DRIVE: Directly Reaching Into Virtual Environment With Bare Hand Manipulation Behind Mobile Display

Seung Wook Kim, Anton Treskunov, Stefan Marti

Samsung Electronics US R&D Center

ABSTRACT

We present DRIVE, an interaction method that allows a user to manipulate virtual content by *reaching behind* a mobile display device such as a cellphone, tablet PC, etc. Unlike prior work that uses front volume as well as front, side, and back surfaces, DRIVE utilizes the back volume of the device. Together with face tracking, out system creates the illusion that the user's hand is co-located with virtual volumetric content.

KEYWORDS: Behind-device-interaction, mobile gesture interaction, 3D interaction.

INDEX TERMS: I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction Techniques; K.8.0 [Personal Computing]: General Games

1 INTRODUCTION

Touch-based input has become a de facto standard for many mobile applications, but it occludes the screen, introducing the "fat finger" problem [18]. Furthermore, surface based touch input is most suitable for two-dimensional content. Meanwhile, following rapid hardware advances in 3D rendering and an emergence of 3D content in games, virtual worlds, and other applications, there is a strong need for spatial interaction methods beyond traditional surface UI metaphors.

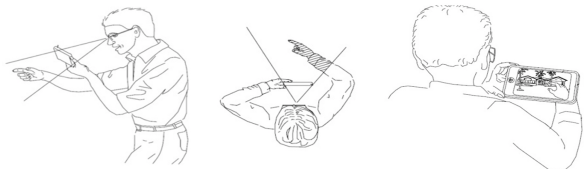

Figure 1. DRIVE concept illustration

In this poster, we propose a novel interaction method called DRIVE (a.k.a., Directly Reaching Into Virtual Environment), whose design addresses the occlusion and fat finger problems of touch input, as well as the conflicting focus problem of traditional gesture interaction, providing an improved user experience for 3D content on mobile devices. Figure 1 illustrates the concept of the DRIVE interaction method.

2 RELATED WORK

To avoid content occlusion by finger touch interaction, researchers have proposed to use the side surfaces of a device [3], the back surface [1,20], and front volume [10]. Surface based methods are two-dimensional in interaction space as well as in the intended content. Front volume based in-air typing work [10] solves the occlusion problem by putting the interacting hands above the display. However, the user's hand in that case is not collocated with manipulated content.

Visual hand tracking research generally uses flat markers [2,12] or multi-colored bands [9] or gloves [19], sometimes using maker-less tracking with explicit initialization [6], requiring outstretched hands [8], or measuring skin tone and stereo disparity [7]. Most of the mentioned works rely on markers [2,9,12,19], or on assumptions of finger position [8] or an initial configuration [6]. The work [7] reports that "most of the users commented on the unstable fingertip tracking which sometimes affected their interaction."

There is also fast growing field of time-of-flight cameras (TOF) [5] and its applications. As mentioned by [4], while "range images of TOF cameras are independent of the texture and lighting conditions, they are somehow affected by the color of the object," which also may explain the noise we encountered in our implementation. Efforts have been put into body part segmentation from range images [4, 14], which is not necessary for us due to the particularities of the DRIVE interaction method.

3 INTERACTION-BEHIND-DISPLAY METHOD

DRIVE is an extension of multiple methods mentioned above. It operates in the volume behind a display. Without touching the back of the device, a user manipulates virtual spatial content by bare hands in space. This approach is particularly useful with object manipulation in 3D space, such as spatial moving, toppling, pushing, grabbing, and so on, for which a 2D touch screen may not provide enough degrees of freedom.

VR head-mounted displays may offer a similar experience, but they require a forced perspective, isolating the user [17]. With a handheld display, by contrast, the user has more direct and ad-hoc control over which part of his field of vision is the real world vs. virtual content by moving the display relative to her head and eyes.

We hypothesize that handheld devices with 3 to 10 inch displays would be ergonomically most suitable for this type of interaction, as the user can reach behind them without any exaggerated arm stretch.

The user's hand is detected and tracked by sensors on the back of the device, and can be represented either as a real or virtual hand on the display. Because interaction happens inside the space volume populated by virtual objects, a new kind of interaction based on collisions between virtual representation of the hand and displayed objects is enabled.

4 IMPLEMENTATION

We developed two prototype systems. The first prototype consists of a 10-inch digital photo frame emulating a tablet display, and an infrared (IR) sensor-emitter configuration mounted on the back of the display, both connected to a PC (Figure 4). We used an array of 880nm IR LEDs whose effective range is long enough to illuminate only the hand, not the background. Accordingly, the system easily detects the hand contour against a uniformly dark background, and looks for the

3000 Orchard Parkway, San Jose, CA 95134, USA
{ seungwook.k | anton.t | s.marti } @sisa.samsung.com

IEEE Symposium on 3D User Interfaces 2011
19-20 March, Singapore
978-1-4577-0064-4/11/$26.00 ©2011 IEEE

leftmost vertex of the convex hull around the contour as the point of reference (e.g., right index fingertip). Fingers are rendered translucently (i.e., "ghost fingers") to provide the user with direct visual feedback without occluding display space. Figure 2 and 3 respectively illustrate a simple 2D GUI control, and a simple 3D object manipulation (e.g., picking and moving around based on the image plane technique [13]).

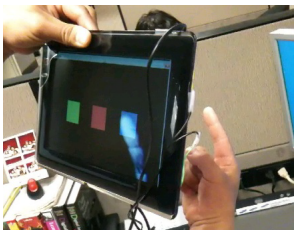

Figure 2. 2D GUI control

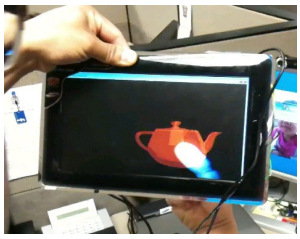

Figure 3. 3D object manipulation

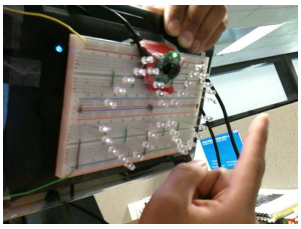

Figure 4. Back of mobile display w/ IR sensor & emitters

Figure 5. TOF depth camera on the stationary display setup

The first prototype demonstrated the feasibility of our reaching-behind approach, but confines the interaction to mostly 2D space. Therefore, we utilize a depth camera [15] (Figure 5) in our second implementation. Because the camera is facing away from the user and sees only his hand, we don't need to detect body parts as in [14]. Moreover, in the case of a static display, to avoid a recognition step, we assume a static background and filter it out in a method similar to z-buffer. After a data frame is filtered based on amplitude (discard pixels with weak signal) and "background-ness", the remaining pixels form a point cloud, and statistical filtering is applied to eliminate isolated outliers [16]. The resulting points are clustered based on distance, and only the largest cluster is kept. We approximate the cluster via an oriented box whose orientation and size is calculated using Principal Component Analysis. Finally, to enable collision based interaction methods, we add two spheres on the opposite sides of the box. Figure 6 shows the point cloud generated by the hand, as well as virtual objects (stack of dices, in Ogre3D) before and after a collision with the hand.

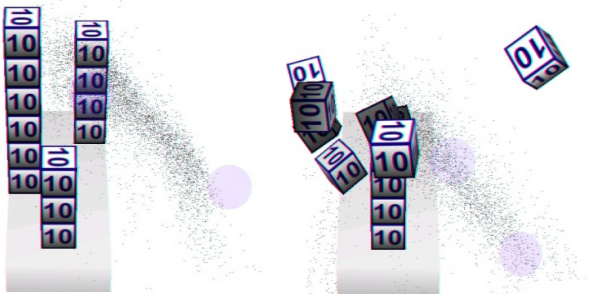

Figure 6. Point cloud and 3D dices, before (left) and after (right) collision with hand; the point cloud bounding box is rendered transparently.

5 "IN-LINE MEDIATION" WITH FACE TRACKING

A crucial component for our system is face tracking. By spatially aligning the user's eyes, the display, and the rendered virtual objects, the device becomes a 3D in-line mediator between the user and the virtual content. Our system detects the user's face from a front-facing Webcam (Figure 5) using OpenCV [11] libraries, and renders content through an asymmetric view frustum. This results in visual coherency where virtual objects are perceived as stationary in the device space, regardless of user movements relative to the device. That makes it easier to manipulate objects, and adds to spatial cues by providing motion parallax.

REFERENCES

[1] Baudisch, P. and Chu, G. (2009). Back-of-Device Interaction Allows Creating Very Small Touch Devices. In *Proc. CHI 2009,* pp. 1923–1932.

[2] Buchmann, V. et al. (2004). FingARtips: gesture based direct manipulation in Augmented Reality. In *Proc. GRAPHITE 2004*, pp. 212-221.

[3] Butler, A. et al. (2008). SideSight: Multi-"touch" Interaction Around Small Devices. In *Proc. UIST 2008,* pp. 201–204.

[4] Ghobadi, S. E. et al. (2007). Hand Segmentation using 2D/3D Images. In *Proc. Image and Vision Computing New Zealand*, pp. 64–69, Hamilton, New Zealand, December 2007.

[5] Gokturk, S. et al. (2004). A time of flight depth sensor, system description, issues and solutions. In *IEEE workshop on Real-Time 3D Sensors*, 2004.

[6] Kölsch, M., and Turk, M. (2005). Hand Tracking with Flocks of Features. In *Video Proc. IEEE CVPR 2005.*

[7] Lee, M. et al. (2008). 3D Natural Hand Interaction for AR Applications. In *Proc. 23rd Int. Conference Image and Vision Computing* New Zealand, 26-28 Nov 2008.

[8] Lee, T. and Höllerer, T. (2007). Handy AR: Markerless Inspection of Augmented Reality Objects Using Fingertip Tracking. In *Proc. IEEE ISWC '07*, Boston, MA

[9] Mistry, P. et al. (2009). WUW - Wear Ur World - A Wearable Gestural Interface. In *Proc. CHI 2009*, pp. 4111-4116.

[10] Niikura, T. et al. (2010). In-air typing interface for mobile devices with vibration feedback. In *Proc. ACM SIGGRAPH 2010 Emerging Technologies*, Article 15.

[11] OpenCV http://opencv.willowgarage.com/wiki/ accessed on Dec. 23, 2010

[12] Piekarski, W. and Thomas, B. H. (2002). Using ARToolKit for 3D Hand Position Tracking in Mobile Outdoor Environments. In *1st Int'l AR Toolkit Workshop*, Darmstadt, Germany.

[13] Pierce, J. et al. (1997). Image Plane Interaction Techniques in 3D Immersive Environments. In *Sym. Interactive 3D Graphics,* pp. 39–44.

[14] Plagemann, C. et al. (2010). Real-time identification and localization of body parts from depth images. In *Proc. ICRA'2010*, pp.3108-3113.

[15] PMDTec CamBoard http://www.pmdtec.com/products-services/reference-design/ accessed on Nov. 19, 2010

[16] Rusu, R. B. et al. (2008). Towards 3D Point Cloud Based Object Maps for Household Environments. In *Robotics and Autonomous Systems Journal, Special Issue on Semantic Knowledge*, 2008.

[17] Sears, A., Jacko, J.A. (2008) The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, Second Edition. CRC ISBN: 0805858709

[18] Siek, K.A. et al. (2005). Fat Finger Worries: How Older and Younger Users Physically Interact with PDAs. In *Proc. INTERACT 2005*, pp. 267–280.

[19] Wang, R.Y., and Popovic, J. (2009). Real-Time Hand-Tracking with a Color Glove. In *ACM ToG SIGGRAPH 2009*, 28(3)

[20] Wigdor, D. et al (2006), "Under the Table Interaction", In *Proc. ACM UIST 2006*, pp. 259-268.