Comparison between Gaze Input and a Hand-Held Controller in AR Interaction

Mie Sato

mie@is.utsunomiya-u.ac.jp
Utsunomiya University, 7-1-2 Yoto, Utsunomiya, Tochigi 321-8585, Japan
Keywords: Eye-tracking, Augmented reality, Human-computer interaction

ABSTRACT

In the interaction with virtual objects, operations that do not require the use of a controller are expected as a convenient way. This study reports comparison between gaze input and a hand-held controller in augmented reality (AR) interaction. Two experiments are conducted: as a cue for gaze input, one uses a wink and the other does the gaze time. From results of the experiments, we summarize the challenges and prospects of gaze input in AR interaction.

1 Introduction

Recently, interaction with virtual objects has gathered attention [1-4]. While controllers are generally used for the interaction, the interaction without using the controllers has been expected to be more convenient because of non-equipment and hands-free. Eye-tracking enables hands-free interaction. Eye-tracking interactions in VR (virtual reality) have been reported as useful methods similar to those using controllers and mice [5]. Gaze is primarily used as a substitute for a mouse or controller pointer. Interaction using gaze input has mainly focused on object selection and simpler selection methods have been examined. However, most studies have suggested the use of the hand or controller as auxiliary input [6, 7], which do not fully allow for hands-free interaction with virtual objects. The studies aimed at hands-free interaction showed the use of the line of sight and the orthogonal plane to move virtual objects [8] and the use of gaze input to rotate virtual objects [9] in VR environments. Those gaze-based interactions have been conducted in VR environments and not enough studied in AR. We have been studying gaze interaction in AR environments. The AR environments have difficulties in accurate spatial mapping between the real and virtual spaces.

The purpose of this study is to compare interaction by gaze input and by a hand-held controller in an AR environment. We conduct two experiments: as a cue for gaze input, one uses a wink and the other does the gaze time. Based on results of the experiments, we discuss the use of gaze input as an interactive way of moving a virtual object in an AR environment.

2 Experiments

We conduct two experiment to compare operations for moving a virtual object in an AR environment. The one is comparison between operations by gaze input with a wink and by a hand-held controller, and the other is one between by input using the gaze time and by a hand-held controller.

2.1 System

We created an AR environment using an HMD with cameras (HTC VIVE Pro Eye). Gaze input from the subject was obtained from an eye-tracker mounted on the HMD. A hand-held controller used here was the HMD accessory. We implemented a system using Unity (2018.4.3f).

We used three operations to move the virtual object:

Gaze input (Wink):

1. To hold the virtual object, gaze at it at the starting position and close the right eye.
2. To place the virtual object at the position where the subject wants to move it, gaze at that position and close the left eye.

Gaze input (Time):

1. To hold the virtual object, gaze at it at the starting position for one second.
2. To place the virtual object at the position where the subject wants to move it, gaze at that position for one second.

Controller:

1. To hold the virtual object, point the controller at the virtual object at the starting position and pull the controller's trigger.
2. To place the virtual object at the position where the subject wants to move it, keeping the trigger pulled, point the controller at that position, and then release the trigger.

Note that during these operations, as shown in Fig. 1, the color of the virtual object was blue at the beginning and it was changed to red when the subject gazed or pointed at the virtual object, and then the virtual object became translucent while the subject was holding it.

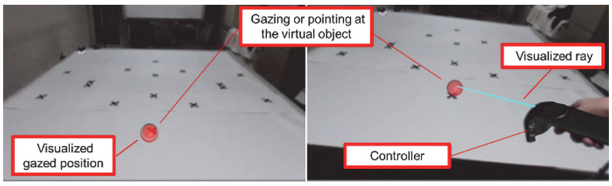


Fig. 1 Subject's point of view.
Gaze input (right) and controller input (left)

2.2 Experimental environments

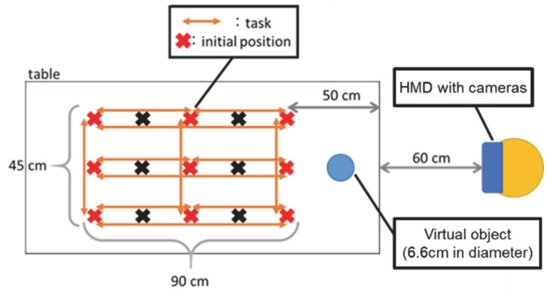
The experimental environments are shown in Fig. 2. Fig. 2(a) is for the comparison between gaze input with a wink and a controller, and Fig. 2(b) are for the comparison between input using the gaze time and a controller. The subject sitting on a chair wears the HMD. The heights of the table and the HMD are 70 cm and approximately 150 cm, respectively. The X marks are written on the table in the real environment. The diameter of the virtual object is 6.6 cm.

2.3 Comparison of Gaze (Wink) and Controller

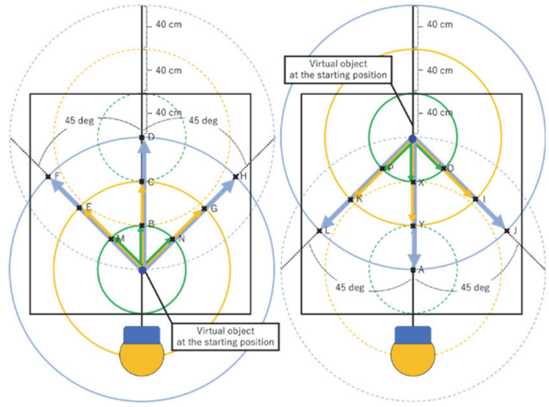
The preliminary experiment suggested that moving the virtual object back and forth may be distinguished from moving it in other ways. Therefore, we mainly focused on the task of moving the virtual object back and forth.

The subject was instructed to move the virtual object from an X mark at the initial position to other X marks. The initial position was chosen from nine red X marks by the experimenter. The nine X marks are shown in red in Fig. 1(a), while the color of the marks was black in the actual experiment. The subjects moved the virtual object at the initial position onto the X mark indicated by the experimenter. The task of moving the virtual object was only back and forth or from side to side. As the arrows shown in Fig. 1(a), there were eighteen tasks for moving back and forth and six tasks for moving from side to side. The total number of tasks was 24. These tasks were performed in random order. Note that each of the subjects performed 48 tasks by the gaze and by the controller. The tasks by the gaze or by the controller were performed by turns. Half of the subjects performed the tasks by the gaze first, while the other half performed the tasks by the controller first.

Fourteen naive volunteers (age range, 21-25 years) participated in this experiment. Most subjects did not have previous experience with VR devices. All had a normal or corrected-to-normal vision. The dominant eye of the four subjects was the left eye and the dominant eye of eight subjects was the right eye. Two subjects did not know their dominant eye.



(a) Comparison of gaze (wink) and controller



(b) Comparison of gaze (time) and controller
Fig. 2 Experimental environment

2.4 Comparison of Gaze (Time) and Controller

We prepared two groups of the task: one is “from the near to the far” and the other is “from the far to the near,” as shown in Fig. 2(b). In Fig. 2(b), the X marks are arranged in concentric circles with three different diameters: near (green), middle (yellow), and far (blue).

The subject was instructed to move the virtual object from an X mark at the starting position to other X marks. For each group, the subjects moved the virtual object at the starting position onto one of the nine X marks indicated by the experimenter. Each group had nine tasks and the total number of tasks was eighteen. These tasks were performed in random order for each group. Note that each of the subjects performed 36 tasks by the gaze and by the controller. The tasks by the gaze or by the controller were performed by turns. Half of the subjects performed the tasks by the gaze first, while the other half performed the tasks by the controller first.

Ten naive volunteers (age range, 20-24 years) participated in this experiment. Most subjects had little experience with VR devices. All had normal or corrected-to-normal visual acuity.

3 Results and Discussion

We measured the error that was the distance between the position where the subject placed the virtual object and the position indicated by the experimenter.

3.1 Results of the Comparison between Gaze (Wink) and Controller

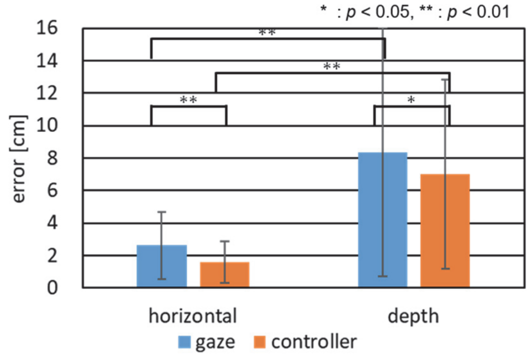
Fig. 3 shows the average and the standard deviation of the data points in which the outliers were excluded. The average error in the gaze condition 9.2 cm was about 1.7 cm larger than that in the controller condition 7.5 cm (*t*-test, $p < 0.01$).

We analyzed the errors separately on the horizontal and depth axes. As shown in Fig. 3(a), the average error on the horizontal axis was approximately 2.6 cm by the gaze and approximately 1.6 cm by the controller, while the average error on the depth axis was approximately 8.3 cm by the gaze and approximately 7.0 cm by the controller. These suggest that there are larger errors on the depth axis and the depth perception is different between in the virtual environment and in the real environment.

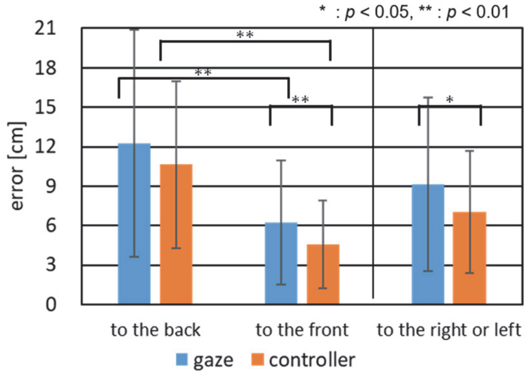
To further analyze the errors, we classified the errors into three groups according to the directions of moving the virtual object. The results are shown in Fig. 3(b). There was a significant difference between moving the virtual object from the front to the back and moving it from the back to the front (*t*-test, $p < 0.01$). As described above, there is a difficulty of the depth perception and this difficulty becomes more when pointing to a distant position. Also, there was a significant difference between the gaze and the controller only for moving the virtual object from the back to the front (*t*-test, $p < 0.01$). This may be because the eyes use more muscle power when looking near than when looking far away. This also suggests that the gaze may be useful as same as the controller in pointing to the distant position. When moving the virtual object from side to side, there was a significant difference between the gaze and the controller. More detailed analyses are desired because this result may be influenced by the dominant eye. We categorized the errors by the subject's dominant eye and compared the means, but there was no significant difference.

3.2 Results of the Comparison between Gaze (Time) and Controller

Fig. 4 shows the average and the box plot for each subject. The average error was approximately 4.6 cm in the gaze condition and approximately 3.9 cm in the controller condition, which means that the gaze condition was about 0.7 cm larger than the controller condition (*t*-test, $p < 0.01$). As shown in Fig. 4, there were two subjects whose average error by the gaze was less than that by the controller.



(a) Horizontal and depth axes



(b) Directions of moving the virtual object

Fig. 3 Results of gaze (wink) and controller

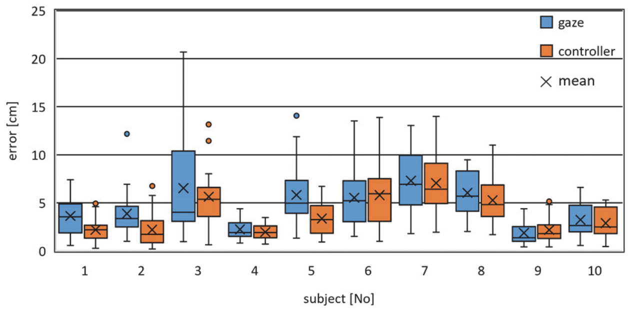


Fig. 4 Results of gaze (time) and controller

We analyzed the errors separately on the horizontal and depth axes and found that there were larger errors on the depth axis both in the gaze and controller conditions. From these results, the accuracy of depth perception needs to be improved in the AR environment. The comparison between the gaze and the controller revealed that on the horizontal axis, the error by the controller was significantly less than that by the gaze (*t*-test, $p < 0.01$). This suggests that the gaze operation on the horizontal axis is more difficult than the controller operation. As for the directions of moving the virtual object, we need more data for further discussion.

3.3 Discussion

Since the comparison between the gaze input with a wink and the controller revealed that the errors on the depth axis and those in the directions of moving the virtual object from the front to the back were larger than the others, the comparison between the input using the gaze time and the controller were focused on the depth direction. The results showed that the input using the gaze time was more accurate than the gaze input with a wink and its average error when moving the tennis ball-sized virtual object was 4.6 cm, only 0.7 cm larger than the controller. As shown in Fig. 4, the differences between the input using the gaze time and the controller were little for most of the subjects. Because the gaze input is still less familiar than the controller, the level of proficiency required for the gaze input needs to be verified.

4 Conclusions

We compared interaction by the gaze input and by the hand-held controller in the AR environment. In the experiments, the three operations were used for moving the virtual object: gaze input with a wink, input using the gaze time, and input using the controller. The results of the experiments showed that the input using the gaze time was more accurate than the gaze input with a wink and almost same as the controller when moving the tennis ball-sized virtual object. From these results, we conclude that the gaze input is useful for moving a virtual object, as a way of hands-free interaction in AR environments.

I would like to thank Ms. Yuki Tanaka for her work on this project.

References

[1] W. Lin, L. Du, C. Harris-Adamson, A. Barr, and D. Rempel, “Design of Hand Gestures for Manipulating Objects in Virtual Reality, ” International Conference on Human-Computer Interaction 2017, Part I, LNCS 10271, pp. 584-592 (2017).

[2] M. Baloup, T. Pietrzak, and G. Casiez, “RayCursor: A 3D Pointing Facilitation Technique based on Raycasting,” 2019 CHI Conference on Human Factors in Computing Systems (CHI '19), Paper No. 101, pp. 1-12 (2019).

[3] Y. Lu, C. Yu, and Y. Shi, “Investigating Bubble Mechanism for Ray-Casting to Improve 3D Target Acquisition in Virtual Reality,” 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 35-43 (2020).

[4] S. Suzuki, H. Suzuki, and Mie Sato, “Grasping a virtual object with a bare hand,” ACM SIGGRAPH 2014 Posters (SIGGRAPH '14), Article 51 (2014).

[5] T. Nukarinen, J. Kangas, J. Rantala, O. Koskinen, and R. Raisamo, “Evaluating ray casting and two gaze-based pointing techniques for object selection in virtual reality,” 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18), Article 86 (2018).

[6] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen, “Gaze + pinch interaction in virtual reality,” 5th Symposium on Spatial User Interaction (SUI '17), pp. 99-108 (2017).

[7] H. J. Kang, J. Shin, and K. Ponto, “A comparative analysis of 3d user interaction: How to move virtual objects in mixed reality,” 2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 275-284 (2020).

[8] C. Liu, A. Plopski, and J. Orlosky, “OrthoGaze: Gaze-based three-dimensional object manipulation using orthogonal planes,” Computers & Graphics, Vol. 89, pp. 1-10 (2020).

[9] C. Liu, J. Orlosky, and A. Plopski, “Eye Gaze-based Object Rotation for Head-mounted Displays,” Symposium on Spatial User Interaction (SUI '20), Article 4, pp.1-9 (2020).