3D Foot Model Construction from Photos, Model Segmentation, and Model Alignment

Wei-Ta Chu
National Cheng Kung University
Tainan, Taiwan
wtchu@gs.ncku.edu.tw

Cheng-Hsi Lin
National Chung Cheng University
Chiayi, Taiwan
abaocc4460@gmail.com

***Abstract*—We present a 3D foot model construction and alignment system. Given photos taken from different viewpoints on a foot, we first construct a 3D foot model by a structure from motion technique. A series of post-processes are proposed to eliminate construction noises, and the primary axes of the foot are determined to facilitate model scaling and alignment. By matching the constructed 3D foot model with the truth model, we show performance variations obtained from different subjects and at different parts. In most cases, the constructed 3D foot model is almost ready for commercial applications, such as on-line shoe shopping and recommendation.**

***Index Terms*—3D foot model, RANSAC, DBSCAN, iterative closest point algorithm**

I. Introduction

Various business models have been proposed for on-line shopping. However, for commodities like clothing and shoe that should match with personal physical characteristics, such as weight, height, size of foot, how to assist consumers to find appropriate commodities is still an on-going problem. Some exciting approaches have been proposed. For example, WANNABY[1] develops mobile AR (augmented reality) applications to makes user virtually try on shoes, jewelry, and nail polish. These AR applications shows the appearance as a user had wore the commodities and narrow down the gap between the virtual world and the physical world. However, these applications don't recommend commodities to individuals according to his/her physical characteristics. In 2017, Invertex[2] develops a 3D foot scanner to connect online and in-store shopping experience. An in-store device scans the feet and sends the constructed 3D foot model to the consumer's mobile phone. A system then guides the consumer to find best-fit shoes using a matching engine. This company was then acquired by Nike in 2018[3]. In May 2019, Nike launched the Nike Fit mobile application that allows users to scan their feet at home and determines the right shoe size[4].

According to the aforementioned news, connecting the virtual world with the physical world would significantly drive the next-generation on-line shopping. However, some issues remain in current solutions. WANNABY's solutions are just for visual appearance but not recommendation, Invertex's solution relies on the in-store 3D scanner to construct 3D foot model, and Nike Fit just measures the size of feet rather than considering rich 3D characteristics. In this work, we would like to develop a system that allows a user to capture multiple photos by their mobile phones at home, constructs a 3D foot model, and then recommends candidate shoes according to his/her 3D foot characteristics.

Contributions of this work are summarized as follows.

- We develop a 3D construction method based on photos captured on the same foot but from different viewpoints. This method is generic to photos captured by different types of cameras.
- In order to remove noise in the constructed 3D model, we propose to adopt the RANSAC (Random Sample Consensus) algorithm to fine the ground plane. After removing points on the ground plane, the DBSCAN clustering algorithm is adopted to cluster remain points. The biggest cluster is usually the foot.
- We propose to match two 3D models based on the ICP (iterative closest point) algorithm. This component is the fundamental for matching the 3D foot model with a given shoe model.

II. System Overview

A. 3D Model Construction

Figure 1 shows the proposed framework for 3D model construction and segmentation. Given a set of foot photos captured by a user's mobile phone from different viewing angles, we adopt the WebODM API[5] to construct the 3D foot model. As shown in Figure 1(a), 16 photos capturing the same foot from 16 different view angles are provided in our experiments. The 3D construction API, i.e., WebODM shown in Figure 1(b), is a free and extendable API originally designed to process aerial images captured by drones. It constructs 3D models and point clouds based on the structure from motion technique. In our case, we can view the mobile phone as a drone, which captures the foot from different perspectives.

Notice that this construction method is not limited to camera models. Quality of construction mainly depends on the relationship (overlapping) between different input images, or

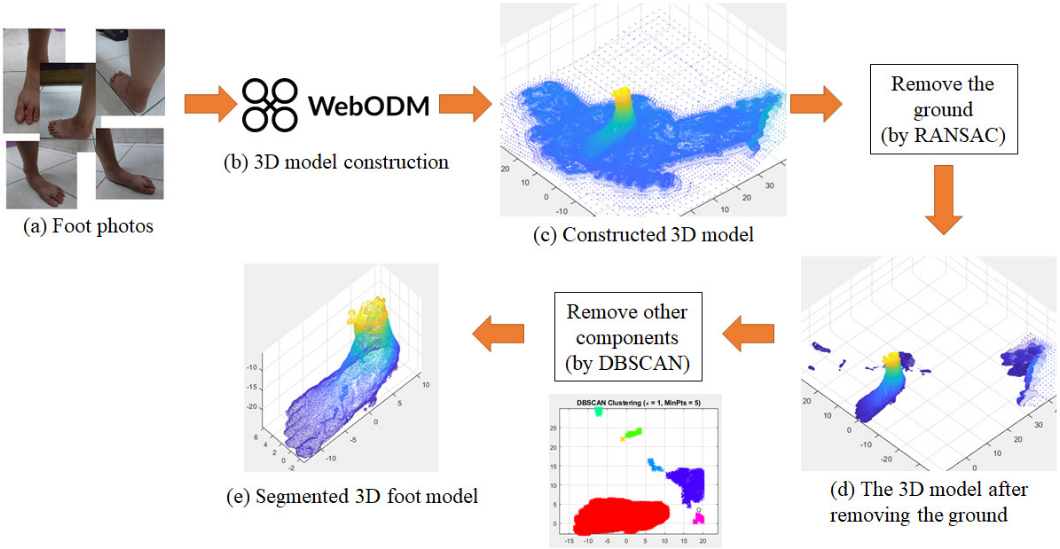
[1] https://wanna.by

[2] www.invertex3d.com

[3] https://news.nike.com/news/nike-invertex-digital-technology

[4] https://www.cnbc.com/2019/05/08/nike-is-launching-nike-fit-to-scan-your-feet-tell-you-your-shoe-size.html

[5] https://www.opendronemap.org/webodm/


WebODM
Remove the ground (by RANSAC)
Remove other components (by DBSCAN)
DBSCAN Clustering (ε = 1, MinPts = 5)

(a) Foot photos

(b) 3D model construction

(c) Constructed 3D model

(d) The 3D model after removing the ground

(e) Segmented 3D foot model

Fig. 1. The framework of 3D model construction and segmentation.

the texture on the target object. The WebODM API is utilized in this work because of its easy implementation. Any other 3D reconstruction methods can also be adopted. 3D reconstruction from multiple images has been a longstanding, challenging, and ill-posed problem in computer vision. Literature survey on this topic can be found in [5] and [6].

B. 3D Model Segmentation

The constructed 3D model is usually quite noisy because the influence of the ground or other objects. We need to automatically remove noisy points and segment the foot, in order to facilitate foot model matching in the succeeding process. To achieve this goal, we first employ the RANSAC (Random Sample Consensus) algorithm [1] to find the equation of the ground. The basic idea is that, as shown in Figure 1(c), lots of points are located at the same plane, which is actually the ground. We employ the RANSAC algorithm to find the "consensus" plane that covers the points the most. This process is plane fitting based on the given points. The RANSAC algorithm is briefly described as follows.

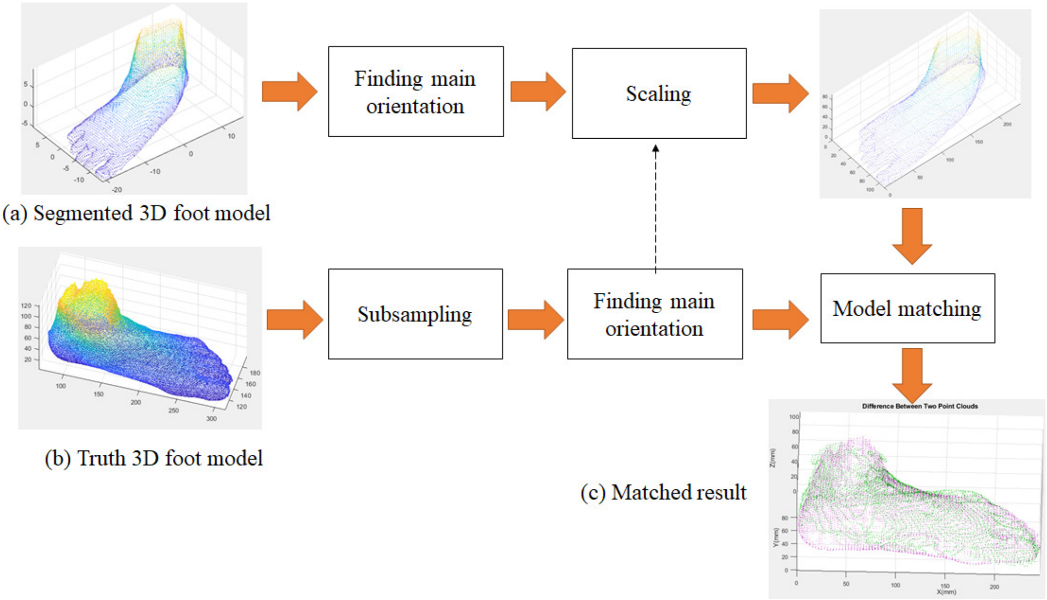
- Step 1: Select a random subset of the original data. We call this subset the *hypothetical inliers*.
- Step 2: A model is fitted to the set of hypothetical inliers. This model conceptually can be viewed as a plane.
- Step 3: All other data are then tested against the fitted model. Those points that fit the estimated model well, according to some model-specific loss function, are considered as part of the consensus set. The loss function we use in this work is simply mean square error between test points and the estimated plane.
- Step 4: The estimated model is reasonably good if sufficiently many points have been classified as part of the consensus set. Generally, if around 80% of the points are classified into the consensus set, we expect that the estimated model well describe the ground plane.
- Repeat Step 1 to Step 4 until some stopping condition is met, e.g., 100 iterations. The estimated model that has the most inliers is taken as the model describing the ground plane.

Assume that the equation describing the ground plane is $ax+by+c=d$, the points located between $ax+by+c=d+\epsilon$ and $ax+by+c=d-\epsilon$ are removed. The parameter ϵ is set empirically. Figure 1(d) shows that many noise on the ground can be removed by the aforementioned process. However, in addition to the foot, noise components out of the plane still remain. We also found that the foot component is actually the largest point cloud. This motivates us to cluster the remaining point clouds by the DBSCAN clustering algorithm [2]. This algorithm automatically determines the number of clusters, based the density information. We pick the point cloud corresponding to the largest cluster as the foot model, as shown Figure 1(e).

C. 3D Model Matching

To facilitate on-line shoe shopping or shoe recommendation, we propose the following application scenario. A user can capture his/her feet by ordinary cameras, and upload photos to the server. A 3D foot model is constructed and segmented by the methods mentioned above. Given this 3D foot model, we would like to recommend shoes that are suitable to this user by considering size, shape, and arch of foot. Assume that the on-line store already has 3D shoe models, matching the foot model with the shoe model is thus the fundamental step to achieve shoe recommendation. Because we don't have digitized shoe models now, in the following we take matching two foot models as the example, one is the segmented foot model from the construction, and one is the truth foot model obtained by a 3D scanner.

Figure 2 shows the 3D model alignment framework. The 3D coordinates of points in the constructed 3D model constitute a



(a) Segmented 3D foot model

(b) Truth 3D foot model

(c) Matched result

Fig. 2. The proposed 3D model alignment framework.

matrix. We apply principal component analysis (PCA) to this matrix, and find the eigenvector associated with the largest eigenvalue. This eigenvector indicates the main orientation of the 3D model, which is usually the orientation of foot length. We apply the same process to the truth 3D model, which was obtained by a high-accuracy 3D scanner. Because the resolution of the truth model is much higher than the constructed model, we first downsample the truth model in order to largely save processing time and required memory. Along the main orientation, we can measure the foot sizes of both models. According to the ratio of the estimated sizes, we scale the constructed 3D model to make it in the same scale as the truth model.

The next step is to find the best rotation and translation parameters to align the two 3D models. Let $\boldsymbol{P} = \{\boldsymbol{p}_i\}$ and $\boldsymbol{Q} = \{\boldsymbol{q}_i\}$ denote the point clouds of the truth foot model and the constructed foot model, respectively. The model matching problem is formulated as an optimization problem, where the loss function we want to minimize is:

$$E = \sum_{i=1}^{N} \|(\boldsymbol{R}\boldsymbol{q}_i + \boldsymbol{t}) - \boldsymbol{p}_i\|, \qquad (1)$$

where N is the number of points in $\boldsymbol{Q}$, $\boldsymbol{R}$ is the rotation matrix, $\boldsymbol{t}$ is the translation vector, and $\boldsymbol{p}_i$ is the point in $\boldsymbol{P}$ that corresponds to $\boldsymbol{q}_i$.

We employ the typical iterative closest point (ICP) algorithm [3] to find the parameters of rotation and translation. For each point $\boldsymbol{q}_j$ in $\boldsymbol{Q}$, find its spatially closest point in $\boldsymbol{P}$. Based on the correspondence between points in $\boldsymbol{P}$ and $\boldsymbol{Q}$, estimate the best parameters such that after rotating and translating $\boldsymbol{Q}$, the transformed $\boldsymbol{Q}'$ is similar to $\boldsymbol{P}$. After this transformation, for each point in $\boldsymbol{Q}'$, we can find its closest point in $\boldsymbol{P}$ again, and proceed the same process. This process iterates until some stop condition meets.

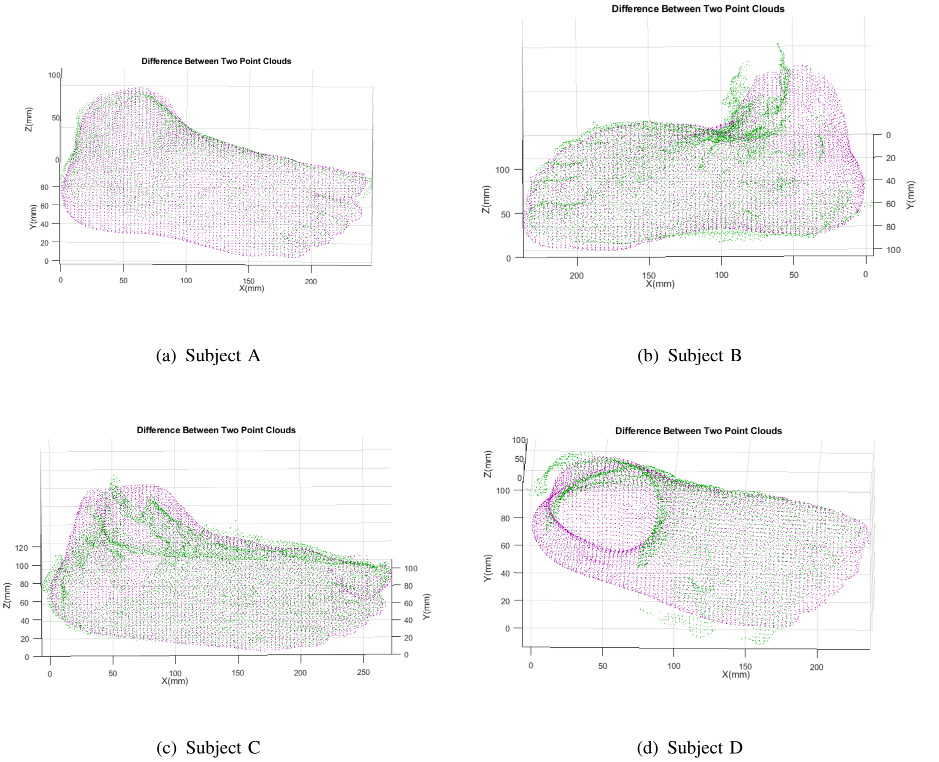
In the aforementioned algorithm, the main computational cost comes from finding the spatially closest point in the reference model for every point in the query model. To make this process efficient, we employ the modified K-D tree algorithm [4] to do closest point computation.

III. Evaluation

As we target at on-line shoe recommendation in the future, we especially evaluate the reconstruction errors at the the widest cross section and the longest cross section of the constructed foot model. The reconstruction errors at these cross sections largely cause uncomfortability of wearing the recommended shoes. In the shoe industry, usually 5mm errors are acceptable. Because we have found the main orientation of each 3D foot model, it is easy to find the two cross sections.

We recruit four subjects for the evaluation. Subjects A, B, and C are male, and Subject D is female. For each subject, we utilize the high-accuracy 3D scanner provided by Footwear & Recreation Technology Research Institute, Taiwan, to obtain the truth 3D foot model. For each subject, cameras equipped in mobile phones were used to capture his/her right foot from 16 different viewing angles. The viewpoints to capture different subjects' feet may be slightly different. Subject A's foot was captured by ASUS Zenfone 4, and other subjects's feet were captured by Apple iPhone SE. The captured 16 images for each subject are jointly considered to construct a 3D foot model, by the processes mentioned in Sec. II.

Table I shows reconstruction errors at the two cross sections in terms of millimeters. Obviously reconstruction errors are highly subject-dependent. As shown in the table, we generally have quite accurate reconstruction in foot length. For subjects C and D, we have problems in removing noise, and have



(a) Subject A

(b) Subject B

(c) Subject C

(d) Subject D

Fig. 3. Sample alignment results of four subjects.

TABLE I
RECONSTRUCTION ERRORS CALCULATED BY COMPARING THE CONSTRUCTED FOOT MODEL AND THE TRUTH FOOT MODEL.

Subjects	Errors in Width	Errors in Length
A	1.68mm	2.94mm
B	0.21mm	2.04mm
C	13.54mm	4.74mm
D	15.94mm	1.11mm

much larger errors in estimating foot width. More robust 3D construction (from photos taken from arbitrary viewpoints) method should be developed in the future.

Figure 3 shows alignment results of these four subjects. The truth foot model's points are in purple, and the constructed foot model's points are in green. As can be seen, the main orientations of two types of foot models are well aligned. However, it is inevitable that the reconstructed foot models have noisy points, especially for Subject C and Subject D.

IV. CONCLUSION

We have presented a series of methods on 3D foot model construction from photos, model segmentation, and model alignment. A structure-from-motion library is used to construct the 3D model from a set of photos capturing the foot from different viewing angles. The constructed 3D model is usually noisy, and thus we propose to utilize the RANSAC algorithm to fine the ground plane, and utilize the DBSCAN algorithm to cluster point clouds. After removing the points nearby the ground plane, and picking the largest cluster, we can get the segmented 3D foot model. Finally, we propose a 3D model matching method by finding the main orientations of different models, and then adopt the ICP algorithm to find the geometric relationship between two foot models. These methods build important foundations to enable on-line shoe recommendation and shopping. In the future, more robust methods for 3D model construction from photos should be devised to make the proposed idea more applicable.

ACKNOWLEDGMENT

This work was partially supported by the Ministry of Science and Technology under the grant 108-2221-E-006-227-MY3, 107-2221-E-006-239-MY2, 107-2923-E-194-003-MY3, 107-2627-H-155-001, and 107-2218-E-002-055.

REFERENCES

[1] Martin A. Fischler and Robert C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," Communications of ACM, vol. 24, no. 6, pp. 381–395, 1981.

[2] Martin Ester, Hans-Peter Kriegel, Jorg Sander, and Xiaowei Xu, "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," Proceedings of International Conference on Knowledge Discovery and Data Mining, pp. 226–231, 1996.

[3] Paul J. Besl and Neil D. McKay, "A Method for Registration of 3-D Shapes," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no. 2, pp. 239–256, 1992.

[4] Zhengyou Zhang, "Iterative Point Matching for Registration of Free-Form Curves and Surfaces," International Journal of Computer Vision, vol. 13, no. 2, pp. 119–152, 1994.

[5] Greg Slabaugh, Ron Schafer, Tom Malzbender, and Bruce Culbertson, "A Survey of Methods for Volumetric Scene Reconstruction from Photographs," Volume Graphics 2001, Eurographics, pp. 226–231, 2001.

[6] Xian-Feng Han, Hamid Laga, and Mohammed Bennamoun, "Image-based 3D Object Reconstruction: State-of-the-Art and Trends in the Deep Learning Era," https://arxiv.org/abs/1906.06543, 2019.