Localizing Anthropometric Landmarks Using 3-D Surface Features

Chang SHU[1], Pengcheng XI[1], Allan KEEFE[2]
[1] National Research Council of Canada, Ottawa, Ontario, Canada;
[2] Defence Research and Development Canada, Toronto, Canada

Abstract

Accurate localization of anthropometric landmarks is crucial for processing and analyzing 3-D anthropometric data. For example, landmarks are used to extract dimensional measurements from 3-D scans of human bodies. They can also be used to fit a template model to the scans such that a correspondence across the scans can be established. From this correspondence, we can perform statistical shape analysis to understand the variabilities of human shapes. In this paper, we propose a new method for localizing anthropometric landmarks using a combination of 3-D surface features and the latest deep learning techniques. The method makes use of geometric features represented as descriptor vectors. We first identify a set of locations that exhibit salient geometric features. Then we use pre-registered 3-D models to train a classifier for each geometric landmark. With the geometric landmarks, we fit a template to the data scan. The full set of anthropometric landmarks can be predicted from the template-fitted model. We validate our method using the 2012 Canadian Forces Anthropometric Survey (CFAS) dataset where 2,200 full-body scans were collected.

Keywords: 3-D anthropometry, anthropometric landmarks, digital human modeling

1. Introduction

As a result of the ever-increasing use of 3-D body scanning technologies, a large amount of 3-D anthropometric data have been produced [10, 20]. 3-D data capture the surface geometry of the human body and contain much more information about the shape of the human body than the traditional 1-D measurement data and is potentially capable of greatly enhancing the design of products that fit the human body. However, 3-D data pose new challenges: the amount and complexity of the 3-D data prevent them to be used effectively. Much processing is necessary before we can make full use of the rich information contained in them. One of the first processing steps is to fit a template mesh to every raw scan. This way, correspondence among the data models is established, making the unordered, noisy, and incomplete 3-D data comparable, paving the way to subsequent processing and analysis, such as creating statistical shape models and extracting physical measurements from the 3-D models [1,4,19,23-25,28]. Anthropometric landmarks are stable corresponding positions on the human body that exist across the population. If we find these points accurately, we will have grasped the structure of the human body from the scans, and other corresponding points can be found by interpolation.

Early attempts of automatically localizing landmarks are rule-based [6, 18]. Each landmark is defined by a set of rules describing its geometric location. For example, the tip of the nose would be defined as the most front point of the face. However, these rules often assume specific coordination systems and postures. Therefore they are not generically applicable to all scans and tend to have many exceptions. Furthermore, the number of exceptions can quickly grow unmanageable.

Several learning-based methods have been proposed. Ben Azouz et al. [2] suggested a method that uses a statistical graphical model to capture the relationships among the landmarks. The parameters of the graphical model are learned from training data. Reasonable results were achieved for small test sets [9]. However, only SPIN image [13] is used for surface feature description.

Creust et al. [5] propose a machine-learning approach in which landmark localization is formulated as a classification problem. Multiple scalar surface descriptors, such as Gaussian and mean curvature are combined to describe the local surface properties at the landmark locations. A classification function is learned from training data. Recently, Xi et al. [29, 30] use a deep learning approach to identify landmarks. Local surface patches are represented as images by simulating a camera taking pictures of the area. A convolutional neural network is trained to compute surface features implicitly and to classify a point as landmark or non-landmark. Both of these learning-based methods have an issue of training: while we know which points are landmarks, it is difficult to decide which points should be used for non-landmark examples for training.

More recently, several authors have proposed solutions to the dense shape correspondence problem [17, 26, 32]. While promising results have been achieved, these methods emphasize global solutions to models with arbitrary postures. Furthermore, no public benchmark data are available for anthropometric landmarks. In the context of ergonomics and design applications, anthropometric landmarks have special importance and require individual treatment.
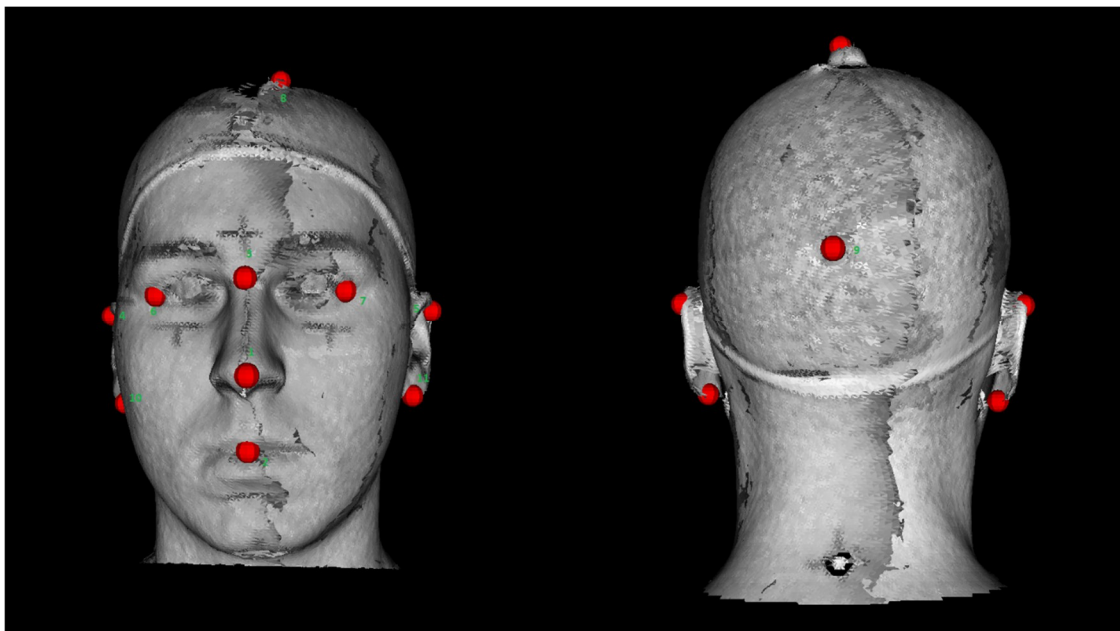
Fig. 1. Geometric landmarks.

In this paper, we re-visit surface features by examining the latest 3-D surface descriptors. These descriptors were developed in computer vision for matching, recognition, and retrieving of 3-D objects. The main feature of these descriptors is that the area upon which a descriptor is based is a variable. In other words, the degree of localness is controllable. This allows us to characterize the local surface patches that surround the landmark using statistics of geometric properties. This information is subsequently used to match points on the surface against a template. We demonstrate that certain landmarks can be identified reliably by simply matching these feature descriptors.

Throughout this paper, we use head models as an example to demonstrate our proposed techniques, but the method applies to data for the full-body or any other parts of the body.

2. Landmark localization

2.1. Geometric landmark and patch features

To locate anthropometric landmarks on the scan models, we need to study the geometric properties of the landmark locations on the surface of the human body. The scan data is a digital representation of the human body surface and it is all the information we have for predicting the landmarks. Unfortunately, not all of the landmarks have salient surface features. Human operators rely on palpating the underlying bone structure to locate the landmarks. Without pre-marking, even human experts cannot locate certain landmarks accurately on the scan model. To solve this problem, we have identified a set of geometric landmarks that correspond to the points that exhibit strong surface features. These landmarks are used to fit a template mesh and the anthropometric landmarks that are not in the set of geometric landmarks can be predicted indirectly by the fitted template [31].

Fig. 1 shows the geometric landmarks we have selected for the head model. When selecting geometric landmarks, we consider two aspects. One is that the local surface should be distinguishable from the nearby surface area. For a point like this, we select it together with a set of surrounding points to make a surface patch. We call this a patch feature. The sizes of these patches vary from landmark to landmark and are learned from training data (see section 2.3.1). Another consideration is the importance of a location. Certain locations are strategically important for fitting a template model. For example, the two landmarks on the ear are such points.

2.2. Surface feature descriptors

In recent years, the computer vision community has developed many effective algorithms for identifying salient geometric features in 3-D scans [12]. These features capture the local surface properties and they have been used successfully for recognizing shapes from 3-D data. We have investigated several feature descriptors, including SHOT (Signature of Histograms of OrienTation), FPFH (Fast Point Feature Histogram) [21], and 3DSC (3D Shape Context) [8]. All these descriptors represent the local surface by statistical distributions of geometric attributes, such as normal and curvature, of the points on the surface. They are invariant to rigid transformation. Through experiments, we found that SHOT produced the best results [22].
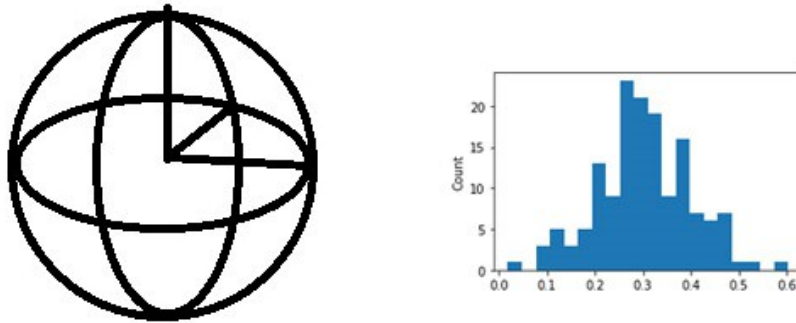
Fig. 2 SHOT surface descriptor.

SHOT descriptor is a point signature which encodes the histograms of the angle between the surface normal in a spherical support region. As shown in Fig. 2, the support region is discretized into small volumes along the radial, azimuth, and elevation dimensions. For each volume, a histogram is computed by counting the number of points that fall into the bins according to the angles between the normal of the points within the volume and the normal of the signature point. All of the histograms are combined together to form a 1344-dimensional vector. In other words, SHOT descriptor represents the local surface geometry compactly. Furthermore, by varying the support region, we can control how 'local' the descriptor needs to be.

2.3 Reference model

To detect landmarks, we choose a scan as the model and manually mark the location of the landmarks. The model serves as a reference for the landmarks whose surface descriptors are compared with in order to detect the landmarks. When choosing the reference model, we avoid those faces that have extreme shapes and find one that is close to the average model. However, it is not necessary to find an optimal reference model. Our algorithm relies on the property that landmarks are geometrically distinguishable from their surrounding points and the majority of the human beings share this property.

We then compute the feature descriptors at the landmark points and these descriptors become the model to check against.

2.4 Landmark detection

2.3.1 Learning for support region

A surface descriptor for a point looks at the surrounding points on the surface and compute statistics about the geometric properties in the region. The size of this region is a parameter that we need to decide. If it is too big, it ceases to be a local descriptor and loss its power to describe the landmark. It is also costly to compute. On the other hand, if it is too small, the descriptor is sensitive to surface noise.

The proper size of the support region for the landmarks are decided by a learning process. A set of scans with ground truth landmarks are set aside and the size of the support is varied in small increments within a range. Then the detection algorithm is run through the training set and the best performing size for each landmark is chosen for the rest of the scans.
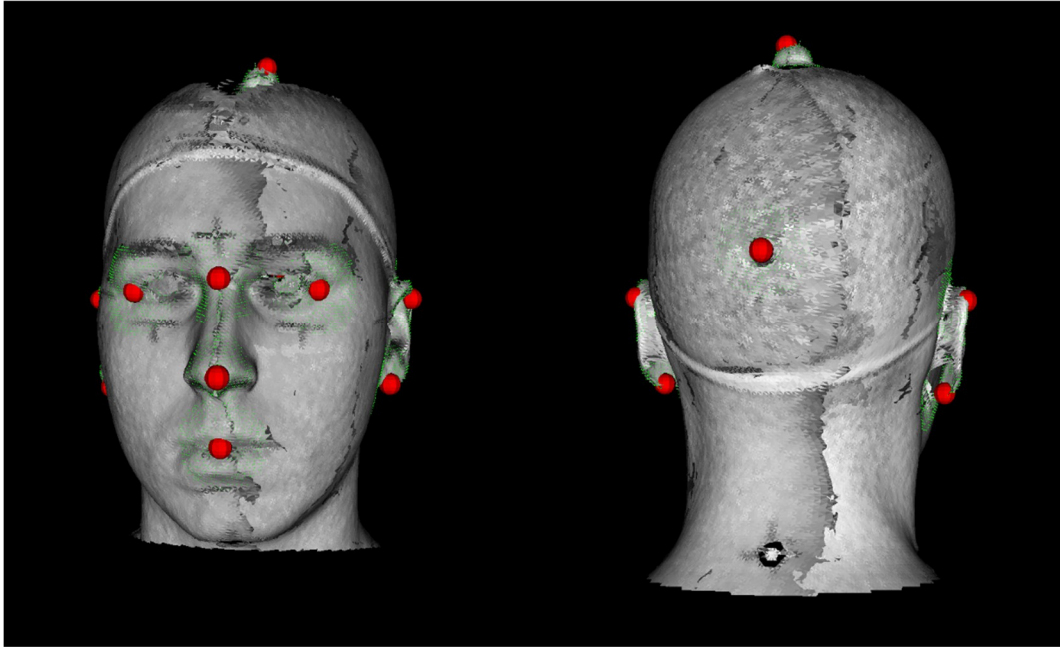
Fig. 3. Support regions for each landmark.

2.3.2 Candidate points

When predicting the landmarks, we find the best matches for the feature descriptors on the target scan. To search for the matching landmark, we first find a candidate region on the target scan. This region can be found by using a rough template fitting based on a few extreme points, such as the top of the head and the tip of the middle fingers [11]. For head models, a simples scaling and an ICP [3] alignment with the reference model is enough to obtain the candidate points.

2.3.3 Search

Finally, the feature descriptor for each point in the candidate region is computed and compared with the model landmark feature. The best match is regarded as the found landmark. In Fig. 3, the descriptor support region is shown in green dots.

The feature descriptor of a single point can be unreliable due to noise. We collect a small patch around each landmark point and compute the feature descriptors. The difference between the model point and the candidate point is evaluated according to the following equation:

$$d(p) = \sum_{q \in N(p)} \|q - \hat{q}\|^2$$

Where $N(p)$ is the collection of descriptors in the patch surrounding the point p and $\hat{q}$ is the nearest neighbor of q in the candidate set.

3. Results

Table 1 shows the results of the landmark prediction.

Table 1. Accuracy of landmark prediction.

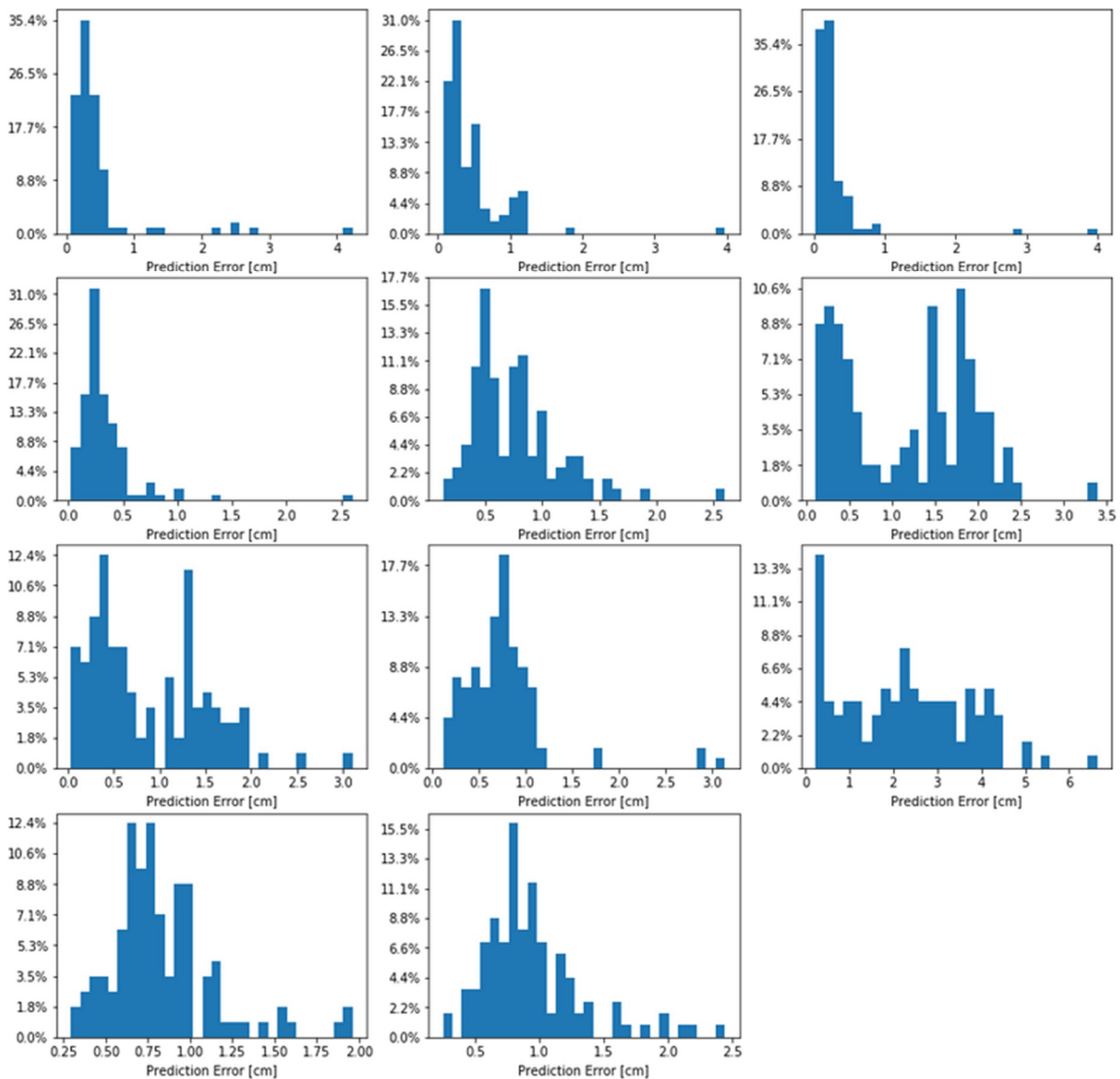| Landmark no.	1	2	3	4	5	6	7	8	9	10	11
Mean (cm)	0.429	0.480	0.275	0.332	0.757	1.154	0.880	0.749	2.276	0.829	0.941
Std	0.570	0.467	0.460	0.299	0.383	0.745	0.624	0.468	1.420	0.308	0.387

Fig. 4. Prediction error distributions.

We validated our method on the CFAS (Canadian Forces Anthropometry Survey) dataset. The CFAS dataset consists of 2,200 scans of Canadian military personnel. In this study, we used the head scans. HumanSolution scanners were used to collect the data. 200 models were set aside for learning the support size and the rest of the models were used for testing.

We used MeshLab to manually pick the landmarks for validation. The programs were implemented in C++. Point Cloud Library (PCL) was used to compute the surface descriptors. The PCL functions assume only point cloud representation. In our case, however, we have triangle meshes to represent the scan models and thus can compute the surface normal more reliably. Therefore, we use the mesh surface normal for computing the surface descriptors. Finally, in the CFAS survey, landmarks were pre-marked. However, in this study, we did not make use of this information. Thus, the results shown here were obtained by using purely geometric information.

4. Conclusions

We have introduced a method for localizing anthropometric landmarks in 3-D scans of humans. We demonstrate that the SHOT surface descriptor can be used effectively to identify certain landmarks. This method is simple to implement and does not involve complicated structures.

Although the recent deep learning techniques [15] offer effective tools for 2-D landmark identification, there are difficulties when it is applied to 3-D scan data. Chief among them is the problem of converting 3-D data to 2-D images to take advantage of the convolution networks. 3-D surface descriptors provide an efficient representation for landmark localization. For future work, we will investigate the possibility of integrating surface feature descriptors with deep learning techniques.

References

[1] B. Allen, B. Curless, and Z. Popovic, The space of human body shapes: reconstruction and parameterization from range scans, ACM Transactions on Graphics, 22(3): 587-594, 2003.

[2] Z. Ben Azouz, C. Shu, and A. Mantel, Automatic Locating of Anthropometric Landmarks on 3D Human Models, Third International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT 2006). Chapel Hill, North Carolina, USA. June 13-16, 2006.

[3] P. J. Besl and N. D. McKay, A method for registration of 3-D shapes, IEEE Trans. Pattern Anal. Mach. Intell., 14(2): 239-256, 1992.

[4] Bradtmiller, B. and Gross, M., "3D Whole Body Scans: Measurement Extraction Software Validation," SAE Technical Paper 1999-01-1892, 1999.

[5] C. Creusot, N. Pears, and J. Austin, A Machine-Learning Approach to Keypoint Detection and Landmarking on 3D Meshes, International Journal of Computer Vision, 102:146-179, 2013.

[6] L. Dekker, I. Douros, B. Buxton, and P. Treleaven, Building symbolic information for 3D human body modeling from range data, Proceedings of the third International Conference on 3-D Digital Imaging and Modeling (3DIM'99), pp. 292-298, 1999.

[7] H. Fan and E. Zhou, Approaching human level facial landmark localization by deep learning, Image and Vision Computing, 47:27-35, 2016. 300W, the First Automatic Facial Landmark Detection in-the-wild Challenge.

[8] A. Frome, D. Huber, R. Kolluri, T. Bulow, and M. Malik, Recognizing objects in range data using regional point descriptors, in Proc. European Conf. Computer Vision, pp. 224-237, 2004.

[9] A. Giachetti, E. Mazzi, F. Piscitelli, M. Aono, A. Ben Hamza, T. Bonis, P. Claes, A. Godil, C. Li, M. Ovsjanikov, V. Patraucean, C Shu, J. Snyders, P. Suetens, A. Tatsuma, D. Vandermeulen, S. Wuhrer, P Xi, SHREC'14 track: automatic location of landmarks used in manual anthropometry, Eurographics Workshop on 3D Object Retrieval, 93-100, 2014.

[10] C.C. Gordon, B. Bradmiller, C.E. Clausen, T. Churchill, J.T. McConville, I. Tebbetts, and R.A. Walker, 1987-1988 Anthropometric survey of US Army personnel. Methods and summary statistics. Natick/TR-89-044. US Army Natick Research Development and Engineering Center, Natic, MA, 1989.

[11] D. A. Hirshberg, M. Loper, E. Rachlin, A. Tsoli, A. Weiss, B. Corner, M. J. Black, Evaluating the Automated Alignment of 3D Human Body Scans, 2nd International Conference on 3D Body Scanning Technologies, 76-86, 2011.

[12] D. Holz, A.E. Ichim, F. Tombari, R.B. Rusu, and S. Behnke, Registration with the Point Cloud Library: A Modular Framework for Aligning in 3-D, IEEE Robotics and Automation Magazine, December 2015.

[13] A. Johnson, Spin-Images: A Representation for 3-D surface matching, PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburg, PA, August 1997.

[14] M. Kouchi and M. Mochimaru, Errors in landmarking and the evaluation of the accuracy of traditional and 3D anthropometry, Applied Ergonomics, 42:518-527, 2011.

[15] A. Krizhevsky, I. Sutskever, G.E. Hinton, "ImageNet classification with deep convolutional neural networks", Advances in Neural Information Processing Systems (NIPS), 2012.

[16] I. Leong, J. Fang, and M. Tsai, Automatic body feature extraction from a marker-less scanned human body, Computer-Aided Design, 39:568-582, 2007.

[17] O. Litany, T. Remez, e. Rodola, A. Bronstein, and M. Bronstein, Deep functional maps: structured prediction for dense shape correspondence, IEEE Int. Conf. Computer Vision, 2017.

[18] J.H. Nurre, Locating landmarks on human body scan data, International Conference on Recent Advances in 3-D digital imaging and modeling, pp. 289-295, 1997.

[19] S. Paquette, J.D. Brantley, B. Corner, P. Li, and T. Oliver, Automated extraction of anthropometric data from 3D images, Procedings of the IEA 2000 / HFES 2000 Congress.

[20] K. Robinette et al., Civilian American and European Anthropometry Resource (CAESAR) Final Report, Volume I: Summary, US Air Force Research Report, AFRL-HE-WP-TR-2002-0169, 2002.

[21] R. B. Rusu, N. Blodow, and M. Beetz, Fast point feature histograms for 3D registration, in Proc. IEEE Int. Conf. Robotics and Automation, pp. 1848-1853, 2009.

[22] S. Salti, F. Tombari, L. Di Stefano, SHOT: Unique signatures of histograms for surface and texture description, Computer Vision and Image Understanding, 125: 251-264, 2014.

[23] C. Shu, S. Wuhrer, and P. Xi, 3D Anthropometric Data Processing, International Journal of Human Factors Modelling and Simulation, 3(2):133-146, 2012.

[24] C. Shu, P. Xi, and A. Keefe, Automatic Extraction of Traditional Anthropometric Measurements from 3-D Body Scans, 5th International Digital Human Modeling Symposium, 2017.

[25] Aggeliki Tsoli, Matthew Loper, and Michael J. Black, Model-based Anthropometry: Predicting Measurements from 3D Human Scans in Multiple Poses, IEEE Winter Conference on Applications of Computer Vision (WACV), 2014.

[26] L. Wei, Q. Huang, D. Ceylan, E. Vouga, and H. Li, Dense human body correspondences using convolutional networks, Proc. CVPR, 2017.

[27] S. Wuhrer, Z. B. Azouz, and C. Shu, Posture invariant surface description and feature extraction. In Computer Vision and Pattern Recognition (CVPR), pages 374–381, June 2010.

[28] P. Xi, W.-S. Lee, and C. Shu, Analysis of segmented human body scans. In Proceedings of Graphics Interface 2007, GI'07, pages 19–26, New York, NY, USA, 2007.

[29] P. Xi, C. Shu, and R. Goubran, Localizing 3-D Anatomical Landmarks Using Deep Convolutional Neural Networks, 14th Conference on Computer and Robot Vision (CRV'17), 197-204, 2017.

[30] P. Xi, C. Shu, and R. Goubran, Comparing 2D Image Features on Viewpoint Independence using 3D Anthropometric Dataset, International Journal of the Digital Human 1(4): 412-425, 2016.

[31] S. Yamazki, M. Kouchi, and M. Mochimaru, Markerless landmark localization on body shape scans by non-rigid model fitting, 2nd International Digital Human Modeling Symposium, 2013.

[32] S. Zuffi and M. J. Black, The stitched puppet: A graphical model of 3D human shape and pose, Proc. CVPR, 2015.