Computers & Graphics 89 (2020) 1–10

Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Technical Section

OrthoGaze: Gaze-based three-dimensional object manipulation using orthogonal planes[☆]

Chang Liu[a,*], Alexander Plopski[b], Jason Orlosky[c]

[a] Graduate School of Information Science and Technology, Osaka University, Japan
[b] Department of Information Science, University of Otago, Dunedin, New Zealand
[c] Cybermedia Center, Osaka University, Japan

ARTICLE INFO

Article history:
Received 17 December 2019
Revised 23 March 2020
Accepted 9 April 2020
Available online 29 April 2020

Keywords:
Human-computer interaction
Eye tracking
Object manipulation
User interface

ABSTRACT

In virtual and augmented reality, gaze-based methods have been explored for decades as effective user interfaces for hands-free interaction. Though several well-known gaze-based methods exist for simple interactions such as selection, no solutions exist for 3D manipulation tasks requiring a higher degree of freedom (DoF). In this paper, we introduce OrthoGaze, a novel user interface that allows users to intuitively manipulate the three-dimensional position of a virtual object using only their eye or head gaze. Our approach makes use of three selectable, orthogonal planes, where each plane not only helps guide the user's gaze in an arbitrary virtual space, but also allows for 2-DoF manipulations of object position. To evaluate our method, we conducted two user studies involving aiming and docking tasks in virtual reality to evaluate the fundamental characteristics of sustained gaze aiming and to determine which type of gaze-based control performs best when combined with OrthoGaze. Results showed that eye gaze was more accurate than head gaze for sustained aiming. Additionally, eye and head gaze-based control for 3D manipulations achieved 78% and 96% performance, respectively, in comparison with a hand-held controller. Subjective results also suggest that gaze-based manipulation can comprehensively cause more fatigue than controller-based. From the experimental results, we expect OrthoGaze to become an effective method for pure hands-free object manipulation in head-mounted displays.

1. Introduction

In recent years, head-mounted displays (HMDs) have enabled the use of virtual and augmented reality (VR and AR). One persisting issue with content in these devices is lack of the ability to author or manipulate content in a hands-free manner.

Over the years, many different manipulation techniques have been developed that enable direct authoring [1], such as gesture-based selection and manipulation [2], the Go-Go technique [3], or avatar representation [4,5] to name a few. Although these have been shown to be efficient, they often require additional hardware and tracking capabilities that may not always be available.

Furthermore, practical use cases exist, e.g. in a crowded public space or small room, where the use of body gestures and peripheral objects could disturb others or occlude surroundings. Research also suggests that performing noticeable movements for interacting with a device may reduce its social acceptability [6]. As such, interaction methods that do not attract extensive attention and allow for discrete manipulation of content are essential.

Handheld devices such as joysticks, mobile phones and tablets can help address this problem [7]. Nevertheless, the requirement for peripheral devices makes interaction difficult when the user's hands are preoccupied or when using an additional device is not an option. Very recently, commercial HMDs such as the Microsoft HoloLens 2, MagicLeap One and HTC Vive Pro-Eye, have begun to include integrated eye tracking, which has great potential for enabling hands-free interaction.

The interest in eye-based interaction is partly due to our tendency to direct our gaze towards objects we are interested in, which makes it a good indicator of the user's intention and focus [8]. Gaze techniques have been widely studied for use in 2D interfaces as a means of estimating the effects of different interfaces on user focus [9], selection of items [10], or to design attentive interfaces that react to the user's gaze [11]. While similar applications have emerged on HMDs as well [12], it is difficult to directly transfer interaction methods from a 2D to a 3D interface due to the higher degrees of freedom. Furthermore, although one can determine the user's gaze point on the screen, it is much more difficult

☆ This article was recommended for publication by H Fu.

* Corresponding author.
E-mail address: liu.chang@lab.ime.cmc.osaka-u.ac.jp (C. Liu).

https://doi.org/10.1016/j.cag.2020.04.005
0097-8493/

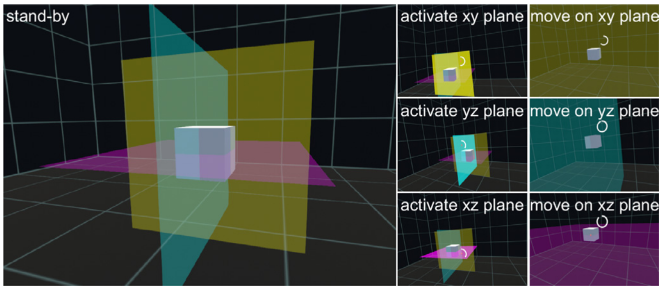


Fig. 1. OrthoGaze enables gaze-based position manipulation of virtual objects in 3D. (left) The user can move the object on the three orthogonal planes displayed around the object. (middle) When the user wants to select a plane, it is highlighted, the other planes are dimmed out, and a dwell timer indicates the selection. (right) After activating a plane, the user can move the object around to adjust its 2-DoF position by looking at the target location and confirming placement through a gaze dwell.

to measure the depth at which the user is focusing on or for users to manipulate the focal depth voluntarily without a reference object [13].

Our goals are to address the need for a hands-free manipulation method that is both discrete and can handle higher DoF operations. In this paper, we present OrthoGaze, a novel interface that allows users to manipulate the 3D position of virtual objects using only eye gaze or head gaze. As shown in Fig. 1, we present users with three semi-transparent orthogonal planes that define three different interactive dimensions. Users can choose a plane on which they want to move the object on and subsequently manipulate the position of the object on that plane using gaze-plane intersection, i.e., the intersection of the user's gaze on the active plane. OrthoGaze is in some ways similar to INSPECT, which allows for 6-DoF control on mobile phones [14]. Our method has the advantage of being hands-free, and does not require external hardware.

In the past, some studies have found that head gaze outperforms eye gaze [15], and others have found contrasting results [16]. Furthermore, these papers focused on selection in 2D tasks and have yet to consider manipulation in 3D. However, we expect that using head gaze for 3D manipulation requires more extensive and accurate head movements than for 2D tasks. Performance can also be affected by constraints of the head's angular motion due to the limitations of neck rotation [17]. As OrthoGaze supports both eye and head gaze, it is necessary to test how well each type of gaze-based control would perform for 3D aiming and manipulation tasks. As such, our first step to evaluate OrthoGaze was to test a user's ability to target different areas on the presented planes to compare how accurately and easily users can adjust the targeting location. Our results show that depending on the targeted plane eye gaze can outperform head gaze and in general was more accurate than head gaze.

Secondly, we evaluated how well users could reposition a virtual object to a target location at different depths. For this purpose, we tested OrthoGaze with eye gaze, head gaze, and raycasting controller for docking tasks as a comprehensive comparison. Though a controller is not hands-free, we chose a controller with raycast and trigger selection as the baseline to determine how well head and eye gaze-based manipulation could match hand-based performance. As expected a controller with raycast performed the best in both qualitative and quantitative measures, but OrthoGaze still enables efficient 3D manipulation for head and eye gaze as well. At the same time, we found contrasting results from those of experiment one. Participants rated eye gaze-based control lower than head gaze-based control, and were more successful with head gaze than eye gaze.

In summary, our contributions in this paper are:

- We present OrthoGaze, a novel approach that enables hands-free adjustment of virtual object position in HMDs.
- We conduct an experiment that evaluates sustained eye gaze and head gaze aiming on planes. The results show that eye gaze outperforms head gaze in terms of accuracy. Furthermore, in some cases larger areas can be covered with eye gaze than head gaze.
- We show that for 3D docking tasks, eye and head gaze-based control with OrthoGaze can achieve 78% and 96% success rates, respectively, when compared to a hand-held controller.

2. Related work

Most of the work related to this research primarily falls into two domains: (a) methods for object position manipulation, and (b) gaze-supported interaction.

2.1. Object position manipulation

As described by Bowman et al. [1] selection and manipulation is a basic element of 3D interaction with a large variety of hardware and software solutions to facilitate it. The most natural interaction method is direct manipulation of virtual objects by picking them up with our hands and manipulating their location. The main limitation of this interaction technique is that it is only applicable within our immediate vicinity. One way to address this limitation is the world-in-miniature technique that presents a miniaturized version of the world in front of the user that replicates any adjustments to virtual objects done on it to their counterpart [18]. Chae et al. [19] applied the same idea in an AR context where they use a wall for supporting the manipulation of distant virtual objects. Another common technique is to use raycasting methods with a depth-manipulation technique, e.g., Go-Go technique [3].

While these techniques require potentially large hand movements, miniature mice [20] and handheld devices [14] can provide discrete manipulation with virtual content. The most basic form is the use of buttons that adjust the location of the virtual object whenever they are activated or entry fields where the user can adjust the position of the object [21]. While this provides the most control, it is time consuming. Instead of buttons, a virtual object can be moved along displacement vectors of a joystick handle [22]. INSPECT [14] extends this idea to mobile devices by combining the orientation tracking and touch-sensitive 2D surface of a handheld device. The orientation of the device defines a virtual plane centered at the object's location, while translation of the user's fingers on the 2D surface is interpreted as displacement of the virtual object on the predefined plane. INSPECT facilitates different control modes where the pivot mode fixes the plane at the original position of the object thus any further rotation results in a displacement of the object, while the free-plane casting mode always places the pivot of the plane at the location of the virtual object, thus any displacement must be initiated through swipe gestures on the handheld device. Piekarski et al. [23] also introduced a plane-based system on mobile devices, enabling 3D modeling and manipulation of distant AR content.

Though existing methods enable high-DoF positioning in the virtual environment, most of them require extensive hardware, e.g. mice or touch panels, which are difficult to apply to hands-free interactions.

2.2. Gaze-supported interaction

When users are interested in an object, they tend to look at that object. This tendency means that eye gaze can be a natural modality for interaction with virtual content. It has commonly been used for selection in 2D and 3D environments [24,25]. Advancement in eye tracking technology has led to a series of studies that compare the performance of these targeting techniques with each other in terms of accuracy and speed. Kyto et al. [26] showed that using only eye-gaze to target and select targets generally performs slower and less accurately than head gaze or a combination of eye gaze with other modalities. These results confirm previous findings by Qian and Teather [15], who found that head-gaze was more reliable than eye-gaze. At the same time, Blattgerste et al. [16] found that eye gaze outperformed head gaze in terms of accuracy, speed, and task load. They also found that this advantage was more dominant in HMDs with a larger field of view. They attributed this to more reliable eye tracking in their evaluation. It is thus still unclear which method will perform better in the long run.

After a target is selected with either eye gaze or head gaze, other methods allow for manipulation of the object through hand gestures [25,27,28] or other controllers. While few 2D interfaces use eye gaze for selection and manipulation [29], to our knowledge no technique exists that allows object manipulation in 3D using only eye-gaze. This can be traced back to a variety of reasons. Compared to interaction with 2D interfaces, AR and VR present additional challenges for object selection and manipulation, for example handling occluded objects or those in the same line of sight. The focus depth can be derived either from the vergence of the user's gaze [13], by selecting the object of interest from a list [30], or combining gaze with other modalities [31]. Although the above methods can disambiguate between objects at different depths, the estimation is too coarse to accurately manipulate the object depth. Furthermore, it is difficult to manipulate the depth of an object after it was selected.

2.3. Our motivation

Unlike other studies that use gaze either as a confirmation method or to support hand interactions, we wanted to come up with something that could be carried out in an entirely hands-free manner. The goal for our study was to test our design to see how well it could help users intuitively manipulate a virtual object. Though OrthoGaze can also potentially be adapted for rotation and scaling, our study thoroughly examines the fundamental properties of our method and its usability for translation tasks. Furthermore, since previous studies found contradicting results on the efficiency of eye and head gaze while focusing on 2D selection scenarios such as menus, it is also necessary to study how these techniques perform in 3D scenarios where users have to interact with objects at different depths.

3. Methodology

In this section, we first discuss the process behind the development of a gaze-only manipulation method. Because of constraints of the human visual system, the design of Orthogaze is fundamentally different from typical 2D or touch-interface methods. We then describe the design and use of OrthoGaze and why this works well for manipulating virtual objects. Note that all units discussed in this section are relative to the world coordinate system.

3.1. Constraints for gaze-based object manipulation

When testing and brainstorming different methods for gaze-based object manipulation, we identified two primary constraints that needed to be accounted for. The first is related to one of the inherent characteristics of gaze fixations. Johansson et al. [24] showed that gaze plays an important role in leading movement during manipulation tasks. While a user is performing a sustained manipulation of an object, as a natural response, his/her gaze will tend to be fixated on the object as it travels to ensure that the manipulation is carried out correctly. Thus, an interface should cause the smallest possible eye gaze offset from the target object during the manipulation process. In other words, the method benefits greatly from synchronously carrying out the manipulation with the user's gaze, rather than with a clutch or secondary mechanism.

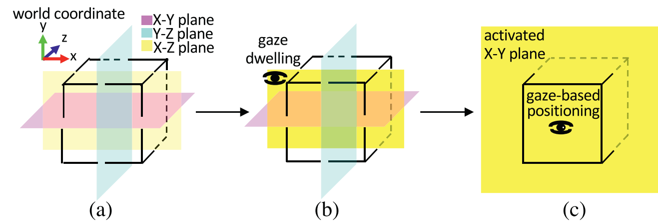


Fig. 2. Images showing an example of the baseline design of OrthoGaze. (a) Three orthogonal planes with different colors intersecting at the geometric center of the object, waiting for activation. (b) Highlighted X-Y plane by user's gaze dwelling. (c) Activated X-Y plane where user can position the object at a 2-DoF level.

Another main constraint is that although it is possible to estimate the approximate gaze depth through eye tracking [13,31], consciously and precisely adjusting one's gaze to an arbitrary depth without guidance is very difficult. This highlights the necessity of providing a clear method to guide the user's eyes, especially to a particular depth in 3D space.

3.2. Orthogonal plane design

OrthoGaze is composed of three orthogonally intersecting square planes, as shown in Fig. 2. Each plane indicates a 2-DoF space for manipulation in virtual world coordinate space. For instance, the yellow plane in Fig. 2 denotes the world xy plane. This is a per-object design, where the three planes are attached to each object at its geometric center.

The planes fulfill two roles during the manipulation. First, as noted in Section 3.1, it is hard for an individual to focus his or her gaze at a particular point in 3D space without a salient feature or object to focus on. As such, the planes provide a surface onto which the user can focus to adjust the position of an object. Second, the plane constrains the object movement to within the plane even if other objects are in the user's view. The three orthogonally intersecting planes allow users to manipulate all 3 positional DoFs where each DoF can be manipulated by two planes. This design also ensures that users always have access to at least one plane even at a poor viewing angle, mainly when it is perpendicular to the user's gaze. Additionally, to maintain general visibility and accessibility, the size of all planes s_{pln} is scaled linearly based on the distance to the object d_{obj} as:

$$s_{\mathrm{pln}} = s_{\mathrm{bsc}} \times (1 + p \times d_{\mathrm{obj}}) \tag{1}$$

where s_{bsc} denotes the size when $d_{\mathrm{obj}} = 0$ and p is a constant scaling factor.

3.3. Manipulation mechanisms

OrthoGaze assists object manipulation through two main functions:

Plane activation. In the default state all planes appear semi-transparent, indicating that the user can trigger an interaction by looking at them. When the gaze ray intersects with a plane, that plane is highlighted to indicate to the user the detected selection. If the user's gaze remains on the plane throughout the dwell period, the system switches into object manipulation on this plane.

This trigger is based on whether or not the gaze ray is intersecting with a given plane. Therefore, as long as the gaze point stays on that plane, the user can still activate the manipulation even if the gaze is jittering or if the estimation is slightly inaccurate.

2-DoF Manipulation. Once a plane is activated, its size expands to better represent the positionable area, and the remaining planes temporarily become transparent to allow seamless manipulation with the selected plane. The object will then follow the intersection point of the user's gaze ray and the plane. This ensures that the object will always be at the location the gaze is focused on.

Placement. Finally, placement at the destination is triggered by another gaze dwell. After this placement is completed, the system switches back into its default state that shows all planes as semi-transparent.

To trigger the activation and selection we detect gaze dwell as follows. In eye gaze mode, gaze dwell at time t_0 is calculated as an angular deviation of eye gaze OE over a time period n:

$$OE_{t_0} = \frac{1}{n} \sum_{t=t_0-\mathrm{n}}^{t_0} \| \arccos(\hat{eye}_t \cdot \hat{head}_t) - \arccos(\hat{eye}_{t-1} \cdot \hat{head}_{t-1}) \| \tag{2}$$

where $\hat{eye}$ is a unit vector of eye gaze, $\hat{head}$ is a unit vector of head gaze and $\cdot$ is an operator of the inner product between vectors. We assume the eye gaze to be fixated at a location if OE is less than a threshold ts. Through initial tests, we found it more robust to detect natural eye gaze dwell using such angular deviation rather than the exact gaze point.

In the head gaze mode, we detect a dwell if the head gaze deviation $OH \leq$ ts, where

$$OH_{t_0} = \frac{1}{n} \sum_{t=t_0-\mathrm{n}}^{t_0} \| \arccos(\hat{head}_t \cdot \hat{head}_{t-1}) \|. \tag{3}$$

Though a number of different selection methods are compatible with OrthoGaze, we chose to use gaze dwell to avoid the Midas touch problem [32] because it is easy to understand and prevalent in research. This decision was also made as a trade off with accuracy to reduce necessary eye movements and maintain intuitiveness.

4. User study

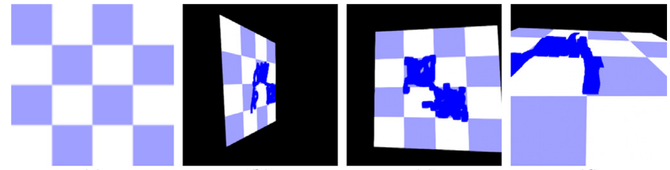
Through the user study, we wanted to investigate the efficiency and the effectiveness of OrthoGaze. In addition, we wanted to test how head, eye, and controller based manipulations would perform using this interaction paradigm. To test this, we implemented two different tasks, painting (A) and docking (B) for simplicity, to evaluate OrthoGaze both fundamentally and practically. When participants entered the experiment room they first received an introduction into the experiment tasks and an explanation of the different control modes of OrthoGaze. After signing a consent they first completed task A followed by task B. During the experiment participants remained seated on a swivel chair and were asked not to stand up or move around. However, local body movements were not physically restricted, and participants could rotate their chair if necessary, which is natural when performing interactions in VR. Participants could take a break between each trial if needed. Participants had a training session before each task where they could practice all of the designated methods. Overall, the experiment took about 1 h. The procedure of the experiment was approved by the institutional review board of Osaka University Review Board.

4.1. Hardware and participants

For the evaluation, we used an HTC Vive Pro-Eye as the HMD, which has integrated eye tracking cameras and provides relatively stable eye gaze data. We set up a virtual environment with the experiment tasks using Unity 2018.3.2f1. The whole system was run on a desktop computer with an Intel Xeon E52690 CPU and an NVIDIA GeForce GTX 970 GPU, at an average frame rate of 60 frames per second. We used R 3.6.2 [33] and coin 1.3-1 [34] for the statistical evaluation.

We recruited 18 students and researchers from the local university, 15 male and 3 female, ranging in age from 21 to 38 (avg. 25.56, stdev. 4.22). 9 of them wore prescription glasses during the experiment. 10 of them had no experience in eye-based human-computer interaction, while the remaining 8 had some experience (less than 5 times in total) before this experiment. All participants received a gift card worth approximately 5 USD as remuneration.

Note that for all eye gaze methods, we did not restrict the participant's neck, which means the participants were allowed to perform natural eye movements supported by head movements.



(a) (b) (c) (d)

Fig. 3. Images showing the content of task A. (a) The chessboard plane used in the experiment. Participants were asked to paint the purple blocks as fully as possible and avoid spilling onto the white blocks. Sample frames of a participants' views during the task are shown and represent the (b) Y-Z plane, (c) X-Y plane, and (d) X-Z plane. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.2. Task A: gaze-based painting

As described in Section 3, OrthoGaze presents multiple planes in the world coordinate system and allows users to translate objects within each plane for a continuous period of time. In other words, the accuracy and speed with which users can sustain gaze action can significantly affect performance. As such we designed a task to compare aiming with head versus eye gaze.

4.2.1. Hypotheses

Some evidence supports the hypothesis that head gaze has better performance than eye gaze for discrete selection tasks in VR environments [15,26]. In contrast to discrete tasks, our method requires continuous gaze movements in a large, dynamic range of distances, i.e. 3D manipulations. We therefore first aim to compare eye and head gaze under continuous targeting. Following the insights from previous work that examined gaze performance for discrete tasks and the expected higher stability of head gaze compared to eye gaze, we hypothesized that:

- **Ha1**. Head gaze will allow for faster performance than eye gaze for distant, continuous aiming tasks.
- **Ha2**. Head gaze will be more accurate than eye gaze for distant, continuous aiming tasks.

4.2.2. Task

The experiment was conducted in a virtual environment with a solid black background. To evaluate the accuracy and the speed of sustained eye gaze and head gaze behaviours, participants had to aim at indicated areas of a target plane while avoiding other areas of the target. We used a 4×4 chessboard pattern with 2 colors as the target plane as shown in Fig. 3(a). The checkerboard appeared at 3 different positions in the world coordinate system for each gaze method, and each position represented a different 2-DoF plane. Detailed specifications of the environmental setup are

Table 1
Environment specifications of task A.

brush size	plane size	position (x, y, z)	rotation (x, y, z)
4×4	120×120	left $(-100, 1, 170)$ front (0,1,170) ground $(0, -21, 60)$	left $(-90, 0, 0)$ front $(0, 0, -90)$ ground (0,0,0)

* Each plane is pivoted on its geometric center.
* Participants' viewing point (head position) is located at (0,1,0).

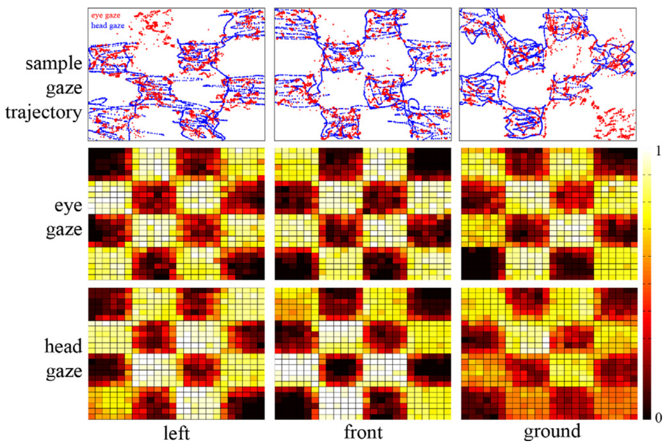


Fig. 4. (top) Sample gaze points from a participant, projected onto 2D planes. (middle and bottom) Images showing heat maps of gaze point coverage results of task A over all participants. Note: to represent frequency of gaze among all participants, each segment containing gaze points is counted only once for each participant regardless of the total gaze points for that participant. The brighter a region, the higher the gaze frequency.

shown in Table 1. Note that the setup of the planes was selected empirically to cover the specific fields of view of the HMD. This avoided severe view point shifts caused by view point changes. For instance if the left plane were placed on the participant's left flank, it might appear the same as the front plane if he or she turned left.

For a single trial, participants were required to use their head gaze or eye gaze to aim at and cover as much as possible of the purple region on the plane within 60 [s], while avoiding aiming at the white region. The intersection point of the gaze and the plane functioned as a square painting brush in blue. Participants could not deactivate the painting during the trial, which means participants would be painting whenever their eye gaze or head gaze intersects with the plane. We conducted the experiment as a 2x3 within-subject study with 2 gaze methods (eye and head gaze) and 3 plane orientations (left, front, and ground) which resulted in 6 trials for each participant. The order of the trials between participants was randomized. Before each trial participants saw an outline of the plane location in the next trial and were informed which gaze technique will be used. The experimenter toggled the next trial at which point the plane became visible and users could paint the plane.

4.2.3. Results

We collected participants' eye gaze point as well as the head gaze point for each frame when the gaze intersected with the target plane. From the collected data we computed trajectories consisting of both eye gaze and head gaze points, as shown in the top row of Fig. 4. For quantitative evaluations, we segmented each plane into 24×24 regions, and evaluated the speed and the accuracy of each gaze method by calculating the cover rate and the ratio of gaze points falling onto target areas versus non-target areas. An Anderson-Darling test showed that our data was not normally distributed, thus we used the Friedman's test to analyze variance when necessary, and the Wilcoxon signed rank test for post-hoc tests. We used a threshold of p=0.05 to determine statistical significance. We report $r = Z/\sqrt{N}$ as the effect size for all (post-hoc) test results, where Z is the statistical value and N is the total sample size [35].

For each participant, the cover rate is given by the ratio of covered correct area compared to the overall target area. In other words, this represents how much of the purple area participants painted blue. As all participants had the same amount of time to cover as much of the target area as possible, the cover ratio represents the speed with which participants can aim and adjust their gaze in a continuous task. As described above, the plane was segmented into 24×24 regions. If at least one gaze point fell into a region we counted it as covered for this participant. The middle and bottom rows in Fig. 4 show results of cover rates of each plane condition for all participants visualized as frequency heat maps. More detailed results of the gaze cover rate are shown in Fig. 5(a). Although we did not find statistically significant difference for cover rate when participants aimed on left ($Z = 0.762, r = 0.13, p > 0.05$) and front plane ($Z = 0.305, r = 0.05, p > 0.05$), participants covered significantly less area with head gaze than eye gaze for the ground plane ($Z = 2.308, r = 0.38, p < 0.05$). In general, no significant difference was found when comparing the cover rate of both gaze methods, regardless of the aimed plane ($Z = 1.270, r = 0.12, p > 0.05$).

We define the accuracy of the gazing as the ratio of gaze samples that fall into the correct areas over all gaze samples for a participant (Fig. 5(b)). There were significant differences between eye gaze and head gaze for the left ($Z = 2.722, r = 0.45, p < 0.01$), front ($Z = 2.896, r = 0.48, p < 0.01$) and ground ($Z = 3.070, r = 0.51, p < 0.01$) plane. In all cases participants were more accurate when using eye gaze than head gaze, which is also supported by the data aggregated by method ($Z = 4.886, r = 0.47, p < 0.001$).

For the results regardless of the gaze method (Fig. 5(c) and (d)), a Friedman's test showed a significant difference between each plane both in the cover rate ($\chi^2(2) = 20.72, p < 0.001$) and the correct rate ($\chi^2(2) = 8.39, p < 0.05$). For the cover rate, a post-hoc Wilcoxon signed rank test with Bonferroni correction showed a significant difference between left and ground plane ($Z = 3.456, r = 0.41, p < 0.001$), and between front and ground plane ($Z = 4.313, r = 0.51, p < 0.001$). For the correct rate, we found a significant difference between left and front plane ($Z = 2.671, r = 0.31, p < 0.05$), and between front and ground plane ($Z = 2.765, r = 0.33, p < 0.05$).

We also asked participants to rate helpfulness, ease of use, and fatigue, but found no statistically significant results.

4.2.4. Discussion

Our results reject our hypotheses **Ha1** and **Ha2**. This result is different from previous findings that showed that head gaze outperforms eye gaze for selecting near-field targets [15]. One potential reason for this is that the longer distance between the target and the user could have led to higher difficulty of head gaze aiming compared to eye gaze, as users had to utilize more muscles. When aiming on targets that are far away, finer control is needed for head gaze since even small angular movement of the head can cause a huge offset projected in the distance, while eye gaze can remain robust in aiming accuracy since users aim with their eye gaze by directly looking at the target position. Additionally, aiming was significantly less accurate on the left and ground planes, as indicated by Fig. 5(d), which could have resulted from the reduced perspective since the front plane was parallel to participants' view and thus had better visibility compared to the other two. We also observed that participants had more difficulty aiming on the ground plane than the other planes, as highlighted by the middle and lower right of Figs. 4 and 5(c) and (d). We believe that this is because it was hard to rotate the neck to cover all areas of

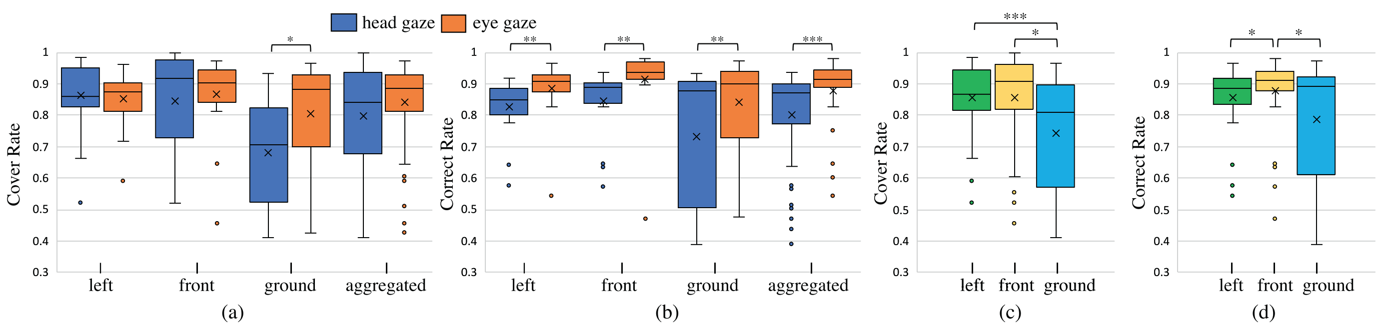


Fig. 5. Boxplots showing the quantitative results of task A. (a) The segmental cover rate of each plane, along with data aggregated by method. This is calculated based on a 24×24 segmentation of the target planes. (b) The correct rate for aiming on each plane, along with data aggregated by method. This is the rate of gaze points that fell into correct segments out of all collected points. (c) The cover rate aggregated by plane. (d) The correct rate aggregated by plane. (***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$).

the plane, as the ground plane was set to occupy the area right at the participant's feet.

4.3. Task B: three-dimensional docking

While task A evaluates the very basic performance of gaze-based methods interacting with orthogonal planes, task B evaluates the actual usability of OrthoGaze for manipulating the 3D position of virtual objects. As OrthoGaze is intended for gaze-based interaction, it generally has good compatibility for raycast-based interaction. We thus decided to compare the performance of OrthoGaze with eye gaze, head gaze and a controller. While we use dwell timing to trigger the different modes for eye gaze and head gaze, we use raycast aiming and button clicks in the controller condition. In that sense, the controller condition serves as a best-case benchmark.

4.3.1. Hypotheses

From our observations in Task A, we expected that participants could utilize OrthoGaze better with eye gaze than head gaze as it was more accurate and faster. Furthermore, as the controller condition did not suffer from the constraint of the dwell time and possible unintentional activation we expected it to outperform other control conditions. Overall we had the following hypotheses:

- **Hb1**. Participants will successfully complete the docking task more often and more quickly when using the hand-held controller than head and eye gaze.
- **Hb2**. Participants will perform the docking task faster with eye gaze than head gaze.

4.3.2. Task

For task B, participants were located at $(0[\mathrm{m}], 1[\mathrm{m}], 0[\mathrm{m}])$ and had to move a white cube, sized $0.5[\mathrm{m}] \times 0.5[\mathrm{m}] \times 0.5[\mathrm{m}]$, from a fixed start position $(-1[\mathrm{m}], 0.5[\mathrm{m}], 5.5[\mathrm{m}])$ to several target positions using all three control methods described above with OrthoGaze. During each trial, a green cube with the same size as the white cube appeared at one of the target locations and participants had to align the white and the green cubes (Fig. 6(b)). The target positions were corners of an imaginary cube with a side size of $2N$[m] whose center coincided with the center of the white cube as shown in Fig. 6(a). We used 8 imaginary cube sizes with $N \in \{0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4\}$. To keep the number of alignments reasonable each offset direction was selected twice and was paired with a different distance in each appearance thus ensuring that each offset distance and direction appeared twice during the experiment, resulting in overall 16 different target positions.

Since it is difficult to perfectly align the white and green cubes, we confirm a successful alignment if the two cubes are less than

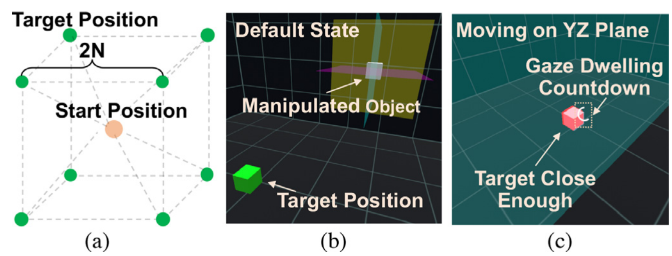


Fig. 6. Images showing the content of the docking task. (a) 8 target positions distributed from the start position equally in world x, y and z axes. Note that we used 16 target positions in total, paired with 8 different distances from the start position. (b) A sample of a start condition of task B. (c) Moving the cube close enough causes a change in the target cube from green to red. Not only moving the cube to the target position but also placing it successfully within 30 [s] counts as a successful trial. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

0.2m apart when the user confirms the placement. As an additional cue we turn the target cube from semitransparent green to red when the two cubes are within the threshold distance of each other (Fig. 6(c)). Note that to succeed in a trial, the participant not only had to move the cube to the target position, but also had to perform the placement successfully. If the trial was a success, the participant received a sound effect as confirmation.

Before starting the experiment, participants had 10 trials for practicing with each control method. During the experiment, participants initially saw an empty room. Each trial started when the white cube and the target location appeared in front of the participant and finished after successful alignment. If participants did not align the cubes within 30 [s] the trial was counted as failed.

For utilizing OrthoGaze with gaze-based methods, we set the dwell time n to 1.3 [s] and the constant threshold ts for the angular gaze offset to 0.005 [rads] for both plane activation and object placement. When using the controller, participants used a raycast to aim on planes and pressed the trigger button for activation and placement. This sets the controller as the standard of utilizing OrthoGaze with least time loss, to which we could also compare the performance of gaze-based methods. We conducted this experiment as a 3x1 within-subjects experiment with the control method as independent variable. We counter-balanced the order of the conditions for all participants using a Latin square. After all trials, participants completed a custom survey related to the task and their experiences.

For evaluating the experimental results quantitatively, we recorded and calculated the following metrics:

- **Success rate**: The success rate is calculated for each participant as the rate of successful trials out of all trials. This evaluates the general efficiency of manipulating objects with OrthoGaze,

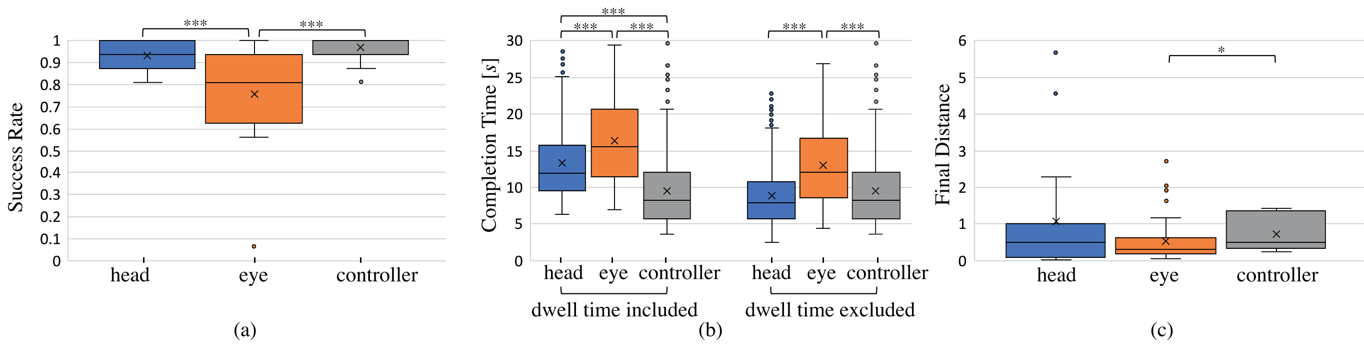


Fig. 7. Box plots showing the quantitative results of Task B. (a) Success rate of the docking task (1 = 100% success) for each control method. (b) Completion time of successful trials for each control method, with the dwell time included (left) and excluded (right) for gaze-based methods. (c) Final distance between the moving cube and the target of failed trials, where each distance is normalized based on the initial distance of that trial. (***: $p < .001$, *: $p < .05$) .

since both accuracy and speed are comprehensively required to successfully complete a trial.

- **Completion time**: Completion time is recorded for each successful trial, and is ignored for failed trials. Note that for gaze-based methods, we recorded the completion time with the time of performing gaze dwell both included and excluded.
- **Final distance**: Final distance is recorded only if the participant fails a trial, normalized as the final distance divided by the initial distance. This normalization stands for how close/far the participant managed to move the object to the target with regards to its initial position.

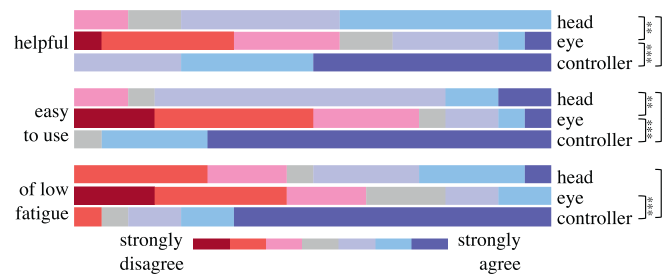


Fig. 8. A 7-point Likert scale chart showing the subjective results regarding the user experience of each control method in task B. (***: $p < .001$, **: $p < .01$).

4.3.3. Results

For the success rate (Fig. 7(a)), both the head gaze mode (avg. 0.934, stdev. 0.067) and the controller mode (avg. 0.969, stdev. 0.052) reached over 0.9 average success rate, while the eye gaze mode (avg. 0.757, stdev. 0.224) reached over 0.7 with a relatively high deviation. An Anderson-Darling test showed that the success rate was not normally distributed, thus we evaluated it with a Friedman's test that showed statistical significance between the different modes ($\chi^2(2) = 21.73, p < 0.001$). A post-hoc Wilcoxon signed rank test with Bonferroni correction showed significant differences between head gaze and eye gaze mode ($Z = 3.271, r = 0.55, p < 0.001$), and between eye gaze and controller mode ($Z = 3.424, r = 0.57, p < 0.001$). This result shows that the controller significantly outperforms eye gaze, which partially supports our hypothesis **Hb1**, and head gaze also significantly outperforms eye gaze, which rejects our hypothesis **Hb2**.

For the completion time including dwell time (Fig. 7(b)), eye gaze mode took the longest (avg. 16.368 [s], stdev. 5.816) over head gaze mode (avg. 13.351 [s], stdev. 5.097) and controller mode (avg. 9.526 [s], stdev. 4.829). Anderson-Darling test showed that the success time was not normally distributed. Since the data include different sample sizes, we used a Kruskal-Wallis test which showed statistical significance between the different modes ($\chi^2(2) = 193.47, p < 0.001$). A post-hoc Mann–Whitney U test with Bonferroni correction showed significant differences between head gaze and eye gaze mode ($Z = 6.065, r = 0.27, p < 0.001$), head gaze and controller mode ($Z = 9.678, r = 0.41, p < 0.001$), and eye gaze and controller mode ($Z = 12.689, r = 0.60, p < 0.001$). We also calculated the reciprocal of completion time as an evaluation of completion speed. On average, head gaze achieved a 71% and eye gaze achieved 58% performance of completion speed in comparison to the controller mode. Interestingly, with the dwell time excluded ($\chi^2(2) = 91.91, p < 0.001$), we only found significant difference between head gaze and eye gaze mode ($Z = 8.929, r = 0.40, p < 0.001$), and eye gaze and controller mode ($Z = 7.855, r = 0.35, p < 0.001$). In this case, head gaze achieved a 107% and eye gaze achieved 74% performance of completion speed compared to the controller.

Fig. 7 (c) shows the final distance between the moved position and the target object of the last frame for all failed trials of each method. A Kruskal–Wallis test showed statistical significance between the different modes ($\chi^2(2) = 11.37, p < 0.01$). A post-hoc Mann–Whitney U test with Bonferroni correction showed significant differences only between the eye gaze and the controller modes ($Z = 2.816, r = 0.23, p < 0.05$).

We show the results of the subjective questionnaire in Fig. 8. A Friedman's test revealed significant differences in helpfulness ($\chi^2(2) = 21.34, p < 0.001$), easiness ($\chi^2(2) = 20.38, p < 0.001$), and fatigue ($\chi^2(2) = 22.53, p < 0.001$). A post-hoc Wilcoxon signed rank test with Bonferroni correction showed that for helpfulness significant differences were found between head gaze and eye gaze ($Z = 2.887, r = 0.48, p < 0.01$), eye gaze and controller ($Z = 3.517, r = 0.59, p < 0.001$), and head gaze and controller modes ($Z = 2.962, r = 0.49, p < 0.01$). For the easiness, we found significant differences between head gaze and eye gaze ($Z = 3.315, r = 0.55, p < 0.01$), eye gaze and controller ($Z = 3.521, r = 0.59, p < 0.001$), and head gaze and controller modes ($Z = 3.358, r = 0.56, p < 0.01$). In regards to fatigue, we found significant differences between eye gaze and controller ($Z = 3.422, r = 0.57, p < 0.001$), and head gaze and controller modes ($Z = 3.316, r = 0.55, p < 0.001$).

As a general result for each item of the survey, the controller mode was significantly considered of highest subjective scores, which also supports our hypothesis **Hb1**, followed by the head gaze mode, while the eye gaze mode had the lowest.

We then asked participants two questions about their experience with OrthoGaze, in particular if it was helpful and if it's visually distracting, along with freeform feedback. Fig. 9 shows the answers. 16 out of 18 participants thought that OrthoGaze helped them complete the task and 17 participants did not think that OrthoGaze was distracting.

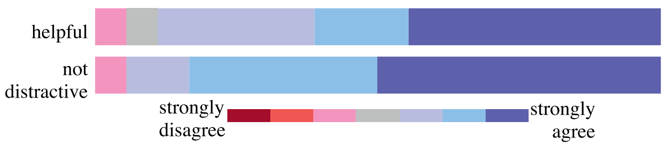


Fig. 9. A 7-point Likert scale chart of the subjective evaluation regarding the usability of OrthoGaze.

4.3.4. Discussion

The results of our second experiment support hypothesis **Hb1**. We believe that this could be in part due to the different confirmation mode utilized in controller mode as participants noted that it was more difficult to keep their gaze fixated on the target for a long time. Note that the threshold we selected to confirm a selection was a very conservative value since we wanted to avoid unintentional triggering and a shorter threshold could make dwell timing more comfortable. The overall positive rating of OrthoGaze being helpful suggests that the improvement in its confirmation mode could potentially further improve the performance with both gaze-based methods.

Interestingly, our results rejected hypothesis **Hb2**. From the participants' comments and the feedback of the questionnaire, we believe that the main limitation in confirming a selection was the gaze dwell. Another explanation could be that each task was rather quick and did not require continuous movement within the plane, but short movements to a new location before confirming the selection. From Fig. 7(c), we can see that for the failed trials, eye gaze-based control had the closest average distance from target. This result reflects the situations in which the participant could correctly move the cube to the target position but could not manage to place it, which was also reported as oral feedback. This supports some previous results that eye gaze is less stable for short-term actions such as selections compared to other methods [15]. One conceivable approach to solve this issue is to use a combination of eye and head gaze, e.g. using eye gaze assisted by head gaze for moving and pure head gaze dwelling for placement.

The subjective results shown in Fig. 8 suggest that using gaze-based control for accomplishing manipulation tasks over a sustained period of time can cause higher fatigue than controller-based control. Subjective feedback indicated that this higher fatigue of using gaze for long-term tasks likely comes from the requirement for higher concentration when consciously controlling the gaze. We also received verbal comments regarding usability such as "*It was difficult to perform the eye gaze dwell*", "*The eye gaze was difficult to fixate, but it's convenient for moving the object*", "*The gaze dwell time was long/short*", and "*The unconscious attempt for the next action sometimes caused an eye movement and thus reset the gaze dwell timer*". It is necessary to further investigate the tasks for which head and eye gaze perform best and how to best combine these gaze techniques to achieve better performance.

5. Implication, limitation and future work

In this section, we discuss in detail the findings, challenges and limitations regarding OrthoGaze and our experimental designs, conceivable future improvements, and directions for further research.

5.1. Implication

The results of task A suggest that although participants can cover areas far away from the user using head gaze, it comes at the cost of lower accuracy as the gaze is more susceptible to minute head movements. At the same time targeting areas at the user's feet could result in lower speed due to a restriction of the neck's movement range. Similarly to head gaze we observed a decrease in the performance of the eye gaze on the ground plane compared to the other planes, although this decrease was less prominent than for the head gaze. For the ground plane the larger movement range of the eyes in combination with head movement allowed users to cover more of the target plane. These results suggest that we need to pay careful attention to the area users will interact in with the virtual content and the required accuracy.

For task B, we observed that some participants could move the cube to the target position, but still failed the trial because they had difficulty gaze dwelling for the final placement. We also received some oral comments from the participants saying that the dwell time was too long or short, which indicates the existence of individual differences on the preference of gaze dwell time that can potentially affect the user experience. The placement method is essential to the performance of manipulation tasks, and the use of gaze dwell likely reduced overall performance and subjective user experience such as fatigue. The completion time excluding gaze dwell time also indicates that the performance of OrthoGaze may have appeared sub-optimal in experiment results and could likely benefit from fine tuning with the method of selection. As a next step, we would test the usability and the efficiency of different windows of gaze dwell time for both activation and placement through user study for developing a method that can achieve more stable placement through eye gaze and suit individual preference for gaze dwell time. We would also like to explore other selection methods that can improve the usability of manipulation using eye gaze.

The results of task B showed that OrthoGaze was able to facilitate eye and head gaze-based manipulations with 78% and 96% success rates in comparison to a hand-held controller, respectively, though there still existed latent improvement in its baseline design. As participants of our experiment were relatively unfamiliar with eye gaze-based user interface, it is also expected that OrthoGaze has the potential to perform better if users are more practiced.

In short, from the results of our user study, we expect OrthoGaze to become an effective method to handle pure hands-free object manipulations. For individuals with handicaps or for users whose hands are constantly occupied with other work, OrthoGaze can provide an effective way to conduct 3D manipulations moving forward.

5.2. Limitations

One critical factor that can affect the performance of OrthoGaze is the viewing angle of the planes. In the docking task, participants were asked to stay seated to exclude the effects of large body movements. However, this constraint sometimes reduced the performance of the gaze-based methods compared to the controller in an unfair way. In situations where the participants had small viewing angles, a few participants managed to finish such trials skillfully by first adjusting on another plane to acquire larger viewing angle, while most participants attempted to move their body to physically change the viewpoint but still failed the trails since they had to stay seated. However, they could still manage the aiming using the controller by reaching out their arms to extend the incident angle of the raycasting. Thus we hypothesize that OrthoGaze can achieve higher usability if the user is able to change his/her viewpoint, or if the viewing angles of the planes are adjustable based on the user perspective, which needs to be validated through further user study.

In addition, in the docking task we only visualized OrthoGaze as a set of semi-transparent planes. We set this in order to test OrthoGaze's naive design. However, the different texturing of the planes in the two tasks could also have affected performance. For example, the grid-textured planes could potentially help with guiding the user's eyes via certain distance cues and hence improve the aiming accuracy of eye gaze. We also observed that in some cases participants had difficulty in activating the plane. This was due to the fact that the complete visualization of all planes might occlude each other and thus prevent the user from properly interacting with the planes. This would also make it difficult to start the manipulation at the original position, and thus make it hard to perform slight refinements of the position. Such limitations can also be prevented by optimizing the visualization, for example by adaptively visualizing the optimal interactive area based on the user's view point or projecting a shadow copy of the moved object for reference. Moreover, a two-step confirmation could also help address this issue, for example defining an area on the plane further away from the object in which the user needs to focus his or her gaze and then adjusting the gaze back to the object for completing an activation. As a next step, we plan to refine these designs and test the effects of different plane visualizations on performance.

The design of our user study for OrthoGaze also has several potential improvements. Since task B was mainly intended for testing the usability of OrthoGaze, we decided to emphasize the evaluation in a task-based manner, i.e., we set a threshold of within 0.2 [m] as successful docking and evaluated the success rate. Though we evaluated the normalized final distance, this evaluation is difficult to make completely fair as it only counts for failed trials and it is based on the assumption that the difficulty of all trials are equivalent. Therefore, the study lacks evaluation of precise accuracy. In addition, we conducted the docking task in a widely open VR environment where little visual occlusions could occur. Considering the visualization of OrthoGaze, the complexity of the environment also could impact the performance, as the planes of OrthoGaze could occlude other objects in the environment and thus affect the user experience. A complex background could affect the visibility of the planes as well. In short, further study needs to be done regarding the accuracy control of OrthoGaze, as well as visual effects in more complex and practical environments.

5.3. Future work

We would like to extend OrthoGaze to hands-free manipulation of even higher DoFs, including but not limited to continuous rotation and scaling, which will be meaningful to achieving hands-free object modeling for large VR and AR environments. However, such functions could also be accompanied with the issue of lost gaze focus. As stated in Section 3.1, if the feedback of the manipulation is not appropriately synchronized with gaze movement, it may result in confusion with the correct recognition of the manipulation result. This is a big challenge for the visualization, proper function, and intuitiveness of manipulations such as rotations.

Additionally, OrthoGaze can also be applied to optical see-through HMDs to achieve hands-free manipulation in AR. In some outdoor AR use cases, such as on a crowded train or in a theatre, it is usually not preferred to use gestures or voices as interaction tools, which meets the requirement for hands-free interaction. To apply OrthoGaze to optical see-through AR, visualization of the planes will likely need to be redesigned in comparison with the VR use cases since it is important to preserve the visibility of the real world.

6. Conclusion

In this paper, we introduce OrthoGaze, a novel approach that allows users to manipulate the 3D position of virtual objects in a virtual environment using eye or head gaze alone. The method is composed of three orthogonal planes, which are affixed to the geometric center of a target object during manipulation. Users can activate each plane using their eye or head gaze, and then move the object on the activated plane by matching the intersection of their gaze with the destination location on that plane. OrthoGaze can be applied not only with eye gaze and head gaze, but also with joysticks and other gesture raycasting methods.

Results of a user study showed that for aiming tasks in VR over a sustained period of time, eye gaze can outperform head gaze for accuracy, especially for distant targets. Results also suggest that OrthoGaze works well for hands-free 3D manipulation. Compared to manipulation with a hand-held controller, eye gaze-based control was able to achieve approximately 78% performance and head gaze-based control achieved 96%. Additionally, subjective results suggest that using both head and eye gaze for sustained 3D manipulation tasks can comprehensively cause more fatigue than a controller. We hope that this method will promote new research on eye gaze manipulation, the development of efficient rotation and scaling functions, and extensions to optical see-through HMDs.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Chang Liu: Conceptualization, Methodology, Software, Investigation, Writing - original draft, Writing - review & editing, Visualization. **Alexander Plopski:** Writing - original draft, Writing - review & editing, Supervision. **Jason Orlosky:** Investigation, Writing - original draft, Writing - review & editing, Supervision.

Acknowledgments

This work was funded in part by the United States Department of the Navy, Office of Naval Research, Grant #N62909–18–1–2036.

References

[1] Bowman D, Kruijff E, LaViola Jr JJ, Poupyrev IP. 3D user interfaces: theory and practice. Addison-Wesley; 2004.
[2] Van den Bergh M, Van Gool L. Combining RGB and TOF cameras for real-time 3d hand gesture interaction. In: Proceedings of the IEEE workshop on applications of computer vision (WACV); 2011. p. 66–72.
[3] Poupyrev I, Billinghurst M, Weghorst S, Ichikawa T. The Go-Go interaction technique: non-linear mapping for direct manipulation in VR. In: Proceedings of the ACM symposium on user interface software and technology; 1996. p. 79–80.
[4] Slater M, Perez-Marcos D, Ehrsson H, Sanchez-Vives M. Towards a digital body: the virtual arm illusion. Front Hum Neurosci 2008;2:6.
[5] Argelaguet F, Hoyet L, Trico M, Lecuyer A. The role of interaction in virtual embodiment: effects of the virtual hand representation. In: Proceedings of the IEEE virtual reality (VR); 2016. p. 3–10.
[6] Alallah F, Neshati A, Sakamoto Y, Hasan K, Lank E, Bunt A, et al. Performer vs. Observer: Whose Comfort Level Should We Consider when Examining the Social Acceptability of Input Modalities for Head-Worn Display?. In: Proceedings of the ACM symposium on virtual reality software and technology; 2018. 10:1-10:9.
[7] Mohr P, Tatzgern M, Langlotz T, Lang A, Schmalstieg D, Kalkofen D. TrackCap: Enabling Smartphones for 3D Interaction on Mobile Head-Mounted Displays. In: Proceedings of the CHI conference on human factors in computing systems; 2019. 585:1-585:11.
[8] Langton SR, Watt RJ, Bruce V. Do the eyes have it? cues to the direction of social attention. Trends Cogn Sci (Regul Ed) 2000;4(2):50–9.
[9] Takagi M, Oyamada H, Abe H, Zee DS, Hasebe H, Miki A, et al. Adaptive changes in dynamic properties of human disparity-Induced vergence. Invest Ophthalmol Visual Sci 2001;42(7):1479–86.
[10] Colombo C, Del Bimbo A. Interacting through eyes. Rob Auton Syst 1997;19(3–4):359–68.
[11] Kumar M, Winograd T, Winograd T, Paepcke A. Gaze-Enhanced Scrolling Techniques. In: Extended abstracts on human factors in computing systems; 2007. p. 2531–6.
[12] Toyama T, Sonntag D, Orlosky J, Kiyokawa K. Attention engagement and cognitive state analysis for augmented reality text display functions. In: Proceedings of the international conference on intelligent user interfaces; 2015. p. 322–32.

[13] Lee Y, Shin C, Plopski A, Itoh Y, Piumsomboon T, Dey A, et al. Estimating gaze depth using multi-layer perceptron. In: Proceedings of the international symposium on ubiquitous virtual reality; 2017. p. 26–9.
[14] Katzakis N, Teather RJ, Kiyokawa K, Takemura H. INSPECT: Extending plane–Casting for 6-DOF control. Hum-Centric Comput Inf Sci 2015;5(1):22.
[15] Qian YY, Teather RJ. The eyes don't have it: an empirical comparison of head--based and eye-based selection in virtual reality. In: Proceedings of the symposium on spatial user interaction; 2017. p. 91–8.
[16] Blattgerste J, Renner P, Pfeiffer T. Advantages of eye-gaze over head-gaze-based selection in virtual and augmented reality under varying field of views. In: Proceedings of the workshop on communication by gaze interaction; 2018. 1:1-1:9.
[17] Kuo C, Fanton M, Wu L, Camarillo D. Spinal constraint modulates head instantaneous center of rotation and dictates head angular motion. J Biomech 2018;76:220–8.
[18] Stoakley R, Conway MJ, Pausch R. Virtual reality on a WIM: interactive worlds in miniature. In: Proceedings of the CHI conference on human factors in computing systems, 95; 1995. p. 265–72.
[19] Chae HJ, Hwang J-i, Seo J. Wall-based space manipulation technique for efficient placement of distant objects in augmented reality. In: Proceedings of the 31st annual ACM symposium on user interface software and technology. UIST 18; 2018. p. 45–52.
[20] Nanayakkara S, Shilkrot R, Yeo KP, Maes P. EyeRing: a finger-worn input device for seamless interactions with our surroundingS. In: Proceedings of the augmented human international conference; 2013. p. 13–20.
[21] Castle RO, Murray DW. Object recognition and localization while tracking and mapping. In: Proceedings of the IEEE international symposium on mixed and augmented reality; 2009. p. 179–80.
[22] Simon A, Doulis M. NOYO: 6DOF elastic rate control for virtual environments. In: Proceedings of the ACM symposium on virtual reality software and technology; 2004. p. 178–81.
[23] Piekarski W, Thomas B. Augmented reality working planes: a foundation for action and construction at a distance. In: Proceedings of the ISMAR 2004: Proceedings of the Third IEEE and ACM international symposium on mixed and augmented reality; 2004. p. 162–71. doi:10.1109/ISMAR.2004.17.
[24] Johansson RS, Westling G, Bäckström A, Flanagan JR. Eye–hand coordination in object manipulation. J. Neurosci. 2001;21(17):6917–32.
[25] Stellmach S, Dachselt R. Still looking: Investigating seamless gaze-supported selection, positioning, and manipulation of distant targets. In: Proceedings of the SIGCHI conference on human factors in computing systems; 2013. p. 285–94.
[26] Kytö M, Ens B, Piumsomboon T, Lee GA, Billinghurst M. Pinpointing: precise head- and eye-based target selection for augmented reality. In: Proceedings of the CHI conference on human factors in computing systems; 2018. 81:1-81:14.
[27] Pfeuffer K, Mayer B, Mardanbegi D, Gellersen H. Gaze + pinch interaction in virtual reality. In: Proceedings of the symposium on spatial user interaction; 2017. p. 99–108.
[28] Song J, Cho S, Baek S-Y, Lee K, Bang H. Gafinc: gaze and finger control interface for 3D model manipulation in CAD application. Comput-Aided Des 2014;46:239– 245.
[29] Hornof A, Cavender A, Hoselton R. EyeDraw: a system for drawing pictures with eye movements. In: Proceedings of the international ACM SIGACCESS conference on computers and accessibility; 2004. p. 86–93.
[30] Piumsomboon T, Lee G, Lindeman RW, Billinghurst M. Exploring natural eye--gaze-based interaction for immersive virtual reality. In: Proceedings of the IEEE symposium on 3d user interfaces; 2017. p. 36–9.
[31] Mardanbegi D, Langlotz T, Gellersen H. Resolving Target Ambiguity in 3D Gaze Interaction through VOR Depth Estimation. In: Proceedings of the CHI conference on human factors in computing systems; 2019. p. 612.
[32] Jacob RJK. Eye tracking in advanced interface design. In: Virtual environments and advanced interface design. Oxford University Press, Inc.; 1995. p. 258–88.
[33] The R Project for Statistical Computing. https://www.r-project.org/; 1993–2020. Accessed on Jan 24, 2020.
[34] Coin: Conditional Inference Procedures in a Permutation Test Framework. http://coin.r-forge.r-project.org/; 2005–2020. Accessed on Jan 24, 2020.
[35] Rosenthal R. Parametric measures of effect size. In: Cooper H, Hedges LV, editors. *The handbook of research synthesis*; 1994. p. 231–44.