Automatic Foot Measurement Extraction from a 3D Point Cloud via a Deep Neural Network

Nastaran NOURBAKHSH KAASHki, Xinxin DAI, Pengpeng HU, Adrian MUNTEANU
Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussels, Belgium

https://doi.org/10.15221/22.47

Abstract

The foot is a vital human body part comprising a complex system of muscles and bones sustaining the human weight, and providing balance and mobility when daily activities are being performed. Extracting accurate foot measurements is of paramount importance in many applications including medical sciences, sports and fashion industry. Traditionally, footwear brands employ contact-based foot measuring methods involving a trained operator to design and produce well-fitted footwear products. However, this process is very time consuming and is prone to human errors. With the advancement of 3D scanning technologies, the foot can be scanned accurately with an affordable 3D scanning device. In this research, we propose, to the best of our knowledge, the first deep neural network (FNet) for automatic foot measurement extraction from a 3D foot point cloud. The proposed FNet is an encoder-decoder neural network which operates on the foot point cloud and outputs the foot reconstruction as well as the corresponding measurements points utilized for measurement extraction. Our study shows that teaching the network to accurately generate the measurement points, performed with the help of the well-designed loss functions, is necessary for automatic and accurate foot measurement extraction. In order to train the proposed neural network, a large dataset of complete foot scans with their corresponding measurement points and measurement values are synthesized. The performance of the proposed method has been evaluated on both synthetic test data as well as the real scans captured by the Occipital Structure Sensor Pro. The results show that our method outperforms the state-of-the-art methods in terms of accuracy and speed.

Keywords: Foot measurement extraction, 3D point Cloud, deep neural network, encoder-decoder

1. Introduction

Foot measurement extraction plays an important role in fashion industry and medical sciences. Medical studies have shown that wearing poor-fitting shoes can be the cause of various health problems, such as stress and pain to the bones and joints [1] [2] [3].

Conventionally, the foot is measured using a measuring tape by a trained operator. However, this process is very time-consuming and unpleasant. Thus, researchers have developed non-contact-based methods to extract the measurements from a digital foot representation [4] [5] [6]. In [6], the subject is asked to put his or her foot on an A4 paper, where a photo of the foot is taken. The authors detect and segment the feet by employing VGG16 neural network [7] for detecting the A4 paper from the input image. Although they can measure the width and height of the foot from the photo, they cannot take the girth measurements as their method performs only on an image taken from the top side of the foot.

Inspired by the success of body and hand measurement extraction methods [8] [9] [10] [11], we propose the first deep neural network, especially developed for foot measurement extraction from a 3D complete foot scan. To this end, a new large dataset of foot samples and measurement values is introduced for training, validating and testing the proposed neural network. Moreover, a novel loss function is employed for foot measurement extraction. The trained deep neural network is evaluated with promising results on real and synthetic foot unseen samples.

The rest of this paper is organized as follows. In Section 2, the proposed method is explained. The experimental results is discussed in Section 3. Finally, Section 4 presents the main conclusions of our work.

2. Method

The framework of the proposed method is given in Fig. 1.

We propose a novel deep encoder-decoder for foot measurement extraction. The network receives a point cloud of a foot scan, fits the template to the foot sample and outputs the deformed template as well as the measurements.

To this end, we first create a foot template using the SMPL-X model [12]. Then, we define 20 different measurement types on the surface of the foot template. Each measurement type is characterized by a set of measurement points and the plane using which the measurement is defined. To train, validate and test our proposed encoder-decoder, we synthesize 400,000 foot mesh samples from the SMPL-X model in various shapes and poses. We transfer the measurements defined on the template to each sample of our synthetic dataset. During the training, we teach the neural network to fit the template as well as the measurements to the input scan using three loss functions [10]:

(i) The mean square error between the output and ground truth points;

(ii) The distance of the output measurement points to their corresponding plane;

(iii) The mean square error of the measurement values.

In the test stage, we scan the foot with the Occipital Structure Sensor Pro. The encoder receives the foot point cloud and outputs the measurements.

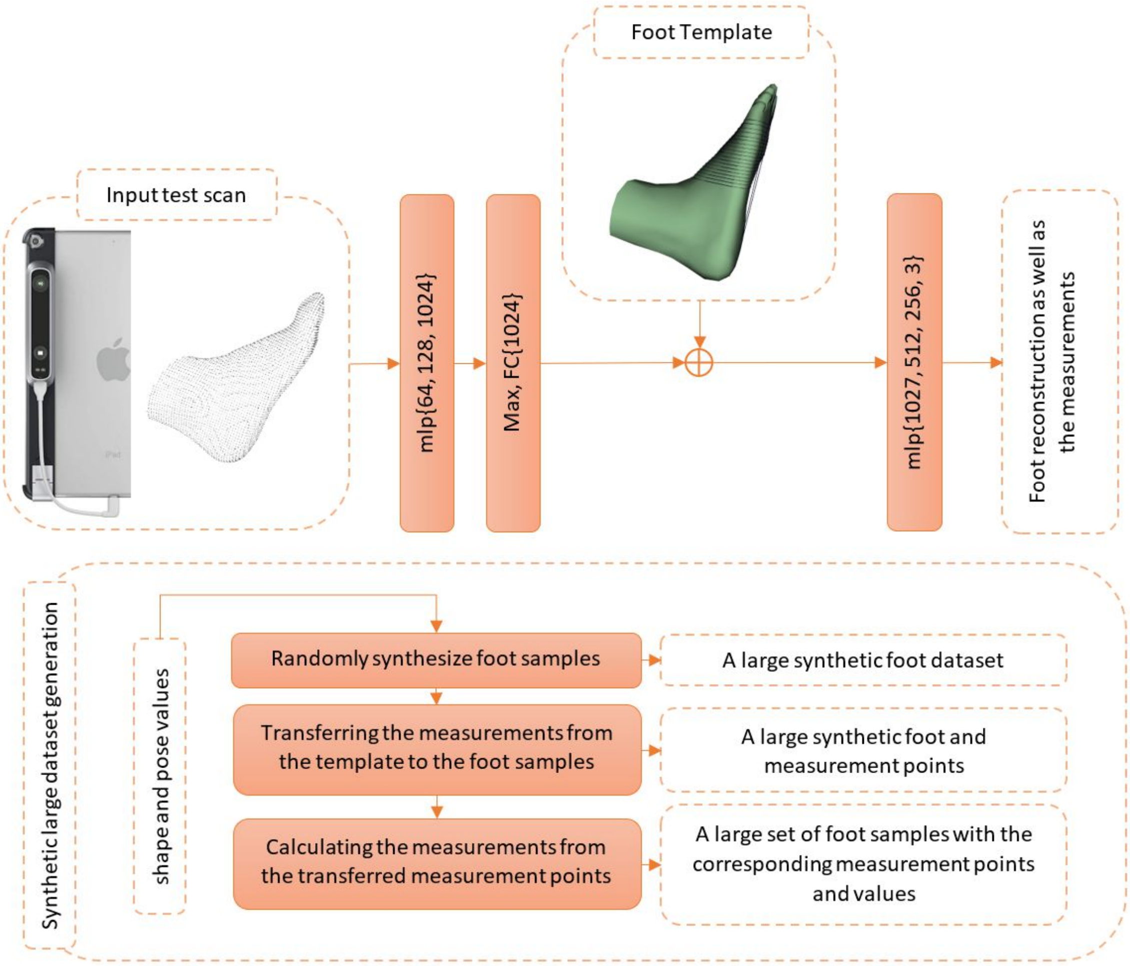

Fig. 1 . The framework of the proposed method.

2. Experiment

In this section, we evaluate the performance of the proposed method on the synthetic foot samples as well as real foot scans.

Our synthetic foot test set includes 200 foot samples with their corresponding ground truth measurement values. To create our real test set, we scanned the feet of 2 different subjects using Occipital Structure Sensor Pro. A professional anthropologist has measured every foot using a measuring tape, resulting in a set of real scans with their corresponding ground truth measurement values.

We calculated absolute and relative measurement errors for 20 different measurement types on each sample of our real and synthetic test set. Table 1 and Fig. 2 compares the performance obtained on the real scans with the results achieved on the synthetic data.

*Table 1. The mean absolute and relative error (mm) on the **synthetic** and **real** test set.*

Method	Mean absolute error on **synthetic** unseen samples	Mean absolute error on **real** unseen samples	Time (s)
Fnet with only loss (i)	9.9	6.9	0.5
Fnet	6.8	6.1	0.5

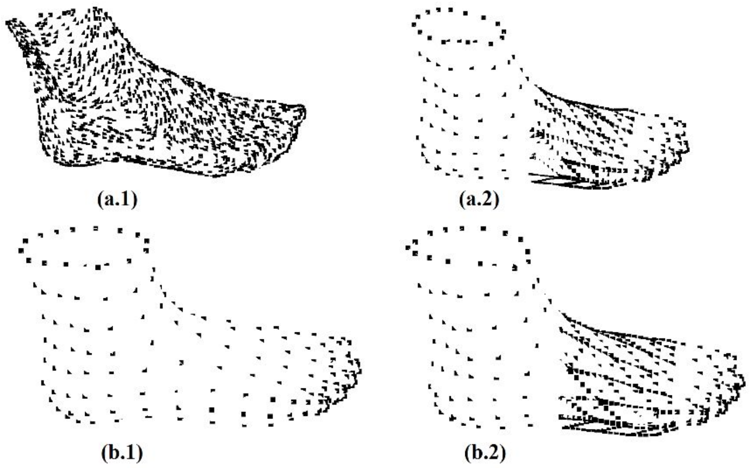

Fig. 2. Visualization results on real and synthetic unseen samples. (a.1) is the real scan captured by Occipital Structure Sensor Pro. (b.1) is the synthetic test sample. (a.2) and (b.2) are the deformed templates and their corresponding measurement points generated by Fnet.

As the results on Fig. 2 illustrate, Fnet performs better on the synthetic test samples compared to real scans. Although adding the measurement points to the template and synthetic samples helps the neural network to measure the foot correctly, the synthetic samples generated from SMPL-X model are not dense enough to represent the scans captured by Occipital Structure Sensor Pro. It can be expected that enhancing the resolution of the synthetic samples will further improve the performance of Fnet.

As it is explained in the section 2, Fnet is trained with a weighted sum of three loss functions. To show the importance of having a loss function especially developed for foot measurement extraction, we have trained our neural network two times. First, we trained it only with the first loss function, the mean square error between the output and ground truth points. Second, we have trained Fnet with the complete loss function. The training and validation errors are depicted in Fig. 4. We depicted the corresponding mean absolute and relative errors on the synthetic foot test set in Fig. 3. As the results in Fig. 3 and Fig. 4. show, Fnet performs significantly better when it is trained using the complete loss function. In other words, it is important to teach the network to keep the measurements planar while minimizing the measurement errors.

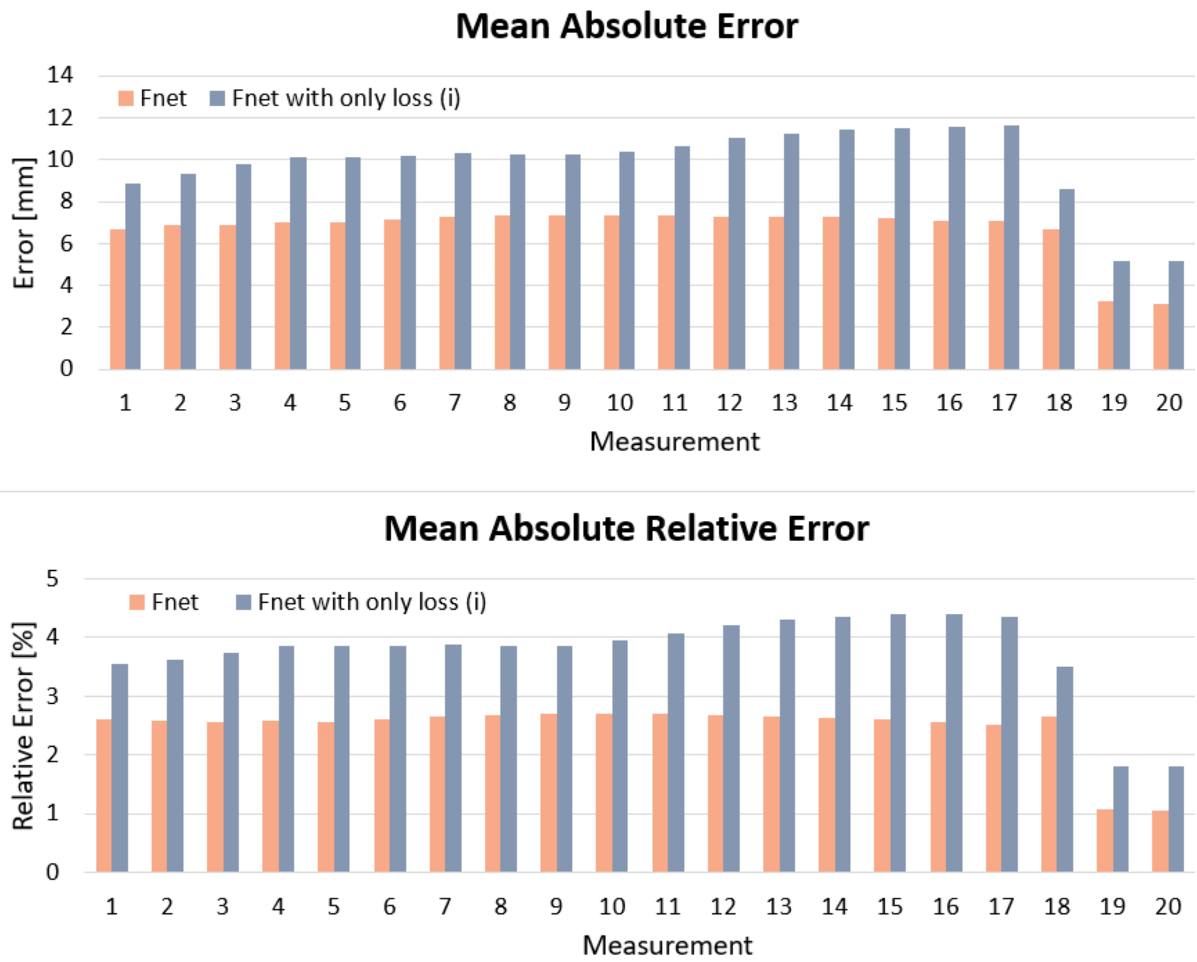

*Fig. 3. The mean absolute and relative error on the **synthetic** test set.*

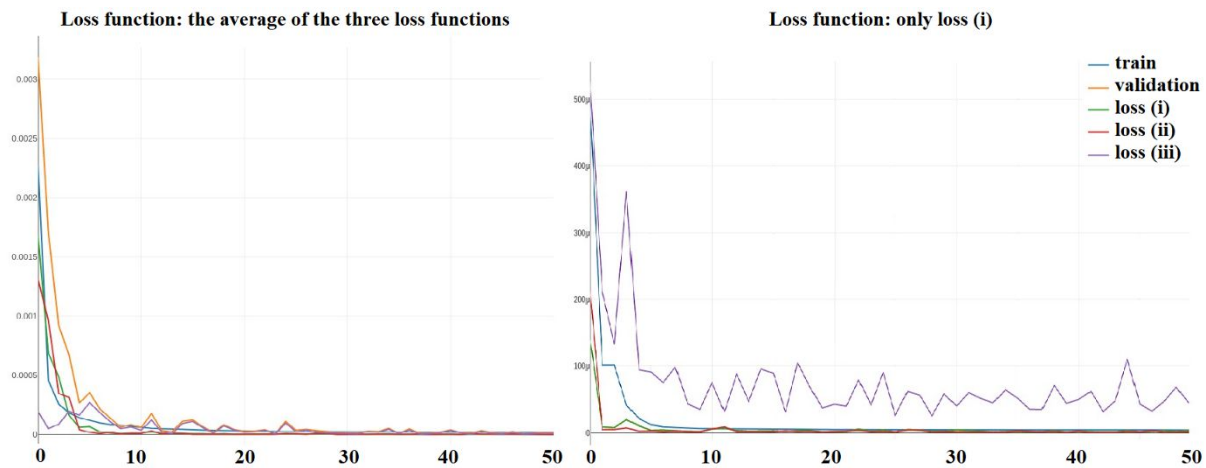

Fig. 4. Training and validation errors while training with different loss functions.

3. Conclusion

We have proposed the first deep neural network (Fnet) for foot measurement extraction from a 3D scan. We have introduced a new dataset of foot samples as well as the corresponding measurement values for training, validating and testing the network. We have shown that foot template-based reconstruction with regular loss functions is not enough for accurate measurement extraction. Rather, an especially-developed loss function is needed to teach the neural network to minimize the measurement error.

The experiments have revealed that although the template and the synthetic data are quite sparse compared to the scans captured by Occipital Structure Sensor Pro, the proposed method can achieve promising results for foot measurement extraction. It is expected that enhancing the resolution of the synthetic samples will further improve the performance of the proposed method.

4. References

[1] F. Mun and A. Choi, "Deep learning approach to estimate foot pressure distribution in walking with application for a cost-effective insole system," *Journal of NeuroEngineering and Rehabilitation,* vol. 19, no. 1, pp. 1-14, 2022.

[2] D. Laidig, A. J. Jocham, B. Guggenberger, K. Adamer, M. Fischer and T. Seel, "Calibration-free gait assessment by foot-worn inertial sensors," *Frontiers in Digital Health,* vol. 3, 2021.

[3] A. Jurca, J. Žabkar and S. Džeroski, "Analysis of 1.2 million foot scans from North America, Europe and Asia," *Scientific reports,* vol. 9, no. 1, 2019.

[4] S. Martedi, H. Saito and M. Servieres, "Shape Measurement System of Foot Sole Surface from Flatbed Scanner Image," *MVA, Citeseer,* pp. 338-341, 2009.

[5] H.-H. Tu and C.-Y. Yu, "Hole-filling and foot measurement extraction of the 3-D foot form data by template matching method," in *The 40th International Conference on Computers & Indutrial Engineering*, IEEE, 2010.

[6] H. Wang, F. Liu and R. Fan, "A research on foot size measurement algorithm based on image," *Journal of Physics: Conference Series,* vol. 1903, no. 1, 2021.

[7] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[8] N. N. Kaashki, P. Hu and A. Munteanu, "Kaashki, Nastaran Nourbakhsh, Pengpeng Hu, and Adrian Munteanu," *IEEE Transactions on Instrumentation and Measurement,* vol. 70, pp. 1-14, 2021.

[9] N. N. Kaashki, P. Hu and A. Munteanu, "Anet: A Deep Neural Network for Automatic 3D Anthropometric Measurement Extraction," *IEEE Transactions on Multimedia,* 2021.

[10] N. N. Kaashki, X. Dai, T. Gyarmathy, P. Hu, B. Iancu and A. Munteanu, "A Deep Learning Approach to Automatically Extract 3D Hand Measurements," in *7th International Conference on Machine Learning Technologies*, Rome, Italy, 2022.

[11] N. N. Kaashki, X. Dai, T. Gyarmathy, P. Hu, B. Iancu and A. Munteanu, "Automatic and Fast Extraction of 3D Hand Measurements using a Deep Neural Network," in *IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*, Canada, 2022.

[12] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, A. A. Osman, D. Tzionas and M. J. Black, "Expressive body capture: 3D hands, face, and body from a single image," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019.