PCA Based 3D Shape Reconstruction of Human Foot Using Multiple Viewpoint Cameras

Edmée Amstutz[1], Tomoaki Teshima[1], Makoto Kimura[2], Masaaki Mochimaru[2], and Hideo Saito[1]

[1] 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
http://www.hvrl.ics.keio.ac.jp

[2] Digital Human Research Center, AIST
2-41-6 Ohmi Koto-ku, Tokyo, 125-0064, Japan
{makoto.kimura,m-mochimaru}@aist.go.jp
http://www.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 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). We insist here on the experiment part where we demonstrate the efficiency of the proposed method on a plastic foot model, and on real human feet with various shapes. We compare different ways to texture the foot, and conclude that using projectors can improve drastically the reconstruction's accuracy. Based on experimental results, we finally propose some improvements regarding to the system integration.

Keywords: Foot shape reconstruction, Multiple cameras, PCA.

1 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 goals is the construction of a database, which could be used to study the variability of people or to design some equipments specifically for a certain type of people. There are many other interesting application fields for anthropometry like human recognition, or medical assistance (design of prosthesis for instance). In this framework, our research focuses on the human foot. 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

A. Gasteratos, M. Vincze, and J.K. Tsotsos (Eds.): ICVS 2008, LNCS 5008, pp. 161–170, 2008.

shoe makers and sellers want to be able to reconstruct especially the 3D foot shape of each client so that they can develop and sell the most adapted shoes to each person. One of our motivating applications is also the possibility for each person to purchase some shoes on the Internet by transmitting his personal foot data, stored on a small ID card.

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, like IN-FOOT system [4], which allows translating a laser along the foot to acquire the exact shape, have already proven their efficiency. However, this method presents two problems: first, the system which consists of a laser and some mechanical devices to allow the translation is very expensive. 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 system to reconstruct the shape of a human foot in 3D by using only multiple USB cameras and computer vision techniques [5].

One of the most interesting methods for 3D shape reconstruction is called "deformable models", studied by McInerney and Terzopoulos [6]. Cootes et al. [7] 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 [8], Wang et al. [9,10] proposed a method to reconstruct the shape of a human foot by using principal component analysis. However, in [10], they use synchronized cameras to achieve the reconstruction.

In the following, we describe our method, which is based on similar principles than [10], 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 compare different ways to texture the foot of interest. Finally, we propose two methods to improve the accuracy of our reconstruction.

2 Method Proposed

Based on the active shape models theory [7], 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 images of the real foot and applying some constraints.

2.1 From the Feet Database

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

Construction of the initial model. Each foot of the n feet database ($n = 397$) is described by the same m vertices ($m = 372$), which have been defined by anatomy's specialists. v_i represents the position of i^{th} vertex in a 3D model $\mathbf{F}$

that represents a foot shape. The initial 3D model ($\mathbf{F}^{average}$) is defined as the average of each database foot($\mathbf{F}^{database}$).

$$\mathbf{F} = [v_1, v_2, ..., v_m] \tag{1}$$

$$\mathbf{F}^{average} = \frac{1}{n}\sum_{k=1}^{n}\mathbf{F}_k^{database} \tag{2}$$

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 a feet database to determine the shape parameters which mostly define any foot.

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 $\mathbf{F}^{new}$ can be obtained by using the following formula:

$$\mathbf{F}^{new} = \mathbf{F}^{average} + \sum_{j=1}^{12} w_j \mathbf{F}_j^{eigen} \tag{3}$$

where $\mathbf{F}_j^{eigen}$ is the j^{th} column of the $3m \times 3m$ matrix which contains the eigenvectors of covariance matrix, and w_j represents the weight of the j^{th} variation mode.

2.2 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 [11]. We use 5 images per camera of a simple marked pattern 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. To solve the problem of small oscillations which can occur between many captures, one of our future goal is to develop an interface to recompute the extrinsic parameters easily before each utilization.

2.3 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.

2.4 Deformation of the Initial 3D Model to Fit the Real Foot

First adjustment of the real foot and the model. The registration between the coordinates system of the calibration and the one in which the initial model is described is done approximately. After applying some 90 rotations to make axes correspond, we search a translation to roughly put the initial model on the foot of interest. To do this, 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 (but only the acquired images' information), we define a 2D gravity center in each binary image which is basically the barycenter of the white pixels (i.e. foot area). 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.

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) by using the camera calibration results. Then, an evaluation function which represents the error between the projected model and the real foot is calculated. Finally, by repeating some optimization processes, this evaluation function is minimized by changing the initial model. 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 PCA (i.e. changing the vector of parameters' weights w defined in formula (3)). There are two different optimizations: the first one concerns the 6 pose parameters and the scale, and the second one, the shape parameters. These optimizations are performed by using a least square method.

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. This 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 foot area. For each camera, the error between the model and the foot is expressed as follows:

$$e_{bin} = \left(1 - \sum_{m} b_m\right) + nb_{outside}\alpha \qquad (4)$$

Where m is the number of vertices, b_m is the pixel' s value at the model point projection in binary images. $nb_{outside}$ is the number of vertices projected outside the foot area and α is the penalization applied to each of those vertices.

Edge images. For each camera, the sum of the 2D distances between each edge point of the adapting model $EP_{model,m}$ and the nearest edge point of the real foot $EP_{foot,m}$ is minimized. (Concerning the adapting model, a point is an edge point if it is not projected in any faces of the foot.) For each camera, the error is expressed with the following formula:

$$e_{edge} = \sum_{m} \text{dist}\left(EP_{model,m} - EP_{foot,m}\right) \tag{5}$$

Color images. The last criterion consists in calculating a variance between pixels' color. First, for each vertex of the adapting model, we calculate the average of pixel's values of this vertex projected in multiple color images. Then, considering one camera, the difference between the color intensity of each projected vertex in this camera and its average intensity calculated previously, is minimized. 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 vertices. Secondly, the foot's shape must be already quite well approximated. For each camera, the error is the sum of each vertex's variance:

$$e_{color} = \sum_{m} p_m - \bar{c}_m \tag{6}$$

Where p_m is the pixel's value at the model vertex's projection, and $\bar{c}_m$ is the average of pixels' values of this same vertex projected in all color images.

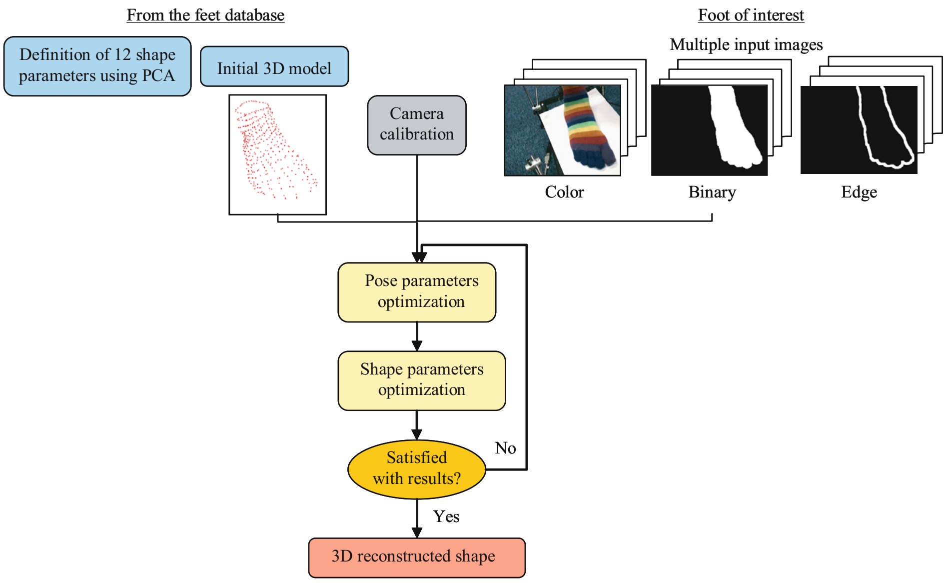

Fig. 1. Overall method for the 3D reconstruction

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 are defined according to the type of optimization and the number of iterations already performed. We explain in more details what is the best combination of the 3 criteria (found empirically) in the experiments and results part.

Occlusion problem. Some vertices can be occluded by other faces, depending on the considered view. To find those occluded vertices, we first calculate if a projected vertex is inside or outside a face, and then if it is inside, we compare the 3D distances from the camera center to the vertex and from the camera center to the face. To avoid the occlusion problem, we simply don't take into account the occluded points in the calculation of the evaluation function.

3 Experiments and Results

Experiments are performed on a plastic foot model and also on 17 human feet. We compare different ways to texture the feet and discuss the influence of criteria weights. Based on the experiments results, we also propose two ways to improve the accuracy of the reconstruction.

3.1 Multi Cameras System

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 box, 10 USB cameras are fixed with sticks so that the entire foot can be captured from each of them. Ten cameras seem to be a good compromise between accuracy (6 are not suffcient to cover the whole foot well) and the computation time. All cameras are connected to one PC, which captures still images of the foot model with a resolution of 640 x 480. Those cameras are not synchronized. The database used is composed of 397 right feet of male and female, which have been captured by laser scanning.

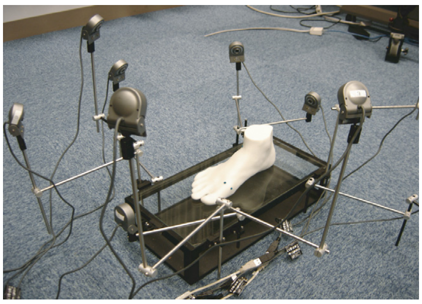

Fig. 2. Photo of the system used

3.2 Different Ways to Texture the Foot of Interest

Those experiments have for objectives to validate the global reconstruction method and to try different ways to texture the foot, needed for a good efficiency of the color criterion. Two different ways are tested: the first one is simply to wear some colored socks during the images' acquisition. The second one consists in using two projectors to apply 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 drawback is the need to use at least two projectors to texture the entire foot and avoid occluded areas.

On plastic foot model. The plastic foot model is captured by 10 cameras. First, we evaluate the influence of criteria' relative weights. Then, we try the different ways to texture the foot of interest, and compare them.

First, this experiment shows that the relative weights between the three criteria have a significant influence on the 3D reconstruction. Fig.(3) shows the reconstructed shape obtained when the binary criterion is too important compared to the edges' criterion. The scale of the adapting 3D model becomes too small due to the important binary constraint. Empirically, we found that the best combination is an edges' criterion with a larger weight than the two other criteria in both optimization processes. The same weights are applied for the binary and color criteria. The relative weights of each criterion can be changed throughout the process.

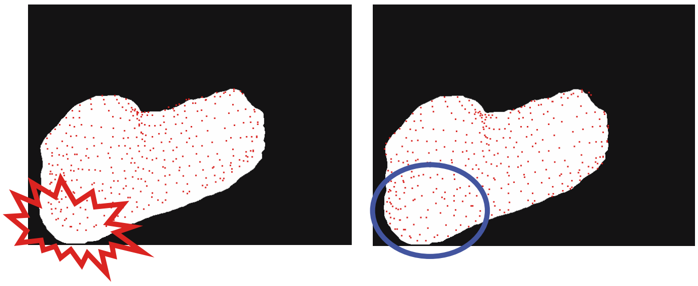

Fig. 3. Illustration of the influence of relative weights between criteria

Secondly, for each texturing method and for the nude foot, the accuracy of the computed shape is calculated. The 3D data of this plastic model are known, so we can calculate a 3D distance error between the computed foot and the real one. By using projectors, we can improve drastically the accuracy: from 3.30mm of error with a nude foot, we obtain 3.06mm with socks and only 0.51mm of error with projectors. Results obtained by using projectors are presented in Fig.(4) where the 3D shape is projected on 6 input images. The reconstruction process, coded in MATLAB, takes 45 minutes when using 10 cameras.

On real human feet. In this second series of experiment, 16 different feet are captured by 6 cameras. 8 of them are textured by projectors whereas the 8 others are wearing colored socks. We check the method's validity while we reconstruct various feet shapes.

Fig.(5) illustrates the final 3D computed shapes of two feet projected in color images. The visual results show a very good adaptation of the model to the real foot, even if we can observe some model points outside the foot. Since we don't have any 3D data of the real feet, we can only define a matching difference (relative surface difference) from the results' images. The average matching differences for feet with colored socks and feet textured by projectors are respectively 6.48% and 5.66%. The first experiment's conclusion is confirmed. Moreover, a last experiment is performed on a human foot with 10 cameras. The accuracy achieved by our method is 1.57mm (obtained by comparison with the 3D data computed

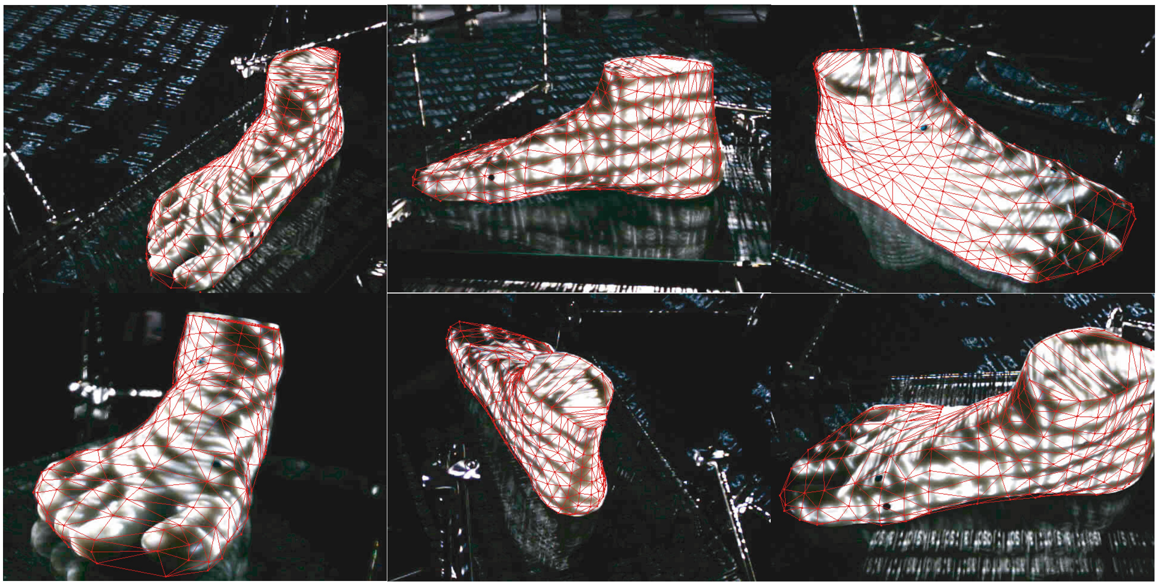

Fig. 4. Final results for the plastic foot model: the 3D computed shape (in red) is projected in 6 cameras

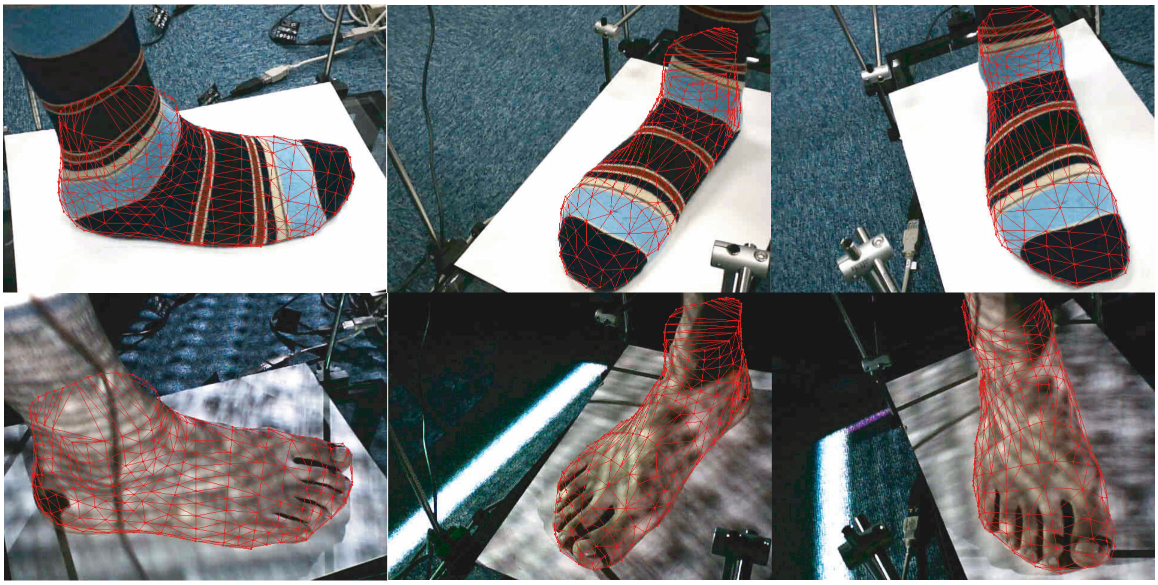

Fig. 5. Final results for 2 human feet: the 3D computed shape (in red) is projected in 3 cameras when using socks (top) or using projectors (bottom)

by a laser based system). From the viewpoint of a computer vision system, the results are very satisfactory, but we still want to improve the reconstruction's accuracy to reach our future objectives.

This experiment makes appear some difficulties which cannot be seen in an ideal experiment: the leg occlusion is one of them. To solve this problem, the leg is manually removed in input images and consequently, vertices which belongs to the upper part of the ankle are not taken into account in the optimizations, so that the manual removal does not induce reconstruction errors. We believe that it is not a major problem considered that most of shoes stop just before the ankle.

3.3 System Integration: Two Ways of Improvement

The error caused in our method can be explained by various reasons. The first one 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 small number of vertices in the initial model (372) can also be an important issue. We are currently developing two solutions to solve those issues.

To improve the system ergonomy, we have placed a small box near the main one, so that the human can put his left foot on it during capture of the right foot. According to humans' anatomy specialists, it is very important to reduce the distance between the 2 feet during the capturing. One drawback of this position is that the left foot will occlude a large part of the right foot. However, we believe that we can overcome this issue by placing the left side cameras very close to the right foot, and reconstruct the foot shape even if we don't see the entire foot in all cameras. We are currently running those experiments.

Moreover, the database used contains male and female feet, where each foot is described by 372 vertices. The sparsity of the model could explain the differences between our computed shape and INFOOT data. Thus, we have decided to create 2 separate databases for male and women, and we have computed a denser model for each foot of the male database by using a software called Homologous Body Modeling, available at the AIST, Japan. This software divides each mesh into 4 new meshes at each iteration, by fitting to some landmarks defined on the foot. We obtain a new model with 1482 vertices instead of only 372. Fig. (6) illustrates the proposed improvements.

(a) Sparse model (b) Dense model (c) Current improved system

Fig. 6. Improvements considering system integration

4 Conclusion and Future Works

Compared to many multiview reconstruction approaches which are based on the matching of features on images, our method takes advantage of the use of a feet database to adapt an initial model and make it fit the captured images. Many experiments have been performed and prove that the method works well even with various human feet, when using unsynchronized USB cameras. Using projectors to texture the foot can improve the reconstruction's accuracy. We are currently working on the densification of database's models and the improvement of the system's ergonomy.

Acknowledgment

The authors would like to thank Makiko Kouchi at Digital Human Research Center for providing the plastic foot model, and also Mr. Kimura, Mr. Utsumi at I-Ware Laboratory Co., Ltd. and Mr. Tsuchiyama at Tachibana Eletech for designing the prototype of the foot system. 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. Robinette, K.M., Daanen, H., Paquet, E.: The Caesar Project: A 3D Surface Anthropometry Survey. In: Proc. in 3D Digital Imaging and Modeling, pp. 380–386 (1999)
3. Weik, S.: A passive full body scanner using shape from silhouettes. In: ICPR 2000, pp. 1750–1753 (2000)
4. Kouchi, M., Mochimaru, M.: Development of a low cost foot-scanner for a custom shoe making system. In: 5th ISB Footwear Biomechanics, Zurich, Switzerland, pp. 58–59 (2001)
5. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2000)
6. McInerney, T., Terzopoulos, D.: Deformable Models in Medical Image Analysis. In: Proc. of MMBIA, pp. 171–180 (1996)
7. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Computer Vision Image Understanding 61(1), 38–59 (1995)
8. Wang, Y., Staib, L.H.: Boundary Finding with Correspondence Using Statistical Shape Models. In: Proc. in CVPR, pp. 338–345 (1998)
9. Wang, J.: Research on 3D surface reconstruction by deformable model-based methods (2005)
10. Wang, J., Saito, H., Kimura, M., Mochimaru, M., Kanade, T.: Human Foot Reconstruction from Multiple Camera Images with Foot Shape Database. IEICE Trans. Inf. and Syst. E89-D(5) (May 2006)
11. `http://www.vision.caltech.edu/bouguetj/calib_doc/`