Towards Automatic Human Body Model Fitting to a 3D Scan

Alexandre SAINT*[1], Abd El Rahman SHABAYEK[1,2], Djamila AOUADA[1],
Björn OTTERSTEN[1], Kseniya CHERENKOVA[3], Gleb GUSEV[3]
[1] Interdisciplinary Centre for Security, Reliability, and Trust, University of Luxembourg, Luxembourg;
[2] Computer Science department, Faculty of Computers and Informatics, Suez Canal University, Egypt;
[3] Artec Europe S.à.r.l., Luxembourg, Luxembourg

Abstract

This paper presents a method to automatically recover a realistic and accurate body shape of a person wearing clothing from a 3D scan. Indeed, in many practical situations, people are scanned wearing clothing. The underlying body shape is thus partially or completely occluded. Yet, it is very desirable to recover the shape of a covered body as it provides non-invasive means of measuring and analysing it. This is particularly convenient for patients in medical applications, customers in a retail shop, as well as in security applications where suspicious objects under clothing are to be detected. To recover the body shape from the 3D scan of a person in any pose, a human body model is usually fitted to the scan. Current methods rely on the manual placement of markers on the body to identify anatomical locations and guide the pose fitting. The markers are either physically placed on the body before scanning or placed in software as a postprocessing step. Some other methods detect key points on the scan using 3D feature descriptors to automate the placement of markers. They usually require a large database of 3D scans. We propose to automatically estimate the body pose of a person from a 3D mesh acquired by standard 3D body scanners, with or without texture. To fit a human model to the scan, we use joint locations as anchors. These are detected from multiple 2D views using a conventional body joint detector working on images. In contrast to existing approaches, the proposed method is fully automatic, and takes advantage of the robustness of state-of-art 2D joint detectors. The proposed approach is validated on scans of people in different poses wearing garments of various thicknesses and on scans of one person in multiple poses with known ground truth wearing close-fitting clothing.

Keywords: 3D body scanning, 3D body modelling, pose fitting, clothing, automation

1. Introduction

An important application of 3D scanners is the measurement of the human body shape as a 3D mesh for use in healthcare, textile, fashion, ergonomics, fitness and security, among other areas. There exist several successful parametric models that represent the pose and shape of the human body in a realistic way via a relatively small set of parameters [1, 2, 3, 4, 5, 6]. The models are trained on a database of 3D scans of people in close-fitting clothing and exhibiting a large enough variation of body shapes (e.g. [7]) and poses (e.g. [1, 2]).

In real-life scenarios people are usually scanned in arbitrary poses and wearing any type of clothing. One way to further process the data realistically and accurately is to fit a 3D human body model to the scan. This makes it possible to recover an accurate estimate of the pose and of the underlying body shape.

To guide the pose fitting, several methods assume the existence of a correspondence of a sparse set of markers on key anatomical locations of the body model and the target scan. These markers are placed manually or stem from a motion capture system [1, 8, 9, 10]. Other works aim at fitting a 3D body model to 2D images. For example, Bogo et al. [11] automatically detect body joint locations in an image and use a body model constrained to a dataset of example plausible poses. The drawbacks are that this constrained model is not publicly available, is tedious to reproduce and its representational power is limited to the example poses.

Hasler et al. [12] and Balan et al. [13, 14] manually define the landmarks or the initial pose before fitting a 3D body model using silhouette information as a constraint.

The existing literature requires either manual intervention or impractical assumptions to fit a body model to a 3D scan. There is a need to free the automatic pose fitting process from restrictive datasets and tedious preprocessing procedures.

---

* alexandre.saint@uni.lu; +352 46 66 44 5649

This paper proposes a method to automatically fit the pose of a human body model to a 3D scan of a person in an arbitrary pose. The body joints are detected robustly from several 2D views using a conventional 2D joint detector. The human body model is then fitted to the joints to obtain an accurate pose estimate. DeeperCut [15] is used to detect joints on 2D images. The method constitutes a basis for further refined shape fitting.

The remainder of this paper is organised as follows. The considered human body model is described in Section 2. In Section 3, the proposed fitting method is presented, and it is evaluated in Section 4. The paper is finally concluded and directions for future works are discussed in Section 5.

2. Background

The 3D human body model named Shape Completion and Animation of PEople (SCAPE) [1] is used to represent the human body shape and pose. With this parametric body model, the body shape is represented with a triangular mesh of fixed tessellation. SCAPE makes it possible to realistically deform an average human body into different shapes and poses. The model used in this paper is trained on publicly available datasets [1, 16].

2.1. Body skeleton and pose

The human body mesh defined above is segmented into rigid body parts to represent the skeleton (Figure 1). The parts are linked by joints in a tree structure.
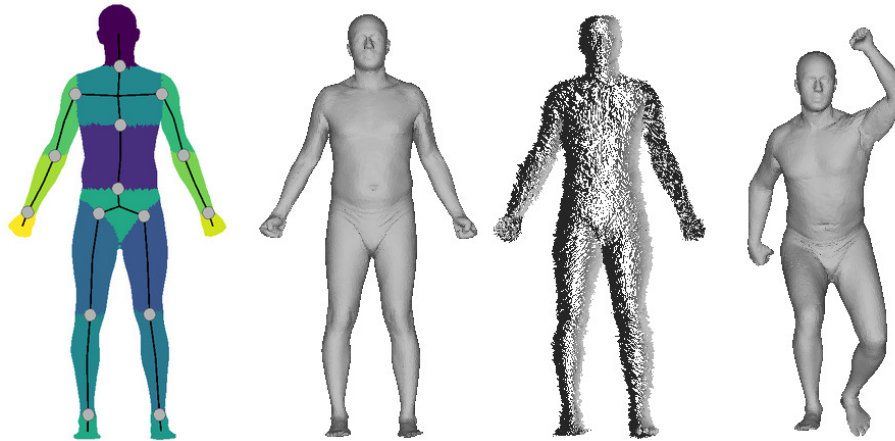

Fig. 1. Illustration of the 3D human body model. From left to right: body parts defining the skeleton; body model in the reference pose; deformed and disconnected triangles (deformation step 1); reconnected triangles (deformation step 2).

The body model is put in a specific pose by rotating each body part independently (Figure 1). The pose is defined by the set of absolute 3D rotation matrices $R = \{R_p\}$ of the P body parts with respect to the reference pose. Similarly, R_k denotes the rotation matrix of the body part associated to the triangle k. Moreover, the rotations of the J joints are defined as the relative rotations from one part to its neighbour by the 3D rotation matrices $\{\Delta R_j\}$. For example, the rotation of joint j, linking parts p and q, is $\Delta R_j = R_p^T R_q$.

2.2. Mesh deformation process

SCAPE proceeds in two steps to deform the reference mesh into a target pose and shape. First, linear transformations, $R_k, Q_k, S_k \in R^{3\times3}$, are applied locally to each triangle k with the equation

$$\Delta y_{k,j} = R_k S_k Q_k \Delta x_{k,j} \qquad (1)$$

where the two 3D vectors representing the edges, $\Delta x_{k,j} = x_{k,j} - x_{k,1}$ for $j = 2,3$, belong to the reference mesh, and, similarly, the edges, $\Delta y_{k,j}$ for $j = 2,3$, belong to the deformed mesh. The triangular deformation process originates from the work of Sumner et al. [17]. The transformations R_k, S_k, Q_k are specific to each triangle k. The rotation R_k articulates the body part to which the triangle k belongs, and is common to all triangles of the body part.

The transformation $S_k(\beta)$ encodes the shape deformation to obtain a specific identity. It depends on the vector of shape parameters β. The matrix $Q_k(R)$ is a refinement of the shape deformation due to the pose R, e.g. the bulging of the biceps for a flexed arm. In this paper, the pose-dependent deformation Q_k depends only on the pose of the two neighbouring parts, as this involves less computations for a comparable deformation quality [1].

Applying the transformations to each triangle independently has the effect of disconnecting the mesh (Figure 1). The second step of the deformation process is thus to reconnect the mesh while preserving the transformations as much as possible. This is done by solving the linear least-squares optimisation problem

$$y = argmin_y \sum_{k,j} \left\| \Delta y_{k,j} - R_k Q_k S_k \Delta x_{k,j} \right\|_2^2 \qquad (2)$$

to find new vertex locations that keep the mesh water-tight while best respecting the transformations applied in equation (1). The new vertex locations are equivalently $\{y_i\}, \{y_{k,j}\}$ or $y \in R^{N \times 3}$, as described in the beginning of Section 2.

3. Proposed automatic fitting scheme

To be fully automatic, the proposed pose fitting scheme relies on the automatic detection of key markers on the body. These markers guide the search for the optimal pose parameters. The considered markers are the joint locations between body parts (Figure 2) because they give reliable information on the overall pose of the body.
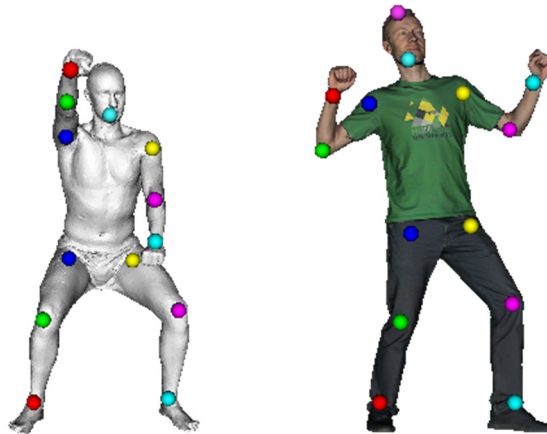

Fig. 2. Detection of joints on the human body. Left: Textureless scan from [16].
Right: 3D scan obtained with Artec Shapify Booth [18].

3.1. Detection of joint locations

The 3D scan is projected onto 2D images from multiple viewpoints. Joint locations are detected in each image using DeeperCut [15]. Figure 2 shows the detected joints on 2D views of two examples scans of the datasets described in Sections 4.1 and 4.2.

From each 2D joint location, the intersections with the 3D scan are found using ray casting. A ray is cast from the camera centre in the direction of the 2D point on the projection plane. Then, the intersections are checked for robustness. If there are exactly two intersection points, the mean of the two points is recorded as the 3D joint location. In all other cases, the joint location is ambiguous, and that joint is discarded for that view. Indeed, with zero or one intersection, the ray does not cross the scan. With more than two intersections, there is occlusion by another body part, and the joint cannot be located reliably. For each joint j, there are V estimates of its 3D location, one in each view.

Joint detectors working on 2D images only are considered to leverage the robustness of the state-of-the-art DeeperCut [15] method. An implementation of DeeperCut is provided by the authors, and comes with pre-trained models or can be trained on commonly available datasets [19]. However, any equivalent 2D joint detector can be used, provided that the detected joints match those of the body model. Moreover, the proposed method can be extended to any joint or marker detector in 3D with minimal adaptation.

3.2. Pose fitting

To fit the pose of the scan, the least-squares solution to equation (1) is found, solving for the vertex locations $y \in R^{N \times 3}$ and the rotation parameters R, while constraining the body pose to match the configuration of the detected 3D joint locations.

The pose fitting is independent of the resolution of the mesh since it only considers the detected joint locations. High-resolution data is thus handled efficiently.

4. Experimental evaluation

The proposed pose fitting algorithm is evaluated quantitatively and qualitatively on two datasets of full-body 3D scans of people in multiple poses. The first dataset contains one person with close-fitting clothing (Section 4.1). The second dataset contains people wearing casual clothing in multiple poses (Section 4.2). The ground-truth poses are available on the first dataset only.

The datasets are considered independently using a common evaluation procedure. The fitting method of Section 3 is applied to each scan of each dataset. The output is a human body model in a specific pose. The pose is defined by the absolute rotations $\{R_p\}$ of the body parts as described in Section 2.1. Following the procedure described in Section 3.1, the joints are detected from 9 views. The views span 180 degrees at equal angular intervals. The middle view faces the front of the body trunk. This setting was found experimentally to provide enough good joint detections. Even though the first dataset (Section 4.1) does not contain texture information, it has been observed that DeeperCut [15] is able to detect the body joints relatively well, as illustrated in Figure 2.

On the first dataset, the pose estimates obtained by the fitting procedure are compared to the ground truth by measuring the geodesic distance between rotation matrices of corresponding body parts [19].

4.1. Pose dataset with close-fitting clothing

The first dataset comes from SCAPE [1]. It contains 3D scans of one person in close-fitting clothing in 70 different poses, including the reference pose. This dataset constitutes part of the data on which the SCAPE model used in this paper has been trained. The scans come registered on the template mesh of Section 2.

The ground-truth poses are required for evaluation. Since they are not provided with the dataset, they are estimated from the data. The body parts are assumed rigid enough such that it is possible to accurately estimate the rotation between poses given the vertex correspondence. This is a reasonable approximation of the reality as each body part is mostly rigid. So, for each scan, the best rotation of each body part with respect to the reference scan is computed by aligning the pair of point clouds as described in [20].
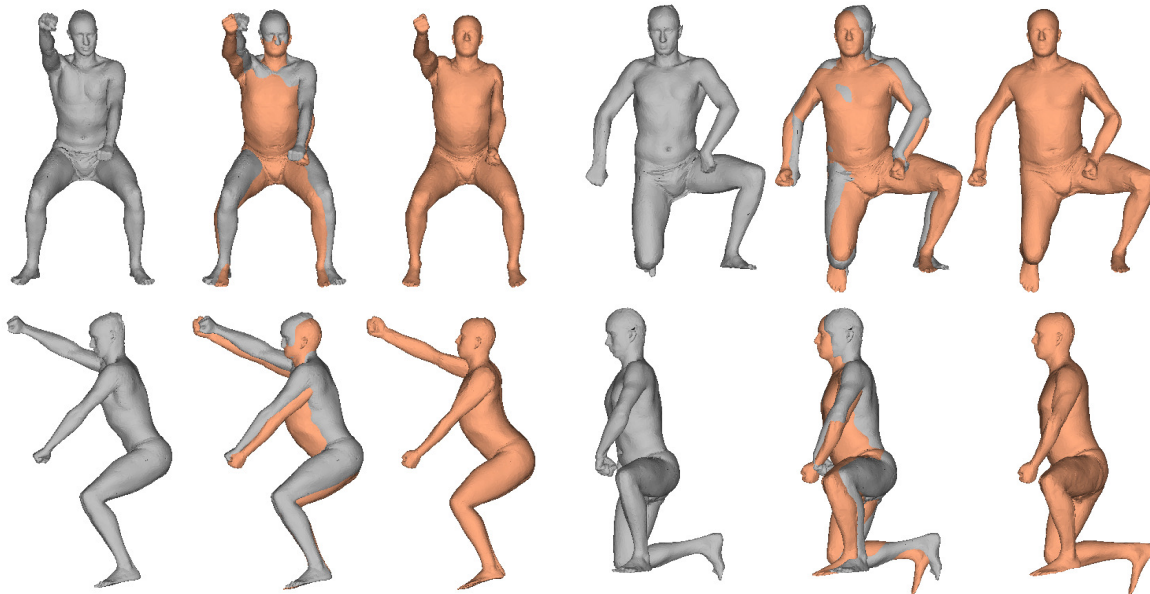

Fig. 3. Results of the proposed pose fitting methods for two scans of the SCAPE pose dataset. Grey: target scan. Orange: fitted scan. Left: scan index 25. Right: scan index 68.

Figure 3 shows the result of pose fitting for two scans of the dataset. The fitted models match the pose of the target scan. The shape does not match because no shape fitting is performed, but this does not prevent to have a correct pose fitting. As can be seen, the overall pose is well recovered. The trunk and the nearest body parts are also well posed. The body parts at the end of the skeleton tree are more difficult to align because the joint locations alone cannot constrain them. For example, the fists in Figure 3 are not well oriented, even though the arms are aligned.
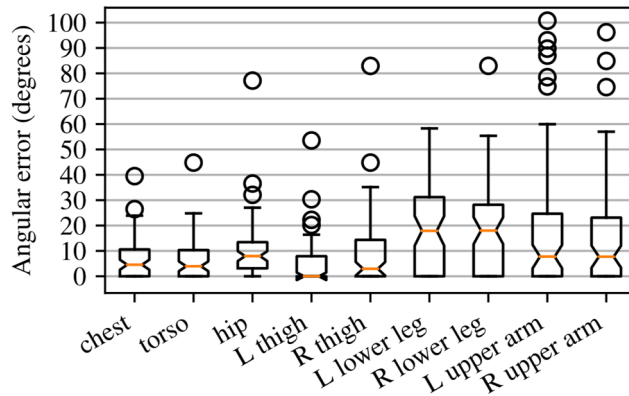
Fig. 4. Distribution of pose errors per part. The error is the geodesic distance between
the ground-truth and estimated rotation matrices of the part on the rotation matrix manifold.
The notches indicate the median values. End body parts are not considered by the algorithm.

Figure 4 shows, for each body part, the distribution of the pose error across the 70 fitted scans. The orientation error of the body parts situated on the trunk is under 10 degrees. The parts extending from the trunk (legs, arms, neck, head) exhibit higher orientation error. The end body parts in the skeleton tree are omitted in Figure 4 because the proposed fitting algorithm focusses on registering the joints. Consequently, the orientation of the end body parts is not fully constrained. This also explains why the lower legs and upper arms exhibit higher pose errors; they are nearer to end body parts and their pose is less constrained than the trunk. It is expected that the pose of the end parts will be corrected by a subsequent shape fitting procedure that will be considered in future works.

4.2. Custom pose dataset with casual clothing

To create the second dataset, scans of different persons in multiple poses wearing casual clothing were acquired using Artec Shapify Booth full-body scanner [18]. The mesh tessellation is random and contains about 300,000 vertices and 600,000 triangles. The ground-truth poses are unknown and difficult to estimate. Therefore, only a qualitative evaluation is performed.
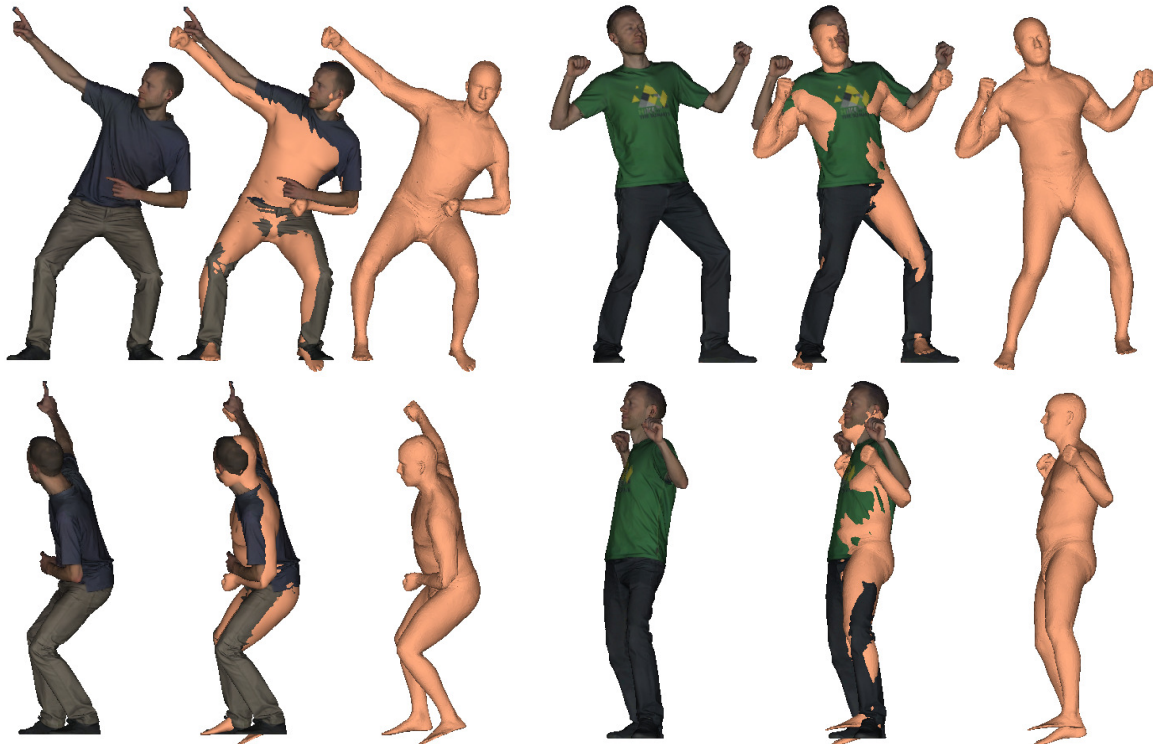

Fig. 5. Results of the proposed pose fitting method on two scans of the custom clothed dataset
acquired with Artec Shapify Booth [18]. Textured scan: target scan. Orange: fitted scan.

Figure 5 shows the result of pose fitting on two example scans of the dataset. Qualitatively, the results are as good as in the previous experiment (Section 4.1). The overall pose is well recovered. The body parts between the joints are well aligned. In contrast, the end body parts show the misalignment due to being constrained by one joint only. As explained in Section 4.1, a subsequent shape fitting, to be considered in future works, is expected to correct the pose of the end body parts.

5. Conclusion

An approach to automatically fit the pose of a 3D scan with a 3D parametric human body model was presented. The automatic algorithm combines SCAPE [1], a state-of-the-art human body model, and DeeperCut [15], a robust 2D joint detector. Joint locations are detected from several 2D views for robustness. Those are then fitted jointly with the body model to recover an accurate pose estimate. The proposed method allows to automate the pose fitting of large databases of 3D scans without manual intervention, as has been found to be required by previous works. Moreover, the fitting of the joints does not increase in complexity with the resolution of the mesh. Even though the approach was presented with a 2D joint detector, it can be extended to 3D joint or marker detectors with minimal adaptations.

The proposed pose fitting method can be used for bootstrapping the more general task of shape fitting. Shape fitting requires an initial pose estimate close to the target pose. Although the proposed method cannot align end body parts well, it should not be necessary for initialising a shape fitting algorithm, since the latter would usually perform a variant of ICP (Iterative Closest Point) to match the shapes and refine the alignment of the end parts at the same time. If accurate alignment of the end body parts is nonetheless required, the proposed approach could, for example, be improved by performing a variant of ICP between the end body parts of the model and the target mesh.

References

[1] 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)*, 2005.

[2] 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*, 2009.

[3] 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),* vol. 34, p. 248, 2015.

[4] A. Jain, T. Thormählen, H.-P. Seidel and C. Theobalt, "Moviereshape: Tracking and reshaping of humans in videos," in *ACM Transactions on Graphics (TOG)*, 2010.

[5] L. Pishchulin, S. Wuhrer, T. Helten, C. Theobalt and B. Schiele, "Building statistical shape spaces for 3d human modeling," *arXiv preprint arXiv:1503.05860,* 2015.

[6] Y. Chen, Z. Liu and Z. Zhang, "Tensor-based human body modeling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013.

[7] 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*, 1999.

[8] 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*, 2016.

[9] N. Hasler, C. Stoll, B. Rosenhahn, T. Thormählen and H.-P. Seidel, "Estimating body shape of dressed humans," *Computers \& Graphics,* vol. 33, pp. 211-216, 2009.

[10] 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,* vol. 127, pp. 31-42, 2014.

[11] 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*, 2016.

[12] 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*, 2010.

[13] A. O. Balan, L. Sigal, M. J. Black, J. E. Davis and H. W. Haussecker, "Detailed human shape and pose from images," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007.

[14] A. O. Bălan and M. J. Black, "The naked truth: Estimating body shape under clothing," in *European Conference on Computer Vision*, 2008.

[15] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka and B. Schiele, "Deepercut: A deeper, stronger, and faster multi-person pose estimation model," in *European Conference on Computer Vision*, 2016.

[16] 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*, 2014.

[17] R. W. Sumner and J. Popović, "Deformation transfer for triangle meshes," *ACM Transactions on Graphics (TOG),* vol. 23, pp. 399-405, 2004.

[18] "Artec Shapify Booth," [Online]. Available: https://www.artec3d.com/3d-scanner/shapifybooth. [Accessed 17 07 2017].

[19] "MPII human pose models," [Online]. Available: http://pose.mpi-inf.mpg.de. [Accessed 02 08 2017].

[20] D. Q. Huynh, "Metrics for 3D rotations: Comparison and analysis," *Journal of Mathematical Imaging and Vision,* vol. 35, pp. 155-164, 2009.

[21] S. Umeyama, "Least-squares estimation of transformation parameters between two point patterns," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 13, pp. 376-380, 1991.