User Study: Concise Network for 3D Human Modeling from Orthogonal Silhouettes

Bin Liu[a], Xiuping Liu[a], Zhi-Xin Yang[b], Charlie C.L. Wang[c,*]

[a] *School of Mathematical Sciences, Dalian University of Technology, Dalian, P.R. China*
[b] *State Key Laboratory of Internet of Things for Smart City and Department of Electromechanical Engineering, University of Macau, Macau, P.R. China*
[c] *Department of Mechanical, Aerospace and Civil Engineering, The University of Manchester, United Kingdom*

Abstract

In this technical report, user study is conducted for evaluating our method presented in [1] on 10 individuals (5 females and 5 males) by using real photos as input. For each individual, photos in different scenarios are given including standard neutral pose, poses with different arm opening angles and bent arms and cameras in extremely wrong positions. For female, we also study the influence of hair. Example photos for these scenarios are given in Fig. 1, where all photos are captured by smartphones. The background of all images are removed by using either the tool provided by smartphone or the publicly available tool at: `https://www.remove.bg`. The body height, hip girth, waist girth and chest girth of each individual were measured as ground truth when taking pictures. We have made this dataset available together with the source code of our implementation.

1. Neutral pose

The first study and comparison is conducted on all users with neutral pose. The measurements are taken for three girths and compared with the ground truth. Among all 30 measurements, 16 most accurate results are generated by our method while the other two methods [2, 3] generate 7 best results each. The input photos and the resultant measurements generated by different methods are given in Fig. 2. Results of 3D models are shown in Fig. 3. It is found that our method performs better.

2. Variation of poses

In order to analyze the robustness of end-to-end network based human model generation approaches, we conduct a study by using poses with different arm opening angles and slightly bent arms. The mean and the range of variation for all individuals are given in Fig. 4. It is found that the methods of Dibra et al. [2], Ji et al. [3] and ours generate the most accurate measurements in the same frequency (i.e., 10 times each) among all 30 measurements. Also, the ranges of errors obtained from all three methods are similar. It can conclude that the robustness to pose variation is similar on these three methods.

3. Incorrect camera positions

We also study the influence of camera position in our experiments. To study the performance in extreme cases, we use photos taken in 'wrong' positions – i.e., the photos are not taken in orthogonal views (see Fig. 5). Analysis is conducted on the methods of Dibra et al. [2], Ji et al. [3] and ours. According to the errors listed in Fig. 5, our method is more sensitive to the incorrect camera positions comparing to the other two methods. Therefore, clear guidance of viewing orientation needs to be given while applying the proposed method in practice (e.g., the smartphone application for customized cloth design as discussed in the following section).

4. Influence by loose hair

Reconstructing 3D human models from silhouettes heavily relies on accurate input of orthogonal silhouettes. Therefore, users are expected to wear tight clothes (as shown in Figs. 1 and 2. Besides of that, it is also interesting to study the influence of loose hair. For all examples conducted in above studies, we require all individuals to either have short hair or long hair coiled up. When photos are taken with loose hairs (as shown in Fig. 6), all methods will be significantly influenced. Among 15 measurements, Dibra et al. [2], Ji et al. [3] and ours generate best results in 6, 5 and 4 times respectively. See the measurements shown in Fig. 6, where the most accurate results are highlighted by bold fonts.

5. Summary

We can make the following conclusions from user study:

1. Our method performs better than other two approach on input with similar neutral poses;
2. The influence of arm poses on our method is similar to the other two approaches;
3. Our method is more sensitive to the correct orthogonal views (i.e., will generate poor results with incorrect position of camera);
4. All methods cannot solve the occlusion problem caused by loose hair.

*Corresponding author
Email address: `changling.wang@manchester.ac.uk` (Charlie C.L. Wang)

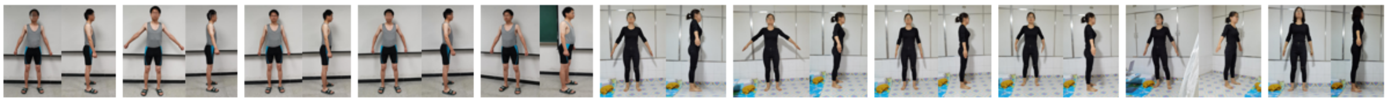
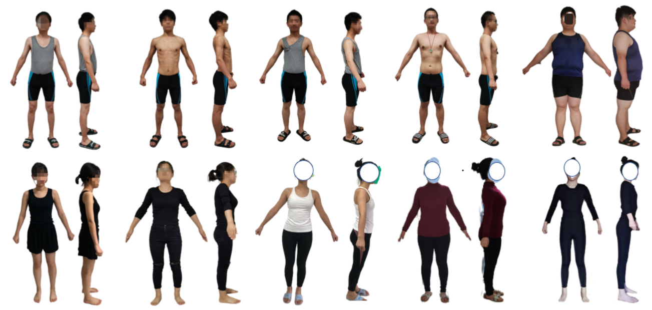


Figure 1: Example photos captured by smartphone for each individual in different scenarios, including standard neutral poses, two poses with different arm opening angles, one pose with slightly bent arms, and extremely wrong positions of camera. For female, one extra photo is taken to study the influence of loose hair.



Users (H, W, B)	Dibra et al. [2]	Ji et al. [3]	Ours
M1 (99, 83, 94)	(101.6, 89.9, 99.5)	(90.8, 79.1, 91.8)	(**97.2**, **83.1**, **99.3**)
M2 (96, 83, 95)	(**97.1**, 85.7, **95.7**)	(92.5, 76.7, 89.7)	(93.9, **80.9**, 98.9)
M3 (96, 82, 88)	(95.3, 88.2, 95.7)	(95.4, **82.8**, 95.0)	(**95.9**, 80.5, **94.4**)
M4 (101, 88, 97)	(97.8, **88.9**, 98.7)	(97.6, 85.7, **97.2**)	(**103.7**, 84.8, 101.6)
M5 (115, 109, 118)	(**112.4**, 105.3, 112.3)	(121.3, 112.9, 121.1)	(118.5, **108.1**, **120.6**)
F1 (91, 71, 86)	(94.8, 84.5, 91.1)	(**92.4**, **79.7**, 89.2)	(96.8, 82.3, **88.6**)
F2 (90, 76, 89)	(92.3, 84.6, 91.6)	(91.6, **79.9**, **88.9**)	(**91.5**, 84.8, 93.4)
F3 (96, 73, 85)	(**97.2**, 86.2, 94.3)	(91.1, 79.3, 89.4)	(90.9, **78.9**, **87.0**)
F4 (99, 82, 90)	(94.9, **85.6**, **92.3**)	(**101.7**, 93.7, 100.3)	(107.4, 102.5, 108.9)
F5 (85, 66, 84)	(93.0, 84.6, 96.1)	(91.4, 77.5, 89.2)	(**85.7**, **70.0**, **86.3**)

Figure 2: Study taken on ten individuals with photos taken in similar neutral poses. The statistic of girth errors (Unit: centimeter) are given on the results generated by Dibra et al. [2], Ji et al. [3] and ours. The most accurate results are highlighted by bold fonts.

These conclusions from user study are very useful for the development of downstream application as we can avoid some extreme cases (e.g., incorrect camera positions, loose hair etc.) by designing a better process of using this technology.

References

[1] B. Liu, X. Liu, Z. Yang, C. C. L. Wang, Concise and effective network for 3d human modeling from orthogonal silhouettes (2020). `arXiv:1912.11616`.

[2] E. Dibra, H. Jain, C. Öztireli, R. Ziegler, M. Gross, Hs-nets: Estimating human body shape from silhouettes with convolutional neural networks, in: 2016 Fourth International Conference on 3D Vision, IEEE, 2016, pp. 108–117.

[3] Z. Ji, X. Qi, Y. Wang, G. Xu, P. Du, X. Wu, Q. Wu, Human body shape reconstruction from binary silhouette images, Computer Aided Geometric Design.

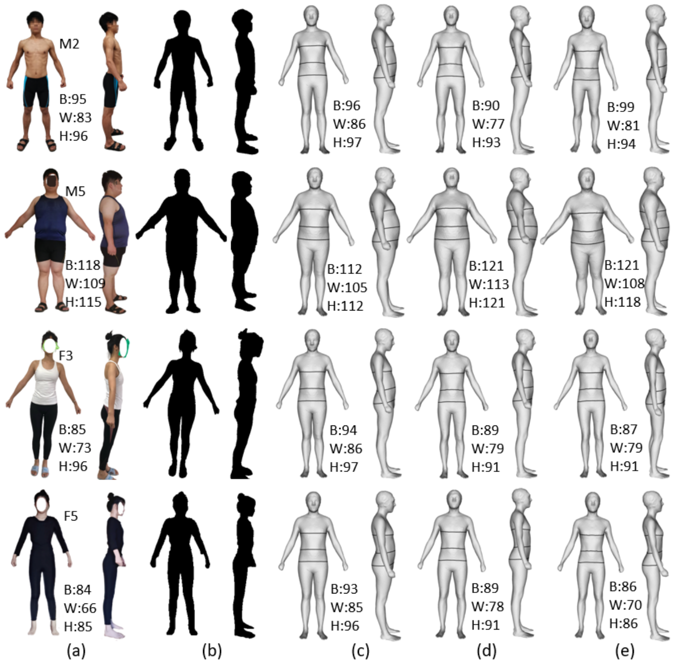


Figure 3: Given mask images (b) generated from real photos (a), results of 3D human models can be generated by Dibra et al. [2] (c), Ji et al. [3] (d) and our method (e). The estimated measurements of three girths (B: Chest, W: Waist, H:Hip) are given for all results to compare with the ground truths listed in (a).

Users (H, W, B)	Dibra et al. [2]			Ji et al. [3]			Our Approach		
	Hip	Waist	Chest	Hip	Waist	Chest	Hip	Waist	Chest
M1 (99, 83, 94)	97.6, [97.3, 98.3]	86.5, [85.7, 87.3]	**95.9**, [95.2, 97.1]	**98.9**, [94.9, 103.1]	**84.5**, [79.9, 88.1]	96.2, [92.8, 99.3]	96.1, [90.7, 99.9]	81.1, [76.9, 84.1]	96.4, [91.3, 99.3]
M2 (96, 83, 95)	93.3, [91.7, 95.4]	**83.8**, [82.0, 85.4]	93.4, [91.2, 94.4]	**94.8**, [93.3, 95.6]	79.8, [78.7, 80.5]	93.3, [92.8, 93.5]	93.5, [90.1, 95.8]	78.4, [76.7, 79.6]	**96.5**, [93.3, 99.1]
M3 (96, 82, 88)	93.5, [92.1, 94.3]	84.6, [82.7, 96.3]	92.6, [91.3, 94.2]	**93.8**, [90.6, 98.5]	**81.2**, [76.1, 86.7]	**91.6**, [87.1, 97.2]	91.1, [90.2, 92.1]	78.4, [77.3, 79.3]	92.7, [92.1, 93.2]
M4 (101, 88, 97)	98.3, [97.6, 99.2]	**87.2**, [85.9, 88.1]	**96.6**, [96.2, 97.7]	98.6, [96.4, 101.4]	83.1, [80, 85.9]	95.2, [92.7, 97.9]	**101.5**, [100, 103.7]	84.4, [83.7, 84.9]	101.9, [101.7, 102.1]
M5 (115, 109, 118)	103.5, [101.7, 108.8]	95.9, [95.1, 102.4]	105, [103.5, 108.9]	123.9, [119.3, 132.2]	**111.9**, [105.5, 116.3]	**121.1**, [113.2, 127.5]	**123.2**, [115.1, 133.6]	114.9, [109.1, 119.7]	127.5, [122.7, 133]
F1 (91, 71, 86)	**91.6**, [90.9, 92.7]	**82.2**, [80.5, 83.1]	90, [87.8, 90.9]	95, [93.6, 96.3]	82.8, [81.9, 83.5]	91.8, [90.8, 92.4]	97.9, [96, 99.8]	82.8, [79.5, 85.0]	**89.9**, [86, 93.3]
F2 (90, 76, 89)	88.3, [86.7, 90.2]	**78.2**, [77.5, 82.5]	86.3, [86.2, 89.4]	93.2, [90, 96.5]	81, [78.8, 83.2]	**91**, [88.9, 93.1]	**91.6**, [87.6, 95.8]	81.8, [76.3, 87.3]	93.8, [87.9, 98]
F3 (96, 73, 85)	**96.2**, [91.7, 99.3]	83.3, [81.4, 87.1]	91.9, [89.7, 95.2]	96.9, [93.4, 100.9]	81.2, [78, 86.1]	92.1, [88.8, 97.1]	91.9, [91.1, 93.1]	**76.4**, [73.4, 78.3]	**86.1**, [85.3, 87.0]
F4 (99, 82, 90)	92.2, [90.7, 93.7]	**81.9**, [80.7, 84.8]	**89.1**, [88.1, 91.2]	**101.3**, [97.7, 107.9]	86.4, [82.4, 92.6]	94.7, [91.5, 99.5]	104.1, [100.9, 107]	95.6, [89, 101.5]	101.4, [94.3, 105.1]
F5 (85, 66, 84)	92.2, [88.8, 94.7]	79.9, [76.7, 82.1]	88.7, [86.4, 90.8]	91.3, [90.9, 91.7]	75.2, [72.9, 76.5]	87, [84.1, 89.5]	**87.3**, [85.6, 88.2]	**70.4**, [69.2, 71.7]	**85.6**, [84.6, 86.3]

Figure 4: Robust study of our method according to the variation of poses – i.e., different arm opening angles (top and middle rows) and slightly bent arms (bottom row). The mean and the range of measurements are reported for different methods applied to all 10 individuals. The most accurate results are highlighted by bold fonts. The measurements are reported in the unit of centimeter.

Users (H, W, B)	Dibra et al. [2]	Ji et al. [3]	Ours
M1 (99, 83, 94)	(96.3, 87.9, 96.5)	(95.7, **84.5**, **95.8**)	(**97.8**, 85.1, 97.5)
M2 (96, 83, 95)	(95.5, 85.9, **94.6**)	(92.1, **81.1**, 92.4)	(**96.1**, 85.6, 100.3)
M3 (96, 82, 88)	(93.5, 82.9, **90.9**)	(**97.5**, 82.4, 91.4)	(93.7, **82.3**, 95.6)
M4 (101, 88, 97)	(97.1, **87.5**, **96.6**)	(**97.8**, 87.0, 98.2)	(107.0, 92.1, 107.8)
M5 (115, 109, 118)	(105.3, 94.1, 101.5)	(**124.6**, **117.3**, **123.8**)	(144.5, 136.8, 136.2)
F1 (91, 71, 86)	(96.6, **78.7**, 93.1)	(**95.0**, 81.9, 91.0)	(96.4, 83.5, **88.8**)
F2 (90, 76, 89)	(**97.3**, **87.1**, 97.4)	(99.8, 89.1, **96.9**)	(115.1, 108.9, 115.5)
F3 (96, 73, 85)	(**101.3**, **89.9**, 99.7)	(102.8, 90.0, **98.1**)	(130.0, 113.7, 115.4)
F4 (99, 82, 90)	(106.7, **89.4**, **97.8**)	(**105.9**, 94.1, 100.1)	(118.9, 108.3, 110.4)
F5 (85, 66, 84)	(97.8, **86.5**, 95.2)	(**96.8**, 86.9, **94.5**)	(102.6, 91.3, 101.8)

Figure 5: Robust study of human reconstruction from photos taken with 'incorrect' positions of cameras – i.e., the contours extracted from photos are not orthogonal silhouettes. None method can generate good results but our method is indeed more sensitive to the position of camera. Unit: centimeter.

Users (H, W, B)	Dibra et al. [2]	Ji et al. [3]	Ours
F1 (91, 71, 86)	(**92.2**, **79.7**, 90.0)	(96.6, 84.7, 92.4)	(98.1, 85.8, **89.4**)
F2 (90, 76, 89)	(92.8, 80.7, **88.7**)	(**90.7**, 78.0, 86.7)	(96.3, **77.9**, 92.6)
F3 (96, 73, 85)	(95.3, 82.8, **90.8**)	(**96.2**, **81.5**, 92)	(97.8, 89.9, 95.2)
F4 (99, 82, 90)	(89.0, **71.3**, **84.5**)	(**100.5**, 96.2, 103.9)	(116.8, 99.8, 104.3)
F5 (85, 66, 84)	(92.7, 79.1, 87.7)	(88.5, 75.0, **85.9**)	(**86.7**, **73.3**, 88.6)

Figure 6: Study on input photos with loose hair – all methods are influenced by this change while ours is most sensitive. Measurements of three girths (Unit: centimeter) are reported and the best results are highlighted by bold fonts.