CALVIS: Chest, wAist and peLVIS circumference from 3D human body meshes as ground truth for deep learning

Yansel Gonzlez Tejeda
Paris Lodron University of Salzburg
yansel.gonzalez-tejeda@stud.sbg.ac.at

Helmut Mayer
Paris Lodron University of Salzburg
helmut@cosy.sbg.ac.at

Abstract

In this paper we present CALVIS, a method to calculate chest, waist and pelvis circumference from 3D human body meshes. Our motivation is to use this data as ground truth for learning with convolutional neural networks (CNN). Previous work had used the large scale CAESAR dataset, determined manually these anthropometrical measurements or directly acquired them from a person. The problem is that acquiring these data is a cost and time consuming endeavor. In contrast our method can be used on 3D meshes automatically. First, we slice cross sections along the Y axis of the mesh. Second, we calculate the cross sections length and establish a signature for the body represented by the mesh. Finally, using this signature we are able to calculate the chest, waist and pelvis circumference by searching for extrema. We conduct two experiments. In the first experiment we synthesize 10 human body meshes. Then we apply CALVIS to calculate chest, waist and pelvis circumference. We evaluate the results qualitatively. We observe that the measurements can indeed be used to estimate the shape of a person. The second experiment asses the plausibility of our approach where we use the calculated human dimensions as a ground truth to train a small CNN. After having trained the network with our data, we demonstrate that learning converges, achieving x percent prediction error. Furthermore, we make the implementation of CALVIS publicly available to advance the field.

1. Introduction

Motivation Predicting 3D human body measurements from images is crucial in several scenarios like virtual try-on, animating, ergonomics, computational forensics and even health and mortality estimation. Researches had named this measurements body intuitive controls ([1]), biometric measurements ([13]), body dimensions ([3]), semantical parameters ([16]), traditional anthropometric measurements ([15]) or only "shape" as in human shape estimation ([7], [2], [8], [4], [11]). In contrast, we assume a more anthropometric approach motivated by comprehensive compendiums like Panero and Zelnik, 1979 [10]. Throughout this paper the term human body dimensions will be used to refer to the above measurements.

The problem of estimating human body dimensions having only an image is an under-constrained (or inverse) problem. Information gets lost when a camera is used to capture the human body in 3D space to 'render' a 2D image.

To tackle this problem a supervised learning approach can be used. This approach demands large amount of human body anthropometric measurements and is certainly one of the biggest challenges in the field. Currently there is only one large-scale dataset, the Civilian American and European Surface Anthropometry Resource (CAESAR) [12] with 3D human body scans and their corresponding anthropometric measurements. This survey was extraordinarily large and resource intensive: around nine thousand people from Europe and the USA where scanned, it was conducted over more than five years and costed more than six million dollars.

In the past decade a noticeably amount of researchers have employed this dataset to investigate human shape estimation. Because the measurement acquiring process is resource intensive and requires large amount of human and material resources, this type of studies are rare and the data they offer is expensive. Therefore, it is important to explore alternative methods where human body measurements derived from real data can be obtained for investigation.

3D human body generative models offers such an alternative. We propose to synthesize 3D human body meshes using the SMPL [8] generative model. Once we have the 3D meshes we can compute chest, waist and pelvis circumference. The next step after obtaining the measurements is to use a camera model to render images. Finally, in possession of this ground truth we can input this images to the learning algorithm and perform inference.

Problem statement: given a 3D human body mesh $\mathcal{M}$ we look for a method capable to automatically output chest, waist and pelvis circumference.

2. Approach

In this work we synthesize 3D human meshes using the SMPL model [8]. This model is at its top level a skinned articulated model, i.e., consists of a surface mesh that mimics the skin and a skeleton related to that mesh. It is defined by a mean template shape represented by a vector of N concatenated vertices $\bar{\mathbf{T}}$ in the zero pose, $\vec{\theta}^*$. In order to synthesize a new human mesh one has to deform the provided template mesh by setting shape parameters $\vec{\beta}$ and pose parameters $\vec{\theta}$. The model provides learned parameters

$$\Phi = \{\bar{\mathbf{T}}, \mathcal{W}, \mathcal{S}, \mathcal{J}, \mathcal{P}\} \quad (1)$$

As mentioned above $\bar{\mathbf{T}}$ is the mean template shape. The weight matrix $\mathcal{W}$ represents how much the rotation of skeleton parts affects the vertices. In addition, the matrices $\mathcal{S}$ and $\mathcal{P}$ define linear functions that are used to deform $\bar{\mathbf{T}}$ and the matrix $\mathcal{J}$ predicts skeleton rest joint locations from vertices in the rest pose. We held fix these parameters during the synthesis.

2.1. Human Body Mesh Signature

Let us consider a human body mesh $\mathcal{M}$. Our method requires that $\mathcal{M}$ is standing with arms raised parallel to the ground at shoulder height at a $90°$ angle. In the line of previous work ([4]), we name this pose the zero (also normalized (rf)) pose $\vec{\theta}^*$. Additionally, we assume that the mesh has LSA orientation, e.g., x, y and z axis are positively directed from right-to-left, inferior-to-superior and posterior-to-anterior, respectively. If the mesh has another orientation we can always apply transformations to LSA-align it.

Intuitively, we would like to measure the chest circumference bellow the arms at the widest part of the torso and the waist circumference at the narrowest part after the chest but above the hips. Similarly, the pelvis circumference is measured often around the widest part of hips and buttocks. We can formalize this intuition by considering the cross-sectional length of the 2D intersection curves ([6]) along the y-axis. Moreover, we can use a plane $\boldsymbol{\pi}$ parallel to the floor to intersect the mesh. Since $\mathcal{M}$ is triangulated, the boundary of this intersection is a collection of segments s_i. Therefore, we can determine the boundary length by specifying a point along the y-axis $j \in \mathbb{R}$ as

$$\mathcal{BL}(\mathcal{M}, \boldsymbol{\pi}, j) = \sum_{i=1}^{i=n} s_i \quad (2)$$

Note that for cross sections where $\boldsymbol{\pi}$ intersects the legs, two disconnected curves will appear. That is by no means a problem because they are still a collection of segment with non-zero length.

Next, we assemble the mesh slide length vector $\vec{\mathcal{L}}$. Starting from the top j_t of the bounding box we sliced iterative mesh $\mathcal{M}$ with plane $\boldsymbol{\pi}$ every m-meters until we reach the bounding box bottom j_b. For every slice at point j, we compute $\mathcal{BL}(j)$.

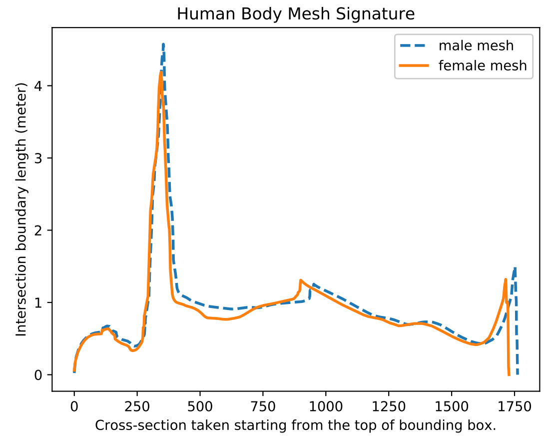

Figure 1. Human Body Mesh Signature for male and female meshes. The function resembles a rotated silhouette of the human body and exhibits several *extrema*. See section 2.2 for discussion of these extrema.

$$\vec{\mathcal{L}}(\mathcal{M}, \boldsymbol{\pi}, m) = \mathcal{BL}(j),\ j \in \{j_t, j_t + m, \cdots, j_b\} \quad (3)$$

Finally, we can construct the human body **mesh signature** $\mathcal{MS} : \mathbb{Z}^+ \to \mathbb{R}$ that maps every slice index $k \in \{0, 1, \cdots, |\mathcal{L}| - 1\}$ to the corresponding boundary length $\mathcal{BL}(j)$ in the mesh slide vector $\vec{\mathcal{L}}$.

$$\mathcal{MS}(k) = \mathcal{BL}(j) \quad (4)$$

Figure 1 shows the $\mathcal{MS}$ of two meshes (male and female) for $m = 0,001$ and plane $\boldsymbol{\pi}$ parallel to the floor (with normal $(0, 1, 0)$) using the library trimesh [9]. Note that the function as defined by equation 4 is bounded and not continuous. It resembles a rotated silhouette of the human body and exhibits several *extrema*.

2.2. Mesh Signature Extrema

By comparing neighboring values of the mesh signature $\mathcal{MS}$ as defined in equation 4 we find several *extrema*. In general, we expect these *extrema* to be adequate features to calculate the human dimensions. More specifically, we assume that:

1. The global maximum M_g represents the length of the largest path around the arms. From a topographic point of view M_g is the highest peak of $\mathcal{MS}$. A horizontal line at height proportion h of the peak intersects it at two points left p_l and right p_r.

2. The local maximum M_{pc} with largest value other than M_g immediately posterior to p_r represents the pelvis circumference.

3. The local minimum M_{wc} posterior to p_r but prior to M_{pc} represents waist circumference.

4. The local maximum M_{cc} posterior to p_r but prior to M_{wc} represents chest circumference. If this local maximum does not exist we set M_{cc} to the value of the path $\mathcal{BL}(j_{cc})$ with largest value in the interval $(p_l; M_{wc})$.

3. Experiments and Results

We conduct two experiments. In the first experiment we synthesize eight (four female and four male) human body meshes using shape parameters provided by SURREAL [14]. The meshes reflect human figures characteristics such as bulky, slim, small and tall subjects. Then we apply CALVIS to calculate chest, waist and pelvis circumference. Since we do not have ground truth, we evaluate the results qualitatively by comparing our method with [5]. The second experiment serves to asses the plausibility of our approach to use the synthetic data for deep learning. We input the calculated human dimensions to an artificial neural network. After having trained the network with our data, we show that learning converges.

4. Conclusion

The method can be optimized. Further research must be conducted.

References

[1] B. Allen, B. Curless, and Z. Popović. The space of human body shapes. In A. P. Rockwood, editor, *ACM SIGGRAPH 2003 Papers on - SIGGRAPH '03*, page 587, New York, New York, USA, 2003. ACM Press. 1

[2] 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 *Computer Vision – ECCV 2016*, Lecture Notes in Computer Science. Springer International Publishing, Oct. 2016. 1

[3] Y. Chen, D. P. Robertson, and R. Cipolla. A practical system for modelling body shapes from single view measurements. In *British Machine Vision Conference, BMVC 2011, Dundee, UK, August 29 - September 2, 2011. Proceedings*, pages 1–11, 2011. 1

[4] E. Dibra, H. Jain, C. Öztireli, R. Ziegler, and M. Gross. Hs-nets: Estimating human body shape from silhouettes with convolutional neural networks. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 108–117. IEEE, 2016. 1, 2

[5] E. Dibra, A. C. Öztireli, R. Ziegler, and M. H. Gross. Shape from selfies: Human body shape estimation using cca regression forests. In *ECCV*, 2016. 3

[6] H. Edelsbrunner and J. Harer. *Computational Topology - an Introduction.* American Mathematical Society, 2010. 2

[7] P. Guan. *Virtual Human Bodies with Clothing and Hair: From Images to Animation*. PhD thesis, Department of Computer Science at Brown University, 2013. 1

[8] M. Loper, N. Mahmood, J. Romero, G. Pons-Moll, and M. J. Black. Smpl: A skinned multi-person linear model. *ACM Trans. Graph.*, 34:248:1–248:16, 2015. 1, 2

[9] Michael Dawson-Haggerty. trimesh. 2

[10] J. Panero. *Human dimension & interior space : a source book of design reference standards*. Whitney Library of Design, New York, 1979. 1

[11] 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

[12] K. M. Robinette, H. Daanen, and E. Paquet. The caesar project: a 3-d surface anthropometry survey. In *Second International Conference on 3-D Digital Imaging and Modeling (Cat. No. PR00062)*, pages 380–386. IEEE, 1999. 1

[13] L. Sigal, A. Balan, and M. J. Black. Combined discriminative and generative articulated pose and non-rigid shape estimation. In J. C. Platt, D. Koller, Y. Singer, and S. T. Roweis, editors, *Advances in Neural Information Processing Systems 20*, pages 1337–1344. Curran Associates, Inc, 2008. 1

[14] G. Varol, J. Romero, X. Martin, N. Mahmood, M. J. Black, I. Laptev, and C. Schmid. Learning from synthetic humans. In *CVPR*, 2017. 3

[15] S. Wuhrer, C. Shu, and P. Xi. Landmark-free posture invariant human shape correspondence. *The Visual Computer*, 27(9):843–852, Sep 2011. 1

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