Popović. The space of human body shapes: reconstruction and parameterization from range scans. *ACM transactions on graphics (TOG)*, 22(3):587–594, 2003. 2

[6] B. Allen, B. Curless, Z. Popović, and A. Hertzmann. Learning a correlated model of identity and pose-dependent body shape variation for real-time synthesis. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 147–156. Eurographics Association, 2006. 1

[7] D. Anguelov, P. Srinivasan, D. Koller, S. Thrun, J. Rodgers, and J. Davis. Scape: shape completion and animation of people. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 408–416. ACM, 2005. 1, 2, 4, 6, 7

[8] 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 *European Conference on Computer Vision*, pages 561–578. Springer, 2016. 1

[9] F. Bogo, J. Romero, M. Loper, and M. J. Black. Faust: Dataset and evaluation for 3d mesh registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3794–3801, 2014. 1, 2, 3, 4

[10] F. Bogo, J. Romero, G. Pons-Moll, and M. J. Black. Dynamic FAUST: Registering human bodies in motion. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, July 2017. 2, 4

[11] A. M. Bronstein, M. M. Bronstein, and R. Kimmel. *Numerical geometry of non-rigid shapes*. Springer Science & Business Media, 2008. 1, 2, 3

[12] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017. 2, 5

[13] L. Campagnola, A. Klein, E. Larson, C. Rossant, and N. P. Rougier. VisPy: Harnessing The GPU For Fast, High-Level Visualization. In K. Huff and J. Bergstra, editors, *Proceedings of the 14th Python in Science Conference*, Austin, Texas, United States, July 2015. 8

[14] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, 2017. 5, 6, 8

[15] K. I. Chang, K. W. Bowyer, and P. J. Flynn. An evaluation of multimodal 2d+ 3d face biometrics. *IEEE transactions on pattern analysis and machine intelligence*, 27(4):619–624, 2005. 2

[16] F. Garcia, D. Aouada, T. Solignac, B. Mirbach, and B. Ottersten. Real-time depth enhancement by fusion for rgb-d cameras. *IET Computer Vision*, 7(5):335–345, 2013. 1

[17] P. Guan, A. Weiss, A. O. Balan, and M. J. Black. Estimating human shape and pose from a single image. In *2009 IEEE 12th International Conference on Computer Vision*, pages 1381–1388. IEEE, 2009. 1

[18] N. Hasler, H. Ackermann, B. Rosenhahn, T. Thormählen, and H.-P. Seidel. Multilinear pose and body shape estimation of dressed subjects from image sets. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1823–1830. IEEE, 2010. 1

[19] N. Hasler, C. Stoll, B. Rosenhahn, T. Thormählen, and H.-P. Seidel. Estimating body shape of dressed humans. *Computers & Graphics*, 33(3):211–216, 2009. 5

[20] N. Hasler, C. Stoll, M. Sunkel, B. Rosenhahn, and H.-P. Seidel. A statistical model of human pose and body shape. In *Computer Graphics Forum*, volume 28, pages 337–346. Wiley Online Library, 2009. 1, 2, 3, 4

[21] D. A. Hirshberg, M. Loper, E. Rachlin, and M. J. Black. Coregistration: Simultaneous alignment and modeling of articulated 3d shape. In *European Conference on Computer Vision*, pages 242–255. Springer, 2012. 5

[22] A. Jain, T. Thormählen, H.-P. Seidel, and C. Theobalt. Moviereshape: Tracking and reshaping of humans in videos. In *ACM Transactions on Graphics (TOG)*, volume 29, page 148. ACM, 2010. 1, 2

[23] V. G. Kim, Y. Lipman, and T. Funkhouser. Blended intrinsic maps. In *ACM Transactions on Graphics (TOG)*, volume 30, page 79. ACM, 2011. 3

[24] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 5

[25] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. Smpl: A skinned multi-person linear model. *ACM Transactions on Graphics (TOG)*, 34(6):248, 2015. 1, 2

[26] O. K. Oyedotun, G. Demisse, A. E. R. Shabayek, D. Aouada, and B. Ottersten. Facial expression recognition via joint deep learning of rgb-depth map latent representations. In *2017 IEEE International Conference on Computer Vision Workshop (ICCVW)*, 2017. 1

[27] L. Pishchulin, S. Wuhrer, T. Helten, C. Theobalt, and B. Schiele. Building statistical shape spaces for 3d human modeling. *Pattern Recognition*, 67:276–286, 2017. 1, 2, 3, 4

[28] G. Pons-Moll, J. Romero, N. Mahmood, and M. J. Black. Dyna: A model of dynamic human shape in motion. *ACM Transactions on Graphics, (Proc. SIGGRAPH)*, 34(4):120:1–120:14, Aug. 2015. 2, 4, 5

[29] R. Poppe. A survey on vision-based human action recognition. *Image and vision computing*, 28(6):976–990, 2010. 1

[30] K. M. Robinette, H. Daanen, and E. Paquet. The caesar project: a 3-d surface anthropometry survey. In *3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on*, pages 380–386. IEEE, 1999. 1, 2, 3, 4

[31] E. Rodolà, S. Rota Bulo, T. Windheuser, M. Vestner, and D. Cremers. Dense non-rigid shape correspondence using random forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4177–4184, 2014. 3

[32] A. Saint, A. E. R. Shabayek, D. Aouada, B. Ottersten, K. Cherenkova, and G. Gusev. Towards automatic human body model fitting to a 3d scan. In *Proceedings of 3DBODY. TECH 2017-8th International Conference and Exhibition on 3D Body Scanning and Processing Technologies, Montreal QC, Canada, 11-12 Oct. 2017*, pages 274–280. Hometrica Consulting, 2017. 5

[33] A. E. R. Shabayek, D. Aouada, A. Saint, and B. Ottersten. Deformation transfer of 3d human shapes and poses on manifolds. In *Image Processing (ICIP), 2017 IEEE International Conference on*, pages 220–224. IEEE, 2017. 1, 5

[34] T. Simon, H. Joo, I. Matthews, and Y. Sheikh. Hand keypoint detection in single images using multiview bootstrapping. In *CVPR*, 2017. 5, 6, 8

[35] R. W. Sumner and J. Popović. Deformation transfer for triangle meshes. *ACM Transactions on Graphics (TOG)*, 23(3):399–405, 2004. 1, 5

[36] G. Varol, J. Romero, X. Martin, N. Mahmood, M. J. Black, I. Laptev, and C. Schmid. Learning from synthetic humans. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2017)*, 2017. 4

[37] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh. Convolutional pose machines. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4724–4732, 2016. 1, 2, 5, 6, 8

[38] A. Weiss, D. Hirshberg, and M. J. Black. Home 3d body scans from noisy image and range data. In *2011 International Conference on Computer Vision*, pages 1951–1958. IEEE, 2011. 5

[39] S. Wuhrer, L. Pishchulin, A. Brunton, C. Shu, and J. Lang. Estimation of human body shape and posture under clothing. *Computer Vision and Image Understanding*, 127:31–42, 2014. 5

[40] Z. Xu, Q. Zhang, and S. Cheng. Multilevel active registration for kinect human body scans: from low quality to high quality. *Multimedia Systems*, pages 1–14, 2017. 1, 2, 3, 5

[41] J. Yang, J.-S. Franco, F. Hétroy-Wheeler, and S. Wuhrer. Estimation of human body shape in motion with wide clothing. In *European Conference on Computer Vision*, pages 439–454. Springer, 2016. 2

[42] Y. Yang, Y. Yu, Y. Zhou, S. Du, J. Davis, and R. Yang. Semantic parametric reshaping of human body models. In *2014 2nd International Conference on 3D Vision*, volume 2, pages 41–48. IEEE, 2014. 1, 2, 3, 4

[43] C. Zhang, S. Pujades, M. Black, and G. Pons-Moll. Detailed, accurate, human shape estimation from clothed 3d scan sequences. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, 2017. 2