

FocalSelect: Improving Occluded Objects Acquisition with Heuristic Selection and Disambiguation in Virtual Reality

Duotun Wang , Linjie Qiu , Boyu Li , Qianxi Liu , Xiaoying Wei , Jianhao Chen , Zeyu Wang , and Mingming Fan

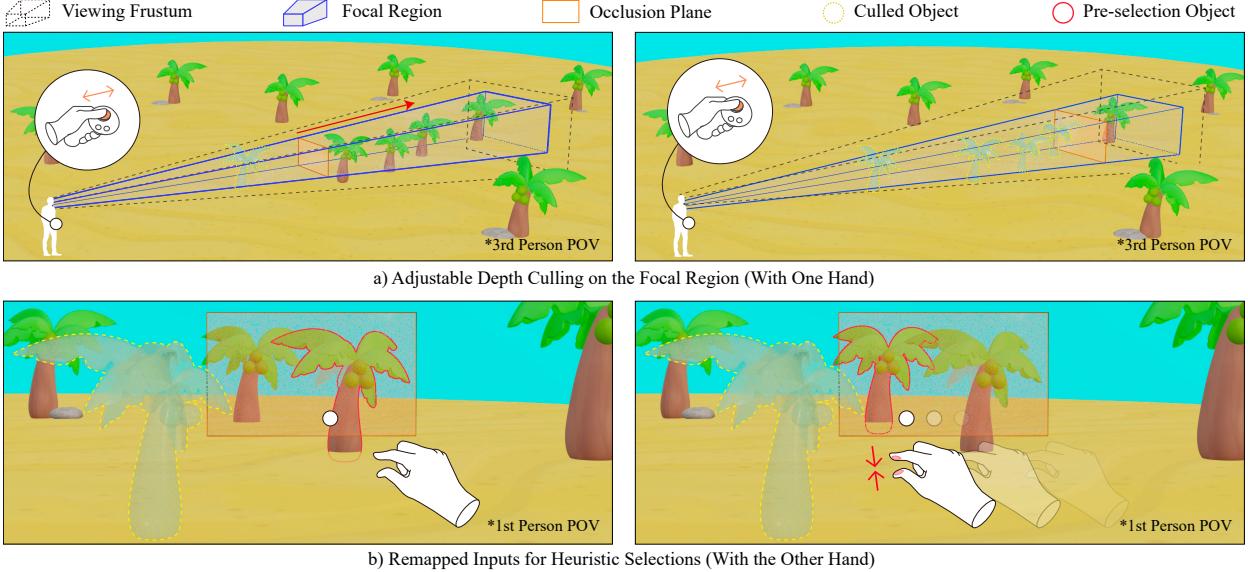


Fig. 1: *FocalSelect* enhances object selection in occluded scenarios through a) adjustable depth culling that reveals selectable areas within the focal region—a subarea of the viewing frustum—and b) remapped inputs for heuristic selections with hybrid inputs: one controller and one bare-hand. Our technique preserves the scene context and objects' layout during the selection process.

Abstract—In recent years, various head-worn virtual reality (VR) techniques have emerged to enhance object selection for occluded or distant targets. However, many approaches focus solely on ray-casting inputs, restricting their use with other input methods, such as bare hands. Additionally, some techniques speed up selection by changing the user's perspective or modifying the scene context, which may complicate interactions when users plan to resume or manipulate the scene afterward. To address these challenges, we present *FocalSelect*, a heuristic selection technique that builds 3D disambiguation through head-hand coordination and scoring-based functions. Our interaction design adheres to the principle that the intended selection range is a small sector of the headset's viewing frustum, allowing optimal targets to be identified within this scope. We also introduce a density-aware adjustable occlusion plane for effective depth culling of rendered objects. Two experiments are conducted to assess the adaptability of *FocalSelect* across different input modalities and its performance against five selection techniques. The results indicate that *FocalSelect* enhances selection experiences in occluded and remote scenarios while preserving the spatial context among objects. This preservation helps maintain users' understanding of the original scene and facilitates further manipulation. We also explore potential applications and enhancements to demonstrate more practical implementations of *FocalSelect*.

Index Terms—Virtual reality, interaction techniques, interaction design

1 INTRODUCTION

The rapid development of VR devices (e.g., Quest3 [42] and Pico4 [51]) enable input techniques tailored to various applications, including pen-

- *Duotun Wang, Linjie Qiu, Boyu Li, Qianxi Liu, Xiaoying Wei, Jianhao Chen, Zeyu Wang, and Mingming Fan are with The Hong Kong University of Science and Technology (Guangzhou). E-mail: dwang866, lqiu250, blibr, qliu930, xweias, jchen628@connect.hkust-gz.edu.cn*
- *Zeyu Wang, and Mingming Fan are also with The Hong Kong University of Science and Technology. E-mail: zeyuwang, mingmingfan@ust.hk*
- *Duotun Wang and Linjie Qiu equally contributed to this work.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

guided interactions [12, 32] and gesture controls [24, 68]. Although ray-casting remains dominant for object selections, it becomes less effective when dealing with distant or occluded objects (e.g., lever effect), making it cumbersome to engage in tasks that require fine-grained controls such as high-dimensional visualizations [34, 61] and 3D authoring [50, 72]. In these cases, addressing challenges like the Heisenberg effect and increased physical workloads becomes crucial [66].

Disambiguation mechanisms have been meticulously examined to identify the intended target among a group of objects in various contexts [4, 38, 71]. Recent studies have proposed two-stage progressive refinements such as *LassoGrid+* [71], *Expand* [8], and *SQUAD* [29] are advantageous for facilitating rapid selections. Nonetheless, these techniques alter the spatial relationships among objects, which can lead to complications. For example, if spatial information cannot be used for differentiation, identifying a specific object from a collection of similarly sized and colored items becomes challenging. Alternative

solutions [39, 71] have been developed to maintain the object location information, albeit at the expense of recovery of scene contexts or selection speed. Another area of improvement focuses on augmenting the degree of freedom (DoF) in ray-casting methods. For instance, *DepthRay* [21] and *RayCursor* [4] incorporate depth-based disambiguation along the ray pointing, enhancing the precision of selection tasks. However, their applicability is limited, as they are not easily adaptable to other input modalities, such as bare hands [37]. Furthermore, to reduce the Heisenberg effect and selection time in ray-casting, heuristic approaches (e.g., *BubbleRay* [35] and *IntenSelect* [11]) have been employed. These methods predict potential optimal selections through scoring-based functions. However, the accuracy of predictions can be significantly compromised by occluded situations. In this work, we aim to refine disambiguation methods that maintain original object locations [38, 39, 71], accommodate non-accuracy selections, and enhance compatibility with various input modalities.

We introduce *FocalSelect*, a heuristic disambiguation technique that leverages the metaphor of focal point and aperture [19] to enhance occlusion revelation and selection accuracy while preserving scene context information. Our design is influenced by the **focus+context way** [16], which suggests that users' attention and the intended selection areas do not encompass the entire rendered scenes. Building upon this concept, *vMirror* [33] extends the focus+context design to ray-casting selections, and *FocalPoint* [38] effectively integrates heuristic disambiguation for selecting nearby objects through the use of a virtual cylinder as the focal region in Augmented Reality (AR). However, *FocalPoint* focuses solely on single bare-hand input. In contrast, our work investigates how this heuristic approach can be applied to ray-casting input and two-hand interactions. Besides, we take an initial step toward adapting the focal region design for VR by incorporating a density-aware adjustable occlusion plane for depth culling. As shown in Fig. 1a, we assume the focal region as the intended selection area, which is the subarea of viewing frustum from the headset's single eye. The location of the focal region strictly follows the movements from head tracking [54, 55]. The fundamental design that enables *FocalSelect* to be effortlessly extended to various input modalities is the effective mapping of input movements, either from controllers or bare hands, to the occlusion plane. Subsequently, this remapped input can execute scoring-based selections to predict the optimal target behind the occlusion plane (Fig. 1b).

To refine the design parameters of our technique, such as depth-culling movement control, focal region size, and visual indicators, we initially conducted a preliminary study among professional VR developers using *FocalSelect* (Fig. 3b). We proceeded with a within-subject experiment to evaluate *FocalSelect*'s performance across different input modalities (Section 5). In the second experiment, we further investigated *FocalSelect*'s performance to assist in selection tasks, comparing it to recent selection techniques that preserve object location information, which are *AlphaCursor* [71], *GravityZone+* [71], *IntenSelect* [11], and *CylinderPIM* [39] (Section 6). Our results indicate that in highly-occluded and remote scenarios, our technique performs on par with *GravityZone+*, while surpassing *CylinderPIM* and *AlphaCursor* in selection accuracy. We further discuss practical applications of our technique. In summary, our work makes the following contributions:

- We propose *FocalSelect*, a heuristic disambiguation technique that facilitates selections in occluded scenarios while preserving scene context and supporting various input modalities.
- We refined the visual design of *FocalSelect* based on a preliminary study and conducted two within-subject experiments to evaluate its performance quantitatively and qualitatively.

2 RELATED WORKS

According to Fitts's law [56] and the Heisenberg effect [66], selection time and accuracy are heavily influenced by the target size, the distance to the target, and disturbed inputs on the tracked devices. Our work is mainly inspired by previous research on improving selection experiences in VR through disambiguation and visualization techniques (Section 2.1). Acknowledging that no single technique is universally

applicable to all interaction scenarios, we also review the literature on experimental design and the establishment of plausible metrics for object selection tasks (Section 2.2).

2.1 Objects Selection and Disambiguation Techniques

To improve immersive experiences with accurate and efficient interactions, many surveys have proposed novel designs of selection techniques in VR [69, 70]. We adopt a two-stage selection strategy: first, users indicate an area of interest, then select a specific item within that area. In this work, we review prior research on improving each stage and derive insights to inform the detailed design of *FocalSelect* (i.e., head-pointing, depth-culling, input mapping, and heuristic selections).

Progressing from the ray-casting method, where the first object intersecting with the ray is chosen, Argelaguet et al. [2] presented cone-based approaches for group selections. Elaborated by Maslych et al. [39], a straightforward strategy for mitigating occluded issues is to change the user's viewing perspectives. *PRECIOUS* [41] can teleport the user closer to the target to bypass occluded areas. To minimize the risk of motion sickness, *vMirror* [33] employed a mirror reflection mechanism to reduce the frequency of teleportation. Some other techniques, such as *BaloonProbe* [15] and *SmashProbe* [71], enhanced selections in dense and occluded scenarios by moving occluding objects away from the cursor's path. Yu et al. [71] suggested *LassoGrid+*, where users can highlight an area of objects to be organized into a grid. However, these designs may inadvertently result in the loss of environmental context. For instance, *Expand* [8] and *SQUAD* [29] transformed target objects. Although *GravityZone+* [71] can maintain the relative positions of objects with quick selections, restoring the entire scene after selections necessitates additional precautions, such as pre-recording the scene context.

Our disambiguation design for the focal region adheres to the widely recognized interaction strategy where users can adjust their position to enhance their view of an intended target [54]. This strategy reduces the demand for users' motor skills (e.g., mitigating unexpected hand tremor [67]). Numerous studies have explored hands-free alternatives to pointing and selection, utilizing integrated head tracking [25, 52]. Head pointing can serve as a stable and controlled method for coarse selections, often followed by gaze-based selections via eye tracking for confirmation [6, 63]. Our design of the focal region identifies users' intended selection areas by **incorporating head tracking for cone-casting** as a preliminary selection step (Fig. 2).

To enhance selection experiences while preserving the scene contexts as much as possible, *CylinderPIM* [39] refined progressive selection approaches by incorporating the occlusion mini-map. Wang et al. [65] followed the *Multiple View ports* [14] design pattern by integrating multiple viewports into one image utilizing graph cameras [47]. This design pattern has been applied in various forms, including *Worlds-in-miniature* [58], *Bird's eye views* [20], and *Worldlets* [17]. Our depth-based occlusion plane builds on insight from the design pattern of *Virtual X-Ray* [14], which has inspired a series of research studies in desktop [46, 59], AR [38], and VR [10, 71] contexts to improve disambiguation. This pattern is achieved by configuring virtual objects that occlude other objects, to appear **invisible or semi-transparent**, thereby providing users with greater awareness of the occluded targets and surrounding environments. Wagner et al. [64] follow this design pattern and propose to combine a pointing-controlled occlusion plane to improve gaze-based selections. *FocalSelect* restricts the movement of the occlusion plane with headset tracking for rendering-guided culling, focusing solely on adjusting the front-back relationship of the occlusion plane and naturally reducing pointing effort. Additionally, we adapt the speed-dependent design proposed by Igarashi et al. [26] to regulate the movement speed of the occlusion plane, as detailed in Section 3.2.

For input modalities, direct manipulation offers enhanced flexibility for inputs beyond ray-casting (e.g., bare-hand [37]). This mechanism is contingent on the virtual representation of input devices colliding with the selectable virtual objects [7]. For instance, *Go-Go* [48] facilitated distant selections by adjusting the position of the virtual hand, and *NearField* [3] allowed for direct manipulations of target replica within the arm's reach. Nevertheless, further modifications are required when

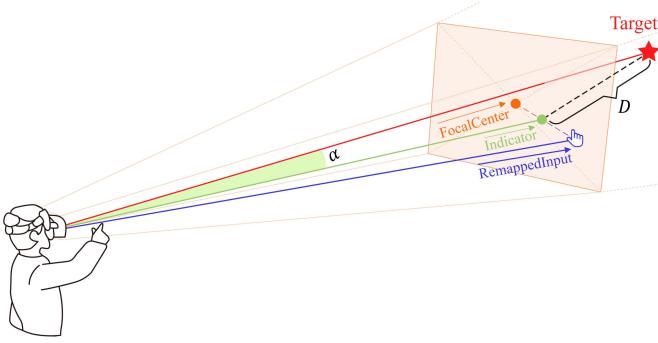


Fig. 2: A visual representation of variables used in computing S_{obj} to determine the optimal target. The occlusion plane is positioned parallel to the near plane of the viewing frustum, allowing the user's inputs to be mapped onto this plane. The resulting indicator assists in selecting targets that are beyond the arm's reach.

targets are partially or fully occluded. In addition to hands-free techniques, our work is heavily inspired by investigations on input mapping to enhance selections beyond the arm stretch [18, 23, 48]. For instance, *Ninja Hands* [53] proposed to map the movement of a single hand to multiple hands, effectively minimizing the distance to the targets. Our work delves further into this mapping concept to eliminate the constraints imposed by input modalities. Consequently, both ray-casting and virtual hand inputs can be seamlessly integrated with *FocalSelect*.

Heuristic methods with scoring-based functions, such as *IntenSelect* [11], *BubbleRay* [35], and *Stick-Ray* [57], have been proposed to identify the optimal target among multiple candidates. *FocalSelect* integrates scoring-based selection to determine the optimal target after adjusting the focal region.

2.2 Evaluation of Object Selection Tasks

In-depth and extensive comparisons with existing methodologies enhance our understanding of a technique's practical usability, accessibility, stability, and versatility [31]. Such analysis enables researchers to engage in informed discussions about the potential advantages of proposed approaches within their specific contexts. For example, when using *BubbleRay* [35] in densely packed environments, its selection accuracy may be significantly compromised, and the snap-to strategy [9] proves to be more effective.

Bergström et al. [5] conducted a reviewed 20 years of studies on VR object selection and manipulation tasks, aiming to establish standards for future evaluations in this area. Poupyrev et al. [49] performed comparative studies to elucidate the relative strengths and weaknesses of various virtual interfaces, such as virtual hands and pointers, in the context of selection and positioning tasks within VR. Other research works have contributed to identifying independent and dependent variables for assessing human performance, including metrics like learning time, miss rates, selection areas, user preferences, physical loads, and feedback types [1, 36, 70].

Additionally, several studies have explored experimental designs for measuring human performance in VR-based interaction tasks through both quantitative and qualitative analysis. Prior research also explored experiment designs for measuring human performances in VR-based interaction tasks with quantitative and qualitative analysis [40, 73]. For instance, Lu et al. [35] built experiments under extreme conditions of high density and occlusion to discuss potentials and limitations in their techniques, while Zhu et al. [74] conducted experiments to evaluate each individual components of *PinchLens*' functionality. Our study focuses on improving selection experiences in occluded and relatively remote scenarios.

3 FOCALSELECT

Building on the derived insights from Section 2, we propose *FocalSelect*, a heuristic selection technique combining user-controlled direct disambiguation (Figs. 1 and 2). It employs a frustum as the primary

selection geometry, aligned with the headset's viewing frustum, to define the focal region. *FocalSelect* incorporates an occlusion plane that allows users to reveal occluded objects through depth-culling. Varying the opacity of objects also aids users in perceiving depth and scene layouts. Candidate objects identified by the focal region are ranked based on head movements and interaction inputs (e.g., virtual hands or ray-casting) remapped to the occlusion plane. The object with the highest ranking is highlighted and snapped to the user's hand upon selection confirmation. The following subsections reflect the sequential steps of *FocalSelect*'s operation pipeline.

3.1 Focal Region Driven by Head Movement

As illustrated in Figs. 1a and 3a, the focal region is represented as a frustum shape in world space, a small sector of the headset's viewing frustum. The far plane of the focal region aligns with that of the viewing frustum, and the occlusion plane, which users can control, determines the near plane (Sections 3.2 and 4.1). This design allows the user's head movements to establish the coarse range of focus in screen space, which is then used for subsequent selection refinement. Objects within and on the periphery of the focal region are considered **selectable candidates**.

3.2 Depth Culling with Occlusion Plane

Our occlusion plane design aims to enhance objects' disambiguation from a depth-based rendering perspective. As shown in Fig. 5, users can use the slider to control the movement of the occlusion plane, causing objects in front of it to be culled and unselectable with a semi-transparent rendering. Once the desired location is set, user input from either a hand or controller is mapped onto this plane along the direction defined by the headset and input. The remapped input and the center of the focal region are used to jointly determine the indicator for selecting the target through heuristic scoring computations. Specifically, the location of the indicator is defined as follows:

$$\overrightarrow{\text{Indicator}} = (1 - \mu)\overrightarrow{\text{RemappedInput}} + \mu\overrightarrow{\text{FocalCenter}}, \quad (1)$$

where μ represents the modulation factor that balances the head-hand coordination under different input modalities. The small value of μ indicates the hand movement is dominant, and the large value demonstrates a more significant impact of hand movement. When using bare-hand input, $\mu \in [0.2, 0.4]$ is adjusted upon the user's feedback during warm-up trials. If users report through the think-aloud way [62] that the indicator moves too quickly, we may increase μ to minimize the impact of hand movements. For hybrid input, $\mu \in [0.4, 0.9]$, reflects the increased importance of head movement. When using ray-casting input, μ is set to 0, indicating that ray-pointing exclusively determines the indicator's location. Section 4.1 further discusses how to refine the design of the occlusion plane for efficient selections based on users' preliminary feedback.

3.3 Selection Disambiguation via Scoring

After identifying a list of candidate objects using the head-movement-based frustum, *FocalSelect* then assigns a score to each object. The highest-ranked object is then highlighted with a yellow contour (Fig. 3) and snapped to the user's hand when a confirmation action is triggered. To achieve this, we combine recent spatial-temporal scoring approaches [11, 35, 38] to develop a new scoring function that continuously evaluates each candidate object. The score S_{obj} is calculated based on the two key factors: the distance D from the object to the indicator and the angle α between the object's center and the indicator relative to the center of the near plane of the viewing frustum at each frame t (Fig. 2). Specifically, the score S_{obj} is defined as follows:

$$G_{obj} = \frac{\mu}{D} + A \cos(K\alpha), \quad (2)$$

$$S_{obj}(t) = \lambda S_{obj}(t-1) + (1 - \lambda) \frac{G_{obj}}{\beta}. \quad (3)$$

G_{obj} computes the spatial scoring influence, which is then integrated into Equation 3. This equation plays a crucial role in retaining temporal

information while calculating the final score S_{obj} . The parameters λ , A , K , and β represent the time-dependent stickiness, angular and scaling coefficients, respectively. Inspired by the configurations of IntenSelect+ [30], λ is initialized to 0.5 to weight the past and current scores equally and can be fine-tuned during each participant's warm-up trials. In our formal experiments, we set A to 10 and K to 5 to help distinguish between objects in occluded cases. β was fixed to 1.1 to match the size of sphere objects used in our experiments. We recommend future works to explore the complexity of balancing the parameterization for diverse interaction scenarios. Notably, when ray-casting is used as input, μ is set to 0. This configuration means that only angular deviations impact the score S_{obj} , resulting in a scoring mechanism that closely resembles that of *IntenSelect* [11].

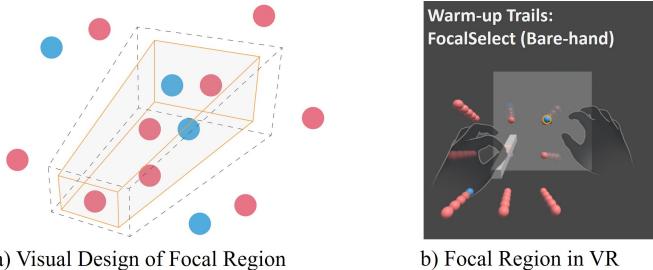


Fig. 3: a) Visual designs for the focal region feature a fixed size and transparency. b) Due to perspective projection, the focal region appears as a semi-transparent plane in VR.

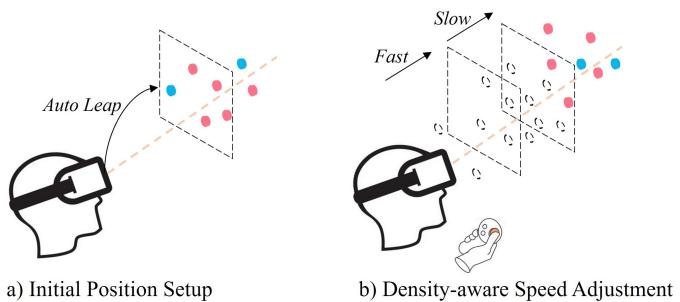


Fig. 4: Designs for the occlusion plane include a) starting position in front of the nearest object relative to the user; b) auto-adapting its movement speed based on the density of surrounding objects.

4 PRELIMINARY STUDY

We conducted a preliminary study to discuss the design choices for the visual presentations of our disambiguation (i.e., the focal region and occlusion plane). Effective visual guidance significantly enhances the feedback for participant to assess selection states and anticipate their subsequent actions, thus improving overall user performance [22]. Inspired by the iterative design process conducted by Lu et al. [35], we invited eight experienced participants (3 female and 5 male), aged 26 to 34 ($\bar{x} = 28.6$, $SD = 1.5$), each with over three years of experience in developing VR applications, to evaluate our disambiguation designs. As shown in Fig. 3b, each participant was asked to wear an Oculus Quest3 (per-eye resolution of 2064×2208 , refresh rate up to 120 Hz, horizontal field of view of 110°) to select 5 blue spheres from 45 red spheres arranged in a fixed grid layout. Participants were free to use either bare-hand or controller inputs. This spatialized layout was used for warm-up trials in the subsequent studies. No data from the preliminary study were included in the formal experiment analysis. We gathered feedback for refinement through semi-structured interviews.

4.1 Empirical Observations

Focal region. The focal region signifies that objects placed within it are likely selected. FocalPoint [38] has introduced the see-through design

of modeling a semi-transparent sphere to indicate the focal region. The benefit of semi-transparent rendering in revealing occluded objects has been demonstrated in *Control-Depth* [60] and *Virtual X-Ray* [14]. It is valuable to discuss the transparency setting to minimize the influence of color differentiation and depth perception [28]. If the focal region indicator is completely transparent, all participants expressed confusion about why sometimes they can not select the intended object. However, if the opaque value is over 0.6, all participants complained that it took them more time to distinguish objects with increased visual disturbance, although the focal region effectively points out a coarse selection range. Two participants also suggested, “*Although being told the appearance of intended object, it is still a little bit hard to confirm my choice under the remote and occluded environments. Adding an outline to indicate the selection may be cool.*” We thus augmented the most potentially selected objects with a visual outline, as presented in many previous works [38, 39, 54]. An opacity value of 0.2 was deemed the most appropriate, as it potentially offers the optimal combination with the depth culling design. The adaptive focal region design has been effectively employed in smartphone-based AR [38]. Our initial implementation adhered to this design, allowing the focal sphere’s size to adapt based on the selection history and the headset’s forward orientation. However, all participants remarked that the continuously changing focal region tended to be distracting and cost them additional time to confirm the selection. We summarized that distraction may stem from the immersive nature of VR, which leads participants to focus more intently on changes in virtual content. Therefore, we fixed the size of the focal region in screen space to $0.4m \times 0.4m$, occupying 16% of the whole rendered screen (Fig. 3b).

Occlusion plane. The design of a controllable Occlusion plane has been validated for its effectiveness in numerous previous studies [27, 38, 39, 44]. However, When targets are sparsely distributed in the scene, significantly more physical effort and time are required to adjust the occlusion plane back and forth to reveal occluded or hidden targets. As one of our invited participants noted, “*It took me too long to wait for the plane to move to the desired location.*” We further enhanced the occlusion plane to be adaptive to the scene context besides user control. Specifically, as illustrated in Fig. 4, the initial position of the occlusion plane can automatically adjust to $0.05m$ in front of the nearest selectable object. Furthermore, the movement speed of the occlusion plane is dynamically adapted based on the density of surrounding objects. Combining the speed-dependent design adapted from Igarashi et al. [26], the speed is dynamically adapted by $wJ_{input}(D_{farthest} - D_{nearest}) / (\log(n+1) + 1)$. w is an adjustable gain parameter fixed to 0.8 in our experiments. n denotes the number of all objects within $0.1m$ in front of and behind the plane. The J_{input} is the coefficient for controlling the moving direction, ranging from -1 to 1 . When $|J_{input}| < 0.3$, the overall speed becomes zero to stabilize the movements.

5 EXPERIMENT 1: INPUT MODALITY EVALUATION

In this section, we aimed to investigate the effectiveness of our disambiguation designs in facilitating target selection in general scenarios, as well as its versatility across various interaction inputs [37], including one hand-held controller, bare-hands, and hybrid inputs, as shown in Fig. 5. The preliminary and following two experiments received approval from the Human and Artefacts Research Ethics Committee (HAREC) at HKUST(GZ). The study had minimal risk, and each participant could choose to halt the study at any time. Oral consent was obtained prior to each experiment, and no identifying information was attached to the collected data.

5.1 Experiment Design and Procedure

Fig. 6 shows the study overview. We followed the layout design of ISO circles in 3D from Wu et al. [67] and adopted a 6 (TECHNIQUE) \times 3 (DEPTH RANGE) \times 3 (REPETITION) with-subject study design based on the guidelines from Bergström et al. [5] and experiments from BubbleRay [39]. The variation of target depth under this spatialized layout affects participants’ distance perception and access capability. Three different depth ranges were employed in this experiment: *near* ($0.8m$, $1.4m$, $2.0m$), *middle* ($3.0m$, $4.2m$, $5.4m$), *far* ($8.4m$, $10.8m$,

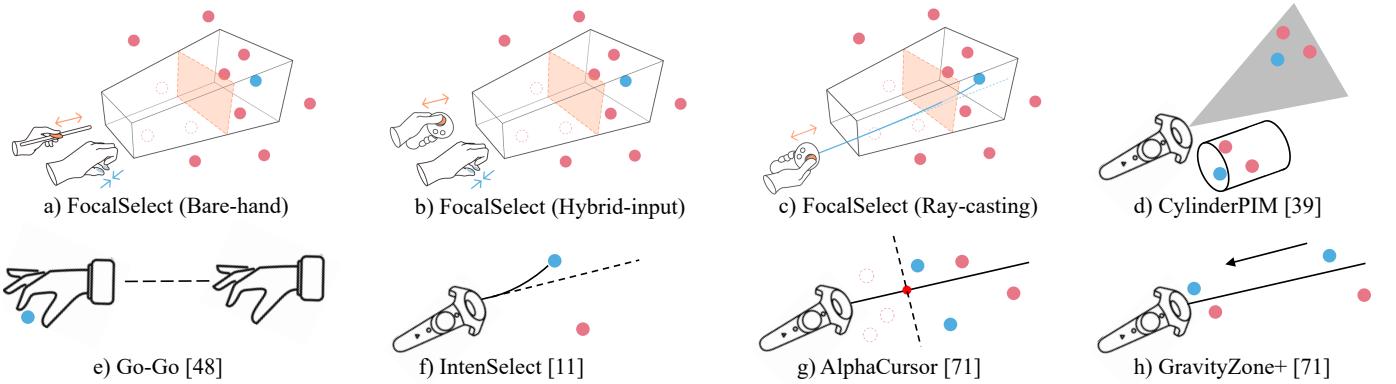


Fig. 5: Selection techniques evaluated in our two experiments.

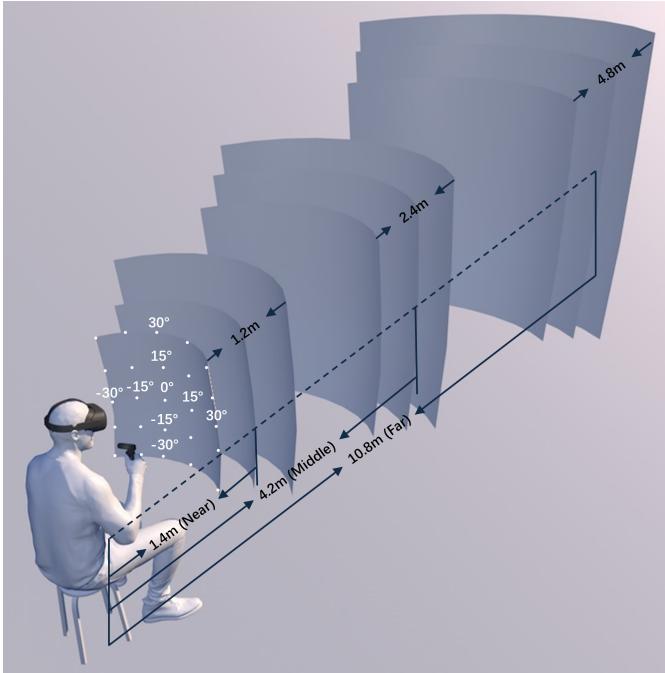


Fig. 6: Designs of spatial layout in Experiment 1. Participants can select objects generated at the intersections of the lines marked on the arc surfaces. The depth ranges of arc surfaces are varied.

13.2m). In each scene that employs a technique, 75 spherical objects of red and blue colors were presented in the ISO 3D layout with the same radius of 0.1m. The vertical and horizontal offsets of each ISO circle from the participant’s perspective range from -30° to 30° with increments of 15° . In Figs. 7 and 9, we denote statistical significance with * for $p < 0.05$, ** for $p < 0.01$, and *** for $p < 0.001$.

Each participant was required to select 10 targets with blue color for each technique. To mitigate potential learning and order effects, we randomized the order in which participants experienced the selection techniques. Within each technique, the order of depth levels was also randomized. Before the formal experiment, participants received instructions and were afforded ample time to practice each interaction technique. They can verbally report their readiness or any confusion they encountered with the techniques. In the formal tests, participants had enough rest time to relax after finishing trials per technique, and they were asked to fill in a NASA-TLX questionnaire to report their experiences subjectively. To address variations in participant height that may affect observed occlusion [39], the initial height of the VR headset is fixed at 1.2m. Participants can complete the experiment trials either

sitting or standing as long as they maintain the same pose throughout each trial. The entire experiment lasts approximately 70 minutes for each participant, including a semi-structured interview to discuss their overall selection experiences. The recorded data on task performance include: (1) *Selection Time* measures the time from the start of each trial till participants confirmed the selection; (2) *Selection Accuracy* records a binary value indicating whether an object has been successfully selected upon confirmation; (3) *Distance of Head Movement*: computes the L2-norm of the displacement vector per frame and sums to represent the total distance traveled by the headset from the starting to the ending location in each trial. Since the focal region aligns with the viewing frustum, the interaction process was inherently influenced by head movement. It is crucial to assess whether this potentially added workload is acceptable to users objectively and subjectively.

5.2 Evaluated Techniques

Besides *FocalSelect* with three input modalities, we chose three other 3D selection techniques for the evaluation (Fig. 5). For the bare-hand input modality, we chose *Go-Go*, while for ray-casting input, we selected *IntenSelect* and *AlphaCursor* as baselines. *IntenSelect* uses score-based heuristics for selection, while *AlphaCursor* employs depth-based culling. *FocalSelect* combines these two strategies through input remapping, making it essential to evaluate whether this integration performs effectively compared to the baseline techniques across different input modalities. All parameters of techniques were well-tuned for best performance.

Go-Go [48]: enables participants to control an extensible virtual hand to grasp the target along the input’s forwarding direction. A target can be selected if it falls within the bounding box of the virtual hand. The *Go-Go* technique represents a category of designs that control a 3D cursor beyond the arm stretch. The position of the virtual hand relative to the user’s body is determined by the input’s position through a nonlinear mapping, as described below:

$$D_{\text{extended}} = \begin{cases} D_{\text{input}} & \text{if } D_{\text{input}} < D_0 \\ D_{\text{input}} + g(D_{\text{input}} - D_0) & \text{otherwise,} \end{cases} \quad (4)$$

where D_{extended} and D_{input} indicate the distance of the extended virtual hand and the controller from the user’s body, D_0 is the threshold, and g is the coefficient associated with the distance conditions, which are 12, 25, and 120 for near, middle, and far case respectively.

IntenSelect [11]: utilizes spatial-temporal information to compute a score for each object in the scene, selecting the one with the highest rank. The scoring function is defined as follows:

$$s_t = s_{t-1}\gamma + (1 - \frac{\eta(t)}{\varepsilon})(1 - \gamma), \quad (5)$$

where s_t denotes the score at time t , $\eta(t)$ is the angle between the ray and the center of the target at time t , ε is the threshold, and γ is

fixed to 0.5 as the time decay coefficient. *IntenSelect* represents a class of heuristic approaches to determine the optimal target. We didn't introduce *IntenSelect+* [30], a recent improvement of *IntenSelect*, as this technique primarily focuses on supporting selections for parts of an object, which falls outside the scope of our work.

AlphaCursor [71]: is a ray-casting technique that allows participants to use a movable cursor for depth-based culling of rendered objects. *Alphacursor* represents a class of selection approaches that provide direct disambiguation controls over objects. The speed of the movable cursor is set to 1.1m/s for adequate occlusion revealing.

FocalSelect (Bare-hand): utilizes the right hand's pinch gesture to confirm selection, while the left-hand controls the movement of the occlusion plane by pushing a virtual slider.

FocalSelect (Hybrid-input): employs the right hand's pinch gesture to confirm the selection and the left hand to adjust the occlusion plane by pushing the joystick forward and backward.

FocalSelect (Ray-casting): uses a single controller to perform heuristic-based selections.

5.3 Participants and Apparatus

We recruited 24 participants (13 females and 11 males, average age 24.6 (SD=1.63)) for this experiment, all self-identifying as right-handed. All individuals exhibited unimpaired vision and color discrimination abilities. 11 of them had professional experience within VR as designers or developers. The experiments were implemented within Unity3D 2022 LTS, ran on Quest 3, and driven by a Windows 10 desktop (CPU: AMD Ryzen 9 7900X, GPU: GeForce RTX 3090). Meta XR All-in-One SDK was used to track the hands and controllers simultaneously.

5.4 Results

We initially removed 189 trials (1.47% of 12960) where the completion time exceeded three standard deviations from the mean for each depth range condition. These outliers were all due to some participants requesting short breaks during one selection trial. The Shapiro-Wilk test indicated that collected data was not normally distributed, leading us to apply the Aligned Rank Transform [13] before conducting an RM-ANOVA. For pairwise comparisons, we performed paired t-tests with Bonferroni correction.

5.4.1 Selection Time

The selection time of every technique under each depth variation was shown in the first row of Fig. 7. An RM-ANOVA test revealed a significant effect of TECHNIQUE in near ($F_{5,115} = 4.76, p < 0.05, \eta^2 = 0.17$), middle ($F_{5,115} = 9.45, p < 0.01, \eta^2 = 0.29$), and far ($F_{5,115} = 14.12, p < 0.01, \eta^2 = 0.43$) scenes. Regarding input modalities of our approach, selection time differed significantly between Ray-casting and Bare-hand ($t_{23} = -3.63, p < 0.002, d = -1.05$). In the **far DEPTH RANGE**, we observed significant differences in selection times over TECHNIQUE: *FocalSelect (Bare-hand)* and *Go-Go* ($t_{23} = -3.59, p < 0.001, d = -1.02$), *FocalSelect (Hybrid-input)* and *Go-Go* ($t_{23} = -3.19, p < 0.01, d = -0.92$), as well as *FocalSelect (Ray-casting)* and *Alphacursor* ($t_{23} = -3.86, p < 0.001, d = -1.11$).

However, the difference between *FocalSelect (Ray-casting)* and *IntenSelect* ($t_{23} = -0.012, p = 0.98$) was not significant. Additionally, a significant difference was found only between near and far DEPTH RANGE ($t_{23} = -4.12, p < 0.001, d = -1.19$). These results supported our expectation that *FocalSelect* performs the best in far scenes, resulting in minimal extra selection time, although it may require more selection steps than baseline techniques.

5.4.2 Selection Accuracy

The selection accuracy of every technique under each depth variation was reported in the second row of Fig. 7. An RM-ANOVA test revealed a significant effect of TECHNIQUE on selection accuracy in near ($F_{5,115} = 3.99, p < 0.05, \eta^2 = 0.14$), middle ($F_{5,115} = 8.13, p < 0.01, \eta^2 = 0.26$), and far ($F_{5,115} = 6.32, p < 0.05, \eta^2 = 0.21$) scenes. Regarding input modalities of our approach, there were no

significant differences in accuracy between Ray-casting and Hybrid-input ($t_{23} = 0.57, p = 0.56$), Ray-casting and Bare-hand ($t_{23} = 0.44, p = 0.65$), and Bare-hand and Hybrid-input ($t_{23} = 0.94, p = 0.35$). For *FocalSelect*, no significant difference was found between near and middle ($t_{23} = 0.027, p = 0.97$), near and far ($t_{23} = 0.89, p = 0.37$), as well as middle and far ($t_{23} = 0.86, p = 0.38$) DEPTH RANGE.

Furthermore, we observed significant differences in selection accuracy over TECHNIQUE: *FocalSelect (Bare-hand)* and *Go-Go* ($t_{23} = 3.88, p < 0.001, d = 1.12$), *FocalSelect (Hybrid-input)* and *Go-Go* ($t_{23} = 3.36, p < 0.01, d = 0.97$), *FocalSelect (Ray-casting)* and *IntenSelect* ($t_{23} = 4.43, p < 0.001, d = 1.26$), as well as *FocalSelect (Ray-casting)* and *AlphaCursor* ($t_{23} = 4.73, p < 0.001, d = 1.37$). In short, *FocalSelect* exhibited the best performance in accuracy over all depth ranges. A significant drop in selection accuracy was observed in *IntenSelect*, *AlphaCursor*, and *Go-Go*, indicating that more retries occurred as depth increased. This decline in accuracy may be attributed to the aggregated Heisenberg Effect, where the increased depth introduces greater uncertainty in selection without proper disambiguation.

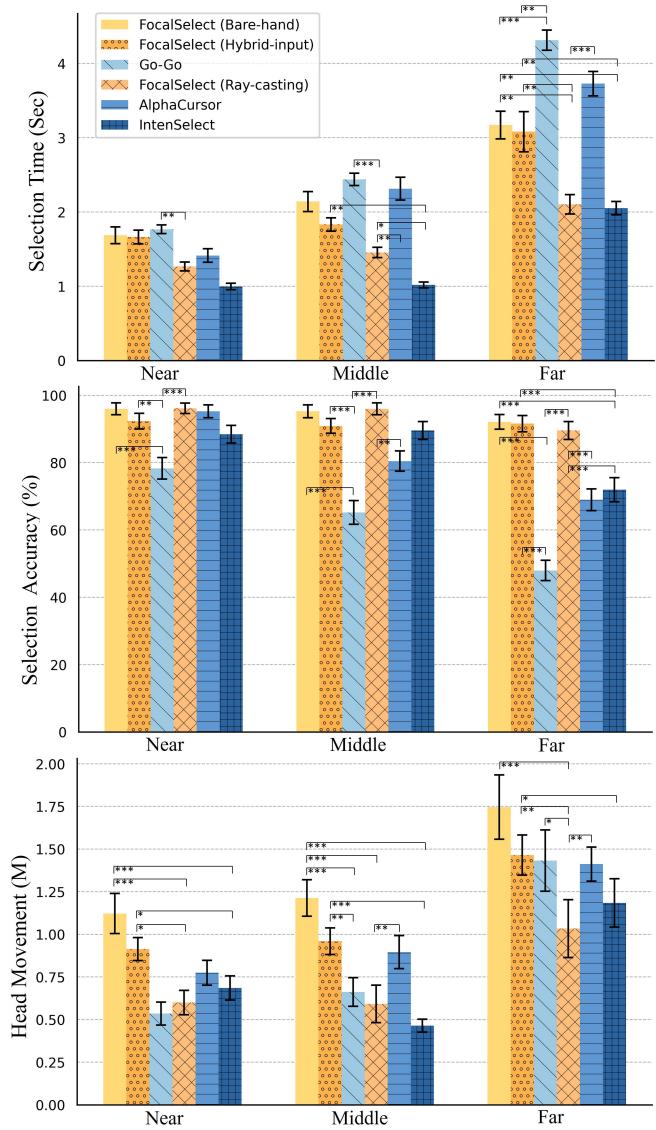


Fig. 7: The performance summary of each technology under each depth range, including selection time, selection accuracy, and head movement. The error bar represents the standard error.

5.4.3 Distance of Head Movement

The head movement of every technique across different depth ranges was presented in the third row of Fig. 7. An RM-ANOVA test revealed a significant effect of TECHNIQUE in near ($F_{5,115} = 10.77, p < 0.01, \eta^2 = 0.31$), middle ($F_{5,115} = 11.27, p < 0.01, \eta^2 = 0.32$), and far ($F_{5,115} = 8.04, p < 0.01, \eta^2 = 0.25$) scenes. We observed significant differences in movement over TECHNIQUE: *FocalSelect (Bare-hand)* and *Go-Go* ($t_{23} = 5.72, p < 0.001, d = 1.65$), *FocalSelect (Bare-hand)* and *FocalSelect (Hybrid-input)* ($t_{23} = 2.56, p < 0.02, d = 0.74$), *FocalSelect (Hybrid-input)* and *Go-Go* ($t_{23} = 3.42, p < 0.002, d = 0.98$), as well as *FocalSelect (Ray-casting)* and *Alphacursor* ($t_{23} = -2.86, p < 0.01, d = -0.81$). In contrast, the difference between *FocalSelect (Ray-casting)* and *IntenSelect* ($t_{23} = -0.042, p = 0.96$), was not significant. Additionally, significant differences in movement were observed between near and far ($t_{23} = -5.49, p < 0.001, d = -1.59$), also middle and far DEPTH RANGE ($t_{23} = -3.64, p < 0.001, d = -1.05$). However, the near and middle ranges did not show a significant difference ($t_{23} = 0.32, p = 0.74$).

In summary, the results aligned with our expectation that the design of the focal regions may increase participants' physical workload to some extent, reflecting the accuracy-workload tradeoff. For bare-hand and hybrid inputs, tracking limitations of all-in-one headsets require hands to stay within the field of view for effective tracking. This increases reliance on head movements to maintain head-hand coordination during selections. For far scenes, *FocalSelect (Ray-casting)* still demonstrated the best performance, indicating that guiding movements from controller pointing may help mitigate head movement workload. Compared to *AlphaCursor*, our heuristic design effectively reduces the physical effort of controlling the occlusion plane.

5.4.4 Subjective Feedback

We conducted a statistical analysis of the collected NASA-TLX feedback, where lower scores indicate that participants perceive the technique as more effective. A Friedman test revealed significant effects of Technique on Performance ($\chi(5) = 74.972, p < 0.001$), Physical Demand ($\chi(5) = 86.295, p < 0.001$), Mental Demand ($\chi(5) = 79.184, p < 0.001$), and Effort ($\chi(5) = 84.011, p < 0.001$). Performance measures how accurate and fast participants feel during selections. Participants rated *FocalSelect (Bare-hand)* and *FocalSelect (Hybrid-input)* as having higher performance compared to *GO-GO*. Physical Demand assesses the physical effort required for successful selections. All three of our techniques were rated less fatiguing than the baseline techniques. Participants reported that they naturally searched for targets by moving their heads and did not feel too bothered by the design of the focal region. Notably, they all expressed a strong preference for *FocalSelect (Hybrid-input)*. *FocalSelect (Bare-hand)* was perceived to have a relatively steep learning curve with a non-dominant hand. When selecting far targets, participants commended *FocalSelect (Ray-casting)*, *FocalSelect (Hybrid-input)* and *IntenSelect*.

5.5 Evaluation and Discussion

The results indicated that *FocalSelect* can improve selection accuracy, particularly in far scenes. No significant differences in selection accuracy were found across input modalities in *FocalSelect*. For selection time, no significant differences were observed between Bare-hand and Hybrid-input. Participants' feedback revealed that most participants appreciated *FocalSelect (Hybrid-input)*. Although *FocalSelect (Bare-hand)* required more head movements statistically, 3 participants suggested that this inconvenience might stem from our experimental setup. They noted that if selections were made in environments where objects are more sparsely distributed, *FocalSelect (Bare-hand)* could be significantly more effective. Two participants mentioned that both virtual hands must be tracked with high precision to facilitate smooth selections. This limitation primarily arises from the nature of all-in-one headsets. When selecting far objects, participants found it significantly more challenging to achieve accuracy with *AlphaCursor*. 5 participants even resorted to using their left hand to help stabilize their selections, which may be attributed to the Heisenberg effect [66]. For the input

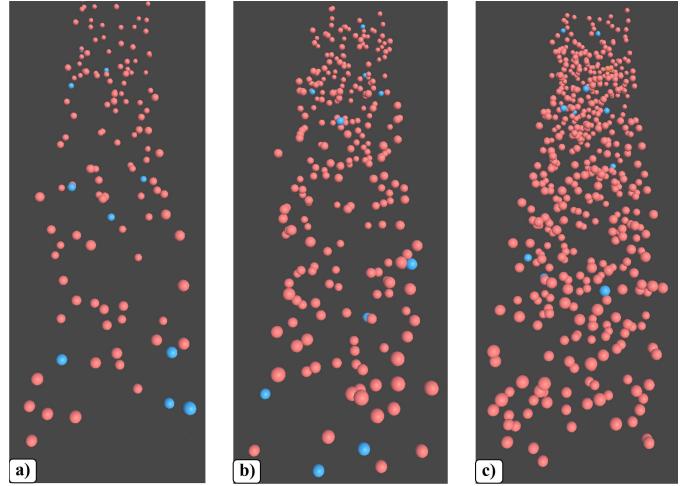


Fig. 8: Top views of experiment environment with three density levels: (a) Low-128 spheres, (b) Medium-256 spheres, (c) High-512 spheres. Positions of target objects (blue) are randomized.

modality, using hybrid input and ray-casting significantly minimizes head movement. We conducted Experiment 2 to further explore *FocalSelect*'s performance with these two inputs.

6 EXPERIMENT 2: PERFORMANCE EVALUATION

The objective of this study was to further evaluate the performance of our heuristic designs with disambiguation in occluded scenarios.

6.1 Experiment Design and Procedure

We followed the spatial setup of DENSITY variations from Maslych et al. [39] and incorporated the scene layout findings from Lu et al. [35]. Fig. 8 illustrates the layout of the testing scene and the variations in density levels (the number of selectable objects in a fixed-size room). All objects are randomly generated within the depth range between 1m to 15m. Our experiment employed a 4 (TECHNIQUE) \times 3 (DENSITY) \times 3 (REPETITION) within-subject design. Similar to Experiment 1, each participant was required to select 10 blue targets for each technique, with the five techniques presented in random order. The order of density levels within each technique was also randomized to minimize potential order effects.

Participants were given warm-up trials to practice each technique, and a calibration process similar to that in Experiment 1 was conducted. At the end of each session, participants received adequate breaks and completed a NASA-TLX questionnaire. Semi-structured interviews were held after all trials to gather deeper insights into their experiences. The entire experiment lasted around 55 minutes for each participant, and we recorded the data for performance analysis in the same manner as in Experiment 1.

6.2 Evaluated Techniques

Building on the findings from the first experiment, we choose to compare *FocalSelect (Hybrid-input)* and *FocalSelect (Ray-casting)* against three baseline selection techniques (*AlphaCursor*, *GravityZone+* and *CylinderPIM*), as shown in Fig. 5. These baseline techniques were picked for their ability to preserve spatial relationships among objects during selection and their explicit design for occluded scenarios. *IntenSelect* and techniques with bare-hand inputs were excluded due to their poor performance in this occluded setup. All parameters of techniques were carefully optimized for best performance.

GravityZone+ [71]: functions by translating all objects in the scene closer or farther from the user. We employed the modified implementation from Maslych et al. [39] to build the experiment of this technique.

CylinderPIM [39]: addresses occlusion by maintaining the relative depth of the minimized objects and arranging them within a 3D cylinder that is 0.48m in diameter and height.

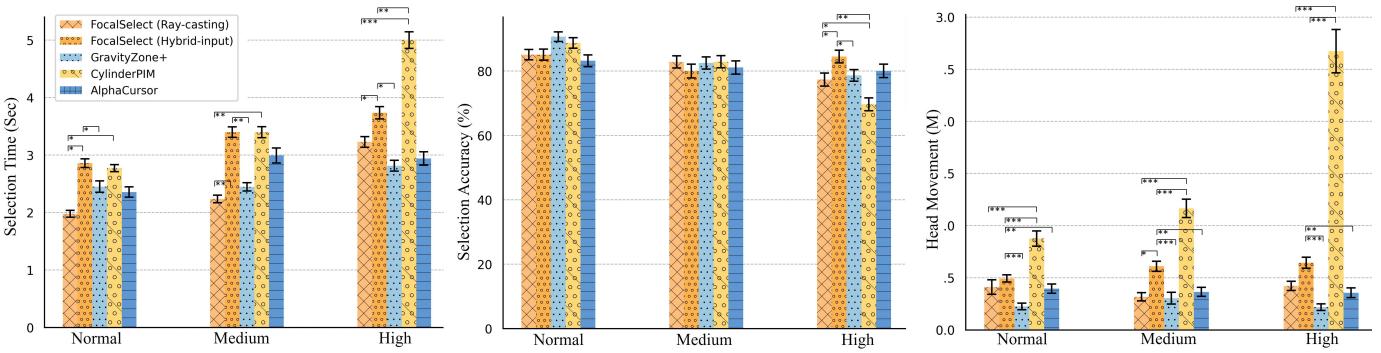


Fig. 9: The performance summary of each technology under each density level, including selection time, selection accuracy, and head movement. The error bar represents the standard error.

6.3 Participants and Apparatus

24 participants (11 females and 13 males, average age 25.1 (SD=3.3)) were recruited, all self-identified as right-handed and exhibiting unimpaired vision and color discrimination abilities. 8 of them had prior experience with VR, and 5 of them developed VR-related applications. We used the same apparatus as in Experiment 1.

6.4 Results

We initially removed 98 trials (0.91% of 10800) where the completion time exceeded three standard deviations from the mean for each density level condition. These outliers were primarily due to some participants requesting short breaks during the experiment (72 trials) and forgetting the steps for selection (26 trials). The Shapiro-Wilk test indicated that collected data was not normally distributed, leading us to apply the Aligned Rank Transform [13] before conducting a two-way RM-ANOVA. We performed paired t-tests for pairwise comparison.

6.4.1 Selection Time

The selection time of every technique under each depth variation was shown in the first column of Fig. 9. An RM-ANOVA test revealed a significant effect of TECHNIQUE in scenes of normal ($F_{4,92} = 7.59, p < 0.05, \eta^2 = 0.24$), medium ($F_{4,92} = 8.58, p < 0.05, \eta^2 = 0.27$), and high ($F_{4,92} = 11.52, p < 0.01, \eta^2 = 0.33$) density levels. The interaction effect of TECHNIQUE \times DENSITY is not significant ($F_{8,184} = 1.80, p = 0.10$). While significant differences were observed among TECHNIQUE: *FocalSelect (Ray-casting)* and *CylinderPIM* ($t_{23} = -2.61, p < 0.05, d = -0.75$), also *FocalSelect (Hybrid-input)* and *GravityZone+* ($t_{23} = 2.29, p < 0.05, d = 0.66$), the difference between *FocalSelect (Ray-casting)* and *GravityZone+* ($t_{23} = -0.70, p = 0.48$), *FocalSelect (Ray-casting)* and *AlphaCursor* ($t_{23} = -0.52, p = 0.60$), also *FocalSelect (Hybrid-input)* and *AlphaCursor* ($t_{23} = 0.99, p = 0.32$) were not significant. For the input modalities of our approach, a significant difference was found between *FocalSelect (Ray-casting)* and *FocalSelect (Hybrid-input)* ($t_{23} = -2.08, p < 0.05, d = -0.60$) with a moderate effect size.

Moreover, significant differences in movement were observed between normal and high ($t_{23} = 6.09, p < 0.001, d = 1.75$), medium and high ($t_{23} = 6.11, p < 0.001, d = 1.76$) DENSITY LEVEL. However, the normal and medium levels did not show a significant difference ($t_{23} = -0.94, p = 0.37$). While *GravityZone+* nearly achieved optimal performance across three density levels, it may hinder scene recovery for further editing. In these cases, pre-recording scene information is essential but can be resource-intensive (e.g., memory usage).

6.4.2 Selection Accuracy

The selection accuracy of every technique under each density variation was reported in the second column of Fig. 9. An RM-ANOVA test revealed a significant effect of TECHNIQUE in scenes of normal ($F_{4,92} = 5.68, p < 0.05, \eta^2 = 0.19$), medium ($F_{4,92} = 4.23, p < 0.05, \eta^2 = 0.15$), and high ($F_{4,92} = 7.17, p < 0.05, \eta^2 = 0.24$) density levels. The interaction effect of TECHNIQUE \times DENSITY is also

significant ($F_{8,184} = 2.29, p < 0.05$). Significant differences in accuracy were observed among TECHNIQUE at the **high DENSITY LEVEL**: *FocalSelect (Ray-casting)* and *FocalSelect (Hybrid-input)* ($t_{23} = -2.53, p < 0.05, d = -0.91$), *FocalSelect (Ray-casting)* and *CylinderPIM* ($t_{23} = 2.69, p < 0.05, d = 0.77$), as well as *FocalSelect (Hybrid-input)* and *CylinderPIM* ($t_{23} = 2.81, p < 0.01, d = 0.81$). In contrast, the difference between *FocalSelect (Hybrid-input)* and *GravityZone+* ($t_{23} = 1.48, p = 0.80$), *FocalSelect (Hybrid-input)* and *AlphaCursor* ($t_{23} = 1.08, p = 0.28$), *FocalSelect (Ray-casting)* and *GravityZone+* ($t_{23} = -1.12, p = 0.0.27$), *FocalSelect (Ray-casting)* and *AlphaCursor* ($t_{23} = -0.22, p = 0.82$) were not significant.

Moreover, significant differences were observed between normal and high ($t_{23} = 2.92, p < 0.01, d = 0.81$), also medium and high DENSITY LEVEL ($t_{23} = 2.11, p < 0.05, d = 0.61$). However, the normal and medium levels did not show a significant difference ($t_{23} = 0.39, p = 0.69$). In short, *FocalSelect* exhibited the best performance, and hybrid inputs may be more suited than ray-casting in highly occluded scenes.

6.4.3 Distance of Head Movement

The head movement of every technique under each density variation was presented in the third column of Fig. 9. An RM-ANOVA test revealed a significant effect of TECHNIQUE in scenes of normal ($F_{4,92} = 12.58, p < 0.001, \eta^2 = 0.35$), medium ($F_{4,92} = 13.15, p < 0.001, \eta^2 = 0.36$), and high ($F_{4,92} = 17.63, p < 0.001, \eta^2 = 0.43$) density levels. The interaction effect of TECHNIQUE \times DENSITY is not significant ($F_{8,184} = 1.04, p = 0.40$). However, significant differences in head movement were observed among TECHNIQUE: *FocalSelect (Ray-casting)* and *CylinderPIM* ($t_{23} = -6.81, p < 0.001, d = -1.96$), *FocalSelect (Hybrid-input)* and *GravityZone+* ($t_{23} = 4.81, p < 0.001, d = 1.39$), *FocalSelect (Hybrid-input)* and *AlphaCursor* ($t_{23} = 2.89, p < 0.01, d = 0.83$), as well as *FocalSelect (Hybrid-input)* and *CylinderPIM* ($t_{23} = -6.25, p < 0.001, d = -1.80$). In contrast, the difference between *FocalSelect (Ray-casting)* and *GravityZone+* ($t_{23} = 1.73, p = 0.089$), also *FocalSelect (Ray-casting)* and *AlphaCursor* ($t_{23} = 0.14, p = 0.88$) was not significant.

For the input modalities of our approach, no significant difference was found between *FocalSelect (Ray-casting)* and *FocalSelect (Hybrid-input)* ($t_{23} = -1.84, p = 0.07$). Moreover, significant differences in movement were observed between normal and high ($t_{23} = -3.42, p < 0.01, d = -0.99$), also medium and high ($t_{23} = -2.72, p < 0.02, d = -0.78$) DENSITY LEVEL. However, the normal and medium levels did not show a significant difference ($t_{23} = -0.83, p = 0.41$). In summary, our approach with ray-casting input resulted in minimal extra head movements, albeit a bit more than *GravityZone+* and *AlphaCursor*. *FocalSelect* outperformed *CylinderPIM*, which also employed a two-stage selection strategy.

6.5 Evaluation and Discussion

The results indicate that *FocalSelect* performs on par with *GravityZone+* and outperforms *CylinderPIM*. In our experiment, which did not include search and repeat tasks [45, 71], the performance of *CylinderPIM*

may be adversely affected in high-density environments. This technique requires users to thoroughly search the entire scene for targets, as occluded objects may remain hidden within the cylinder mini-map. Conversely, experiments by Maslych et al. [39] indicate that *Cylinder-PIM* outperforms *GravityZone+* when the target region is prompted. In our setup, while *GravityZone+* may perform better in selection time, 18 participants raised concerns about its practicality. They noted that this technique could disrupt the original scene context, even though relative spatial relationships are preserved. Three participants suggested that *GravityZone+* might be most suitable for applications like virtual treasure hunting, where the primary objective is to locate a valuable item as quickly as possible. Furthermore, the NASA-TLX analysis revealed no significant differences among the four techniques across all measures except Physical Demand ($\chi(4) = 51.94, p < 0.001$). Regarding input modalities of *FocalSelect*, we found that although the hybrid input increases selection time compared to ray-casting, the difference in head movements is not significant. Another noteworthy finding is that *FocalSelect* (*Hybrid-input*) outperforms *FocalSelect* (*Ray-casting*) in terms of accuracy in the high DENSITY LEVEL. This result supports the effectiveness of our scoring-based function, which relies on distance and angle offsets, demonstrating its efficacy in occluded selections.

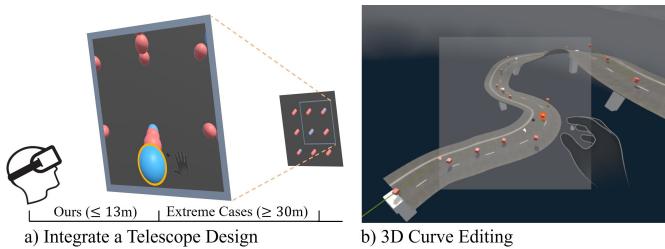


Fig. 10: Practical Demonstrations. a) *FocalSelect* can be extended to support extreme distant cases. b) *FocalSelect* can be integrated with the SplineMesh plugin [43] to enhance 3D editing in VR.

7 DISCUSSIONS

The results from two experiments indicate that *FocalSelect* performs comparably to baseline techniques across all three input modalities. Notably, it achieved the **highest selection accuracy** in far and highly-occluded scenarios. Despite requiring additional steps and a slight increase in physical workload (e.g., focal region adjustments), *FocalSelect* still resulted in competitive selection time. This claim is supported by subjective feedback, as approximately half of the participants preferred investing more time and efforts to achieve accurate selections rather than increasing the number of retries. This leads us to summarize that *FocalSelect* performs well in occluded environments and tasks without strict time constraints. Its effectiveness is especially evident in scene editing scenarios, where preserving spatial context is crucial. By maintaining object relationships within their surroundings, *FocalSelect* excels at managing spatial dynamics. These findings align with those reported in *FocalPoint* [38].

Our work has several limitations. Users may experience fatigue from repeatedly raising their hands in coordination with their head movements. However, this issue primarily stems from the tracking limitations of all-in-one headsets for bare-hand inputs [42, 51]. Future research could explore designing an offset mapping function to minimize this tiring action. Additionally, the current focal region is fixed and aligned with the head movement to mitigate distractions. Eye-tracking [54] may be useful to determine whether the focal region should be disabled, restored or distorted. For instance, the focal region could disappear if the user's gaze remains outside it in screen space for more than 5 seconds. We employed fixed depth cue displays, leveraging established design patterns to effectively reveal occluded objects through semi-transparent rendering [14, 60, 64]. This approach implicitly conveys the dynamic position of the occlusion plane. Future work could comprehensively evaluate how varying depth cue displays affect selection performance across diverse interaction scenarios.

All participants strongly supported the hybrid input in *FocalSelect*. 6 participants with professional VR development experiences commented that this design effectively simulates the collaborative usage of a mouse and keyboard in desktop environments, which is particularly beneficial for 3D modeling software such as MAYA and Blender. Future research could explore this hybrid-input to enhance authoring tasks [50] in VR.

7.1 Application and Potential Improvement

In this section, we demonstrate the practical functionalities of *FocalSelect* with two implementations in VR.

7.1.1 Bézier Curve Manipulation

The Bézier Curve, a parametric curve, is extensively utilized in both academia and industry for computer-aided design, facilitating the creation of 2D sketches and 3D models. Our first demonstration application, illustrated in Fig. 10b, highlights how *FocalSelect* enhances the selection and manipulation of Bézier Curve control points in VR. With *FocalSelect*, users can intuitively edit the shape of a road by selecting and adjusting the control points of the Bézier Curve, enabling precise and efficient modifications.

7.1.2 Telescope Design for Enhanced Target Selection

Inspired by *PinchLens* [74] and *vMirror* [33], we integrated a telescope design into *FocalSelect* for selecting targets at extreme distances, as shown in Fig. 10a. This design helps reduce physical loads and motion sickness by minimizing the amount of movement and the number of teleports required.

7.2 Future Work

FocalSelect employs a remapping mechanism to extend user inputs beyond the reach of their arms. However, as discussed in Section 2.2, this remapping may exacerbate the Heisenberg effects due to the increased offsets between the target locations and the remapped inputs. Future research could explore how this effect varies with distance and input modality. For the evaluations, our study did not examine the performance of *FocalSelect* when using the non-dominant hand. In addition, three participants remarked that having control over when to use the focal region could help reduce head movements. For example, when only three objects are rendered in a scene and are sparsely distributed, disabling the focal region allows for score computations for all three objects, enabling the selection of the optimal target without requiring users to face it directly. Future studies could explore this adaptive control of the focal region and integrate scene context to facilitate this coarse pre-selection feature, ultimately reducing physical workloads on users. Specifically, the focal region could be activated only when densely packed objects are rendered in the scene. Furthermore, Ma et al. [38] have demonstrated that heuristic methods can be effectively applied in AR, where preserving scene context is crucial. It would be valuable to investigate how *FocalSelect* can assist in editing Gaussian Splatting [27], a point-cloud-based 3D representation.

8 CONCLUSION

In this paper, we present *FocalSelect*, a heuristic-based selection technique, and iteratively refine its visual representations. We then conduct two experiments to demonstrate that *FocalSelect* enhances selection experience in highly occluded and remote scenarios, without being limited to a specific input modality. Based on the findings from experiments, we discuss variants, applications, and potential improvements to boost *FocalSelect*'s practicality in real-world use.

ACKNOWLEDGMENTS

This work is partially supported by the Guangzhou-HKUST(GZ) Joint Funding Project (No.2024A03J0617), Guangzhou Science and Technology Program City-University Joint Funding Project (No.2023A03J0001), the Project of DEGP (No.2023KCXTD042), and Guangdong Provincial Key Lab of Integrated Communication, Sensing and Computation for Ubiquitous Internet of Things (No.2023B1212010007).

REFERENCES

- [1] F. Argelaguet and C. Andujar. A Survey of 3D Object Selection Techniques for Virtual Environments. *Computers & Graphics*, 37(3):121–136, 2013. 3
- [2] F. Argelaguet, C. Andujar, and R. Trueba. Overcoming Eye-hand Visibility Mismatch in 3D Pointing Selection. In *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*, pp. 43–46, 2008. 2
- [3] S. V. Babu, W. Hsieh, and J. Chuang. The Benefits of Near-field Manipulation and Viewing to Distant Object Manipulation in VR. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 408–417. IEEE Computer Society, Los Alamitos, CA, USA, mar 2024. 2
- [4] M. Baloup, T. Pietrzak, and G. Casiez. RayCursor: A 3D Pointing Facilitation Technique based on Raycasting. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI ’19, 12 pages, p. 1–12. Association for Computing Machinery, New York, NY, USA, 2019. 1, 2
- [5] J. Bergström, T.-S. Dalsgaard, J. Alexander, and K. Hornbæk. How to Evaluate Object Selection and Manipulation in VR? Guidelines from 20 Years of Studies. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–20. ACM, Yokohama Japan, may 2021. 3, 4
- [6] J. Blattgerste, P. Renner, and T. Pfeiffer. Advantages of Eye-Gaze over Head-Gaze-based Selection in Virtual and Augmented Reality under Varying Field of Views. In *Proceedings of the Workshop on Communication by Gaze Interaction*, COGAIN ’18, article no. 1, 9 pages. Association for Computing Machinery, New York, NY, USA, 2018. 2
- [7] D. A. Bowman and L. F. Hodges. An Evaluation of Techniques for Grabbing and Manipulating Remote Objects in Immersive Virtual Environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics*, pp. 35–ff, 1997. 2
- [8] J. Cashion, C. Wingrave, and J. J. LaViola Jr. Dense and Dynamic 3D Selection for Game-Based Virtual Environments. *IEEE Transactions on Visualization and Computer Graphics*, 18(4):634–642, 2012. 1, 2
- [9] O. Chapuis, J.-B. Labrune, and E. Pietriga. DynaSpot: Speed-Dependent Area Cursor. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’09, 10 pages, p. 1391–1400. Association for Computing Machinery, New York, NY, USA, 2009. 3
- [10] Y. Chen, J. Sun, Q. Xu, E. Lank, P. Irani, and W. Li. Global Scene Filtering, Exploration, and Pointing in Occluded Virtual Space. In *IFIP Conference on Human-Computer Interaction*, pp. 156–176. Springer, 2021. 2
- [11] G. de Haan, M. Koutek, and F. H. Post. IntenSelect: Using Dynamic Object Rating for Assisting 3D Object Selection. In *Proceedings of the 11th Eurographics Conference on Virtual Environments*, EGVE’05, 9 pages. Eurographics Association, Goslar, DEU, 2005. 2, 3, 4, 5
- [12] T. Drey, J. Gugenheimer, J. Karlbauer, M. Milo, and E. Rukzio. VRSketchin: Exploring the Design Space of Pen and Tablet Interaction for 3D Sketching in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2020. 1
- [13] L. A. Elkin, M. Kay, J. J. Higgins, and J. O. Wobbrock. An Aligned Rank Transform Procedure for Multifactor Contrast Tests. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, UIST ’21, 15 pages, p. 754–768. Association for Computing Machinery, New York, NY, USA, 2021. 6, 8
- [14] N. Elmqvist and P. Tsigas. A Taxonomy of 3D Occlusion Management Techniques. In *2007 IEEE Virtual Reality Conference*, pp. 51–58, 2007. 2, 4, 9
- [15] N. Elmqvist and M. E. Tudoreanu. Evaluating the Effectiveness of Occlusion Reduction Techniques for 3D Virtual Environments. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST ’06, 10 pages, p. 9–18. Association for Computing Machinery, New York, NY, USA, 2006. 2
- [16] N. Elmqvist and M. E. Tudoreanu. Occlusion Management in Immersive and Desktop 3D Virtual Environments: Theory and Evaluation. *Int. J. Virtual Real.*, 6(2):21–32, 2007. 2
- [17] T. T. Elvins, D. R. Nadeau, and D. Kirsh. Worldlets—3D Thumbnails for Wayfinding in Virtual Environments. In *Proceedings of the 10th Annual ACM Symposium on User Interface Software and Technology*, UIST ’97, 10 pages, p. 21–30. Association for Computing Machinery, New York, NY, USA, 1997. 2
- [18] T. Feuchtnar and J. Müller. Extending the Body for Interaction with Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI ’17, 13 pages, p. 5145–5157. Association for Computing Machinery, New York, NY, USA, 2017. 3
- [19] A. Forsberg, K. Herndon, and R. Zelezniak. Aperture based Selection for Immersive Virtual Environments. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology*, UIST ’96, 2 pages, p. 95–96. Association for Computing Machinery, New York, NY, USA, 1996. 2
- [20] S. Fukatsu, Y. Kitamura, T. Masaki, and F. Kishino. Intuitive Control of “Bird’s Eye” Overview Images for Navigation in an Enormous Virtual Environment. In *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, VRST ’98, 10 pages, p. 67–76. Association for Computing Machinery, New York, NY, USA, 1998. 2
- [21] T. Grossman and R. Balakrishnan. 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*, UIST ’06, 10 pages, p. 3–12. Association for Computing Machinery, New York, NY, USA, 2006. 2
- [22] M. Guillon, F. Leitner, and L. Nigay. Investigating Visual Feedforward for Target Expansion Techniques. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, CHI ’15, 10 pages, p. 2777–2786. Association for Computing Machinery, New York, NY, USA, 2015. 4
- [23] A. Guterstam, V. I. Petkova, and H. H. Ehrsson. The Illusion of Owning a Third Arm. *PloS one*, 6(2):e17208, 2011. 3
- [24] N. Halabi, E. Jones, and P. Mirza-Babaei. Unleashing Immersive Experiences: The Power of Gesture-Based VR Interaction. *Interactions*, 30(6):52–55, 2023. 1
- [25] J. P. Hansen, V. Rajanna, I. S. MacKenzie, and P. Bækgaard. A Fitts’ Law Study of Click and Dwell Interaction by Gaze, Head and Mouse with a Head-Mounted Display. In *Proceedings of the Workshop on Communication by Gaze Interaction*, COGAIN ’18, article no. 7, 5 pages. Association for Computing Machinery, New York, NY, USA, 2018. 2
- [26] T. Igarashi and K. Hinckley. Speed-Dependent Automatic Zooming for Browsing Large Documents. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology*, UIST ’00, 10 pages, p. 139–148. Association for Computing Machinery, New York, NY, USA, 2000. 2, 4
- [27] B. Kerbl, G. Kopanas, T. Leimkuehler, and G. Drettakis. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.*, 42(4), article no. 139, 14 pages, jul 2023. 4, 9
- [28] F. L. Kooi and A. Toet. Visual Comfort of Binocular and 3D Displays. *Displays*, 25(2-3):99–108, 2004. 4
- [29] R. Kopper, F. Bacim, and D. A. Bowman. Rapid and Accurate 3D Selection by Progressive Refinement. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, 2011. 1, 2
- [30] M. Krüger, T. Gerrits, T. Römer, T. Kuhlen, and T. Weissker. IntenSelect+: Enhancing Score-Based Selection in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, 30(5):2829–2838, 10 pages, mar 2024. 4, 6
- [31] J. J. LaViola Jr, E. Kruijff, R. P. McMahan, D. Bowman, and I. P. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley Professional, 2017. 3
- [32] N. Li, T. Han, F. Tian, J. Huang, M. Sun, P. Irani, and J. Alexander. Get a Grip: Evaluating Grip Gestures for VR Input Using a Lightweight Pen. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2020. 1
- [33] N. Li, Z. Zhang, C. Liu, Z. Yang, Y. Fu, F. Tian, T. Han, and M. Fan. 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*, 2021. 2, 9
- [34] R. Liu, M. Gao, L. Wang, X. Wang, Y. Xiang, A. Zhang, J. Xia, Y. Chen, and S. Chen. Interactive Extended Reality Techniques in Information Visualization. *IEEE Transactions on Human-Machine Systems*, 52(6):1338–1351, 2022. 1
- [35] Y. Lu, C. Yu, and Y. Shi. 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, 2020. 2, 3, 4, 7
- [36] J. Lucas, D. Bowman, J. Chen, and C. Wingrave. *Design and Evaluation of 3D Multiple Object Selection Techniques*. PhD thesis, Virginia Tech, Jan. 2005. 3
- [37] T. Luong, Y. F. Cheng, M. Möbus, A. Fender, and C. Holz. Controllers or Bare Hands? A Controlled Evaluation of Input Techniques on Interaction

- Performance and Exertion in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, 2023, **2**, 4
- [38] J. Ma, J. Qian, T. Zhou, and J. Huang. FocalPoint: Adaptive Direct Manipulation for Selecting Small 3D Virtual Objects. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 7(1), article no. 22, 26 pages, mar 2023. **1, 2, 3, 4, 9**
- [39] M. Maslych, Y. Hmaiti, R. Ghamandi, P. Leber, R. K. Kattoju, J. Belga, and J. J. LaViola. Toward Intuitive Acquisition of Occluded VR Objects through an Interactive Disocclusion Mini-map. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 460–470. IEEE, 2023. **2, 4, 5, 7, 9**
- [40] D. Mendes, F. M. Caputo, A. Giachetti, A. Ferreira, and J. Jorge. A Survey on 3D Virtual Object Manipulation: From the Desktop to Immersive Virtual Environments. *Computer Graphics Forum*, 38(1):21–45, 2019. **3**
- [41] D. Mendes, D. Medeiros, E. Cordeiro, M. Sousa, A. Ferreira, and J. Jorge. PRECIOUS! Out-of-reach Selection Using Iterative Refinement in VR. In *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 237–238, 2017. **2**
- [42] I. Meta Platforms. Quest 3, 2023. Accessed on: July 10th, 2024. **1, 9**
- [43] Meth. Why Learn to Draw When You Can Trace?, 2017. Accessed Aug 28, 2024. **9**
- [44] R. A. Montano-Murillo, C. Nguyen, R. H. Kazi, S. Subramanian, S. Di-Verdi, and D. Martinez-Plasencia. Slicing-Volume: Hybrid 3D/2D Multi-target Selection Technique for Dense Virtual Environments. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2020. **4**
- [45] J. Petford, M. A. Nacenta, and C. Gutwin. Pointing all around you: Selection performance of mouse and ray-cast pointing in full-coverage displays. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI ’18, 14 pages, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2018. **8**
- [46] K. Pietroszek, J. R. Wallace, and E. Lank. Tiltcasting: 3D Interaction on Large Displays using a Mobile Device. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, UIST ’15, 6 pages, p. 57–62. Association for Computing Machinery, New York, NY, USA, 2015. **2**
- [47] V. Popescu, P. Rosen, and N. Adamo-Villani. The Graph Camera. In *ACM SIGGRAPH Asia 2009 Papers*, SIGGRAPH Asia ’09, article no. 158, 8 pages. Association for Computing Machinery, New York, NY, USA, 2009. **2**
- [48] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. The Go-Go Interaction Technique: Non-Linear Mapping for Direct Manipulation in VR. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology*, UIST ’96, 2 pages, p. 79–80. Association for Computing Machinery, New York, NY, USA, 1996. **2, 3, 5**
- [49] I. Poupyrev, T. Ichikawa, S. Weghorst, and M. Billinghurst. Egocentric Object Manipulation in Virtual Environments: Empirical Evaluation of Interaction Techniques. *Computer Graphics Forum*, 17(3):41–52, aug 1998. **3**
- [50] H. . programme. GravitySketch. <https://www.gravitysketch.com/>, 2021. Accessed Aug 28, 2024. **1, 9**
- [51] P. I. Pte.ltd. Pico 4, 2023. Accessed on: July 10th, 2024. **1, 9**
- [52] Y. Y. Qian and R. J. Teather. The Eyes Don’t Have It: an Empirical Comparison of Head-based and Eye-based Selection in Virtual Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction*, SUI ’17, 8 pages, p. 91–98. Association for Computing Machinery, New York, NY, USA, 2017. **2**
- [53] J. Schjørlund, K. Hornbæk, and J. Bergström. Ninja Hands: Using Many Hands to Improve Target Selection in VR. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI ’21, article no. 130, 14 pages. Association for Computing Machinery, New York, NY, USA, 2021. **3**
- [54] L. Sidenmark, C. Clarke, X. Zhang, J. Phu, and H. Gellersen. Outline Pursuits: Gaze-assisted Selection of Occluded Objects in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2020. **2, 4, 9**
- [55] L. Sidenmark and H. Gellersen. Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality. *ACM Trans. Comput.-Hum. Interact.*, 27(1), article no. 4, 40 pages, dec 2019. **2**
- [56] R. W. Soukoreff and I. S. MacKenzie. Towards a Standard for Pointing Device Evaluation, Perspectives on 27 Years of Fitts’ Law Research in HCI. *International Journal of Human-Computer Studies*, 61(6):751–789, 2004. **2**
- [57] F. Steinicke, T. Ropinski, and K. Hinrichs. Object Selection in Virtual Environments Using an Improved Virtual Pointer Metaphor. In *Computer Vision and Graphics: International Conference, ICCVG 2004*, Warsaw, Poland, September 2004, Proceedings, pp. 320–326. Springer, 2006. **3**
- [58] R. Stoakley, M. J. Conway, and R. Pausch. Virtual Reality on a WIM: Interactive Worlds in Miniature. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’95, 8 pages, p. 265–272. ACM Press/Addison-Wesley Publishing Co., USA, 1995. **2**
- [59] J. Sun and W. Stuerzlinger. Selecting and Sliding Hidden Objects in 3D Desktop Environments. In *Proceedings of the 45th Graphics Interface Conference on Proceedings of Graphics Interface 2019*, GI’19, article no. 8, 8 pages. Canadian Human-Computer Communications Society, Waterloo, CAN, 2019. **2**
- [60] J. Sun and W. Stuerzlinger. Selecting and Sliding Hidden Objects in 3D Desktop Environments. In *Graphics Interface*, vol. 8, pp. 1–8, 2019. **4, 9**
- [61] J. J. Valdes and A. J. Barton. Visualizing High Dimensional Objective Spaces for Multi-objective Optimization: A Virtual Reality Approach. In *2007 IEEE Congress on Evolutionary Computation*, pp. 4199–4206, 2007. **1**
- [62] M. Van Someren, Y. F. Barnard, and J. Sandberg. The Think Aloud Method: A Practical Approach to Modelling Cognitive. *London: AcademicPress*, 11(6), 1994. **3**
- [63] O. Špakov and P. Majaranta. Enhanced Gaze Interaction Using Simple Head Gestures. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing, UbiComp ’12*, 6 pages, p. 705–710. Association for Computing Machinery, New York, NY, USA, 2012. **2**
- [64] U. Wagner, M. Albrecht, A. A. Jacobsen, H. Wang, H. Gellersen, and K. Pfeuffer. Gaze, Wall, and Racket: Combining Gaze and Hand-Controlled Plane for 3D Selection in Virtual Reality. 8(ISS), article no. 534, 25 pages, Oct 2024. **2, 9**
- [65] L. Wang, J. Chen, Q. Ma, and V. Popescu. Disocclusion Headlight for Selection Assistance in VR. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, 2021. **2**
- [66] D. Wolf, J. Gugenheimer, M. Combosch, and E. Rukzio. Understanding the Heisenberg Effect of Spatial Interaction: A Selection Induced Error for Spatially Tracked Input Devices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI ’20, 10 pages, p. 1–10. Association for Computing Machinery, New York, NY, USA, 2020. **1, 2, 7**
- [67] Z. Wu, D. Wang, S. Zhang, Y. Huang, Z. Wang, and M. Fan. Toward Making Virtual Reality (VR) More Inclusive for Older Adults: Investigating Aging Effect on Target Selection and Manipulation Tasks in VR. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2024. **2, 4**
- [68] L. Yang, J. Huang, T. Feng, W. Hong-An, and D. Guo-Zhong. Gesture Interaction in Virtual Reality. *Virtual Reality & Intelligent Hardware*, 1(1):84–112, 2019. **1**
- [69] D. Yu, H.-N. Liang, K. Fan, H. Zhang, C. Fleming, and K. Papangelis. Design and Evaluation of Visualization Techniques of Off-screen and Occluded Rargets in Virtual Reality Environments. *IEEE Transactions on Visualization and Computer Graphics*, 26(9):2762–2774, 2019. **2**
- [70] D. Yu, H.-N. Liang, F. Lu, V. Nanjappan, K. Papangelis, W. Wang, et al. Target Selection in Head-Mounted Display Virtual Reality Environments. *J. Univers. Comput. Sci.*, 24(9):1217–1243, 2018. **2, 3**
- [71] D. Yu, Q. Zhou, J. Newn, T. Dingler, E. Veloso, and J. Goncalves. Fully-Occluded Target Selection in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3402–3413, 2020. **1, 2, 6, 7, 8**
- [72] Z. Yu, D. Zeidler, V. Victor, and M. McGinity. Dynascape: Immersive Authoring of Real-World Dynamic Scenes with Spatially Tracked RGB-D Videos. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*, VRST ’23, article no. 10, 12 pages. Association for Computing Machinery, New York, NY, USA, 2023. **1**
- [73] S. Zhai. *Human Performance in Six Degree of Freedom Input Control [Microform]*. PhD thesis, University of Toronto, Jan. 1995. **3**
- [74] F. Zhu, L. Sidenmark, M. Sousa, and T. Grossman. PinchLens: Applying Spatial Magnification and Adaptive Control-Display Gain for Precise Selection in Virtual Reality. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 1221–1230, 2023. **3, 9**