Measurement of 3D Foot Shape Deformation in Motion

Makoto Kimura[*][*]   Masaaki Mochimaru[*][†]   Takeo Kanade[*][◇][‡]

[*]Digital Human Research Center   [◇]The Robotics Institute

National Institute of Advanced Industrial Science and Technology, Japan   Carnegie Mellon University
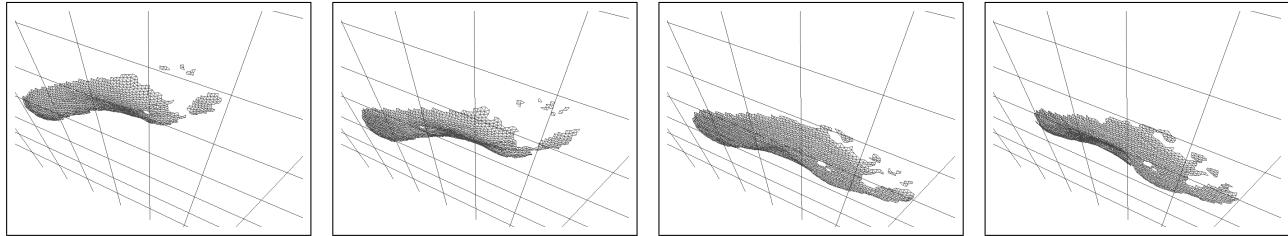
Figure 1: *Examples of the foot sole shape while the subject is running.*

Abstract

Recently, techniques for measuring and modeling of human body are recieving attention, because human models are useful for ergonomic design in manufacturing. We aim to accurately measure the dynamic shape of human foot in motion (i.e. walking or running). Such measurement is profitable for shoe design and sports analysis. In this paper, a projector-camera system is proposed to measure the shape of the naked foot while walking or running. A characteristic pattern is set on the projector, so that correspondence between the projection pattern and the camera captured image can be solved easily. Because pattern switching is not required, the system can measure foot shape even when the foot is in motion. The proposed method trades "density of measurement" for "stability of matching", but the reduced density is sufficient for our purpose.

CR Categories: C.4 [Measurement techniques]: ;— [I.5.4]: Computer vision—Applications

Keywords: measurement system, structured light, 3D deformation

1 Introduction

In recent years, anatomy and biomechanics have been widely used in criminological and medical applications, and in selective trial of products. They also play an important part in ergonomic design[Treleaven 2004]. For such purposes, accurate measurement and modeling of human shape[International Organization for Standardization 1997] is necessary. Nowadays, 3D scanners for human measurement (body, head, foot, and so on)[Treleaven 2004; Daa-

---

[*]e-mail: makoto.kimura@aist.go.jp

[†]e-mail: m-mochimaru@aist.go.jp

[‡]e-mail: tk@cs.cmu.edu

nen et al. 1998; Boehnen and Flynn 2005; Kouchi and Mochimaru 2001] are available.

For example, shoes should be designed in consideration of a user's feet[Kouchi and Mochimaru 2001]. Ideally, the design should consider not only the static shape, but also the dynamic shape while walking, running, and so on. We researched a technique for the measurement of anatomical feature cross-sections using multiple-view stereo matching, which can measure foot shape while the subject is walking[Kimura et al. 2005]. In this way, we aim to accurately measure the dynamic shape of the human foot in motion (i.e. walking or running). Our previous system measured only feature cross-sections, and the next step is the measurement of the whole 3D foot shape. Because our goal is measurement for product design and motion analysis, accuracy is important issue. Our desired accuracy is less than 1mm error, which is similar to the existing 3D scanners for static foot measurement.

As previously mentioned, many 3D scanners are available as commercial products, which can measure 3D shape with high accuracy. Unfortunately, most of them cannot measure dynamic transformation of an object, because they use laser scanning or pattern switching approaches. In other words, they fundamentally require a time period for measurement, and their target is only static object. Recently, fast scanners have been developed by several researchers and companies. Some of them mainly focus on dealing with incomplete stillness, because humans cannot stand completely still for a couple of seconds. They do not assume that the target shape undergoes deformation and motion, and they are not designed to make measurements in a sequence. The rest mainly focus on measuring dynamic shape, and the human face is a popular target[Zhang and Huang 2006; Jones et al. 2006]. They assume that the target shape does not change in a short time, and it is acceptable in their cases. However, this approach is not valid in our case, because the human foot moves much faster than the human face (e.g. in our experiences of slow walking – not running –, foot can move about 5mm in 10msec.)

Motion capture systems are also a popular method for 3D measurement. However, they are not suitable for shape measurement, because they have strong limitations about the number and density of markers. Figure 2 shows an example of human gait analysis using a motion capture system. In this figure, about 20 markers are set on the foot and leg, and this number is the maximum possible in our experience. Thus, motion capture system cannot measure "shape".

There has also been research into the stereo matching method[Hartley and Zisserman 2004]. Using synchronized
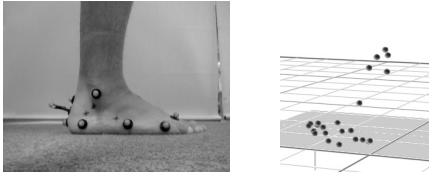
Figure 2: *An example of motion capture experiments.*

Projection Patterns
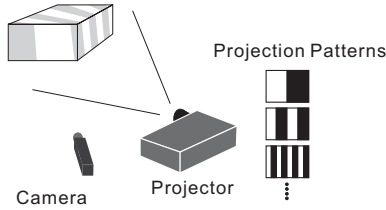
Camera   Projector

Figure 3: *The principle of coded structured light method.*

multiple video cameras, stereo matching can produce 3D shape frame by frame. However, the stereo matching method requires accurate correspondence between images, and it cannot be solved easily.

In this paper, a projector-camera system is proposed to measure the shape of a naked foot while walking or running. Because this method relies on the triangulation between the projection pattern and the camera image, its basic concept is just like stereo matching between multiple cameras. Using a projector instead of a camera has several advantages. Camera captured images always have noise, and video cameras produce more noisy images than still cameras. Generally, the projected texture's noise in the scene is lower than noise in the captured image. This means stereo matching between the projection pattern and the camera image is less affected by noise than stereo matching between multiple camera images. In stereo matching using cameras, the accuracy and stability of the process highly depends on the target scene, and texture on the target objects is one of the most important elements. In the projector-camera system, an arbitrary projection pattern can be set on the projector, so that detection of correspondence can be easier. Thus, the correspondence between projector and camera can be solved easier than the correspondence between multiple cameras.

In the following sections, a projection pattern is proposed, which magnifies the above mentioned advantages. Generally, a striped pattern is often used in existing projector-camera systems. A striped pattern is characteristic in the vertical direction of the stripe, but is non-characteristic in the non-vertical direction. The proposed projection pattern has many small rectangles, which have a characteristic color distribution in it. Assuming that a rectangle is the unit of the matching process, all units have individual features, so that correspondence between camera image and the pattern can be solved easily. Thus, the proposed method trades density of measurement for stability of matching. In our experiments, reconstructed points are in an approximately 3mm interval, which examples are shown in Figure 1, which is sufficient for analysis of foot shape.

2 Related Works

2.1 Coded Structured Light

In making a projector-camera system to measure 3D shape, some kind of structured light is always used. Especially, binary pattern

switching is a very popular approach. Using multiple camera images with pattern switching, the correspondence between the projection pattern geometry and the camera image geometry can be obtained easily, because the variation of a pixel's values in the camera image sequence reveals the projection pattern geometry. Thus, the pattern switching method is equal to embedding 2D location information in the variation of projector's brightness (Figure 3).

Such kind of method requires a time to measure in its principle, and so they cannot measure any dynamic scenes. However, this "embedding" concept has some relation with our approach, which will be described later.

2.2 Dynamic Face Scanner

Recently, several fast measurement systems have been developed. Face scanner systems can measure a human's dynamic face shape[Zhang and Huang 2006; Jones et al. 2006]. They are based on the strong precondition that face shape does not change in short time, so that the pattern switching principle can be applied. Therefore, they are using several frames in a video sequence, which is captured by a high-speed camera.

Because both the proposed method and the dynamic face scanners are targeting human shape, it seems to be a good idea to use the same technique in our system. However, our target is the foot, which can move about 5mm in 10msec even in slow walking. Thus, we cannot accept the precondition like face scanners. [1] In theory, faster high-speed scanners and faster high-speed projectors might a solution. Such kind of approach requires customized devices, which are not easy to make, and are expensive.

From a different point of view, the face scanners have several advantages compared with the proposed method. They can acquire more dense points than our system, and acquire not only 3D geometry but also texture information on the surface. The proposed method gives up the density of measurement and texture information, and it trades them for the ability to measure a fast moving target.

3 Proposed Method

We propose a 3D measurement method using a projector-camera system, which does not require pattern switching. The proposed method uses a fixed characteristic pattern instead of pattern switching. The projection pattern has small areas which have certain uniqueness, so that correspondence between the pattern and captured images can be solved easily.

In this paper, the measurement target is a foot in motion. The proposed method is not specific to foot, however it implicitly uses preconditions about target. The surface of the object must have uniform reflectance property, and must not be bumpy in shape.

Hardware equipment must be selected appropriately. The projector should be of the LCD type, because DLP type projectors use time-multiplexed projection. The camera must have a function to be set fast shutter speed. (In our experiments, shutter speed is always faster than 2msec.)

The calibration of the projector and the camera is assumed to be known.

---

[1]Strictly, the assumption "shape does not change in very short time" is required in this paper, too. Camera captures an image by opening its shutter at short time, so camera always requires such an assumption. In our experiments, shutter speed is always faster than 2msec.
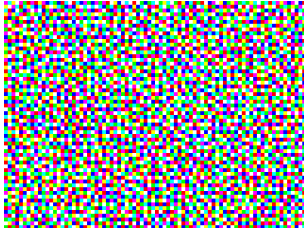
Figure 4: *Color coded pattern. Seven colors (combination of [R,G,B] × [0 or 255], except for black) are arranged to make a unique color distribution.*
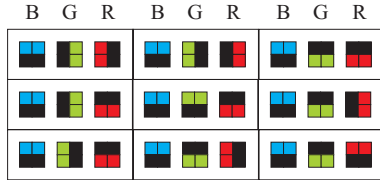

Figure 5: *Projection cells. Each cell has a 2×2 characteristic color pattern. 36 type patterns (9 type patterns with 4 rotations) can be arranged in the projection pattern.*
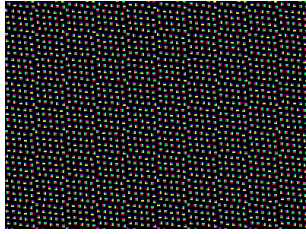

Figure 6: *An example of the projection pattern. The pattern has a number of cells. Cells are arranged like skewed grid.*
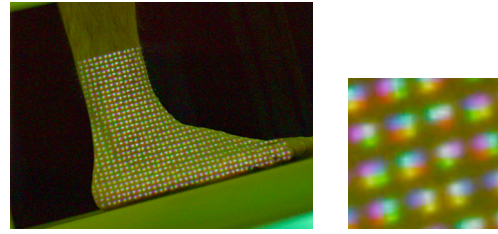
3.1 Projection Pattern

In most of the 3D measurement systems, the measurement principle is triangulation between multiple sensors. Such a triangulation is equivalent to calculating the intersection of two known vectors – projections from cameras (or projector). Thus, triangulation between a calibrated projector and a camera is equivalent to finding the correspondence between the captured image and the projection pattern.

Figure 4 shows an ideal pattern for our basic concept. In this color pattern, any 3×3 region has unique color distribution. So, we can identify 2D geometry in the pattern from any 3×3 color region in it. Projecting such a pattern onto the target object, we can capture the color textured object by a camera. Ideally, the matching problem can be solved by just checking the color distribution surrounding the interest point. However, color sensitivities of the projector and the camera are individual. Furthermore, camera image always contains noise, so color identification in the image is not an easy issue.

Therefore, we gave up some of the measurement density, and designed the pattern like Figure 6. This pattern is composed by the grouping of small 2×2 pixel areas, which we call "*cells*" in this paper. Each cell has a 2×2 color pattern, which is randomly chosen from variations in Figure 5. and is made to satisfy all of these conditions:

- Cell size is 2×2 pixels.

<div style="text-align:center">(a)   (b)</div>

Figure 7: *An example of the foot image with the pattern projection: (a) Original image. (b) Magnified.*

- In each cell, all pixels must have some color. No black pixel exists inside it.

- In each color plane of RGB, adjoined two pixels always have the same value.

As shown in Figure 5, the number of variations of cells is limited. So, lots of duplicated cells exist in the entire pattern in Figure 6. Unlike Figure 4, the correspondence to the interest point cannot be solved by color distribution. A procedure like usual stereo matching process is used at the following.

In Figure 6, cells are arranged like a skewed grid. This satisfies two requirements. The uniformity of the density should be kept in the whole pattern. At the same time, the existence of similar patterns on any epipolar line should be avoided as far as possible, because correspondence is always searched on the epipolar line[Hartley and Zisserman 2004]. Thus, this pattern has both regularity and randomness. For example, if the projector and the camera are horizontally set in a parallel direction, and if the projection pattern is just a grid of cells, then many cells exist in every epipolar line. Therefore, the projection pattern is generated with consideration for the arrangement of the camera.

3.2 Cell Detection

Projecting the proposed pattern onto the foot, the textured foot is captured as shown in Figure 7(a). At first, cell regions are detected from the captured image. Naturally, each cell area is brighter than the surrounding area as shown in Figure 7(b). They can be detected by threshold. However, the intensity distribution is different according to the position in the image, so a fixed threshold is not effective. This problem can be solved by a dynamic threshold, which is simple in implementation.

In this paper, the environment of the target scene is roughly known. Thus, the size of each cell in the captured image is also roughly known (e.g. "about 4×4 pixels to 8×8 pixels"). Using this condition, the dynamic threshold th is decided by the following:

$$th(x,y) = s(M_{max}(x,y) - M_{min}(x,y)) + M_{min}(x,y) \quad (1)$$

$M_{max}(x,y)$ is the maximum value in the local area ($\pm m$ in x and y), and $M_{min}(x,y)$ is the minimum value. Parameters m and s are decided experimentally. For example, m is set to be large enough to include several cells inside the $M_{max,min}$ area. Thus, foreground pixels in the captured image are divided from background pixels. At this point, the color of the cell is not used.
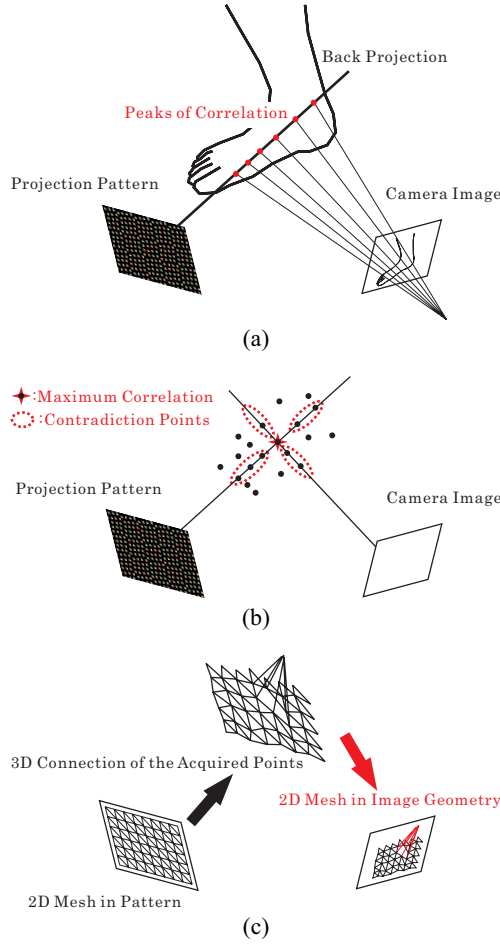
Figure 9: *Catwalk of our system*

Figure 10: *The arrangement of the projector and the camera: (a) Foot side measurement (b) Foot sole measurement.*

Figure 8: *Concept of correspondence estimation between image and pattern: (a) A lot of local maximum of points using normalized cross correlation. (b) The highest correlation point is adopted, and the contradiction points are deleted. (c) 3D Mesh is made based on the projection pattern, and its topology is checked in camera image's 2D geometry.*

3.3 Correspondence Estimation

Just like usual stereo matching process, the corresponding point is searched along the epipolar line. It searches for the correspondence point in the camera image's foreground area, based on a center position of each cell in the projection pattern.

The matching process is done by three steps.

In the first step, it looks for the local maximum of points using normalized cross correlation. Generally, two or more points have local maximum values for a cell in the projection pattern as shown in Figure 8(a). This process is applied to all cells.

The second step is iterative. The highest correlation point is adopted from all of the candidate points. Then, the candidate points that contradict the adopted point are deleted as shown in Figure 8(b). Next, go back to the beginning of second step loop, so that the highest correlation point can be adopted from the left over candidates. Thus, the loop continues until no more points are detected.

The final step is the reduction of mismatching points from the output of the second step. Even if the proposed projection pattern is used, some mismatching is unavoidable. Thanks to the discontin-

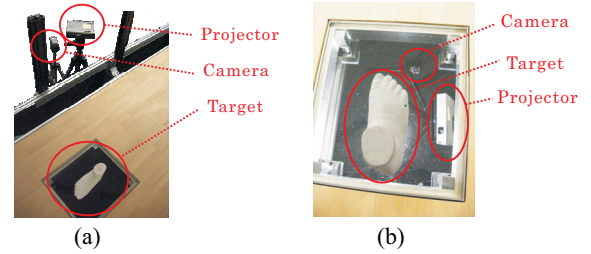uous projection pattern, 2D Delaunay structure can be calculated easily. With reference of the 2D Delaunay in the projection pattern, all points output by the second step are connected as shown in Figure 8 (c), and the 3D mesh is re-projected onto camera image's geometry. As shown in Figure 8 (c), a mismatching point breaks the mesh topology in camera image's geometry. Thus, mismatching points are removed based on mesh connection's topology.

4 Experiments

Figure 9 shows the catwalk of our system. This catwalk is 1m in height, and 10m in length. Thus, it is possible to measure it from the side of foot while walking. A glass board of 40×50cm is buried on the floor of the catwalk. Therefore, it is also possible to measure the sole of the foot from under the glass board.

4.1 Measurement of Foot Side while Walking

Figure 11 shows examples of input images while the subject is walking. In this experiment, input images were captured by 1024×768 pixel resolution at 14 frames per second. The projector-camera system was set as shown in Figure 10(a). In this experiment, the measured points were spaced at about 5mm intervals. Figure 12 shows acquired 3D shapes from the images in Figure 11. Figure 13 shows moving view of one of the shapes in Figure 12. Thus, plausible shape is measured overall.

4.2 Measurement of Foot Sole while Running

As shown in Figure 10(b), the projector-camera system for foot sole measurement was set under a glass board. In this case, 640×480 pixel high-speed camera was used at 200 fps, and the subject was running. In this experiment, the measured points were spaced at

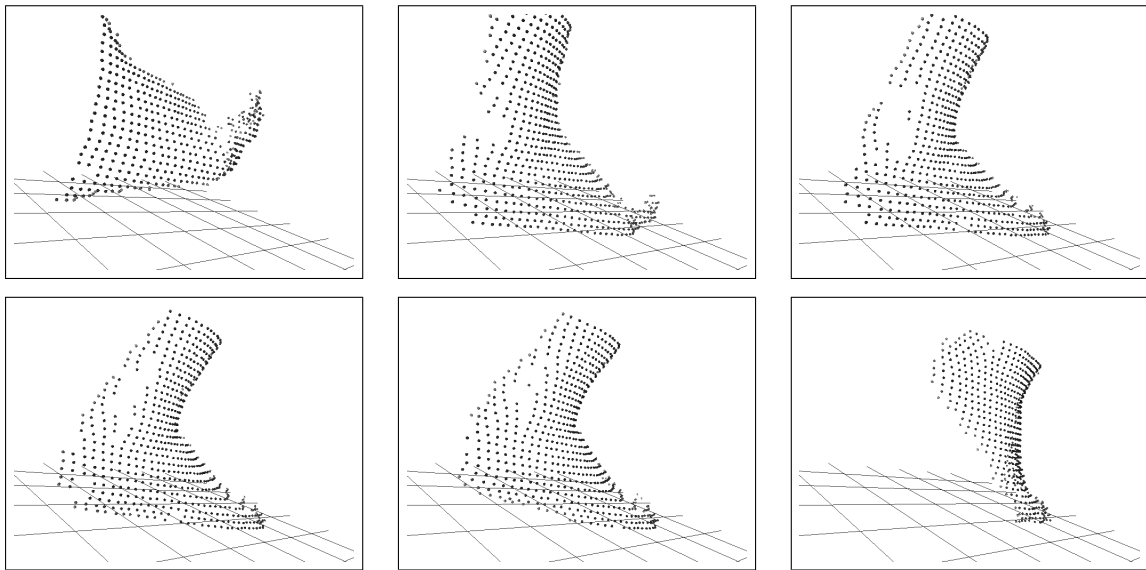**Figure 11:** *Input images for the foot side measurement. The subject is walking.*

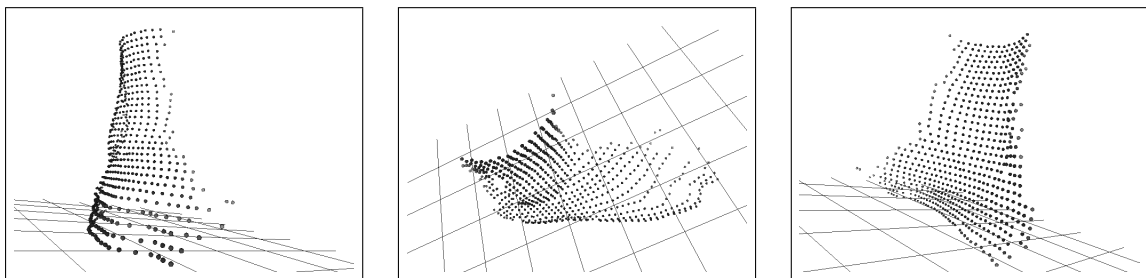**Figure 12:** *The acquired shapes of the foot from images in Figure 11.*

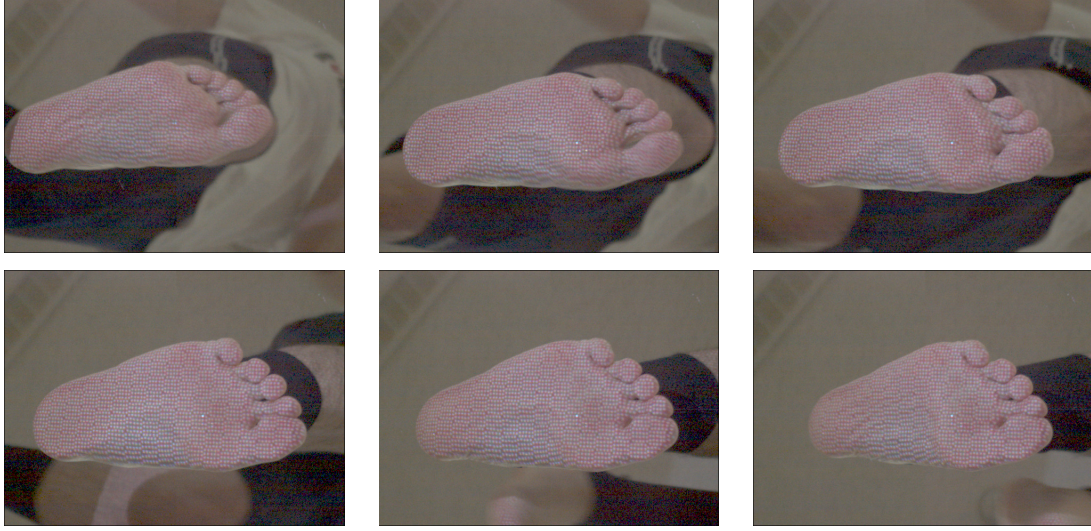**Figure 13:** *The acquired shape of the foot (moving view).*

Figure 14: *Input images for the foot sole measurement. The subject is running.*

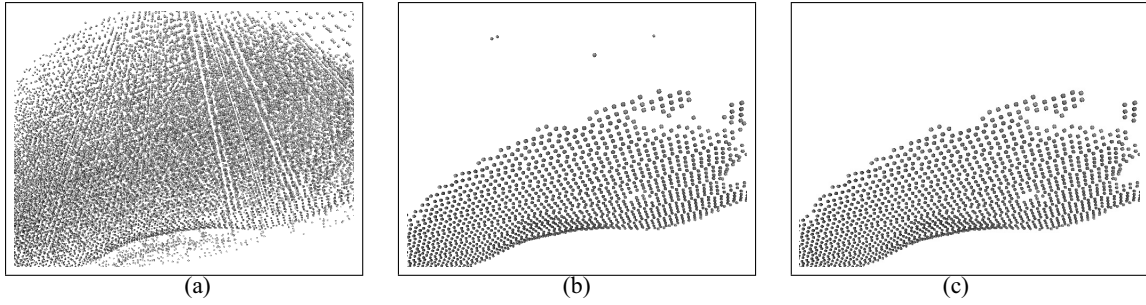(a)   (b)   (c)

Figure 15: *Process of the correspondence estimation: (a) All correspondence candidates for each cell in the projection pattern. (b) Adopted correspondence points without contradiction. (c) Final adopted points after checking mesh topology in the image geometry.*
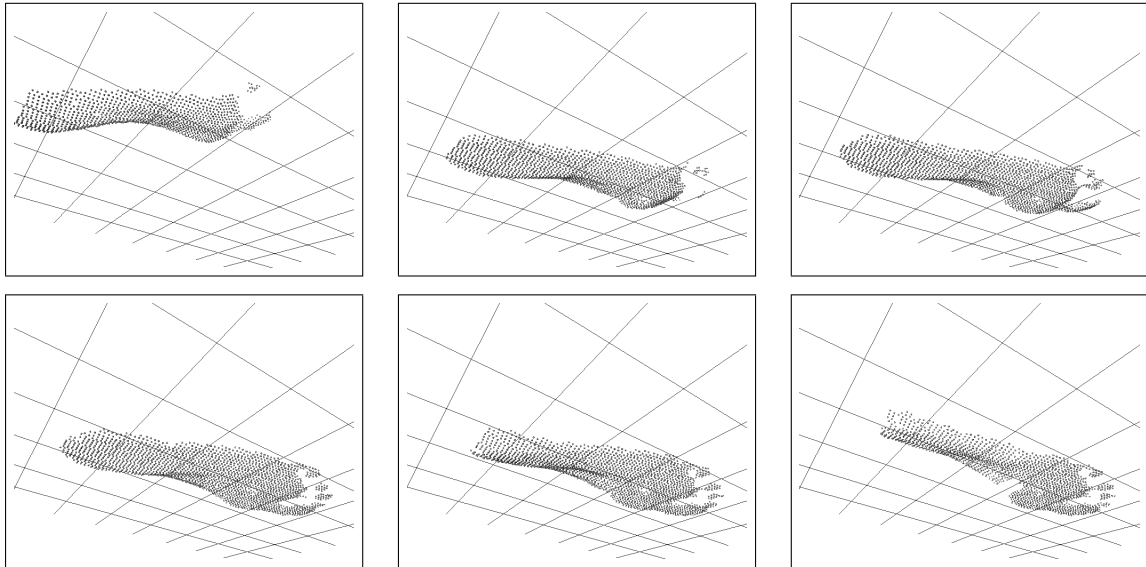

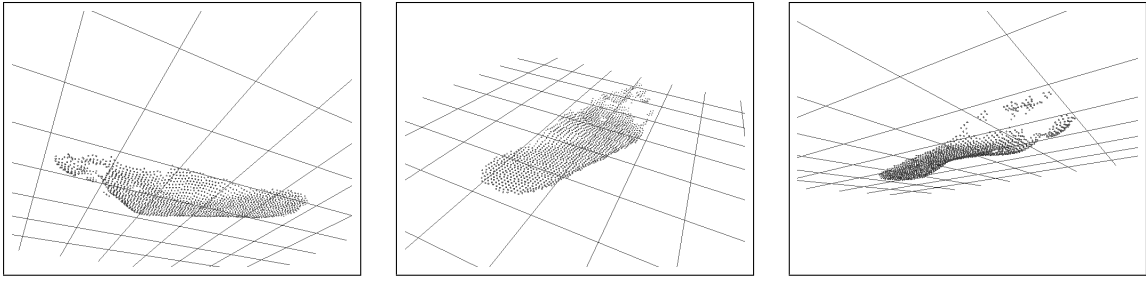**Figure 16:** *The acquired shapes of the foot from images in Figure 14.*

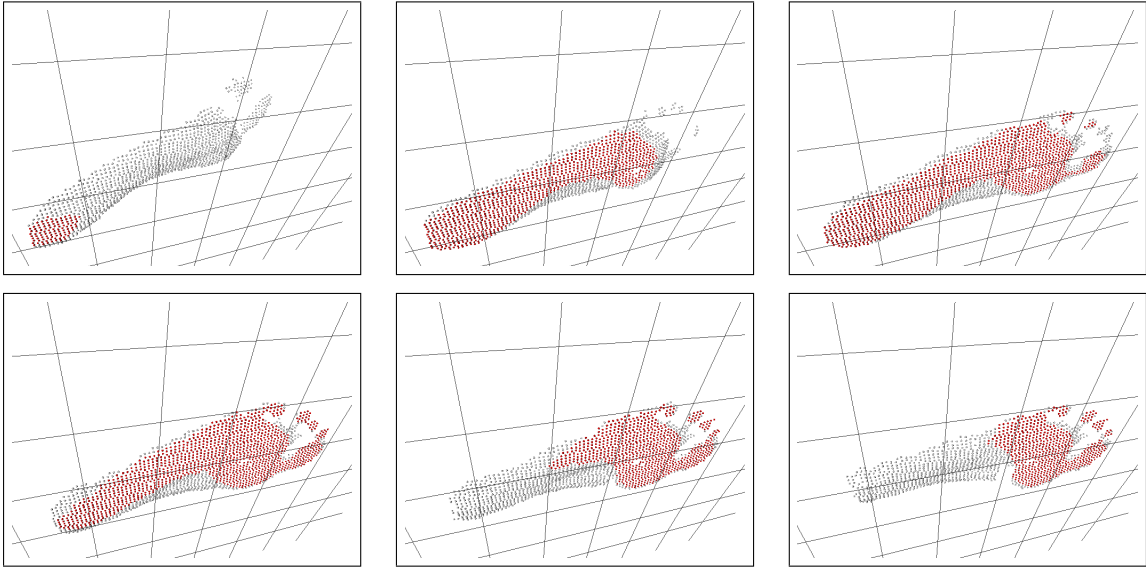**Figure 17:** *The acquired shape of the foot (moving view).*

Figure 18: *Contact area of the foot sole while running.*

about 3mm intervals. Figure 14 shows examples of input images. Figure 15 shows the process described in Section 3.3. In Figure 15(a), there are many candidates of correspondence for each cell in the projection pattern. In Figure 15(b), measurement points are acquired, but include some noise caused by mismatching. In Figure 15(c), noise is removed. As shown in Figure 16, 3D shapes are acquired frame by frame. Thus, the shape of the foot sole (e.g. its arch) is acquired. Figure 17 shows moving view of one of the acquired shapes. Figure 1 shows examples of the meshed foot sole shape.

Figure 18 shows a simple example of foot shape analysis. Needless to say, the ground contact area of the foot is changing while the subject is running. With consideration of the measurement error, we set a heuristic threshold "0.5mm" in height, so that the contact points are visualized. Because we measure the whole foot sole, we also have arch shape while running.

In this way, interesting data was measured by the proposed method. The measurement and the observation of such deformation of foot sole while landing are very difficult in the existing systems, because any markers cannot be attached on the foot sole.

4.3 Evaluation of Accuracy

To evaluate the accuracy of the proposed method, we measured the sole of a plaster foot in several poses as shown in Figure 19, and
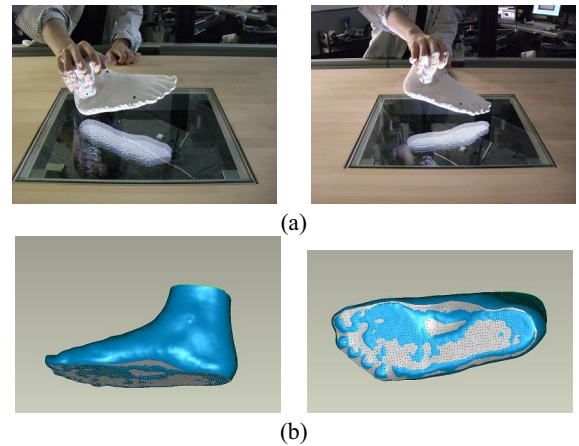

Figure 19: *Evaluation of the accuracy of the measured points: (a) Plaster foot in several poses. (b) An example of comparison with 3D shape acquired by the 3D scanner for static foot.*

compared them to the 3D shape measured by an existing 3D scanner for static foot. It is difficult to evaluate the accuracy with deformable shapes, because there is no reference 3D data of deformation. In comparison with 3D shape acquired by the 3D scanner, the average distance was always less than 0.2mm. As long as this value can be assumed to be the error level of the proposed method, the proposed method achieved our desired accuracy (less than 1mm error).

5 Conclusion

In this paper, a method to measure dynamic 3D shape is proposed. The proposed method requires only a camera and a LCD projector, that are commercially available anywhere. Thus, the system costs are very low.

Although the proposed method is not specific to the foot, it is not a general 3D measurement system. In the proposed method, the surface of the object must have a uniform reflectance property, and must not be bumpy in shape. Additionally, the proposed method acquires sparse points, because it trades "density of measurement" for "stability of matching".

In our experiments, the measured points are spaced at about 3mm intervals, which is sufficient for analysis of foot shape. The effectiveness of the proposed method is shown by the experiments using real feet and the plaster foot. The proposed method worked stably in our experiments of 10 subjects × 10 sequences. The desired accuracy (less than 1mm error) was achieved in spite of the motion of the target object. We consider that the proposed technique is useful for dynamic human body (e.g. arm, leg, belly, and so on) measurement in the same way.

As described in this paper, the proposed method was effective to measure the foot shape from one direction. However, ideal measurement data is circumference shape of the whole foot, which is just like existing 3D scanners for a static foot. Thus, we now aim to make simultaneous measurement by multiple pairs of projector-camera systems.

References

BOEHNEN, C., AND FLYNN, P. 2005. Accuracy of 3d scanning technologies in a face scanning scenario. In *Proceedings of The 5th International Conference on 3D Digital Imaging and Modeling (3DIM05)*, 310–317.

DAANEN, H. A. M., J., G., AND VAN DE WATER. 1998. Whole body scanners. *Displays 19*, 3, 111–120.

HARTLEY, R., AND ZISSERMAN, A. 2004. *Multiple View Geometry in Computer Vision second edition*. Cambridge University Press.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 1997. Iso 7250: Basic human body measurements of technological design.

JONES, A., GARDNER, A., BOLAS, M., MCDOWALL, I., AND DEBEVEC, P. 2006. Simulating spatially varying lighting on a live performance. In *3rd European Conference on Visual Media Production (CVMP)*, The Institution of Engineering and Technology, 127–133.

KIMURA, M., MOCHIMARU, M., KOUCHI, M., AND KANADE, T. 2005. 3d cross-sectional shape measurement of the foot while walking. In *Proceedings of The 7th Symposium on Footwear Biomechanics*, International Society of Biomechanics, 34–35.

KOUCHI, M., AND MOCHIMARU, M. 2001. Development of a low cost foot–scanner for a custom shoe making system. In *Proceedings of the 5th Symposium on Footwear Biomechanics*, International Society of Biomechanics, 58–59.

TRELEAVEN, P. 2004. Sizing us up – new 3-d body scanners are reshaping clothing, car seats and more. *IEEE Spectrum 41*, 17–19.

ZHANG, S., AND HUANG, P. S. 2006. High-resolution, real-time three-dimensional shape measurement. *Optical Engineering 45*, 12 (December), 123601.