卒業論文

Parallax-controlled finger pointing interface for mixed reality environment

氏　　名：REN JIMIN
学籍番号：2600130451-7
指導教員：柴田 史久 教授
提出年月：2017年1月30日

立命館大学 情報理工学部 情報コミュニケーション学科

Abstract

Selecting an object far away from the standing position in mixed reality environment using finger is challenging. Because of the existence of relative binocular parallax between user's hand and the object of the target, the pointing directions of the finger from the left eye and the right eye differ slightly. This difference makes two variable paths passing the finger from both eyes, resulting in ambiguity for selections. Therefore, it makes selecting objects far away from the user becomes a difficult task and significantly impacts the user's performance.

In this paper, a novel method was introduced. This approach improved the user's performance on such tasks utilizing diminished reality techniques with stereo RGB cameras. By adjusting the binocular parallax of hand, the approach in this paper made the pointing position of the finger from both eyes the same. Thus, it kept pointing positions of the finger from both eyes match the same location. As a result, the performance and accurateness of pointing and selection task can be improved.

Keywords

Mixed Reality
Diminished Reality
Parallax
User-interface
Video-see-through Head-mounted-display

Table of Contents

Chapter 1 Introduction

In mixed reality environments, interacting with real objects is one of the most common tasks to do. For enhancement of the ability and experience in these scenarios, interacting objects not only near the user but also far away from the user becomes a sensible practice. Since physical haptic feedback in 3D user-interfaces usually absent, pointing to an element has a better experience comparing to touching. Such functionality can be achieved by several methods including body gesture recognition [1-3] and pointer devices [4, 5].

Such interfaces can be found in recent virtual-reality and mixed-reality devices like the Microsoft HoloLens; the Meta 2 Augmented Reality Development Kit, the Oculus Touch and et cetera. These devices use gesture recognition or sensor systems as a pointing input source. HoloLens, for instance, uses the user's head position and direction as a pointer and hand gestures or clicker devices as activators (see **Figure 1,2**). In user's perspective, a circle cursor appears in the center point of view and works as a pointer (see **Figure 3**). This method is relatively simple to implement, but the efficiency of pointing and selecting is limited because of the gaze direction and the direction that user wants to point differs. To pointing at user-interface elements such as menus and buttons, the user need to move their head, which lower user's performance of pointing and selecting.

Figure 1 HoloLens Input Devices [6]

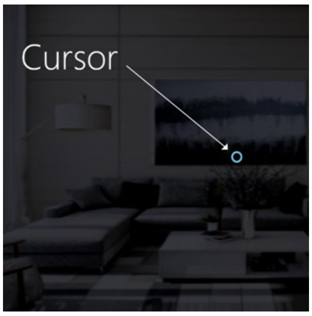

Figure 2 HoloLens Gaze Cursor [7]

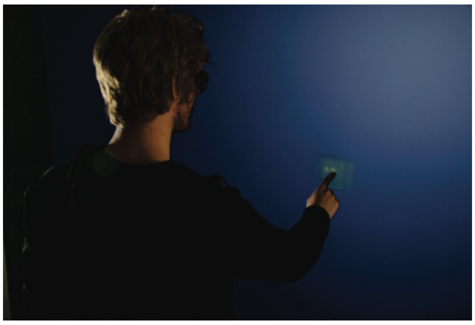

Figure 3 Hand as a 2D pointer in 3D Environment [8]

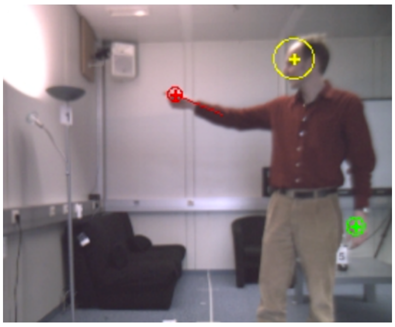

Figure 4 Arm as pointer [1]

Some systems use a hand as a pointer to operate their user-interfaces (see **Figure 3**). These systems use a 2D surface input device to obtain relative direction from user's perspective to interact with 3D user-interfaces. These systems need physical input surface to work. Thus, the user required to stay at one place for operation. Other systems use arm direction to point at elements of user-interface (see **Figure 4**). These systems used only the arm's direction and ignored user's perspective, which differs from the daily experience of pointing.

A better solution can be using a finger in mid-air within user's perspective to create a natural experience of pointing objects. Nevertheless, visual conflicts significantly impact user's performance when selecting an object that is far away from the user using their finger as a pointer. For example, with the existence of binocular parallax, either the hand or the background appears in two slightly different positions depend on user's focus point. This phenomenon creates an ambiguity of selecting direction, which decreases both the performance and the accuracy of 3D mid-air selection tasks.

In this paper, we presented a novel method with diminished reality techniques controlling binocular parallax to improve user's performance of pointing and selecting tasks, designed experiments to test our method, and analyzed the result.

In the related work section, the history of pointing methods in mixed reality environment was depicted. The performance, advancements, and drawbacks of these methods were compared. In the third section, the method introduced in this paper was discussed in details. In experiment section, experiments evolved in this research was compared, and then been showed the result and discuss what can we learned from the result. In the fifth section, the content of this paper was summed up and the future direction of this research was illustrated.

Chapter 2 Related work

In the last section, the background of our research was interpreted. In this section, other researches related to 3D mid-air selection using hand were depicted.

2.1 Direct selection of virtual objects

2.1.1 Fish tank system

Recently, many approaches for 3D mid-air selection using hand have been proposed [9, 10]. For instance, J. Lee has presented a system for spatial 3D user-interface operation [9]. They have introduced a system using a 2D surface touch input to manipulate floating 3D elements including typing, drawing and clicking (see **Figure 5**). In this mechanism, the user sees through a transparent display showing user-interface elements, and operating these elements with hands behind the display. With face tracking techniques, the user could move their head and still get a result without visual conflicts. Since this system is designed as a fish tank system, the scale of the device is limited, which means it could not be used in scenes other than desktop environments. Also, this system did not resolve the ambiguity of selection from binocular parallax.

2.1.2 Other systems

Several researchers have worked on direct selection method of virtual objects with other types of systems. These methods are usually enabled by tracking technologies, such as optical marker systems, the Leap Motion, or the Microsoft Kinect. However, lacking physical feedbacks and suffering from visual conflicts, users tend to have an ambiguity of depth perception and object interrelations, which eventually lead to a lower performance of pointing and selecting because of a significant number of overshoot errors [11]. L. Chan et al. [12] have created a

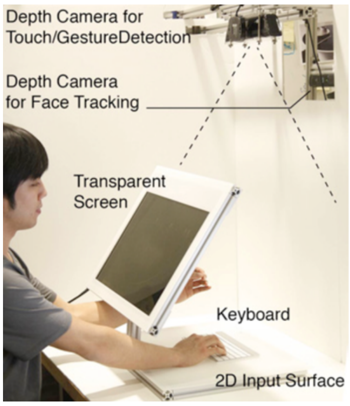

Figure 5 Device of J. Lee et al [9]

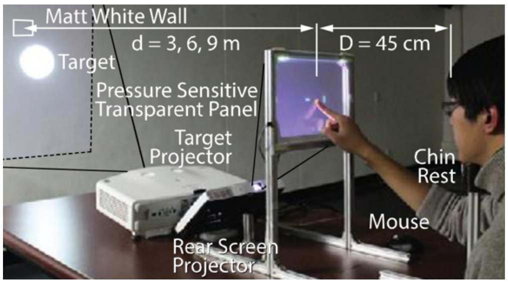

Figure 6 Device of J.H. Lee et al [10]

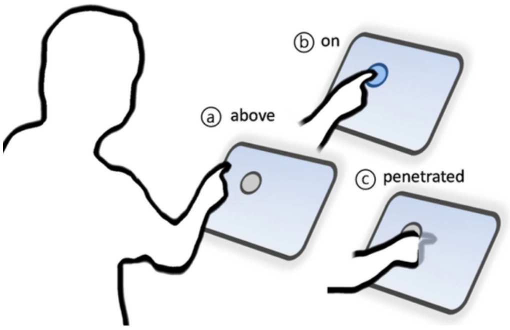

Figure 7 Selecting a mid-air target without physical feedbacks [12]

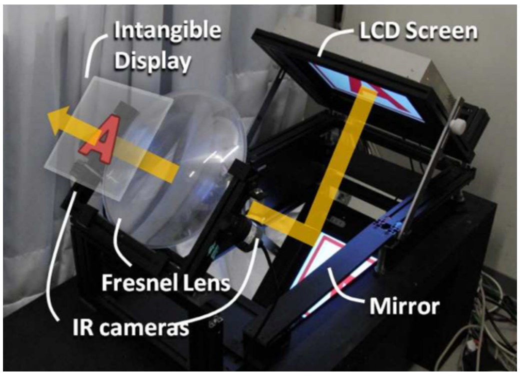

Figure 8 [12] System built by L. Chan et al.

system resolving such problem by setting an intangible display in front of the standing point of the user. The display moves to the surface aligned with the virtual object and when the user's hand approaches the surface, the intangible display simulates the shadow normally appears in such situation to give the user a more accurate and precise perception of depth. The experiment results showed a significant increase in user's performance of pointing and selecting virtual objects. J.H. Lee et al. [10] have presented a system with a transparent display to aid the user selecting objects beyond the display. They have developed a mechanism to resolve binocular selection ambiguity by visualizing the correct selection point (see **Figure 6**) instead of preventing binocular parallax from showing. Although their experiments showed a positive result of improving user's performance of selecting, by avoiding the negative influence of binocular parallax, they did not solve visual conflicts. Moreover, the setup of their experiments shows that the system needed a fixed transparent display to work, which limits the availability in other scenarios.

2.2 Offset-based Selection of Virtual Objects

Offset-based selection of virtual objects is a method that has been researched for over a decade. Mine et al. [11] have investigated the difference between interaction with objects that directly by hand and that with a fixed or mutable offset, and shown that an offset significantly decreased user's performance. I. Poupyrev et al. [16] have shown a method called the Go-Go technique that by extending the length of the virtual arm in virtual-reality environment, the user could interact with objects beyond the range of the user's arm (see **Figure 9**). Later they have investigated the difference of user's performance between the Go-Go technique and the ray-casting technique (see **Figure 10**) and found the performance are basically identical. A. Paljic et al. [14] used a tracked stylus held by the user to control a virtual cursor for selecting objects (see **Figure 11**). They found no significant difference in user's performance of pointing between when the offset distance is set to 0cm and 20cm, but found the user operate slowly when the offset is set to 40cm or 50cm. G. Bruder et al. [15] have investigated the effect of visual conflicts for mid-air 3D selection performance using the hand as a pointer directly and using an offset hand as a cursor (see **Figure 12**). They found that the moving time of former method is significantly less than latter scenario, yet with a higher error rate.

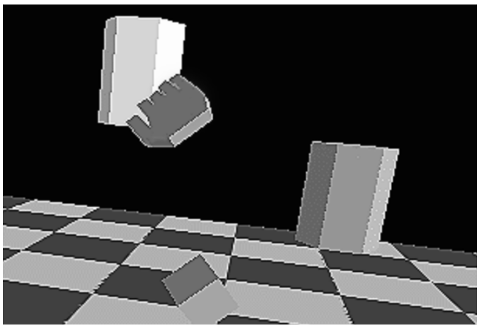

Figure 9 Interact with virtual object using the Go-Go technique [13]

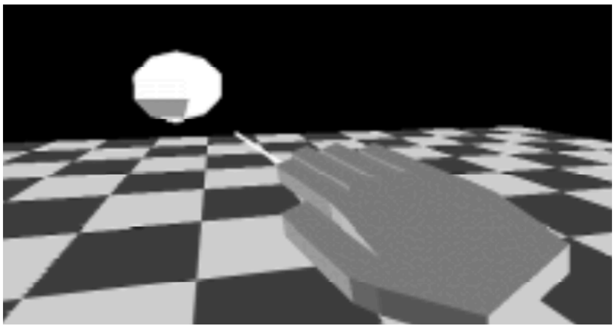

Figure 10 Interact with virtual object using ray-casting technique [13]

Figure 11 Interact with virtual object using a virtual crosshair cursor [14]

Figure 12 Interact with virtual object using an offset hand [15]

Chapter 3 Parallax-controlled finger pointing interface

In the last section, various 3D mid-air selection methods and compared these implementations were depicted. In this section our novel method to improve the user's performance on 3D mid-air selection tasks was discussed.

3.1 Introduction

A novel method was introduced for repositioning finger in images from a pair of stereo cameras. This method used diminished-reality techniques to control the binocular parallax of the finger. The images were captured from RGB stereo cameras, then were processed with a computer and were showed the developed images on a video-see-through head-mounted-display. **Figure 13** shows an overall flowchart of our method.

In practice, fixed positions of the user and the background were used to simplify the optical model and implementation, and the direction that the user faces was at right angles to the background. Therefore, the distance between the user and the background was known, and binocular parallax of the background of different position was relatively equal.

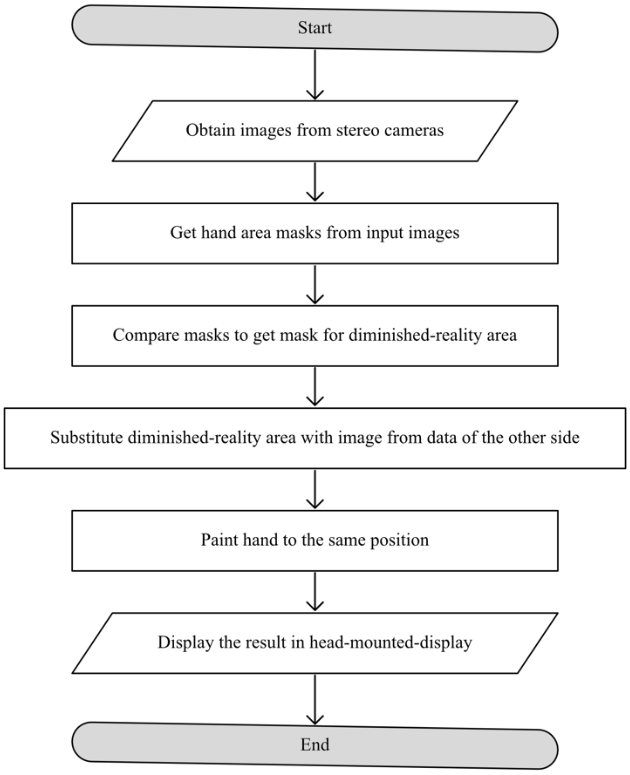

Figure 13 Parallax control with diminished reality technique

3.2 Equipment configuration

Stereo cameras were attached in front of a video-see-through head-mounted display to obtain binocular images. After processed these images with our program, the result was shown in the video-see-through head-mounted display. For communication with real objects, the main computer was connected to a router, which also connects a controller module. There were several LED lights attached to this module.

3.3 Hand area detection

As shown in **Figure 14**, color selection was the approach to detect hand areas in images from stereo cameras. Simply selecting a color range was practical in simple scenarios, in complicated scenes, however, did not work well because there were similar color blocks in the background (see **Figure 14(c)**). To achieve a better hand area selection, an additional mask was added (see **Figure 14(d)**). This mask came from another color range selection but with converted HSV color space. Then, the common pixels from these two masks were filtered to create an enhanced mask (see **Figure 14(e)**). The accuracy of hand area detection was pushed further by eroding and dilating this enhanced mask to get rid of irrelevant blocks in the background (see **Figure 14(f)**). Finally, image segmentation and labeling were applied and only the cluster with the biggest area is selected (see **Figure 14(g)**).

3.4 Diminished Reality

After getting hand area masks of both sides from stereo cameras, these masks were compared to find the inpainting area that the background was visible from one side but blocked by hand from the other side (see **Figure 15(e)**). To paint these areas without visual glitches, an offset of the input image was employed to cover the binocular parallax of the background. Since the surface of the background was flat and the angle between the background and the direction that the user's faces was a right angle, the binocular parallax did not change much between the central area and peripheral area. Then the compensate masks were applied (see **Figure 15(e)**) to offset-adjusted images for obtaining compensate images (see **Figure 15(f)**). Next, the inpainting area from input images was substituted with these compensate images for removal of the area that were blocked by hand from exact one eye (see **Figure 15(g)**).

Finally, the hand image was painted to both sides at the same position for binocular parallax control. The hand image was created from one of hand masks depend on the target hand position to obtain hand images from input images (see **Figure 15(c)**). For instance, **Figure 15** shows the procedure to align hand positions to the right eye, which used the image from the right camera to obtain the hand image. This image was used after diminished-reality technique application.

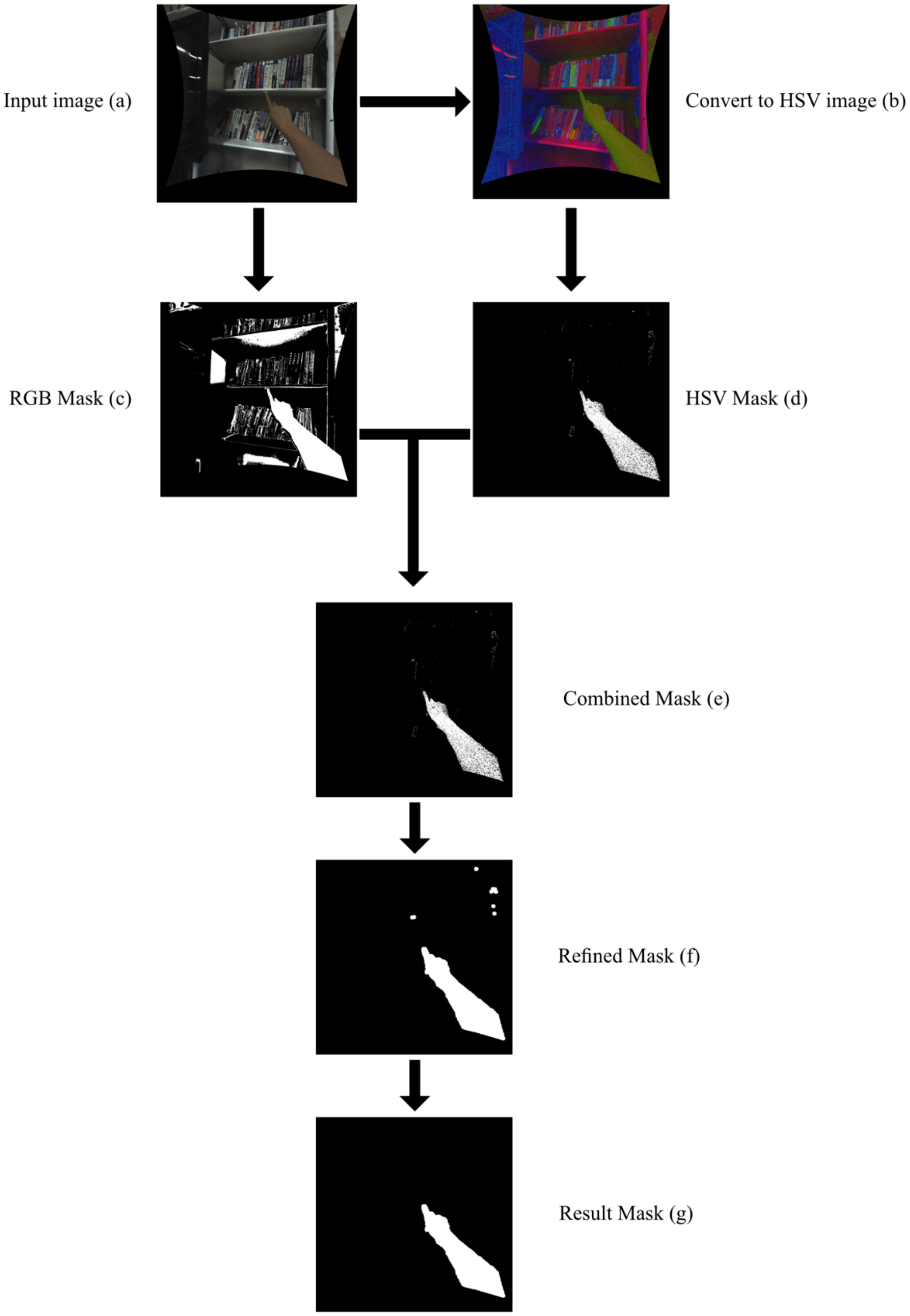

Figure 14 Hand area detection

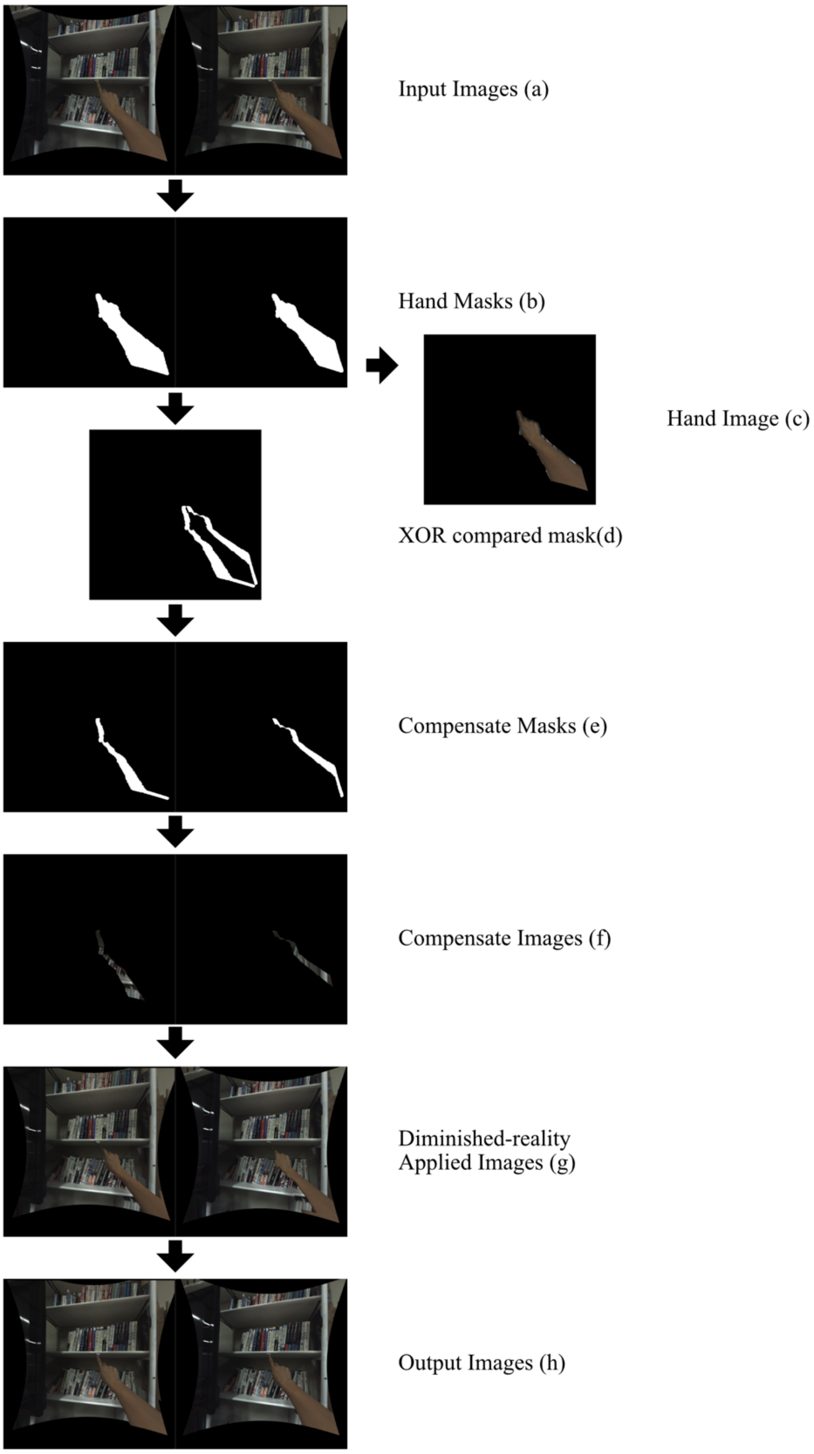

Figure 15 Input images

Figure 16 Obtaining finger position

3.5 Finger position detection

Because a single finger was used as a pointer, and the stereo cameras were attached in front of the video-see-through displays, when the user performed selecting gesture, the shape of hand area was predictable. To obtain the position of the finger used as a pointer, the point with the longest distance to the centroid of the cluster of hand was searched from the top side and the opposite side of the dominant hand (see **Figure 16**). For instance, if the user is dexterous (right-handed), the point will be chosen from the top side and left the side that with the longest distance to the centroid in the hand area cluster we obtained that introduced in the last chapter.

3.6 Interact with real objects

To interact with real objects, a mechanism was introduced using several LEDs as signal and get the position with image processing techniques. RGB color selection was the method to get the brighter area when LED is light up, then the centroid of the brighter area was recognized as the target position. Then the position of the target and the finger was compared to determine whether the finger was pointing at the target. The target would seem as hovered by the finger when the distance was less than a specific value. After a specific period of time, the item seemed as selected.

Chapter 4 Experiment

In the last section, the novel method to improve the user's performance of 3D mid-air selection was described. In this section, the experiments in this research, showed the result and discussed what we learnt from the result was illustrated.

4.1 Objective

The objective of the experiment was to inspect whether the method of this research improved user's performance of 3D mid-air selection task. Specifically, the time consumption of moving to a given selection task were measured. The seven LEDs lighted up one by one in a fixed order (see **Figure 20**). The participant needed to use their finger to point at the LED which was turned on and keep pointing it for one second to select one target. When all of seven targets were selected, the time would be measured. Four scenarios were compared: Scenario without parallax control, Scenario with parallax aligned to dominant eye, Scenario with parallax aligned to the center of two eyes, Scenario with parallax aligned to the non-dominant eye. Thus, we could check if the method works and if it works, and found the most optimized configuration of our method.

4.2 Experiment environment

The experiment environment is shown in **Figure 17** and **18**. We installed 7 LED lights on a bookshelf in three layers, which lights up in a fixed sequence (see **Figure 20**). The distance between the standing point and the bookshelf was 150 centimeters.

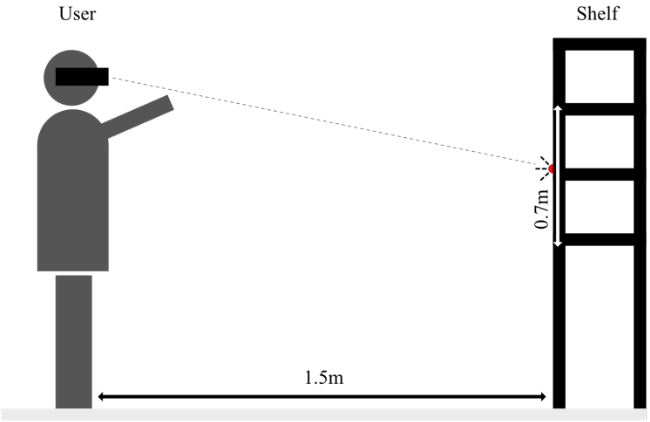

Figure 17 Experiment environment

Figure 18 Shelf configuration

4.3 Environment configuration

As shown in **Figure 19**, we adopt OVR Vision Pro and Oculus Development Kit 2 as our video-see-through head-mounted-display, and a high-performance computer for processing image. The specifications of the computer were listed in table 1 and the specification of the head-mounted display were listed in table 2. To implement the interaction with real world object, we used a Raspberry Pi A+ to control the LEDs and to communicate with the main computer via TCP communication over a router.

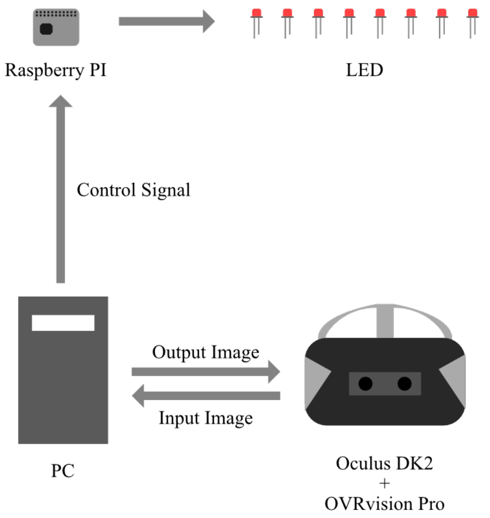

Figure 19 Environment configuration

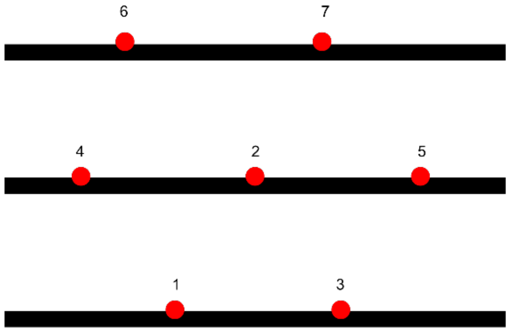

Figure 20 LED Sequence

Table 1 Computer specifications

CPU	Intel Core i7-6700
GPU	NVidia GeForce GTX 1060
RAM	16GB
Storage	256GB SATA III Solid state drive
Network interface	1000Mbps

Table 2 Head-mounted display specification

Camera resolution	960x950 (each eye)
Camera framerate	60 fps
Camera view angle	Horizontal 100° Vertical 98°
Camera Aperture	F1.8
Display resolution	960x1080 (each eye)

4.4 Experiment procedure

There were two patterns with different orders of target selection task for unfair factors cancellation. The alternative experiment pattern had the order from Step 4 to Step 7 reversed. Everything else was identical. 50 percent of participants were tested with the alternative experiment order. The default experiment procedure is listed below.

Step 1: Stand at the experiment position, wear the head-mounted display.
Step 2: Move right hand to the center of the view of the stereo camera.
Step 3: Capture one frame for setting threshold parameters.
Step 4: Select all seven targets without parallax control.
Step 5: Select all seven targets with parallax aligned to right eye.
Step 6: Select all seven targets with parallax aligned to the center of two eyes.
Step 7: Select all seven targets with parallax aligned to left eye.
Step 8: Take dominant eye test (will be discussed in **4.6**).

The alternative experiment procedure is listed below.

Step 1: Stand at the experiment position, wear the head-mounted display.
Step 2: Move right hand to the center of the view of the stereo camera.
Step 3: Capture one frame for setting threshold parameters.
Step 4: Select all seven targets with parallax aligned to left eye.
Step 5: Select all seven targets with parallax aligned to the center of two eyes.
Step 6: Select all seven targets with parallax aligned to right eye.
Step 7: Select all seven targets without parallax control.
Step 8: Take dominant eye test (will be discussed in **4.6**).

The moving time consumption and the number of errors occurred of every target were measured.

4.5 Participants

10 male subjects (ages 18-26) participated in the experiments. Subjects were students of media information department and information communication department.

4.6 Dominant eye determination

The point-a-finger test (Porta Test) was used to determine the user's dominant eye [17, 18]. The instructor asked the participant to hold a pen with both eyes open. Then, they were requested to point the pen to an object that was 6 meters away from the participant. The participant was asked to alternately close each eye, and tell from which eye, the direction from the eye through the pen tip to the target was correct. The eye which was viewing the target correctly was determined to be the dominant eye.

4.7 Result

The time consumption of moving and the number of errors occurred of every target selected were collected. The time consumption was the duration of moving to the pointing position and selecting each object. The number of error was defined as the number of frames when the position of user's finger was around of the target but not hovering.

According to some participants' comments, the overall experience of target selection of scenarios with parallax control were better than the scenario without parallax control due to the ambiguity of selection caused by binocular parallax, and the experience when the parallax had aligned to one eye was better than when the parallax had aligned to center of two eyes, because some additional visual glitches were introduced. This problem was caused by our hand area selection method, which dilated the area to gain a better mask. This method, however, expanded the mask by several pixels, thus covered a little area of background surrounding of hand. When the parallax was aligned to a position between two eyes, both hand and this portion of background were copied to the target position.

4.7.1 Scenario without parallax control

As shown in **Figure 21** and **Figure 22**, the experiment result from scenario without parallax showed variant moving times and errors occurred per target. The horizontal axis from N1 to N7 indicates the order of the target and the vertical axis indicates the time consumption in second (Figure 21) or errors occurred per target (**Figure 22**).

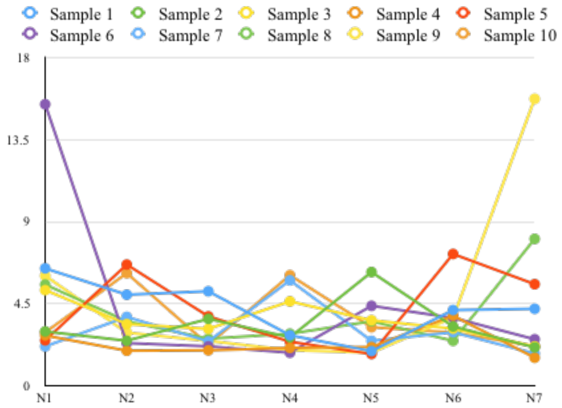

Figure 21 Time consumption per target

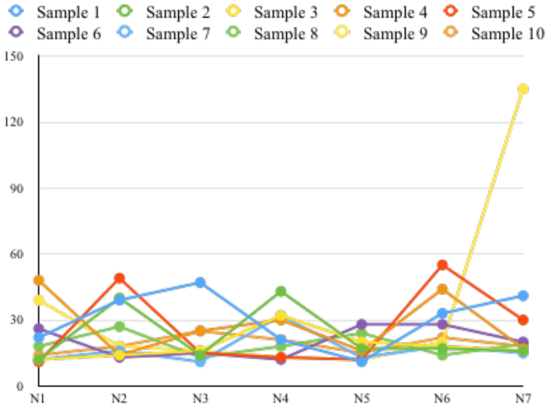

Figure 22 Errors occurred per target

4.7.2 Scenario with parallax aligned to dominant eye

Figure 23 and **Figure 24** showed a slightly converged and stable time consumption and errors occurred per target from the scenario with parallax aligned to dominant eye.

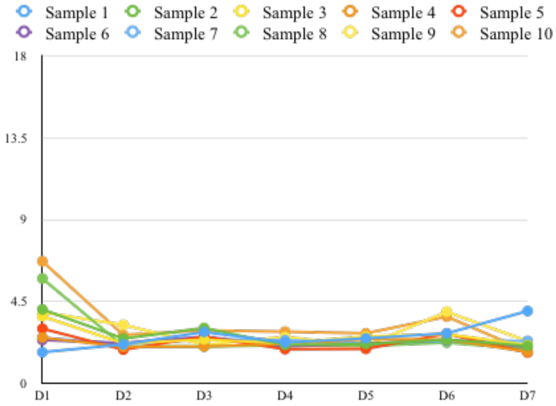

Figure 23 Time consumption per target

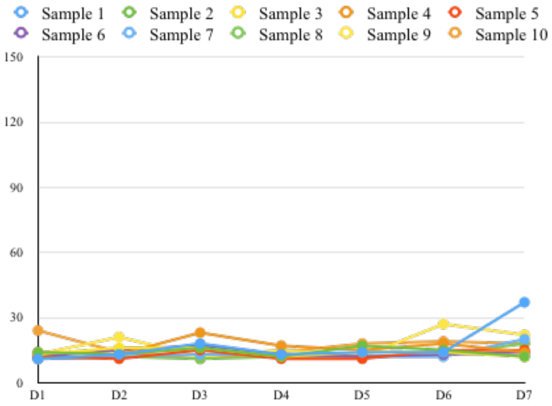

Figure 24 Errors occurred per target

4.7.3 Scenario with parallax aligned to center of two eyes

From the results shown in **Figure 25** and **Figure 26**, the time consumption records in scenario with parallax aligned to center of two eyes were more deviated than the scenario with parallax aligned to the dominant eye but less fluctuated than the scenario without parallax control, and the error occurred per target records were similar to the scenario with parallax aligned to dominant eye.

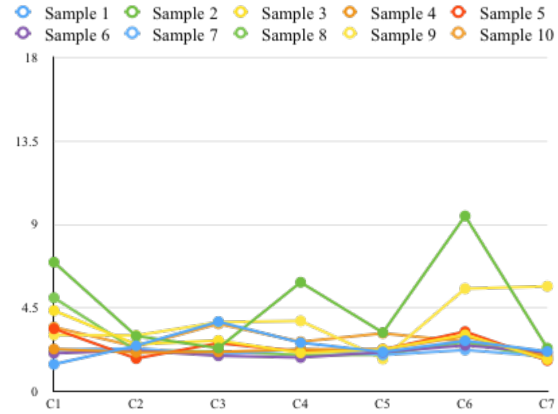

Figure 25 Time consumption per target

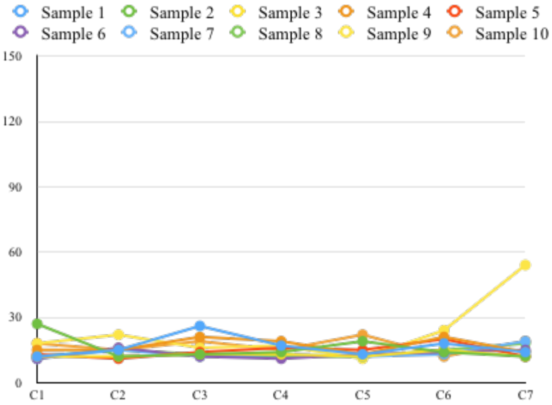

Figure 26 Errors occurred per target

4.7.4 Scenario with parallax aligned to non-dominant eye

Figure 27 and **Figure 28** showed a similar result as the scenario with parallax aligned to dominant eye with stable time consumptions and errors occurred per target.

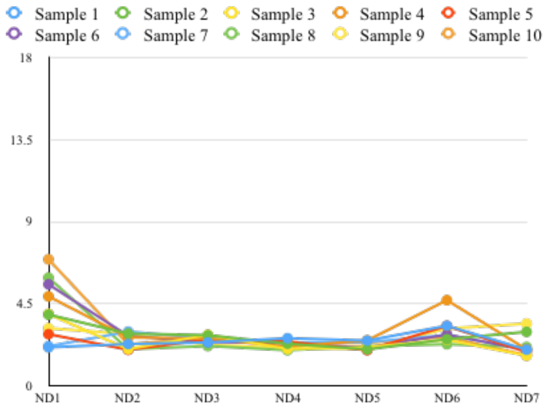

Figure 27 Time consumption per target

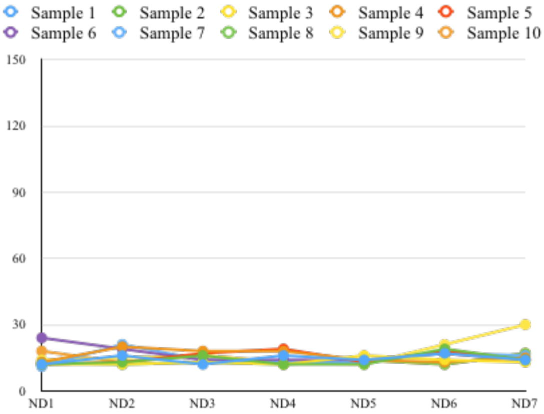

Figure 28 Errors occurred per target

4.8 Discussion

From the result of experiments, we learned that the user's performance of selection task with parallax control was significantly better than scenarios without parallax control (see **Figure 29, 30**). Specifically, the average time consumption of selecting one target with parallax aligned to the dominant eye was 2.6 seconds, which was 26.8% quicker than without parallax control. At the same time, errors occurred in the parallax controlled scenario was 33.8% less than scenario without parallax control. Moreover, according to analysis of variance with post-hoc Tukey-Kramer test, the differences between the scenario without parallax control and other scenario were significant. Therefore, the result indicated that our interface with parallax control could increase the user's performance of 3D mid-air selection.

On the other hand, although the results of scenarios with parallax control showed that when the parallax was aligned to either eye, the average time consumption of moving hands and errors occurred per target were less than when the parallax was aligned to the center of both eyes, according to analysis of variance with post-hoc Tukey-Kramer test, the difference was not significant. It was the same between the scenario with parallax aligned to dominant eye and the scenario with parallax aligned to non-dominant eye. Also, according to comments from some participants, when the parallax was set to the center of two eyes, because of the existence of error in finger positions, the rendered finger was not stable and there were some visual conflicts at the surrounding of the hand.

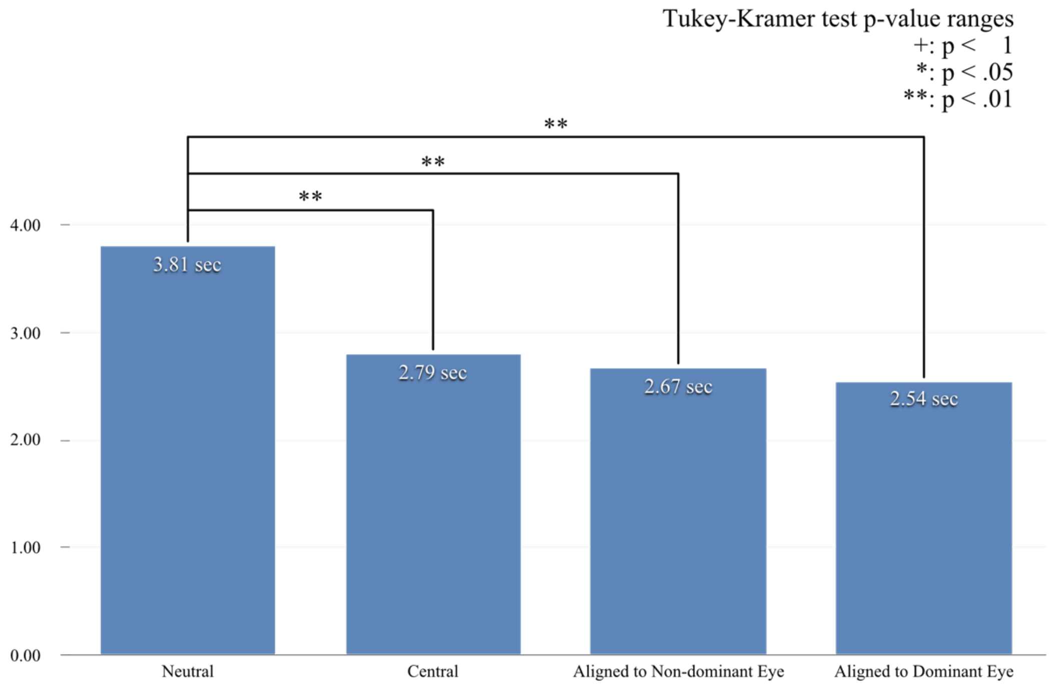

Figure 29 Average time consumption per target (less is better)

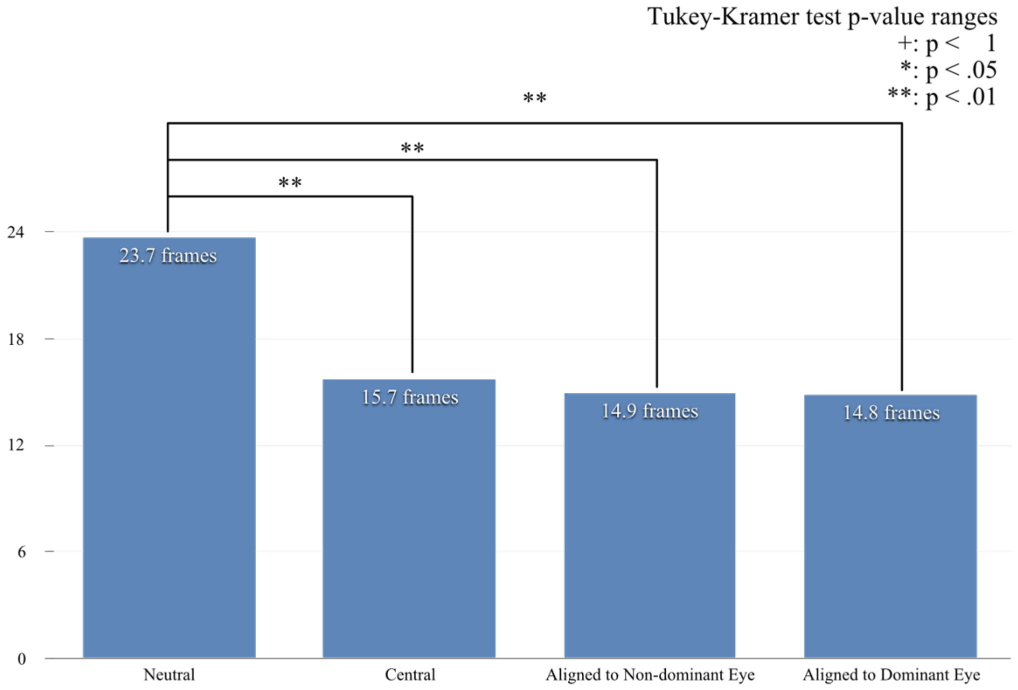

Figure 30 Average errors occurred per target (less is better)

Chapter 5 Conclusion

In this paper, a novel approach was introduced, this method controlled binocular parallax of finger in mixed-reality environments to improve the user's performance of 3D mid-air pointing and selection task with diminished-reality techniques, interpreted our experiments and discussed the results of experiments.

From the experiment results, we learned that by controlling binocular parallax and aligning the parallax to either dominant eye or non-dominant eye, the user's performance could be improved significantly. Therefore, the performance of 3D mid-air pointing and selection task could be improved with this new method.

On the contrary, several drawbacks were found. The overall performance was not ideal and the framerate of our system needed to be improved to reduce lag of moving and discomfort of user experience. Also, the visual glitch happened when the hand position was aligned to the middle between hand positions from two eyes, which impacted both the user's performance and experience. Furthermore, since the color selection method in this paper was fixed, to obtain masks of user's hand, the illumination condition must be consequent, and the capacity of moving was limited. In the future, we want to resolve these problems and improve our system.

Finally, we expect our system to be and enhance the user's performance of pointing and selecting in utilized in a variety of scenarios.

Acknowledgements

The author would like to thank Fumihisa Shibata whose comments and suggestions were of inestimable value for my study. The author would also like to thank Sei Ikeda, Asako Kimura, Hideyuki Tamura from the Faculty of Information Science and Engineering, Ritsumeikan University. who gave me invaluable comments and warm encouragements. This work was supported by Mobile Computing Laboratory and Reality Media Laboratory.

References

[1] Kai Nickel, R. Stiefelhagen: “Real-Time Recognition of 3D-Pointing Gestures for Human-Machine-Interaction,” DAGM 2003, Pattern Recognition, pp. 557 – 565, 2003.

[2] G. Yepeng, Z. Mingen: “Real-time 3D pointing gesture recognition for human HCI,” WCICA 2008, Intelligent Control and Automation, pp. 2433 – 2436, 2008.

[3] E. Hosoya, H. Sato, M. Kitabata, I. Harada, H. Nojima, A. Onozawa: “Arm-Pointer: 3D Pointing Interface for Real-World Interaction,” CVHCI 2004, Computer Vision in Human-Computer Interaction, pp. 72 – 82, 2004.

[4] K. Zidek, J. Pitel: “Smart 3D Pointing Device Based on MEMS Sensor and Bluetooth Low Energy,” 2013 IEEE Symposium on CICA, pp. 207 – 211, 2013.

[5] P. Baudisch, M. Sinclair, A. Wilson: “Soap: a Pointing Device that Works in Mid-Air,” UIST 2006, Proceedings of the 19th annual ACM symposium on User interface software and technology, pp. 43 – 46, 2006.

[6] Hardware details, Windows Dev Center, https://developer.microsoft.com/en-us/windows/holographic/hardware_details (Jan 12, 2017)

[7] Gaze, Windows Dev Center, https://developer.microsoft.com/en-us/windows/holographic/gaze/ (Jan 13, 2017)

[8] D. Valkov, F. Steinicke, G. Bruder, K. Hinrichs: “2D Touching of 3D Stereoscopic Objects,”, CHI 2011, 3D Interaction, pp. 1353 – 1362, 2011.

[9] J. Lee, A. Olwal, H. Ishii, C. Boulanger: “SpaceTop: Integrating 2D and Spatial 3D Interactions in a See-through Desktop Environment,” CHI 2013, 3D User Interfaces, pp.189 – 192, 2013.

[10] J.H. Lee, S. Bae: “Binocular Cursor: Enabling Selection on Transparent Displays Troubled by Binocular Parallax,” CHI 2013, Spatial Interfaces, pp. 3169 – 3172, 2013.

[11] M. Mine, F. Brooks Jr., C. Sequin: “Moving objects in space: exploiting proprioception in virtual-environment interaction,” SIGGRAPH 1997, pp. 19 – 26, 1997.

[12] L. Chan, H. Kao, M. Chen, M. Lee, Y. Hsu, Y. Hung: “Touching the void: direct-touch interaction for intangible displays,” CHI 2010, pp. 2625 – 2634, 2010.

[13] I. Poupyrev, S. Weghorst, M. Billinghurst, T. Ichikawa: “Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques,” Computer Graphics Forum, September 1998, Volume 17, Issue 3, pp. 41 – 52, 1998.

[14] A. Paljic, S. Coquillart, J. Burkhardt, P. Richard: “A Study of Distance of Manipulation on the Responsive Workbench,” In Proc. Of IPT 2002, pp. 1 – 8, 2002.

[15] G. Bruder, F. Steinicke, W. Stuzlinger: “Effects of Visual Conflicts on 3D Selection Task Performance in Stereoscopic Display Environments,” IEEE 2013, Symposium on 3D User Interfaces, pp. 115 – 118, 2013.

[16] I. Poupyrev, M. Billinghurst, S. Weghorst, T. Ichikawa: “The go-go interaction technique: non-linear mapping for direct manipulation in VR,” UIST 1996, ACM symposium on User Interface software and technology, pp. 79 – 80, 1996.

[17] L. Roth, N. Lora, M. Heilman: “Effects of monocular viewing and eye dominance on spatial attention,” Brain 2002, Volume 125, Issue 9, pp. 2023 – 2035, 2002.

[18] Z. Khan, J. Douglas: “Ocular dominance reverses as a function of horizontal gaze angle,” Vision Research 2001, Volume 41, Issue 14, pp. 1743 – 1748, 2001.