Hands-Free Is Fine: Gaze-Dominant Object Manipulation in Virtual Reality

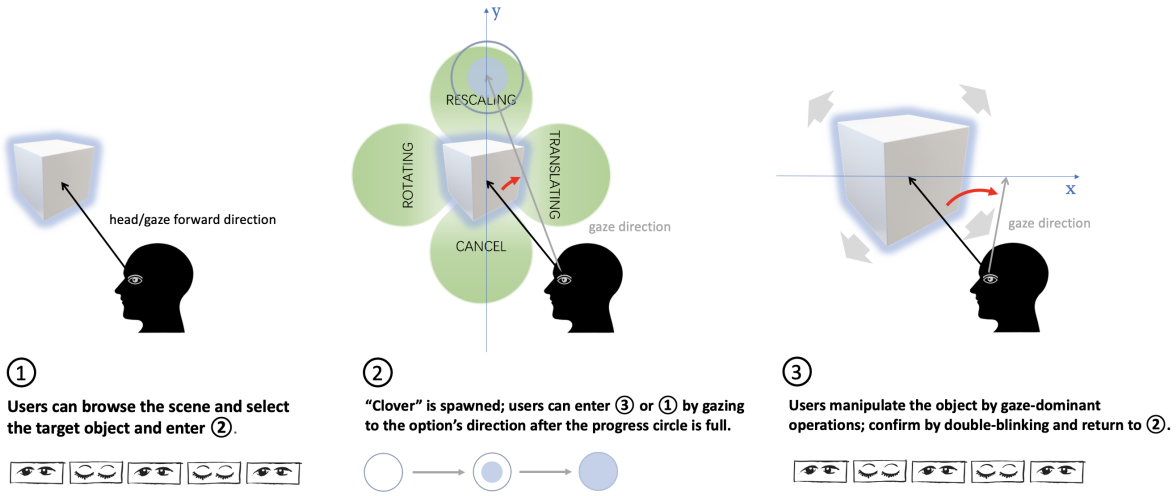

Figure 1: ① is in the initial state (IDLE) of the system; users can browse the scene and aim the target object based on the head forward direction. At aiming, the object would highlight; user can now double-blink to select and enter ②. ② is in the OBJECT_SELECTED state of the system, where the "Clover" Mode Switching Menu is spawned at the center of the interface. Users can enter ③ by gazing to the direction of a manipulation mode or enter ① by gazing at "CANCEL", but only after the progress circle is full in case of mishandling. ③ is in the manipulating state of the system (e.g., OBJECT_RESCALING), where users manipulates the selected object with gaze-dominant operations. Users can confirm the manipulation by double-blinking and return to ②.

Abstract

Efficient object manipulation is critical to VR interaction, and hands-free is a method worth discussing. We introduce a hands-free gaze-dominant manipulating pipeline that incorporates a mode-switching user interface menu, Clover. This menu empowers users to effortlessly switch between different manipulation modes using only their eye movements. By exploring alternative methods that eliminate the reliance on physical hand gestures or controllers, this method has the potential to offer expanded possibilities in various contexts, including situations where individuals have physical disabilities or their hands are occupied or restrained. And we conducted two user studies. The results show that our approach significantly improves efficiency (success rate, task completion time, and final distance) and user experience (SSQ, SUS, and NASA-TLX) compared to current state-of-the-art methods.

Index Terms: Virtual Reality—3D Object Manipulation—Gaze Input—Hands-free Manipulation

1 Introduction

Object manipulation is one of the fundamental interaction methods in virtual reality (VR). There are several comprehensive reviews [21, 23] available that provide detailed descriptions of manipulation techniques. Interested readers can refer to these reviews for in-depth reading on the subject.

Currently, the primary method for object manipulation in virtual reality (VR) involves using hands [4, 7, 10, 14, 15, 17, 21, 26, 28, 40, 48, 52–54]. The Virtual Hand [4, 14, 26, 28, 40, 52, 54], a predominant input method in VR, allows users to manipulate objects based on hand tracking; while criticized for inefficiency and lack of precision, approaches like speed enhancement [52], scaling [14, 26, 40, 54], control-display ratio adjustment [7, 53], DoF separation [10, 15, 22, 48], viewpoint quality [51], and emerging concepts like gain [19], MGF [18], and techniques like VR-HandNet [9] aim to improve manipulation control in VR. However, there are scenarios where hand-based object manipulation is not feasible in VR. For example, in many scenarios, hand-based manipulation may not be optimal due to factors such as physical disabilities or situations where hands are occupied or restrained. Exploring alternative methods that alleviate the need for physical hand gestures or controllers can unlock new opportunities and improve accessibility. A comprehensive review on hand-free interaction in virtual reality [23] provides detailed descriptions of hand-free VR interfaces. OrthoGaze method [16] has been proposed for object manipulation through gaze, but it only considers translating and is tedious as it frequently requires re-selecting orthogonal planes. Regarding eye movements, there are three challenges as follows: (1) the limited availability of signals obtained through eye tracking; (2) the instability of eye movement signals and the difficulty in resolving them due to interference from instinctive actions such as blinking; and (3) the significant cognitive load imposed by eye movement manipulation, as existing methods often require complex procedures and intense visual focus, leading to ocular fatigue.

To address these challenges, we introduce a comprehensive pipeline that incorporates a mode-switching 3D user interface menu, Clover. This menu empowers users to effortlessly switch between different manipulation modes using only their eye movements. It provides an inclusive and versatile means of interacting with virtual reality environments, enabling a broader range of users to engage with immersive experiences. To evaluate the performance of our

method, we conducted two user studies. Compared to current state-of-the-art methods, our method significantly improves in efficiency (success rate, task completion time, and final distance) and user experience (SSQ, SUS, and NASA-TLX). Figure 1 illustrates a complete flow of manipulation.

In summary, the contributions of our method are as follows:

- We proposed a fully hands-free object manipulation method based on gaze-dominant interaction, which significantly outperforms the current state-of-the-art gaze-based hands-free object manipulation method.
- We introduced Clover, a Mode Switching Menu, to provide smooth manipulation mode switching, thereby establishing a complete closed-loop manipulation process.
- We designed a user study with the task of block-building, facilitating a quantitative evaluation of the efficiency of the proposed method.

2 Related Work

Efficient object manipulation is crucial for VR interactions. Many researchers have dedicated over 20 years to studying this topic. Currently, the primary focus in VR object manipulation methods revolves around hand-based manipulation. However, there is an increasing emphasis on hand-free human-computer interaction in VR, and recent developments have introduced gaze-based object manipulation methods. In this section, we primarily review the relevant literature on object manipulation methods through manipulation based on hands and controllers (Sect. 2.1) and manipulation supported by gaze (Sect. 2.2).

2.1 Manipulation based on Hands and Controllers

The Virtual Hand, based on mid-air interaction, is a predominant input method utilized in contemporary virtual reality (VR) systems [8, 21]. By tracking the spatial positions of the hand, typically with six degrees-of-freedom (DoF), users can directly manipulate and rotate objects in virtual environments, mimicking their actions in the physical world [30]. Despite criticisms of its inefficiency and lack of precision [5, 21], the Virtual Hand remains widely adopted in various VR applications due to its simplicity and intuitive control. To enhance the capabilities of the Virtual Hand, additional approaches have been employed. For instance, Go-Go [28] and its recent extension [52] increase the speed of the virtual hand, enabling users to reach distant targets, even at potentially infinite distances [4]. Raycasting offers an alternative solution for interacting with distant objects, but precise rotation may be challenging as the hand is attached to the end of the ray [4]. Other methods [14, 26, 40, 54] scale down the virtual world to facilitate interaction with objects that are out of the user's reach.

In order to provide more precise manipulation control, several interaction techniques employ a decrease in the control-display ratio based on hand velocity [7, 53]. Another promising approach is the separation of degrees of freedom (DoF) [15, 22, 48], wherein only one or two DoF are manipulated at a time, rather than all six simultaneously. For example, recent research attempted to reduce DoF during object manipulation by confining it to the shape of a point, ray, or plane, thereby enhancing precision [10]. Viewpoint quality is also proposed for enhancing manipulation efficiency [51]. Recently, the concept of gain [19] and manipulation guidance field [18] has been proposed. Besides, VR-HandNet [9] employs a neural network to perform dexterous hand manipulation.

However, many mid-air interaction techniques encounter limitations when it comes to supporting prolonged manipulation due to cumulative muscle fatigue in the user's arm, commonly known as the "gorilla arm" effect [12]. This issue is particularly problematic in interaction scenarios like 3D modelling in VR, which demand precise and extended usage of mid-air interfaces. To address these challenges, incorporating indirect mappings [17] or integrating less physically demanding input modalities, such as gaze, into VR object manipulation techniques holds the potential to provide relief.

2.2 Manipulation Supported by Gaze

The utilization of gaze for object manipulation has been extensively explored in various contexts beyond virtual reality (VR). Generally, while gaze provides quick and intuitive pointing, it faces challenges related to imprecise selection and the difficulty of confirming a choice. To address these challenges, several techniques have combined gaze with additional modalities, such as the "gaze select, hands manipulate" principle [6, 24, 39, 49]. For instance, Pfeufer et al. introduced Gaze-touch [24], which allowed users to control gaze-selected targets through multi-touch gestures on interactive surfaces indirectly. Another approach, proposed by Turner et al. [46], involves mapping the object that the user is looking at to the touch/cursor position, enabling further manipulation. Conversely, alternative approaches [32, 38, 44–47, 50] for transferring content between different displays have integrated gaze movement into the translation process. These prototypes typically necessitate the use of a hand trigger to "attach" the object to the gaze direction and subsequently release the trigger to "drop" it. In a subsequent study, Turner et al. [43] expanded on this concept by developing techniques that maintain concurrent rotation and scaling operations when performing translation tasks using gaze and touch.

Eye and head synergetic pointing and selection are also extensively researched in recent years, such as Sidenmark et al.'s work [33], [34], and [35]. However, limited research has explored the use of gaze input for object manipulation in virtual reality (VR) or 3D virtual space. Simeone et al. [36] combined bi-manual touch gestures with gaze input to enable object scaling along the XYZ-axis within a touchscreen. Liu et al. presented OrthoGaze [16], where gaze was employed to move an object along three orthogonal planes in VR. Others have utilized eye gaze for object selection and employed indirect freehand gestures for manipulation [25, 27, 31, 37]. These approaches still adhere to the "gaze select, hands manipulate" concept. In contrast, our work incorporates gaze input not only for object selection but also for the entire process of manipulating the target, requiring continuous actions rather than discrete selection operations [43]. To achieve efficient transitions between each manipulating mode, we incorporate a switching menu, Clover; this idea is spurred by StickyPie [1], a recent menu design for AR/VR introduced by Ahn et al. Our objective is to investigate how various methods of integrating, coordinating, and transitioning between eye and head movements can enhance user performance and provide a seamless manipulating flow.

3 Method

Firstly, we present a pipeline (Sect. 3.1) for object manipulation. Subsequently, we elaborate on the specific aspects of scene browsing and target selection (Sect. 3.2), followed by an in-depth discussion of the "Clover" Mode Switching Menu (Sect. 3.3) that we have developed. Finally, we provide detailed insights into the techniques and procedures involved in object manipulation(Sect. 3.4).

3.1 Pipeline

The method pipeline of the object manipulation interaction system can be represented with a finite state machine, as shown in Figure 2.

Scene Browsing and Target Selection: Corresponding to `IDLE`. During target selection, the user initially aligns the head forward, directing a ray toward the target. Subsequently, a double-blinking serves as a confirmation signal for selecting the object. The advantages of this selection and confirmation method were discussed by the Yuan Yuan Qian team in 2017 [29]. Upon receiving the confirmation signal, the system transits to the next state.

Mode Selection: Corresponding to `OBJECT_SELECTED`. A "Clover" Mode Switching Menu is generated at the center of the

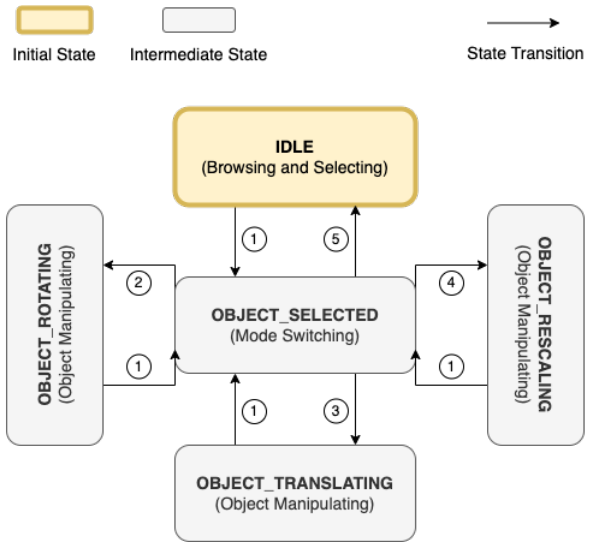

Figure 2: The finite state machine with transitions: ① eye-based confirmation signal; on Clover (Figure 1 ②), selected ② "ROTATING" or ③ "TRANSLATING" or ④ or "RESCALING" or ⑤ "CANCEL".

3D user interface, allowing users to gaze at a specific option to enter the corresponding manipulation mode. Users can also gaze at the "return" option to go back to state 1 (refer to Section 3.3 for details).

Object Manipulation: Corresponding to three intermediate states: `OBJECT_TRANSLATING`, `OBJECT_ROTATING`, or `OBJECT_RESCALING` in the finite state machine. Our proposed manipulation system supports six degrees of freedom (6DOF) manipulation. In different object manipulation modes, users can perform three degrees of freedom displacement, two degrees of freedom rotation, and one degree of freedom scaling on the object (refer to Section 3.4 for details).

Confirming Manipulation: A rapid double-blinking serves as a confirmation signal to validate the current manipulation status and return to `OBJECT_SELECTED`.

3.2 Scene Browsing and Target Selection

The browsing and selection methods of this interaction system eliminate the need for complex controllers or other input devices, allowing users to engage with the virtual environment in a more natural manner. In addition, to cope with the challenge of unstable and noisy eye-tracking data, we employ a filtering algorithm [2].

3.2.1 Scene Browsing

During scene exploration, the system utilizes the user's head movement to control the virtual camera, providing a real-time view of their head-mounted display. The real-time viewpoint calculation considers the user's field of view and viewpoint position, which are determined by their head forward direction and virtual environment location, respectively.

For object selection, we employ the ray-casting method, a widely used technique in virtual reality. A ray is cast forward from the user's head to detect collisions with objects in the scene, ensuring accurate focal point determination. This method offers simplicity, intuitiveness, efficiency, and low learning cost and reduces the potential for dizziness or confusion. Objects intersecting with the focal point are highlighted, eliminating ambiguity during selection. Additionally, a pointer appears at the intersection of the forward ray and the user interface to assist in targeting the desired object, specifically during `IDLE`.

3.2.2 Target Selection

In `IDLE`, users can utilize eye movements as confirmation signals for target selection. The main challenges in this part include accurately capturing and analyzing the user's eye movement data and providing appropriate feedback and cues.

The most common and natural active eye movement signals are single-eye blink and quick double-eye blink. Therefore, we consider these two eye movement behaviours as candidates for the final eye movement confirmation signal pool. We conducted a pilot study to determine the most efficient and least burdensome eye movement confirmation signal from these two options. Detailed information on this pilot study will be provided in the subsequent section on experimental design. By analyzing the experimental results of the two eye movement behaviours using weighting analysis, we ultimately determined that the quick double-eye blink is the optimal choice for target selection and confirmation.

After the selection is confirmed, the system transitions to the `OBJECT_SELECTED`state and provides the user with auditory feedback. This feedback, a modality different from visual feedback, aims to enhance the reliability of the interaction system and reduce user confusion.

3.3 "Clover" Mode Switching Menu

3.3.1 Rationale

During the design phase of our interaction system, we took into consideration that most real-world manipulation processes involve not only the selection of a specific manipulation mode but also the need to switch between multiple modes. However, existing gaze-based and eye-tracking interaction methods do not adequately address convenient mode switching. In these methods, if a user needs to switch to a different manipulation mode after completing one manipulation, they would have to revert back to the initial state and repeat the entire selection-to-manipulation process, resulting in significant redundancy and efficiency loss. Hence, considering this limitation, we have designed a "Clover" Mode Switching Menu.

The "Clover" Mode Switching Menu is intended to provide users with a convenient way to select or switch to a particular interaction state using gaze-based actions.

3.3.2 Arrangement and Detailed Designs

The "Clover" Mode Switching Menu is generated on the user interface only when the system's finite state machine is in `OBJECT_SELECTED`. Users can use this menu to choose a specific interaction mode: spatial displacement, spatial rotation, and spatial scaling, corresponding to `OBJECT_TRANSLATING`, `OBJECT_ROTATING`, and `OBJECT_RESCALING` in the system's finite state machine, respectively. Users can also use the menu to deselect an object and return to `IDLE`. This process is reflected in our defined system's finite state machine, as shown in Figure 2.

The menu provides four options in four directions: displacement, rotation, scaling, and cancel. Users can select the corresponding option by looking in the respective direction. To determine the specific mapping between each direction and option, we conducted a questionnaire survey distributed to a sample of 30 randomly selected individuals to gather their perceived frequency rankings of the four options. Additionally, we referred to the findings by Maxwell et al. in 2006 (which suggested that the burden of horizontal eye movements is lower than that of vertical movements [20]) to position the two most frequently selected interaction mode options (displacement and rotation) on the left and right sides, while placing the remaining two options (scaling and cancel) on the top and bottom sides.

To avoid accidental triggers and the classic "Midas touch" problem in human-computer interaction [11], we defined a selection confirmation time. By default, this time is set to 1 second, but users can adjust it according to their preferences before starting. When users make a selection, a progress circle (Figure 1 ②) is displayed on

the interface, indicating the countdown for selection confirmation. The progress dial gradually fills up as the user's gaze ray remains focused on the selected option. Users can cancel the selection by redirecting their gaze before the progress dial is completely filled. When the selection confirmation countdown ends, and the progress dial is full, the user immediately transitions to the corresponding state associated with the option indicated by their gaze ray.

3.4 Object Manipulation based Gaze

Our interactive system supports complete 6DOF (Degrees of Freedom) object manipulation, allowing users to perform spatial translation, spatial rotation, and uniform rescaling on target objects. These specific manipulations correspond to `OBJECT_TRANSLATING`, `OBJECT_ROTATING`, and `OBJECT_RESCALING` in the finite state machine of our system. We continue employing the filtering algorithm [2] to maintain stability.

For each object manipulation mode, we strive to employ a linear mapping strategy and introduce a Gaze Adaptation Function, where v represents the original offset of gaze, and A represents the adapted value for manipulation; we can consider the value used for manipulation based on original gaze would simply follow $A = v$):

$$A = 5v^5, -1 \leq v \leq 1 \tag{1}$$

The eye-tracking device captures gaze data and transforms it into a uniform linear distribution between -1 and 1. Our interactive system applies an adaptation function based on this, allowing fine adjustments during small-scale motions and rapid movements during large-scale motions. Figure 3 demonstrates the enhanced effects, where small-range eye movements are attenuated while others are enhanced.

To minimize operational burden and learning difficulty, we primarily rely on eye movements for manipulation, avoiding the introduction of additional modalities. A non-relative three-dimensional Cartesian coordinate system is established for explaining manipulation methods in the object manipulation space, where the positive X-axis points to the right, the positive Y-axis points upwards, and the positive Z-axis points towards the viewer (out of the screen). The origin (0, 0, 0) is the point where the X, Y, and Z axes intersect. Objects in the scene are positioned and oriented based on their coordinates within this coordinate system.

To ensure a natural interaction flow and enable subtle actions, we combine signals from eye movements and head movements using a collaborative processing equation; by optimizing the calculation of fixation, this equation reduces cognitive load and creates a smoother

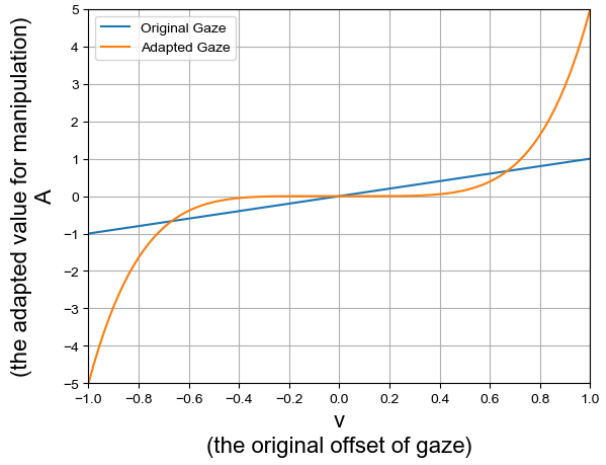

Figure 3: Gaze Adaptation Function

interaction process. We define the eye movement manipulation fixation dwell OE at time t_0 as the combined angular offset of the eye movement forward ray and the head movement forward ray within a time interval n.

$$OE_{t_0} = \frac{1}{n} \sum_{t=t_0-n}^{t_0} \left| \hat{eye}_t \cdot \hat{head}_t - \hat{eye}_{t-1} \cdot \hat{head}_{t-1} \right| \tag{2}$$

Among them, $\hat{eye}$ is the unit vector representing the line of sight, $\hat{head}$ is the unit vector indicating the direction of head gaze, and $\cdot$ denotes the dot product operator between vectors. If OE is less than a certain threshold, it indicates that the user is attempting to focus their gaze. Our pilot experiment demonstrated the necessity of this optimization, as it leads to a smoother and more natural interaction process with reduced cognitive load.

3.4.1 Translation

During spatial displacement, for movement in the X-Y plane, the object responds accordingly to the projection distance of the eye gaze forward ray onto the X-Y plane. Assuming the projected coordinates of the eye gaze forward ray on the X-Y plane are (x, y), the object's movement along the X-axis and Y-axis can be represented as:

$$\delta_x = \begin{cases} 0, & if\ |x| < T \\ x \cdot C, & default \end{cases} \tag{3}$$

$$\delta_y = \begin{cases} 0, & if\ |y| < T \\ y \cdot C, & default \end{cases} \tag{4}$$

Where T and C are pre-defined thresholds and scaling factors, respectively. For the movement along the Z-axis, the object responds to the head's rotational movement around the Z-axis by mapping the angle-distance relationship accordingly. Assuming the rotational angle of head movement around the Z-axis is denoted as ω, the object's movement along the Z-axis can be expressed as follows:

$$\delta_z = \begin{cases} 0, & if\ |\omega| < T \\ \frac{\pi}{180} \cdot \omega \cdot C, & default \end{cases} \tag{5}$$

Where T and C represent predetermined threshold values and scaling coefficients, respectively.

3.4.2 Rotation

During spatial rotation, for rotation around the X-axis, the object responds to the distance-angle mapping of the forward eye gaze ray's projection on the Y-axis. Similarly, for rotation around the Y-axis, the object responds to the distance-angle mapping of the forward eye gaze ray's projection on the X-axis. Assuming the projected coordinates of the forward eye gaze ray on the X-Y plane are denoted as (x, y), the object undergoes rotation around the X-axis and Y-axis as follows:

$$\delta_x = \begin{cases} 0, & if\ |x| < T \\ \frac{180}{\pi} \cdot x \cdot C, & default \end{cases} \tag{6}$$

$$\delta_y = \begin{cases} 0, & if\ |y| < T \\ \frac{180}{\pi} \cdot y \cdot C, & default \end{cases} \tag{7}$$

Where T and C represent predetermined threshold values and scaling coefficients, respectively.

3.4.3 Rescaling

When performing spatial scaling, the object responds to the eye gaze by mapping the distance between the eye gaze ray projection on the X-axis and the scaling factor. Assuming the projected coordinates of the eye gaze ray on the X-axis are $(x,0)$, the scaling factor K of the object is computed as follows:

$$K = \begin{cases} 0, & if\ |y| < T\ or\ x \leq -1 \\ 2, & if\ \frac{x}{C} \geq 1 \\ 1+\frac{x}{C}, & default \end{cases} \tag{8}$$

Where T and C represent predetermined threshold values and scaling coefficients, respectively. Based on these scaling coefficients, the specific manifestation of object scaling is given by $Scale' = Scale \cdot K$, where K denotes the scaling factor. We consider the boundary we set ($K \in [0,2]$) to be necessary because the K value would be meaningless if it is less than 0, considering it is to multiply with the current scale; if it is greater than 2, users would likely lose control of the enlarging process, introducing extra task load and manipulating time to repeatedly adjust.

4 Pilot Study

All our studies (incl. pilot studies and user studies) followed a within-subjects design and were conducted under the environmental conditions specified in Table 1. And we referred to Triantafyllidis et al.'s and Bergström et al.'s papers [3,42] for selecting metrics.

4.1 Pilot Study 1: Eye-based Confirmation

4.1.1 Design and Procedure

This pilot study aims to determine the most efficient and least burdensome eye movement confirmation signal between winking and double-blinking. Participants completed the task twice, using each eye movement as the confirmation signal. They eliminated 20 small balls in a virtual environment by aiming and issuing the designated eye movement. Results were compared based on rules that ensured comparability. Participants recorded the number of signal attempts by pressing the space key. The scene and procedure are shown in Figure 4. Afterward, participants completed a separate NASA-TLX workload assessment form for each confirmation signal.

Participants. We recruited six participants for this study to maintain a balanced gender ratio to ensure diversity. Before the study, participants were required to fill out a demographic questionnaire, which included questions about gender, occupation, educational background, age, and ability to perform winking naturally. According to the questionnaire, the experimental group consisted of three males and three females, whose ages ranged from 20 to 22 with a variance of 0.567, and three individuals were unable to perform winking naturally; 3 participants are during their undergraduate study and 3 are during their post-graduate study.

Table 1: Study Environmrnt

Category	Item	Spec
VR	HMD	Vive Focus 3
	Eye Tracking Device	Vive Focus 3 Eye tracker
PC	OS	Windows 11
	GPU	GTX 3060
	CPU	i7-9900KF
	RAM	16GB
Software	Unity Editor	2021.3.16f1
	Vive Business Streaming	1.10.11
	SteamVR	1.24.7
Participants	Gesture	Seated

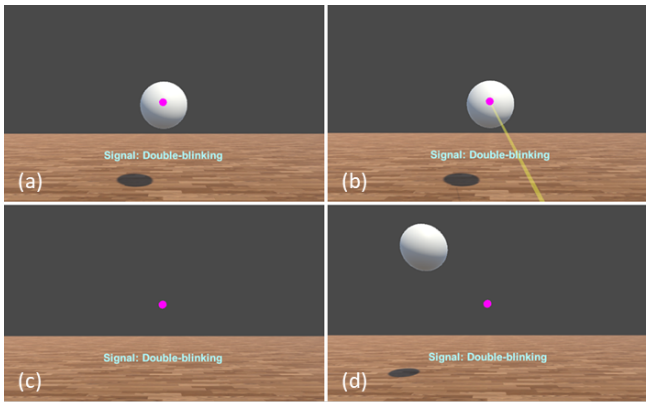

Figure 4: Scene and Procedure. There would constantly be a pointer in the middle of the interface. A ball is randomly spawned (a). The user eliminates the ball with the designated eye confirmation signal (e.g., double-blinking); there would be a ray indicating the signal (b). Then, the user waits for the next ball to spawn (c). The next ball is randomly spawned, and the user repeats the previous actions (d).

Metrics. We evaluate the results based on one subjective metric (NASA-TLX) and the following two objective metrics:

- Task completion time (in seconds);
- Feedback Accuracy Index (FAI). In a single experiment, let the number of balls successfully eliminated by a participant be denoted as N, and the total number of signal attempts made be denoted as M. The FAI is calculated as follows:

$$FAI = \frac{N}{M} \tag{9}$$

Hypothesis. The final eye confirmation signal should outperform the other in at least one metric.

4.1.2 Results

The results are reported in Figure 5. We employ the Mann-Whitney U test to analyze the significance of the data differences. This test assumes that the two samples come from two populations that are identical except for the difference in population means. Its purpose is to test whether the means of these two populations differ significantly. For task completion time, the test result is $U = 18,\ p < 0.05$; for task workload, the test result is $U = 42,\ p < 0.05$; for FAI, the test result is $U = 15.5,\ p < 0.05$. It can be observed that there are significant differences in all three sets of results. Moreover, the mean performance of the rapid double blink is more favourable, indicating that the rapid double blink is significantly superior to the one-eye blink, supporting our hypothesis.

4.2 Pilot Study 2: Gaze Fixation Optimization

4.2.1 Design and Procedure

The aim of this pilot study is to determine the necessity of introducing optimization for gaze fixation calculation.

Each participant will complete the same task twice. In both instances, we will randomly introduce optimization in one of the tasks to eliminate subjective psychological interference. The task content is identical to that of pilot study 1. For this pilot study, we specify that double-blinking will be used as the confirmation signal for both tasks.

After the completion of the experiments, each participant filled out a NASA-TLX workload assessment form for both tasks. Since

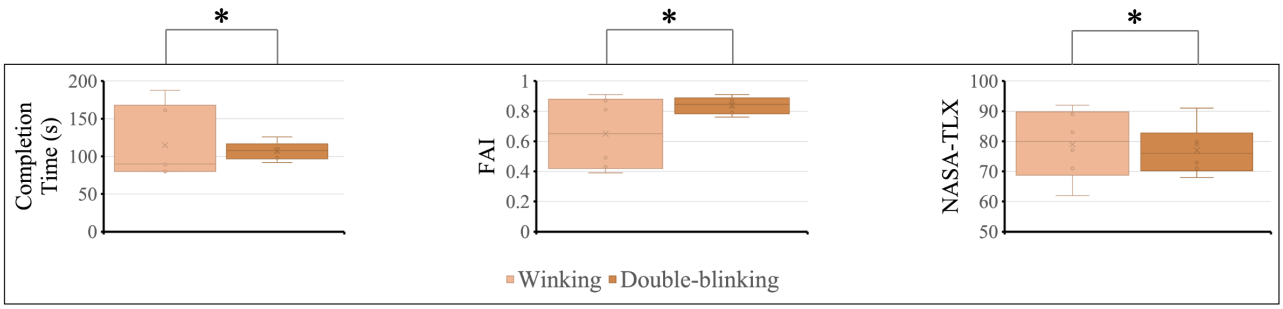

Figure 5: Completion Time (in seconds), NASA-TLX, FAI. Significant differences are denoted as $*$ if exist.

the users were unaware of whether optimization was introduced in a particular task during the experiment, we can consider the results to be objective.

Participants. The experimental group in this study is consistent with Pilot Study 1 (refer to Section 4.1).

Metrics. We evaluated the results based on NASA-TLX. The report can be found in Figure 6.

Hypothesis. The introduction of optimization should bring improvement to NASA-TLX.

4.2.2 Results

We used the Mann-Whitney U test to analyze the significance of the differences. For the workload, the test result was $U = 32,\ p < 0.05$. It can be observed that there is a significant difference between the two groups of results, and there is no apparent correlation between the introduction of optimization groups and NASA-TLX results. The mean workload of the group with optimization introduced performed more favourably, indicating that the introduction of optimization is necessary, supporting our hypothesis.

5 User Study

While the user studies conducted in this research did not specifically involve individuals with disabilities or encumbered hands, the findings shed light on the viability and potential benefits of the proposed method. Future investigations should aim to include participants from diverse backgrounds, including those with physical disabilities or constrained hand movements, to comprehensively assess the applicability and impact in such contexts.

5.1 User Study 1: Single Object Translating

5.1.1 Design and Procedure

In this study, we employed the OrthoGaze method, which is currently considered the optimal method based on head-eye coordination, as a control condition (CC) and baseline. We also reused one of the object displacement user studies from OrthoGaze [16]. The user study was designed as follows: Participants were positioned at the starting position $(0, 1[m], 0)$ and were tasked with moving a white cube of size $0.5[m] \times 0.5[m] \times 0.5[m]$ from a fixed starting position $(-1\ [m],\ 0.5\ [m],\ 5.5\ [m])$ to multiple target positions. In each task, a translucent cube of the same size as the white cube appeared at the target position. Participants were required to align the white cube with the target cube, as shown in Figure 8. The target positions were always located at the corners of a cubic space with side length $2N\,[m]$, where the center coincided with the initial position of the white cube. We used 8 different sizes of cubic spaces, where $N \in \{0.5,\ 1,\ 1.5,\ 2,\ 2.5,\ 3,\ 3.5,\ 4\}$.

To ensure a reasonable number of cube docking scenarios, each directional offset in the cubic space was selected twice with different distances, generating a total of 16 target positions. Successful docking required the distance between the white and green cubes to be less than 0.2m when confirmed by the user. The target cube turned red when within the threshold distance. Participants practiced with ten docking attempts before the formal experiment. If cubes were not aligned within 20 seconds, the task was considered a failure. After completing all tasks, participants filled out the Simulator Sickness Questionnaire (SSQ), System Usability Scale (SUS), and NASA Task Load Index (NASA-TLX) questionnaires.

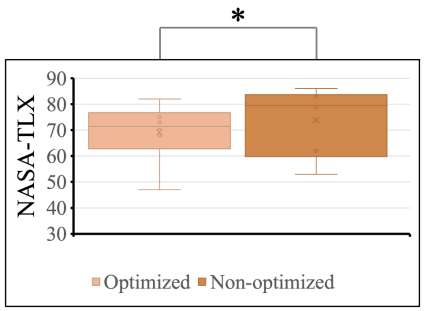

Figure 6: NASA-TLX. A significant difference is denoted as $*$ if exists.

Participants. We recruited 14 participants and made efforts to maintain a balanced gender ratio. Prior to the experiment, participants were required to complete a personal information questionnaire, which inquired about their gender, age, educational background, VR usage experience (subjective measure), and more. The results revealed that the experimental group consisted of eight males and six females, with ages ranging from 20 to 38, with a variance of 24.273. Among them, four participants had no prior VR experience (novices), 6 had some exposure to VR (intermediates), and four were very familiar with VR usage (experts). Among them, 10 are during their undergraduate study and 4 are during their post-graduate study.

Metrics. We evaluate the experimental results of each participant using three subjective and three objective metrics. The subjective metrics include: sense of presence (SSQ), usability (SUS), and task load (NASA-TLX), and the objective metrics include:

- **Success Rate**: The success rate is calculated for each participant and represents the ratio of successful tasks to the total number of tasks. This evaluates the overall efficiency of manipulating objects, as successfully completing a task requires considering both accuracy and speed. If a task is successful, the "Success" column in the table records "Y"; otherwise, it records "N".
- **Completion Time**: For each successfully completed task, the completion time is recorded. Failed tasks are not taken into account for this metric.
- **Final Distance**: This metric only applies to failed tasks. The final distance refers to the Euclidean distance between the white cube and the target position at the moment of failure.

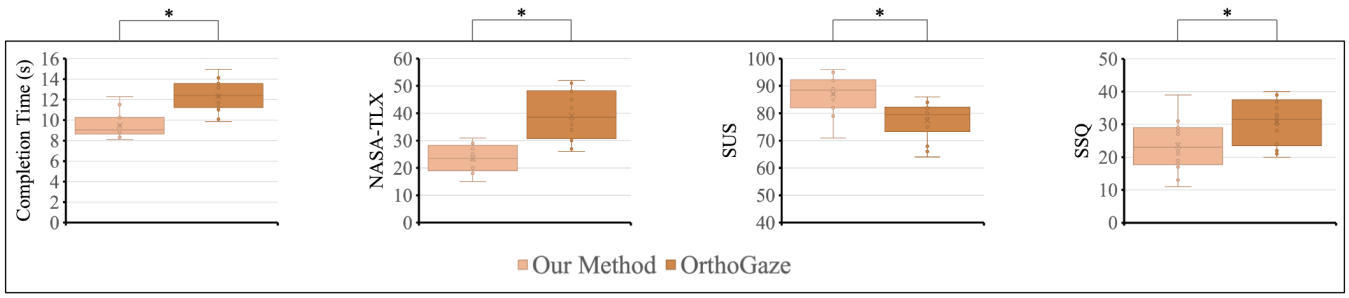

Figure 7: Completion Time (in seconds), SSQ, SUS, NASA-TLX. Significant differences are denoted as $*$ if they exist.

This reflects the proximity or distance of the object to its initial position when moving it to the target location.

Hypothesis 1. Our method would significantly outperform OrthoGaze in at least one metric.

Hypothesis 2. A high learnability should induce no difference in completion time with different VR experiences.

5.1.2 Results and Discussion

The mean comparison of the results for this user study is presented in Figure 7. Since all tasks were successful, we no longer consider the final distance metric.

It can be observed that the EC (Experimental Condition) has overall superior effects compared to the CC (Control Condition). In this experiment, there is one independent variable (factor) and one dependent variable (response variable): the independent variable is the manipulation method, which is a categorical variable, and

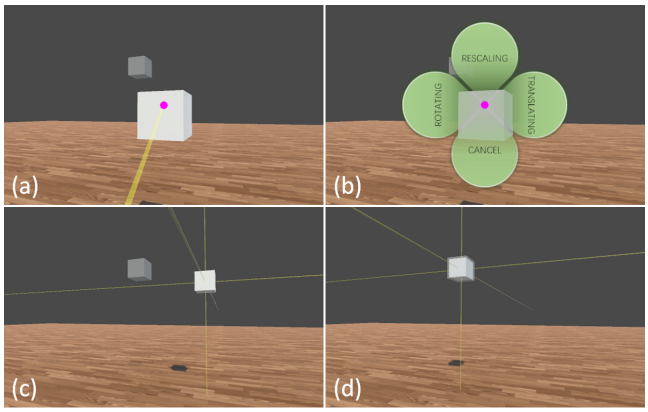

Figure 8: Scene and Procedure. (a) A translucent cube is spawned at a random target position, and the user selects the white cube to manipulate; (b) the user enters `OBJECT_TRANSLATING` with Clover; (c) the user manipulates the white cube to reach the target position; (d) the white cube is docked successfully.

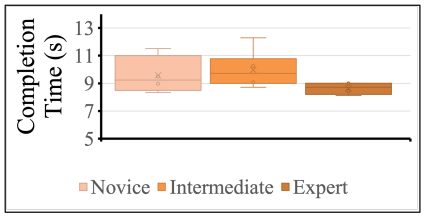

Figure 9: Completion times of difference VR experiences. Significant differences are denoted as $*$ if they exist.

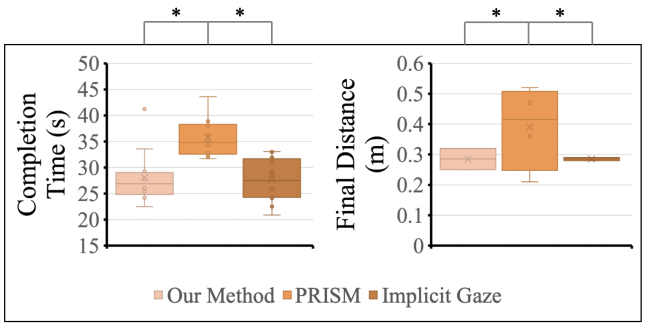

Figure 10: Completion Time and Final Distance. Significant differences are denoted as $*$ if they exist.

the dependent variable is the completion time and other outcome data, which is a continuous variable. Additionally, all samples are independent, and the Shapiro-Wilk test revealed that the data follows a normal distribution ($p_1 = 0.192 > 0.05, p_2 = 0.886 > 0.05, p_3 = 0.958 > 0.05, p_4 = 0.284 > 0.05, p_5 = 0.443 > 0.05, p_6 = 0.083 > 0.05, p_7 = 0.671 > 0.05, p_8 = 0.821 > 0.05$). Furthermore, Levene's test indicated that the assumption of homogeneity of variances among the compared data is valid ($p_{12} = 0.061 > 0.05, p_{34} = 0.745 > 0.05, p_{56} = 0.969 > 0.05, p_{78} = 0.076 > 0.05$). Therefore, we employed a one-way analysis of variance (ANOVA) to determine the significance of the differences between the two comparison methods.

The results of the statistical tests reveal significant differences between EC and CC in terms of completion time ($F_{12}[1,26] = 39.445, p_{12} = 1.20 \times 10^{-6} < 0.05$), SSQ (Satisfaction Rating Scale) ($F_{34}[1,26] = 7.119, p_{34} = 0.0130 < 0.05$), SUS (System Usability Scale) ($F_{56}[1,26] = 13.995, p_{56} = 0.000915 < 0.05$), and NASA-TLX (NASA Task Load Index) ($F_{78}[1,26] = 33.440, p_{78} = 4.32 \times 10^{-6} < 0.05$), supporting our first hypothesis. Consequently, we can conclude that our method significantly outperforms the state-of-the-art method, OrthoGaze, in terms of efficiency and user experience.

The reason for our advantages in efficiency and user experience could be: (1) we simplify the process of translating an object by minimizing selection-related actions down to once and applying a natural eye-dominant mapping strategy; (2) we apply the gaze adaptation function, making the interaction more fluent and intuitive. Furthermore, to assess the learnability of our method, we analyzed whether the experimental results produced by our method would exhibit significant differences due to variations in participants' VR experiences. Therefore, we extracted the completion time data corresponding to three levels of VR experience (none, novice, proficient) and presented it in Figure 9. Based on one-way ANOVA, there were no significant differences in completion time data among the three levels of VR experience ($F[2,11] = 2.31, p = 0.1477 > 0.05$),

supporting our second hypothesis. Hence, we can also conclude that our method possesses high learnability.

5.2 User Study 2: Multiple Object Manipulation

5.2.1 Design and Procedure

In this user study, participants are asked to complete a "block-building" task using our gaze-dominant method and two comparison methods. In each "block-building" task, we ensure that the necessary manipulation actions include translation, scaling, and rotation. The virtual scene of the experiment consists of a fixed long table, randomly generated rigid brown target blocks without collision detection on the left side of the table, and manipulable rigid white blocks with collision detection on the right side. The number of manipulable blocks is fixed at 3, and all blocks are cube-shaped because cubes provide stability and allow for an intuitive representation of scaling and rotation effects. In each task, participants need to construct the target shape on the left side by manipulating the blocks on the right side using translation, rotation, and scaling (Figure 11).

Participants will perform three independent tasks using our method (EC), the PRISM method (CC_1) [13], and the current state-of-the-art method based on gaze and hand movements, Implicit Gaze (CC_2) [55]. Each task has a completion time of 60 seconds. Meanwhile, the system continuously calculates the similarity between the constructed shape and the target shape in real-time using the Hausdorff Distance [41]. If the Hausdorff Distance falls below a certain threshold (set to 0.2), the task is considered successful, and the completion time is recorded, marking the end of the task. If the completion time exceeds the allotted time, the task is considered unsuccessful, and the current Hausdorff Distance is recorded as the final distance. After all participants have completed the tasks, we calculate the success rate, analyze the completion times (only for successful tasks), analyze the final distances (only for unsuccessful tasks), and ultimately evaluate the study's results based on these three objective metrics.

Participants. User Study 2 continues with the same set of participants as User Study 1.

Metrics. We evaluate the results of each participant using metrics which are the same as User Study 1, only in this study, we record Hausdorff distance as the Final Distance instead of Euclidean distance, considering the nature of the study design.

Hypothesis. Our method (EC) would outperform PRISM (CC_1) in at least one metric and parallel Implicit Gaze (CC_2).

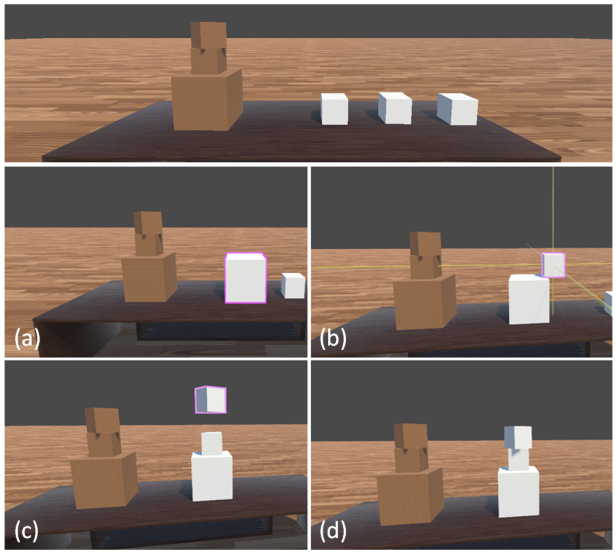

Figure 11: Scene and procedure. The top figure shows the scene at the start-up. The user (a) rescales the first block; (b) translates the second block; (c) translates and rotates the third block; (d) finishes building blocks.

5.2.2 Results and Discussion

The results of this user study are reported in Figure 10, which supports our hypothesis by showing two pairs of significant differences.

For objective metrics, we conducted a significance analysis of the completion times generated by the three comparative methods. The Shapiro-Wilk test revealed that the completion times for all three groups followed a normal distribution ($p_1 = 0.155 > 0.05, p_2 = 0.139 > 0.05, p_3 = 0.843 > 0.05$). Additionally, the Levene test indicated that the assumption of homogeneity of variances among the compared data was valid ($p_{123} = 0.788 > 0.05$). Therefore, we employed a one-way analysis of variance (ANOVA) to test the significance of differences among the data. The analysis demonstrated that there were significant differences between at least two of the groups ($F_{123}[2,31] = 7.119, p_{123} = 0.000219 < 0.05$). Subsequently, a Tukey's honestly significant difference (HSD) post hoc test was performed, revealing that the main sources of difference were between EC and CC_1, as well as between CC_2 and CC_1. Based on the mean performance, we can conclude that our method significantly outperforms PRISM in manipulation efficiency, and there is no significant difference between our method and Implicit Gaze. Furthermore, an analysis of success rates and final distances was conducted. The success rate for EC was 85.7%, higher than CC_1 with 71.4%, and consistent with CC_2. Regarding the final distance, a one-way ANOVA showed no significant differences among the three methods ($F_{456}[2,5] = 0.933, p_{456} = 0.453 > 0.05$). Therefore, we conclude that although there is no significant difference in the maximum attempt level among the three methods, our method and Implicit Gaze are superior to PRISM with considerable success rates.

For subjective metrics, there are no significant differences in SSQ, SUS, and NASA-TLX, for both methods effectively and intuitively allow users to interact with objects in a seamless and natural manner. Currently, the optimal object manipulation methods are hand-based, with PRISM being one of the most diverse hand-based methods, while Implicit Gaze represents the SOTA hand-based object manipulation method [23]. In conclusion, we can assert that our method is comparable to any hand-based method and significantly superior to the PRISM method in terms of efficiency. Compared with two of the most prevalent hand-based methods, the reason for our advantages in efficiency could be: our gaze-dominant manipulation aligns with the natural human behavior of using gaze to focus attention and interact with the environment. This intuitive interaction modality reduces the learning curve for users and enhances overall efficiency.

6 Conclusion, Limitations, and Future Work

We proposed a fully hands-free object manipulation method based on gaze-dominant interaction, which significantly outperforms the current state-of-the-art gaze-based hands-free object manipulation method and provides the potential to offer expanded possibilities in various contexts, including situations where individuals have physical disabilities or their hands are occupied or restrained. We also introduced a "Clover" Mode Switching Menu to address the inconsistency issue in object manipulation research, thereby establishing a complete closed-loop manipulation process. To enhance user experience, we apply an adaptation function for gaze signals and optimize the calculation of gaze fixation. Lastly, we presented a comprehensive "block-building" user study, facilitating a quantitative evaluation of the effectiveness of the proposed methods.

Our approach has the following three limitations. We assume that the user will always remain fixed in one position while operating the interactive system. Therefore, we have not taken into account the impact of user position changes in our mapping between head-eye

signals and object behaviors. The second limitation is the difficulty for users to continuously observe the state of the objects while interacting using eye gaze signals, resulting in the need for multiple adjustments in many cases. This is the primary source of usability loss in the system. We also have not considered the issue of occlusion when selecting target objects. Consequently, our interaction system is currently unable to handle situations where an obstacle obstructs a target object quickly.

To address the fixed mapping that affects user experience, we will introduce the concept of a relative coordinate system to generate mapping relationships for head-eye signal-object behaviors based on the user's position. To overcome the problem of not being able to observe objects during gaze manipulation, we will place a fixed virtual camera at the target object and add a visual window at a certain distance along the user's gaze ray to provide real-time feedback on the target object's state. To tackle occlusion issues during object selection, a hypothetical solution is to establish a pre-selection set containing all objects that the user's gaze ray penetrates, allowing the user to make a second selection from this set. Furthermore, we could also explore the usability of our method in other virtual environments, such as augmented and mixed reality.

REFERENCES

[1] S. Ahn, S. Santosa, M. Parent, D. Wigdor, T. Grossman, and M. Giordano. Stickypie: A gaze-based, scale-invariant marking menu optimized for ar/vr. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445297

[2] S. Z. Bagherzadeh and S. Toosizadeh. Eye tracking algorithm based on multi model kalman filter. *HighTech and Innovation Journal*, 3(1):15–27, 2022.

[3] J. Bergström, T.-S. Dalsgaard, J. Alexander, and K. Hornbæk. How to evaluate object selection and manipulation in vr? guidelines from 20 years of studies. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445193

[4] D. A. Bowman and L. F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pp. 35–ff, 1997.

[5] D. A. Bowman, R. P. McMahan, and E. D. Ragan. Questioning naturalism in 3d user interfaces. *Communications of the ACM*, 55(9):78–88, 2012.

[6] I. Chatterjee, R. Xiao, and C. Harrison. Gaze+ gesture: Expressive, precise and targeted free-space interactions. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pp. 131–138, 2015.

[7] S. Frees and G. D. Kessler. Precise and rapid interaction through scaled manipulation in immersive virtual environments. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, pp. 99–106. IEEE, 2005.

[8] G. Ganias, C. Lougiakis, A. Katifori, M. Roussou, Y. Ioannidis, and l. P. Ioannidis. Comparing different grasping visualizations for object manipulation in vr using controllers. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2369–2378, 2023. doi: 10.1109/TVCG.2023.3247039

[9] D. Han, R. Lee, K. Kim, and H. Kang. Vr-handnet: A visually and physically plausible hand manipulation system in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–12, 2023. doi: 10.1109/TVCG.2023.3255991

[10] D. Hayatpur, S. Heo, H. Xia, W. Stuerzlinger, and D. Wigdor. Plane, ray, and point: Enabling precise spatial manipulations with shape constraints. In *Proceedings of the 32nd annual ACM symposium on user interface software and technology*, pp. 1185–1195, 2019.

[11] R. Jacob and S. Stellmach. What you look at is what you get: Gaze-based user interfaces. *Interactions*, 23(5):62–65, aug 2016. doi: 10.1145/2978577

[12] S. Jang, W. Stuerzlinger, S. Ambike, and K. Ramani. Modeling cumulative arm fatigue in mid-air interaction based on perceived exertion and kinetics of arm motion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 3328–3339, 2017.

[13] K. Kim, R. L. Lawrence, N. Kyllonen, P. M. Ludewig, A. M. Ellingson, and D. F. Keefe. Anatomical 2d/3d shape-matching in virtual reality: A user interface for quantifying joint kinematics with radiographic imaging. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 243–244, 2017. doi: 10.1109/3DUI.2017.7893362

[14] C.-Y. Lee, W.-A. Hsieh, D. Brickler, S. V. Babu, and J.-H. Chuang. Design and empirical evaluation of a novel near-field interaction metaphor on distant object manipulation in vr. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*, SUI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3485279.3485296

[15] C. Lim, J. Kim, and M. J. Kim. Thumble: One-handed 3d object manipulation using a thimble-shaped wearable device in virtual reality. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology*, UIST '22 Adjunct. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3526114.3558703

[16] C. Liu, A. Plopski, and J. Orlosky. Orthogaze: Gaze-based three-dimensional object manipulation using orthogonal planes. *Computers*

& Graphics, 89:1–10, 2020.

[17] M. Liu, M. Nancel, and D. Vogel. Gunslinger: Subtle arms-down mid-air interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pp. 63–71, 2015.

[18] X. Liu, S. Luan, L. Wang, and C.-T. Lam. Manipulation guidance field for collaborative object manipulation in vr. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 713–714, 2023. doi: 10.1109/VRW58643.2023.00199

[19] X. Liu, L. Wang, S. Luan, X. Shi, and X. Liu. Distant object manipulation with adaptive gains in virtual reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 739–747, 2022. doi: 10.1109/ISMAR55827.2022.00092

[20] J. S. Maxwell and C. M. Schor. The coordination of binocular eye movements: Vertical and torsional alignment. *Vision Research*, 46(21):3537–3548, 2006. doi: 10.1016/j.visres.2006.06.005

[21] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge. A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments. In *Computer graphics forum*, vol. 38, pp. 21–45. Wiley Online Library, 2019.

[22] D. Mendes, F. Relvas, A. Ferreira, and J. Jorge. The benefits of dof separation in mid-air 3d object manipulation. In *Proceedings of the 22nd ACM conference on virtual reality software and technology*, pp. 261–268, 2016.

[23] P. Monteiro, G. Gonçalves, H. Coelho, M. Melo, and M. Bessa. Hands-free interaction in immersive virtual reality: A systematic review. *IEEE Transactions on Visualization and Computer Graphics*, 27(5):2702–2713, 2021. doi: 10.1109/TVCG.2021.3067687

[24] K. Pfeuffer, J. Alexander, M. K. Chong, and H. Gellersen. Gaze-touch: combining gaze with multi-touch for interaction on the same surface. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pp. 509–518, 2014.

[25] K. Pfeuffer, B. Mayer, D. Mardanbegi, and H. Gellersen. Gaze+ pinch interaction in virtual reality. In *Proceedings of the 5th symposium on spatial user interaction*, pp. 99–108, 2017.

[26] J. S. Pierce, B. C. Stearns, and R. Pausch. Voodoo dolls: seamless interaction at multiple scales in virtual environments. In *Proceedings of the 1999 symposium on Interactive 3D graphics*, pp. 141–145, 1999.

[27] M. Pouke, A. Karhu, S. Hickey, and L. Arhippainen. Gaze tracking and non-touch gesture based interaction method for mobile 3d virtual spaces. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, pp. 505–512, 2012.

[28] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. The go-go interaction technique: non-linear mapping for direct manipulation in vr. In *Proceedings of the 9th annual ACM symposium on User interface software and technology*, pp. 79–80, 1996.

[29] Y. Y. Qian and R. J. Teather. The eyes don't have it: An empirical comparison of head-based and eye-based selection in virtual reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI '17, p. 91–98. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3131277.3132182

[30] W. Robinett and R. Holloway. Implementation of flying, scaling and grabbing in virtual worlds. In *Proceedings of the 1992 symposium on Interactive 3D graphics*, pp. 189–192, 1992.

[31] K. Ryu, J.-J. Lee, and J.-M. Park. Gg interaction: a gaze–grasp pose interaction for 3d virtual object selection. *Journal on Multimodal User Interfaces*, 13(4):383–393, 2019.

[32] M. Serrano, B. Ens, X.-D. Yang, and P. Irani. Gluey: Developing a head-worn display interface to unify the interaction experience in distributed display environments. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 161–171, 2015.

[33] L. Sidenmark and H. Gellersen. Eye, head and torso coordination during gaze shifts in virtual reality. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 27(1):1–40, 2019.

[34] L. Sidenmark and H. Gellersen. Eye&head: Synergetic eye and head movement for gaze pointing and selection. In *Proceedings of the 32nd annual ACM symposium on user interface software and technology*, pp. 1161–1174, 2019.

[35] L. Sidenmark, D. Mardanbegi, A. R. Gomez, C. Clarke, and H. Gellersen. Bimodalgaze: Seamlessly refined pointing with gaze and filtered gestural head movement. In *ACM Symposium on Eye Tracking Research and Applications*, pp. 1–9, 2020.

[36] A. L. Simeone, A. Bulling, J. Alexander, and H. Gellersen. Three-point interaction: Combining bi-manual direct touch with gaze. In *Proceedings of the international working conference on advanced visual interfaces*, pp. 168–175, 2016.

[37] D. Slambekova, R. Bailey, and J. Geigel. Gaze and gesture based object manipulation in virtual worlds. In *Proceedings of the 18th ACM symposium on Virtual reality software and technology*, pp. 203–204, 2012.

[38] S. Stellmach and R. Dachselt. 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*, pp. 285–294, 2013.

[39] S. Stellmach, S. Stober, A. Nürnberger, and R. Dachselt. Designing gaze-supported multimodal interactions for the exploration of large image collections. In *Proceedings of the 1st conference on novel gaze-controlled applications*, pp. 1–8, 2011.

[40] R. Stoakley, M. J. Conway, and R. Pausch. Virtual reality on a wim: interactive worlds in miniature. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 265–272, 1995.

[41] A. A. Taha and A. Hanbury. An efficient algorithm for calculating the exact hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(11):2153–2163, 2015. doi: 10.1109/TPAMI.2015.2408351

[42] E. Triantafyllidis, W. Hu, C. McGreavy, and Z. Li. Metrics for 3d object pointing and manipulation in virtual reality: The introduction and validation of a novel approach in measuring human performance. *IEEE Robotics Automation Magazine*, 29(1):76–91, 2022. doi: 10.1109/MRA.2021.3090070

[43] J. Turner, J. Alexander, A. Bulling, and H. Gellersen. Gaze+ rst: integrating gaze and multitouch for remote rotate-scale-translate tasks. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 4179–4188, 2015.

[44] J. Turner, J. Alexander, A. Bulling, D. Schmidt, and H. Gellersen. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. In *Human-Computer Interaction–INTERACT 2013: 14th IFIP TC 13 International Conference, Cape Town, South Africa, September 2-6, 2013, Proceedings, Part II 14*, pp. 170–186. Springer, 2013.

[45] J. Turner, A. Bulling, J. Alexander, and H. Gellersen. Eye drop: an interaction concept for gaze-supported point-to-point content transfer. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*, pp. 1–4, 2013.

[46] J. Turner, A. Bulling, J. Alexander, and H. Gellersen. Cross-device gaze-supported point-to-point content transfer. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 19–26, 2014.

[47] J. Turner, A. Bulling, and H. Gellersen. Combining gaze with manual interaction to extend physical reach. In *Proceedings of the 1st international workshop on pervasive eye tracking & mobile eye-based interaction*, pp. 33–36, 2011.

[48] M. Veit, A. Capobianco, and D. Bechmann. Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, pp. 51–58, 2009.

[49] E. Velloso, J. Turner, J. Alexander, A. Bulling, and H. Gellersen. An empirical investigation of gaze selection in mid-air gestural 3d manipulation. In *Human-Computer Interaction–INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14-18, 2015, Proceedings, Part II 15*, pp. 315–330. Springer, 2015.

[50] S. Voelker, S. Hueber, C. Holz, C. Remy, and N. Marquardt. Gaze-conduits: Calibration-free cross-device collaboration through gaze and touch. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–10, 2020.

[51] L. Wang, X. Liu, and X. Li. Vr collaborative object manipulation based on viewpoint quality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 60–68, 2021. doi: 10.1109/ISMAR52148.2021.00020

[52] J. Wentzel, G. d'Eon, and D. Vogel. Improving virtual reality er-

gonomics through reach-bounded non-linear input amplification. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2020.

[53] C. Wilkes and D. A. Bowman. Advantages of velocity-based scaling for distant 3d manipulation. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pp. 23–29, 2008.

[54] D. Yu, H.-N. Liang, K. Fan, H. Zhang, C. Fleming, and K. Papangelis. Design and evaluation of visualization techniques of off-screen and occluded targets in virtual reality environments. *IEEE transactions on visualization and computer graphics*, 26(9):2762–2774, 2019.

[55] D. Yu, X. Lu, R. Shi, H.-N. Liang, T. Dingler, E. Velloso, and J. Goncalves. Gaze-supported 3d object manipulation in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2021.