

# A Dual-Stick Controller for Enhancing Raycasting Interactions with Virtual Objects

Nianlong Li\* Tong Wu† Zhenxuan He† Luyao Shen§ Tianren Luo† Teng Han†  
BoYu Gao¶ Yu Zhang\* Liuxin Zhang\* Feng Tian† Qianying Wang\*‡

\*Lenovo Research

†Institute of Software, Chinese Academy of Sciences

§The Hong Kong University of Science and Technology (Guangzhou)

¶Jinan University

## ABSTRACT

This work presents *Dual-Stick*, a novel controller with two sticks connected at the end that innovates a *Dual-Ray* interaction paradigm to enrich raycasting input in Virtual Reality (VR). *Dual-Stick* leverages the inherent human dexterity in using everyday tools such as clamps and tweezers to adjust the relative angle between two sticks. This design supports *Dual-Ray* interactions that provide with a heuristics-based enhanced mechanism. It also offers more flexible manipulation by taking advantages of additional degrees of freedom provided by clamping angle. We conducted two studies to evaluate the effectiveness of *Dual-Ray* in target selection and manipulation tasks. The results indicated that *Dual-Ray* significantly improved efficiency in target selection compared to single-ray input but did not outperform the enhanced single-ray technique. In terms of manipulation, *Dual-Ray* effectively reduced completion time and mode switching compared to single-ray input.

**Index Terms:** Handheld controller, virtual reality, raycasting, 3D target selection and manipulation.

## 1 INTRODUCTION

Handheld controllers are an integral part of virtual reality (VR) system and user experience [29]. They enable users to intuitively interact with the virtual environment, translating hand actions into VR and providing precise control and tactility. An emerging and promising direction is switching over from traditional bulky controllers to more lightweight and sleek designs. Notable examples can be seen in recent research and industrial efforts dedicated to developing pen-like [17, 27, 34] and wand-like prototypes [64, 8] for VR controllers, facilitating interactions in a more accurate and intuitive way. However, lightweight devices maintain a ray-button input paradigm, and the limited degrees of freedom (DoF) in a single-ray system restrict 3D manipulations, such as scaling [63] and depth adjustments [11, 47]. Furthermore, the compact size of lightweight controllers prevents them from having as many buttons and haptic devices as traditional controllers [55], leading to reduced mode-switching options and diminished feedback, and the fact that selection accuracy can be hindered due to factors like muscle fatigue or hand tremors [48].

Previous research has primarily focused on enhancing raycasting input capabilities [14, 10, 48, 4, 76, 79, 45] and expanding device input channels [62, 19, 43, 1, 29, 78, 15]. Numerous

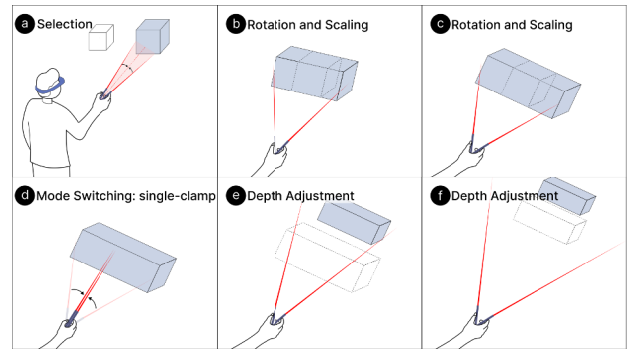


Figure 1: *Dual-Stick* with the novel *Dual-Ray* paradigm enables clamp selection (a) and adds DoF for scaling (b-c) or depth adjustment (e-f) via mode switching (d) during manipulation.

heuristic-based selection techniques [14, 10, 48] as well as multiple rays manipulation techniques [4, 76, 79, 45] have been proposed. These advancements enable users to perform more efficiently in complex tasks, such as occlusion target selection [14, 79, 76] and different trigger actions [45]. However, these enhancements are not innate capabilities of controllers and require different configuration parameters. Furthermore, expansion of input channels has been investigated by adding additional sensors such as capacitive sensors [62, 1], deformation-detecting [19], and cameras [43] to the device or by combining these with multimodality [78, 15]. These designs support rich grip and gestures input to replace buttons, although they require additional sensing units which might not always be available or convenient to use.

In this work, we drew inspiration from the dexterity of everyday tools like chopsticks and pliers to propose a controller called *Dual-Stick*, characterized by a dual-stick structure combined with a *Dual-Ray* interaction paradigm (Fig. 1). This design leverages finger dexterity, allowing for independent adjustments to each stick's position and the relative angle between them. The innovation of the dual-ray interaction paradigm lies in its selection mechanism, which replaces the traditional pointer-triggered selection with a buttonless clamp. Coupled with the added degree of freedom, this greatly enhances the flexibility of object manipulation with dual-stick tools.

We conducted two user studies to evaluate the performance of *Dual-Ray*. The first study focused on distant target selection, and the results showed that *Dual-Ray* was more efficient than traditional single-ray input, with selection time significantly reduced by 23.5% and error rates significantly lowered by 31.9%. However, it did not surpass a heuristically enhanced single-ray technique, where selection time was comparable, but the error rates were 9.4% points higher. The second study assessed distant object manipulation, and the results showed that *Dual-Ray* significantly reduced completion

\*e-mail: {linl1, zhangyu29, zhanglx2, wangqya}@lenovo.com

†Tong Wu and Tianren Luo are also with the School of Computer Science and Technology, while Teng Han and Feng Tian are with the School of Artificial Intelligence, University of Chinese Academy of Sciences.

‡corresponding author

time by 21.6% and the number of mode switches by 65.8% compared to single-ray input.

This paper makes the following contributions: i) *Dual-Stick* as a novel design of VR controller that leverages the extra input dimensions of dual rays, innovating a buttonless clamping mechanism. ii) design of the interaction techniques for *Dual-Ray* to support target selection, manipulation and mode switching; iii) evaluation of *Dual-Ray* that demonstrated its advantages in distant target selections and flexible object manipulation.

## 2 RELATED WORK

The *Dual-Stick* represents an innovative attempt to expand raycasting input capabilities and reduce button dependency, in the face of increasingly lightweight input devices. This effort closely aligns with ongoing research in VR input devices, raycasting innovations, and mode switching techniques.

### 2.1 VR Input Devices

Current VR input methods include vision-based tracking [49], wearable devices [30], and handheld controllers [73]. Vision-based devices, like Leap Motion and Oculus Quest, use computer vision for free-hand interactions but are susceptible to illumination and occlusion issues [50]. Wearable devices, such as data gloves, offer precise tracking but face challenges in fabrication and calibration [28]. Alternative modalities like eye gaze [12] and speech [15] show promise but require further refinement for intuitive 3D input [13]. Despite these advancements, mainstream VR still relies on handheld controllers (e.g., HTC Vive, Oculus Rift) that offer six degrees of freedom but can be bulky, costly, and lack precision in complex interactions [13].

This has spurred research into thin, lightweight controllers inspired by everyday tools, like pen-shaped controllers [34, 55], which enable precise input and are popular in VR sketching systems [16]. Other designs explore spatial input using wands [9], chopsticks [8], and mobile phones [44]. However, these compact controllers often compromise on buttons and haptic feedback. While sensors like capacitive [62] and deformation-detecting [19] technologies have expanded input channels, their practicality and availability vary. Furthermore, these devices are often akin to single stick tools, yet there are other familiar tools with potential for VR input which remain unexplored. This paper investigated dual-stick tools like chopsticks and tweezers. Existing literature that paid attention to these tools primarily focused on investigating their haptic design and applications, including VR Grabbers [77] and HapLinkage [35], while neglecting the exploration of their input capabilities.

### 2.2 Raycasting-based Selection and Manipulation

Target selection and manipulation are essential interactions in VR [32, 5], with raycasting techniques [26, 38] being widely used for interacting with distant objects. Raycasting enables users to manipulate a 5 DoF ray, allowing object selection beyond physical reach with minimal movement. However, traditional raycasting has limitations, particularly when selecting small or distant targets due to hand tremors [52]. To address this, researchers proposed using a cone-shaped ray, or "spotlight," to improve stability and increase selection range [38]. Techniques like manual disambiguation and heuristic approaches, such as the SQUAD [31] and IntenSelect [14], have been developed to resolve issues when multiple objects are present within the selection volume.

Another challenge with traditional raycasting is its difficulty in accurately designating specific locations in 3D space [57], leading to challenges with occluded target selection and restricted manipulation. Various techniques [21, 3, 57] have been proposed to overcome these limitations, often involving a fixed cursor controlled by VR controllers or touchscreens. Raycasting-based target manipulation often requires multiple DoF. Integrating higher DoF op-

erations allows users to manipulate objects more naturally, mimicking real-world movements. However, this can also introduce greater complexity and may not always lead to improved performance [71, 22]. Researchers also explored multi-finger [45, 4] and both hands [76, 79, 18] input methods to enhance ray expressiveness, such as the iSith technique [76], which calculates the shortest distance between two rays from each hand. Other multi-model techniques like gaze-based target manipulation [78] were also proposed, allowing for more complex VR interactions. In this work, *Dual-Stick* employs dual rays emitted from the two sticks to augment the DoF of input, with which we investigate users' ability to manipulate the rays for VR interactions.

### 2.3 Mode Switching Techniques

*Mode* was defined as a distinct setting within an interface where the same user input produces results different from those it would produce in other settings [56]. In typical interactive systems, some common activation modes include *Quasi Mode*, which relies on dynamic, context-based switching, and *Absolute Mode*, which ensures a stable and continuous mapping between user input and system response. [56]. *Mode switching* refers to transitioning from one mode to another [67]. Early mode-switching techniques focused on pucks, mice, and especially pens (styli) [67], with a significant body of research examining short stroke gestures [24], pressing firmly or lightly [37], stylus rolling [6], pen-holding postures [62, 25], and non-dominant hand usage [7] in pen and touch tabletop systems. In contrast, for VR/AR mid-interfaces, various hand gestures [61, 67] and head-based gestures [60] explored by researchers enable mode switching in mid-air, yet conventional VR applications continue to rely upon controller buttons and menu switching, as observed in Microsoft Maquette [40] and Oculus Medium [46].

Mode switching is crucial for target manipulation tasks, encapsulating multiple modes for object selection, movement, rotation, and scaling [72, 20]. This paper seeks to introduce a buttonless mode switching method, utilizing controller buttons as a baseline for evaluating mode switching efficiency.

## 3 DESIGN AND IMPLEMENTATION

We first systematically examined the design and usability of everyday hand tools with stick-like structures that could potentially inspire the design of our new controller, and categorized them based on their structural types and grip postures. This gave us an understanding of the common dual-stick structures and the ways users typically hold these devices. We then developed the *Dual-Stick* prototype and designed the *Dual-Ray* interaction paradigm, a shift in how user interaction is conducted for better target selection, manipulation, and mode switching.

### 3.1 Design Space of Stick-Like Handheld Tools

Considering the need for VR controllers to be rigid, directional, lightweight, and portable, we focused on stick-like everyday handheld tools, reviewing 64 tools from literature [35, 33, 70], commerce platforms (e.g., Amazon), and image sources (e.g., Pinterest). These tools were categorized by structure and grip posture using affinity diagramming (Fig. 2.a). We identified four *structure types*: *Single-Stick*, common in VR controllers (e.g., screwdrivers, wrenches); *Parallel Dual-Stick*, requiring precise finger control (e.g., chopsticks); *End-cross Dual-Stick*, enhancing stability for gripping (e.g., tweezers); and *Middle-cross Dual-Stick*, useful for cutting and clamping (e.g., pruners). Five *grip postures* were summarized: *Grab* (thumb and fingers wrap around, e.g., wrenches), *Claw* (similar to Grab but perpendicular), *Tripod* (fine operations, e.g., writing), *Clench* (fists control sticks, e.g., fitness tools), and *Stab* (one stick in each hand for balanced tasks).

It can be seen from the design space that although all three *Dual-Stick* structure tools are commonly used in life, *Parallel Dual-Stick*

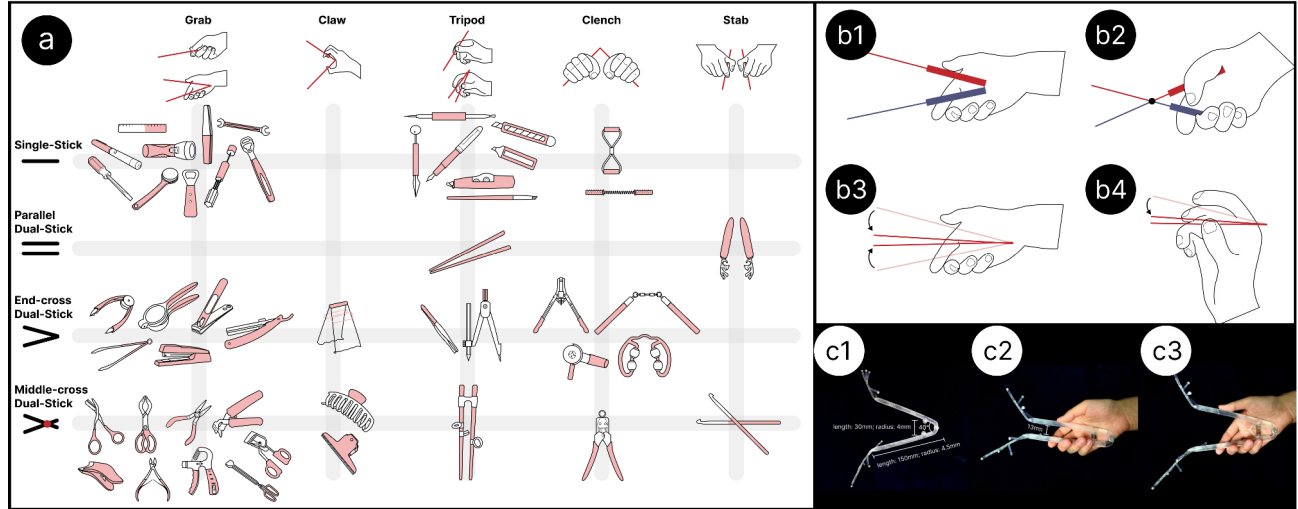


Figure 2: (a) A design space for stick-shaped handheld tools in daily life includes five grip postures (horizontal) and four structure types (vertical); Differences in spatial correspondence between structures: *End-cross Dual-Stick* (b1), *Middle-cross Dual-Stick* (b2), and variations in movement patterns for grip postures: *Grab* (b3), *Tripod* (b4); 3D-printed *Dual-Stick* prototype: natural state (c1), clamped in hand (c2), relaxed in hand (c3).

tools like chopstick often require skill training to master. Comparing the *End-cross Dual-Stick* with the *Middle-cross Dual-Stick*, we found that the *Middle-cross Dual-Stick* structure forms a Class I lever (fulcrum between effort and load) [74] to save force, while the *End-cross Dual-Stick* is classified as a Class III lever (effort between fulcrum and load) [74] for speed and range of motion [68]. When being used as a VR controller, the latter is more important. Moreover, considering the raycasting input, the *Middle-cross Dual-Stick* structure presents a reverse spatial correspondence between the user's actual manipulation end and the virtual ray end, which is against the user's intuition, but the *End-cross Dual-Stick* structure provides consistent ray control and perception (Fig. 2.b1-b2). Therefore, we decided to develop the prototype of the *Dual-Stick* controller based on the *End-cross Dual-Stick* structure.

Among the candidate grip postures for this structure, both the *Grab* posture and *Tripod* posture provide good pointing and do not require an additional gripping handle design. However, the *Grab* posture has the advantage of making use of the palm of the hand as much as possible to ensure stable support with the forearm naturally relaxed. In contrast, the *Tripod* posture involves more finger dexterity and its usage requires the forearm to be lifted for input. This could potentially lead to fatigue in long-duration tasks [34]. Additionally, we noticed differences in the movement patterns of the front end of the *Dual-Stick* when performing the clamping action with these two different postures (Fig. 2.b3-b4). The asymmetric movement pattern of the *Tripod* posture requires extra effort to reorient the *Dual-Stick* for accurate target clamping. Therefore, we designate the *Grab* posture as the default posture for subsequent design and experiment.

### 3.2 Prototype

Fig. 2.c1-c3 illustrate our 3D-printed prototype in *Formlabs Clear Resin (V4)*, a transparent SLA 3D printing material. It composed of two sticks, each measuring 150 mm in length and 9 mm in diameter, paralleling the size reference of the Apple Pencil [54]. This size provides a comfortable gripping experience. The sticks are connected together at one end by an M3\*13 screw. Through trial and error, we incorporated an 8 mm diameter spring (153.3 N/m), 30 mm in length and positioned 30 mm from the end of the sticks, to enable automatic recovery after clamping and provide primary haptic feedback similar to that of tweezers and barbecue tongs. This design allows the sticks to achieve a maximum opening angle of 40°

in their natural state, while maintaining a 13 mm distance from each other when closed. The average maximum angle of a relaxed grip is 32°. To track the device's location and orientation in space, we 3d printed two structurally asymmetric supports to ensure that the tracking process is unobstructed, and affixed five reflective spheres. The assembled prototype has a total weight of 38.03 grams.

### 3.3 Dual-Ray Interaction Design

Traditional controllers have typically used a ray-casting to achieve distant object pointing, together with buttons for selection and manipulation. In contrast, the *Dual-Stick* prototype adopts an end connection structure formed by two sticks, introducing an innovative interaction paradigm known as the *Dual-Ray*. The concept is inspired by how tools like pliers or tweezers clamp objects in the real world, which is extended to allow selection, manipulation, and mode-switching of distant objects within a VR environment.

#### 3.3.1 Target Selection

The initial design of the *Dual-Ray* selection mechanism involved two rays clamping and forming a 0-degree angle and simultaneously targeting the object. To evaluate its efficiency, we conducted a preliminary study with 8 participants between age of 24 to 35 ( $M = 28.5$ ,  $SD = 3.71$ ) from a local university, using a Fitts' law task. Participants were required to interact with two small spheres placed 2 meter in front of them and 1 meters apart; one sphere functioned as the start button and the other as the target. We varied six target sizes (0.143m, 0.097m, 0.067m, 0.046m, 0.032m, 0.023m) to represent six different Fitts' ID (index of difficulty), ranging from 3 to 5.5 in 0.5 increments—a common range for Fitts' ID settings [65, 69, 34]. The goal was to assess error rates associated with these difficulty levels. Our findings indicated that for IDs of 4 or higher, the error rate exceeded 40%, and for IDs of 5 or higher, participants were even unable to complete the selection task. These results led us to reconsider and seek improvements in our selection strategies.

Unlike real-world scenarios with two sticks, using two rays to "grab" a virtual object is impractical due to the absence of force resistance, causing the rays to easily pass through the object. This direct selection approach demands users to precisely and simultaneously maneuver the two rays to reach a target, increasing control complexity and making it more challenging than selecting with a single ray. Conversely, the clamping action of *Dual-Ray* introduces an additional projection cone (formed by the two rays and

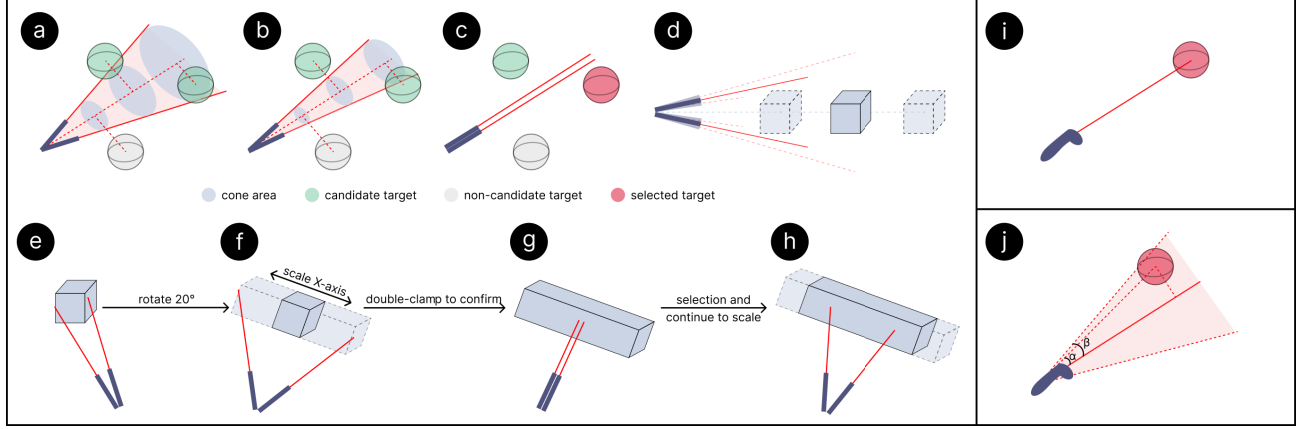


Figure 3: Target selection (a-c) and manipulation mechanism of *Dual-Ray*, including angle-based manipulation with scaling (e-h) and depth adjustment(d), where the object's movement and rotation are bound to the midline between the two rays. Traditional single-ray input (i) and the IntenSelect technique (j) compared in study 1.

their clamping angle) beyond pointing, which can be incorporated into the selection process. Therefore, an enhanced selection mechanism is under consideration, where the clamping action is utilized to eliminate the selection ambiguity by assigning likelihood scores to targets that update based on how much the clamped area overlaps with the target region.

This approach is common among various heuristic-based single-ray enhancement techniques [14, 53, 66]. For instance, the *IntenSelect* technique [14] employs spatial and temporal functions to assign a score to each target  $T_i$  (Fig. 3.j), ultimately selecting the one with the highest score:

$$T_i\text{score}(t) = T_i\text{score}(t-1)\lambda + (1 - \frac{\alpha(t)}{\beta})(1 - \lambda) \quad (1)$$

where  $\alpha(t)$  represents the angle between the ray and the center of the target at time  $t$ ,  $\beta$  is the threshold, and  $\lambda$  denotes the weight given to time decay. It allows for control over the selection's *stickiness* and *snappiness* by adjusting the  $\lambda$  value.

Getting inspired, we define the cone-shaped volume formed by the two rays emitted from the same point as the valid selection region. Targets within this conic volume qualify as candidate targets. This qualification is determined by comparing each target's radius with the projected distance from the target's center point to the cone's centerline (Fig. 3.a). In our current proof-of-concept, we limit our test to the bounding sphere of an object [14]. For each time step, the score of candidate target  $T_i$  is calculated by Eq.2:

$$T_i\text{score}(t) = T_i\text{score}(t-1) + \Delta S(t) \quad (2)$$

where  $i$  represents the candidate target's number,  $t$  is the current time step. The weight  $\Delta S(t)$  is calculated by Eq.3:

$$\Delta S(t) = \frac{S_{\text{overlap}}(t)}{S_{\text{cone}}(t)} \quad (3)$$

where  $S_{\text{cone}}(t)$  represents the cone-shaped volume's circular projection area on the target's depth plane (blue circular area in Fig. 3.a-b),  $S_{\text{overlap}}(t)$  is the overlapping area between  $S_{\text{cone}}(t)$  and the target with the same projection (dark green overlapping area in Fig. 3.a-b). Consequently, when the cone-shaped volume wholly encompasses the candidate target,  $\Delta S$  reaches its peak value of 1.

The process of target selection involves dynamically adjusting the score of each candidate target as the cone progressively narrows, eventually triggering the selection when the dual-stick angle reaches zero and selecting the target with the highest score

(Fig. 3.b-c). Since the scoring is a cumulative process, the timing of computation commencement holds significant importance. It necessitates a balance between *snappiness* – the urgency in responding to real-time changes; and *stickiness* – the score's resistance to short term fluctuations. Our pilot study showed that starting the calculation when the conic circle's diameter at the target depth is twice the target circle's diameter yields agreeable outcomes.

### 3.3.2 Target Manipulation

A conventional approach for manipulating objects with a single ray involves binding the object to the ray to perform translation and rotation. *Dual-Ray* extends this operation by enabling adjustments to an object's size or depth by changing the angle between two rays, and it facilitates switching between these two manipulation modes depending on task requirements. While separating the DoF enables more precise control [71, 22], our design integrates higher DoF, driven by the characteristics of the dual-stick tool and users' familiarity with such tools in daily use, thus enabling the simultaneous execution of multiple actions.

**Manipulation with Scaling:** Moving, rotating, and scaling are often done simultaneously on desktop devices due to the flexibility of multi-finger touch. In VR, however, these tasks typically require both hands or frequent mode switching. *Dual-Ray* enables users to perform these operations simultaneously with one hand. Object movement and rotation are bound to the *Dual-Ray*'s centerline, while changes in the ray's angle control scaling. *Dual-Ray* also allows axis-specific scaling based on the object's orientation (Fig. 3.e-f). When the sticks reach their angular limits, a reset can be performed through mode switching, allowing continued scaling (Fig. 3.g-h).

**Manipulation with Depth Adjustment:** Adjusting the depth of an object in VR can be challenging due to the limitations of single ray freedom, often requiring the use of joysticks or trackpads on the controller. *Dual-Ray* allows the user to adjust an object's distance along the centerline of the *Dual-Ray* through manipulation of the angle between the two rays. The larger the angle between the rays, the farther the object is (Fig. 3.d). The object's movement and rotation are also tied to the *Dual-Ray*'s centerline.

### 3.3.3 Mode Switching

In our design, we define *single-clamp* and *double-clamp* operations instead of button mode switching, similar to the single and double clicks of a mouse. A *single-clamp* action is recognized when the



user's instantaneous acceleration exceeds a predetermined threshold and the clamp angle returns to zero within 500ms [23, 51]. A *double-clamp*, on the other hand, is recognized when the clamp angle reaches zero twice within the same timeframe.

In the selection mode, the user automatically transitions to the target manipulation mode after completing the target selection through a determined *single-clamp* action. While in the target manipulation mode, the *Dual-Ray* begins to adjust when it touches the target edge or when the angle surpasses a predefined threshold ( $15^\circ$  for our setting). Users can confirm the current adjustment and revert to the selection mode using a *double-clamp* action. Additionally, a *single-clamp* action facilitates the swap between scaling manipulation and depth adjustment, which prevents conflict with the *double-clamp* exit action.

#### 4 STUDY 1: PERFORMANCE IN VR TARGET SELECTION

The objective of this study was to evaluate the effectiveness of the *Dual-Ray* clamping object selection mechanism in the context of distant and dense target selection tasks. We compared our method against the *IntenSelect* technique [14], which similarly uses scoring heuristic-based enhancement. Meanwhile, we employed the conventional single-ray of commercial controllers as the baseline.

##### 4.1 Participants and Apparatus

We recruited 12 participants (7 females) between the age of 21 to 30 years ( $M = 24.08$ ,  $SD = 2.1$ ) from a university via mailing lists and posters. Their technical backgrounds included industrial design, computer applications, and user experience design. All participants reported to be right-handed. They all had experience using *End-Cross Dual-Stick* structure tools like tweezers and clamps in daily life, but not very often. Each participant was rewarded with USD20 for their participation. Eight of them had experience with ray-casting techniques in virtual reality before. The studies in this work obtained ethics approval from a local university.

The *Dual-Stick* was tracked by four OptiTrack Flex 13 cameras (120 FPS). We used Oculus Quest 2 headset as the device to render the virtual environment for users. The software was implemented in C# in Unity 3D (2021.3.20) with the optitrack plugin (1.4.0), which was driven by a Windows 10 desktop (CPU: i7-6700, 16 GB, GPU: Geforce GTX 1080).

##### 4.2 Design and Procedure

We designed a within-subject experiment with three independent variables: *Techniques*: *Single-Ray*, *IntenSelect* (*IntenS*), *Dual-Ray*; *Densities*: *Low* (15), *High* (40); and *Sizes*: *Big* (8 cm), *Small* (4 cm). The order of techniques was balanced between participants using a Latin square. The *Densities* and *Sizes* were set following Baloup et al. [3]. *High* and *Low* densities are comparative to each other. All the targets were spread out at pseudo-random positions in a 2m diameter sphere without visual occlusion, whose center was 2m in front of the participants. Participants were allowed to rest between blocks. The study design was: 12 *Participants*  $\times$  3 *Techniques*  $\times$  2 *Densities*  $\times$  2 *Sizes*  $\times$  3 *Blocks*  $\times$  10 *Targets* = 4,320 trials in total. The sample size was ensured by conducting an adequate number of trials per participant.

At first, we introduced the goal of the experiment and guided the participant to use the device. They were provided sufficient time to get familiar with the tasks and techniques before the formal study started. The virtual scene was set in a confined space measuring 7 square meters. They were standing and maintaining a natural and comfortable position for their arms (Fig. 4.a-b). The two rays of the *Dual-Ray* are set to red and green for differentiation. The *IntenS* originally had an additional curve connected to the selected target as an indicator. While this provided helpful cues, it also introduced potential visual interference during selection, which fell outside the

scope of our analysis. Additionally, this feedback could lead to inconsistent selection strategies, such as selecting prematurely based on predictions or waiting for more precise aiming, resulting in varied outcomes. Therefore, we excluded the visual curve from *IntenS* in this study. Prior to the initiation of each condition, a red start button was presented to the participants. Upon selection of this button, the condition was initiated. The desired target was highlighted in blue while other targets were displayed in white. Participants were required to complete the trials as fast and as accurately as possible, and were not permitted to proceed to the next target until the current one had been accurately selected. Rest periods were allowed between blocks to minimize the effects of fatigue. The experiment lasted around 30 min per participant.

To ensure equitable comparisons of completion times and error rates across conditions, a singular sequence of targets was generated for all participants, techniques, and blocks, prior to the commencement of the experiment. After the experiment, the participant was required to fill a questionnaire based on a 7-point Likert scale (the higher score the better) to give subjective scores for all techniques. To reduce participants' burden, we referenced [39] and employed several key metrics instead of the standard SUS questionnaire to assess the following: **Performance**: *how at ease you feel and how efficiently you can select targets*; **Non-Fatigue**: *the degree to which the technique minimizes physical or cognitive strain over time*; **Sense of Control**: *the perceived ability to effectively manage and direct your actions during interaction* [42]; **Preference**: *the techniques you find most intuitive or enjoyable to use*.

##### 4.3 Results

To analyze the collected data, we first discarded the outliers that deviated more than three standard deviations from the mean value ( $mean \pm 3std.$ ) in each condition (49 trials, 1.1%). We measured the completion time and error rates of the task. A repeated-measures ANOVA was conducted, followed by post-hoc pairwise comparisons with Bonferroni correction. The results are shown in Fig. 4.c.

###### 4.3.1 Completion Time

**Main Effects.** Completion time, defined as the interval between two selections, was analyzed using a Shapiro-Wilk test, which indicated a non-normal distribution and necessitated the Aligned Rank Transform (ART) [75]. Repeated-measures ANOVA revealed significant main effects of *Technique* ( $F_{2,11} = 8.994$ ,  $p = 0.001$ ), *Size* ( $F_{2,11} = 66.395$ ,  $p < 0.001$ ), and *Density* ( $F_{2,11} = 61.036$ ,  $p < 0.001$ ). Post hoc test revealed that both *Dual-Ray* ( $M = 1588.91ms$ ,  $SD = 126.41ms$ ) and *IntenS* ( $M = 1622.58ms$ ,  $SD = 40.63ms$ ) were significantly faster than *Single-Ray* ( $M = 2079.40ms$ ,  $SD = 154.94ms$ ) ( $p = 0.044$  and  $p < 0.001$ , respectively). However, no significant differences were found between *Dual-Ray* and *IntenS* ( $p = 0.542$ ).

**Interaction Effects.** The analysis revealed a significant interaction effect of *Technique* $\times$ *Size* ( $F_{2,11} = 38.687$ ,  $p < 0.001$ ). Post hoc tests indicated that no significant differences were observed between *Small* and *Large* targets for *Dual-Ray* ( $p = 0.147$ ). Moreover, both *Dual-Ray* and *IntenS* were significantly faster than *Single-Ray* only for *Small* targets ( $p = 0.017$  and  $p < 0.001$ , respectively). A significant interaction effect of *Technique* $\times$ *Density* was also found ( $F_{2,11} = 7.177$ ,  $p = 0.004$ ). Post hoc tests revealed significant differences between *Low* and *High* densities for both *Dual-Ray* ( $p < 0.001$ ) and *IntenS* ( $p = 0.002$ ). Additionally, significant differences were observed between *Dual-Ray* and *Single-Ray* for both *Low* and *High* densities (both  $p < 0.001$ ).

###### 4.3.2 Error Rates

**Main Effects.** Error rates, defined as the ratio of failed selections before a successful one to total selections, were normally distributed according to a Shapiro-Wilks test at the 5% level. Repeated-measures ANOVA revealed significant main effects of

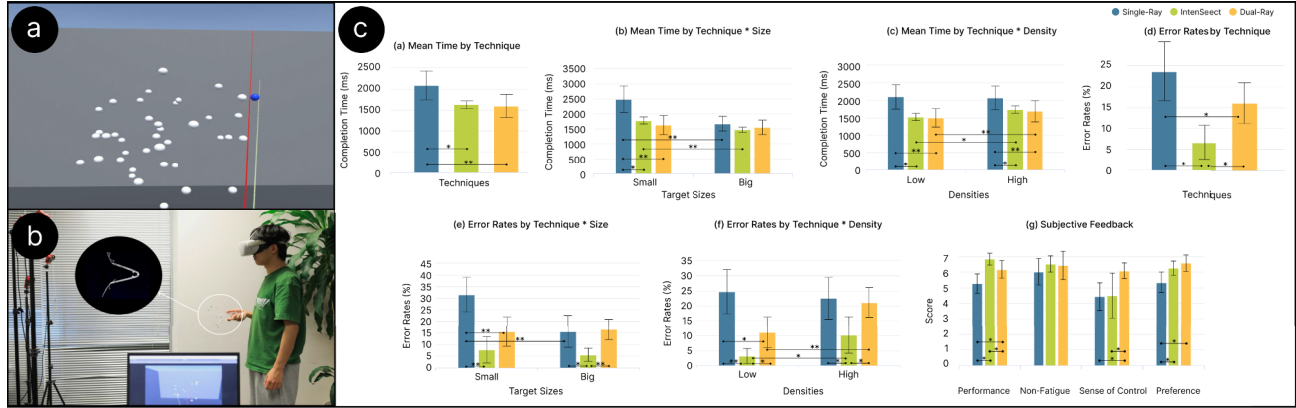


Figure 4: (a-b) User interface for selection with *Dual-Ray* and study 1 settings; (c) Mean times and error rates, with 95% confidence intervals error bars, and subjective feedback, where error bars indicate the standard deviation. Significant effects are marked (\* =  $p < .05$  and \*\* =  $p < .001$ ).

*Technique* ( $F_{2,11} = 20.260$ ,  $p < 0.001$ ), *Size* ( $F_{2,11} = 18.709$ ,  $p = 0.001$ ), and *Density* ( $F_{2,11} = 36.384$ ,  $p < 0.001$ ). Post hoc tests showed that the error rates of *Dual-Ray* ( $M = 16.05\%$ ,  $SD = 2.22\%$ ) and *IntenS* ( $M = 6.64\%$ ,  $SD = 1.87\%$ ) were significantly lower than those of *Single-Ray* ( $M = 23.57\%$ ,  $SD = 3.1\%$ ) ( $p = 0.044$  and  $p = 0.001$ , respectively). Additionally, *IntenS* showed significantly lower error rates than *Dual-Ray* ( $p = 0.003$ ).

**Interaction Effects.** The analysis revealed a significant interaction effect of *Technique* × *Size* ( $F_{2,11} = 39.036$ ,  $p < 0.001$ ). Post hoc tests revealed significant differences between *Dual-Ray* and *Single-Ray* only for *Small* targets ( $p < 0.001$ ). *IntenS* had lower error rates than both *Dual-Ray* and *Single-Ray* for *Large* targets ( $p < 0.001$  and  $p = 0.032$ , respectively), but for *Small* targets, no significant differences were found between *IntenS* and *Dual-Ray* ( $p = 0.052$ ). Additionally, no significant differences were observed between *Small* and *Large* targets for either *Dual-Ray* ( $p = 0.658$ ) and *IntenS* ( $p = 0.173$ ). The analysis also revealed a significant interaction effect of *Technique* × *Density* ( $F_{2,11} = 10.997$ ,  $p < 0.001$ ). Post hoc tests showed that *Dual-Ray* had lower error rates than *Single-Ray* only for *Low* densities ( $p = 0.001$ ). *IntenS* exhibited lower error rates than both *Dual-Ray* and *Single-Ray* for *Low* ( $p = 0.014$  and  $p < 0.001$ , respectively) and *High* ( $p = 0.007$  and  $p = 0.02$ , respectively) densities. Significant differences were also observed between *Low* and *High* densities for both *Dual-Ray* ( $p < 0.001$ ) and *IntenS* ( $p = 0.005$ ).

#### 4.3.3 Subjective Feedback

A Friedman test revealed significant effects of *Technique* on *Performance* ( $\chi^2(2) = 21.571$ ,  $p < 0.001$ ), *Sense of Control* ( $\chi^2(2) = 10.905$ ,  $p = 0.004$ ), and *Preference* ( $\chi^2(2) = 15.048$ ,  $p = 0.001$ ), but not on *Non-Fatigue* ( $\chi^2(2) = 2.114$ ,  $p = 0.347$ ) (Fig. 4.c). Post-hoc Wilcoxon Signed-Rank tests (Bonferroni corrected) showed that *Dual-Ray* outperformed *Single-Ray* in *Performance* ( $Z = -3.317$ ,  $p = 0.001$ ) but was rated lower than *IntenS* ( $Z = -2.530$ ,  $p = 0.011$ ). For *Sense of Control*, *Dual-Ray* was rated higher than both *IntenS* ( $Z = -2.547$ ,  $p = 0.011$ ) and *Single-Ray* ( $Z = -2.795$ ,  $p = 0.005$ ), although it did not reduce *Fatigue*. Participants appreciated the spring feedback in *Dual-Ray* but desired more tactile feedback. In terms of *Preference*, both *Dual-Ray* and *IntenS* were favored over *Single-Ray* ( $p < 0.05$ ), with no significant difference between *Dual-Ray* and *IntenS* ( $Z = -1.414$ ,  $p = 0.157$ ).

#### 4.4 Study Discussion

Results showed that *Dual-Ray*'s performance was unaffected by *Size*, but was significantly impacted by *Density*, with a significant decline under high-density conditions. This could be attributed

to the heuristic mechanism that effectively reduces target mis-selection, but cannot mitigate occlusion effects at varying depths. Integrating techniques such as *DepthRay* [21] and *vMirror* [36] could further enhance performance in high-density environments.

**Dual-Ray vs. Single-Ray.** The *Dual-Ray* enhanced target selection by using the conical area between two rays to increase the active selection area, resulting in a significant reduction in overall selection time and error rates compared to *Single-Ray*. However, as the target size increased, the benefit of the enlarged active area diminished, leading to no significant differences between the two techniques for large targets. We also noticed high error rates for *Single-Ray*, although we reminded the participant to select the target as quickly and accurately as possible before starting the experiment. By further analyzing the data, we found that in *Small* and *Low* condition, there were cases where users selected them multiple times, e.g., 73 trials were selected 2-3 times, and 10 trials were selected 4-5 times, resulting in higher overall error rates.

**Dual-Ray vs. IntenS.** The results indicated that *Dual-Ray* and *IntenS* yielded similar overall task completion times, but *IntenS* exhibited significantly lower error rates across almost all conditions, except for small targets. One possible explanation is that both techniques utilized heuristic methods to accumulate each object's score, but differ in user interface. *IntenS*, as a single-ray system with a precise pointing marker, requires more time for aiming but results in lower error rates. In contrast, *Dual-Ray* lacks a precise indicator, relying on user intuition, which speeds up the process but increases errors. Future work could explore adding explicit selection indicators for *Dual-Ray*, such as visualizing the conical selection area or midline, to reduce errors. Another contributing factor may be device stability: the commercial-grade *Oculus* VR device provides superior ergonomics and tracking stability, while the *Dual-stick* still in its prototype phase, resulted in higher error rates compared to *IntenS*, even approaching *Single-Ray* under high-density conditions.

**User Feedback.** The *Dual-Stick* offers variable force feedback via a spring mechanism during object selection, simulating the sensation of grasping a real object and affording users a heightened *Sense of Control* compared to two *Oculus*-based techniques. Users noted that the fatigue associated with *Single-Ray* and *IntenS* arose from the cognitive effort required for precise aiming, while the fatigue with *Dual-Ray* was due to the physical strain of pressing on the spring. Consequently, no significant differences in *Non-fatigue* were observed across the three techniques. This feedback suggests that future iterations of *Dual-Stick* could benefit from adaptive spring force adjustments or active motors to reduce physical strain.

## 5 STUDY 2: PERFORMANCE IN VR TARGET MANIPULATION

Beyond the clamp selection of the target, another advantage of the *Dual-Ray* is to adjust the angle between the two rays as an additional input, enhancing the freedom of manipulation. To verify this, in the second study, we evaluated the performance of the *Dual-Ray* technique for remote object manipulation tasks with mode switching mechanism, and compared it to the *Single-Ray* technique used in commercial controllers. *IntenS* was not included in the comparison as it is not designed for target manipulation.

### 5.1 Participants and Apparatus

We recruited 12 participants (6 females) between the ages of 22 and 32 years ( $M = 24.7$ ,  $SD = 2.5$ ) from the university. Their technical backgrounds included industrial design, computer applications, and user experience design. They did not participate in Study 1. All participants reported being right-handed and received USD20 as compensation for their participation. The same apparatus as the first study was used.

### 5.2 Design and Procedure

We designed a typical object manipulation task where participants were asked to move a cuboid to a target location while maintaining its size and orientation to match the target. This was a within-subjects study with three independent variables: *Techniques*: *Single-Ray*, *Dual-Ray*; *Sizes*: *Small* (50%), *EqualSize* (100%), *Large* (200%), relative to the initial cuboid; and *Depths*: *Near* (-2 m), *EqualDepth* (0 m), *Far* (+2 m) from the user on a sphere relative to the initial cuboid.

We referred the guidelines by Bergstrom et al. [5]. For each condition, the initial size of the cuboid to be manipulated was 0.5 m × 1 m × 0.5 m, large enough to avoid selection difficulties, located in the centre of the scene at a depth of 5 m from the user. The target cuboid was positioned at 10 random points on the sphere plane at the specified depth, within a yaw and pitch angle of 45°. The positions were the same for all participants, but appeared in a random order. The target cuboid was of the specified size and rotated by a random angle in the z-direction. The order of the *Dual-Ray* and *Single-Ray* techniques was counterbalanced. Thus, the experiment design was: 12 Participants × 2 Techniques × 3 Sizes × 3 Depths × 10 Positions = 2160 trials in total. The sample size was ensured by conducting an adequate number of trials per participant.

The operations of the *Dual-Ray* technique included *Selection*, *Manipulation with Scaling*, *Manipulation with Depth Adjustment*, and *Mode Switching* via *single-clamp* and *double-clamp*. *Manipulation with Depth Adjustment* was set as the default mode after selecting the target, with adjustment rates of 0.5 m/° for depth and 0.1 times/° for scaling. For the *Single-Ray* technique, we used the operations of commercial applications like *ShapeXR* [58] for Oculus. Users selected the target with the *IndexTrigger*, attached to it to the handle for translation and rotation, and used the *HandTrigger* to switch between scaling and depth adjustment modes. Adjustments were made by altering the y-value of the *ThumbStick*, and manipulation was confirmed with the *A* button. The current manipulation mode was displayed in the interface for both techniques.

Participants were briefed on the study setup and wore an Oculus Quest 2 headset while standing (Fig. 5.b). They were given time to familiarize themselves with the tasks and techniques before the experiment began. Each trial started with a white initial cuboid and a semi-transparent orange target cuboid. Participants manipulated the initial cuboid to match the target's position, rotation, and size. In the *Dual-Ray* technique, the two rays were differentiated by red and blue colors, with a dotted green line connecting the *Dual-Stick* to the object to indicate the manipulated target. To assist with depth and size perception, 3D arrows were used as indicators (Fig. 5.a). The degree of overlap between cuboids was shown by the target

cuboid's gradient color, with darker shades (closer to yellow) indicating greater overlap. A trial was successful when the cuboids' position, angle, and scale discrepancies were within 10 cm, 10 degrees, and 30%, respectively, at which point the target cuboid turned green, and the experiment proceeded to the next trial. Each participant's session lasted about 30 minutes.

### 5.3 Measures

For the quantitative assessment, we measured *completion time* and *number of mode switches* for each trial. *Completion time* included selection, manipulation, and mode switching times. Selection time was measured from trial start to target selection, achieved via *single-clamp* or button press. For *Dual-Ray*, manipulation time included scaling and depth adjustment, measured from mode entry to just before a clamp switch or mode exit. For *Single-Ray*, manipulation time was the sum of translation, rotation, scaling, and depth adjustment times, measured from button press to release. Mode switching time covered all non-selection and non-manipulation periods: the duration of *single-clamp* and *double-clamp* actions for *Dual-Ray*, and cumulative time between key presses for *Single-Ray*. Mode switching frequency was tracked as the total number of transitions between operation modes. For subjective assessment, we asked participants to fill a questionnaire based on the same 7-point Likert scale as in Study 1 to give subjective scores for two techniques after the experiment.

### 5.4 Results

To analyze the collected data, we still excluded the outliers that deviated more than three standard deviations from the mean value ( $mean \pm 3std.$ ) in each condition (40 trials, 1.9%). We measured the completion time and the number of mode switching, the results were shown in Fig. 5.c and Fig. 5.e. A repeated-measures ANOVA was conducted, followed by post-hoc pairwise comparisons with Bonferroni correction. No instances of incorrect target selection prior to manipulation were observed during the experiment.

#### 5.4.1 Completion Time

**Main Effects.** A Shapiro-Wilk test indicated that the data were non-normally distributed and underwent through ART [75]. Our Analysis showed significant main effects of *Technique* ( $F_{1,11} = 26.51$ ,  $p < 0.001$ ), *Size* ( $F_{2,22} = 136.05$ ,  $p < 0.001$ ), and *Depth* ( $F_{2,22} = 76.29$ ,  $p < 0.001$ ) on task completion time. Post hoc tests revealed that *Dual-Ray* ( $M = 9.97s$ ,  $SD = 0.53s$ ) was significantly faster than *Single-Ray* ( $M = 7.82s$ ,  $SD = 0.54s$ ) ( $p < 0.001$ ). We further split the completion time using the same analysis and found that *Technique* had a significant effect on manipulation time ( $F_{1,11} = 7.89$ ,  $p = 0.017$ ) and mode switching time ( $F_{1,11} = 37.55$ ,  $p < 0.001$ ). Post hoc tests showed that *Dual-Ray* significantly reduced manipulation time ( $p = 0.017$ ) and mode switching time ( $p < 0.001$ ) compared to *Single-Ray*.

**Interaction Effects.** The interaction effects were found between *Technique* × *Size* ( $F_{2,22} = 17.18$ ,  $p < 0.001$ ). Post hoc tests revealed significant differences across sizes for both *Dual-Ray* (*Small* vs. *Large*:  $p = 0.018$ ; the others:  $p < 0.001$ ) and *Single-Ray* (all  $p < 0.001$ ). Moreover, *Dual-Ray* demonstrated superior speed compared to *Single-Ray* for *Small* ( $p < 0.001$ ), *EqualSize* ( $p = 0.041$ ), and *Large* ( $p = 0.004$ ) target cuboids. Interaction effects were also observed for *Technique* × *Depth* ( $F_{2,22} = 35.18$ ,  $p < 0.001$ ). Post hoc tests revealed significant differences across depths for *Single-Ray* (*Near* vs. *Far*:  $p = 0.001$ ; the others:  $p < 0.001$ ), and only between *Near* and *EqualDepth* for *Dual-Ray* ( $p = 0.03$ ). Compared to *Single-Ray*, *Dual-Ray* demonstrated significantly faster performance for *Near* and *Far* target cuboids (both  $p < 0.001$ ), with no significant differences observed for *EqualDepth* ( $p = 0.972$ ).



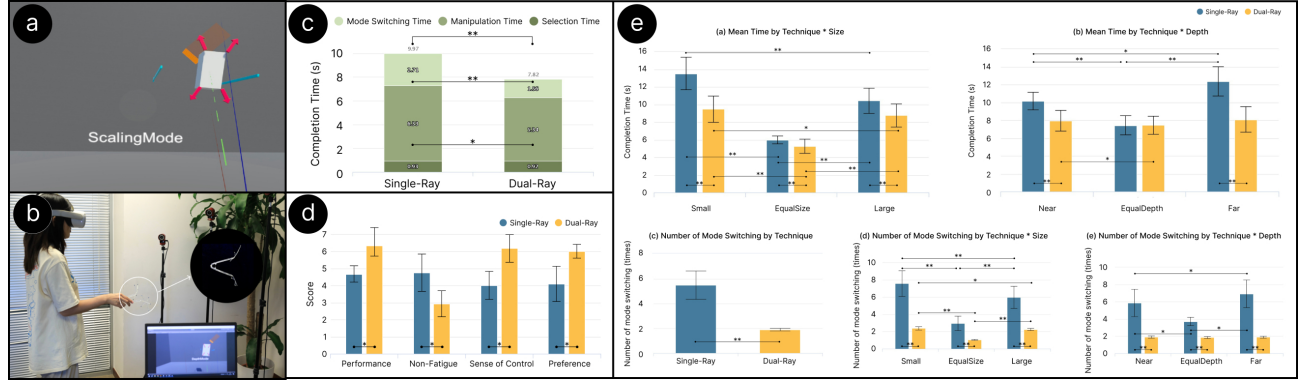


Figure 5: (a-b) User interface for manipulation with *Dual-Ray* and study 2 settings; (c-e) Mean completion time and number of mode switching, with 95% confidence intervals error bars, and subjective feedback, where error bars indicate standard deviation. Significant effects are marked (\* =  $p < .05$  and \*\* =  $p < .001$ ).

#### 5.4.2 Number of Mode Switching

**Main Effects.** A Shapiro-Wilk test indicated non-normal distribution, so we applied ART [75]. Our analysis showed significant effects of *Technique* ( $F_{1,11} = 53.31, p < 0.001$ ), *Size* ( $F_{2,22} = 130.34, p < 0.001$ ), and *Depth* ( $F_{2,22} = 19.78, p < 0.001$ ) on mode switches. Post hoc tests revealed that *Dual-Ray* ( $M = 1.88, SD = 0.05$ ) had significantly fewer switches than *Single-Ray* ( $M = 5.50, SD = 0.52$ ) ( $p < 0.001$ ).

**Interaction Effects.** Interaction effects were observed between *Technique* × *Size* ( $F_{2,22} = 48.03, p < 0.001$ ). Post hoc analyses revealed significant differences across *Size* for both *Dual-Ray* (Small vs. Large:  $p = 0.022$ ; the others:  $p < 0.001$ ) and *Single-Ray* (all  $p < 0.001$ ), with *Dual-Ray* consistently exhibiting lower error rates than *Single-Ray* for all *Size* conditions (all  $p < 0.001$ ). Interaction effects were also significant for *Technique* × *Depth* ( $F_{2,22} = 18.89, p < 0.001$ ). Post hoc tests showed significant differences across *Depth* for *Single-Ray* (Near vs. EqualDepth:  $p = 0.018$ ; Near vs. Far:  $p = 0.007$ ; EqualDepth vs. Far:  $p = 0.001$ ), but no significant differences for *Dual-Ray* (all  $p > 0.05$ ). Furthermore, *Dual-Ray* demonstrated significantly lower error rates than *Single-Ray* across all *Depth* conditions (all  $p < 0.001$ ).

#### 5.4.3 Subjective Feedback

We asked all participants to rate from 1 to 7 from the same four perspectives as in Study 1, the higher score the better (Fig. 5.d). A Wilcoxon Signed-Rank showed a significant differences between *Techniques* for *Performance* ( $Z = -3.12, p = 0.002$ ), *Preference* ( $Z = -2.98, p = 0.003$ ), *Non-Fatigue* ( $Z = -3.09, p = 0.002$ ) and *Sense of Control* ( $Z = -3.11, p = 0.002$ ). The results showed that users preferred the *Dual-Ray*, which performed better and provided users with a sense of control. “*Dual-Ray is very flexible, I can adjust the size of the object according to the arrows while moving.*” (P1). However, prolonged use led to fatigue, as noted by one participant, “*my palms felt tired after clamping many times.*” (P9).

#### 5.5 Study Discussion

*Dual-Ray* leveraged additional DoF, reducing three manipulation modes into two, which significantly decreased overall task completion time and mode-switching frequency. Detailed analysis showed that *Dual-Ray* minimized both manipulation and mode-switching time, while the sufficiently large target sizes resulted in no significant differences in selection times between the two techniques. The analysis of interaction effects showed that *Dual-Ray* was less influenced by changes in depth, both in terms of completion time and the number of mode switches. Compared to single-ray, no significant differences were observed under the *EqualDepth* condition as well.

This may be due to *Dual-Ray*’s default manipulation mode, which involves depth adjustment, requiring users to recalibrate depth each time they begin manipulation, even for targets at the same depth. We can assume that if the default mode were scaling, a similar pattern would emerge across different target sizes. Additionally, making objects smaller or more distant presents greater challenges for both techniques, as users’ perception and judgment become less accurate. We also observed that the especially high frequency of mode switches in *Single-Ray* could be attributed to users’ inability to complete transformations in a single attempt, necessitating multiple switches for depth or size adjustments.

Despite its efficiency, *Dual-Ray* caused more user fatigue, especially during prolonged use, due to challenges in maintaining grip stability and dealing with resistance from the springs. As *Dual-Ray* is currently a simple prototype, future iterations should focus on improving ergonomic design to enhance user comfort. For example, reshaping the handles to fit the user’s palm more naturally and incorporating automatic angle adjustments between the sticks could help reduce fatigue during operation.

#### 6 DISCUSSION AND FUTURE WORK

As VR controllers become lighter, traditional single-ray and button-based interactions face challenges in input accuracy and flexibility. This research explores a *Dual-Stick* tool, inspired by everyday tools like clamps and tweezers, using a *Dual-Ray* input paradigm to enhance clamp selection and offer additional DoF for manipulation. Our first study evaluated *Dual-Ray* in a target selection task, showing improved efficiency through a clamping mechanism that expands the selection range, with selection times comparable to the single-ray enhancement technique. The second study confirmed that controlling the angle between the *Dual-Rays* reduced task completion time and mode switching in manipulation tasks. These findings suggest that the *Dual-Stick* structure and *Dual-Ray* paradigm offer innovative directions for future VR input devices. However, our paper has limitations that require further research.

##### 6.1 Dual-Ray as A New Extension of Raycasting Input

*Dual-Ray* represents a novel extension of VR raycasting input, altering the conventional single-ray with button supplement paradigm. Through the collaboration of two rays to execute selection and manipulation tasks, it offers a unique avenue for object interaction based on clamping mechanism. This is the first exploration of this paradigm, and our findings offer a preliminary validation of *Dual-Ray*’s potential in facilitating distant VR target selection, manipulation, as well as real-world object interaction. However, in our first study, we only demonstrated that *Dual-Ray* outperforms



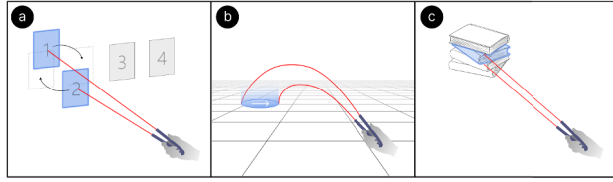


Figure 6: Other benefits of *Dual-Ray*: (a) independent control of two rays; (b) enhanced teleportation; (c) select narrow objects.

traditional single-ray and achieves close performance to similar heuristic-based enhancement techniques, and it is unclear how this performance compares to more sophisticated techniques. Future work could compare advanced single-ray selection enhancement techniques, such as Bubble Cursor [39] and Quad Cone [31, 38].

In addition to assisting in basic target selection and manipulation tasks, *Dual-Ray* has more interesting benefits that deserve further exploration. As shown in Fig. 6.a, future research could investigate the ability of the user to independently manipulate two rays, e.g., to enable rapidly exchange the position of two paintings on the wall. When using teleportation techniques in VR, the post-transportation orientation is controlled by changing the relative positions of the *Dual-Ray* (Fig. 6.b). Taking advantage of the *Dual-Ray*'s capability to select small targets, it can also be used in VR office and gaming scenarios to clamp narrow items such as books, paper, bows and arrows to enhance the immersive experience (Fig. 6.c).

However, the *Dual-Ray* input mechanism also brings some new challenges. The first is the user's habit and acceptance. The single ray has decades of history and is very familiar to users, but the operation of the *Dual-Ray* needs to be re-learned and become naturalized. Secondly, user needs can vary, with certain tasks requiring less rather than more manipulation freedom [22], such as rotate a spotlight about a single axis or making a tree taller while not moving or rotating it. *Dual-Ray* should allow the user to combine or limit the DoF for different tasks in the future. The third challenge is the higher requirements for mode switching. *Dual-Ray* requires more finger involvement than single-ray, so the mode switching should minimize finger movement. This informs our preference for a buttonless mode switching design.

## 6.2 Alternative Mode Switching Approaches

Our research demonstrated that the clamping mechanism works well to support mode switching in basic target selection and manipulation tasks. However, when frequent switching between multiple options is required in VR applications, using clamp switching may be inefficient and fatiguing. Future work could explore other buttonless mode switching approaches such as pen-based gestures [6, 34] or micro-gestures [59]. Pen-based gestures encompassing mid-air tilting, poking, and rolling can be easily recognized and have been investigated in VR context [34]. Micro-gestures necessitate only minor finger movements and minimally interrupt the primary task. [2]. Although we advocate the use of a buttonless mode switching mechanism, it does not mean that we are completely against buttons. A small number (1 or 2) of button designs could still be compatible in the future. The key lies in the design of the location and form of the buttons. The button locations should be ergonomically planned, and the buttons should favor touch-sensitive rather than physical activation to ensure stable input.

## 6.3 Other Device Formats that Support Dual-Ray

The implementation of the *Dual-Ray* interaction mechanism hinges on a suitable controller form. While the *Dual-Stick* controller represents one such design, a variety of alternative devices may be considered. Future research could explore the adaptation of existing single-stick controllers, introducing flexibility or foldable properties, thereby facilitating seamless transition between single and

dual stick forms. In this way, the single ray continues to promote stable and continuous input for basic input or sketching tasks, while dual ray allows for more freedom and fine control, ideal for tasks that require precise selection or complex manipulation. Alternatively, transforming real-world dual-stick objects (e.g., scissors or pliers) into VR input tools by integrating a tracking module could expand the range of available devices. This approach could reduce reliance on specialized VR controllers, allowing users to repurpose everyday objects as VR input tools when needed.

## 6.4 Multimodal Feedback Enhancement

The use of multimodal feedback in VR can enhance the user experience and improve the interaction efficiency [41]. In terms of visual feedback, *Dual-Ray* incorporates an additional centerline connecting the *Dual-Stick* to the selected object alongside the two rays, facilitating intuitive identification of the object under manipulation. It would be beneficial in the future to visualize the plane that the rays clamp onto, or the cone they encompass, thereby enhancing the visualization of the selection range. The haptic feedback of *Dual-Ray* is derived from spring resistance. While this has been shown to enhance immersion and enjoyment, it does not provide a tactile cue indicating object contact during the clamping process. Future research could enhance the user experience by integrating a simple vibration motor and audible cues into the *Dual-Stick*. However, the added weight and altered shape might make the controller less comfortable to hold, leading to increased fatigue. Therefore, it is important to strike a balance between enhancing the immersive experience and preserving comfort. Such features were omitted from our prototype to maintain the validity of the comparison with traditional VR controllers.

## 6.5 Handheld Input Device for Spatial Interaction

As spatial interaction and computation continue to evolve, the future direction for handheld controllers is gravitating towards lightweight designs and widespread compatibility. The lightweight design facilitates ease of portability and operation of devices in various settings. The design of *Dual-Ray* is a novel attempt towards the goal, with its advantages and challenges being aware of and investigated. Compatibility is due to the fact that future spatial interaction scenarios will encompass desktop interaction, virtual reality, augmented reality, large-screen interaction, and cross-platform interaction amongst these devices. Leveraging its lightweight structure and pen-like ability to input directly into 2D desktop devices, the *Dual-Stick* is well-positioned to meet these demands. In the future, it could be possible to explore cross-device and cross-reality interaction capabilities of the *Dual-Stick*.

## 7 CONCLUSION

This research introduced the *Dual-Stick* tool with a *Dual-Ray* input paradigm, designed for clamp selection and enhanced manipulation DoF. Two studies were conducted to evaluate *Dual-Ray*'s performance in target selection and manipulation tasks, using traditional single-ray as a baseline. The first study showed a 23.5% reduction in selection time and a 31.9% decrease in error rates with *Dual-Ray*. Compared to the heuristic *IntenSelect*, *Dual-Ray* had similar selection times but 9.4% higher error rates. The second study demonstrated that *Dual-Ray*'s added DoF reduced task completion time by 21.6% and mode switching by 65.8%. We also discussed the limitations of *Dual-Ray* and future research opportunities.

## 8 ACKNOWLEDGMENT

We thank all who participated in the user study and those putting valuable suggestions during the review process. This project was funded by the Natural Science Foundation of China (62132010), and Key Project of the Institute of Software Chinese Academy of Sciences (ISCAS-ZD-202401).

## REFERENCES

- [1] K. Arimatsu and H. Mori. Evaluation of machine learning techniques for hand pose estimation on handheld device with proximity sensor. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2020. 1
- [2] D. L. Ashbrook. *Enabling mobile microinteractions*. Georgia Institute of Technology, 2010. 9
- [3] 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*, pp. 1–12, 2019. 2, 5
- [4] A. Banerjee, J. Burstyn, A. Girouard, and R. Vertegaal. Multipoint: Comparing laser and manual pointing as remote input in large display interactions. *International Journal of Human-Computer Studies*, 70(10):690–702, 2012. 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, 2021. 2, 7
- [6] X. Bi, T. Moscovich, G. Ramos, R. Balakrishnan, and K. Hinckley. An exploration of pen rolling for pen-based interaction. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, pp. 191–200, 2008. 2, 9
- [7] P. Brandl, C. Forlines, D. Wigdor, M. Haller, and C. Shen. Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces. In *Proceedings of the working conference on Advanced visual interfaces*, pp. 154–161, 2008. 2
- [8] M. A. Brown and W. Stuerzlinger. Exploring the throughput potential of in-air pointing. In *Human-Computer Interaction. Interaction Platforms and Techniques: 18th International Conference, HCI International 2016, Toronto, ON, Canada, July 17-22, 2016. Proceedings, Part II 18*, pp. 13–24. Springer, 2016. 1, 2
- [9] X. Cao and R. Balakrishnan. Visionwand: interaction techniques for large displays using a passive wand tracked in 3d. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pp. 173–182, 2003. 2
- [10] 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
- [11] H. J. Chae, J.-i. Hwang, and J. Seo. Wall-based space manipulation technique for efficient placement of distant objects in augmented reality. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pp. 45–52, 2018. 1
- [12] M. Choe, Y. Choi, J. Park, and H. K. Kim. Comparison of gaze cursor input methods for virtual reality devices. *International Journal of Human-Computer Interaction*, 35(7):620–629, 2019. 2
- [13] P. Cipresso, I. A. C. Giglioli, M. A. Raya, and G. Riva. The past, present, and future of virtual and augmented reality research: a network and cluster analysis of the literature. *Frontiers in psychology*, p. 2086, 2018. 2
- [14] G. De Haan, M. Koutek, and F. H. Post. Intenselect: Using dynamic object rating for assisting 3d object selection. In *Ipt/egve*, pp. 201–209, 2005. 1, 2, 4, 5
- [15] D. V. Dorozhkin and J. M. Vance. Implementing speech recognition in virtual reality. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 36215, pp. 61–65, 2002. 1, 2
- [16] T. Drey, J. Gugenheimer, J. Karlbauer, M. Milo, and E. Rukzio. Vrs-ketchin: 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. 2
- [17] L. V. I. P. Edition, 2019. Retrieved July 31, 2019 from <https://www.logitech.com/en-us/promo/vr-ink.html>. 1
- [18] A. O. S. Feiner. The flexible pointer: An interaction technique for selection in augmented and virtual reality. In *Proc. UIST*, vol. 3, pp. 81–82, 2003. 2
- [19] N. Fellion, T. Pietrzak, and A. Girouard. Flexstylus: Leveraging bend input for pen interaction. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 375–385, 2017. 1, 2
- [20] J. G. Grandi, H. G. Debarba, L. Nedel, and A. Maciel. Design and evaluation of a handheld-based 3d user interface for collaborative object manipulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, pp. 5881–5891, 2017. 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*, pp. 3–12, 2006. 2, 6
- [22] D. Hayatpur, S. Heo, H. Xia, W. Stuerzlinger, and D. Wigdor. Plane, ray, and point: Enabling precise spatial manipulations with shape constraints. In *Proceedings of the 32nd annual ACM symposium on user interface software and technology*, pp. 1185–1195, 2019. 2, 4, 9
- [23] S. Heo, J. Gu, and G. Lee. Expanding touch input vocabulary by using consecutive distant taps. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 2597–2606, 2014. 5
- [24] K. Hinckley, P. Baudisch, G. Ramos, and F. Guimbretiere. Design and analysis of delimiters for selection-action pen gesture phrases in scriboli. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 451–460, 2005. 2
- [25] K. Hinckley, M. Pahud, H. Benko, P. Irani, F. Guimbretière, M. Gavrilu, X. Chen, F. Matulic, W. Buxton, and A. Wilson. Sensing techniques for tablet+ stylus interaction. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*, pp. 605–614, 2014. 2
- [26] K. Hinckley, R. Pausch, J. C. Goble, and N. F. Kassell. A survey of design issues in spatial input. In *Proceedings of the 7th annual ACM symposium on User interface software and technology*, pp. 213–222, 1994. 2
- [27] Holo-Stylus, 2017. Retrieved July 31, 2023 from <https://www.holo-stylus.com/>. 1
- [28] Y. Jiang, C. Zhang, H. Fu, A. Cannavò, F. Lamberti, H. Y. Lau, and W. Wang. Handpainter-3d sketching in vr with hand-based physical proxy. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2021. 2
- [29] D. Kim, K. Park, and G. Lee. Atatouch: Robust finger pinch detection for a vr controller using rf return loss. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–9, 2021. 1
- [30] J.-H. Kim, N. D. Thang, and T.-S. Kim. 3-d hand motion tracking and gesture recognition using a data glove. In *2009 IEEE International Symposium on Industrial Electronics*, pp. 1013–1018. IEEE, 2009. 2
- [31] 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)*, pp. 67–74. IEEE, 2011. 2, 9
- [32] 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. 2
- [33] J. Li, A. Samoylov, J. Kim, and X. Chen. Roman: Making everyday objects robotically manipulable with 3d-printable add-on mechanisms. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–17, 2022. 2
- [34] 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, 2, 3, 9
- [35] N. Li, H.-J. Kim, L. Shen, F. Tian, T. Han, X.-D. Yang, and T.-J. Nam. Haplinkage: Prototyping haptic proxies for virtual hand tools using linkage mechanism. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 1261–1274, 2020. 2
- [36] 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*, pp. 1–11, 2021. 6
- [37] Y. Li, K. Hinckley, Z. Guan, and J. A. Landay. Experimental analysis of mode switching techniques in pen-based user interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 461–470, 2005. 2
- [38] J. Liang and M. Green. Geometric modeling using six degrees of

- freedom input devices. In *Proc. 3rd International Conference on CAD and Computer Graphics*, pp. 217–222. Citeseer, 1993. 2, 9
- [39] 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. IEEE, 2020. 5, 9
- [40] M. Maquette, 2018. Retrieved Sep 6, 2023 from <https://learn.microsoft.com/en-us/windows/mixed-reality/design/maquette>. 2
- [41] D. Martin, S. Malpica, D. Gutierrez, B. Masia, and A. Serrano. Multimodality in vr: A survey. *ACM Computing Surveys (CSUR)*, 54(10s):1–36, 2022. 9
- [42] P. I. C. Martinez, S. De Piro, C. T. Vi, and S. Subramanian. Agency in mid-air interfaces. In *CHI*, pp. 2426–2439, 2017. 5
- [43] F. Matulic, R. Arakawa, B. Vogel, and D. Vogel. Pensight: Enhanced interaction with a pen-top camera. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–14, 2020. 1
- [44] F. Matulic, A. Ganesan, H. Fujiwara, and D. Vogel. Phonetroller: Visual representations of fingers for precise touch input with mobile phones in vr. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2021. 2
- [45] F. Matulic and D. Vogel. Multiray: multi-finger raycasting for large displays. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2018. 1, 2
- [46] O. Medium, 2018. Retrieved Sep 6, 2023 from [https://www.meta.com/experiences/pcvr/1336762299669605/?utm\\_source=www.google.com&utm\\_medium=oculusredirect](https://www.meta.com/experiences/pcvr/1336762299669605/?utm_source=www.google.com&utm_medium=oculusredirect). 2
- [47] M. R. Mine. Virtual environment interaction techniques. *UNC Chapel Hill CS Dept*, 1995. 1
- [48] A. G. Moore, J. G. Hatch, S. Kuehl, and R. P. McMahan. Vote: A ray-casting study of vote-oriented technique enhancements. *International Journal of Human-Computer Studies*, 120:36–48, 2018. 1
- [49] L. Motion, 2019. Retrieved July 31, 2019 from <https://www.leapmotion.com/>. 2
- [50] F. Mueller, M. Davis, F. Bernard, O. Sotnychenko, M. Verschoor, M. A. Otaduy, D. Casas, and C. Theobalt. Real-time pose and shape reconstruction of two interacting hands with a single depth camera. *ACM Transactions on Graphics (ToG)*, 38(4):1–13, 2019. 2
- [51] N. Nishida, K. Ikematsu, J. Sato, S. Yamanaka, and K. Tsubouchi. Single-tap latency reduction with single-or double-tap prediction. *Proceedings of the ACM on Human-Computer Interaction*, 7(MHCI):1–26, 2023. 5
- [52] D. R. Olsen Jr and T. Nielsen. Laser pointer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 17–22, 2001. 2
- [53] M. Ortega. Hook: Heuristics for selecting 3d moving objects in dense target environments. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 119–122. IEEE, 2013. 4
- [54] A. Pencil, 2019. Retrieved July 31, 2019 from <https://www.apple.com/apple-pencil/>. 3
- [55] D.-M. Pham and W. Stuerzlinger. Is the pen mightier than the controller? a comparison of input devices for selection in virtual and augmented reality. In *Proceedings of the 25th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–11, 2019. 1, 2
- [56] J. Raskin. *The humane interface: new directions for designing interactive systems*. Addison-Wesley Professional, 2000. 2
- [57] H. Ro, S. Chae, I. Kim, J. Byun, Y. Yang, Y. Park, and T. Han. 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. IEEE, 2017. 2
- [58] ShapesXR, 2023. Retrieved Dec 9, 2023 from <https://www.shapesxr.com/>. 7
- [59] A. Sharma, J. S. Roo, and J. Steimle. Grasping microgestures: Eliciting single-hand microgestures for handheld objects. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2019. 9
- [60] R. Shi, N. Zhu, H.-N. Liang, and S. Zhao. Exploring head-based mode-switching in virtual reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 118–127. IEEE, 2021. 2
- [61] J. Smith, I. Wang, W. Wei, J. Woodward, and J. Ruiz. Evaluating the scalability of non-preferred hand mode switching in augmented reality. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pp. 1–9, 2020. 2
- [62] H. Song, H. Benko, F. Guimbretiere, S. Izadi, X. Cao, and K. Hinckley. Grips and gestures on a multi-touch pen. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pp. 1323–1332, 2011. 1, 2
- [63] P. Song, W. B. Goh, W. Hutama, C.-W. Fu, and X. Liu. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1297–1306, 2012. 1
- [64] Sony, 2010. Retrieved July 31, 2023 from <https://www.playstation.com/en-sg/accessories/playstation-move-motion-controller/>. 1
- [65] 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. 3
- [66] A. Steed. Towards a general model for selection in virtual environments. In *3D User Interfaces (3DUI’06)*, pp. 103–110. IEEE, 2006. 4
- [67] H. B. Surale, F. Matulic, and D. Vogel. Experimental analysis of bare-hand mid-air mode-switching techniques in virtual reality. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–14, 2019. 2
- [68] F. Tabassum and M. Mondal. Physical education scientific aspects of lever. 03 2016. 3
- [69] R. J. Teather and W. Stuerzlinger. Pointing at 3d targets in a stereo head-tracked virtual environment. In *2011 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 87–94. IEEE, 2011. 3
- [70] C. tools, 2020. [https://en.wikipedia.org/wiki/Category:Hand\\_tools](https://en.wikipedia.org/wiki/Category:Hand_tools). 2
- [71] M. Veit, A. Capobianco, and D. Bechmann. Influence of degrees of freedom’s manipulation on performances during orientation tasks in virtual reality environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, pp. 51–58, 2009. 2, 4
- [72] P. Wacker, O. Nowak, S. Voelker, and J. Borchers. Arpen: Mid-air object manipulation techniques for a bimanual ar system with pen & smartphone. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–12, 2019. 2
- [73] T.-Y. Wei, H.-R. Tsai, Y.-S. Liao, C. Tsai, Y.-S. Chen, C. Wang, and B.-Y. Chen. Elastilinks: Force feedback between vr controllers with dynamic points of application of force. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*, pp. 1023–1034, 2020. 2
- [74] Wikipedia contributors. Lever — Wikipedia, the free encyclopedia, 2023. [Online; accessed 29-November-2023]. 3
- [75] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 143–146, 2011. 5, 7, 8
- [76] H. P. Wyss, R. Blach, and M. Bues. isith-intersection-based spatial interaction for two hands. In *3D User Interfaces (3DUI’06)*, pp. 59–61. IEEE, 2006. 1, 2
- [77] J. Yang, H. Horii, A. Thayer, and R. Ballagas. Vr grabbers: Ungrounded haptic retargeting for precision grabbing tools. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology*, pp. 889–899, 2018. 2
- [78] D. Yu, X. Lu, R. Shi, H.-N. Liang, T. Dingler, E. Velloso, and J. Goncalves. Gaze-supported 3d object manipulation in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–13, 2021. 1, 2
- [79] F. Zhang, K. Katsuragawa, and E. Lank. Conductor: Intersection-based bimanual pointing in augmented and virtual reality. *Proceedings of the ACM on Human-Computer Interaction*, 6(ISS):103–117, 2022. 1, 2