Estimation of 3-D Foot Parameters Using Hand-Held RGB-D Camera

Y.S. Chen, Y.C. Chen, P.Y. Kao, S.W. Shih, and Y.P. Hung

Dept. of CSIE, National Taiwan University, Taiwan
Dept. of CSIE, National Chi Nan University, Taiwan

Abstract. Most people choose shoes mainly based on their foot sizes. However, a foot size only reflects the foot length which does not consider the foot width. Therefore, some people use both width and length of their feet to select shoes, but those two parameters cannot fully characterize the 3-D shape of a foot and are certainly not enough for selecting a pair of comfortable shoes. In general, the ball-girth is also required for shoe selection in addition to the width and the length of a foot. In this paper, we propose a foot measurement system which consists of a low cost Intel Creative Senz3D RGB-D camera, an A4-size reference pattern, and a desktop computer. The reference pattern is used to provide video-rate camera pose estimation. Therefore, the acquired 3-D data can be converted into a common reference coordinate system to form a set of complete foot surface data. Also, we proposed a markerless ball-girth estimation method which uses the lengthes of two toes gaps to infer the joint locations of the big/little toes and the metatarsals. Results from real experiments show that the proposed method is accurate enough to provide three major foot parameters for shoe selection.

1 Introduction

Shoes were invented to help people get to their goal by protecting and comforting their feet and joints. They have long been a necessity in human being life. They keep evolving according to requirements from fashion, rehabilitation, and sports activities. Nowadays, there are so many choices out there for shoes that choosing a suitable one is usually not easy. In general, people will use their foot sizes to screen out unfitted ones. However, the foot size only reflects the length of one's foot, which is not enough to determine a pair of suitable shoes without repetitive try-on. Therefore, selecting a pair of well-fitted shoes can be very time consuming. Also, the requirement of trying on shoes makes online shoe shopping not convenient. The solution to make choosing a pair of well-fitted shoes easier can be revealed by inspecting how shoe producers design their products. They produce each product for prospects with a specific foot shape which is parameterized with a set of foot parameters. Techniques for measuring those foot parameters can help people choose well-fitted shoes more conveniently and are strongly demanded.

Foot measurement methods can be classified into two categories according to their purposes. The first category includes methods for shoe last design [1]

[2], which require accurate 3-D foot scanners and are out of the scope of this paper. The second category includes methods for measuring foot parameters for assisting shoes selection. Most of the methods in the second category use a 2-D footprint to estimate the length and the width of a foot [3] [4] [5] [6] [7], which are not good enough because a 2-D footprint only provide partial information of a 3-D foot. In 2006, Witana *et al.* proposed a method to measure foot parameters from 3-D foot scanning data [8]. However, their measurement method requires to manually mark 10 anatomical landmarks on the foot surface which requires a foot expert to identify the landmarks correctly.

To help consumers choosing comfortable sports shoes, Asics [9] developed a foot measurement system which uses lasers and cameras to scan the foot surface. To improve the system robustness, their foot measurement system also requires to manually mark three anatomical landmarks on the foot surface. Ildiko Gal Company also provides a 3-D foot scanner consists of a single RGB camera and a turn table. Instead of projecting structure light patterns on the foot surface, they ask users to wear a sock with mesh patterns. Detailed information about their proprietary foot measurement method is unknown.

Shoefitr [10], on the other hand, developed an interesting method for recommending shoes to customers. Their method does not require any hardware device but to ask a user to provide the brand name and the type of the shoes that he/she is wearing. They use the size parameters of the old shoes to search for suitable new shoes. However, when the input brand name of the old shoes is not in their database, their method will fail to provide useful feedback.

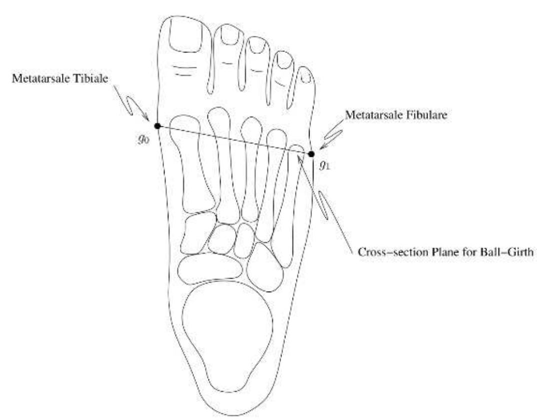


Fig. 1. The ball-girth's cross-section plane.

In this paper, we describe a system, which consists of a low cost Intel Creative Senz3D RGB-D camera, an A4-size printed reference pattern, and a desktop computer, to estimate three foot parameters for shoe selection. The three foot parameters include two basic and one advanced parameters. The two basic parameters are the length and the width of a foot. The advanced parameter is the ball-girth which is defined to be the circumference of the cross-section curve of the foot whose cross-section plane is perpendicular to the ground plane and passing through the Metatarsale Tibiale and the Metatarsale Fibulare as shown in Fig. 1. The two basic parameters can be used to screen out shoes which are either too short or too narrow. However, those two 2-D parameters do not guarantee that the selected shoes are comfortable. Therefore, the foot size is represented with a 3-D vector which includes the length, the width, and the ball-girth.

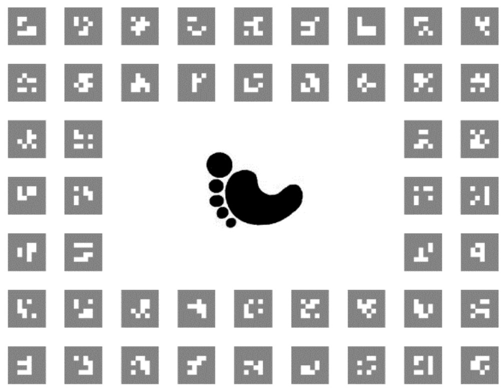


Fig. 2. The reference pattern consists of a set of AR codes.

The reference pattern is an array of AR codes (see Fig. 2). Each AR code is a square pattern encoding a number which can be easily recognized with ArToolKit [11]. This reference pattern can facilitate real-time camera pose tracking. Therefore, we can use a hand-held camera to scan user's foot from different directions. The depth data acquired with the RGB-D camera are transformed to the coordinate system of the reference pattern for 3-D data integration. The proposed method does not require manually marked anatomical landmarks. Therefore, it is more flexible than existing methods and can be operated by users without foot measurement knowledge.

The remainder of this paper is organized as follows. Section 2 introduces the foot measurement system which includes the camera pose estimation, definition of the foot coordinate system, estimation of the foot width/length, estimation of the girth cutting points, and estimation of the ball-girth. Section 3 shows the experimental results of the proposed system. Section 4 concludes this paper.

2 Hand-Held Foot Measurement System

Figure 3 shows the schematic diagram of using a hand-held RGB-D camera to obtain 3-D surface data of the foot. Notably, since the AR codes captured in the

RGB image can be easily identified as shown in Fig. 4, the 3-D coordinates of the four corners of the AR codes can be used to estimate the camera pose with the method described in Section 2.1. Thanks to the depth information provided by the RGB-D camera, the pose estimation can be accomplished in video rate. Therefore, one can transform the acquired 3-D data into a common reference coordinate system so as to perform 3-D data fusion. However, since the RGB-D camera generates about 300,000 3-D data points per 1/30 sec, fusing such large amount of 3-D data will slow down the system reaction time considerably. In fact, we only need a few key RGB-D frames acquired at a few specific locations/orientations to reconstruct the 3-D surface of the foot. Therefore, we propose to use the video-rate camera pose estimation results to guide the user moving the camera toward one of the m pre-specified locations. When the system detects that the camera pose is close enough to the i-th pre-specified camera pose, $1 \leq i \leq m$, an RGB-D image is automatically captured. Then, the system will guide the user to move the camera to the $(i+1)$-th location. After all m RGB-D images are captured. The system automatically perform foot parameter estimation procedure described in Sections 2.2-2.5.

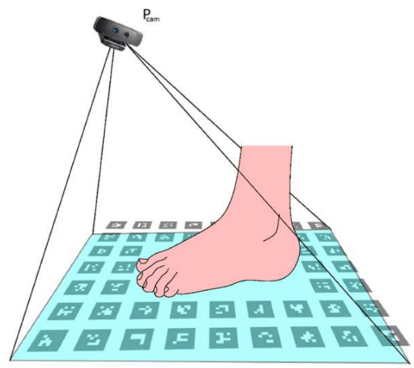


Fig. 3. The Schematic diagram of the proposed foot measurement system using a hand-held camera.

2.1 Camera Pose Estimation

In this work, we adopt the camera model proposed by Tsai [11]. The relationship between a 3-D point, denoted as $\mathbf{p}_c = [x\ y\ z]^\top$, in the camera coordinate system (CCS) and its corresponding image point, denoted as $\mathbf{q}_c = [u\ v]^\top$, is given by

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \sim \mathbf{A} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{1}$$

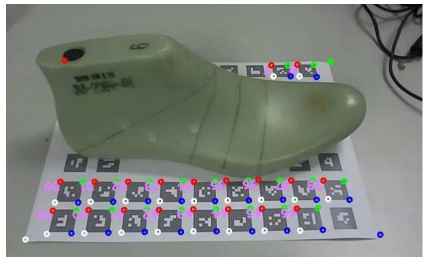


Fig. 4. The identified AR code fiducial marks.

where

$$\mathbf{A} = \begin{bmatrix} \alpha_u & 0 & u_o \\ 0 & \alpha_v & v_o \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

is the calibration matrix which contains the intrinsic parameters, i.e., α_u, α_v, u_0, and v_0, of the Intel Creative Senz3D RGB-D camera all provided by the manufacturer. Nonlinear lens distortion is not considered in this work.

The extrinsic parameters of the camera consist of a rotation matrix $\mathbf{R}$ and a translation vector $\mathbf{t}$ transforming coordinates from the reference coordinate system (RCS) to the CCS. Let $\mathbf{E} = [\mathbf{R} \quad \mathbf{t}]$. Then, the complete camera projection matrix can be represented as follows.

$$\mathbf{C} = \mathbf{A} \cdot \mathbf{E} \tag{3}$$

Camera pose estimation is equivalent to calibrating the extrinsic parameters which are the coordinate transformation matrix from the RCS to the CCS. The RCS is defined with a calibration pattern placed on the foot standing area. The calibration pattern is printed on a planar object and is composed of a set of AR code patterns. For each recognized AR code, we can extract the image coordinates of its four exterior corners of the square AR code pattern. Therefore, we can obtain n image points which are denoted by $\mathbf{q}_{ci}$'s, where $1 \leq i \leq n$. By using the image positions of the recognized AR codes, we can compute an ROI which contains the whole calibration pattern. Then, for each image point inside the ROI, its 3-D coordinates are computed using the depth map. If the 3-D coordinates are valid (a valid depth value is greater than zero), then the 3-D coordinates are recorded. We then use the RANSAC algorithm to find a reliable plane equation in the CCS. For each $\mathbf{q}_{ci}$, we back project the image point to form a 3-D ray as follows

$$\mathbf{r}_{ci} = \mathbf{A}^{-1} \cdot \mathbf{q}_{ci}. \tag{4}$$

The 3-D coordinates of the calibration points in the CCS can be computed by finding the intersection points $\mathbf{p}_{ri}$'s of the 3-D ray $\mathbf{r}_{ci}$'s and the plane π. Since the calibration pattern is designed and printed by ourselves, the coordinates of the corners of every AR code pattern are known. Therefore, we have the corresponding 3-D coordinates of the n image points which are denoted by

$\mathbf{p}_{ti}$s, where $1 \leq i \leq n$. Therefore, for each 3-D point in the CCS, i.e., $\mathbf{p}_{ri}$, we can determine its corresponding point in the RCS, i.e., $\mathbf{p}_{ti}$. The coordinate transformation matrix from the RCS to the CCS can be estimated by using the least-square fitting algorithm proposed by Arun et al. [12] which completes the camera pose estimation process.

2.2 Foot Coordinate System

After capturing m RGB-D images around the m specified locations, the depth maps are converted to point clouds. In the first step, we perform noise removal to delete points belonging to the background. The remaining points are called the foot point cloud Pt_{ft}. In order to measure the correct width and length of the foot, we need to orthogonally project all points in Pt_{ft} onto the x-y plane of the RCS which is the surface of the foot standing area. Then, we perform a 2-D principal component analysis (PCA) to find the long-axis and the short-axis of the 2-D foot print (see Fig. 5). The long-axis and the short-axis of the foot print are used to define the y-axis and the x-axis of a new 3-D coordinate system, called the foot coordinate system (FCS).

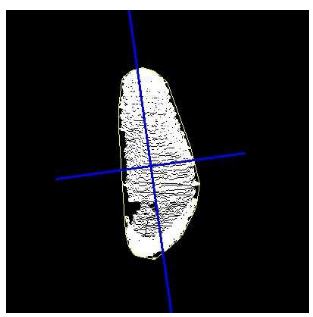


Fig. 5. The Foot axes computed with PCA of the 2-D foot print points.

2.3 Estimation of the Foot Width and the Foot Length

By transforming the 2-D points in the foot print area to the FCS, we have Pt'_{ft}. The width and the length of the foot is estimated by evaluating the width and length of the bounding box of Pt'_{ft}. Figure 6 shows a schematic diagram of the foot width and the foot height estimation method. The x-axis and the y-axis of the FCS shown in Fig. 6 are denoted by X_{ft} and Y_{ft}, respectively. Also, the leftmost, the rightmost, the topmost, and the bottommost points of the foot in the FCS is marked with Pt_{W0}, Pt_{W1}, Pt_{H0} and Pt_{H1}, respectively.

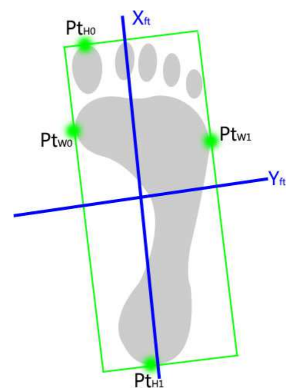


Fig. 6. The evaluation of the width and the length of the footprint.

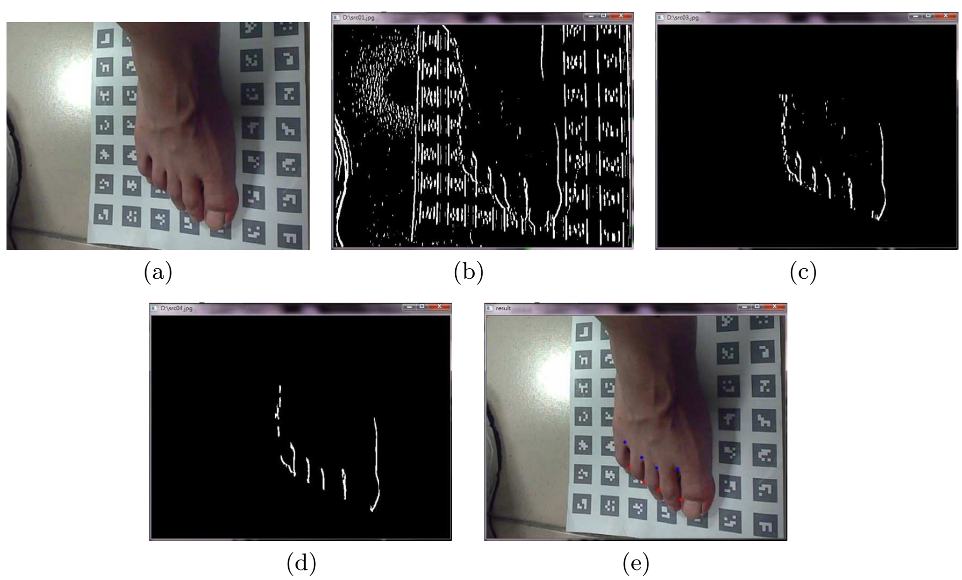


Fig. 7. The process of finding the length of the toe gaps: (a) the original RGB image, (b) the Canny edge detection result, (c) clear edge pixels outside ROI, (d) noise removal result, and (e) estimated endpoints of toes gaps.

2.4 Estimation of the Girth Cut Points

According to the definition of the ball-girth, before we can compute it, we need to determine the girth cut points $\mathbf{g}_0$ and $\mathbf{g}_1$, which are the joints of the metatarsals and the big/little toes, respectively. The main challenge to find those two points is that the foot bone positions are not visible. Therefore, existing methods require an operator to manually mark those two points. In this work, we propose to infer the girth cut points by using the length of the toe gap next to the big/little toe. Since the toe gap can be very narrow, it is unreliable to estimate the toe gap length. Therefore, we determine the toe gap length solely with RGB image. First, Canny edge detection algorithm is applied to the intensity of the RGB image (see Fig. 7(b)). Then, all the edge pixels outside the ROI of the foot stand area are cleared (see Fig. 7(c)). Perform connected component analysis and remove those blobs smaller than an empirically determined threshold value (see Fig. 7(d)). Finally, we can determine the end points of the toes gaps as shown in Fig. 7(e). The end points of the big/little toe gaps are denoted as V_0 and V_1 (see Fig. 8), respectively.

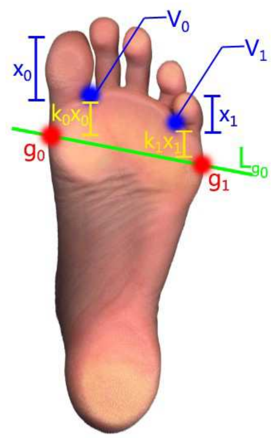


Fig. 8. The estimation of the girth cutting points.

Let x_0 and x_1 denotes the length of the toe gaps next to the big toe and the little toe, respectively. Points V_0 and V_1 are offseted toward the heel by k_0x_0 and k_1x_1 where k_0 and k_1 are the self-defined constants. Finally, a straight line connecting the two offset points is computed. The intersection points of the footprint contour and the straight line are the girth cutting points (see Fig. 8).

2.5 Estimation of the Ball-Girth

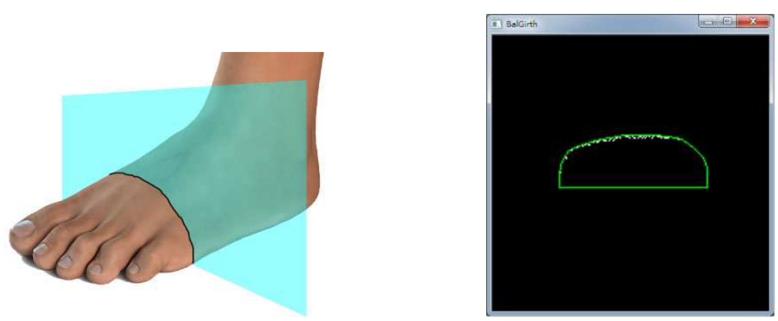
To compute the ball-girth value defined in Section 1, we transform Pt_{ft} into the FCS. Assume that we have known the stable girth cut points $\mathbf{g}_0$ and $\mathbf{g}_1$, we need to find a plane equation that includes those two cut points and is perpendicular to the x-y plane of the FCS (see Fig. 9). The normal vector of the cutting plane is given by

$$\boldsymbol{g}_{pl} = \begin{bmatrix} g_x & g_y & g_z \end{bmatrix} = \boldsymbol{g} \times \boldsymbol{z} \tag{5}$$

where $\boldsymbol{g} = \mathbf{g}_0 - \mathbf{g}_1$ and $\boldsymbol{z}$ is the z-axis vector of the FCS. With the plane normal vector, we have the following ball-girth's plane equation parameterized by using $\boldsymbol{g_{pl}}$ and $\mathbf{g}_0$

$$G_1 = \begin{bmatrix} \boldsymbol{g_{pl}} & c \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = g_x\, x + g_y\, y + g_z\, z + c \tag{6}$$

where $c = \mathbf{g}_0 \cdot \boldsymbol{g_{pl}^T}$. The point-to-plane distance between each point in Pt_{ft} and the cutting plane is computed and if it is smaller than a specified threshold δ, the point is projected to the Since the 3-D data projected onto the cutting plane is still noisy, the contour points are smoothed and connected by using the snake algorithm [13][14]. The length of the snake curve is computed as the ball-girth value.



(a) (b)

Fig. 9. The Estimation of the ball-grith: (a) the cross-section plane computed with girth cutting points, and (b) the extracted cross-section curve for determining the ball-girth.

In summary, the point clouds from multiple cameras are transformed into the RCS. Points which belong to the background objects are removed to yield an integrated point cloud of the foot surface. Then, we find the rotation invariant FCS using the PCA technique. After obtaining the FCS, we can compute the width and the length of the foot. For computing the girth value, we need to process the image to find the toe position and use the toe position to infer the girth cut points $\mathbf{g}_0$ and $\mathbf{g}_1$. Finally, we can calculate the ball-girth value using the above-mentioned method.

3 Experiment

To test the proposed method, we invite five male adults to measure their left and right foot parameters. Each subject was asked to repeat the measurement for 10 times. The following tables list the standard deviations of the estimations of all foot parameters. The results show that the repeatabilities of the foot length and the foot width estimations are good enough for selecting shoes. Their standard deviations are less than 3.5 mm. Conversely, the girth estimations are slightly noisier than the first two parameters. However, the variations of the estimated values are still within the 6 mm which are good enough for shoe selection.

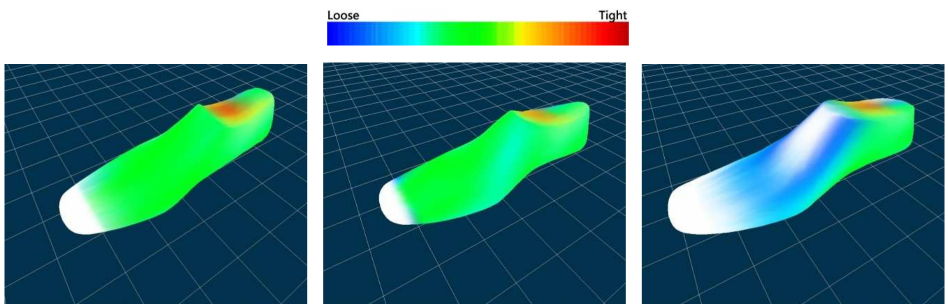
Table 1. The standard deviations of the estimated left foot parameters.

Subject	Length(mm)	Width(mm)	Girth(mm)
Subject1	2.17	1.45	2.76
Subject2	2.10	2.42	3.10
Subject3	3.50	2.17	4.25
Subject4	2.85	3.11	3.09
Subject5	2.25	2.14	3.04

Table 2. The standard deviations of the estimated right foot parameters.

Subject	Length(mm)	Width(mm)	Girth(mm)
Subject1	1.18	1.32	2.16
Subject2	2.51	2.15	3.57
Subject3	1.93	1.81	3.03
Subject4	2.59	2.04	5.69
Subject5	2.34	1.94	3.42

With the computed foot parameters, we can search for a suitable shoes from a database provided by a shoe producer. The user can then choose a shoe to perform further analysis. First, the 3-D model of the selected shoe is retrieved from the database. Second, the 3-D shoe model is projected to the x-y plane to compute its long and short axes using the PCA method. Third, align the heels, the long axes and the short axes of both the shoe and user's footprint. Fourth, adjust the foot orientation to maximize the overlapping area. Finally, signed distances between the foot surface point cloud and the shoe surfaces are evaluated and are converted to a subjective loose/tight index using a predetermined lookup table. Fig. 10 shows the degree of fitness of three shoes using the pseudo color technique. Fig. 10(a) shows that the shoe perfectly match user's foot. If we replace the shoe with a slightly larger one, then the computed pseudo colors shift toward blue as shown in Fig. 10(b). Also, Fig. 10(c) shows that the shoe is too large for the user.



(a) (b) (c)

Fig. 10. Estimation of shoe fitness: (a) A suitable shoe, (b) a slightly larger one, and (c) an oversize shoe for the same foot.

Results of the two experiments show that the 3-D foot reconstruction results using a hand-held RGB-D camera is accurate enough for providing satisfactory foot measurements.

4 Conclusion

In this paper, we propose a method to reconstruct the 3-D foot surface using a hand-held RGB-D camera. A reference pattern consists of an array of AR codes is used to provide registration information so that real-time camera pose estimation can be achieved. The estimated camera pose is used to guide the users to move the camera to a sequence of predetermined poses. When the camera pose is close enough to a predetermined one, the system will acquire an RGB-D image automatically. Upon the completion of image acquisition, the system will automatically evaluate three foot parameters for finding suitable shoes. Results from real experiments show that the repeatability of the proposed method is good enough and can be used to recommend well-fitted shoes.

References

1. Wang, C.S., Chang, T.R., Lin, M.C.: A systematic approach in shoe last design for human feet. In: IEEE International Conference on Industrial Engineering and Engineering Management, Ieee (2008) 1204–1208
2. Zhao, J., Xiong, S., Bu, Y., Goonetilleke, R.S.: Computerized girth determination for custom footwear manufacture. Computers and Industrial Engineering **54** (2008) 359–373
3. Kouchi, M.: Analysis of foot shape variation based on the medial axis of foot outline. Ergonomics **38** (1995) 1911–1920
4. Kouchi, M.: Foot dimensions and foot shape: Differences due to growth, generation and ethnic origin. Anthropological Science **106** (1998) 161–188
5. Oda, T., Sato, N., Nakano, I., Kaneko, Y., Ota, T.: System and method for assisting shoe selection. United States Patent 7089152 (2006)
6. Nguyen, P., Hong, B.: Method and system for sizing feet and fitting shoes. United States Patent 20040073407 (2004)
7. White, J.P.: Foot measurement and footwear sizing system. United States Patent 5128880 (1992)
8. Witana, C.P., Xiong, S., Zhao, J., Goonetilleke, R.S.: Foot measurements from 3-dimensional scans : A comparison and evaluation of different methods. International Journal of Industrial Ergonomics **36** (2006) 789–807
9. Asics: (http://www.asics.com.tw/)
10. Shoefitr: (http://shoefitr.com/)
11. Tsai, R.Y.: A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. IEEE Jornal of Robotics and Automation **RA-3** (1987) 323–344
12. Arun, K.S., Huang, T.S., Blostein, S.D.: Least-squares fitting of two 3-d point sets. IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-9** (1987) 698 – 700
13. Kass, M., Witkin, A., Terzopoulous, D.: Snakes: Active contour models. International Journal of Computer Vision (1988) 321–331
14. Cohen, L.D.: On active contour models and balloons. **53** (1991) 211–218