Align-to-Scale: Mode Switching Technique for Unimanual 3D Object Manipulation with Gaze-Hand-Object Alignment in Extended Reality

MIN-YUNG KIM, Graduate School of Culture Technology, KAIST, Republic of Korea
JINWOOK KIM, Institute of Information Electronics, KAIST, Republic of Korea
KEN PFEUFFER, Department of Computer Science, Aarhus University, Denmark
SANG HO YOON, Graduate School of Culture Technology, KAIST, Republic of Korea

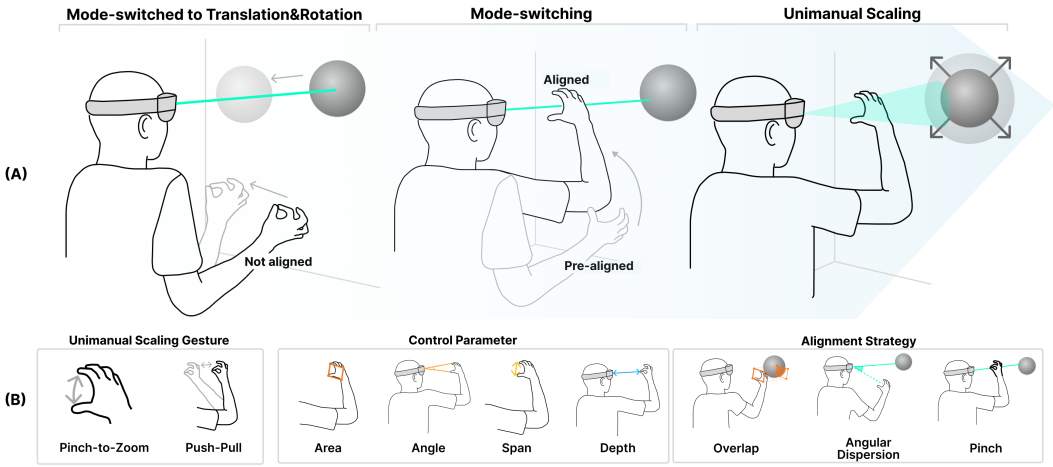

Fig. 1. We propose a unimanual object manipulation by enabling single-handed scaling with gaze-hand-object alignment-based mode-switching. When the hand, gaze, and virtual object do not align, the Gaze+Pinch-based translation mode is activated (A-left). On the other hand, when the hand, gaze, and the object are aligned, in line-of-sight (A-middle), scaling mode is activated (A-right) with unimanual scaling gestures (B-left), by utilizing control parameters to calculate scaling factors (B-middle). Utilizing the proposed gaze-hand-virtual object alignment strategies (B-right) enables us to perform all manipulation features using a single hand.

As extended reality (XR) technologies rapidly become as ubiquitous as today's mobile devices, supporting one-handed interaction becomes essential for XR. However, the prevalent Gaze + Pinch interaction model partially supports unimanual interaction, where users select, move, and rotate objects with one hand, but scaling typically requires both hands. In this work, we leverage the spatial alignment between gaze and hand as a mode switch to enable single-handed pinch-to-scale. We design and evaluate several techniques geared for one-handed scaling and assess their usability in a compound translate-scale task. Our findings show that all proposed methods effectively enable one-handed scaling, but each method offers distinct advantages and trade-offs. To this end, we derive design guidelines to support futuristic 3D interfaces with unimanual interaction. Our work helps make eye-hand 3D interaction in XR more mobile, flexible, and accessible.

Authors' Contact Information: Min-yung Kim, min.kim@kaist.ac.kr, Graduate School of Culture Technology, and KAIST, Daejeon, Republic of Korea; Jinwook Kim, jinwook.kim31@kaist.ac.kr, Institute of Information Electronics, and KAIST, Daejeon, Republic of Korea; Ken Pfeuffer, ken@cs.au.dk, Department of Computer Science, and Aarhus University, Aarhus, Denmark; Sang Ho Yoon, sangho@kaist.ac.kr, Graduate School of Culture Technology, and KAIST, Daejeon, Republic of Korea.

2026. ACM 2577-6193/2026/6-ART24
https://doi.org/10.1145/3803538

CCS Concepts: • **Human-centered computing** → **Interaction techniques**; **Mixed / augmented reality**; **Virtual reality**; **User studies**; **Interaction design theory, concepts and paradigms**.

Additional Key Words and Phrases: Virtual Reality, Mode Switch, Scaling, Gaze, Gestures, Eye-Hand interaction

ACM Reference Format:
Min-yung Kim, Jinwook Kim, Ken Pfeuffer, and Sang Ho Yoon. 2026. Align-to-Scale: Mode Switching Technique for Unimanual 3D Object Manipulation with Gaze-Hand-Object Alignment in Extended Reality. *Proc. ACM Comput. Graph. Interact. Tech.* 9, 2, Article 24 (June 2026), 18 pages. https://doi.org/10.1145/3803538

1 Introduction

As extended reality (XR) and glasses-based form factors like the Meta Ray-Ban Display, Snap Spectacles, and XReal Pro emerge, researchers and practitioners are envisioning their use in everyday contexts [Grubert et al. 2016; Manakhov et al. 2024; Rasch et al. 2025]. To achieve this goal, a key capability is to operate with only a single hand [Karlson et al. 2008]. According to Microsoft's Persona Spectrum, unimanual interactions are essential for permanent or temporary impairments of the upper limbs, as well as for situational impairments such as a parent holding a baby [Microsoft 2016]. One-handed situations are also common in daily life such as when multitasking with a pen, navigating on-the-go with a bag, or simply holding everyday objects like a cup [Karlson et al. 2008; Ng et al. 2013]. Accordingly, for mobile phones, alternative unimanual or single-finger techniques have been explored when established bimanual or multi-finger approaches exist [Boring et al. 2012; Esteves et al. 2022; Holman et al. 2013]. Given the nature of wearable devices [Pascoe et al. 2000], the same necessity of alternative unimanual interaction also applies to XR.

However, unimanual control remains limited in current XR hand interactions. To operate XR user interfaces (UIs), devices increasingly support gaze as a fast pointing mechanism, coupled with pinch gestures for object manipulation. Users select and move objects by looking at them and performing a pinch with the index and thumb. Similarly, object rotation is supported by default with one hand through rotating the hand. In contrast, to scale an object, users pinch with both hands to change the distance between them. This is a standard model for atomic rotate-scale-translate (RST) tasks as proposed in prior scientific research [Chatterjee et al. 2015; Pfeuffer et al. 2017] and used, e.g., in the Apple Vision Pro XR headset. Yet, for basic scaling tasks such as resizing a window or an image, and zooming within a map, there is currently no default option for one-handed control.

Therefore, we explore a mode-switching method, Align-to-Scale, that exploits an eye-hand spatial coordination to enable unimanual manipulation in XR. The idea is to activate the scaling mode when the hand and gaze are aligned (Fig. 1(A-right)), and deactivate when they are separate (Fig. 1(A-left)). This is inspired by the gaze-hand alignment concept for a selection [Lystbæk et al. 2022], which we adopt to a framework where the aligned state serves as a mode-switch cue from translate-rotate to scaling. The main advantage is that, instead of switching between hands, users can perform the whole manipulation with the same hand, enabling complete one-handed manipulation.

The main research question is how users perform unimanual scaling and mode-switching with Align-to-Scale. We designed four unimanual scaling techniques by combining gesture type, gaze-hand alignment strategies, and control parameters. For scaling gestures, we adopted familiar Pinch-to-Zoom (varying the distance between thumb and index) and Push-Pull (moving the hand back and forth) (Fig. 1(B-left)). Secondly, we examine alignment strategies for robust mode-switching (e.g., Overlap, Angular Dispersion, and Pinch) (Fig. 1(B-right)). The conditions are studied in a translate-scale task, where users first move an object, then switch mode, and scale the target.

Our results indicate that, despite performance costs associated with transitioning from bimanual to unimanual, participants were able to understand and execute unimanual manipulation and

alignment-based mode-switching. Each technique had pros and cons regarding robust mode separation, scaling accuracy, and perceived intuitiveness. Here, we emphasize that our comparison with the baseline of bimanual scaling is not about outperforming, but providing qualitative benefits of an alternative unimanual option. Based on the results, we derive design guidelines for selecting the most suitable unimanual scaling gesture, alignment strategy, and control parameter based on user contexts and tasks, and suggest potential directions for improvement. Furthermore, we open-sourced our implementations on GitHub. Our contributions are as follows:

- We present interactions integrating gaze and hand as a mode-switching cue. Extending the familiar scaling gestures, we provide alternative unimanual scaling for one-handed manipulation in XR.
- We present a user study evaluating the techniques in a translation-scaling task, revealing trade-offs between mode-switching and scaling performance, and perceived intuitiveness, alongside the performance cost of transitioning from bimanual to unimanual.
- We present guidelines informing the techniques' suitability for practical use cases.

2 Related Work

2.1 Mode-switching Techniques for XR hand interactions

Bimanual mode-switching is known to be efficient, due to its asymmetric division of labor where the non-dominant hand provides mode-switching cues while the dominant hand does the primary task [Guiard 1987]. Examples are raising the non-dominant hand [Kim et al. 2014; Tan et al. 2013; Vatavu 2013], making a gesture [Hayatpur et al. 2019; Kim et al. 2018], or touching a body part with it [Hayashi et al. 2014; Löcken et al. 2012; Walter et al. 2013]. For unimanual mode-switching, performing distinctive gestures [Freeman et al. 2012; Hayatpur et al. 2019; Kim et al. 2025b; Vogel and Balakrishnan 2005], making a pinch with different fingers [Shi et al. 2024], and moving or rotating the hand [Smith et al. 2019] were suggested. Surale et al. compared VR hand-based mode-switching and reported that bimanual mode-switching was less error-prone. They also identified that making a pinch with different fingers is a promising option, whereas distinctive hand gestures for mode-switching cause user confusion [Surale et al. 2019]. However, their experiment was limited to a single task type, line drawing, lacking practical insights on whether unimanual mode-switching can also support different manipulation modes, including scaling, which we aim to explore.

These hand-based mode-switching also suffer from increased cognitive load as the number of gestures and complexity increase [LaViola Jr et al. 2017]. Thus, there have been approaches to introduce new modalities alongside hand input, including feet [Velloso et al. 2015], voice [Bolt 1980], head [Shi et al. 2021], or using external tools [Kim et al. 2023; Stoakley et al. 1995]. In this work, we focus on gaze, a modality that has become a common input in XR. Recent work has proposed the alignment between gaze and hand in 3D space as a new interaction cue [Lystbæk et al. 2022]. This gaze-hand alignment has been used for UI [Lystbæk et al. 2022; Wagner et al. 2023] or region selection [Shi et al. 2023], or accessing remote objects [Liu et al. 2025]. Building on this, as shown in Tab. 1, we address that our alignment-based mode-switching is the first attempt to integrate gaze and hand for mode-switching to enable unimanual scaling.

2.2 Unimanual Object Manipulation in XR

Unimanual interaction enables multitasking [Boring et al. 2012; Li and Fu 2013], minimizes attentional load [Karlson and Bederson 2007; Pascoe et al. 2000] and lowers barriers of acceptability [Serrano et al. 2014] and accessibility [Yamagami et al. 2022]. However, current scaling in XR is done by varying the distance between two hands [Lee et al. 2024; Pierce et al. 1997; Song et al. 2012] based on a metaphor of 'stretch and squeeze' [Mendes et al. 2019; Van Dam 1997].

In case of unimanual scaling, based on a natural tendency to use symmetrical thumb-index movement for resizing [Brouet et al. 2013], a Pinch-to-Zoom (PTZ) has been utilized in 2D UI [Avery et al. 2014; Hinckley et al. 1998; Käser et al. 2011; Malacria et al. 2010]. However, when PTZ is applied within the Gaze+Pinch framework, gesture misclassifications can occur between the full pinch gesture and the PTZ. To avoid this, Dewitz et al. employed non-dominant hand pinching to activate PTZ, making the interaction bimanual [Dewitz et al. 2021]. Another option is 'Push-Pull' gesture, which involves moving the hand back and forth [Büschel et al. 2019; Nancel et al. 2011; Stellmach et al. 2012; Yoo et al. 2010]. Stellmach et al. implemented it by raising the non-dominant hand [Stellmach et al. 2012], and Yoo et al. required users to make a distinctive gesture [Yoo et al. 2010] when making the Push-Pull gesture. A double-pinch gesture was also used for zooming, and a normal pinch for panning [Büschel et al. 2019], yet the double-pinch is only compatible with Push-Pull and not PTZ, as it requires the thumb and index finger to be in contact.

In Tab. 1, we compared previous mode-switching methods for supporting these unimanual gestures. Hand Position/rotation-based mode-switching was the only method compatible with both. Shi et al. used wrist rotation to switch between selection modes; however, it was physically uncomfortable [Shi et al. 2024]. Smith et al. investigated hand-depth-based switching during translation tasks, which proved to be the fastest but also the most inaccurate [Smith et al. 2019]. Therefore, building on these approaches, we adopt the gaze-hand alignment concept as a mode-switching to enable these intuitive unimanual gestures in XR. As mode-switching mostly refers to an implicit cue rather than an explicit visual UI [Raskin 2000], and also considering its visually disruptive manner [Guimbretiere and Winograd 2000; Kurtenbach and Buxton 1994], we do not consider visual UI-based unimanual manipulations like 3D bounding boxes [Microsoft 2024].

3 Design Space for Unimanual Object Manipulation

The design space for unimanual object manipulation involves three design factors: *gesture type*, *alignment strategy*, and *control parameter* (Fig. 2). The *alignment strategy* is a factor that determines the alignment between gaze, hand, and the object, whereas the *gesture type* and *control parameter* were selected as representative factors highly relevant to scaling interactions [Nancel et al. 2011; Schubert et al. 2023]. By combining the elements, we derived four interactions: PTZ-Area, PTZ-Angle, PTZ-Span, and Push-Pull-Depth. All interaction parameters are chosen via pilot testing and detailed in Tab. 2. All subsequent mentions of the variables refer to those in this table. For reproducibility, we open-sourced the implementations [1].

3.1 Gesture Type

Pinch-to-Zoom (PTZ). PTZ gesture is a familiar unimanual zooming gesture on 2D UI. Increasing the span between the thumb and index finger tips scales the object up, while decreasing it scales the object down (Fig. 2(A-1)). However, it has not been adopted in XR due to an absence of a robust mode-switching method that distinguishes PTZ from the normal pinch gesture for selection or translation in Gaze+Pinch. Thus, we propose utilizing gaze-hand alignment as a mode-switching cue; when aligned, thumb and index movement is mapped to scaling through PTZ, and when not aligned, their movement would be interpreted as the normal pinch.

Push-Pull. Moving the hand back-and-forth corresponds to pushing the object away to make the size smaller and pulling it closer to the user's body to make it bigger (Fig. 2(A-2)). Unlike PTZ, this gesture is compatible with unimanual hand-based mode-switching (Distinctive Gesture, Different Finger Pinch, or Double-Pinch in Tab. 1). Thus, we compare the two gestures to explore the trade-off between the intuitiveness of the gesture and the robustness of mode-switching.

[1] https://github.com/MinKim242/ETRA_Align-to-Scale.git

Table 1. Comparison of Align-to-Scale with previous methods. The methods are analyzed based on their support for unimanual interactions, integration of different modalities, and compatibility with scaling gestures. A ✓ indicates that the mode-switching supports the dimension, while a - indicates a lack of support.

Mode-switching	Unimanual	Integration of Gaze & Hand	Gesture Compatibility		
			PTZ	Push-Pull	Bimanual
Non-Dominant Hand [Dewitz et al. 2021; Stellmach et al. 2012]	-	-	✓	✓	-
Distinctive Gesture [Hayatpur et al. 2019; Yoo et al. 2010]	✓	-	-	✓	-
Different Finger Pinch [Shi et al. 2024; Surale et al. 2019]	✓	-	-	✓	-
Hand Position/Rotation [Shi et al. 2024; Smith et al. 2019]	✓	-	✓	✓	-
Bimanual Scaling [Lee et al. 2024; Song et al. 2012]	-	-	-	-	✓
Double-Pinch [Büschel et al. 2019]	✓	-	-	✓	-
Align-to-Scale Mode-switching (Ours)	✓	✓	✓	✓	-

3.2 Alignment Strategy

Alignment Strategy denotes the type of cue to determine whether the hand is in line-of-sight with gaze and the object. We derived three variations to distinguish between normal and scaling modes.

Overlap. We used an overlap-based gaze-hand alignment (Fig. 2(B-1)) with the PTZ gesture. Mode-in to scaling is triggered when the stereoscopic view area overlaps with the object, and while being gazed at. The stereoscopic view area is defined as the rectangular region formed by projecting the thumb and index fingertips from the left and right eyes. We calculate how much of the view area the object covers, and vice versa. If either ratio exceeds its overlap threshold, the system assumes the object and the view area are overlapping. The stereoscopic view area is illustrated as a purple rectangular area, while the overlap is denoted by a purple fill inside the view area (Fig. 2(B-1)).

The overlap-based alignment is driven by two motivations. First, we expect the method to reproduce the sensation of direct scaling in XR, as when performing PTZ on 2D screens [Hinckley et al. 1998; Pfeuffer et al. 2016]. Second, we aim to improve the conventional Head Crusher technique, which utilizes a similar concept of framing the target within the space between fingers [Pierce et al. 1997]. Prior experiments indicated the conventional logic of ray-casting from eye through the thumb-index midpoint resulted in a low selection performance [Wagner et al. 2023].

Angular Dispersion. It is a conventional cue of gaze-hand alignment [Lystbæk et al. 2022]. The angular dispersion is measured by the angle between a gaze ray and a vector from the gaze origin to the hand. The scaling mode is activated when the angular dispersion between gaze and the hand

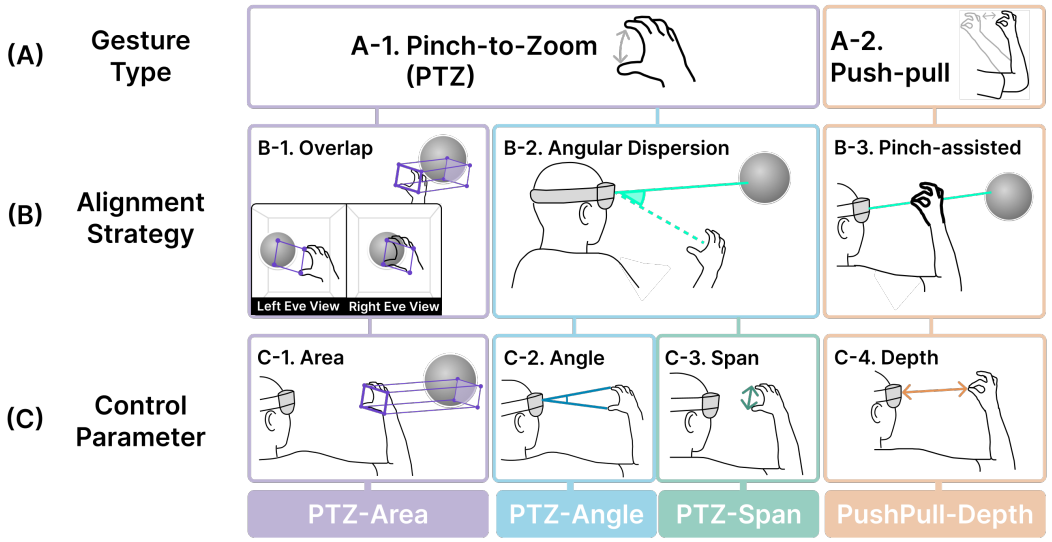

Fig. 2. Design space of unimanual scaling interactions. We derived four interactions: PTZ-Area, -Angle, -Span, and Push-Pull-Depth, by combining design factors of scaling gesture type, alignment strategy, and control parameter. We chose two unimanual scaling gestures: Pinch-to-Zoom (PTZ) (A-1) and Push-Pull (A-2). Three gaze-hand alignment strategies for mode-switch: stereoscopic view area overlap (purple rectangular area, B-1), angular dispersion between gaze–object and gaze–hand (cyan angle, B-2), and pinch gestures performed when the gaze, hand, and object were aligned(B-3) were tested. (C) Each input depicted in purple, blue, green, and orange represents control parameters used to compute scaling factors.

falls below a mode-in threshold (cyan angle in Fig. 2(B-2)), and the mode-out threshold was set bigger than the mode-in threshold to prevent unintentional mode-outs.

Pinch-assisted. We adopted pinch-assisted gaze-hand alignment (Fig. 2(B-3)) as a representative from hand gesture-based mode-switching [Surale et al. 2019]. When users make a pinch with their hands raised near the gaze, mode-in to scaling occurs. Moving the hand for-/backward will then scale. When they make a pinch gesture without the alignment, translation mode is triggered. The same thresholds for mode-in and out were used with the angular dispersion.

3.3 Control Parameter

The scaling factor is a multiplier that determines how much an object's size changes relative to its original size. The scaling factor is calculated as a ratio between the new and the initial object size. We calculated s_t, the current scale of the object at time t, with the following equation of $s_t = s_0 \times \frac{I_t}{I_0}$. s_0 is an object scale at the start of the scaling mode, I_t is a current input from the scaling gesture at time t, and I_0 is an initial input at the start of the scaling mode. Here, the scaling factor is $\frac{I_t}{I_0}$. By the term control parameter, we refer to the input I used for the computation of the scaling factor.

By proposing **stereoscopic view area** (Fig. 2(C-1)) and **angle** as control parameters, we explore a new style of scaling where scaling is influenced by both a span between the thumb and the index tips, and the depth of the hand. The same stereoscopic view area used for overlap-based alignment is reused as a control parameter. The angle for the scaling factor refers to the angle between two vectors from the gaze origin to the tips of the thumb and index finger, respectively. Both the overlap area and the angle increase as the finger span widens or the hand moves closer in depth, resulting in a scale-up, and a scale-down for the opposite direction. We also included

Span-only as a familiar control parameter used in 2D PTZ (Fig. 2(C-3)). For the Push-Pull scaling, we implement the **depth** of the hand as a control parameter (Fig. 2(C-4)). The depth of the hand is defined as the perpendicular distance from the head to the hand. To prevent noises in input and ensure a comfortable range of movements [de la Fuente and Bix 2010], the control parameters were clamped between minimum and maximum values of each input.

Table 2. Interaction parameters used in the study. All the mentioned thresholds in Sec. 3 are listed here. Note that the unit of the minimum and maximum clamping for PTZ-Area is a proportion of the stereoscopic view area on the HMD display between 0 to 100%.

Design Factor	Technique	Parameter	Value
Alignment Strategy: Overlap	PTZ-Area	Overlap Threshold	Min. 25% of Stereoscopic View Area OR Min. 15% of Object is covered by each other
Alignment Strategy: Angular dispersion, Pinch-assisted	PTZ-Angle, -Span, Push-Pull-Depth	Mode-in Threshold	15°
		Mode-out Threshold	17°
Control Parameter	PTZ-Area	Min. & Max. of Clamping	0.1%, 100%
	PTZ-Angle		3°, 40°
	PTZ-Span		0.01m, 0.15m
	Push-Pull-Depth		0.1m, 0.5m
	Bimanual		0.01m, 0.8m

4 User Study

4.1 Task Design

We conducted a within-subjects design to evaluate the interactions. Each trial required participants to first translate a virtual object to a specified position, mode-switch to scaling mode, and then scale the object to match a target size. Our analysis focused on mode-switching events: mode-in to translation, mode-in to scaling, and mode-out from scaling mode to idle state. The mode-out for translation was excluded as it occurs with the instant release of the pinch. The rotation task was excluded because it occurs within the same mode as translation for hand interactions [Mendes et al. 2019], requiring no explicit mode-switch. Standard frameworks like Virtual Hand [LaViola Jr et al. 2017], and Gaze+Pinch [Pfeuffer et al. 2017] also combine both tasks into a single mode.

There were three independent variables: 5 techniques (including the baseline), 4 target positions (up, down, left, right), and 4 target scales (×0.4, ×0.67, ×1.5, ×2.5). We position bimanual scaling as a relevant baseline for both scaling and mode-switching, as use of the non-dominant hand was also frequently suggested as a cue for mode-switch [Ruiz et al. 2008; Surale et al. 2019]. A single session consisted of 16 combinations of target scales and positions, and participants repeated three sessions, with the first session as a practice. Thus, the total number of trials for analysis was 160 (5 techniques × 4 target scales × 4 target positions × 2 sessions). The order of the techniques was randomized among participants to prevent learning effects.

4.2 Implementation

The study environment was developed in Unity 2022.3.39f1 with Varjo XR-3 HMD (90Hz, 115° FoV) embedded with an eye tracker (200 Hz). We used Ultraleap SDK's hand-tracking data, which was smoothed with a 1€ filter [Casiez et al. 2012], with parameters empirically set at $f_{c_{min}} = 1.0$ and $\beta = 90$ [Wagner et al. 2024]. Following the previous works [Kim et al. 2025a], [Pfeuffer et al. 2017], we also adopted the change in the object outline color as an indicator of current mode: orange in translation mode, yellow in scaling mode, and white when gazed at.

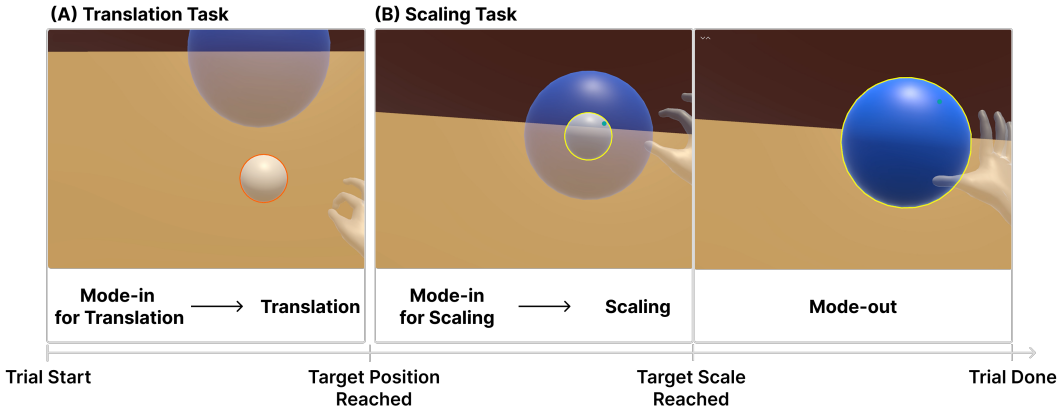

Fig. 3. User study overview. Participants manipulated a white sphere to match the position and scale of a semi-transparent blue sphere. Participants first translated the white sphere toward the blue sphere. The outline of the white sphere turned orange when switched to translation mode (A). Then, participants mode-in to scaling mode. When they switched to scaling mode, the outline color of the white sphere turned yellow (B-left). When the white sphere's scale reached the target scale, it turned opaque blue, indicating participants to switch out of the scaling mode (B-right). The trial ended when the participant mode-out of scaling.

4.3 Procedure

We recruited 20 participants ($M = 24.85, SD = 3.08, 7Female$), and none reported visual or motor impairments (i.e., color deficiencies) that would affect performing the tasks. The study was approved by the Institutional Review Board (IRB), and participants received $15 for their participation. After signing a consent form, we introduced the interactions and the tasks for 15 minutes. For each technique, participants completed one practice session followed by two main sessions, all of them with the same number of trials. Each main session was followed by subjective ratings and a 5-minute rest. After all the sessions, participants ranked the five techniques and took a short interview.

In the main sessions, participants manipulated the white sphere to match the blue sphere, indicating a target position and scale (Fig. 3). Both spheres appeared at 2 m depth. The white sphere was centered in front, while the blue spheres were offset by 35° in one of four target directions. The white sphere had a diameter of 14°, and the blue sphere was scaled to one of the four target scales.

Participants first started with the translation task (Fig. 3(A)). They translated the white sphere to the position of the blue sphere using Gaze+Pinch [Pfeuffer et al. 2017]. The sphere snapped to the position when its center was within 0.15 m of the target position to indicate task completion and start scaling the task. The snapping feature was utilized to ensure center-to-center alignment between the white sphere, which participants are controlling, and the blue sphere indicating the target scale. Then, participants started scaling the white sphere using different techniques assigned for each session (Fig. 3(B-left)). The color of the white sphere changed to blue at the moment the sphere reached the target scale (Fig. 3(B-right)). This informs participants to mode-out of the scaling mode and move on to the next trial. We considered the target scale to be achieved once the diameter difference between the white and blue spheres was under 0.1. In the case of the scaling task, participants had to stop the scaling on their own while trying to keep the scale of the sphere as close to the target scale. To measure mode-switching time, we controlled the scaling task such that participants could only perform a single scaling attempt without clutching. Trials in which participants failed to reach the target scale on the first try were recorded as scaling error trials.

4.4 Evaluation Metrics

We have two error categories: mode-switching and scaling errors. Mode-switching errors are divided into mode-in errors for translation and scaling tasks. A mode-in error is a transition to a mode that does not align with the current task, including gesture misclassifications between the pinch gesture for translation and PTZ. Specifically, during translation, errors are logged when the system switches to scaling instead of translation, and vice versa for scaling tasks. Scaling errors occur when participants fail to reach the target scale. Trials are marked as 1 if an error occurred and 0 otherwise. The error rate is computed by averaging across trials for each technique [Surale et al. 2017, 2019]. We also calculated the overall mode-switching error rate by logging the error trials regardless of the mode-in error types. We converted the error rate to a percentile, and a closer to 100 % indicates a greater likelihood of committing that error type.

We also measure mode-in time, defined as the duration to transition from the idle state to the mode corresponding to the current task. For scaling performance, mode-out time, which is the duration from scaling mode back to idle, was analyzed. Furthermore, we measure the scaling difference, the deviation between the final object scale after mode-out and the target scale. Subjective evaluations were conducted using NASA Task Load Index (NASA-TLX) [Hart 2006], our own usability questionnaires, and preference rankings on the techniques. We used a raw NASA-TLX on a 7-point scale as in previous XR gaze experiments [Wagner et al. 2023; Yu et al. 2021].

The overall mode-switching error rate from mode-in errors to translation and scaling, mode-in error rates and time for each task were used to assess the robustness of mode-switching. Scaling error rate, scale difference, and mode-out time for scaling were used to evaluate the scaling performance.

5 Result

While all trials were included in the error rate analysis, analyses of completion time and scale difference excluded trials that failed to reach the target scale. We removed 50 trials (1.8%) identified as outliers due to tracking loss, exceeding 3 SDs from the mean task completion time for each condition. The Shapiro-Wilk test indicated non-normal distributions; thus, we conducted a nonparametric two-way repeated-measures ANOVA using the Aligned Rank Transform (ART) [Wobbrock et al. 2011], with Technique and Target Scale as within-subject factors. When significant effects were found, post-hoc pairwise comparisons were performed using the ART-C procedure[Elkin et al. 2021] with Holm correction. For brevity, we only report the significant pairs within the same Technique when interaction effects are significant, and both Technique and the Technique × Target Scale interaction in our figures to provide an overview. Detailed results are in the supplementary material.

5.1 Mode-switching Performance

5.1.1 Overall Mode-Switching Error Rate. The Overall Error Rate was significantly influenced only by the Technique factor ($F^4_{361} = 23.0,\ p < .001,\ \eta_p^2 = .20$) (Fig. 4(A)). The post-hoc test revealed that Bimanual scaling ($M = 20.5,\ SD = 17.5$) resulted in a lower Overall Mode-switching Error Rate than PTZ-Angle ($M = 42.7,\ SD = 26.5;\ t(361) = -7.58,\ p < .001,\ d = 0.80$), PTZ-Span ($M = 39.8,\ SD = 25.1;\ t(361) = -7.03,\ p < .001,\ d = 0.74$), and Push-Pull scaling ($M = 34.2,\ SD = 22.8;\ t(361) = -5.26,\ p < .001,\ d = 0.55$). PTZ-Area ($M = 24.2,\ SD = 21.5$) also led to a lower Overall Mode-switching Error Rate than PTZ-Angle ($t(361) = -6.18,\ p < .001,\ d = 0.65$), PTZ-Span ($t(361) = -5.63,\ p < .001,\ d = 0.59$), and Push-Pull scaling ($t(361) = -3.86,\ p < .001,\ d = 0.41$).

5.1.2 Mode-in Error Rate for Translation. The Technique was the only significant factor ($F^4_{361} = 38.5,\ p < .001,\ \eta_p^2 = .30$) (Fig. 4(B)). Post-hoc comparisons revealed that PTZ-Angle ($M = 28.4$, $SD = 27.2$) and PTZ-Span ($M = 29.1$, $SD = 25.6$) yielded significantly higher error rates than all

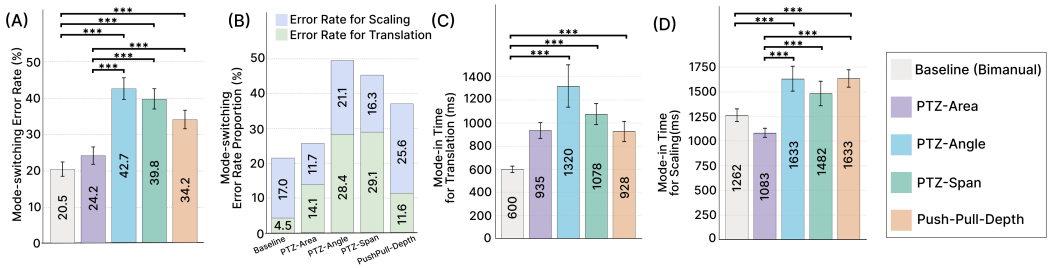

Fig. 4. Results on mode-switching performance. (A) Overall mode-switching error rates by Technique. Overall mode-switching error includes both the mode-in for translation and the scaling mode. Error rates closer to 100% indicate more frequent incorrect mode activations. (B) Proportions of mode-switching error rates by Technique. Multiple error types can occur from a single trial. (C) The mean mode-switching time it took for mode-in to translation. (D) The mean mode-switching time during the mode-in for scaling mode. Significance levels are indicated as $^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$; error bars represent standard error.

other techniques. PTZ-Angle had significantly higher error rates than the baseline ($M = 4.5,\ SD = 8.0;\ t(361) = 9.60,\ p < .001,\ d = 1.01$), Push-Pull-Depth ($M = 1.6,\ SD = 16.7;\ t(361) = -6.04,\ p < .001,\ d = 0.64$), and PTZ-Area ($M = 14.1,\ SD = 18.2;\ t(361) = 5.52,\ p < .001,\ d = 0.58$). PTZ-Span led to significantly higher error rates than the baseline ($t(361) = 9.60,\ p < .001,\ d = 1.01$), Push-Pull-Depth ($t(361) = 6.86,\ p < .001,\ d = 0.72$), and PTZ-Area ($t(361) = 6.33,\ p < .001,\ d = 0.67$). Additionally, the baseline resulted in significantly lower error rates than PTZ-Area ($t(361) = -4.08,\ p = .001,\ d = 0.43$) and Push-Pull-Depth ($t(361) = -3.56,\ p < .001,\ d = 0.37$).

5.1.3 Mode-in Error Rate for Scaling. There was significant main effect of Technique ($F^4_{361} = 7.41,\ p < .001,\ \eta^2_p = .076$) and Target Scales ($F^3_{361} = 4.34,\ p = .005$). The interaction effect was also significant ($F^{12}_{361} = 3.40,\ p < .001,\ \eta^2_p = .10$) (Fig. 4(B)). The significant pairs of interaction effects were observed for PTZ-Span. Specifically, the Target Scale of ×2.5 yielded significantly higher values than ×0.4 ($t(361) = -4.96,\ p < .001,\ d = 0.52$) and ×0.67 ($t(361) = -4.21,\ p = .006,\ d = 0.44$).

5.1.4 Mode-in Time for Translation. We found a significant effect of Technique on the metric (Fig. 4(C)) ($F^4_{361} = 17.1,\ p < .001,\ \eta^2_p = .16$). From the post-hoc test, the Bimanual method ($M = 600,\ SD = 251$) resulted in significantly faster mode-in time compared to other techniques; Push-Pull-Depth ($M = 928,\ SD = 773,\ t(361) = -5.27,\ p < .001,\ d = 0.55$), PTZ-Angle ($M = 1320,\ SD = 1636,\ t(361) = -7.31,\ p < .001,\ d = 0.77$), PTZ-Span ($M = 1078,\ SD = 800,\ t(361) = -6.92,\ p < .001,\ d = 0.73$), and PTZ-Area ($M = 935,\ SD = 624,\ t(361) = -5.34,\ p < .001,\ d = 0.56$).

5.1.5 Mode-in Time for Scaling. Main effects were significant, on Technique ($F^4_{361} = 13.0,\ p < .001,\ \eta^2_p = .13$) and Target Scales ($F^3_{361} = 5.22,\ p = .002,\ \eta^2_p = .042$). The interaction effect was also significant ($F^{12}_{361} = 2.33,\ p = .007,\ \eta^2_p = .072$) (Fig. 5(D)), however, there were no significant pairs.

5.2 Scaling Performance

5.2.1 Scaling Error Rate. There was a significant effect of Technique ($F^4_{361} = 32.8,\ p < .001,\ \eta^2_p = .27$) and Target Scale ($F^3_{361} == 60.4,\ p < .001,\ \eta^2_p = .33$), and interaction effect ($F^{12}_{361} == 6.72,\ p < .001,\ \eta^2_p = .18$) (Fig. 5(A, B)).

Except for the baseline, all the Techniques included pairs of the Target Scale with significant differences (Fig. 5(B)). Within the PTZ-Area, ×0.4 led to a higher Scaling Error Rate than ×0.67 ($t(361) = 5.41,\ p < .001,\ d = 0.57$) and ×1.5 ($t(361) = 3.86,\ p = .016,\ d = 0.41$). ×2.5 resulted in

a higher error rate than ×0.67 ($t(361) = 5.65,\ p < .001,\ d = 0.60$) and ×1.5 ($t(361) = 4.10,\ p = .006,\ d = 0.43$). For the PTZ-Angle, ×2.5 induced higher error rate than ×0.4 ($t(361) = 5.19,\ p < .001,\ d = 0.55$), ×0.67 ($t(361) = 6.41,\ p < .001,\ d = 0.67$), and ×1.5 ($t(361) = 5.12,\ p < .001,\ d = 0.54$). The PTZ-Span included the same significant pairs: ×2.5 with a higher error rate than ×0.4 ($t(361) = -5.31,\ p < .001,\ d = 0.56$), ×0.67 ($t(361) = 5.31,\ p < .001,\ d = 0.56$), and ×1.5 ($t(361) = 3.56,\ p = .049,\ d = 0.38$). Lastly, ×2.5 of Push-Pull-Depth scaling also resulted in a higher error rate than ×0.4 ($t(361) = 5.84,\ p < .001,\ d = 0.61$), ×0.67 ($t(361) = 6.90,\ p < .001,\ d = 0.73$), and ×1.5 ($t(361) = 3.76,\ p = .024,\ d = 0.40$).

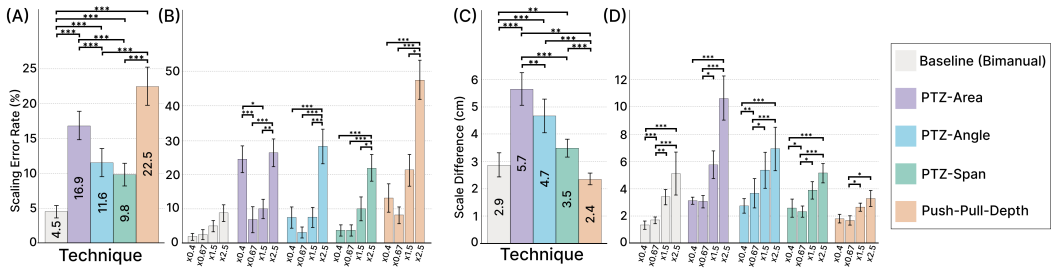

Fig. 5. Results on scaling performance. (A) Scaling error rate by technique. (B) Scaling error rate by Technique × Target Scale. (C) The mean scale difference shows the deviation between the final object and the target scale after mode-out. (D) The mean scale difference by Technique×Target Scale. Significance levels are indicated as *p < .05, **p < .01, ***p < .001; error bars show standard error.

5.2.2 Scale Difference. There were significant effects of Technique ($F^4_{361} = 35.5,\ p < .001,\ \eta^2_p = .28$) and Target Scale ($F^3_{361} = 75.1,\ p < .001,\ \eta^2_p = .38$), and interaction effect ($F^{12}_{361} = 4.30,\ p < .001,\ \eta^2_p = .13$) (Fig. 5(C, D)).

All Techniques showed significant differences between Target Scales (Fig. 5(D)). Within the PTZ-Area, ×2.5 led to a bigger scale difference than that of ×0.4 ($t(361) = 5.35,\ p < .001,\ d = 0.56$) and ×0.67 ($t(361) = 6.62,\ p < .001,\ d = 0.70$). ×1.5 showed bigger difference than the target scale of ×0.67 ($t(361) = 3.68,\ p = .030,\ d = 0.39$). For the PTZ-Angle, ×2.5 showed a bigger scale difference than ×0.4 ($t(361) = 6.38,\ p < .001,\ d = 0.67$) and ×0.67 ($t(361) = 5.72,\ p < .001,\ d = 0.60$). Scale Difference of ×1.5 was significantly bigger than ×0.4 ($t(361) = 4.43,\ p = .002,\ d = 0.47$) and ×0.67 ($t(361) = 3.77,\ p = .022,\ d = 0.40$). ×2.5 of PTZ-Span resulted in bigger Scale difference than ×0.4 ($t(361) = 5.46,\ p < .001,\ d = 0.57$) and ×0.67 ($t(361) = 5.43,\ p < .001,\ d = 0.57$). ×1.5 of PTZ-Span also resulted in a bigger difference than ×0.4 ($t(361) = 3.71,\ p = .027,\ d = 0.39$) and ×0.67 ($t(361) = 3.68,\ p = .030,\ d = 0.39$). The Push-Pull scaling exhibited bigger scale difference at ×2.5 than ×0.67 ($t(361) = 3.95,\ p = .011,\ d = 0.42$). ×1.5 had higher difference than ×0.67 ($t(361) = 3.67,\ p = .031,\ d = 0.39$). The baseline resulted in bigger scale difference at ×2.5 than ×0.4 ($t(361) = 6.48,\ p < .001,\ d = 0.68$) and ×0.67 ($t(361) = 4.85,\ p < .001,\ d = 0.51$), and ×1.5 higher than ×0.4 ($t(361) = 5.75,\ p < .001,\ d = 0.60$) and ×0.67 ($t(361) = 4.12,\ p = .006,\ d = 0.51$).

5.2.3 Mode-out Time for Scaling. The influences of Technique ($F^4_{361} = 20.3,\ p < .001,\ \eta^2_p = .18$) and Target Scale ($F^3_{361} = 5.83,\ p < .001,\ \eta^2_p = 0.046$) were statistically meaningful. Their interaction effect was also significant ($F^{12}_{361} = 5.50,\ p < .001,\ \eta^2_p = .15$) (Fig. 6(A)) There were significant pairwise differences across target scales within PTZ-Area. Specifically, the smaller target scales showed significantly faster mode-out times: ×0.4 was faster than both ×1.5 ($t(361) = -7.23,\ p < .001,\ d = 0.76$) and ×2.5 ($t(361) = -6.06,\ p < .001,\ d = 0.64$), while ×0.67 was also faster than both ×1.5 ($t(361) = -5.23,\ p < .001,\ d = 0.55$) and ×2.5 ($t(361) = -4.06,\ p = .009,\ d = 0.43$).

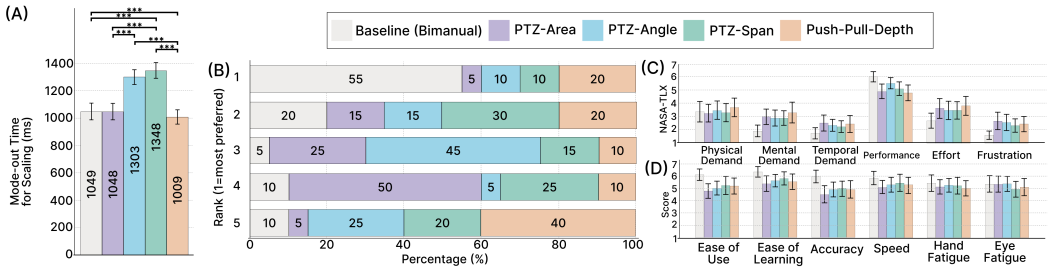

Fig. 6. (A) The mean of mode-out time for scaling, the duration from scaling mode back to the idle state, by Technique. (B) Results on the ranking of user preferences. (C) Results of raw NASA-TLX score measured in 7-point scale by Techniques. (D) Subjective ratings of Techniques. Significance levels are indicated as *p < .05, **p < .01, ***p < .001; error bars represent standard error.

5.3 Subjective Evaluation

From the preference ranking (Fig. 6(B)), the bimanual method was the most preferred technique. Among the unimanual techniques, Push-Pull-Depth scaling was the most preferred. All results showed no significant effects on NASA-TLX factors and subjective questionnaires (Fig. 6(C, D)).

5.4 User Feedback

5.4.1 Intuitiveness of Scaling Gestures. The PTZ scaling gesture was the most mentioned gesture perceived as intuitive ($N = 9$). P4 reported being already familiar with the it. On the other hand, depth-based Push-Pull scaling was perceived as less intuitive, requiring participants to anticipate the direction of hand movement before scaling ($N = 8$).

5.4.2 Perceived Range of Scaling. Participants noted that bimanual scaling enabled the widest scaling range ($N = 5$). Regarding PTZ gestures, incorporating both span and depth (PTZ-Area and PTZ-Angle) was perceived to afford a broader scaling range ($N = 2$) than PTZ-Span, which was criticized for its limited range ($N = 3$). P5 described that coarse scaling was performed using span and refinement with depth for the PTZ-Angle. In the Push-Pull scaling, participants reported hand tracking loss when their hand was too close to the HMD, causing scaling errors ($N = 8$).

5.4.3 Relationship between Scaling Accuracy and Mode-out Timing. Participants noted that the bimanual allowed accurate mode-out at the intended timing, resulting in precise scaling ($N = 5$). Specifically, P19 preferred techniques that could accurately reach the target scale through precise mode-out timing. In contrast, participants commented that it was difficult to mode-out at the intended timing with PTZ gestures ($N = 9$).

5.4.4 Physical Fatigue. While the bimanual baseline was most preferred, it was most cited as physically fatiguing ($N = 5$). P4 explained that the method may not be practical for daily life, while P10 mentioned that as the target scale gets bigger, their hand movement also gets bigger.

6 Discussion

6.1 Comparison with Bimanual Scaling: Performance Costs of Transitioning from Bimanual to Unimanual

While the results indicate that unimanual scaling did not outperform the baseline of bimanual scaling, we emphasize that the purpose of the research is not to beat the bimanual, but to provide

a necessary unimanual alternative. The overlap and angular dispersion-based alignment especially suffered from low scaling precision, because unintended scale change occurred when users moved their hand away from the gaze during mode-out. Instead, the pinch-assisted mode-switch of bimanual and Push-Pull scaling enabled an instant mode-out, resulting in accurate scaling. Furthermore, this performance cost of transitioning from bimanual to unimanual techniques was also observed in previous studies [Guiard 1987; Stellmach et al. 2012; Surale et al. 2019]. In this sense, our contribution lies in accommodating users with limited hand availability by demonstrating that Align-to-Scale is learnable and enables intuitive scaling, while simultaneously cautioning designers to account for this inherent performance gap when developing unimanual alternatives.

6.2 Comparisons among Unimanual Techniques

6.2.1 PTZ scaling. Among the unimanual techniques, PTZ-Area showed the most robust mode-in performance, followed by PTZ-Angle and -Span. Compared to the conventional angular dispersion cue, the overlap-based alignment may capture users' intention to interact with a target more quickly and accurately. Similar to area cursors, where selection occurs when the interaction area overlaps with the target, it allows faster and easier selection of small or grouped objects [Choi et al. 2020; Kabbash and Buxton 1995]. These benefits, combined with contextual cues from the stereoscopic view area [Lee et al. 2023; Wang and Cooper 2022], may explain its fast and accurate performance.

However, using the stereoscopic view area as a control parameter was shown to be less robust, due to its poor scaling performance. We attribute this to a fixed threshold of the overlap ratio. When objects became smaller during scale-down, the overlap ratio tends to fall below the threshold more easily before reaching the target scale. In contrast, during ×2.5 scale-up tasks, as the object got bigger, users had to move their hands farther to reduce overlap, causing the scale value to fluctuate during the mode-out. We believe this is why span-based scaling resulted in the lowest scale difference among the control parameters of PTZ, as it was influenced only by span, not depth. Considering the feedback, where participants still enjoyed scaling with both span and depth, this conflicting tendency highlights the potential of combining span and depth as a control parameter.

6.2.2 Push-Pull Scaling. The most preferred unimanual scaling was Push-Pull-Depth-based scaling, which used the pinch-assisted alignment strategy. This was due to its high scaling performance, enabled by precise temporal control over releasing the pinch, afforded by tactile feedback [Huang et al. 2016; Waugh et al. 2022]. However, frequent hand-tracking loss near the HMD caused abrupt mode-out during scale-up (×2.5), increasing scaling errors. This highlights a limitation of depth as a control parameter, where the system's detectable range and the user's perceived movement range do not match. Furthermore, participants also mentioned that the Push-pull gesture was not as intuitive as PTZ. Despite the limitations, Push-Pull scaling was still the most preferred unimanual technique, implying that guaranteeing accurate scaling capability is crucial.

6.3 Design Guidelines for Align-to-Scale

6.3.1 Unimanual Techniques for Physically Constrained Conditions. Our unimanual techniques offer an alternative to bimanual scaling when users are physically constrained to one-handed interaction. This encompasses not only people with permanent or temporary impairments, but also situational impairments. Carrying bags, boxes, cups or infants with one hand, restricts hand availability, necessitating unimanual interaction [Wobbrock 2019] and driving preference for such techniques [Ng et al. 2013]. Unimanual scaling also enables multitasking to enhance productivity, such as taking notes with one hand while scaling an object with the other [Li and Fu 2013].

In these scenarios, we recommend choosing a technique driven by task requirements: PTZ scaling when rapid task completion is prioritized over scaling precision, and pinch-assisted Push-Pull

gesture for tasks demanding high precision. For instance, when users need to quickly zoom into a map while on the move, PTZ-Area can be used due to the low cognitive load of its intuitive gesture. Similarly, for exploratory tasks like taking a look at distant or minute objects that do not require pixel-wise precision, PTZ scaling can also be used. Instead, for precision-critical scaling such as 3D modeling or technical drawing [Jiang et al. 2021; Lee et al. 2024], we recommend the Push-Pull scaling, as it was the only technique to achieve precision comparable to the bimanual baseline.

6.3.2 Integrating PTZ with Semi-Pinch Quasi-Modes. To expand quasi-mode interactions specified with the 'semi-pinch' gesture [Lee et al. 2025; Zhu et al. 2023], a pre-pinch state similar to PTZ, can be integrated with Align-to-Scale. For example, in a multi-selection task, users can expand the selection area when the hand aligns with the gaze, and select multiple objects at once when not, effectively adding a scaling dimension to the existing quasi-mode [Kim et al. 2025a].

6.4 Limitations & Future Directions

First, our controlled experiment prohibited clutching to isolate mode-switching performance. Because repeated clutching is a common practice when resizing [Avery et al. 2014], this constraint may have contributed to higher scaling errors. Future work should investigate clutching-enabled performance to offer more practical insights. Secondly, while we employed visual feedback to indicate the current mode, the feedback method was not tested, as it is beyond the scope of our research. To enhance accessibility, other modalities such as auditory or tactile [Jang et al. 2024] can be considered. Moreover, our design space does not encompass all potential unimanual components. While we focused on gaze-hand alignment as a mode-switch cue, future research could further extend by combining and comparing our Align-to-Scale with other components in Tab.1. Lastly, our method may induce arm fatigue, leading to the gorilla arm effect [Jang et al. 2017]. While we acknowledge that this contrasts with the current Gaze+Pinch design trend in commercial HMDs, which is to reduce arm fatigue, we consider this as a compensation to enable unimanual interactions.

To improve the scaling accuracy of PTZ, combining different alignment strategies and control parameters could be an option. Future designs might, for instance, pair overlap-based alignment with a span parameter with a PTZ gesture. To make users aware of when the mode-switching occurs, visualizing the stereoscopic view area can also help.

7 Conclusion

We explore how gaze and hand can be used together to enable unimanual object manipulation, with a focus on the one-handed mode-switching and scaling. We adopted the unimanual scaling gestures: PTZ and Push-Pull. As these gestures are not fully compatible with the solely hand-based mode-switching to separate scaling and other object manipulation actions, we propose gaze-hand-object alignment-based mode-switching. By aligning the hand, gaze, and the object, users can start scaling, and finish by separating them. The techniques were evaluated to prevent user confusion across different modes while maintaining scaling accuracy. We explain the strengths and weaknesses of each technique and provide design guidelines for unimanual object scaling interactions.

Acknowledgments

This work was supported by the IITP(Institute of Information & Communications Technology Planning & Evaluation)-ITRC(Information Technology Research Center) grant funded by the Korea government(Ministry of Science and ICT)(IITP-2026-RS-2024-00436398). This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2026-25470200).

References

Jeff Avery, Mark Choi, Daniel Vogel, and Edward Lank. 2014. Pinch-to-zoom-plus: an enhanced pinch-to-zoom that reduces clutching and panning. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 595–604.

Richard A Bolt. 1980. "Put-that-there" Voice and gesture at the graphics interface. In *Proceedings of the 7th annual conference on Computer graphics and interactive techniques*. 262–270.

Sebastian Boring, David Ledo, Xiang'Anthony' Chen, Nicolai Marquardt, Anthony Tang, and Saul Greenberg. 2012. The fat thumb: using the thumb's contact size for single-handed mobile interaction. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*. 39–48.

Rémi Brouet, Renaud Blanch, and Marie-Paule Cani. 2013. Understanding hand degrees of freedom and natural gestures for 3D interaction on tabletop. In *IFIP Conference on Human-Computer Interaction*. Springer, 297–314.

Wolfgang Büschel, Annett Mitschick, Thomas Meyer, and Raimund Dachselt. 2019. Investigating smartphone-based pan and zoom in 3D data spaces in augmented reality. In *Proceedings of the 21st International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–13.

Géry Casiez, Nicolas Roussel, and Daniel Vogel. 2012. 1€ filter: a simple speed-based low-pass filter for noisy input in interactive systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2527–2530.

Ishan Chatterjee, Robert Xiao, and Chris Harrison. 2015. Gaze+ gesture: Expressive, precise and targeted free-space interactions. In *Proceedings of the 2015 ACM on international conference on multimodal interaction*. 131–138.

Myungguen Choi, Daisuke Sakamoto, and Tetsuo Ono. 2020. Bubble gaze cursor+ bubble gaze lens: Applying area cursor technique to eye-gaze interface. In *ACM Symposium on Eye Tracking Research and Applications*. 1–10.

J. de la Fuente and L. Bix. 2010. User-pack interaction: Insights for Designing Inclusive Child-resistant Packaging. In *Designing Inclusive Interactions*, Patrick Martin Langdon, Peter John Clarkson, and Peter Robinson (Eds.). Springer London, London, 89–100.

Bastian Dewitz, Chris Geiger, Frank Steinicke, and Calvin Huhn. 2021. Virtuality between my Fingers–Investigation of Zoom Mechanisms for Visual Exploration of Virtual Environments. In *GI VR/AR Workshop*. Gesellschaft für Informatik eV, 10–18420.

Lisa A Elkin, Matthew Kay, James J Higgins, and Jacob O Wobbrock. 2021. An aligned rank transform procedure for multifactor contrast tests. In *The 34th annual ACM symposium on user interface software and technology*. 754–768.

Augusto Esteves, Elizabeth Bouquet, Ken Pfeuffer, and Florian Alt. 2022. One-handed input for mobile devices via motion matching and orbits controls. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–24.

Dustin Freeman, Ramadevi Vennelakanti, and Sriganesh Madhvanath. 2012. Freehand pose-based gestural interaction: Studies and implications for interface design. In *2012 4th International Conference on Intelligent Human Computer Interaction (IHCI)*. IEEE, 1–6.

Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. 2016. Towards pervasive augmented reality: Context-awareness in augmented reality. *IEEE transactions on visualization and computer graphics* 23, 6 (2016), 1706–1724.

Yves Guiard. 1987. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Journal of motor behavior* 19, 4 (1987), 486–517.

François Guimbretiere and Terry Winograd. 2000. FlowMenu: combining command, text, and data entry. In *Proceedings of the 13th annual ACM symposium on User interface software and technology*. 213–216.

Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

Eiji Hayashi, Manuel Maas, and Jason I Hong. 2014. Wave to me: user identification using body lengths and natural gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3453–3462.

Devamardeep Hayatpur, Seongkook Heo, Haijun Xia, Wolfgang Stuerzlinger, and Daniel Wigdor. 2019. 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*. 1185–1195.

Ken Hinckley, Mary Czerwinski, and Mike Sinclair. 1998. Interaction and modeling techniques for desktop two-handed input. In *Proceedings of the 11th annual ACM symposium on User interface software and technology*. 49–58.

David Holman, Andreas Hollatz, Amartya Banerjee, and Roel Vertegaal. 2013. Unifone: Designing for auxiliary finger input in one-handed mobile interactions. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*. 177–184.

Da-Yuan Huang, Liwei Chan, Shuo Yang, Fan Wang, Rong-Hao Liang, De-Nian Yang, Yi-Ping Hung, and Bing-Yu Chen. 2016. Digitspace: Designing thumb-to-fingers touch interfaces for one-handed and eyes-free interactions. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 1526–1537.

Hyunyoung Jang, Jinwook Kim, and Jeongmi Lee. 2024. Effects of congruent multisensory feedback on the perception and performance of virtual reality hand-retargeted interaction. *IEEE Access* (2024).

Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. 2017. Modeling cumulative arm fatigue in mid-air interaction based on perceived exertion and kinetics of arm motion. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 3328–3339.

Ying Jiang, Congyi Zhang, Hongbo Fu, Alberto Cannavò, Fabrizio Lamberti, Henry YK Lau, and Wenping Wang. 2021. Handpainter-3d sketching in vr with hand-based physical proxy. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–13.

Paul Kabbash and William AS Buxton. 1995. The "prince" technique: Fitts' law and selection using area cursors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 273–279.

Amy K Karlson and Benjamin B Bederson. 2007. ThumbSpace: generalized one-handed input for touchscreen-based mobile devices. In *IFIP Conference on Human-Computer Interaction*. Springer, 324–338.

Amy K Karlson, Benjamin B Bederson, and Jose L Contreras-Vidal. 2008. Understanding one-handed use of mobile devices. In *Handbook of research on user interface design and evaluation for mobile technology*. IGI Global, 86–101.

Dominik P Käser, Maneesh Agrawala, and Mark Pauly. 2011. FingerGlass: efficient multiscale interaction on multitouch screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1601–1610.

Jina Kim, Minyung Kim, Woo Suk Lee, and Sang Ho Yoon. 2023. VibAware: Context-Aware Tap and Swipe Gestures Using Bio-Acoustic Sensing. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction*. 1–12.

Jinwook Kim, Sangmin Park, Qiushi Zhou, Mar Gonzalez-Franco, Jeongmi Lee, and Ken Pfeuffer. 2025a. PinchCatcher: Enabling Multi-selection for Gaze+ Pinch. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–16.

Jina Kim, Yang Zhang, and Sang Ho Yoon. 2025b. T2IRay: Design of Thumb-to-Index based Indirect Pointing for Continuous and Robust AR/VR Input. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–16.

Yongkwan Kim, Sang-Gyun An, Joon Hyub Lee, and Seok-Hyung Bae. 2018. Agile 3D sketching with air scaffolding. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–12.

Yeji Kim, Sohyun Sim, Seoungjae Cho, Woon-woo Lee, Young-Sik Jeong, Kyungeun Cho, and Kyhyun Um. 2014. Intuitive nui for controlling virtual objects based on hand movements. In *Future Information Technology: FutureTech 2014*. Springer, 457–461.

Gordon Kurtenbach and William Buxton. 1994. User learning and performance with marking menus. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 258–264.

Joseph J LaViola Jr, Ernst Kruijff, Ryan P McMahan, Doug Bowman, and Ivan P Poupyrev. 2017. *3D user interfaces: theory and practice*. Addison-Wesley Professional.

Jihyeon Lee, Jinwook Kim, and Jeongmi Lee. 2025. Facilitating the Exploration of Linearly Aligned Objects in Controller-Free 3D Environment with Gaze and Microgestures. *IEEE Transactions on Visualization and Computer Graphics* (2025).

Joon Hyub Lee, Taegyu Jin, Sang-Hyun Lee, Seung-Jun Lee, and Seok-Hyung Bae. 2023. Stereoscopic viewing and monoscopic touching: selecting distant objects in VR through a mobile device. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–7.

Sang-Hyun Lee, Joon Hyub Lee, and Seok-Hyung Bae. 2024. Bimanual Interactions for Surfacing Curve Networks in VR. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.

Wing Ho Andy Li and Hongbo Fu. 2013. BezelCursor: bezel-initiated cursor for one-handed target acquisition on mobile touch screens. In *SIGGRAPH Asia 2013 Symposium on Mobile Graphics and Interactive Applications*. 1–1.

Yang Liu, Thorbjørn Mikkelsen, Zehai Liu, Gengchen Tian, Diako Mardanbegi, Qiushi Zhou, Hans Gellersen, and Ken Pfeuffer. 2025. At a Glance to Your Fingertips: Enabling Direct Manipulation of Distant Objects Through SightWarp. *arXiv preprint arXiv:2508.04821* (2025).

Andreas Löcken, Tobias Hesselmann, Martin Pielot, Niels Henze, and Susanne Boll. 2012. User-centred process for the definition of free-hand gestures applied to controlling music playback. *Multimedia systems* 18, 1 (2012), 15–31.

Mathias N Lystbæk, Peter Rosenberg, Ken Pfeuffer, Jens Emil Grønbæk, and Hans Gellersen. 2022. Gaze-hand alignment: Combining eye gaze and mid-air pointing for interacting with menus in augmented reality. *Proceedings of the ACM on Human-Computer Interaction* 6, ETRA (2022), 1–18.

Sylvain Malacria, Eric Lecolinet, and Yves Guiard. 2010. Clutch-free panning and integrated pan-zoom control on touch-sensitive surfaces: the cyclostar approach. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2615–2624.

Pavel Manakhov, Ludwig Sidenmark, Ken Pfeuffer, and Hans Gellersen. 2024. Gaze on the Go: Effect of Spatial Reference Frame on Visual Target Acquisition During Physical Locomotion in Extended Reality. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 373, 16 pages. doi:10.1145/3613904.3642915

Daniel Mendes, Fabio Marco Caputo, Andrea Giachetti, Alfredo Ferreira, and Joaquim Jorge. 2019. A survey on 3d virtual object manipulation: From the desktop to immersive virtual environments. In *Computer graphics forum*, Vol. 38. Wiley Online Library, 21–45.

Microsoft. 2016. Inclusive Design Toolkit. https://www.microsoft.com/design/inclusive/. Accessed: 2026-01-22.

Microsoft. 2024. App bar and bounding box. https://learn.microsoft.com/en-us/windows/mixed-reality/design/app-bar-and-bounding-box. Accessed: 2026-02-09.

Mathieu Nancel, Julie Wagner, Emmanuel Pietriga, Olivier Chapuis, and Wendy Mackay. 2011. Mid-air pan-and-zoom on wall-sized displays. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 177–186.

Alexander Ng, Stephen A Brewster, and John Williamson. 2013. The impact of encumbrance on mobile interactions. In *IFIP Conference on Human-Computer Interaction*. Springer, 92–109.

Jason Pascoe, Nick Ryan, and David Morse. 2000. Using while moving: HCI issues in fieldwork environments. *ACM Transactions on Computer-Human Interaction (TOCHI)* 7, 3 (2000), 417–437.

Ken Pfeuffer, Jason Alexander, and Hans Gellersen. 2016. Partially-indirect bimanual input with gaze, pen, and touch for pan, zoom, and ink interaction. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 2845–2856.

Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze+ pinch interaction in virtual reality. In *Proceedings of the 5th symposium on spatial user interaction*. 99–108.

Jeffrey S Pierce, Andrew S Forsberg, Matthew J Conway, Seung Hong, Robert C Zeleznik, and Mark R Mine. 1997. Image plane interaction techniques in 3D immersive environments. In *Proceedings of the 1997 symposium on Interactive 3D graphics*. 39–ff.

Julian Rasch, Matthias Wilhalm, Florian Müller, and Francesco Chiossi. 2025. AR You on Track? Investigating Effects of Augmented Reality Anchoring on Dual-Task Performance While Walking. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1217, 21 pages. doi:10.1145/3706598.3714258

Jef Raskin. 2000. *The humane interface: new directions for designing interactive systems*. Addison-Wesley Professional.

Jaime Ruiz, Andrea Bunt, and Edward Lank. 2008. A model of non-preferred hand mode switching. In *Proceedings of Graphics Interface 2008*. 49–56.

Ryan Schubert, Gerd Bruder, and Greg Welch. 2023. Intuitive User Interfaces for Real-Time Magnification in Augmented Reality. In *Proceedings of the 29th ACM Symposium on Virtual Reality Software and Technology*. 1–10.

Marcos Serrano, Barrett M Ens, and Pourang P Irani. 2014. Exploring the use of hand-to-face input for interacting with head-worn displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3181–3190.

Rongkai Shi, Yushi Wei, Xuning Hu, Yu Liu, Yong Yue, Lingyun Yu, and Hai-Ning Liang. 2024. Experimental analysis of freehand multi-object selection techniques in virtual reality head-mounted displays. *Proceedings of the ACM on Human-Computer Interaction* 8, ISS (2024), 93–111.

Rongkai Shi, Yushi Wei, Xueying Qin, Pan Hui, and Hai-Ning Liang. 2023. Exploring gaze-assisted and hand-based region selection in augmented reality. *Proceedings of the ACM on Human-Computer Interaction* 7, ETRA (2023), 1–19.

Rongkai Shi, Nan Zhu, Hai-Ning Liang, and Shengdong Zhao. 2021. Exploring head-based mode-switching in virtual reality. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 118–127.

Jesse Smith, Isaac Wang, Julia Woodward, and Jaime Ruiz. 2019. Experimental Analysis of Single Mode Switching Techniques in Augmented Reality.. In *Graphics Interface*. 20–1.

Peng Song, Wooi Boon Goh, William Hutama, Chi-Wing Fu, and Xiaopei Liu. 2012. A handle bar metaphor for virtual object manipulation with mid-air interaction. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1297–1306.

Sophie Stellmach, Markus Jüttner, Christian Nywelt, Jens Schneider, and Raimund Dachselt. 2012. Investigating Freehand Pan and Zoom. In *Mensch & Computer 2012: interaktiv informiert – allgegenwärtig und allumfassend!?* Oldenbourg Verlag, München, 303–312.

Richard Stoakley, Matthew J Conway, and Randy Pausch. 1995. Virtual reality on a WIM: interactive worlds in miniature. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 265–272.

Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2017. Experimental analysis of mode switching techniques in touch-based user interfaces. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 3267–3280.

Hemant Bhaskar Surale, Fabrice Matulic, and Daniel Vogel. 2019. Experimental analysis of barehand mid-air mode-switching techniques in virtual reality. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–14.

Justin H Tan, Cherng Chao, Mazen Zawaideh, Anne C Roberts, and Thomas B Kinney. 2013. Informatics in radiology: Developing a touchless user interface for intraoperative image control during interventional radiology procedures. *Radiographics* 33, 2 (2013), E61–E70.

Andries Van Dam. 1997. *1997 Symposium on Interactive 3D Graphics*. Association for Computing Machinery (ACM).

Radu-Daniel Vatavu. 2013. A comparative study of user-defined handheld vs. freehand gestures for home entertainment environments. *Journal of Ambient Intelligence and Smart Environments* 5, 2 (2013), 187–211.

Eduardo Velloso, Dominik Schmidt, Jason Alexander, Hans Gellersen, and Andreas Bulling. 2015. The feet in human–computer interaction: A survey of foot-based interaction. *ACM Computing Surveys (CSUR)* 48, 2 (2015), 1–35.

Daniel Vogel and Ravin Balakrishnan. 2005. Distant freehand pointing and clicking on very large, high resolution displays. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*. 33–42.

Uta Wagner, Andreas Asferg Jacobsen, Tiare Feuchtner, Hans Gellersen, and Ken Pfeuffer. 2024. Eye-Hand Movement of Objects in Near Space Extended Reality. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–13.

Uta Wagner, Mathias N Lystbæk, Pavel Manakhov, Jens Emil Sloth Grønbæk, Ken Pfeuffer, and Hans Gellersen. 2023. A fitts' law study of gaze-hand alignment for selection in 3d user interfaces. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

Robert Walter, Gilles Bailly, and Jörg Müller. 2013. StrikeAPose: revealing mid-air gestures on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 841–850.

Minqi Wang and Emily A Cooper. 2022. Perceptual guidelines for optimizing field of view in stereoscopic augmented reality displays. *ACM Transactions on Applied Perception* 19, 4 (2022), 1–23.

Kieran Waugh, Mark McGill, and Euan Freeman. 2022. Push or pinch? Exploring slider control gestures for touchless user interfaces. In *Nordic Human-Computer Interaction Conference*. 1–10.

Jacob O Wobbrock. 2019. Situationally aware mobile devices for overcoming situational impairments. In *Proceedings of the ACM SIGCHI symposium on engineering interactive computing systems*. 1–18.

Jacob O Wobbrock, Leah Findlater, Darren Gergle, and James J Higgins. 2011. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 143–146.

Momona Yamagami, Sasa Junuzovic, Mar Gonzalez-Franco, Eyal Ofek, Edward Cutrell, John R Porter, Andrew D Wilson, and Martez E Mott. 2022. Two-in-one: A design space for mapping unimanual input into bimanual interactions in vr for users with limited movement. *ACM Transactions on Accessible Computing (TACCESS)* 15, 3 (2022), 1–25.

ByungIn Yoo, Jae-Joon Han, Changkyu Choi, Kwonju Yi, Sungjoo Suh, Dusik Park, and Changyeong Kim. 2010. 3D user interface combining gaze and hand gestures for large-scale display. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems*. 3709–3714.

Difeng Yu, Xueshi Lu, Rongkai Shi, Hai-Ning Liang, Tilman Dingler, Eduardo Velloso, and Jorge Goncalves. 2021. Gaze-supported 3d object manipulation in virtual reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–13.

Fengyuan Zhu, Ludwig Sidenmark, Mauricio Sousa, and Tovi Grossman. 2023. 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)*. IEEE, 1221–1230.