Binocular Interface: Interaction Techniques Considering Binocular Parallax for a Large Display

Keigo Yoshimura*
Graduate school of Interdisciplinary Information Studies,
The University of Tokyo

Takefumi Ogawa†
Information Technology Center,
The University of Tokyo

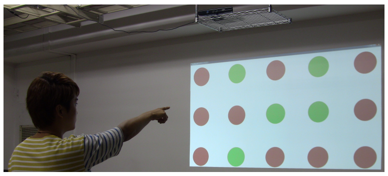

Figure 1: Overview of *Binocular Interface*

(a) (b) (c)

Figure 2: Operation methods; (a)*select*, (b)*move*, (c)*resize*

ABSTRACT

There have been many studies on intuitive user interfaces for large displays by using pointing movements. However, if a user cannot reach a display, object manipulations on the display are difficult because the user will see duplicate fingers due to binocular parallax. We propose *Binocular Interface*, which enables interactions with an object by using two pseudo fingers. In a prototype, pointing positions on the display are estimated on the basis of the positions of eyes and a finger detected by an RGB-D camera. We implemented three basic operations (*select*, move, and *resize*) using duplicate fingers and evaluated each operation.

Keywords: binocular parallax, pointing interface, large display

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction Styles, User Interfaces—Prototyping; D.2.2 [Software Engineering]: Design Tools and Techniques—User Interfaces

1 INTRODUCTION

With the recent advances in display technologies, large displays are now being used in an increasing number of fields. For example, we can easily obtain various types of information from large displays in shopping malls because most of these displays are equipped with touch sensitive screens.

As a display becomes larger, it becomes more difficult to access data on the upper part of the display. Naturally, if we cannot physically reach a display, we cannot access the data. While most of the existing interactive techniques assume users will be standing close to a large display [5], there has been some research proposing a user interface in which users can point to a display with a finger [6] or gesture toward it using a pen-shaped device [3], laser pointer [1] or other device [2]. Although it is easy for users to intuitively understand the pointing gesture, it is difficult to manipulate objects on a large display by pointing because a finger does not properly match up with objects due to binocular parallax. Our left and right eyes see the world from slightly different positions, so the images sensed by each eye are slightly different. This difference in the sensed image, called binocular parallax, is what the human visual system uses to detect depth cues. When users focus on a near object, they see distant objects in duplicate, and vice versa. Therefore, if users focus on the display, they cannot point at only one position because they see duplicated fingertips. On the other hand, if they focus on their fingertip—in other words, if they do not keep their eyes fixed on the target object— they cannot manipulate the target object. To solve the problem of binocular parallax, the pointing method for a large screen at a distance has been proposed [3]. This method required a transparent display in front of the user, and the restrictions for building the system are large.

* e-mail: k.yoshimura@cnl.t.u-tokyo.ac.jp
† e-mail: ogawa@nc.u-tokyo.ac.jp

IEEE Virtual Reality Conference 2015
23 - 27 March, Arles, France
978-1-4799-1727-3/15/$31.00 ©2015 IEEE

2 BINOCULAR INTERFACE

2.1 Approach

We have developed a novel interface called *Binocular Interface* that actually takes advantage of the duplicate finger images (called *binocular fingers*) created by binocular parallax. The proposed interface provides intuitive and intelligible methods for large display environments. When users focus on a display and move a finger back and forth between their head and the display, the interval between binocular fingers changes depending on the distance between their finger and eyes. Users can hold, move, or warp an object on the large display by using binocular fingers without the need for any pointing device such as a virtual glove or 3D mouse. *Binocular Interface* allows users to manipulate objects by binocular fingers as if they were using a touch display.

2.2 Three Basic Operation Methods

Our interface provides three basic operations. *Select* is realized by holding an object on a large display by binocular fingers as if actually holding something, like using chopsticks, as shown in Fig. 2(a). With the *move* operation, a user moves the selected object to somewhere on the display. A user can move the object by manipulating the binocular fingers in the same way something can be moved by using chopsticks, as shown in Fig. 2(b). *Resize* is realized by stretching or shortening an object, as shown in Fig. 2(c). To increase or decrease the size, the user moves the edges away from or toward the center of the object, similar to the pinching zoom in/out operation on a touch screen.

3 PROTOTYPE

Our prototype estimates the coordinates of a user's fingertip and eyes by way of a sensor installed in the environment. We implemented a prototype in Processing under Windows 7 using

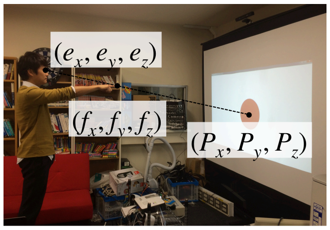

Figure 4: Estimation of pointing position

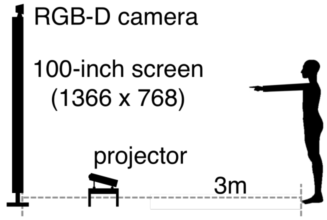

Figure 5: Experimental configuration

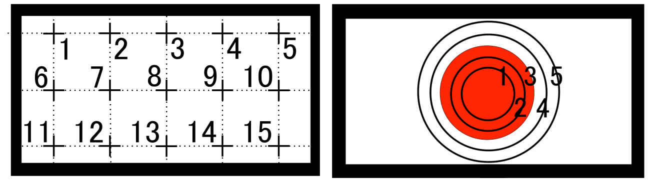

Figure 6: All targets: (a) 15 lattice points, (b) four targets and a basic circle.

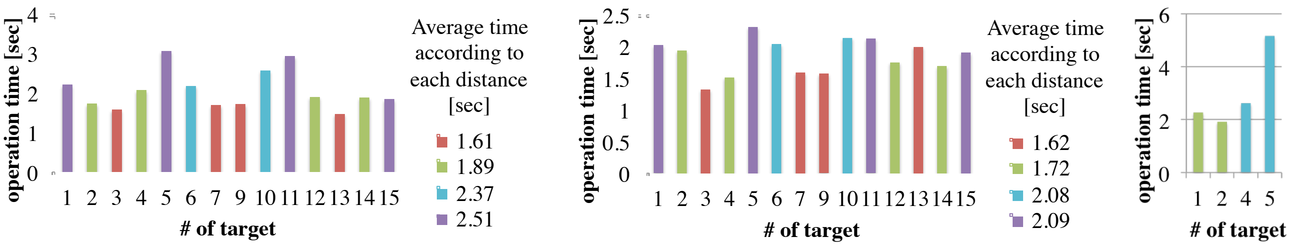

Figure 7: Average time of each operation: (a) *select*, (b) *move*, (c) *resize*

Kinect for Windows. In the system coordinates with their origin at the RGB-D camera, a user's pointing position on a display (P_x, P_y, P_z) is calculated by

$$P_x = f_x - f_z \cdot (e_x - f_x)/(e_x - f_z) \tag{1}$$

$$P_y = f_y - f_z \cdot (e_y - f_y)/(e_y - f_z) \tag{2}$$

$$P_z = 0 \tag{3}$$

where (f_x, f_y, f_z) and (e_x, e_y, e_z) denote the position of the fingertip and each of the eyes, respectively.

These three operations (*select*, *move*, and *resize*) are realized by referencing the changing distances of the binocular fingers. If the binocular fingers enter within 50 pixels from the edges of the object when a user is stretching his or her arm, the *select* operation is complete. During the *move* operation, the object is put on the midpoint of the binocular fingertips. The *move* operation is complete after the selection of the object is cleared. In the *resize* operation, if the user moves the finger back and forth, the size of the object is changed to fit the binocular fingers.

4 Experiment

To evaluate the detection accuracy of pointing positions, we computed their expected value and jitter when 5 participants overlapped binocular fingers on a target for three seconds. The experimental configurations are shown in Fig. 5. The average distance between target and estimated position of binocular fingers were 223 and 446 pixels under conditions in which participants stretched and bent their arm, respectively. The average of jitters under the stretching and bending conditions were 337 and 498 pixels, respectively. We also conducted usability tests to evaluate the three basic operations. A target circle was randomly presented at 14 lattice points excepting point no. 8 (starting point) on the display, as shown in Fig. 6(a). We let participants select the target after putting binocular fingers on the starting point or move the circle from the starting point to the target. The average amount of time it took a *select* operation to be performed is shown in Fig. 7(a). Although some variation exists, the average operation time increased with the distance from the center of the display. The operating time for selection depends on the travel distance of a finger from the above results. The average amount of time it took a *move* operation to be performed is shown in Fig. 7(b). When a participant moved the circle toward target no. 1, 6, and 10, those times were longer than others. That is, the task was more difficult when participants moved the circle toward the left part of the screen than to other parts. To investigate the operability of the *resize* operation, we let participants select circle no. 3 and resize it to fit circles no. 1, 2, 4, and 5, as shown in Fig. 6(b). The average amount of time it took to perform *resize* operations is shown in Fig. 7(c). As shown in the graph, the enlarge operation required more time than the shrink operation. The fact that the jitter is large when the hand is close to the head has an influence on the results in this experiment.

5 Conclusions

In this paper, we proposed a *Binocular Interface* for a pointing interface using large displays in which we make active use of two finger images created by binocular parallax. We described the three basic operations (*select*, *move*, and *resize*) of the proposed interface and carried out experiments to evaluate the basic performance and operability of these operations. The results show that users are able to manipulate an object using binocular fingers as directed. In future work, we plan to improve the accuracy of the pointing position estimation and examine additional operation techniques. This research was supported in part by a Grant-in-Aid for Scientific Research (C) numbered 25330227 by the Japan Society for the Promotion of Science (JSPS).

References

[1] Cheng, K., & Pulo, K.: Direct Interaction with Large-Scale Display Systems Using Infrared Laser tracking Devices, *Proc. of APVIS*, volume 24, pp.67-74, 2003.

[2] Baudisch, P., Sinclair, M., & Wilson, A.: Soap: a Pointing Device that Works in Mid-Air, *Proc. of UIST*, pp.43-46, 2006.

[3] Guimbretière, F., Stone, M., & Winograd, T.: Fluid Interaction with High-Resolution Wall-Size Displays, *Proc. of UIST*, pp.21-30, 2001.

[4] Lee, J. & Bae, S.: Binocular Cursor: Enabling Selection on Transparent Display Troubled by Binocular Parallax, *Proc. of SIGCHI*, pp.3169-3172, 2013.

[5] Ringel, M., *et al.*: Barehands: Implement-Free Interaction with a Wall-Mounted Display, *Proc. of CHI*, pp.367-368, 2001.

[6] Vogel, D. & Balakrishnan, R.: Distant Freehand Pointing and Clicking on Very Large, High Resolution Displays, *Proc. of UIST*, pp.33-42, 2005.