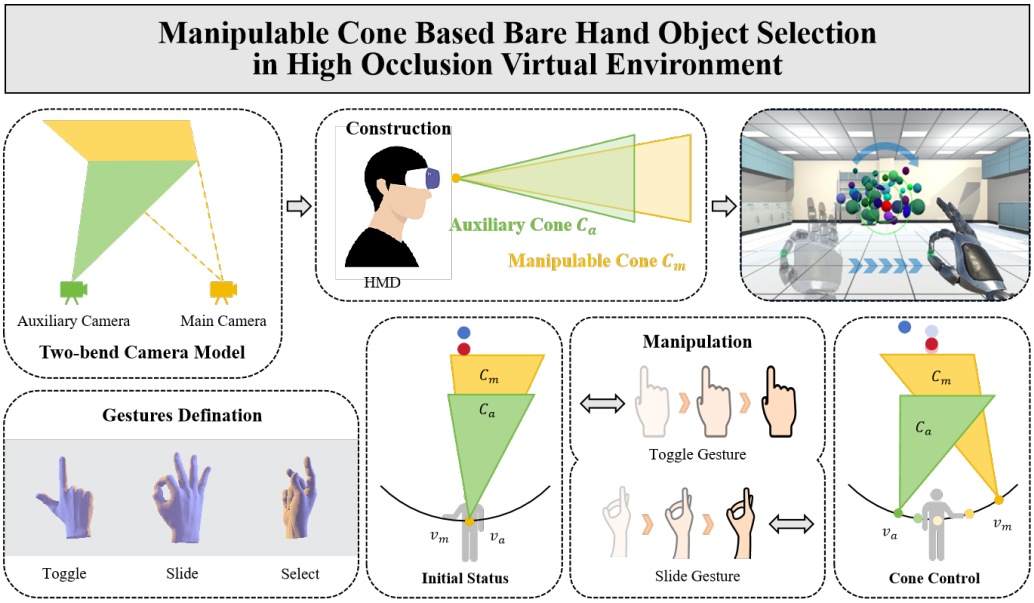
Graphical Abstract

Manipulable Cone Based Bare Hand Object Selection

in High Occlusion Virtual Environment

Jiaqi Zhou, Jian Wu, Sio Kei Im, Runze Fan, Lili Wang



Highlights

Manipulable Cone Based Bare Hand Object Selection in High Occlusion Virtual Environment

Jiaqi Zhou, Jian Wu, Sio Kei Im, Runze Fan, Lili Wang

- We propose the manipulable cone which is constructed based on the two-bend camera model. The user is allowed to manipulate the objects inside the cone, removing the occlusion before the target object in high occlusion scenes.
- We propose a new bare hand interaction design. We define three gestures: Toggle, Slide, and Select. We define five states: MCBOS OFF, Initial Status, Cone Control, Inter Status, and Selection Process. The user can use gestures to smoothly control the switching between states and select the occluded object.
- Two user studies to evaluate the efficiency, utility, and adaptability of our proposed method. The results indicate that in high occlusion environments, MCBOS significantly reduces Task Completion Time, Total Rotation, Total Translation, Error Rate, Task Load, and Simulator Sickness Questionnaire scores.

Manipulable Cone Based Bare Hand Object Selection in High Occlusion Virtual Environment

Jiaqi Zhou[a,b], Jian Wu[a], Sio Kei Im[c], Runze Fan[a], Lili Wang[a,b,*]

[a]*State key laboratory of virtual reality technology and systems, Beihang University, Beijing, 100191, Beijing, China*
[b]*Peng Cheng Laboratory, Shenzhen, 518052, Guangzhou, China*
[c]*Macau Polytechnic University, Macau, 999078, Macau, China*

Abstract

We propose a manipulable cone based bare hand object selection method in a high occlusion virtual environment. First, a new concept of the manipulable cone is introduced, which can be used to control the manipulable cone to remove the occlusion before the target object. Next, we propose its construction and manipulation methods. It includes initialization and manipulation of its motion using the bare hand. Finally, a bare hand interaction design based on manipulable cones for object selection is proposed. The user does not need a handheld device and can realize the task of object selection using only his/her hands. During the selection process, three gestures, and five states are defined, allowing the user to use the three gestures to smoothly switch between the five states and complete the selection task. We designed two user studies to evaluate the efficiency, utility, and adaptability of our method. The results proved that our method significantly reduces task completion time, total rotations, total translations, error rate, task

*Corresponding author
Email addresses: zhoujq@buaa.edu.cn (Jiaqi Zhou), lanayawj@buaa.edu.cn (Jian Wu), marcusim@mpu.edu.mo (Sio Kei Im), by2106131@buaa.edu.cn (Runze Fan), wanglily@buaa.edu.cn (Lili Wang)

load, and Simulator Sickness Questionnaire (SSQ) scores in high occlusion environments compared to existing state-of-the-art (SOTA) bare hand methods, and also significantly improves the User Experience Questionnaire (UEQ) score and performance ranking.
Keywords: Virtual reality, Object selection, Interaction design, Bare hand, High occlusion scene

1. Introduction

With the continuous advancement of technology, virtual reality (VR) has been widely applied in fields such as smart manufacturing, medicine, and gaming. Object selection, a fundamental operation that enables users to acquire objects for further manipulation or interaction, is the most commonly used interaction method in VR. For instance, in smart manufacturing, workers can use VR to select and operate various machines and equipment for production line planning and simulation. The efficiency and utility of object selection methods significantly impact the user interaction experience during VR tasks, with high selection efficiency and utility enhancing the fluidity and naturalness of VR operations.

Currently, many researchers have carried out studies on the task of object selection in VR environments. Yang Li et al. proposed a target selection method using handheld controllers in VR environments (Li et al., 2022). This method utilizes real-life sewing as a design metaphor, allowing users to select targets through a single spatial motion of the handheld controller penetrating the target, thereby enhancing selection efficiency. To improve the accuracy of ray selection, Kaiqi Chen et al. backtracked the ray interaction history to when the desired target was highlighted and selected the target object (Chen et al., 2023). With the gradual

development of head-mounted display (HMD) technology, bare hand selection methods have been widely studied. Hai-Ning Liang et al. addressed a bare hand selection method for VR environments, proposing and developing two bare hand directional selection methods to achieve faster and more accurate selection performance (Shi et al., 2023). To date, numerous object selection methods have been proposed and proven effective. However, research on bare-hand object selection remains limited, especially in highly occluded environments. Existing methods such as 3D-Bubble (Vanacken et al., 2007) and Multi-Bubble (Delamare et al., 2022), while providing solutions, often struggle with precision in complex scenes. These methods either lead to unintentional selection of nearby objects or require the user to make frequent selections, which imposes a cognitive load and destroys immersion. To address these issues, there is much room for improvement in the selection efficiency and utility of bare hand selection techniques.

In this paper, we propose a manipulable cone based bare hand object selection method (MCBOS) in a high occlusion virtual environment. First, we introduce a new concept of the manipulable cone, which can be used to manipulate the subspace and the objects inside the cone to reduce occlusion. We also propose its construction and manipulation methods. Second, we designed a bare hand interaction process, which the user does not need a handheld device and can realize the task of object selection using only their hands based on manipulable cone. We used three main gestures to switch between the five states during the selection process. In the cone control state, hand translation is used to manipulate the apex of the manipulable cone, thus manipulating the virtual objects in the cone, exposing the target objects occluded by these objects, and then performing the target selection. We conducted two user studies to evaluate the efficiency, utility, and adaptabil-

ity of MCBOS. Compared to the existing SOTA methods, 3D-Bubble (Vanacken et al., 2007), Multi-Bubble (Delamare et al., 2022), our MCBOS achieves significant reductions in the task completion time, total rotation, total translation, error rate, task load and SSQ score in high occlusion environments. Our MCBOS also significantly improves the UEQ score and performance ranking. Thus, MCBOS significantly increases the efficiency, utility, and adaptability in high occlusion environments. Figure 1 compares the 3D-Bubble, Multi-Bubble, and our MCBOS in a selection process under high occlusion conditions in a practical application scenario. By manipulating the manipulable Cone, MCBOS realizes a fast occlusion removal of the object and completes the selection task of the corresponding object, which shows a remarkably excellent performance compared to the previous methods.

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

- We introduce a new concept, the manipulable cone, for manipulating space and objects in space and removing occlusions in target selection tasks.
- We design a manipulable cone based object selection interaction with bare hands to help users efficiently select virtual objects in highly occluded virtual environments.
- We conduct two user studies to evaluate our method's efficiency, utility, and adaptability.

2. Related work

In this section, we first review controller based object selection methods. Then, we describe bare hand based object selection methods. Finally, we discuss the se-

Figure 1: Comparison of 3D-Bubble, Multi-Bubble and MCBOS under high occlusion selection condition. When the target orange box is partially obscured by the brown box, making only a small portion visible, different selection methods exhibit varying performance. **3D-Bubble (left):** The user needs to finely control the position of the 3D-Bubble with his hand to realize the accurate selection of the visible region, and thus select the orange box. **Multi-Bubble (middle):** The four bubbles controlled by the user's four fingers are applied to both the occluder brown box and the occluded orange box as candidates. However, due to the density of the objects, it is difficult to see the position and connectivity of the bubbles controlled by different fingers, making it hard to select the target. **MCBOS (right):** The user can manipulate the manipulable cones on the screen through gestures, thereby exposing the orange box hidden between the two brown boxes to a larger area of our field of view during the selection process, which facilitates the selection.

lection method in highly occluded virtual scenes, which is currently more complex in the selection tasks, and explore whether existing methods can efficiently handle selection tasks in highly occluded VR environments. For a more comprehensive understanding of object selection, we recommend that readers refer to the survey paper (Bergström et al., 2021).

2.1. Controller based object selection

Controller based selection methods are a popular and dominant component of VR target selection design, primarily employing the controller as the point of selection and the rational operation of the controller as the trigger for selection (de Haan et al., 2005; Grossman and Balakrishnan, 2006a; Kopper et al., 2011; Cashion et al., 2012; Ro et al., 2017). Lu et al. proposed a bubble mechanism for ray-casting selection (Lu et al., 2020), which allows the user to minimize errors due to hand shaking when selecting objects. Baloup et al. proposed RayCursor for selection (Baloup et al., 2019), which enables the user to move the cursor through the controller's touchpad to select objects at a deeper level. Wu et al. introduced ClockRay, an improved version of RayCursor, which controls the depth of the cursor in virtual space by adding wrist rotation to select occluded objects (Wu et al., 2023). Chen et al. sought to improve the accuracy of ray selection by backtracking the history of ray interactions to the time when the desired target was highlighted and selecting that target (Chen et al., 2023). Additionally, the combination of controller and gaze is also one of the current mainstream selection methods. Specifically, the user performs the pointing of the selection through the controller, while further correcting the pointing in conjunction with the gaze data, thus realizing the object selection task (Lystbæk et al., 2022b,a; Wagner et al., 2023; Wei et al., 2023). Despite the controller's advantages in terms of accuracy

and depth control, it has disadvantages in terms of immersion, learning cost, device burden, and naturalness compared to bare hand interaction. Currently, with the advancement of technology, the application of bare hand techniques has gradually replaced controllers to reduce dependency on external devices. However, algorithm research on object selection using controllers has provided significant insights for bare hand studies.

2.2. Bare hand based object selection

With the technological advancement and widespread adoption of HMD, providing users with more natural and intuitive interaction methods has become one of the core objectives. Particularly, the progress in barehanded recognition technology enables users to interact with the virtual world directly through their hand movements and gestures without the need for any external devices, significantly enhancing the sense of immersion and user experience. Consequently, an increasing number of bare hand selection methods are being proposed by researchers (Shi et al., 2023; Kim and Xiong, 2022). Bhowmick et al. proposed a technique based on continuous multi-finger zooming, where the selection of the target object is accomplished through the cooperation of the right and left hands (Bhowmick et al., 2023). Schjerlund et al. developed a target selection technique that reduces the distance to the selected target (Schjerlund et al., 2021). Delamare et al. improved upon the technique of Vanacken et al. by selecting a volume that includes multiple targets, thus increasing the effective width of the target to improve the selection performance (Delamare et al., 2022; Vanacken et al., 2007). Kim et al. proposed virtual panel interaction, which enables faster and more accurate target selection (Kim and Xiong, 2022). Although these bare hand object selection methods have demonstrated good performance in object selection tasks and have been validated,

there is still significant potential for improvement in their selection efficiency and utility.

2.3. Object selection in high occlusion scenes

Selection tasks in high-occlusion environments represent a common challenge in target selection within VR and have garnered considerable research effort. For selection tasks involving partially occluded object scenes, the BalloonProbe technique allows users to place a balloon among the objects, which moves the objects towards its surface along the radius, thereby reducing the occlusion of VR object groups(Elmqvist, 2005; Elmqvist and Tudoreanu, 2007). Grossman et al. and Baloup et al. have utilized ray-enhancement methods, employing depth markers fixed at certain positions on the ray or controlled via a touchpad, to select the object closest to the depth marker (Grossman and Balakrishnan, 2006b; Baloup et al., 2019). Sidnmark et al. introduced the "Contour Tracking" technique, enabling the selection of partially occluded VR objects by tracking their moving contours with gaze tracking(Sidenmark et al., 2020). For selection tasks in highly occluded or completely invisible scenes, Wang and Wu proposed two rendering techniques for detecting potential occlusion views and rendering occlusion views revealed by user selection(Wang et al., 2019; Wu et al., 2022). Li et al. introduced vMirror, an interactive component that facilitates the observation of completely occluded objects by correctly positioning the component(Li et al., 2021). Maslych et al. combined methods including cone projection, mini-world, and the grasping metaphor to enhance the selection efficiency of semi-occluded and fully occluded objects(Maslych et al., 2023). The aforementioned methods involve using controllers to interact with VR spaces in highly occluded environments. Liang et al. designed a dual bare-hand selection method for object selection in highly oc-

cluded environments (Shi et al., 2023). Compared to dual hand selection method, single hand selection method allows the other hand to remain free, which is advantageous when other tasks, such as holding objects or operating additional tools or equipment, are required. Our method is a single bare hand selection method designed for object selection tasks in highly occluded virtual environments.

3. Design rationale

Our goal is to design a bare hand selection method in highly occluded environments that makes the task of object selection in highly occluded environments more efficient and accurate, and the bare-handed interaction more natural and smooth. We will introduce our design rationales in detail regarding the naturalness of bare hand interaction and the efficiency and accuracy of the selection process, respectively.

3.1. Natural bare hand interaction with a smooth selection process

This rationale is based on three core elements: bare-handed operation, natural and convenient gesture use, and easy gesture differentiation and transition. Together, these elements ensure that using MCBOS is as intuitive and fluid as possible.

Our design emphasizes interaction without the use of any external devices. Users do not need to wear special gloves or hold any physical devices and can perform all operations with their bare hands alone. This approach not only reduces the cost of use, but also makes the approach more generalizable and acceptable because it rejects the physical constraints of traditional interaction interfaces and allows the user to naturally interact with the virtual environment at any time. In designing our control gestures, we drew on common intuitive gestures people use

in their daily lives, such as tapping on the screen and pinching objects, as inputs to ensure that the gestures are easy to manipulate. This design allows users to familiarize themselves with our gestures with very little learning. Finally, we should try to ensure that the differences between the transformed gesture changes are small enough to enable smooth transitions between different gestures.

3.2. More efficiency and accuracy in the selection process

To accomplish this, we approached the design of our approach from three perspectives: increasing the visibility of occluded objects, reducing the number of gestures used in the selection process, and reducing user body and head movement during selection.

We found that by increasing the occluded object's exposed area, users can easily accomplish the task of selecting occluded objects in complex environments. For this reason, our design ensures that the user can realize the function of removing the occlusion in front of the object, so that the object to be selected has a larger visible area in the field of view, and the selection operation can be completed in that state. Our design emphasizes the use of as few gestures as possible to accomplish selection tasks in order to increase the efficiency and intuitiveness of the operation. For example, users can initiate our method with a simple tap action, or quickly achieve the de-occlusion effect by sliding a pinch gesture. We recognize that frequent body or head movements while making object selections can lead to user fatigue and even discomfort. Therefore, our design ensures that the user moves their body and head as little as possible when making selections. Complex selections can be accomplished with simple gestures with little or no movement of the user's body or head.

4. MCBOS

This section describes the basic concept, interaction design, and visual cues of the MCBOS method, a new approach for rapidly selecting occluded objects in VR environments. Firstly, we introduce the detailed definition, construction, and manipulation of the manipulable cone. Secondly, we outline its interaction design including the gesture definition, mapping, and interaction process. Finally, we present the visual cue component, designed to aid the navigation and selection process by providing clear, color-coded cue messages.

4.1. Manipulable Cone

4.1.1. Definition

Based on our design rationales, we define the manipulable cone C_m (p_m,r_m,v_m) to provide the user with a larger imaging footprint of the target object in the user's view. The manipulable cone C_m includes three parameters: p_m,r_m, and v_m. p_m denotes the center point of the cone base of C_m. r_m denotes the radius of the cone base of C_m. v_m represents the apex position of the C_m.

4.1.2. Construction and Manipulation

Construction. To ensure that objects within the manipulable cone can be manipulated, we adopted the two-bend camera model (Popescu et al., 2009; Wang et al., 2021). The two-bend camera model has two view frustums, corresponding to our manipulable cone C_m and the auxiliary cone C_a. The positions of the two cameras correspond to the apex positions of the manipulable cone C_m and the auxiliary cone C_a, respectively. The auxiliary cone C_a also includes three parameters. p_a denotes the center point of the cone base of C_a. r_a denotes the radius of the cone base of C_a. v_a represents the apex position of the auxiliary cone C_a.

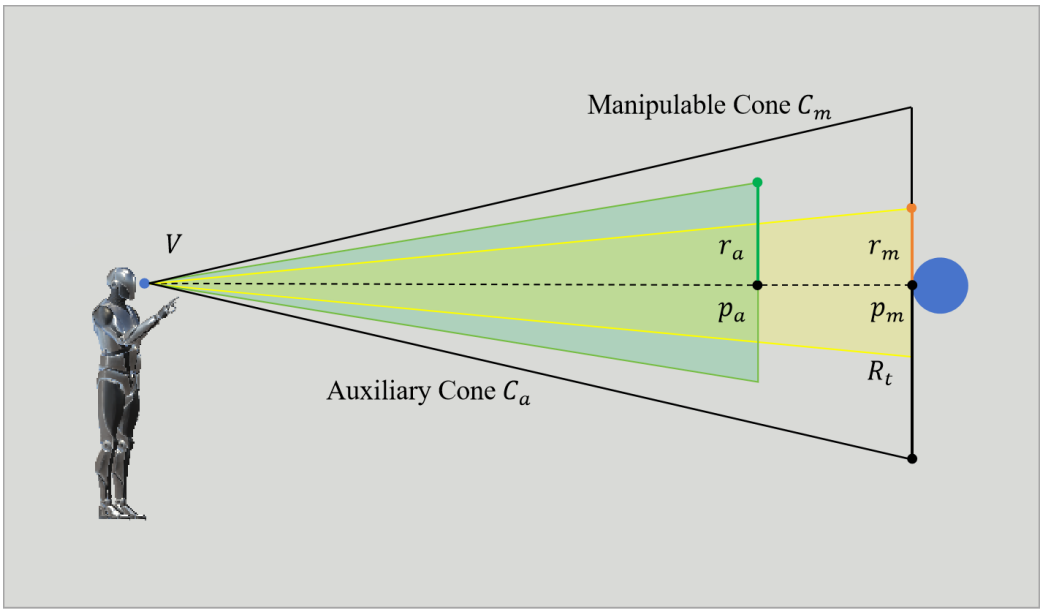


Figure 2: Definition and initialization of the manipulable cone. The yellow region represents the Manipulable Cone C_m, while the green region denotes the Auxiliary Cone C_a. The manipulable cone C_m is characterized by three key parameters: p_m (position), r_m (radius), v_m (vertex). Similarly, the auxiliary cone C_a is defined by three parameters: p_a (position), r_a (radius), v_a (vertex).

The initialization and construction of C_m and C_a are illustrated in Figure 2. In the Figure 2, V is our main viewpoint, and R_t is the maximum radius of the light path bending region. In the initial state, v_m and v_a are positioned at the same location as the main viewpoint V. By default, p_m and p_a are constructed 20 meters and 16 meters directly in front of the main viewpoint, respectively. r_m is set to 0.08 screen height by default and R_t is set to 0.14 screen height by default. The value of r_a is calculated and is described in the following sections.

Manipulation. After the manipulable cone C_m and the auxiliary cone C_a are initialized, we need to control the movement of the apex v_m of C_m to achieve the desired de-occlusion effect. When v_m is moved, v_a also moves accordingly. The relationship between the positions of v_m and v_a is illustrated in Figure 3.

Π_p is the plane perpendicular to the direction of the main viewpoint passing through the point of p_m. Π_{1r} is the plane containing the main viewpoint V and the apices of the two cones C_m and C_a. Π_{2r} and Π_a are two intermediate planes located between Π_p and Π_{1r}, both perpendicular to the direction of the main viewpoint. The relationship between the distances of these two planes to the main viewpoint V and p_m is shown in Equation 1.

$$\begin{cases} d(\Pi_a, V) = 0.8 \times \|p_m V\| \\ d(\Pi_{2r}, V) = 0.5 \times \|p_m V\| \end{cases} \tag{1}$$

In this equation. $\|p_m V\|$ denotes the distance from p_m to V. $d(\Pi_a, V)$ denotes the distance from V to the plane Π_a. $d(\Pi_{2r}, V)$ denotes the distance from V to the plane Π_{2r}. r_m is a user-specific parameter that can be adjusted by the user while using MCBOS. Once p_m, r_m, and v_m are determined, C_m can be successfully constructed.

When C_m is determined, the position of p_a can be determined by p_m, v_m, and

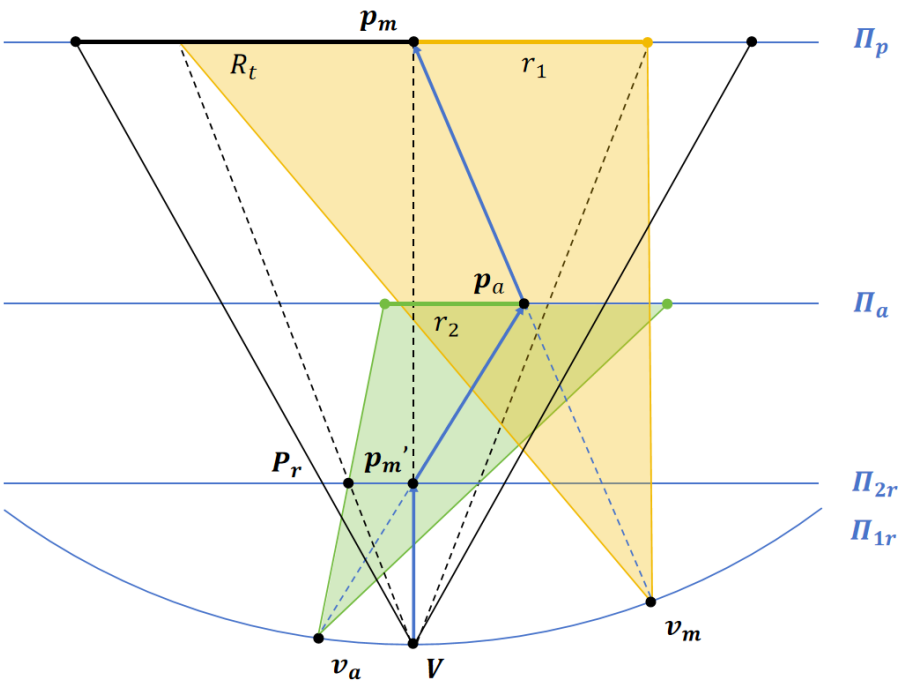


Figure 3: Construction of the manipulable cone C_m and the auxiliary cone C_a in MCBOS for practical applications. The yellow area indicates the actual construction area of C_m after the apex v_m moves. The green area indicates the constructed area of C_a. Through the movement of C_m and C_a, the light propagation path of p_m in the figure changes from the original emission of p_m to V to emission from p_m via p_a, p_m' and finally to V. Thus, the bend of the light path is realized.

the Π_a plane. The position of p_a is the projected position of p_m in the C_m region in the Π_a plane. Once the position of p_a is determined, by the projection p'_m of p_m on Π_{2r} and p_a, we can calculate the position v_a of the C_a apex using Equation 2.

$$\begin{cases} \mathbf{v} = p_a - p'_m \\ \mathbf{r}(t) = p'_m + t\mathbf{v} \\ a = \mathbf{v} \cdot \mathbf{v} \\ b = 2\mathbf{v} \cdot (p'_m - p_m) \\ h = p_m - V \\ c = (p'_m - p_m) \cdot (p'_m - p_m) - h^2 \\ t = \mathrm{QES}(a,b,c) \\ v_a = p'_m + t\mathbf{v} \end{cases} \tag{2}$$

where $\mathbf{v}$ is the direction vector from p_a to p'_m. $\mathbf{r}(t)$ denotes the position of any point on the parameterized line segment. t is a variable, usually varying in some range (e.g., from 0 to 1), that controls movement from the start point to the endpoint. When $t = 0$, the position is at the start point; when $t = 1$, the position is at the endpoint. After bringing the parameterized line segment into the spherical equation, we obtain a quadratic equation in terms of t. QES denotes the quadratic equation solver for t. Finally the apex position v_a of our C_a is derived.

Using the positional relations among the C_a apex v_a, P_r, p'_m, and p_a, we can calculate the radius r_a of C_a using Equation 3.

$$r_a = \frac{d(p'_m, P_r) \cdot d(p'_m, p_a)}{d(v_a, p'_m)} \tag{3}$$

In the formula, P_r is the point where the edge of C_m is projected onto the Π_{2r} plane. $d(p'_m, P_r)$ denotes the distance from p'_m to P_r, $d(p'_m, p_a)$ denotes the

distance from p'_m to p_a, and $d(v_a, p'_m)$ denotes the distance from v_a to p'_m.

As described above, we have established the correspondence between the manipulable cone C_m and the auxiliary cone C_a. Given the position of C_m, we can determine the position of C_a in real-time. Next, we introduce how to obtain the position parameters of C_m during the operation process.

The position of p_m is the first parameter to be determined when we use the manipulable cone through gestures. The relationship between the position of p_m in space and the position of the hand P_w is shown in Figure 4.

In the Figure 4, P_w is the position of the hand-wrist joint, which follows the position of the hand in space. P_c is the collision point detected by collision in the direction of the main viewpoint towards the line of sight. h' is the distance between the collision point P_c and the main viewpoint V. Π_p is the plane perpendicular to the direction of the main viewpoint through the point P_c. P'_w is the projection point of the hand joint position P_w on Π_p. δ is the offset of p_m compared to P'_w on the Π_p plane, a value we provide in the subsequent section 5.2. Using Equation 4, we can derive the position of p_m on the Π_p plane.

$$p_m = P'_w + \delta \tag{4}$$

When the position of p_m is determined, we can construct Π_{1r}. The construction of the apex plane Π_{1r} for MCBOS in practical applications and the effect of distorting the light paths of the vertices of different objects are shown in Figure 5.

In space, Π_{1r} is part of a sphere centered at p_m with radius h. h denotes the distance from the main viewpoint V to p_m. Once the sphere is determined and V, p_m, and the maximum rotation angle θ_m are known, we can achieve Π_{1r} by using Equation 5.

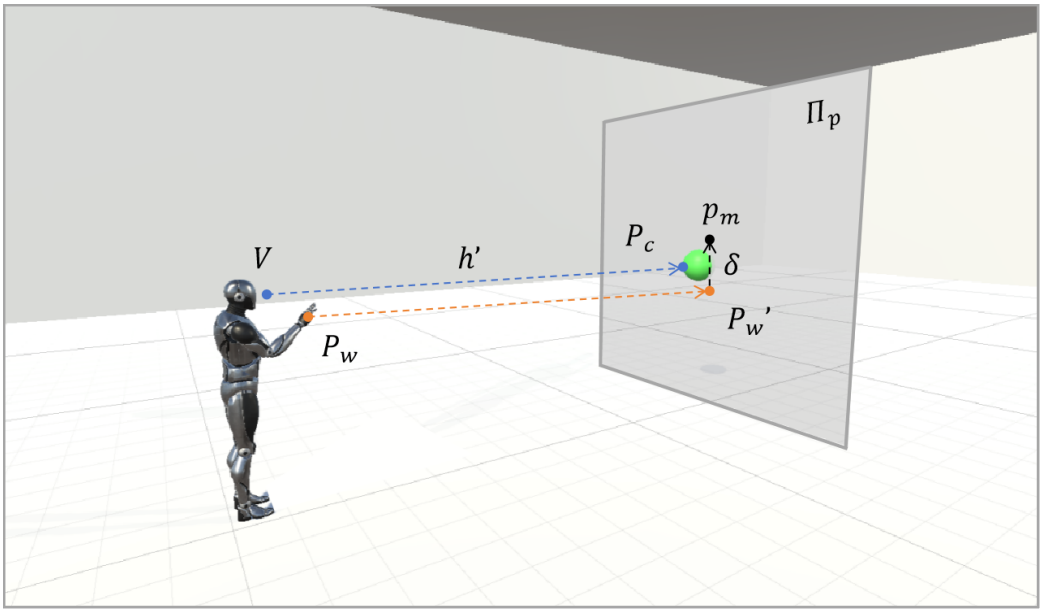


Figure 4: Determination of the position of p_m in space. The green object represents target along the user's line of sight in the 3D space. The spatial position of p_m in space is determined by a combination of the hand-wrist joint position P_w and the hyperparameter δ, which fine-tunes the projection of p_m on the plane Π_p. The blue dashed line indicates the line of sight from the user's viewpoint V to the target object, while the orange dashed line represents the projection path influenced by the user's hand movements.

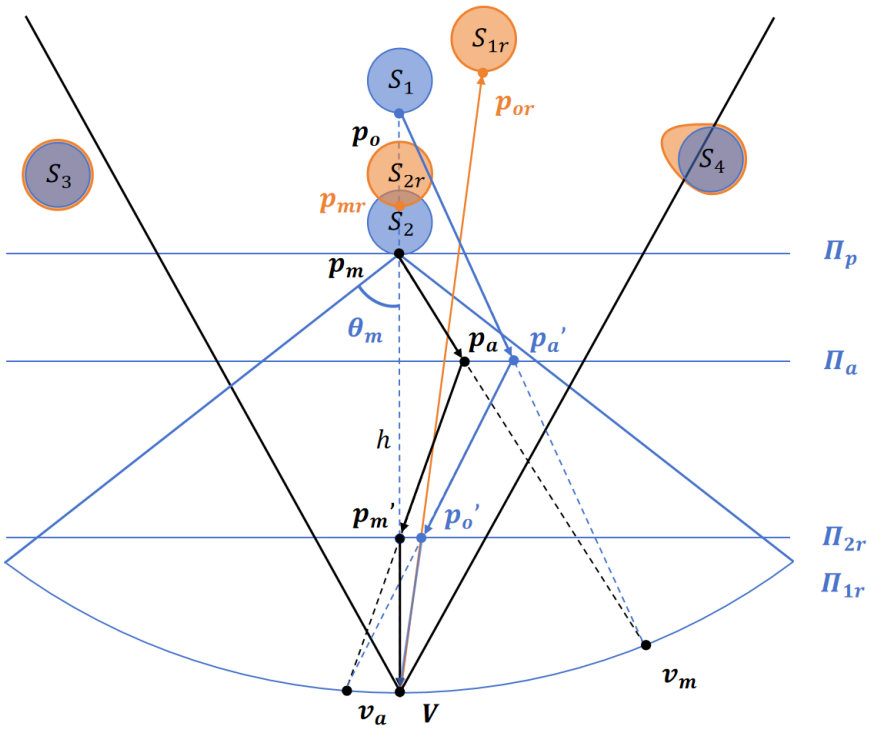


Figure 5: MCBOS optical path bending implementation on Π_{1r}. This figure demonstrates how MCBOS facilitates dis-occlusion by bending the optical path to reveal occluded objects. Without MCBOS, the light path of the fully occluded object S_1 would be completely blocked by the occluding object S_2. However, by utilizing MCBOS, the light path from S_1 is redirected through p'_a and projected onto the Π_{2r} plane at position p'_o, bypassing the obstruction caused by S_2.

$$\Pi_{1r} = C(p_m, V, \theta_m) \cap S(p_m, h) \tag{5}$$

In the formula, C denotes the cone, and S denotes the sphere. θ_m is also determined experimentally in the subsequent section 5.1. The construction of Π_{1r} can be understood as finding the intersection plane of a cone and a sphere. When Π_{1r} is successfully constructed, the user can control the movement of v_m apex on Π_{1r} by gestures to achieve the function of removing the occlusion in front of the target object. The relationship between the position of v_m on Π_{1r} and gestures is described in Section 4.2.3.

Figure 5 demonstrates the distortion of the object's light path and the change in its position after occlusion removal using the manipulable cone. In Figure 5, object S_1 is completely occluded by object S_2. As an example, we consider the light bending of two vertices on S_1, S_2. Using the manipulable cone, p_o on object S_1 is first bent on the Π_a plane with the bent point p'_a. Subsequently, p'_a is again bent on the Π_{2r} plane to the bent point p'_o. Thus, from the main viewpoint, the point p'_o that we see on the Π_{2r} plane has bypassed the occlusion of the object S_2 and is presented to the right of the bent point p'_m of S_2 on Π_{2r}. Using the Equation 6, we can derive the new position p_{or} of the point p_o in space after it has been bent, and determine the new position S_{1r} of the object S_1.

$$d(p_{or}, V) = d(p_o, p'_a) + d(p'_a, p'_o) + d(p'_o, V) \tag{6}$$

The bent point p'_m of object S_2 on Π_{2r} is aligned with p_m and V, ensuring that the planar position of object S_2 does not change in visual space. However, due to the bending of the light path, the new position S_{2r} of object S_2 after bending becomes farther away from the main viewpoint. Object S_3 is outside the range of

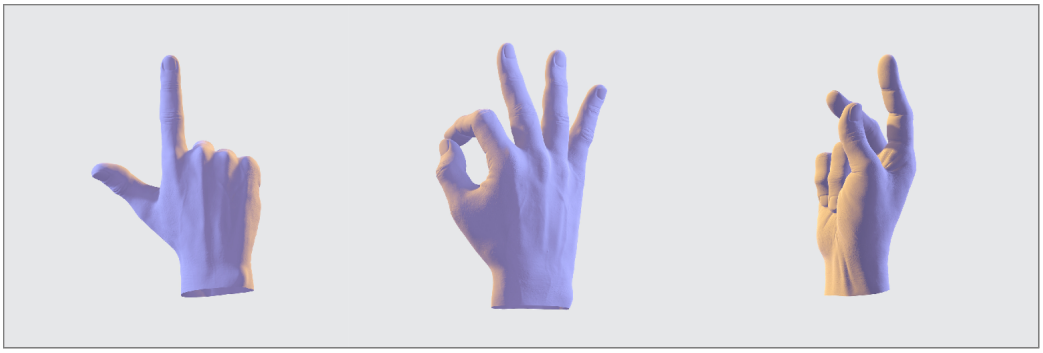


Figure 6: Gesture definitions for MCBOS manipulation. This figure illustrates the three distinct hand gestures designed for controlling the MCBOS. Toggle (left): Used to start and stop the MCBOS interaction process, functioning as an activation switch. Slide (middle): Enables the user to perform dis-occlusion by adjusting the manipulable cone, exposing hidden objects. Select (right): Controls the object selection process, allowing the user to confirm and finalize their selection.

the bent region we have set and therefore does not change. Object S_4 lies partially inside the bent range and partially outside, thus, only the vertices inside the range are elongated by the bending.

4.2. Manipulable cone based interaction design

4.2.1. Gesture definition

According to our design rationales, to achieve smooth control of MCBOS and a more natural gesture-switching flow, we define three gestures based on the pinching and moving of objects in the real world. These gestures are named Toggle, Slide, and Select, as shown in Figure 6.

- Toggle: Turn on/off the MCBOS

The Toggle gesture is used to control the start and stop of MCBOS. This gesture is defined as the user's thumb and index finger remaining straight, while the

other three fingers are curled inwards. When this gesture is detected for the first time, the manipulable cone is constructed. The positions of v_m, and v_a are aligned with V. The position of p_m changes with the movement of the hand position P_w, as shown in Figure 4. Conversely, MCBOS turns off when the gesture is detected again. This design allows the user to quickly turn MCBOS on or off as needed.

- Slide: Controls the apex position of v_m

The Slide gesture is used to lock the position of p_m and to control the apex position of v_m, as shown in Figure 3. This gesture is defined as the user's thumb and index finger being in a pinched state. When the Slide gesture is detected for the first time, it locks the position of p_m, i.e., it locks the area where we want to remove the occlusion. When the Slide gesture is consistently detected and the user maintains the gesture by swiping in a certain direction on the screen, the apex position v_m of C_m changes accordingly, thus removing the occlusion of the target region. The correspondence between the translation of the Slide gesture and the movement of the apex v_m is introduced in section 4.2.3.

- Select: Selection control for MCBOS

The Select gesture is used to complete the final selection interaction process. This gesture is defined as the pinching of the user's thumb and middle finger. When the Select gesture is detected, MCBOS enters the selection process. At this point, a selection reference point appears on the screen. We need to control this point to be located on our target object and then release the gesture to complete the selection.

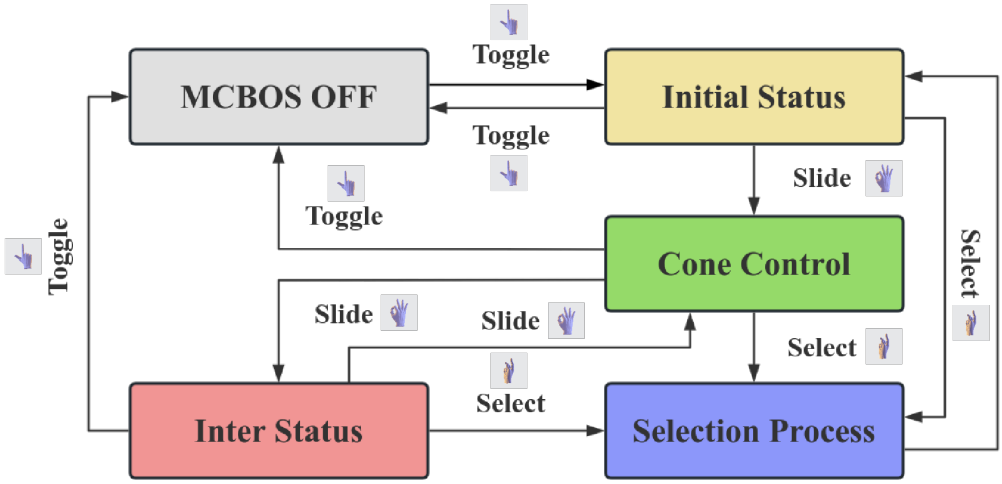


Figure 7: State transitions in the interaction process of MCBOS. This figure illustrates the transition conditions between various states in the MCBOS interaction workflow. The different colors represent distinct control states of MCBOS, with one-way arrows indicating the possible transitions between these states. Each transition is triggered by a specific gesture, as shown above the corresponding arrows.

4.2.2. Selection process

By combining the three gestures we defined with the functions corresponding to each gesture, we can achieve a complete MCBOS control flow. To facilitate better transitions between different gestures, we defined five states for MCBOS during the selection process: MCBOS OFF, Initial Status, Cone Control, Inter Status, and Selection Process. Through different gestures, we can achieve transitions between these states. The state transition diagram is shown in Figure 7.

In the figure, MCBOS OFF indicates that MCBOS is turned off, and the user cannot manipulate the manipulable cone. Initial Status indicates the initializa-

tion state, where the user can move their hand to change the position p_m of the manipulable cone C_m. Cone Control indicates the manipulation state, where the user is manipulating objects within the selected manipulable cone. Inter Status represents an intermediate state between the Cone Control and Selection Process, allowing the user to observe and adjust the effects within the manipulable cone. The Selection Process indicates the selection state, where the user completes the final selection of the target object.

In the MCBOS OFF state, the user can only use the Toggle gesture to switch between the Initial Status and this state. When in the Initial Status state, the user has three subsequent choices: 1) Switch to MCBOS OFF status with the Toggle gesture; 2) Enter the Cone Control status with the Slide gesture; 3) Enter the Selection Process status directly with the Select gesture. In the Cone Control state, the user can directly transition to the MCBOS OFF state using the Toggle gesture. By releasing the Slide gesture, the user can move to the Inter Status. Furthermore, the user can enter the Selection Process state by utilizing the Select gesture. While in the Inter Status state, the user can resume the Cone Control state by continuing to use the Slide gesture. Additionally, the user can employ the Select gesture to enter the Selection Process. The MCBOS can also be closed by utilizing the Toggle gesture. While in the Selection Process state, the user can exit this state solely by releasing the Select gesture. Upon exiting the Selection Process state, the control flow transitions directly to the Initial Status.

The basic process that users need to follow to complete a selection task using MCBOS is illustrated in the Figure 8. First, in the MCBOS OFF state, the user activates MCBOS by using the Toggle gesture. Next, in the Initial Status state, the user moves the manipulable cone to the desired area by adjusting their hand posi-

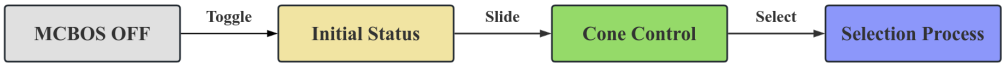


Figure 8: Basic selection process flow in MCBOS. This figure outlines the fundamental steps involved in the MCBOS selection process. The process follows a sequential flow, with each state transition triggered by specific gestures. By executing the process once, the user can realize the selection of a target object.

tion. Then, in the Cone Control state, the user performs dis-occlusion operations on the area using the Slide gesture. Finally, in the Selection Process state, the user completes the selection of the occluded target object by using the Select gesture.

4.2.3. Mapping

When manipulating the MCBOS to remove the occlusion, there exists a corresponding movement relationship between the Slide gesture and the apex v_m of the manipulable cone C_m. Through this correspondence, the user can use the Slide gesture to manipulate the position of the v_m to complete the de-occlusion manipulation of the target object. The correspondence of the Slide gesture to the movement of the apex v_m is shown in Figure 9.

When the Slide gesture is first detected, the error margin appears at the pinch point. r is the radius of the error margin region. P_i denotes the position where the Slide gesture was first detected. P_e denotes the positions that the Slide gestures subsequently moved to. $\overrightarrow{v}$ denotes the displacement velocity of the apex after being controlled by the gesture at that position. We can use Equation 7 and Equation 8 to determine the corresponding movement speed $\overrightarrow{v}$ of the v_m apex.

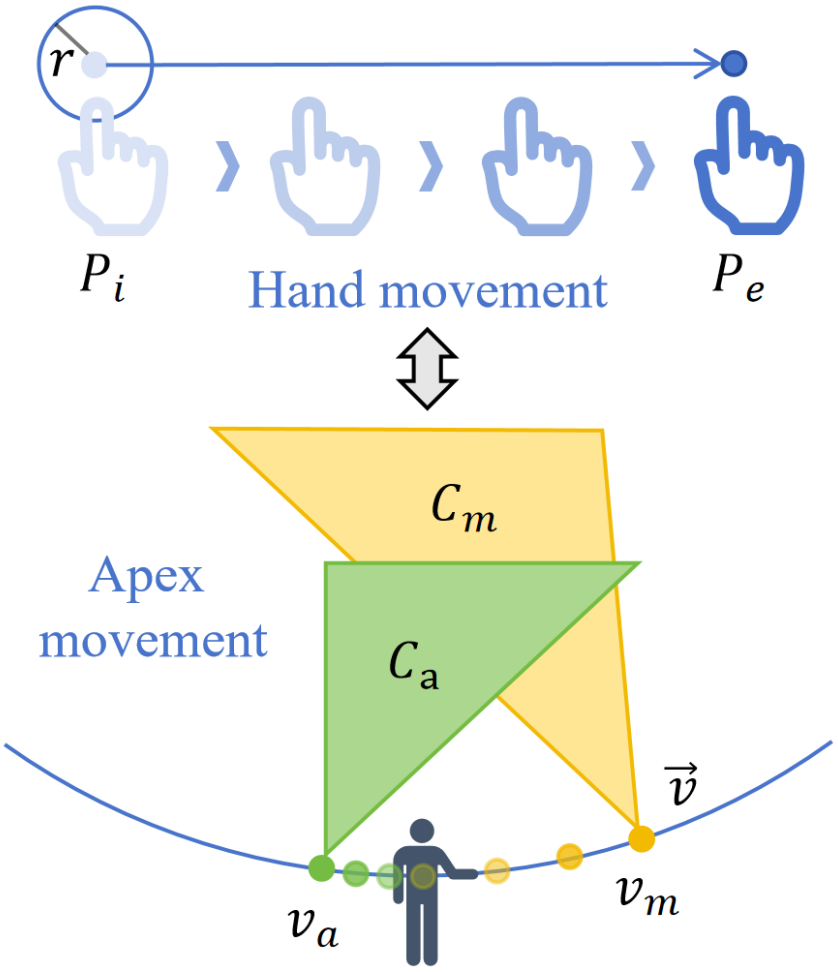


Figure 9: Mapping of Slide gesture movement to v_m apex movement. This figure illustrates the correspondence between the Slide gesture movement and the resulting movement of the v_m apex within the manipulable cone C_m. As the user slides their hand from the initial position P_i to the end position P_e, it triggers the movement of the apex V_m. The apex v_m shifts in the same direction as the hand movement, dynamically adjusting the shape and range of the manipulable cone C_m. The auxiliary cone C_a assists in refining the interaction by aligning with the apex movement.

$$\mathbf{d} = P_e - P_i \tag{7}$$

$$\begin{cases} \overrightarrow{v} = 0 & |\mathbf{d}| \leq r \\ \overrightarrow{v} = \alpha \cdot \frac{\mathbf{d}}{|\mathbf{d}|} & |\mathbf{d}| > r \end{cases} \tag{8}$$

In Equation 7, d represents the vector from the initial position P_i to the end position P_e. In Equation 8. α is a velocity increment constant that is used to represent the distance traveled by the apex in each frame. This parameter is used for user personalization and has a default value of 0.18. Users can personalize this parameter in a test environment before using MCBOS for the first time. Please refer to section 6.3 for more details about the personalization process.

4.3. Visual cues for MCBOS

Within the interaction process of VR, visual cues play an indispensable role. Optimal visual prompts, such as highlighting, outline rendering or halos, can significantly enhance the operational experience during use, facilitating users in swiftly completing each interaction task (Bowman et al., 2004). To this end, we have developed a set of visual cues for MCBOS. These visual cues provide information to the user by changing the color of the visualized area on the screen and the hand color according to the current state of MCBOS. As shown in Figure 10, we have designed four types of visual cues for MCBOS.

In the figure, when MCBOS is in the Initial Status state, the manipulable area on the screen is visualized in yellow, and the user's hand is also visualized in yellow. When MCBOS is in the Cone Control state, the manipulable area is visualized in green, and the user's hand is also visualized in green. Similarly, when MCBOS is in the Inter Status state, the manipulable area and the user's hand are

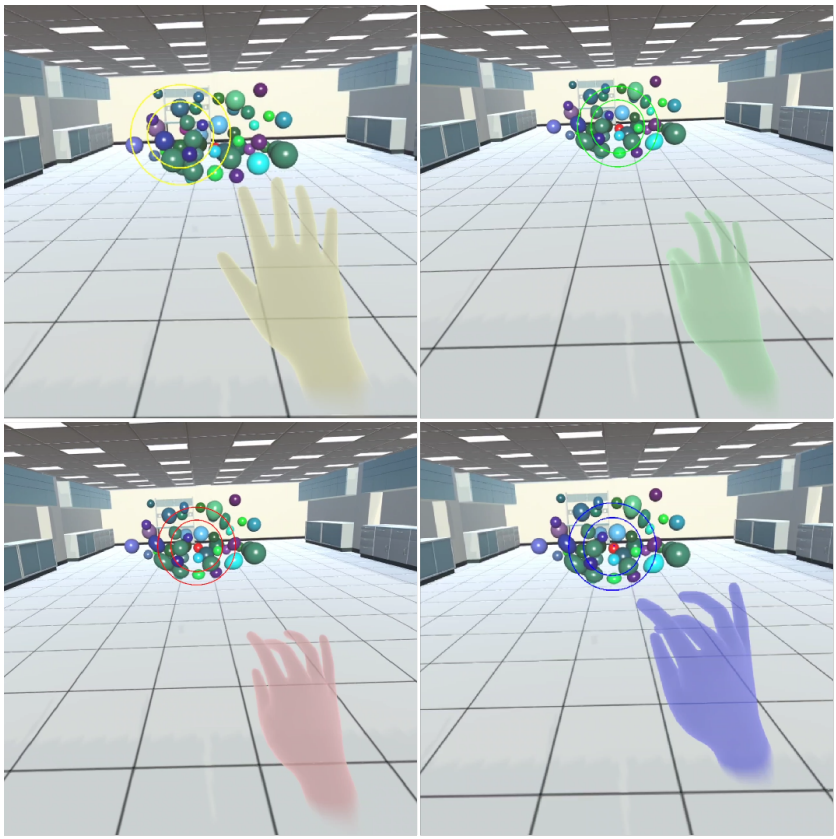


Figure 10: Visual cues corresponding to different MCBOS states. This figure illustrates the visual feedback provided in the four key states of the MCBOS interaction process, helping the user understand their current mode of operation. Initial Status (Top-Left): The hand appears in yellow, indicating the system is in its initial state, ready for activation but not yet manipulating objects. Cone Control (Top-Right): The hand changes to green, signifying that the user is actively manipulating the manipulable cone to adjust the field of view or dis-occlude hidden objects. Inter Status (Bottom-Left): The hand turns red, representing an intermediate state where the system awaits further input to transition to selection or cone control. Selection Process (Bottom-Right): The hand appears in blue, indicating the system is in the selection mode, allowing the user to finalize their target selection.

visualized in red. When MCBOS is in the Selection Process state, the manipulable area and the user's hand are visualized in blue.

5. Pilot user study

In this section, we define two subtasks to determine the maximum movement angle θ_m and the hyperparameter δ, as proposed in 4.1.2.

5.1. Pilot user study 1

This pilot user study aimed to determine the maximum movement angle θ_m, as shown in Figure 5. This parameter was used to construct the movement plane Π_{1r} for the apex v_m of C_m and to ensure that the user had an optimal de-occlusion effect when manipulating the apex v_m while keeping the original scene as tear-free as possible.

Participants. We recruited 16 participants for the experiment (8 males and 8 females, aged between 22 and 28 years, $M = 24.62$, $SD = 1.78$). All participants had normal vision without color blindness (14 of them wore glasses for short-sighted correction). Fourteen of them had experience with HMD VR applications, and two others were also in VR-related industries.

Apparatus. The experiment system was implemented based on the PICO 4 hand recognition technology and Unity version 2021.3.32 LTS. The workstation is equipped with a 5.60 GHz 13th Gen Intel(R) Core(TM) i9-13900F CPU, 64GB of RAM, and an NVIDIA GeForce GTX 4080 graphics card.

Task. The task is that participants have 5 buttons in front of them, indicating the 5 different maximum angles of movement of the manipulable cone, $\theta_m \in \{ 15°, 30°,$

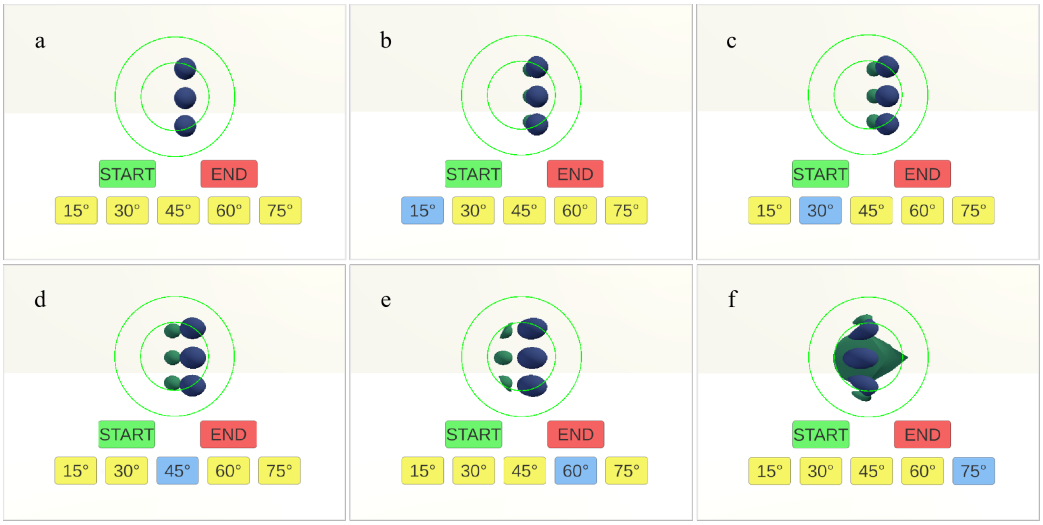


Figure 11: Experimental scenarios for pilot user study. **(a):** Shows the initial setup where three blue spheres are aligned side by side, with a green sphere positioned behind each blue sphere. Participants begin or end the experiment by clicking the START and END buttons, respectively. **(b, c, d, e, f):** Demonstrate the de-occlusion effect at the maximum rotation angles of 15°, 30°, 45°, 60°, and 75°, respectively.

45°, 60°, 75° }. The purpose is to select the maximum rotation angle θ_m, thus, our experimental scene is designed as a fully occluded scene and the manipulable area is also fixed to this occlusion area. The specific experimental scenario design is shown in Figure 11.

Procedure. First, to account for the different hand sizes of various users, our test system needs to initialize the hand recognition before use. PICO 4 provides this function internally. When participants are ready to start the experiment, they spread the five fingers of both hands with the palms of the hands facing the device and keep them still for 3 seconds, allowing the test system to complete the recog-

nition of the new hand binding. Next, participants clicked a button to rotate the manipulable cone to the appropriate angle and achieve occlusion removal. Participants were given five minutes to change the rotation angle of the manipulable cone in the experimental environment by pressing the button. Finally, Participants were asked to combine the de-occlusion effect with the overall effect of the image at each of the five angles and to fill out a questionnaire for each of the five angles after the experiment was completed. This task took approximately 10-15 minutes per participant.

Metric. The metric of the pilot user study is a specific questionnaire, as shown in Table 1. Participants rated each question on a 7-point Likert scale (1: strongly disagree - 7: strongly agree). The HMD device was placed next to the participant throughout the whole experiment, allowing participants to view the different de-occlusion effects of different angles repeatedly through the HMD device while filling out the questionnaire.

Statistical analysis. We collect and statistically analyze the scores from the user survey. For all aspects, the analysis uses the average score for each category. First, we perform the Grubbs' outlier test to identify any outliers in the data. Next, we conduct the Kolmogorov-Smirnov normality test to verify the normal distribution of the data. Finally, the ANOVA test is applied to analyze these data.

Result. Figure 12 illustrates the data from the integration of our questionnaire for 16 participants. The results of ANOVA test showed that different DEGREE ($F_{4,75}$ = 361.136, $p < .001$, η_p^2 = .951) had a significant effect on participants' questionnaire scores. The two line graphs in the figure shows that the total score of the questionnaire is the highest at 45°, in addition to which the users get a more

Table 1: Questions to be answered in the questionnaire.

No.	Question
Q_1	Satisfied with the results of dis-occlusion at this rotation angle?
Q_2	Are you satisfied with the continuity of the picture at this rotation angle?
Q_3	Satisfied with the tearing of the picture at that rotation angle?
Q_4	Are you satisfied with the overall presentation of the image at this rotation angle?
Q_5	Is it agreed that this angle will be the maximum movement angle of the MCBOS?

satisfactory de-occlusion effect. By analyzing the results of the Q_1 questionnaire at this point, we can find that the de-occlusion effect at angles of $60°(p = .153$, $\eta_p^2 = .027)$ and $75°(p = .087$, $\eta_p^2 = .039)$ is not significantly improved compared to the de-occlusion effect at 45°, but the total score is greatly decreased compared to the 45°. Therefore, we choose 45°as our maximum movement angle θ_m.

5.2. Pilot user study 2

This pilot user study is used to determine the value of the hyperparameter δ required for p_m position determination during the MCBOS construction process, as shown in Figure 4. This ensures that the build position of the manipulable cone is offset from the hand in the process of using MCBOS, avoiding the occlusion of the hand on the objects in the manipulable cone area, and also providing a good manipulation experience for the user. The participants and apparatus for this pilot user study are the same as those used in section 5.1.

Task. The task of this pilot is that participants need to use the MCBOS method to complete the operation of removing occlusion from the occluded object within the manipulable region determined by different δ parameters. δ is a two-dimensional vector, which in this context can be understood as the offset of the center of the

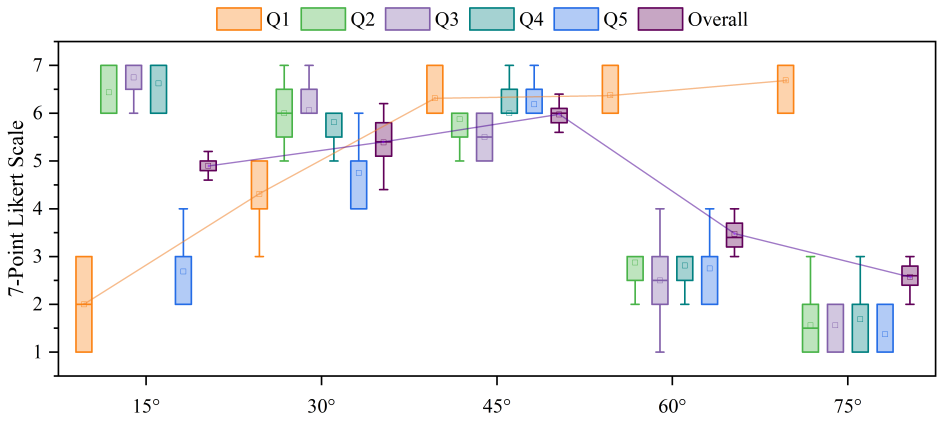


Figure 12: Boxplot of 7-Point Likert Scale Scores Across Different Angles. This figure presents the questionnaire scores for five questions (Q1–Q5) and the Overall scores across different maximum rotation angles (15°, 30°, 45°, 60°, and 75°). Each boxplot is color-coded to represent the corresponding question or overall score. The Overall scores have been normalized by dividing them by 5 to facilitate observation and comparison. The two lines in the graph connect the median values of the scores for Q1 and Overall, illustrating trends in user feedback across different angles.

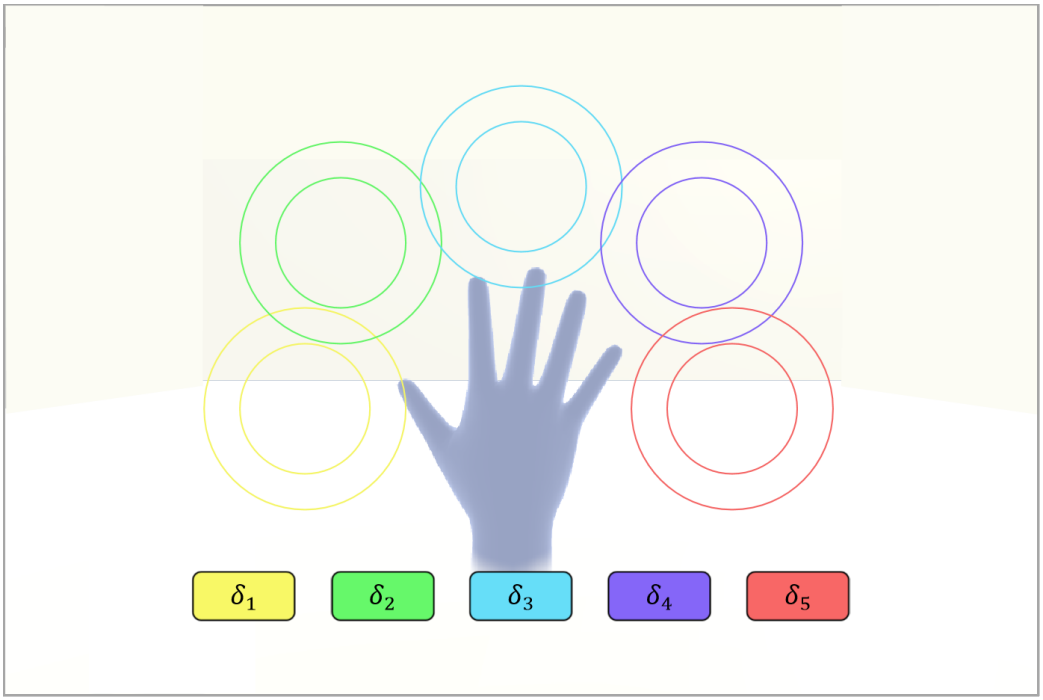


Figure 13: Mapping of δ Parameters to Manipulable Cone Positions. This figure illustrates how each of the five candidate δ parameters (δ_1, δ_2, δ_3, δ_4, and δ_5) corresponds to a specific position of the manipulable cone. The colored buttons at the bottom match the colors of the rings in the figure, indicating the corresponding position associated with each δ parameter.

manipulable region concerning the hand. We provided participants with the location of the manipulable region determined by five different $\delta \in \{\ \delta_1\ (-0.20, 0.00)$, $\delta_2\ (-0.18, 0.18)$, $\delta_3\ (0.00, 0.20)$, $\delta_4\ (0.18, 0.18)$, $\delta_5\ (0.20, 0.00)\ \}$. Specifically, the locations of the manipulable regions corresponding to the different δ parameters are shown in Figure 13.

Procedure. In the procedure, there are five buttons in front of the participant, each corresponding to five δ parameters. When the participant pressed a button, the ma-

nipulable region would go and change its position according to this δ parameter. Participants were asked to repeat the de-occlusion task using the five δ parameter. After completing the experiment, participants needed to fill out a questionnaire to evaluate the operating comfort of the different manipulable region positions.

Metric. In the metric, we designed a 10-point Likert scale questionnaire to measure the manipulable comfort score for this experiment. Participants were asked to complete the questionnaire, rating the manipulable comfort of the manipulable area for each δ parameter. The ratings range from 1 to 10, representing bad to excellent comfort. The statistical analysis methods for this pilot user study are consistent with those used in section 5.1.

Result. Figure 14 shows the data we obtained after integrating the questionnaires from 16 participants. In the figure, we have calculated the average of the scores of the five parameter values. The average score of δ_2 is $M = 7.688$, which is much larger than the average score of the remaining four δ_1 $(M = 5.438)$, δ_3 $(M = 5.562)$, δ_4 $(M = 3.312)$, δ_5 $(M = 1.938)$. The results of the RM-ANOVA test showed that the δ_2 scores were all significantly higher compared to δ_1 $(SE = 0.233, P < .001)$, δ_3 $(SE = 0.202, P < .001)$, δ_4 $(SE = 0.272, P < .001)$ and δ_5 $(SE = 0.266, P < .001)$. This consistent pattern of significant differences emphasizes the robustness of the effects observed across comparisons. Therefore, we chose δ_2 $(-0.18, 0.18)$ as the offset constant δ for the manipulable cone.

6. User study - selection efficiency and utility validation

We conducted a user study to validate the efficiency and utility of the MCBOS method for object selection in high occlusion VR environments. In this study,

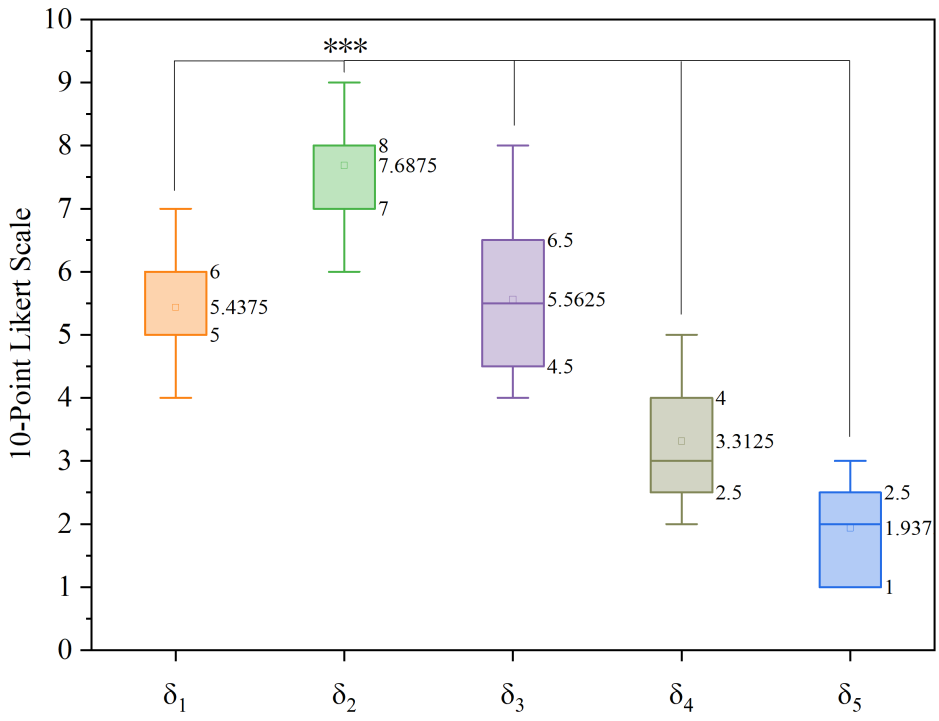


Figure 14: Box plots of score statistics for different δ parameters. This figure presents the score distribution across different δ parameters using a 10-point Likert scale. Each box plot visualizes the upper and lower bounds (whiskers), the interquartile range (box), and the mean value (indicated by a square marker) for each parameter. Significant differences in pairwise comparisons are marked with ***, representing a Bonferroni-adjusted significance level of .001.

we investigated the effects of various factors on the performance of MCBOS for object selection in high occlusion VR environments. Based on this user study, we propose the following hypotheses:

- **H1.** MCBOS can significantly improve the selection efficiency in high occlusion VR environments.

- **H2.** MCBOS can significantly improve its utility in high occlusion VR environments.

6.1. Participants and apparatus

Twenty-four participants (12 males, 12 females, aged between 22-29 years, M = 24.08, SD = 2.15, all right-handed) volunteered to participate in this study. None of them had experience with HMDs and handheld controllers prior to the study. All participants had normal vision and were not color blind (10 of them wore corrective glasses for myopia). The configuration of the experimental system is identical to that used in section 5.

6.2. Task and Scene design

Based on the study of previous target selection experiments (Lu et al., 2020), we designed an experiment for MCBOS, as shown in Figure 15. The following three factors were considered in this experiment: *selection method*, *environment design* (related to the number of spheres and occlusion), and *experiment distance* (related to the distance from the participant to the selected target).

Three methods of bare hand object selection were evaluated, Control Condition 1: 3D-Bubble (CC1), Control Condition 2: Multi-Bubble(CC2), and Experimental Condition: MCBOS (EC). The 3D-Bubble method allows the user to

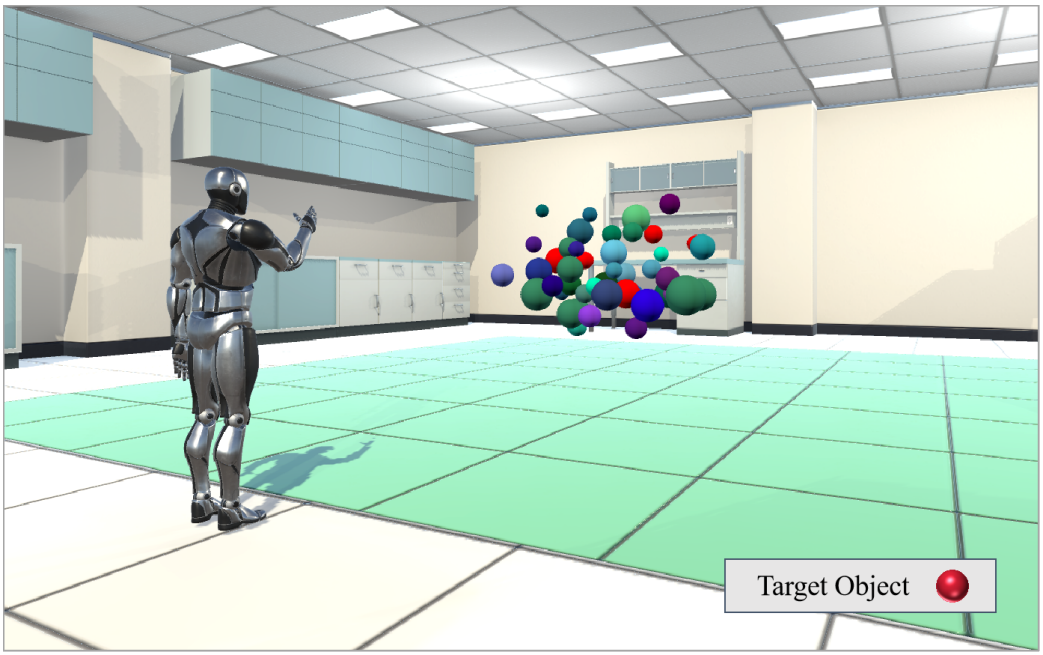


Figure 15: Experimental scenario for the three selection methods. This figure illustrates the experimental setup used to evaluate our three selection methods. The red objects represent the five target objects that participants must select during the experiment. At any given time, only one target object is visible for selection. After a participant successfully selects the current target, the next target is chosen randomly from the remaining four. The initial target object is also randomly selected from the pool of five red objects.

control a semi-transparent sphere by mapping the position of the hand to the position of the sphere. The sphere adjusts its radius to always select the closest target. The Multi-Bubble method allows the user to include multiple objects within the sphere manipulated by the palm of the main hand. This method associates these candidate objects with fingers. The user can move their palm and fingers to adjust the selection. Finally, the user selects the mapped, desired object by flexing the appropriate finger. MCBOS allows users to control an area with their hands, use gestures to remove occlusion before the target object, and finally select it by gesture manipulation. MCBOS essentially solves the occlusion problem in the selection process.

Environment Density is defined as the number of objects contained in the same region. *experiment distance* is defined as the distance from the participant's position to the center point of the selected object region. To evaluate the effect of different *environment density* and *experiment distance* on the speed and the accuracy of object selection, we designed a combination of 2 *environment density* (low density, high density) × 2 *experiment distance* (near, far). The specific experimental arrangements for different combinations are shown in Figure 16.

Therefore, in this experiment, we used a total of 12 combinations (3 methods × 2 density × 2 distance) to accurately evaluate our method.

6.3. Procedure, metrics and statistical analysis

The duration of the experiment was approximately 40 minutes per participant. We first introduced the experiment to the participants and guided them to stand in a fixed position in the center of the experimental area while facing the experimental area and wearing an HMD on their heads. Participants defaulted to using their right hand to control the different selection methods. Prior to the start of

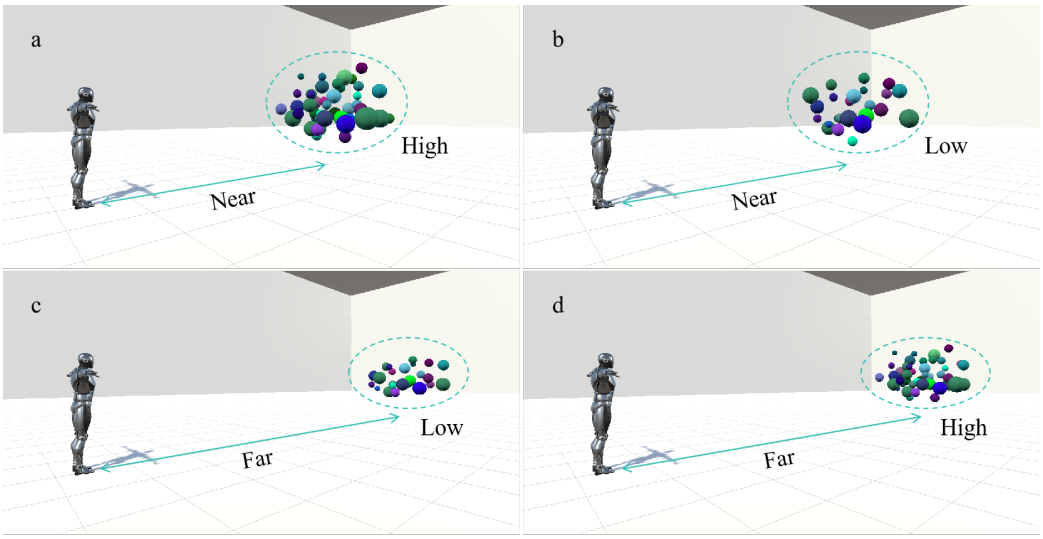
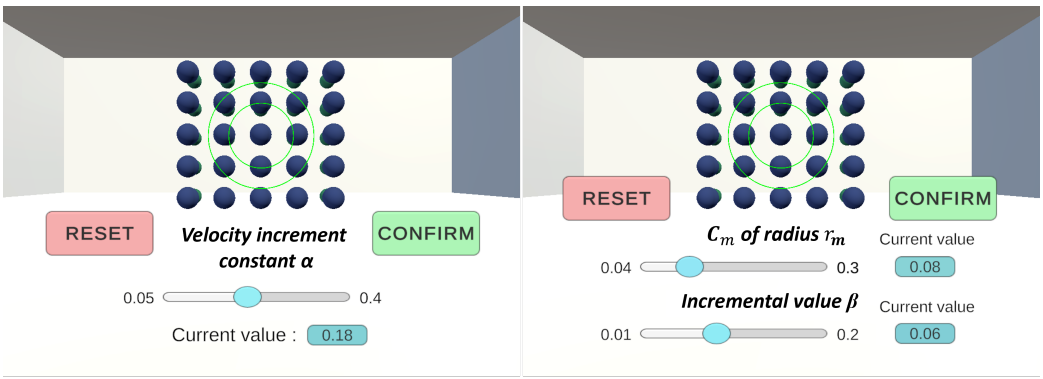


Figure 16: Scene pairings for different *environment density* and *experiment distance*. This figure illustrates the four experimental scenarios based on combinations of environment density (High or Low) and experiment distance (Near or Far): **(a):** Near - High Density; **(b):** Near - Low Density; **(c):** Far - Low Density; **(d):** Far - High Density. In each scenario, the participant interacts with objects at varying distances and object densities. High density indicates a greater number of objects in the cluster, while Low density features fewer objects. Near and Far refer to the spatial distance between the user and the object cluster.

the experiment, participants were asked to complete an SSQ questionnaire that was used as a baseline for the SSQ analysis. We provided per participant with 15 minutes to get used to the VR system and to practice the three selection methods described above. Before the exercise started with the MCBOS method, we set up three personalized parameters for the participants to adjust, which are shown in Figure 3: 1. the velocity increment constant α for the apices of the manipulable cone C_m; 2. radius r_m of the manipulable cone C_m; and 3. the maximum radius of the light path bending region R_t. To allow participants to personalize the three parameters of MCBOS, we set up two tuning environments for these three parameters. Participants could adjust the parameters in these two environments, and then use MCBOS to perform selection exercises. The specific parameter adjustment environments are shown in Figure 17.

After familiarizing themselves with all of the selection methods, participants were asked to select a total of five red-highlighted objects in the environment as quickly as possible using the three selection methods. At the beginning of the experiment, a partially or fully occluded red highlighted sphere appeared in the environment, and participants were required to move their heads and turn their heads to find the target and complete a selection. Participants needed to select the correct target to trigger subsequent choices. When the red target is successfully selected, the target returns to its original color and the next target to be selected turns red. The next target did not turn red until the previous target was successfully selected. Participants were required to select all 5 selected targets in a single experiment correctly. Each participant took a five-minute break after conducting a set of experiments, in order to alleviate any potential feelings of cybersickness. In addition, participants were asked if there was any physical discomfort that would



(a) (b)

Figure 17: Personalized parameter adjustment environment. (a) shows the velocity increment constant α adjustment environment, the default value is 0.18, and participants can slide the slider in front of them to make adjustments to the value. Pressing the Reset button restores the default value, and pressing the Confirm button completes this adjustment. (b) shows the adjustment environment of radius r_m and the maximum radius of the light path bending region R_t. Where R_t is adjusted by using the incremental parameter β instead. R_t equals r_m plus β. The specific adjustment process is the same as that of (a).

allow them to perform the next experiment. Each participant was required to use three different methods in each scenario and complete five experiments for each method. In addition, to avoid the influence of method test order on the final test results, we used the Latin square order to prioritize the participants' experiments for the three methods. In total, we collected data from 1440 experiments (= 24 participants × 3 methods × 4 scenarios × 5 repetitions).

We collected both participants' objective data and subjective data. In each trial, we recorded the following parameters: *completion time*, the time duration of a trial from start to finish; *total rotation*, the total angle of rotation of the participant's head during a trial from start to finish; *total translation*, the total distance traveled by the participant from the start to the end of a trial; *error rate*, the ratio of the number of correct selections to the total number of selections in the experiment. After participants completed all trials in each method condition, we gave them a UEQ (Schrepp et al., 2017), a NASA-TLX questionnaire (Hart, 2006), and a SSQ (Kennedy et al., 1993) to measure user experience, workload, and sickness for each method. At the end of the experiment, we also asked participants to rank these methods according to their overall preference and conducted an interview to obtain additional feedback and comments.

In statistical analysis, for each participant and each condition, We first identified and removed outliers where completion time exceeded $M \pm 3 \cdot std.$ in each condition (28 trials, ≈1.94%). Shapiro-Wilk tests and Q-Q plots indicated that all performance metrics obeyed a normal distribution. We then performed the RM-ANOVA test and the Holm-Bonferroni corrected post hoc paired t-test for each condition. Finally, We quantified the effect sizes for each metric using Cohen's d (Cohen, 2013). The d values were translated to qualitative effect size estimates

of Huge ($d > 2.0$), Very Large ($2.0 > d > 1.2$), Large ($1.2 > d > 0.8$), Medium ($0.8 > d > 0.5$), Small ($0.5 > d > 0.2$), and Very Small ($0.2 > d > 0.01$). The statistical analysis was performed using the SPSS software (IBM, n.d.).

6.4. Results

6.4.1. Efficiency results

Table 2: Completion time for each condition. Asterisks denote statistical significance.

Scene	**Method**	**Avg. ± std.**(s)	$(M_i - M_{MCBOS})/M_i$	**p**	**Cohen's d**	**Effect size**
Near High	CC1	42.78 ± 3.86	43.4%	$< .001$*	6.17	Huge
	CC2	34.64 ± 2.18	30.2%	$< .001$*	5.21	Huge
	EC	24.19 ± 1.80				
Near Low	CC1	23.69 ± 1.64	51.3%	$< .001$*	8.53	Huge
	CC2	14.98 ± 1.73	22.9%	0.002*	2.33	Huge
	EC	11.54 ± 1.17				
Far High	CC1	90.08 ± 8.42	61.3%	$< .001$*	8.64	Huge
	CC2	54.78 ± 4.04	36.4%	$< .001$*	5.42	Huge
	EC	34.83 ± 3.28				
Far Low	CC1	29.81 ± 3.91	58.1%	$< .001$*	6.00	Huge
	CC2	15.39 ± 1.76	18.9%	0.003*	1.95	Very Large
	EC	12.48 ± 1.16				

Completion time. Table 2 showed the results of the post-hoc test for the completion time metric. First, the RM-ANOVA test revealed that METHOD ($F_{2,72}$ = 482.698, $p < .001$, η_p^2 = .931), DISTANCE ($F_{1,36}$ = 483.374, $p < .001$, η_p^2 = .931), and DENSITY ($F_{1,36}$ = 1984.747, $p < .001$, η_p^2 = .982) all had a significant effect on completion time. We also found a cross effect between METHOD and DISTANCE ($F_{2,72}$ = 85.396, $p < .001$, η_p^2 = .703). Additionally, a cross effect between METHOD and DENSITY ($F_{2,72}$ = 86.818, $p < .001$, η_p^2 = .707) was

observed. Moreover, by analyzing the results of the post-hoc test on completion time, EC drastically reduced this metric compared to CC1 and CC2 in all four scenarios. Specifically, in the Near High scene, we observed significant improvements $(43.4\%,\ p < .001;\ 30.2\%,\ p < .001)$ compared to CC1 and CC2. Cohen's d-value further emphasized the significant improvement in EC compared to CC1 and CC2. Similarly, in Near Low and Far Low scenarios, EC demonstrated significant improvements relative to CC1 and CC2 $(51.3\%,\ p < .001,\ 22.9\%,\ p = .002;\ 58.1\%,\ p < .001,\ 18.9\%,\ p = .003)$, and Cohen's d-value confirmed the significant enhancement of EC. Most notably, in the Far High scenario, EC exhibited the largest enhancement $(61.3\%,\ p < .001;\ 36.4\%,\ p < .001)$ compared to CC1 and CC2, with Cohen's d-value also highlighted this significant difference.

Table 3: Error rate for each condition. Asterisks denote statistical significance.

Scene	**Method**	**Avg. ± std.**	$(M_i - M_{MCBOS})/M_i$	**p**	**Cohen's d**	**Effect size**
Near High	CC1	0.28 ± 0.07	88.2%	$< .001$*	4.76	Huge
	CC2	0.13 ± 0.04	75.2%	0.004*	2.77	Huge
	EC	0.03 ± 0.03				
Near Low	CC1	0.15 ± 0.03	89.4%	$< .001$*	5.58	Huge
	CC2	0.03 ± 0.03	45.6%	0.373	0.67	Medium
	EC	0.02 ± 0.01				
Far High	CC1	0.34 ± 0.04	80.4%	$< .001$*	5.09	Huge
	CC2	0.26 ± 0.08	74.0%	0.006*	2.64	Huge
	EC	0.07 ± 0.06				
Far Low	CC1	0.30 ± 0.08	88.0%	$< .001$*	4.57	Huge
	CC2	0.08 ± 0.04	54.6%	0.095	1.34	Very Large
	EC	0.04 ± 0.03				

Error rate. Table 3 illustrated the results derived from the post-hoc test for the error rate metric. To begin with, the RM-ANOVA test revealed that METHOD

($F_{2,72}$ = 210.102, $p < .001$, η_p^2 = .854), DISTANCE ($F_{1,36}$ = 52.306, $p < .001$, η_p^2 = .592), and DENSITY ($F_{1,36}$ = 70.013, $p < .001$, η_p^2 = .660) all had a significant effect on error rate. We also found a cross effect between METHOD and DISTANCE ($F_{2,72}$ = 6.451, $p < .003$, η_p^2 = .152). Additionally, a cross effect between METHOD and DENSITY ($F_{2,72}$ = 13.380, $p < .001$, η_p^2 = .271) was observed. The post-hoc test results for error rate indicated that EC outperformed both CC1 and CC2 in reducing this metric across all four scenarios. More precisely, compared to the CC1 method, EC showed a significant improvement in the error rate of selection over the four scenarios (88.2%, $p < .001$; 89.4%, $p < .001$; 80.4%, $p < .001$; 88.0%, $p < .001$, respectively). The value of Cohen's d also verified this statement, with an improvement level of Huge for all scenarios. Compared to the CC2 method, EC showed a significant improvement in the selection error rate in Near High and Far High scenarios (75.2%, $p = .004$; 74.0%, $p = .006$, respectively). There was also an improvement of more than 50% in the Far Low scenario compared to CC2 (54.6%, $p = .095$). The improvement in the Near Low scene compared to the CC2 (45.6%, $p = .373$) method was the smallest but still exceeded 40%. The Cohen's d values supported these assertions as well.

6.4.2. Utility results

Total rotation. Table 4 presented the outcomes of the post-hoc test analysis conducted on the total rotation metric. Initially, the RM-ANOVA test revealed that METHOD ($F_{2,72}$ = 1437.310, $p < .001$, η_p^2 = .976), DISTANCE ($F_{1,36}$ = 201.520, $p < .001$, η_p^2 = .848), and DENSITY ($F_{1,36}$ = 4960.289, $p < .001$, η_p^2 = .993) all had a significant effect on total rotation. We also found a cross effect between METHOD and DISTANCE ($F_{2,72}$ = 62.507, $p < .001$, η_p^2 = .635). Additionally, a cross effect between METHOD and DENSITY ($F_{2,72}$ = 413.161, $p <$

Table 4: Total rotation for each condition. Asterisks denote statistical significance.

Scene	Method	Avg. ± std.(°)	$(M_i - M_{MCBOS})/M_i$	p	Cohen's d	Effect size
Near High	CC1	286.78 ± 18.67	75.8%	< .001*	15.26	Huge
	CC2	186.38 ± 19.12	62.8%	< .001*	8.05	Huge
	EC	69.40 ± 7.57				
Near Low	CC1	74.74 ± 5.18	66.1%	< .001*	11.77	Huge
	CC2	42.05 ± 3.57	39.7%	< .001*	5.14	Huge
	EC	25.35 ± 2.89				
Far High	CC1	351.56 ± 22.89	78.8%	< .001*	16.49	Huge
	CC2	290.51 ± 15.61	74.4%	< .001*	18.12	Huge
	EC	74.49 ± 6.37				
Far Low	CC1	132.27 ± 17.93	94.7%	< .001*	9.83	Huge
	CC2	21.61 ± 2.48	67.7%	< .001*	6.59	Huge
	EC	6.98 ± 1.92				

.001, η_p^2 = .920) was observed. Furthermore, the post-hoc test data revealed that EC significantly diminished this metric, outstripping CC1 and CC2 in all scenarios. Precisely, in the Near High scene, we observed that EC showed significant improvements (75.8%, $p < .001$; 62.8%, $p < .001$) compared to CC1 and CC2. Cohen's d-value further emphasized the significant improvement in EC compared to CC1 and CC2. Similarly, in the Near Low scenario, EC showed significant improvements relative to CC1 (66.1%, $p < .001$) and CC2 (39.7%, $p < .001$). In the Far High scenario, EC showed the most improvement compared to CC2 (74.4%, $p < .001$), and also improved significantly compared to CC1(78.8%, $p < .001$). Finally, in the Far Low scenario, EC exhibited the largest enhancement compared to CC1 (94.7%, $p < .001$). For CC2 (67.7%, $p < .001$), there was also a substantial improvement in that scenario. Cohen's d-value also highlighted these significant differences.

Table 5: Total translation for each condition. Asterisks denote statistical significance.

Scene	Method	Avg. ± std.(m)	$(M_i - M_{MCBOS})/M_i$	p	Cohen's d	Effect size
Near High	CC1	8.25 ± 1.27	86.5%	$< .001$*	7.86	Huge
	CC2	4.51 ± 0.91	75.4%	$< .001$*	5.21	Huge
	EC	1.11 ± 0.17				
Near Low	CC1	5.97 ± 1.11	82.2%	$< .001$*	6.46	Huge
	CC2	2.15 ± 0.40	58.8%	$< .001$*	4.24	Huge
	EC	0.89 ± 0.12				
Far High	CC1	18.73 ± 2.29	79.1%	$< .001$*	8.80	Huge
	CC2	11.10 ± 1.31	64.7%	$< .001$*	6.99	Huge
	EC	3.92 ± 0.63				
Far Low	CC1	3.82 ± 0.96	79.8%	0.001*	4.44	Huge
	CC2	1.65 ± 0.33	53.2%	0.002*	3.50	Huge
	EC	0.77 ± 0.11				

Total translation. Table 5 delineated the findings from the post-hoc test evaluation pertaining to the total translation metric. Firstly, the RM-ANOVA test revealed that METHOD ($F_{2,72} = 468.742$, $p < .001$, $\eta_p^2 = .929$), DISTANCE ($F_{1,36} = 247.411$, $p < .001$, $\eta_p^2 = .873$), and DENSITY ($F_{1,36} = 886.464$, $p < .001$, $\eta_p^2 = .961$) all had a significant main effect on total translation. We also found a cross effect between METHOD and DISTANCE ($F_{2,72} = 16.489$, $p < .001$, $\eta_p^2 = .314$). Additionally, a cross effect between METHOD and DENSITY ($F_{2,72} = 99.644$, $p < .001$, $\eta_p^2 = .735$) was observed. According to the post-hoc analysis of total translation, EC notably decreased this measurement when compared to CC1 and CC2 in each of the four scenarios. Particularly, in the Near High scenario, EC demonstrated the largest improvement compared to CC1 (86.5%, $p < .001$) and CC2 (75.4%, $p < .001$). In the Near Low scenario, the improvement compared to the CC2 (58.8%, $p < .001$) method was reduced, but still a substantial im-

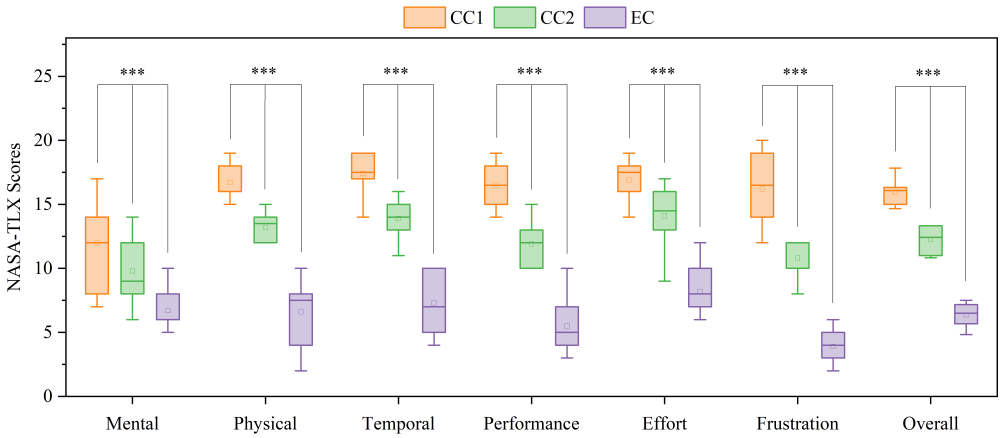


Figure 18: NASA-TLX Scores box plot (lower scores are better). The scores for the NASA-TLX questionnaire are tallied and the scores are stacked to calculate a total score. The individual scores for each dimension are tallied to calculate a total score, which is divided by 6 for analysis. Significant differences in pairwise comparisons are marked with ***, representing a Bonferroni-adjusted significance level of .001.

provement compared to the CC1 (82.2%, $p < .001$) method. In the Far High and Far Low scenarios, EC showed a very close improvement compared to the CC1 (79.1%, $p < .001$; 79.8%, $p < .001$) method, and the improvement was above 50% compared to the CC2 (64.7%, $p < .001$; 53.2%, $p = .002$) method. The values of Cohen's d also confirmed our statement above.

NASA-TLX. Figure 18 presented the results of the NASA-TLX questionnaire. The RM-ANOVA test indicated that METHOD ($F_{2,27} = 268.099$, $p < .001$, $\eta_p^2 = .952$) had a significant effect on the results. Specifically, we explored the effect of different methods on the relationship between these variables using Mental, Physical, Temporal, Performance, Effort, Frustration, and Overall as dependent variables. The results of the One-way ANOVA test showed that EC had significant

advantages in Mental ($p < .001$, $\eta_p^2 = .486$; $p = .004$, $\eta_p^2 = .263$), Physical ($p < .001$, $\eta_p^2 = .858$; $p < .001$, $\eta_p^2 = .721$), Temporal ($p < .001$, $\eta_p^2 = .840$; $p < .001$, $\eta_p^2 = .691$), Performance ($p < .001$, $\eta_p^2 = .853$; $p < .001$, $\eta_p^2 = .663$), Effort ($p < .001$, $\eta_p^2 = .753$; $p < .001$, $\eta_p^2 = .556$), Frustration ($p < .001$, $\eta_p^2 = .873$; $p < .001$, $\eta_p^2 = .690$), and Overall ($p < .001$, $\eta_p^2 = .951$; $p < .001$, $\eta_p^2 = .881$, respectively) metrics compared to CC1 and CC2.

UEQ. Figure 19 showed the results of the UEQ questionnaire, which was summarized into six different evaluation metrics. The RM-ANOVA test showed a significant correlation between METHOD and the overall results ($F_{2,27} = 212.620$, $p < .001$, $\eta_p^2 = .940$), with EC showing a significant improvement in the overall results. We performed an ANOVA test and a post-hoc test for the relationship between the results and METHOD for each group separately. The results showed that EC showed significant improvement compared to CC1 and CC2 in terms of Attractiveness ($p < .001$, $\eta_p^2 = .915$; $p < .001$, $\eta_p^2 = .844$), Efficiency ($p < .001$, $\eta_p^2 = .948$; $p < .001$, $\eta_p^2 = .926$), Dependability ($p < .001$, $\eta_p^2 = .954$; $p < .001$, $\eta_p^2 = .926$), Simulation ($p < .001$, $\eta_p^2 = .898$; $p < .001$, $\eta_p^2 = .795$), and Novelty ($p < .001$, $\eta_p^2 = .908$; $p < .001$, $\eta_p^2 = .826$, respectively) metrics. In Perspicuity ($p = .022$, $\eta_p^2 = .181$; $p = .099$, $\eta_p^2 = .097$, respectively), EC was slightly worse than CC1 but still better than CC2. Figure 20 further summarized the results of the UEQ scales as Attractiveness, Pragmatic Quality, and Hedonic Quality and showed the results. The RM-ANOVA test indicated that significant differences existed between the three methods ($F_{2,54} = 567.161$, $p < .001$, $\eta_p^2 = .955$). Specifically, the results of the post-hoc test showed that EC produced significant improvements in Attractiveness ($p < .001$), Pragmatic Quality ($p < .001$), and Hedonic Quality ($p < .001$) compared to CC1 and CC2.

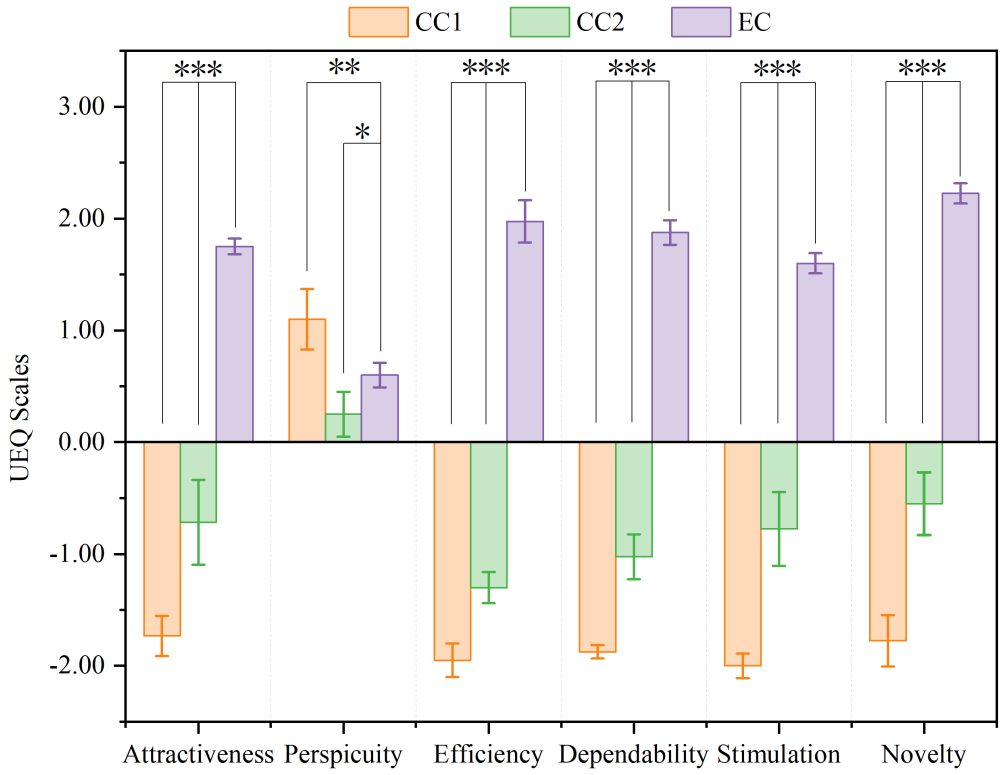


Figure 19: UEQ Scales bar chart. This bar chart shows the 26 questions of the UEQ questionnaire summarized into 7 relevant outcomes ranging from -3 (horribly bad) to 3 (extremely good). The error bars in this figure represent standard errors. Significant differences in pairwise comparisons are marked with *, **, and ***, representing a Bonferroni-adjusted significance level of .05, .01, and .001, respectively.

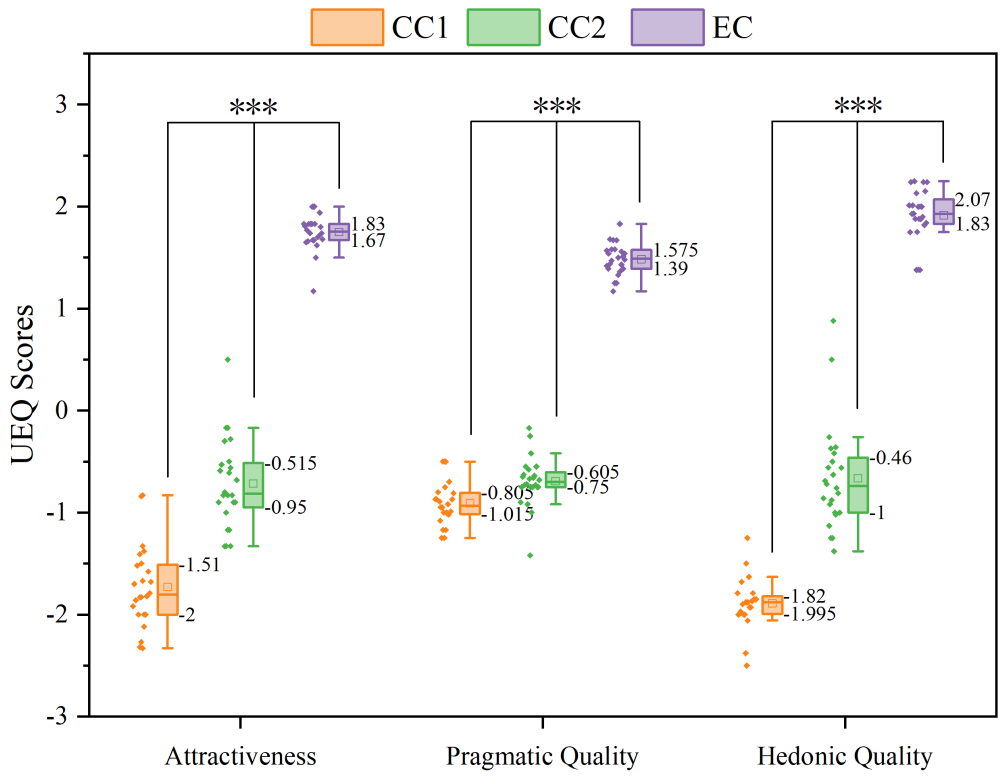


Figure 20: UEQ Scores box plot across different quality dimensions. This figure presents the results of the UEQ scores across three key dimensions: Attractiveness, Pragmatic quality (Perspicuity, Efficiency, Dependability), and Hedonic quality (Stimulation, Originality). Each box plot displays the median value, interquartile range, and individual data points for the corresponding dimension. Significant differences in pairwise comparisons are marked with ***, representing a Bonferroni-adjusted significance level of .001.

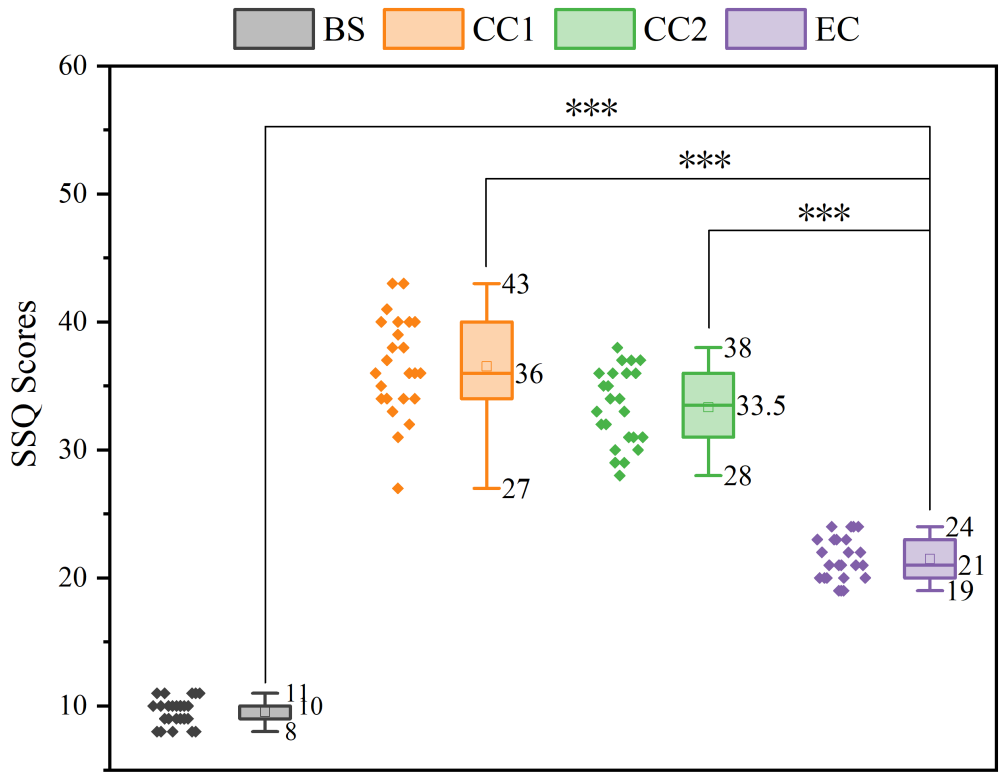


Figure 21: SSQ scores box plot (lower scores are better). This figure presents box plots of the SSQ scores for three methods. BS indicates the participants' SSQ questionnaire before the start of the experiment. On the left, individual data points illustrate the distribution of SSQ scores for each method. On the right, the box plots show the interquartile range and median scores for each method. Significant differences in pairwise comparisons are marked with ***, representing a Bonferroni-adjusted significance level of .001.

SSQ. Figure 21 showed the results of the SSQ scores. The RM-ANOVA test showed that METHOD ($F_{3,92}$ = 519.067, $p < .001$, η_p^2 = .944) had a significant effect on the SSQ scores. Specifically, the One-way ANOVA test showed that there was a significant difference in the SSQ scores for EC compared to CC1 and CC2 ($p < .001$, η_p^2 = .744; $p < .001$, η_p^2 = .643, respectively). All three methods also showed significant differences compared to baseline BS: CC1 ($p < .001$, η_p^2 = .932), CC2 ($p < .001$, η_p^2 = .913), EC ($p < .001$, η_p^2 = .727). The results indicate that, although EC induces a certain degree of sickness, it demonstrates a significant improvement compared to the CC1 and CC2 methods.

Overall Preference. Figure 22 showed the voting results of the participants' preferences. The results also showed that participants' preferences differed. In the first ranking, 92% of the participants (22) considered EC to be the top-ranked method, and 8% of the participants (2) considered CC2 to be the top-ranked method. In the second ranking, 4% of the participants (1) considered EC to be the second-ranked method, 83% of the participants (20) considered CC2 to be the second-ranked method, and 13% of the participants (3) considered CC1 to be the second-ranked method. In the third ranking, 13% of the participants (3) considered CC2 to be the third-ranked method and 88% of the participants (21) considered CC1 to be the third-ranked method.

7. User study - adaptation in practical application scenarios

This user study aims to verify the applicability of MCBOS in practical scenarios and the ability of MCBOS to expand in applications. Based on this user study, we propose the following hypothesis:

- **H3.** MCBOS demonstrates excellent adaptability in practical application

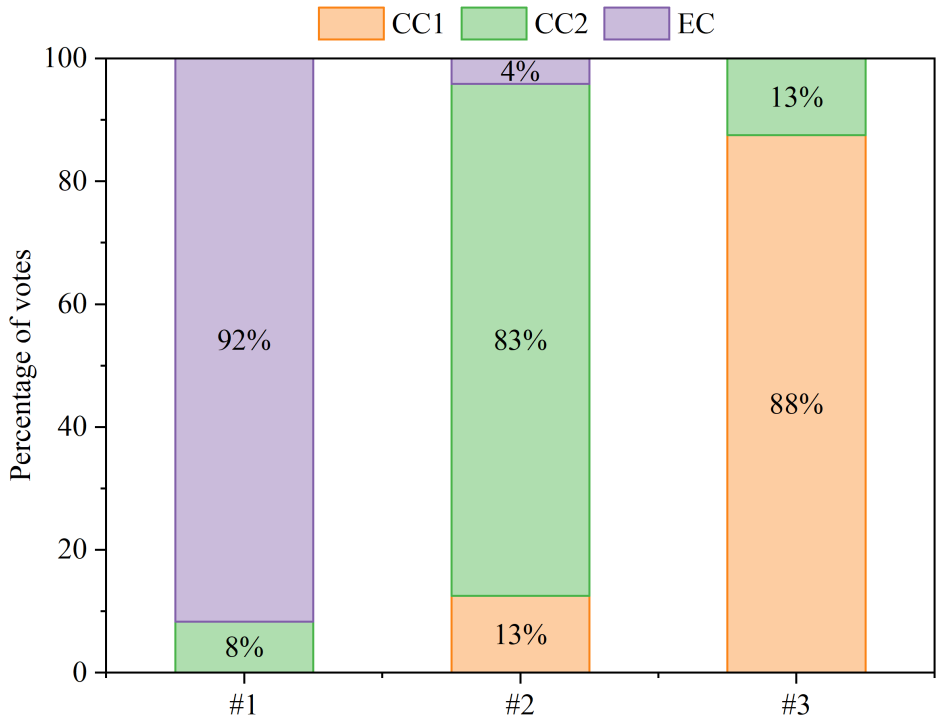


Figure 22: Users' rankings (#1, #2, and #3) of the three methods. This figure presents the final user rankings of the three tested methods (CC1, CC2, and EC), with rankings labeled as #1, #2, and #3. The height of each colored segment indicates the percentage of participants who voted for that method in the respective ranking. This visualization highlights user preferences, showing that EC is overwhelmingly preferred as the top choice, while CC1 is mostly ranked as the least preferred method.

scenarios.

7.1. Participants and apparatus

Twenty-four participants (12 males, 12 females, aged between 22-28 years, M = 24.67, SD = 1.84, all right-handed) volunteered to participate in this study. The participants are all professionals in the field of VR or have experience using VR devices. All participants had normal vision and were not color blind (10 of them wore corrective glasses for myopia).

The configuration of the experimental system in this experiment is consistent with those described in section 6.

7.2. Task

We designed an object selection task in a VR factory. Participants select two occluded objects according to our prompts in the environment, and completing the selection of all two objects is regarded as completing an experiment. As in section 6, we also used the three methods of 3D-Bubble, Multi-Bubble, and MCBOS for the evaluation.

7.3. Procedure, metrics and statistical analysis

The duration of the experiment was approximately 60 minutes for each participant. We introduced the experiment to the participants and guided them to stand in a fixed position in the center of the experimental area while facing the experimental area and wearing an HMD on their heads. Participants defaulted to using their right hand to control the different selection methods. Prior to the start of the experiment, participants were asked to fill out an SSQ questionnaire to be used as a baseline for SSQ analysis. Once the experiment began, we first

gave participants 5 minutes to familiarize themselves with our testing environment and the subsequent walking route. We then provided each participant with 15 minutes to familiarize themselves with our three-choice methods. As in section 6, each participant could personalize the MCBOS before starting. At the start of the experiment, participants were asked to select two objects in sequence as quickly as possible. Since the positions of the objects in the scene were initially invisible, participants had to move around to locate the chosen items. To help participants approach and select the objects more efficiently, we provided hints in the experimental environment regarding the current locations of the objects, as shown in Figure 23. The red arrow indicates the path hint, the yellow arrow shows the direction hint, and the green arrow points to the object position hint. These visual cues ensured that participants could quickly navigate the environment and accurately localize the target object for selection.

The path for the second object, along with the cue indicating its location, was presented only after the participant had successfully completed the selection of the first object. Each participant was required to complete six separate choice experiments using three different methods. In addition, to avoid the influence of method test order on the final test results, we used the Latin square order to prioritize the participants' experiments for the three methods. Whenever a set of methods was completed, participants were asked to take a 5-minute break to minimize the effect of vertigo on the results of subsequent experiments. In total, we collected data from 432 experiments (= 24 participants × 3 methods × 1 scenario × 6 repetitions).

We collected both participants' objective data and subjective data. In each trial, we recorded the following parameters: *completion time*, the time duration of a trial from start to finish; *total rotation*, the total angle of rotation of the partic-

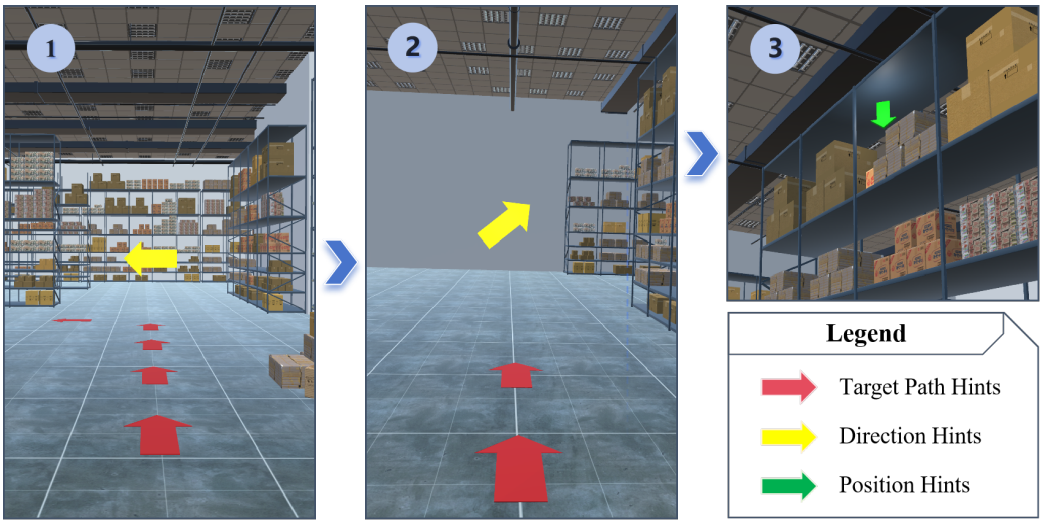


Figure 23: Visual cues provided to participants in the practical scene experiment. At the beginning of the experiment, participants were given a hint about the first target object's location. The following visual cues were provided to guide participants efficiently through the scene: **Target Path Hints (Red Arrows):** Indicate the path participants should follow to approach the target object. **Direction Hints (Yellow Arrows):** Help guide participants in the correct direction as they approach the target area. **Position Hints (Green Arrows):** Pinpoint the precise location of the target object for selection.

ipant's head during a trial from start to finish; *total translation*, the total distance traveled by the participant from the start to the end of a trial; After participants completed all trials in each method condition, we gave them a UEQ (Schrepp et al., 2017), a NASA-TLX questionnaire (Hart, 2006), and a SSQ (Kennedy et al., 1993) to measure user experience, workload, and sickness for each method. At the end of the experiment, we also asked participants to rank these methods according to their overall preference and conducted an interview to obtain additional feedback and comments.

In statistical analysis, for each participant and each condition, We first identified and removed outliers where completion time exceeded $M \pm 3 \cdot std.$ in each condition (10 trials, $\approx$2.31%). Shapiro-Wilk tests and Q-Q plots indicated that all performance metrics obeyed a normal distribution. We then performed the RM-ANOVA test and the Holm-Bonferroni corrected post hoc paired t-test for each condition. Finally, We quantified the effect sizes for each metric using Cohen's d (Cohen, 2013). The d values were translated to qualitative effect size estimates of Huge ($d > 2.0$), Very Large ($2.0 > d > 1.2$), Large ($1.2 > d > 0.8$), Medium ($0.8 > d > 0.5$), Small ($0.5 > d > 0.2$), and Very Small ($0.2 > d > 0.01$). The statistical analysis was performed using the SPSS software (IBM, n.d.).

7.4. Results

7.4.1. Objective results

We analyzed the experimental data and presented the results using box plots and tables. Figure 24 illustrates the overall trends of the three objective metrics. As shown in Figure 24, the EC group exhibits significant improvements in all three key metrics: completion time, total rotation, and total translation. Next, we will conduct a more detailed mathematical analysis of these objective metrics.

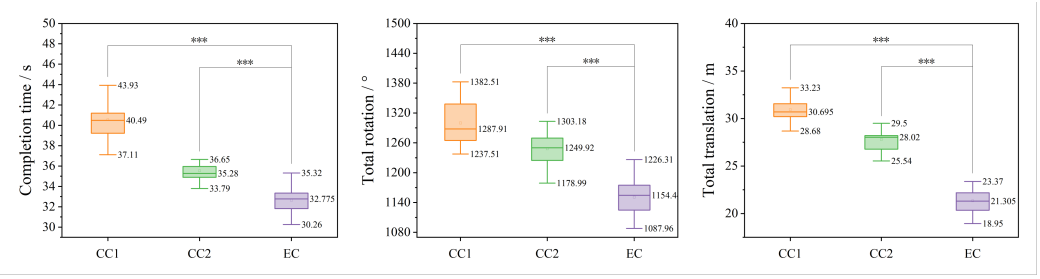


Figure 24: Comparison of Task Performance Metrics Across Methods (CC1, CC2, EC). The box plots illustrate completion time (left), total rotation (center), and total translation (right) across three conditions: CC1, CC2, and EC. Lower values indicate better performance. The EC condition consistently achieves the lowest scores across all metrics. Significant differences in pairwise comparisons are marked with ***, representing a Bonferroni-adjusted significance level of .001.

Table 6: Comparison of methods across metrics. Asterisks denote statistical significance.

Metric	**Method**	**Avg. ± std.**	$(M_i - M_{MCBOS})/M_i$	**p**	**Cohen's d**	**Effect size**
Completion time	CC1	40.54 ± 2.06	19.5%	$< .001$*	4.67	Huge
	CC2	35.54 ± 1.97	8.2%	$< .001$*	1.78	Very Large
	EC	32.63 ± 1.21				
Total rotation	CC1	1299.91 ± 45.04	11.5%	$< .001$*	3.72	Huge
	CC2	1248.12 ± 31.36	7.8%	$< .001$*	2.95	Huge
	EC	1150.83 ± 34.49				
Total translation	CC1	30.93 ± 2.23	31.0%	$< .001$*	5.43	Huge
	CC2	27.80 ± 1.71	23.2%	$< .001$*	4.46	Huge
	EC	21.34 ± 1.12				

Completion time. Table 6 presents the results of the post-hoc test for completion time. The results indicate that EC shows a significant improvement compared to CC1 ($p < 0.001$) and CC2 ($p < 0.001$), with increases of 19.5% and 8.2%, respectively. The values of Cohen's d and the results of the effect size also support this finding. Furthermore, the results of the ANOVA test ($F_{2,69} = 114.948$, $p <$

.001, η_p^2 = .769) indicate that there are significant differences between the different methods. EC demonstrated a significant improvement compared to both CC1 ($p < .001$, η_p^2 = .765) and CC2 ($p < .001$, η_p^2 = .305).

Total rotation. Table 6 reveals the results of the post-hoc test for total rotation. The findings reveal that EC significantly outperforms CC1 ($p < 0.001$) and CC2 ($p < 0.001$), with enhancements of 11.5% and 7.8%, respectively. Additionally, the Cohen's d value and the effect size results further corroborate this conclusion. The ANOVA test results ($F_{2,69}$ = 94.085, $p < .001$, η_p^2 = .732) indicate significant differences among the various methods. EC exhibited a notable improvement over both CC1 ($p < .001$, η_p^2 = .726) and CC2 ($p < .001$, η_p^2 = .530).

Total translation. Table 6 exhibits the results of the post-hoc test for total translation. The results demonstrate that EC exhibits a significant improvement over CC1 ($p < 0.001$) and CC2 ($p < 0.001$), achieving increases of 31.0% and 23,2%, respectively. Moreover, the Cohen's d value and effect size results provide further evidence for this finding. The ANOVA test results ($F_{2,69}$ = 180.059, $p <$ .001, η_p^2 = .839) reveal significant differences among the different methods. EC showed a significant enhancement compared to both CC1 ($p < .001$, η_p^2 = .834) and CC2 ($p < .001$, η_p^2 = .695).

7.4.2. Subjective results

NASA-TLX. Figure 25 illustrates the results of the NASA-TLX questionnaire. We investigated how different methods influenced the relationship between these variables, using Mental, Physical, Temporal, Performance, Effort, Frustration, and Overall as dependent variables. The results of the RM-ANOVA test indicated that there were significant differences in the results produced by the different methods in terms of Mental ($F_{2,69}$ = 69.046, $p < .001$, η_p^2 = .667), Physical ($F_{2,69}$ =

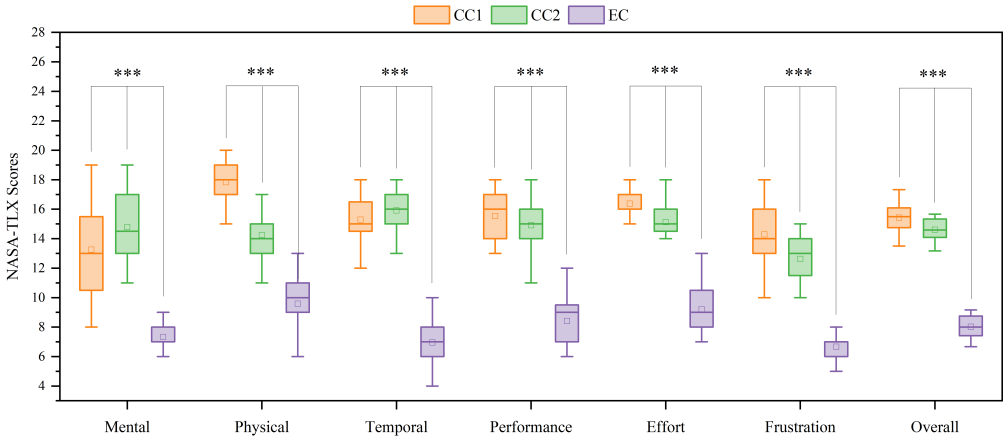


Figure 25: NASA-TLX Scores box plot (lower scores are better). The scores for the NASA-TLX questionnaire are tallied and the scores are stacked to calculate a total score. The individual scores for each dimension are tallied to calculate a total score, which is divided by 6 for analysis. Significant differences in pairwise comparisons are marked with ***, representing a Bonferroni-adjusted significance level of .001.

163.185, $p < .001$, $\eta_p^2 = .825$), Temporal ($F_{2,69} = 244.209$, $p < .001$, $\eta_p^2 = .876$), Performance ($F_{2,69} = 138.883$, $p < .001$, $\eta_p^2 = .801$), Effort ($F_{2,69} = 110.209$, $p < .001$, $\eta_p^2 = .762$), Frustration ($F_{2,69} = 137.234$, $p < .001$, $\eta_p^2 = .799$), and Overall ($F_{2,69} = 551.229$, $p < .001$, $\eta_p^2 = .941$). Specifically, the results of the one-way ANOVA test indicated that EC showed significant improvements in the Mental ($p < .001$, $\eta_p^2 = .530$; $p < .001$, $\eta_p^2 = .642$), Physical ($p < .001$, $\eta_p^2 = .825$; $p < .001$, $\eta_p^2 = .601$), Temporal ($p < .001$, $\eta_p^2 = .831$; $p < .001$, $\eta_p^2 = .850$), Performance ($p < .001$, $\eta_p^2 = .766$; $p < .001$, $\eta_p^2 = .732$), Effort ($p < .001$, $\eta_p^2 = .737$; $p < .001$, $\eta_p^2 = .656$), Frustration ($p < .001$, $\eta_p^2 = .783$; $p < .001$, $\eta_p^2 = .687$), and Overall ($p < .001$, $\eta_p^2 = .930$; $p < .001$, $\eta_p^2 = .913$) compared to CC1 and CC2.

UEQ. Figure 26 presents the results of the UEQ questionnaire, which have been categorized into six distinct evaluation metrics. The results of the RM-ANOVA test indicate significant effects among the different methods in six areas: Attractiveness ($F_{2,69} = 284.012$, $p < .001$, $\eta_p^2 = .892$), Perspicuity ($F_{2,69} = 9.723$, $p < .001$, $\eta_p^2 = .220$), Efficiency ($F_{2,69} = 468.177$, $p < .001$, $\eta_p^2 = .931$), Dependability ($F_{2,69} = 374.543$, $p < .001$, $\eta_p^2 = .916$), Stimulation ($F_{2,69} = 309.249$, $p < .001$, $\eta_p^2 = .900$), and Novelty ($F_{2,69} = 249.512$, $p < .001$, $\eta_p^2 = .879$). Specifically, EC shows significant improvements over CC1 and CC2 in Attractiveness ($p < .001$, $\eta_p^2 = .872$; $p < .001$, $\eta_p^2 = .845$), Efficiency ($p < .001$, $\eta_p^2 = .915$; $p < .001$, $\eta_p^2 = .906$), Dependability ($p < .001$, $\eta_p^2 = .852$; $p < .001$, $\eta_p^2 = .909$), Stimulation ($p < .001$, $\eta_p^2 = .878$; $p < .001$, $\eta_p^2 = .861$), and Novelty ($p < .001$, $\eta_p^2 = .878$; $p < .001$, $\eta_p^2 = .708$). In perspicuity, EC demonstrates a significant enhancement compared to CC2 ($p < .001$, $\eta_p^2 = .171$) while being on par with CC1. Figure 27 further consolidates the results of the UEQ scales into Attractiveness, Pragmatic Quality, and Hedonic Quality, presenting the findings accordingly. The

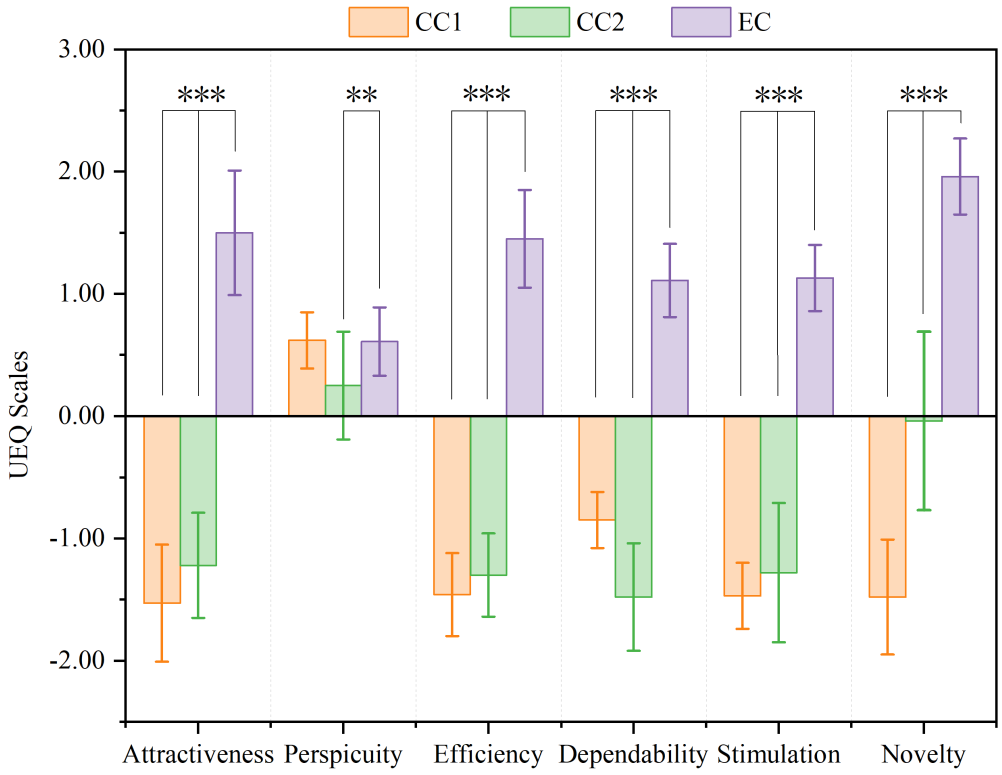


Figure 26: UEQ Scales bar chart. This bar chart shows the 26 questions of the UEQ questionnaire summarized into 7 relevant outcomes ranging from -3 (horribly bad) to 3 (extremely good). The error bars in this figure represent standard errors. Significant differences in pairwise comparisons are marked with **, and ***, representing a Bonferroni-adjusted significance level of .01, and .001, respectively.

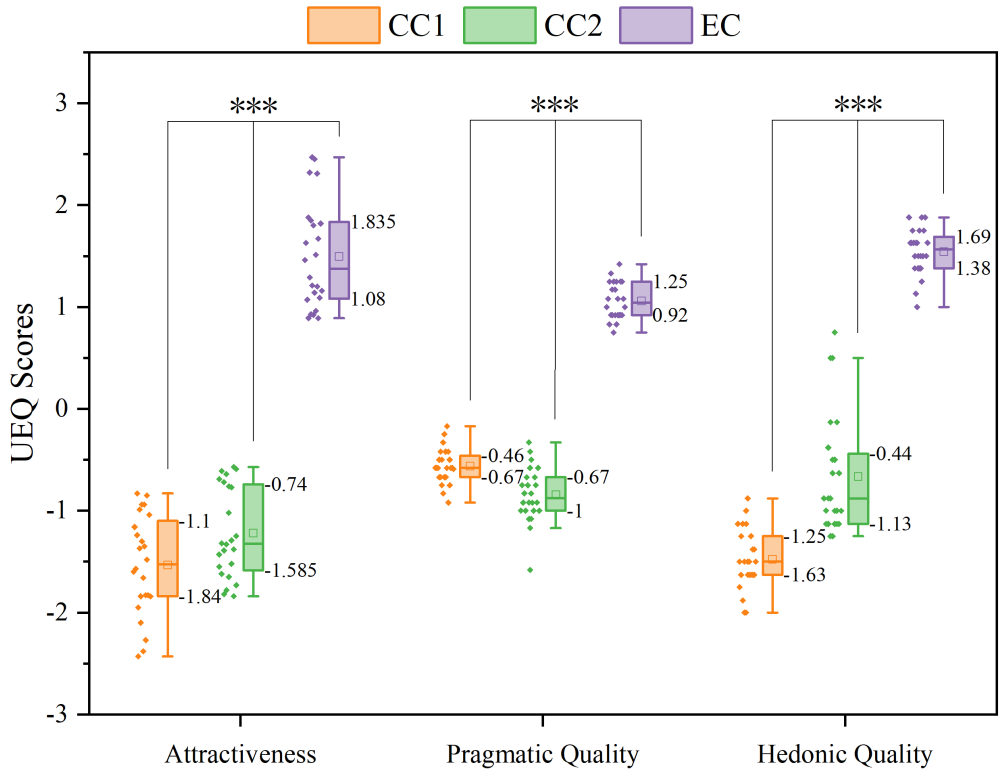


Figure 27: UEQ Scores box plot across different quality dimensions. This figure presents the results of the UEQ scores across three key dimensions: Attractiveness, Pragmatic quality (Perspicuity, Efficiency, Dependability), and Hedonic quality (Stimulation, Originality). Each box plot displays the median value, interquartile range, and individual data points for the corresponding dimension. Significant differences in pairwise comparisons are marked with ***, representing a Bonferroni-adjusted significance level of .001.

findings from the RM-ANOVA test reveal significant variations among the different methods regarding Attractiveness ($F_{2,69}$ = 284.012, $p < .001$, η_p^2 = .892), Pragmatic Quality ($F_{2,69}$ = 560.220, $p < .001$, η_p^2 = .942), and Hedonic Quality ($F_{2,69}$ = 369.481, $p < .001$, η_p^2 = .915). Specifically, EC has achieved significant improvements over CC1 and CC2 in Attractiveness ($p < .001$, η_p^2 = .872; $p < .001$, η_p^2 = .845), Pragmatic Quality ($p < .001$, η_p^2 = .910; $p < .001$, η_p^2 = .933), and Hedonic Quality ($p < .001$, η_p^2 = .909; $p < .001$, η_p^2 = .842).

SSQ. Figure 28 presents the results of the SSQ scores. The results of the ANOVA test showed that significant differences were exhibited in the SSQ results of different methods ($F_{3,92}$ = 458.593, $p < .001$, η_p^2 = .937). The results of the one-way ANOVA test indicate that, although EC does induce some sickness compared to BS, it demonstrates a significant improvement in reducing sickness when compared to both CC1 ($p < .001$, η_p^2 = .715) and CC2 ($p < .001$, η_p^2 = .776).

8. Discussion

In the first user study, we evaluated the efficiency and utility of the MCBOS method in high occlusion VR environments through two hypotheses **H1** and **H2**. In the second user study, we validated the applicability and scalability of MCBOS in practical application scenarios in relation to hypothesis **H3**. We proved these hypotheses using both objective and subjective data, providing detailed supporting evidence.

H1 posited that MCBOS would significantly enhance selection efficiency in high-occlusion VR environments. From the completion time data in Table 2, MCBOS demonstrated significant efficiency improvements across all high density scenarios. Error rate data (Table 3) also supported this hypothesis, with MCBOS

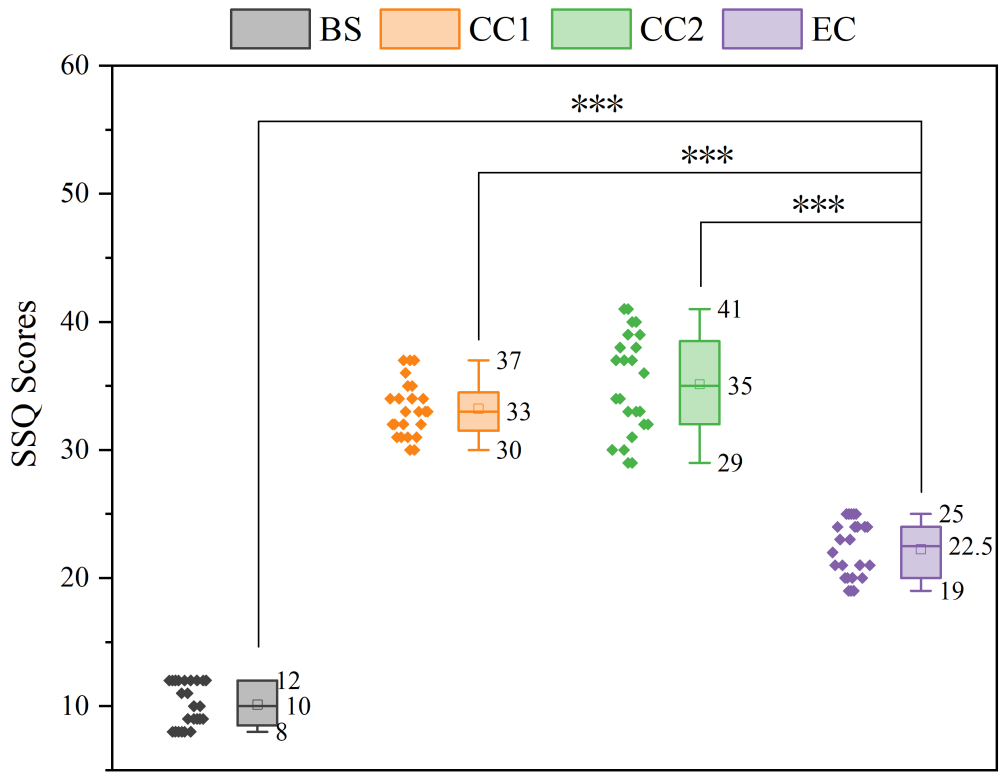


Figure 28: SSQ scores box plot (lower scores are better). This figure presents box plots of the SSQ scores for three methods. BS indicates the participants' SSQ questionnaire before the start of the experiment. On the left, individual data points illustrate the distribution of SSQ scores for each method. On the right, the box plots show the interquartile range and median scores for each method. Significant differences in pairwise comparisons are marked with ***, representing a Bonferroni-adjusted significance level of .001.

showing an error rate of 0.03 in the ”Near High” scenario, significantly lower than the 0.28 and 0.13 of CC1 and CC2. Cohen's d-values and p-values further indicated the statistical significance and large effect size of these results. The improvement in these two metrics is expected. MCBOS assists users in achieving de-occlusion of objects in complex, highly occluded environments, allowing for a greater visible area of the occluded objects within our field of view. As a result, users can complete target selection tasks more easily without interference from the occluding objects, reducing selection time and significantly lowering the error rate in the process. Similarly, our data also reflect this trend. These data demonstrated that MCBOS significantly improves the selection efficiency in high-occlusion VR environments, thus proving our proposed hypothesis **H1**.

H2 posited that MCBOS would significantly improve its utility in high-occlusion VR environments. From the data in Table 4 and Table 5, MCBOS significantly reduced the total rotation and total translation value during the selection task across all four scenarios. Next, the NASA-TLX scores (Figure 18) indicated that users experienced significantly lower task load across all dimensions when using MCBOS compared to the CC1 and CC2. Then, from the UEQ data in Figure 19 and Figure 20, MCBOS showed a significant advantage over CC1 and CC2 in the dimensions of Attractiveness, Efficiency, Dependability, Stimulation, and Novelty without a loss of Perspicuity. Furthermore, the SSQ data (Figure 21) showed that users reported significantly fewer symptoms of sickness and discomfort when using MCBOS. This indicated that the design of MCBOS effectively reduced unnecessary head and body movements, thereby minimizing users' sickness when they were using HMD. Finally, In the rankings of the three methods (Figure 22), 92% of the participants considered MCBOS to be the first ranking of the three methods.

The improvement in the aforementioned metrics is also within our expectations. MCBOS enables users to achieve de-occlusion of obstructed objects, allowing them to make selections by viewing only a smaller portion of the objects. This reduces the need for extensive translation and head rotation to expose the objects in their field of view. Additionally, the reduction in physical movement, coupled with a more novel interaction method, lowers the users' task load, decreases SSQ scores, and enhances UEQ performance. All the above data demonstrated that MCBOS significantly improved utility in high-occlusion VR environments, thus proving our proposed hypothesis **H2**.

H3 posits that MCBOS has strong applicability in practical application scenarios. According to the data from the user experiments, MCBOS demonstrated excellent performance in selection tasks within actual industrial environments. Both objective and subjective metrics showed significant improvements for MCBOS compared to the other two methods. In terms of objective metrics Table 6, the experimental results revealed that MCBOS significantly improved Completion time, Total rotation, and Total translation. Users were able to complete selection tasks more accurately and efficiently using MCBOS, which validates its applicability in practical application scenarios. For subjective metrics, MCBOS significantly reduced task load Figure 25 and SSQ scores Figure 28, while also enhancing UEQ ratings Figure 27. This indicates that MCBOS performs well in theory and provides a user-friendly interaction experience in practical applications, supporting its applicability across different environments. These results can be attributed to the de-occlusion mechanism and the novel interaction design we implemented for MCBOS. Since our method can remove obstructions in front of the target object, users of MCBOS can start selecting as soon as they approximate the location of

the target. Unlike other methods that require precise localization and exposure of the object's surface area for selection, our method reduces the time needed to locate the object and decreases the additional translation distance and rotation angle required to expose the object's visible area. The data presented by objective results also corroborate this conclusion. Furthermore, the reductions in translation distance and rotation angle result in lower scores on the NASA-TLX and SSQ indices. With the decrease in selection time, translation distance, and rotation angle, there is a corresponding increase in the UEQ indices. Our subjective results also reveal this trend. The data strongly supports our proposed hypothesis **H3**, confirming the applicability of MCBOS in practical applications.

9. Limitations and future work

MCBOS provides a way to accurately select occluded objects in high occlusion environments, but there are still many limitations and future work for improvement. First, selecting distant objects poses a challenge. Although MCBOS can fully expose distant objects in the screen space, these objects occupy a small area on the screen, making accurate selection difficult. In the future, we can address the challenge of selecting distant objects by constructing a gesture-operated three-dimensional space. This would enable MCBOS to operate in 3D space, achieving a magnification effect for distant objects. Second, the hand gesture is detected using the HMD, thus the hand must be within the detectable area of the HMD, limiting the area in which the hand can move. As a follow-up, we can add a third-party camera to the scene to detect the hand pose, thereby increasing the hand's activity area in the scene. Third, similar to other bare hand interaction techniques, our approach is limited by the hand recognition accuracy of the

HMD, which affects the accuracy of our selection interaction. To overcome this limitation, we can introduce new recognition modalities, such as eye tracking or large-scale language models, to assist our choices and compensate for the lack of bare hand recognition accuracy. Fourth, all of our participants in this experiment were right-handed, which may have resulted in some of the experimental data being less persuasive. In future work, we will also consider introducing left-handed participants and studying their adaptation and performance to the method. Finally, we recognize that the proposed party may have a certain learning curve, i.e., the user needs to absorb a large amount of information in a short period of time. This complexity may cause some difficulties to the user's operation in practical applications and affect its usability in practical application scenarios. To mitigate this problem, we propose to provide more detailed user guidance and training materials or develop more intuitive user interfaces in future designs to help users grasp the core functionality more quickly, so as to enhance the operability and acceptance of the method.

10. Conclusion

We have proposed a manipulable cone based bare hand object selection method in a high occlusion virtual environment. We introduce a new concept of the manipulable cone, which can be used to manipulate the subspace and the objects inside the cone to achieve dis-occlusion and propose its construction and manipulation methods. Then, we design a bare hand interaction based on manipulable cones for object selection. Three main gestures are used to switch between the five states during the selection process. During the selection process, we use hand translation to manipulate the apex of the manipulable cone, thus manipulating

the virtual objects in the cone, exposing the target objects occluded by these objects, and then performing the target selection. Next, we prove the efficiency and utility of our method through a user study. Compared to the existing SOTA methods, our method significantly reduces task completion time, total rotations, total translations, error rate, task load, and SSQ scores in high occlusion environments. Additionally, our method significantly improves the UEQ scores and performance ranking. Finally, we validated the applicability of our method through a user study in a practical application scenario. Our approach maintained high efficiency and adaptability under complex conditions, effectively handling occluded object selection tasks without performance degradation. These results confirm its robustness and practicality for practical applications. The results show that, compared to SOTA methods, our approach consistently improves task efficiency, reduces cognitive load, and enhances user experience, demonstrating its potential for broader VR applications.

11. Acknowledgment

This work is supported by the National Natural Science Foundation of China through Projects 61932003 and 62372026, Beijing Science and Technology Plan Project Z221100007722004, and the National Key R&D plan 2019YFC1521102.

References

Baloup, M., Pietrzak, T., Casiez, G., 2019. Raycursor: A 3d pointing facilitation technique based on raycasting, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery,

New York, NY, USA. p. 1–12. URL: https://doi.org/10.1145/3290605.3300331, doi:10.1145/3290605.3300331.

Bergström, J., Dalsgaard, T.S., Alexander, J., Hornbæk, K., 2021. 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, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3411764.3445193, doi:10.1145/3411764.3445193.

Bhowmick, S., Biswas, N., Kalita, P.C., Sorathia, K., 2023. Design and evaluation of amaze: A multi-finger approach to select small and distant objects in dense virtual environments. Displays 80, 102539. URL: https://www.sciencedirect.com/science/article/pii/S0141938223001725, doi:https://doi.org/10.1016/j.displa.2023.102539.

Bowman, D.A., Kruijff, E., Laviola, J.J., Poupyrev, I., 2004. 3d user interfaces: Theory and practice. URL: https://api.semanticscholar.org/CorpusID:60454016.

Cashion, J., Wingrave, C., LaViola Jr., J.J., 2012. Dense and dynamic 3d selection for game-based virtual environments. IEEE Transactions on Visualization and Computer Graphics 18, 634–642. doi:10.1109/TVCG.2012.40.

Chen, K., Wan, H., Zhao, S., Liu, X., 2023. Backtracer: Improving ray-casting 3d target acquisition by backtracking the interaction history. Int. J. Hum.-Comput. Stud. 176. URL: https://doi.org/10.1016/j.ijhcs.2023.103045, doi:10.1016/j.ijhcs.2023.103045.

Cohen, J., 2013. Statistical power analysis for the behavioral sciences. Routledge.

Delamare, W., Daniel, M., Hasan, K., 2022. Multifingerbubble: A 3d bubble cursor variation for dense environments, in: Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3491101.3519692, doi:10.1145/3491101.3519692.

Elmqvist, N., 2005. Balloonprobe: Reducing occlusion in 3d using interactive space distortion, in: Proceedings of the ACM symposium on Virtual reality software and technology, pp. 134–137.

Elmqvist, N., Tudoreanu, M.E., 2007. Occlusion management in immersive and desktop 3d virtual environments: Theory and evaluation. Int. J. Virtual Real. 6, 21–32.

Grossman, T., Balakrishnan, R., 2006a. The design and evaluation of selection techniques for 3d volumetric displays, in: Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology, Association for Computing Machinery, New York, NY, USA. p. 3–12. URL: https://doi.org/10.1145/1166253.1166257, doi:10.1145/1166253.1166257.

Grossman, T., Balakrishnan, R., 2006b. The design and evaluation of selection techniques for 3d volumetric displays, in: Proceedings of the 19th annual ACM symposium on User interface software and technology, pp. 3–12.

de Haan, G., Koutek, M., Post, F.H., 2005. Intenselect: using dynamic object rating for assisting 3d object selection, in: Proceedings of the 11th Eurographics

Conference on Virtual Environments, Eurographics Association, Goslar, DEU. p. 201–209.

Hart, S.G., 2006. Nasa-task load index (nasa-tlx); 20 years later, in: Proceedings of the human factors and ergonomics society annual meeting, Sage publications Sage CA: Los Angeles, CA. pp. 904–908.

Kennedy, R.S., Lane, N.E., Berbaum, K.S., Lilienthal, M.G., 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. The international journal of aviation psychology 3, 203–220.

Kim, W., Xiong, S., 2022. Viewfindervr: configurable viewfinder for selection of distant objects in vr. Virtual Reality 26, 1573–1592.

Kopper, R., Bacim, F., Bowman, D.A., 2011. Rapid and accurate 3d selection by progressive refinement, in: 2011 IEEE Symposium on 3D User Interfaces (3DUI), pp. 67–74. doi:10.1109/3DUI.2011.5759219.

Li, N., Zhang, Z., Liu, C., Yang, Z., Fu, Y., Tian, F., Han, T., Fan, M., 2021. Vmirror: Enhancing the interaction with occluded or distant objects in vr with virtual mirrors, in: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–11.

Li, Y., Sarcar, S., Kim, K., Tu, H., Ren, X., 2022. Designing successive target selection in virtual reality via penetrating the intangible interface with handheld controllers. Int. J. Hum.-Comput. Stud. 165. URL: https://doi.org/10.1016/j.ijhcs.2022.102835, doi:10.1016/j.ijhcs.2022.102835.

Lu, Y., Yu, C., Shi, Y., 2020. Investigating bubble mechanism for ray-casting to improve 3d target acquisition in virtual reality, in: 2020 IEEE Conference

on Virtual Reality and 3D User Interfaces (VR), pp. 35–43. doi:10.1109/VR46266.2020.00021.

Lystbæk, M.N., Pfeuffer, K., Grønbæk, J.E.S., Gellersen, H., 2022a. Exploring gaze for assisting freehand selection-based text entry in ar. Proc. ACM Hum.-Comput. Interact. 6. URL: https://doi.org/10.1145/3530882, doi:10.1145/3530882.

Lystbæk, M.N., Rosenberg, P., Pfeuffer, K., Grønbæk, J.E., Gellersen, H., 2022b. Gaze-hand alignment: Combining eye gaze and mid-air pointing for interacting with menus in augmented reality. Proc. ACM Hum.-Comput. Interact. 6. URL: https://doi.org/10.1145/3530886, doi:10.1145/3530886.

Maslych, M., Hmaiti, Y., Ghamandi, R., Leber, P., Kattoju, R.K., Belga, J., LaViola, J.J., 2023. Toward intuitive acquisition of occluded vr objects through an interactive disocclusion mini-map, in: 2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR), IEEE. pp. 460–470.

Popescu, V., Rosen, P., Adamo-Villani, N., 2009. The graph camera. ACM Trans. Graph. 28, 1–8. URL: https://doi.org/10.1145/1618452.1618504, doi:10.1145/1618452.1618504.

Ro, H., Chae, S., Kim, I., Byun, J., Yang, Y., Park, Y., Han, T., 2017. A dynamic depth-variable ray-casting interface for object manipulation in ar environments, in: 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 2873–2878. doi:10.1109/SMC.2017.8123063.

Schjerlund, J., Hornbæk, K., Bergström, J., 2021. Ninja hands: Using many hands to improve target selection in vr, in: Proceedings of the 2021 CHI Con-

ference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3411764.3445759, doi:10.1145/3411764.3445759.

Schrepp, M., Hinderks, A., Thomaschewski, J., 2017. Design and evaluation of a short version of the user experience questionnaire (ueq-s). Int. J. Interact. Multim. Artif. Intell. 4, 103–108. URL: https://api.semanticscholar.org/CorpusID:2315053.

Shi, R., Zhang, J., Yue, Y., Yu, L., Liang, H.N., 2023. Exploration of bare-hand mid-air pointing selection techniques for dense virtual reality environments, in: Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3544549.3585615, doi:10.1145/3544549.3585615.

Sidenmark, L., Clarke, C., Zhang, X., Phu, J., Gellersen, H., 2020. Outline pursuits: Gaze-assisted selection of occluded objects in virtual reality, in: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. p. 1–13. URL: https://doi.org/10.1145/3313831.3376438, doi:10.1145/3313831.3376438.

Vanacken, L., Grossman, T., Coninx, K., 2007. Exploring the effects of environment density and target visibility on object selection in 3d virtual environments, in: 2007 IEEE Symposium on 3D User Interfaces. doi:10.1109/3DUI.2007.340783.

Wagner, U., Lystbæk, M.N., Manakhov, P., Grønbæk, J.E.S., Pfeuffer, K., Gellersen, H., 2023. A fitts' law study of gaze-hand alignment for selection in 3d user interfaces, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3544548.3581423, doi:10.1145/3544548.3581423.

Wang, L., Chen, J., Ma, Q., Popescu, V., 2021. Disocclusion headlight for selection assistance in vr, in: 2021 IEEE Virtual Reality and 3D User Interfaces (VR), pp. 216–225. doi:10.1109/VR50410.2021.00043.

Wang, L., Wu, J., Yang, X., Popescu, V., 2019. Vr exploration assistance through automatic occlusion removal. IEEE transactions on visualization and computer graphics 25, 2083–2092.

Wei, Y., Shi, R., Yu, D., Wang, Y., Li, Y., Yu, L., Liang, H.N., 2023. Predicting gaze-based target selection in augmented reality headsets based on eye and head endpoint distributions, in: Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, Association for Computing Machinery, New York, NY, USA. URL: https://doi.org/10.1145/3544548.3581042, doi:10.1145/3544548.3581042.

Wu, H., Sun, X., Tu, H., Zhang, X., 2023. Clockray: A wrist-rotation based technique for occluded-target selection in virtual reality. IEEE Transactions on Visualization and Computer Graphics , 1–12doi:10.1109/TVCG.2023.3239951.

Wu, J., Wang, L., Zhang, H., Popescu, V., 2022. Quantifiable fine-grain occlusion removal assistance for efficient vr exploration. IEEE Transactions on Visu-

alization and Computer Graphics 28, 3154–3167. doi:10.1109/TVCG.2021.3053287.