Multiple cameras based method to reconstruct the 3D shape of humans' foot

Edmee Amstutz, Tomoaki Teshima and Hideo Saito
Graduate School of Science and Technology
Keio University
3-14-1 Hiyoshi Kohoku-ku, Yokohama 223-8522, Japan
{eamstutz, tomoaki, saito}@ozawa.ics.keio.ac.jp

Masaaki Mochimaru
Digital Human RC, AIST
2-41-6 Ohmi Koto-ku, Tokyo
125-0064, Japan
m-mochimaru@aist.go.jp

***Abstract*—This article describes a multiple camera based method to reconstruct a 3D shape of a human foot. From a feet database, an initial 3D model of the foot represented by a cloud of points is built. In addition, some shape parameters, which characterize any foot at more than 92%, are defined by using Principal Component Analysis. Then, the initial 3D model is adapted to the foot of interest captured in multiple images based on "active shape models" methods by applying some constraints (edge points' distance, color variance for example). In this article, we also describe the experimental system that is developed for demonstrating the effectiveness of the proposed method. Using the experimental system, we demonstrate that the texture pattern projection on the surface of the object foot is efficient to improve the accuracy of the reconstructed shape by applying the proposed method to a plastic foot model and several real humans' foot.We finally perform a last experiment where the number of cameras is increased, and propose a method to quantify the accuracy of the reconstructed shape.**

I. Introduction

In the past 70 years, there had been a quick and significant change in the humans' height evolution, which is called "secular change" [1]. This has led to an important development of anthropometric methods and projects. The CAESAR project [2], a survey of body measurements for people in three countries, is one of them. One of their objectives is the construction of a database, which could be used to study the variability of people or to design some clothes or equipment specifically for a certain type of people. Actually, the applications of anthropometry are very huge: the most obvious one is effectively the design of things which interact with the human body, but we can also imagine many other interesting application fields like human recognition, or medical assistance (design of prosthesis for instance). In a further future, one of our motivating applications is the possibility for each person to purchase some shoes on the Internet by transmitting his personal foot data, stored on a small ID card.

In this framework, our research focuses on the human foot. Actually, the secular change has particularly led to some huge differences between feet' size and shape among people of various generations. This is the reason why some shoe makers and sellers want to be able to reconstruct especially the 3D foot's shape of each client so that they can develop and sell the most adapted shoes to each person.

One of the most classical methods to measure the shape of the human body is the use of laser, like the method described by Weik [3]. Some systems, which allow translating this type of laser along the foot to acquire the exact shape, are already available in some shoe shops. However, this method presents two problems: first, the system which consists of a laser and some servo to allow the translation is very expensive and difficult to maintain. The second problem is a psychological one: some people are afraid of laser and refuse to use this kind of system. To solve these two issues, our goal is the development of a new method to reconstruct the shape of a human foot in 3D by using only multiple USB cameras and computer vision techniques [4].

Many different methods have been already proposed to reconstruct shapes of various objects. One of them is the silhouette-based volumetric intersection, which consists in intersecting the volumetric cones of space in which the object is constrained to be by each silhouette image. Generally, the results obtained by applying this method are rough because of the difficulty to reconstruct concave shapes. Another way for shape reconstruction is the use of structured light techniques, which change the aspect of the object's surface by projecting a special light pattern on it. One of the most interesting methods for our purpose is called "deformable models", studied by McInerney and Terzopoulos [5]. Cootes et al. [6] proposed to combine the deformable models theory with a statistical approach, which consists in defining the ways the model can be deformed by analyzing the variability of shapes among a training set. Based on this idea and the work on statistical approach developed by Wang and Staib [7], Wang et al. [8], [9] proposed a method to reconstruct the shape of a human foot by using principal component analysis. However, in [9], they use some sophisticated cameras to achieve the reconstruction and they have shown results on modeled data and on one real foot only.

In the following, we describe our method, which is based on similar principles than [9], but can work with simple USB cameras. Then, we insist on the experiment part, where we test our method on a plastic foot model, but also on 17 different human feet. We show the ability of the method to reconstruct feet with various shapes, and we discuss two issues: first, the relative importance of the three criteria used to deform the model. Second, we propose different ways to apply a texture on the foot of interest and compare their efficiencies. Finally, we try to increase the number of cameras and define a method to quantify the accuracy of the computed shape.

II. Method Proposed

Based on the active shape models theory [6], the general approach consists in adapting an initial 3D model of a foot to the real foot, by analyzing the initial model's projection in the 2D images of the real foot and taking into account some constraints.

A. From the feet database

From a feet database, we define the initial 3D model and we determine the 12 most important shape parameters which characterize a foot, by using Principal Component Analysis.

1) Construction of the initial model: The initial 3D model is simply defined as the average of all the feet of the database.

$$AverageFoot = \frac{1}{n}\sum_{n} Foot_{database} \tag{1}$$

Where n is the number of feet in the database.

2) Definition of the shape parameters: Principal Component Analysis is a powerful method to describe the statistical relationship among a set of objects. Here, this method is used among all feet of the database to determine the twelve shape parameters which define any foot at more than 92%.

First, the covariance matrix, which basically represents the variation of each foot from the average foot, is calculated and the eigenvectors of this matrix are computed. Then, the most important variation modes can be found easily: they correspond to the eigenvectors associated to the largest eigenvalues. By applying this theory to our feet database, we have found that the 12 first eigenvectors define the statistical relationship among the feet at 92%, which is the best compromise between computation cost and information gained.

A new foot can be obtained by using the following formula:

$$NewFoot = AverageFoot + PB \tag{2}$$

P is the matrix of the 12 eigenvectors defined earlier in columns, and B is a vector which represents the weights of each variation mode.

B. Camera calibration

We use a multiple cameras system to acquire images of the human foot we want to reconstruct in 3D. The camera calibration is done by using the Calibration Toolbox provided by MATLAB [10]. We use 5 images per camera of a simple marked pattern (viewed from different orientations) to obtain the intrinsic parameters. Secondly, we fixed the multiple cameras on their definitive position, and we take a last image per camera of the pattern in one fixed position, in order to obtain the extrinsic parameters. Since the pattern cannot be seen entirely in some cameras, a translation is applied to correct the origin of the calibration system.

C. Input images of the human foot we want to reconstruct

During the image acquisition, color images (24 bit color depth) are taken. For the 3D shape reconstruction, the binary (foot/background) and edge images of the foot are also necessary. The binary images are obtained using a background segmentation program. Some images sometimes require manual correction. The background segmentation program begins with a background subtraction and uses some morphological operations to smoothen the results. The edge images are directly obtained from the binary images.

D. Deformation of the 3D initial model to fit the real foot in the multiple images

1) First adjustment of the real foot and the model: An affine transformation is applied to make the coordinate system of the calibration correspond with the system in which the initial 3D model is described.

To roughly put the initial model on the foot of interest, an initial rotation and translation are necessary. To calculate the translation, the idea is to make the two gravity centers (initial model and real foot) correspond. Since we don't have any 3D data of the real foot (we have only the information contained in acquired images), we cannot calculate the 3D world coordinates of its gravity center. That is why we define only a 2D gravity center in each image which is basically the barycenter of the white pixels (i.e. foot area in binary images). Then, the distance between projected center of the initial 3D model and 2D gravity center of the foot of interest, for each image, is minimized.

2) Adaptation by successive optimizations: The 3D shape reconstruction of the foot of interest is done as follows: first, the initial 3D model is projected on the 3 types of input images (color, binary and edge images) by using the result of the camera calibration. Secondly, an evaluation function which represents the error between the projected model and the real foot is calculated. Finally, by using some optimization techniques, this evaluation function is minimized by changing the initial model and then repeating these two steps. Modifying the initial model consists in applying a translation, a rotation and a scale and changing the 12 most important shape parameters defined by the Principal

Component Analysis (changing the shape parameters means changing the vector of parameters' weights B defined in formula (2)).

There are two different optimization processes: the first one concerns the 7 pose parameters (3 for the translation, 3 for the rotation and 1 for the scale), and the second concerns the 12 shape parameters. These optimizations are performed by using a least square method.

3) Criteria used in the evaluation function: Three different criteria are used to represent the error between the adapting 3D model and the real foot. Each of them uses one of the three types of input images.

Binary images

The simplest criterion consists in counting the number of vertices of the adapting model which are projected in the foot area in the multiple binary images. A penalization is applied when a vertex is projected outside the image or outside the foot area. For each camera, the first error between the model and the foot of interest is expressed as follows:

$$ErrorBin = \left(1 - \sum_m p_{m/binary}\right) + penalty \qquad (3)$$

Where m is the number of vertices which describe a foot in the database, $p_{m/binary}$ is the pixelfs value at the model point projection in binary images. $Penalty$ is the penalization applied when the model point is projected outside the foot.

Edge images

The distance between edge point of the adapting model (cloud of points) and the nearest edge point of the real foot is minimized. The edge points of the adapting model are found by using the theory of faces and occlusions developed in [8]. Basically, a point is an edge point if it is not projected in any faces of the foot. The multiple edge images directly give the edge points of the real foot. For this criterion, for each camera, the error is expressed with the following formula:

$$ErrorEdges = \sum_m dist\left(EP_{model} - nearestEP_{foot}\right) \qquad (4)$$

Where EP means edge point.

Color images

The last criterion consists in calculating a variance between pixels' color. First, for each point of the adapting model, we calculate the average of pixel's values of the projected point in multiple color images. Then, the difference between the color intensity of projected model points and the average intensity calculated previously is minimized.

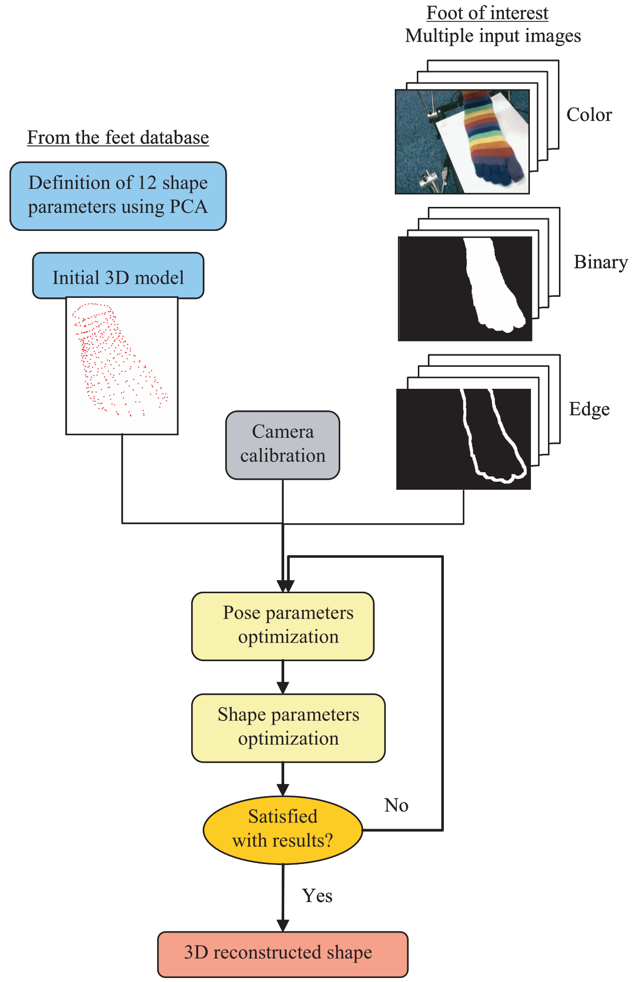


Fig. 1. Overall method for the 3D reconstruction

There are two conditions to use this criterion efficiently. First, the foot has to be textured in color images, so that there can be some differences of pixels' values according to the localization of projected model points. Secondly, this criterion can be applied only if the foot's shape is already quite well approximated. That is the reason why it is applied only for the optimization of the shape parameters.

$$ErrorColor = \sum_m variance_m \qquad (5)$$

$$variance_m = p_{m/color} - average_m \qquad (6)$$

Where $p_{m/color}$ is the pixel's value at the model point's projection, and $average_m$ is the average intensity of pixels' values of the projection of this same point is all color images.

Weights according to the criterion

Some weights are applied to each criterion, so that their influence can be changed to obtain a better result. Those weights $w_{bin}, w_{edges}, w_{color}$ are defined according to the type of optimization, the number of iterations already performed, and the resulted images of the adaptation at the previous step.

$$\begin{aligned} \textit{Total Error} &= w_{bin}\textit{Error Bin} \\ &\quad + w_{edges}\textit{Error Edges} \\ &\quad + w_{color}\textit{Error Color} \end{aligned} \tag{7}$$

4) Occlusion problem: Some vertices can be occluded by other faces, depending the considered view. In this case, the theory developed in [8] is used to find those occluded vertices of the foot according to the considered camera. To avoid the occlusion problem, we simply don't take into account the occluded points in the calculation of the evaluation function.

III. Experiments and Results

The acquisition system is composed of a small box on top of which a human can put his foot Fig.(2). On the four sides of this same box, there are 6 USB cameras which are fixed with long sticks so that the entire foot can be captured from each of them. Each camera is connected to a PC. A central PC allows the user to control all the others.

The database used is composed of 397 right feet of male and female. In this database, each foot is defined by 372 vertices specified by specialists of humans' anatomy.

Three different types of experiments are performed: the first one concerns the shape's reconstruction of a plastic foot model, whereas the second one has for objective to demonstrate the efficiency of the proposed method on human feet with various shapes. By taking into account the analysis of the previous ones, we perform a last experiment with an increased number of cameras and quantify the accuracy of the reconstruction process.

Fig. 2. Photo of the system used

A. On plastic foot model

The first experiment has for objective to test the validity of the general method proposed, and evaluate the influence of criteria' relative weights. This experiment is performed on a plastic foot model, with colored socks so that the color criterion will be effective. The plastic foot model is captured by 5 cameras. The Fig.(4) illustrates the 3D reconstructed shape obtained with our proposed method. Visually, the results of the adaptation are very good. The cloud of red points fits well the plastic foot.

We show that the relative weights between the three criteria have a significant influence on the accuracy of the 3D reconstructed shape obtained. For example, the Fig.(3) shows the reconstructed shape obtained when the binary images criterion (which maximizes the number of model points projected in the white area) is much larger than the edge's criterion (which minimizes the distance between the edge points of the model and the foot). In this case, the scale of the adapting 3D model becomes too small because of the important constraint applied by the binary criterion.

Empirically, we found that the best combination is an edges' criterion with a larger weight than the two other criteria in both optimization processes. For the shape optimization, applying the same weights for the binary and color criteria allows us obtaining good results. The relative weights of each criterion can be changed throughout the reconstruction process, based on the resulted images obtained after each optimization.

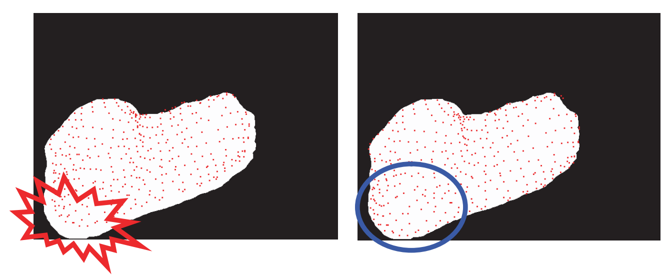


Fig. 3. Illustration of the influence of relative weights between criteria. On the left, a too important binary criterion implies a too small scale. On the right, since the importance is reduced, the results are better.

B. On real human feet

This second experiment is much more significant than the first one on plastic foot model because it deals with many different feet which have various shapes. Moreover, another relevant aspect of this second experiment is that it makes appear some difficulties which cannot be seen in an ideal experiment. For instance, the leg occlusion problem is obviously one of those difficulties which appear in the real

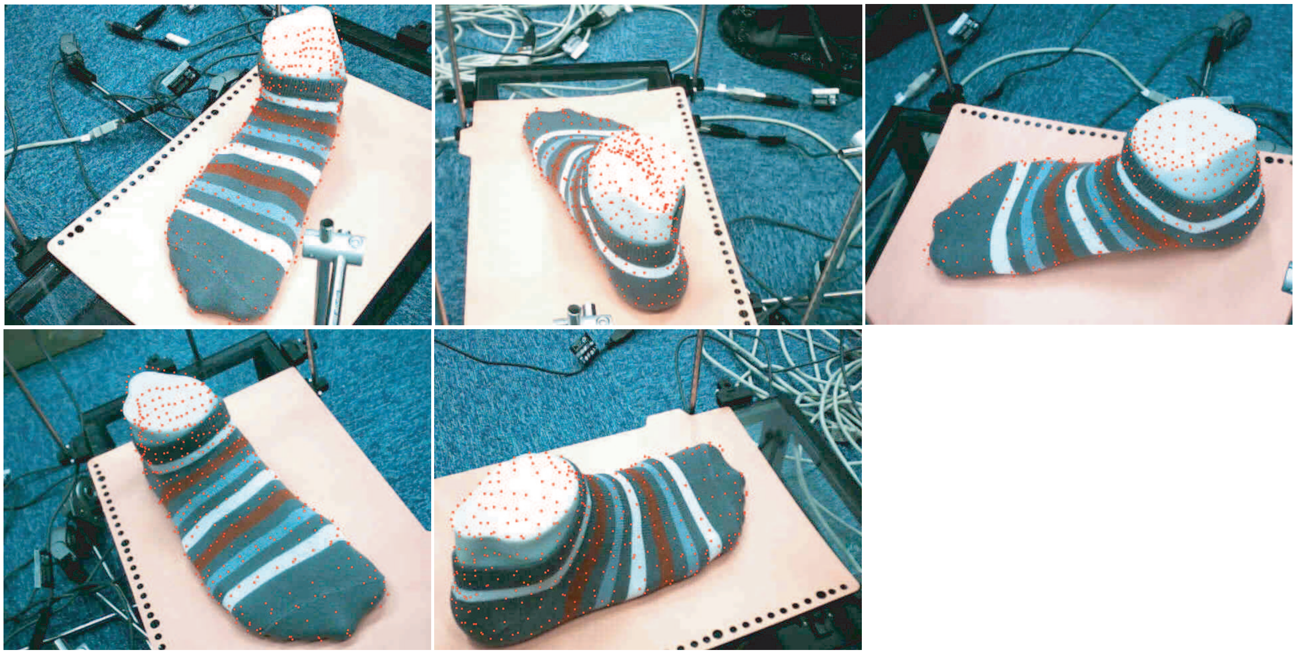


Fig. 4. Results' images of the adaptation of the 3D model (cloud of red points) to the plastic foot model, for the 5 cameras

situation of human feet' reconstruction.

In some cameras, the person's leg occludes a part of his foot. To solve this problem, the leg is manually removed in the multiple input images. Since we cannot remove the exact same part of the leg manually in all images, this process creates some differences in the upper part of the ankle of the foot in the multiple input images. To avoid some errors due to those small differences between images, the points of the initial 3D model which belongs to this upper part of the ankle are not taken into account in the optimization processes. It is not an important problem considered the fact that most of shoes stop just before the ankle.

In this experiment, 16 different persons' feet are captured by 6 cameras. 8 of them are textured by projectors whereas the 8 other persons are wearing colored socks. The results of the adaptation of the initial 3D model to the human foot for one camera are shown in Fig.(5). Fig.(6) illustrates the final results of the reconstruction for one foot by projecting the final 3D model computed in multiple color images.

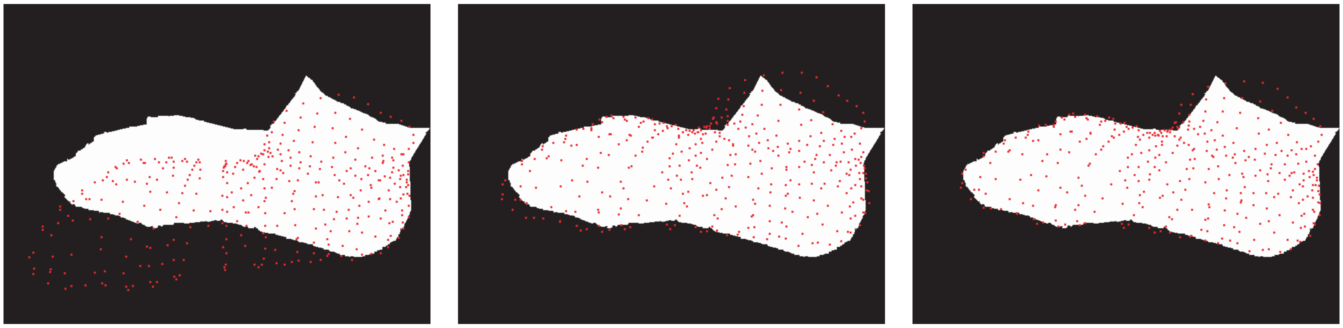
1) Different methods to texture the foot of interest: In this experiment, two different ways are tested to apply a texture on the foot, needed for a good efficiency of the color criterion. The first one is simply to wear some colored socks during the images' acquisition. The second one consists in using two projectors to project a pattern on both sides of the foot. The advantage of using projectors is that the nude foot can be captured, so that the reconstruction can be more accurate. However, the drawbacks are the need to use at least two projectors to texture the entire foot and avoid occluded areas, and the difficulty to choose a good pattern to project.

2) Results' analysis: The visual results (Fig.(6)) show a very good general adaptation of the initial model to the real foot, even if we can observe some model points outside the foot which was not the case for the plastic foot model.

One objective of this experiment is to quantify the accuracy of the computed shapes, because the visual results are not sufficient to conclude on the method's efficiency. Since we don't have any 3D data of the real feet, we can only define a "silhouette difference" from the 2D results' images. We compute binary images which represent the difference between the real foot's binary images and projection of the computed foot on those images. From those binary images, we define a matching difference between the two feet' silhouettes.

The average silhouette differences for feet with colored socks and feet textured by projectors are respectively 6.48% and 5.66%. The average silhouette difference for all the 16 tested human feet is 6.07%. Nevertheless, those silhouette differences can be used only as references, but do not always represent the exact 3D shape error.

This matching difference between silhouettes can be explained by various reasons. The first reason concerns the human's posture. Since the system is not really ergonomic, the person tends to relax his foot and the angle between the foot and the leg becomes too important. The number of cameras may be not sufficient to achieve a very accurate reconstruction, and their position can also be an important issue.



(a) Image obtained after the initial adjustment of the two feet

(b) Image obtained after 1 iteration

(c) Final image obtained after 6 iterations

Fig. 5. Adaptation of the 3D model (cloud of red points) during iteration (1 iteration = optimization of pose and shape parameters once each)

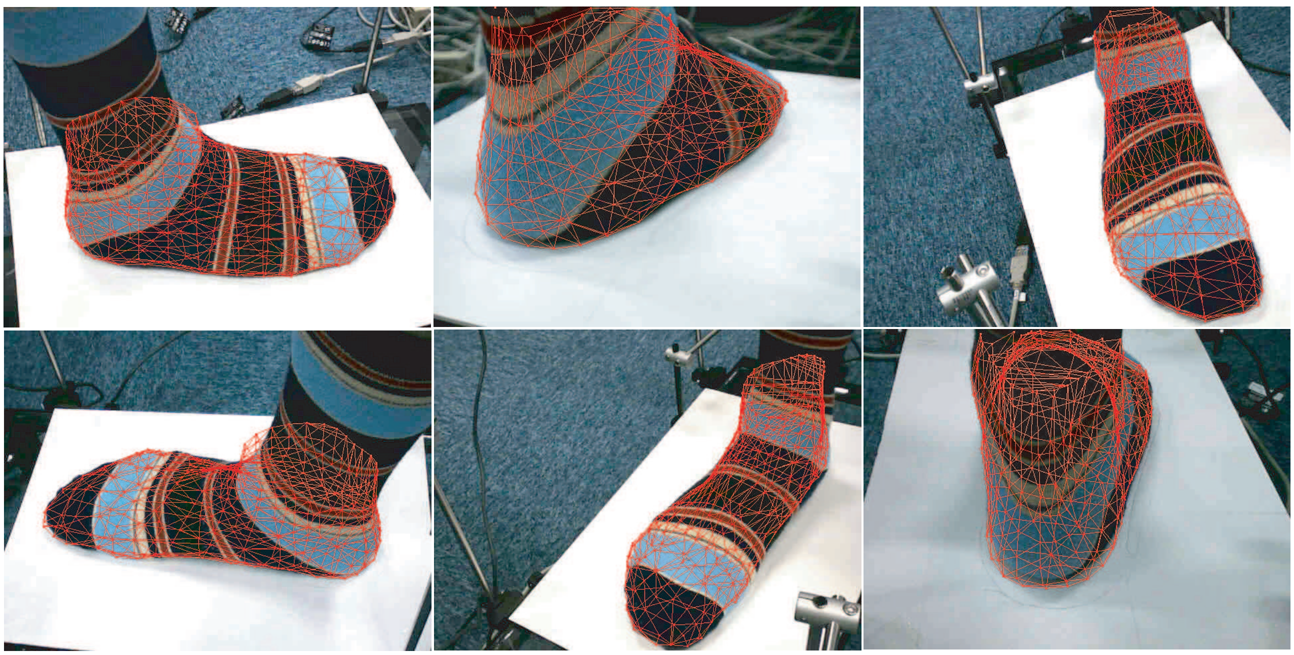


Fig. 6. Final results for a human foot: the 3D computed shape (in red) is projected in the 6 cameras

C. With an increased number of cameras

Based on the results' analysis and the experience of the two previous experiments, we expose the results of a last experiment with an increased number of cameras and a particular attention on the human's posture. We try again the different ways to texture the foot of interest, and compare with the nude foot to evaluate the difference. The plastic foot model and one human foot are captured by 9 cameras. The Fig.(7) shows the 3D shape projected on 4 input images.

The accuracy of the plastic foot model reconstruction can be defined easily: the 3D data of this plastic model are known, so we can calculate a distance error in 3D between the two feet. The accuracy for the plastic foot model reconstruction is very good: from 0.5 to 3 (mm) according to the texturing method. The different results show that pattern projection by the projector significantly improves the shape recovery accuracy.

IV. Conclusion

Based on deformable models and statistical approach, we propose an efficient method to reconstruct the 3D shape of a human foot. Besides, we study different ways to apply a texture on the foot of interest and compare them. Compared to [9], many various experiments have been performed and prove that the method works well even with various human feet, when using simple USB cameras.

Acknowledgment

The authors would like to thank Makiko Kouchi at Digital Human Research Center for providing the plastic foot model. The authors would also like to thank Mr. Kimura, Mr. Utsumi at I-Ware Laboratory Co., Ltd. and Mr. Tsuchiyama at Tachibana Eletech for designing the prototype of the foot

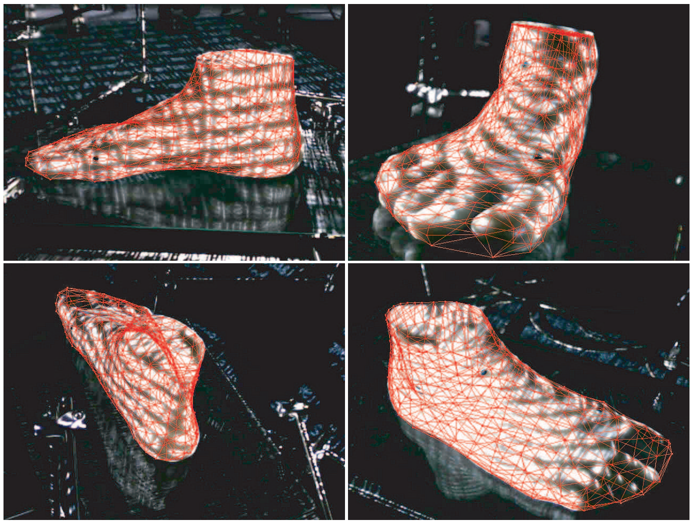


Fig. 7. Final results for a human foot: the 3D computed shape (in red) is projected in the 6 cameras

system. Also, the authors would like to thank to Dr. Jihahui Wang for giving authors many helpful advices.

This work was supported by Grant-in-Aid for Scientific Research (C) (17500119).

REFERENCES

[1] http://www.dh.aist.go.jp/research/centered/index.php.en
[2] K.M. Robinette, H. Daanen, and E. Paquet, "The Caesar Project: A 3D Surface Anthropometry Survey", Proc. in 3D Digital Imaging and Modeling, pp.380-386, 1999.
[3] S. Weik, "A passive full body scanner using shape from silhouettes", ICPR00, pp.1750-1753, 2000.
[4] R. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge Univ. Press, 2000.
[5] T. McInerney and D. Terzopoulos, "Deformable Models in Medical Image Analysis", Proc. of MMBIA, pp. 171-180, 1996.
[6] T.F. Cootes, C.J. Taylor, D.H. Cooper, and J. Graham, "Active shape models-their training and application", Computer Vision Image Understanding, Vol.61, No.1, pp.38-59, Jan. 1995.
[7] Y. Wang and L.H. Staib, "Boundary Finding with Correspondence Using Statistical Shape Models", Proc. in CVPR, pp.338-345, 1998.
[8] J. Wang, "Research on 3D surface reconstruction by deformable model-based methods", 2005.
[9] J. Wang, H. Saito, M. Kimura, M. Mochimaru, and T. Kanade, "Human Foot Reconstruction from Multiple Camera Images with Foot Shape Database", IEICE Trans. Inf. and Syst., Vol.E89-D, No.5, May 2006.
[10] http://www.vision.caltech.edu/bouguetj/calib_ doc/