Automatic and Fast Extraction of 3D Hand Measurements using a Deep Neural Network

Kaashki, N., Dai, X., Gyarmathy, T., Hu, P., Iancu, B. & Munteanu, A.

Original citation & hyperlink:

Kaashki, N, Dai, X, Gyarmathy, T, Hu, P, Iancu, B & Munteanu, A 2022, Automatic and Fast Extraction of 3D Hand Measurements using a Deep Neural Network. in IEEE International Instrumentation and Measurement Technology Conference. IEEE, pp. 1-6, 2022 IEEE International Instrumentation and Measurement Technology Conference , Ottawa, Ontario, Canada, 16/05/22.
https://doi.org/10.1109/I2MTC48687.2022.9806686

DOI 10.1109/I2MTC48687.2022.9806686
ISSN 2642-2069
ESSN 2642-2077

Publisher: IEEE

Automatic and Fast Extraction of 3D Hand Measurements using a Deep Neural Network

1st Nastaran Nourbakhsh Kaashki
Department of Electronics and Informatics
Vrije Universiteit Brussel
Brussels, Belgium
nknourba@etrovub.be

2nd Xinxin Dai
Department of Electronics and Informatics
Vrije Universiteit Brussel
Brussels, Belgium
xdai@etrovub.be

3rd Timea Gyarmathy
Computer Science Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
gyarmathy.timi@gmail.com

4th Pengpeng Hu
Department of Electronics and Informatics
Vrije Universiteit Brussel
Brussels, Belgium
phu@etrovub.be

5th Bogdan Iancu
Computer Science Department
Technical University of Cluj-Napoca
Cluj-Napoca, Romania
bogdan.iancu@cs.utcluj.ro

6th Adrian Munteanu
Department of Electronics and Informatics
Vrije Universiteit Brussel
Brussels, Belgium
acmuntea@etrovub.be

***Abstract*—Recent advancements in 3D scanning technologies enable us to acquire the hand geometry represented as a three-dimensional point cloud. Providing accurate 3D hand scanning and accurately extracting its biometrics are of crucial importance for a number of applications in medical sciences, fashion industry, augmented and virtual reality (AR/VR). Traditional methods for hand measurement extraction require manual intervention using a measuring tape, which is time-consuming and highly dependent on the operator's expertise. In this paper, we propose, to the best of our knowledge, the first deep neural network for automatic hand measurement extraction from a single 3D scan (H-Net). The proposed network follows an encoder-decoder architecture design, taking a point cloud of the hand as input and outputting the reconstructed hand mesh as well as the corresponding measurement values. In order to train the proposed deep model, a novel synthetic dataset of hands in various shapes and poses and their corresponding measurements is proposed. Experimental results on both synthetic data and real scans captured by Occipital Mark I structure sensor demonstrate that the proposed method outperforms the state-of-the-art methods in terms of accuracy and speed.**

***Index Terms*—hand measurement extraction, template fitting, deep learning, point cloud, 3D scanning, structure sensor Mark I.**

I. Introduction

With the advancements of 3D scanning, providing anthropometric measurements [1] from 3D scanned human bodies became of paramount importance in numerous applications. In this context, *automatic* extraction of specific human biometrics from 3D scans is a challenging problem that has drawn the attention of the research community in the past decade. Although recent studies [2]–[4] were proposed to automatically extract body measurements, they do not address automatic extraction of hand measurements. Since hand measurements play an essential role in many studies ranging from medical science [5]–[8] and culture [9], to glove fabrication [10], [11] and AR/VR applications [12], there is a demand for measuring the hands automatically, reliably, and fast.

Traditional hand measurement extraction methods involve direct contact with the subject because the hand has to be measured by a trained operator using a measuring tape. Manual interventions required in these methods slow down the process and increase the cost. Moreover, the accuracy of such methods is highly dependent on the expertise of the operator. In other words, if the operator does not know well how to find the hand joints and take the measurements, the resulting measurements might not be correct, especially when numerous hands are subjected to get measured. To speed up this process, some studies focused on developing digital tools for measuring the hands based on their corresponding 2D or 3D scans [13], [14]. Although these methods incorporate 2D cameras and 3D sensors, reducing the direct contact with the hand, they still entail a trained operator either for manually adding landmarks to the subject's hands before getting scanned [13], or for manually taking the measurements on the input scan using a digital caliper and measuring tape [13], [14]. The authors in [15] propose an automatic hand measurement extraction method for customized glove production. Although this method is automatic, it cannot produce girth measurements as it only depends on a 2D image captured from the front side of the hand.

With the advent of 3D sensors, several researchers have addressed the automatic biometrics extraction problem with the use of 3D information captured from human bodies [16], [17]. Their results have revealed that, despite of the inherent acquisition noise, accurate body biometrics can be extracted from 3D human scans. In [18], the authors validated the effectiveness of vision-based hand measurement methods by comparing the manual measurements with the measurements taken from the 3D scans captured by the Occipital Structure Sensor and Artec Leo. They found that the results achieved

FWO (G084117), Innoviris (AI43D) research projects, corresponding author: Pengpeng Hu

by both scanners were comparable to those collected using traditional methods.

Deep learning methods have contributed to incredible performance improvements in many domains in computer vision and 3D processing. More recently, with the rise of geometric deep learning, these methods have found their way to solve complex 3D computer vision problems [19], [20]. An essential requirement for training a deep neural network is the use of a large training dataset which cannot be easily gathered. The human hand is among the most complex mechanical human body parts: it consists of 34 muscles and 27 bones, equaling a quarter of the bones in human body, which leads to the fact that the human hand can have various shapes and complicated poses. Scanning and measuring hundreds of thousands of hands in order to gather a training dataset would be prohibitive: extremely time-consuming, expensive, and unpleasant. Moreover, the errors incurred by the 3D registration and processing algorithms as well as the inherent measurement errors introduced by human operators makes such data acquisition paradigm unreliable to be employed for capturing ground truth data. Our solution to this problem is to employ a synthetic dataset based on the parametrized SMPL-X hand model [21]. In this sense, a first contribution of this work is a novel dataset of 3D hand samples in many variant poses and shapes. In addition, we generate a set of measurement points and measurement values for each hand sample in our dataset. This dataset is utilized for training and validating the proposed measurement extraction method.

Inspired by the success of deep neural networks, we propose an encoder-decoder neural network for hand measurement extraction. To the best of our knowledge, the proposed method is the first deep neural network which takes a point cloud of the hand as input and outputs a complete hand mesh as well as its corresponding measurements. The trained deep model is experimentally evaluated with promising results on unseen synthetic samples and real scans captured by Occipital Structure sensor Mark I.

The key contributions of this work are as follows:

- The first deep neural network for automatic extraction of hand measurements from a single hand scan.
- A novel dataset of synthetic hand samples in various shapes and poses with the labeled ground-truth measurement points and values.
- A thorough experimental analysis with promising results of the performance of the proposed method on both unseen synthetic data samples and real scans captured by Occipital Structure Sensor Mark I.

The remainder of this paper is organized as follows: Section II explains the proposed method. The experimental results and our analysis are reported in Section III . Finally, Section IV draws the conclusions of our work.

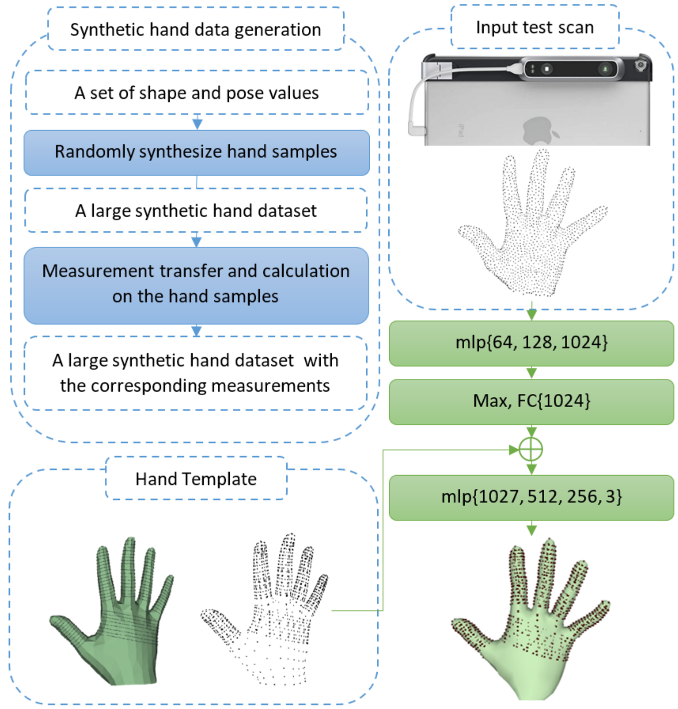

Fig. 1. The framework of the proposed method.

II. Proposed Method

A. Overview

The framework of the proposed method is given in Fig. 1. Synthetic data generation, encoding and decoding are the main three parts of this framework. Before starting training the encoder and the decoder, numerous human hand scans are synthesized in various shapes and poses using the SMPL-X model. Then, an arbitrarily large set of measurements are defined on every synthetic hand sample. Finally, the encoder and the decoder are trained in an end-to-end manner on the proposed dataset.

In the testing phase, given a point cloud of the hand captured by a scanner (e.g. Occipital structure sensor Mark I), the trained encoder extracts the features from the scan and the trained decoder deforms a hand template to reconstruct the hand mesh and the measurement points. The final measurement values are calculated based on the Euclidean distances between the corresponding measurement points. The details of each step are given in the following sections.

B. Data acquisition and template definition

In the data acquisition step, the human hand is scanned by a structure sensor Mark I, as shown in Fig. 2. This sensor is attached to an iPad and adds precise 3D vision to the device. It projects a unique pattern of dots to the object using the infrared laser projector incorporated in the device. The displacement of the dots permits an accurate retrieval of the 3D coordinates of the object surface. The captured 3D scan of the hand is employed as the input to the proposed deep neural network.

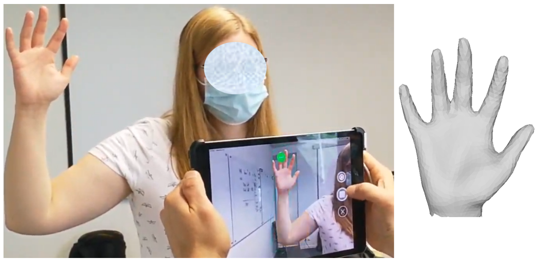

Fig. 2. The hand is getting scanned by the Occipital structure sensor Mark I.

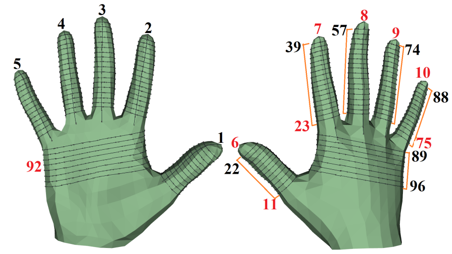

Fig. 3. The hand template accompanied by the measurements and their IDs. The measurements chosen for the experiment analysed in Section III-C are shown with a red index.

To get the measurement values on the input hand scan, a generic hand template onto which the measurements are defined is fitted to the input hand point cloud. Fig. 3 shows the hand template and the corresponding measurement types as defined by the user. To generate this template, a hand mesh is synthesized from the SMPL-X model [21]. SMPL-X is a unified 3D model of the human body, which includes 10475 vertices and 54 joints for the neck, jaw, eyeballs, and fingers. In this study, we keep only the hand which includes 778 vertices. There are 15x3 joint pose parameters θ for each hand. The 3 values per joint represent the rotation for the respective joint, and it usually ranges between -1 and 1. The values outside this range might result in unnatural poses for some joints. The SMPL-X shape parameters β are a vector of 10 values that interact with the model to change the appearance of the body. Their range is [-5, 5] in order to obtain natural looking human shapes. We set the shape and pose parameters of the SMPL-X model in order to synthesize an average-shaped hand mesh $(\boldsymbol{V}, E)$ in the "open-palm-and-finger" pose, as illustrated in Fig. 3. Subsequently, k different measurement types are defined by the user on the template mesh. These measurements are defined once by the user and represent the measurements that are to be automatically extracted from each 3D scan. Each measurement type is defined using an auxiliary plane $(\boldsymbol{n}_i, \boldsymbol{o}_i)$ which intersects the template. The points where the template edges E meet the plane are called measurement points. Formally, the template can be described as:

$$T = \{\boldsymbol{V}, E, \boldsymbol{M}, \boldsymbol{N}, \boldsymbol{O}, \boldsymbol{q}\}, where: \\ \boldsymbol{V} \in \mathbb{R}^{n\times 3}; E \subseteq \boldsymbol{V} \times \boldsymbol{V}; \boldsymbol{M}, \boldsymbol{N}, \boldsymbol{O} \in \mathbb{R}^{m\times 3}; \boldsymbol{q} \in \mathbb{R}^{k}. \tag{1}$$

Here, T is the template, $\boldsymbol{V}$ is a matrix of n 3D vertices, E is the set of edges, and $\boldsymbol{M}$ is a matrix constructed by concatenating the measurement points in all measurement types. The total number of measurement points in $\boldsymbol{M}$ defined on the template is m. To provide a one-to-one correspondence between each measurement point and the normal and origin of the plane onto which the measurement point lays, two matrices of size $(m \times 3)$ are built by duplicating the normal and origin vectors of every measurement type for every corresponding measurement point; these matrices are denoted in (1) as $\boldsymbol{N}$ and $\boldsymbol{O}$ respectively. Finally, $\boldsymbol{q}$ is a vector of k measurement values computed from k measurement types. Each measurement value is the sum of the Euclidean distances between the consecutive measurement points defining a measurement type.

Training the proposed encoder-decoder necessitates a large training dataset. Creating this dataset from real scans is very time-consuming and expensive. To solve this problem, a novel *synthetic* hand dataset is created using the SMPL-X model: (i) 200K training meshes are synthesized by randomly selecting a pose from the pose values β and choosing a shape from the shape values θ ranging from -5 to 5. Similarly, 200 validation and 200 test meshes are synthesized in different shapes and poses; (iv) finally, the measurements are transferred from the template to each mesh sample. Therefore, each sample in our dataset can be described in a way similar to the template in Eq. (1), that is: $S = \{\boldsymbol{V}, E, \boldsymbol{M}, \boldsymbol{N}, \boldsymbol{O}, \boldsymbol{q}\}$.

C. Measurement extraction

As mentioned above, the proposed method consumes the point cloud and deforms the hand template to fit the input accordingly. In the training stage, the input point cloud $\boldsymbol{X}_I$ is an unordered set of points randomly sub-sampled from the vertices $\boldsymbol{V}_I$ and measurement points $\boldsymbol{M}_I$. In the validation and testing stages, the input point cloud $\boldsymbol{X}_I$ is an unordered set of points randomly sub-sampled only from the vertices $\boldsymbol{V}_I$.

The encoder which is a modified version of PointNet [19] consumes the input point cloud $\boldsymbol{X}_I$ and returns a feature vector $\boldsymbol{f}$ describing the input hand sample. To deform the template vertices and measurement points, a new matrix is created by concatenating the feature vector to every points on the template:

$$\boldsymbol{X}_D = \boldsymbol{f} \oplus_c (\boldsymbol{V}_T \oplus_r \boldsymbol{M}_T) \tag{2}$$

where $\oplus_r$ and $\oplus_c$ are the row-wise and column-wise concatenation operators, respectively. The new matrix $\boldsymbol{X}_D$ is the input to the first layer of the decoder. The encoder-decoder learns to deform the template vertices and measurement points using a loss function specially developed for measurement extraction:

$$L(S_I, S_O) = \alpha L_1(S_I, S_O) + \beta L_2(S_I, S_O) + \gamma L_3(S_I, S_O) \tag{3}$$

where S_I and S_O are the input and output samples, respectively; L_1 is the mean square error between the vertices and measurement points of the input and output samples:

$$L_1(S_I, S_O) = \frac{1}{n+m} \sum_{i=1}^{n+m} \|(\boldsymbol{V}_I \oplus_r \boldsymbol{M}_I)_i - (\boldsymbol{V}_O \oplus_r \boldsymbol{M}_O)_i\|_2^2 \quad (4)$$

L_2 is the distance between the output measurement points and their ground truth plane which teaches the network to keep the measurement points planar:

$$L_2(S_I, S_O) = \frac{1}{m} \sum_{i=1}^{m} \|(N_I.M_O)_i - (N_I.O_I)_i\|_2^2 \quad (5)$$

Finally, L_3 is the mean square error between the output and ground truth measurement values:

$$L_3(S_I, S_O) = \frac{1}{k} \sum_{i=1}^{k} \|q_{I,i} - q_{O,i}\|_2^2 \quad (6)$$

Unlike our method proposed in [4], which depends on the relative measurement errors, the proposed loss function gives higher priority to the measurements taken from larger areas. This ensures that we will not have high measurement errors for the larger body areas such as the palm.

III. Experimental Results

A. Datasets

1) Synthetic data: Following the algorithm explained in Section II-B, we have produced 200.000 training data samples, 200 validation, and 200 test samples. The proposed method has been trained using the training and validation sets. During the training process, the data samples are corrupted with white noise of fixed variance to simulate the noise encountered in the 3D scanning process. We define 96 different measurement types on the hand samples, as shown in Fig. 3. Manually measuring real hands for this number of measurements and for thousands of subjects is very difficult, time-consuming, expensive and prone to errors. The proposed synthetic data set alleviates this problem, helping us to train the proposed method using an arbitrarily large number of measurements and to test it on an arbitrarily large set of data samples.

2) Real-world scans: To validate the generalization of the proposed method on real-world data, we scanned 20 hands from 12 men and 8 women using an affordable 3D scanner, namely Occipital structure sensor Mark I. An example of these scans is given in Fig. 2. We chose 11 different measurement types from the measurements defined in [8]. These measurements, shown with a red index in Fig. 3, are listed in Table III. A professional anthropometrist has measured every hand subject. The measurements taken by the anthropometrist serve as ground truth in our experiments on real world data.

B. Evaluating the performance on unseen synthetic data

In this experiment, we evaluate the performance of the proposed method on unseen synthetic data for many hand measurement types. To do so, we trained the proposed network on the synthetic data (Section III-A1) with the training setup given in Table I.

TABLE I
Training Setup

Parameter	Value
Optimizer	Adam
Initial learning rate	0.001
Updating learning rate	Decay in epoch 80 and 90 by 0.1
Weight decay	0
Maximum epoch	100
b (Batch size)	32
n (number of points)	2700

Additionally, we retrained two relevant benchmark networks on the same synthetic dataset: 3D-CODED-TM and AM-DL [3]. The former employs 3D-CODED [22] to deform the hand template according to the input scan. This method uses the hand template that includes only the hand vertices without the measurement points. Once the template is fitted to the input, we transfer the measurements from the template to the deformed template in a post processing step.

The later [3] is an anthropometric measurement extraction method which (i) determines a feature vector ($\boldsymbol{f}$) by performing a multiscale analysis of the input, (ii) deforms the template mesh according to the extracted feature vector, and (iii) extracts the measurements after transferring and refining the measurements from the template to the deformed template. We have tested these methods as well as our proposed method on a synthetic hand test set, which is not included in the training phase.

In order to compare and evaluate the performance of the methods, we utilize two metrics, the signed error e and the signed relative error e_r:

$$e = q_o - q_t \quad (7)$$

$$e_r = ((q_o - q_t)/q_t) * 100 \quad (8)$$

where q_o and q_t are the output and target measurement values.

Table II reports the performance and complexity of our proposed method in comparison to 3D-CODED-TM and AM-DL. For every measurement, the mean absolute error and relative error of the methods are illustrated in Fig. 4. Moreover, the interquartile range (IQR) of signed errors and relative errors are depicted in Fig. 5.

As the results show, our method outperforms the state-of-the-art methods. Thanks to the employed Multi-scale DGCNN encoder, AM-DL performs slightly better than 3D-CODED-TM (reducing the mean absolute error from an average of 5.1mm to 4.5mm and the mean absolute relative error from 8.3% to 7.0%). One notes that the deep neural network used in both methods only deforms the template vertices without involving the measurement points, meaning that both methods require an additional post processing step for measurement extraction and refinement. In contrast, our deep network learns to simultaneously deform both the template vertices and the measurement points. Therefore, it does not require any further

TABLE II
THE ERROR AND COMPLEXITY ON THE SYNTHETIC DATA

metric	3D-CODED-TM	AM-DL	H-Net
Mean Absolute [mm]	5.1	4.5	1.0
Standard Deviation [mm]	6.0	5.7	1.1
Max Absolute [mm]	47.5	45.4	5.6
Mean Relative [%]	8.3	7.0	1.5
Standard Deviation Relative [%]	8.7	8.1	3.0
Max Relative [%]	38.0	36.8	9.1
Time [s]	20.6	27.9	0.8

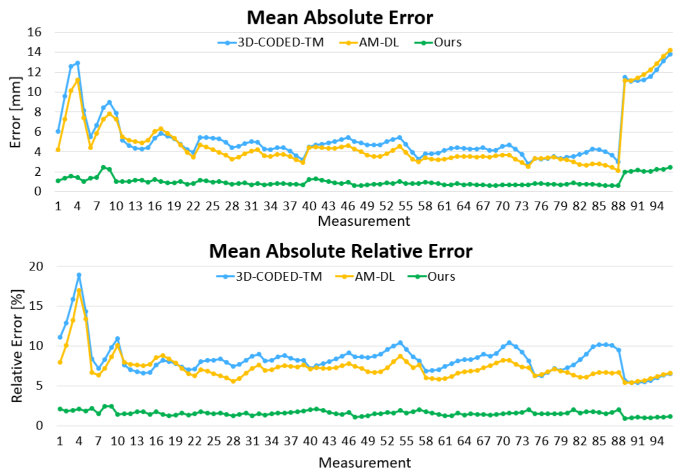

Fig. 4. Mean absolute relative and simple errors on the synthesized dataset.

measurement transfer or refinement after the template deformation; this results in lower complexity (25 times faster than 3D-CODED-TM) and avoids potential error accumulation in sequential template deformation and measurement extraction steps. Moreover, training the network to keep the measurement points planar and to reduce the measurement error leads to higher accuracy for the proposed method (reducing the mean absolute error from an average of 4.5mm to 1.0mm and the mean absolute relative error from 7.0% to 1.5%).

Fig. 4-5 indicate that the proposed method performs well for all the measurements while 3D-CODED-TM and AM-DL lead to high error for the larger areas (palm girths and finger lengths). This is mainly because of the third loss function (Eq. 6) teaching the network to reduce the measurement squared errors for all the measurement types. Since it considers the squared errors and not the relative ones, it focuses on the large errors corresponding to the larger areas more than the small errors belonging to smaller areas (finger girths).

C. Evaluating the performance on real-world scans

In this experiment, we compare the proposed method against 3D-CODED-TM and AM-DL [3] by measuring 11 types of different measurements, shown with a red index in Fig. 3, on the real-world scans. The results are given in Table III.

As the results show, our method outperforms 3D-CODED-TM and AM-DL. As illustrated in Fig. 6, it can be noted that

TABLE III
THE MEAN ABSOLUTE ERROR [MM] ON THE REAL SCANS

Measurement	3D-CODED-TM	AM-DL [3]	H-Net
Thumb finger length (6)	3.0	**0.4**	2.6
Index finger length (7)	4.7	**3.2**	3.8
Middle finger length (8)	5.8	**1.8**	3.4
Ring finger length (9)	3.1	**0.7**	2.8
Little finger length (10)	11.2	9.5	**8.8**
Thumb finger girth (12)	4.1	0.8	**1.7**
Index finger girth (23)	4.3	1.7	**1.5**
Middle finger girth (40)	2.3	11.6	**1.9**
Ring finger girth (58)	**7.3**	8.6	7.4
Little finger girth (75)	7.9	12.4	**6.6**
Palm girth (92)	11.1	13.3	**8.8**
Average	5.9	5.8	**4.5**

the proposed method infers the hand measurements of the real scans correctly, but it performs better on the synthetic test scans than on the real scans. The reasons is that the proposed synthetic dataset contains many complicated poses for the purpose of better generalization. However, when testing on the real-world scans, the subjects are usually asked to perform similar open poses (see Figure 2) to avoid self occlusions during data acquisition. Our model trained on the complicated posed hands may not yield optimal performance for such open hand poses. In future work, one can improve the dataset by removing unwanted poses and adding more shapes to close the performance gap between the results obtained on real scans and on the synthetic data.

IV. CONCLUSIONS AND FUTURE WORK

This paper proposes the first deep neural network for automatic hand measurement extraction. The network takes hand point clouds as input and outputs the reconstructed hand mesh as well as the corresponding measurements. A novel dataset has been developed for training and evaluating the proposed neural network. Each sample in the dataset is accompanied by a set of measurements that serve as ground truth. The experimental results on synthetic test data and real scans show that the proposed method leads to higher accuracy and speed compared to state-of-the-art methods.

In future work, it can be explored if it is possible to further improve the proposed hand dataset used to train the proposed neural network. It is expected that removing unwanted poses and adding more various shapes will further increase the accuracy obtained on real-world 3D scans.

REFERENCES

[1] K. Bartol, D. Bojanić, T. Petković, and T. Pribanić, “A review of body measurement using 3D scanning,” *IEEE Access*, vol. 9, pp. 67 281–67 301, 2021.

[2] L. Markiewicz, M. Witkowski, R. Sitnik, and E. Mielicka, “3D anthropometric algorithms for the estimation of measurements required for specialized garment design,” *Expert Systems with Applications*, vol. 85, 2017.

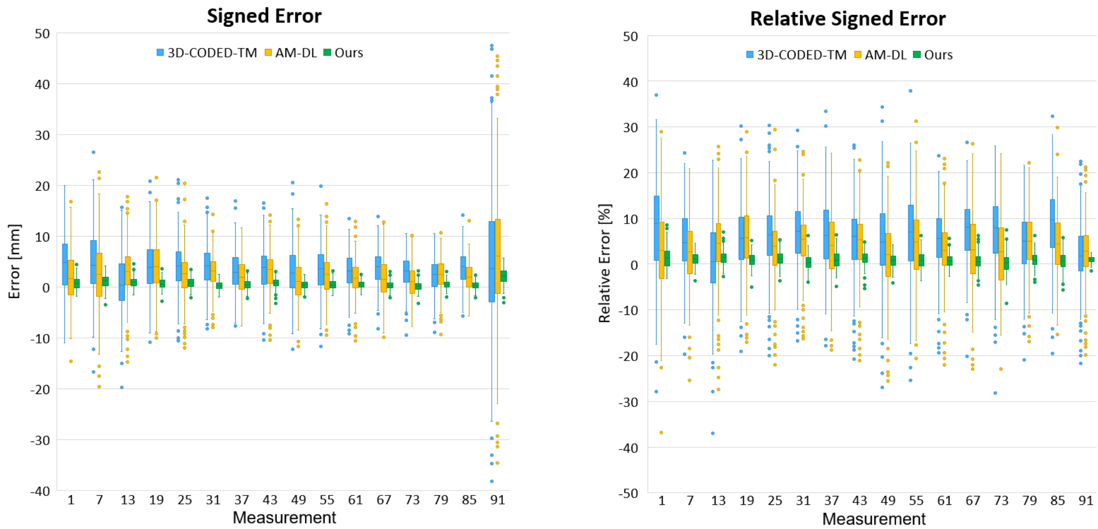

Fig. 5. Relative percent error and signed error on the synthetic dataset.

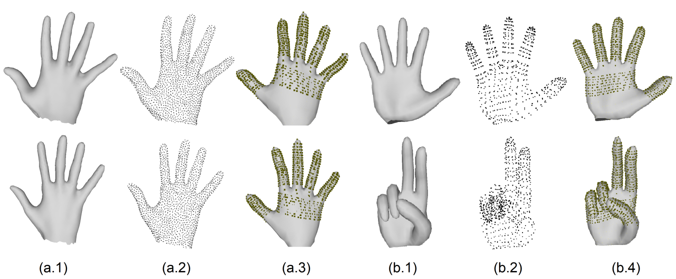

Fig. 6. Visualization results. (a.1) The input ***real* scans**. (a.2) The vertices of (a.1). (a.3) The deformed templates generated by the proposed deep network fed by (a.2). (b.1) The input ***synthetic* meshes**. (b.2) The vertices of (b.1). (b.3) The deformed templates generated by the proposed deep network fed by (b.2).

[3] N. N. Kaashki, P. Hu, and A. Munteanu, "Deep learning-based automated extraction of anthropometric measurements from a single 3-D scan," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2021.

[4] ——, "Anet: A deep neural network for automatic 3D anthropometric measurement extraction," *IEEE Transactions on Multimedia*, 2021, early access.

[5] W. Zhang, J. Robertson, S. Doherty, J. Liu, R. Maciewicz, K. Muir, and M. Doherty, "Index to ring finger length ratio and the risk of osteoarthritis," *Arthritis & Rheumatism*, vol. 58, no. 1, pp. 137–144, 2008.

[6] S. M. Hussain, Y. Wang, D. C. Muller, A. E. Wluka, G. G. Giles, J. T. Manning, S. Graves, and F. M. Cicuttini, "Association between index-to-ring finger length ratio and risk of severe knee and hip osteoarthritis requiring total joint replacement," *Rheumatology*, vol. 53, no. 7, pp. 1200–1207, 2014.

[7] D. Magu, A. Aggarwal, P. Behera, and A. Khurana, "Use of finger length ratio as a marker for knee osteoarthritis: A case-control study of 2,456 patients," *medRxiv*, 2020.

[8] C.-C. Kuo, H.-Y. Kung, H.-C. Wu, and M.-J. Wang, "Developing a hand sizing system for a hand exoskeleton device based on the kansei engineering method," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–13, 2020.

[9] K. R. Hasan, S. Ara, and F. A. M. H. Banna, "Correlation of index finger length (2D) with height, weight and bmi in adult bangladeshi male," *Journal of Enam Medical College*, vol. 7, no. 2, pp. 90–94, 2017.

[10] D. J. Tan, T. Cashman, J. Taylor, A. Fitzgibbon, D. Tarlow, S. Khamis, S. Izadi, and J. 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.

[11] M.-y. Kwan, K.-l. Yick, L. Chow, A. Yu, S.-p. Ng, and J. Yip, "Impact of postural variation on hand measurements: Three-dimensional anatomical analysis," *PloS one*, vol. 16, no. 4, 2021.

[12] C. Wan, T. Probst, L. Van Gool, and A. Yao, "Dual grid net: Hand mesh vertex regression from single depth maps," in *European Conference on Computer Vision*, 2020, pp. 442–459.

[13] W.-W. Wang, K.-W. Lee, S.-Y. Lin, C.-H. Lin, L.-C. Fu, J.-S. Lai, J.-J. Luh, W.-S. Chen, and T.-G. Wang, "A joint localizer for finger length measurements," in *IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, 2013, pp. 111–115.

[14] B. Dunbar and P. Chapates, "Comparison of 3D photogrammetric and laser hand scans to manual measurement methods for eva glove fabrication," in *IEEE Aerospace Conference*, 2019, pp. 1–11.

[15] H. S. Han and C. K. Park, "Automatic hand measurement system from 2D hand image for customized glove production," *Fashion & Textile Research Journal*, vol. 18, no. 4, pp. 468–476, 2016.

[16] N. N. Kaashki and R. Safabakhsh, "3D constrained local model-based face recognition using kinect under variant conditions," in *International Symposium on Telecommunications (IST'2014)*, 2014, pp. 361–366.

[17] ——, "RGB-D face recognition under various conditions via 3D constrained local model," *Journal of Visual Communication and Image Representation*, vol. 52, pp. 66–85, 2018.

[18] E. Seifert and L. Griffin, "Comparison and validation of traditional and 3D scanning anthropometric methods to measure the hand," in *Int. Conference and Exhibition on 3D Body Scanning and Processing Technologies*, 2020.

[19] C. R. Qi, H. Su, K. Mo, and L. 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.

[20] P. Hu, N. N. Kaashki, V. Dadarlat, and A. Munteanu, "Learning to estimate the body shape under clothing from a single 3-D scan," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 6, pp. 3793–3802, 2020.

[21] G. Pavlakos, V. Choutas, N. Ghorbani, T. Bolkart, A. A. A. Osman, D. Tzionas, and M. J. Black, "Expressive body capture: 3D hands, face, and body from a single image," in *Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[22] T. Groueix, M. Fisher, V. G. Kim, B. C. Russell, and M. Aubry, "3D-CODED: 3D correspondences by deep deformation," in *European Conference on Computer Vision (ECCV)*, 2018, pp. 230–246.