Z-touch: An Infrastructure for 3D gesture interaction in the proximity of tabletop surfaces

Yoshiki Takeoka
The University of Tokyo
7-3-1 Hongo
Bunkyo-ku, Tokyo, JAPAN
yoshikinoko@gmail.com

Takashi Miyaki
The University of Tokyo
miyaki@acm.org

Jun Rekimoto
The University of Tokyo,
Sony Computer Science
Laboratories, Inc.
rekimoto@acm.org

ABSTRACT
Sensing the depth (distance from the surface) of fingers/hands near a tabletop is very important. It allows us to use three-dimensional (3D) gesture interaction in multi-touch applications as we do in the real world. We introduce Z-touch, a multi-touch table that can sense the approximate postures of fingers or hands in the proximity of the tabletop's surface. Z-touch uses a vision-based posture sensing system. Multilayered infrared (IR) laser planes are synchronized with shutter signals from a high-speed camera, which captures each layer of the laser images. A depth map is obtained by using the captured image. Our prototype works at $\sim$30 fps. Z-touch not only uses with the finger/hand contact points but also the angle of the hovering fingers. The interaction with the finger angles is unique and allows users to control multiple parameters by using a single finger. In this study, we introduce the principle of the method of finger detection and its applications (e.g., drawing, map zooming viewer, Bezier curve control).

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. - Input devices and strategies.

General terms: Multi-touch, Interactive Surface, Gesture Interaction.

Keywords: Multi-touch Table, Interactive Surface, 3D Hand Gesture Interaction.

INTRODUCTION
Most surface computing systems are capable of tracking multiple finger positions [3]. Users of these systems can use both hands and multiple fingers for interacting with various objects on an interactive surface.

While these systems are useful, we think these interactions can be greatly improved if the system can detect a finger's three-dimensional (3D) shape and its distance from the surface. For example, if the system can correctly distinguish between finger contact and finger hovering, interaction techniques similar to the mouse-over mode used in the graphic user interfaces (GUIs) would become available. This would enable the user to inspect the object's information by approaching it with his/her finger before actually touching it. Similarly, if the system can detect the inclination and direction of a finger, they can be used to control objects. For example, a user would be able to rotate an onscreen dial knob using a finger by controlling the direction of the finger and manipulate the shapes of onscreen objects by controlling their contour curves by using finger position and angle. We may also use various 3D hand gestures on a tabletop.

To enable these features, several attempts have been made to use depth information obtained from surface computing systems. A pen-based interaction system that uses the angle of the pen has been proposed [7]. Capacitive sensor-based depth detection systems, such as SmartSkin [6], detect finger proximity and can be used with LCD displays. Some vision-based systems TouchLight [9], Perceptive Workbench [5], and BiDi Screen [4] use correlation-based stereo matching of finger images obtained from two cameras or multiple image sensors and detect the hand's distance from the surface. However, capacitive-based systems measure the distance inaccurately because of noise and effects from the user's body. Stereo-based systems often require more computational power than systems using a single camera. They also require very accurate camera calibration and are less robust against planar texture regions. Although systems using a depth camera [1, 10] have been proposed, these are still very expensive, and the obtained depth accuracy is inadequate especially for clearly distinguishing between touching and hovering. A method for estimating angles of fingers by using an image of the contact fingers [2] has been proposed. However, it cannot detect hovering fingers.

To overcome these limitations, we introduce Z-touch, a multi-touch tabletop system that detects finger positions and distances from an interactive surface [8]. Z-Touch uses multiple laser layers to detect the distance of each finger from the surface (Figure 2). These multiple layers are illuminated in turn, and a high-speed camera installed under the surface captures an image of each layer independently by synchronizing with the timing of laser layers. This system mainly provides two features. The first one is hand posture detection. This feature is useful for detection of the vertical/horizontal orientation of the fingers touching near the surface. The other feature is

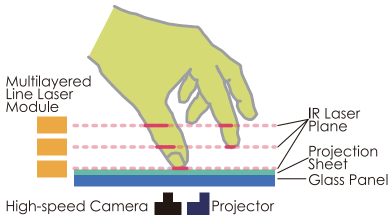

Figure 1: Z-Touch sensing principle. Multiple infrared laser planes are projected for recognizing finger distance from the interaction surface. Infrared line laser modules are synchronized with a high-speed camera's shutter signal.

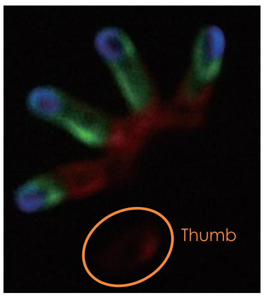

Figure 2: Captured depth map of a hand near the top panel of Z-touch. RGB channels correspond to the laser plane: R, highest; G, middle; B, lowest (very close to the tabletop's surface). The image at the bottom is that of the thumb not touching the surface; hence, the silhouette is pale red in color.

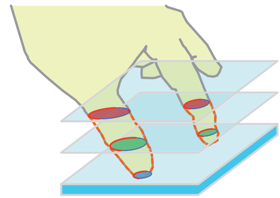

Figure 3: Finger angle detection. The system can detect the finger angles by matching the appearance of finger postures in laser plane images obtained at different heights.

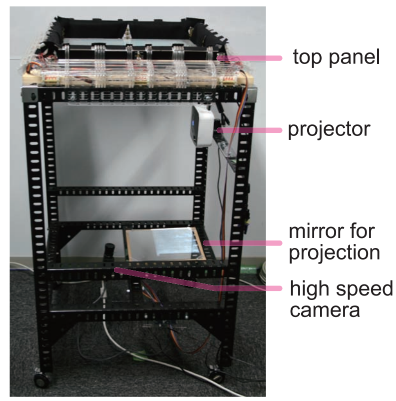

Figure 4: Z-touch hardware setup. The dimensions of Z-touch are 1.0 m $\times$ 0.6 m $\times$ 0.6 m. The framework is a 40 mm wide steel angle frame.

hover detection. The system can detect finger positions while the finger is within proximity if not touching to the surface. In this paper, we describe the sensing architecture of Z-touch and propose several techniques of interaction for the system using 3D finger depth information.

PROPOSED APPROACH

Principle of Z-touch

Z-touch is a multi-touch table infrastructure for 3D gesture interaction in the proximity of tabletop surfaces. Figure 1 shows its system architecture. The key components are multilayered infrared (IR) line laser modules and a high-speed camera. Our proposed approach expands the multi-touch implementation method of the laser light plane (LLP). However, only a single laser plane has been used in the LLP, whereas Z-touch uses a multilayered laser plane. These laser planes are synchronized with a high-speed camera's shutter signal.

Hardware setup

To prevent the flickering effects of environmental illumination and projection light, we used IR line laser modules and an IR band-pass filter. The power of IR line laser is 5 mW, and its wavelength is 780 $\pm$ 10 nm. Peak transmission of the IR band-pass filter is 780 nm. We choose this wavelength parameter, based on the spectacle response of the image sensor of high speed camera. Eight sets of triple-line laser modules are placed at the center of the edge and at the vertex of the top panel. The IR laser planes are calibrated such that the distance between two laser planes is 11.5 mm, and they are parallel to the top panel. The maximum detectable height is almost the same as the typical distance between the tip of the finger and the second knuckle. The lowest laser plane is the closest to the surface of the top panel. The top panel (control area) is a 420 mm wide and 6 mm thick square glass covered with a transparent projection sheet, so that only visible light emitted from the projector diffuses in the surface of the top panel but reflected IR light passes through (Figure 5). Figure 4 shows the Z-touch hardware setup. Its dimensions are 1 m $\times$ 0.6 m $\times$ 0.6 m. For safety, a small black felt curtain is attached to the edge of the top glass panel to prevent the IR laser from reflecting. We used a 40 mm steel angle frame for the framework. The projector was placed in the middle frame, and the mirror used for the projection is in the bottom frame. The high-speed camera was installed in the bottom frame. We used a Point Grey Grasshopper, a high-speed camera which can capture 8-bit gray-scale VGA images at 200 fps. This camera generates external trigger signal to laser control circuit for every capturing.

Finger tracking

Figure 2 shows a depth map obtained by Z-touch. A blob will appear at different positions in the captured frames, until the surface of the fingers/hand is completely perpendicular to the laser plane. Therefore, the system can detect the finger angles by matching the appearance of finger postures in the laser plane images at different heights (Figure 3). We used three-step image processing to track and estimate the posture of the fingers/hands as follows.

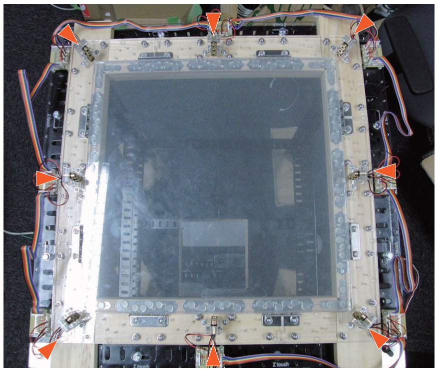

Figure 5: Top panel of Z-touch. Eight sets of multi-layered line laser modules are placed and directed as indicated by the orange triangles. Transparent screen sheet is attached to the glass panel.

1. Image binarization.
2. Blob detection in the binarized image.
3. Silhouette matching of the detected blob regions with that in layers at different heights.

The silhouette matching process finds a match for the blob by using the following equations. a and b are blobs detected in at different heights, $a \in Blob_{height}$, and $b \in Blob_{height+1}$. $s(a, b)$ is the size of the set-theoretic difference between a and b, $d(a, b)$ is the Euclidean distance between the blobs, and T is the distance threshold.

$$match(a) = \arg\min_{b} c(a, b) \quad (1)$$

$$c(a,b) = \begin{cases} s(a,b) \times k + d(a,b) & d(a,b) < T, k > 0 \\ \infty & \text{else.} \end{cases}$$

This matching method detects the finger heights at the three levels. Because adjusting the angle and the power of the laser is very hard, we do not estimate the granularity of finger height. We used the process described above for ease of implementation. More efficient processes (such as particle filter tracking) could also be used. Figure 6 is an application that demonstrates the touch/hover of fingers/hands near the surface. The position of a finger near the tabletop is shown by the red and green circles. The circles change color when the finger contacts the surface, and a message "Touch!" is displayed near the point where contact was established. The position of the rectangle indicates where a blob has been detected, and the color shows the laser layer images of the blob.

APPLICATIONS ON Z-touch

To demonstrate the effectiveness of the proposed depth sensing architectures, we implemented three types of interaction techniques on the basis of Z-touch.

Drawing

Figure 7 illustrates the drawing application. To evaluate the

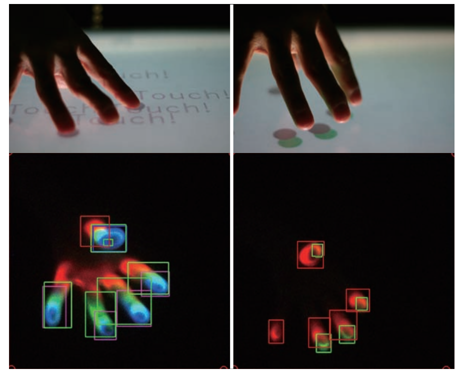

Figure 6: Touch detection. The images in the bottom row are depth maps. Colors representing the depth are the same as in Figure 2. Rectangles show the positions of detected blobs in each laser layer image. While the hand on the right is hovering over the surface, no blob is detected in the lowest laser plane (shown in blue on the left).

Figure 7: Drawing. The pixels' drawing colors corresponding to the depth map values.

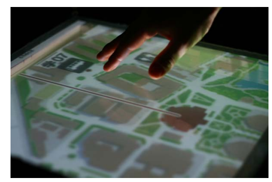

Figure 8: Map zooming. The zoom level of the map is changed according to the lowest height of fingers/hands detected.

accuracy of depth sensing, color information of the drawing pixels is corresponded to the values of the depth map.

Map zooming

Figure 8 illustrates the map zooming viewer. Users can simultaneously control zooming and scrolling by pointing a finger and varying its height. The zooming level of the map is controlled by the proximity of the finger to the surface.

Bezier curve control

Figure 9 illustrates an application for Bezier curve control. Users can control multiple curve parameters by using just one finger. The direction of a Bezier curve's handle corresponds to the horizontal angle of the finger, θ, and the length corresponds to its vertical angle ϕ. The endpoint of the curve is the contact point of the user's finger $F(x_f, y_f)$. The position of the curve handle $H(x_h, y_h)$ is calculated as follows.

$$H(x_h, y_h) = F(x_f, y_f) + \frac{-l}{\tan\phi}(\sin\theta, \cos\theta), l > 0. \quad (2)$$

DISCUSSION

Limitations on finger tracking

Some limitations exist on finger tracking. Because it is difficult to align the laser, we cannot estimate the granularity of finger's height. The algorithm for angle estimation is af-

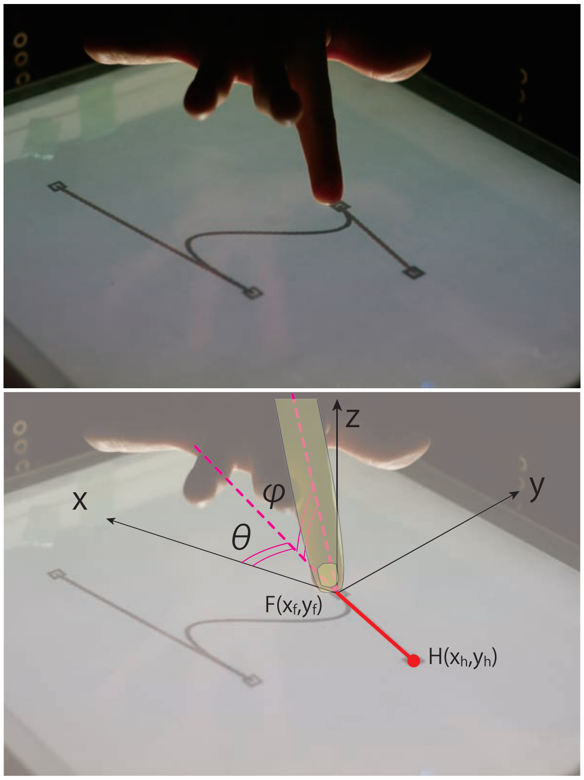

Figure 9: Bezier curve control. Endpoint of the curve is the contact position of the finger $F(x_f, y_f)$. The direction of the curve handle corresponds to the horizontal angle θ of the finger. The length of the curve handle corresponds to the vertical angle ϕ. Equation(2) relates the position of the endpoint $F(x_f, y_f)$ and the position of the curve handle $H(x_h, y_h)$.

fected by changes in illumination such as occlusion by fingers/hands. When large objects like the user's arm are placed on the tabletop, the silhouette of the fingers changes dramatically, and finger detection may fail temporarily. We deployed a laser module at 20 cm intervals on the bezel to prevent the occlusion effects (Figure 5). Our suggested algorithm used matching for different laser layer also has a limitation for tracking. This is due to the coarse matching nature found in the silhouette matching algorithm. So any mismatching of blob region will happen temporarily, and will produce incorrect finger's angle estimation result. However we could improve the accuracy by using gradation of image for matching.

CONCLUSION AND DIRECTION FOR FUTURE WORK

In this paper, we presented Z-touch, a multi-touch table that can sense the distance and posture of fingers near the tabletop's surface even if the fingers/hands are not touching the surface. Z-touch uses a multilayered IR laser plane synchronized with a high-speed camera's shutter signal. By combining multiple levels of laser plane images, the proximity and tilt angle of an approaching finger can be extracted using a vision-based technique in real time. We also developed applications that use Z-touch such as map zooming and multiple parameter control by using a single finger. Our proposed architecture and these interaction techniques will enhance the possibilities of existing user interfaces for tabletop computing. Users can intuitively manipulate detailed information on a tabletop. Further work on hardware configuration adjustment or vision recognition can be done to develop practical applications.

REFERENCES

1. H. Benko and A. D. Wilson. DepthTouch: Using Depth-Sensing Camera to Enable Freehand Interactions On and Above the Interactive Surface. Technical Report MSR-TR-2009-23, Microsoft Research, 2009.
2. C. T. Dang, M. Straub, and E. André. Hand distinction for multi-touch tabletop interaction. In *ITS '09: Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pages 101–108. ACM, 2009.
3. J. Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *UIST '05: Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 115–118. ACM, 2005.
4. M. Hirsch, D. Lanman, H. Holtzman, and R. Raskar. BiDi screen: a thin, depth-sensing LCD for 3D interaction using light fields. In *SIGGRAPH Asia '09: ACM SIGGRAPH Asia 2009 papers*, pages 1–9. ACM, 2009.
5. B. Leibe, T. Starner, W. Ribarsky, Z. Wartell, D. Krum, B. Singletary, and L. Hodges. The perceptive workbench: Towards spontaneous and natural interaction in semi-immersive virtual environments. In *VR '2000: IEEE Virtual Reality 2000 Conference*, pages 13–20. IEEE Computer Society, 2000.
6. J. Rekimoto. SmartSkin: an infrastructure for freehand manipulation on interactive surfaces. In *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 113–120. ACM, 2002.
7. S. Subramanian, D. Aliakseyeu, and A. Lucero. Multi-layer interaction for digital tables. In *UIST '06: Proceedings of the 19th annual ACM symposium on User interface software and technology*, pages 269–272. ACM, 2006.
8. Y. Takeoka, T. Miyaki, and J. Rekimoto. Z-touch: a multi-touch system for detecting spatial gestures near the tabletop. In *SIGGRAPH '10: ACM SIGGRAPH 2010 Talks*. ACM, 2010.
9. A. D. Wilson. TouchLight: an imaging touch screen and display for gesture-based interaction. In *ICMI '04: Proceedings of the 6th international conference on Multimodal interfaces*, pages 69–76. ACM, 2004.
10. A. D. Wilson. Depth-Sensing Video Cameras for 3D Tangible Tabletop Interaction. In *TABLETOP '07: IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, pages 201–204. IEEE Computer Society, 2007.