DOI 10.1109/ICIP49359.2023.10222114
ISBN 978-1-7281-9836-1
EISN 978-1-7281-9835-4

Publisher: IEEE

MEASURE4DHAND: DYNAMIC HAND MEASUREMENT EXTRACTION FROM 4D SCANS

Xinxin Dai⋆ *Ran Zhao*⋆ *Pengpeng Hu*†,‡ *Vasile Palade*† *Adrian Munteanu*⋆

⋆ Department of Electronics and Informatics, Vrije Universiteit Brussel, Brussels, Belgium
†Centre for Computational Science and Mathematical Modelling, Coventry University, Coventry, UK
‡corresponding author(ae1133@coventry.ac.uk)

ABSTRACT

Hand measurement is vital for hand-centric applications such as glove design, immobilization design, protective gear design, to name a few. Vision-based methods have been previously proposed but are limited in their ability to only extract hand dimensions in a static and standardized posture (open-palm hand). However, dynamic hand measurements should be considered when designing these wearable products since the interaction between hands and products cannot be ignored. Unfortunately, none of the existing methods are designed for measuring dynamic hands. To address this problem, we propose a user-friendly and fast method dubbed Measure4DHand, which automatically extracts dynamic hand measurements from a sequence of depth images captured by a single depth camera. Firstly, the ten dimensions of the hand are defined. Secondly, a deep neural network is developed to predict landmark sequences for the ten dimensions from partial point cloud sequences. Finally, a method is designed to calculate dimension values from landmark sequences. A novel synthetic dataset consisting of 234K hands in various shapes and poses, along with their corresponding ground truth landmarks, is proposed for training the proposed methods. The experiment based on real-world data captured by a Kinect illustrates the evolution of the ten dimensions during hand movement, while the mean ranges of variation are also reported, providing valuable information for the hand wearable product design. (The video abstract is available here.)

Index Terms— hand measurement, point cloud processing, dynamic hand, landmark, partial scan

1. INTRODUCTION

The advent of 4D scanning technology has propelled dynamic anthropometry measurements [1] to the forefront of various applications. Although existing methods prioritize analyzing dynamic body measurements to ensure the ideal fit and ergonomic comfort of clothing products, dynamic hand measurement is under-researched [2, 3]. Dynamic hand measurement values are essential in a multitude of fields, such as protective gear [4, 5], glove fabrication [6, 7, 8], hand-centric entertainment [9, 10], to name a few. Furthermore, several studies have demonstrated that different hand postures can cause skin deformation, thereby affecting the fit of hand-centric appliances. Therefore, there is a growing need to comprehend how the human hand measurement alters while it is moved. Such knowledge could provide important information in achieving optimal fitting, comfort, and performance in hand-centric wearable product design.

Traditional hand measurements are manually extracted by means of a measuring tape by an experienced anthropometrist. However, it requires subjects to stretch their hands and keep still during measurement, which is not suitable for static hands in complex postures or dynamic hands. Furthermore, the precision of the measurement highly depends on the anthropometrist's expertise. With the development of 3D scanning technology, researchers have proposed the automatic extraction of hand measurements from 3D scans [11, 12]. However, these methods can only work for the static hand in an open-palm pose, which does not reflect the dynamic interaction between the agonist and antagonist muscles during movement. Moreover, they require a complete hand scan as input, which is not always available. Consequently, hand measurement values estimated based on these static postures are not sufficient for developing hand-centric appliances. Given that the hand is in constant motion, interacting with the environment, the designed hand appliances should not be designed as rigid shells that restrict hand movement. This realization has prompted the research community to turn their attention to the technological evolution from 3D to 4D scanning systems in recent years. Klepser *et al* [2, **?**] revealed that the proportions of the human body vary depending on whether the subject is in a dynamic or static state, and they also investigated how the range of dynamic measurements affected the fit and comfort of garments worn by athletic individuals. However, the study overlooked the deformation of the hand shape during movement. The human hand is a highly intricate structure comprising 34 muscles and 27 bones, accounting for a quarter of the bones in the body. This unique anatomy allows the hand to achieve a wide range of shapes and complex poses, making the analysis of

This work was supported in part by Innoviris under project AI43D and in part by Fonds Wetenschappelijk Onderzoek (FWO) under Project G094122N.

dynamic hand measurements a considerably more complex issue. Nasir and Kwan *et al*[13, 7] found that postural variation significantly affects hand skin deformation. However, they only studied three static hand postures: relaxed, grip and power grip.

In this paper, we proposed a novel deep learning-based method dubbed Measure4DHand, which enables the extraction of dynamic hand measurements from 4D partial scans. Although leveraging the real-world dynamic hand sequence to train the proposed model is the ideal strategy, it is significantly expensive and time-consuming to collect such a dataset. Therefore, we trained our model based on synthetic single scans but the trained model generalizes well to sequences. Specifically, in the training phase, Measure4DHand takes a partial hand point cloud as input and generates a set of designed landmarks and their corresponding measurement values. While in the testing phase, Measure4DHand directly consumes partial point cloud sequences and produces a sequence of measurement landmarks and corresponding measurement values. Unlike previous methods aiming to predict the complete shapes from partial scans [14], we strive to predict landmarks rather than complete hand shapes. Such a strategy enables the network to focus on measurement landmark prediction, resulting in better estimations of hand measurements.

The main contributions in this paper can be summarized as follows:

- We proposed, to the best of our knowledge, the first deep learning-based method, dubbed Measure4DHand, for automatic extraction of dynamic hand measurements from partial hand point cloud sequences.
- We proposed a novel large-scale synthetic dataset consisting of 234K hands with a wide variety of hand shapes and poses, corresponding ground truth landmarks, and measurement values. Besides, we also collected 5 real hands via Kinect to validate the efficacy of the proposed method. To facilitate the related study, we will make the real-world dataset public when the paper is published.
- We tested our proposed method on real scans and illustrated the evolution of the ten dimensions during hand movement.

2. PROPOSED METHOD

2.1. Problem Statement

The proposed method mainly contains three steps: landmark definition, landmark extraction, and measurement value estimation, as the following is introduced. Given a sequence of partial point clouds of a hand $\mathcal{X} = \{\mathbf{S}^n\}_{n=1}^N$, where $\mathbf{S}^n = \{s_i^n \in \mathbb{R}^3 | i = 1, 2, ..., I^n\}$ denotes the set of points with I^n points captured from n^{th} frame and N is the number of frames and can be set to an arbitrary number in this study. The hand is moving when it is scanned, so the hand poses of N frames are a set of coherent movements and the point numbers of each $\mathbf{S}^n$ can be different. Our target is to devise a user-friendly method to automatically extract a sequence of hand measurement landmarks $\mathcal{Y} = \{\mathbf{L}^n\}_{n=1}^N$ and estimate measurement values $\mathcal{Z} = \{\mathbf{M}^n\}_{n=1}^N$ from $\mathcal{Y}$, where $\mathbf{L}^n = \{l_j^n \in \mathbb{R}^3 | j = 1, 2, ..., J\}$ denotes the set of hand landmarks with J points corresponding to $\mathbf{S}^n$ and $\mathbf{M}^n$ represents the measurement values of $\mathbf{S}^n$. To this end, we first leverage a neural network to learn a mapping $\mathcal{M} : \mathcal{X} \longmapsto \mathcal{Y}$. Then the measurement values are estimated by a measurement function $\mathcal{F} : \mathcal{Y} \longmapsto \mathcal{Z}$.

2.2. Proposed Datasets

2.2.1. Proposed synthetic dataset

2.2.1.1. Hand model generation

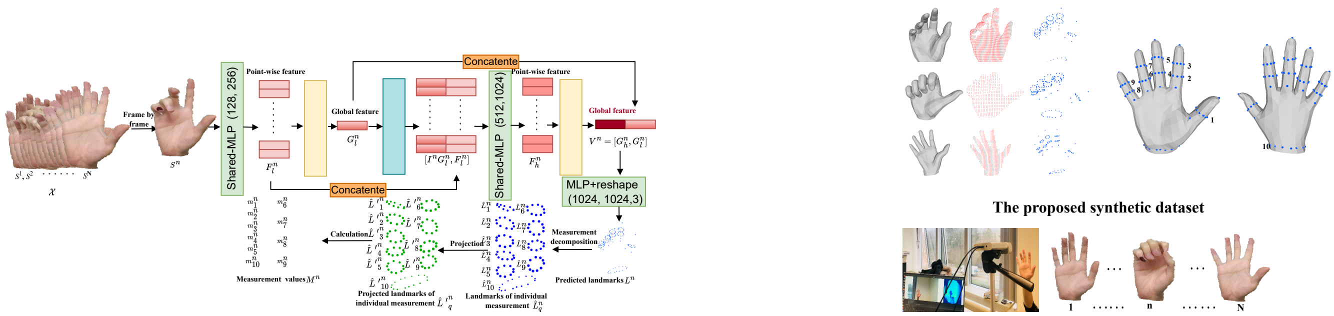
Generating a large-scale dataset (e.g., 100K samples) by scanning and measuring physical subjects is an incredibly time-consuming, expensive, and tedious task. To address this problem, we proposed a novel synthetic dataset by means of the SMPL-X model [15]. SMPL-X is a unified 3D model of the human body that encompasses the body, the face and the hand. In this paper, we only focus on the hand part, which consists of 778 vertices and 1538 triangles. each hand model is controlled by 15×3 pose parameters θ and 10×1 shape parameters β. We input thousands of shape and pose parameters that are extracted with FrankMocap [16] into the SMPL-X model to generate 234K hand meshes with hand shape and pose variations, some examples are illustrated in the first column at the top of Fig. 1b. Subsequently, We rendered partial scans from the obtained hand models via the open-source Blender Sensor [17], as shown in the second column at the top of Fig. 1b. The rendered scan of the hand is employed as the input to our proposed neural network.

2.2.1.2. landmark definition

To get the measurement values of the hands, it is necessary to define the measurements in advance. In this study, we defined 10 types of measurement: thumb girth, index-IP girth, index-DI girth, middle-IP girth, middle-DI girth, ring-IP girth, ring-DI girth, little-IP girth, little-DI girth and wrist girth (DI - the joint between Distal and Intermediate phalanges, IP - the joint between Intermediate and Proximal phalanges). The last two columns at the top of Fig. 1b shows examples of the proposed hand models and defined landmarks. With the exception of wrist girth, which comprises 16 landmarks, all other measurements incorporate 10 landmarks. Additionally, a fingertip landmark and two fingerroot landmarks are preserved for each finger, totaling 121 landmarks are defined.

2.2.2. Collected real-world dataset

The proposed method was trained based on a synthetic dataset via a frame-wise manner but should generalize well to the real-world partial point cloud sequence. To validate its performance, we created a real-world dataset by utilizing a Kinect



(a) Overview of the proposed method. Measure4DHand processes a sequence of hand partial point clouds frame by frame and outputs measurement values.

(b) The proposed synthetic and real-world datasets

Fig. 1. (a) The architecture of the proposed method. (b) The proposed datasets

to scan the hand of a subject, as shown at the bottom in Fig. 1b. The hand performs a closed-loop coherent motion from an open palm to a palm grip to a fully open palm, which represents the maximum range of motion that can be achieved by the hand, as illustrated at the bottom in Fig. 1b.

2.3. Landmark extraction

As aforementioned, the proposed neural network consumes the partial point cloud and deforms the hand measurement landmarks to fit the pose of input accordingly. In the training phase, the input is a single partial point cloud $\mathbf{S}^n$ and the output is landmarked $\mathbf{L}^n$. In the testing phase, a sequence of partial point clouds $\mathcal{X}$ is utilized as input, and a sequence of anticipated measurement landmarks $\mathcal{Y}$ is output.

The proposed neural network follows the encoder-decoder framework, as shown in Fig. 1a. The encoder is designed by stacking two simplified PointNet [18]. The first PointNet with a shared MLP consisting of two hidden layers of low dimensions to convert the coordinates of $\mathbf{S}^n$ into point-wise feature matrix F_l^n, which is further extracted into a global feature G_l^n by point-wise max-pooling operation. Following the similar processing of the first PointNet, the combination of G_l^n and F_l^n as input of the second PointNet with two hidden layers of high dimensions output a high-dimension global feature G_h^n. To the end, G_l^n and G_h^n are concatenated, forming the combined latent vector V^n. The decoder with three fully connected layers is responsible for generating the landmarks $\mathbf{L}^n$ from the combined latent vector V^n.

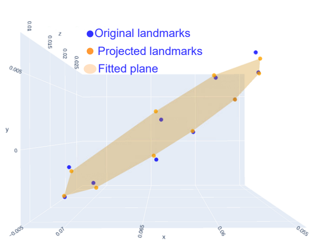


Fig. 2. Illustration of projecting landmarks to the same plane.

2.4. Measurement estimation

As described in Sec. B(1), the landmarks are selected from the vertices of the hand mesh, so the order of landmarks can be fixed. The extracted landmarks usually are not on the same plane, especially during the movement, which can lead to an increase in measurement errors. Therefore, We adopt a point-projection method to project all the landmarks onto the same plane, as shown in Fig. 2. Let $\hat{L}_q^n \subset \mathbf{L}^n$ denote the landmarks of q^{th} dimension, where $\hat{L}_q^n = \{\hat{l}_{q,k}^n \in \mathbb{R}^3 | q = 1,2,...,Q, k = 1,2,...,K\}$. We first compute the centroid μ_q^n and normal vector w_q^n of $\hat{L}_q^n$ as:

$$\mu_q^n = \frac{1}{K}\sum_{k=1}^{K}\hat{l}_{q,k}^n \tag{1}$$

$$\begin{aligned}[W_q^n, -, -] &= SVD((\hat{l}_{q,k}^n - \mu_q^n)\cdot((\hat{l}_{q,k}^n - \mu_q^n))^T) \\ w_q^n &= W_q^n[:, -1]\end{aligned} \tag{2}$$

where SVD is the Singular Value Decomposition that can find the main dimensions of the data distribution. The normal vector w is the last column of basis vector W. The fitted plane can be represented as:

$$(w_q^n)^T \mu_q^n = b_q^n \tag{3}$$

where b is the distance between the fitted plane and the original point. The point $\hat{l}_k$ can be projected on the fitted plane as follow:

$$\hat{l}\,'^n_{q,k} = \hat{l}_{q,k}^n - \frac{(\hat{l}_{q,k}^n \cdot w_q^n - (w_q^n)^T\mu)}{(w_q^n \cdot w_q^n)w_q^n} \tag{4}$$

In the end, each measurement value can be calculated as follow:

$$m_q^n = \sum_{k=1}^{K}||\hat{l}\,'^n_{q,k+1} - \hat{l}\,'^n_{q,k}||_2^2 \tag{5}$$

2.5. Loss Functions

To train the proposed model, the loss function is defined as the mean square error:

$$\mathcal{L}(\mathbf{L}, \mathbf{L}_{GT}) = \frac{1}{J}\sum_{j=1}^{J}||l_j - l_j^{GT}||_2^2, J = 121 \tag{6}$$

where $\mathbf{L}$ is the predicted measurement landmarks, $\mathbf{L}_{GT}$ is the ground truth measurement landmarks directly extracted from the vertices of hand models.

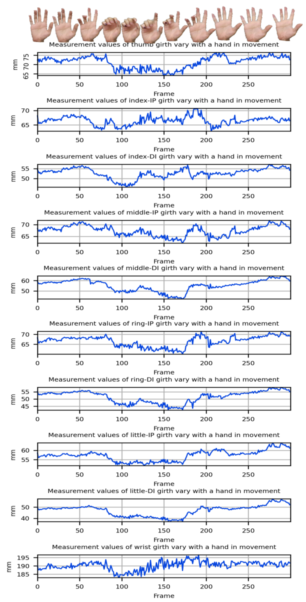


Fig. 3. The evolution of each hand measurement along the movement (frames) of a hand. The first row is partial point clouds captured by Kinect, while the following rows illustrate variation plots of measurement values.

3. EXPERIMENTAL RESULTS

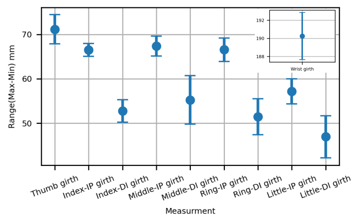
The proposed model was trained on the synthetic dataset but generalizes well to the unseen real-world partial point cloud sequence. We measure the hand of each frame to qualitatively and quantitatively analyze the measurement range while it is moved.

3.1. Qualitative evaluation

Fig. 3 visualizes the evolution of each hand measurement along the movement (frames) of a hand. The figure is useful to illustrate the pattern and the magnitude of contraction and expansion of each dimension during movement. As described in Sec. B(2), the captured motion performs a closed-loop coherent motion from an open palm to a palm grip to a fully open palm, which constitutes a single cyclic motion. As shown in Fig. 3, it can be seen that such motion can lead to a clear cyclic variation in measurement values. Specifically, the variation of measurement values for the open palm and the fully open palm is a similar trend with slight fluctuation and the measurement values of the fully open palm are slightly greater than those of the open palm. While the hand postures are during grip movements, the measurement values show significant fluctuations and are much lower than those of the open palm. Additionally, while the finger-IP and finger-DI exhibit similar variations throughout the entire motion, the

Table 1: Mean range and standard deviations of the ten measurements (Unit:mm)

Measurements	Mean $\pm St.Dev$
Thumb girth	71.2±3.3
Index-IP girth	66.5±1.5
Index-DI girth	52.8±2.5
Middle-IP girth	67.4±2.2
Middle-DI girth	55.2±5.4
Ring-IP girth	66.5±2.7
Ring-DI girth	51.5 ±4.0
Little-IP girth	57.2±2.8
Little-DI girth	47.0±4.8
Wrist girth	190.3±2.6



former display more intense fluctuations in comparison to the latter.

3.2. Quantitative evaluation

For quantifying the magnitude of the variation of each measurement throughout a movement, the mean and standard deviation of each dimension is calculated, as shown in Table 1. The range represents the maximum value minus the minimum value, which is the total variation. Depending on the motion performed, we can see the variation of nine finger girths ranges from 4.2 cm to 7.5 cm, while the range of wrist girth varies from 18.7 cm to 19.3 cm. We also draw an error bar to enhance the visualization of the range, as shown in Fig. 1. Due to the significant difference in measurement values between the wrist and fingers, the results of the wrist and fingers are presented separately. It clearly shows the variation range of joint DI is larger than that of joint IP.

4. CONCLUSION

This paper presents a novel user-friendly and fast method for extracting dynamic hand measurements automatically. Compared with existing methods, the proposed method can work well for static hands in complex postures and dynamic hands. Specifically, it takes partial hand point clouds as input and outputs dynamic hand measurement values. A novel synthetic dataset consisting of 234K hands with corresponding ground truth measurement is generated for training, and a real-world dataset consisting of 5 hands has been collected to evaluate the proposed method. Experimental results on the real-world data illustrate that our method is able to well record the variation of hand measurement values when the hand is moving. It facilitates the analysis of the range of measurement values due to skin deformations caused by different postures, providing valuable information for hand-centric wearable product design.

In the future, we will explore the evolution of finger length and palm region, which are also important information for hand hand-centric applications. Moreover, our method relies on a neural network approach that processes the depth image frame by frame, which may result in measurement fluctuating from one frame to the next. This limitation can be mitigated by considering the temporal sequence information.

5. REFERENCES

[1] Inga Dabolina, Eva Lapkovska, and Ausma Vilumsone, "Dynamic anthropometry for investigation of body movement comfort in protective jacket," in *Functional Textiles and Clothing*, pp. 241–259. Springer, 2019.

[2] Anke Klepser and Simone Morlock, "4d scanning-dynamic view on body measurements," *Communications in Development and Assembling of Textile Products*, vol. 1, no. 1, pp. 30–38, 2020.

[3] Jordi Uriel, Ana Ruescas, Sofia Iranzo, Alfredo Ballester, Eduardo Parrilla, Alfredo Remón, and Sandra Alemany, "A methodology to obtain anthropometric measurements from 4d scans," in *Proceedings of the 7th International Digital Human Modeling Symposium*. University of Iowa, 2022, vol. 7.

[4] Justin Kok Soon Tan, Shang Wei Song, Jialiu Zeng, and Chih Hung Lo, "A hand-targeted auxiliary personal protective equipment for intervention of fomite transmission of viruses," *Bioengineering & Translational Medicine*, p. e10411, 2022.

[5] P Saradhamani, KP Keerthana, K Karthikeyan, et al., "Personal protective equipment on hand gloves-a review study," in *2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE, 2022, vol. 1, pp. 1697–1701.

[6] Linsey Griffin, Susan L Sokolowski, Elisheva Savvateev, and Arif UI Anwar Bhuyan, "Comparison of glove specifications, 3d hand scans, and sizing of sports gloves for athletes," in *3DBODY. TECH 2019-10th International Conference and Exhibition on 3D Body Scanning and Processing Technologies*, 2019, pp. 109–118.

[7] Mei-ying Kwan, Kit-lun Yick, Lung Chow, Annie Yu, Sun-pui Ng, and Joanne Yip, "Impact of postural variation on hand measurements: Three-dimensional anatomical analysis," *PloS one*, vol. 16, no. 4, pp. e0250428, 2021.

[8] David Joseph Tan, Thomas Cashman, Jonathan Taylor, Andrew Fitzgibbon, Daniel Tarlow, Sameh Khamis, Shahram Izadi, and Jamie Shotton, "Fits like a glove: Rapid and reliable hand shape personalization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 5610–5619.

[9] Siyou Pei, Alexander Chen, Jaewook Lee, and Yang Zhang, "Hand interfaces: Using hands to imitate objects in ar/vr for expressive interactions," in *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 2022, pp. 1–16.

[10] Merwan Achibet, Maud Marchal, Ferran Argelaguet, and Anatole Lécuyer, "The virtual mitten: A novel interaction paradigm for visuo-haptic manipulation of objects using grip force," in *2014 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 2014, pp. 59–66.

[11] Nastaran Nourbakhsh Kaashki, Xinxin Dai, Timea Gyarmathy, Pengpeng Hu, Bogdan Iancu, and Adrian Munteanu, "Automatic and fast extraction of 3d hand measurements using a deep neural network," in *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, 2022, pp. 1–6.

[12] Nastaran Nourbakhsh Kaashki, Xinxin Dai, Timea Gyarmathy, Pengpeng Hu, Bogdan Iancu, and Adrian Munteanu, "A deep learning approach to automatically extract 3d hand measurements," in *2022 7th International Conference on Machine Learning Technologies (ICMLT)*, 2022, pp. 141–146.

[13] Siti Hana Nasir and Olga Troynikov, "Influence of hand movement on skin deformation: A therapeutic glove design perspective," *Applied Ergonomics*, vol. 60, pp. 154–162, 2017.

[14] Nastaran Nourbakhsh Kaashki, Pengpeng Hu, and Adrian Munteanu, "Anet: A deep neural network for automatic 3d anthropometric measurement extraction," *IEEE Transactions on Multimedia*, 2021.

[15] Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed AA Osman, Dimitrios Tzionas, and Michael 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, pp. 10975–10985.

[16] Yu Rong, Takaaki Shiratori, and Hanbyul Joo, "Frankmocap: A monocular 3d whole-body pose estimation system via regression and integration," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 1749–1759.

[17] Michael Gschwandtner, Roland Kwitt, Andreas Uhl, and Wolfgang Pree, "Blensor: Blender sensor simulation toolbox," in *Advances in Visual Computing: 7th International Symposium, ISVC 2011, Las Vegas, NV, USA, September 26-28, 2011. Proceedings, Part II 7*. Springer, 2011, pp. 199–208.

[18] Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas, "Pointnet: Deep learning on point sets for 3d classification and segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 652–660.