

A Simple 3D Scanning System of the Human Foot Using a Smartphone with Depth Camera

Takumi KOBAYASHI^{*1}, Naoto IENAGA¹, Yuta SUGIURA¹,
Hideo SAITO¹, Natsuki MIYATA², Mitsumori TADA²

¹ Keio University, Yokohama, Japan;

² Digital Human Research Group, Human Informatics Research Institute,
National Institute of Advanced Industrial Science and Technology, Japan

DOI: 10.15221/18.161 <http://dx.doi.org/10.15221/18.161>

Abstract

In recent years, online purchasing of clothes and shoes has become increasingly common. Although this is convenient, it can be difficult to choose the correct shoe size. While 3D foot scanners can accurately measure foot size and shape, this expensive and large scale equipment is not generally accessible for personal use, and there is a need for some simple and accurate means of measuring the foot in 3D. Recently developed smartphones with depth cameras enable easier measurement of 3D shapes, and this paper describes a method for measuring foot shape using a 3D point cloud captured from multiple directions by such a camera. As a 3D point cloud can potentially include noise or may omit occluded parts of the foot, we propose the use of a dataset of 3D foot shapes collected by a precise 3D shape scanner. We show how a deformable model can be generated by performing a principal component analysis on this dataset, minimizing error to recover a complete and high-accuracy 3D profile of the entire foot. We tested this method by comparing the 3D shape so produced to the 3D shape measured by the 3D scanner. The proposed method was found to scan foot shape with an error of about 1.13 mm. As demonstrated experimentally, the contribution of our work is in introducing the deformable model of 3D foot shapes based on principal component analysis, so that accurate shape models can be calculated from noisy and occluded 3D point clouds obtained via smartphone input.

Keywords: 3D foot scanning, 3D reconstruction, foot measurement, PCA, smartphone camera, depth camera

1. Introduction

With the rapid growth of E-commerce services, online purchasing of clothes and shoes has become increasingly common, and virtual try-on applications are also emerging. Although these services are convenient, users experience problems in selecting the correct size, and existing systems cannot reliably simulate fittings for various body shapes. In particular, it can be difficult to choose the correct shoe size, as dimensions differ with brand or shoe type; there is evidence that 30% of shoes sold through online shops are returned because the size is incorrect, creating a high demand for a system that can accurately measure the foot's 3D shape. While some shops now provide a 3D scanner to select suitable shoes or to make custom-made products, this technology is not widely used in small stores or for personal use because it is large-scale and very expensive. For this reason, there is a need for some simple and accurate means of measuring the foot's size and shape in 3D.

Smartphones have become an indispensable part of daily life, and along with high-performance cameras, some smartphones now include a depth camera that enables measurement of 3D distances. Additionally, devices equipped with a motion tracking camera can automatically estimate self-position to instantly acquire a wide ranging 3D point cloud. It is expected that practical applications to measure 3D foot shape using a smartphone camera will soon be available. This paper presents a method for measuring foot shape from a 3D point cloud as input captured from multiple directions using a smartphone depth camera, allowing users to measure their own foot shape at home. However, as such 3D point clouds can potentially include noise or may omit parts of the foot due to occlusion, we propose to use a dataset of precise 3D foot shapes collected by a 3D shape scanner to generate a deformable model by performing principal component analysis (PCA). By minimizing any shape error in the deformable model and the 3D point cloud, we aimed to reconstruct the complete 3D shape of the foot with high accuracy. To test this method, we compared the shape so produced to the shape measured by the 3D scanner. The results indicate that the proposed method can scan foot shape with an error of about 1.13 mm.

2. Previous work

Several studies have been conducted to develop instruments for measuring the 3D shape of the human body or foot [1]. However, these 3D scanners are expensive and large-scale, making simple measurement impossible for many users. Estimation of 3D shape from a single human silhouette scanned by a camera is one reported method [2, 3], applying deep learning methods to estimate shape parameters. Similarly, Lunsher et al. reported 3D shape estimation of the human foot using deep learning methods [4]. This method is simple, using only depth maps from a single viewpoint as input. However, one problem of deep learning is the need to prepare a large volume of image or depth map data associated with 3D shape parameters.

In [5], the 3D shape of the foot was measured on a flat scanner using the photometric stereo method. In the same way, [6] performed 3D measurement using depth data captured from the sole of the foot by a depth camera. As these methods could only reconstruct the shape of the sole, it was not possible to measure the 3D shape of the entire foot. Other research has estimated the shape of the human body [7] or the feet [8, 9] using multi-view images as input. Wang et al. [8] and Amstutz et al. [9] made use of deformable models generated by performing PCA on a dataset of 3D foot shapes. These 3D shape data were defined to ensure that each data point had the same anatomical meaning, making it possible to statistically analyze the 3D shape data. In these studies, 3D foot shape was estimated by fitting the deformable model to the actual foot shape. It was expected that sufficient accuracy could be achieved without a large amount of data, but it was necessary to fix multiple cameras in appropriate positions for camera calibration. As this camera calibration process is difficult for ordinary users, this cannot be described as a simple system.

In the method proposed here, we generated the deformable model using the same approach as Amstutz et al. [9], but rather than multiple viewpoint images, we used a 3D point cloud as input. By using a smartphone camera that automatically estimates self-position, it was also possible to acquire a point cloud as input using a handheld device.

3. Proposed method

To realize our goal of designing a simple system to accurately estimate 3D foot shape, we used a deformable model and a 3D point cloud acquired by a smartphone depth camera. The system overview is shown in Fig. 1. For preprocessing, we performed a PCA on the 3D foot shape dataset, in which one of the principal components indicated a particular foot shape. This enabled us to reconstruct the foot shape from principal component parameters to generate the deformable model. To compare the deformable model and the point cloud of the 3D foot shape, the foot portion of the acquired point cloud was extracted by removing the floor and any parts above the ankle. After processing, we aligned the deformable model and the captured point cloud. Finally, we defined the evaluation function referred to in section 3.5, and the actual 3D foot shape was estimated by optimizing the shape parameters.

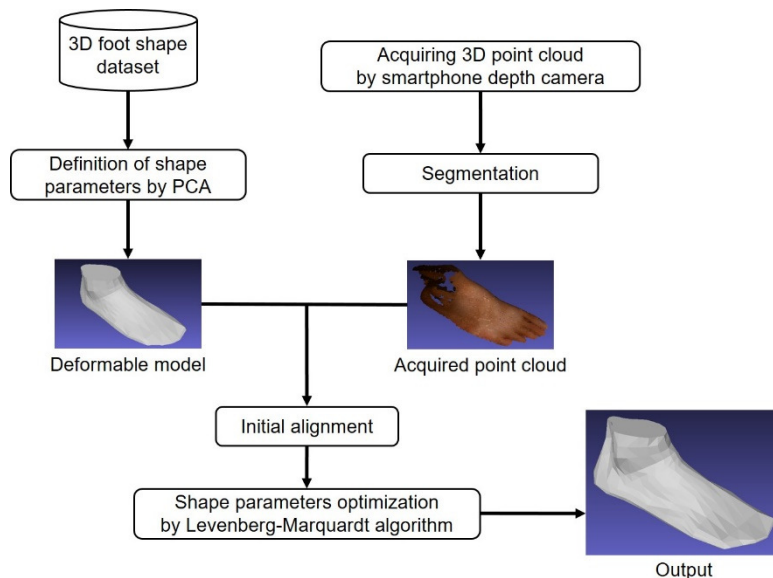


Fig. 1. System overview

3.1. Segmentation

Having acquired a point cloud of the real foot using a smartphone camera, excluding objects other than the foot, the point cloud still contained additional elements such as the floor and the area above the ankle as shown in Fig. 2(a), making point cloud segmentation necessary. To begin, we detected the point cloud of the floor plane and removed it (Fig. 2(b)), using the RANSAC algorithm [10] for plane parameter estimation. Next, we extracted the point cloud within 10 cm of the floor and isolated the point cloud of the foot shape. As outliers were included in the acquired point cloud, we performed the following filtering. For each point, we calculated the average distance to a certain number of neighborhood points and assumed a Gaussian distribution. Any point at an average distance larger than the interval defined by the average value and the standard deviation of a Gaussian distribution was regarded as an outlier and was excluded from the acquired point cloud. This process eventually yielded the point cloud of the foot (Fig. 2(c)).



Fig. 2. (a) Point cloud acquired by smartphone depth camera; (b) after removing plane; (c) after filtering and extracting foot shape

3.2. Dataset

To generate a deformable foot model, we used the dataset created by [11]. Each foot shape in the dataset included mesh information, and the coordinate system was set to a size measurement standard. A sample from the dataset is shown in Fig. 3.

Each foot X_i in the dataset of m feet was described by the same n vertices, as defined by anatomy specialists. The initial shape of the deformable model \bar{X} was defined as the average shape of each 3D foot shape in the dataset as follows:

$$X_i = (x_1, y_1, z_1, x_2, y_2, z_2, \dots, x_n, y_n, z_n)^T, i = 1, 2, \dots, m \quad (1)$$

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_i \quad (2)$$

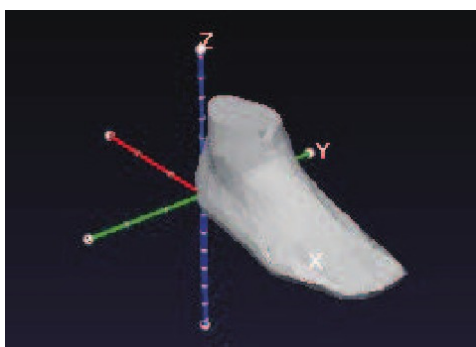


Fig. 3. A sample from the dataset of 3D foot shapes

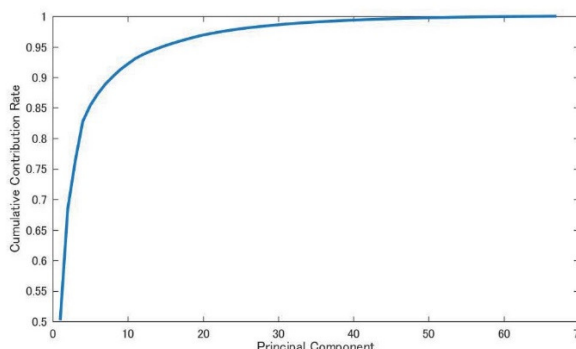


Fig. 4. Cumulative contribution rate

3.3. Definition of shape parameters by PCA

We performed a PCA on the 3D foot shape dataset, making it possible to reproduce each data point in low dimensions while suppressing information loss as far as possible. Here, the i -th principal component is denoted as P_i ($1 \leq i \leq k$) and $k = \min(n, m)$.

The rate of variance of each data point on the i -th principal component to the sum of variance of P_1, P_2, \dots, P_k is called the contribution rate; the cumulative contribution rate can be calculated by adding the contribution rate in order from the first principal component. Fig. 4 traces the cumulative contribution rate for a dataset of 68 men's foot shapes used in this study. In the proposed method, principal components up to the 12th principal component were adopted as shape parameters of the deformable model, following Amstutz et al. [9].

In Fig. 4, the cumulative rate to the 12th principal component exceeds 93%, which means that foot shape variations for 93% of the dataset can be described in the 12th principal component space. The 3D foot shape in the model can be deformed using the following formula:

$$\mathbf{X}' = \bar{\mathbf{X}} + \sum_{j=1}^{12} \omega_j \mathbf{P}_j \quad (3)$$

where each element of the principal component score vector $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_{12})^T$ is the shape parameter coefficient, and the model can be deformed from the average shape by changing the value of $\boldsymbol{\omega}$. In the proposed method, 3D foot shape is estimated by optimizing the principal component score vector $\boldsymbol{\omega}$ so that the difference between the model and the acquired point cloud is minimized.

3.4. Initial alignment

Before deforming the shape of the model, it was first necessary to align the 3D point cloud with the coordinate system of the deformable model. When using the ICP algorithm [12], which is widely employed as a method of 3D point cloud alignment, a nearest point from the other point cloud is searched as a corresponding point for each point of a point cloud and to estimate a rigid transformation that minimizes the distance between corresponding points. However, the ICP algorithm is used in case each point cloud represents the same shape object; it cannot estimate a rigid transformation for sparse point clouds or in cases where initial alignment has failed. For that reason, the ICP algorithm was not used here.

Instead, we performed the initial alignment by rigid transformation of the acquired point cloud according to the coordinate system of the foot shape model. The flow of the initial alignment is shown in Fig. 5. First, PCA was performed on the acquired point cloud \mathbf{V}_0 to obtain a vector that indicates the direction from heel to toe. In section 3.3, PCA was performed on the foot shape dataset to reduce the dimension of shape parameters of the deformable model; in this case, it was used to calculate a vector in the direction of largest variance. The first principal component of the acquired point cloud is denoted as \mathbf{p}_1 . The normal vector of the floor plane estimated in section 3.1 is denoted as \mathbf{z}_{axis} . The acquired point cloud \mathbf{V}_0 is rotated to the same posture as the deformable model \mathbf{F} using the rotation matrix \mathbf{R} as follows:

$$\mathbf{R} = (\mathbf{x}_{axis}, \mathbf{y}_{axis}, \mathbf{z}_{axis})^{-1} \quad (4)$$

Here, $\mathbf{y}_{axis} = \mathbf{z}_{axis} \times \mathbf{p}_1$, $\mathbf{x}_{axis} = \mathbf{y}_{axis} \times \mathbf{z}_{axis}$. The center of gravity of the model \mathbf{F} is then denoted as (x_c^F, y_c^F, z_c^F) , the center of gravity of the rotated point cloud \mathbf{V}_1 is denoted as $(x_c^{V_1}, y_c^{V_1}, z_c^{V_1})$, and the respective centers of gravity are matched by the translation vector \mathbf{t}_1 as follows:

$$\mathbf{t}_1 = \begin{pmatrix} x_c^F - x_c^{V_1} \\ y_c^F - y_c^{V_1} \\ z_c^F - z_c^{V_1} \end{pmatrix} \quad (5)$$

The x coordinate of the point with the minimum x coordinate in the translated acquired point cloud \mathbf{V}_2 is then denoted as $x_m^{V_2}$. The respective heel positions are matched by the transformation vector \mathbf{t}_2 as follows:

$$\mathbf{t}_2 = \begin{pmatrix} -x_m^{V_1} \\ 0 \\ 0 \end{pmatrix} \quad (6)$$

Finally, the scaling transformation is performed on the model F . The x coordinate of the points with the maximum x coordinate in the model F and in the point cloud V_3 are denoted as x_M^F , $x_M^{V_3}$ and $s = x_M^{V_3}/x_M^F$. The respective lengths from heel to toe are rendered equal by the scaling matrix S as follows:

$$S = \begin{pmatrix} s & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & s \end{pmatrix} \quad (7)$$

The model after scaling transformation is denoted as F' .

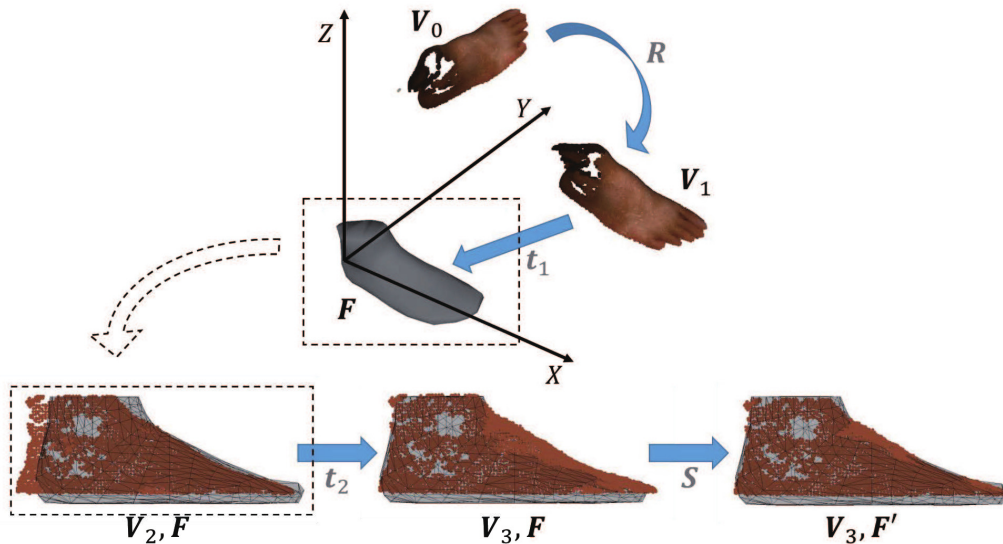


Fig. 5. Flow of initial alignment

3.5. Evaluation function

The evaluation function is defined as the distance error between the model F' and the point cloud V_3 following initial alignment. The 3D foot shape is estimated by deforming the model to minimize the error. The deformable model comprises N meshes. First, the point in the acquired point cloud closest to the inner center $v^{IC}(M_i)$ of each mesh $M_i (1 \leq i \leq N)$ constituting the model is determined as the corresponding point v_i of each mesh. The distance error e_i between the corresponding point and the mesh plane is shown in the following equation:

$$e_i = |(\mathbf{v}_i - \mathbf{v}^{IC}(M_i)) \cdot \mathbf{n}(M_i)| \quad (8)$$

Here, $\mathbf{n}(M_i)$ is the normal vector of each mesh M_i of the deformable model. As the vertex of the deformable model is determined by the principal component score vector ω , the inner point and the normal vector of each mesh is also regarded as a function of ω . The distance error can therefore be described as a function of ω . The evaluation function is defined as the sum of squares of the distance error calculated by equation (8) for all meshes whose corresponding point distance meets a certain threshold. Optimization is performed using the Levenberg-Marquardt method [13].

$$E = \sum_{M_i \in M} e_i^2 \quad (9)$$

3.4.1. Occlusion problems

As shown by the red ellipse in Fig. 6, the point cloud of the sole and around the ankle are omitted in the acquired point cloud due to occlusion. If corresponding points are sought for all meshes without considering occlusion, distances to the wrong corresponding points are included in the optimization calculation. Because the mesh of the deformable model is regular, it is possible to judge whether a mesh is of a certain area, and the meshes constituting the sole part of the deformable model are therefore excluded in advance from calculation of the evaluation function E . In addition, even in cases where the corresponding point distance exceeds a certain threshold, it is regarded as the mesh that corresponds to the point of the missing part.

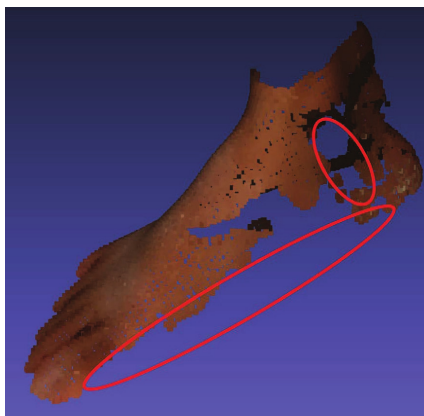


Fig. 6. Parts omitted due to occlusion

4. Experiments and results

The 3D foot shape dataset [11] to generate a deformable model in the experiment comprises 68 male feet, each composed of 295 vertices. When acquiring a 3D point cloud for a foot by oneself, the smartphone is moved to capture the shape of the entire foot. The actual experimental condition is shown in Fig. 7. As the ASUS ZenFoneAR smartphone used in this experiment can automatically perform motion tracking while taking depth map images, it is possible to acquire a wide range of point cloud by linking depth map images to device movement. In this experiment, 3D foot shape is estimated by fitting the deformable model to the input point cloud of the subject's own foot as acquired by the smartphone. For evaluation purposes, we compared the resulting foot shape quantitatively with the high-accuracy measurement of the actual foot shape as captured by a 3D scanner. For measurement of the ground truth, we used Eyewear Laboratory's INFOOT 3D foot scanner [14].



Fig. 7. Experiments in self-acquisition of a point cloud using a smartphone camera

4.1 Initial alignment

Following alignment of the acquired point cloud and the deformable model, the scaling transformation was performed on the model so that the each foot length was equal. Fig. 8 shows the position and posture of each foot shape to be identically aligned.

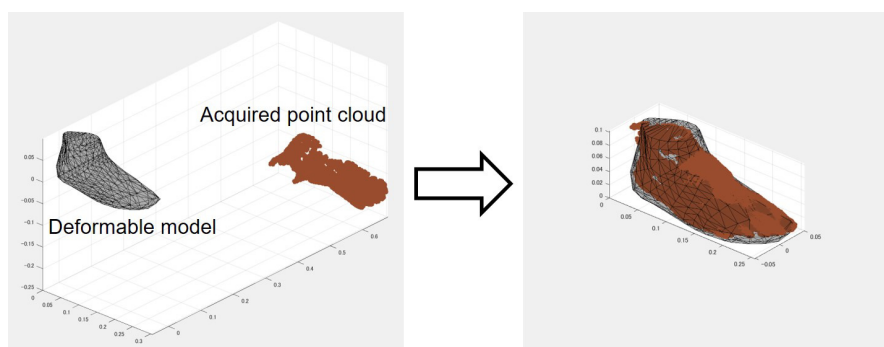
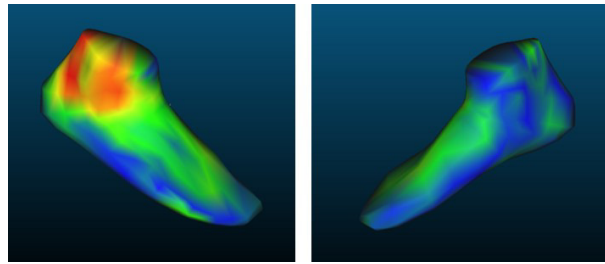
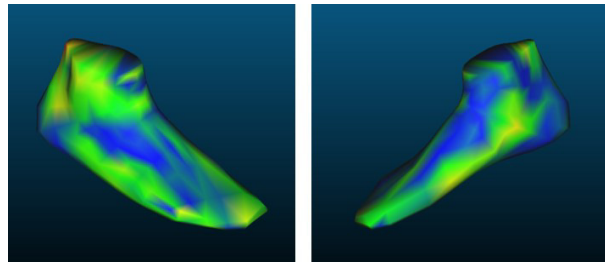


Fig. 8. Alignment outcome



(a)



(b)

Fig. 9. Error distribution (a) before deformation and (b) after deformation

4.2 Reconstruction of 3D foot shape

The result of model deformation is shown in Fig. 9. The mesh color is displayed in the order red, yellow, green and blue. Fig. 9 shows that some parts of the distance error become smaller following deformation. The distribution error of the finally estimated foot shape is visualized in Fig. 10. The entire foot shape is estimated by compensating for the part omitted by occlusion in the acquired point cloud. In addition, the shape of the deformable model is not collapsed by fitting but is deformed while maintaining the natural foot shape.

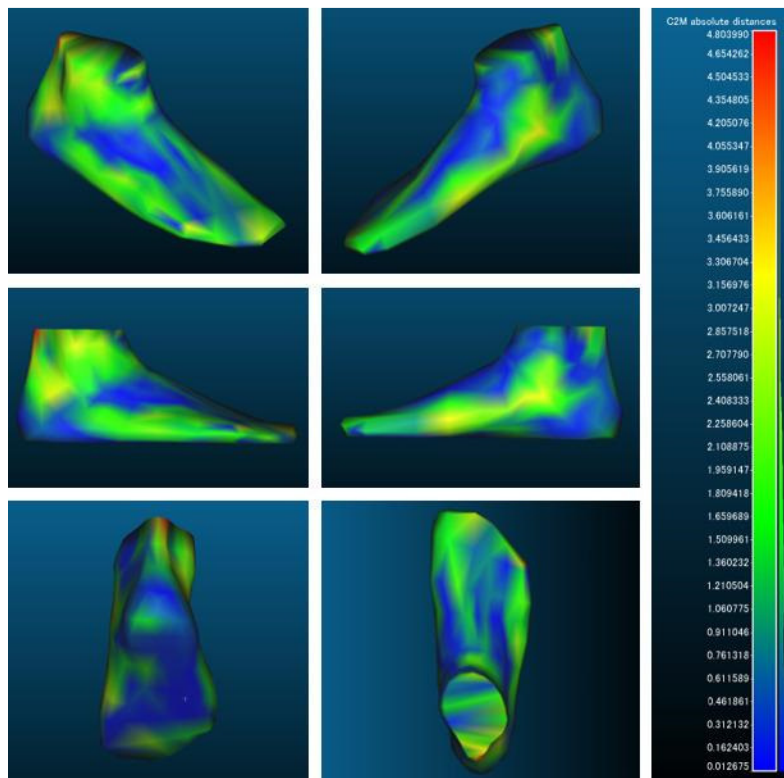


Fig. 10. Error distribution of the computed 3D foot shape in 6 views. Upper left: outer instep. Upper right: inner instep. Middle left: outside of the foot. Middle right: inside of the foot. Lower left: heel and sole. Lower right: upper side of the foot.

5. Discussion

The average distance between the meshes of two 3D shapes was defined as measurement error, and the 3D shape reconstructed by the proposed method was compared to the 3D shape scanned by the 3D foot scanner. The resulting measurement error was 1.13 mm.

As shown in Fig. 10, the top part of the foot was estimated with high accuracy, but the side part incurred a large error. In particular, the error was large for the inner side surface. This is because there are more omitted parts of the acquired point cloud on the inner side surface of the foot than on the instep part, and the correct shape cannot be estimated. Fig. 9 also shows that the error for the side part following deformation becomes larger. In fact, because it is necessary to move a certain distance from the target object to take a depth map image, it is difficult to acquire a finely shaped point cloud. It follows that since the surface of the point cloud in the range approximately 1 cm above the floor has not been acquired due to occlusion, it is transformed and differs from the real shape following optimization. In the experiment, the estimation result depended on the 3D point cloud acquired by the smartphone depth camera.

6. Conclusion

This paper describes a simple 3D scanning system that estimates 3D foot shape by using a 3D point cloud captured by a smartphone depth camera. This system uses a deformable model generated by PCA of a dataset of 3D foot shape data and employs optimization to minimize the distance between the acquired point cloud and the mesh plane of the deformable model. The findings confirm that it is possible to deform the model shape while maintaining a foot-like shape even where there is occlusion or noise in the input point cloud. The evaluation experiment yielded a measurement error of 1.13 mm, confirming that foot shape could be estimated with sufficient accuracy to select the correct shoe size. It was also shown that the entire foot shape can be estimated even when the input is compromised by missing data or noise. Additionally, by acquiring a 3D point cloud using a smartphone depth camera, the required input data for shape estimation can readily be obtained by this simple 3D scanning system.

References

- [1] H. Peng, X. Shang, C. Guo, G. Xiong, T. R. Nyberg, D. Fan and Y. Wang, "A survey on big data for human body shape", in *Service Operations and Logistics, and Informatics*, 2016, pp.145-150, <https://doi.org/10.1109/SOLI.2016.7551677>
- [2] E. Dibra, H. Jain, C. Oztireli, R. Ziegler and M. Gross, "Hs-nets : Estimating human body shape from silhouettes with convolutional neural networks", in *2016 Fourth International Conference on 3D Vision*, 2016, pp.108-117, <https://doi.org/10.1109/3DV.2016.19>
- [3] E. Dibra, H. Jain, C. Oztireli, R. Ziegler and M. Gross, "Human shape from silhouettes using generative hks descriptors and cross-modal neural networks", in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol.5, 2017, <https://doi.org/10.1109/CVPR.2017.584>
- [4] N. Lunscher, and Z. John, "Deep learning anthropomorphic 3D point clouds from a single depth map camera viewpoint", in *IEEE International Conference on Computer Vision Workshops (ICCVW)*, 2017, pp. 689-696, <https://doi.org/10.1109/ICCVW.2017.87>.
- [5] S. Martedi, H. Saito, and Myriam Servieres, "Shape measurement system of foot sole surface from flatbed scanner image", in *Proc. of the 11th IAPR Conference on Machine Vision Applications*, 2009, pp. 338-341.
- [6] S. Chun, S. Kong, K. R. Mun and J. Kim, "foot-arch parameter measurement system using a RGB-D camera", in *Sensors*, Vol.17, No.8, 2017, <https://doi.org/10.3390/s17081796>
- [7] S. Saito, M. Kouchi, M. Mochimaru and Y. Aoki, "Simple system for 3D body shape estimation", in *IEEE 1st Global Conference on Consumer Electronics*, 2012, pp.203-205, <https://doi.org/10.1109/GCCE.2012.6379580>
- [8] J. Wang, H. Saito, M. Kimura, M. Mochimaru and T. Kanade, "Shape reconstruction of human foot from multi-camera images based on PCA of human shape database", in *Fifth International Conference on 3-D Digital Imaging and Modeling (3DIM2005)*, 2005, pp.424-431, <https://doi.org/10.1109/3DIM.2005.73>

- [9] E. Amstutz, T. Teshima, M. Kimura, M. Mochimaru and H. Saito, "PCA-based 3D shape reconstruction of human foot using multiple viewpoint cameras", in *International Journal of Automation and Computing*, Vol.5, No.3, 2008, pp.217-225, https://doi.org/10.1007/978-3-540-79547-6_16
- [10] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography", in *Communications of the ACM*, Vol.24, No.6, 1981, pp.381-395, <https://doi.org/10.1145/358669.358692>.
- [11] M. Mochimaru and M. Kouchi, "Statistics for 3D Human Body Forms", in *Human Factors and Ergonomics Society Annual Meeting*, Vol.44, No.38, 2000, pp.852-855, <https://dx.doi.org/10.1177/154193120004403846>
- [12] Paul J. Besl and Neil D. McKay, "A method for registration of 3-D shapes", in *IEEE Transactions on pattern analysis and machine intelligence*, Vol.14, No.2, 1992, pp.239-256, <https://doi.org/10.1109/34.121791>
- [13] Jorge J. More., "The Levenberg-Marquardt algorithm: implementation and theory", in *International Journal of Computer Vision*, Vol.65, 2005, pp.43-72, <https://dx.doi.org/10.1007/BFb0067700>
- [14] M. Kouchi and M. Mochimaru., "Development of a low cost foot-scanner for a custom shoe making system", in *5th ISB Footwear Biomechanics*, 2001, pp.58-59.